数据仓库服务
9.1.2
SQL 语法参考

文档版本	01
发布日期	2026-11-24



华为云计算技术有限公司

版权所有 © 华为云计算技术有限公司 2026。 保留一切权利。

非经本公司书面许可，任何单位和个人不得擅自摘抄、复制本文档内容的部分或全部，并不得以任何形式传播。

商标声明


和其他华为商标均为华为技术有限公司的商标。
本文档提及的其他所有商标或注册商标，由各自的所有人拥有。

注意

您购买的产品、服务或特性等应受华为云计算技术有限公司商业合同和条款的约束，本文档中描述的全部或部分产品、服务或特性可能不在您的购买或使用范围之内。除非合同另有约定，华为云计算技术有限公司对本文档内容不做任何明示或暗示的声明或保证。

由于产品版本升级或其他原因，本文档内容会不定期进行更新。除非另有约定，本文档仅作为使用指导，本文档中的所有陈述、信息和建议不构成任何明示或暗示的担保。

目 录

1 DWS SQL 概述 1
2 DWS SQL 语法格式说明 3
3 与 PostgreSQL 的差异 4
4 关键字 8
5 数据类型 41
5.1 数值类型 41
5.2 货币类型 45
5.3 布尔类型 46
5.4 字符类型 47
5.5 二进制类型 53
5.6 日期/时间类型 55
5.7 几何类型 61
5.8 数组类型 63
5.9 枚举类型 67
5.10 网络地址类型 68
5.11 位串类型 70
5.12 文本搜索类型 71
5.13 UUID 类型 75
5.14 JSON 类型 76
5.15 RoaringBitmap 类型 79
5.16 HLL 数据类型 79
5.17 对象标识符类型 81
5.18 伪类型 83
5.19 范围类型 84
5.20 复合类型 88
5.21 XML 类型 90
5.22 列存表支持的数据类型 91
6 常量与宏 93
7 函数和操作符 94
7.1 字符串处理函数和操作符 94
7.1.1 字符串处理函数和操作符总体介绍 94

 版权所有 © 华为云计算技术有限公司

7.1.2 获取字符串长度 95
7.1.3 截取字符串 99
7.1.4 拼接字符串 110
7.1.5 替换和填充字符串 112
7.1.6 删除指定字符 117
7.1.7 分割字符串 118
7.1.8 转换字符串 120
7.1.9 比较字符串 128
7.1.10 获取某一个字符或子字符串的位置 128
7.1.11 判断某个字符串是否符合某种匹配模式(字符串相关的模式匹配函数) 130
7.2 二进制字符串函数和操作符 132
7.3 位串函数和操作符 136
7.4 数学函数和运算符 138
7.4.1 数学运算符 138
7.4.2 数学函数 141
7.5 时间日期函数和操作符 150
7.5.1 时间和日期操作符 150
7.5.2 时间和日期函数 152
7.5.2.1 时间和日期函数概述 152
7.5.2.2 时间和日期基本函数 156
7.5.2.3 时间和日期运算函数 164
7.5.2.4 时间和日期截取函数 171
7.5.2.5 获取当前时间和日期函数 177
7.5.2.6 时间和日期格式化函数 180
7.5.2.7 其他时间和日期处理函数 183
7.6 SEQUENCE 函数 185
7.7 数组函数和操作符 188
7.7.1 数组操作符 188
7.7.2 数组函数 191
7.8 逻辑操作符 198
7.9 比较操作符 199
7.10 模式匹配操作符 200
7.11 聚集函数 204
7.12 窗口函数 221
7.13 类型转换函数 226
7.14 JSON/JSONB 函数和操作符 237
7.14.1 JSON/JSONB 操作符 237
7.14.2 JSON/JSONB 函数 240
7.15 安全函数 260
7.16 条件表达式函数 269
7.17 范围函数和操作符 272
7.17.1 范围操作符 273

 版权所有 © 华为云计算技术有限公司

7.17.2 范围函数 275
7.18 数据脱敏函数 277
7.19 位图（Roaring Bitmap）函数和操作符 279
7.19.1 Roaring Bitmap 操作符 279
7.19.2 Roaring Bitmap 函数 282
7.19.3 Roaring Bitmap 聚合函数 289
7.19.4 使用场景 291
7.20 UUID 函数 293
7.21 文本检索函数和操作符 295
7.21.1 文本检索操作符 295
7.21.2 文本检索函数 297
7.21.3 文本检索调试函数 300
7.22 HLL 函数和操作符 302
7.22.1 HLL 操作符 302
7.22.2 哈希函数 304
7.22.3 精度函数 307
7.22.4 聚合函数 309
7.22.5 功能函数 311
7.22.6 内置函数 313
7.23 返回集合的函数 314
7.23.1 序列号生成函数 314
7.23.2 下标生成函数 316
7.24 几何函数和操作符 317
7.24.1 几何操作符 317
7.24.2 几何函数 323
7.24.3 几何类型转换函数 326
7.25 网络地址函数和操作符 330
7.25.1 cidr 和 inet 操作符 330
7.25.2 网络地址函数 333
7.26 系统信息函数 336
7.26.1 会话信息函数 336
7.26.2 访问权限查询函数 342
7.26.3 模式可见性查询函数 345
7.26.4 系统表信息函数 347
7.26.5 系统函数信息函数 353
7.26.6 注释信息函数 354
7.26.7 事务 ID 和快照 355
7.26.8 计算子集群函数 357
7.26.9 锁信息函数 358
7.26.10 缓存信息函数 358
7.27 系统管理函数 366
7.27.1 配置设置函数 366

 版权所有 © 华为云计算技术有限公司

7.27.2 通用文件访问函数.....367
7.27.3 服务器信号函数.....369
7.27.4 快照同步函数.....373
7.27.5 咨询锁函数.....374
7.27.6 复制函数.....377
7.27.7 资源管理函数.....385
7.27.8 内存管理函数.....398
7.27.9 缓存管理函数.....403
7.27.10 plan management 管理函数.....404
7.27.11 其他函数.....409
7.28 备份恢复控制函数.....439
7.28.1 备份控制函数.....439
7.28.2 恢复控制函数.....444
7.29 数据库对象函数.....448
7.29.1 数据库对象尺寸函数.....448
7.29.2 数据库对象位置函数.....452
7.29.3 分区管理函数.....452
7.29.4 排序规则版本函数.....453
7.29.5 冷热表用户函数.....454
7.30 残留文件管理函数.....456
7.30.1 获取残留文件列表函数.....456
7.30.2 扫描残留文件函数.....458
7.30.3 验证残留文件函数.....464
7.30.4 删除残留文件函数.....467
7.30.5 残留文件管理函数应用.....469
7.31 统计信息函数.....472
7.32 触发器函数.....511
7.33 XML 函数.....511
7.33.1 产生 XML 内容.....511
7.33.2 XML 谓词.....515
7.33.3 处理 XML.....517
7.33.4 将表映射到 XML.....519
7.34 调用栈记录函数.....521
7.35 Hudi 系统函数.....523
7.36 漏斗函数和留存函数.....530
7.37 EXTERNAL SCHEMA 系统函数.....535
7.38 存算分离函数.....536
7.38.1 数据共享函数.....536
7.38.2 异步 IO 限流.....537
7.38.3 LStore 存储引擎函数.....545
7.39 Iceberg 表服务函数.....556
7.40 时序函数.....562

 版权所有 © 华为云计算技术有限公司

7.41 实时数仓函数 569
8 表达式 578
8.1 简单表达式 578
8.2 条件表达式 579
8.3 子查询表达式 584
8.4 数组表达式 586
8.5 行表达式 588
9 类型转换 589
9.1 概述 589
9.2 操作符 590
9.3 函数 592
9.4 值存储 594
9.5 UNION，CASE 和相关构造 595
10 全文检索 599
10.1 介绍 599
10.1.1 全文检索概述 599
10.1.2 文档概念 600
10.1.3 基本文本匹配 600
10.1.4 分词器 601
10.2 表和索引 602
10.2.1 搜索表 602
10.2.2 创建索引 603
10.2.3 索引使用约束 604
10.3 控制文本搜索 605
10.3.1 解析文档 605
10.3.2 解析查询 606
10.3.3 排序查询结果 607
10.3.4 高亮搜索结果 608
10.4 附加功能 610
10.4.1 处理 tsvector 610
10.4.2 处理查询 610
10.4.3 查询重写 611
10.4.4 收集文献统计 612
10.5 解析器 613
10.6 词典 617
10.6.1 词典概述 617
10.6.2 停用词 618
10.6.3 Simple 词典 618
10.6.4 Synonym 词典 619
10.6.5 Thesaurus 词典 621
10.6.6 Ispell 词典 622

 版权所有 © 华为云计算技术有限公司

10.6.7 Snowball 词典 623
10.7 配置示例 623
10.8 测试和调试文本搜索 625
10.8.1 分词器测试 625
10.8.2 解析器测试 626
10.8.3 词典测试 626
10.9 限制约束 627
11 系统操作 628
12 事务管理 629
13 DDL 语法 635
13.1 DDL 语法一览表 635
13.2 数据库&模式&数据类型 641
13.2.1 CREATE DATABASE 642
13.2.2 ALTER DATABASE 646
13.2.3 DROP DATABASE 648
13.2.4 CREATE SCHEMA 650
13.2.5 ALTER SCHEMA 653
13.2.6 DROP SCHEMA 656
13.2.7 ALTER EXTERNAL SCHEMA 657
13.2.8 CREATE EXTERNAL SCHEMA 658
13.2.9 CREATE TYPE 662
13.2.10 ALTER TYPE 672
13.2.11 DROP TYPE 675
13.3 用户和角色 676
13.3.1 CREATE USER 676
13.3.2 ALTER USER 682
13.3.3 DROP USER 688
13.3.4 CREATE ROLE 689
13.3.5 ALTER ROLE 697
13.3.6 SET ROLE 703
13.3.7 DROP ROLE 704
13.3.8 CREATE GROUP 705
13.3.9 ALTER GROUP 706
13.3.10 DROP GROUP 707
13.3.11 DROP OWNED 708
13.4 表和外部数据源 709
13.4.1 CREATE TABLE 710
13.4.2 CREATE TABLE PARTITION 749
13.4.3 CREATE TABLE AS 779
13.4.4 ALTER TABLE 785
13.4.5 ALTER TABLE PARTITION 804

 版权所有 © 华为云计算技术有限公司

13.4.6 RENAME TABLE 818
13.4.7 DROP TABLE 819
13.4.8 CREATE FOREIGN TABLE (SQL on OBS or Hadoop) 820
13.4.9 CREATE FOREIGN TABLE (OBS 导入导出) 850
13.4.10 CREATE FOREIGN TABLE (GDS 导入导出) 864
13.4.11 ALTER FOREIGN TABLE (For HDFS or OBS) 883
13.4.12 ALTER FOREIGN TABLE (GDS 导入导出) 885
13.4.13 DROP FOREIGN TABLE 887
13.4.14 CREATE SERVER 888
13.4.15 ALTER SERVER 892
13.4.16 DROP SERVER 895
13.5 视图和物化视图 896
13.5.1 CREATE VIEW 896
13.5.2 ALTER VIEW 899
13.5.3 DROP VIEW 901
13.5.4 CREATE MATERIALIZED VIEW 903
13.5.5 ALTER MATERIALIZED VIEW 912
13.5.6 REFRESH MATERIALIZED VIEW 915
13.5.7 DROP MATERIALIZED VIEW 917
13.6 索引&序列&文件目录 918
13.6.1 CREATE INDEX 918
13.6.2 ALTER INDEX 930
13.6.3 REINDEX 935
13.6.4 DROP INDEX 937
13.6.5 CREATE SEQUENCE 938
13.6.6 ALTER SEQUENCE 942
13.6.7 DROP SEQUENCE 944
13.6.8 CREATE TEXT SEARCH CONFIGURATION 945
13.6.9 ALTER TEXT SEARCH CONFIGURATION 948
13.6.10 DROP TEXT SEARCH CONFIGURATION 951
13.6.11 CREATE TEXT SEARCH DICTIONARY 952
13.6.12 ALTER TEXT SEARCH DICTIONARY 956
13.6.13 DROP TEXT SEARCH DICTIONARY 958
13.6.14 CLUSTER 959
13.7 存储过程&函数&触发器&游标 961
13.7.1 CREATE FUNCTION 961
13.7.2 ALTER FUNCTION 977
13.7.3 DROP FUNCTION 980
13.7.4 CREATE PROCEDURE 981
13.7.5 DROP PROCEDURE 987
13.7.6 CREATE TRIGGER 988
13.7.7 ALTER TRIGGER 993

 版权所有 © 华为云计算技术有限公司

13.7.8 DROP TRIGGER....994
13.7.9 FETCH....995
13.7.10 MOVE....998
13.7.11 CURSOR....999
13.7.12 CLOSE....1001
13.8 数据脱敏和安全相关....1002
13.8.1 CREATE ROW LEVEL SECURITY POLICY....1002
13.8.2 ALTER ROW LEVEL SECURITY POLICY....1009
13.8.3 DROP ROW LEVEL SECURITY POLICY....1010
13.8.4 CREATE REDACTION POLICY....1012
13.8.5 ALTER REDACTION POLICY....1016
13.8.6 DROP REDACTION POLICY....1019
13.8.7 ALTER CLUSTER....1020
13.9 资源池和过滤规则....1020
13.9.1 CREATE RESOURCE POOL....1020
13.9.2 ALTER RESOURCE POOL....1026
13.9.3 DROP RESOURCE POOL....1028
13.9.4 CREATE EXCEPT RULE....1028
13.9.5 ALTER EXCEPT RULE....1031
13.9.6 DROP EXCEPT RULE....1033
13.9.7 CREATE BLOCK RULE....1034
13.9.8 ALTER BLOCK RULE....1042
13.9.9 DROP BLOCK RULE....1046
13.10 性能和执行计划....1047
13.10.1 CREATE STATISTICS....1047
13.10.2 DROP STATISTICS....1048
13.10.3 CREATE OUTLINE....1049
13.10.4 DROP OUTLINE....1050
13.11 连接&会话&资源清理....1051
13.11.1 CLEAN CONNECTION....1051
13.11.2 ALTER SYSTEM KILL SESSION....1052
13.11.3 ALTER SESSION....1053
13.11.4 DISCARD....1056
13.11.5 TRUNCATE....1058
13.11.6 VACUUM....1061
13.11.7 PURGE....1069
13.11.8 TIMECAPSULE....1070
13.12 设置与注释....1072
13.12.1 SHOW....1072
13.12.2 SET....1072
13.12.3 RESET....1075
13.12.4 SET CONSTRAINTS....1076

 版权所有 © 华为云计算技术有限公司

13.12.5 SET SESSION AUTHORIZATION 1077
13.12.6 COMMENT 1079
13.13 发布和订阅 1081
13.13.1 CREATE PUBLICATION 1081
13.13.2 ALTER PUBLICATION 1083
13.13.3 DROP PUBLICATION 1085
13.13.4 CREATE SUBSCRIPTION 1086
13.13.5 ALTER SUBSCRIPTION 1088
13.13.6 DROP SUBSCRIPTION 1090
13.14 其他 1091
13.14.1 CREATE SYNONYM 1091
13.14.2 ALTER SYNONYM 1093
13.14.3 DROP SYNONYM 1094
13.14.4 ALTER LARGE OBJECT 1095
13.14.5 CREATE BARRIER 1096
14 DML 语法 1097
14.1 DML 语法一览表 1097
14.2 CALL 1098
14.3 COPY 1100
14.4 DELETE 1118
14.5 EXPLAIN 1122
14.6 EXPLAIN PLAN 1129
14.7 LOCK 1131
14.8 MERGE INTO 1135
14.9 INSERT 和 UPSERT 1139
14.9.1 INSERT 1139
14.9.2 UPSERT 1149
14.10 UPDATE 1154
14.11 VALUES 1162
15 DCL 语法 1164
15.1 DCL 语法一览表 1164
15.2 ALTER DEFAULT PRIVILEGES 1164
15.3 ANALYZE | ANALYSE 1167
15.4 DEALLOCATE 1172
15.5 DO 1173
15.6 EXECUTE 1174
15.7 EXECUTE DIRECT 1175
15.8 GRANT 1176
15.9 PREPARE 1185
15.10 REASSIGN OWNED 1186
15.11 REVOKE 1187

 版权所有 © 华为云计算技术有限公司

16 DQL 语法.....1191
16.1 DQL 语法一览表.....1191
16.2 SELECT.....1191
16.3 SELECT INTO.....1219
17 TCL 语法.....1221
17.1 TCL 语法一览表.....1221
17.2 ABORT.....1221
17.3 BEGIN.....1223
17.4 CHECKPOINT.....1224
17.5 COMMIT | END.....1225
17.6 COMMIT PREPARED.....1225
17.7 PREPARE TRANSACTION.....1227
17.8 SAVEPOINT.....1230
17.9 SET TRANSACTION.....1232
17.10 START TRANSACTION.....1233
17.11 ROLLBACK.....1235
17.12 RELEASE SAVEPOINT.....1235
17.13 ROLLBACK PREPARED.....1236
17.14 ROLLBACK TO SAVEPOINT.....1238

 版权所有 © 华为云计算技术有限公司

1 DWS SQL 概述

什么是 SQL

SQL是用于访问和处理数据库的标准计算机语言。

SQL提供了各种任务的语句，包括：

- 查询数据。
- 在表中插入、更新和删除行。
- 创建、替换、更改和删除对象。
- 控制对数据库及其对象的访问。
- 保证数据库的一致性和完整性。

SQL语言由用于处理数据库和数据库对象的命令和函数组成。该语言还会强制实施有关数据类型、表达式和文本使用的规则。因此在SQL参考章节，除了SQL语法参考外，还会看到有关数据类型、表达式、函数和操作符等信息。

SQL 发展简史

SQL发展简史如下：

- 1986年，ANSI X3.135-1986，ISO/IEC 9075:1986，SQL-86
- 1989年，ANSI X3.135-1989，ISO/IEC 9075:1989，SQL-89
- 1992年，ANSI X3.135-1992，ISO/IEC 9075:1992，SQL-92（SQL2）
- 1999年，ISO/IEC 9075:1999，SQL:1999（SQL3）
- 2003年，ISO/IEC 9075:2003，SQL:2003（SQL4）
- 2011年，ISO/IEC 9075:200N，SQL:2011（SQL5）

DWS 支持的 SQL 标准

DWS兼容Postgres-XC，默认支持SQL2、SQL3和SQL4的主要特性、SQL5的部分特性。

DWS 支持的语言扩展

DWS支持的语言扩展包括PL/pgSQL和PL/Java。

 版权所有 © 华为云计算技术有限公司

SQL 示例说明

手册中的部分SQL示例是基于TPC-DS模型开发的，如果需要运行手册中的示例，请先参考官网说明（**https://www.tpc.org/tpcds/**），安装TPC-DS benchmark。

版权所有 © 华为云计算技术有限公司

2 DWS SQL 语法格式说明

为了方便对语法使用的理解，在文档中对SQL语法文本按如下格式进行表述。

表 2-1 SQL 语法格式说明

格式	意义
大写	语法关键字（语句中保持不变、必须与语法格式一致的部分）采用大写表示。
小写	参数（语句中必须由实际值进行替代的部分）采用小写表示。
[ ]	可选语法项。表示用“[ ]”括起来的部分是可选的。
{ }	必选语法项。
...	表示前面的元素可重复出现。
[x | y | ...]	表示从两个或多个选项中选取一个或者不选。
{ x | y | ... }	表示从两个或多个选项中选取一个。
[x | y | ...] [...]	表示可选多个参数或者不选，如果选择多个参数，则参数之间用空格分隔。
[x | y | ...] [,...]	表示可选多个参数或者不选，如果选择多个参数，则参数之间用逗号分隔。
{ x | y | ... } [...]	表示可选多个参数，至少选一个，如果选择多个参数，则参数之间以空格分隔。
{ x | y | ... } [,...]	表示可选多个参数，至少选一个，如果选择多个参数，则参数之间用逗号分隔。

 版权所有 © 华为云计算技术有限公司

3 与 PostgreSQL 的差异

DWS与PostgreSQL的差异基于PostgreSQL 9.X版本整理，具体差异如下：

客户端差异

DWS gsql相较于PostgreSQL psql做了如下安全加固变更：

- 取消通过元命令\password设置用户密码。
- 新增元命令\i+、 \ir+、\include_relative+和输入输出参数-k，以支持给导入导出的文件加密。
- 取消打印命令行历史到文件的元命令\s。
- 涉及敏感操作SQL历史不再记录，如含有密码操作。即用户通过翻页/上下键查阅SQL历史将不能查到对应的记录。
- 支持连接后在屏幕上给出用户密码过期提示以及版本信息。

gsql在psql基础上还增加了如下功能：

- 新增输出格式参数-r。支持用户输入命令时的tab补齐和方向键调整焦点。
- 新增并行操作元命令\parallel，以提升执行性能。
- 新增\set RETRY支持语句出错重试。
- 新增PLSQL默认结束符功能，将“/”作为PLSQL语句（ create or replace function/procedure ）的默认结束符，增加便利性。

libpq：

DWS在开发某些功能，如客户端连接工具gsql时，对PostgreSQL libpq进行了较大修改，但并未对此接口在应用程序开发场景下的使用做验证。因此对使用此接口进行应用程序开发存在的风险未知，推荐用户使用ODBC或JDBC接口来替代。

SQL 差异

表 3-1 DWS 不支持的 PostgreSQL 数据库语言

分类	DWS不支持	说明
数据类型	几何类型line	DWS所支持的数据类型参见**数据类型**。

版权所有 © 华为云计算技术有限公司

分类	DWS不支持	说明
	pg_node_tree	
函数	枚举支持函数： • enum_first(anyenum) • enum_last(anyenum) • enum_range(anyenum) • enum_range(anyenum, anyenum)	DWS所支持的函数参见**函数和操作符**。
	访问权限查询函数： • has_sequence_privilege(user, sequence, privilege) • has_sequence_privilege(sequence, privilege)	
	系统目录信息函数： • pg_get_triggerdef(trigger_oid) • pg_get_triggerdef(trigger_oid, pretty_bool)	
	几何类型转换函数： line(point, point)	
	pg_node_tree函数	
SQL语法	CREATE TABLE子句： INHERITS (parent_table [, ...])	继承表。
	CREATE TABLE的列约束： REFERENCES reftable [(refcolumn)] [MATCH FULL | MATCH PARTIAL | MATCH SIMPLE] [ON DELETE action] [ON UPDATE action]	列约束中用REFERENCES reftable [(refcolumn)] [MATCH FULL | MATCH PARTIAL | MATCH SIMPLE] [ON DELETE action] [ON UPDATE action] 为表创建外键约束。
	CREATE TABLE的表约束： EXCLUDE [USING index_method] (exclude_element WITH operator [, ...])	表约束中用EXCLUDE [USING index_method] (exclude_element WITH operator [, ...])为表创建排除约束。
	CREATE/ALTER/DROP EXTENSION	扩展的加载、修改和删除。
	CREATE/ALTER/DROP AGGREGATE	聚集函数的定义、修改和删除。
	CREATE/ALTER/DROP OPERATOR	操作符（OPERATOR）的创建、修改和删除。

版权所有 © 华为云计算技术有限公司

分类	DWS不支持	说明
	CREATE/ALTER/DROP OPERATOR CLASS	操作符类（OPERATOR CLASS）的创建、修改和删除。
	CREATE/ALTER/DROP OPERATOR FAMILY	操作符族（OPERATOR FAMILY）的创建、修改和删除。
	CREATE/ALTER/DROP TEXT SEARCH PARSER	文本检索解析器（TEXT SEARCH PARSER）的创建、修改和删除。
	CREATE/ALTER/DROP TEXT SEARCH TEMPLATE	文本检索模板（TEXT SEARCH TEMPLATE）的创建、修改和删除。
	CREATE/ALTER/DROP COLLATION	排序规则（COLLATION）的创建、修改和删除。
	CREATE/ALTER/DROP CONVERSION	字符集编码转换（CONVERSION）的定义、修改和删除。
	CREATE/ALTER/DROP RULE	规则（RULE）的创建、修改和删除。
	CREATE/ALTER/DROP LANGUAGE	过程语言（LANGUAGE）的注册、修改和删除。
	CREATE/ALTER/DROP DOMAIN	域（DOMAIN）的创建、修改和删除。
	CREATE/DROP CAST	类型转换（CAST）的定义和删除。
	CREATE/ALTER/DROP USER MAPPING	用户映射（USER MAPPING）的定义、修改和删除。
	SECURITY LABEL	定义或更改对象的安全标签。
	NOTIFY	生成一个通知。
	LISTEN	监听一个通知。
	UNLISTEN	停止监听通知信息。
	LOAD	加载或重新加载一个共享库文件。
	MOVE BACKWARD	反向移动游标。
	COPY的COPY FROM FILE和COPY TO FILE	为了权限的隔离，DWS禁用COPY FROM FILE和COPY TO FILE。

 版权所有 © 华为云计算技术有限公司

分类	DWS不支持	说明
其他	自定义C函数	DWS支持的用户自定义函数可参见**用户自定义函数**章节。

 版权所有 © 华为云计算技术有限公司

4 关键字

在进行数据库开发时，用户需要为表、列、索引、函数等数据库对象命名。然而，SQL语言本身定义了一批具有特殊含义的关键字（如SELECT、CREATE、FROM等），如果将这些关键字用作对象名称，SQL解析器会将其识别为关键字而非标识符，导致语法错误。

DWS提供了系统函数pg_get_keywords()查询全部关键字及其分类，并支持保留关键字与非保留关键字的区分。用户可在建表前查阅关键字列表，避免使用保留关键字作为对象名称；如确需使用，可通过双引号括起标识符来绕过冲突。本文档列出了DWS支持的全部关键字及其在DWS、SQL:1999、SQL:1992标准中的分类，供开发参考。

基本概念

关键字

关键字（Keyword）是SQL语法中预定义的、具有特殊含义的保留词。例如SELECT表示查询操作、CREATE表示创建对象、WHERE表示条件过滤。SQL标准将关键字分为两类：

- **保留关键字（Reserved Keyword）**：不能用作标识符（表名、列名、函数名等）。如果直接使用保留关键字作为对象名称，SQL将报语法错误。例如，SELECT是保留关键字，不能用SELECT作为表名。
- **非保留关键字（Non-reserved Keyword）**：仅在特定语法环境中有特殊含义，在其他位置可用作标识符。例如，ABORT是非保留关键字，可用作列名。但为避免代码歧义，建议尽量避免将非保留关键字用作标识符。

此外，DWS中部分非保留关键字标注为”非保留(不能是函数或类型)”，表示该关键字虽可用作普通标识符（如表名、列名），但不能用作函数名或数据类型名。

标识符

标识符（Identifier）是用户为数据库对象定义的名称，用于在SQL语句中引用这些对象。例如CREATE TABLE orders (id int);中，orders是表名标识符，id是列名标识符。标识符需遵守命名规范（见下文”约束与限制”），且不能与保留关键字冲突。

查询 DWS 关键字

可以使用**SELECT * FROM pg_get_keywords();**查询DWS的关键字。

版权所有 © 华为云计算技术有限公司

步骤1 连接DWS数据库集群。

步骤2 执行以下SQL查询全部关键字及其分类。

```
SELECT word, catcode, catdesc
FROM pg_get_keywords()
ORDER BY word;
```

步骤3 如果只需查询保留关键字（即不能用作标识符的关键字），可添加过滤条件。

```
SELECT word
FROM pg_get_keywords()
WHERE catcode = 'R'
ORDER BY word;
```

步骤4 如果需要确认某个特定词是否为保留关键字，可精确查询。

```
SELECT word, catcode, catdesc
FROM pg_get_keywords()
WHERE word = 'authorization';
```

----结束

验证方法： 查询应返回关键字名称（word）、分类代码（catcode）和分类描述（catdesc）三列。catcode列的取值含义如下：

表 4-1 catcode 列

catcode值	含义	对应下表标注
R	保留（reserved）	保留
N	非保留（non-reserved）	非保留
T	非保留（不能是函数或类型）	非保留(不能是函数或类型)

举例：“AUTHORIZATION”为保留关键字，当使用“AUTHORIZATION”作为表名执行建表语句时系统会报错，不允许创建名称为“AUTHORIZATION”的表。

但是有例外情况，“AUTHORIZATION”可以作为函数或类型名称使用（下表有说明）。

图 4-1 使用关键字作为表名



 版权所有 © 华为云计算技术有限公司

图 4-2 使用关键字作为函数名



标识符命名规范

数据库标识符是用来唯一标识数据库中对象（如表、列、索引等）的名称。在数据库中，标识符在SQL查询语句中经常被用来引用数据库对象，确保数据库操作的准确性和一致性。合理命名和使用标识符非常重要，DWS标识符的命名需遵守如下规范：

- 标识符需要由字母（a-z）、下划线（_）、数字（0-9）或$组成。
- 标识符建议以字母（a-z）或下划线（_）开头。
- 标识符长度不超过63个字节。
- 标识符不能是保留的SQL关键字，非保留关键字请慎用。
- 数据库字符集为GBK、UTF8、SQL_ASCII时标识符支持中文汉字，其中UTF8、SQL_ASCII字符集下，一个中文汉字占3个字符，最多支持21个中文汉字；GBK字符集下，一个中文汉字占2个字符，最多支持31个中文汉字。Latin1字符集不支持中文汉字命名。字符集格式在创建数据库时指定，详情参见**CREATE DATABASE**。

表 4-2 SQL 关键字

关键字	DWS	SQL:1999	SQL:1992
ABORT	非保留	-	-
ABS	-	非保留	-
ABSOLUTE	非保留	保留	保留
ACCESS	非保留	-	-
ACCOUNT	非保留	-	-
ACTION	非保留	保留	保留
ADA	-	非保留	非保留
ADD	非保留	保留	保留
ADMIN	非保留	保留	-
AFTER	非保留	保留	-

版权所有 © 华为云计算技术有限公司

关键字	DWS	SQL:1999	SQL:1992
AGGREGATE	非保留	保留	-
ALIAS	-	保留	-
ALL	保留	保留	保留
ALLOCATE	-	保留	保留
ALLOW_CONNECTIONS	非保留 9.1.1.200集群版本新增。	-	-
ALSO	非保留	-	-
ALTER	非保留	保留	保留
ALWAYS	非保留	-	-
ANALYSE	保留	-	-
ANALYZE	保留	-	-
AND	保留	保留	保留
ANY	保留	保留	保留
APP	非保留	-	-
ARE	-	保留	保留
ARRAY	保留	保留	-
AS	保留	保留	保留
ASC	保留	保留	保留
ASENSITIVE	-	非保留	-
ASSERTION	非保留	保留	保留
ASSIGNMENT	非保留	非保留	-
ASYMMETRIC	保留	非保留	-
AT	非保留	保留	保留
ATOMIC	-	非保留	-
ATTRIBUTE	非保留	-	-
AUTHID	保留	-	-
AUTHINFO	非保留	-	-
AUTHORIZATION	保留(可以是函数或类型)	保留	保留
AUTO	非保留	-	-

版权所有 © 华为云计算技术有限公司

关键字	DWS	SQL:1999	SQL:1992
AUTOEXTEND	非保留	-	-
AUTOMAPPED	非保留	-	-
AUTOREFRESH	非保留 9.1.0.210集群版本新增。	-	-
AUTOSCALE	非保留	-	-
AVG	-	非保留	保留
BACKWARD	非保留	-	-
BARRIER	非保留	-	-
BEFORE	非保留	保留	-
BEGIN	非保留	保留	保留
BETWEEN	非保留(不能是函数或类型)	非保留	保留
BIGINT	非保留(不能是函数或类型)	-	-
BINARY	保留(可以是函数或类型)	保留	-
BINARY_DOUBLE	非保留(不能是函数或类型)	-	-
BINARY_INTEGER	非保留(不能是函数或类型)	-	-
BIT	非保留(不能是函数或类型)	保留	保留
BITVAR	-	非保留	-
BIT_LENGTH	-	非保留	保留
BLOB	非保留	保留	-
BLOCK	非保留	-	-
BOOLEAN	非保留(不能是函数或类型)	保留	-
BOTH	保留	保留	保留
BUCKETS	保留	-	-
BUILD	非保留	-	-
BREADTH	-	保留	-

 版权所有 © 华为云计算技术有限公司

关键字	DWS	SQL:1999	SQL:1992
BY	非保留	保留	保留
C	-	非保留	非保留
CACHE	非保留	-	-
CALL	非保留	保留	-
CALLED	非保留	非保留	-
CARDINALITY	-	非保留	-
CASCADE	非保留	保留	保留
CASCADED	非保留	保留	保留
CASE	保留	保留	保留
CAST	保留	保留	保留
CATALOG	非保留	保留	保留
CATALOG_NAME	-	非保留	非保留
CHAIN	非保留	非保留	-
CHANGE	非保留	-	-
CHAR	非保留(不能是函数或类型)	保留	保留
CHARACTER	非保留(不能是函数或类型)	保留	保留
CHARACTERISTICS	非保留	-	-
CHARACTER_LENGTH	-	非保留	保留
CHARACTER_SET_CATALOG	-	非保留	非保留
CHARACTER_SET_NAME	-	非保留	非保留
CHARACTER_SET_SCHEMA	-	非保留	非保留
CHAR_LENGTH	-	非保留	保留
CHECK	保留	保留	保留
CHECKED	-	非保留	-
CHECKPOINT	非保留	-	-
CLASS	非保留	保留	-
CLEAN	非保留	-	-

版权所有 © 华为云计算技术有限公司

关键字	DWS	SQL:1999	SQL:1992
CLASS_ORIGIN	-	非保留	非保留
CLOB	非保留	保留	-
CLOSE	非保留	保留	保留
CLUSTER	非保留	-	-
COALESCE	非保留(不能是函数或类型)	非保留	保留
COBOL	-	非保留	非保留
COLLATE	保留	保留	保留
COLLATION	保留(可以是函数或类型)	保留	保留
COLLATION_CATALOG	-	非保留	非保留
COLLATION_NAME	-	非保留	非保留
COLLATION_SCHEMA	-	非保留	非保留
COLUMN	保留	保留	保留
COLUMNS	非保留	-	-
COLUMN_NAME	-	非保留	非保留
COMMAND_FUNCTION	-	非保留	非保留
COMMAND_FUNCTION_CODE	-	非保留	-
COMMENT	非保留	-	-
COMMENTS	非保留	-	-
COMMIT	非保留	保留	保留
COMMITTED	非保留	非保留	非保留
COMPACT	保留(可以是函数或类型)	-	-
COMPATIBLE_ILLEGAL_CHARS	非保留	-	-
COMPLETE	非保留	-	-
COMPRESS	非保留	-	-
COMPLETION	-	保留	-

 版权所有 © 华为云计算技术有限公司

关键字	DWS	SQL:1999	SQL:1992
CONCURRENTLY	保留(可以是函数或类型)	-	-
CONDITION	-	-	-
CONDITION_NUMBER	-	非保留	非保留
CONFIGURATION	非保留	-	-
CONFLICT	非保留	-	-
CONNECT	-	保留	保留
CONNECTION	非保留	保留	保留
CONNECTION_NAME	-	非保留	非保留
CONSTRAINT	保留	保留	保留
CONSTRAINTS	非保留	保留	保留
CONSTRAINT_CATALOG	-	非保留	非保留
CONSTRAINT_NAME	-	非保留	非保留
CONSTRAINT_SCHEMA	-	非保留	非保留
CONSTRUCTOR	-	保留	-
CONTAINS	-	非保留	-
CONTENT	非保留	-	-
CONTINUE	非保留	保留	保留
CONVERSION	非保留	-	-
CONVERT	-	非保留	保留
COORDINATOR	非保留	-	-
COPY	非保留	-	-
CORRESPONDING	-	保留	保留
COST	非保留	-	-
COUNT	-	非保留	保留
CREATE	保留	保留	保留
CROSS	保留(可以是函数或类型)	保留	保留

 版权所有 © 华为云计算技术有限公司

关键字	DWS	SQL:1999	SQL:1992
CSN_NUM	保留(可以是函数或类型) 9.1.1.100集群版本新增。	-	-
CSV	非保留	-	-
CUBE	非保留	保留	-
CURRENT	非保留	保留	保留
CURRENT_CATALOG	保留	-	-
CURRENT_DATE	保留	保留	保留
CURRENT_PATH	-	保留	-
CURRENT_ROLE	保留	保留	-
CURRENT_SCHEMA	保留(可以是函数或类型)	-	-
CURRENT_TIME	保留	保留	保留
CURRENT_TIMESTAMP	保留	保留	保留
CURRENT_USER	保留	保留	保留
CURSOR	非保留	保留	保留
CURSOR_NAME	-	非保留	非保留
CYCLE	非保留	保留	-
DATA	非保留	保留	非保留
DATE_FORMAT	非保留	-	-
DATABASE	非保留	-	-
DATAFILE	非保留	-	-
DATETIME_INTERVAL_CODE	-	非保留	非保留
DATETIME_INTERVAL_PRECISION	-	非保留	非保留
DAY	非保留	保留	保留
DBCOMPATIBILITY	非保留	-	-
DBCOMPATIBILITY_BEHAVIOR	非保留	-	-
DEALLOCATE	非保留	保留	保留

 版权所有 © 华为云计算技术有限公司

关键字	DWS	SQL:1999	SQL:1992
DEC	非保留(不能是函数或类型)	保留	保留
DECIMAL	非保留(不能是函数或类型)	保留	保留
DECLARE	非保留	保留	保留
DECRYPT	非保留 9.1.1.200集群版本新增。	-	-
DECODE	非保留(不能是函数或类型)	-	-
DEDICATED	非保留 9.1.1.100集群版本新增。	-	-
DEFAULT	保留	保留	保留
DEFAULTS	非保留	-	-
DEFERRABLE	保留	保留	保留
DEFERRED	非保留	保留	保留
DEFINED	-	非保留	-
DEFINER	非保留	非保留	-
DELETE	非保留	保留	保留
DELIMITER	非保留	-	-
DELIMITERS	非保留	-	-
DELTA	非保留	-	-
DELTAMERGE	保留(可以是函数或类型)	-	-
DEMAND	非保留	-	-
DEPTH	-	保留	-
DEREF	-	保留	-
DESC	保留	保留	保留
DESCRIBE	-	保留	保留
DESCRIPTOR	-	保留	保留
DESTROY	-	保留	-
DESTRUCTOR	-	保留	-

版权所有 © 华为云计算技术有限公司

关键字	DWS	SQL:1999	SQL:1992
DETERMINISTIC	非保留	保留	-
DIAGNOSTICS	-	保留	保留
DICTIONARY	非保留	保留	-
DIRECT	非保留	-	-
DIRECTORY	非保留	-	-
DISABLE	非保留	-	-
DISCARD	非保留	-	-
DISCONNECT	-	保留	保留
DISPATCH	-	非保留	-
DISTINCT	保留	保留	保留
DISTRIBUTE	非保留	-	-
DISTRIBUTION	非保留	-	-
DO	保留	-	-
DOCUMENT	非保留	-	-
DOMAIN	非保留	保留	保留
DOUBLE	非保留	保留	保留
DROP	非保留	保留	保留
DYNAMIC	-	保留	-
DYNAMIC_FUNCTION	-	非保留	非保留
DYNAMIC_FUNCTION_CODE	-	非保留	-
EACH	非保留	保留	-
ELASTIC	非保留	-	-
ELSE	保留	保留	保留
ENABLE	非保留	-	-
ENCODING	非保留	-	-
ENCRYPT	非保留 9.1.1.200集群版本新增。	-	-
ENCRYPTED	非保留	-	-

 版权所有 © 华为云计算技术有限公司

关键字	DWS	SQL:1999	SQL:1992
ENCRYPTION	非保留 9.1.1.200集群版本新增。	-	-
END	保留	保留	保留
END-EXEC	-	保留	保留
ENFORCED	非保留	-	-
ENUM	非保留	-	-
EOL	非保留	-	-
EQUALS	-	保留	-
ERRORS	非保留	-	-
ESCAPE	非保留	保留	保留
ESCAPING	非保留	-	-
EVERY	非保留	保留	-
EXCEPT	保留	保留	保留
EXCEPTION	-	保留	保留
EXCHANGE	非保留	-	-
EXCLUDE	非保留	-	-
EXCLUDING	非保留	-	-
EXCLUSIVE	非保留	-	-
EXEC	-	保留	保留
EXECUTE	非保留	保留	保留
EXISTING	-	非保留	-
EXISTS	非保留(不能是函数或类型)	非保留	保留
EXPIRATION	非保留	-	-
EXPLAIN	非保留	-	-
EXTENSION	非保留	-	-
EXTERNAL	非保留	保留	保留
EXTRACT	非保留(不能是函数或类型)	非保留	保留
FALSE	保留	保留	保留

版权所有 © 华为云计算技术有限公司

关键字	DWS	SQL:1999	SQL:1992
FAMILY	非保留	-	-
FAST	非保留	-	-
FENCED	保留	-	-
FETCH	保留	保留	保留
FILEHEADER	非保留	-	-
FILL_AVG	非保留(不能是函数或类型)	-	-
FILL_FIRST	非保留(不能是函数或类型)	-	-
FILL_LAST	非保留(不能是函数或类型)	-	-
FILL_MISSING_FIELDS	非保留	-	-
FILTER	非保留	-	-
FINAL	-	非保留	-
FIRST	非保留	保留	保留
FIXED	非保留	保留	保留
FLOAT	非保留(不能是函数或类型)	保留	保留
FOLLOWING	非保留	-	-
FOR	保留	保留	保留
FORCE	非保留	-	-
FOREIGN	保留	保留	保留
FORMATTER	非保留	-	-
FORTRAN	-	非保留	非保留
FORWARD	非保留	-	-
FOUND	-	保留	保留
FREE	-	保留	-
FREEZE	保留(可以是函数或类型)	-	-
FROM	保留	保留	保留
FULL	保留(可以是函数或类型)	保留	保留

版权所有 © 华为云计算技术有限公司

关键字	DWS	SQL:1999	SQL:1992
FUNCTION	非保留	保留	-
FUNCTIONS	非保留	-	-
G	-	非保留	-
GENERAL	-	保留	-
GENERATED	-	非保留	-
GET	-	保留	保留
GLOBAL	非保留	保留	保留
GO	-	保留	保留
GOTO	-	保留	保留
GRANT	保留	保留	保留
GRANTED	非保留	非保留	-
GREATEST	非保留(不能是函数或类型)	-	-
GROUP	保留	保留	保留
GROUPING	非保留(不能是函数或类型)	保留	-
HANDLER	非保留	-	-
HAVING	保留	保留	保留
HDFSDIRECTORY	保留(可以是函数或类型)	-	-
HEADER	非保留	-	-
HIERARCHY	-	非保留	-
HOLD	非保留	非保留	-
HOST	-	保留	-
HOT	保留 9.1.0.210集群版本新增。	-	-
HOUR	非保留	保留	保留
IDENTIFIED	非保留	-	-
IDENTITY	非保留	保留	保留
IF	非保留(不能是函数或类型)	-	-

 版权所有 © 华为云计算技术有限公司

关键字	DWS	SQL:1999	SQL:1992
IFNULL	非保留(不能是函数或类型)	-	-
IGNORE	非保留	保留	-
IGNORE_EXTRA_DATA	非保留	-	-
ILIKE	保留(可以是函数或类型)	-	-
IMMEDIATE	非保留	保留	保留
IMMUTABLE	非保留	-	-
IMPLEMENTATION	-	非保留	-
IMPLICIT	非保留	-	-
IN	保留	保留	保留
INCLUDING	非保留	-	-
INCREMENT	非保留	-	-
INDEX	非保留	-	-
INDEXES	非保留	-	-
INDICATOR	-	保留	保留
INFIX	-	非保留	-
INHERIT	非保留	-	-
INHERITS	非保留	-	-
INITIAL	非保留	-	-
INITIALIZE	-	保留	-
INITIALLY	保留	保留	保留
INITRANS	非保留	-	-
INLINE	非保留	-	-
INNER	保留(可以是函数或类型)	保留	保留
INOUT	非保留(不能是函数或类型)	保留	-
INPUT	非保留	保留	保留
INSENSITIVE	非保留	非保留	保留
INSERT	非保留	保留	保留

版权所有 © 华为云计算技术有限公司

关键字	DWS	SQL:1999	SQL:1992
INSTANCE	-	非保留	-
INSTANTIABLE	-	非保留	-
INSTEAD	非保留	-	-
INT	非保留(不能是函数或类型)	保留	保留
INTEGER	非保留(不能是函数或类型)	保留	保留
INTERNAL	保留	-	-
INTERSECT	保留	保留	保留
INTERVAL	非保留(不能是函数或类型)	保留	保留
INTO	保留	保留	保留
INVALID	非保留	-	-
INVOKER	非保留	非保留	-
IS	保留	保留	保留
ISNULL	非保留 ● 9.1.1.200以下集群版本中不能是函数或类型 ● 9.1.1.200及以上集群版本中可以是函数或类型。	-	-
ISOLATION	非保留	保留	保留
ITERATE	-	保留	-
JOIN	保留(可以是函数或类型)	保留	保留
K	-	非保留	-
KEEP	非保留 9.1.0.210集群版本新增。	-	-
KEEPALIVE	非保留	-	-
KEY	非保留	保留	保留
KEY_MEMBER	-	非保留	-
KEY_TYPE	-	非保留	-

版权所有 © 华为云计算技术有限公司

关键字	DWS	SQL:1999	SQL:1992
LABEL	非保留	-	-
LANGUAGE	非保留	保留	保留
LARGE	非保留	保留	-
LAST	非保留	保留	保留
LATERAL	-	保留	-
LC_COLLATE	非保留	-	-
LC_CTYPE	非保留	-	-
LEADING	保留	保留	保留
LEAKPROOF	非保留	-	-
LEAST	非保留(不能是函数或类型)	-	-
LEFT	保留(可以是函数或类型)	保留	保留
LENGTH	-	非保留	非保留
LESS	保留	保留	-
LEVEL	非保留	保留	保留
LIGHT	非保留	-	-
LIKE	保留(可以是函数或类型)	保留	保留
LIMIT	保留	保留	-
LIST	非保留	-	-
LISTEN	非保留	-	-
LOAD	非保留	-	-
LOCAL	非保留	保留	保留
LOCALTIME	保留	保留	-
LOCALTIMESTAMP	保留	保留	-
LOCATION	非保留	-	-
LOCATOR	-	保留	-
LOCK	非保留	-	-
LOG	非保留	-	-
LOGGING	非保留	-	-

 版权所有 © 华为云计算技术有限公司

关键字	DWS	SQL:1999	SQL:1992
LOGIN	非保留	-	-
LOOP	非保留	-	-
LOWER	-	非保留	保留
M	-	非保留	-
MAP	-	保留	-
MAPPING	非保留	-	-
MASKED	非保留	-	-
MATCH	非保留	保留	保留
MATCHED	非保留	-	-
MATERIALIZED	非保留	-	-
MAX	-	非保留	保留
MAXEXTENTS	非保留	-	-
MAXSIZE	非保留	-	-
MAXTRANS	非保留	-	-
MAXVALUE	保留	-	-
MERGE	非保留	-	-
MESSAGE_LENGTH	-	非保留	非保留
MESSAGE_OCTET_LENGTH	-	非保留	非保留
MESSAGE_TEXT	-	非保留	非保留
METAADDRESS	非保留	-	-
METHOD	-	非保留	-
MIN	-	非保留	保留
MINEXTENTS	非保留	-	-
MINUS	保留	-	-
MINUTE	非保留	保留	保留
MINVALUE	非保留	-	-
MOD	-	非保留	-
MODE	非保留	-	-
MODIFIES	-	保留	-

 版权所有 © 华为云计算技术有限公司

关键字	DWS	SQL:1999	SQL:1992
MODIFY	保留	保留	-
MODULE	-	保留	保留
MONTH	非保留	保留	保留
MORE	-	非保留	非保留
MOVE	非保留	-	-
MOVEMENT	非保留	-	-
MUMPS	-	非保留	非保留
NAME	非保留	非保留	非保留
NAMES	非保留	保留	保留
NATIONAL	非保留(不能是函数或类型)	保留	保留
NATURAL	保留(可以是函数或类型)	保留	保留
NBUCKETS	非保留	-	-
NCHAR	非保留(不能是函数或类型)	保留	保留
NCLOB	-	保留	-
NEW	-	保留	-
NEXT	非保留	保留	保留
NLSSORT	保留	-	-
NO	非保留	保留	保留
NOCOMPRESS	非保留	-	-
NOCYCLE	非保留	-	-
NODE	非保留	-	-
NOLOGGING	非保留	-	-
NOLOGIN	非保留	-	-
NOMAXVALUE	非保留	-	-
NOMINVALUE	非保留	-	-
NONE	非保留(不能是函数或类型)	保留	-
NOT	保留	保留	保留
NOTHING	非保留	-	-

 版权所有 © 华为云计算技术有限公司

关键字	DWS	SQL:1999	SQL:1992
NOTIFY	非保留	-	-
NOTNULL	保留(可以是函数或类型)	-	-
NOWAIT	非保留	-	-
NULL	保留	保留	保留
NULLABLE	-	非保留	非保留
NULLIF	非保留(不能是函数或类型)	非保留	保留
NULLS	非保留	-	-
NUMBER	非保留(不能是函数或类型)	非保留	非保留
NUMERIC	非保留(不能是函数或类型)	保留	保留
NUMSTR	非保留	-	-
NVARCHAR2	非保留(不能是函数或类型)	-	-
NVL	非保留(不能是函数或类型)	-	-
OBJECT	非保留	保留	-
OCTET_LENGTH	-	非保留	保留
OF	非保留	保留	保留
OFF	非保留	保留	-
OFFLOADING_RATE	非保留 9.1.1.100集群版本新增。	-	-
OFFLOADING_STRATEGY	非保留 9.1.1.100集群版本新增。	-	-
OFFSET	保留	-	-
OIDS	非保留	-	-
OLD	-	保留	-
ON	保留	保留	保留
ONLINE	非保留	非保留	非保留
ONLY	保留	保留	保留

版权所有 © 华为云计算技术有限公司

关键字	DWS	SQL:1999	SQL:1992
OPEN	-	保留	保留
OPERATION	-	保留	-
OPERATOR	非保留	-	-
OPTIMIZATION	非保留	-	-
OPTION	非保留	保留	保留
OPTIONS	非保留	非保留	-
OR	保留	保留	保留
ORDER	保留	保留	保留
ORDINALITY	非保留	保留	-
OUT	非保留(不能是函数或类型)	保留	-
OUTER	保留(可以是函数或类型)	保留	保留
OUTLINE	非保留	-	-
OUTPUT	-	保留	保留
OVER	非保留	-	-
OVERLAPS	保留(可以是函数或类型)	非保留	保留
OVERLAY	非保留(不能是函数或类型)	非保留	-
OVERRIDING	-	非保留	-
OVERWRITE	非保留	-	-
OWNED	非保留	-	-
OWNER	非保留	-	-
PACKAGE	非保留	-	-
PAD	-	保留	保留
PARAMETER	-	保留	-
PARAMETERS	-	保留	-
PARAMETER_MODE	-	非保留	-
PARAMETER_NAME	-	非保留	-
PARAMETER_ORDINAL_POSITION	-	非保留	-

版权所有 © 华为云计算技术有限公司

关键字	DWS	SQL:1999	SQL:1992
PARAMETER_SPECIFIC_CATALOG	-	非保留	-
PARAMETER_SPECIFIC_NAME	-	非保留	-
PARAMETER_SPECIFIC_SCHEMA	-	非保留	-
PARENT_NAME	非保留	-	-
PARSER	非保留	-	-
PARTIAL	非保留	保留	保留
PARTITION	非保留	-	-
PARTITIONS	非保留	-	-
PASCAL	-	非保留	非保留
PASSING	非保留	-	-
PASSWORD	非保留	-	-
PATH	-	保留	-
PCTFREE	非保留	-	-
PER	非保留	-	-
PERM	非保留	-	-
PERCENT	非保留	-	-
PERFORMANCE	保留	-	-
PLACING	保留	-	-
PLAN	保留	-	-
PLANS	非保留	-	-
PLI	-	非保留	非保留
POLICY	非保留	-	-
POOL	非保留	-	-
POSITION	非保留(不能是函数或类型)	非保留	保留
POSTFIX	-	保留	-
PRECEDING	非保留	-	-
PRECISION	非保留(不能是函数或类型)	保留	保留

版权所有 © 华为云计算技术有限公司

关键字	DWS	SQL:1999	SQL:1992
PREDICATE	非保留	-	-
PREFERRED	非保留	-	-
PREFIX	非保留	保留	-
PREORDER	-	保留	-
PREPARE	非保留	保留	保留
PREPARED	非保留	-	-
PRESERVE	非保留	保留	保留
PRESERVE_BLANKS	非保留	-	-
PRIMARY	保留	保留	保留
PRIOR	非保留	保留	保留
PRIVATE	非保留	-	-
PRIVILEGE	非保留	-	-
PRIVILEGES	非保留	保留	保留
PROCEDURAL	非保留	-	-
PROCEDURE	保留	保留	保留
PROFILE	非保留	-	-
PUBLIC	-	保留	保留
PUBLICATION	非保留	-	-
PURGE	非保留 9.1.0.210集群版本新增。	-	-
QUERY	非保留	-	-
QUOTE	非保留	-	-
RANGE	非保留	-	-
RAW	非保留	-	-
READ	非保留	保留	保留
READS	-	保留	-
REAL	非保留(不能是函数或类型)	保留	保留
REASSIGN	非保留	-	-
REBUILD	非保留	-	-

版权所有 © 华为云计算技术有限公司

关键字	DWS	SQL:1999	SQL:1992
RECHECK	非保留	-	-
RECURSIVE	非保留	保留	-
RECYCLEBIN	保留(可以是函数或类型) 9.1.0.210集群版本新增。	-	-
REF	非保留	保留	-
REFRESH	非保留	-	-
REFERENCES	保留	保留	保留
REFERENCING	-	保留	-
REINDEX	非保留	-	-
REJECT	保留	-	-
RELATIVE	非保留	保留	保留
RELEASE	非保留	-	-
RELOPTIONS	非保留	-	-
REMOTE	非保留	-	-
RENAME	非保留	-	-
REPEATABLE	非保留	非保留	非保留
REPLACE	非保留	-	-
REPLICA	非保留	-	-
RESET	非保留	-	-
RESIZE	非保留	-	-
RESOURCE	非保留	-	-
RESTART	非保留	-	-
RESTRICT	非保留	保留	保留
RESULT	-	保留	-
RETURN	非保留	保留	-
RETURNED_LENGTH	-	非保留	非保留
RETURNED_OCTET_LENGTH	-	非保留	非保留
RETURNED_SQLSTATE	-	非保留	非保留

 版权所有 © 华为云计算技术有限公司

关键字	DWS	SQL:1999	SQL:1992
RETURNING	保留	-	-
RETURNS	非保留	保留	-
REUSE	非保留	-	-
REVOKE	非保留	保留	保留
REWRITE	非保留	-	-
RIGHT	保留(可以是函数或类型)	保留	保留
ROLE	非保留	保留	-
ROLLBACK	非保留	保留	保留
ROLLUP	非保留	保留	-
ROTATION	非保留 9.1.1.200集群版本新增。	-	-
ROUTINE	-	保留	-
ROUTINE_CATALOG	-	非保留	-
ROUTINE_NAME	-	非保留	-
ROUTINE_SCHEMA	-	非保留	-
ROW	非保留(不能是函数或类型)	保留	-
ROWS	非保留	保留	保留
ROW_COUNT	-	非保留	非保留
RULE	非保留	-	-
SAVEPOINT	非保留	保留	-
SCALE	-	非保留	非保留
SCHEMA	非保留	保留	保留
SCHEMA_NAME	-	非保留	非保留
SCOPE	-	保留	-
SCROLL	非保留	保留	保留
SEARCH	非保留	保留	-
SECOND	非保留	保留	保留
SECTION	-	保留	保留

版权所有 © 华为云计算技术有限公司

关键字	DWS	SQL:1999	SQL:1992
SECURITY	非保留	非保留	-
SELECT	保留	保留	保留
SELF	-	非保留	-
SENSITIVE	-	非保留	-
SEPARATOR	非保留	-	-
SEQUENCE	非保留	保留	-
SEQUENCES	非保留	-	-
SERIALIZABLE	非保留	非保留	非保留
SERVER	非保留	-	-
SERVER_NAME	-	非保留	非保留
SESSION	非保留	保留	保留
SESSION_USER	保留	保留	保留
SET	非保留	保留	保留
SETOF	非保留(不能是函数或类型)	-	-
SETS	非保留	保留	-
SET_RESTORED	非保留 9.1.1.100集群版本新增。	-	-
SHARE	非保留	-	-
SHIPPABLE	非保留	-	-
SHOW	非保留	-	-
SIMILAR	保留(可以是函数或类型)	非保留	-
SIMPLE	非保留	非保留	-
SIZE	非保留	保留	保留
SLOT	非保留	-	-
SMALLDATETIME	非保留(不能是函数或类型)	-	-
SMALLDATETIME_FORMAT	非保留	-	-
SMALLINT	非保留(不能是函数或类型)	保留	保留

版权所有 © 华为云计算技术有限公司

关键字	DWS	SQL:1999	SQL:1992
SNAPSHOT	非保留	-	-
SOME	保留	保留	保留
SOURCE	非保留	非保留	-
SPACE	非保留	保留	保留
SPECIFIC	-	保留	-
SPECIFICTYPE	-	保留	-
SPECIFIC_NAME	-	非保留	-
SPILL	非保留	-	-
SPLIT	非保留	-	-
SQL	非保留	保留	保留
SQLCODE	-	-	保留
SQLERROR	-	-	保留
SQLEXCEPTION	-	保留	-
SQLSTATE	-	保留	保留
SQLWARNING	-	保留	-
STABLE	非保留	-	-
STANDALONE	非保留	-	-
START	非保留	保留	-
STATE	-	保留	-
STATEMENT	非保留	保留	-
STATEMENT_ID	非保留	-	-
STATIC	-	保留	-
STATISTICS	非保留	-	-
STDIN	非保留	-	-
STDOUT	非保留	-	-
STORAGE	非保留	-	-
STORE	非保留	-	-
STRICT	非保留	-	-
STRIP	非保留	-	-
STRUCTURE	-	保留	-

 版权所有 © 华为云计算技术有限公司

关键字	DWS	SQL:1999	SQL:1992
STYLE	-	非保留	-
SUBCLASS_ORIGIN	-	非保留	非保留
SUBLIST	-	非保留	-
SUBSCRIPTION	非保留	-	-
SUBSTRING	非保留(不能是函数或类型)	非保留	保留
SUM	-	非保留	保留
SYMMETRIC	保留	非保留	-
SYNONYM	非保留	-	-
SYS_REFCURSOR	非保留	-	-
SYSDATE	保留	-	-
SYSID	非保留	-	-
SYSTEM	非保留	非保留	-
SYSTEM_USER	-	保留	保留
TABLE	保留	保留	保留
TABLES	非保留	-	-
TABLESAMPLE	保留(可以是函数或类型)	-	-
TABLE_NAME	-	非保留	非保留
TEMP	非保留	-	-
TEMPLATE	非保留	-	-
TEMPORARY	非保留	保留	保留
TERMINATE	-	保留	-
TEXT	非保留	-	-
THAN	非保留	保留	-
THEN	保留	保留	保留
TIME	非保留(不能是函数或类型)	保留	保留
TIME_FILL	非保留(不能是函数或类型)	-	-
TIME_FORMAT	非保留	-	-

 版权所有 © 华为云计算技术有限公司

关键字	DWS	SQL:1999	SQL:1992
TIMECAPSULE	保留(可以是函数或类型) 9.1.0.210集群版本新增。	-	-
TIMESTAMP	非保留(不能是函数或类型)	保留	保留
TIMESTAMPADD	非保留(不能是函数或类型)	-	-
TIMESTAMPDIFF	非保留(不能是函数或类型)	-	-
TIMESTAMP_FORMAT	非保留	-	-
TIMEZONE_HOUR	-	保留	保留
TIMEZONE_MINUTE	-	保留	保留
TINYINT	非保留(不能是函数或类型)	-	-
TO	保留	保留	保留
TRAILING	保留	保留	保留
TRANSACTION	非保留	保留	保留
TRANSACTIONS_COMMITTED	-	非保留	-
TRANSACTIONS_ROLLED_BACK	-	非保留	-
TRANSACTION_ACTIVE	-	非保留	-
TRANSFORM	-	非保留	-
TRANSFORMS	-	非保留	-
TRANSLATE	-	非保留	保留
TRANSLATION	-	保留	保留
TREAT	非保留(不能是函数或类型)	保留	-
TRIGGER	非保留	保留	-
TRIGGER_CATALOG	-	非保留	-
TRIGGER_NAME	-	非保留	-
TRIGGER_SCHEMA	-	非保留	-

 版权所有 © 华为云计算技术有限公司

关键字	DWS	SQL:1999	SQL:1992
TRIM	非保留(不能是函数或类型)	非保留	保留
TRUE	保留	保留	保留
TRUNCATE	非保留	-	-
TRUSTED	非保留	-	-
TRY_CAST	非保留	-	-
TSTAG	保留，该字段仅在实时数仓中使用	-	-
TSTIME	保留，该字段仅在实时数仓中使用	-	-
TSFIELD	保留，该字段仅在实时数仓中使用	-	-
TYPE	非保留	非保留	非保留
TYPES	非保留	-	-
UESCAPE	-	-	-
UNBOUNDED	非保留	-	-
UNCOMMITTED	非保留	非保留	非保留
UNDER	-	保留	-
UNENCRYPTED	非保留	-	-
UNION	保留	保留	保留
UNIQUE	保留	保留	保留
UNKNOWN	非保留	保留	保留
UNLIMITED	非保留	-	-
UNLISTEN	非保留	-	-
UNLOCK	非保留	-	-
UNLOGGED	非保留	-	-
UNNAMED	-	非保留	非保留
UNNEST	-	保留	-
UNTIL	非保留	-	-
UNUSABLE	非保留	-	-
UPDATE	非保留	保留	保留
UPPER	-	非保留	保留

版权所有 © 华为云计算技术有限公司

关键字	DWS	SQL:1999	SQL:1992
USAGE	-	保留	保留
USER	保留	保留	保留
USER_DEFINED_TYPE_CATALOG	-	非保留	-
USER_DEFINED_TYPE_NAME	-	非保留	-
USER_DEFINED_TYPE_SCHEMA	-	非保留	-
USING	保留	保留	保留
VACUUM	非保留	-	-
VALID	非保留	-	-
VALIDATE	非保留	-	-
VALIDATION	非保留	-	-
VALIDATOR	非保留	-	-
VALUE	非保留	保留	保留
VALUES	非保留(不能是函数或类型)	保留	保留
VARCHAR	非保留(不能是函数或类型)	保留	保留
VARCHAR2	非保留(不能是函数或类型)	-	-
VARIABLE	-	保留	-
VARIADIC	保留	-	-
VARYING	非保留	保留	保留
VCGROUP	非保留	-	-
VERBOSE	保留(可以是函数或类型)	-	-
VERIFY	非保留	-	-
VERSION	非保留	-	-
VGROUPID	非保留 9.1.1.100集群版本新增。	-	-
VIEW	非保留	保留	保留

版权所有 © 华为云计算技术有限公司

关键字	DWS	SQL:1999	SQL:1992
VOLATILE	非保留	-	-
WARMUP	保留 9.1.0.210集群版本新增。	-	-
WHEN	保留	保留	保留
WHENEVER	-	保留	保留
WHERE	保留	保留	保留
WHITESPACE	非保留	-	-
WINDOW	保留	-	-
WITH	保留	保留	保留
WITHIN	非保留	-	-
WITHOUT	非保留	保留	-
WORK	非保留	保留	保留
WORKLOAD	非保留	-	-
WRAPPER	非保留	-	-
WRITE	非保留	保留	保留
XML	非保留	-	-
XMLATTRIBUTES	非保留(不能是函数或类型)	-	-
XMLCONCAT	非保留(不能是函数或类型)	-	-
XMLELEMENT	非保留(不能是函数或类型)	-	-
XMLEXISTS	非保留(不能是函数或类型)	-	-
XMLFOREST	非保留(不能是函数或类型)	-	-
XMLNAMESPACES	非保留(不能是函数或类型)	-	-
XMLPARSE	非保留(不能是函数或类型)	-	-
XMLPI	非保留(不能是函数或类型)	-	-

版权所有 © 华为云计算技术有限公司

关键字	DWS	SQL:1999	SQL:1992
XMLROOT	非保留(不能是函数或类型)	-	-
XMLSERIALIZE	非保留(不能是函数或类型)	-	-
XMLTABLE	非保留(不能是函数或类型)	-	-
YEAR	非保留	保留	保留
YES	非保留	-	-
ZONE	非保留	保留	保留

版权所有 © 华为云计算技术有限公司

5 数据类型

5.1 数值类型

数值类型也叫数字类型。由1、2、4或8字节的整数以及4或8字节的浮点数和可选精度小数组成。

对应的操作符和相关函数，请参见**数学函数和运算符**。

DWS支持的数值类型按精度可以分为：整数类型，任意精度型，浮点类型和序列整型。

整数类型

TINYINT、SMALLINT、INTEGER、BINARY_INTEGER和BIGINT类型存储整个数值（不带有小数部分），也就是整数。如果尝试存储超出范围以外的数值将会导致错误。

常用的类型是INTEGER，一般只有取值范围确定不超过SMALLINT的情况下，才会使用SMALLINT类型。而只有在INTEGER的范围不够的时候才使用BIGINT，因为前者相对快得多。

表 5-1 整数类型

名称	描述	存储空间	范围
TINYINT	微整数，别名为INT1。	1字节	0 ~ 255
SMALLINT	小范围整数，别名为INT2。	2字节	-32,768 ~ +32,767
INTEGER	常用的整数，别名为INT4。	4字节	-2,147,483,648 ~ +2,147,483,647
BINARY_INTEGER	常用的整数INTEGER的别名，为兼容Oracle类型。	4字节	-2,147,483,648 ~ +2,147,483,647

版权所有 © 华为云计算技术有限公司

名称	描述	存储空间	范围
BIGINT	大范围的整数，别名为INT8。	8字节	-9,223,372,036,854,775,808 ~ 9,223,372,036,854,775,807

示例：

创建带有TINYINT、INTEGER、BIGINT类型数据的表。

```
CREATE TABLE int_type_t1
(
    a TINYINT,
    b TINYINT,
    c INTEGER,
    d BIGINT
);
```

插入数据。

```
INSERT INTO int_type_t1 VALUES(100, 10, 1000, 10000);
```

查看数据。

```
SELECT * FROM int_type_t1;
 a | b |  c   |   d
---+---+------+-------
 100 | 10 | 1000 | 10000
(1 row)
```

任意精度型

NUMBER类型能够用于存储对于精度位数没有限制的数字，并且可以用于执行精确计算。当要求高精确度时，推荐使用这种类型来存储货币总量和其他类型的数量值。与整数类型相比，任意精度类型需要更大的存储空间，其存储效率、运算效率以及压缩比效果都要差一些。

NUMBER类型数值的范围是小数点右边部分的小数位数。NUMBER类型数值的精度是指整个数值包含的所有数字，也就是小数点左右两边的所有数字。所以，可以说数值23.1234的精度为6，范围是4。可以认为整数的范围是0。

使用Numeric/Decimal进行列定义时，建议指定该列的精度p（总位数）以及范围s（小数位数）。

如果数值的精度或者范围大于列的数据类型所声明的精度和范围，那么系统将会试图对这个值进行四舍五入。如果不能对数值进行四舍五入的处理来满足数据类型的限制，则会报错。

表 5-2 任意精度型

名称	描述	存储空间	范围
NUMERIC[(p[,s])], DECIMAL[(p[,s])]	精度p取值范围为[1,1000]，标度s取值范围为[0,p]。	用户声明精度。每四位（十进制位）占用两个字节，然后在整个数据上加上八个字节的额外开销。	未指定精度的情况下，小数点前最大131,072位，小数点后最大16,383位。

版权所有 © 华为云计算技术有限公司

名称	描述	存储空间	范围
NUMBER[(p[,s])]	NUMERIC类型的别名，为兼容Oracle数据类型。	用户声明精度。每四位（十进制位）占用两个字节，然后在整个数据上加上八个字节的额外开销。	未指定精度的情况下，小数点前最大131,072位，小数点后最大16,383位。

示例：

创建带有DECIMAL数值类型的表。

```
CREATE TABLE decimal_type_t1 (DT_COL1 DECIMAL(10,4));
```

插入数据。

```
INSERT INTO decimal_type_t1 VALUES(123456.122331);
INSERT INTO decimal_type_t1 VALUES(123456.452399);
```

查看数据。

```
SELECT * FROM decimal_type_t1;
   dt_col1
-------------
 123456.1223
 123456.4524
(2 rows)
```

浮点类型

浮点类型属于非精确，可变精度的数值类型。实际上，这些类型通常是对于二进制浮点算术（分别是单精度和双精度）的IEEE标准754的具体实现，在一定范围内由特定的处理器，操作系统和编译器所支持。

表 5-3 浮点类型

名称	描述	存储空间	范围
REAL, FLOAT4	单精度浮点数，不精准。	4字节	6位十进制数字精度。
DOUBLE PRECISION, FLOAT8	双精度浮点数，不精准。	8字节	1E-307~1E+308， 15位十进制数字精度。
FLOAT[(p)]	浮点数，不精准。精度p取值范围为[1,53]。 **说明** p为精度，表示总位数。	4字节或8字节	根据精度p不同选择REAL或DOUBLE PRECISION作为内部表示。如不指定精度，内部用DOUBLE PRECISION表示。

版权所有 © 华为云计算技术有限公司

名称	描述	存储空间	范围
BINARY_DOUBLE	是DOUBLE PRECISION的别名，为兼容Oracle类型。	8字节	1E-307~1E+308， 15位十进制数字精度。
DEC[(p[,s])]	精度p取值范围为[1,1000]，标度s取值范围为[0,p]。 **说明** p为总位数，s为小数位位数。	用户声明精度。每四位（十进制位）占用两个字节，然后在整个数据上加上八个字节的额外开销。	未指定精度的情况下，小数点前最大131,072位，小数点后最大16,383位。
INTEGER[(p[,s])]	精度p取值范围为[1,1000]，标度s取值范围为[0,p]。	用户声明精度。每四位（十进制位）占用两个字节，然后在整个数据上加上八个字节的额外开销。	未指定精度的情况下，小数点前最大131,072位，小数点后最大16,383位。

示例：

创建带有浮点类型的表。

```
CREATE TABLE float_type_t2
(
    FT_COL1 INTEGER,
    FT_COL2 FLOAT4,
    FT_COL3 FLOAT8,
    FT_COL4 FLOAT(3),
    FT_COL5 BINARY_DOUBLE,
    FT_COL6 DECIMAL(10,4),
    FT_COL7 INTEGER(6,3)
) DISTRIBUTE BY HASH ( ft_col1);
```

插入数据。

```
INSERT INTO float_type_t2 VALUES(10,10.365456,123456.1234,10.3214, 321.321, 123.123654, 123.123654);
```

查看数据。

```
SELECT * FROM float_type_t2;
 ft_col1 | ft_col2 |   ft_col3   | ft_col4 | ft_col5 | ft_col6  | ft_col7
---------+---------+-------------+---------+---------+----------+---------
      10 | 10.3655 | 123456.1234 | 10.3214 | 321.321 | 123.1237 | 123.124
(1 row)
```

序列整型

SMALLSERIAL，SERIAL和BIGSERIAL类型不是真正的类型，只是为在表中设置唯一标识而存在的概念。因此，创建一个整数字段，并且把它的缺省数值安排为从一个序列发生器读取。应用了一个NOT NULL约束以确保NULL不会被插入。在大多数情况下用户可能还希望附加一个UNIQUE或PRIMARY KEY约束避免意外地插入重复的数值。最

 版权所有 © 华为云计算技术有限公司

后，将序列发生器从属于那个字段，这样当该字段或表被删除的时候也一并删除该序列。目前只支持在创建表时指定SERIAL列，不可以在已有的表中增加SERIAL列。另外临时表也不支持创建SERIAL列。因为SERIAL不是真正的类型，也不可以将表中存在的列类型转化为SERIAL。

表 5-4 序列整型

名称	描述	存储空间	范围
SMALLSERIAL	二字节序列整型。	2字节	1 ~ 32,767
SERIAL	四字节序列整型。	4字节	1 ~ 2,147,483,647
BIGSERIAL	八字节序列整型。	8字节	1 ~ 9,223,372,036,854,775,807

示例：

创建带有序列类型的表。

```
CREATE TABLE smallserial_type_tab(a SMALLSERIAL);
```

插入数据。

```
INSERT INTO smallserial_type_tab VALUES(default);
```

再次插入数据。

```
INSERT INTO smallserial_type_tab VALUES(default);
```

查看数据。

```
SELECT * FROM smallserial_type_tab;
 a
---
 1
 2
(2 rows)
```

插入NULL值会报错。

```
INSERT INTO smallserial_type_tab VALUES(NULL);
ERROR:  dn_6001_6002: null value in column "a" violates not-null constraint
```

5.2 货币类型

货币类型存储带有固定小数精度的货币金额。表5-5中显示的范围假设有两位小数。可以以任意格式输入，包括整型、浮点型或者典型的货币格式（如“$1,000.00”）。根据区域字符集，输出一般是最后一种形式。

表 5-5 货币类型

名字	存储容量	描述	范围
money	8 字节	货币金额	-92233720368547758.08 到 +92233720368547758.07

版权所有 © 华为云计算技术有限公司

numeric，int和bigint类型的值可以转化为money类型。如果从real和double precision类型转换到money类型，可以先转化为numeric类型，再转化为money类型，例如：

```
SELECT '12.34'::float8::numeric::money;
```

这种用法是不推荐使用的。浮点数不应该用来处理货币类型，因为小数点的位数可能会导致错误。

money类型的值可以转换为numeric类型而不丢失精度。转换为其他类型可能丢失精度，并且必须通过以下两步来完成：

```
SELECT '52093.89'::money::numeric::float8;
```

当一个money类型的值除以另一个money类型的值时，结果是double precision（也就是，一个纯数字，而不是money类型）；在运算过程中货币单位相互抵消。

5.3 布尔类型

表 5-6 布尔类型

名称	描述	存储空间	取值
BOOLEAN	布尔类型	1字节。	• true：真 • false：假 • null：未知（unknown）

“真”值的有效文本值是：

TRUE、't'、'true'、'y'、'yes'、'1'。

“假”值的有效文本值是：

FALSE、'f'、'false'、'n'、'no'、'0'。

使用TRUE和FALSE是比较规范的用法（也是SQL兼容的用法）。

示例

显示用字母t和f输出boolean值。

步骤1 创建表。

```
DROP TABLE IF EXISTS bool_type_t1;
CREATE TABLE bool_type_t1
(
    BT_COL1 BOOLEAN,
    BT_COL2 TEXT
) DISTRIBUTE BY HASH(BT_COL2);
```

步骤2 插入数据。

```
INSERT INTO bool_type_t1 VALUES (TRUE, 'sic est');
INSERT INTO bool_type_t1 VALUES (FALSE, 'non est');
```

步骤3 查询数据。

```
SELECT * FROM bool_type_t1;
```

 版权所有 © 华为云计算技术有限公司

```
gaussdb=> SELECT * FROM bool_type_t1;
 bt_col1 | bt_col2
---------+---------
 t       | sic est
 f       | non est
(2 rows)
```

```
SELECT * FROM bool_type_t1 WHERE bt_col1 = 't';
```

```
gaussdb=> SELECT * FROM bool_type_t1 WHERE bt_col1 = 't';
 bt_col1 | bt_col2
---------+---------
 t       | sic est
(1 row)
```

步骤4 删除表。

```
DROP TABLE bool_type_t1;
```

----结束

5.4 字符类型

在创建表定义字段时，需要根据所存储数据的特征选择合适的字符类型。字符类型选择不当会直接影响存储空间占用和查询性能，甚至引发数据截断等业务问题。DWS提供了多种字符类型，按长度是否固定可分为**定长字符串**和**变长字符串**两大类：

- **定长字符串（如 CHAR、NCHAR）**：适合长度基本固定的数据，如固定位数的编码、性别、国家代码等。不足长度会自动填充空格，超出会截断。适用于数据长度可预期的场景。
- **变长字符串（如 VARCHAR、TEXT）**：只占用实际数据的存储空间，适合长度不固定的数据，如姓名、地址、描述等。为DWS推荐的常用类型。

一般而言，除非明确知道数据类型为固定长度字符串，否则建议使用变长字符串类型（VARCHAR/TEXT），因为定长字符类型在DWS中会存在额外的存储和内存开销。

本文介绍DWS支持的各类字符类型及其长度规则、存储特性，帮助您为字段选择最合适的类型。

字符类型一览

DWS支持的字符类型请参见**表5-7**。字符串操作符和相关的内置函数请参见**字符串处理函数和操作符**。

表 5-7 字符类型

名称	描述	长度	存储空间
• CHAR(n)或BPCHAR或CHARACTER(n) • NCHAR(n)	定长字符串，超出会截断，不足填充空格。	n是指字节长度，最小值1，最大10485760。如不带精度n，默认精度为1。	最大为10MB。

版权所有 © 华为云计算技术有限公司

名称	描述	长度	存储空间
• VARCHAR(n)或CHARACTER VARYING(n) • VARCHAR2(n)	变长字符串。 VARCHAR2(n)是VARCHAR(n)类型的别名，为兼容Oracle类型。	n是指字节长度，最小值1，最大10485760。	最大为10MB。
NVARCHAR2(n)	变长字符串。	n是指字符长度，最小值1，最大10485760。	最大为10MB。
CLOB	变长字符串。文本大对象。是TEXT类型的别名，为兼容Oracle类型。	-	最大为1GB-8203B（即1073733621B）。
TEXT	变长字符串。	-	最大为1GB-8203B（即1073733621B）。

说明

- 除了每列的大小限制以外，每个元组的总大小也不可超过1GB-8203B（即1073733621B）。
- 对于字符串数据，建议使用变长字符串数据类型，并指定最大长度。请务必确保指定的最大长度大于需要存储的最大字符数，避免超出最大长度时出现字符截断现象。除非明确知道数据类型为固定长度字符串，否则，不建议使用CHAR(n)、NCHAR(n)、CHARACTER(n)等定长字符类型。在DWS中，定长字符类型的运算会存在额外的存储和内存开销。
- 一般情况下，如果含有中文字符，用NCHAR/NVARCHAR2，如果纯英文和数字，用CHAR/VARCHAR。

概念说明

为便于理解各类型名称之间的关系，先说明以下概念：

- **CHAR 与 BPCHAR**：BPCHAR 是 CHAR 的别名（内部类型名称为 bpchar，即 blank-padded char）。建表时使用 CHAR(n) 或 BPCHAR(n) 效果完全相同，均表示长度为 n 的定长字符串。
- **CHAR 与 CHARACTER**：CHARACTER(n) 是 SQL 标准中 CHAR(n) 的完整名称，二者等价；同理，CHARACTER VARYING(n) 是 SQL 标准中 VARCHAR(n) 的完整名称。DWS 同时支持短名称与标准名称。
- **CHAR 与 NCHAR**：
  - CHAR(n) 中的 n 表示**字节长度**。
  - NCHAR(n) 对应 SQL 标准的 NATIONAL CHARACTER，其 n 表示**字符长度**，适用于需要按字符个数（而非字节数）限定长度的多语言场景。
  - 在多字节编码下（如 UTF-8 中汉字占 3 字节），同样长度的 CHAR(n) 和 NCHAR(n) 能存放的汉字个数不同，需加以区分。

 版权所有 © 华为云计算技术有限公司

- **VARCHAR2 与 NVARCHAR2**：VARCHAR2(n) 中 n 为**字节长度**（兼容 Oracle）；NVARCHAR2(n) 中 n 为**字符长度**。详见**字节数和字符数**。

特殊字符类型

在DWS里另外还有两种定长字符类型。在**表5-8**里显示。

表 5-8 特殊字符类型

名称	描述	存储空间
name	用于对象名的内部类型，只用在内部系统表中，作为存储标识符。不建议普通用户使用。	64字节（63可用字符加结束符）。
"char"	单字节内部类型，它在系统内部主要用于系统表，主要作为简单化的枚举类型使用。	1字节。

name 类型的对齐与截断机制

name类型长度固定为64字节（**63 个可用字符 + 1 字节结束符**）。当name类型与其他字符类型参与集合运算（如 UNION/UNION ALL）或表达式求值（如 COALESCE、ARRAY）时，结果类型可能被统一对齐为name类型，导致超过63个字符的数据被**静默截断**（不报错），造成数据丢失。

以下示例基于DWS实测结果，完整演示截断现象及规避方法。

步骤1 创建带name 类型字段的测试表，并插入一条数据。

```
DROP TABLE IF EXISTS t_name;
CREATE TABLE t_name (id int, name_col name);
INSERT INTO t_name VALUES (1, 'short_name');
```

步骤2 构造一个长度为 70（大于 63）的 text 字符串用于验证，确认其原始长度。

```
SELECT lengthb('ABCDEFGHIJABCDEFGHIJABCDEFGHIJABCDEFGHIJABCDEFGHIJABCDEFGHIJABCDEFGHIJ') 
AS origin_len,
       length('ABCDEFGHIJABCDEFGHIJABCDEFGHIJABCDEFGHIJABCDEFGHIJABCDEFGHIJABCDEFGHIJ') AS 
origin_chars;
```

预期输出（原始长度70 字节/ 70个字符）：

```
origin_len | origin_chars
------------+--------------
         70 |           70
(1 row)
```

步骤3 使用UNION ALL，将name类型列放在第一个查询中、超长text放在第二个查询中。此时结果类型对齐为name，超长字符串被截断为 63 个字符。

```
SELECT pg_typeof(x)::text AS result_type
FROM (SELECT name_col AS x FROM t_name WHERE id = 1
      UNION ALL
      SELECT 'ABCDEFGHIJABCDEFGHIJABCDEFGHIJABCDEFGHIJABCDEFGHIJABCDEFGHIJABCDEFGHIJ'::text) s
LIMIT 1;
```

预期输出（结果类型为 name）：

```
 result_type
-------------
```

 版权所有 © 华为云计算技术有限公司

```
 name
(1 row)
SELECT x, lengthb(x::text) AS len
FROM (SELECT name_col AS x FROM t_name WHERE id = 1
      UNION ALL
      SELECT 'ABCDEFGHIJABCDEFGHIJABCDEFGHIJABCDEFGHIJABCDEFGHIJABCDEFGHIJABCDEFGHIJ'::text) s
ORDER BY len;
```

预期输出（超长字符串被截断为63个字符，数据丢失）：

```
                               x                               | len
---------------------------------------------------------------------+-----
 short_name                                                    |  10
 ABCDEFGHIJABCDEFGHIJABCDEFGHIJABCDEFGHIJABCDEFGHIJABCDEFGHIJABC |  63
(2 rows)
```

说明： 当UNION/UNION AL的第一个查询返回name类型时，结果类型向name对齐，超出63个字符的部分被静默截断，需重点关注。

步骤4 规避方法：将 name 类型列显式转换为 text（如 name_col::text）参与集合运算，使结果类型对齐为 text，从而完整保留超长字符串。

```
SELECT pg_typeof(x)::text AS result_type
FROM (SELECT name_col::text AS x FROM t_name WHERE id = 1
      UNION ALL
      SELECT 'ABCDEFGHIJABCDEFGHIJABCDEFGHIJABCDEFGHIJABCDEFGHIJABCDEFGHIJABCDEFGHIJ'::text) s
LIMIT 1;
```

预期输出（结果类型为 text）：

```
 result_type
-------------
 text
(1 row)
SELECT x, lengthb(x::text) AS len
FROM (SELECT name_col::text AS x FROM t_name WHERE id = 1
      UNION ALL
      SELECT 'ABCDEFGHIJABCDEFGHIJABCDEFGHIJABCDEFGHIJABCDEFGHIJABCDEFGHIJABCDEFGHIJ'::text) s
ORDER BY len;
```

预期输出（超长字符串完整保留 70 个字符）：

```
                               x                                | len
-----------------------------------------------------------------------+-----
 short_name                                                     |  10
 ABCDEFGHIJABCDEFGHIJABCDEFGHIJABCDEFGHIJABCDEFGHIJABCDEFGHIJABCDEFGHIJ |  70
(2 rows)
```

步骤5 直接向 name 类型列写入超长字符串时，同样会发生静默截断（不报错）。

```
INSERT INTO t_name VALUES (2,
'ABCDEFGHIJABCDEFGHIJABCDEFGHIJABCDEFGHIJABCDEFGHIJABCDEFGHIJABCDEFGHIJ');
SELECT id, name_col::text AS v, lengthb(name_col::text) AS len FROM t_name ORDER BY id;
```

预期输出（id=2 的数据被截断为 63 个字符）：

```
 id |                               v                               | len
----+---------------------------------------------------------------------+-----
  1 | short_name                                                    |  10
  2 | ABCDEFGHIJABCDEFGHIJABCDEFGHIJABCDEFGHIJABCDEFGHIJABCDEFGHIJABC |  63
(2 rows)
```

步骤6 清理测试表。

```
DROP TABLE IF EXISTS t_name;
```

----结束

实测表明，COALESCE(name_col, text)的结果类型为name，ARRAY[name_col, text]的结果类型为 name[]，均会向name对齐。而CASE WHEN在name与text混用时会被

 版权所有 © 华为云计算技术有限公司

解析为text，不会截断（但为兼容不同版本，仍建议显式转换）。因此，凡涉及name类型与其他字符类型混用的场合（UNION、COALESCE、ARRAY、函数返回值等），**务必显式将name转换为text（如 name_col::text）**，避免按 63 个字符对齐截断导致数据丢失。

长度

如果把一个字段定义为char(n)或者varchar(n)， 代表该字段最大可容纳n个长度的数据。无论哪种类型，可设置的最大长度都不得超过10485760（即10MB）。

当数据长度超过指定的长度n时，会抛出错误"value too long"。也可通过指定数据类型，使超过长度的数据自动截断。

示例：

1. 创建表t1，指定其字段的字符类型。
```
CREATE TABLE t1 (a char(5),b varchar(5));
```
2. 向表t1插入数据时超过指定的字节长度报错。
```
INSERT INTO t1 VALUES('bookstore','123');
ERROR:  value too long for type character(5)
CONTEXT:  referenced column: a
```
3. 向表t1插入数据并明确超过指定字节长度后自动截断。
```
INSERT INTO t1 VALUES('bookstore'::char(5),'12345678'::varchar(5));
INSERT 0 1

SELECT a,b FROM t1;
   a   |   b
-------+-------
 books | 12345
(1 row)
```

定长与变长

所有字符类型根据长度是否固定可以分为**定长字符串**与**变长字符串**两大类。

- 对于定长字符串，长度必须确定，如果不指定长度，则默认长度1；如果数据长度不足，会在尾部自动填充空格，用于存储和显示；但这部分填充的数据是无意义的，实际使用中会被忽略，如比较、排序或类型转换。
- 对于变长字符串，若指定长度，则为最大可存储数据长度；如果不指定长度，则认为该字段支持任意长度。

示例：

1. 创建表t2，指定其字段的字符类型。
```
CREATE TABLE t2 (a char(5),b varchar(5));
```
2. 向表t2插入数据并查询字段a的字节长度。因建表时指定a的字符类型为char(5)且是定长字符串，长度不足，填充空格，所以查询的字节长度为5。
```
INSERT INTO t2 VALUES('abc','abc');
INSERT 0 1

SELECT a,lengthb(a),b FROM t2;
   a   | lengthb |  b
-------+---------+-----
 abc   |       5 | abc
(1 row)
```
3. 用函数转换后查询字段a的实际字节长度为3。
```
SELECT a = b from t2;
 ?column?
----------
```

 版权所有 © 华为云计算技术有限公司

```
 t
(1 row)

SELECT cast(a as text) as val,lengthb(val) FROM t2;
 val | lengthb
-----+---------
 abc |       3
(1 row)
```

字节数和字符数

VARCHAR2(n)和NVARCHAR2(n)中长度含义不同，需区别对待。

- VARCHAR2中n为**字节长度**。
- NVARCHAR2中n为**字符长度**。

  说明

  以UTF8编码的数据库为例，字母占1个字节，汉字占3个字节，VARCHAR2(6)可以存放6个字母或2个汉字，而NVARCHAR2(6)可以存放6个字母或6个汉字。

示例：

1. 创建表t3，指定其字段的字符类型。
   ```
   CREATE TABLE t3 (a varchar2(6),b nvarchar2(6));
   ```
2. 向表t3插入数据，超出长度报错。（以UTF8字符编码的数据库中操作为例）
   ```
   INSERT INTO t3 values('产品名','auto');
   ERROR:  value too long for type character varying(6)
   CONTEXT:  referenced column: a

   INSERT INTO t3 values('auto','产品名abcd');
   ERROR:  dn_6003_6004: value too long for type nvarchar2(6)
   CONTEXT:  referenced column: b
   ```
3. 调整插入数据长度后向表t3插入数据成功。
   ```
   INSERT INTO t3 values('产品','产品名abc');
   INSERT 0 1

   SELECT a,b from t3;
    a   |    b
   ------+-----------
    产品 | 产品名abc
   ```

空串与 NULL

Oracle兼容模式下，不区分空串与NULL，执行语句查询或数据导入时会将空串处理为NULL。

由于空串默认被处理为NULL，那就不能使用 = '' 作为查询条件，也不能用is ''。虽然不会有语法错误，但是不会有结果集返回。正确的用法是is null，不等于就是is not null 。

示例：

1. 创建表t4，指定其字段的字符类型。
   ```
   CREATE TABLE t4 (a text);
   ```
2. 向表t4插入数据，插入值中包含空串和NULL。
   ```
   INSERT INTO t4 VALUES('abc'),(''),(null);
   INSERT 0 3
   ```
3. 查询表t4中是否存在空值。
   ```
   SELECT a,a isnull FROM t4;
    a  | ?column?
   ```

 版权所有 © 华为云计算技术有限公司

```
-----+----------
     | t
     | t
 abc | f
(3 rows)

SELECT a,a isnull FROM t4 WHERE a is null;
 a | ?column?
---+----------
   | t
   | t
(2 rows)
```

TD与MySQL兼容模式下，区分空串与null。

- TD兼容模式下示例：

```
SELECT  '' is null , null is null;
 isnull |  isnull
--------+----------
 f      | t
(1 row)
```

- MySQL兼容模式下示例：

```
SELECT  '' is null , null is null;
 isnull |  isnull
--------+----------
 f      | t
(1 row)
```

常见问题

- **Q：CHAR 和 VARCHAR 应该怎么选?**

  A：优先使用 VARCHAR（变长）。除非数据长度严格固定（如固定位数编码），否则使用 CHAR 会因补齐空格而占用更多存储空间，且 DWS 中定长类型运算存在额外开销。含中文字符建议用 NCHAR/NVARCHAR2，纯英文和数字用 CHAR/VARCHAR。

- **Q：插入数据报错 “value too long” 是什么原因?**

  A：表示插入的数据长度超过了字段定义的长度 n。需检查字段定义长度是否足够，或使用 ::char(n)、::varchar(n) 进行显式转换（注意会造成静默截断和数据丢失）。

- **Q：VARCHAR2(n) 能存几个汉字?**

  A：取决于数据库编码：UTF-8 下汉字占 3 字节，VARCHAR2(6) 可存 2 个汉字；GBK 下汉字占 2 字节，VARCHAR2(6) 可存 3 个汉字。若需按字符个数控制长度，请使用 NVARCHAR2。

相关文档

- **字符串处理函数和操作符**：字符串截取、拼接、长度计算等相关内置函数。
- **二进制类型**：BYTEA 等二进制数据类型，适用于存储非文本的原始字节数据。
- **布尔类型**：BOOLEAN 类型，用于存储真/假值。
- **数值类型**：整数、浮点、定点等数值类型，与字符类型转换时需注意。

5.5 二进制类型

DWS支持的二进制类型请参见**表5-9**。

 版权所有 © 华为云计算技术有限公司

表 5-9 二进制类型

名称	描述	存储空间
BLOB	二进制大对象 目前BLOB支持的外部存取接口仅为： ● DBMS_LOB.GETLENGTH ● DBMS_LOB.READ ● DBMS_LOB.WRITE ● DBMS_LOB.WRITEAPPEND ● DBMS_LOB.COPY ● DBMS_LOB.ERASE 这些接口详细说明请参见DBMS_LOB。 **说明** 列存不支持BLOB类型	最大为1G-8203B（即1073733621B）。
RAW	变长的十六进制类型 **说明** 列存不支持RAW类型	4字节加上实际的十六进制字符串。最大为1G-8203B（即1073733621B）。
BYTEA	变长的二进制字符串	4字节加上实际的二进制字符串。最大为1G-8203B（即1073733621B）。

说明

除了每列的大小限制以外，每个元组的总大小也不可超过1G-8203字节。

示例

```
--创建表。
CREATE TABLE blob_type_t1
(
    BT_COL1 INTEGER,
    BT_COL2 BLOB,
    BT_COL3 RAW,
    BT_COL4 BYTEA
) DISTRIBUTE BY REPLICATION;

--插入数据。
INSERT INTO blob_type_t1 VALUES(10,empty_blob(),
HEXTORAW('DEADBEEF'),E'\\xDEADBEEF');

--查询表中的数据。
SELECT * FROM blob_type_t1;
 bt_col1 | bt_col2 | bt_col3  |  bt_col4
---------+---------+----------+------------
      10 |         | DEADBEEF | \xdeadbeef
(1 row)

--删除表。
DROP TABLE blob_type_t1;
```

 版权所有 © 华为云计算技术有限公司

5.6 日期/时间类型

DWS支持的日期/时间类型请参见**表5-10**。该类型的操作符和内置函数请参见**时间日期函数和操作符**。

说明

如果其他的数据库时间格式和DWS的时间格式不一致，可通过修改配置参数**DateStyle**的值来保持一致。

表 5-10 日期/时间类型

名称	描述	存储空间
DATE	Oracle兼容模式下等价于timestamp(0)，记录日期和时间。 其他模式下，记录日期。	Oracle兼容模式下，占存储空间8字节 其他模式下，占存储空间4字节
TIME [(p)] [WITHOUT TIME ZONE]	只用于一日内时间。 p表示小数点后的精度，取值范围为0~6。默认精度为6，表示精确到微秒。	8字节
TIME [(p)] [WITH TIME ZONE]	只用于一日内时间，带时区。 p表示小数点后的精度，取值范围为0~6。默认精度为6，表示精确到微秒。	12字节
TIMESTAMP[(p)] [WITHOUT TIME ZONE]	日期和时间。 p表示小数点后的精度，取值范围为0~6。默认精度为6，表示精确到微秒。	8字节
TIMESTAMP[(p)] [WITH TIME ZONE]	日期和时间，带时区。TIMESTAMP的别名为TIMESTAMPTZ。 p表示小数点后的精度，取值范围为0~6。默认精度为6，表示精确到微秒。	8字节
SMALLDATETIME	日期和时间，不带时区。 精确到分钟，秒位大于等于30秒进一位。	8字节
INTERVAL DAY (l) TO SECOND (p)	时间间隔，X天X小时X分X秒。 • l：天数的精度，取值范围为0~6。为适配Oracle语法，未实现具体功能。 • p：秒数的精度，取值范围为0~6。小数末尾的零不显示。	16字节

版权所有 © 华为云计算技术有限公司

名称	描述	存储空间
INTERVAL [FIELDS] [(p)]	时间间隔。 • fields：可以是YEAR，MONTH，DAY，HOUR，MINUTE，SECOND，DAY TO HOUR，DAY TO MINUTE，DAY TO SECOND，HOUR TO MINUTE，HOUR TO SECOND，MINUTE TO SECOND。 • p：秒数的精度，取值范围为0~6，且fields为SECOND，DAY TO SECOND，HOUR TO SECOND或MINUTE TO SECOND时，参数p才有效。小数末尾的零不显示。 **说明** 9.1.0.200及以上集群版本的MySQL兼容性模式下，当fields为YEAR，MONTH，DAY，HOUR，MINUTE，SECOND时，入参可以为不含正负号的整数或浮点数，例如INTERVAL 1 YEAR, INTERVAL 2.5 MONTH等。	12字节
reltime	相对时间间隔。格式为： X years X mons X days XX:XX:XX。 • 采用儒略历计时，规定一年为365.25天，一个月为30天，计算输入值对应的相对时间间隔，输出采用POSTGRES格式。	4字节

示例：

```
--创建表。
CREATE TABLE date_type_tab(coll date);

--插入数据。
INSERT INTO date_type_tab VALUES (date '12-10-2010');

--查看数据。
SELECT * FROM date_type_tab;
        coll
---------------------
 2010-12-10 00:00:00
(1 row)

--删除表。
DROP TABLE date_type_tab;

--创建表。
CREATE TABLE time_type_tab (da time without time zone ,dai time with time zone,dfgh timestamp without time zone,dfga timestamp with time zone, vbg smalldatetime);

--插入数据。
INSERT INTO time_type_tab VALUES ('21:21:21','21:21:21 pst','2010-12-12','2013-12-11 pst','2003-04-12 04:05:06');
```

 版权所有 © 华为云计算技术有限公司

```
--查看数据。
SELECT * FROM time_type_tab;
    da    |     dai     |        dfgh         |          dfga          |         vbg
----------+-------------+---------------------+------------------------+---------------------
 21:21:21 | 21:21:21-08 | 2010-12-12 00:00:00 | 2013-12-11 16:00:00+08 | 2003-04-12 04:05:00
(1 row)

--删除表。
DROP TABLE time_type_tab;

--创建表。
CREATE TABLE day_type_tab (a int,b INTERVAL DAY(3) TO SECOND (4));

--插入数据。
INSERT INTO day_type_tab VALUES (1, INTERVAL '3' DAY);

--查看数据。
SELECT * FROM day_type_tab;
 a |   b
---+--------
 1 | 3 days
(1 row)

--删除表。
DROP TABLE day_type_tab;

--创建表。
CREATE TABLE year_type_tab(a int, b interval year (6));

--插入数据。
INSERT INTO year_type_tab VALUES(1,interval '2' year);

--查看数据。
SELECT * FROM year_type_tab;
 a |    b
---+---------
 1 | 2 years
(1 row)

--删除表。
DROP TABLE year_type_tab;
```

日期输入

日期和时间的输入几乎可以是任何合理的格式，包括ISO-8601格式、SQL-兼容格式、传统POSTGRES格式或者其他的形式。系统支持按照日、月、年的顺序自定义日期输入。如果把DateStyle参数设置为MDY就按照“月-日-年”解析，设置为DMY就按照“日-月-年”解析，设置为YMD就按照“年-月-日”解析。

日期的文本输入需要加单引号包围，语法如下：

type [(p)] 'value'

可选的精度声明中的p是一个整数，表示在秒域中小数部分的位数。表5-11显示了date类型的输入方式。

表 5-11 日期输入方式

例子	描述
1999-01-08	ISO 8601格式（建议格式），任何方式下都是1999年1月8号。

 版权所有 © 华为云计算技术有限公司

例子	描述
January 8, 1999	在任何datestyle输入模式下都无歧义。
1/8/1999	有歧义，在MDY模式下是一月八号，在DMY模式下是八月一号。
1/18/1999	MDY模式下是一月十八日，其他模式下被拒绝。
01/02/03	• MDY模式下的2003年1月2日。 • DMY模式下的2003年2月1日。 • YMD模式下的2001年2月3日。
1999-Jan-08	任何模式下都是1月8日。
Jan-08-1999	任何模式下都是1月8日。
08-Jan-1999	任何模式下都是1月8日。
99-Jan-08	YMD模式下是1月8日，否则错误。
08-Jan-99	一月八日，除了在YMD模式下是错误的之外。
Jan-08-99	一月八日，除了在YMD模式下是错误的之外。
19990108	ISO 8601；任何模式下都是1999年1月8日。
990108	ISO 8601；任何模式下都是1999年1月8日。
1999.008	年和年里的第几天。
J2451187	儒略日。
January 8, 99 BC	公元前99年。

示例：

```
--创建表。
CREATE TABLE date_type_tab(coll date);

--插入数据。
INSERT INTO date_type_tab VALUES (date '12-10-2010');

--查看数据。
SELECT * FROM date_type_tab;
        coll
---------------------
 2010-12-10 00:00:00
(1 row)

--查看日期格式。
SHOW datestyle;
 DateStyle
-----------
 ISO, MDY
(1 row)

--设置日期格式。
SET datestyle='YMD';
SET
```

 版权所有 © 华为云计算技术有限公司

```
--插入数据。
INSERT INTO date_type_tab VALUES(date '2010-12-11');

--查看数据。
SELECT * FROM date_type_tab;
        coll
---------------------
 2010-12-10 00:00:00
 2010-12-11 00:00:00
(2 rows)

--删除表。
DROP TABLE date_type_tab;
```

时间

时间类型包括time [(p)] without time zone和time [(p)] with time zone。如果只写time等效于time without time zone。

如果在time without time zone类型的输入中声明了时区，则会忽略这个时区。

时间输入类型的详细信息请参见表5-12，时区输入类型的详细信息请参见表5-13。

表 5-12 时间输入

例子	描述
05:06.8	ISO 8601
4:05:06	ISO 8601
4:05	ISO 8601
40506	ISO 8601
4:05 AM	与04:05一样，AM不影响数值
4:05 PM	与16:05一样，输入小时数必须<= 12
04:05:06.789-8	ISO 8601
04:05:06-08:00	ISO 8601
04:05-08:00	ISO 8601
040506-08	ISO 8601
04:05:06 PST	缩写的时区

表 5-13 时区输入

例子	描述
PST	太平洋标准时间（Pacific Standard Time）
-8:00	ISO 8601与PST的偏移
-800	ISO 8601与PST的偏移

版权所有 © 华为云计算技术有限公司

例子	描述
-8	ISO 8601与PST的偏移

示例：

```
SELECT time '04:05:06';
   time
----------
 04:05:06
(1 row)

SELECT time '04:05:06 PST';
   time
----------
 04:05:06
(1 row)

SELECT time with time zone '04:05:06 PST';
   timetz
-------------
 04:05:06-08
(1 row)
```

特殊值

DWS支持几个特殊值，在读取的时候将被转换成普通的日期/时间值，请参考**表5-14**。

表 5-14 特殊值

输入字符串	适用类型	描述
epoch	date，timestamp	1970-01-01 00:00:00+00 （Unix系统零时）
infinity	timestamp	比任何其他时间戳都晚
-infinity	timestamp	比任何其他时间戳都早
now	date，time，timestamp	当前事务的开始时间
today	date，timestamp	今日午夜
tomorrow	date，timestamp	明日午夜
yesterday	date，timestamp	昨日午夜
allballs	time	00:00:00.00 UTC

时间段输入

reltime的输入方式可以采用任何合法的时间段文本格式，包括数字形式（含负数和小数）及时间形式，其中时间形式的输入支持SQL标准格式、ISO-8601格式、POSTGRES格式等。另外，文本输入需要加单引号。

时间段输入的详细信息请参考**表6 时间段输入**。

版权所有 © 华为云计算技术有限公司

表 5-15 时间段输入

输入示例	输出结果	描述
60	2 mons	采用数字表示时间段，默认单位是day，可以是小数或负数。特别的，负数时间段，在语义上，可以理解为“早于多久”。
31.25	1 mons 1 days 06:00:00	
-365	-12 mons -5 days	
1 years 1 mons 8 days 12:00:00	1 years 1 mons 8 days 12:00:00	采用POSTGRES格式表示时间段，可以正负混用，不区分大小写，输出结果为将输入时间段计算并转换得到的简化POSTGRES格式时间段。
-13 months -10 hours	-1 years -25 days -04:00:00	
-2 YEARS +5 MONTHS 10 DAYS	-1 years -6 mons -25 days -06:00:00	
P-1.1Y10M	-3 mons -5 days -06:00:00	采用ISO-8601格式表示时间段，可以正负混用，不区分大小写，输出结果为将输入时间段计算并转换得到的简化POSTGRES格式时间段。
-12H	-12:00:00	

示例：

```
--创建表。
CREATE TABLE reltime_type_tab(col1 character(30), col2 reltime);

--插入数据。
INSERT INTO reltime_type_tab VALUES ('90', '90');
INSERT INTO reltime_type_tab VALUES ('-366', '-366');
INSERT INTO reltime_type_tab VALUES ('1975.25', '1975.25');
INSERT INTO reltime_type_tab VALUES ('-2 YEARS +5 MONTHS 10 DAYS', '-2 YEARS +5 MONTHS 10 DAYS');
INSERT INTO reltime_type_tab VALUES ('30 DAYS 12:00:00', '30 DAYS 12:00:00');
INSERT INTO reltime_type_tab VALUES ('P-1.1Y10M', 'P-1.1Y10M');

--查看数据。
SELECT * FROM reltime_type_tab;
             col1              |               col2
--------------------------------+-----------------------------------
 1975.25                        | 5 years 4 mons 29 days
 -2 YEARS +5 MONTHS 10 DAYS     | -1 years -6 mons -25 days -06:00:00
 P-1.1Y10M                      | -3 mons -5 days -06:00:00
 -366                           | -1 years -18:00:00
 90                             | 3 mons
 30 DAYS 12:00:00               | 1 mon 12:00:00
(6 rows)

--删除表。
DROP TABLE reltime_type_tab;
```

5.7 几何类型

DWS支持的几何类型请参见表5-16。最基本的类型：点，是其他类型的基础。

版权所有 © 华为云计算技术有限公司

表 5-16 几何类型

名字	存储空间	说明	表现形式
point	16字节	平面中的点	(x,y)
lseg	32字节	（有限）线段	((x1,y1),(x2,y2))
box	32字节	矩形	((x1,y1),(x2,y2))
path	16+16n字节	闭合路径（与多边形类似）	((x1,y1),...)
path	16+16n字节	开放路径	[(x1,y1),...]
polygon	40+16n字节	多边形（与闭合路径相似）	((x1,y1),...)
circle	24 字节	圆	<(x,y),r>（圆心和半径）

DWS提供了一系列的函数和操作符用来进行各种几何计算，如拉伸、转换、旋转、计算相交等。详细信息请参考**几何函数和操作符**。

点

点是几何类型的基本二维构造单位。用下面语法描述point的数值：

```
( x , y )
x , y
```

x和y是用浮点数表示的点的坐标。

点输出使用第一种语法。

线段

线段（lseg）是用一对点来代表的。用下面的语法描述lseg的数值：

```
[ ( x1 , y1 ) , ( x2 , y2 ) ]
( ( x1 , y1 ) , ( x2 , y2 ) )
( x1 , y1 ) , ( x2 , y2 )
x1 , y1   ,   x2 , y2
```

(x1,y1)和(x2,y2)表示线段的端点。

线段输出使用第一种语法。

矩形

矩形是用一对对角点来表示的。用下面的语法描述box的值：

```
( ( x1 , y1 ) , ( x2 , y2 ) )
( x1 , y1 ) , ( x2 , y2 )
x1 , y1   ,   x2 , y2
```

(x1,y1)和(x2,y2)表示矩形的一对对角点。

矩形的输出使用第二种语法。

任何两个对角都可以出现在输入中，但按照那样的顺序，右上角和左下角的值会被重新排序以存储。

 版权所有 © 华为云计算技术有限公司

路径

路径由一系列连接的点组成。路径可能是开放的，也就是认为列表中第一个点和最后一个点没有连接，也可能是闭合的，这时认为第一个和最后一个点连接起来。

用下面的语法描述path的数值：

```
[ ( x1 , y1 ) , ... , ( xn , yn ) ]
( ( x1 , y1 ) , ... , ( xn , yn ) )
( x1 , y1 ) , ... , ( xn , yn )
( x1 , y1   , ... ,  xn , yn )
x1 , y1   , ... ,  xn , yn
```

- 点，表示组成路径的线段的端点。
- 方括弧（ []），表示开放路径。
- 圆括号（ ()），表示闭合路径。

如果省略最外层的括号（如简化语法场景），则默认视为闭合路径。系统在输出路径时，会统一使用带括号的语法（即使用方括号或圆括号）。

多边形

多边形由一系列点代表（多边形的顶点）。多边形可以认为与闭合路径一样，但是存储方式不一样而且有自己的一套支持函数。

用下面的语法描述polygon的数值：

```
( ( x1 , y1 ) , ... , ( xn , yn ) )
( x1 , y1 ) , ... , ( xn , yn )
( x1 , y1   , ... ,  xn , yn )
x1 , y1   , ... ,  xn , yn
```

点表示多边形的端点。

多边形输出使用第一种语法。

圆

圆由一个圆心和半径标识。用下面的语法描述circle的数值：

```
< ( x , y ) , r >
( ( x , y ) , r )
( x , y ) , r
x , y   , r
```

(x,y)表示圆心，r表示半径。

圆的输出用第一种格式。

5.8 数组类型

数组是一组数据的集合。数组类型允许在单个数据库字段中存储多个值。数组类型通常用于存储和处理具有相似属性的数据。

语法格式

```
ARRAY [ param ]
```

或

```
'{ param }'
```

 版权所有 © 华为云计算技术有限公司

其中参数param说明如下：

- param ： 数组包含的值，允许出现零个或多个，多个值之间用逗号分隔，没有值可填写为NULL。
- 以'{ param }' 这种格式作为数组常量时，其中的字符串类型的元素不能再以单引号开始和结束，需要使用双引号。两个连续单引号转换为一个单引号。
- 以第一个元素的数据类型作为数组的数据类型，要求数组中所有元素的类型相同，或者能够相互转换。

数组类型的定义

一个数组数据类型可以通过在数组元素的数据类型名称后面加上方括号（ [] ）来命名。

例如，创建表books，其中表示书本价格的列price的类型为一维integer类型数组，表示书本标签的列tag的类型为二维text类型数组。

```
CREATE TABLE books (id SERIAL PRIMARY KEY, title VARCHAR(100), price_by_quarter int[], tags TEXT[][]);
```

CREATE TABLE语法可以指定数组的大小，例如：

```
CREATE TABLE test ( a int[3]);
```

当前的数据库实现会忽略语句中数组的大小限制，即其行为与未指定长度的数组相同。同时，也不会强制所声明的维度数。一个特定元素类型的数组全部被当作是相同的类型，而忽略其大小或维度数。

也可以使用关键词ARRAY来定义一维数组。表books中的列price使用ARRAY定义并指定数组大小，如下所示：

```
price_by_quarter int ARRAY[4]
```

使用ARRAY定义，不指定数组尺寸：

```
price_by_quarter int ARRAY
```

数组值输入

输入数组值时要把一个数组值写成一个文字常数，将元素值用花括号包围并用逗号分隔。一个数组常量的一般格式如下：

```
'{ val1 delim val2 delim ... }'
```

其中，delim是类型的定界符，每个val可以是数组元素类型的一个常量或子数组。

一个数组常量的例子如下：

```
'{{1,2,3},{4,5,6},{7,8,9}}'
```

该常量是一个二维的，3乘3数组，它由3个整数子数组构成。

向表books插入数据并查询表books：

```
INSERT INTO books
  VALUES (1, 'One Hundred Years of Solitude','{25,25,25,25}','{{"fiction"}, {"adventure"}}'),
         (2, 'Robinson Crusoe', '{30,32,32,32}', '{{"adventure"}, {"fiction"}}'),
         (3, 'Gone with the Wind', '{27,27,29,28}', '{{"romance"}, {"fantasy"}}');

SELECT * FROM books;
 id |            title            | price_by_quarter |          tags
----+-----------------------------+------------------+-------------------------
  1 | One Hundred Years of Solitude | {25,25,25,25}    | {{fiction},{adventure}}
```


版权所有 © 华为云计算技术有限公司

```
 2 | Robinson Crusoe              | {30,32,32,32}    | {{adventure},{fiction}}
 3 | Gone with the Wind           | {27,27,29,28}    | {{romance},{fantasy}}
(3 rows)
```

注意

插入多维数组数据时，多维数组的每一维都必须有相匹配的长度。

使用ARRAY关键字插入数据：

```
INSERT INTO books
  VALUES (1, 'One Hundred Years of Solitude', ARRAY[25,25,25,25], ARRAY['fiction', 'adventure']),
         (2, 'Robinson Crusoe', ARRAY[30,32,32,32], ARRAY['adventure', 'fiction']),
         (3, 'Gone with the Wind', ARRAY[27,27,29,28], ARRAY['romance', 'fantasy']);
```

访问数组

访问数组元素

查询在第二季度价格发生变化的书籍的名称：

```
SELECT title FROM books WHERE price_by_quarter[1] <> price_by_quarter[2];
      title
-----------------
 Robinson Crusoe
(1 row)
```

查询检索所有书籍第三季度的价格：

```
SELECT price_by_quarter[3] FROM books;
 price_by_quarter
------------------
               29
               25
               32
(3 rows)
```

访问数组切片

可以访问一个数组的任意矩形切片或者子数组。一个数组切片可以通过在一个或多个数组维度上指定[下界:上界]来定义。

查询Gone with the Wind的第二个标签：

```
SELECT tags[2:2] FROM books WHERE title = 'Gone with the Wind';
   tags
-----------
 {fantasy}
(1 row)
```

使用函数访问数组

使用array_dims函数获得数组值的当前维度：

```
SELECT array_dims(tags) FROM books WHERE title = 'Robinson Crusoe';
 array_dims
------------
 [1:2]
(1 row)
```

也可以使用array_upper和array_lower来获得数组维度，它们将分别返回一个指定数组的上界和下界：

```
SELECT array_upper(tags, 1) FROM books WHERE title = 'Robinson Crusoe';
 array_upper
```

 版权所有 © 华为云计算技术有限公司

```
-------------
           2
(1 row)
```

array_length函数将返回一个指定数组维度的长度：

```
SELECT array_length(tags, 1) FROM books WHERE title = 'Robinson Crusoe';
 array_length
--------------
            2
(1 row)
```

修改数组

更新数组

更新整个数组数据：

```
UPDATE books SET price_by_quarter = '{30,30,30,30}'
    WHERE title = 'Robinson Crusoe';
```

使用ARRAY表达式语法更新整个数组数据：

```
UPDATE books SET price_by_quarter = ARRAY[30,30,30,30]
    WHERE title = 'Robinson Crusoe';
```

更新数组中的一个元素：

```
UPDATE books SET price_by_quarter[4] = 35
    WHERE title = 'Robinson Crusoe';
```

更新数组中的一个切片元素：

```
UPDATE books SET price_by_quarter[1:2] = '{27,27}'
    WHERE title = 'Robinson Crusoe';
```

一个已存储的数组值可以被通过对其还不存在的元素赋值来扩大大小。任何位于已存在的元素和新元素之间的位置都将被空值填充。例如，如果数组myarray目前有4个元素，使用UPDATE对myarray[6]赋值后它将有6个元素，其中myarray[5]为空值。目前，采用这种方式扩大数组只允许使用在一维数组上。

构建新数组

新的数组值也可以通过串接操作符“||”构建。串接操作符允许把一个单独的元素加入到一个一维数组的开头或末尾。也可接受两个N维数组，或者一个N维数组和一个N+1维数组。

```
SELECT ARRAY[1,2] || ARRAY[3,4];
 ?column?
-----------
 {1,2,3,4}
(1 row)

SELECT ARRAY[5,6] || ARRAY[[1,2],[3,4]];
     ?column?
---------------------
 {{5,6},{1,2},{3,4}}
(1 row)
```

使用函数array_prepend、array_append或array_cat构建数组。

```
SELECT array_prepend(1, ARRAY[2,3]);
 array_prepend
---------------
 {1,2,3}
(1 row)
```

 版权所有 © 华为云计算技术有限公司

```
SELECT array_append(ARRAY[1,2], 3);
 array_append
--------------
 {1,2,3}
(1 row)

SELECT array_cat(ARRAY[1,2], ARRAY[3,4]);
 array_cat
-----------
 {1,2,3,4}
(1 row)

SELECT array_cat(ARRAY[[1,2],[3,4]], ARRAY[5,6]);

     array_cat
---------------------
 {{1,2},{3,4},{5,6}}
(1 row)

SELECT array_cat(ARRAY[5,6], ARRAY[[1,2],[3,4]]);
     array_cat
---------------------
 {{5,6},{1,2},{3,4}}
(1 row)
```

5.9 枚举类型

枚举（enum）类型是由一个静态、值的有序集合构成的数据类型。它们等效于很多编程语言所支持的enum类型。枚举类型可以是一周中的日期，或者一个数据的状态值集合。

枚举类型的声明

枚举类型可以使用CREATE TYPE命令创建，例如：

```
CREATE TYPE mood AS ENUM ('sad', 'ok', 'happy');
```

枚举类型被创建后，可以在表和函数定义中使用：

```
CREATE TYPE mood AS ENUM ('sad', 'ok', 'happy');
CREATE TABLE person (name text, current_mood mood);
INSERT INTO person VALUES ('Moe', 'happy');
SELECT * FROM person WHERE current_mood = 'happy';
 name | current_mood
------+--------------
 Moe  | happy
(1 row)
```

排序

枚举类型值的排序是该类型被创建时所列出的值的顺序。

```
INSERT INTO person VALUES ('Larry', 'sad');
INSERT INTO person VALUES ('Curly', 'ok');

SELECT * FROM person WHERE current_mood > 'sad';
 name  | current_mood
-------+--------------
 Moe   | happy
 Curly | ok
(2 rows)

SELECT * FROM person WHERE current_mood > 'sad' ORDER BY current_mood;
 name  | current_mood
-------+--------------
```


版权所有 © 华为云计算技术有限公司

```
 Curly | ok
 Moe   | happy
(2 rows)

SELECT name FROM person WHERE current_mood = (SELECT MIN(current_mood) FROM person);
 name
-------
 Larry
(1 row)
```

枚举类型的安全性

每一种枚举数据类型都是独立的并且不能和其他枚举类型相比较。

```
CREATE TYPE happiness AS ENUM ('happy', 'very happy', 'ecstatic');
CREATE TABLE holidays (num_weeks integer, happiness happiness);
INSERT INTO holidays(num_weeks,happiness) VALUES (4, 'happy');
INSERT INTO holidays(num_weeks,happiness) VALUES (6, 'very happy');
INSERT INTO holidays(num_weeks,happiness) VALUES (8, 'ecstatic');

INSERT INTO holidays(num_weeks,happiness) VALUES (2, 'sad');
ERROR:  invalid input value for enum happiness: "sad"

SELECT person.name, holidays.num_weeks FROM person, holidays
      WHERE person.current_mood = holidays.happiness;
ERROR:  operator does not exist: mood = happiness
```

如果需要作比较，可以使用自定义的操作符或者在查询中加上显式类型：

```
 SELECT person.name, holidays.num_weeks FROM person, holidays
      WHERE person.current_mood::text = holidays.happiness::text;
 name | num_weeks
------+-----------
 Moe  |         4
(1 row)
```

注意事项

- 枚举标签是大小写敏感的，因此'happy'与'HAPPY'是不同的。标签中的空格也是有意义的。
- 尽管枚举类型的主要目的是用于值的静态集合，但也有方法在现有枚举类型中增加新值和重命名值（见**ALTER TYPE**）。不能从枚举类型中去除现有的值，也不能更改这些值的排序顺序，除非删除并且重建枚举类型。
- 从内部枚举值到文本标签的翻译保存在系统表PG_ENUM中。

5.10 网络地址类型

DWS提供用于存储IPv4、IPv6、MAC地址的数据类型。

网络地址类型提供输入错误检查和特殊的操作和功能（请参见**网络地址函数和操作符**），比纯文本类型更适合存储IPv4、IPv6、MAC地址的数据类型。

表 5-17 网络地址类型

名字	存储空间	描述
cidr	7或19字节	IPv4或IPv6网络
inet	7或19字节	IPv4或IPv6主机和网络

版权所有 © 华为云计算技术有限公司

名字	存储空间	描述
macaddr	6字节	MAC地址

在对inet或cidr数据类型进行排序的时候，IPv4地址总是排在IPv6地址前面，包括那些封装或者是映射在IPv6地址里的IPv4地址，例如::10.2.3.4或::ffff:10.4.3.2。

cidr

cidr（无类别域间路由，Classless Inter-Domain Routing）类型，保存一个IPv4或IPv6网络地址。声明网络格式为address/y，address表示IPv4或者IPv6地址，y表示子网掩码的二进制位数。如果省略y，则掩码部分使用已有类别的网络编号系统进行计算，但要求输入的数据已经包括了确定掩码所需的所有字节。

- 示例一：CIDR格式换算为IP地址网段

  例如，10.0.0.0/8换算为32位二进制地址：00001010.00000000.00000000.00000000。其中/8表示8位网络ID，即32位二进制地址中前8位是固定不变的，对应网段为：00001010.00000000.00000000.00000000~00001010.11111111.11111111.11111111。则换算为十进制后，10.0.0.0/8表示：子网掩码为255.0.0.0，对应网段为10.0.0.0~10.255.255.255。
- 示例二：IP地址网段换算为CIDR格式

  例如，192.168.0.0~192.168.31.255，后两段IP换算为二进制地址：00000000.00000000~00011111.11111111，可以得出前19位（8*2+3）是固定不变的，则换算为CIDR格式后，表示为：192.168.0.0/19。

表 5-18 cidr 类型输入举例

cidr输入	cidr输出	abbrev（cidr）
192.168.100.128/25	192.168.100.128/25	192.168.100.128/25
192.168/24	192.168.0.0/24	192.168.0/24
192.168/25	192.168.0.0/25	192.168.0.0/25
192.168.1	192.168.1.0/24	192.168.1/24
192.168	192.168.0.0/24	192.168.0/24
10.1.2	10.1.2.0/24	10.1.2/24
10.1	10.1.0.0/16	10.1/16
10	10.0.0.0/8	10/8
10.1.2.3/32	10.1.2.3/32	10.1.2.3/32
2001:4f8:3:ba::/64	2001:4f8:3:ba::/64	2001:4f8:3:ba::/64
2001:4f8:3:ba:2e0:81ff:fe22:d1f1/128	2001:4f8:3:ba:2e0:81ff:fe22:d1f1/128	2001:4f8:3:ba:2e0:81ff:fe22:d1f1

 版权所有 © 华为云计算技术有限公司

inet

inet类型在一个数据区域内保存主机的IPv4或IPv6地址，以及一个可选子网。主机地址中网络地址的位数表示子网（“子网掩码”）。如果子网掩码是32并且地址是IPv4，则这个值不表示任何子网，只表示一台主机。在IPv6里，地址长度是128位，因此128位表示唯一的主机地址。

该类型的输入格式是address/y，address表示IPv4或者IPv6地址，y是子网掩码的二进制位数。如果省略/y，则子网掩码对IPv4是32，对IPv6是128，所以该值表示只有一台主机。如果该值表示只有一台主机，/y将不会显示。

inet和cidr类型之间的基本区别是inet接受子网掩码，而cidr不接受。

macaddr

macaddr类型存储MAC地址，也就是以太网卡硬件地址（尽管MAC地址还用于其他用途）。可以接受下列格式：

```
'08:00:2b:01:02:03'
'08-00-2b-01-02-03'
'08002b:010203'
'08002b-010203'
'0800.2b01.0203'
'08002b010203'
```

这些示例都表示同一个地址。对于数据位a到f，大小写都行。输出时都是以第一种形式展示。

5.11 位串类型

位串就是一串1和0的字符串。它们可以用于存储位掩码。

DWS支持两种位串类型：bit(n)和bit varying(n)，其中n是一个正整数。

bit类型的数据必须准确匹配长度n，如果存储短或者长的数据都会报错。bit varying类型的数据是最长为n的变长类型，超过n的类型会被拒绝。一个没有长度的bit等效于bit(1)，没有长度的bit varying表示没有长度限制。

说明

如果显式地把一个位串值转换成bit(n)，则此位串右边的内容将被截断或者在右边补齐零，直到刚好n位，而且不会抛出任何错误。类似地，如果显式地把一个位串数值转换成bit varying(n)，如果它超过了n位，则它的右边将被截断。

位串类型使用示例：

1. 创建示例表bit_type_t1：

```
CREATE TABLE bit_type_t1
(
    BT_COL1 INTEGER,
    BT_COL2 BIT(3),
    BT_COL3 BIT VARYING(5)
) DISTRIBUTE BY REPLICATION;
```

2. 插入数据：

```
INSERT INTO bit_type_t1 VALUES(1, B'101', B'00');
```

插入数据的长度不符合类型的标准会报错。

```
INSERT INTO bit_type_t1 VALUES(2, B'10', B'101');
ERROR:  bit string length 2 does not match type bit(3)
CONTEXT:  referenced column: bt_col2
```

 版权所有 © 华为云计算技术有限公司

3. 将不符合类型长度的数据进行转换：

```
INSERT INTO bit_type_t1 VALUES(2, B'10'::bit(3), B'101');
```

4. 查看数据：

```
SELECT * FROM bit_type_t1;
 bt_col1 | bt_col2 | bt_col3
---------+---------+---------
       1 | 101     | 00
       2 | 100     | 101
(2 rows)
```

5.12 文本搜索类型

在海量文本数据中快速检索包含特定关键词的记录，是数据仓库应用中的常见需求。例如，在日志分析场景中，使用tsvector和tsquery可以高效检索包含特定关键词的日志记录；在电商系统中，根据商品描述进行全文搜索以匹配用户查询；在文档管理系统中，对大量文档内容进行关键词检索和相关性排序。传统的关系型数据库使用LIKE或~（正则表达式）进行文本匹配，但在大数据量场景下性能较差，且无法进行语义层面的词干匹配和相关性排序。

DWS提供了tsvector和tsquery两种数据类型用于支持全文检索。tsvector类型表示为文本搜索优化的文档分词格式，tsquery类型表示文本查询条件。两者配合使用，通过@@操作符进行匹配，能够高效地实现全文检索功能。

基本概念

- **全文检索（Full-Text Search）**：一种在文本数据中检索包含特定关键词或语义的记录的技术。与简单的LIKE模式匹配不同，全文检索会对文本进行分词和词干提取，支持语义层面的匹配和相关性排序，适用于大数据量的文本检索场景。
- **分词器（Parser）**：负责将原始文本拆分为多个词条（lexeme）的组件。分词器按空格、标点等分隔符将文本拆分，并识别词条的边界。DWS内置了多种分词器配置（如english、default），可通过to_tsvector('配置名', 文本)指定使用的分词器。
- **词干（Lexeme）**：词条经过词干提取后的规范化形式。例如，英文单词”Rats”的词干为“rat”，“running”的词干为“run”。词干提取使得检索能够匹配同一词的不同形式（如单复数、时态变化），提高检索的召回率。

tsvector

tsvector类型表示一个检索单元，通常是一个数据库表中的一行文本字段或者这些字段的组合。

tsvector类型的值是唯一分词的分类列表，把一句话的词格式化为不同的词条，在进行分词处理的时候tsvector会按照一定的顺序录入，并自动去掉分词中重复的词条。

to_tsvector函数通常用于解析和标准化文档字符串。

基本分词

通过tsvector把一个字符串按照空格进行分词，分词的顺序是按照字母和长短排序的，请看以下例子：

```
SELECT 'a fat cat sat on a mat and ate a fat rat'::tsvector;
                    tsvector
-----------------------------------------------------
 'a' 'and' 'ate' 'cat' 'fat' 'mat' 'on' 'rat' 'sat'
(1 row)
```

 版权所有 © 华为云计算技术有限公司

如果词条中包含空格或标点符号，可以用引号包围：

```
SELECT $$the lexeme '    ' contains spaces$$::tsvector;
                  tsvector
-------------------------------------------
 '    ' 'contains' 'lexeme' 'spaces' 'the'
(1 row)
```

使用常规的单引号引起来的字符串，字符串中嵌入的单引号(')和反斜杠(\)必须双写进行转义：

```
SELECT $$the lexeme 'Joe''s' contains a quote$$::tsvector;
                    tsvector
-------------------------------------------------
 'Joe''s' 'a' 'contains' 'lexeme' 'quote' 'the'
(1 row)
```

位置常量

词条**位置常量**也可以放到词汇中，表示该词条在原文中的位置：

```
SELECT 'a:1 fat:2 cat:3 sat:4 on:5 a:6 mat:7 and:8 ate:9 a:10 fat:11 rat:12'::tsvector;
                                  tsvector
-------------------------------------------------------------------------------
 'a':1,6,10 'and':8 'ate':9 'cat':3 'fat':2,11 'mat':7 'on':5 'rat':12 'sat':4
(1 row)
```

位置常量通常表示文档中源字的位置。位置信息可以用于进行排名。位置常量的范围是1到16383，最大值默认是16383。相同词的重复位会被忽略掉。

权重

拥有位置的词汇可以进一步地被标注一个权重，它可以是A，B，C或D。 D是默认权重，因此输出中不会显示：

```
SELECT 'a:1A fat:2B,4C cat:5D'::tsvector;
          tsvector
----------------------------
 'a':1A 'cat':5 'fat':2B,4C
(1 row)
```

表 5-19 权重参数说明

权重	说明	是否显示	典型用途
A	最高权重	显示（标记为A）	标题、重要关键词
B	中高权重	显示（标记为B）	摘要、小标题
C	中低权重	显示（标记为C）	正文重要段落
D	默认权重	不显示	正文普通内容

权重通常被用来反映文档结构，如：将标题标记成与正文词不同。文本检索排序函数可以为不同的权重标记分配不同的优先级。

规范化处理

tsvector类型标准用法示例如下：

 版权所有 © 华为云计算技术有限公司

```
SELECT 'The Fat Rats'::tsvector;
      tsvector
--------------------
 'Fat' 'Rats' 'The'
(1 row)
```

但是对于英文全文检索应用来说，上面的单词会被认为非规范化的，所以需要通过to_tsvector函数对这些单词进行规范化处理：

```
SELECT to_tsvector('english', 'The Fat Rats');
   to_tsvector
-----------------
 'fat':2 'rat':3
(1 row)
```

to_tsvector('english', ...)使用english分词器配置，将”The Fat Rats”分词并提取词干：停用词”the”被去除，”Fat”转换为小写”fat”，”Rats”提取词干为”rat”，并保留了位置常量2和3。

tsquery

tsquery类型表示一个检索条件，存储用于检索的词汇，并且使用布尔操作符将这些词汇进行组合。

布尔操作符

tsquery支持以下布尔操作符，优先级规则：如果没有圆括号，（NOT）的优先级最高，其次是&（AND），最后是|（OR）。使用圆括号可以改变操作符的分组优先级。

表 5-20 tsquery 操作符

操作符	含义	说明	优先级
&	AND（与）	两边的词条需同时匹配。	中
|	OR（或）	两边的词条至少一个匹配。	低
!	NOT（非）	后面的词条不匹配。	高

```
SELECT 'fat & rat'::tsquery;
    tsquery
----------------
 'fat' & 'rat'
(1 row)

SELECT 'fat & (rat | cat)'::tsquery;
          tsquery
---------------------------
 'fat' & ( 'rat' | 'cat' )
(1 row)

SELECT 'fat & rat & ! cat'::tsquery;
        tsquery
------------------------
 'fat' & 'rat' & !'cat'
(1 row)
```

 版权所有 © 华为云计算技术有限公司

权重标记

tsquery中的词汇可以用一个或多个权重字母来标记，这些权重字母限制词汇只能与匹配权重的tsvector词汇进行匹配。

```
SELECT 'fat:ab & cat'::tsquery;
     tsquery
------------------
 'fat':AB & 'cat'
(1 row)

--'fat:ab'表示该词条仅在tsvector中对应词条的权重为A或B时才匹配。
```

前缀匹配

tsquery中的词汇可以用*标记来指定前缀匹配。例如：这个查询可以匹配tsvector中以“super”开始的任意单词。

```
SELECT 'super:*'::tsquery;
  tsquery
-----------
 'super':*
(1 row)
```

需注意，前缀匹配会首先被文本搜索分词器处理。例如：postgres中提取的词干是postgr，匹配到了postgraduate，也就意味着下面的结果为真：

```
SELECT to_tsvector( 'postgraduate' ) @@ to_tsquery( 'postgres:*' ) AS RESULT;
  result
----------
 t
(1 row)
SELECT to_tsquery('postgres:*');
 to_tsquery
------------
 'postgr':*
(1 row)
```

@@操作符（全文检索匹配）

@@是全文检索匹配操作符，用于判断tsquery查询条件是否匹配tsvector文档分词。

语法：

```
tsvector @@ tsquery
```

返回类型： boolean（true表示匹配成功，false表示不匹配）

表 5-21 参数说明

操作数	类型	说明
左操作数	tsvector	文档的分词列表，通常通过to_tsvector生成。
右操作数	tsquery	查询条件，通常通过to_tsquery或plainto_tsquery生成。

匹配规则： 判断tsquery中的查询条件（通过AND/OR/NOT组合）是否能在tsvector中找到满足的词条。如果所有条件均满足，返回true，否则返回false。

```
-- 查询文档中是否同时包含"fat"和"rat"
SELECT to_tsvector('english', 'The Fat Rats') @@ to_tsquery('fat & rat') AS result;
```

 版权所有 © 华为云计算技术有限公司

```
-- 预期结果:
--  result
-- --------
--  t
-- (1 row)
-- 查询文档中是否包含"cat"（不包含，返回false）
SELECT to_tsvector('english', 'The Fat Rats') @@ to_tsquery('cat') AS result;
-- 预期结果:
--  result
-- --------
--  f
-- (1 row)
```

规范化处理

to_tsquery函数会将单词转换为tsquery类型前进行规范化处理。'Fat:ab & Cats'规范化转为tsquery类型结果如下：

```
--to_tsquery对输入进行了词干提取和标准化：Fat转换为小写fat，Cats提取词干为cat，权重标记ab转换为大写AB。
SELECT to_tsquery('Fat:ab & Cats');
    to_tsquery
------------------
 'fat':AB & 'cat'
(1 row)
```

5.13 UUID 类型

UUID：通用唯一识别码（Universally Unique Identifier）是用于计算机体系中以识别信息的一个128位标识符。

UUID的作用是让分布式系统中的所有元素都能有唯一的辨识信息，而不需要通过中央控制端来做辨识信息的指定。很多应用场景需要一个ID，仅用来标识一个对象。常见的例子有数据库表的ID字段。另一个例子是前端的各种UI库，因为它们通常需要动态创建各种UI元素，这些元素需要唯一的ID，这时就需要使用UUID。

DWS支持的UUID函数及UUID应用示例，可参考**UUID函数**。

UUID 格式

UUID由开放软件基金会标准化，作为分布式计算环境的一部分，在互联网工程任务组（IETF）公布的RFC 4122标准中对UUID进行了标准化。标准的UUID由36个字符组成，其中包括32个16进制数字和4个连字符‘-’，形式为8-4-4-4-12，标准的UUID示例如下：

```
a0eebc99-9c0b-4ef8-bb6d-6bb9bd380a11
```

除了标准型的UUID，DWS同样支持以其他方式输入：大写字母和数字、由花括号包围的标准格式、省略部分或所有连字符、在任意一组四位数字之后加一个连字符。示例：

```
A0EEBC99-9C0B-4EF8-BB6D-6BB9BD380A11
{a0eebc99-9c0b-4ef8-bb6d-6bb9bd380a11}
a0eebc999c0b4ef8bb6d6bb9bd380a11
a0ee-bc99-9c0b-4ef8-bb6d-6bb9-bd38-0a11
```

UUID 组成

UUID采用时间信息和空间信息确保全局唯一，通常由时间戳信息、时钟序列和节点ID（MAC地址）组成。但在DWS分布式集群中，多个节点可能部署在同一台机器上，其

 版权所有 © 华为云计算技术有限公司

MAC地址相同会导致生成的UUID可能出现冲突。因此DWS将最后48位的MAC地址替换为生成UUID的CN或DN的序号以及当前的线程ID，从而确保分布式集群内UUID的全局唯一性。

5.14 JSON 类型

JSON数据类型可以用来存储JSON（JavaScript Object Notation）数据。

可以是单独的一个标量，也可以是一个数组，也可以是一个键值对象，其中数组和对象可以统称容器(container)：

1. 标量(scalar)：单一的数字、bool、string、null都可以叫做标量。
2. 数组(array)：[]结构，里面存放的元素可以是任意类型的JSON，并且不要求数组内所有元素都是同一类型。
3. 对象(object)：{}结构，存储key:value的键值对，其键只能是用“”包裹起来的字符串，值可以是任意类型的JSON，对于重复的键，按最后一个键值对为准。

DWS支持使用：json数据类型和jsonb数据类型存储JSON数据。其中：

- json对输入的字符串进行完整复制，使用时再去解析，所以它会保留输入的空格，重复键以及顺序等。
- jsonb解析输入后保存的二进制，它在解析时会删除语义无关的细节和重复的键，对键值也会进行排序，使用时不用再次解析。

因此可以发现，两者其实都是JSON，它们接受相同的字符串作为输入。它们实际的主要差别是效率。json数据类型存储输入文本的精确复制，处理函数必须在每个执行上重新解析； 而jsonb数据以分解的二进制格式存储， 这使得它由于添加了转换机制而在输入上稍微慢些，但是在处理上明显更快， 因为不需要重新解析。同时由于jsonb类型存在解析后的格式归一化等操作，同等的语义下只会有一种格式，因此可以更好地支持很多其他额外的操作，比如按照一定的规则进行大小比较等。jsonb也支持索引，这也是一个明显的优势。

输入格式

json和jsonb输入必须是一个符合JSON数据格式的字符串，此字符串用单引号''声明。

null (null-json)：仅null，全小写。

```
SELECT 'null'::json;   -- suc
SELECT 'NULL'::jsonb;  -- err
```

数字 (num-json)：正负整数、小数、0，支持科学计数法。

```
SELECT '1'::json;
SELECT '-1.5'::json;
SELECT '-1.5e-5'::jsonb, '-1.5e+2'::jsonb;
SELECT '001'::json, '+15'::json, 'NaN'::json;  -- 不支持多余的前导0，正数的+号，以及NaN和infinity。
```

布尔(bool-json)：仅true、false，全小写。

```
SELECT 'true'::json;
SELECT 'false'::jsonb;
```

字符串(str-json)：必须是加双引号的字符串。

```
SELECT '"a"'::json;
SELECT '"abc"'::jsonb;
```

 版权所有 © 华为云计算技术有限公司

数组(array-json)：使用中括号[]包裹，满足数组书写条件。数组内元素类型可以是任意合法的JSON，且不要求类型一致。

```
SELECT '[1, 2, "foo", null]'::json;
SELECT '[]'::json;
SELECT '[1, 2, "foo", null, [[]], {}]'::jsonb;
```

对象(object-json)：使用大括号{}包裹，键必须是满足JSON字符串规则的字符串，值可以是任意合法的JSON。

```
SELECT '{}'::json;
SELECT '{"a": 1, "b": {"a": 2,  "b": null}}'::json;
SELECT '{"foo": [true, "bar"], "tags": {"a": 1, "b": null}}'::jsonb;
```

⚠ 注意

- 'null'::json和null::json是两个不同的概念，类似于字符串str=“ ”和str=null。
- 对于数字，当使用科学计数法的时候，jsonb类型会将其展开，而json会精准复制输入。

jsonb 高级特性

json和jsonb的主要差异在于存储方式上的不同，jsonb存储的是解析后的二进制，能够体现JSON的层次结构，更便于直接访问等，因此jsonb较json具有很多高级特性。

格式归一化

- 对于输入的object-json字符串，解析成jsonb二进制后，会天然的丢弃语义上无关紧要的细节，比如空格：

```
SELECT '   [1, " a ", {"a"   :1    }]  '::jsonb;
   jsonb
----------------------
[1, " a ", {"a": 1}]
(1 row)
```

- 对于object-json，会删除重复的键值，只保留最后一个出现的，例如：

```
SELECT '{"a" : 1, "a" : 2}'::jsonb;
jsonb
----------
{"a": 2}
(1 row)
```

- 对于object-json，键值会重新进行排序，排序规则：长度长的在后、长度相等则ascii码大的在后，例如：

```
SELECT '{"aa" : 1, "b" : 2, "a" : 3}'::jsonb;
      jsonb
---------------------------
{"a": 3, "b": 2, "aa": 1}
(1 row)
```

大小比较

由于经过了格式归一化，保证了同一种语义下的jsonb只会有一种存在形式，因此按照指定的规则，可以比较大小。

1. 首先比较类型：object-jsonb > array-jsonb > bool-jsonb > num-jsonb > str-jsonb > null-jsonb
2. 同类型则比较内容：
   - str-json类型：依据text比较的方法，使用数据库默认排序规则进行比较，返回值正数代表大于，负数代表小于，0表示相等。

版权所有 © 华为云计算技术有限公司

- num-json类型：数值比较
- bool-json类型：true > false
- array-jsonb类型：长度长的 > 长度短的，长度相等则依次比较每个元素。
- object-jsonb类型：长度长的 > 长度短的，长度相等则依次比较每个键值对，先比较键，再比较值。

⚠ 注意

object-jsonb类型内比较时使用的是格式整理后的最终结果进行比较，因此相对于直接的输入未必会很直观。

创建索引

行存jsonb类型支持创建B-Tree索引和GIN索引，列存jsonb类型仅支持创建cbtree索引。

如果整个JSONB列使用B-Tree索引，则能使用的运算符是=、<、<=、>、>=。

示例：创建表test，并插入数据。

```
CREATE TABLE test(id bigserial, data JSONB, PRIMARY KEY (id));
INSERT INTO test(data) VALUES('{"name":"Jack", "age":10, "nick_name":["Jacky","baobao"], "phone_list":["1111","2222"]}'::jsonb);
```

- 创建B-Tree索引。
  ```
  CREATE INDEX idx_test_data_age ON test USING btree(((data->>'age')::int));
  ```
  使用B-Tree索引查询age>1。
  ```
  SELECT * FROM test WHERE (data->>'age')::int>1;
  ```
- 创建GIN索引。
  ```
  CREATE INDEX idx_test_data ON test USING gin (data);
  ```
  使用GIN索引查询顶层关键词是否存在。
  ```
  SELECT * FROM test WHERE data ? 'id';
  SELECT * FROM test WHERE data ?| array['id','name'];
  ```
- 使用GIN索引查询非顶层关键词是否存在。
  ```
  CREATE INDEX idx_test_data_nick_name ON test USING gin((data->'nick_name'));
  SELECT * FROM test WHERE data->'nick_name' ? 'Jacky';
  ```
- 使用@>查询JSON中是否包含子json对象。
  ```
  SELECT * FROM test WHERE data @> '{"age":10, "nick_name":["Jacky"]}';
  ```

包含存在

查询一个JSON之中是否包含某些元素，或者某些元素是否存在于某个JSON中是jsonb的一个重要能力。

- 简单的标量/原始值只包含相同的值。
  ```
  SELECT '"foo"'::jsonb @> '"foo"'::jsonb;
  ```
- 左侧数组包含了右侧字符串。
  ```
  SELECT '[1, "aa", 3]'::jsonb ? 'aa';
  ```
- 左侧数组包含了右侧的数组所有元素，顺序、重复不重要。
  ```
  SELECT '[1, 2, 3]'::jsonb @> '[1, 3, 1]'::jsonb;
  ```
- 左侧object-json包含了右侧object-json的所有键值。
  ```
  SELECT '{"product": "PostgreSQL", "version": 9.4, "jsonb":true}'::jsonb @> '{"version":9.4}'::jsonb;
  ```
- 左侧数组并没有包含右侧的数组所有元素，因为左侧数组的三个元素为1、2、[1,3]，右侧的为1、3。

版权所有 © 华为云计算技术有限公司

```
SELECT '[1, 2, [1, 3]]'::jsonb @> '[1, 3]'::jsonb; -- false
```

- 相似的这种也没有包含。
```
SELECT '{"foo": {"bar": "baz"}}'::jsonb @> '{"bar": "baz"}'::jsonb; -- false
```

5.15 RoaringBitmap 类型

DWS自8.1.3集群版本开始，支持RoaringBitmap数据类型，用于存储位图数据集。

roaringbitmap数据类型支持行存，列存表。

表 5-22 RoaringBitmap 类型

名字	存储容量	描述	范围
RoaringBitmap	64字节	存储位图数据集	-9,223,372,036,854,775,808 ~ 9,223,372,036,854,775,807

示例：创建带有roaringbitmap数据类型的表。

```
CREATE TABLE r_row (a int ,b text, c roaringbitmap);
CREATE TABLE r_col (a int ,b text, c roaringbitmap) with (orientation=column);
```

5.16 HLL 数据类型

HLL（HyperLoglog）是统计数据集中唯一值个数的高效近似算法。它有着计算速度快，节省空间的特点，不需要直接存储集合本身，而是存储一种名为HLL的数据结构。每当有新数据加入进行统计时，只需要把数据经过哈希计算并插入到HLL中，最后根据HLL就可以得到结果。

HLL与其他算法的比较请参见**表5-23**。

表 5-23 HLL 与其他算法比较

项目	Sort算法	Hash算法	HLL
时间复杂度	O(nlogn)	O(n)	O(n)
空间复杂度	O(n)	O(n)	1280 bytes
误差率	0	0	≈2%
所需存储空间	原始数据大小	原始数据大小	1280 bytes

HLL在计算速度和所占存储空间上都占优势。在时间复杂度上，Sort算法需要排序至少O(nlogn)的时间，虽说Hash算法和HLL一样扫描一次全表O(n)的时间就可以得出结果，但是存储空间上，Sort算法和Hash算法都需要先把原始数据存起来再进行统计，会导致存储空间消耗巨大，而对HLL来说不需要存原始数据，只需要维护HLL数据结构，故占用空间始终是1280bytes常数级别。

 版权所有 © 华为云计算技术有限公司

说明

- 当前默认规格下可计算最大distinct值的数量为1.6e+12个，误差率最大仅2.3%。用户应注意如果计算结果超过当前规格下distinct最大值会导致计算结果误差率变大，或导致计算结果失败并报错。
- 用户在首次使用该特性时，应该对业务的distinct value做评估，选取适当的配置参数并做验证，以确保精度符合要求：
  - 当前默认参数下，可以计算的distinct value值为1.6e+12，如果计算得到的distinct value值为NaN，需要调整log2m和regwidth来容纳更多的distinct value。
  - 虽然hash算法存在极低的hash collision概率，但是建议用户在首次使用时，选取2-3个hash seed验证，如果得到的distinct value相差不大，则可以从该组seed中任选一个作为hash seed。

HLL中主要的数据结构，请参见表5-24。

表 5-24 HyperLogLog 中主要数据结构

数据类型	功能描述
hll	大小为确定的1280 bytes，可直接计算得到distinct值。

HLL 的应用场景

- 使用hll数据类型场景
  a. 创建带有hll类型的表并向表中插入空的hll。
  ```
  CREATE TABLE helloworld (id integer, col hll);
  INSERT INTO helloworld(id, col) VALUES (1, hll_empty());
  ```
  b. 把整数经过哈希计算加入到hll中。
  ```
  UPDATE helloworld SET col = hll_add(col, hll_hash_integer(12345)) WHERE id = 1;
  ```
  c. 把字符串经过哈希计算加入到hll中。
  ```
  UPDATE helloworld SET col = hll_add(col, hll_hash_text('hello world')) WHERE id = 1;
  ```
  d. 得到hll中的distinct值。
  ```
  SELECT hll_cardinality(col) FROM helloworld WHERE id = 1;
   hll_cardinality
  -----------------
                2
  (1 row)
  ```
- 使用hll进行网站访客统计场景
  a. 创建原始数据表facts，记录用户访问网站时间。
  ```
  CREATE TABLE facts (
          date            date,
          user_id         integer
  );
  ```
  b. 插入用户访问过网站的数据。
  ```
  INSERT INTO facts VALUES ('2019-02-20', generate_series(1,100));
  INSERT INTO facts VALUES ('2019-02-21', generate_series(1,200));
  INSERT INTO facts VALUES ('2019-02-22', generate_series(1,300));
  INSERT INTO facts VALUES ('2019-02-23', generate_series(1,400));
  INSERT INTO facts VALUES ('2019-02-24', generate_series(1,500));
  INSERT INTO facts VALUES ('2019-02-25', generate_series(1,600));
  INSERT INTO facts VALUES ('2019-02-26', generate_series(1,700));
  INSERT INTO facts VALUES ('2019-02-27', generate_series(1,800));
  ```
  c. 创建表并指定列为hll。根据日期把数据分组，并把数据插入到hll中。
  ```
  CREATE TABLE daily_uniques (
     date            date UNIQUE,
  ```

 版权所有 © 华为云计算技术有限公司

```
    users           hll
);

INSERT INTO daily_uniques(date, users)
    SELECT date, hll_add_agg(hll_hash_integer(user_id))
    FROM facts
    GROUP BY 1;
```

d. 计算每一天访问网站不同用户数量。

```
SELECT date, hll_cardinality(users) FROM daily_uniques ORDER BY date;
        date         | hll_cardinality
---------------------+------------------
 2019-02-20 00:00:00 |              100
 2019-02-21 00:00:00 | 203.813355588808
 2019-02-22 00:00:00 | 308.048239950384
 2019-02-23 00:00:00 | 410.529188080374
 2019-02-24 00:00:00 | 513.263875705319
 2019-02-25 00:00:00 | 609.271181107416
 2019-02-26 00:00:00 | 702.941844662509
 2019-02-27 00:00:00 | 792.249946595237
(8 rows)
```

e. 计算在2019.02.20到2019.02.26一周中有多少不同用户访问过网站。

```
SELECT hll_cardinality(hll_union_agg(users)) FROM daily_uniques WHERE date >=
'2019-02-20'::date AND date <= '2019-02-26'::date;
 hll_cardinality
------------------
 702.941844662509
(1 row)
```

f. 计算昨天访问过网站而今天没访问网站的用户数量。

```
SELECT date, (#hll_union_agg(users) OVER two_days) - #users AS lost_uniques FROM
daily_uniques WINDOW two_days AS (ORDER BY date ASC ROWS 1
PRECEDING);
        date         | lost_uniques
---------------------+--------------
 2019-02-20 00:00:00 |            0
 2019-02-21 00:00:00 |            0
 2019-02-22 00:00:00 |            0
 2019-02-23 00:00:00 |            0
 2019-02-24 00:00:00 |            0
 2019-02-25 00:00:00 |            0
 2019-02-26 00:00:00 |            0
 2019-02-27 00:00:00 |            0
(8 rows)
```

- 插入数据不满足hll数据结构要求时报错场景

  当用户给hll类型的字段插入数据的时候，必须保证插入的数据满足hll数据结构要求，如果解析后不满足就会报错。

  例如： 插入数据'E\\1234'时，该数据不满足hll数据结构，不能解析成功因此失败报错。

```
CREATE TABLE test(id integer, col hll);
INSERT INTO test VALUES(1, 'E\\1234');
ERROR:  invalid input syntax for integer: "E\\1234"
```

5.17 对象标识符类型

DWS在内部使用对象标识符（OID）作为各种系统表的主键。系统不会给用户创建的表增加一个OID系统字段，OID类型代表一个对象标识符。

目前OID类型用一个四字节的无符号整数实现。因此不建议在创建的表中使用OID字段做主键。

 版权所有 © 华为云计算技术有限公司

表 5-25 对象标识符类型

名称	引用	描述	示例
OID	-	数字化的对象标识符。	564182
CID	-	命令标识符。它是系统字段cmin和cmax的数据类型。命令标识符是32位的量。	-
XID	-	事务标识符。它是系统字段xmin和xmax的数据类型。事务标识符也是32位的量。	-
TID	-	行标识符。它是系统表字段ctid的数据类型。行ID是一对数值（块号，块内的行索引），它标识该行在其所在表内的物理位置。	-
REGCONFIG	pg_ts_config	文本搜索配置。	english
REGDICTIONARY	pg_ts_dict	文本搜索字典。	simple
REGOPER	pg_operator	操作符名。	+
REGOPERATOR	pg_operator	带参数类型的操作符。	*(integer,integer)或-(NONE,integer)
REGPROC	pg_proc	函数名称。	sum
REGPROCEDURE	pg_proc	带参数类型的函数。	sum(int4)
REGCLASS	pg_class	关系名。	pg_type
REGTYPE	pg_type	数据类型名。	integer

OID类型：主要作为数据库系统表中字段使用。

示例：

```
SELECT oid FROM pg_class WHERE relname = 'pg_type';
 oid
------
 1247
(1 row)
```

OID别名类型REGCLASS：主要用于对象OID值的简化查找。

示例：

```
SELECT attrelid,attname,atttypid,attstattarget FROM pg_attribute WHERE attrelid = 'pg_type'::REGCLASS;
 attrelid |    attname     | atttypid | attstattarget
----------+----------------+----------+---------------
     1247 | xc_node_id     |       23 |             0
```

 版权所有 © 华为云计算技术有限公司

```
  1247 | tableoid       |       26 |           0
  1247 | cmax           |       29 |           0
  1247 | xmax           |       28 |           0
  1247 | cmin           |       29 |           0
  1247 | xmin           |       28 |           0
  1247 | oid            |       26 |           0
  1247 | ctid           |       27 |           0
  1247 | typname        |       19 |          -1
  1247 | typnamespace   |       26 |          -1
  1247 | typowner       |       26 |          -1
  1247 | typlen         |       21 |          -1
  1247 | typbyval       |       16 |          -1
  1247 | typtype        |       18 |          -1
  1247 | typcategory    |       18 |          -1
  1247 | typispreferred |       16 |          -1
  1247 | typisdefined   |       16 |          -1
  1247 | typdelim       |       18 |          -1
  1247 | typrelid       |       26 |          -1
  1247 | typelem        |       26 |          -1
  1247 | typarray       |       26 |          -1
  1247 | typinput       |       24 |          -1
  1247 | typoutput      |       24 |          -1
  1247 | typreceive     |       24 |          -1
  1247 | typsend        |       24 |          -1
  1247 | typmodin       |       24 |          -1
  1247 | typmodout      |       24 |          -1
  1247 | typanalyze     |       24 |          -1
  1247 | typalign       |       18 |          -1
  1247 | typstorage     |       18 |          -1
  1247 | typnotnull     |       16 |          -1
  1247 | typbasetype    |       26 |          -1
  1247 | typtypmod      |       23 |          -1
  1247 | typndims       |       23 |          -1
  1247 | typcollation   |       26 |          -1
  1247 | typdefaultbin  |      194 |          -1
  1247 | typdefault     |       25 |          -1
  1247 | typacl         |     1034 |          -1
(38 rows)
```

5.18 伪类型

DWS数据类型中包含一系列特殊用途的类型，这些类型按照类别被称为伪类型。伪类型不能作为字段的数据类型，但是可以用于声明函数的参数或者结果类型。

当一个函数不仅仅是简单地接受并返回某种SQL数据类型，伪类型能起到很大的作用。表5-26列出了所有的伪类型。

表 5-26 伪类型

名字	描述
any	表示函数接受任何输入数据类型。
anyelement	表示函数接受任何数据类型。
anyarray	表示函数接受任意数组数据类型。
anynonarray	表示函数接受任意非数组数据类型。
anyenum	表示函数接受任意枚举数据类型。
anyrange	表示函数接受任意范围数据类型。

版权所有 © 华为云计算技术有限公司

名字	描述
cstring	表示函数接受或者返回一个空结尾的C字符串。
internal	表示函数接受或者返回一种服务器内部的数据类型。
language_handler	声明一个过程语言调用句柄返回language_handler。
fdw_handler	声明一个外部数据封装器返回fdw_handler。
record	标识函数返回一个未声明的行类型。
trigger	声明一个触发器函数返回trigger。
void	表示函数不返回数值。
opaque	一个已经过时的类型，以前用于所有上面这些用途。

声明用C编写的函数（不管是内置的还是动态装载的）都可以接受或者返回任何这样的伪数据类型。当伪类型作为参数类型使用时，用户需要保证函数的正常运行。

用过程语言编写的函数只能使用实现语言允许的伪类型。目前，过程语言不允许使用伪类型作为参数类型的，只允许使用void和record作为结果类型。一些多态的函数还支持使用anyelement，anyarray，anynonarray anyenum和anyrange类型。

伪类型internal用于声明只能在数据库系统内部调用的函数，这些函数不能直接在SQL查询里调用。如果某函数至少有一个internal类型的参数，则不能从SQL里调用。建议不要创建任何声明返回internal的函数，除非该函数至少有一个internal类型的参数。

示例：

创建或替换函数showall()。

```
CREATE OR REPLACE FUNCTION showall() RETURNS SETOF record
AS $$ SELECT count(*) from tpcds.store_sales where ss_customer_sk = 9692; $$
LANGUAGE SQL;
```

调用函数showall()。

```
SELECT showall();
 showall
---------
 (35)
(1 row)
```

删除函数。

```
DROP FUNCTION showall();
```

5.19 范围类型

范围类型是表示某些元素类型（称为范围的子类型）的值范围的数据类型。例如，时间戳范围可用于表示保留会议室的时间范围。在这种情况下，数据类型为tsrange（“时间戳范围”的缩写），时间戳是子类型。子类型必须具有总体的顺序，以便很好地定义元素值是在值的范围内、之前还是之后。

范围类型可以表达单一范围值中的多个元素值，并且可以很清晰地表达诸如范围重叠等概念。用于计划安排的时间和日期范围，也可以用于价格范围或仪器的测量范围等。

 版权所有 © 华为云计算技术有限公司

内置范围类型

DWS带有下列内置范围类型：

- int4range — integer的范围
- int8range — bigint的范围
- numrange — numeric的范围
- tsrange — 不带时区的timestamp的范围
- tstzrange — 带时区的timestamp的范围
- daterange — date的范围

说明

DWS暂不支持用户自定义范围类型。

创建表reservation并插入数据，其中during字段类型为tsrange：

```
CREATE TABLE reservation (room int, during tsrange);
INSERT INTO reservation VALUES (1108, '[2010-01-01 14:30, 2010-01-01 15:30)');
```

查询范围是否包含：

```
SELECT int4range(10, 20) @> 3;
 ?column?
----------
 f
(1 row)
```

查询范围是否重叠：

```
SELECT numrange(11.1, 22.2) && numrange(20.0, 30.0);
 ?column?
----------
 t
(1 row)
```

抽取范围的上界：

```
SELECT upper(int8range(15, 25));
 upper
-------
    25
(1 row)
```

计算范围的交集：

```
SELECT int4range(10, 20) * int4range(15, 25);
 ?column?
----------
 [15,20)
(1 row)
```

查询范围是否为空：

```
SELECT isempty(numrange(1, 5));
 isempty
---------
 f
(1 row)
```


版权所有 © 华为云计算技术有限公司

包含和排除边界

每一个非空范围都有两个界限：下界和上界。这些值之间的所有点都被包括在范围内。包含界限意味着边界点本身也被包括在范围内，而排除边界意味着边界点不被包括在范围内。

在范围的文本形式中，包含下界用“[”表示，排除下界用“(”表示。同样，包含上界用“]”表示，排除上界用“)”表示。

函数**lower_inc(anyrange)**和**upper_inc(anyrange)**分别测试一个范围值的上下界。

无限（无界）范围

范围的下界可以省略，这意味着所有小于上界的值都包括在范围中，例如(,3]。同样，范围的上界被省略，则所有大于下界的值都包括在范围中。如果下界和上界都被省略，则该元素类型的所有值都被认为在范围内。如果缺失的边界指定为包含则自动将包含转换为排除，例如[,]将转换为(,)。可以将这些缺失的值视为+/-无穷大，但它们是特殊的范围类型值，并且被视为超出任何范围元素类型的+/-无穷大值。

具有“无穷大”概念的元素类型可以将其作为显式边界值。例如，在时间戳范围，[today,infinity)不包括特殊的timestamp值infinity，尽管[today,infinity]包括它，就好比 [today,)和[today,]。

函数**lower_inf(anyrange)**和**upper_inf(anyrange)**分别测试一个范围的无限上下界。

范围的输入/输出

范围值的输入必须遵循下列模式之一：

```
(lower-bound,upper-bound)
(lower-bound,upper-bound]
[lower-bound,upper-bound)
[lower-bound,upper-bound]
empty
```

圆括号或方括号指示上下界是否为排除的或者包含的。最后一个模式是empty，它表示一个空范围（一个不包含点的范围）。

lower-bound可以作为子类型的合法输入的一个字符串，或者是空，表示没有下界。同样，upper-bound可以是作为子类型的合法输入的一个字符串，或者是空，表示没有上界。

每个界限值可以使用双引号引用。如果界限值包含圆括号、方括号、逗号、双引号或反斜线时，则必须使用双引号引用，否则这些符号会被认作范围语法的一部分。要想把双引号或反斜线放在被引用的界限值中，需要在双引号或反斜线前面加一个反斜线（在双引号引用的界限值中的一对双引号表示一个双引号字符，这与SQL字符串中的单引号规则类似）。此外，可以避免引用并且使用反斜线转义来保护所有数据字符，否则它们会被当做返回语法的一部分。什么都不写则表示一个无限界限，因此，要表示空字符串的界限值，可以写成""。

范围值前后允许有空格，但是圆括号或方括号之间的任何空格会被当做上下界值的一部分（取决于元素类型，它可能是有意义的也可能是无意义的）。

示例

查询包括3，不包括7，并且包括3和7之间的所有点：

```
SELECT '[3,7)'::int4range;
 int4range
```


版权所有 © 华为云计算技术有限公司

```
-----------
 [3,7)
(1 row)
```

查询既不包括3也不包括 7，但是包括之间的所有点：

```
SELECT '(3,7)'::int4range;
 int4range
-----------
 [4,7)
(1 row)
```

查询只包括单独一个点4：

```
SELECT '[4,4]'::int4range;
 int4range
-----------
 [4,5)
(1 row)
```

查询不包括点（并且将被标准化为 '空'）：

```
SELECT '[4,4)'::int4range;
 int4range
-----------
 empty
(1 row)
```

构造范围

每一种范围类型都有一个与其同名的构造器函数。使用构造器函数比写一个范围文字常数更方便，因为它避免了对界限值的额外引用。构造器函数接受两个或三个参数。两个参数的形式以标准的形式构造一个范围（包含下界，排除上界），而三个参数的形式按照第三个参数指定的界限形式构造一个范围。第三个参数必须是下列字符串之一：“()”、“(]”、“[)”或者“[]”。例如：

完整形式是：下界、上界以及指示界限包含性/排除性的文本参数：

```
SELECT numrange(1.0, 14.0, '(]');
  numrange
------------
 (1.0,14.0]
(1 row)
```

如果第三个参数被忽略，则假定为 '[)'：

```
SELECT numrange(1.0, 14.0);
  numrange
------------
 [1.0,14.0)
(1 row)
```

尽管这里指定了'(]'，但返回结果时该值将被转换成标准形式，因为int8range是一种**离散范围类型**：

```
SELECT int8range(1, 14, '(]');
 int8range
-----------
 [2,15)
(1 row)
```

界限使用NULL导致范围是无界的：

```
SELECT numrange(NULL, 2.2);
 numrange
----------
 (,2.2)
(1 row)
```


版权所有 © 华为云计算技术有限公司

离散范围类型

离散范围是指其元素类型具有定义明确的“步长”的范围，如integer或date。在这些类型中，当两个元素之间没有有效值时，它们可以被说成是相邻。

离散范围类型的每个元素值都有一个明确的“下一个”或“上一个”值。这样就可以通过选择下一个或上一个元素值，在范围界限的包含和排除表达之间转换。例如，在整数范围类型中，[4,8]和(3,9)表示相同的值集合，但对于超过numeric的范围，情况并非如此。

离散范围类型应具有识别元素类型所需步长的规范函数。规范化函数负责将范围类型的等价值转换为具有相同的表示，特别是与包含或者排除界限一致。如果未指定规范化函数，则具有不同格式的范围将始终被视为不相等，即使它们实际上是表达相同的一组值。

内置范围类型int4range、int8range和daterange都使用规范形式，该形式包括下界并且排除上界，也就是[)。但是，用户定义的范围类型可以使用其他约定。

自定义范围类型

用户可以定义范围类型。例如，要创建一个subtype float8的范围类型：

```
CREATE TYPE floatrange AS RANGE (
        subtype = float8,
        subtype_diff = float8mi
    );

SELECT '[1.234, 5.678]'::floatrange;
```

索引

可以为范围类型的表列创建GiST索引。例如：

```
CREATE TABLE reservation (room int, during tsrange);
CREATE INDEX reservation_idx ON reservation USING gist (during);
```

GiST索引可以加速涉及以下范围操作符的查询： =、 &&、 <@、 @>、 <<、 >>、 -|-、 &<以及 &> 。详见**范围操作符**。

此外，也可以在范围类型的表列上创建B-tree索引。对于这些索引类型，有用的范围操作就是等值。范围类型的B-tree支持主要是为了允许在查询内部进行排序，而不是创建真正的索引。

5.20 复合类型

复合类型表示行或记录的结构，它本质上就是字段名及其数据类型的列表。DWS允许支持将表的列声明为复合类型。复合类型本质上和表的行类型相同，但是如果只想定义一种类型，使用**CREATE TYPE**可避免创建一个实际的表。单独的复合类型也是很有用的，例如可以作为函数的参数或者返回类型。

复合类型的声明

DWS支持用户使用**CREATE TYPE**定义复合类型：

```
CREATE TYPE complex AS (
        r       double precision,
        i       double precision
    );
```


版权所有 © 华为云计算技术有限公司

```
CREATE TYPE inventory_item AS (
    name            text,
    supplier_id     integer,
    price           numeric
);
```

定义复合类型之后，可用来创建表或函数：

```
CREATE TABLE on_hand (
    item      inventory_item,
    count     integer
);

INSERT INTO on_hand VALUES (ROW('fuzzy dice', 42, 1.99), 1000);
CREATE FUNCTION price_extension(inventory_item, integer) RETURNS numeric
    AS 'SELECT $1.price * $2' LANGUAGE SQL;

SELECT price_extension(item, 10) FROM on_hand;
```

构造复合值

要把复合值写作文字常量，可以将字段值括在圆括号中，并用逗号分隔。可以在任何字段值加上双引号，如果字段值包含逗号或括号则必须这样做。复合常量的一般格式如下：

```
'( val1 , val2 , ... )'
```

上文中的'("fuzzy dice",42,1.99)'便属于inventory_item类型的一个合法值。

要让一个字段为NULL，在列表中对应位置上空出即可。如果需要一个字段为空字符串，使用引号即可。例如下列示例，第一个字段是非NULL空字符串，第三个是NULL：

```
'("",42,)'
```

ROW表达式也能被用来构建组合值。例如：

```
ROW('fuzzy dice', 42, 1.99)
ROW('', 42, NULL)
```

访问组合类型

要访问复合列的字段，可以写成一个点和字段的名称，就像从表名中选择字段一样。为了避免混淆，必须使用括号进行区分。例如，尝试从示例表on_hand中选择一些子域：

```
SELECT item.name FROM on_hand WHERE item.price > 9.99;
ERROR:  missing FROM-clause entry for table "item"
```

这样写是会与从表中选择字段混淆的，因为名称item会被当成是一个表名，而不是on_hand的一个列名。必须写成这样：

```
SELECT (item).name FROM on_hand WHERE (item).price > 9.99;
```

或者还需要使用表名（例如在一个多表查询中）：

```
SELECT (on_hand.item).name FROM on_hand WHERE (on_hand.item).price > 9.99;
```

现在加上括号的对象就被正确地解释为对item列的引用，然后可以从中选出子域。

只要从一个组合值中选择一个字段，相似的语法问题就适用。例如，要从返回组合值的函数的结果中选取一个字段，需要这样写：

```
SELECT (my_func(...)).field FROM ...
```


版权所有 © 华为云计算技术有限公司

如果没有额外的圆括号，这将生成一个语法错误。

5.21 XML 类型

XML数据类型可以被用来存储XML（eXtensible Markup Language）数据。XML数据也可以存储为text类型，但是XML数据类型的优势在于会检查每个存储的值是不是结构良好的XML值。XML可以存储由XML标准定义的格式良好的文档，以及由XML标准中定义的“内容”片段，内容片段中可以有多个顶级元素或字符节点。

XML类型相关的支持函数请参见**XML函数**。

设置 XML 参数

语法格式如下：

```
SET XML OPTION { DOCUMENT | CONTENT };
SET xmloption TO { DOCUMENT | CONTENT };
```

当一个字符串值在没有通过XMLPARSE或XMLSERIALIZE函数与XML类型进行转换时，由XML OPTION会话配置参数来决定选择DOCUMENT还是CONTENT。

默认为CONTENT，表示所有形式的xml数据都被允许。

示例：

```
SET XML OPTION DOCUMENT;
SET
SET xmloption TO DOCUMENT;
SET
```

设置二进制数据的编码格式

语法格式：

```
SET xmlbinary TO { base64 | hex};
```

示例：

```
SET xmlbinary TO base64;
SET

SELECT xmlelement(name foo, bytea 'bar');
xmlelement
-----------------
<foo>YmFy</foo>
(1 row)

SET xmlbinary TO hex;
SET

SELECT xmlelement(name foo, bytea 'bar');
xmlelement
-------------------
<foo>626172</foo>
(1 row)
```

访问 XML 值

XML数据类型比较特殊，它不提供任何比较操作符，这是因为对于XML数据没有通用的比较算法，所以无法通过比较一个XML值和一个搜索值来检索数据行。XML数据通常应该伴随一个ID值用于检索数据。另一种比较XML值的方案是将XML值转换成字符串，但字符串的比较并不能解决常见的XML值比较场景。

版权所有 © 华为云计算技术有限公司

5.22 列存表支持的数据类型

列存表支持的普通数据类型如表5-27所示。

表 5-27 列存表支持的数据类型

类别	数据类型	描述	长度
数值类型	smallint	小范围整数，别名为INT2。	2
	integer	常用的整数，别名为INT4。	4
	bigint	大范围的整数，别名为INT8。	8
	decimal	任意精度型。	可变长度
	numeric	任意精度型。	可变长度
	real	单精度浮点数。	4
	double precision	双精度浮点数。	8
	smallserial	二字节序列整型。	2
	serial	四字节序列整型。	4
	bigserial	八字节序列整型。	8
货币类型	money	货币金额。	8
字符类型	character varying(n), varchar(n)	变长字符串。	可变长度
	character(n), char(n)	定长字符串。	n
	character、char	单字节内部类型。	1
	text	变长字符串。	可变长度
	nvarchar2	变长字符串。	可变长度
	clob	文本大对象。	可变长度
日期/时间类型	timestamp with time zone	日期和时间，带时区。	8
	timestamp without time zone	日期和时间。	8

版权所有 © 华为云计算技术有限公司

类别	数据类型	描述	长度
	date	Oracle兼容模式下记录日期和时间；其他兼容模式下，记录日期。	Oracle兼容模式下，占存储空间8字节；其他兼容模式下，占存储空间4字节。
	time without time zone	只用于一日内时间。	8
	time with time zone	只用于一日内时间，带时区。	12
	interval	时间间隔。	16
JSON类型	json	-	-
	jsonb	-	-
XML类型	xml	-	-
RoaringBitmap	-	高效位图，支持int类型数据集位图计算。	可变长度

数组类型支持情况与GUC参数enable_vectorization_array相关，具体如下：

- 当GUC参数enable_vectorization_array（该参数仅9.1.1.100及以上集群版本支持）打开时，列存表可支持的数组类型包括bool, tiny int, small int, int, bigint, float4, float8, numeric,text, char, varchar,nvarchar2,bpchar, tinterval, smalldatetime,date,time, timestamp, timestamptz, interval, timetz。
- enable_vectorization_array关闭时，不支持列存表创建数组类型。
- 其他未列举的数据类型，暂不支持。

 版权所有 © 华为云计算技术有限公司

6 常量与宏

DWS支持的常量和宏请参见表6-1。

表 6-1 常量和宏

参数	描述	相关SQL语句示例
CURRENT_CATALOG	当前数据库	SELECT CURRENT_CATALOG;
CURRENT_ROLE	当前角色	SELECT CURRENT_ROLE;
CURRENT_SCHEMA	当前数据库模式	SELECT CURRENT_SCHEMA;
CURRENT_USER	当前用户	SELECT CURRENT_USER;
LOCALTIMESTAMP	当前会话时间（无时区）	SELECT LOCALTIMESTAMP;
NULL	空值	-
SESSION_USER	当前系统用户	SELECT SESSION_USER;
SYSDATE	当前系统日期	SELECT SYSDATE; 或 SELECT now()::DATE;
USER	当前用户，与CURRENT_USER一致。	SELECT USER;

版权所有 © 华为云计算技术有限公司

7 函数和操作符

7.1 字符串处理函数和操作符

7.1.1 字符串处理函数和操作符总体介绍

DWS提供的字符处理函数和操作符主要用于字符串处理，例如获取字符串长度、数据截取、拼接、转化、比较等操作。

在了解字符串处理函数前，您可以先简单了解**位**、**字节**、**字符**的相关概念：

- **位（bit）：** 计算机中最小的信息单位，表示二进制中的0或1，只有两种状态（对应电路的通 / 断、磁介质的正 / 反向磁化）。
- **字节（byte）：** 1字节=8位，是计算机存储和传输数据的基本单位。
- **字符：** 人类可读的文本符号，例如字母、数字、标点符号或汉字，而**数据库字符编码**决定了字符如何转换为字节。数据库字符编码在创建数据库语法中定义，详情请参见**CREATE DATABASE**。

例如在DWS的**UTF-8编码**中：一个英文字符占1个字节，而一个中文字符占3个字节，若字段定义为VARCHAR(100)，理论上最大占用100字符 * 3字节/字符 = 300字节。而在**GBK编码**中，一个中文字符占2个字节。

字符串处理函数按使用场景不同可分为以下几类：

- **获取字符串长度**：查询某个字符串的位数、字节长度等，例如查询“World”的位数。
- **截取字符串**：从一长串的字符串中截取某个子字符串，例如从“HelloWorld”中截取“Hello”。
- **拼接字符串**：将两串单独的字符串进行拼接，例如拼接“DATA”和“BASE”，得到“DATABASE”。
- **替换和填充字符串**：将字符串中的某个字符或子字符串替换成别的字符串，或者将多余空格删除，例如将“Database”后面的空格清空。
- **删除指定字符**：从字符串中删除掉指定的字符，例如删除“DataxBasexx”后面多余的“xx”。
- **分割字符串**：将字符串按需要分割成几块，并用分隔符分割，例如将“hello world”分成“hello， world”。

版权所有 © 华为云计算技术有限公司

- **转换字符串**：将字符串以某种规则进行转换，例如字母大小写转换。
- **比较字符串**：将两个字符串进行比较。
- **获取某一个字符或子字符串的位置**：例如，获取子字符串“bcd”在主字符串“abcdabcdabcd”中，从第2个字符开始，第2次匹配的位置。
- **判断某个字符串是否符合某种匹配模式(字符串相关的模式匹配函数)**：例如，判断“a”是否在“[ac]”的正则表达式里面。

7.1.2 获取字符串长度

DWS支持以下函数计算字符串的长度，包括获取**位数**、**字节数**和**字符数**。关于位数、字节数和字符数的区别可参见**字符串处理函数和操作符总体介绍**。

表 7-1 获取字符串长度常用函数

分类	功能	对应函数	简单示例	使用区别
获取位数	获取字符串的位数。	**bit_length(string)**	SELECT bit_length('world'); bit_length ------------ 40 (1 row)	-
获取字节数	获取所有字符串类型的字节数。	**octet_length(string)**	SELECT octet_length('world'); octet_length ------------ 5 (1 row)	octet_length(string)与lengthb(string)在功能上完全一致，在处理MySQL/DWS等跨平台数据库开发时，**优先使用octet_length(string)**，**lengthb(string)**是为了兼容Oracle而设定的一个映射函数，**仅在Oracle迁移业务使用**。 lengthb(text/bpchar)与lengthb(string)一致，只是特指入参为text或bpchar类型的情况。
	获取所有字符串类型的字节数，仅为兼容Oracle。	**lengthb(string)**	SELECT lengthb('abc'::text); lengthb ------------ 3 (1 row)	
	获取text、bpchar类型的字节数，仅为兼容Oracle。	**lengthb(text/bpchar)**	SELECT lengthb('abc'::CHAR(5)); lengthb ------------ 5 (1 row)	
获取字符数	获取字符串的字符数。	**char_length(string)或character_length(string)** **length(string)**	SELECT length('database'); length -------- 8 (1 row)	length(string)功能与char_length(string)完全等价，写法更简洁。

版权所有 © 华为云计算技术有限公司

分类	功能	对应函数	简单示例	使用区别
	获取指定编码格式的字符串的字符数。	length(string bytea, encoding name)	-	-

bit_length(string)

描述：字符串的位数。返回结果为字节数乘以8，得到总位数。字符串的位数取决于数据库编码方式，例如GBK字符编码下，一个中文为2个字节，UTF8和SQL_ASCII字符编码下，一个中文字为3个字节。

返回值类型：integer

示例：

```
SELECT bit_length('world');
 bit_length
------------
         40
(1 row)
```

GBK字符编码下，示例：

```
SELECT bit_length('我们');
 bit_length
------------
         32
(1 row)
```

UTF8和SQL_ASCII字符编码下，示例：

```
SELECT bit_length('我们');
 bit_length
------------
         48
(1 row)
```

octet_length(string)

描述：获取字符串中的字节数。该函数返回字符串的字节数，对于处理多字节字符（如 UTF-8 编码的汉字、特殊符号等）非常重要，它能准确获取字符串的存储大小。

返回值类型：integer

示例：

```
SELECT octet_length('data');
 octet_length
--------------
            4
(1 row)

SELECT octet_length('数据库');
 octet_length
--------------
            9
(1 row)
```


版权所有 © 华为云计算技术有限公司

lengthb(string)

描述：获取字符串的字节长度。**功能与octet_length(string)没有任何差别**，只是为兼容Oracle而存在，如果是从Oracle迁移过来的场景，请使用lengthb(string)，但**如果业务需要跨数据库运行（如同时兼容DWS和MySQL），请使用octet_length(string)**，因为MySQL没有lengthb函数。

返回结果与字符集有关，同样的中文字符，在GBK与UTF8中，返回的字节数不同。

返回值类型：integer

示例：

```
SELECT lengthb('hello');
 lengthb
---------
       5
(1 row)
```

lengthb(text/bpchar)

描述：获取指定字符串的字节数，入参支持text、bpchar（等同于CHAR(n)），其中:

- lengthb(text)表示计算实际字符的字节长度（无填充空格），适合绝大多数业务场景。
- lengthb(bpchar)表示计算固定长度字符串的字节长度（包含填充空格），适合严格定长场景（如固定长度的编码、证件号），不推荐日常使用（空格填充易引发隐式问题）。

返回值类型：integer

说明

- 若字符串中存在换行符，如字符串由一个换行符和一个空格组成，在DWS中LENGTH和LENGTHB的值为2。
- 此函数获取的为指定字符串的字节数，对于多字节字符编码，一个字符可能占用多个字节（例如，UTF-8编码中，中文字符通常占用3个字节）。

示例：

```
SELECT lengthb('hello');
 lengthb
---------
       5
(1 row)
```

入参为text类型和bpchar类型的示例对比如下：

```
-- 案例1: 使用 text 类型
SELECT lengthb('abc'::text);    -- 'abc'占用3字节（UTF-8），返回3
-- 案例2: 使用 bpchar (CHAR(5)) 类型
SELECT lengthb('abc'::CHAR(5)); -- 存储为 'abc  '， 返回5（每个空格占1字节）
```


版权所有 © 华为云计算技术有限公司

```
utf8=> SELECT lengthb('abc'::text);
 lengthb
---------
       3
(1 row)

utf8=> SELECT lengthb('abc'::CHAR(5));
 lengthb
---------
       5
(1 row)
```

UTF8编码中，示例：

```
SELECT lengthb('数据库');
 lengthb
---------
       9
(1 row)
```

char_length(string)或 character_length(string)

描述：获取字符串中的字符个数。

返回值类型：integer

说明

对于多字节字符编码，一个字符可能占用多个字节（例如，UTF-8编码中，中文字符通常占用3个字节）。

示例：

```
SELECT char_length('hello');
 char_length
-------------
           5
(1 row)

SELECT char_length('数据库');
 char_length
-------------
           9
(1 row)
```

length(string)

描述：获取字符串的字符长度（字符数量），length(string)功能与**char_length(string)**完全等价，写法更简洁。

返回值类型：integer

示例：

```
SELECT length('database');
 length
--------
      8
(1 row)
```


版权所有 © 华为云计算技术有限公司

length(string bytea, encoding name)

描述：获取指定编码格式的字符串的字符数。在指定的编码格式中，string必须是有效的。

返回值类型：integer

示例：

```
SELECT length('database', 'UTF8');
 length
--------
      8
(1 row)
```

7.1.3 截取字符串

在数据仓库多维分析场景中，**字符串截取**是支撑业务分析的关键技术能力。

典型场景包括：

- **提取核心信息**：对复合订单号进行智能分割，例如将"EC2023-BJ-0129A"拆解出核心序列号"0129A"，在18位身份证编码中使用substring()函数同时捕获行政区划代码、出生年月信息等。
- **脱敏数据**：用户隐私数据处理，例如通过right()函数精准提取手机号末4位实现数据脱敏。
- **精准提取信息**：支持**按正则表达式**进行匹配提取，从而精准提取有效信息，例如通过regexp_matches函数匹配出多个以b字母开头的字符串。
- **清除冗余数据**：例如清除某个订单号信息前后的多余空格或空字符。

DWS提供多种截取字符串的函数供用户选择。本章节中函数所使用的示例大部分以英文字符串示例为主，如果示例中存在中文字符串，则一般指UTF-8字符编码下的场景示例，字符编码在创建数据库时指定，参见**CREATE DATABASE**。

表 7-2 截取字符串常用函数

分类	功能	对应函数	简单示例	使用区别
去除两端字符	去除字符串两端指定的字符或空格。	**btrim(string text [, characters text])**	SELECT btrim(' sting '); btrim ------- sting (1 row)	-
提取子串，单位字节	提取子串，从第几个字节开始提取，一共提取几个字节。	**substrb(string,from,count)**	SELECT substrb('ABCDEF',2,2); substrb --------- BC (1 row)	四个函数的功能很相似，区别是提取剩余完整部分，还是只提取中间几个字节，入参是text类型，还是所有类字符串。

版权所有 © 华为云计算技术有限公司

分类	功能	对应函数	简单示例	使用区别
	提取子串，从第几个字节开始提取，一共提取几个字节。功能同上，区别是强调入参类型是text。	substrb(text,int,int)	-	
	提取子串，从第几个字节开始提取，提取剩余完整部分，当from是负数时，可以表示截取后边n个字节。	substrb(string,from)	SELECT substrb('ABCDEF',-2); substrb --------- EF (1 row)	
	提取子串，从第几个字节开始提取，提取剩余完整部分。功能同上，区别是强调入参类型是text。	substrb(text,int)	SELECT substrb('string',2); substrb --------- tring (1 row)	
提取子串，单位字符	提取子串，从第几个字符开始提取，一共提取几个字符。	substr(string,from,count)	SELECT substr('database',0,4); substr -------- data (1 row)	是提取剩余完整部分，还是只提取中间几个字符，substr(string,from,count)与substring(string [from int] [for int])高度相似，后者是标准SQL，前者更简洁。
	提取子串，从第几个字符开始提取，提取剩余完整部分。	substr(string,from)	SELECT substr('database',5); substr -------- base (1 row)	
	提取子串，从第几个字符开始提取，一共提取几个字符。与substr(string,from,count)功能一致。	substring(string [from int] [for int])	-	
	提取前几位字符。	left(str text, n int)	-	-
	提取后几位字符。	right(str text, n int)	-	-

版权所有 © 华为云计算技术有限公司

分类	功能	对应函数	简单示例	使用区别
按正则表达式提取子串	提取子串，匹配某种表达式，符合了才截取成功。	**substring(string from pattern)**	单击左侧链接跳转查看。	regexp_substr函数与substring函数类似，substring函数是SQL标准，regexp_substr属于扩展函数，正则能力更强。 regexp_matches返回类型为文本数组，适合提取分组信息的场景。
	提取子串，匹配某种表达式，符合了才截取成功，支持自定义转义符。	**substring(string from pattern for escape)**		
	功能同上类似，但是正则能力强。	**regexp_substr(text,text)**		
	功能同上类似，但是正则能力强。	**regexp_substr(source_char, pattern)**		
	提取子串，但是返回类型是文本数组，尤其适合需要多条件匹配或提取分组信息的场景。	**regexp_matches(string text, pattern text [, flags text])**		

btrim(string text [, characters text])

描述：去除字符串（string）两端指定的字符。仅处理两端字符，中间字符即使匹配也不会被移除。

返回值类型：text

参数说明：

- string，必选，目标字符串。
- characters，可选，要在目标字符串两端移除的字符。characters参数缺省时，默认去除所有的空白字符。

示例：

```
SELECT btrim('string' , 'ing');
 btrim
-------
 str
(1 row)

SELECT btrim('   string   ');  --不指定characters，默认去除多余空格
 btrim
-------
 string
(1 row)

SELECT btrim('xxdataxxbasexx','x');   --注意只去除两端的x，中间有x也不会去除。
```


版权所有 © 华为云计算技术有限公司

```
 btrim
--------
 dataxxbase
(1 row)
```

substrb(string,from,count)

描述：从字符串中截取指定长度的子字符串。from表示截取的起始位置，count表示截取的子字符串长度。

该函数和**substr(string,from,count)**函数功能一致，但是计算单位为字节。

- from为0时，按1处理。
- from为正数时，截取从from开始的count个字节。
- from为负数时，截取从倒数第n个开始的count个字节，n为from的绝对值。
- count小于1时，返回null。

返回值类型：bytea

示例：

```
SELECT substrb('ABCDEF',2,2);
 substrb
---------
 BC
(1 row)
```

substrb(text,int,int)

描述：从字符串中提取子字符串，第一个int表示从第几个字节位置开始提取，第二个int表示提取几个字节。

返回值类型：text

示例：

```
SELECT substrb('string',2,3);     ---从第2个字节“t”开始，提取3个字节。
 substrb
---------
 tri
(1 row)

SELECT substrb('学习数据库知识',7,9);     --UTF8编码下，一个中文占3个字节，从第7个字节位置开始，即跳过2个中文字，从“数”开始，截取9个字节（3个中文）
 substrb
---------
 数据库
(1 row)

SELECT substrb('学习数据库知识',7,9);     --GBK编码下，一个中文占2个字节，从第7个字节位置开始，即跳过3个中文字，从“据”开始，截取9个字节（4个中文）
 substrb
---------
 据库知识
(1 row)
```

substrb(string,from)

描述：从字符串中截取指定的子字符串，from表示截取的起始位置。该函数和**substr(string,from)**函数功能一致，但是计算单位为字节。

版权所有 © 华为云计算技术有限公司

- from为0时，按1处理。
- from为正数时，截取从from到末尾的所有字节。
- from为负数时，截取字符串的后n个字节，n为from的绝对值。

返回值类型：bytea

示例：

```
SELECT substrb('ABCDEF',-2);
 substrb
---------
 EF
(1 row)
```

substrb(text,int)

描述：从字符串中提取子字符串，int表示从第几个字节开始提取，提取后面剩余的完整部分。

返回值类型：text

示例：

```
SELECT substrb('string',2);
 substrb
---------
 tring
(1 row)

SELECT substrb('学习数据库知识',7);    --UTF8编码下，一个中文占3个字节，从第7个字节位置开始，即跳过2个中文字，从“数”开始截取后面完整部分
 substrb
---------
 数据库知识
(1 row)
```

substr(string,from,count)

描述：从字符串中截取指定长度的子字符串。

from表示截取的起始位置，count表示截取的子字符串长度。

- ORA和TD兼容模式下，from为0时，按1处理；MySQL兼容模式下，from为0时，返回空串。
- from为正数时，截取从from开始的count个字符。
- from为负数时，截取从倒数第n个开始的count个字符，n为from的绝对值。
- count小于1时，返回null。

说明

from为0或负数时，MySQL兼容模式下该函数的返回值与数据库GUC参数behavior_compat_options的配置项disable_compat_substr_mysql也相关，具体可参考**behavior_compat_options**参数兼容性配置项disable_compat_substr_mysql说明。

返回值类型：varchar

示例：

ORA和TD兼容模式下，from为0时，截取“database”从第1个字符开始的4个字符。

```
ora_compatibility_db=# SELECT substr('database',0,4);
 substr
```

 版权所有 © 华为云计算技术有限公司

```
--------
 data
(1 row)

td_compatibility_db=# SELECT substr('database',0,4);
 substr
--------
 data
(1 row)
```

MySQL兼容下，from为0时，返回空串。

```
mysql_compatibility_db=# SELECT substr('database',0,4);
 substr
--------

(1 row)
```

from为正数时，截取“database”从第5个字符开始的4个字符。

```
SELECT substr('database',5,4);
 substr
--------
 base
(1 row)
```

from为负数时，截取“database”从倒数第4个字符开始的3个字符。

```
SELECT substr('database',-4,3);
 substr
--------
 bas
(1 row)
```

substr(string,from)

描述：从字符串中截取指定的子字符串。

from表示截取的起始位置。

- ORA和TD兼容模式下，from为0时，按1处理；MySQL兼容模式下，from为0时，返回空串。
- from为正数时，截取从from到末尾的所有字符。
- from为负数时，截取字符串的后n个字符，n为from的绝对值。

> **说明**
>
> from为0或负数时，MySQL兼容模式下该函数的返回值与数据库GUC参数behavior_compat_options的配置项disable_compat_substr_mysql也相关，具体可参考**behavior_compat_options**参数兼容性配置项disable_compat_substr_mysql说明。

返回值类型：varchar

示例：

- ORA和TD兼容模式下，from为0时，截取“database”从第1个字符至末尾的所有字符。

```
ora_compatibility_db=# SELECT substr('database',0);
  substr
----------
 database
(1 row)

td_compatibility_db=# SELECT substr('database',0);
  substr
```

 版权所有 © 华为云计算技术有限公司

```
----------
 database
(1 row)
```

- MySQL兼容下，from为0时，返回空串。

```
mysql_compatibility_db=# SELECT substr('database',0);
 substr
--------

(1 row)
```

- from为正数时，截取“database”从第5个字符至末尾的所有字符。

```
SELECT substr('database',5);
 substr
--------
 base
(1 row)
```

- from为负数时，截取“database”从倒数第4个字符开始至末尾的所有字符。

```
SELECT substr('database',-4);
 substr
--------
 base
(1 row)
```

substring(string [from int] [for int])

描述：截取子字符串，from int表示从第几个字符开始截取，for int表示截取几个字符。

返回值类型：text

示例：

从字符串“Thomas”的第2个字符“h”开始连续截取3个字符，即返回“hom”。

```
SELECT substring('Thomas' from 2 for 3);
 substring
-----------
 hom
(1 row)
```

left(str text, n int)

描述：返回字符串的前n个字符。

- ORA和TD兼容模式下，当n是负数时，返回除最后|n|个字符以外的所有字符。
- MySQL兼容模式下，当n是负数时，返回空串。

返回值类型：text

示例：

```
 SELECT left('database', 4);
 left
------
 data
(1 row)
```

right(str text, n int)

描述：返回字符串中的后n个字符。

- ORA和TD兼容模式下，当n是负数时，返回除前|n|个字符以外的所有字符。

 版权所有 © 华为云计算技术有限公司

- MySQL兼容模式下，当n是负数时，返回空串。

返回值类型：text

示例：

```
SELECT right('abcde', 2);
 right
-------
 de
(1 row)

SELECT right('abcde', -2);
 right
-------
 cde
(1 row)
```

substring(string from *pattern*)

描述：截取匹配POSIX正则表达式（也称为模式）的子字符串，其中string表示原始字符串，pattern表示待匹配的模式。如果该字符串没有匹配上模式则返回空值，否则返回文本中匹配到模式的那一部分。

返回值类型：text

示例：

```
SELECT substring('Thomas' from '...$');    ---匹配末尾3个连续字符
 substring
-----------
 mas
(1 row)

SELECT substring('Thomas' from '^...');     ---匹配开头3个连续字符
 substring
-----------
 Tho
(1 row)

SELECT substring('Tom' from '....$');      ---匹配末尾4个连续字符，Tom仅3个字符，无法匹配条件，结果为空
 substring
-----------

(1 row)

SELECT substring('order_20251120_12345' from '_([0-9]+)_');      ---获取两个“_”之间的订单号
 substring
-----------
 20251120
(1 row)


SELECT substring('tel:13812345678,status:valid' from 'tel:([0-9]{11})');     ---提取手机号信息
 substring
-----------
 13812345678
(1 row)


SELECT substring('foobar' from 'o(.)b');     --匹配o(.)b规则， o表示匹配第一个字母o， (.)表示匹配任意一个字符，并用括号捕获它， b表示匹配到字母b。 'foobar' 中，'oob' 是匹配的子串，括号位置取的是子串的中间字符，所以结果是第二个o。
 substring
--------
 o
(1 row)
```

 版权所有 © 华为云计算技术有限公司

```
SELECT substring('foobar' from '(o(.)b)');  --嵌套捕获规则，外层第一个捕获组，匹配到oob，内层第二个捕获组，匹配到第3个o，函数结果仅返回第一个捕获组，因此结果为oob
 substring
--------
 oob
(1 row)
```

说明

如果POSIX正则表达式模式包含任何圆括号，那么将返回匹配第一对子表达式（对应第一个左圆括号的）的文本。如果想在表达式里使用圆括号而又不想导致这个例外，那么可以在整个表达式外边加上一对圆括号。

substring(string from *pattern* for *escape*)

描述：截取匹配SQL正则表达式的子字符串，其中string表示原始字符串，pattern表示匹配的模式，escape表示转义符，通过for进行指定。如果没有匹配到模式则返回空值。

为了标识在成功的时候应该返回的模式部分，模式必须包含转义字符的两次出现，并且每个转义字符后面要跟上双引号（"）。如果匹配这两个标记之间的模式的文本将被返回。

返回值类型：text

示例：

```
SELECT substring('Thomas' from '%#"o_a#"_' for '#');
 substring
-----------
 oma
(1 row)
```

示例解读：

1. **%#"o_a#"_** 为匹配的模式，**#**为转义符。
2. **%**表示匹配任意数量的任意前缀的通配符，该例子匹配到前缀**Tho**。
3. #"为一个标记，该语法的目的是要匹配两个#"之间包含的文本，即**o_a**。
4. 最后一个_表示匹配到后缀字符**s**。

综上，整个匹配模式体**%#"o_a#"**_最后展开来是 % o_a _，含义变为匹配「任意前缀」+ o +「单个任意字符」+ a +「单个任意字符」。

字符串Thomas如果需要匹配上这个模式，只截取两个#'之间的文本，即o_a，其中_可以表示通配任何字符，所以结果是**oma**。

拓展示例：

已知快递单号FX**5892**20251124，结构为FX + 4位单号 + 8位时间戳，实际系统只需要捕捉FX后面的4位，则可以通过以下写法获取。

```
SELECT substring('FX589220251124' from 'FX#"____#"%' for '#');
 substring
-----------
 5892
(1 row)

SELECT substring('FX564320241024' from 'FX#"____#"%' for '#');
 substring
-----------
```

 版权所有 © 华为云计算技术有限公司

```
 5643
(1 row)
```

regexp_substr(text,text)

描述：正则表达式的抽取子串函数。第一个text表示待处理的原始字符串，第二个text为定义的正则表达式。与substr功能相似，正则表达式出现多个并列的括号时，也全部处理。

返回值类型：text

示例：

```
SELECT regexp_substr('str','[ac]');   ---从“str”字符串中，匹配“ac”中的任意一个字符，结果为空
 regexp_substr
---------------


(1 row)
```

示例：从混合文本中提取第一个手机号。

```
SELECT regexp_substr('我的手机号是13812345678，备用号13987654321', '\d{11}');  -- \d{11}表示找11位连续数字
 regexp_substr
---------------
13812345678
(1 row)
```

示例：从快递单号中取前面10位订单号，10位后面的其他数字为时间戳，可以忽略。

```
SELECT regexp_substr('FX58920235822025112404', '\d{10}');   --\d{10} 表示找10位连续数字，匹配过程，从左到右扫描字符串，取第一个符合“10位连续数字”的子串。
 regexp_substr
---------------
 5892023582
(1 row)
```

regexp_substr(source_char, pattern)

描述：正则表达式的抽取子串函数，source_char表示原始字符串，pattern表示定义的正则表达式。

返回值类型：varchar

示例：

```
--匹配【一个逗号】+【至少1个非逗号字符】+【一个逗号】的连续子串
SELECT regexp_substr('500 Hello World, Redwood Shores, CA', ',[^,]+,') "REGEXPR_SUBSTR";
  REGEXPR_SUBSTR
-------------------
 , Redwood Shores,
(1 row)
```

regexp_matches(string text, pattern text [, flags text])

描述：返回string中所有匹配POSIX正则表达式的子字符串，返回结果是**文本数组**。如果pattern不匹配，该函数不返回行。这个函数在处理复杂文本解析时非常灵活，尤其适合需要多条件匹配或提取分组信息的场景。

- string为原始字符串。
- pattern为匹配模式。pattern可以是一个包含多个带圆括号的子表达式，例如**(bar)(beque)**，匹配成功后返回对应的含有多个元素的文本数组，例如

 版权所有 © 华为云计算技术有限公司

{bar,beque}；如果pattern不包含带有圆括号的子表达式，则返回一个单一元素的文本数组，例如**{barbeque}**。

- flags为可选参数，取值为i表示大小写不敏感，g表示替换每一个匹配的子字符串而不仅仅是第一个，即全局匹配。

须知

如果提供了最后一个参数flags，但参数值是空字符串（''），且数据库SQL兼容模式设置为ORA的情况下，会导致返回结果为空集。这是因为ORA兼容模式将''作为NULL处理，避免此类行为的方式有如下几种：

- 将数据库SQL兼容模式改为TD。
- 不提供最后一个参数，或最后一个参数不为空字符串。

返回值类型：多行文本数组 setof text[]

```
--匹配规则包含两个捕获分组，第一个(bar),匹配字面bar， 第二个(beque)，匹配字面beque，结果为包含2个元素的文本数组
SELECT regexp_matches('foobarbequebaz', '(bar)(beque)');
 regexp_matches
----------------
 {bar,beque}
(1 row)


--不带括号的匹配规则，结果为只包含1个元素的文本数组
SELECT regexp_matches('foobarbequebaz', 'barbeque');
 regexp_matches
----------------
 {barbeque}
(1 row)

--匹配b开头+若干非b字符，匹配结果分前后两组，即匹配到bar、beque ，但是不带g，所以只匹配了第一行
SELECT regexp_matches('foobarbequebazilbarfbonk', '(b[^b]+)(b[^b]+)');
    regexp_matches
--------------
 {bar,beque}
(1 row)


--相比于上面的示例，带了g，则继续匹配，然后显示到第二行文本数组中
SELECT regexp_matches('foobarbequebazilbarfbonk', '(b[^b]+)(b[^b]+)', 'g');
    regexp_matches
--------------
 {bar,beque}
 {bazil,barf}
(2 rows)

--从文本中提取所有11位手机号，全局匹配，返回所有行
SELECT regexp_matches('13812345678,13987654321', '\d{11}', 'g');
 regexp_matches
----------------
 {13812345678}
 {13987654321}
(2 rows)

--忽略大小写匹配，加参数i
SELECT regexp_matches('fooBarBeQuebaz', 'barbeque', 'i');
 regexp_matches
----------------
 {BarBeQue}
(1 row)
```

 版权所有 © 华为云计算技术有限公司

```
--不带i，则区分大小写，匹配为空
SELECT regexp_matches('fooBarBeQuebaz', 'barbeque');
 regexp_matches
----------------
(0 rows)
```

⚠ 注意

如果没有子查询，当regexp_matches函数没有匹配上时，不会输出表中的数据，这通常不是所需的返回结果，应避免这种写法，建议可使用**regexp_substr(text,text)**函数来实现相同的功能。

```
SELECT * FROM tab;
 c1  | c2
-----+-----
 dws | dws
(1 row)

SELECT c1, regexp_matches(c2, '(bar)(beque)') FROM tab;
 c1 | regexp_matches
----+----------------
(0 rows)

SELECT c1, c2, (SELECT regexp_matches(c2, '(bar)(beque)')) FROM tab;
 c1  | c2  | regexp_matches
-----+-----+----------------
 dws | dws |
(1 row)
```

7.1.4 拼接字符串

rawcat(raw,raw)

描述：字符串拼接函数。

返回值类型：raw

示例：

```
SELECT rawcat('ab','cd');
 rawcat
--------
 ABCD
(1 row)
```

repeat(string text, number int)

描述：按照设置的次数将字符串重复输出。如果指定的重复次数为零或负数，函数返回空字符串。

返回值类型：text

示例：

```
SELECT repeat('abc', 4);
    repeat
--------------
 abcabcabcabc
(1 row)
```

 版权所有 © 华为云计算技术有限公司

string || string

描述：连接字符串。

返回值类型：text

示例：

```
SELECT 'DA'||'TABASE' AS RESULT;
 result
----------
 DATABASE
(1 row)
```

string || non-string 或 non-string || string

描述：连接字符串和非字符串。

返回值类型：text

示例：

```
SELECT 'Value: '||123 AS RESULT;
   result
------------
 Value: 123
(1 row)
```

concat(str1,str2)

描述：将字符串str1和str2连接并返回。

- ORA和TD兼容模式下，返回结果为所有非NULL字符串的连接。
- MySQL兼容模式下，入参中存在NULL时，返回结果为NULL。

返回值类型：varchar

示例：

```
SELECT concat('Hello', ' World!');
    concat
--------------
 Hello World!
(1 row)
```

concat_ws(sep text, str"any" [, str"any" [, ...]])

描述：以第一个参数为分隔符，连接第二个以后的所有参数。

返回值类型：text

示例：

```
SELECT concat_ws(',', 'ABCDE', 2, NULL, 22);
 concat_ws
------------
 ABCDE,2,22
(1 row)
```

 版权所有 © 华为云计算技术有限公司

7.1.5 替换和填充字符串

lpad(string text, length int [, fill text])

描述：通过填充字符fill（缺省时为空白），把string填充为length长度。如果string已经比length长则将其尾部截断。

返回值类型：text

示例：

```
SELECT lpad('data', 8, 'xy');
  lpad
----------
 xyxydata
(1 row)
```

overlay(string placing string FROM int [for int])

描述：用指定的字符串替换原字符串中从指定位置开始的字符串部分，它可以灵活地处理字符串的替换操作。

参数说明：

FROM int，表示从第一个string的第几个字符开始替换。

for int，可选参数，表示替换第一个string的字符数目。如果不指定，默认替换从from开始的所有字符。

返回值类型：text

示例：从位置3开始，用“xyz”替换原字符串中的部分，结果为“abxyzf”。

```
SELECT overlay('abcdef' placing 'xyz' FROM 3);
 overlay
---------
 abxyzf
(1 row)
```

在字符串“hello”中，从第2个字符开始，连续替换3个字符，用“world”来覆盖这些位置，结果为“hworldo”。

```
SELECT overlay('hello' placing 'world' from 2 for 3 );
 overlay
---------
 hworldo
(1 row)
```

replace(string text, from text, to text)

描述：将原字符串中所有匹配from参数的部分替换为to参数指定的内容。

说明

replace函数大小写敏感，需确保from text和to text的大小写与原字符串匹配。

返回值类型：text

示例：

```
SELECT replace('databasedata', 'data', 'xxx');
  replace
```

 版权所有 © 华为云计算技术有限公司

```
------------
 xxxbasexxx
(1 row)
```

rpad(string text, length int [, fill text])

描述：使用填充字符fill，把string填充到length长度。如果string已经比length长则将其从尾部截断。

参数说明：

- string，要处理的原字符串。
- length，目标字符串的长度。
- fill（可选参数）：用于填充的字符或字符串。缺省时为单个空格 ' '。

返回值类型：text

示例：

使用填充字符“xy”将长度为4的原字符串“data”填充到目标长度10。

```
SELECT rpad('data', 10, 'xy');
    rpad
------------
 dataxyxyxy
(1 row)
```

使用字符“xy”填充原字符串“database”时，因string（database）比length（4）长，则将其从尾部截断，返回字符串data。

```
SELECT rpad('database',4,'xy');
 rpad
------
 data
(1 row)
```

rtrim(string text [, characters text])

描述：去除字符串右侧（末尾）的指定字符。

参数说明：

- string，要处理的原字符串。
- characters（可选参数），指定要从字符串右侧去除的字符集。缺省时，去除所有空白字符（例如，空格等）。

返回值类型：text

示例：

未指定characters参数时默认去除字符串“Database”右侧的空白字符。

```
SELECT rtrim('Database    ');
  rtrim
----------
 Database
(1 row)
```

去除字符串中的指定字符“x”。

```
SELECT rtrim('trimxxxx', 'x');
 rtrim
-------
```

 版权所有 © 华为云计算技术有限公司

```
 trim
(1 row)
```

rpad(string varchar, length int [, fill varchar])

描述：在字符串右侧填充指定字符，使字符串达到指定长度。如果string已经比length长则将其从尾部截断。

参数说明：

- string，要处理的原字符串。
- length，目标字符串的长度。length参数在DWS中表示字符长度。一个汉字长度计算为一个字符。
- fill（可选参数）：用于填充的字符或字符串。缺省时为单个空格' '。

返回值类型：varchar

示例：

```
SELECT rpad('data',10,'xyz');
 rpad
------------
 dataxyzxyz
(1 row)
SELECT rpad('hi',5,'abcdefg');
 rpad
-------
 hiabc
(1 row)
```

replace(string varchar, search_string varchar, replacement_string varchar)

描述：用于在字符串中查找特定子字符串（search_string）并将其替换为另一个字符串（replacement_string）。

参数说明：

- string，要处理的原字符串。
- search_string，需要被替换的子字符串。
- replacement_string：用于替换search_string的新字符串。

返回值类型：varchar

示例：

将字符串“ababab and mabab”中的所有子字符串“ab”替换为“cd”。

```
SELECT replace('ababab and mabab','ab','cd');
     replace
------------------
 cdcdcd and mcdcd
(1 row)
```

lpad(string varchar, length int[, repeat_string varchar])

描述：在字符串的左侧填充指定字符，使得字符串达到指定的长度。如果原字符串的长度已经大于目标长度（length参数），则从右侧将目标字符串截断为长度为length的字符串。

参数说明：

 版权所有 © 华为云计算技术有限公司

- string，需要填充的原字符串。
- length，目标字符串的长度。
- repeat_string，可选参数，用于指定填充的字符。缺省时使用空格填充。

返回值类型：varchar

示例：

原字符串“base”长度为4，目标长度为8，使用字符串“*”填充4个字符，填充后为“****base”。

```
SELECT lpad('base',8,'*');
   lpad
----------
 ****base
(1 row)
```

原字符串“database”长度为8，目标长度为6，不需要填充，从右侧将字符串“database”截断为长度为6的字符串，返回“databa”。

```
SELECT lpad('database',6,'data');
  lpad
--------
 databa
(1 row)
```

regexp_replace(string, pattern, replacement [,flags])

描述：替换匹配POSIX正则表达式的子字符串。 如果没有匹配pattern，那么返回不加修改的string串。 如果有匹配，则返回的string串里面的匹配子串将被replacement串替换掉。

replacement串可以包含\n， 其中\n是1到9， 表明string串里匹配模式里第n个圆括号子表达式的子串应该被插入， 并且它可以包含\&表示应该插入匹配整个模式的子串。

可选的flags参数包含零个或多个改变函数行为的单字母标记，见下表。

表 7-3 flags 参数的可选项

选项	描述
g	表示替换每一个匹配的子字符串而不仅仅是第一个（默认仅替换第一个匹配的子字符串）
B	默认情况下使用Henry Spencer正则库及其正则语法。指定B选项后，表示优先选用boost regex正则库及其正则语法。 以下两种情况在指定了B选项时，也会自动转换为选择Henry Spencer正则库及其正则语法： ● flags同时指定了p、q、w、x中的任意个字符。 ● string或pattern参数中含有多字节字符。
b	表示按照BRE(POSIX Basic Regular Expression)匹配模式的规则进行匹配。
c	大小写敏感匹配

 版权所有 © 华为云计算技术有限公司

选项	描述
e	表示按照ERE(POSIX Extended Regular Expression)匹配模式的规则进行匹配。当b和e都未指定时，如果选用的是Henry Spencer正则库，则按照ARE(Advanced Regular Expression，类似于Perl Regular Expression)匹配模式的规则进行匹配；如果选用的是boost regex正则库，则按照Perl Regular Expression匹配模式的规则进行匹配。
i	大小写不敏感匹配
m	换行敏感匹配，与选项n同义。
n	换行敏感匹配。此选项生效时，换行符影响元字符(.、^、$和[^)的匹配。
p	部分换行敏感匹配，此选项生效时，换行符影响元字符(.和[^)的匹配。部分是相对选项n而言。
q	重置正则表达式为加双引号的文本字符串，所有都是普通字符。
s	非换行敏感匹配。
t	紧凑语法（缺省）。该选项生效时，所有字符都很重要。
w	反部分换行敏感匹配，此选项生效时，换行符影响元字符(^和$)的匹配。部分是相对选项n而言。
x	扩展语法。与紧凑语法相对，在扩展的语法中，正则表达式中的空白字符被忽略。空白字符包括空格、水平制表符、新行和任何属于space字符表的字符。

返回值类型：varchar

示例：

```
SELECT regexp_replace('Thomas', '.[mN]a.', 'M');
 regexp_replace
----------------
 ThM
(1 row)
SELECT regexp_replace('foobarbaz','b(..)', E'X\\1Y', 'g') AS RESULT;
   result
-------------
 fooXarYXazY
(1 row)
```

SPACE(n int)

描述：返回由n个空格组成的字符串。若参数含有NULL值，返回NULL。

返回值类型：text

示例：

```
SELECT SPACE(2) as result;
 result
--------

(1 row)
```

 版权所有 © 华为云计算技术有限公司

7.1.6 删除指定字符

trim([leading |trailing |both] [characters] from string)

描述：删除字符串首部、尾部或者两端的特定字符。

参数说明：

- leading，仅删除字符串开头的空格或指定字符。
- trailing，仅删除字符串结尾的空格或指定字符。
- both，删除字符串开头和结尾的空格或指定字符（默认行为）。
- characters：可选参数，指定要去除的特定字符。如果不指定，默认去除空格。
- string：要处理的原字符串。

返回值类型：varchar

示例：

删除字符串“xTomxx”开头和结尾的字符“x”。

```
SELECT trim(BOTH 'x' FROM 'xTomxx');
 btrim
-------
 Tom
(1 row)

SELECT trim('x' FROM 'xTomxx');
 btrim
-------
 Tom
(1 row)
```

删除字符串“xTomxx”开头的字符“x”。

```
SELECT trim(LEADING 'x' FROM 'xTomxx');
 ltrim
-------
 Tomxx
(1 row)
```

删除字符串“xTomxx”结尾的字符“x”。

```
SELECT trim(TRAILING 'x' FROM 'xTomxx');
 rtrim
-------
 xTom
(1 row)
```

rtrim(string [, characters])

描述：从字符串尾部删除指定的字符。

参数说明：

- string：要处理的原字符串。
- characters（可选参数），指定要从字符串尾部删除的字符集。缺省时，删除所有空白字符（例如，空格等）。

返回值类型：varchar

 版权所有 © 华为云计算技术有限公司

示例：

```
SELECT rtrim('DataxBasexx','x');
   rtrim
-----------
 DataxBase
(1 row)

SELECT rtrim('    Data   ');
  rtrim
---------
    Data
(1 row)
```

ltrim(string [, characters])

描述：从字符串的开头删除指定的字符串。

参数说明：

- string：要处理的原字符串。
- characters（可选参数），指定要从字符串开头删除的字符集。缺省时，删除所有空白字符（例如，空格等）。

返回值类型：varchar

示例：

```
SELECT ltrim('xxDataxBase','x');
   ltrim
-----------
 DataxBase
(1 row)

SELECT ltrim('    Data   ');
 ltrim
--------
 Data
(1 row)
```

7.1.7 分割字符串

regexp_split_to_array(string text, pattern text [, flags text])

描述：用POSIX正则表达式作为分隔符，分隔string。和regexp_split_to_table相同，不过regexp_split_to_array会把它的结果以一个text数组的形式返回。

返回值类型：text[]

示例：

```
SELECT regexp_split_to_array('hello world', E'\\s+');
 regexp_split_to_array
-----------------------
 {hello,world}
(1 row)
```

regexp_split_to_table(string text, pattern text [, flags text])

描述：用POSIX正则表达式作为分隔符，分隔string。如果没有与pattern的匹配，该函数返回string。如果有至少有一个匹配，对每一个匹配它都返回从上一个匹配的末尾

 版权所有 © 华为云计算技术有限公司

（或者串的开头）到这次匹配开头之间的文本。当没有更多匹配时，它返回从上一次匹配的末尾到串末尾之间的文本。

flags参数包含零个或多个改变函数行为的单字母标记。i表示进行大小写无关的匹配，g表示替换每一个匹配的子字符串而不仅仅是第一个。

返回值类型：setof text

示例：

```
SELECT regexp_split_to_table('hello world', E'\\s+');
 regexp_split_to_table
-----------------------
 hello
 world
(2 rows)
```

⚠ 注意

如果没有子查询，当regexp_split_to_table函数没有匹配上时，会导致原始行被过滤掉。如果业务需要保留原始行，应避免这种写法。

```
SELECT * FROM tab;
 c1  | c2
-----+-----
 dws |
(1 row)

SELECT c1, regexp_split_to_table(c2, E'\\s+') FROM tab;
 c1 | regexp_split_to_table
----+-----------------------
(0 rows)

SELECT c1, (select regexp_split_to_table(c2, E'\\s+')) FROM tab;
 c1  | regexp_split_to_table
-----+-----------------------
 dws |
(1 row)
```

split_part(string text, delimiter text, field int)

描述：将原字符串根据指定的分隔符（delimiter参数）分割成若干个部分，然后提取指定位置（field参数）的部分。该函数适用于从复合字符串中提取特定字段。

参数说明：

- string，要分割的原字符串。
- delimiter，用于分割字符串的分隔符。
- field，分割后要获取的部分，从1开始计数。

返回值类型：text

示例：

将字符串“abc~@~def~@~ghi”按照“~@~”分隔符进行分割，得到3个部分（abc，def，ghi），提取第2部分内容，即“def”。

```
SELECT split_part('abc~@~def~@~ghi', '~@~', 2);
 split_part
------------
```

 版权所有 © 华为云计算技术有限公司

```
 def
(1 row)
```

7.1.8 转换字符串

upper(string)

描述：把字符串转化为大写。

返回值类型：varchar

示例：

```
SELECT upper('database');
  upper
----------
 DATABASE
(1 row)
```

ucase(string)

描述：把字符串转化为大写。

返回值类型：varchar

示例：

```
SELECT ucase('database');
  ucase
----------
 DATABASE
(1 row)
```

lower(string)

描述：把字符串转化为小写。

返回值类型：varchar

示例：

```
SELECT lower('DATABASE');
  lower
----------
 database
(1 row)
```

lcase(string)

描述：把字符串转化为小写。

返回值类型：varchar

示例：

```
SELECT lcase('DATABASE');
  lcase
----------
 database
(1 row)
```

 版权所有 © 华为云计算技术有限公司

initcap(string)

描述：将字符串中的每个单词的首字母转化为大写，其他字母转化为小写。

返回值类型：text

示例：

```
SELECT initcap('hi THOMAS');
  initcap
-----------
 Hi Thomas
(1 row)
```

quote_ident(string text)

描述：返回适用于SQL语句的标识符形式（使用适当的引号进行界定）。只有在必要的时候才会添加引号（字符串包含非标识符字符或者会转换大小写的字符）。返回值中嵌入的引号都写了两次。

返回值类型：text

示例：

```
SELECT quote_ident('hello world');
 quote_ident
--------------
 "hello world"
(1 row)
```

quote_literal(string text)

描述：返回适用于在SQL语句里当作文本使用的形式（使用适当的引号进行界定）。

返回值类型：text

示例：

```
SELECT quote_literal('hello');
 quote_literal
---------------
 'hello'
(1 row)
```

如果出现如下写法，text文本将进行转义。

```
SELECT quote_literal(E'O\'hello');
 quote_literal
---------------
 'O''hello'
(1 row)
```

如果出现如下写法，反斜杠会写入两次。

```
SELECT quote_literal('O\hello');
 quote_literal
---------------
 E'O\\hello'
(1 row)
```

如果参数为NULL，返回空。如果参数可能为null，通常使用函数quote_nullable更适用。

```
SELECT quote_literal(NULL);
 quote_literal
```

 版权所有 © 华为云计算技术有限公司

```
---------------

(1 row)
```

quote_literal(value anyelement)

描述：将给定的值强制转换为text，加上引号作为文本。

返回值类型：text

示例：

```
SELECT quote_literal(42.5);
 quote_literal
---------------
 '42.5'
(1 row)
```

如果出现如下写法，定值将进行转义。

```
SELECT quote_literal(E'O\'42.5');
 quote_literal
---------------
 'O''42.5'
(1 row)
```

如果出现如下写法，反斜杠会写入两次。

```
SELECT quote_literal('O\42.5');
 quote_literal
---------------
 E'O\\42.5'
(1 row)
```

quote_nullable(string text)

描述：返回适用于在SQL语句里当作字符串使用的形式（使用适当的引号进行界定）。

返回值类型：text

示例：

```
SELECT quote_nullable('hello');
 quote_nullable
----------------
 'hello'
(1 row)
```

如果出现如下写法，text文本将进行转义。

```
SELECT quote_nullable(E'O\'hello');
 quote_nullable
----------------
 'O''hello'
(1 row)
```

如果出现如下写法，反斜杠会写入两次。

```
SELECT quote_nullable('O\hello');
 quote_nullable
----------------
 E'O\\hello'
(1 row)
```

如果参数为NULL，返回NULL。

 版权所有 © 华为云计算技术有限公司

```
SELECT quote_nullable(NULL);
 quote_nullable
----------------
 NULL
(1 row)
```

quote_nullable(value anyelement)

描述：将给定的参数值转化为text，加上引号作为文本。

返回值类型：text

示例：

```
SELECT quote_nullable(42.5);
 quote_nullable
----------------
 '42.5'
(1 row)
```

如果出现如下写法，定值将进行转义。

```
SELECT quote_nullable(E'O\'42.5');
 quote_nullable
----------------
 'O''42.5'
(1 row)
```

如果出现如下写法，反斜杠会写入两次。

```
SELECT quote_nullable('O\42.5');
 quote_nullable
----------------
 E'O\\42.5'
(1 row)
```

如果参数为NULL，返回NULL。

```
SELECT quote_nullable(NULL);
 quote_nullable
----------------
 NULL
(1 row)
```

BIN(n bigint)

描述：将bigint类型从十进制转换成二进制，并以字符串的形式返回结果值。若参数含有NULL值，则返回NULL。

返回值类型：text

示例：

```
SELECT BIN(16) as result;
 result
--------
 10000
(1 row)
```

HEX(n)

描述：n可以是int类型也可以是字符串。返回n的十六进制字符串。若参数含有NULL值，返回NULL。

返回值类型：text

 版权所有 © 华为云计算技术有限公司

示例：

```
SELECT HEX(255) as result;
 result
--------
   FF
(1 row)
SELECT HEX('abc') as result;
 result
--------
 616263
(1 row)
```

UNHEX(n)

描述：执行HEX(n)的反向操作，n可以是int类型也可以是字符串，将参数中的每一对十六进制数字理解为一个数字，并将其转化为该数字代表的字符。若参数含有NULL值，返回NULL。该函数仅8.2.0及以上集群版本支持。

返回值类型：bytea

示例：

```
SELECT UNHEX('abc') as result;
 result
--------
 \x0abc
(1 row)
```

reverse(str)

描述：返回倒序排列的字符串。

返回值类型：text

示例：

```
SELECT reverse('abcde');
 reverse
---------
 edcba
(1 row)
```

to_hex(number int or bigint)

描述：将数字转换成十六进制的表现形式。

返回值类型：text

示例：

```
SELECT to_hex(2147483647);
  to_hex
----------
 7fffffff
(1 row)
```

translate(string text, from text, to text)

描述：根据from和to参数定义的字符映射关系，将原字符串中的每个字符替换为目标字符。如果from比to长，删掉在from中出现的额外的字符。

参数说明：

 版权所有 © 华为云计算技术有限公司

- string：要处理的原字符串。
- from：指定要替换的字符。
- to：指定替换后的字符。

返回值类型：text

示例：

根据from参数（143）和to参数（ax）的映射关系将原字符“12345”中的1替换为a，4替换为x， 3无对应替换字符所以删除。返回替换后的结果“a2x5”。

```
SELECT translate('12345', '143', 'ax');
 translate
-----------
 a2x5
(1 row)
```

ascii(string)

描述：返回字符串中第一个字符的ASCII码。

返回值类型：integer

示例：

返回字符串“xyz”中第一个字符“x”的ASCII码。

```
SELECT ascii('xyz');
 ascii
-------
   120
(1 row)
```

convert(string bytea, src_encoding name, dest_encoding name)

描述：以dest_encoding指定的目标编码方式转化字符串bytea。src_encoding指定源编码方式，在该编码下，string必须是合法的。

返回值类型：bytea

示例：

```
SELECT convert('text_in_utf8', 'UTF8', 'GBK');
         convert
----------------------------
 \x746578745f696e5f75746638
(1 row)
```

 版权所有 © 华为云计算技术有限公司

说明

如果源编码格式到目标编码格式的转化规则不存在，则字符串不进行任何转换直接返回，如GBK和LATIN1之间的转换规则是不存在的，具体转换规则可以通过查看系统表pg_conversion获得。

示例：

```
show server_encoding;
 server_encoding
-----------------
 LATIN1
(1 row)

SELECT convert_from('some text', 'GBK');
 convert_from
--------------
 some text
(1 row)

db_latin1=# SELECT convert_to('some text', 'GBK');
      convert_to
----------------------
 \x736f6d652074657874
(1 row)

db_latin1=# SELECT convert('some text', 'GBK', 'LATIN1');
       convert
----------------------
 \x736f6d652074657874
(1 row)
```

convert_from(string bytea, src_encoding name)

描述：以数据库的编码方式转化字符串bytea。

src_encoding指定源编码方式，在该编码下，string必须是合法的。

返回值类型：text

示例：

```
SELECT convert_from('text_in_utf8', 'UTF8');
 convert_from
--------------
 text_in_utf8
(1 row)
SELECT convert_from('\x6461746162617365','gbk');
 convert_from
--------------
 database
(1 row)
```

convert_to(string text, dest_encoding name)

描述：将字符串转化为dest_encoding的编码格式。

返回值类型：bytea

示例：

```
SELECT convert_to('some text', 'UTF8');
      convert_to
----------------------
 \x736f6d652074657874
(1 row)
SELECT convert_to('database', 'gbk');
    convert_to
```

 版权所有 © 华为云计算技术有限公司

```
--------------------
 \x6461746162617365
(1 row)
```

chr(integer)

描述：将整数转换为对应的ASCII码的字符。

返回值类型：varchar

示例：

```
SELECT chr(65);
 chr
-----
 A
(1 row)
```

format(formatstr text [, str"any" [, ...]])

描述：格式化字符串。

返回值类型：text

示例：

```
SELECT format('Hello %s, %1$s', 'World');
       format
--------------------
 Hello World, World
(1 row)
```

md5(string)

描述：将字符串使用MD5加密，并以16进制数作为返回值。

说明

MD5的安全性较低，不建议使用。

返回值类型：text

示例：

```
SELECT md5('data');
               md5
----------------------------------
 8d777f385d3dfec8815d20f7496026dc
(1 row)
```

decode(string text, format text)

描述：将二进制数据从文本数据中解码。

返回值类型：bytea

示例：

```
SELECT decode('ZGF0YWJhc2U=', 'base64');
    decode
---------------
 \x6461746162617365
(1 row)
```

 版权所有 © 华为云计算技术有限公司

```
SELECT convert_from('\x6461746162617365','utf-8');
 convert_from
--------------
 database
(1 row)
```

encode(data bytea, format text)

描述：将二进制数据编码为文本数据。

返回值类型：text

示例：

```
SELECT encode('database', 'base64');
  encode
----------
 ZGF0YWJhc2U=
(1 row)
```

CONV(n, fromBase, toBase)

描述： 将给定的数值或者字符串转换成目标进制，并按照字符串的形式输出结果。若参数含有NULL值，返回NULL。进制表示范围为[-36, -2]&[2, 36]。

返回值类型：text

示例：

```
SELECT CONV(-1, 10, 16) as result;
      result
------------------
 FFFFFFFFFFFFFFFF
(1 row)
```

7.1.9 比较字符串

STRCMP(text, text)

描述：比较两个字符串大小，若所有的字符串均相同，则返回0，若根据当前分类次序，第一个字符串小于第二个，则返回-1，其他情况返回1。若参数含有NULL值，返回NULL。

返回值类型：text

示例：

```
SELECT STRCMP('AA', 'AA'), STRCMP('AA', 'AB'), STRCMP('AA', 'A');
STRCMP  |  STRCMP  |  STRCMP
------------------------------
   0  |   -1   |    1
(1 row)
```

7.1.10 获取某一个字符或子字符串的位置

instr(text,text,int,int)

描述：获取子字符串在目标字符串中的位置。如果没有获取到匹配项，则返回0。

 版权所有 © 华为云计算技术有限公司

- 第一个text：必选，目标字符串。
- 第二个text，必选，待匹配的子字符串。
- 第一个int，表示开始匹配起始位置，若未配置该参数，默认从头开始匹配。
- 第二个int，表示匹配的次数，若未配置该参数，默认是第一次出现的位置。

返回值类型：integer

示例：

获取子字符串“bcd”在主字符串“abcdabcdabcd”中，从第2个字符开始，第2次匹配的位置。

```
SELECT instr( 'abcdabcdabcd', 'bcd', 2, 2 );
 instr
-------
     6
(1 row)

SELECT instr( 'abcdabcdabcd', 'lmk', 2);
 instr
-------
     0
(1 row)
```

position(substring in string)

描述：返回指定子字符串在字符串中的位置。若string中没有substring，则返回0。

返回值类型：integer

示例：

```
SELECT position('ing' in 'string');
 position
----------
        4
(1 row)

SELECT position('ing' in 'strin');
 position
----------
       0
(1 row)
```

strpos(string, substring)

描述：获取指定的子字符串（ substring ）在目标字符串（ string ）中首次出现的位置。子字符串不存在于目标字符串中时，返回0。

返回值类型：integer

示例：

```
SELECT strpos('source', 'rc');
 strpos
--------
      4
(1 row)

SELECT strpos('source', 'data');
 strpos
--------
      0
(1 row)
```

 版权所有 © 华为云计算技术有限公司

instr(string,substring[,position,occurrence])

描述：从字符串string的position（缺省时为1）所指的位置开始查找并返回第occurrence（缺省时为1）次出现子串substring的位置的值。

- 当position为0时，返回0。
- 当position为负数时，从字符串倒数第n个字符往前逆向搜索。n为position的绝对值。

本函数以字符为计算单位，如一个汉字为一个字符。

返回值类型：integer

示例：

```
SELECT instr('corporate floor','or', 3);
 instr
-------
     5
(1 row)
SELECT instr('corporate floor','or',-3,2);
 instr
-------
     2
(1 row)
```

locate(substring,string[,position])

描述：从字符串string的position（缺省时为1）所指的位置开始查找并返回第一次出现子串substring位置的值。以字符为计算单位。当string中不存在substring时，返回0。

返回值类型：integer

示例：

```
SELECT locate('ball','football');
 locate
--------
     5
(1 row)
SELECT locate('er','soccerplayer','6');
 locate
--------
    11
(1 row)
```

7.1.11 判断某个字符串是否符合某种匹配模式(字符串相关的模式匹配函数)

什么是模式匹配函数

模式匹配函数（LIKE）是一种字符串比较函数，用于判断**目标字符串**是否符合**指定的模式规则**，返回布尔值（true/false，示例中用 t/f 表示）。

它与**操作符=**的区别如下：

- **=**：严格匹配整个字符串（字符数、内容完全一致）。
- **LIKE**：模糊匹配（通过通配符忽略部分字符差异，只关注「关键模式」）。

适用场景：这类函数主要用于**模糊查询**，例如：

 版权所有 © 华为云计算技术有限公司

- 查找姓名中包含“张”的用户：name LIKE '%张%'；
- 查找手机号以“138”开头的用户：phone LIKE '138%'；
- 查找邮箱后缀为@abc.com且用户名长度为5位的用户：email LIKE '_____@abc.com'（5 个下划线）。

regexp_like(text,text,text)

描述：正则表达式的模式匹配函数。

返回值类型：bool

示例：

```
SELECT regexp_like('str','[ac]');
 regexp_like
-------------
 f
(1 row)
```

string [NOT] LIKE pattern [ESCAPE escape-character]

描述：模式匹配函数。

如果pattern不包含百分号或者下划线，该模式只代表它本身，这时候LIKE的行为就像等号操作符。在pattern里的下划线（_）匹配任何单个字符；而一个百分号（%）匹配零或多个任何字符。

要匹配下划线或者百分号本身，在pattern里相应的字符必须前导逃逸字符。缺省的逃逸字符是反斜杠，但是用户可以用ESCAPE子句指定一个。要匹配逃逸字符本身，写两个逃逸字符。

返回值类型：boolean

示例：

```
SELECT 'AA_BBCC' LIKE '%A@_B%' ESCAPE '@' AS RESULT;
 result
--------
 t
(1 row)
SELECT 'AA_BBCC' LIKE '%A@_B%' AS RESULT;
 result
--------
 f
(1 row)
SELECT 'AA@_BBCC' LIKE '%A@_B%' AS RESULT;
 result
--------
 t
(1 row)
```

REGEXP_LIKE(source_string, pattern [, match_parameter])

描述：正则表达式的模式匹配函数。

source_string为源字符串，pattern为正则表达式匹配模式。 match_parameter为匹配选项，可取值为：

- 'i'：大小写不敏感。

 版权所有 © 华为云计算技术有限公司

- 'c'：大小写敏感。
- 'n'：允许正则表达式元字符“.”匹配换行符。
- 'm'：将source_string视为多行。

若忽略match_parameter选项，默认为大小写敏感，“.”不匹配换行符，source_string视为单行。

返回值类型：boolean

示例：

```
SELECT regexp_like('ABC', '[A-Z]');
 regexp_like
-------------
 t
(1 row)
SELECT regexp_like('ABC', '[D-Z]');
 regexp_like
-------------
 f
(1 row)
SELECT regexp_like('abc', '[A-Z]','i');
 regexp_like
-------------
 t
(1 row)
SELECT regexp_like('abc', '[A-Z]');
 regexp_like
-------------
 f
(1 row)
```

7.2 二进制字符串函数和操作符

在数据仓库的数据处理过程中，用户经常需要操作二进制字符串（bytea类型）数据。例如，处理包含特殊字符或非文本数据的字段时，需要截取二进制子串、查找特定字节序列的位置、替换部分内容，或对二进制数据进行逐位/逐字节的精确操作。此外，从外部系统导入的数据可能包含分隔符分隔的复合字符串，需要按分隔符提取子串。

SQL标准定义了一部分二进制字符串函数，这些函数使用关键字（如from、placing、in）而不是逗号来分隔参数。DWS在此基础上提供了更多使用常规逗号分隔参数语法的扩展函数，覆盖更丰富的二进制字符串操作场景。本文档将系统介绍DWS提供的二进制字符串函数和操作符，帮助用户在实际业务中正确选择和使用这些函数。

基本概念

- **bytea类型**：DWS中的二进制数据类型，用于存储原始字节数据。bytea类型的数据以\x开头的十六进制格式显示，例如\x546f6d表示字符串”Tom”的字节序列。在SQL语句中，可以使用E'\\xxx'::bytea语法定义bytea字面量，其中\\xxx为八进制转义表示。
- **SQL关键字语法**：SQL标准定义的部分二进制字符串函数使用关键字而非逗号分隔参数。例如substring(string from int for int)使用from和for关键字指定起始位置和长度，position(substring in string)使用in关键字指定搜索范围。这些函数是SQL标准的一部分，兼容性较好。
- **函数调用语法**：DWS扩展的二进制字符串函数使用常规的逗号分隔参数语法。例如get_bit(string, offset)、btrim(string, bytes)等，与普通函数调用方式一致。

 版权所有 © 华为云计算技术有限公司

- **二进制字符串与普通字符串的区别**：普通字符串（text/varchar）按字符处理，受字符集编码影响；二进制字符串（bytea）按字节处理，不受字符集编码影响，适用于存储和操作任意二进制数据。

二进制字符串函数分类

DWS二进制字符串函数按功能可分为以下5类：

表 7-4 函数分类

分类	包含函数	说明
长度查询函数	octet_length(string)	查询二进制字符串的字节数。
查找与截取函数	position(substring in string)、substring(string [from int] [for int])、substring_index(string, delim, count)	查找子串位置、截取子串。
替换函数	overlay(string placing string from int [for int])	替换二进制字符串中的子串。
裁剪函数	trim([both] bytes from string)、btrim(string bytea, bytes bytea)	从首尾删除指定字节。
位与字节操作函数	get_bit(string, offset)、get_byte(string, offset)、set_bit(string, offset, newvalue)、set_byte(string, offset, newvalue)	逐位/逐字节读取和修改。

长度查询函数

octet_length(string)

描述：二进制字符串中的字节数。

返回值类型：integer

示例：

```
SELECT octet_length(E'jo\\000se'::bytea) AS RESULT;
 result
--------
      5
(1 row)
```

查找与截取函数

position(substring in string)

描述：返回特定子字符串在二进制字符串中的起始位置（从1开始计数）。未找到时返回0。

返回值类型：integer

版权所有 © 华为云计算技术有限公司

示例：

```
--返回子串\000om在Th\000omas中的起始位置（从1开始计数）,结果为3，表示\000om从第3个字节开始（T=第1字节, h=第2字节, \000=第3字节）。
SELECT position(E'\\000om'::bytea in E'Th\\000omas'::bytea) AS RESULT;
 result
--------
      3
(1 row)
```

substring(string [from int] [for int])

描述：从二进制字符串中截取子串。from指定起始位置（从1开始计数），for指定截取长度。省略from时从第1个字节开始，省略for时截取到字符串末尾。

返回值类型：bytea

示例：

```
--截取第2~4字节，即 68 00 6F，结果为\x68006f，十六进制表示截取到的3个字节（68='h', 00=空字节, 6f='o'）
SELECT substring(E'Th\\000omas'::bytea from 2 for 3) AS RESULT;
  result
----------
 \x68006f
(1 row)
```

截取时间，获取小时数：

```
SELECT substring('2022-07-18 14:38:15',12,2)AS RESULT;
 result
-----------
 14
(1 row)
```

substring_index(string, delim, count)

描述：按照区分大小写匹配查找delimiter分隔符，返回string字符串中第count次出现delim分隔符之前的子串。若count为负数，则从末尾向前查找delim分隔符。若参数含有NULL值，返回NULL。该函数仅8.2.0及以上集群版本支持。

返回值类型：text

示例：按照区分大小写匹配查找delimiter分隔符".wWw."，返回string字符串"www.wWw.cloud.wWw.com"中第2次出现delimiter分隔符之前的子串"www.wWw.cloud"。

```
SELECT SUBSTRING_INDEX('www.wWw.cloud.wWw.com', '.wWw.', 2) AS RESULT;
    result
---------------
 www.wWw.cloud
(1 row)
```

替换函数

overlay(string placing string from int [for int])

描述：替换二进制字符串中的子串。将string中从from指定位置开始的子串替换为placing指定的字符串。for指定被替换子串的长度，省略时替换长度等于placing字符串的长度。

返回值类型：bytea

示例：

 版权所有 © 华为云计算技术有限公司

```
--从第2个字节开始，用替换串(02 03)替换原串中3个字节(68 00 6F)的内容，结果为\x5402036d6173，原串第2~4字节(68 00 6F)被替换为(02 03)，结果为 54 02 03 6D 61 73
SELECT overlay(E'Th\\000omas'::bytea placing E'\\002\\003'::bytea from 2 for 3) AS RESULT;
    result
----------------
 \x5402036d6173
(1 row)
```

裁剪函数

trim([both] bytes from string)

描述：从string的开头和结尾删除只包含bytes中字节的最长字符串。both关键字可省略，表示同时裁剪首尾。

返回值类型：bytea

示例：

```
--删除原串开头和结尾的\000空字节，输出：\x546f6d，即"Tom"（54='T', 6f='o', 6d='m'），首尾的\000空字节已删除
SELECT trim(E'\\000'::bytea from E'\\000Tom\\000'::bytea) AS RESULT;
  result
----------
 \x546f6d
(1 row)
```

btrim(string bytea,bytes bytea)

描述：从string的开头和结尾删除只包含bytes中字节的最长的字符串。功能与**trim([both] bytes from string)**相同，但使用逗号分隔参数的函数调用语法。

返回值类型：bytea

示例：

```
--删除原串开头和结尾的\000空字节，功能同trim但使用逗号分隔参数语法，输出：\x7472696d，即"trim"（74='t', 72='r', 69='i', 6d='m'），首尾的\000空字节已删除
SELECT btrim(E'\\000trim\\000'::bytea, E'\\000'::bytea) AS RESULT;
   result
------------
 \x7472696d
(1 row)
```

位与字节操作函数

以下函数支持对二进制字符串进行逐位或逐字节的精确读取和修改。offset参数从0开始计数。

get_bit(string, offset)

描述：从字符串中抽取指定位置的一个位（bit），返回值为0或1。

返回值类型：integer

示例：

```
--获取第45位的值（0或1），输出：1，表示第45位的值为1（第5字节0x61=01100001，第5位=1）
SELECT get_bit(E'Th\\000omas'::bytea, 45) AS RESULT;
 result
--------
      1
(1 row)
```

 版权所有 © 华为云计算技术有限公司

get_byte(string, offset)

描述：从字符串中抽取指定位置的一个字节，返回值为该字节的ASCII码（0-255）。

返回值类型：integer

示例：

```
-- 输入：E'Th\\000omas'（字节序列：54 68 00 6F 6D 61 73），offset=4（第5个字节，0开始计数）
-- 功能：获取offset=4位置的字节值
SELECT get_byte(E'Th\\000omas'::bytea, 4) AS RESULT;
 result
--------
    109
(1 row)
-- 输出：109，即字符'm'的ASCII码值（0x6D = 十进制109）
```

set_bit(string,offset, newvalue)

描述：设置字符串中指定位置的位值。newvalue为0或1。返回修改后的新字符串，原字符串不变。

返回值类型：bytea

示例：

```
-- 输入：E'Th\\000omas'（54 68 00 6F 6D 61 73），offset=45（第5字节第5位），newvalue=0
-- 功能：将第45位的值从1设置为0
SELECT set_bit(E'Th\\000omas'::bytea, 45, 0) AS RESULT;
      result
------------------
 \x5468006f6d4173
(1 row)
-- 输出：\x5468006f6d4173，第5字节由0x61('a')变为0x41('A')，即第5位从1变为0
```

set_byte(string,offset, newvalue)

描述：设置字符串中指定位置的字节值。newvalue为0-255的整数。返回修改后的新字符串，原字符串不变。

返回值类型：bytea

示例：

```
-- 输入：E'Th\\000omas'（54 68 00 6F 6D 61 73），offset=4（第5个字节），newvalue=64
-- 功能：将第5个字节的值从0x6D(109)修改为64(0x40='@')
SELECT set_byte(E'Th\\000omas'::bytea, 4, 64) AS RESULT;
      result
------------------
 \x5468006f406173
(1 row)
-- 输出：\x5468006f406173，第5字节由6D变为40，即'm'变为'@'
```

7.3 位串函数和操作符

除了常用的比较操作符之外，还可以使用以下的操作符。&，|和#的位串操作数必须等长。在位移的时候，保留原始的位串长度（并以0填充）。

||

描述：位串之间进行连接。

示例：

 版权所有 © 华为云计算技术有限公司

```
SELECT B'10001' || B'011' AS RESULT;
  result
----------
 10001011
(1 row)
```

&

描述：位串之间进行“与”操作。

示例：

```
SELECT B'10001' & B'01101' AS RESULT;
 result
--------
 00001
(1 row)
```

|

描述：位串之间进行“或”操作。

示例：

```
SELECT B'10001' | B'01101' AS RESULT;
 result
--------
 11101
(1 row)
```

#

描述：位串之间如果不一致进行“或”操作。如果两个位串中对应位置都为1或者0，则该位置返回为0。

示例：

```
SELECT B'10001' # B'01101' AS RESULT;
 result
--------
 11100
(1 row)
```

~

描述：位串之间进行“非”操作。

示例：

```
SELECT ~B'10001'AS RESULT;
 result
----------
 01110
(1 row)
```

<<

描述：位串进行左移操作。

示例：

```
SELECT B'10001' << 3 AS RESULT;
 result
----------
```

 版权所有 © 华为云计算技术有限公司

```
 01000
(1 row)
```

>>

描述：位串进行右移操作。

示例：

```
SELECT B'10001' >> 2 AS RESULT;
 result
----------
 00100
(1 row)
```

下列SQL标准函数除了可以用于字符串之外，也可以用于位串：length，bit_length，octet_length，position，substring，overlay。

下列的函数用于位串和二进制字符串：get_bit，set_bit。当用于位串时，函数位数从字符串的第一位（最左边）作为0位。

另外，在整数和bit之间可以相互转换。示例：

```
SELECT 44::bit(10) AS RESULT;
   result
------------
 0000101100
(1 row)

SELECT 44::bit(3) AS RESULT;
 result
--------
 100
(1 row)

SELECT cast(-44 as bit(12)) AS RESULT;
    result
--------------
 111111010100
(1 row)

SELECT '1110'::bit(4)::integer AS RESULT;
 result
--------
     14
(1 row)
```

说明

只是转换为“bit”的意思是转换成bit(1)，因此只会转换成整数的最低位。

7.4 数学函数和运算符

7.4.1 数学运算符

数学运算操作符用于执行各种操作（如数学计算、位操作）的特殊符号。

 版权所有 © 华为云计算技术有限公司

操作符	描述	用法示例
+	加	计算2加3，返回结果5。 SELECT 2+3 AS RESULT; result -------- 5 (1 row)
-	减	计算2减3，返回结果-1。 SELECT 2-3 AS RESULT; result -------- -1 (1 row)
*	乘	计算2乘以3，返回结果6。 SELECT 2*3 AS RESULT; result -------- 6 (1 row)
/	除（除法操作符不会取整）	计算4除以2返回结果2。 SELECT 4/2 AS RESULT; result -------- 2 (1 row) 计算4除以3返回结果1.3333333333333333。 SELECT 4/3 AS RESULT; result ------------------- 1.3333333333333333 (1 row)
+/-	正/负	带有负数的运算。 SELECT -2-5 AS RESULT; result -------- -2 (1 row)
%	模（求余）	计算两个数相除后的余数。5除以4的余数，返回结果1。 SELECT 5%4 AS RESULT; result -------- 1 (1 row)
@	绝对值	获取-5.0的绝对值。 SELECT @ -5.0 AS RESULT; result -------- 5.0 (1 row)

 版权所有 © 华为云计算技术有限公司

操作符	描述	用法示例
^	幂（指数运算）	MySQL兼容模式下，作用为异或，参见**位串函数和操作符**章节的操作符“#”。 计算2的3次幂，返回结果8。 SELECT 2.0^3.0 AS RESULT; result ---------- 8.0000000000000000 (1 row)
|/	平方根	计算25的平方根，返回结果5。 SELECT |/ 25.0 AS RESULT; result -------- 5 (1 row)
||/	立方根	计算27的立方根，返回结果3。 SELECT ||/ 27.0 AS RESULT; result -------- 3 (1 row)
!	阶乘	计算5的阶乘，返回结果120。 SELECT 5! AS RESULT; result -------- 120 (1 row)
!!	阶乘（前缀操作符）	计算5的阶乘，返回结果120。 SELECT !!5 AS RESULT; result -------- 120 (1 row)
&	二进制AND	对91和15进行按位与（ AND）运算，返回结果11。 SELECT 91&15 AS RESULT; result -------- 11 (1 row)
|	二进制OR	对32和3进行按位或（ OR）运算，返回结果35。 SELECT 32|3 AS RESULT; result -------- 35 (1 row)
#	二进制XOR	对17和5进行按位异或（ XOR）运算，返回结果20。 SELECT 17#5 AS RESULT; result -------- 20 (1 row)

版权所有 © 华为云计算技术有限公司

操作符	描述	用法示例
~	二进制NOT	对1进行按位非（NOT）运算，返回结果-2。 SELECT ~1 AS RESULT; result -------- -2 (1 row)
<<	二进制左移	将1的二进制表示（例如，00000001）向左移动4位数后转为十进制，返回结果16。 SELECT 1<<4 AS RESULT; result -------- 16 (1 row)
>>	二进制右移	将8的二进制表示（例如，00001000）向右移动2位数后转为十进制，返回结果2。 SELECT 8>>2 AS RESULT; result -------- 2 (1 row)

7.4.2 数学函数

abs(x)

描述：计算x的绝对值。

输入参数x支持bigint、double precision、integer、numeric、real或smallint类型。

返回值类型：和输入相同。

示例：

```
SELECT abs(-17.4);
 abs
------
 17.4
(1 row)
```

acos(x)

描述：计算x的反余弦。

输入参数为double precision类型。取值范围为[-1,1]。

返回值类型：double precision

示例：

```
SELECT acos(-1);
       acos
------------------
 3.14159265358979
(1 row)
```

 版权所有 © 华为云计算技术有限公司

asin(x)

描述：计算x的反正弦。

输入参数为double precision类型。取值范围为[-1,1]。

返回值类型：double precision

示例：

```
SELECT asin(0.5);
       asin
------------------
 0.523598775598299
(1 row)
```

atan(x)

描述：计算x反正切函数。

输入参数为double precision类型。

返回值类型：double precision

示例：

```
SELECT atan(1);
       atan
------------------
 0.785398163397448
(1 row)
```

atan2(y, x)

描述：计算y/x的反正切函数。

输入参数为double precision类型。

返回值类型：double precision

示例：

```
SELECT atan2(2, 1);
      atan2
------------------
 1.10714871779409
(1 row)
```

bitand(integer, integer)

描述：计算两个数字与运算(&)的结果。

返回值类型：bigint

示例：

```
SELECT bitand(127, 63);
 bitand
--------
     63
(1 row)
```

 版权所有 © 华为云计算技术有限公司

sqrt(double precision or numeric)

描述：计算给定数值的平方根。

返回值类型：与输入类型相同。

示例：

```
SELECT sqrt(16.0);
       sqrt
-------------------
 4.000000000000000
(1 row)
```

cbrt(double precision)

描述：计算给定数值的立方根。

返回值类型：double precision

示例：

```
SELECT cbrt(27.0);
 cbrt
------
    3
(1 row)
```

ceil(double precision or numeric)

描述：对给定值向上取整。x为正数，则朝远离0的方向取整。x为负数，则朝向0的方向取整。

返回值类型：与输入类型相同。

示例：

```
SELECT ceil(-42.8);
 ceil
------
  -42
(1 row)
```

ceiling(double precision or numeric)

描述：对给定值向上取整（**ceil(double precision or numeric)**函数的别名）。

返回值类型：与输入类型相同。

示例：

```
SELECT ceiling(-95.3);
 ceiling
---------
     -95
(1 row)
```

floor(double precision or numeric)

描述：对给定值向下取整。x为正数，则朝向0的方向取整；x为负数，则朝远离0的方向取整。

返回值类型：与输入类型相同。

 版权所有 © 华为云计算技术有限公司

示例：

```
SELECT floor(-42.8);
 floor
-------
   -43
(1 row)
```

cos(x)

描述：计算x的余弦。

输入参数为double precision类型。

返回值类型：double precision

示例：

```
SELECT cos(-3.1415927);
        cos
--------------------
 -0.999999999999999
(1 row)
```

sin(x)

描述：计算x的正弦。

输入参数为double precision类型。

返回值类型：double precision

示例：

```
SELECT sin(1.57079);
       sin
------------------
 0.999999999979986
(1 row)
```

tan(x)

描述：计算x的正切。

返回值类型：double precision

示例：

```
SELECT tan(20);
       tan
------------------
 2.23716094422474
(1 row)
```

cot(x)

描述：计算x的余切。

输入参数为double precision类型。

返回值类型：double precision

示例：

 版权所有 © 华为云计算技术有限公司

```
SELECT cot(1);
       cot
------------------
 0.642092615934331
(1 row)
```

degrees(double precision)

描述：将弧度转为角度。

返回值类型：double precision

示例：

```
SELECT degrees(0.5);
     degrees
------------------
 28.6478897565412
(1 row)
```

radians(double precision)

描述：将角度转为弧度。

返回值类型：double precision

示例：

```
SELECT radians(45.0);
     radians
------------------
 0.785398163397448
(1 row)
```

div(y numeric, x numeric)

描述：y除以x的商的整数部分。

返回值类型：numeric

示例：

```
SELECT div(9,4);
 div
-----
   2
(1 row)
```

exp(double precision or numeric)

描述：计算给定值的自然指数。

返回值类型：与输入类型相同。

示例：

```
SELECT exp(1.0);
        exp
--------------------
 2.7182818284590452
(1 row)
```

 版权所有 © 华为云计算技术有限公司

random()

描述：返回0.0到1.0之间的随机数。

返回值类型：double precision

示例：

```
SELECT random();
      random
------------------
 0.824823560658842
(1 row)
```

rand()

描述：返回0.0到1.0之间的随机数。此函数为Mysql兼容性函数，仅8.2.0及以上集群版本支持。

返回值类型：double precision

示例：

```
SELECT rand();
      rand
------------------
 0.824823560658842
(1 row)
```

ln(double precision or numeric)

描述：自然对数。

返回值类型：与输入类型相同。

示例：

```
SELECT ln(2.0);
        ln
-------------------
 0.6931471805599453
(1 row)
```

log(double precision or numeric)

描述：计算以10为底的对数。

- ORA和TD兼容模式下，表现为以10为底的对数。
- MySQL兼容模式下，表现为自然对数。

返回值类型：与输入类型相同。

示例：

ORA和TD兼容模式下计算给定数值100以10为底的对数。

```
SELECT log(100.0);
        log
--------------------
 2.0000000000000000
(1 row)
```

MySQL兼容模式下计算给定数值100以自然数为底的对数。

 版权所有 © 华为云计算技术有限公司

```
SELECT log(100.0);
       log
--------------------
 4.6051701859880914
(1 row)
```

log(b numeric, x numeric)

描述：计算x以b为底数的对数。

返回值类型：numeric

示例：计算64.0以2.0为底的对数

```
SELECT log(2.0, 64.0);
       log
--------------------
 6.0000000000000000
(1 row)
```

mod(x,y)

描述：计算x/y的余数（即取模运算）。如果x是0，则返回0。如果y是0，则返回x。

返回值类型：与参数类型相同。

示例：

```
SELECT mod(9,4);
 mod
-----
   1
(1 row)
SELECT mod(9,0);
 mod
-----
   9
(1 row)
```

pi()

描述：返回“π”值，精确到小数点后15位。

返回值类型：double precision

示例：

```
SELECT pi();
       pi
------------------
 3.141592653589793
(1 row)
```

power(a double precision, b double precision)

描述：计算a的b次幂。

返回值类型：double precision

示例：

```
SELECT power(9.0, 3.0);
       power
----------------------
```

 版权所有 © 华为云计算技术有限公司

```
 729.0000000000000000
(1 row)
```

round(double precision| numeric)

描述：将给定数值进行四舍五入。

返回值类型：与输入类型相同。

示例：

```
SELECT round(42.4);
 round
-------
    42
(1 row)

SELECT round(42.6);
 round
-------
    43
(1 row)
```

说明

当调用round函数时，数值类型将舍入到零，而（在大多数计算机上）实数和双精度类型，以最接近的偶数为结果。

round(v numeric, s int)

描述：将给定数值进行四舍五入，保留小数点后s位。

返回值类型：numeric

示例：

```
SELECT round(42.4382, 2);
 round
-------
 42.44
(1 row)
```

setseed(double precision)

描述：为随后的random()调用设置种子(-1.0到1.0之间，包含边界值)。

返回值类型：void

示例：

```
SELECT setseed(0.54823);
 setseed
---------

(1 row)
```

sign(double precision or numeric)

描述：返回给定值的符号，通过1、0、-1表示。

返回值-1表示给定值为负数，返回值0表示给定值为0，返回值1表示给定值为正数。

示例：

 版权所有 © 华为云计算技术有限公司

```
SELECT sign(-8.4);
 sign
------
   -1
(1 row)
```

trunc(double precision or numeric)

描述：截断给定值的小数部分（即取整数部分）。

返回值类型：与输入类型相同。

示例：

```
SELECT trunc(42.8);
 trunc
-------
    42
(1 row)
```

trunc(v numeric, s int)

描述：截断给定值v的小数部分，保留小数点后s位。

返回值类型：numeric

示例：

```
SELECT trunc(42.4382, 2);
 trunc
-------
 42.43
(1 row)
```

truncate(v numeric, s int)

描述：截断给定值v的小数部分，保留小数点后s位。

v除了可以是任意精度型，还可以是整型和浮点型，返回值与入参v类型相同。s为负数时截断整数部分。该函数仅8.2.0及以上集群版本支持。

返回值类型：numeric

示例：

```
SELECT truncate(42.4382, 2);
 truncate
-------
 42.4300
(1 row)
```

width_bucket(operand numeric, b1 numeric, b2 numeric, count int)

描述：设定分组范围的最小值、最大值和分组个数，构建指定个数的大小相同的分组，返回指定字段值落入的分组编号。b1为分组范围的最小值，b2为分组范围的最大值，count为分组的个数。

返回值类型：integer

示例：

```
SELECT width_bucket(5.35, 0.024, 10.06, 5);
 width_bucket
```

 版权所有 © 华为云计算技术有限公司

```
--------------
             3
(1 row)
```

width_bucket(operand double precision, b1 double precision, b2 double precision, count int)

描述：设定分组范围的最小值、最大值和分组个数，构建指定个数的大小相同的分组，返回指定字段值落入的分组编号。b1为分组范围的最小值，b2为分组范围的最大值，count为分组的个数。

返回值类型：integer

示例：

```
SELECT width_bucket(5.35, 0.024, 10.06, 5);
 width_bucket
--------------
             3
(1 row)
```

7.5 时间日期函数和操作符

7.5.1 时间和日期操作符

说明

在使用时间和日期操作符时，对应的操作数请使用明确的类型前缀修饰，以确保数据库在解析操作数的时候能够与用户预期一致，不会产生非预期的结果。

如下示例，没有明确数据类型就会出现异常错误。

```
SELECT date '2001-10-01' - '7' AS RESULT;
```

表 7-5 时间和日期操作符

操作符	用法示例	示例结果
+	date + integer → timestamp 给日期加上天数。 `SELECT date '2001-09-28' + integer '7' AS RESULT;`	date类型参数与integer参数相加，获取时间间隔为7天后的时间： `2001-10-05 00:00:00`

版权所有 © 华为云计算技术有限公司

操作符	用法示例	示例结果
+	date + interval → timestamp 为日期添加时间间隔。 SELECT date '2001-09-28' + interval '1 hour' AS RESULT; SELECT date '2021-02-28' + interval '1 month' AS RESULT; SELECT date '2021-01-31' + interval '1 month' AS RESULT;	date类型参数与interval参数相加。 获取时间间隔为1小时后的时间： 2001-09-28 01:00:00 date函数对于日期相加减超过月份的日期范围，会对齐到对应月份最后一天，不超过则不处理。 2021-03-28 00:00:00 获取时间间隔为1个月的时间： 2021-02-28 00:00:00
+	date + time → timestamp 在日期中添加一天中的时间。 SELECT date '2001-09-28' + time '03:00' AS RESULT;	date类型参数与time类型参数相加，获取具体的日期和时间结果： 2001-09-28 03:00:00
+	interval + interval → interval 添加时间间隔。 SELECT interval '1 day' + interval '1 hour' AS RESULT;	interval参数相加，获取两个时间间隔之和： 1 day 01:00:00
+	timestamp + interval → timestamp 在时间戳中添加一个时间间隔。 SELECT timestamp '2001-09-28 01:00' + interval '23 hours' AS RESULT;	timestamp时间类型参数与interval参数相加，获取间隔23小时后的时间： 2001-09-29 00:00:00
+	time + interval → time 为时间添加时间间隔。 SELECT time '01:00' + interval '3 hours' AS RESULT;	time类型参数与interval参数相加，获取间隔时间为3小时后的时间： 04:00:00
-	date - date → integer 减去日期，生成日期时间差。 SELECT date '2001-10-01' - date '2001-09-28' AS RESULT;	date类型参数相减，获取两个日期的时间差： 3 days
-	date - integer → timestamp 从日期中减去天数。 SELECT date '2001-10-01' - integer '7' AS RESULT;	date类型参数与integer参数相减，获取两者的时间差： 2001-09-24 00:00:00
-	date - interval → timestamp 从日期中减去时间间隔。 SELECT date '2001-09-28' - interval '1 hour' AS RESULT;	date类型参数与interval参数相减，获取两者的日期、时间差： 2001-09-27 23:00:00
-	time - time → interval 减去时间。 SELECT time '05:00' - time '03:00' AS RESULT;	time类型参数相减，获取两参数的时间差： 02:00:00

版权所有 © 华为云计算技术有限公司

操作符	用法示例	示例结果
-	time - interval → time 从时间中减去时间间隔。 SELECT time '05:00' - interval '2 hours' AS RESULT;	time类型参数与interval相减，获取两参数的时间差： 03:00:00
-	timestamp - interval → timestamp 从时间戳中减去时间间隔。 SELECT timestamp '2001-09-28 23:00' - interval '23 hours' AS RESULT;	timestamp类型参数与interval相减，获取两者的日期时间差： 2001-09-28 00:00:00
-	interval - interval → interval 减去时间间隔。 SELECT interval '1 day' - interval '1 hour' AS RESULT;	interval参数相减，获取两者的时间差： 23:00:00
-	timestamp - timestamp → interval 减去时间戳。 SELECT timestamp '2001-09-29 03:00' - timestamp '2001-09-27 12:00' AS RESULT;	timestamp类型参数相减，获取两者的日期时间差： 1 day 15:00:00
-	select now() - interval → timestamp with time zone 获取当前日期的前一天： select now() - interval '1 day' AS RESULT;	获取当前日期的前一天： 2025-07-29 10:54:29.116501+08
*	interval * double precision → interval 将时间间隔乘以数量： SELECT 900 * interval '1 second' AS RESULT; SELECT 21 * interval '1 day' AS RESULT; SELECT double precision '3.5' * interval '1 hour' AS RESULT;	间隔1秒乘以900。 00:15:00 间隔1天乘以21天。 21 days 间隔1小时乘以3.5。 03:30:00
/	interval / double precision → interval SELECT interval '1 hour' / double precision '1.5' AS RESULT;	用时间间隔除以数量，获取一段时间中的某一段： 00:40:00

7.5.2 时间和日期函数

7.5.2.1 时间和日期函数概述

DWS的时间和日期函数用于在数据业务中加减日期，获取当前时间、格式化日期显示、提取年月日等操作，广泛应用于数据分析、日程管理、财务报表等领域，帮助用户提高数据处理效率和准确性。

表 7-6 时间和日期函数列表

类型	函数	功能
时间和日期基本函数	**day(date | timestamp with time zone)**	获取给定日期时间中的日。

版权所有 © 华为云计算技术有限公司

类型	函数	功能
	weekday(date | timestamp with time zone)	获取给定日期时间对应的星期值，星期一作为一周的开始日。
	week(date[, mode])	根据模式返回指定日期时间所处年份中对应的周数。
	month(date | timestamp with time zone)	获取给定日期时间的月份部分。
	quarter(date | timestamp with time zone)	获取给定日期时间所属的季度。
	year(date | timestamp with time zone)	获取给定日期时间所处的年份。
	yearweek(date[, mode])	获取给定日期时间在本年中对应的年份和周数
	dayofweek(date | timestamp with time zone)	获取给定日期时间对应的星期值，星期日作为一周的开始日。
	dayofmonth(date | timestamp with time zone)	获取给定日期时间所处月份中的天数。
	dayofyear(date | timestamp with time zone)	获取给定日期在本年中是第几天。
	weekofyear(date | timestamp with time zone)	获取给定日期时间在本年中是第几周，等价于week(date, 3)。
	makedate(year, dayofyear)	根据给定的年份和一年中的第几天生成对应的日期值。
	maketime(hour, minute, second)	根据给定的小时、分钟和秒数生成时间值。
	hour(timestamp with time zone)	获取时间中的小时值。
	minute(timestamp with time zone)	获取时间中的分钟值。
	second(timestamp with time zone)	获取时间的秒数值。
	microsecond(timestamp with time zone)	获取时间中的微秒值。

版权所有 © 华为云计算技术有限公司

类型	函数	功能
	from_unixtime(unix_timestamp[,format])	格式串缺省时，将unix时间戳转换为日期时间类型输出。
	unix_timestamp([timestamp with time zone])	用于将日期时间值转换为 Unix时间戳（即自1970-01-01 00:00:00 UTC以来经过的秒数）。
	now([fsp])	当前事务开始的日期及时间。
	last_day(d)	获取日期或者日期时间中对应月份的最后一天。
	next_day(x,y)	计算给定时间（x）之后第一个匹配目标星期几（y）的日期。
	from_days(days)	根据给定的天数，返回对应的日期值。
	to_days(timestamp)	根据指定日期，返回自0年开始的天数。
时间和日期运算函数	age(timestamp, timestamp)	将两个参数相减，并以年、月、日作为返回值。
	age(timestamp)	将两个参数相减，并以年、月、日作为返回值。
	adddate(date, interval | int)	返回给定日期时间加上指定单位的时间间隔的结果。
	subdate(date, interval | int)	返回给定日期时间减去指定单位的时间间隔的结果。
	addtime(timestamp | time | text, interval | text | numeric)	返回给定日期/时间加上指定时间间隔的结果。
	subtime(timestamp | time | text, interval | text | numeric)	返回给定日期/时间减去指定时间间隔的结果。
	date_add(date, interval)	返回给定日期时间加上指定单位的时间间隔的结果。
	date_sub(date, interval)	返回给定日期时间减去指定单位的时间间隔的结果。
	timestampadd(field, numeric | text, timestamp)	将以单位field的整数时间间隔（秒数可以带小数）添加到日期时间表达式中。
	timestampdiff(field, timestamp1, timestamp2)	将两个日期参数相减(timestamp2 - timestamp1)，并以单位field作为返回值。

版权所有 © 华为云计算技术有限公司

类型	函数	功能
	timediff(timestamp | time | text, timestamp | time | text)	将两个日期参数相减。
	datediff(date1, date2)	返回给定日期之间相差的天数值。
	period_add(P, N)	返回给定时期加上N个月后的日期。
	period_diff(P1, P2)	返回给定日期之间的月数差值。
	add_months(d,n)	用于计算时间点d再加上n个月的时间。
时间和日期截取函数	date_part(text, timestamp | interval)	从一个日期时间表达式中提取指定的部分。
	date_trunc(text, timestamp with time zone | timestamp without time zone | interval)	将时间日期数据截断到指定的精度。
	trunc(timestamp)	默认按天截取。
	extract(field from timestamp | time | interval)	从日期时间的表达式中提取指定的时间部分。
获取当前时间和日期函数	clock_timestamp()	实时时钟的当前时间戳。
	current_date	返回当前日期。
	curdate()	返回当前日期。此函数为MySQL兼容性函数。
	current_time	返回当前时间。
	curtime([fsp])	返回当前时间。
	current_timestamp	返回当前日期和时间（当前事务的开始时间）。
	localtime	返回当前时间。
	localtimestamp	返回当前日期及时间。
	statement_timestamp()	返回当前日期及时间（当前语句的开始时间）
	sysdate	返回当前系统的日期及时间。
	timeofday()	返回当前日期及时间。
	transaction_timestamp()	返回当前事务开始的系统日期及时间。

版权所有 © 华为云计算技术有限公司

类型	函数	功能
时间和日期格式化函数	date_format(timestamp, fmt)	将日期参数按照fmt指定的格式转换为字符串。
	time_format(time, fmt)	将日期参数按照fmt指定的格式转换为字符串。
	str_to_date(str, format)	将日期时间格式的字符串，按照所提供的显示格式转换为日期类型的值。
	str_to_time(str, format)	将时间格式的字符串，按照所提供的显示格式转换为时间类型的值。
其他时间和日期处理函数	justify_days(interval)	将时间间隔以30天为单位。
	justify_hours(interval)	将时间间隔以24小时为单位。
	justify_interval(interval)	结合justify_days和justify_hours，调整时间间隔。
	numtodsinterval(num, text)	将数字加时间间隔单位转换为INTERVAL DAY TO SECOND类型的值。
	convert_tz(timestamp, from_tz, to_tz)	将时间日期值timestamp从*from_tz*(text)给出的时区转到*to_tz*(text)给出的时区，返回转换后的时区时间。
	isfinite(date)	测试是否为有限日期。
	isfinite(timestamp)	测试判断是否为有限时间。
	isfinite(interval)	测试是否为有限区间。
	pg_sleep(seconds)	服务器线程延迟时间。

7.5.2.2 时间和日期基本函数

day(date | timestamp with time zone)

描述：获取给定日期时间中的日（即月份中的第几天）。与**dayofmonth(date)**函数相同。

取值范围：1~31

返回值类型：integer

示例：

```
SELECT day('2025-06-28');
 day
-----
  28
(1 row)
```

 版权所有 © 华为云计算技术有限公司

```
SELECT day('2025-06-28 09:00:00+08');
 day
-----
  28
(1 row)
```

weekday(date | timestamp with time zone)

描述：获取给定日期时间对应的星期值，星期一作为一周的开始日。

取值范围：0~6

返回值类型：integer

示例：

```
SELECT weekday('2020-11-06');
 weekday
---------
       4
(1 row)
SELECT weekday('2020-11-06 09:00:00+08');
 weekday
---------
       4
(1 row)
```

week(date[, mode])

描述：根据模式（mode）返回指定日期时间所处年份中对应的周数，默认模式为0。

- date，表示要获取周数的日期。
- mode，可选参数，指定不同的模式来计算一周的开始和一年的第一周。默认一周的第一天是星期日，也可以指定一周是从星期一还是星期日开始。mode支持的格式如下表所示。

返回值类型：integer

表 7-7 week 函数中 mode 模式的工作原理

模式	一周的第一天	周数范围	第一周的判断规则
0	星期日	0-53	从本年的第一个星期日开始，是第一周。前面的计算为第0周。
1	星期一	0-53	若1月1日到第一个周一的天数超过3天，则计算为本年的第一周。否则为第0周。
2	星期日	1-53	从本年的第一个星期日开始，是第一周。前面的计算为上年度的第5x周。
3	星期一	1-53	若1月1日到第一个周一的天数超过3天，则计算为本年的第一周。否则为上年度的第5x周。
4	星期日	0-53	若1月1日到第一个周日的天数超过3天，则计算为本年的第一周。否则为第0周。
5	星期一	0-53	从本年的第一个星期一开始，是第一周。前面的计算为第0周。

 版权所有 © 华为云计算技术有限公司

模式	一周的第一天	周数范围	第一周的判断规则
6	星期日	1-53	若1月1日到第一个周日的天数超过3天，则计算为本年的第一周。否则为上年度的第5x周。
7	星期一	1-53	从本年的第一个星期一开始，是第一周。前面的计算为上年度的第5x周。

示例：

获取指定日期的对应周数，默认模式0（周日为一周的第一天）。

```
SELECT WEEK('2026-01-01');
 week
------
    0     -- 由于1月1日不在2026年的第一周，根据模式0的规则，它属于上一年的最后一周，因此返回0。
(1 row)
```

获取指定日期的对应周数，默认模式1（ISO 标准周一为一周的第一天）。

```
SELECT WEEK('2026-01-01', 1);
 week
------
    1   -- 根据模式1的规则，2026年1月1日属于当年的第一周。
(1 row)
```

month(date | timestamp with time zone)

描述：获取给定日期时间的月份部分。

返回值类型：integer

示例：

```
SELECT month('2025-11-30');
 month
-------
    11
(1 row)
SELECT month('2025-11-30 09:00:00+08');
 month
-------
    11
(1 row)
```

quarter(date | timestamp with time zone)

描述：获取给定日期时间所属的季度。季度范围为1~4。

返回值类型：integer

示例：

```
SELECT quarter(date '2025-12-13');
 quarter
---------
       4
(1 row)
SELECT quarter('2025-03-20 09:00:00+08');
 quarter
```

 版权所有 © 华为云计算技术有限公司

```
---------
       1
(1 row)
```

year(date | timestamp with time zone)

描述：获取给定日期时间所处的年份。

返回值类型：integer

示例：

```
SELECT year('2025-11-13');
 year
------
 2025
(1 row)
SELECT year('2025-11-13 09:00:00+08');
 year
------
 2025
(1 row)
```

yearweek(date[, mode])

描述：获取给定日期时间在本年中对应的年份和周数，周数范围为1~53。

参数说明：

- date，输入的日期或时间戳。
- mode，可选参数，整数类型，用于确定周数计算的逻辑，标准的mode取值范围为0~7。参考**表7-7**。

  该函数支持更广泛的整数mode输入。输入的mode值会按照公式 (mode % 8 + 8) % 8 映射到 0~7 的标准模式范围内。例如，输入8等同于0，输入-1等同于7，输入-2 等同于6。支持输入任意32位整数（如2147483647）。

返回值类型：integer

示例：

返回给定日期对应的年份和周数。

```
SELECT yearweek('2025-11-13');
 yearweek
----------
   202545
(1 row)
SELECT yearweek('2025-1-1');
 yearweek
----------
   202452
(1 row)
```

指定mode值为-2（-2映射为6），表示周从星期日开始，2024年1月1日（星期一）属于第1周的第二天，2024年1月1日属于第1周。

```
SELECT yearweek('2024-1-1', -2);
 yearweek
----------
   202401
(1 row)
```

指定mode值为-6（-6映射为2），表示周从星期三开始，2024年1月1日（星期一）属于上一年的最后一周（即2023年的第53周）。

 版权所有 © 华为云计算技术有限公司

```
SELECT yearweek('2024-1-1', -6);
 yearweek
----------
   202353
(1 row)
```

指定mode值为2147483647（2147483647映射为7），表示本年的第一个星期一开始是第一周，2024年1月1日（星期一）属于第1周的第一天，2024年1月1日属于第1周。

```
SELECT yearweek('2024-1-1', 2147483647);
 yearweek
----------
   202401
(1 row)
```

dayofweek(date | timestamp with time zone)

描述：获取给定日期时间对应的星期值，星期日作为一周的开始日。

取值范围：1~7

返回值类型：integer

示例：

```
SELECT dayofweek('2025-11-22');
 dayofweek
-----------
         7
(1 row)
SELECT dayofweek('2025-11-22 09:00:00+08');
 dayofweek
-----------
         7
(1 row)
```

dayofmonth(date | timestamp with time zone)

描述：获取给定日期时间所处月份中的天数。

取值范围：1~31

返回值类型：integer

示例：

```
SELECT dayofmonth('2025-10-28');
 dayofmonth
------------
         28
(1 row)
SELECT dayofmonth('2025-10-28 09:00:00+08');
 dayofmonth
------------
         28
(1 row)
```

dayofyear(date | timestamp with time zone)

描述：获取给定日期时间在本年中是第几天。1月1日为第1天，12月31日为365或366天。

取值范围：1~366

 版权所有 © 华为云计算技术有限公司

返回值类型：integer

示例：

```
SELECT dayofyear('2025-10-28');
 dayofyear
-----------
       301
(1 row)
SELECT dayofyear('2025-10-28 09:00:00+08');
 dayofyear
-----------
       301
(1 row
```

weekofyear(date | timestamp with time zone)

描述：获取给定日期时间在本年中是第几周，等价于week(date, 3)。

取值范围：1~53

返回值类型：integer

示例：

```
SELECT weekofyear('2025-11-30');
 weekofyear
------------
         48
(1 row)
SELECT weekofyear('2025-11-30 09:00:00+08');
 weekofyear
------------
         48
(1 row)
```

makedate(year, dayofyear)

描述：根据给定的年份和一年中的第几天生成对应的日期值。

返回值类型：date

示例：

```
SELECT makedate(2025, 60);
  makedate
------------
 2025-03-01
(1 row)
```

maketime(hour, minute, second)

描述：根据给定的小时、分钟和秒数生成时间值。

由于GaussDB(DWS)中的time类型的取值范围为00:00:00到24:00:00, 故不支持hour大于24时和hour小于0时的场景。

返回值类型：time

示例：

```
SELECT maketime(12, 15, 30.12);
  maketime
-------------
```

 版权所有 © 华为云计算技术有限公司

```
 12:15:30.12
(1 row)
```

hour(timestamp with time zone)

描述：获取时间中的小时值。

返回值类型：integer

示例：

```
SELECT hour(timestamptz '2025-12-13 12:11:15+06');
 hour
------
   14
(1 row)
```

minute(timestamp with time zone)

描述：获取时间中的分钟值。

返回值类型：integer

示例：

```
SELECT minute(timestamptz '2025-12-13 12:11:15+06');
 minute
--------
     11
(1 row)
```

second(timestamp with time zone)

描述：获取时间的秒数值。

返回值类型：integer

示例：

```
SELECT second(timestamptz '2025-12-13 12:11:15+06');
 second
--------
     15
(1 row)
```

microsecond(timestamp with time zone)

描述：获取时间中的微秒值。

返回值类型：integer

示例：

```
SELECT microsecond(timestamptz '2025-12-13 12:11:15.123634+06');
 microsecond
-------------
      123634
(1 row)
```

from_unixtime(unix_timestamp[,format])

描述：格式串缺省时，将unix时间戳转换为日期时间类型输出。格式串指定时，将unix时间戳转换为指定格式的字符串输出。

 版权所有 © 华为云计算技术有限公司

返回值类型：timestamp（格式串缺省）/ text（格式串指定）

示例：

```
SELECT from_unixtime(875996580);
    from_unixtime
---------------------
 1997-10-04 20:23:00
(1 row)
SELECT from_unixtime(875996580, '%Y %D %M %h:%i:%s');
       from_unixtime
----------------------------
 1997 5th October 04:23:00
(1 row)
```

unix_timestamp([timestamp with time zone])

描述：用于将日期时间值转换为 Unix时间戳（即自1970-01-01 00:00:00 UTC以来经过的秒数）。无入参时，指定为当前时间。

返回值类型：bigint（无入参）/numeric（有入参）

示例：

```
SELECT unix_timestamp();
 unix_timestamp
----------------
     1768793418
(1 row)
SELECT unix_timestamp('2025-09-08 12:11:13+06');
    unix_timestamp
-------------------
 1757311873.000000
(1 row)
```

now([fsp])

描述：当前事务开始的日期及时间。入参确定微秒输出精度，缺省时为6。

返回值类型：timestamp with time zone

示例：

```
SELECT now();
              now
-------------------------------
 2026-01-19 11:35:09.082256+08
(1 row)
SELECT now(3);
             now
----------------------------
 2026-01-19 11:35:31.839+08
(1 row)
```

last_day(d)

描述：获取日期或者日期时间中对应月份的最后一天。

返回值类型：

- ORA和TD兼容模式下，返回值类型为timestamp。
- MySQL兼容模式下，返回值类型为date。

示例：

 版权所有 © 华为云计算技术有限公司

```
SELECT last_day(to_date('2026-01-01', 'YYYY-MM-DD')) AS cal_result;
    cal_result
---------------------
 2026-01-31 00:00:00
(1 row)
```

next_day(x,y)

描述：用于计算给定时间（x）之后第一个匹配目标星期几（y）的日期。

返回值类型：

- ORA和TD兼容模式下，返回值类型为timestamp。
- MySQL兼容模式下，返回值类型为date。

示例：

```
SELECT next_day(timestamp '2025-05-21 00:00:00','Sunday')AS cal_result;
    cal_result
---------------------
 2025-05-25 00:00:00
(1 row)
```

from_days(days)

描述：根据给定的天数，返回对应的日期值。

返回值类型：date

示例：

```
SELECT from_days(730669);
 from_days
------------
 2000-07-03
(1 row)
```

to_days(timestamp)

描述：根据指定日期，返回自0年开始的天数。

返回值类型：integer

示例：

```
SELECT to_days(timestamp '2025-10-07');
 to_days
---------
  739896
(1 row)
```

7.5.2.3 时间和日期运算函数

age(timestamp, timestamp)

描述：将两个参数相减，并以年、月、日作为返回值。若相减值为负，则函数返回亦为负。

返回值类型：interval

示例：

 版权所有 © 华为云计算技术有限公司

```
SELECT age(timestamp '2001-04-10', timestamp '1957-06-13');
          age
-------------------------
 43 years 9 mons 27 days
(1 row)
```

age(timestamp)

描述：当前时间和参数相减。

返回值类型：interval

示例：

```
SELECT age(timestamp '1957-06-13');
          age
-------------------------
 68 years 7 mons 7 days
(1 row)
```

adddate(date, interval | int)

描述：返回给定日期时间加上指定单位的时间间隔的结果。默认单位（即第二个参数为整型时）为天数。

返回值类型：timestamp

示例：

当入参为text类型时：

```
SELECT adddate('2020-11-13', 10);
  adddate
------------
 2020-11-23
(1 row)

SELECT adddate('2020-11-13', interval '1' month);
  adddate
------------
 2020-12-13
(1 row)

SELECT adddate('2020-11-13 12:15:16', interval '1' month);
       adddate
---------------------
 2020-12-13 12:15:16
(1 row)


SELECT adddate('2020-11-13', interval '1' minute);
       adddate
---------------------
 2020-11-13 00:01:00
(1 row)
```

当入参为date类型时：

```
SELECT adddate(current_date, 10);
  adddate
------------
 2021-09-24
(1 row)

SELECT adddate(date '2020-11-13', interval '1' month);
       adddate
---------------------
```

 版权所有 © 华为云计算技术有限公司

```
 2020-12-13 00:00:00
(1 row)
```

subdate(date, interval | int)

描述：返回给定日期时间减去指定单位的时间间隔的结果。默认单位（即第二个参数为整型时）为天数。

返回值类型：timestamp

示例：

当入参为text类型时：

```
SELECT subdate('2020-11-13', 10);
  subdate
------------
 2020-11-03
(1 row)

SELECT subdate('2020-11-13', interval '2' month);
  subdate
------------
 2020-09-13
(1 row)

SELECT subdate('2020-11-13 12:15:16', interval '1' month);
       subdate
---------------------
 2020-10-13 12:15:16
(1 row)

SELECT subdate('2020-11-13', interval '2' minute);
       subdate
---------------------
 2020-11-12 23:58:00
(1 row)
```

当入参为date类型时：

```
SELECT subdate(current_date, 10);
  subdate
------------
 2021-09-05
(1 row)

SELECT subdate(current_date, interval '1' month);
       subdate
---------------------
 2021-08-15 00:00:00
(1 row)
```

addtime(timestamp | time | text, interval | text | numeric)

描述：返回给定日期/时间加上指定时间间隔的结果。该函数仅8.2.0及以上集群版本支持。

- 如果第一个参数为timestamp，函数返回timestamp。
- 如果第一个参数为time或text，函数返回text或time（取决于具体实现和兼容性配置，通常表现为字符串格式的时间），且支持超过24 小时的时间表示（类似于MySQL 行为）。
- 支持第二个参数为numeric类型（通常表示秒数）。
- 支持第二个参数为text类型（解析为时间格式或数值）。

 版权所有 © 华为云计算技术有限公司

- 当输入文本包含无效字符时，会尝试截断并处理首字符负号的情况。

返回值类型：与第一个入参类型相同。

示例：

给定时间（Timestamp类型）加指定时间间隔 （Interval类型）。

```
SELECT ADDTIME('2020-11-13 01:01:01', '23:59:59');
addtime
---------------------
2020-11-14 01:01:00
(1 row)
```

给定时间（text类型）加指定时间间隔（text类型）返回结果可超过24小时。

```
SELECT ADDTIME('20:1:1', '20:59:59+8');
addtime
----------
41:01:00
(1 row)
```

给定时间（time类型）加指定时间间隔数值（numeric类型，通常表示秒数）。

```
SELECT addtime('10:00:00', 1.999999);
addtime
-----------------
10:00:01.999999
(1 row)
```

给定时间（text类型）加带有无效字符的文本时，会截断无效字符。

```
SELECT addtime('12:30:45', '00:15:30xyz');
addtime
----------
12:46:15
(1 row)
```

subtime(timestamp | time | text, interval | text | numeric)

描述：返回给定日期/时间减去指定时间间隔的结果。该函数仅8.2.0及以上集群版本支持。

返回值类型：与第一个入参类型相同。

示例：

timestamp减interval。

```
SELECT subtime('2020-11-13 01:01:01', '23:59:59');
subtime
---------------------
2020-11-12 01:01:02
(1 row)
```

time减数值 (数值视为秒)。

```
SELECT subtime('10:00:00', 1.999999);
subtime
-----------------
09:59:58.000001
(1 row)
```

处理负数文本输入。

```
SELECT subtime('10:00:00', '-59.9999999');
subtime
----------
```

 版权所有 © 华为云计算技术有限公司

```
10:01:00
(1 row)
```

文本输入包含负号及无效字符。

```
SELECT subtime('12:30:45', '-00:10:15abc');
subtime
----------
12:41:00
(1 row)
```

date_add(date, interval)

描述：返回给定日期时间加上指定单位的时间间隔的结果。等效于**adddate(date, interval | int)**。

返回值类型：timestamp

date_sub(date, interval)

描述：返回给定日期时间减去指定单位的时间间隔的结果，等效于**subdate(date, interval | int)**。

返回值类型：timestamp

timestampadd(field, numeric | text, timestamp)

描述：将以单位field的整数时间间隔（秒数可以带小数）添加到日期时间表达式中。若数值为负，则表示从给定的时间日期时间表达式中减去对应的时间间隔。

field 支持的参数：year，month，quarter，day，week，hour，minute，second，microsecond。

截断逻辑：

- 当第二个参数为 **数值 (numeric)** 时：支持小数运算（如 1.5 year）。
- 当第二个参数为 **文本 (text)** 时：
  - 若 field 为 second：保留小数部分进行计算。
  - 若 field 为其他单位（如 day, year 等）：**小数部分会被截断**（取整），仅使用整数部分进行计算。这是为了解决与 MySQL 行为不一致的问题。

说明

当函数传入的时间表达式格式为yyyy-mon-day+{value}的格式时，其中{value}字段将被解析为时区表达式，表示当前时间的时区，如'2023-01-31+1'将被解析为东一区时区的2023年01月31日。DWS这种表达式传入格式的解析与MySQL解析规则不一致，MySQL数据库仅当开启宽松解析规则才能支持此类格式，且解析结果为系统时区的2023年01月31日1时。

返回值类型：timestamp

示例：

基础整数加法。

```
SELECT timestampadd(year, 1, timestamp '2020-2-29');
   timestampadd
---------------------
 2021-02-28 00:00:00
(1 row)
```

数值类型支持小数 (Month 2.8，约等于2个月+0.8个月)。

版权所有 © 华为云计算技术有限公司

```
SELECT timestampadd(month, 2.8, '2020-01-31'::timestamp);
   timestampadd
---------------------
 2020-04-30 00:00:00
(1 row)
```

文本类型在非second 单位下的截断行为（ Day '30.9' 被截断为30 ）。

```
SELECT timestampadd(day, '30.9', '2016-02-29');
   timestampadd
---------------------
 2016-03-30 00:00:00
(1 row)
```

文本类型在second单位下保留小数。

```
SELECT timestampadd(second, '-4.999999', '2016-03-28 10:00:00');
       timestampadd
----------------------------
 2016-03-28 09:59:55.000001
(1 row)
```

timestampdiff(field, timestamp1, timestamp2)

描述：将两个日期参数相减(timestamp2 - timestamp1)，并以单位field作为返回值。若相减值为负，则函数返回值为负。field支持的参数为year、month、quarter、day、week、hour、minute、second和microsecond。

返回值类型：bigint

示例：

```
SELECT timestampdiff(day, timestamp '2001-02-01', timestamp '2003-05-01 12:05:55');
 timestampdiff
---------------
     819
(1 row)
```

timediff(timestamp | time | text, timestamp | time | text)

描述：将两个日期参数相减。若相减值为负，则函数返回值为负。该函数仅8.2.0及以上集群版本支持。

- 两个入参类型应当一致，或者可以隐式转换为一致的类型（ 均支持 text 格式输入 ）。
- 若第一个参数小于第二个参数，返回值为负数。
- 该函数支持计算结果超过24小时的差值。

返回值类型：time

示例：

timestamp相减。

```
SELECT TIMEDIFF('2022-7-5 1:1:1'::timestamp, '2022-7-5 2:2:2');
 timediff
-----------
 -01:01:01
(1 row)
```

time相减 (结果为负)。

```
SELECT TIMEDIFF('1:1:1'::TIME, '23:22:59');
 timediff
```

 版权所有 © 华为云计算技术有限公司

```
-----------
 -22:21:58
(1 row)
```

text格式相减。

```
SELECT TIMEDIFF('1:1:1', '23:22:59');
 timediff
-----------
 -22:21:58
(1 row)
```

datediff(date1, date2)

描述：返回给定日期之间相差的天数值。

返回值类型：bigint

示例：

```
SELECT datediff(date '2020-11-13', date '2012-10-16');
 datediff
----------
    2950
(1 row)
```

period_add(P, N)

描述：返回给定时期加上N个月后的日期。

返回值类型：integer

示例：

```
SELECT period_add(200801, 2);
 period_add
------------
    200803
(1 row)
```

period_diff(P1, P2)

描述：返回给定日期之间的月数差值。

返回值类型：integer

```
SELECT period_diff(200802, 200703);
 period_diff
-------------
          11
(1 row)
```

add_months(d,n)

描述：用于计算日期date加上n个月后的日期值。

返回值类型：timestamp

示例：

```
SELECT add_months(to_date('2017-5-29', 'yyyy-mm-dd'), 11) FROM dual;
     add_months
---------------------
 2018-04-29 00:00:00
(1 row)
```

 版权所有 © 华为云计算技术有限公司

7.5.2.4 时间和日期截取函数

date_part(text, timestamp | interval)

date_part函数用于从一个日期时间表达式（timestamp或interval）中提取指定的部分（text），等效于extract(field from timestamp)函数。

语法：

```
date_part('field', source)
```

- source，必须是一个timestamp、time或interval类型的值表达式（类型为date的表达式转换为timestamp，因此也可以用）。
- field，是一个标识符或者字符串，它指定从源数据中抽取的时间部分。field的取值见**表7-8**。

返回值类型：double precision

示例：

获取指定日期的小时数。

```
SELECT date_part('hour', timestamp '2001-02-16 20:38:40');
 date_part
-----------
        20
(1 row)
```

从时间间隔2 years 3 months中提取月份部分。

```
SELECT date_part('month', interval '2 years 3 months');
 date_part
-----------
         3
(1 row)
```

date_trunc(text, timestamp with time zone | timestamp without time zone | interval)

描述：将时间日期数据截断到指定的（text）精度。

text支持microseconds、milliseconds、second、minute、hour、day、week、month、quarter、year、decade、century、millennium。

返回值类型：timestamp with time zone、timestamp without time zone、interval

示例：

将输入的时间截断到小时。

```
SELECT date_trunc('hour', timestamp '2025-02-16 20:38:40');
     date_trunc
---------------------
 2025-02-16 20:00:00
(1 row)
```

将输入的时间截断到天。

```
SELECT date_trunc('day', timestamptz'2025-02-16 20:38:40+08');
       date_trunc
-------------------------
 2025-02-16 00:00:00-05
(1 row)
```

 版权所有 © 华为云计算技术有限公司

将输入的时间截断到月。

```
SELECT date_trunc('month', timestamp '2025-02-16 20:38:40');
    date_trunc
---------------------
 2025-02-01 00:00:00
(1 row)
```

将输入时间戳截断至一个季度的第一天。

```
SELECT DATE_TRUNC('quarter', TIMESTAMP '20250430 04:05:06.789');
    date_trunc
---------------------
 2025-04-01 00:00:00
(1 row)
```

截取当前月的1号12点。

```
SELECT date_trunc('month',now()) +interval '12h';
       ?column?
------------------------
 2025-09-01 12:00:00-04
(1 row)
```

截取当日的9点。

```
SELECT date_trunc('day',now()) + interval '9h';
       ?column?
------------------------
 2025-09-22 09:00:00-04
(1 row)
```

截取今年的第一天。

```
SELECT date_trunc('year',CURRENT_DATE);
      date_trunc
------------------------
 2025-01-01 00:00:00-05
(1 row)
```

date_bin(interval, source [, origin])

描述：将输入的时间戳“分桶”到指定间隔（步长）内，桶的边界与指定的原点对齐。该函数仅9.1.2及以上集群版本支持。

- interval，表示分桶的步长间隔。支持microseconds、milliseconds、second、minute、hour、day、week，不支持month或更长的时间间隔。
- source，表示待对齐的时间值，支持timestamp with time zone和timestamp without time zone类型。
- origin，表示桶的对齐原点，类型必须和source一致。

返回值类型：timestamp with time zone、timestamp without time zone、interval

示例：

1. 标准对齐：

   原点（origin）为2001-01-01 00:00:00，步长（interval）为15分钟，桶边界为15:30、15:45等整15分钟时刻。输入时间15:44:17（source）落在15:30~15:45的桶内，返回桶起始时间15:30:00。

   ```
   SELECT date_bin('15 minutes', TIMESTAMP '2020-02-11 15:44:17', TIMESTAMP '2001-01-01');
       date_bin
   ---------------------
    2020-02-11 15:30:00
   (1 row)
   ```

 版权所有 © 华为云计算技术有限公司

未指定origin时，使用默认原点2000-01-01 00:00:00，桶边界为整15分钟时刻。输入时间15:44:17落在15:30开始的桶内，返回15:30:00。

```
SELECT date_bin('15 minutes', TIMESTAMP '2020-02-11 15:44:17');
     date_bin
---------------------
 2020-02-11 15:30:00
(1 row)
```

2. 偏移对齐：

   原点（origin）为2001-01-01 00:02:30，步长（interval）仍为15分钟，但桶边界整体偏移2分30秒，变为15:32:30、15:47:30等时刻。输入时间15:44:17（source）落在15:32:30~15:47:30的桶内，返回桶起始时间15:32:30。

   当步长为完整时间单位（1 分钟、1 小时等）时，date_bin与对应的date_trunc查询结果相同。两者的区别在于：date_bin可以按任意间隔进行截断对齐。

```
SELECT date_bin('15 minutes', TIMESTAMP '2020-02-11 15:44:17', TIMESTAMP '2001-01-01 00:02:30');
     date_bin
---------------------
 2020-02-11 15:32:30
(1 row)
```

trunc(timestamp)

描述：默认按天截取。

返回值类型：timestamp

示例：

```
SELECT trunc(timestamp '2001-02-16
20:38:40');
                trunc
---------------------
2001-02-16 00:00:00
(1 row)
```

extract(field from timestamp | time | interval)

extract函数用于从日期时间的表达式中提取指定的时间部分，例如提取年份、月份、小时等。

语法：

```
EXTRACT(field FROM source)
```

- source，必须是一个timestamp、time或interval类型的值表达式（类型为date的表达式转换为timestamp，因此也可以用）。
- field，是一个标识符或者字符串，它指定从源数据中抽取的时间部分。field的取值见**表7-8**。

extract函数返回值类型：double precision

 版权所有 © 华为云计算技术有限公司

表 7-8 field 的取值

field	用法示例	提取结果
century	从给定的时间戳中提取世纪部分数值： SELECT EXTRACT(CENTURY FROM TIMESTAMP '2000-12-16 12:21:13'); 第一个世纪从0001-01-01 00:00:00 AD开始。这个定义适用于所有使用阳历的国家。没有0世纪，直接从公元前1世纪到公元1世纪。	返回结果如下： date_part ----------- 20
millennium	从指定的时间戳中提取千年部分的数值： SELECT EXTRACT(MILLENNIUM FROM TIMESTAMP '2001-02-16 20:38:40'); 20世纪（19xx年）里面的年份在第二个千年里。第三个千年从2001年1月1日零时开始。	返回结果如下： date_part ----------- 3 (1 row)
decade	从指定的时间戳中提取十年编号的数值： SELECT EXTRACT(DECADE FROM TIMESTAMP '2001-02-16 20:38:40');	返回结果如下： date_part ----------- 200 (1 row)
year	从指定的时间戳中提取年份部分。 SELECT EXTRACT(YEAR FROM TIMESTAMP '2001-02-16 20:38:40');	返回结果如下： date_part ----------- 2001 (1 row)
quarter	从指定的时间戳中提取日期所在的季度（1~4个季度）。 SELECT EXTRACT(QUARTER FROM TIMESTAMP '2001-02-16 20:38:40');	返回结果如下： date_part ----------- 1 (1 row)
month	• 从指定的时间戳中提取月份部分。如果source为timestamp，表示一年里的月份数（1~12）。 SELECT EXTRACT(MONTH FROM TIMESTAMP '2001-02-16 20:38:40'); • 从时间间隔中提取月份部分。如果source为interval，表示月的数目，自动将月份超过12的部分转换为年份，剩余部分作为月份。 SELECT EXTRACT(MONTH FROM INTERVAL '2 years 13 months');	• 返回结果如下： date_part ----------- 2 (1 row) • 返回结果如下： date_part ----------- 1 (1 row)
week	遵循ISO 8601标准：一年的第一个星期四在第一周；每周从星期一开始，到星期日结束；范围：1-53。 获取目标日期是在一年中的第几周。 SELECT EXTRACT(WEEK FROM TIMESTAMP '2001-02-16 20:38:40');	返回结果如下： date_part ----------- 7 (1 row)

版权所有 © 华为云计算技术有限公司

field	用法示例	提取结果

day	• 从指定的时间戳中提取天（一个月中的第几天）的部分。
	如果source为timestamp，表示月份里的日期（1~31）。
	`SELECT EXTRACT(DAY FROM TIMESTAMP '2001-02-16 20:38:40');`
	• 从时间间隔中提取天数部分。
	如果source为interval，表示天数。
	`SELECT EXTRACT(DAY FROM INTERVAL '40 days 1 minute');`
		• 返回结果如下：
		`date_part`
		`-----------`
		`16`
		`(1 row)`
		• 返回结果如下：
		`date_part`
		`-----------`
		`40`
		`(1 row)`

hour	从指定的时间戳中提取小时部分（0~23）。
	`SELECT EXTRACT(HOUR FROM TIMESTAMP '2001-02-16 20:38:40');`
		返回结果如下：
		`date_part`
		`-----------`
		`20`
		`(1 row)`

minute	从指定的时间戳中提取分钟部分。
	`SELECT EXTRACT(MINUTE FROM TIMESTAMP '2001-02-16 20:38:40');`
		返回结果如下：
		`date_part`
		`-----------`
		`38`
		`(1 row)`

second	从指定的时间戳中提取秒部分（包含小数部分）。
	`SELECT EXTRACT(SECOND FROM TIME '17:12:28.5');`
		返回结果如下：
		`date_part`
		`-----------`
		`28.5`
		`(1 row)`

milliseconds	从指定的时间戳中提取毫秒部分。返回的是整个秒字段的毫秒表示，即秒部分（包括小数秒）乘以 1000。
	`SELECT EXTRACT(MILLISECONDS FROM TIME '17:12:28.5');`
		返回结果如下：
		`date_part`
		`-----------`
		`28500`
		`(1 row)`

microseconds	从指定的时间戳中提取微秒部分。返回的是整个秒字段的微秒表示，即秒部分（包括小数秒）乘以 1,000,000。
	`SELECT EXTRACT(MICROSECONDS FROM TIME '17:12:28.5');`
		返回结果如下：
		`date_part`
		`-----------`
		`28500000`
		`(1 row)`

版权所有 © 华为云计算技术有限公司

field	用法示例	提取结果

epoch	● 如果source为timestamp with time zone，表示自1970-01-01 00:00:00-00 UTC以来的秒数（结果可能是负数）；
	`SELECT EXTRACT(EPOCH FROM TIMESTAMP WITH TIME ZONE '2001-02-16 20:38:40.12-08');`
	● 如果source为date和timestamp，表示自1970-01-01 00:00:00-00当地时间以来的秒数；
	● 如果source为interval，表示时间间隔的总秒数。
	`SELECT EXTRACT(EPOCH FROM INTERVAL '5 days 3 hours');`
	将epoch值转换为时间戳的方法。
	`SELECT TIMESTAMP WITH TIME ZONE 'epoch' + 982384720.12 * INTERVAL '1 second' AS RESULT;`
	`result`
	`---`
	`2001-02-17 12:38:40.12+08`
	`(1 row)`
		返回结果如下：
		`date_part`
		`---`
		`982384720.12`
		`(1 row)`
		返回结果如下：
		`date_part`
		`---`
		`442800`
		`(1 row)`

doy	获取指定时间戳中的日期是一年中的第几天（1~365/366）。
	`SELECT EXTRACT(DOY FROM TIMESTAMP '2001-02-16 20:38:40');`
		返回结果如下：
		`date_part`
		`---`
		`47`
		`(1 row)`

isodow	获取指定时间戳中的日期是一周中的第几天（1~7，星期一为1，星期天为7）。
	`SELECT EXTRACT(ISODOW FROM TIMESTAMP '2001-02-18 20:38:40');`
		返回结果如下：
		`date_part`
		`---`
		`7`
		`(1 row)`

dow	获取指定时间戳中的日期是一周中的第几天（0~6，星期天为0，星期六为6）
	`SELECT EXTRACT(DOW FROM TIMESTAMP '2001-02-16 20:38:40');`
		返回结果如下：
		`date_part`
		`---`
		`5`
		`(1 row)`

isoyear	获取指定日期在ISO 8601标准下的年份（不适用于interval）
	每个带有星期一开始的周中包含1月4日的ISO年，所以在年初的1月或12月下旬的ISO年可能会不同于阳历的年。可参见week的描述。
	`SELECT EXTRACT(ISOYEAR FROM DATE '2006-01-01');`
	`SELECT EXTRACT(ISOYEAR FROM DATE '2006-01-02');`
		返回结果如下：
		`date_part`
		`---`
		`2005`
		`(1 row)`
		返回结果如下：
		`date_part`
		`---`
		`2006`
		`(1 row)`

timezone	与UTC的时区偏移量，单位为秒。正数对应UTC东边的时区，负数对应UTC西边的时区。
	`SELECT EXTRACT(timezone FROM TIMETZ '17:12:28');`
		返回结果如下：
		`date_part`
		`---`
		`28800`
		`(1 row)`

timezone_hour	从带时区的时间值中提取时区偏移量的小时部分。
	`SELECT EXTRACT(timezone_hour FROM TIMETZ '17:12:28');`
		返回结果如下：
		`date_part`
		`---`
		`8`
		`(1 row)`

版权所有 © 华为云计算技术有限公司

field	用法示例	提取结果
timezone_minute	从带时区的时间值中提取时区偏移量的分钟部分。 SELECT EXTRACT(timezone_minute FROM TIMETZ '17:12:28');	返回结果如下： date_part ----------- 0 (1 row)

7.5.2.5 获取当前时间和日期函数

clock_timestamp()

描述：实时时钟的当前时间戳。

返回值类型：timestamp with time zone

示例：

```
SELECT clock_timestamp();
       clock_timestamp
-------------------------------
 2026-01-14 15:18:30.766999+08
(1 row)
```

current_date

描述：返回当前日期。

返回值类型：date

示例：

```
SELECT current_date;
    date
------------
 2026-01-14
(1 row)
```

curdate()

描述：返回当前日期。此函数为MySQL兼容性函数。该函数仅8.2.0及以上集群版本支持。

返回值类型：date

示例：

```
SELECT curdate();
  curdate
------------
 2026-01-14
(1 row)
```

current_time

描述：返回当前时间。

返回值类型：time with time zone

示例：

 版权所有 © 华为云计算技术有限公司

```
SELECT current_time;
      timetz
--------------------
 16:58:07.086215+08
(1 row)
```

curtime([fsp])

描述：返回当前时间。

fsp为选填参数，参数类型为整型，表示指定返回的小数秒精度。该函数仅8.2.0及以上集群版本支持。

返回值类型：time with time zone

示例：

```
SELECT curtime();
      timetz
--------------------
 16:58:07.086215+08
(1 row)
SELECT curtime(2);
      timetz
--------------------
 16:58:07.08+08
(1 row)
```

current_timestamp

描述：返回当前日期和时间（当前事务的开始时间）。

返回值类型：timestamp with time zone

示例：

```
SELECT current_timestamp;
       pg_systimestamp
-------------------------------
 2026-01-14 15:30:51.717927+08
(1 row)
```

localtime

描述：返回当前时间。

返回值类型：time

示例：

```
SELECT localtime;
     time
-----------------
 15:40:54.885857
(1 row)
```

localtimestamp

描述：返回当前日期及时间。

返回值类型：timestamp

示例：

 版权所有 © 华为云计算技术有限公司

```
SELECT localtimestamp;
        timestamp
----------------------------
 2026-01-14 15:41:49.111089
(1 row)
```

statement_timestamp()

描述：返回当前日期及时间（当前语句的开始时间）。

返回值类型：timestamp with time zone

示例：

```
SELECT statement_timestamp();
     statement_timestamp
-------------------------------
 2026-01-14 15:43:14.907913+08
(1 row)
```

sysdate

描述：返回当前系统的日期及时间。

返回值类型：timestamp

示例：

```
SELECT sysdate;
       sysdate
---------------------
 2026-01-14 15:47:58
(1 row)
```

timeofday()

描述：返回当前日期及时间（类似**clock_timestamp()**，但是返回为text。）

返回值类型：text

示例：

```
SELECT timeofday();
              timeofday
--------------------------------------
 Wed Jan 14 15:51:04.053480 2026 CST
(1 row)
```

transaction_timestamp()

描述：返回当前事务开始的系统日期及时间，与current_timestamp等效。

返回值类型：timestamp with time zone

示例：

```
SELECT transaction_timestamp();

    transaction_timestamp
-------------------------------
 2026-01-14 16:59:52.045873+08
(1 row)
```

 版权所有 © 华为云计算技术有限公司

7.5.2.6 时间和日期格式化函数

date_format(timestamp, fmt)

date_format函数将日期参数按照fmt指定的格式转换为字符串。

示例：

```
SELECT date_format('2009-10-04 22:23:00', '%M %D %W');
   date_format
--------------------
 October 4th Sunday
(1 row)
SELECT date_format('2021-02-20 08:30:45', '%Y-%m-%d %H:%i:%S');
     date_format
---------------------
 2021-02-20 08:30:45
(1 row)
SELECT date_format('2021-02-20 18:10:15', '%r-%T');
     date_format
----------------------
 06:10:15 PM-18:10:15
(1 row)
```

表7-9显示了可以用于将日期参数格式化输出的格式类型，这些格式类型适用于函数date_format、time_format、str_to_date、str_to_time和from_unixtime。

表 7-9 date_format 支持的输出格式

格式	说明	取值
%a	缩写星期名	Sun...Sat
%b	缩写月份名	Jan...Dec
%c	月份	0...12
%D	带英文后缀的月份日期	0th, 1st, 2nd, 3rd, …
%d	一个月里的日，2位	00...31
%e	一个月里的日	0...31
%f	微秒	000000...999999
%H	小时，24小时制	00...23
%h	小时，12小时制	01...12
%I	小时，12小时制，同%h	01...12
%i	分钟	00...59
%j	一年里的日	001...366
%k	小时，24小时制，同%H	0...23
%l	小时，12小时制，同%h	1...12
%M	月份名	January...December
%m	月份，两位	00...12

版权所有 © 华为云计算技术有限公司

格式	说明	取值
%p	上下午	AM PM
%r	时间，12小时制	hh:mm:ss AM/PM
%S	秒	00...59
%s	秒，同%S	00...59
%T	时间，24小时制	hh:mm:ss
%U	周 (00-53) 星期日是一周的第一天	00...53
%u	周 (00-53) 星期一是一周的第一天	00...53
%V	周 (01-53) 星期日是一周的第一天，与%X搭配使用	01...53
%v	周 (01-53) 星期一是一周的第一天，与%x搭配使用	01...53
%W	星期名	Sunday...Saturday
%w	一周的日，周日为0	0...6
%X	年份，其中的星期日是周的第一天，4 位，与%V搭配使用	-
%x	年份，其中的星期一是周的第一天，4 位，与%v搭配使用	-
%Y	年份，4位	-
%y	年份，2位	-
%%	字符'%'	字符'%'
%*x*	'*x*'，上述未列出的任意字符	字符'*x*'

⚠ 注意

date_format支持的输出格式中，%U、%u、%V、%v、%X、%x暂不支持。

time_format(time, fmt)

描述：time_format函数将日期参数按照fmt指定的格式转换为字符串。与date_format函数类似，但格式字符串只能包含小时、分钟、秒和微秒的格式说明符，如果包含其他说明符则会返回NULL值或0。

返回值类型：text

示例：

版权所有 © 华为云计算技术有限公司

```
SELECT time_format('2009-10-04 22:23:00', '%M %D %W');
   time_format
--------------------

(1 row)
SELECT time_format('2021-02-20 08:30:45', '%Y-%m-%d %H:%i:%S');
    time_format
---------------------
 0000-00-00 08:30:45
(1 row)
SELECT time_format('2021-02-20 18:10:15', '%r-%T');
    time_format
----------------------
 06:10:15 PM-18:10:15
(1 row)
```

须知

time_format仅支持时间相关的格式输出（%f、%H、%h、%I、%i、%k、%l、%p、%r、%S、%s、%T），不支持日期相关格式，其他情况处理为普通字符。

str_to_date(str, format)

描述：将日期/时间格式的字符串（str），按照所提供的显示格式（format）转换为日期类型的值。

返回值类型：timestamp

示例：

```
SELECT str_to_date('01,5,2021','%d,%m,%Y');
    str_to_date
---------------------
 2021-05-01 00:00:00
(1 row)
SELECT str_to_date('01,5,2021,09,30,17','%d,%m,%Y,%h,%i,%s');
    str_to_date
---------------------
 2021-05-01 09:30:17
(1 row)
```

适用于str_to_date的格式化输入的格式类型参考**表7-9**。这里仅支持“日期”格式、“日期+时间”格式的输入转换，对于仅“时间”格式的输入场景请使用str_to_time。

str_to_time(str, format)

描述：将时间格式的字符串（str），按照所提供的显示格式（format）转换为时间类型的值。

返回值类型：time

示例：

```
SELECT str_to_time('09:30:17','%h:%i:%s');
 str_to_time
-------------
 09:30:17
(1 row)
```

适用于str_to_time的格式化输入的格式类型参考**表7-9**，这里仅支持“时间”格式的输入转换，对于“日期”格式、“日期+时间”格式的输入场景请使用str_to_date。

 版权所有 © 华为云计算技术有限公司

7.5.2.7 其他时间和日期处理函数

justify_days(interval)

描述：调整间隔，使得30天时间周期表示为月。

返回值类型：interval

示例：

```
SELECT justify_days(interval '35 days');
 justify_days
--------------
 1 mon 5 days
(1 row)
```

justify_hours(interval)

描述：调整时间间隔，使得24小时时间周期表示为日。

返回值类型：interval

示例：

```
SELECT JUSTIFY_HOURS(INTERVAL '27 HOURS');
 justify_hours
----------------
 1 day 03:00:00
(1 row)
```

justify_interval(interval)

描述：结合justify_days和justify_hours，调整时间间隔。

返回值类型：interval

示例：

```
SELECT JUSTIFY_INTERVAL(INTERVAL '1 MON -1 HOUR');
 justify_interval
------------------
 29 days 23:00:00
(1 row)
```

numtodsinterval(num, text)

描述：将数字加时间间隔单位转换为INTERVAL DAY TO SECOND类型的值。可以用来对一个日期时间值进行加减计算。可以通过设置参数IntervalStyle为oracle，兼容该函数在Oracle中的interval输出格式。

参数：

- num，为numeric类型数字，表示要转换的数字。
- text，为指定间隔单位，支持DAY、 HOUR、 MINUTE和SECOND

示例：

返回当前时间加100小时后的时间。

```
SELECT sysdate + numtodsinterval(100, 'HOUR') FROM dual;
      ?column?
---------------------
```

 版权所有 © 华为云计算技术有限公司

```
 2026-01-18 18:49:00
(1 row)
```

设置参数IntervalStyle为oracle，兼容该函数在Oracle中的interval输出格式。将100小时转换为INTERVAL DAY TO SECOND类型数据。

```
SET intervalstyle = oracle;
SET
SELECT numtodsinterval(100, 'HOUR');
       numtodsinterval
------------------------------
 +000000004 04:00:00.000000000
(1 row)
```

convert_tz(timestamp, from_tz, to_tz)

描述：将时间日期值timestamp从*from_tz* (text)给出的时区转到*to_tz*(text)给出的时区，返回转换后的时区时间。该函数仅8.2.0及以上集群版本支持。

返回值类型：timestamp without time zone

示例：

```
SELECT convert_tz('2018-12-25 13:25:00', '+02:00', '+08:00');
    convert_tz
--------------------
 2018-12-25 19:25:00
(1 row)
SELECT convert_tz('2000-02-28 23:00:00', 'GMT', 'MET');
    convert_tz
--------------------
 2000-02-29 00:00:00
(1 row)
```

isfinite(date)

描述：测试是否为有限日期。

返回值类型：boolean

示例：

```
SELECT isfinite(date '2001-02-16');
 isfinite
----------
 t
(1 row)
SELECT isfinite(date 'infinity');
 isfinite
----------
 f
(1 row)
```

isfinite(timestamp)

描述：测试判断是否为有限时间戳。

返回值类型：boolean

示例：

```
SELECT isfinite(timestamp '2001-02-16 21:28:30');
 isfinite
----------
```

 版权所有 © 华为云计算技术有限公司

```
 t
(1 row)
SELECT isfinite(timestamp 'infinity');
 isfinite
----------
 f
(1 row)
```

isfinite(interval)

描述：测试是否为有限时间间隔。

返回值类型：boolean

示例：

```
SELECT isfinite(interval '4 hours');
 isfinite
----------
 t
(1 row)
```

pg_sleep(seconds)

描述：服务器线程延迟时间，以秒计。

返回值类型：void

示例：

```
SELECT pg_sleep(10);
 pg_sleep
----------

(1 row)
```

7.6 SEQUENCE 函数

在数据库应用开发中，唯一且递增的序号生成是一项基础而常见的需求。例如，在电商系统中为每笔订单生成唯一的订单号、在用户注册时分配自增主键ID、在日志系统中为每条记录生成唯一流水号等。这些场景都需要在多用户并发环境下安全地获取不重复的序列值。

DWS提供了SEQUENCE函数（包括nextval、currval、lastval、setval），为用户从序列对象中获取后续的序列值提供了简单的多用户安全的方法。相比应用层自行生成序号，SEQUENCE函数通过全局事务管理器（GTM）统一协调序列值分配，能够在分布式集群环境中保证序列值的唯一性和连续性，避免并发冲突。

基本概念

- **regclass**：DWS中的数据类型，表示系统表pg_class中某行的OID（对象标识符）。在序列函数中，regclass参数用于指定序列对象的名称或OID，系统会自动将序列名称（如'seqDemo'）解析为对应的OID。
- **is_called标志**：序列对象的内部状态标志，控制下一次nextval的返回行为。当is_called为true时，表示当前序列值已被“调用过”，下一次nextval将返回当前值+增量；当is_called为false时，表示当前序列值还未被“调用过”，下一次nextval将返回当前值本身。

 版权所有 © 华为云计算技术有限公司

- **下推（Pushdown）**：DWS查询优化机制，指将SQL语句从CN下发给DN直接执行，减少CN与DN之间的数据传输。当nextval被下推到DN时，各DN会各自连接GTM请求序列值，可能消耗较多GTM连接数。

前置操作：创建序列

在使用SEQUENCE函数前，需要先创建序列对象。执行以下SQL语句创建示例序列：

```
-- 创建示例序列seqDemo，起始值为1，步长为1
CREATE SEQUENCE seqDemo START 1;
-- 创建示例序列seq1，起始值为1，步长为1
CREATE SEQUENCE seq1 START 1;
```

CREATE SEQUENCE的完整语法和参数说明请参见**CREATE SEQUENCE**。创建序列后，即可使用以下SEQUENCE函数操作该序列对象。

nextval(regclass)

递增序列并返回新值。

说明

- 为了避免从同一个序列获取值的并发事务被阻塞， nextval操作不会回滚；也就是说，一旦获取了一个值，它就会被认为已被使用，并且不会再次返回。 即使该操作处于事务中，当事务随后中断，或者调用查询最终没有使用该值，也会出现这种情况。这种情况将在指定值的顺序中留下未使用的“空洞”。 因此，DWS序列对象不能用于获取“无间隙”序列。
- 如果nextval被下推到DN上时，各个DN会自动连接GTM，请求next values值，例如（INSERT INTO t1 SELECT xxx，t1某一列需要调用nextval函数），由于GTM上有最大连接数为8192的限制，而这类下推语句会导致消耗过多的GTM连接数，因此对于这类语句的并发数目限制为7000（其他语句需要占用部分连接）/集群DN数目。

返回类型：bigint

nextval函数有两种调用方式（其中第二种调用方式兼容Oracle的语法，目前不支持Sequence命名中有特殊字符"."的情况），如下：

示例1：

```
SELECT nextval('seqDemo');
 nextval
---------
       2
(1 row)
```

示例2：

```
SELECT seqDemo.nextval;
 nextval
---------
       2
(1 row)
```

currval(regclass)

返回当前会话里最近一次nextval返回的指定的sequence的数值。如果当前会话还没有调用过指定的sequence的nextval，那么调用currval将会报错。需要注意的是，这个函数在默认情况下是不支持的，需要通过设置enable_beta_features为true之后，才能使用这个函数。同时在设置enable_beta_features为true之后，nextval()函数将不支持下推。

 版权所有 © 华为云计算技术有限公司

返回类型：bigint

currval函数有两种调用方式（其中第二种调用方式兼容Oracle的语法，目前不支持Sequence命名中有特殊字符"."的情况），如下：

示例1：

```
SELECT currval('seq1');
 currval
---------
       2
(1 row)
```

示例2：

```
SELECT seq1.currval seq1;
 currval
---------
       2
(1 row)
```

lastval()

描述：返回当前会话里最近一次nextval返回的数值。这个函数等效于currval，只是它不用序列名为参数，它抓取当前会话里面最近一次nextval使用的序列。如果当前会话还没有调用过nextval，那么调用lastval将会报错。

需要注意的是，这个函数在默认情况下是不支持的，需要通过设置enable_beta_features或者lastval_supported为true之后，才能使用这个函数。同时这种情况下，nextval()函数将不支持下推。

返回类型：bigint

示例：

```
SELECT lastval();
 lastval
---------
       2
(1 row)
```

setval(regclass, bigint)

描述：设置序列的当前数值。

返回类型：bigint

示例：

```
SELECT setval('seqDemo',1);
 setval
--------
      1
(1 row)
```

setval(regclass, bigint, boolean)

描述：设置序列的当前数值以及is_called标志。

返回类型：bigint

示例：

 版权所有 © 华为云计算技术有限公司

```
SELECT setval('seqDemo',1,true);
 setval
--------
      1
(1 row)
```

说明

Setval后当前会话及GTM上会立刻生效，但如果其他会话有缓存的序列值，只能等到缓存值用尽才能感知Setval的作用。所以为了避免序列值冲突，setval要谨慎使用。

因为序列是非事务的，setval造成的改变不会由于事务的回滚而撤销。

7.7 数组函数和操作符

7.7.1 数组操作符

数组比较是使用默认的B-tree比较函数对所有元素逐一进行比较的。多维数组的元素按照行顺序进行访问。如果两个数组的内容相同但维数不等，决定排序顺序的首要因素是维数。

DWS支持的数组操作符如表7-10所示

表 7-10 数组操作符

分类	功能	对应操作符	简单示例
比大小	比较两个数组是否相等。	=	SELECT ARRAY[1.1,2.1,3.1]::int[] = ARRAY[1,2,3] AS RESULT;
	判断两个数组是否不相等。	<>	SELECT ARRAY[1,2,3] <> ARRAY[1,2,4] AS RESULT;
	判断一个数组是否小于另一个数组。	<	SELECT ARRAY[1,2,3] < ARRAY[1,2,4] AS RESULT;
	判断一个数组是否大于另一个数组。	>	SELECT ARRAY[1,4,3] > ARRAY[1,2,4] AS RESULT;
	判断一个数组是否小于或等于另一个数组。	<=	SELECT ARRAY[1,2,3] <= ARRAY[1,2,3] AS RESULT;
	判断一个数组是否大于或等于另一个数组。	>=	SELECT ARRAY[1,4,3] >= ARRAY[1,4,3] AS RESULT;
包含或重叠	判断一个数组是否包含另一个数组。	@>	SELECT ARRAY[1,4,3] @> ARRAY[3,1] AS RESULT;

版权所有 © 华为云计算技术有限公司

分类	功能	对应操作符	简单示例
	判断一个数组是否被包含于另一个数组。	<@	SELECT ARRAY[2,7] <@ ARRAY[1,7,4,2,6] AS RESULT;
	判断一个数组是否和另一个数组重叠（有共同元素）。	&&	SELECT ARRAY[1,4,3] && ARRAY[2,1] AS RESULT;
连接	数组与数组进行连接。	||	SELECT ARRAY[1,2,3] || ARRAY[4,5,6] AS RESULT;
	元素与数组进行连接。	||	SELECT 3 || ARRAY[4,5,6] AS RESULT;
	数组与元素进行连接。	||	SELECT ARRAY[4,5,6] || 7 AS RESULT;

=

描述：两个数组是否相等

示例：

```
SELECT ARRAY[1.1,2.1,3.1]::int[] = ARRAY[1,2,3] AS RESULT;
 result
--------
 t
(1 row)
```

<>

描述：两个数组是否不相等

示例：

```
SELECT ARRAY[1,2,3] <> ARRAY[1,2,4] AS RESULT;
 result
--------
 t
(1 row)
```

<

描述：一个数组是否小于另一个数组

示例：

```
SELECT ARRAY[1,2,3] < ARRAY[1,2,4] AS RESULT;
 result
--------
 t
(1 row)
```

 版权所有 © 华为云计算技术有限公司

>

描述：一个数组是否大于另一个数组

示例：

```
SELECT ARRAY[1,4,3] > ARRAY[1,2,4] AS RESULT;
 result
--------
 t
(1 row)
```

<=

描述：一个数组是否小于或等于另一个数组

示例：

```
SELECT ARRAY[1,2,3] <= ARRAY[1,2,3] AS RESULT;
 result
--------
 t
(1 row)
```

>=

描述：一个数组是否大于或等于另一个数组

示例：

```
SELECT ARRAY[1,4,3] >= ARRAY[1,4,3] AS RESULT;
 result
--------
 t
(1 row)
```

@>

描述：一个数组是否包含另一个数组

示例：

```
SELECT ARRAY[1,4,3] @> ARRAY[3,1] AS RESULT;
 result
--------
 t
(1 row)
```

<@

描述：一个数组是否被包含于另一个数组

示例：

```
SELECT ARRAY[2,7] <@ ARRAY[1,7,4,2,6] AS RESULT;
 result
--------
 t
(1 row)
```

&&

描述：一个数组是否和另一个数组重叠（有共同元素）

 版权所有 © 华为云计算技术有限公司

示例：

```
SELECT ARRAY[1,4,3] && ARRAY[2,1] AS RESULT;
 result
--------
 t
(1 row)
```

||

描述：数组与数组进行连接

示例：

```
SELECT ARRAY[1,2,3] || ARRAY[4,5,6] AS RESULT;
    result
---------------
 {1,2,3,4,5,6}
(1 row)
SELECT ARRAY[1,2,3] || ARRAY[[4,5,6],[7,8,9]] AS RESULT;
          result
-----------------------------
 {{1,2,3},{4,5,6},{7,8,9}}
(1 row)
```

||

描述：元素与数组进行连接

示例：

```
SELECT 3 || ARRAY[4,5,6] AS RESULT;
  result
-----------
 {3,4,5,6}
(1 row)
```

||

描述：数组与元素进行连接

示例：

```
SELECT ARRAY[4,5,6] || 7 AS RESULT;
  result
-----------
 {4,5,6,7}
(1 row)
```

7.7.2 数组函数

数组操作函数对输入的数组数据进行一些基本操作，如添加元素、查找元素，转换数组并返回操作后的数组数据。

DWS支持的数组函数参考**表7-11**。

表 7-11 数组函数

函数名称	功能	使用示例
array_append(anyarray, anyelement)	向数组末尾添加元素。	SELECT array_append(ARRAY[1,2], 3) AS RESULT;

版权所有 © 华为云计算技术有限公司

函数名称	功能	使用示例
array_prepend(anyelement, anyarray)	向数组开头添加元素	SELECT array_append(ARRAY[1,2], 3) AS RESULT;
array_cat(anyarray, anyarray)	连接两个数组。	SELECT array_cat(ARRAY[1,2,3], ARRAY[4,5]) AS RESULT;
array_ndims(anyarray)	返回数组的维数。	SELECT array_dims(ARRAY[[1,2,3], [4,5,6]]) AS RESULT;
array_dims(anyarray)	返回数组维数的文本表示。	SELECT array_dims(ARRAY[[1,2,3], [4,5,6]]) AS RESULT;
array_length(anyarray, int)	返回数组维度的长度。	SELECT array_length(array[1,2,3], 1) AS RESULT;
array_lower(anyarray, int)	返回数组维数的下界。	SELECT array_lower('[0:2]={1,2,3}'::int[], 1) AS RESULT;
array_upper(anyarray, int)	返回数组维数的上界。	SELECT array_upper(ARRAY[1,8,3,7], 1) AS RESULT;
array_to_string(anyarray, text [, text])	将数组转换为字符串。	SELECT array_to_string(ARRAY[1, 2, 3, NULL, 5], ',', '*') AS RESULT;
string_to_array(text, text [, text])	将字符串转换为数组。	SELECT string_to_array('xx~^~yy~^~zz', '~^~', 'yy') AS RESULT;
array_flatten(anyarray)	将多维数组转换为一维数组。	SELECT array_flatten(ARRAY[1, 2], [3,4]]) AS RESULT;
array_distinct(anyarray)	返回去重之后的数组，不统计NULL值。	SELECT array_distinct(ARRAY[1, 2, 3, 1, NULL, 5]) AS RESULT;
array_uniq(anyarray)	返回去重之后的数组元素个数。	SELECT array_uniq(ARRAY[1, 2, 3, 1, NULL, 5]) AS RESULT;
array_sort(anyarray)	返回按升序排列后的数组。	SELECT array_sort(ARRAY[1, 2, 3, 1, NULL, 5]) AS RESULT;
array_reversesort(anyarray)	返回按降序排列后的数组。	SELECT array_reversesort(ARRAY[1, 2, 3, 1, NULL, 5]) AS
unnest(anyarray)	扩展一个数组为一组行。	SELECT unnest(ARRAY[1,2]) AS RESULT;
interval(N, N1, N2, N3 ...)	从输入的整数数组中，查找返回最后一个小于等于目标参数n的数组索引。	SELECT INTERVAL(10, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 10, 11) AS RESULT;
split(text, text)	将字符串按分隔符进行分隔，并返回数组。	SELECT SPLIT('a-b-c-d-e', '-') AS RESULT;

版权所有 © 华为云计算技术有限公司

array_append(anyarray, anyelement)

描述：向数组末尾添加元素，只支持一维数组。

返回类型：anyarray

示例：向数组[1,2]末尾添加元素3。

```
SELECT array_append(ARRAY[1,2], 3) AS RESULT;
 result
---------
 {1,2,3}
(1 row)
```

array_prepend(anyelement, anyarray)

描述：向数组开头添加元素，只支持一维数组。

返回类型：anyarray

示例：向数组[2,3]开头添加元素1。

```
SELECT array_prepend(1, ARRAY[2,3]) AS RESULT;
 result
---------
 {1,2,3}
(1 row)
```

array_cat(anyarray, anyarray)

描述：连接两个数组，支持多维数组。

返回类型：anyarray

示例：连接数组[1,2,3]和数组[4,5]。

```
SELECT array_cat(ARRAY[1,2,3], ARRAY[4,5]) AS RESULT;
   result
-------------
 {1,2,3,4,5}
(1 row)
```

连接二维数组[[1,2],[4,5]]和一维数组[6,7]。

```
SELECT array_cat(ARRAY[[1,2],[4,5]], ARRAY[6,7]) AS RESULT;
       result
---------------------
 {{1,2},{4,5},{6,7}}
(1 row)
```

array_ndims(anyarray)

描述：返回数组的维数。

返回类型：int

示例：查看数组[[1,2,3], [4,5,6]]的维数。

```
SELECT array_ndims(ARRAY[[1,2,3], [4,5,6]]) AS RESULT;
 result
--------
      2
(1 row)
```

 版权所有 © 华为云计算技术有限公司

array_dims(anyarray)

描述：返回数组维数的文本表示。

返回类型：text

示例：获取数组的维度，返回结果表示该数组是一个2行3列的二维数组。

```
SELECT array_dims(ARRAY[[1,2,3], [4,5,6]]) AS RESULT;
   result
------------
 [1:2][1:3]
(1 row)
```

array_length(anyarray, int)

描述：返回数组维度的长度。

返回类型：int

示例：计算数组[1,2,3]在第一维的长度。

```
SELECT array_length(array[1,2,3], 1) AS RESULT;
 result
--------
      3
(1 row)
```

array_lower(anyarray, int)

描述：返回数组维数的下界。

返回类型：int

示例：查询指定范围的数组[0:2]={1,2,3}，在第一维的下界。

```
SELECT array_lower('[0:2]={1,2,3}'::int[], 1) AS RESULT;
 result
--------
      0
(1 row)
```

array_upper(anyarray, int)

描述：返回数组维数的上界。

返回类型：int

示例：查询数组 [1, 8, 3, 7]在第一维的上界。

```
SELECT array_upper(ARRAY[1,8,3,7], 1) AS RESULT;
 result
--------
      4
(1 row)
```

array_to_string(anyarray, text [, text])

描述：将数组转换为字符串，并用指定的分隔符连接，同时支持定义如何处理数组中的NULL。

第一个text作为数组的新分隔符，第二个text替换数组值为NULL的值。

 版权所有 © 华为云计算技术有限公司

返回类型：text

示例：将数组[1, 2, 3, NULL, 5]转换为字符串，用“,”分隔元素，并用“*”替换NULL值。

```
SELECT array_to_string(ARRAY[1, 2, 3, NULL, 5], ',', '*') AS RESULT;
  result
-----------
 1,2,3,*,5
(1 row)
```

说明

该函数的第二个参数text，控制如何处理数组中的NULL元素。如果省略此参数或显式指定为NULL，函数将忽略数组中所有NULL元素，且不会出现在最终输出的字符串中。

例如：

省略第二个参数text的场景。

```
SELECT array_to_string(ARRAY[1, NULL, 3, NULL, 5], ',') AS result;
 result
--------
 1,3,5
(1 row)
```

第二个参数text显式指定为NULL场景。

```
SELECT array_to_string(ARRAY[1, NULL, 3, NULL, 5], ',',NULL) AS RESULT;
 result
--------
 1,3,5
(1 row)
```

string_to_array(text, text [, text])

描述：将字符串按指定的分隔符拆分成数组，同时支持替换特定值为NULL。

第二个text为指定的分隔符。第三个可选的text替换特定值为NULL，如果分隔后的子串与第三个可选的text完全匹配，则将其替换为NULL。

返回类型：text[]

示例：将字符串xx~^~yy~^~zz按照分隔符“~^~”拆分成数组，并替换 'yy' 值为NULL。

```
SELECT string_to_array('xx~^~yy~^~zz', '~^~', 'yy') AS RESULT;
    result
--------------
 {xx,NULL,zz}
(1 row)
```

字符串xx~^~yy~^~zz分隔后的子串yy与参数“y”未完全匹配，不会将其替换为NULL。

```
SELECT string_to_array('xx~^~yy~^~zz', '~^~', 'y') AS RESULT;
   result
------------
 {xx,yy,zz}
(1 row)
```

 版权所有 © 华为云计算技术有限公司

说明

- 该函数的第三个参数text，控制替换特定值为NULL。如果省略此参数或显式指定为NULL，函数会将原字符串中连续两个分隔符之间的空子串替换为NULL。

  例如：

  省略第三个参数text的场景。

  ```
  SELECT string_to_array('a,,b', ',') AS result;
   result
  ----------
   {a,"",b}
  (1 row)
  ```

  第三个参数text显式指定为NULL场景。

  ```
  SELECT string_to_array('a,,b', ',', NULL) AS result;
   result
  ----------
   {a,"",b}
  (1 row)
  ```

- 在string_to_array中，如果分隔符参数是NULL，输入字符串中的每个字符将在结果数组中变成一个独立的元素。如果分隔符是一个空字符串，则整个输入的字符串将变为一个元素的数组。否则输入字符串将在每个分隔字符串处分开。

  例如：

  分隔符参数是NULL的场景。

  ```
  SELECT string_to_array('abc', NULL) AS result;
   result
  ---------
   {a,b,c}
  (1 row)
  ```

  分隔符是一个空字符串的场景。

  ```
  SELECT string_to_array('abcde', ' ')AS result;
   result
  ---------
   {abcde}
  (1 row)
  ```

array_flatten(anyarray)

描述：将多维数组转换为一维数组。该函数仅9.1.1.100及以上集群版本支持。

返回类型：anyarray

示例：

```
SELECT array_flatten(ARRAY[1, 2], [3,4]]) AS RESULT;
  result
-----------
 1,2,3,4
(1 row)
```

array_distinct(anyarray)

描述：返回去重之后的数组，不统计NULL值。该函数仅9.1.1.100及以上集群版本支持。

返回类型：anyarray

示例：

```
SELECT array_distinct(ARRAY[1, 2, 3, 1, NULL, 5]) AS RESULT;
  result
-----------
```

 版权所有 © 华为云计算技术有限公司

```
 1,2,3,5
(1 row)
```

注意

array_distinct()、array_uniq()、array_sort()、array_reversesort()函数只支持一维数组。

array_uniq(anyarray)

描述：返回去重之后的数组元素个数。如果数组元素包含NULL，也会统计。该函数仅9.1.1.100及以上集群版本支持。

返回类型：int

示例：

```
SELECT array_uniq(ARRAY[1, 2, 3, 1, NULL, 5]) AS RESULT;
  result
---------
       5
(1 row)
```

array_sort(anyarray)

描述：返回按升序排列后的数组。如果数组元素包含NULL，则输出的排序数组会将NULL排在最后。该函数仅9.1.1.100及以上集群版本支持。

返回类型：anyarray

示例：

```
SELECT array_sort(ARRAY[1, 2, 3, 1, NULL, 5]) AS RESULT;
  result
---------
 1,1,2,3,5,NULL
(1 row)
```

array_reversesort(anyarray)

描述：返回按降序排列后的数组。如果数组元素包含NULL，则输出的排序数组会将NULL排在最后。该函数仅9.1.1.100及以上集群版本支持。

返回类型：anyarray

示例：

```
SELECT array_reversesort(ARRAY[1, 2, 3, 1, NULL, 5]) AS RESULT;
  result
---------
 5,3,2,1,1,NULL
(1 row)
```

unnest(anyarray)

描述：扩展一个数组为一组行。

返回类型：setof anyelement

示例：

 版权所有 © 华为云计算技术有限公司

```
SELECT unnest(ARRAY[1,2]) AS RESULT;
 result
--------
      1
      2
(2 rows)
```

unnest函数配合**string_to_array(text, text [, text])**数组使用。数组转列，先将字符串按逗号分割成数组，然后再把数组转成列：

```
SELECT unnest(string_to_array('a,b,c,d',',')) AS RESULT;
 result
--------
 a
 b
 c
 d
(4 rows)
```

interval(N, N1, N2, N3 ...)

描述：从输入的整数数组中，查找返回最后一个小于等于目标参数n的数组索引。

若n为NULL，返回-1。interval函数不支持interval(N, N1)的场景。

返回类型：int

示例：查找数字10位于后面一串数字划分的哪个区间，并返回对应的区间索引。

```
SELECT INTERVAL(10, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 10, 11) AS RESULT;
 result
--------
     11
(1 row)
```

split(text, text)

描述：将字符串按分隔符进行分隔，并返回数组。

第一个参数text为字符串，第二个参数text为分隔符。

返回类型：text[]

示例：将字符串a-b-c-d-e按照分隔符“-”拆分成多个部分，并返回一个数组。

```
SELECT SPLIT('a-b-c-d-e', '-') AS RESULT;
   result
-------------
 {a,b,c,d,e}
(1 row)
```

将字符串a-b-c-d-e按照分隔符“-”拆分并提取第四个元素。

```
SELECT SPLIT('a-b-c-d-e', '-')[4] AS RESULT;
 result
--------
 d
(1 row)
```

7.8 逻辑操作符

常用的逻辑操作符有AND、OR和NOT，其运算结果有三个值，分别为TRUE、FALSE和NULL，其中NULL代表未知。运算优先级顺序为：NOT>AND>OR。

 版权所有 © 华为云计算技术有限公司

运算规则请参见表7-12，表中的a和b代表逻辑表达式。

表 7-12 运算规则表

a	b	a AND b的结果	a OR b的结果	NOT a的结果
TRUE	TRUE	TRUE	TRUE	FALSE
TRUE	FALSE	FALSE	TRUE	FALSE
TRUE	NULL	NULL	TRUE	FALSE
FALSE	FALSE	FALSE	FALSE	TRUE
FALSE	NULL	FALSE	NULL	TRUE
NULL	NULL	NULL	NULL	NULL

说明

操作符AND和OR具有交换性，即交换左右两个操作数，不影响其结果。

7.9 比较操作符

在数据仓库的日常使用中，用户经常需要从海量数据中筛选出满足特定条件的记录。例如，在员工表中查找薪资高于5000元的员工、在订单表中筛选金额不等于0的有效订单、在产品表中查询库存量小于安全阈值的商品等。这些场景都离不开比较操作符的使用。

比较操作符是SQL查询中最基础也是最常用的条件判断工具，广泛应用于以下典型场景：

- **数据筛选**：在SELECT语句的WHERE子句中，使用比较操作符筛选满足条件的记录，如SELECT * FROM employees WHERE salary > 5000。
- **多条件组合**：在WHERE子句中结合AND、OR等逻辑操作符，使用多个比较操作符构建复合筛选条件。
- **数据更新与删除**：在UPDATE和DELETE语句中通过比较操作符精确定位需要操作的行。
- **连接条件**：在JOIN语句的ON子句中使用比较操作符指定表间关联条件。
- **子查询判断**：在子查询中使用比较操作符进行存在性判断或值比较。

了解比较操作符的使用方法、数据类型转换规则以及NULL值的特殊处理方式，有助于编写高效、准确的SQL查询语句。

比较操作符原理描述

比较操作符用于对两个表达式进行比较，并返回布尔类型（TRUE、FALSE或NULL）的结果。所有比较操作符都是双目操作符，即需要两个操作数参与运算。

被比较的两个数据类型必须是相同的数据类型，或者是可以进行隐式转换的类型。例如，1 < 2 < 3这样的表达式为非法的，因为1 < 2的结果是布尔值，布尔值和3之间不能做比较。

 版权所有 © 华为云计算技术有限公司

DWS提供的比较操作符请参见表7-13。

表 7-13 比较操作符

操作符	描述	语法示例
<	小于	查询薪资低于5000的员工： SELECT * FROM employees WHERE salary < 5000;
>	大于	查询薪资高于5000的员工： SELECT * FROM employees WHERE salary > 5000;
<=	小于或等于	查询薪资不超过5000的员工： SELECT * FROM employees WHERE salary <= 5000;
>=	大于或等于	查询薪资不低于5000的员工： SELECT * FROM employees WHERE salary >= 5000;
=	等于	查询部门编号为10的员工： SELECT * FROM employees WHERE department_id = 10;
<> 或 !=	不等于	查询部门编号不为10的员工： SELECT * FROM employees WHERE department_id <> 10; SELECT * FROM employees WHERE department_id != 10;

NULL 值比较行为

在SQL标准中，NULL表示“未知值”，与任何值（包括NULL自身）进行比较的结果均为NULL（即unknown），而不是TRUE或FALSE。

- NULL = 1 的结果为 NULL，而不是 FALSE。
- NULL = NULL 的结果也为 NULL，而不是 TRUE。
- NULL <> 1 的结果同样为 NULL。

在WHERE子句中，只有比较结果为TRUE的行才会被返回。结果为NULL或FALSE的行均不会被返回。因此，**不能使用=或<>操作符来判断一个值是否为NULL**，必须使用IS NULL或IS NOT NULL操作符。

7.10 模式匹配操作符

数据库提供了三种实现模式匹配的方法：SQL LIKE操作符、SIMILAR TO操作符和POSIX-风格的正则表达式。除了这些基本的操作符外，还有一些函数可用于提取或替换匹配子串并在匹配位置分离一个串。

 版权所有 © 华为云计算技术有限公司

LIKE

判断字符串是否能匹配上LIKE后的模式字符串。如果字符串与提供的模式匹配，则LIKE表达式返回为真（NOT LIKE表达式返回假），否则返回为假（NOT LIKE表达式返回真）。

- 匹配规则
  a. 此操作符只有在它的模式匹配整个串的时候才能成功。如果要匹配在串内任何位置的序列，该模式必须以百分号开头和结尾。
  b. 下划线（_）代表（匹配）任何单个字符； 百分号（%）代表任意串的通配符。
  c. 要匹配文本里的下划线（_）或者百分号（%），在提供的模式里相应字符必须前导逃逸字符。逃逸字符的作用是禁用元字符的特殊含义，缺省的逃逸字符是反斜线，也可以用ESCAPE子句指定一个不同的逃逸字符。
  d. 要匹配逃逸字符本身，需写两个逃逸字符。例如要写一个包含反斜线的模式常量，那就要在SQL语句里写两个反斜线。

     说明

     参数standard_conforming_strings设置为off时，在字符串常量中写的任何反斜线都需要被双写。因此写一个匹配单个反斜线的模式实际上要在语句里写四个反斜线。可通过用ESCAPE选择一个不同的逃逸字符来避免这种情况，这样反斜线就不再是LIKE的特殊字符了。但仍然是字符文本分析器的特殊字符，所以还是需要两个反斜线。也可通过写ESCAPE ''的方式不选择逃逸字符，这样可以有效地禁用逃逸机制，但是无法关闭下划线和百分号在模式中的特殊含义。
  e. 关键字ILIKE可以用于替换LIKE，区别是LIKE大小写敏感，ILIKE大小写不敏感。
  f. 操作符~~等效于LIKE，操作符~~*等效于ILIKE。
- 示例

```
SELECT 'abc' LIKE 'abc' AS RESULT;
 result
-----------
 t
(1 row)
SELECT 'abc' LIKE 'a%' AS RESULT;
 result
-----------
 t
(1 row)
SELECT 'abc' LIKE '_b_' AS RESULT;
 result
-----------
 t
(1 row)
SELECT 'abc' LIKE 'c' AS RESULT;
 result
-----------
 f
(1 row)
```

SIMILAR TO

SIMILAR TO操作符根据特定模式判断字符串匹配情况并返回真假值。SIMILAR TO在功能上与LIKE非常类似，只不过它使用SQL标准定义的正则表达式解析模式。

- 匹配规则
  a. 和LIKE一样，SIMILAR TO操作符只有在它的模式匹配整个串的时候才返回真。如果要匹配在串内任何位置的序列，该模式必须以百分号开头和结尾。

 版权所有 © 华为云计算技术有限公司

b. 下划线（ _ ）代表（匹配）任何单个字符； 百分号（ % ）代表任意串的通配符。

c. SIMILAR TO也支持下面这些从POSIX正则表达式借用的模式匹配元字符。

表 7-14 模式匹配元字符

元字符	含义
|	表示选择（两个候选之一）。
*	表示重复前面的项零次或更多次。
+	表示重复前面的项一次或更多次。
?	表示重复前面的项零次或一次。
{m}	表示重复前面的项刚好m次。
{m,}	表示重复前面的项m次或更多次。
{m,n}	表示重复前面的项至少m次并且不超过n次。
()	把多个项组合成一个逻辑项。
[...]	声明一个字符类，就像POSIX正则表达式一样。

d. 前导逃逸字符可以禁止所有这些元字符的特殊含义。逃逸字符的使用规则和LIKE一样。

- 注意事项

  如果SIMILAR TO正则表达式重复匹配字符数量非常庞大，由于受递归大小限制，执行语句会失败并报错invalid regular expression: regular expression is too complex，可尝试调大GUC参数**max_stack_depth**。

- 正则表达式函数

  支持使用函数**substring(string from pattern for escape)**截取匹配SQL正则表达式的子字符串。

- 示例

```
SELECT 'abc' SIMILAR TO 'abc' AS RESULT;
 result
------------
 t
(1 row)
SELECT 'abc' SIMILAR TO 'a' AS RESULT;
 result
------------
 f
(1 row)
SELECT 'abc' SIMILAR TO '%(b|d)%' AS RESULT;
 result
------------
 t
(1 row)
SELECT 'abc' SIMILAR TO '(b|c)%'  AS RESULT;
 result
------------
 f
(1 row)
```

 版权所有 © 华为云计算技术有限公司

POSIX 正则表达式

正则表达式是一个字符序列，它是定义一个串集合（一个正则集）的缩写。 如果一个串是正则表达式描述的正则集中的一员时， 那么就说这个串匹配该正则表达式。 POSIX正则表达式提供了比LIKE和SIMILAR TO操作符更强大的含义。表7-15列出了所有可用于POSIX正则表达式模式匹配的操作符。

表 7-15 正则表达式匹配操作符

操作符	描述	例子
~	匹配正则表达式，大小写敏感	'thomas' ~ '.*thomas.*'
~*	匹配正则表达式，大小写不敏感	'thomas' ~* '.*Thomas.*'
!~	不匹配正则表达式，大小写敏感	'thomas' !~ '.*Thomas.*'
!~*	不匹配正则表达式，大小写不敏感	'thomas' !~* '.*vadim.*'

- 匹配规则
  a. 与LIKE不同，正则表达式允许匹配串里的任何位置，除非该正则表达式显式地挂接在串的开头或者结尾。
  b. 除了上文提到的元字符外， POSIX正则表达式还支持下表的模式匹配元字符。

  表 7-16 模式匹配元字符

元字符	含义
^	表示串开头的匹配。
$	表示串末尾的匹配。
.	匹配任意单个字符。

- 正则表达式函数
  POSIX正则表达式支持下面函数。
  - **substring(string from pattern)**函数提供了抽取一个匹配POSIX正则表达式模式的子串的方法。
  - **regexp_replace(string, pattern, replacement [,flags])**函数提供了将匹配POSIX正则表达式模式的子串替换为新文本的功能。
  - **regexp_matches(string text, pattern text [, flags text])**函数返回一个文本数组，该数组由匹配一个POSIX正则表达式模式得到的所有被捕获子串构成。
  - **regexp_split_to_table(string text, pattern text [, flags text])**函数把一个POSIX正则表达式模式当作一个定界符来分离一个串。

版权所有 © 华为云计算技术有限公司

- **regexp_split_to_array(string text, pattern text [, flags text])**和regexp_split_to_table类似，是一个正则表达式分离函数，不过它的结果以一个text数组的形式返回。

  说明

  正则表达式分离函数会忽略零长度的匹配，这种匹配发生在串的开头或结尾或者正好发生在前一个匹配之后。这和正则表达式匹配的严格定义是相悖的，后者由regexp_matches实现，但是通常前者是实际中最常用的行为。

- 示例

```
 SELECT 'abc' ~ 'Abc' AS RESULT;
result
--------
 f
(1 row)
SELECT 'abc' ~* 'Abc' AS RESULT;
 result
--------
 t
(1 row)
SELECT 'abc' !~ 'Abc' AS RESULT;
 result
--------
 t
(1 row)
SELECT 'abc'!~* 'Abc' AS RESULT;
 result
--------
 f
(1 row)
SELECT 'abc' ~ '^a' AS RESULT;
 result
--------
 t
(1 row)
SELECT 'abc' ~ '(b|d)'AS RESULT;
 result
--------
 t
(1 row)
SELECT 'abc' ~ '^(b|c)'AS RESULT;
 result
--------
 f
(1 row)
```

  虽然大部分的正则表达式搜索都能很快地执行，但是正则表达式仍可能被人为地控制，通过任意长的时间和任意量的内存进行处理。不建议接受非安全来源的正则表达式搜索模式，如果必须这样做，建议加上语句超时限制。使用SIMILAR TO模式的搜索具有同样的安全性危险， 因为SIMILAR TO提供了很多和POSIX-风格正则表达式相同的能力。LIKE搜索比其他两种选项简单得多，因此在接受非安全来源的搜索模式时要更安全些。

7.11 聚集函数

聚集函数是数据库中常用的函数，它将多行数据合并计算为单个统计值，帮助用户快速从大量数据中获取关键数据概览，如总数、平均数、最大值和去重计数等，广泛应用于数据报表和业务分析等场景。

 版权所有 © 华为云计算技术有限公司

表 7-17 聚集函数

函数分类	函数名称	功能
基础统计计算	sum(expression)	计算表达式的总和。
	max(expression)	返回表达式的最大值。
	min(expression)	返回表达式的最小值。
	avg(expression)	计算表达式的平均值。
	count(expression)	返回表中非空（NOT NULL）的行数。
	count(*)	返回表中的总行数（包括含NULL值的行）。
分位数与分布分析	median(expression)	计算中位数。
	percentile_cont(const) within group(order by expression)	连续分布百分位数。通过线性插值计算指定百分位的值。
	percentile_disc(const) within group(order by expression)	离散百分位数。返回数据集中最接近指定百分位的实际值。
集合聚合与字符串拼接	array_agg(expression)	将多行值聚合到一个数组中。
	string_agg(expression, delimiter)	将多行字符串用指定分隔符连接成一个字符串。
	listagg(expression [, delimiter]) WITHIN GROUP(ORDER BY order-list)	将多行字符串按指定顺序连接成一个字符串。
	group_concat(expression [ORDER BY {col_name | expr} [ASC | DESC]] [SEPARATOR str_val])	将分组中的多行字符串值连接成一个字符串。
统计学高级分析	var_pop(expression)	计算总体方差。
	var_samp(expression)	计算样本方差。
	variance(expression)	计算样本方差。
	stddev_pop(expression)	计算总体标准差。
	stddev_samp(expression)	计算样本标准差。

版权所有 © 华为云计算技术有限公司

函数分类	函数名称	功能
	stddev(expression)	计算样本标准差。
	covar_pop(Y, X)	计算Y和X的总体协方差。
	covar_samp(Y, X)	计算Y和X的样本协方差。
	corr(Y, X)	计算Y和X的相关系数。
线性回归分析	regr_avgx(Y, X)	计算X的平均值。
	regr_avgy(Y, X)	计算Y的平均值。
	regr_count(Y, X)	计算非空对 (X, Y) 的数量。
	regr_r2(Y, X)	计算相关系数的平方。
	regr_slope(Y, X)	根据输入点(X, Y)拟合成一个线性方程，然后计算该直线的斜率。
	regr_intercept(Y, X)	根据输入点(X, Y)拟合成一个线性方程，然后计算该直线的Y轴截距。
	regr_sxx(Y, X)	计算X的平方和。
	regr_syy(Y, X)	计算Y的平方和。
	regr_sxy(Y, X)	计算X和Y的乘积和。
逻辑与位运算聚合	bit_and(expression)	对所有行的值进行按位与（AND）运算。
	bit_or(expression)	对所有行的值进行按位或（OR）运算。
	bool_and(expression)	如果所有输入值都是真，则返回真（TRUE）。
	bool_or(expression)	如果所有输入值只要有一个为真，则返回真（TRUE）
	every(expression)	等效于bool_and()，如果所有输入值都是真，则返回真（TRUE）。
去重与校验	checksum(expression)	计算数据的校验和，通过对比结果判断操作前后表中的数据是否一致。
	approx_count_distinct(col_name)	近似计算唯一值的数量
	UNIQ(col_name)	计算唯一值的数量。

sum(expression)

描述：计算所有输入行的expression总和。

返回类型：

 版权所有 © 华为云计算技术有限公司

通常情况下输入数据类型和输出数据类型是相同的，但以下情况会发生类型转换：

- 对于SMALLINT或INT输入，输出类型为BIGINT。
- 对于BIGINT输入，输出类型为NUMBER 。
- 对于浮点数输入，输出类型为DOUBLE PRECISION。

示例：

```
SELECT SUM(ss_ext_tax) FROM tpcds.STORE_SALES;
  sum
--------------
 213267594.69
(1 row)
```

max(expression)

描述：返回所有输入行中expression的最大值。

参数类型：任意数组、数值、字符串、日期/时间类型。

返回类型：与参数数据类型相同。

示例：

```
SELECT MAX(inv_quantity_on_hand) FROM tpcds.inventory;
   max
---------
 1000000
(1 row)
```

min(expression)

描述：返回所有输入行中expression的最小值。

参数类型：任意数组、数值、字符串、日期/时间类型。

返回类型：与参数数据类型相同。

示例：

```
SELECT MIN(inv_quantity_on_hand) FROM tpcds.inventory;
 min
-----
   0
(1 row)
```

avg(expression)

描述：计算所有输入值的算数平均值。

返回类型：

- 对于任何整数类型输入，结果都是NUMBER类型。
- 对于任何浮点输入，结果都是DOUBLE PRECISION类型。
- 其他，和输入数据类型相同。

示例：

```
SELECT AVG(inv_quantity_on_hand) FROM tpcds.inventory;
        avg
----------------------
```

 版权所有 © 华为云计算技术有限公司

```
 500.0387129084044604
(1 row)
```

median(expression)

描述：计算所有输入值的中位数。当前仅支持数值类型和interval类型。空值不参与计算。

返回类型：

- 对于任何整型数据输入，结果都是NUMERIC类型。否则与输入数据类型相同。
- Teradata兼容模式下，如果输入为整型，则返回的数据精度只有整数位。

示例：

```
SELECT MEDIAN(inv_quantity_on_hand) FROM tpcds.inventory;
 median
--------
    500
(1 row)
```

percentile_cont(const) within group(order by expression)

描述：返回一个对应于目标列排序中指定分位数的值，如有必要就在相邻的输入项之间插入值。空值不参与计算。

参数说明：

- const，指定百分位数，取值范围为0~1之间的数值。
- expression，指定排序依据的列或表达式，当前仅支持数值类型和interval类型。

返回类型：

- 对于任何整型数据输入，结果都是NUMERIC类型。否则与输入数据类型相同。
- Teradata兼容模式下，如果输入为整型，则返回的数据精度只有整数位。

示例：

```
SELECT percentile_cont(0.3) within group(order by x) FROM (SELECT generate_series(1,5) AS x) AS t;
percentile_cont
-----------------
2.2
(1 row)
SELECT percentile_cont(0.3) within group(order by x desc) FROM (SELECT generate_series(1,5) AS x) AS t;
percentile_cont
-----------------
3.8
(1 row)
```

percentile_disc(const) within group(order by expression)

描述：返回第一个在排序中位置等于或者超过指定分数的输入值。

参数说明：

- const，指定百分位数，取值范围为0~1之间的数值。
- expression，指定排序依据的列或表达式，当前仅支持数值类型和interval类型。空值不参与计算。

返回类型：对于任何整型数据输入，结果都是NUMERIC类型。否则，与输入数据类型相同。

 版权所有 © 华为云计算技术有限公司

示例：

```
SELECT percentile_disc(0.3) within group(order by x) FROM (SELECT generate_series(1,5) AS x) AS t;
percentile_disc
-----------------
2
(1 row)
SELECT percentile_disc(0.3) within group(order by x desc) FROM (SELECT generate_series(1,5) AS x) AS t;
percentile_disc
-----------------
4
(1 row)
```

count(expression)

描述：返回表中expression不为NULL的行数。

返回类型：BIGINT

示例：

```
SELECT COUNT(inv_quantity_on_hand) FROM tpcds.inventory;
  count
----------
 11158087
(1 row)
```

count(*)

描述：返回表中的总行数（包括含NULL值的行）。

返回类型：BIGINT

示例：

```
SELECT COUNT(*) FROM tpcds.inventory;
  count
----------
 11745000
(1 row)
```

array_agg(expression)

描述：将所有输入值（包括空）连接成一个数组。

9.1.1.300及以上集群版本该函数入参支持数组类型，最大支持6维数组。当入参为数组时，要求数组的维度一致，聚合后结果将提升1维。

返回类型：参数类型的数组。

示例：

创建表employeeinfo，并插入数据：

```
CREATE TABLE employeeinfo (empno smallint, ename varchar(20), job varchar(20), projects varchar(20)[],
hiredate date,deptno smallint);
INSERT INTO employeeinfo VALUES (7155, 'JACK', 'SALESMAN', '{ "ERP", "CRM" }', '2018-12-01', 30);
INSERT INTO employeeinfo VALUES (7003, 'TOM', 'FINANCE', '{ "Tax_Refactor","Audit" }', '2016-06-15', 20);
INSERT INTO employeeinfo VALUES (7357, 'MAX', 'SALESMAN', '{ "HR_System", "ERP" }', '2020-10-01', 30);

SELECT * FROM employeeinfo;
 empno | ename |   job    |       projects       |      hiredate       | deptno
-------+-------+----------+----------------------+---------------------+--------
  7155 | JACK  | SALESMAN | {ERP,CRM}            | 2018-12-01 00:00:00 |     30
  7003 | TOM   | FINANCE  | {Tax_Refactor,Audit} | 2016-06-15 00:00:00 |     20
```

 版权所有 © 华为云计算技术有限公司

```
  7357 | MAX   | SALESMAN | {HR_System,ERP}      | 2020-10-01 00:00:00 |     30
(3 rows)
```

查询部门编号为30的所有员工姓名。

```
SELECT array_agg(ename) FROM employeeinfo where deptno = 30;
 array_agg
------------
 {JACK,MAX}
(1 row)
```

查询属于同一个部门的所有员工。

```
SELECT deptno, array_agg(ename) FROM employeeinfo group by deptno;
 deptno | array_agg
--------+------------
     30 | {JACK,MAX}
     20 | {TOM}
(2 rows)

SELECT distinct array_agg(ename) OVER (PARTITION BY deptno) FROM employeeinfo;
 array_agg
------------
 {TOM}
 {JACK,MAX}
(2 rows)
```

查询所有的部门编号且去重。

```
SELECT array_agg(distinct deptno) FROM employeeinfo group by deptno;
 array_agg
-----------
 {20}
 {30}
(2 rows)
```

查询所有的部门编号，去重后按降序排列。

```
SELECT array_agg(distinct deptno order by deptno desc) FROM employeeinfo;
 array_agg
-----------
 {30,20}
(1 row)
```

查询部门编号为30的所有项目名称。

```
SELECT array_agg(projects) FROM employeeinfo where deptno = 30;
          array_agg
-----------------------------
 {{ERP,CRM},{HR_System,ERP}}
(1 row)
```

查询属于同一个部门的项目。

```
SELECT deptno, array_agg(projects) FROM employeeinfo group by deptno;
 deptno |           array_agg
--------+-------------------------------
     20 | {{Tax_Refactor,Audit}}
     30 | {{ERP,CRM},{HR_System,ERP}}
(2 rows)
```

查询所有的项目列表并按部门降序排列。

```
SELECT array_agg(projects order by deptno desc) FROM employeeinfo;
                    array_agg
--------------------------------------------------
 {{ERP,CRM},{HR_System,ERP},{Tax_Refactor,Audit}}
(1 row)
```

 版权所有 © 华为云计算技术有限公司

string_agg(expression, delimiter)

描述：将输入值用指定的分隔符连接成一个字符串。

返回类型：和参数数据类型相同。

示例：

基于创建的表employeeinfo，查询属于同一个部门的所有员工：

```
SELECT deptno, string_agg(ename,',') FROM employeeinfo group by deptno;
 deptno | string_agg
--------+------------
     30 | JACK,MAX
     20 | TOM
(2 rows)
```

查询工号小于7156的员工：

```
SELECT string_agg(ename,',') FROM employeeinfo where empno < 7156;
 string_agg
------------
 TOM,JACK
(1 row)
```

listagg(expression [, delimiter]) WITHIN GROUP(ORDER BY order-list)

描述：将聚集列数据按WITHIN GROUP指定的排序方式排列，并用delimiter指定的分隔符拼接成一个字符串。

参数说明：

- expression：必选。指定聚集列名或基于列的有效表达式，不支持DISTINCT关键字和VARIADIC参数。
- delimiter：可选。指定分隔符，可以是字符串常数或基于分组列的确定性表达式，缺省时表示分隔符为空。
- order-list：必选。指定分组内的排序方式。

返回类型：text

说明

listagg是兼容Oracle 11g2的列转行聚集函数，可以指定OVER子句用作窗口函数。为了避免与函数本身WITHIN GROUP子句的ORDER BY造成二义性，listagg用作窗口函数时，OVER子句不支持ORDER BY的窗口排序或窗口框架。

示例：

聚集列是文本字符集类型：

```
SELECT deptno, listagg(ename, ',') WITHIN GROUP(ORDER BY ename) AS employees FROM emp GROUP
BY deptno;
 deptno |              employees
--------+---------------------------------------
     10 | CLARK,KING,MILLER
     20 | ADAMS,FORD,JONES,SCOTT,SMITH
     30 | ALLEN,BLAKE,JAMES,MARTIN,TURNER,WARD
(3 rows)
```

聚集列是整型：

```
SELECT deptno, listagg(mgrno, ',') WITHIN GROUP(ORDER BY mgrno NULLS FIRST) AS mgrnos FROM emp
GROUP BY deptno;
```


版权所有 © 华为云计算技术有限公司

```
 deptno |            mgrnos
--------+------------------------------
     10 | 7782,7839
     20 | 7566,7566,7788,7839,7902
     30 | 7698,7698,7698,7698,7698,7839
(3 rows)
```

聚集列是浮点类型：

```
SELECT job, listagg(bonus, '($); ') WITHIN GROUP(ORDER BY bonus DESC) || '($)' AS bonus FROM emp
GROUP BY job;
   job     |                      bonus
-----------+-------------------------------------------------
 CLERK     | 10234.21($); 2000.80($); 1100.00($); 1000.22($)
 PRESIDENT | 23011.88($)
 ANALYST   | 2002.12($); 1001.01($)
 MANAGER   | 10000.01($); 2399.50($); 999.10($)
 SALESMAN  | 1000.01($); 899.00($); 99.99($); 9.00($)
(5 rows)
```

聚集列是时间类型：

```
SELECT deptno, listagg(hiredate, ', ') WITHIN GROUP(ORDER BY hiredate DESC) AS hiredates FROM emp
GROUP BY deptno;
 deptno |                                                 hiredates
--------
+-----------------------------------------------------------------------------------------------------------------------------
--
     10 | 1982-01-23 00:00:00, 1981-11-17 00:00:00, 1981-06-09 00:00:00
     20 | 2001-04-02 00:00:00, 1999-12-17 00:00:00, 1987-05-23 00:00:00, 1987-04-19 00:00:00, 1981-12-03
00:00:00
     30 | 2015-02-20 00:00:00, 2010-02-22 00:00:00, 1997-09-28 00:00:00, 1981-12-03 00:00:00, 1981-09-08
00:00:00, 1981-05-01 00:00:00
(3 rows)
```

聚集列是时间间隔类型：

```
SELECT deptno, listagg(vacationTime, '; ') WITHIN GROUP(ORDER BY vacationTime DESC) AS vacationTime
FROM emp GROUP BY deptno;
 deptno |                                vacationtime
--------+-------------------------------------------------------------------------------
     10 | 1 year 30 days; 40 days; 10 days
     20 | 70 days; 36 days; 9 days; 5 days
     30 | 1 year 1 mon; 2 mons 10 days; 30 days; 12 days 12:00:00; 4 days 06:00:00; 24:00:00
(3 rows)
```

分隔符缺省时，默认为空：

```
SELECT deptno, listagg(job) WITHIN GROUP(ORDER BY job) AS jobs FROM emp GROUP BY deptno;
 deptno |                  jobs
--------+-----------------------------------------
     10 | CLERKMANAGERPRESIDENT
     20 | ANALYSTANALYSTCLERKCLERKMANAGER
     30 | CLERKMANAGERSALESMANSALESMANSALESMANSALESMAN
(3 rows)
```

listagg作为窗口函数时，OVER子句不支持ORDER BY的窗口排序，listagg列为对应分组的有序聚集：

```
SELECT deptno, mgrno, bonus, listagg(ename,'; ') WITHIN GROUP(ORDER BY hiredate) OVER(PARTITION
BY deptno) AS employees FROM emp;
 deptno | mgrno |  bonus   |              employees
--------+-------+----------+-----------------------------------------
     10 |  7839 | 10000.01 | CLARK; KING; MILLER
     10 |       | 23011.88 | CLARK; KING; MILLER
     10 |  7782 | 10234.21 | CLARK; KING; MILLER
     20 |  7566 |  2002.12 | FORD; SCOTT; ADAMS; SMITH; JONES
     20 |  7566 |  1001.01 | FORD; SCOTT; ADAMS; SMITH; JONES
     20 |  7788 |  1100.00 | FORD; SCOTT; ADAMS; SMITH; JONES
     20 |  7902 |  2000.80 | FORD; SCOTT; ADAMS; SMITH; JONES
```

 版权所有 © 华为云计算技术有限公司

```
    20 |  7839 |   999.10 | FORD; SCOTT; ADAMS; SMITH; JONES
    30 |  7839 |  2399.50 | BLAKE; TURNER; JAMES; MARTIN; WARD; ALLEN
    30 |  7698 |     9.00 | BLAKE; TURNER; JAMES; MARTIN; WARD; ALLEN
    30 |  7698 |  1000.22 | BLAKE; TURNER; JAMES; MARTIN; WARD; ALLEN
    30 |  7698 |    99.99 | BLAKE; TURNER; JAMES; MARTIN; WARD; ALLEN
    30 |  7698 |  1000.01 | BLAKE; TURNER; JAMES; MARTIN; WARD; ALLEN
    30 |  7698 |   899.00 | BLAKE; TURNER; JAMES; MARTIN; WARD; ALLEN
(14 rows)
```

group_concat(expression [ORDER BY {col_name | expr} [ASC | DESC]] [SEPARATOR str_val])

描述：将列数据使用指定的str_val分隔符，按照ORDER BY子句指定的排序方式拼接成字符串，ORDER BY子句必须指定排序方式，不支持ORDER BY 1的写法。

参数说明：

- expression：必选。指定列名或基于列的有效表达式，不支持DISTINCT关键字和VARIADIC参数。
- str_val：可选。指定的分隔符，可以是字符串常数或基于分组列的确定性表达式。缺省时表示分隔符为逗号。

返回类型：text

示例：

默认分隔符为逗号：

```
SELECT group_concat(sname) FROM group_concat_test;
              group_concat
-------------------------------------------
 ADAMS,FORD,JONES,KING,MILLER,SCOTT,SMITH
(1 row)
```

group_concat函数支持自定义分隔符：

```
SELECT group_concat(sname separator ';') from group_concat_test;
              group_concat
-------------------------------------------
 ADAMS;FORD;JONES;KING;MILLER;SCOTT;SMITH
(1 row)
```

group_concat函数支持ORDER BY子句，将列数据进行有序拼接：

```
SELECT group_concat(sname order by snumber separator ';') FROM group_concat_test;
              group_concat
-------------------------------------------
 MILLER;FORD;SCOTT;SMITH;KING;JONES;ADAMS
(1 row)
```

covar_pop(Y, X)

描述：计算总体协方差。

返回类型：double precision

示例：

```
SELECT COVAR_POP(sr_fee, sr_net_loss) FROM tpcds.store_returns WHERE sr_customer_sk < 1000;
    covar_pop
------------------
 829.749627587403
(1 row)
```

 版权所有 © 华为云计算技术有限公司

covar_samp(Y, X)

描述：计算样本协方差。

返回类型：double precision

示例：

```
SELECT COVAR_SAMP(sr_fee, sr_net_loss) FROM tpcds.store_returns WHERE sr_customer_sk < 1000;
    covar_samp
------------------
 830.052235037289
(1 row)
```

stddev_pop(expression)

描述：计算总体标准差（总体方差的平方根）。

返回类型：对于浮点类型的输入返回double precision，其他输入返回numeric。

示例：

```
SELECT STDDEV_POP(inv_quantity_on_hand) FROM tpcds.inventory WHERE inv_warehouse_sk = 1;
    stddev_pop
------------------
 289.224294957556
(1 row)
```

stddev_samp(expression)

描述：计算样本标准差（样本方差的平方根）。

返回类型：对于浮点类型的输入返回double precision，其他输入返回numeric。

示例：

```
SELECT STDDEV_SAMP(inv_quantity_on_hand) FROM tpcds.inventory WHERE inv_warehouse_sk = 1;
   stddev_samp
------------------
 289.224359757315
(1 row)
```

var_pop(expression)

描述：计算总体方差（总体标准差的平方）。

返回类型：对于浮点类型的输入返回double precision类型，其他输入返回numeric类型。

示例：

```
SELECT VAR_POP(inv_quantity_on_hand) FROM tpcds.inventory WHERE inv_warehouse_sk = 1;
      var_pop
--------------------
 83650.692793695475
(1 row)
```

var_samp(expression)

描述：计算样本方差（样本标准差的平方）。

返回类型：对于浮点类型的输入返回double precision类型，其他输入返回numeric类型。

 版权所有 © 华为云计算技术有限公司

示例：

```
SELECT VAR_SAMP(inv_quantity_on_hand) FROM tpcds.inventory WHERE inv_warehouse_sk = 1;
     var_samp
--------------------
 83650.730277028768
(1 row)
```

bit_and(expression)

描述：对所有非NULL输入值按位与（AND）运算。如果全部输入值皆为NULL，则结果也为NULL 。

返回类型：和参数数据类型相同。

示例：

```
SELECT BIT_AND(inv_quantity_on_hand) FROM tpcds.inventory WHERE inv_warehouse_sk = 1;
 bit_and
---------
       0
(1 row)
```

bit_or(expression)

描述：对所有非NULL输入值按位或（OR）运算。如果全部输入值皆为NULL，则结果也为NULL。

返回类型：和参数数据类型相同

示例：

```
SELECT BIT_OR(inv_quantity_on_hand) FROM tpcds.inventory WHERE inv_warehouse_sk = 1;
 bit_or
--------
   1023
(1 row)
```

bool_and(expression)

描述：如果所有输入值都是真，则返回真（TRUE），否则返回假（FALSE）。

返回类型：bool

示例：

```
SELECT bool_and(100 <2500);
 bool_and
----------
 t
(1 row)
```

bool_or(expression)

描述：如果所有输入值只要有一个为真，则返回真（TRUE），否则返回假（FALSE）。

返回类型：bool

示例：

```
SELECT bool_or(100 <2500);
 bool_or
```

 版权所有 © 华为云计算技术有限公司

```
----------
 t
(1 row)
```

corr(Y, X)

描述：计算相关系数（协方差除以两个变量标准差的乘积）。

返回类型：double precision

示例：

```
SELECT CORR(sr_fee, sr_net_loss) FROM tpcds.store_returns WHERE sr_customer_sk < 1000;
       corr
--------------------
 0.0381383624904186
(1 row)
```

every(expression)

描述：等效于**bool_and(expression)**。如果所有输入值都是真，则返回真（TRUE），否则返回假（FALSE）。

返回类型：bool

示例：

```
SELECT every(100 <2500);
 every
-------
 t
(1 row)
```

regr_avgx(Y, X)

描述：计算自变量X的平均值 (sum(X)/N)。

返回类型：double precision

示例：

```
SELECT REGR_AVGX(sr_fee, sr_net_loss) FROM tpcds.store_returns WHERE sr_customer_sk < 1000;
    regr_avgx
------------------
 578.606576740795
(1 row)
```

regr_avgy(Y, X)

描述：计算因变量Y的平均值 (sum(Y)/N)。

返回类型：double precision

示例：

```
SELECT REGR_AVGY(sr_fee, sr_net_loss) FROM tpcds.store_returns WHERE sr_customer_sk < 1000;
    regr_avgy
------------------
 50.0136711629602
(1 row)
```

 版权所有 © 华为云计算技术有限公司

regr_count(Y, X)

描述：计算两个表达式都不为NULL的输入行数。

返回类型：bigint

示例：

```
SELECT REGR_COUNT(sr_fee, sr_net_loss) FROM tpcds.store_returns WHERE sr_customer_sk < 1000;
 regr_count
------------
       2743
(1 row)
```

regr_intercept(Y, X)

描述：根据所有输入的点(X, Y)按照最小二乘法拟合成一个线性方程，然后返回该直线的Y轴截距。

返回类型：double precision

示例：

```
SELECT REGR_INTERCEPT(sr_fee, sr_net_loss) FROM tpcds.store_returns WHERE sr_customer_sk < 1000;
  regr_intercept
------------------
 49.2040847848607
(1 row)
```

regr_r2(Y, X)

描述：计算相关系数的平方。

返回类型：double precision

示例：

```
SELECT REGR_R2(sr_fee, sr_net_loss) FROM tpcds.store_returns WHERE sr_customer_sk < 1000;
      regr_r2
--------------------
 0.00145453469345058
(1 row)
```

regr_slope(Y, X)

描述：根据所有输入的点(X, Y)按照最小二乘法拟合成一个线性方程， 然后返回该直线的斜率。

返回类型：double precision

示例：

```
SELECT REGR_SLOPE(sr_fee, sr_net_loss) FROM tpcds.store_returns WHERE sr_customer_sk < 1000;
     regr_slope
--------------------
 0.00139920009665259
(1 row)
```

regr_sxx(Y, X)

描述：计算辅助统计量sum(X^2) - sum(X)^2/N （自变量X的平方和）。

返回类型：double precision

 版权所有 © 华为云计算技术有限公司

示例：

```
SELECT REGR_SXX(sr_fee, sr_net_loss) FROM tpcds.store_returns WHERE sr_customer_sk < 1000;
    regr_sxx
------------------
 1626645991.46135
(1 row)
```

regr_sxy(Y, X)

描述：计算辅助统计量sum(X*Y) - sum(X) * sum(Y)/N （自变量X和因变量Y的乘方积）。

返回类型：double precision

示例：

```
SELECT REGR_SXY(sr_fee, sr_net_loss) FROM tpcds.store_returns WHERE sr_customer_sk < 1000;
    regr_sxy
------------------
 2276003.22847225
(1 row)
```

regr_syy(Y, X)

描述：计算sum(Y^2) - sum(Y)^2/N（因变量Y的平方和）。

返回类型：double precision

示例：

```
SELECT REGR_SYY(sr_fee, sr_net_loss) FROM tpcds.store_returns WHERE sr_customer_sk < 1000;
   regr_syy
------------------
 2189417.6547314
(1 row)
```

stddev(expression)

描述：**stddev_samp(expression)**的别名，计算样本标准差。

返回类型：对于浮点类型的输入返回double precision，其他输入返回numeric。

示例：

```
SELECT STDDEV(inv_quantity_on_hand) FROM tpcds.inventory WHERE inv_warehouse_sk = 1;
     stddev
------------------
 289.224359757315
(1 row)
```

variance(expression)

描述：**var_samp(expression)**的别名，计算样本方差。

返回类型：对于浮点类型的输入返回double precision类型，其他输入返回numeric类型。

示例：

```
SELECT VARIANCE(inv_quantity_on_hand) FROM tpcds.inventory WHERE inv_warehouse_sk = 1;
     variance
--------------------
```

 版权所有 © 华为云计算技术有限公司

```
 83650.730277028768
(1 row)
```

checksum(expression)

描述：返回所有输入值的CHECKSUM值。使用该函数可以用来验证DWS数据库（不支持DWS之外的其他数据库）的备份恢复或者数据迁移操作前后表中的数据是否一致。在备份恢复或者数据迁移操作前后，都需要用户手动执行SQL命令获取结果，通过对比结果判断操作前后表中的数据是否一致。

说明

- 对于数据量大的表，执行CHECKSUM()函数可能耗时较长。
- 如果两个表的CHECKSUM值不同，则表明两个表的内容不同。由于CHECKSUM函数使用散列函数，不能保证无冲突，因此两个不同内容的表可能会得到相同的CHECKSUM值，但这种情况的可能性较小。对于列进行的CHECKSUM也存在相同的情况。
- 对于时间类型timestamp、timestamptz和smalldatetime，计算CHECKSUM值时请确保时区设置一致。

- 若计算某列的CHECKSUM值，且该列类型可以默认转为TEXT类型，则expression为列名。
- 若计算某列的CHECKSUM值，且该列类型不能默认转为TEXT类型，则expression为列名::TEXT。
- 若计算所有列的CHECKSUM值，则expression为表名::TEXT。

可以默认转换为TEXT类型的类型包括：char、name、int8、int2、int1、int4、raw、pg_node_tree、float4、float8、bpchar、varchar、nvarchar2、date、timestamp、timestamptz、numeric和smalldatetime，其他类型需要强制转换为TEXT。

返回类型：numeric

示例：

表中可以默认转为TEXT类型的某列的CHECKSUM值：

```
SELECT CHECKSUM(inv_quantity_on_hand) FROM tpcds.inventory;
    checksum
-------------------
 24417258945265247
(1 row)
```

表中不能默认转为TEXT类型的某列的CHECKSUM值（注意此时CHECKSUM参数是列名::TEXT）：

```
SELECT CHECKSUM(inv_quantity_on_hand::TEXT) FROM tpcds.inventory;
    checksum
-------------------
 24417258945265247
(1 row)
```

表中所有列的CHECKSUM值。注意此时CHECKSUM参数是表名::TEXT，且表名前不加Schema：

```
SELECT CHECKSUM(inventory::TEXT) FROM tpcds.inventory;
    checksum
-------------------
 25223696246875800
(1 row)
```

 版权所有 © 华为云计算技术有限公司

approx_count_distinct(col_name)

描述：使用HyperLogLog++ (HLL++) 算法估算某一列去重后的行数（即基数，Cardinality）。该函数仅8.3.0及以上集群版本支持。

入参说明：col_name，需要估算基数的列名。

说明

可通过GUC参数approx_count_distinct_precision调整误差率。

- 参数取值范围为[10,20]，默认值为17，理论误差率为千分之三。
- 该参数表示HyperLogLog++ (HLL++)算法中分桶个数，参数值越大，分桶数越多，理论误差率则越小。
- 该参数值越大，相应的计算时间和内存资源开销越大，但仍远小于精确的count distinct语句对应的开销。推荐在估算基数较大时使用该函数替换count distinct。

示例：

```
CREATE TABLE employeeinfo (empno smallint, ename varchar(20), job varchar(20), hiredate date,deptno smallint) WITH (ORIENTATION = COLUMN);
INSERT INTO employeeinfo VALUES (7155, 'JACK', 'SALESMAN', '2018-12-01', 30);
INSERT INTO employeeinfo VALUES (7003, 'TOM', 'FINANCE', '2016-06-15', 20);
INSERT INTO employeeinfo VALUES (7357, 'MAX', 'SALESMAN', '2020-10-01', 30);

SELECT APPROX_COUNT_DISTINCT(empno) from employeeinfo;
 approx_count_distinct
-----------------------
                     3
(1 row)

SELECT COUNT(DISTINCT empno) FROM employeeinfo GROUP BY ename;
 count
-------
     1
     1
     1
(3 rows)
```

UNIQ(col_name)

描述：计算某一列去重后的行数，结果返回一个去重值。与COUNT DISTINCT功能类似。该函数仅8.3.0及以上集群版本支持。

入参说明：col_name，表示需要计算去重后行数的列。支持SMALLINT、INTEGER、BIGINT、REAL、DOUBLE PRECISION、TEXT、VARCHAR、TIMESTAMP、TIMESTAMPTZ、DATE、TIMETZ、UUID类型。

说明

- DWS目前仅列存表支持UNIQ()函数。
- 使用UNIQ()函数时，SQL需要包含GROUP BY，为了取得更好性能所选GROUP BY字段分布需要尽量均匀。
- 建议当SQL中需要去重的列大于等于3列时，使用UNIQ()函数去重，以取得比COUNT DISTINCT更好的效果。
- 一般情况下UNIQ()函数内存消耗低于COUNT DISTINCT，如果使用COUNT DISTINCT时遇到内存超限，可以使用UNIQ()函数进行替换。

示例：

```
CREATE TABLE employeeinfo (empno smallint, ename varchar(20), job varchar(20), hiredate date,deptno smallint) WITH (ORIENTATION = COLUMN);
```

 版权所有 © 华为云计算技术有限公司

```
INSERT INTO employeeinfo VALUES (7155, 'JACK', 'SALESMAN', '2018-12-01', 30);
INSERT INTO employeeinfo VALUES (7003, 'TOM', 'FINANCE', '2016-06-15', 20);
INSERT INTO employeeinfo VALUES (7357, 'MAX', 'SALESMAN', '2020-10-01', 30);

SELECT UNIQ(deptno) FROM employeeinfo GROUP BY ename;
 uniq
------
    1
    1
    1
(3 rows)
```

7.12 窗口函数

普通的聚集函数将多行数据聚集为一行结果，而窗口函数可以跨行计算，并将计算结果填充到每一行中。

窗口函数基于查询结果集的子集（即“窗口”）进行计算，实现了与聚集函数类似的聚集功能，因此聚集函数也可以作为窗口函数使用。 窗口函数可以扫描所有的行的同时，将原始数据和聚集计算结果一并输出。

注意事项

- 列存表目前只支持窗口函数rank(expression)、row_number(expression)以及聚集函数的sum，count，avg，min和max；行存表没有限制。
- 单个查询中可以包含一个或多个窗口函数表达式。
- 窗口函数仅能出现在输出列中。如果需要使用窗口函数的值进行条件过滤，需要将窗口函数嵌套在子查询中，在外层使用窗口函数表达式的别名进行条件过滤。例如：
  ```
  SELECT classid, id, score FROM(SELECT *, avg(score) OVER(PARTITION BY classid) as avg_score FROM score) WHERE score >= avg_score;
  ```
- 窗口函数所在查询块中支持使用GROUP BY表达式进行分组去重，但要求窗口函数中的PARTITION BY子句中必须是GROUP BY表达式的子集，以保证窗口函数在GROUP BY列去重后的结果上进行窗口运算，同时ORDER BY子句的表达式也需要是GROUP BY表达式的子集，或聚集运算的聚集函数。例如：
  ```
  SELECT classid,rank() OVER(PARTITION BY classid ORDER BY sum(score)) as avg_score FROM score GROUP BY classid, id;
  ```
- 9.1.1.200及以上集群版本中部分函数支持在窗口函数场景下使用DISTINCT去重计算，支持的函数包括：count、sum、min、max、avg、array_agg、stddev/stddev_samp、variance/var_samp。使用时，OVER子句中不能含有ORDER BY与frame_clause选项，且函数的参数数量只能为1。

  OVER子句外的ORDER BY仅用于对查询最终输出结果进行排序，不影响函数计算顺序。例如：
  ```
  SELECT classid, id, count(DISTINCT score) OVER(PARTITION BY classid) as num_distinct_score FROM score ORDER BY id;
  ```

语法格式

窗口函数通过OVER语句来定义窗口。OVER语句用于对数据进行分组（PARTITION BY）和组内排序（ORDER BY），从而确定窗口函数的计算范围。

```
function_name ([expression [, expression ... ]]) OVER ( window_definition )
function_name ([expression [, expression ... ]]) OVER window_name
function_name ( * ) OVER ( window_definition )
function_name ( * ) OVER window_name
```

其中window_definition子句option为：

 版权所有 © 华为云计算技术有限公司

```
[ existing_window_name ]
[ PARTITION BY expression [, ...] ]
[ ORDER BY expression [ ASC | DESC | USING operator ] [ NULLS { FIRST | LAST } ] [, ...] ]
[ frame_clause ]
```

PARTITION BY选项将具有相同PARTITION BY表达式值的行划分为一组。

ORDER BY选项用于控制窗口函数处理行的顺序。ORDER BY后面必须跟字段名，若ORDER BY后面跟数字，该数字会被按照常量处理，对目标列没有起到排序的作用。

frame_clause子句option为：

```
[ RANGE | ROWS ] frame_start
[ RANGE | ROWS ] BETWEEN frame_start AND frame_end
```

当需要指定一个窗口对分组内所有行结果进行计算时，需要指定窗口区间开始的行和结束的行。窗口区间支持RANGE、ROWS两种模式，ROWS以物理单位（行）指定窗口，RANGE将窗口指定为逻辑偏移量。

RANGE、ROWS中可以使用BETWEEN frame_start AND frame_end指定边界可取值。如果仅指定frame_start，则frame_end默认为CURRENT ROW。

frame_start和frame_end取值为：

- CURRENT ROW，当前行。
- N PRECEDING，当前行向前第n行。
- UNBOUNDED PRECEDING，当前PARTITION的第1行。
- N FOLLOWING，当前行向后第n行。
- UNBOUNDED FOLLOWING，当前PARTITION的最后1行。

需要注意，frame_start不能为UNBOUNDED FOLLOWING，frame_end不能为UNBOUNDED PRECEDING，并且frame_end选项不能比上面取值中出现的frame_start选项早。例如RANGE BETWEEN CURRENT ROW AND N PRECEDING是不被允许的。

RANK()

描述：为一个组内值进行排序，相同的值会获得相同排序序号，且后续排序会跳过并列的序号（例如：1，2，3，3，5），因此排序可能不连续。

返回值类型：BIGINT

示例：

示例表score的各字段分别表示学生id、所在班级id以及考试成绩。使用RANK函数对学生成绩进行排序：

```
CREATE TABLE score(id int,classid int,score int);
INSERT INTO score VALUES(1,1,95),(2,2,95),(3,2,85),(4,1,70),(5,2,88),(6,1,70);

SELECT id, classid, score,RANK() OVER(ORDER BY score DESC) FROM score;
 id | classid | score | rank
----+---------+-------+------
  1 |       1 |    95 |    1
  2 |       2 |    95 |    1
  6 |       1 |    70 |    5
  4 |       1 |    70 |    5
  5 |       2 |    88 |    3
  3 |       2 |    85 |    4
(6 rows)
```


版权所有 © 华为云计算技术有限公司

ROW_NUMBER()

描述：ROW_NUMBER函数为各组内值生成连续排序序号，相同的值其序号也不相同，即序号唯一。

返回值类型：BIGINT

示例：

```
SELECT id, classid, score,ROW_NUMBER() OVER(ORDER BY score DESC) FROM score ORDER BY score DESC;
 id | classid | score | row_number
----+---------+-------+------------
  1 |       1 |    95 |          1
  2 |       2 |    95 |          2
  5 |       2 |    88 |          3
  3 |       2 |    85 |          4
  6 |       1 |    70 |          5
  4 |       1 |    70 |          6
(6 rows)
```

DENSE_RANK()

描述：DENSE_RANK函数为各组内值生成连续排序序号，其中相同的值具有相同序号，相同值只占用一个编号。

返回值类型：BIGINT

示例：

```
SELECT id, classid, score,DENSE_RANK() OVER(ORDER BY score DESC) FROM score;
 id | classid | score | dense_rank
----+---------+-------+------------
  1 |       1 |    95 |          1
  2 |       2 |    95 |          1
  5 |       2 |    88 |          2
  3 |       2 |    85 |          3
  6 |       1 |    70 |          4
  4 |       1 |    70 |          4
(6 rows)
```

PERCENT_RANK()

描述：PERCENT_RANK函数为各组内对应值生成相对序号，即根据公式 (rank - 1) / (total rows - 1)计算所得的值。其中rank为该值依据RANK函数所生成的对应序号，totalrows为该分组内的总元素个数。

返回值类型：DOUBLE PRECISION

示例：

```
SELECT id, classid, score,PERCENT_RANK() OVER(ORDER BY score DESC) FROM score;
 id | classid | score | percent_rank
----+---------+-------+--------------
  1 |       1 |    95 |            0
  2 |       2 |    95 |            0
  3 |       2 |    85 |           .6
  4 |       1 |    70 |           .8
  5 |       2 |    88 |           .4
  6 |       1 |    70 |           .8
(6 rows)
```


版权所有 © 华为云计算技术有限公司

CUME_DIST()

描述：CUME_DIST函数为各组内对应值生成累积分布序号。即根据公式(小于等于当前值的数据行数)/(该分组总行数totalrows)计算所得的相对序号。

返回值类型：DOUBLE PRECISION

示例：

```
SELECT id,classid,score,CUME_DIST() OVER(ORDER BY score DESC) FROM score;
 id | classid | score |    cume_dist
----+---------+-------+-----------------
  1 |       1 |    95 | .333333333333333
  2 |       2 |    95 | .333333333333333
  5 |       2 |    88 |               .5
  3 |       2 |    85 | .666666666666667
  4 |       1 |    70 |                1
  6 |       1 |    70 |                1
(6 rows)
```

NTILE(num_buckets integer)

描述：NTILE函数根据num_buckets integer将有序的数据集合平均分配到num_buckets所指定数量的桶中，并将桶号分配给每一行。分配时应尽量做到平均分配。

返回值类型：INTEGER

示例：

```
SELECT id,classid,score,NTILE(3) OVER(ORDER BY score DESC) FROM score;
 id | classid | score | ntile
----+---------+-------+-------
  1 |       1 |    95 |     1
  2 |       2 |    95 |     1
  5 |       2 |    88 |     2
  3 |       2 |    85 |     2
  4 |       1 |    70 |     3
  6 |       1 |    70 |     3
(6 rows)
```

LAG(value any [, offset integer [, default any]])

描述：LAG函数为各组内对应值生成滞后值。即当前值对应的行数往前偏移offset位后所得行的value值作为序号。若经过偏移后行数不存在，则对应结果取为default值。若无指定，在默认情况下，offset取为1，default值取为NULL。

返回值类型：与参数数据类型相同。

示例：

```
SELECT id,classid,score,LAG(id,3) OVER(ORDER BY score DESC) FROM score;
 id | classid | score | lag
----+---------+-------+-----
  1 |       1 |    95 |
  2 |       2 |    95 |
  5 |       2 |    88 |
  3 |       2 |    85 |   1
  4 |       1 |    70 |   2
  6 |       1 |    70 |   5
(6 rows)
```


版权所有 © 华为云计算技术有限公司

LEAD(value any [, offset integer [, default any]])

描述：LEAD函数为各组内对应值生成提前值。即当前值对应的行数向后偏移offset位后所得行的value值作为序号。若经过向后偏移后行数超过当前组内的总行数，则对应结果取为default值。若无指定，在默认情况下，offset取为1，default值取为NULL。

返回值类型：与参数数据类型相同。

示例：

```
SELECT id,classid,score,LEAD(id,3) OVER(ORDER BY score DESC) FROM score;
 id | classid | score | lead
----+---------+-------+------
  1 |       1 |    95 |    3
  2 |       2 |    95 |    4
  5 |       2 |    88 |    6
  3 |       2 |    85 |
  4 |       1 |    70 |
  6 |       1 |    70 |
(6 rows)
```

FIRST_VALUE(value any)

描述：取各组内的第一个数据值作为返回结果。

返回值类型：与参数数据类型相同。

示例：

```
SELECT id,classid,score,FIRST_VALUE(id) OVER(ORDER BY score DESC) FROM score;
 id | classid | score | first_value
----+---------+-------+-------------
  1 |       1 |    95 |           1
  2 |       2 |    95 |           1
  5 |       2 |    88 |           1
  3 |       2 |    85 |           1
  4 |       1 |    70 |           1
  6 |       1 |    70 |           1
(6 rows)
```

LAST_VALUE(value any)

描述：LAST_VALUE函数取各组内的最后一个值作为返回结果。

返回值类型：与参数数据类型相同。

LAST_VALUE函数支持IGNORE NULLS语法，该语法返回非NULL窗口中的最后一个值，如果所有值都为NULL，则返回NULL，具体格式为：

```
LAST_VALUE (expression) [IGNORE NULLS] OVER (window_definition)
```

当前IGNORE NULLS仅支持ROWS between CURRENT ROW and UNBOUNDED FOLLOWING和ROWS BETWEEN UNBOUNDED PRECEDING and CURRENT ROW两种窗口区间。

示例：

```
SELECT id,classid,score,LAST_VALUE(id) OVER(ORDER BY score DESC) FROM score;
 id | classid | score | last_value
----+---------+-------+------------
  1 |       1 |    95 |          2
  2 |       2 |    95 |          2
  5 |       2 |    88 |          5
  3 |       2 |    85 |          3
```


版权所有 © 华为云计算技术有限公司

```
 4 |     1 |   70 |        6
 6 |     1 |   70 |        6
(6 rows)
```

NTH_VALUE(value any, nth integer)

描述：返回该组内的第n行作为结果。若该行不存在，则默认返回NULL。

返回值类型：与参数数据类型相同。

示例：返回按成绩降序排序后的结果集中第3行的id值。

```
SELECT id,classid,score,NTH_VALUE(id,3) OVER(ORDER BY score DESC) FROM score;
 id | classid | score | nth_value
----+---------+-------+-----------
 1 |     1 |   95 |
 2 |     2 |   95 |
 5 |     2 |   88 |         5
 3 |     2 |   85 |         5
 4 |     1 |   70 |         5
 6 |     1 |   70 |         5
(6 rows)
```

7.13 类型转换函数

在数据仓库的日常使用中，用户经常需要在不同数据类型之间进行转换。例如，从外部数据源导入数据时，日期可能以字符串形式存储；进行数据分析时，需要将数值格式化为特定精度的字符串用于报表展示；处理日志数据时，需要将十六进制编码还原为原始二进制内容。DWS提供了丰富的类型转换函数，覆盖日期时间转换、数值格式化、字符集转换、二进制与十六进制互转等场景。

然而，不同函数的语法格式、适用版本和返回值类型存在差异，用户需要准确选择合适的函数并正确指定格式模板，否则可能出现转换失败或结果不符合预期的情况。本文档将系统介绍DWS各类型转换函数的语法、参数、返回值和使用示例，帮助用户在实际业务中正确完成类型转换操作。

类型转换函数分类

DWS类型转换函数按转换方向和用途可分为以下8类：

表 7-18 类型转换函数分类

分类	包含函数	说明
通用类型转换函数	cast(x as y)、cast(x, y)、try_cast(x as type)	任意类型间的通用转换，无需格式模板。
数值格式化函数（数值转字符串）	to_char(double precision, text)、to_char(integer/number [, fmt])、to_char(int, text)、to_char(numeric, text)	将数值按格式模板转为字符串。
日期时间格式化函数（日期时间转字符串）	to_char(datetime/interval [, fmt])、to_char(interval, text)、to_char(timestamp, text)	将日期时间按格式模板转为字符串。

版权所有 © 华为云计算技术有限公司

分类	包含函数	说明
字符串类型转换函数	to_char(string)、to_clob(...)	字符串类型间的转换。
字符串转数值函数	to_number(expr [, fmt])、to_number(text, text)	将字符串按格式转为数值。
字符串转日期时间函数	to_date(text)、to_date(text, text)、to_date(string, fmt)、to_timestamp(double precision)、to_timestamp(string [,fmt])、to_timestamp(text, text)	将字符串或UNIX时间戳转为日期/时间戳。
二进制与十六进制互转函数	hextoraw(string)、rawtohex(string)	十六进制字符串与二进制互转。
其他转换函数	numtoday(numeric)、pg_systimestamp()	数值转时间、获取系统时间戳。

通用类型转换函数

通用类型转换函数适用于任意数据类型之间的转换，无需指定格式模板。

cast(x as y)

描述：类型转换函数，将x转换成y指定的类型。

示例：

```
SELECT cast('22-oct-1997' as timestamp);
     timestamp
---------------------
 1997-10-22 00:00:00
(1 row)
```

cast(x, y)

描述：类型转换函数，将x转换成y指定的类型。该函数仅8.2.0及以上集群版本支持。

示例：

```
SELECT cast('22-oct-1997', timestamp);
     timestamp
---------------------
 1997-10-22 00:00:00
(1 row)
```

try_cast(x as type)

描述：将x转换成给定的type类型值，若转换失败且当前类型转换为DWS允许的转换，则返回NULL，否则报错。该函数仅8.2.0及以上集群版本支持。

示例：

```
SELECT try_cast('a' as int4);
     int4
```


版权所有 © 华为云计算技术有限公司

```
---------------

(1 row)
SELECT try_cast('22-oct-1997' as timestamp);
      timestamp
---------------------
 1997-10-22 00:00:00
(1 row)
```

数值格式化函数（数值转字符串）

以下函数将数值类型（整数、浮点数、双精度等）按指定格式模板转换为字符串。

to_char(double precision, text)

描述：将双精度类型的值转换为指定格式的字符串。

返回值类型：text

示例：

```
SELECT to_char(125.8::real, '999D99');
 to_char
---------
  125.80
(1 row)
```

to_char (integer/number[, fmt])

描述：将一个整型或者浮点类型的值转换为指定格式的字符串。

- 可选参数fmt可以为以下几类：十进制字符、“分组”符、正负号和货币符号，每类都可以有不同的模板，模板之间可以合理组合，常见的模板有：9、0、,（千分隔符）、.（小数点），可参考表 数值格式化的模板模式。
- 模板可以有类似FM的修饰词，但FM不抑制由模板0指定而输出的0。
- 要将整型类型的值转换成对应16进制值的字符串，使用模板“x”或“X”。

返回值类型：varchar

示例：

```
SELECT to_char(1485,'9,999');
 to_char
---------
  1,485
(1 row)
SELECT to_char( 1148.5,'9,999.999');
  to_char
------------
  1,148.500
(1 row)
SELECT to_char(148.5,'990999.909');
   to_char
-------------
    0148.500
(1 row)
SELECT to_char(123,'XXX');
 to_char
---------
   7B
(1 row)
```

 版权所有 © 华为云计算技术有限公司

to_char(int, text)

描述：将整数类型的值转换为指定格式的字符串。

返回值类型：text

示例：

```
SELECT to_char(125, '999');
 to_char
---------
  125
(1 row)
```

to_char(numeric, text)

描述：将数字类型的值转换为指定格式的字符串。

返回值类型：text

示例：

```
SELECT to_char(-125.8, '999D99S');
 to_char
---------
 125.80-
(1 row)
```

日期时间格式化函数（日期时间转字符串）

以下函数将日期时间类型（DATE、TIMESTAMP、INTERVAL等）按指定格式模板转换为字符串。

to_char (datetime/interval [, fmt])

描述：将一个DATE、TIMESTAMP、TIMESTAMP WITH TIME ZONE或者TIMESTAMP WITH LOCAL TIME ZONE类型的DATETIME或者INTERVAL值按照fmt指定的格式转换为VARCHAR类型。

- 可选参数fmt可以为以下几类：日期、时间、星期、季度和世纪。每类都可以有不同的模板，模板之间可以合理组合，常见的模板有：HH、MI、SS、YYYY、MM、DD，可参考**表7-20**。
- 模板可以有修饰词，常用的修饰词是FM，可以用来抑制前导的零或尾随的空白。

返回值类型：varchar

示例：

```
SELECT to_char(current_timestamp,'HH12:MI:SS');
 to_char
----------
 10:19:26
(1 row)
SELECT to_char(current_timestamp,'FMHH12:FMMI:FMSS');
 to_char
----------
 10:19:46
(1 row)
```

to_char(interval, text)

描述：将时间间隔类型的值转换为指定格式的字符串。

 版权所有 © 华为云计算技术有限公司

返回值类型：text

示例：

```
SELECT to_char(interval '15h 2m 12s', 'HH24:MI:SS');
 to_char
----------
 15:02:12
(1 row)
```

to_char(timestamp, text)

描述：将时间戳类型的值转换为指定格式的字符串。

返回值类型：text

示例：

```
SELECT to_char(current_timestamp, 'HH12:MI:SS');
 to_char
----------
 10:55:59
(1 row)
```

字符串类型转换函数

以下函数在不同字符串类型（CHAR、VARCHAR、CLOB等）之间进行转换。

to_char (string)

描述：将CHAR、VARCHAR、VARCHAR2、CLOB类型转换为VARCHAR类型。

如使用该函数对CLOB类型进行转换，且待转换CLOB类型的值超出目标类型的范围，则返回错误。

返回值类型：varchar

示例：

```
SELECT to_char('01110');
 to_char
---------
 01110
(1 row)
```

to_clob(char/nchar/varchar/nvarchar/varchar2/nvarchar2/text/raw)

描述：将RAW类型或者文本字符集类型CHAR、NCHAR、VARCHAR、VARCHAR2、NVARCHAR2、TEXT转成CLOB类型。

返回值类型：clob

示例：

```
SELECT to_clob('ABCDEF'::RAW(10));
 to_clob
---------
 ABCDEF
(1 row)
SELECT to_clob('hello111'::CHAR(15));
 to_clob
----------
 hello111
(1 row)
```

 版权所有 © 华为云计算技术有限公司

```
SELECT to_clob('gauss123'::NCHAR(10));
 to_clob
----------
 gauss123
(1 row)
SELECT to_clob('gauss234'::VARCHAR(10));
 to_clob
----------
 gauss234
(1 row)
SELECT to_clob('gauss345'::VARCHAR2(10));
 to_clob
----------
 gauss345
(1 row)
SELECT to_clob('gauss456'::NVARCHAR2(10));
 to_clob
----------
 gauss456
(1 row)
SELECT to_clob('World222!'::TEXT);
  to_clob
-----------
 World222!
(1 row)
```

字符串转数值函数

以下函数将字符串按指定格式模板转换为数值类型。

to_number (expr [, fmt])

描述：将expr按指定格式转换为一个NUMBER类型的值。

类型转换格式请参考**表7-19**。

转换十六进制字符串为十进制数字时，最多支持16个字节的十六进制字符串转换为无符号数。

转换十六进制字符串为十进制数字时，格式字符串中不允许出现除'x'或'X'以外的其他字符，否则报错。

返回值类型：number

示例：

```
SELECT to_number('12,454.8-', '99G999D9S');
 to_number
-----------
  -12454.8
(1 row)
```

to_number(text, text)

描述：将字符串类型的值转换为指定格式的数字。

返回值类型：numeric

示例：

```
SELECT to_number('12,454.8-', '99G999D9S');
 to_number
-----------
  -12454.8
(1 row)
```

 版权所有 © 华为云计算技术有限公司

字符串转日期时间函数

以下函数将字符串或UNIX时间戳转换为日期或时间戳类型。

to_date(text)

描述：将文本类型的值转换为指定格式的时间戳。

返回值类型：timestamp

示例：

```
SELECT to_date('2015-08-14');
      to_date
---------------------
 2015-08-14 00:00:00
(1 row)
```

to_date(text, text)

描述：将字符串类型的值转换为指定格式的日期。

返回值类型：timestamp

示例：

```
SELECT to_date('05 Dec 2000', 'DD Mon YYYY');
      to_date
---------------------
 2000-12-05 00:00:00
(1 row)
```

to_date(string, fmt)

描述：将字符串string按fmt指定格式转化为DATE类型的值。fmt格式可参考**表7-20**。

该函数不能直接支持CLOB类型，但是CLOB类型的参数能够通过隐式转换实现。

返回值类型：date

示例：

```
SELECT TO_DATE('05 Dec 2010','DD Mon YYYY');
      to_date
---------------------
 2010-12-05 00:00:00
(1 row)
```

to_timestamp(double precision)

描述：把UNIX纪元转换成时间戳。

返回值类型：timestamp with time zone

示例：

```
SELECT to_timestamp(1284352323);
     to_timestamp
------------------------
 2010-09-13 12:32:03+08
(1 row)
```

 版权所有 © 华为云计算技术有限公司

to_timestamp(string [,fmt])

描述：将字符串string按fmt指定的格式转换成时间戳类型的值。不指定fmt时，按参数nls_timestamp_format所指定的格式转换。fmt格式可参考**表7-20**。

DWS的to_timestamp中，

- 如果输入的年份YYYY=0，系统报错。
- 如果输入的年份YYYY<0，在fmt中指定SYYYY，则正确输出公元前绝对值n的年份。

fmt中出现的字符必须与日期/时间格式化的模式相匹配，否则报错。

返回值类型：timestamp without time zone

示例：

```
SHOW nls_timestamp_format;
    nls_timestamp_format
-----------------------------
 DD-Mon-YYYY HH:MI:SS.FF AM
(1 row)

SELECT to_timestamp('12-sep-2014');
    to_timestamp
---------------------
 2014-09-12 00:00:00
(1 row)
SELECT to_timestamp('12-Sep-10 14:10:10.123000','DD-Mon-YY HH24:MI:SS.FF');
      to_timestamp
-------------------------
 2010-09-12 14:10:10.123
(1 row)
SELECT to_timestamp('-1','SYYYY');
      to_timestamp
-----------------------
 0001-01-01 00:00:00 BC
(1 row)
SELECT to_timestamp('98','RR');
    to_timestamp
---------------------
 1998-01-01 00:00:00
(1 row)
SELECT to_timestamp('01','RR');
    to_timestamp
---------------------
 2001-01-01 00:00:00
(1 row)
```

to_timestamp(text, text)

描述：将字符串类型的值转换为指定格式的时间戳。

返回值类型：timestamp

示例：

```
SELECT to_timestamp('05 Dec 2000', 'DD Mon YYYY');
    to_timestamp
---------------------
 2000-12-05 00:00:00
(1 row)
```

二进制与十六进制互转函数

以下函数在十六进制字符串与二进制类型之间进行转换。

 版权所有 © 华为云计算技术有限公司

hextoraw(string)

描述：将一个十六进制构成的字符串转换为二进制。

返回值类型：raw

示例：

```
SELECT hextoraw('7D');
 hextoraw
----------
 7D
(1 row)
```

rawtohex(string)

描述：将一个二进制构成的字符串转换为十六进制的字符串。

结果为输入字符的ASCII码，以十六进制表示。

返回值类型：varchar

示例：

```
SELECT rawtohex('1234567');
    rawtohex
----------------
 31323334353637
(1 row)
```

其他转换函数

numtoday(numeric)

描述：将数字类型的值转换为指定格式的时间戳。

返回值类型：timestamp

示例：

```
SELECT numtoday(2);
 numtoday
----------
 2 days
(1 row)
```

pg_systimestamp()

描述：获取系统时间戳。

返回值类型：timestamp with time zone

示例：

```
SELECT pg_systimestamp();
        pg_systimestamp
-------------------------------
 2015-10-14 11:21:28.317367+08
(1 row)
```

格式模板参考

表7-19显示了可以用于格式化数值的模板模式，适用于函数to_number。

 版权所有 © 华为云计算技术有限公司

表 7-19 数值格式化的模板模式

模式	描述
9	带有指定数值位数的值
0	带前导零的值
.（句点）	小数点
,（逗号）	分组（千）分隔符
PR	尖括号内负值
S	带符号的数值（使用区域设置）
L	货币符号（使用区域设置）
D	小数点（使用区域设置）
G	分组分隔符（使用区域设置）
MI	在指明的位置的负号（如果数字 < 0）
PL	在指明的位置的正号（如果数字 > 0）
SG	在指明的位置的正/负号
RN	罗马数字（输入在 1 和 3999 之间）
TH或th	序数后缀
V	移动指定位（小数）

表7-20显示了可以用于格式化日期和时间值的模板，这些模式适用于函数to_date、to_timestamp、to_char和参数nls_timestamp_format。

表 7-20 用于日期/时间格式化的模式

类别	模式	描述
小时	HH	一天的小时数（01-12）
	HH12	一天的小时数（01-12）
	HH24	一天的小时数（00-23）
分钟	MI	分钟（00-59）
秒	SS	秒（00-59）
	FF	微秒（000000-999999）
	SSSSS	午夜后的秒（0-86399）
上、下午	AM或A.M.	上午标识
	PM或P.M.	下午标识

版权所有 © 华为云计算技术有限公司

类别	模式	描述
年	Y,YYY	带逗号的年（4和更多位）
	SYYYY	公元前四位年
	YYYY	年（4和更多位）
	YYY	年的后三位
	YY	年的后两位
	Y	年的最后一位
	IYYY	ISO年（4位或更多位）
	IYY	ISO年的最后三位
	IY	ISO年的最后两位
	I	ISO年的最后一位
	RR	年的后两位（可在21世纪存储20世纪的年份） 规则如下： ● 输入的两位年份在00~49之间： 当前年份的后两位在00~49之间，返回值年份的前两位和当前年份的前两位相同； 当前年份的后两位在50~99之间，返回值年份的前两位是当前年份的前两位加1。 ● 输入的两位年份在50~99之间： 当前年份的后两位在00~49之间，返回值年份的前两位是当前年份的前两位减1； 当前年份的后两位在50~99之间，返回值年份的前两位和当前年份的前两位相同。
	RRRR	可接收4位年或两位年。若是两位，则和RR的返回值相同，若是四位，则和YYYY相同。
	● BC或B.C. ● AD或A.D.	纪元标识。BC（公元前），AD（公元后）。
月	MONTH	全长大写月份名（空白填充为9字符）
	MON	大写缩写月份名（3字符）
	MM	月份数（01-12）
	RM	罗马数字的月份（I-XII； I=JAN）（大写）
天	DAY	全长大写日期名（空白填充为9字符）
	DY	缩写大写日期名（3字符）
	DDD	一年里的日（001-366）
	DD	一个月里的日（01-31）

版权所有 © 华为云计算技术有限公司

类别	模式	描述
	D	一周里的日（1-7；周日是 1）
周	W	一个月里的周数（1-5）（第一周从该月第一天开始）
	WW	一年里的周数（1-53）（第一周从该年的第一天开始）
	IW	ISO一年里的周数（第一个星期四在第一周里）
世纪	CC	世纪（2位）（21 世纪从 2001-01-01 开始）
儒略日	J	儒略日（自公元前 4712 年 1 月 1 日来的天数）
季度	Q	季度

7.14 JSON/JSONB 函数和操作符

7.14.1 JSON/JSONB 操作符

表 7-21 json 和 jsonb 通用操作符

操作符	左操作数类型	右操作数类型	返回类型	描述	示例
->	Array-json(b)	int	json(b)	获得array-json元素。下标不存在返回空。	SELECT '[{"a":"foo"}, {"b":"bar"}, {"c":"baz"}]'::json->2; ?column? ---------- {"c":"baz"} (1 row)
->	object-json(b)	text	json(b)	通过键获得值。不存在则返回空。	SELECT '{"a":{"b":"foo"}}'::json->'a'; ?column? ---------- {"b":"foo"} (1 row)
->>	Array-json(b)	int	text	获得array-json元素。下标不存在返回空。	SELECT '[{"a":"foo"}, {"b":"bar"}, {"c":"baz"}]'::json->>2; ?column? ---------- {"c":"baz"} (1 row)
->>	object-json(b)	text	text	通过键获得值。不存在则返回空	SELECT '{"a":{"b":"foo"}}'::json->>'a'; ?column? ---------- {"b":"foo"} (1 row)

版权所有 © 华为云计算技术有限公司

操作符	左操作数类型	右操作数类型	返回类型	描述	示例
#>	container-json (b)	text[]	json	获取在指定路径的JSON对象，路径不存在则返回空。 **说明** DWS对象标识符支持以符号"#"结尾，为避免a#>b解析过程出现歧义，因此操作符"#>"前后需要增加空格，否则解析报错。	SELECT '{"a":{"b":{"c":1}}}'::json #> '{a, b}'; ?column? ---------- {"c":1} (1 row)
#>>	container-json (b)	text[]	text	获取在指定路径的JSON对象，路径不存在则返回空。	SELECT '{"a":{"b":{"c":1}}}'::json #>> '{a, b}'; ?column? ---------- {"c":1} (1 row)

表 7-22 jsonb 支持的操作符

操作符	右操作类型	返回类型	描述	示例
=	jsonb	bool	判断两个jsonb的大小关系。同函数jsonb_eq。	SELECT '{"a":{"b":{"c":1}}}'::jsonb = '{"a":{"b":{"c":1}}}'::jsonb; ?column? ---------- t (1 row)
<>	jsonb	bool	判断两个jsonb的大小关系。同函数jsonb_ne。	SELECT '{"a":{"b":{"c":1}}}'::jsonb <> '{"a":{"b":{"c":1}}}'::jsonb; ?column? ---------- f (1 row)
<	jsonb	bool	判断两个jsonb的大小关系。同函数jsonb_lt。	SELECT '{"a":{"b":{"c":2}}}'::jsonb < '{"a":{"b":{"c":1}}}'::jsonb; ?column? ---------- f (1 row)
>	jsonb	bool	判断两个jsonb的大小关系。同函数jsonb_gt。	SELECT '{"a":{"b":{"c":2}}}'::jsonb > '{"a":{"b":{"c":1}}}'::jsonb; ?column? ---------- t (1 row)

版权所有 © 华为云计算技术有限公司

操作符	右操作类型	返回类型	描述	示例
<=	jsonb	bool	判断两个jsonb的大小关系。同函数jsonb_le。	SELECT '{"a":{"b":{"c":2}}}'::jsonb <= '{"a":{"b":{"c":1}}}'::jsonb; ?column? ---------- f (1 row)
>=	jsonb	bool	判断两个jsonb的大小关系。同函数jsonb_ge。	SELECT '{"a":{"b":{"c":2}}}'::jsonb >= '{"a":{"b":{"c":1}}}'::jsonb; ?column? ---------- t (1 row)
?	text	bool	键/元素的字符串是否存在JSON值的顶层。	SELECT '{"a":1, "b":2}'::jsonb ? 'b'; ?column? ---------- t (1 row)
?|	text[]	bool	这些数组字符串中的任何一个是否作为顶层键存在。	SELECT '{"a":1, "b":2, "c":3, "d":4}'::jsonb ?| '{a, b, e}'::text[]; ?column? ---------- t (1 row)
?&	text[]	bool	是否所有这些数组字符串都作为顶层键存在。	SELECT '{"a":1, "b":2, "c":3, "d":4}'::jsonb ?& '{a, b, c}'::text[]; ?column? ---------- t (1 row)
<@	jsonb	bool	左边的JSON的所有项是否全部存在于右边JSON的顶层。	SELECT '{"b":3}'::jsonb <@ '{"a":{"b":{"c":2}}, "b":3}'::jsonb; ?column? ---------- t (1 row)
@>	jsonb	bool	左边的JSON的顶层是否包含右边JSON的顶层所有项。	SELECT '{"a":{"b":{"c":2}}, "b":3}'::jsonb @> '{"b":3}'::jsonb; ?column? ---------- t (1 row)
||	jsonb	jsonb	两个jsonb对象合并成一个。	SELECT '{"a":1, "b":2}'::jsonb || '{"c":3, "d":4}'::jsonb; ?column? ---------- {"a": 1, "b": 2, "c": 3, "d": 4} (1 row)
-	text	jsonb	删除jsonb对象删除指定的键值对。	SELECT '{"a":1, "b":2}'::jsonb - 'a'; ?column? ---------- {"b": 2} (1 row)

版权所有 © 华为云计算技术有限公司

操作符	右操作类型	返回类型	描述	示例
-	text	jsonb	删除jsonb对象删除指定的键值对。	SELECT '{"a":1, "b":2, "c":3, "d":4}'::jsonb - '{a, b}'::text[]; ?column? ---------- {"c": 3, "d": 4} (1 row)
-	int	jsonb	删除jsonb数组中下标对应的元素。	SELECT '["a", "b", "c"]'::jsonb - 2; ?column? ---------- ["a", "b"] (1 row)
#-	text[]	jsonb	删除jsonb对象中路径对应的键值对。	SELECT '{"a":{"b":{"c":{"d":1}}}, "e":2, "f":3}'::jsonb #- '{a, b}'::text[]; ?column? ---------- {"a": {}, "e": 2, "f": 3} (1 row)

7.14.2 JSON/JSONB 函数

JSON/JSONB函数表示可以用于JSON类型（请参考**JSON类型**）数据的函数。

说明

JSON/JSONB函数只提供基础功能，如有其他复杂或特殊性能要求场景，请联系技术工程师进行评估。

array_to_json(anyarray [, pretty_bool])

描述：返回JSON类型的数组。一个多维数组成为一个JSON数组的数组。如果pretty_bool设置为true，将在一维元素之间添加换行符。

返回类型：json

示例：

```
SELECT array_to_json('{{1,5},{99,100}}'::int[]);
array_to_json
---------------
[[1,5],[99,100]]
(1 row)
```

row_to_json(record [, pretty_bool])

描述：返回JSON类型的行。如果pretty_bool设置为true，将在第一级元素之间添加换行符。

返回类型：json

示例：

```
SELECT row_to_json(row(1,'foo'));
    row_to_json
---------------------
```


版权所有 © 华为云计算技术有限公司

```
 {"f1":1,"f2":"foo"}
(1 row)
```

json_agg(any)

描述：将值聚集为json数组。

返回类型：array-json

示例：

```
SELECT * FROM classes;
 name | score
 -----+-------
 A    |     2
 A    |     3
 D    |     5
 D    |
(4 rows)
SELECT name, json_agg(score) score FROM classes group by name order by name;
 name |       score
 -----+------------------
 A    | [2, 3]
 D    | [5, null]
      | [null]
(3 rows)
```

json_object_agg(any, any)

描述：将值聚集为json对象。

返回类型：json

示例：

```
SELECT * FROM classes;
 name | score
 -----+-------
 A    |     2
 A    |     3
 D    |     5
 D    |
(4 rows)

SELECT json_object_agg(name, score) FROM classes group by name order by name;
     json_object_agg
--------------------------
 { "A" : 2, "A" : 3 }
 { "D" : 5, "D" : null }
(2 rows)
```

json_build_array(VARIADIC "any")

描述：从可变参数列表中构建一个可能是异构类型的JSON数组。

返回类型：json

示例：

```
SELECT json_build_array(1,2,'3',4,5);
 json_build_array
-------------------
 [1, 2, "3", 4, 5]
(1 row)
```

 版权所有 © 华为云计算技术有限公司

json_build_object(VARIADIC "any")

描述：从可变参数列表中构建JSON对象。参数列表由交替的键和值组成。其入参必须为偶数个，两两一组组成键值对。注意键不可为null。

返回类型：json

示例：

```
SELECT json_build_object('foo',1,'bar',2);
  json_build_object
------------------------
 {"foo" : 1, "bar" : 2}
(1 row)
```

json_object(text[])、json_object(text[], text[])

描述：从文本数组中构建JSON对象。

这是个重载函数，当入参为一个文本数组的时候，其数组长度必须为偶数，成员被当做交替出现的键/值对。两个文本数组的时候，第一个数组被视为键，第二个被视为值，两个数组长度必须相等。键不可为null。

返回类型：json

示例：

```
SELECT json_object('{a, 1, b, "def", c, 3.5}');
              json_object
-----------------------------------------
 {"a" : "1", "b" : "def", "c" : "3.5"}
(1 row)

SELECT json_object('{{a, 1},{b, "def"},{c, 3.5}}');
              json_object
-----------------------------------------
 {"a" : "1", "b" : "def", "c" : "3.5"}
(1 row)

SELECT json_object('{a,b,"a b c"}', '{a,1,1}');
               json_object
 ----------------------------------------
 {"a" : "a", "b" : "1", "a b c" : "1"}
(1 row)
```

to_json(anyelement)

描述：把参数转换为json。

返回类型：json

示例：

```
SELECT to_json('Fred said "Hi."'::text);
       to_json
---------------------
 "Fred said \"Hi.\""
(1 row)
---将列存表json_tbl_2转换为json
SELECT * FROM json_tbl_2;
  a |  b
 ---+-----
  1 | aaa
  1 | bbb
  2 | ccc
```

 版权所有 © 华为云计算技术有限公司

```
 2 | ddd
(4 rows)
SELECT to_json(t.*) FROM json_tbl_2 t;
      to_json
-------------------
 {"a":1,"b":"bbb"}
 {"a":2,"b":"ddd"}
 {"a":1,"b":"aaa"}
 {"a":2,"b":"ccc"}
(4 rows)
```

json_strip_nulls(json)

描述：所有具有空值的对象字段被忽略，其他值保持不变。

返回类型：json

示例：

```
SELECT json_strip_nulls('[{"f1":1,"f2":null},2,null,3]');
  json_strip_nulls
---------------------
 [{"f1":1},2,null,3]
(1 row)
```

json_object_field(json, text)

描述：同操作符->, 返回对象中指定键对应的值。

返回类型：json

示例：

```
SELECT json_object_field('{"a": {"b":"foo"}}','a');
 json_object_field
-------------------
 {"b":"foo"}
(1 row)
```

json_object_field_text(object-json, text)

描述：同操作符->>, 返回对象中指定键对应的值。

返回类型：text

示例：

```
SELECT json_object_field_text('{"a": {"b":"foo"}}','a');
 json_object_field_text
------------------------
 {"b":"foo"}
(1 row)
```

json_array_element(array-json, integer)

描述：同操作符->, 返回数组中指定下标的元素。

返回类型：json

示例：

```
SELECT json_array_element('[1,true,[1,[2,3]],null]',2);
 json_array_element
--------------------
```

 版权所有 © 华为云计算技术有限公司

```
 [1,[2,3]]
(1 row)
```

json_array_element_text(array-json, integer)

描述：同操作符->>, 返回数组中指定下标的元素。

返回类型：text

示例：

```
SELECT json_array_element_text('[1,true,[1,[2,3]],null]',2);
 json_array_element_text
-------------------------
 [1,[2,3]]
(1 row)
```

json_extract_path(json, VARIADIC text[])

描述：同操作符#>，返回$2所指路径的JSON值。

返回类型：json

示例：

```
SELECT json_extract_path('{"f2":{"f3":1},"f4":{"f5":99,"f6":"stringy"}}', 'f4','f6');
 json_extract_path
-------------------
 "stringy"
(1 row)
```

json_extract_path_text(json, VARIADIC text[])

描述：同操作符#>>，返回$2所指路径的text值。

返回类型：text

示例：

```
SELECT json_extract_path_text('{"f2":{"f3":1},"f4":{"f5":99,"f6":"stringy"}}', 'f4','f6');
 json_extract_path_text
------------------------
 stringy
(1 row)
```

json_array_elements(array-json)

描述：拆分数组，每一个元素返回一行。

返回类型：json

示例：

```
SELECT json_array_elements('[1,true,[1,[2,3]],null]');
 json_array_elements
---------------------
 1
 true
 [1,[2,3]]
 null
(4 rows)
```

 版权所有 © 华为云计算技术有限公司

json_array_elements_text(array-json)

描述：拆分数组，每一个元素返回一行。

返回类型：text

示例：

```
SELECT * FROM json_array_elements_text('[1,true,[1,[2,3]],null]');
   value
-----------
 1
 true
 [1,[2,3]]

(4 rows)
```

json_array_length(array-json)

描述：返回数组长度。

返回类型：integer

示例：

```
SELECT json_array_length('[1,2,3,{"f1":1,"f2":[5,6]},4,null]');
 json_array_length
-------------------
         6
(1 row)
```

json_object_keys(object-json)

描述：返回对象中顶层的所有键。

返回类型：text

示例：

```
SELECT json_object_keys('{"f1":"abc","f2":{"f3":"a", "f4":"b"}, "f1":"abcd"}');
 json_object_keys
------------------
 f1
 f2
 f1
(3 rows)
```

json_each(object-json)

描述：将对象的每个键值对拆分转换成一行两列。

返回类型：setof(key text, value json)

示例：

```
SELECT * FROM json_each('{"f1":[1,2,3],"f2":{"f3":1},"f4":null}');
 key |  value
-----+----------
 f1  | [1,2,3]
 f2  | {"f3":1}
 f4  | null
(3 rows)
```

 版权所有 © 华为云计算技术有限公司

json_each_text(object-json)

描述：将对象的每个键值对拆分转换成一行两列。

返回类型：setof(key text, value text)

示例：

```
SELECT * FROM json_each_text('{"f1":[1,2,3],"f2":{"f3":1},"f4":null}');
 key |  value
-----+----------
 f1  | [1,2,3]
 f2  | {"f3":1}
 f4  |
(3 rows)
```

json_populate_record(anyelement, object-json [, bool])

描述：$1必须是一个复合类型的参数。将会把object-json里的每个对键值进行拆分，以键当做列名，与$1中的列名进行匹配查找，并填充到$1的格式中。

说明

JSON/JSONB函数中入参为复合类型时，可以使用CREATE TYPE或CREATE TABLE定义复合类型，例如：

```
CREATE TYPE jpop AS (a text, b INT, c timestamp);
CREATE TABLE jpop2(a text, b INT, c timestamp);
```

返回类型：anyelement

示例：

```
CREATE TYPE jpop AS (a text, b INT, c timestamp);
SELECT * FROM json_populate_record(null::jpop,'{"a":"blurfl","x":43.2}');
   a    | b | c
--------+---+---
 blurfl |   |
(1 row)
```

json_populate_recordset(anyelement, array-json [, bool])

描述：参考函数json_populate_record、jsonb_populate_record，对$2数组的每一个元素进行上述参数函数的操作，因此这也要求$2数组的每个元素都是object-json类型。

返回类型：setof anyelement

示例：

```
CREATE TYPE jpop AS (a text, b INT, c timestamp);
SELECT * FROM json_populate_recordset(null::jpop, '[{"a":1,"b":2},{"a":3,"b":4}]');
 a | b | c
---+---+---
 1 | 2 |
 3 | 4 |
(2 rows)
```

json_to_record(object-json)

描述：正如所有返回record的函数一样，调用者必须用一个AS子句显式地定义记录的结构。会将object-json的键值对进行拆分重组，把键当做列名，去匹配填充AS显示指定的记录的结构。

 版权所有 © 华为云计算技术有限公司

返回类型：record

示例：

```
SELECT * FROM json_to_record('{"a":1,"b":"foo","c":"bar"}'::json) as x(a int, b text, d text);
 a |  b  | d
---+-----+---
 1 | foo |
(1 row)
```

json_to_recordset(array-json)

描述：参考函数json_to_record，对数组内的每个元素，执行上述函数的操作，因此这要求数组内的每个元素都得是object-json。

返回类型：setof record

示例：

```
SELECT * FROM json_to_recordset('[{"a":1,"b":{"d":"foo"},"c":true},{"a":2,"c":false,"b":{"d":"bar"}}]') AS x(a
INT, b json, c BOOLEAN);
 a |      b      | c
---+-------------+---
 1 | {"d":"foo"} | t
 2 | {"d":"bar"} | f
(2 rows)

SELECT * FROM json_to_recordset('[{"a":1,"b":"foo","d":false},{"a":2,"b":"bar","c":true}]') AS x(a INT, b text, c
BOOLEAN);
 a |  b  | c
---+-----+---
 1 | foo |
 2 | bar | t
(2 rows)
```

json_typeof(json)

描述：检测json类型。

返回类型：text

示例：

```
SELECT value, json_typeof(value) from (values (json '123.4'), (json '"foo"'), (json 'true'), (json 'null'), (json
'[1, 2, 3]'), (json '{"x":"foo", "y":123}'), (NULL::json)) as data(value);
        value          | json_typeof
-----------------------+-------------
 123.4                 | number
 "foo"                 | string
 true                  | boolean
 null                  | null
 [1, 2, 3]             | array
 {"x":"foo", "y":123} | object
                       |
(7 rows)
```

jsonb_object(text[])

描述：从一个文本数组构造一个object-jsonb。这是个重载函数，当入参为一个文本数组的时候，其数组长度必须为偶数，成员被当做交替出现的键/值对。

返回类型：jsonb

示例：

 版权所有 © 华为云计算技术有限公司

```
SELECT jsonb_object('{a,1,b,2,3,NULL,"d e f","a b c"}');
                    jsonb_object
-----------------------------------------------------
 {"3": null, "a": "1", "b": "2", "d e f": "a b c"}
(1 row)
```

jsonb_object(text[], text[])

描述：两个文本数组的时候，第一个数组认为是键，第二个认为是值，两个数组长度必须相等。键不可为null。

返回类型：jsonb

示例：

```
SELECT jsonb_object('{a,b,"a b c"}', '{a,1,1}');
            jsonb_object
-------------------------------------
 {"a": "a", "b": "1", "a b c": "1"}
(1 row)
```

to_jsonb(anyment)

描述：将其他类型转换成对应的jsonb类型。

返回类型：jsonb

示例：

```
SELECT to_jsonb(1.1);
 to_jsonb
----------
 1.1
(1 row)
```

jsonb_agg

描述：将jsonb对象聚合成jsonb数组。

返回类型：jsonb

示例：

```
SELECT * FROM json_tbl_2;
 a |  b
---+-----
 1 | aaa
 1 | bbb
 2 | ccc
 2 | ddd
(4 rows)

SELECT a, jsonb_agg(b) FROM json_tbl_2 GROUP BY a ORDER BY a;
 a |   jsonb_agg
---+----------------
 1 | ["aaa", "bbb"]
 2 | ["ccc", "ddd"]
(2 rows)
```

jsonb_object_agg

描述：将键/值对聚集成一个JSON对象。

返回类型：jsonb

 版权所有 © 华为云计算技术有限公司

示例：

```
SELECT * FROM json_tbl_3;
 a |  b  | c
---+-----+----
 1 | aaa | 10
 1 | bbb | 20
 2 | ccc | 30
 2 | ddd | 40
(4 rows)
SELECT a, jsonb_object_agg(b, c) FROM json_tbl_3 GROUP BY a ORDER BY a;
 a |     jsonb_object_agg
---+-------------------------
 1 | {"aaa": 10, "bbb": 20}
 2 | {"ccc": 30, "ddd": 40}
(2 rows)
```

jsonb_build_array([VARIADIC "any"])

描述：从一个可变参数列表构造一个可能包含异质类型的JSON数组。

返回类型：jsonb

示例：

```
SELECT jsonb_build_array('a',1,'b',1.2,'c',true,'d',null,'e',json '{"x": 3, "y": [1,2,3]}','');
                              jsonb_build_array
-------------------------------------------------------------------------------
 ["a", 1, "b", 1.2, "c", true, "d", null, "e", {"x": 3, "y": [1, 2, 3]}, null]
(1 row)
```

jsonb_build_object([VARIADIC "any"])

描述：从一个可变参数列表构造出一个JSON对象，其入参必须为偶数个，两两一组组成键值对。注意键不可为null。

返回类型：jsonb

示例：

```
SELECT jsonb_build_object(1,2);
 jsonb_build_object
--------------------
 {"1": 2}
(1 row)
```

jsonb_strip_nulls(jsonb)

描述：所有具有空值的对象字段均被省略。其他空值保持不变。

返回类型：jsonb

示例：

```
SELECT jsonb_strip_nulls('[{"f1":1,"f2":null},2,null,3]');
    jsonb_strip_nulls
--------------------------
 [{"f1": 1}, 2, null, 3]
(1 row)
```

jsonb_object_field(jsonb, text)

描述：同操作符->, 返回对象中指定键对应的值。

 版权所有 © 华为云计算技术有限公司

返回类型：jsonb

示例：

```
SELECT jsonb_object_field('{"a": {"b":"foo"}}','a');
 jsonb_object_field
--------------------
 {"b": "foo"}
(1 row)
```

jsonb_object_field_text(jsonb, text)

描述：同操作符->>, 返回对象中指定键对应的值。

返回类型：text

示例：

```
SELECT jsonb_object_field_text('{"a": {"b":"foo"}}','a');
 jsonb_object_field_text
-------------------------
 {"b": "foo"}
(1 row)
```

jsonb_array_element(array-jsonb, integer)

描述：同操作符->, 返回数组中指定下标的元素。

返回类型：jsonb

示例：

```
SELECT jsonb_array_element('[1,true,[1,[2,3]],null]',2);
 jsonb_array_element
---------------------
 [1, [2, 3]]
(1 row)
```

jsonb_array_element_text(array-jsonb, integer)

描述：同操作符->>, 返回数组中指定下标的元素。

返回类型：text

示例：

```
SELECT jsonb_array_element_text('[1,true,[1,[2,3]],null]',2);
 jsonb_array_element_text
--------------------------
 [1, [2, 3]]
(1 row)
```

jsonb_extract_path(jsonb, VARIADIC text[])

描述：等价于操作符#>，返回$2所指路径的值。

返回类型：jsonb

示例：

```
SELECT jsonb_extract_path('{"f2":{"f3":1},"f4":{"f5":99,"f6":"stringy"}}', 'f4','f6');
 jsonb_extract_path
--------------------
```

 版权所有 © 华为云计算技术有限公司

```
 "stringy"
(1 row)
```

jsonb_extract_path_text(jsonb, VARIADIC text[])

描述：等价于操作符#>>，返回$2所指路径的值。

返回类型：text

示例：

```
SELECT jsonb_extract_path_text('{"f2":{"f3":1},"f4":{"f5":99,"f6":"stringy"}}', 'f4','f6');
 jsonb_extract_path_text
-------------------------
 stringy
(1 row)
```

jsonb_extract(jsonb, VARIADIC text[])

描述：输入任意的object-jsonb类型、array-jsonb类型，返回$2所指路径的值。该函数仅9.1.0及以上集群版本支持。

返回类型：SETOF jsonb

示例：

```
SELECT jsonb_extract('{"f2":{"f3":1},"f4":[{"f5":99},{"f6":"stringy"}]}','f2','f3');
 jsonb_extract
---------------
 1
(1 row)

SELECT jsonb_extract('{"f2":{"f3":1},"f4":[{"f5":99},{"f5":"stringy"}]}', 'f4','f5');
 jsonb_extract
---------------
 99
 "stringy"
(2 rows)
```

jsonb_extract_text(jsonb, VARIADIC text[])

描述：输入任意的object-jsonb类型、array-jsonb类型，返回$2所指路径的值。该函数仅9.1.0及以上集群版本支持。

返回类型：SETOF text

示例：

```
SELECT jsonb_extract_text('{"f2":{"f3":1},"f4":[{"f5":99},{"f6":"stringy"}]}','f2','f3');
 jsonb_extract_text
--------------------
 1
(1 row)

SELECT jsonb_extract_text('{"f2":{"f3":1},"f4":[{"f5":99},{"f5":"stringy"}]}', 'f4','f5');
 jsonb_extract_text
--------------------
 99
 stringy
(2 rows)
```

jsonb_array_elements(array-jsonb)

描述：拆分数组，每一个元素返回一行。

 版权所有 © 华为云计算技术有限公司

返回类型：jsonb

示例：

```
SELECT jsonb_array_elements('[1,true,[1,[2,3]],null]');
 jsonb_array_elements
----------------------
 1
 true
 [1, [2, 3]]
 null
(4 rows)
```

jsonb_array_elements_text(array-jsonb)

描述：拆分数组，每一个元素返回一行。

返回类型：text

示例：

```
SELECT * FROM jsonb_array_elements_text('[1,true,[1,[2,3]],null]');
   value
-------------
 1
 true
 [1, [2, 3]]

(4 rows)
```

jsonb_array_length(array-jsonb)

描述：返回数组长度。

返回类型：integer

示例：

```
SELECT jsonb_array_length('[1,2,3,{"f1":1,"f2":[5,6]},4,null]');
 jsonb_array_length
--------------------
          6
(1 row)
```

jsonb_object_keys(object-jsonb)

描述：返回对象中顶层的所有键。

返回类型：SETOF text

示例：

```
SELECT jsonb_object_keys('{"f1":"abc","f2":{"f3":"a", "f4":"b"}, "f1":"abcd"}');
 jsonb_object_keys
-------------------
 f1
 f2
(2 rows)
```

jsonb_each(object-jsonb)

描述：将对象的每个键值对拆分转换成一行两列。

返回类型：setof(key text, value jsonb)

 版权所有 © 华为云计算技术有限公司

示例：

```
SELECT * FROM jsonb_each('{"f1":[1,2,3],"f2":{"f3":1},"f4":null}');
 key |   value
-----+-----------
 f1  | [1, 2, 3]
 f2  | {"f3": 1}
 f4  | null
(3 rows)
```

jsonb_each_text(object-jsonb)

描述：将对象的每个键值对拆分转换成一行两列。

返回类型：setof(key text, value text)

示例：

```
SELECT * FROM jsonb_each_text('{"f1":[1,2,3],"f2":{"f3":1},"f4":null}');
 key |   value
-----+-----------
 f1  | [1, 2, 3]
 f2  | {"f3": 1}
 f4  |
(3 rows)
```

jsonb_populate_record(anyelement, object-jsonb [, bool])

描述：$1必须是一个复合类型的参数。将会把object-json里的每个对键值进行拆分，以键当做列名，与$1中的列名进行匹配查找，并填充到$1的格式中。

返回类型：anyelement

示例：

```
SELECT * FROM jsonb_populate_record(null::jpop,'{"a":"blurfl","x":43.2}');
   a    | b | c
--------+---+---
 blurfl |   |
(1 row)
```

jsonb_populate_recordset(anyelement, array-jsonb [, bool])

描述：参考上述函数json_populate_record、jsonb_populate_record，对$2数组的每一个元素进行上述参数函数的操作，因此这也要求$2数组的每个元素都是object-json类型。

返回类型：setof anyelement

示例：

```
SELECT * FROM jsonb_populate_recordset(null::jpop, '[{"a":1,"b":2},{"a":3,"b":4}]');
 a | b | c
---+---+---
 1 | 2 |
 3 | 4 |
(2 rows)
```

jsonb_to_record(object-json)

描述：正如所有返回record的函数一样，调用者必须用一个AS子句显式地定义记录的结构。会将object-json的键值对进行拆分重组，把键当做列名，去匹配填充AS显示指定的记录的结构。

 版权所有 © 华为云计算技术有限公司

返回类型：record

示例：

```
SELECT * FROM jsonb_to_record('{"a":1,"b":"foo","c":"bar"}'::jsonb) as x(a int, b text, d text);
 a |  b  | d
---+-----+---
 1 | foo |
(1 row)
```

jsonb_to_recordset(array-json)

描述：参考函数jsonb_to_record，对数组内每个元素，执行上述函数的操作，因此这要求数组内的每个元素都得是object-jsonb。

返回类型：setof record

示例：

```
SELECT * FROM jsonb_to_recordset('[{"a":1,"b":"foo","d":false},{"a":2,"b":"bar","c":true}]') AS x(a INT, b text,
c boolean);
 a |  b  | c
---+-----+---
 1 | foo |
 2 | bar | t
(2 rows)
```

jsonb_typeof(jsonb)

描述：检测jsonb类型。

返回类型：text

示例：

```
SELECT jsonb_typeof(to_jsonb(1.1));
 jsonb_typeof
--------------
 number
(1 row)
```

jsonb_ne(jsonb, jsonb)

描述：同操作符<>，比较两个值的大小。

返回类型：bool

示例：

```
SELECT jsonb_ne('{"a":1, "b":2}'::jsonb, '{"a":1, "b":3}'::jsonb);
 jsonb_ne
----------
 t
(1 row)
```

jsonb_lt(jsonb, jsonb)

描述：同操作符<，比较两个值的大小。

返回类型：bool

示例：

 版权所有 © 华为云计算技术有限公司

```
SELECT jsonb_lt('{"a":1, "b":2}'::jsonb, '{"a":1, "b":3}'::jsonb);
 jsonb_lt
----------
 t
(1 row)
```

jsonb_gt(jsonb, jsonb)

描述：同操作符>，比较两个值的大小。

返回类型：bool

示例：

```
SELECT jsonb_gt('{"a":1, "b":2}'::jsonb, '{"a":1, "b":3}'::jsonb);
 jsonb_gt
----------
 f
(1 row)
```

jsonb_le(jsonb, jsonb)

描述：同操作符<=，比较两个值的大小。

返回类型：bool

示例：

```
SELECT jsonb_le('["a", "b"]', '{"a":1, "b":2}');
 jsonb_le
----------
 t
(1 row)
```

jsonb_ge(jsonb, jsonb)

描述：同操作符>=，比较两个值的大小。

返回类型：bool

示例：

```
SELECT jsonb_ge('["a", "b"]', '{"a":1, "b":2}');
 jsonb_ge
----------
 f
(1 row)
```

jsonb_eq(jsonb, jsonb)

描述：同操作符=，比较两个值的大小。

返回类型：bool

示例：

```
SELECT jsonb_eq('["a", "b"]', '{"a":1, "b":2}');
 jsonb_eq
----------
 f
(1 row)
```

 版权所有 © 华为云计算技术有限公司

jsonb_cmp(jsonb, jsonb)

描述：比较大小，正数表示大于，负数表示小于，0表示相等。

返回类型：integer

示例：

```
SELECT jsonb_cmp('["a", "b"]', '{"a":1, "b":2}');
 jsonb_cmp
-----------
 -1
(1 row)
```

jsonb_exists(jsonb, text)

描述：同操作符?，字符串$2是否存在$1的顶层以key\elem\scalar的形式存在。

返回类型：bool

示例：

```
SELECT jsonb_exists('["1",2,3]', '1');
 jsonb_exists
--------------
 t
(1 row)
```

jsonb_exists_any(jsonb, text[])

描述：同操作符?|，字符串数组$2里面是否存在的元素，在$1的顶层以key\elem\scalar的形式存在。

返回类型：

示例：

```
SELECT jsonb_exists_any('["1","2",3]', '{1, 2, 4}');
 jsonb_exists_any
------------------
 t
(1 row)
```

jsonb_exists_all(jsonb, text[])

描述：同操作符?&，字符串数组$2里面是否所有的元素，都在$1的顶层以key\elem\scalar的形式存在。

返回类型：

bool

示例：

```
SELECT jsonb_exists_all('["1","2",3]', '{1, 2}');
 jsonb_exists_all
------------------
 t
(1 row)
```

jsonb_contained(jsonb, jsonb)

描述：同操作符<@, 判断$1中的所有元素是否在$2的顶层存在。

 版权所有 © 华为云计算技术有限公司

返回类型：bool

示例：

```
SELECT jsonb_contained('[1,2,3]', '[1,2,3,4]');
 jsonb_contained
-----------------
 t
(1 row)
```

jsonb_contains(jsonb, jsonb)

描述：同操作符@>, 判断$1中的顶层所有元素是否包含在$2的所有元素。

返回类型：bool

示例：

```
SELECT jsonb_contains('{"a":1, "b":2, "c":3}'::jsonb, '{"a":1}');
 jsonb_contains
-----------------
 t
(1 row)
```

jsonb_concat(jsonb, jsonb)

描述：连接两个jsonb对象为一个jsonb。

返回类型：jsonb

示例：

```
SELECT jsonb_concat('{"a":1, "b":2}'::jsonb, '{"c":3, "d":4}'::jsonb);
          jsonb_concat
-----------------------------------
 {"a": 1, "b": 2, "c": 3, "d": 4}
(1 row)
```

jsonb_delete(jsonb, text)

描述：删除jsonb中的key值对应的键值对。

返回类型：jsonb

示例：

```
SELECT jsonb_delete('{"a":1, "b":2}'::jsonb, 'a');
 jsonb_delete
--------------
 {"b": 2}
(1 row)
```

jsonb_delete_idx(jsonb, text)

描述：删除数组下标对应的元素。

返回类型：jsonb

示例：

```
SELECT jsonb_delete_idx('[0,1,2,3,4]'::jsonb, 2);
 jsonb_delete_idx
------------------
```

 版权所有 © 华为云计算技术有限公司

```
 [0, 1, 3, 4]
(1 row)
```

jsonb_delete_array(jsonb, VARIADIC text[])

描述：删除jsonb数组中的多个元素。

返回类型：jsonb

示例：

```
SELECT jsonb_delete_array('["a", "b", "c"]'::jsonb , 'a', 'b');
 jsonb_delete_array
--------------------
 ["c"]
(1 row)
```

jsonb_delete_path(jsonb, text[])

描述：删除jsonb数组中指定路径的元素。

返回类型：jsonb

示例：

```
SELECT jsonb_delete_path('{"a":{"b":{"c":1, "d":2}}, "e":3}'::jsonb , array['a', 'b']);
 jsonb_delete_path
-------------------
 {"a": {}, "e": 3}
(1 row)
```

jsonb_set(target jsonb, path text[], new_value jsonb [, create_missing boolean])

描述：返回target，用path指定的部分被new_value替换，或者如果create_missing为true（默认值为true）且path指定的项不存在，则添加new_value。与面向路径的运算符一样，path中出现的负整数从JSON数组的末尾开始计数。

返回类型：jsonb

示例：

```
SELECT jsonb_set('[{"f1":1,"f2":null},2,null,3]', '{0,f1}','[2,3,4]', false);
                jsonb_set
-----------------------------------------------
 [{"f1": [2, 3, 4], "f2": null}, 2, null, 3]
(1 row)
```

jsonb_pretty(jsonb)

描述：以缩进的JSON文本形式返回。

返回类型：jsonb

示例：

```
SELECT jsonb_pretty('{"a":{"b":{"c":1, "d":2}}, "e":3}'::jsonb);
   jsonb_pretty
----------------------
 {                   +
    "a": {           +
       "b": {        +
          "c": 1,+
          "d": 2 +
```

 版权所有 © 华为云计算技术有限公司

```
      }         +
   },           +
   "e": 3        +
}
(1 row)
```

jsonb_insert(target jsonb, path text[], new_value jsonb [, insert_after boolean])

描述：返回被插入了new_value的target。如果path指定的target部分位于JSONB数组中，则new_value将被插入到目标之前（insert_after为false，默认值）或者之后（insert_after为true）。如果path指定的target部分位于一个JSONB对象内，则仅当target不存在时才插入new_value。与面向路径的运算符一样，path中出现的负整数表示从JSON数组的末尾开始计数。

返回类型：jsonb

示例：

```
SELECT jsonb_insert('{"a": [0,1,2]}', '{a, 1}', '"new_value"');
        jsonb_insert
--------------------------------
 {"a": [0, "new_value", 1, 2]}
(1 row)
```

ts_headline([config regconfig,] document jsonb, query tsquery [, options text])

描述：高亮jsonb搜索结果。

返回类型：jsonb

示例：

```
SELECT ts_headline('english',
'[{"id":9928,"user_id":4562,"user_name":"9LOHR4","create_time":"2021-06-22T16:28:16.504518+08:00"},
{"id":9959,"user_id":5524,"user_name":"YID07D","create_time":"2021-06-22T16:28:16.557228+08:00"},
{"id":9962,"user_id":7991,"user_name":"7C6QOM","create_time":"2021-06-22T16:28:16.56234+08:00"}]'::jsonb,
 to_tsquery('english', '9LOHR4'), 'StartSel = <, StopSel = >');

ts_headline

--------------------------------
```

```
 'b':2 'c':4
(1 row)
```

7.15 安全函数

gs_password_deadline()

描述：显示当前账户距离密码过期的时间。密码过期后提示用户修改密码。与GUC参数password_effect_time相关。

返回值类型：interval

示例：

```
SELECT gs_password_deadline();
  gs_password_deadline
------------------------
 83 days 17:44:32.196094
(1 row)
```

gs_password_expiration()

描述：显示当前账户距离密码过期的时间。密码过期后用户无法登录数据库。与创建用户的DDL语句PASSWORD EXPIRATION period相关，函数返回值大于等于-1，如果创建用户时未指定PASSWORD EXPIRATION period，该函数的缺省值为-1，表示没有过期限制。

返回值类型：interval

示例：

```
SELECT gs_password_expiration();
  gs_password_expiration
-------------------------
 29 days 23:59:49.731482
(1 row)
```

login_audit_messages(flag boolean)

描述：查看登录用户的登录信息。

返回值类型：元组

示例：

- 查看上一次登录认证通过的日期、时间和IP等信息：

```
SELECT * FROM login_audit_messages(true);
  username  | database |       logintime        |     type      | result |  client_conninfo   |
session_id
------------+----------+------------------------+---------------+--------+--------------------
+----------------------------------------
    dbadmin    | gaussdb | 2017-06-02 15:28:34+08 | login_success | ok     | gsql@[local]    |
1663745305.140192538154752.coordinator1
(1 row)
```

- 查看上一次登录认证失败的日期、时间和IP等信息：

```
SELECT * FROM login_audit_messages(false) ORDER BY logintime desc limit 1;
  username  | database |       logintime        |     type      | result |     client_conninfo     | session_id
------------+----------+------------------------+---------------+--------+-------------------------+------------
(0 rows)
```

- 查看自从最后一次认证通过以来失败的尝试次数、日期和时间：

 版权所有 © 华为云计算技术有限公司

```
SELECT * FROM login_audit_messages(false);
 username | database |      logintime      |    type    | result |    client_conninfo     | session_id
----------+----------+---------------------+------------+--------+------------------------+------------
(0 rows)
```

login_audit_messages_pid(flag boolean)

描述：查看登录用户的登录信息。与login_audit_messages的区别在于结果基于当前backendid向前查找。所以不会因为同一用户的后续登录，而影响本次登录的查询结果。也就是查询不到该用户后续登录的信息。

返回值类型：元组

示例：

- 查看上一次登录认证通过的日期、时间和IP等信息：

```
SELECT * FROM login_audit_messages_pid(true);
 username | database |      logintime      |    type    | result |  client_conninfo | backendid
|             session_id
----------+----------+---------------------+------------+--------+-------------------
+----------------------------------------
    dbadmin    | postgres | 2017-06-02 15:28:34+08 | login_success | ok     | gsql@[local] |
140311900702464 | 1663745305.140192538154752.coordinator1
(1 row)
```

- 查看上一次登录认证失败的日期、时间和IP等信息：

```
SELECT * FROM login_audit_messages_pid(false) ORDER BY logintime desc limit 1;
 username | database |      logintime      |    type    | result |    client_conninfo    | backendid |
session_id
----------+----------+---------------------+------------+--------+------------------------+-----------
(0 rows)
```

- 查看自从最后一次认证通过以来失败的尝试次数、日期和时间：

```
SELECT * FROM login_audit_messages_pid(false);
 username | database |      logintime      |    type    | result |    client_conninfo    | backendid |
session_id
----------+----------+---------------------+------------+--------+------------------------+-----------
(0 rows)
```

pg_query_audit()

描述：查看当前CN节点审计日志。

返回值类型：record

函数返回字段如下：

表 7-23 pg_query_audit()函数返回字段

名称	类型	描述
begintime	timestamp with time zone	操作的执行开始时间。
endtime	timestamp with time zone	操作的执行结束时间。
operation_type	text	操作类型，具体类型见**表7-24**。
audit_type	text	审计类型，具体类型见**表7-25**。
result	text	操作结果。

版权所有 © 华为云计算技术有限公司

名称	类型	描述
username	text	执行操作的用户名。
database	text	数据库名称。
client_conninfo	text	客户端连接信息，即gsql，jdbc或odbc。
object_name	text	操作对象名称。
object_details	text	记录语句中涉及的列及其使用类型。
command_text	text	操作的执行命令。
detail_info	text	执行操作详细信息。
transaction_xid	text	事务ID。
query_id	text	查询ID。
node_name	text	节点名称。
session_id	text	会话ID。
local_port	text	本地端口。
remote_port	text	远端端口。
result_rows	text	语句执行返回的记录数。
error_code	text	语句执行错误时的错误码。

表 7-24 operation_type 操作类型项

操作类型	描述
audit_switch	表示对用户打开和关闭审计日志操作场景进行审计。
login_logout	表示对用户登录和登出操作场景进行审计。
system	表示对系统的启停、实例切换操作场景进行审计。
sql_parse	表示对SQL语句解析场景进行审计。
user_lock	表示对用户锁定和解锁操作的场景进行审计。
grant_revoke	表示对用户权限授予和回收操作场景进行审计。
violation	表示对用户访问存在越权的场景进行审计。
ddl	表示对DDL操作场景进行审计，因为DDL操作会根据操作对象进行更细粒度控制，仍然沿用审计开关audit_system_object，即由audit_system_object控制对哪些对象的DDL操作进行审计（此处不配置ddl，只要配置了audit_system_object，审计也会生效）。
dml	表示对DML操作场景进行审计。

版权所有 © 华为云计算技术有限公司

操作类型	描述
select	表示对SELECT操作场景进行审计。
internal_event	表示对内部事件操作场景进行审计。
user_func	表示对用户自定义函数、存储过程、匿名块操作场景进行审计。 **说明** 如果自定义函数、存储过程中有fetch语句，则审计fetch语句时，其中common_text字段记录的为其对应的CURSOR内容。
special_func	表示对特殊函数调用操作场景进行审计，特殊函数包括：pg_terminate_backend和pg_cancel_backend。
copy	表示对COPY操作场景进行审计。
set	表示对SET操作场景进行审计。
transaction	表示对事务操作场景进行审计。
vacuum	表示对VACUUM操作场景进行审计。
analyze	表示对ANALYZE操作场景进行审计。
cursor	表示对游标操作的场景进行审计。
anonymous_block	表示对匿名块操作场景进行审计。
explain	表示对EXPLAIN操作场景进行审计。
show	表示对SHOW操作场景进行审计。
lock_table	表示对锁表操作场景进行审计。
comment	表示对COMMENT操作场景进行审计。
preparestmt	表示对PREPARE、EXECUTE、DEALLOCATE操作场景进行审计。
cluster	表示对CLUSTER操作场景进行审计。
constraints	表示对CONSTRAINTS操作场景进行审计。
checkpoint	表示对CHECKPOINT操作场景进行审计。
barrier	表示对BARRIER操作场景进行审计。
cleanconn	表示对CLEAN CONNECTION操作场景进行审计。
seclabel	表示对安全标签操作进行审计。
notify	表示对通知操作进行审计。
load	表示对加载操作进行审计。
discard	表示对清理当前会话所有的全局临时表信息场景进行审计。

 版权所有 © 华为云计算技术有限公司

表 7-25 audit_type 审计类型项

审计类型	描述
audit_open/audit_close	表示审计类型为打开和关闭审计日志操作。
user_login/user_logout	表示审计类型为用户登录/退出成功的操作和用户。
system_start/system_stop/ system_recover/system_switch	表示审计类型为系统的启停、实例切换操作。
sql_wait/sql_parse	表示审计类型为SQL语句解析。
lock_user/unlock_user	表示审计类型为用户锁定和解锁成功的操作。
grant_role/revoke_role	表示审计类型为用户权限授予和回收的操作。
user_violation	表示审计类型为用户访问存在越权的操作。
ddl_数据库对象	表示审计类型为DDL操作，因为DDL操作由会根据操作对象进行更细粒度控制，仍然沿用审计开关audit_system_object，即由audit_system_object控制对哪些对象的DDL操作进行审计（此处不配置ddl，只要配置了audit_system_object，审计也会生效）。 例如：ddl_sequence表示审计类型为序列相关操作。
dml_action_insert/ dml_action_delete/ dml_action_update/ dml_action_merge/ dml_action_select	表示审计类型为INSERT、DELETE、UPDATE、MERGE等DML操作。
internal_event	表示审计类型为内部事件。
user_func	表示审计类型为用户自定义函数、存储过程、匿名块操作。 **说明** 如果自定义函数、存储过程中有fetch语句，则审计fetch语句时，其中common_text字段记录的为其对应的CURSOR内容。
special_func	表示审计类型为特殊函数调用操作，特殊函数包括：pg_terminate_backend和pg_cancel_backend。
copy_to/copy_from	表示审计类型为COPY相关操作。
set_parameter	表示审计类型为SET操作。

 版权所有 © 华为云计算技术有限公司

审计类型	描述
trans_begin/trans_commit/ trans_prepare/ trans_rollback_to/ trans_release/trans_savepoint/ trans_commit_prepare/ trans_rollback_prepare/ trans_rollback	表示审计类型为事务相关操作。
vacuum/vacuum_full/ vacuum_merge	表示审计类型为VACUUM相关操作。
analyze/analyze_verify	表示审计类型为ANALYZE相关操作。
cursor_declare/cursor_move/ cursor_fetch/cursor_close	表示审计类型为游标相关操作。
codeblock_execute	表示审计类型为匿名块。
explain	表示审计类型为EXPLAIN操作。
show	表示审计类型为SHOW操作。
lock_table	表示审计类型为锁表操作。
comment	表示审计类型为COMMENT操作。
prepare/execute/deallocate	表示审计类型为PREPARE、EXECUTE或DEALLOCATE操作。
cluster	表示审计类型为CLUSTER操作。
constraints	表示审计类型为CONSTRAINTS操作。
checkpoint	表示审计类型为CHECKPOINT操作。
barrier	表示审计类型为BARRIER操作。
cleanconn	表示审计类型为CLEAN CONNECTION操作。
seclabel	表示审计类型为安全标签操作。
notify	表示审计类型为通知操作。
load	表示审计类型为加载操作。
discard	表示审计类型为DISCARD操作。

pgxc_query_audit()

描述：查看所有CN节点审计日志。

返回值类型：record

函数返回字段同pg_query_audit函数。

 版权所有 © 华为云计算技术有限公司

pg_query_audit_details()

描述：查看审计日志并将审计日志中的object_name和object_details字段由json格式解析出来。该函数仅8.2.1.100及以上集群版本支持。

返回值类型：record

函数返回字段如下：

表 7-26 pg_query_audit()函数返回字段

名称	类型	描述
begintime	timestamp with time zone	操作的执行开始时间。
endtime	timestamp with time zone	操作的执行结束时间。
operation_type	text	操作类型，具体类型见**表7-24**。
audit_type	text	审计类型，具体类型见**表7-25**。
result	text	操作结果。
username	text	执行操作的用户名。
database	text	数据库名称。
client_conninfo	text	客户端连接信息，即gsql，jdbc或odbc。
transaction_xid	text	事务ID。
query_id	text	查询ID。
node_name	text	节点名称。
session_id	text	会话ID。
local_port	text	本地端口。
remote_port	text	远端端口。
object_name	text	表名、函数名、视图名。
column_name	text	列名。

版权所有 © 华为云计算技术有限公司

名称	类型	描述
type_of_use	text	对象的使用类型： 1：仅涉及（在实际使用中，此标志位暂不存在） 2：执行过程中访问（语句中出现的列以及在解析过程中访问到的列） 4：条件中发现（条件类型语句中，不包括条件表达式及函数） 8：inner join中发现 16：outer join中发现 32：聚合节点中发现（包括distinct、group by、聚集函数） 64：full join中发现 该列数值为叠加显示的数值。
use_type	text	type_of_use解析出的具体类型： 1：Reference only 2：Access 4：Conditional 8：Inner join 16：Outer join 32：Sum 64：Full join
command_text	text	操作的执行命令。

示例：

1. 查询审计语句中所有对象的列及其在语句中使用的类型：

```
SET audit_object_details = on;
SELECT object_name,object_details,result_rows,error_code,command_text FROM
pg_query_audit('2023-05-12-03 8:00:00','2023-05-12 22:55:00') where command_text like '%student%';
```

查询结果如下：

```
                        object_name                         |
object_details                                | result_rows | error_code
|                                                    command_text
-------------------------------------------------------------------
+-----------------------------------------------------------------------------------------+-------------+-----------
+-----------------------------------------------------------------------------------------------------------------
-----------------------------------
 student
|                                                                 | 0           |           | CREATE TABLE
student(stuNo int, stuName TEXT);
 studentscore
|                                                                 | 0           |           | CREATE TABLE
studentscore(stuNo int, stuscore int);
 ["public.student_view01","public.studentscore","public.student"]
|                                                                 | 0           |           | CREATE OR
REPLACE VIEW student_view01 AS SELECT * FROM student t1 where t1.stuNo in (select stuNo from
studentscore t2 where t1.stuNo = t2.stuNo);
 ["public.student_view01","public.student","public.studentscore"] | {"public.student":[{"stuno":"6"},
```

 版权所有 © 华为云计算技术有限公司

```
{"stuname":"2"}],"public.studentscore":[{"stuno":"6"}]} | 0           |           | SELECT * FROM
student_view01
```

2. 查询结果object_details中显示部分语句执行过程中，涉及到的列及其使用类型，使用json格式进行记录。

   使用pg_query_audit_details函数对object_name和object_details列进行解析：

```
SELECT database,object_name,column_name,type_of_use,use_type FROM
pg_query_audit_details('2021-02-03 8:00:00','2024-02-03 22:55:00','current') where command_text like
'%student%';
```

   查询结果如下：

```
 database |      object_name      | column_name | type_of_use |      use_type
----------+-----------------------+-------------+-------------+--------------------
 gaussdb  | student               |             | 2           | Access
 gaussdb  |                       |             | 2           | Access
 gaussdb  | studentscore          |             | 2           | Access
 gaussdb  |                       |             | 2           | Access
 gaussdb  | public.student_view01 |             | 2           | Access
 gaussdb  | public.studentscore   |             | 2           | Access
 gaussdb  | public.student        |             | 2           | Access
 gaussdb  |                       |             | 2           | Access
 gaussdb  | public.student        | stuno       | 6           | Access,Conditional
 gaussdb  | public.student        | stuname     | 2           | Access
 gaussdb  | public.studentscore   | stuno       | 6           | Access,Conditional
 gaussdb  | public.student_view01 |             | 2           | Access
 gaussdb  | public.student        |             | 2           | Access
 gaussdb  | public.studentscore   |             | 2           | Access
 gaussdb  |                       |             | 2           | Access
(15 rows)
```

pg_delete_audit()

描述：删除指定时间段的审计日志。8.1.3及以上集群版本中该函数已废弃。 返回值类型：void

说明

基于数据库安全考虑，不提供删除指定时间段的审计日志的函数接口，调用该函数将直接报“ERROR: For security purposes, it is not allowed to manually delete audit logs”。

create_audit_csv_foregion_table()

描述：创建读审计日志的分区外表。用户调用该函数生成审计日志外表pgxc_audit_logs，通过读取pgxc_audit_logs可以读取放在obs上的审计日志信息。该函数仅8.3.0及以上集群支持。

入参：

- obs_server：text类型，obs server名称
- file_path：text类型，obs文件路径

返回值类型：record

示例：

```
--创建obs server
CREATE SERVER obs_server FOREIGN DATA WRAPPER DFS_FDW OPTIONS (
  address 'obs.example.xxx:xxx',
  encrypt 'true',
  access_key 'xxxxxxxxx',
  secret_access_key 'xxxxxxxxxxxxxx',
  type 'obs' );
--调用该函数生成审计日志外表
SELECT * FROM pg_catalog.create_audit_csv_foregion_table('obs_server','/obs-audit/test/');
```

 版权所有 © 华为云计算技术有限公司

```
 create_audit_csv_foregion_table
---------------------------------
 t
(1 row)
--读取放在obs上的审计日志信息
SELECT * FROM pgxc_audit_logs where year=2023 and month=7 and date=24;
 year | month | date |          begintime          |          endtime            | operation_type |    audit_type     | result
| username | database | client_conninfo |          object_name          | object_details |command_text|
detail_info | transaction_xid |      query_id      | node_name |             session_id              | local_port |
remote_port | result_rows | error_code
------+-------+------+-----------------------------+-----------------------------+----------------+-------------------
+--------+----------+----------+-----------------+---------------------------------+----------------
+-------------------------------------------------------------------------------------------------------------------------------------------------------------
-------------------------------------------------------------------------------------------------------------------------------------------------------------
-------------------------------------------------------------------+------------+----------------+------------+-------------------
+-----------+-----------------------------------+------------+------------+------------+------------
 2023 | 7     | 24   | 2023-07-24 09:56:40.727+08 | 2023-07-24 09:56:42.215+08 | dml            |
dml_action_select | ok     | dbadmin  | gaussdb  | gsql@[local]    | public.pgxc_audit_logs         |               |
select * f
rom pgxc_audit_logs where year=2023 and date=24 and month=7;
```

7.16 条件表达式函数

coalesce(expr1, expr2, ..., exprn)

描述：返回参数列表中第一个非NULL的参数值。

COALESCE(expr1, expr2) 等价于CASE WHEN expr1 IS NOT NULL THEN expr1 ELSE expr2 END。

示例：

```
SELECT coalesce(NULL,'hello');
 coalesce
----------
 hello
(1 row)
```

说明

- 如果表达式列表中的所有表达式都等于NULL，则本函数返回NULL。
- 它常用于在显示数据时用缺省值替换NULL。
- 和CASE表达式一样，COALESCE不会计算不需要用来判断结果的参数；即在第一个非空参数右边的参数不会被计算。

decode(base_expr, compare1, value1, Compare2,value2, … default)

描述：把base_expr与后面的每个compare(n) 进行比较，如果匹配返回相应的value(n)。如果没有发生匹配，则返回default。

示例：

```
SELECT decode('A','A',1,'B',2,0);
 case
------
 1
(1 row)
```

if(bool_expr, expr1, expr2)

描述：当bool_expr为true时，返回expr1，否则返回expr2。

 版权所有 © 华为云计算技术有限公司

if(bool_expr, expr1, expr2) 等价于CASE WHEN bool_expr = true THEN expr1 ELSE expr2 END。

示例：

```
SELECT if(1 < 2, 'yes', 'no');
 if
-----
 yes
(1 row)
```

说明

参数expr1和expr2可以为任意类型，返回结果类型规则请参考**UNION，CASE和相关构造**。

ifnull(expr1, expr2)

描述：当expr1不为NULL时，返回expr1，否则返回expr2。

ifnull(expr1, expr2) 逻辑上等价于CASE WHEN expr1 IS NOT NULL THEN expr1 ELSE expr2 END。

示例：

```
SELECT ifnull(NULL,'hello');
 ifnull
--------
 hello
(1 row)
```

说明

参数expr1和expr2可以为任意类型，返回结果类型规则请参考**UNION，CASE和相关构造**。

isnull(expr)

描述：当expr为NULL时，返回true，否则返回false。

isnull(expr) 逻辑上等价于expr IS NULL。

示例：

```
SELECT isnull(NULL), isnull('abc');
 isnull | isnull
--------+--------
 t      | f
(1 row)
```

nullif(expr1, expr2)

描述：当且仅当expr1和expr2相等时，NULLIF才返回NULL，否则它返回expr1。

nullif(expr1, expr2) 逻辑上等价于CASE WHEN expr1 = expr2 THEN NULL ELSE expr1 END。

示例：

```
SELECT nullif('hello','world');
 nullif
--------
 hello
(1 row)
```

备注：

 版权所有 © 华为云计算技术有限公司

如果两个参数的数据类型不同，则：

- 两种数据类型之间存在隐式转换，则以其中优先级较高的数据类型为基准将另一个参数隐式转换成该类型，转换成功则进行计算，转换失败则返回错误。如：

```
SELECT nullif('1234'::VARCHAR,123::INT4);
 nullif
--------
   1234
(1 row)
SELECT nullif('1234'::VARCHAR,'2012-12-24'::DATE);
ERROR:  invalid input syntax for type timestamp: "1234"
```

- 两种数据类型之间不存在隐式转换，则返回错误 。如：

```
SELECT nullif(TRUE::BOOLEAN,'2012-12-24'::DATE);
ERROR:  operator does not exist: boolean = timestamp without time zone
LINE 1: SELECT nullif(TRUE::BOOLEAN,'2012-12-24'::DATE) FROM DUAL;
^
HINT:  No operator matches the given name and argument type(s). You might need to add explicit type casts.
```

nvl(expr1 , expr2)

描述：如果expr1为NULL则返回expr2。如果expr1非NULL，则返回expr1。

示例：

```
SELECT nvl('hello','world');
  nvl
-------
 hello
(1 row)
```

说明

参数expr1和expr2可以为任意类型，当NVL的两个参数不属于同类型时，看第二个参数是否可以向第一个参数进行隐式转换，如果可以则返回第一个参数类型。如果第二个参数不能向第一个参数进行隐式转换而第一个参数可以向第二个参数进行隐式转换，则返回第二个参数的类型。如果两个参数之间不存在隐式类型转换并且也不属于同一类型则报错。

sys_context('namespace' , 'parameter')

描述：获取并返回指定namespace下参数parameter的值。

返回值类型：text

示例：

```
SELECT sys_context('USERENV', 'CURRENT_SCHEMA');
 sys_context
-------------
 public
(1 row)
```

根据当前所在的实际schema而变化。

说明

目前仅支持SYS_CONTEXT('USERENV', 'CURRENT_SCHEMA') 和SYS_CONTEXT('USERENV', 'CURRENT_USER')两种格式。

greatest(expr1 [, ...])

描述：获取并返回参数列表中值最大的表达式的值。

 版权所有 © 华为云计算技术有限公司

- ORA和TD兼容模式下，返回结果为所有非null参数的最大值。
- MySQL兼容模式下，入参中存在null时，返回结果为null。

示例：

```
SELECT greatest(1*2,2-3,4-1);
 greatest
----------
        3
(1 row)
SELECT greatest('ABC', 'BCD', 'CDE');
 greatest
----------
 CDE
(1 row)
```

least(expr1 [, ...])

描述：获取并返回参数列表中值最小的表达式的值。

- ORA和TD兼容模式下，返回结果为所有非null参数的最小值。
- MySQL兼容模式下，入参中存在null时，返回结果为null。

示例：

```
SELECT least(1*2,2-3,4-1);
 least
-------
    -1
(1 row)
SELECT least('ABC','BCD','CDE');
 least
--------
 ABC
(1 row)
```

EMPTY_BLOB()

描述：使用EMPTY_BLOB在INSERT或UPDATE语句中初始化一个BLOB变量，取值为NULL。

返回值类型：BLOB

示例：

```
--新建表
CREATE TABLE blob_tb(b blob,id int) DISTRIBUTE BY REPLICATION;
--插入数据
INSERT INTO blob_tb VALUES (empty_blob(),1);
--删除表
 DROP TABLE blob_tb;
```

说明

使用DBMS.GETLENGTH求得的长度为0。

7.17 范围函数和操作符

 版权所有 © 华为云计算技术有限公司

7.17.1 范围操作符

范围操作符用于处理范围类型数据的运算（如 @>、&& ）。范围操作符支持直接应用于查询表达式，快速筛选特定区间内的数据，常用于范围比较、查询限制及区间界定。

说明

- 比较操作符<、>、<=和>=先比较下界，只有下界相等时才比较上界。
- <<、>>和-|-操作符当包含空范围时也会返回false，即不认为空范围在其他范围之前或之后。
- 并集和差集操作符的执行结果无法包含两个不相交的子范围。

操作符	描述	用法示例
=	等于	比较两个范围是否相等，返回t，表示相等。 SELECT int4range(1,5) = '[1,4]'::int4range AS RESULT; result -------- t (1 row)
<>	不等于	比较两个范围是否不相等，返回t，表示不相等。 SELECT numrange(1.1,2.2) <> numrange(1.1,2.3) AS RESULT; result -------- t (1 row)
<	小于	比较(1,10)是否小于(2,3)，先比较下界，且只有在下界相等时才比较上界。 SELECT int4range(1,10) < int4range(2,3) AS RESULT; result -------- t (1 row)
>	大于	比较(1,10)是否大于(1,5)，先比较下界，且只有在下界相等时才比较上界。 SELECT int4range(1,10) > int4range(1,5) AS RESULT; result -------- t (1 row)
<=	小于或等于	比较(1.1,2.2) 是否小于或等于(1.1,2.2)。 SELECT numrange(1.1,2.2) <= numrange(1.1,2.2) AS RESULT; result -------- t (1 row)
>=	大于或等于	比较(1.1,2.2) 是否大于等于(1.1,2.0)，先比较下界，且只有在下界相等时才比较上界。 SELECT numrange(1.1,2.2) >= numrange(1.1,2.0) AS RESULT; result -------- t (1 row)

版权所有 © 华为云计算技术有限公司

操作符	描述	用法示例
@>	包含范围、包含元素	(2,4)是否包含(2,3)，返回t，表示包含。 SELECT int4range(2,4) @> int4range(2,3) AS RESULT; result -------- t (1 row) [2011-01-01,2011-03-01)是否包含2011-01-10，返回t，表示包含。 SELECT '[2011-01-01,2011-03-01)'::tsrange @> '2011-01-10'::timestamp AS RESULT; result -------- t (1 row)
<@	范围包含于、元素包含于	(2,4)是否包含于(1,7)，返回t，表示包含。 SELECT int4range(2,4) <@ int4range(1,7) AS RESULT; result -------- t (1 row) 元素42是否包含于(1,7)，返回f，表示不包含。 SELECT 42 <@ int4range(1,7) AS RESULT; result -------- f (1 row)
&&	重叠（有共同点）	判断(3,7)和(4,12)的范围是否重叠，返回t，表示存在重叠部分。 SELECT int8range(3,7) && int8range(4,12) AS RESULT; result -------- t (1 row)
<<	范围值是否比另一个范围值的最小值还小（没有交集）	判断左边的范围值是否比右边范围值的最小值还小，且两者没有交集。 SELECT int8range(1,10) << int8range(100,110) AS RESULT; result -------- t (1 row)
>>	范围值是否比另一个范围值的最大值还大（没有交集）	判断左边的范围值是否比右边范围值的最大值还大，且两者没有交集。 SELECT int8range(50,60) >> int8range(20,30) AS RESULT; result -------- t (1 row)
&<	不超过右边范围上界。	判断左边的范围值是否不超过右边范围值的上界。 SELECT int8range(1,20) &< int8range(18,20) AS RESULT; result -------- t (1 row)

版权所有 © 华为云计算技术有限公司

操作符	描述	用法示例
&>	不超过左边范围上界。	判断右边的范围值是否不超过左边范围值的上界。 SELECT int8range(7,20) &> int8range(5,10) AS RESULT; result -------- t (1 row)
-|-	相邻	判断范围是否相邻，返回t，表示相邻。 SELECT numrange(1.1,2.2) -|- numrange(2.2,3.3) AS RESULT; result -------- t (1 row)
+	并集	计算两个区间的并集，结果返回新的区间[5,20)。 SELECT numrange(5,15) + numrange(10,20) AS RESULT; result -------- [5,20) (1 row)
*	交集	计算两个区间的交集，结果返回新的区间[10,15)。 SELECT int8range(5,15) * int8range(10,20) AS RESULT; result --------- [10,15) (1 row)
-	差集	计算两个区间的差集，结果返回新的区间 [5,10)。 SELECT int8range(5,15) - int8range(10,20) AS RESULT; result -------- [5,10) (1 row)

7.17.2 范围函数

范围函数用于定义范围的上下限以及判断某个值是否在范围限制内。

lower(anyrange)

描述：获取范围的下界。如果范围是空或者下界为无限，则返回null。

返回类型：范围元素类型

示例：

```
SELECT lower(numrange(1.1,2.2)) AS RESULT;
 result
--------
    1.1
(1 row)
```

upper(anyrange)

描述：获取范围的上界。如果范围是空或者上界为无限，则返回null。

返回类型：范围元素类型

 版权所有 © 华为云计算技术有限公司

示例：

```
SELECT upper(numrange(1.1,2.2)) AS RESULT;
 result
--------
    2.2
(1 row)
```

isempty(anyrange)

描述：判断范围是否为空。

返回类型：boolean

示例：

```
SELECT isempty(numrange(1.1,2.2)) AS RESULT;
 result
--------
 f
(1 row)
```

lower_inc(anyrange)

描述：判断范围的下界是否被包含在内。

返回类型：boolean

示例：

```
SELECT lower_inc(numrange(1.1,2.2)) AS RESULT;
 result
--------
 t
(1 row)
```

upper_inc(anyrange)

描述：判断范围的上界是否被包含在内。

返回类型：boolean

示例：

numrange(1.1, 2.2)默认使用左闭右开[1.1, 2.2) 的形式。上界(2.2)由于默认是不包含的，所以返回false。

```
SELECT upper_inc(numrange(1.1,2.2)) AS RESULT;
 result
--------
 f
(1 row)
```

lower_inf(anyrange)

描述：判断范围的下界是否为无限。

返回类型：boolean

示例：

```
SELECT lower_inf('(,)'::daterange) AS RESULT;
 result
--------
```

 版权所有 © 华为云计算技术有限公司

```
 t
(1 row)
```

upper_inf(anyrange)

描述：判断范围的上界是否为无限。

返回类型：boolean

示例：

```
SELECT upper_inf('(,)'::daterange) AS RESULT;
 result
--------
 t
(1 row)
```

说明

lower_inc、upper_inc、lower_inf和upper_inf函数均对空范围返回false。

7.18 数据脱敏函数

数据脱敏函数提供一系列不同脱敏形式的函数接口，可以覆盖常见的脱敏场景。通常结合数据脱敏语法，与脱敏列绑定使用，不推荐直接作用在查询语句上。

mask_none(column_name)

描述：不作任何脱敏处理，仅内部测试用。

返回值类型：与入参column_name数据类型相同

mask_full(column_name)

描述：全脱敏成固定值。脱敏列的数据类型不同，脱敏的固定值不同。

返回值类型：与入参column_name数据类型相同。

mask_partial(column_name, mask_digital, mask_from[, mask_to])

描述：针对数值类型数据，将第mask_from到mask_to位的数字部分脱敏成mask_digital对应的数字。其中，参数mask_to允许缺省，缺省时即脱敏到数据结束位置。参数mask_digital只能取[0,9]区间内的数字。

返回值类型：与入参column_name数据类型相同。

mask_partial(column_name [, input_format, output_format], mask_char, mask_from[, mask_to])

描述：针对字符类型数据，对照指定的输入输出格式，将第mask_from到mask_to位的数字部分脱敏成mask_char指定的字符。

参数说明：

- input_format

 版权所有 © 华为云计算技术有限公司

输入格式是由V和F组成的字符序列，与脱敏列数据长度相同。V对应位置的字符可能会被脱敏，F对象位置的字符会被忽略跳过，V字符序列标识脱敏范围。输入输出格式参数适用于定长数据，比如，银行卡号、身份证号、手机号等。

- output_format

  输出格式是由V和其他任意字符组成的字符序列，与脱敏列数据长度相同。V字符位置与input_format的V位置对应，其他字符位置与input_format的F位置对应，且不会脱敏，通常为数据分隔符。

  input_format和output_format可以缺省或指定为空串""，此时，无输入输出格式要求，原始字符序列范围即为脱敏范围。

- mask_char

  脱敏字符，仅允许长度为1的任意字符。场景的脱敏字符包括"*"，"#"等。

- mask_from

  脱敏范围的起始位置，要求大于0。

- mask_to

  脱敏范围的结束位置，允许缺省。缺省时，即脱敏到原始数据结束位置。

返回值类型：与入参column_name数据类型相同。

mask_partial(column_name, mask_field1, mask_value1, mask_field2, mask_value2, mask_field3, mask_value3)

描述：按指定三个时间域做部分脱敏，仅适用于日期或时间类型数据。若mask_value取-1，即此mask_field不脱敏。其中，mask_field可以取"month"、"day"、"year"、"hour"、"minute"、"second"六个时间域之一。各域的取值范围需满足实际时间单位的取值范围。

返回值类型：与入参column_name数据类型相同。

说明

- 脱敏函数可以覆盖常见敏感信息的脱敏场景，推荐用户优先使用脱敏函数创建脱敏策略。
- 数据脱敏函数的使用方法，请参考**数据脱敏**。

自定义脱敏函数

支持用户使用PL/pgSQL语言自定义脱敏函数。

自定义脱敏函数需要严格遵循如下要求：

- 返回值与脱敏列类型一致。
- 函数必须可下推。
- 参数列表除脱敏格式外，只能包含一个脱敏列。
- 函数仅实现针对特定数据类型的格式化改写功能，不能涉及与其他表对象的复杂关联操作。

不满足前两项中任一项时，创建脱敏策略会报错。不满足后两项中任一项时，可成功创建脱敏策略，但查询执行结果可能会出现不可预知的问题。

如果绑定的自定义脱敏函数被删除，查询会因为脱敏策略失效而执行报错。

 版权所有 © 华为云计算技术有限公司

7.19 位图（Roaring Bitmap）函数和操作符

7.19.1 Roaring Bitmap 操作符

8.1.3集群版本开始，DWS支持高效的位图处理操作符，位图操作符在数据库中用于高效地进行集合运算，如交集、并集、差集和异或，通过合理使用位图操作符，可以显著提升数据库的处理效率，满足各种复杂的数据处理需求。适用于需要频繁进行集合运算的场景，如数据挖掘、机器学习中的特征工程，以及需要快速数据筛选和聚合的场景。

⚠ 注意

- 9.1.1.100集群版本开始支持64位整数的位图。
- 9.1.1.100集群版本中的|、-、@>、<@位图操作符，使用bigint作为输入时，不支持使用nvarchar2类型直接作为bigint输入，需要使用::bigint操作进行数据类型强转换。例如：

```
select '2022'::nvarchar2::bigint | '\x3a300000010000000000020010000000010002000300'::roaringbitmap;
                 ?column?
----------------------------------------------------
 \x3a300000010000000000030010000000010002000300e607
```

- 9.1.1.200及以上集群版本中支持使用nvarchar2类型直接作为bigint输入，无需再进行强转换。

表 7-27 Roaring Bitmap 操作符列表

操作符	描述	返回值类型	示例

=	比较两个roaringbitmap是否相等。	bool	两个位图完全相同，返回t，表示它们相等。
			`SELECT rb_build('{1,2,3}') = rb_build('{1,2,3}');`
			`?column?`
			`----------`
			`t`
			`(1 row)`
			两个位图包含的元素不同，返回f，表示它们不相等。
			`SELECT rb_build('{2,3}') = rb_build('{1,2,3}');`
			`?column?`
			`----------`
			`f`
			`(1 row)`

版权所有 © 华为云计算技术有限公司

操作符	描述	返回值类型	示例
<>	比较两个roaringbitmap是否不相等。	bool	两个位图完全相同，返回f，表示它们相等。 SELECT rb_build('**{1,2,3}**') <> rb_build('**{1,2,3}**'); ?column? ---------- f (1 row) 两个位图包含的元素不同，返回t，表示它们不相等 SELECT rb_build('**{2,3}**') <> rb_build('**{1,2,3}**'); ?column? ---------- t (1 row)
&	计算两个roaringbitmap交集以后的结果。	roaringbitmap	将两个位图交集后的结果，返回共同元素{2,3}。 SELECT rb_to_array(rb_build('**{2,3}**') & rb_build('**{1,2,3}**')); rb_to_array ---------- {2,3} (1 row)
|	计算两个roaringbitmap并集后的结果。	roaringbitmap	计算两个位图的并集结果，返回所有不重复的元素{1,2,3}。 SELECT rb_to_array(rb_build('**{2,3}**') | rb_build('**{1,2,3}**')); rb_to_array ---------- {1,2,3} (1 row)
|	计算一个roaringbitmap增加一个元素后的结果。	roaringbitmap	计算包含元素{2,3}的位图添加元素4后的结果，返回{2,3,4}。 SELECT rb_to_array(rb_build('**{2,3}**') | **4**); rb_to_array ---------- {2,3,4} (1 row)
#	计算两个roaringbitmap进行异或后的结果。	roaringbitmap	计算两个位图的异或，返回两个集合中不相同的元素结果{1}。 SELECT rb_to_array(rb_build('**{2,3}**') # rb_build('**{1,2,3}**')); rb_to_array ---------- {1} (1 row)
-	计算属于第一个roaringbitmap，但是不属于第二个roaringbitmap的结果集。	roaringbitmap	计算属于第一个位图{2,3,4}的元素，但这些元素不在第二个位图{1,2,3}中，返回结果 {4} ，表示元素4只在第一个位图。 SELECT rb_to_array(rb_build('**{2,3,4}**') - rb_build('**{1,2,3}**')); rb_to_array ---------- {4} (1 row)

版权所有 © 华为云计算技术有限公司

操作符	描述	返回值类型	示例
-	从roaringbitmap中去掉一个指定ID后的结果集。	roaringbitmap	从位图{2,3,4}中去掉一个指定元素3，返回结果集{2,4}。 SELECT rb_to_array(rb_build('**{2,3,4}**') - **3**); rb_to_array ------------- {2,4} (1 row)
@>	操作符前面的roaringbitmap是否包含后面的roaringbitmap。	bool	判断操作符@>前面的位图{2,3,4}是否包含后面的位图{2,3}，返回t，表示包含。 SELECT rb_build('**{2,3,4}**') @> rb_build('**{2,3}**'); ?column? ---------- t (1 row) 判断操作符@>前面的位图{2,3,4}是否包含后面的位图5，返回f，表示不包含。 SELECT rb_build('**{2,3,4}**') @> **5**; ?column? ---------- f (1 row)
<@	操作符前面的roaringbitmap是否被后面的roaringbitmap包含。	bool	判断操作符<@前面的4是否被后面的位图{2,3,4}包含，返回t，表示包含。 SELECT **4** <@ rb_build('**{2,3,4}**'); ?column? ---------- t (1 row) 判断操作符<@前面的位图{2,3,4}是否被后面的位图{2,3}包含，返回f，表示不包含。 SELECT rb_build('**{2,3,4}**') <@ rb_build('**{2,3}**'); ?column? ---------- f (1 row)
&&	两个roaringbitmap如果有交集返回true，否则返回false。	bool	判断两个位图是否有交集，返回t，表示存在交集**{2,3}**。 SELECT rb_build('**{2,3,4}**') && rb_build('**{2,3}**'); ?column? ---------- t (1 row) 判断两个位图是否有交集，返回f，表示不存在任何交集。 SELECT rb_build('**{2,3,4}**') && rb_build('**{7,8,9}**'); ?column? ---------- f (1 row)

 版权所有 © 华为云计算技术有限公司

7.19.2 Roaring Bitmap 函数

8.1.3集群版本开始，DWS支持高效的位图处理函数，可用于用户画像，精准营销等场景，极大地提高了查询性能。

说明

9.1.1.100集群版本开始支持64位整数的位图。

rb_build(array)

描述：将int数组转成roaringbitmap类型。

返回值类型：roaringbitmap

示例：

```
SELECT rb_build('{1,2,3}');
rb_build
--------------------------------------------------
\x3a300000010000000000020010000000010002000300
(1 row)

select rb_build('{1, 5000000000, 5000000001}');
                                                rb_build
-------------------------------------------------------------------------------------------------------------------

\x060200000000000000000000003a300000010000000000000010000000010001000003a300000010000000
52a01001000000000f201f2
(1 row)
CREATE TABLE r_row (a int, b text, c roaringbitmap);
NOTICE:  The 'DISTRIBUTE BY' clause is not specified. Using round-robin as the distribution mode by default.
HINT:  Please use 'DISTRIBUTE BY' clause to specify suitable data distribution column.
CREATE TABLE

INSERT INTO r_row values (1, 'a', rb_build('{1,2,3}'));
INSERT 0 1

SELECT * FROM r_row;
 a | b |                         c
---+---+--------------------------------------------------
 1 | a | \x3a300000010000000000020010000000010002000300
(1 row)

INSERT INTO r_row values (2, 'b', rb_build('{}'));
INSERT 0 1

SELECT * FROM r_row;
 a | b |                         c
---+---+--------------------------------------------------
 2 | b | \x3a30000000000000
 1 | a | \x3a300000010000000000020010000000010002000300
(2 rows)
```

rb_iterate(roaringbitmap)

描述：把roaringbitmap数据转成int，按照多行输出。

返回值类型：record类型(多行int值)

示例：

```
SELECT rb_iterate(c) FROM r_row;
rb_iterate
------------
```

 版权所有 © 华为云计算技术有限公司

```
1
2
3
(3 rows)
```

rb_to_array(roaringbitmap)

描述：rb_build的逆向操作，把roaringBitmap转成int数组。

返回值类型: array

示例：

```
SELECT rb_to_array(c) FROM r_row;
rb_to_array
-------------
{1,2,3}
(1 row)

SELECT rb_to_array('\x3a300000010000000000020010000000010002000300');
rb_to_array
-------------
{1,2,3}
(1 row)
```

rb_to_bigint_array(roaringbitmap)

描述：rb_build的逆向操作，把roaringBitmap转成bigint数组。该函数仅9.1.1.100及以上集群版本支持。

返回值类型：array

示例：

```
SELECT
rb_to_bigint_array('\x06020000000000000000000003a30000001000000000000001000000001000100000003a30000001000000052a01001000000000f201f2');
rb_to_bigint_array
-------------
{1, 5000000000, 5000000001}
(1 row)
```

rb_and(roaringbitmap, roaringbitmap)

描述：计算两个roaringbitmap的交集。

返回值类型：roaringbitmap

示例：

```
SELECT rb_to_array(rb_and(rb_build('{1,2,3}'), rb_build('{2,3,4}')));
rb_to_array
-------------
{2,3}
(1 row)
```

rb_or(roaringbitmap, roaringbitmap)

描述：计算两个roaringbitmap的并集。

返回值类型：roaringbitmap

示例：

 版权所有 © 华为云计算技术有限公司

```
SELECT rb_to_array(rb_or(rb_build('{1,2,3}'), rb_build('{2,3,4}')));
rb_to_array
-------------
{1,2,3,4}
(1 row)
```

rb_xor(roaringbitmap, roaringbitmap)

描述：计算两个roaringbitmap的异或。

返回值类型：roaringbitmap

示例：

```
SELECT rb_to_array(rb_xor(rb_build('{1,2,3}'), rb_build('{2,3,4}')));
rb_to_array
-------------
{1,4}
(1 row)
```

rb_andnot(roaringbitmap, roaringbitmap)

描述：在第一个roaringbitmap集合中，但是不在第二个roaringbitmap中的集合。

返回值类型：roaringbitmap

示例：

```
SELECT rb_to_array(rb_andnot(rb_build('{1,2,3}'), rb_build('{2,3,4}')));
rb_to_array
-------------
{1}
(1 row)
```

rb_cardinality(roaringbitmap)

描述：计算一个roaringbitmap的基数。

返回值类型：bigint

示例：

```
SELECT rb_cardinality(rb_build('{1,2,3}'));
rb_cardinality
----------------
3
(1 row)
```

rb_and_cardinality(roaringbitmap, roaringbitmap)

描述：计算两个roaringbitmap的交集的基数。

返回值类型：bigint

示例：

```
SELECT rb_and_cardinality(rb_build('{1,2,3}'), rb_build('{2,3,4}'));
rb_and_cardinality
--------------------
2
(1 row)
```

 版权所有 © 华为云计算技术有限公司

rb_or_cardinality(roaringbitmap, roaringbitmap)

描述：计算两个roaringbitmap的并集的基数。

返回值类型：bigint

示例：

```
SELECT rb_or_cardinality(rb_build('{1,2,3}'), rb_build('{2,3,4}'));
rb_or_cardinality
-------------------
4
(1 row)
```

rb_xor_cardinality(roaringbitmap, roaringbitmap)

描述：计算两个roaringbitmap异或以后的基数。

返回值类型：bigint

示例：

```
SELECT rb_xor_cardinality(rb_build('{1,2,3}'), rb_build('{2,3,4}'));
rb_xor_cardinality
--------------------
2
(1 row)
```

rb_andnot_cardinality(roaringbitmap, roaringbitmap)

描述：计算两个roaringbitmap按照andnot计算结果以后的基数。

返回值类型：bigint

示例：

```
SELECT rb_andnot_cardinality(rb_build('{1,2,3}'), rb_build('{2,3,4}'));
rb_andnot_cardinality
-----------------------
1
(1 row)
```

rb_is_empty(roaringbitmap)

描述：判断一个roaringbitmap是否为空。

返回值类型：bool

示例：

```
SELECT rb_is_empty(rb_build('{1,2,3}'));
rb_is_empty
-------------
f
(1 row)
```

rb_equals(roaringbitmap, roaringbitmap)

描述：判断两个roaringbitmap是否相等。

返回值类型：bool

示例：

 版权所有 © 华为云计算技术有限公司

```
SELECT rb_equals(rb_build('{1,2,3}'), rb_build('{2,3,4}'));
rb_equals
-----------
f
(1 row)
```

rb_intersect(roaringbitmap, roaringbitmap)

描述：判断两个roaringbitmap是否相交。

返回值类型：bool

示例：

```
SELECT rb_intersect(rb_build('{1,2,3}'), rb_build('{2,3,4}'));
rb_intersect
--------------
t
(1 row)
```

rb_min(roaringbitmap)

描述：返回roaringbitmap中的最小值。

返回值类型：bigint

示例：

```
SELECT rb_min(rb_build('{1,2,3}'));
rb_min
--------
1
(1 row)
```

rb_max(roaringbitmap)

描述：返回roaringbitmap中的最大值。

返回值类型：bigint

示例：

```
SELECT rb_max(rb_build('{1,2,3}'));
rb_max
--------
3
(1 row)
```

rb_add(roaringbitmap, bigint)

描述：在roaringbitmap中增加一个元素。

返回值类型：roaringbitmap

示例：

```
SELECT rb_to_array(rb_add(rb_build('{1,3}'), 2));
rb_to_array
-------------
{1,2,3}
(1 row)
```

 版权所有 © 华为云计算技术有限公司

⚠ 注意

rb_add、rb_added、rb_contain、rb_containedby、 rb_remove函数，使用bigint作为输入时，不支持使用nvarchar2类型直接作为bigint输入，需要使用::bigint操作进行数据类型强转换。例如：

```
SELECT rb_to_array(rb_add(rb_build('{1,3}'), '2'::nvarchar2::bigint));
rb_to_array
-------------
 {1,2,3}
```

rb_added(bigint, roaringbitmap)

描述：在roaringbitmap中增加一个元素。

返回值类型：roaringbitmap

示例：

```
SELECT rb_to_array(rb_added(2, rb_build('{1,3}')));
rb_to_array
-------------
{1,2,3}
(1 row)
```

rb_contain(roaringbitmap,bigint)

描述：判断roaringbitmap是否包含指定的元素。

返回值类型：bool

示例：

```
SELECT rb_contain(rb_build('{1,3}'), 2);
rb_contain
------------
f
(1 row)
```

rb_containedby(bigint,roaringbitmap)

描述：判断给定的元素是否被给定的roaringbitmap包含。

示例：

```
SELECT rb_containedby(2,rb_build('{1,3}'));
rb_containedby
----------------
f
(1 row)
```

rb_contain_rb(roaringbitmap,roaringbitmap)

描述：判断第一个roaringbitmap是否包含第二个roaringbitmap。

返回值类型：bool

示例：

```
SELECT rb_contain_rb(rb_build('{1,3}'), rb_build('{2,3}'));
rb_contain_rb
---------------
```


版权所有 © 华为云计算技术有限公司

```
f
(1 row)
```

rb_containedby_rb(roaringbitmap,roaringbitmap)

描述：判断给定的第二个roaringbitmap是否包含第一个roaringbitmap。

返回值类型：bool

示例：

```
SELECT rb_containedby_rb(rb_build('{1,3}'), rb_build('{2,3}'));
rb_containedby_rb
---------------
f
(1 row)
```

rb_remove(roaringbitmap,bigint)

描述：从roaringbitmap中移除指定的元素。

返回值类型：roaringbitmap

示例：

```
SELECT rb_to_array(rb_remove(rb_build('{1,3}'),1));
rb_to_array
-------------
{3}
(1 row)
```

rb_clear(roaringbitmap,bigint,bigint)

描述：从roaringbitmap中清除指定范围内的元素。

返回值类型：roaringbitmap

示例：

```
SELECT
rb_to_array(rb_clear(rb_build('{1,2,3}'),1,2));

rb_to_array

-------------

{2,3}
                                                                (1 row)
```

rb_flip(roaringbitmap,bigint,bigint)

描述：反转指定范围的元素。

示例：

```
SELECT rb_to_array(rb_flip(rb_build('{1,2,3,7,9}'), 1,10));
rb_to_array
--------------
{4,5,6,8,10}
(1 row)
```

rb_rank(roaringbitmap,bigint)

描述：返回小于指定值的集合的基数。

 版权所有 © 华为云计算技术有限公司

返回值类型：bigint

示例：

```
SELECT rb_rank(rb_build('{1,10,100}'),99);
rb_rank
---------
2
(1 row)
```

7.19.3 Roaring Bitmap 聚合函数

8.1.3集群版本开始，DWS支持高效的位图聚合函数，这些函数在大规模数据处理和复杂集合运算的场景中能够显著提升数据处理的效率和灵活性。

rb_build_agg(bigint)

描述：将分组内的bigint值聚合成一个roaringbitmap值。也支持使用int作为入参。

返回值类型：roaringbitmap

示例：

```
CREATE TABLE t1 (a int ,b int);
INSERT INTO t1 SELECT generate_series(1,10),generate_series(1,20,2);
SELECT rb_iterate(rb_build_agg(b)) FROM t1;
rb_iterate
------------
1
3
5
7
9
11
13
15
17
19
(10 rows)
```

rb_and_agg(roaringbitmap)

描述：将分组内的roaringbitmap数据按照交的操作聚合成一个roaringbitmap集合。

示例：

```
CREATE TABLE r1(a int ,b roaringbitmap);
INSERT INTO r1 SELECT a, rb_build_agg(b) FROM t1 GROUP BY a;
INSERT INTO t1 SELECT generate_series(1,10),generate_series(1,20,4);
INSERT INTO r1 SELECT a, rb_build_agg(b) FROM t1 GROUP BY a;
SELECT a, rb_to_array(rb_and_agg(b)) FROM r1 GROUP BY a ORDER BY a;
 a  | rb_to_array
----+-------------
 1  | {1}
 2  | {3}
 3  | {5}
 4  | {7}
 5  | {9}
 6  | {11}
 7  | {13}
 8  | {15}
 9  | {17}
10  | {19}
(10 rows)
```

 版权所有 © 华为云计算技术有限公司

rb_or_agg(roaringbitmap)

描述：将分组内的roaringbitmap按照并的逻辑组合成一个roaringbitmap。

示例：

```
SELECT a, rb_to_array(rb_or_agg(b)) FROM r1 GROUP BY a ORDER BY a;
 a  | rb_to_array
----+-------------
 1  | {1}
 2  | {3,5}
 3  | {5,9}
 4  | {7,13}
 5  | {9,17}
 6  | {1,11}
 7  | {5,13}
 8  | {9,15}
 9  | {13,17}
 10 | {17,19}
(10 rows)
```

rb_xor_agg(roaringbitmap)

描述：将分组内的roaringbitmap按照异或的逻辑组合成一个roaringbitmap。

示例：

```
SELECT a, rb_to_array(rb_xor_agg(b)) FROM r1 GROUP BY a ORDER BY a;
a   | rb_to_array
----+-------------
 1  | {}
 2  | {5}
 3  | {9}
 4  | {13}
 5  | {17}
 6  | {1}
 7  | {5}
 8  | {9}
 9  | {13}
 10 | {17}
(10 rows)
```

rb_and_cardinality_agg(roaringbitmap)

描述：分组内的roaringbitmap按照交集计算后的基数。

返回值类型：bigint

示例：

```
SELECT a, rb_and_cardinality_agg(b) FROM r1 GROUP BY a ORDER BY 1;
 a  | rb_and_cardinality_agg
----+------------------------
 1  |                      1
 2  |                      1
 3  |                      1
 4  |                      1
 5  |                      1
 6  |                      1
 7  |                      1
 8  |                      1
 9  |                      1
 10 |                      1
(10 rows)
```

 版权所有 © 华为云计算技术有限公司

rb_or_cardinality_agg(roaringbitmap)

描述：将分组内的roaringbitmap按照并集计算后的基数。

返回值类型：bigint

示例：

```
SELECT a, rb_or_cardinality_agg(b) FROM r1 GROUP BY a ORDER BY 1;
 a  | rb_or_cardinality_agg
----+-----------------------
 1  |                     1
 2  |                     2
 3  |                     2
 4  |                     2
 5  |                     2
 6  |                     2
 7  |                     2
 8  |                     2
 9  |                     2
 10 |                     2
(10 rows)
```

rb_xor_cardinality_agg(roaringbitmap)

描述：将分组内的roaringbitmap按照异或的逻辑合并后的基数。

返回值类型：bigint

示例：

```
SELECT a, rb_xor_cardinality_agg(b) FROM r1 GROUP BY a ORDER BY 1;
 a  | rb_xor_cardinality_agg
----+------------------------
 1  |                      0
 2  |                      1
 3  |                      1
 4  |                      1
 5  |                      1
 6  |                      1
 7  |                      1
 8  |                      1
 9  |                      1
 10 |                      1
(10 rows)
```

7.19.4 使用场景

背景

目前在互联网、教育、游戏等行业都有实时精准营销的需求。通过系统生成用户画像，在营销时通过条件组合筛选用户，快速提取目标群体。例如：

- 在电商行业中，商家在进行营销活动前，需要根据活动的目的，圈选一批满足特定特征的目标用户群体进行广告推送。
- 在教育行业中，需要根据学生不同的特征，推送有针对性的练习题目，帮助学生查漏补缺。
- 在搜索、视频、门户网站中，根据用户关注的热点，推送不同的内容。

这些业务场景都有一些共同的特点：

- 数据量庞大，运算量极大。

 版权所有 © 华为云计算技术有限公司

- 用户规模庞大，标签多，字段多，占用存储空间也多。
- 圈选的特征条件多样化，很难找到固定索引，如果每个字段一个索引，存储空间又会暴增。
- 性能要求高，因为实时营销要求秒级响应。
- 数据更新时效要求高，用户画像几乎要求实时更新。

针对上述业务场景特点，DWS的roaringbitmap可以高效生成、压缩、解析位图数据，支持最常见的位图聚合操作（与、或、非、异或），满足用户在亿级以上、千万级标签的大数据量下实时精准营销、快速圈选用户的需求。

roaringbitmap 使用示例

假设有一张用户浏览网页的流水信息表userinfo，表中的字段如下：

```
CREATE TABLE userinfo
(userid int,
age int,
gender text,
salary int,
hobby  text
)WITH (orientation=column);
```

userinfo表中的数据会随着用户信息的变化不断增长，比如用户有多个hobby属性，那么就有多条记录。

如果用户需要筛选出所有“收入大于10000 USD的男性，年龄大于30岁，爱好钓鱼”的群体，向这些目标群体推送特定的消息。

传统的方法是直接在原表上执行查询，语句如下：

```
SELECT distinct userid FROM userinfo WHERE salary > 10000 AND age > 30 AND gender ='m' AND hobby ='fishing';
```

当userinfo表的数据量不大的时候，可以通过在salary，age，gender，hobby列上建立索引来满足需求。但是如果userinfo表的数据量非常大，同时一张表的标签数非常多的时候，上述语句就不能满足诉求，因为如下原因：

- 需要创建的索引会非常多。
- count (distinct)的性能比较差。

这种场景下使用roaringbitmap就会有比较好的效果。

1. 新建一张Roaringbitmap表：
   ```
   CREATE TABLE userinfoset
   ( age int,
   gender text,
   salary int,
   hobby  text,
   userset roaringbitmap,
   PRIMARY KEY(age,gender,salary,hobby)
   )WITH (orientation=column);
   ```
2. 所有userinfo表中的数据要通过标签列聚合到userinfoset表中。可以采用对全量数据进行聚合的方法（如下命令所示）。也可以采用只对增量数据进行聚合的方法。只对增量数据进行聚合即对含有相同的标签的用户集合放到表的一条记录中，通常可以通过upsert来实现。考虑到其中频繁的update操作可能产生大量的脏数据，因此对增量数据进行聚合的方法，建议将userinfoset表创建为行存表。
   ```
   INSERT INTO userinfoset
   SELECT age, gender, salary, hobby, rb_build_agg(userid)
   FROM
   ```

 版权所有 © 华为云计算技术有限公司

```
userinfo
GROUP BY age, gender, salary, hobby;
```

3. 直接查询userinfoset表获得用户筛选信息。

```
SELECT rb_iterate(rb_or_agg(userset)) FROM userinfoset WHERE salary > 10000 AND age > 30 AND
gender ='m' AND hobby ='fishing';
```

数据进行聚合后的userinfoset的数据量相比源表小了很多，基表scan的性能会快很多，同时基于Roaringbitmap的优势，计算rb_or_agg和rb_iterate的性能也很好，相比传统的方法，性能明显提升。

7.20 UUID 函数

UUID函数表示可以用于生成UUID类型（请参考**UUID类型**）数据的函数。

uuid_generate_v1()

描述：生成一个UUID类型的序列号。

返回类型：UUID

示例：

```
SELECT uuid_generate_v1();
          uuid_generate_v1
---------------------------------------
 c71ceaca-a175-11e9-a920-797ff7000001
(1 row)
```

说明

uuid_generate_v1函数根据时间信息、集群节点编号和生成该序列的线程号生成UUID，该UUID在单个集群内是全局唯一的，但在多个集群间的时间信息、集群节点编号、线程号和时钟序列仍然存在同时相等的可能性，因此多个集群间生成的UUID仍然存在极低概率的重复风险。

uuid()

描述：生成一个UUID类型的序列号。此函数为MySQL兼容性函数，仅8.2.0及以上集群版本支持。

返回类型：UUID

示例：

```
SELECT uuid();
             uuid
-----------------------------------
 6327dc96-f0e7-0100-f2f2-6c9ff700fffe
(1 row)
```

说明

uuid函数内部生成原理同**uuid_generate_v1()**函数，即根据时间信息、集群节点编号和生成该序列的线程号生成UUID，该UUID在单个集群内是全局唯一的，但在多个集群间的时间信息、集群节点编号、线程号和时钟序列仍然存在同时相等的可能性，因此多个集群间生成的UUID仍然存在极低概率的重复风险。

sys_guid()

描述：生成Oracle的GUID序列号，类似UUID。此函数为Oracle兼容性函数。

返回类型：text

 版权所有 © 华为云计算技术有限公司

示例：

```
SELECT sys_guid();
             sys_guid
----------------------------------
 4EBD3C74A17A11E9A1BF797FF7000001
(1 row)
```

说明

sys_guid函数内部生成原理同uuid_generate_v1函数。

UUID 函数应用示例

UUID全局唯一的特点，可以作为数据表生成主键，也可以作为数据表的分布列，uuid_generate_v1()作为数据表分布列的默认值时，通过Hash分布可以将数据均匀分布到各个DN上，防止数据倾斜。

说明

UUID的显著优点就是全局唯一，不需要中心节点，单个节点独立生成。但是也存在缺点，UUID较INT占用更多的存储空间，索引效率低，生成的ID随机，没有递增的特性，所以辨识困难。因此，在应用中，要根据实际情况选择UUID还是Sequence作为数据表主键。

示例如下：

1. INT类型作为分布列。

   创建示例哈希表mytable01，int类型作为分布列，插入数据后，查询数据存在数据倾斜。

   ```
   CREATE TABLE mytable01(a INT, b INT) DISTRIBUTE BY hash(a);
   CREATE TABLE
   INSERT INTO mytable01 VALUES(1, 10);
   INSERT 0 1
   INSERT INTO mytable01 VALUES(1, 10);
   INSERT 0 1
   INSERT INTO mytable01 VALUES(1, 10);
   INSERT 0 1
   INSERT INTO mytable01 VALUES(1, 10);
   INSERT 0 1
   INSERT INTO mytable01 VALUES(1, 10);
   INSERT 0 1
   SELECT * FROM mytable01;
    a | b
   ---+----
    1 | 10
    1 | 10
    1 | 10
    1 | 10
    1 | 10
   (5 rows)

   SELECT table_skewness('mytable01');
             table_skewness
   ---------------------------------------
    ("dn_6003_6004         ",5,100.000%)
    ("dn_6001_6002         ",0,0.000%)
    ("dn_6005_6006         ",0,0.000%)
   (3 rows)
   ```

2. UUID类型作为分布列。

   创建示例哈希表mytable02，UUID类型作为分布列，插入数据后，查询数据分布正常。

   ```
   CREATE TABLE mytable02 (id UUID default uuid_generate_v1(), a INT, b INT) DISTRIBUTE BY hash(id);
   CREATE TABLE
   ```


版权所有 © 华为云计算技术有限公司

```
INSERT INTO mytable02(a, b) VALUES(1, 10);
INSERT 0 1
INSERT INTO mytable02(a, b) VALUES(1, 10);
INSERT 0 1
INSERT INTO mytable02(a, b) VALUES(1, 10);
INSERT 0 1
INSERT INTO mytable02(a, b) VALUES(1, 10);
INSERT 0 1
INSERT INTO mytable02(a, b) VALUES(1, 10);
INSERT 0 1

SELECT * FROM mytable02;
                  id                  | a | b
--------------------------------------+---+----
 63e45c14-cc74-0e00-e9aa-0a2c3fa0fffe | 1 | 10
 63e45c1f-4d18-0700-e9ab-0a2c3fa0fffe | 1 | 10
 63e45c26-f859-0b00-e9ad-0a2c3fa0fffe | 1 | 10
 63e45c23-9e5d-0300-e9ac-0a2c3fa0fffe | 1 | 10
 63e45c2a-5825-0600-e9ae-0a2c3fa0fffe | 1 | 10
(5 rows)

SELECT table_skewness('mytable02');
           table_skewness
-------------------------------------
 ("dn_6001_6002         ",3,60.000%)
 ("dn_6003_6004         ",2,40.000%)
 ("dn_6005_6006         ",0,0.000%)
(3 rows)
```

7.21 文本检索函数和操作符

7.21.1 文本检索操作符

@@

描述：tsvector类型的词汇与tsquery类型的词汇是否匹配

示例：

```
SELECT to_tsvector('fat cats ate rats') @@ to_tsquery('cat & rat') AS RESULT;
 result
--------
 t
(1 row)
```

@@@

描述：@@的同义词

示例：

```
SELECT to_tsvector('fat cats ate rats') @@@ to_tsquery('cat & rat') AS RESULT;
 result
--------
 t
(1 row)
```

&&

描述：将两个tsquery类型的词汇进行“与”操作

 版权所有 © 华为云计算技术有限公司

示例：

```
SELECT 'fat | rat'::tsquery && 'cat'::tsquery AS RESULT;
          result
---------------------------
 ( 'fat' | 'rat' ) & 'cat'
(1 row)
```

||

描述：将两个tsquery类型的词汇进行“或”操作

示例：

```
SELECT 'fat | rat'::tsquery || 'cat'::tsquery AS RESULT;
          result
----------------------------
 ( 'fat' | 'rat' ) | 'cat'
(1 row)
SELECT 'a:1 b:2'::tsvector || 'c:1 d:2 b:3'::tsvector AS RESULT;
          result
----------------------------
 'a':1 'b':2,5 'c':3 'd':4
(1 row)
```

!!

描述：tsquery类型词汇的非关系

示例：

```
SELECT !! 'cat'::tsquery AS RESULT;
 result
--------
 !'cat'
(1 row)
```

@>

描述：一个tsquery类型的词汇是否包含另一个tsquery类型的词汇

示例：

```
SELECT 'cat'::tsquery @> 'cat & rat'::tsquery AS RESULT;
 result
--------
 f
(1 row)
```

<@

描述：一个tsquery类型的词汇是否被包含另一个tsquery类型的词汇

示例：

```
SELECT 'cat'::tsquery <@ 'cat & rat'::tsquery AS RESULT;
 result
--------
 t
(1 row)
```

除了上述的操作符，还为tsvector类型和tsquery类型的数据定义了普通的B-tree比较操作符（=，<等）。

 版权所有 © 华为云计算技术有限公司

7.21.2 文本检索函数

get_current_ts_config()

描述：获取文本检索的默认配置。

返回类型：regconfig

示例：

```
SELECT get_current_ts_config();
 get_current_ts_config
-----------------------
 english
(1 row)
```

length(tsvector)

描述：tsvector类型词汇的单词数。

返回类型：integer

示例：

```
SELECT length('fat:2,4 cat:3 rat:5A'::tsvector);
 length
--------
      3
(1 row)
```

numnode(tsquery)

描述：tsquery类型的单词加上操作符的数量。

返回类型：integer

示例：

```
SELECT numnode('(fat & rat) | cat'::tsquery);
 numnode
---------
       5
(1 row)
```

plainto_tsquery([config regconfig ,] query text)

描述：产生tsquery类型的词汇，并忽略标点。

返回类型：tsquery

示例：

```
SELECT plainto_tsquery('english', 'The Fat Rats');
 plainto_tsquery
-----------------
 'fat' & 'rat'
(1 row)
```

querytree(query tsquery)

描述：获取tsquery类型的词汇可加索引的部分。

 版权所有 © 华为云计算技术有限公司

返回类型：text

示例：

```
SELECT querytree('foo & ! bar'::tsquery);
 querytree
-----------
 'foo'
(1 row)
```

setweight(tsvector, "char")

描述：给tsvector类型的每个元素分配权值。

返回类型：tsvector

示例：

```
SELECT setweight('fat:2,4 cat:3 rat:5B'::tsvector, 'A');
           setweight
--------------------------------
 'cat':3A 'fat':2A,4A 'rat':5A
(1 row)
```

strip(tsvector)

描述：删除tsvector类型单词中的position和权值。

返回类型：tsvector

示例：

```
SELECT strip('fat:2,4 cat:3 rat:5A'::tsvector);
       strip
--------------------
 'cat' 'fat' 'rat'
(1 row)
```

to_tsquery([config regconfig ,] query text)

描述：标准化单词，并转换为tsquery类型。

返回类型：tsquery

示例：

```
SELECT to_tsquery('english', 'The & Fat & Rats');
  to_tsquery
---------------
 'fat' & 'rat'
(1 row)
```

to_tsvector([config regconfig ,] document text)

描述：去除文件信息，并转换为tsvector类型。

返回类型：tsvector

示例：

```
SELECT to_tsvector('english', 'The Fat Rats');
   to_tsvector
-----------------
```

 版权所有 © 华为云计算技术有限公司

```
 'fat':2 'rat':3
(1 row)
```

ts_headline([config regconfig,] document text, query tsquery [, options text])

描述：高亮显示查询的匹配项。

返回类型：text

示例：

```
SELECT ts_headline('x y z', 'z'::tsquery);
 ts_headline
--------------
 x y <b>z</b>
(1 row)
```

ts_rank([weights float4[],] vector tsvector, query tsquery [, normalization integer])

描述：文档查询排名。

返回类型：float4

示例：

```
SELECT ts_rank('hello world'::tsvector, 'world'::tsquery);
 ts_rank
----------
 .0607927
(1 row)
```

ts_rank_cd([weights float4[],] vector tsvector, query tsquery [, normalization integer])

描述：排序文件查询使用覆盖密度。

返回类型：float4

示例：

```
SELECT ts_rank_cd('hello world'::tsvector, 'world'::tsquery);
 ts_rank_cd
------------
          0
(1 row)
```

ts_rewrite(query tsquery, target tsquery, substitute tsquery)

描述：替换目标tsquery类型的单词。

返回类型：tsquery

示例：

```
SELECT ts_rewrite('a & b'::tsquery, 'a'::tsquery, 'foo|bar'::tsquery);
       ts_rewrite
--------------------------
 'b' & ( 'foo' | 'bar' )
(1 row)
```

 版权所有 © 华为云计算技术有限公司

ts_rewrite(query tsquery, select text)

描述：使用SELECT命令的结果替代目标中tsquery类型的单词。

返回类型：tsquery

示例：

```
SELECT ts_rewrite('world'::tsquery, 'select ''world''::tsquery, ''hello''::tsquery');
 ts_rewrite
------------
 'hello'
(1 row)
```

7.21.3 文本检索调试函数

ts_debug([config regconfig,] document text, OUT alias text, OUT description text, OUT token text, OUT dictionaries regdictionary[], OUT dictionary regdictionary, OUT lexemes text[])

描述：测试一个配置。

返回类型：setof record

示例：

```
SELECT ts_debug('english', 'The Brightest supernovae');
                                ts_debug
-------------------------------------------------------------------------------
 (asciiword,"Word, all ASCII",The,{english_stem},english_stem,{})
 (blank,"Space symbols"," ",{},,)
 (asciiword,"Word, all ASCII",Brightest,{english_stem},english_stem,{brightest})
 (blank,"Space symbols"," ",{},,)
 (asciiword,"Word, all ASCII",supernovae,{english_stem},english_stem,{supernova})
(5 rows)
```

ts_lexize(dict regdictionary, token text)

描述：测试一个数据字典。

返回类型：text[]

示例：

```
SELECT ts_lexize('english_stem', 'stars');
 ts_lexize
-----------
 {star}
(1 row)
```

ts_parse(parser_name text, document text, OUT tokid integer, OUT token text)

描述：测试一个解析。

返回类型：setof record

示例：

```
SELECT ts_parse('default', 'foo - bar');
 ts_parse
-----------
```

 版权所有 © 华为云计算技术有限公司

```
 (1,foo)
 (12," ")
 (12,"- ")
 (1,bar)
(4 rows)
```

ts_parse(parser_oid oid, document text, OUT tokid integer, OUT token text)

描述：测试一个解析。

返回类型：setof record

示例：

```
SELECT ts_parse(3722, 'foo - bar');
 ts_parse
-----------
 (1,foo)
 (12," ")
 (12,"- ")
 (1,bar)
(4 rows)
```

ts_token_type(parser_name text, OUT tokid integer, OUT alias text, OUT description text)

描述：获取分析器定义的记号类型。

返回类型：setof record

示例：

```
SELECT ts_token_type('default');
                      ts_token_type
---------------------------------------------------------------
 (1,asciiword,"Word, all ASCII")
 (2,word,"Word, all letters")
 (3,numword,"Word, letters and digits")
 (4,email,"Email address")
 (5,url,URL)
 (6,host,Host)
 (7,sfloat,"Scientific notation")
 (8,version,"Version number")
 (9,hword_numpart,"Hyphenated word part, letters and digits")
 (10,hword_part,"Hyphenated word part, all letters")
 (11,hword_asciipart,"Hyphenated word part, all ASCII")
 (12,blank,"Space symbols")
 (13,tag,"XML tag")
 (14,protocol,"Protocol head")
 (15,numhword,"Hyphenated word, letters and digits")
 (16,asciihword,"Hyphenated word, all ASCII")
 (17,hword,"Hyphenated word, all letters")
 (18,url_path,"URL path")
 (19,file,"File or path name")
 (20,float,"Decimal notation")
 (21,int,"Signed integer")
 (22,uint,"Unsigned integer")
 (23,entity,"XML entity")
(23 rows)
```

ts_token_type(parser_oid oid, OUT tokid integer, OUT alias text, OUT description text)

描述：获取分析器定义的记号类型。

 版权所有 © 华为云计算技术有限公司

返回类型：setof record

示例：

```
SELECT ts_token_type(3722);
                      ts_token_type
---------------------------------------------------------------
 (1,asciiword,"Word, all ASCII")
 (2,word,"Word, all letters")
 (3,numword,"Word, letters and digits")
 (4,email,"Email address")
 (5,url,URL)
 (6,host,Host)
 (7,sfloat,"Scientific notation")
 (8,version,"Version number")
 (9,hword_numpart,"Hyphenated word part, letters and digits")
 (10,hword_part,"Hyphenated word part, all letters")
 (11,hword_asciipart,"Hyphenated word part, all ASCII")
 (12,blank,"Space symbols")
 (13,tag,"XML tag")
 (14,protocol,"Protocol head")
 (15,numhword,"Hyphenated word, letters and digits")
 (16,asciihword,"Hyphenated word, all ASCII")
 (17,hword,"Hyphenated word, all letters")
 (18,url_path,"URL path")
 (19,file,"File or path name")
 (20,float,"Decimal notation")
 (21,int,"Signed integer")
 (22,uint,"Unsigned integer")
 (23,entity,"XML entity")
(23 rows)
```

ts_stat(sqlquery text, [weights text,] OUT word text, OUT ndoc integer, OUT nentry integer)

描述：获取tsvector列的统计数据。

返回类型：setof record

示例：

```
SELECT ts_stat('select ''hello world''::tsvector');
   ts_stat
-------------
 (world,1,1)
 (hello,1,1)
(2 rows)
```

7.22 HLL 函数和操作符

7.22.1 HLL 操作符

HLL类型支持如下操作符：

 版权所有 © 华为云计算技术有限公司

表 7-28 HLL 操作符

HLL操作符	描述	返回值类型	示例
=	比较hll或hll_hashval的值是否相等。	bool	• hll SELECT (hll_empty() || hll_hash_integer(1)) = (hll_empty() || hll_hash_integer(1)); ?column? ---------- t (1 row) • hll_hashval SELECT hll_hash_integer(1) = hll_hash_integer(1); ?column? ---------- t (1 row)
<> or !=	比较hll或hll_hashval是否不相等。	bool	• hll SELECT (hll_empty() || hll_hash_integer(1)) <> (hll_empty() || hll_hash_integer(2)); ?column? ---------- t (1 row) • hll_hashval SELECT hll_hash_integer(1) <> hll_hash_integer(2); ?column? ---------- t (1 row)
||	可代表hll_add，hll_union，hll_add_rev三个函数的功能。	hll	• hll_add SELECT hll_empty() || hll_hash_integer(1); ?column? -------------------------- \x128b7f8895a3f5af28cafe (1 row) • hll_add_rev SELECT hll_hash_integer(1) || hll_empty(); ?column? -------------------------- \x128b7f8895a3f5af28cafe (1 row) • hll_union SELECT (hll_empty() || hll_hash_integer(1)) || (hll_empty() || hll_hash_integer(2)); ?column? --- \x128b7f8895a3f5af28cafeda0ce907e4355b60 (1 row)
#	计算出hll的distinct值，同hll_cardinality函数。	integer	SELECT #(hll_empty() || hll_hash_integer(1)); ?column? ---------- 1 (1 row)

 版权所有 © 华为云计算技术有限公司

7.22.2 哈希函数

hll_hash_boolean(bool)

描述：对bool类型数据计算哈希值。

返回值类型：hll_hashval

示例：

```
SELECT hll_hash_boolean(FALSE);
  hll_hash_boolean
---------------------
 5048724184180415669
(1 row)
```

hll_hash_boolean(bool, int32)

描述：设置hash seed（即改变哈希策略）并对bool类型数据计算哈希值。

返回值类型：hll_hashval

示例：

```
SELECT hll_hash_boolean(FALSE, 10);
  hll_hash_boolean
---------------------
 391264977436098630
(1 row)
```

hll_hash_smallint(smallint)

描述：对smallint类型数据计算哈希值。

返回值类型：hll_hashval

示例：

```
SELECT hll_hash_smallint(100::smallint);
  hll_hash_smallint
---------------------
 4631102066694327276
(1 row)
```

说明

数值大小相同的参数使用不同数据类型的哈希函数计算，最后结果会不一样，因为不同类型哈希函数会选取不同的哈希计算策略。

hll_hash_smallint(smallint, int32)

描述：设置hash seed（即改变哈希策略）同时对smallint类型数据计算哈希值。

返回值类型：hll_hashval

示例：

```
SELECT hll_hash_smallint(100::smallint, 10);
  hll_hash_smallint
---------------------
 8349353095166695771
(1 row)
```

 版权所有 © 华为云计算技术有限公司

hll_hash_integer(integer)

描述：对integer类型数据计算哈希值。

返回值类型：hll_hashval

示例：

```
SELECT hll_hash_integer(0);
  hll_hash_integer
----------------------
 -3485513579396041028
(1 row)
```

hll_hash_integer(integer, int32)

描述：对integer类型数据计算哈希值，并设置hashseed（即改变哈希策略）。

返回值类型：hll_hashval

示例：

```
 SELECT hll_hash_integer(0, 10);
  hll_hash_integer
--------------------
 183371090322255134
(1 row)
```

hll_hash_bigint(bigint)

描述：对bigint类型数据计算哈希值。

返回值类型：hll_hashval

示例：

```
SELECT hll_hash_bigint(100::bigint);
  hll_hash_bigint
---------------------
 8349353095166695771
(1 row)
```

hll_hash_bigint(bigint, int32)

描述：对bigint类型数据计算哈希值，并设置hashseed（即改变哈希策略）。

返回值类型：hll_hashval

示例：

```
SELECT hll_hash_bigint(100::bigint, 10);
  hll_hash_bigint
---------------------
 4631120266694327276
(1 row)
```

hll_hash_bytea(bytea)

描述：对bytea类型数据计算哈希值。

返回值类型：hll_hashval

示例：

 版权所有 © 华为云计算技术有限公司

```
SELECT hll_hash_bytea(E'\\x');
 hll_hash_bytea
----------------
 0
(1 row)
```

hll_hash_bytea(bytea, int32)

描述：对bytea类型数据计算哈希值，并设置hashseed（即改变哈希策略）。

返回值类型：hll_hashval

示例：

```
SELECT hll_hash_bytea(E'\\x', 10);
   hll_hash_bytea
---------------------
 6574525721897061910
(1 row)
```

hll_hash_text(text)

描述：对text类型数据计算哈希值。

返回值类型：hll_hashval

示例：

```
SELECT hll_hash_text('AB');
    hll_hash_text
---------------------
 5365230931951287672
(1 row)
```

hll_hash_text(text, int32)

描述：对text类型数据计算哈希值，并设置hashseed（即改变哈希策略）。

返回值类型：hll_hashval

示例：

```
SELECT hll_hash_text('AB', 10);
    hll_hash_text
---------------------
 7680762839921155903
(1 row)
```

hll_hash_any(anytype)

描述：对任意类型数据计算哈希值。

返回值类型：hll_hashval

示例：

```
SELECT hll_hash_any(1);
     hll_hash_any
----------------------
 -8604791237420463362
(1 row)

SELECT hll_hash_any('08:00:2b:01:02:03'::macaddr);
     hll_hash_any
```

 版权所有 © 华为云计算技术有限公司

```
----------------------
 -4883882473551067169
(1 row)
```

hll_hash_any(anytype, int32)

描述：对任意类型数据计算哈希值，并设置hashseed（即改变哈希策略）。

返回值类型：hll_hashval

示例：

```
SELECT hll_hash_any(1, 10);
    hll_hash_any
----------------------
 -1478847531811254870
(1 row)
```

hll_hashval_eq(hll_hashval, hll_hashval)

描述：比较两个hll_hashval类型数据是否相等。

返回值类型：bool

示例：

```
SELECT hll_hashval_eq(hll_hash_integer(1), hll_hash_integer(1));
 hll_hashval_eq
----------------
 t
(1 row)
```

hll_hashval_ne(hll_hashval, hll_hashval)

描述：比较两个hll_hashval类型数据是否不相等。

返回值类型：bool

示例：

```
SELECT hll_hashval_ne(hll_hash_integer(1), hll_hash_integer(1));
 hll_hashval_ne
----------------
 f
(1 row)
```

7.22.3 精度函数

HLL（HyperLogLog）主要存在三种模式Explicit，Sparse，Full。当数据规模比较小的时候会使用Explicit模式和Sparse模式， 这两种模式在计算结果上基本上没有误差。随着distinct值越来越多，就会转换成Full模式，但结果也会存在一定误差。下列函数用于查看HLL中精度参数。

hll_schema_version(hll)

描述：查看当前hll中的schema version。

返回值类型：integer

示例：

```
SELECT hll_schema_version(hll_empty());
 hll_schema_version
```

 版权所有 © 华为云计算技术有限公司

```
--------------------
         1
(1 row)
```

hll_type(hll)

描述：查看当前hll的类型。

返回值类型：integer

示例：

```
SELECT hll_type(hll_empty());
 hll_type
----------
       1
(1 row)
```

hll_log2m(hll)

描述：查看当前hll的log2m数值，此值会影响最后hll计算distinct误差率，误差率计算公式为：

$$\pm 1.04/\sqrt{2 \wedge log2m}$$

返回值类型：integer

示例：

```
SELECT hll_log2m(hll_empty());
 hll_log2m
-----------
       11
(1 row)
```

hll_regwidth(hll)

描述：查看hll数据结构中桶的位数大小。

返回值类型：integer

示例：

```
SELECT hll_regwidth(hll_empty());
 hll_regwidth
--------------
          5
(1 row)
```

hll_expthresh(hll)

描述：得到当前hll中expthresh大小，hll通常会由Explicit模式到Sparse模式再到Full模式，这个过程称为promotion hierarchy策略。可以通过调整expthresh值的大小改变策略，比如expthresh为0的时候就会跳过Explicit模式而直接进入Sparse模式。当显式指定expthresh的取值为1-7之间时，该函数得到的是 $2^{expthresh}$。

返回值类型：record

示例：

```
SELECT hll_expthresh(hll_empty());
 hll_expthresh
```

 版权所有 © 华为云计算技术有限公司

```
---------------
 (-1,160)
(1 row)

SELECT hll_expthresh(hll_empty(11,5,3));
 hll_expthresh
---------------
 (8,8)
(1 row)
```

hll_sparseon(hll)

描述：是否启用sparse模式，0是关闭，1是开启。

返回值类型：integer

示例：

```
SELECT hll_sparseon(hll_empty());
 hll_sparseon
--------------
            1
(1 row)
```

7.22.4 聚合函数

hll_add_agg(hll_hashval)

描述：把哈希后的数据按照分组放到hll中。

返回值类型：hll

示例：

1. 准备数据。
   ```
   CREATE TABLE t_id(id int);
   INSERT INTO t_id VALUES(generate_series(1,500));
   CREATE TABLE t_data(a int, c text);
   INSERT INTO t_data SELECT mod(id,2), id FROM t_id;
   ```
2. 创建表并指定列为hll。
   ```
   CREATE TABLE t_a_c_hll(a int, c hll);
   ```
3. 根据a列group by对数据分组，把各组数据加到hll中。
   ```
   INSERT INTO t_a_c_hll SELECT a, hll_add_agg(hll_hash_text(c)) FROM t_data GROUP BY a;
   ```
4. 得到每组数据中hll的Distinct值。
   ```
   SELECT a, #c as cardinality FROM t_a_c_hll order by a;
    a |   cardinality
   ---+------------------
    0 | 250.741759091658
    1 | 250.741759091658
   (2 rows)
   ```

hll_add_agg(hll_hashval, int32 log2m)

描述：把哈希后的数据按照分组放到hll中。并指定参数log2m，取值范围为10~16。

返回值类型：hll

示例：

```
SELECT hll_cardinality(hll_add_agg(hll_hash_text(c), 10)) FROM t_data;
 hll_cardinality
------------------
```

 版权所有 © 华为云计算技术有限公司

```
 503.932348927339
(1 row)
```

hll_add_agg(hll_hashval, int32 log2m, int32 regwidth)

描述：把哈希后的数据按照分组放到hll中。依次指定参数log2m，regwidth。regwidth取值范围为1~5。

返回值类型：hll

示例：

```
SELECT hll_cardinality(hll_add_agg(hll_hash_text(c), NULL, 1)) FROM t_data;
 hll_cardinality
------------------
 496.628982624022
(1 row)
```

hll_add_agg(hll_hashval, int32 log2m, int32 regwidth, int64 expthresh)

描述：把哈希后的数据按照分组放到hll中，依次指定参数log2m、regwidth、expthresh。expthresh的取值范围是-1~7之间的整数，该参数可以用来设置从Explicit模式到Sparse模式的阈值大小。-1表示自动模式，0表示跳过Explicit模式，取1~7表示在基数到达 $2^{expthresh}$时切换模式。

返回值类型：hll

示例：

```
 SELECT hll_cardinality(hll_add_agg(hll_hash_text(c), NULL, 1, 4)) FROM t_data;
 hll_cardinality
------------------
 496.628982624022
(1 row)
```

hll_add_agg(hll_hashval, int32 log2m, int32 regwidth, int64 expthresh, int32 sparseon)

描述：把哈希后的数据按照分组放到hll中，依次指定参数log2m、regwidth、expthresh、sparseon。sparseon取值范围为0或者1。

返回值类型：hll

示例：

```
 SELECT hll_cardinality(hll_add_agg(hll_hash_text(c), NULL, 1, 4, 0)) FROM t_data;
 hll_cardinality
------------------
 496.628982624022
(1 row)
```

hll_union_agg(hll)

描述：将多个hll类型数据union成一个hll。

返回值类型：hll

示例：

将各组中的hll数据union成一个hll，并计算distinct值。

```
SELECT #hll_union_agg(c) as cardinality FROM t_a_c_hll;
  cardinality
```

 版权所有 © 华为云计算技术有限公司

```
------------------
 496.628982624022
(1 row)
```

说明

当两个或者多个hll数据结构执行union时，必须要保证其中每一个hll里面的精度参数一样，否则不能进行union。同样的约束也适用于函数hll_union(hll,hll)。

7.22.5 功能函数

hll_print(hll)

描述：打印hll的一些debug参数信息。

返回值类型：cstring

示例：

```
SELECT hll_print(hll_empty());
                          hll_print
------------------------------------------------------------
 EMPTY, nregs=2048, nbits=5, expthresh=-1(160), sparseon=1
(1 row)
```

hll_empty()

描述：创建一个空的hll。

返回值类型：hll

示例：

```
SELECT hll_empty();
 hll_empty
-----------
 \x118b7f
(1 row)
```

hll_empty(int32 log2m)

描述：创建空的hll并指定参数log2m，取值范围是10到16。

返回值类型: hll

示例：

```
SELECT hll_empty(10);
 hll_empty
-----------
 \x118a7f
(1 row)
```

hll_empty(int32 log2m, int32 regwidth)

描述：创建空的hll并依次指定参数log2m、regwidth。regwidth取值范围是1到5。

返回值类型: hll

示例：

```
SELECT hll_empty(10, 4);
 hll_empty
```

 版权所有 © 华为云计算技术有限公司

```
-----------
 \x116a7f
(1 row)
```

hll_empty(int32 log2m, int32 regwidth, int64 expthresh)

描述：创建空的hll并依次指定参数log2m、regwidth、expthresh。expthresh取值范围是-1到7之间的整数。该参数可以用来设置从Explicit模式到Sparse模式的阈值大小。-1表示自动模式，0表示跳过Explicit模式，取1-7表示在基数到达$2^{expthresh}$时切换模式。

返回值类型：hll

示例：

```
SELECT hll_empty(10, 4, 7);
 hll_empty
-----------
 \x116a48
(1 row)
```

hll_empty(int32 log2m, int32 regwidth, int64 expthresh, int32 sparseon)

描述：创建空的hll并依次指定参数log2m、regwidth、expthresh、sparseon。sparseon取0或者1。

返回值类型：hll

示例：

```
SELECT hll_empty(10,4,7,0);
 hll_empty
-----------
 \x116a08
(1 row)
```

hll_add(hll, hll_hashval)

描述：把hll_hashval加入到hll中。

返回值类型：hll

示例：

```
SELECT hll_add(hll_empty(), hll_hash_integer(1));
        hll_add
--------------------------
 \x128b7f8895a3f5af28cafe
(1 row)
```

hll_add_rev(hll_hashval, hll)

描述：把hll_hashval加入到hll中，和hll_add功能一样，只是参数位置进行了交换。

返回值类型：hll

示例：

```
SELECT hll_add_rev(hll_hash_integer(1), hll_empty());
      hll_add_rev
--------------------------
 \x128b7f8895a3f5af28cafe
(1 row)
```

 版权所有 © 华为云计算技术有限公司

hll_eq(hll, hll)

描述：比较两个hll是否相等。

返回值类型：bool

示例：

```
SELECT hll_eq(hll_add(hll_empty(), hll_hash_integer(1)), hll_add(hll_empty(), hll_hash_integer(2)));
 hll_eq
--------
 f
(1 row)
```

hll_ne(hll, hll)

描述：比较两个hll是否不相等。

返回值类型：bool

示例：

```
SELECT hll_ne(hll_add(hll_empty(), hll_hash_integer(1)), hll_add(hll_empty(), hll_hash_integer(2)));
 hll_ne
--------
 t
(1 row)
```

hll_cardinality(hll)

描述：计算hll的distinct值。

返回值类型：integer

示例：

```
SELECT hll_cardinality(hll_empty() || hll_hash_integer(1));
 hll_cardinality
-----------------
               1
(1 row)
```

hll_union(hll, hll)

描述：把两个hll数据结构union成一个。

返回值类型：hll

示例：

```
SELECT hll_union(hll_add(hll_empty(), hll_hash_integer(1)), hll_add(hll_empty(), hll_hash_integer(2)));
                                hll_union
-------------------------------------------------------------------------
 \x128b7f8895a3f5af28cafeda0ce907e4355b60
(1 row)
```

7.22.6 内置函数

HLL（HyperLogLog）有一系列内置函数用于内部对数据进行处理，一般情况下不建议用户使用。

 版权所有 © 华为云计算技术有限公司

表 7-29 内置函数

函数名称	功能描述
hll_in	以string格式接收hll数据。
hll_out	以string格式发送hll数据。
hll_recv	以bytea格式接收hll数据。
hll_send	以bytea格式发送hll数据。
hll_trans_in	以string格式接收hll_trans_type数据。
hll_trans_out	以string格式发送hll_trans_type数据。
hll_trans_recv	以bytea形式接收hll_trans_type数据。
hll_trans_send	以bytea形式发送hll_trans_type数据。
hll_typmod_in	接收typmod类型数据。
hll_typmod_out	发送typmod类型数据。
hll_hashval_in	接收hll_hashval类型数据。
hll_hashval_out	发送hll_hashval类型数据。
hll_add_trans0	类似于hll_add所提供的功能，通常在分布式聚合运算的第一阶段DN上使用。
hll_union_trans	类似hll_union所提供的功能，在分布式聚合运算的第一阶段DN上使用。
hll_union_collect	类似于hll_union所提供的功能，在分布式聚合运算第二阶段CN上使用，汇总各个DN上的结果。
hll_pack	在分布式聚合运算第三阶段CN上使用，把自定义hll_trans_type类型最后转换成hll类型。
hll	用于hll类型转换成hll类型，根据输入参数会设定指定参数。
hll_hashval	用于bigint类型转换成hll_hashval类型。
hll_hashval_int4	用于int4类型转换成hll_hashval类型。

7.23 返回集合的函数

7.23.1 序列号生成函数

generate_series()函数根据指定的开始值(start)、结束值(stop)和步长(step)返回一个基于系列的集合。

generate_series()函数的入参中，当step是正数且start大于stop，则返回零行。相反，当step是负数且start小于stop，则返回零行。如果任何输入为NULL也会返回零行。如果step为零则会报错。

 版权所有 © 华为云计算技术有限公司

generate_series(start, stop)

描述：生成一个数值序列，从start到stop，步长默认为1。

参数类型：int、bigint、numeric

返回值类型：setof int、setof bigint、setof numeric（与参数类型相同）

示例：

```
SELECT * FROM generate_series(2,4);
 generate_series
-----------------
               2
               3
               4
(3 rows)

SELECT * FROM generate_series(4,3);
 generate_series
-----------------
(0 rows)

SELECT * FROM generate_series(1,NULL);
 generate_series
-----------------
(0 rows)
```

generate_series(start, stop, step)

描述：生成一个数值序列，从start到stop，步长为step。

参数类型：int、bigint、numeric

返回值类型：setof int、setof bigint、setof numeric（与参数类型相同）

示例：

```
SELECT * FROM generate_series(5,1,-2);
 generate_series
-----------------
               5
               3
               1
(3 rows)

SELECT * FROM generate_series(4,6,-5);
 generate_series
-----------------
(0 rows)

SELECT * FROM generate_series(4,3,0);
ERROR:  step size cannot equal zero
```

generate_series(start, stop, step interval)

描述：生成一个数值序列，从start到stop，步长为step。

参数类型：timestamp或timestamp with time zone

返回值类型：setof timestamp或setof timestamp with time zone（与参数类型相同）

示例：

 版权所有 © 华为云计算技术有限公司

```
--这个示例应用于date-plus-integer操作符。
SELECT current_date + s.a AS dates FROM generate_series(0,14,7) AS s(a);
   dates
------------
 2017-06-02
 2017-06-09
 2017-06-16
(3 rows)

SELECT * FROM generate_series('2008-03-01 00:00'::timestamp, '2008-03-04 12:00', '10 hours');
   generate_series
---------------------
 2008-03-01 00:00:00
 2008-03-01 10:00:00
 2008-03-01 20:00:00
 2008-03-02 06:00:00
 2008-03-02 16:00:00
 2008-03-03 02:00:00
 2008-03-03 12:00:00
 2008-03-03 22:00:00
 2008-03-04 08:00:00
(9 rows)
```

7.23.2 下标生成函数

generate_subscripts(array anyarray, dim int)

描述：生成一系列包括给定数组的下标。

返回值类型：setof int

示例：

```
SELECT generate_subscripts('{NULL,1,NULL,2}'::int[], 1) AS s;
 s
---
 1
 2
 3
 4
(4 rows)
```

generate_subscripts(array anyarray, dim int, reverse boolean)

描述：生成一系列包括给定数组的下标。当reverse为真时，该系列则以相反的顺序返回。

返回值类型：setof int

示例：

```
SELECT generate_subscripts('{NULL,1,NULL,2}'::int[], 1,TRUE) AS s;
 s
---
 4
 3
 2
 1
(4 rows)
```

generate_subscripts是为给定数组的指定维度生成有效下标集的函数。如果数组中没有所请求的维度或者任何输入NULL数组，均返回零行（但是会给数组元素为空的返回有效下标）。示例：

```
--unnest一个2D数组。
CREATE OR REPLACE FUNCTION unnest2(anyarray)
```

 版权所有 © 华为云计算技术有限公司

```
RETURNS SETOF anyelement AS $$
SELECT $1[i][j]
   FROM generate_subscripts($1,1) g1(i),
        generate_subscripts($1,2) g2(j);
$$ LANGUAGE sql IMMUTABLE;

SELECT * FROM unnest2(ARRAY[[1,2],[3,4]]);
 unnest2
---------
       1
       2
       3
       4
(4 rows)
```

7.24 几何函数和操作符

7.24.1 几何操作符

+

描述：平移，即从第一个参数的每个点的坐标中加上第二个point的坐标。

示例：

```
SELECT box '((0,0),(1,1))' + point '(2.0,0)' AS RESULT;
   result
-------------
 (3,1),(2,0)
(1 row)
```

-

描述：平移，从第一个参数的每个点的坐标中减去第二个point的坐标。

示例：

```
SELECT box '((0,0),(1,1))' - point '(2.0,0)' AS RESULT;
    result
---------------
 (-1,1),(-2,0)
(1 row)
```

*****

描述：伸展/旋转。

示例：

```
SELECT box '((0,0),(1,1))' * point '(2.0,0)' AS RESULT;
   result
-------------
 (2,2),(0,0)
(1 row)
```

/

描述：收缩/旋转。

示例：

 版权所有 © 华为云计算技术有限公司

```
SELECT box '((0,0),(2,2))' / point '(2.0,0)' AS RESULT;
   result
-------------
 (1,1),(0,0)
(1 row)
```

#

描述：两个图形交点或者交面。

示例：

```
SELECT box'((1,-1),(-1,1))' # box'((1,1),(-1,-1))' AS RESULT;
 result
--------
(1,1),(-1,-1)
(1 row)
```

#

描述：图形的路径数目或多边形顶点数。

示例：

```
SELECT # path'((1,0),(0,1),(-1,0))' AS RESULT;
 result
--------
      3
(1 row)
```

@-@

描述：图形的长度或者周长。

示例：

```
SELECT @-@ path '((0,0),(1,0))' AS RESULT;
 result
--------
      2
(1 row)
```

@@

描述：图形的中心。

示例：

```
SELECT @@ circle '((0,0),10)' AS RESULT;
 result
--------
 (0,0)
(1 row)
```

##

描述：第一个图形相对第二个图形的最近点。

示例：

```
SELECT point '(0,0)' ## box '((2,0),(0,2))' AS RESULT;
 result
--------
```

 版权所有 © 华为云计算技术有限公司

```
 (0,0)
(1 row)
```

<->

描述：两个图形之间的距离。

示例：

```
SELECT circle '((0,0),1)' <-> circle '((5,0),1)' AS RESULT;
 result
--------
      3
(1 row)
```

&&

描述：两个图形是否重叠（有一个共同点就为真）。

示例：

```
SELECT box '((0,0),(1,1))' && box '((0,0),(2,2))' AS RESULT;
 result
--------
 t
(1 row)
```

<<

描述：图形是否全部在另一个图形的左边（没有相同的横坐标）。

示例：

```
SELECT circle '((0,0),1)' << circle '((5,0),1)' AS RESULT;
 result
--------
 t
(1 row)
```

>>

描述：图形是否全部在另一个图形的右边（没有相同的横坐标）。

示例：

```
SELECT circle '((5,0),1)' >> circle '((0,0),1)' AS RESULT;
 result
--------
 t
(1 row)
```

&<

描述：图形的最右边是否不超过在另一个图形的最右边。

示例：

```
SELECT box '((0,0),(1,1))' &< box '((0,0),(2,2))' AS RESULT;
 result
--------
 t
(1 row)
```

 版权所有 © 华为云计算技术有限公司

&>

描述：图形的最左边是否不超过在另一个图形的最左边。

示例：

```
SELECT box '((0,0),(3,3))' &> box '((0,0),(2,2))' AS RESULT;
 result
--------
 t
(1 row)
```

<<|

描述：图形是否全部在另一个图形的下边（没有相同的纵坐标）。

示例：

```
SELECT box '((0,0),(3,3))' <<| box '((3,4),(5,5))' AS RESULT;
 result
--------
 t
(1 row)
```

|>>

描述：图形是否全部在另一个图形的上边（没有相同的纵坐标）。

示例：

```
SELECT box '((3,4),(5,5))' |>> box '((0,0),(3,3))' AS RESULT;
 result
--------
 t
(1 row)
```

&<|

描述：图形的最上边是否不超过另一个图形的最上边。

示例：

```
SELECT box '((0,0),(1,1))' &<| box '((0,0),(2,2))' AS RESULT;
 result
--------
 t
(1 row)
```

|&>

描述：图形的最下边是否不超过另一个图形的最下边。

示例：

```
SELECT box '((0,0),(3,3))' |&> box '((0,0),(2,2))' AS RESULT;
 result
--------
 t
(1 row)
```

<^

描述：图形是否低于另一个图形（允许两个图形有接触）。

 版权所有 © 华为云计算技术有限公司

示例：

```
SELECT box '((0,0),(-3,-3))' <^ box '((0,0),(2,2))' AS RESULT;
 result
--------
 t
(1 row)
```

>^

描述：图形是否高于另一个图形（允许两个图形有接触）。

示例：

```
SELECT box '((0,0),(2,2))' >^ box '((0,0),(-3,-3))'  AS RESULT;
 result
--------
 t
(1 row)
```

?#

描述：两个图形是否相交。

示例：

```
SELECT lseg '((-1,0),(1,0))' ?# box '((-2,-2),(2,2))' AS RESULT;
 result
--------
 t
(1 row)
```

?-

描述：图形是否处于水平位置。

示例：

```
SELECT ?- lseg '((-1,0),(1,0))' AS RESULT;
 result
--------
 t
(1 row)
```

?-

描述：图形是否水平对齐。

示例：

```
SELECT point '(1,0)' ?- point '(0,0)' AS RESULT;
 result
--------
 t
(1 row)
```

?|

描述：图形是否处于竖直位置。

示例：

```
SELECT ?| lseg '((-1,0),(1,0))' AS RESULT;
 result
```

 版权所有 © 华为云计算技术有限公司

```
--------
 f
(1 row)
```

?|

描述：图形是否竖直对齐。

示例：

```
SELECT point '(0,1)' ?| point '(0,0)' AS RESULT;
 result
--------
 t
(1 row)
```

?-|

描述：两条线是否垂直。

示例：

```
SELECT lseg '((0,0),(0,1))' ?-| lseg '((0,0),(1,0))' AS RESULT;
 result
--------
 t
(1 row)
```

?||

描述：两条线是否平行。

示例：

```
SELECT lseg '((-1,0),(1,0))' ?|| lseg '((-1,2),(1,2))' AS RESULT;
 result
--------
 t
(1 row)
```

@>

描述：图形是否包含另一个图形。

示例：

```
SELECT circle '((0,0),2)' @> point '(1,1)' AS RESULT;
 result
--------
 t
(1 row)
```

<@

描述：图形是否被包含于另一个图形。

示例：

```
SELECT point '(1,1)' <@ circle '((0,0),2)' AS RESULT;
 result
--------
 t
(1 row)
```

 版权所有 © 华为云计算技术有限公司

~=

描述：两个图形是否相同?

示例：

```
SELECT polygon '((0,0),(1,1))' ~= polygon '((1,1),(0,0))' AS RESULT;
 result
--------
 t
(1 row)
```

7.24.2 几何函数

area(object)

描述：计算图形的面积。

返回类型：double precision

示例：

```
SELECT area(box '((0,0),(1,1))') AS RESULT;
 result
--------
      1
(1 row)
```

center(object)

描述：计算图形的中心。

返回类型：point

示例：

```
SELECT center(box '((0,0),(1,2))') AS RESULT;
 result
---------
 (0.5,1)
(1 row)
```

diameter(circle)

描述：计算圆的直径。

返回类型：double precision

示例：

```
SELECT diameter(circle '((0,0),2.0)') AS RESULT;
 result
--------
      4
(1 row)
```

height(box)

描述：矩形的竖直高度。

返回类型：double precision

示例：

 版权所有 © 华为云计算技术有限公司

```
SELECT height(box '((0,0),(1,1))') AS RESULT;
 result
--------
      1
(1 row)
```

isclosed(path)

描述：图形是否为闭合路径。

返回类型：boolean

示例：

```
SELECT isclosed(path '((0,0),(1,1),(2,0))') AS RESULT;
 result
--------
 t
(1 row)
```

isopen(path)

描述：图形是否为开放路径。

返回类型：boolean

示例：

```
SELECT isopen(path '[(0,0),(1,1),(2,0)]') AS RESULT;
 result
--------
 t
(1 row)
```

length(object)

描述：计算图形的长度。

返回类型：double precision

示例：

```
SELECT length(path '((-1,0),(1,0))') AS RESULT;
 result
--------
      4
(1 row)
```

npoints(path)

描述：计算路径的顶点数。

返回类型：int

示例：

```
SELECT npoints(path '[(0,0),(1,1),(2,0)]') AS RESULT;
 result
--------
      3
(1 row)
```

 版权所有 © 华为云计算技术有限公司

npoints(polygon)

描述：计算多边形的顶点数。

返回类型：int

示例：

```
SELECT npoints(polygon '((1,1),(0,0))') AS RESULT;
 result
--------
      2
(1 row)
```

pclose(path)

描述：把路径转换为闭合路径。

返回类型：path

示例：

```
SELECT pclose(path '[(0,0),(1,1),(2,0)]') AS RESULT;
       result
---------------------
 ((0,0),(1,1),(2,0))
(1 row)
```

popen(path)

描述：把路径转换为开放路径。

返回类型：path

示例：

```
SELECT popen(path '((0,0),(1,1),(2,0))') AS RESULT;
       result
---------------------
 [(0,0),(1,1),(2,0)]
(1 row)
```

radius(circle)

描述：计算圆的半径。

返回类型：double precision

示例：

```
SELECT radius(circle '((0,0),2.0)') AS RESULT;
 result
--------
      2
(1 row)
```

width(box)

描述：计算矩形的水平尺寸。

返回类型：double precision

示例：

 版权所有 © 华为云计算技术有限公司

```
SELECT width(box '((0,0),(1,1))') AS RESULT;
 result
--------
      1
(1 row)
```

7.24.3 几何类型转换函数

box(circle)

描述：将圆转换成矩形。

返回类型：box

示例：

```
SELECT box(circle '((0,0),2.0)') AS RESULT;
                                result
-----------------------------------------------------------------------
 (1.41421356237309,1.41421356237309),(-1.41421356237309,-1.41421356237309)
(1 row)
```

box(point, point)

描述：将点转换成矩形。

返回类型：box

示例：

```
SELECT box(point '(0,0)', point '(1,1)') AS RESULT;
   result
-------------
 (1,1),(0,0)
(1 row)
```

box(polygon)

描述：将多边形转换成矩形。

返回类型：box

示例：

```
SELECT box(polygon '((0,0),(1,1),(2,0))') AS RESULT;
   result
-------------
 (2,1),(0,0)
(1 row)
```

circle(box)

描述：矩形转换成圆。

返回类型：circle

示例：

```
SELECT circle(box '((0,0),(1,1))') AS RESULT;
            result
--------------------------------
 <(0.5,0.5),0.707106781186548>
(1 row)
```

 版权所有 © 华为云计算技术有限公司

circle(point, double precision)

描述：将圆心和半径转换成圆。

返回类型：circle

示例：

```
SELECT circle(point '(0,0)', 2.0) AS RESULT;
  result
-----------
 <(0,0),2>
(1 row)
```

circle(polygon)

描述：将多边形转换成圆。

返回类型：circle

示例：

```
SELECT circle(polygon '((0,0),(1,1),(2,0))') AS RESULT;
                  result
-----------------------------------------------
 <(1,0.333333333333333),0.924950591148529>
(1 row)
```

lseg(box)

描述：矩形对角线转化成线段。

返回类型：lseg

示例：

```
SELECT lseg(box '((-1,0),(1,0))') AS RESULT;
    result
----------------
 [(1,0),(-1,0)]
(1 row)
```

lseg(point, point)

描述：点转换成线段。

返回类型：lseg

示例：

```
SELECT lseg(point '(-1,0)', point '(1,0)') AS RESULT;
    result
----------------
 [(-1,0),(1,0)]
(1 row)
```

path(polygon)

描述：多边形转换成路径。

返回类型：path

示例：

 版权所有 © 华为云计算技术有限公司

```
SELECT path(polygon '((0,0),(1,1),(2,0))') AS RESULT;
      result
---------------------
 ((0,0),(1,1),(2,0))
(1 row)
```

point(double precision, double precision)

描述：结点。

返回类型：point

示例：

```
SELECT point(23.4, -44.5) AS RESULT;
   result
--------------
 (23.4,-44.5)
(1 row)
```

point(box)

描述：矩形的中心。

返回类型：point

示例：

```
SELECT point(box '((-1,0),(1,0))') AS RESULT;
 result
--------
 (0,0)
(1 row)
```

point(circle)

描述：圆心。

返回类型：point

示例：

```
SELECT point(circle '((0,0),2.0)') AS RESULT;
 result
--------
 (0,0)
(1 row)
```

point(lseg)

描述：线段的中心。

返回类型：point

示例：

```
SELECT point(lseg '((-1,0),(1,0))') AS RESULT;
 result
--------
 (0,0)
(1 row)
```

 版权所有 © 华为云计算技术有限公司

point(polygon)

描述：多边形的中心。

返回类型：point

示例：

```
SELECT point(polygon '((0,0),(1,1),(2,0))') AS RESULT;
       result
----------------------
 (1,0.333333333333333)
(1 row)
```

polygon(box)

描述：矩形转换成4点多边形。

返回类型：polygon

示例：

```
SELECT polygon(box '((0,0),(1,1))') AS RESULT;
         result
----------------------------
 ((0,0),(0,1),(1,1),(1,0))
(1 row)
```

polygon(circle)

描述：圆转换成12点多边形。

返回类型：polygon

示例：

```
SELECT polygon(circle '((0,0),2.0)') AS RESULT;

result

------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------
 ((-2,0),(-1.73205080756888,1),(-1,1.73205080756888),(-1.22464679914735e-16,2),(1,1.73205080756888),
(1.73205080756888,1),(2,2.44929359829471e-16),(1.73205080756888,-0.999999999999999),
(1,-1.73205080756888),(3.67394039744206e-16,-2),(-0.999999999999999,-1.73205080756888),
(-1.73205080756888,-1))
(1 row)
```

polygon(npts, circle)

描述：圆转换成npts点多边形。

返回类型：polygon

示例：

```
SELECT polygon(12, circle '((0,0),2.0)') AS RESULT;

result

------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------
 ((-2,0),(-1.73205080756888,1),(-1,1.73205080756888),(-1.22464679914735e-16,2),(1,1.73205080756888),
```

 版权所有 © 华为云计算技术有限公司

```
(1.73205080756888,1),(2,2.44929359829471e-16),(1.73205080756888,-0.999999999999999),
(1,-1.73205080756888),(3.67394039744206e-16,-2),(-0.999999999999999,-1.73205080756888),
(-1.73205080756888,-1))
(1 row)
```

polygon(path)

描述：路径转换成多边形。

返回类型：polygon

示例：

```
SELECT polygon(path '((0,0),(1,1),(2,0))') AS RESULT;
      result
---------------------
 ((0,0),(1,1),(2,0))
(1 row)
```

7.25 网络地址函数和操作符

7.25.1 cidr 和 inet 操作符

操作符<<，<<=，>>，>>=对子网包含进行测试。它们只考虑两个地址的网络部分（忽略任何主机部分），然后判断其中一个网络是等于另外一个网络，还是另外一个网络的子网。

<

描述：小于

示例：

```
SELECT inet '192.168.1.5' < inet '192.168.1.6' AS RESULT;
 result
--------
 t
(1 row)
```

<=

描述：小于或等于

示例：

```
SELECT inet '192.168.1.5' <= inet '192.168.1.5' AS RESULT;
 result
--------
 t
(1 row)
```

=

描述：等于

示例：

```
SELECT inet '192.168.1.5' = inet '192.168.1.5' AS RESULT;
 result
--------
```

 版权所有 © 华为云计算技术有限公司

```
 t
(1 row)
```

>=

描述：大于或等于

示例：

```
SELECT inet '192.168.1.5' >= inet '192.168.1.5' AS RESULT;
 result
--------
 t
(1 row)
```

>

描述：大于

示例：

```
SELECT inet '192.168.1.5' > inet '192.168.1.4' AS RESULT;
 result
--------
 t
(1 row)
```

<>

描述：不等于

示例：

```
SELECT inet '192.168.1.5' <> inet '192.168.1.4' AS RESULT;
 result
--------
 t
(1 row)
```

<<

描述：包含于

示例：

```
SELECT inet '192.168.1.5' << inet '192.168.1/24' AS RESULT;
 result
--------
 t
(1 row)
```

<<=

描述：包含于或等于

示例：

```
SELECT inet '192.168.1/24' <<= inet '192.168.1/24' AS RESULT;
 result
--------
 t
(1 row)
```

 版权所有 © 华为云计算技术有限公司

>>

描述：包含

示例：

```
SELECT inet '192.168.1/24' >> inet '192.168.1.5' AS RESULT;
 result
--------
 t
(1 row)
```

>>=

描述：包含或等于

示例：

```
SELECT inet '192.168.1/24' >>= inet '192.168.1/24' AS RESULT;
 result
--------
 t
(1 row)
```

~

描述：位非

示例：

```
SELECT ~ inet '192.168.1.6' AS RESULT;
    result
---------------
 63.87.254.249
(1 row)
```

&

描述：两个网络地址的每一位都进行“与”操作。

示例：

```
SELECT inet '192.168.1.6' & inet '10.0.0.0' AS RESULT;
 result
---------
 0.0.0.0
(1 row)
```

|

描述：两个网络地址的每一位都进行“或”操作。

示例：

```
SELECT inet '192.168.1.6' | inet '10.0.0.0' AS RESULT;
   result
-------------
 202.168.1.6
(1 row)
```

+

描述：加

 版权所有 © 华为云计算技术有限公司

示例：

```
SELECT inet '192.168.1.6' + 25 AS RESULT;
    result
--------------
 192.168.1.31
(1 row)
```

-

描述：减

示例：

```
SELECT inet '192.168.1.43' - 36 AS RESULT;
   result
-------------
 192.168.1.7
(1 row)
SELECT inet '192.168.1.43' - inet '192.168.1.19' AS RESULT;
 result
--------
     24
(1 row)
```

7.25.2 网络地址函数

函数abbrev，host，text主要是为了提供可选的显示格式。

任何cidr值都能以显式或者隐式的方式转换为inet值，因此能够操作inet值的函数也同样能够操作cidr值。inet值也可以转换为cidr值，此时inet子网掩码右侧的所有位都将转换为零，以创建一个有效的cidr值。另外，用户还可以使用常规的类型转换语法将一个文本字符串转换为inet或cidr值。例如：inet(expression)或colname::cidr。

abbrev(inet)

描述：缩写显示格式文本。

返回类型：text

示例：

```
SELECT abbrev(inet '10.1.0.0/16') AS RESULT;
   result
-------------
 10.1.0.0/16
(1 row)
```

abbrev(cidr)

描述：缩写显示格式文本。

返回类型：text

示例：

```
SELECT abbrev(cidr '10.1.0.0/16') AS RESULT;
 result
---------
 10.1/16
(1 row)
```

 版权所有 © 华为云计算技术有限公司

broadcast(inet)

描述：网络广播地址。

返回类型：inet

示例：

```
SELECT broadcast('192.168.1.5/24') AS RESULT;
      result
------------------
 192.168.1.255/24
(1 row)
```

family(inet)

描述：抽取地址族，4为IPv4，6为IPv6。

返回类型：int

示例：

```
SELECT family('::1') AS RESULT;
 result
--------
      6
(1 row)
```

host(inet)

描述：将主机地址类型抽出为文本。

返回类型：text

示例：

```
SELECT host('192.168.1.5/24') AS RESULT;
   result
-------------
 192.168.1.5
(1 row)
```

hostmask(inet)

描述：为网络构造主机掩码。

返回类型：inet

示例：

```
SELECT hostmask('192.168.23.20/30') AS RESULT;
 result
---------
 0.0.0.3
(1 row)
```

masklen(inet)

描述：抽取子网掩码长度。

返回类型：int

示例：

 版权所有 © 华为云计算技术有限公司

```
SELECT masklen('192.168.1.5/24') AS RESULT;
 result
--------
     24
(1 row)
```

netmask(inet)

描述：为网络构造子网掩码。

返回类型：inet

示例：

```
SELECT netmask('192.168.1.5/24') AS RESULT;
    result
---------------
 255.255.255.0
(1 row)
```

network(inet)

描述：抽取地址的网络部分。

返回类型：cidr

示例：

```
SELECT network('192.168.1.5/24') AS RESULT;
     result
-----------------
 192.168.1.0/24
(1 row)
```

set_masklen(inet, int)

描述：为inet数值设置子网掩码长度。

返回类型：inet

示例：

```
SELECT set_masklen('192.168.1.5/24', 16) AS RESULT;
     result
-----------------
 192.168.1.5/16
(1 row)
```

set_masklen(cidr, int)

描述：为cidr数值设置子网掩码长度。

返回类型：cidr

示例：

```
SELECT set_masklen('192.168.1.0/24'::cidr, 16) AS RESULT;
     result
-----------------
 192.168.0.0/16
(1 row)
```

 版权所有 © 华为云计算技术有限公司

text(inet)

描述：把IP地址和掩码长度抽取为文本。

返回类型：text

示例：

```
SELECT text(inet '192.168.1.5') AS RESULT;
     result
----------------
 192.168.1.5/32
(1 row)
```

trunc(macaddr)

函数trunc(macaddr)返回一个MAC地址，该地址的最后三个字节设置为零。

描述：把后三个字节置为零。

返回类型：macaddr

示例：

```
SELECT trunc(macaddr '12:34:56:78:90:ab') AS RESULT;
      result
-------------------
 12:34:56:00:00:00
(1 row)
```

macaddr类型还支持标准关系操作符（>，<=等）用于词法排序，和按位运算符（~，&和|）非，与和或。

7.26 系统信息函数

7.26.1 会话信息函数

current_catalog

描述：当前数据库的名字（在标准SQL中称“catalog”），与current_database()同义。

返回值类型：name

示例：

```
SELECT current_catalog;
 current_database
------------------
 gaussdb
(1 row)
```

current_database()

描述：当前数据库的名字。

返回值类型：name

示例：

 版权所有 © 华为云计算技术有限公司

```
SELECT current_database();
 current_database
------------------
 gaussdb
(1 row)
```

pg_client_encoding()

描述：用于获取当前客户端的字符编码方式。当遇到字符显示问题时，通过检查客户端编码可快速定位问题。

返回值类型：name

示例：

```
SELECT pg_client_encoding();
 pg_client_encoding
--------------------
 UTF8
(1 row)
```

current_query()

描述：由客户端提交的当前执行语句（可能包含多个声明）。

返回值类型：text

示例：

```
SELECT current_query();
      current_query
--------------------------
 SELECT current_query();
(1 row)
```

current_schema[()]

描述：current_schema返回在搜索路径中第一个顺位有效的模式名。（如果搜索路径为空则返回NULL，没有有效的模式名也返回NULL）。如果创建表或者其他命名对象时没有声明目标模式，则将使用这些对象的模式。

返回值类型：name

示例：

```
SELECT current_schema();
 current_schema
----------------
 public
(1 row)
```

current_schemas(boolean)

描述：current_schemas(boolean)返回搜索路径中所有模式名字的数组。布尔选项决定像pg_catalog这样隐含包含的系统模式是否包含在返回的搜索路径中。

返回值类型：name[]

示例：

```
SELECT current_schemas(true);
   current_schemas
---------------------
```

 版权所有 © 华为云计算技术有限公司

```
 {pg_catalog,public}
(1 row)
```

说明

搜索路径可以通过运行时设置更改。命令是：

```
SET search_path TO schema [, schema, ...]
```

current_user

描述：当前执行环境下的用户名。current_user是用于权限检查的用户标识。通常用来表示会话用户，但是可以通过**SET ROLE**改变这个设置。在函数执行的过程中随着属性SECURITY DEFINER的改变，其值也会改变。

返回值类型：name

示例：

```
SELECT current_user;
 current_user
--------------
 dbadmin
(1 row)
```

inet_client_addr()

描述：显示当前连接的客户端IP信息。

说明

- 此函数只有在远程连接模式下有效。
- 如果是通过本地连接，使用此接口显示为空。

返回值类型：inet

示例：

```
SELECT inet_client_addr();
 inet_client_addr
------------------
 10.10.0.50
(1 row)
```

inet_client_port()

描述：显示当前连接的客户端的端口号。

说明

此函数只有在远程连接模式下有效。

返回值类型：integer

示例：

```
SELECT inet_client_port();
 inet_client_port
------------------
            33143
(1 row)
```

 版权所有 © 华为云计算技术有限公司

inet_server_addr()

描述：显示当前服务器的IP信息。

说明

- 此函数只有在远程连接模式下有效。
- 如果是通过本地连接，使用此接口显示为空。

返回值类型：inet

示例：

```
SELECT inet_server_addr();
 inet_server_addr
------------------
 10.10.0.13
(1 row)
```

inet_server_port()

描述：显示当前服务器的端口。如果是通过Unix-domain socket连接的，则所有这些函数都返回NULL。

说明

此函数只有在远程连接模式下有效。

返回值类型：integer

示例：

```
SELECT inet_server_port();
 inet_server_port
------------------
 8000
(1 row)
```

pg_backend_pid()

描述：当前会话连接的服务进程的进程ID。

返回值类型：integer

示例：

```
SELECT pg_backend_pid();
 pg_backend_pid
-----------------
 140229352617744
(1 row)
```

pg_conf_load_time()

描述：配置加载时间。pg_conf_load_time返回最后加载服务器配置文件的时间戳。

返回值类型：timestamp with time zone

示例：

```
SELECT pg_conf_load_time();
     pg_conf_load_time
```

 版权所有 © 华为云计算技术有限公司

```
-------------------------------
 2017-09-01 16:05:23.89868+08
(1 row)
```

pg_my_temp_schema()

描述：pg_my_temp_schema返回当前会话中临时模式的OID。如果不存在（没有创建临时表），则返回0。如果给定的OID是其他会话中临时模式的OID，pg_is_other_temp_schema则返回true。

返回值类型：oid

示例：

```
SELECT pg_my_temp_schema();
 pg_my_temp_schema
-------------------
                 0
(1 row)
```

pg_is_other_temp_schema(oid)

描述：是否为另一个会话的临时模式。

返回值类型：boolean

示例：

```
SELECT pg_is_other_temp_schema(25356);
 pg_is_other_temp_schema
-------------------------
 f
(1 row)
```

pg_postmaster_start_time()

描述：数据库实例启动时间。pg_postmaster_start_time返回数据库实例启动时的timestamp with time zone。

返回值类型：timestamp with time zone

示例：

```
SELECT pg_postmaster_start_time();
   pg_postmaster_start_time
-------------------------------
 2017-08-30 16:02:54.99854+08
(1 row)
```

pg_trigger_depth()

描述：触发器的嵌套层次。

返回值类型：integer

示例：

```
SELECT pg_trigger_depth();
 pg_trigger_depth
-------------------
                 0
(1 row)
```

 版权所有 © 华为云计算技术有限公司

pgxc_version()

描述：Postgres-XC版本信息。

返回值类型：text

示例：

```
SELECT pgxc_version();
                                                pgxc_version
-----------------------------------------------------------------------------------------------------------
 Postgres-XC 1.1 on x86_64-unknown-linux-gnu, based on PostgreSQL 9.2.4, compiled by g++ (GCC) 5.4.0,
64-bit
(1 row)
```

session_user

描述：会话用户名。session_user通常是连接当前数据库的初始用户，不过系统管理员可以用**SET SESSION AUTHORIZATION**修改这个设置。

返回值类型：name

示例：

```
SELECT session_user;
 session_user
--------------
 dbadmin
(1 row)
```

user

描述：等价于current_user。

返回值类型：name

示例：

```
SELECT user;
 current_user
--------------
 dbadmin
(1 row)
```

version()

描述：版本信息。version返回一个描述服务器版本信息的字符串。

返回值类型：text

示例：

```
SELECT version();
                                                version
-----------------------------------------------------------------------------------------------------------------------------------------
---------
 PostgreSQL 9.2.4 gsql ((GaussDB 8.2.1 build 39137c2d) compiled at 2022-09-23 15:43:11 commit 3629 last
mr 5138 release) on x86_64-unknown-linux-gnu, compiled by g++ (GCC) 5.4.0, 64-bit
(1 row)
```

get_kae_accelerate()

描述：查询当前集群使用KAE加速能力的状态。该函数仅9.1.1.210及以上集群版本支持。

 版权所有 © 华为云计算技术有限公司

返回值类型：text

示例：

```
SELECT get_kae_accelerate();
 get_kae_accelerate
------------------------------------------------------------
 encryption: on
(1 row)
```

7.26.2 访问权限查询函数

has_any_column_privilege(user, table, privilege)

描述：指定用户是否有访问表任何列的权限。

返回类型：boolean

has_any_column_privilege(table, privilege)

描述：当前用户是否有访问表任何列的权限。

返回类型：boolean

备注：has_any_column_privilege检查用户是否以特定方式访问表的任何列。其参数可能与has_table_privilege类似，除了访问权限类型必须是SELECT、INSERT、UPDATE或REFERENCES的一些组合。

说明

拥有表的表级别权限则隐含的拥有该表每列的列级权限，因此如果与has_table_privilege参数相同，has_any_column_privilege总是返回true。但是如果授予至少一列的列级权限也返回成功。

has_column_privilege(user, table, column, privilege)

描述：指定用户是否有访问列的权限。

返回类型：boolean

has_column_privilege(table, column, privilege)

描述：当前用户是否有访问列的权限。

返回类型：boolean

备注：has_column_privilege检查用户是否以特定方式访问一列。其参数类似于has_table_privilege，可以通过列名或属性号添加列。想要的访问权限类型必须是SELECT、INSERT、UPDATE或REFERENCES的一些组合。

说明

拥有表的表级别权限则隐含的拥有该表每列的列级权限。

has_database_privilege(user, database, privilege)

描述：指定用户是否有访问数据库的权限。

返回类型：boolean

 版权所有 © 华为云计算技术有限公司

has_database_privilege(database, privilege)

描述：当前用户是否有访问数据库的权限。

返回类型：boolean

备注：has_database_privilege检查用户是否能以在特定方式访问数据库。其参数类似has_table_privilege。访问权限类型必须是CREATE、CONNECT、TEMPORARY或TEMP（等价于TEMPORARY）的一些组合。

has_foreign_data_wrapper_privilege(user, fdw, privilege)

描述：指定用户是否有访问外部数据封装器的权限。

参数fdw，表示外部数据封装器，取值范围为外部数据封装器名字或ID。

返回类型：boolean

has_foreign_data_wrapper_privilege(fdw, privilege)

描述：当前用户是否有访问外部数据封装器的权限。

返回类型：boolean

备注：has_foreign_data_wrapper_privilege检查用户是否能以特定方式访问外部数据封装器。其参数类似has_table_privilege。访问权限类型必须是USAGE。

has_function_privilege(user, function, privilege)

描述：指定用户是否有访问函数的权限。

返回类型：boolean

has_function_privilege(function, privilege)

描述：当前用户是否有访问函数的权限。

返回类型：boolean

备注：has_function_privilege检查一个用户是否能以指定方式访问函数。其参数类似has_table_privilege。使用文本字符而不是OID声明一个函数时，允许输入的类型和regprocedure数据类型一样（请参考**对象标识符类型**）。访问权限类型必须是EXECUTE。

has_language_privilege(user, language, privilege)

描述：指定用户是否有访问语言的权限。

返回类型：boolean

has_language_privilege(language, privilege)

描述：当前用户是否有访问语言的权限。

返回类型：boolean

备注：has_language_privilege检查用户是否能以特定方式访问过程语言。其参数类似has_table_privilege。访问权限类型必须是USAGE。

 版权所有 © 华为云计算技术有限公司

has_schema_privilege(user, schema, privilege)

描述：指定用户是否有访问模式的权限。

返回类型：boolean

has_schema_privilege(schema, privilege)

描述：当前用户是否有访问模式的权限。

返回类型：boolean

备注：has_schema_privilege检查用户是否能以特定方式访问模式。其参数类似has_table_privilege。访问权限类型必须是CREATE或USAGE的一些组合。

has_server_privilege(user, server, privilege)

描述：指定用户是否有访问外部服务的权限。

返回类型：boolean

has_server_privilege(server, privilege)

描述：当前用户是否有访问外部服务的权限。

返回类型：boolean

备注：has_server_privilege检查用户是否能以指定方式访问外部服务器。其参数类似has_table_privilege。访问权限类型必须是USAGE。

has_table_privilege(user, table, privilege)

描述：指定用户是否有访问表的权限。

返回类型：boolean

has_table_privilege(table, privilege)

描述：当前用户是否有访问表的权限。

返回类型：boolean

备注：has_table_privilege检查用户是否以特定方式访问表。用户可以通过名字或OID（pg_authid.oid）来指定，public表明PUBLIC伪角色，如果缺省该参数，则使用current_user。该表可以通过名字或者OID声明。如果用名字声明，则在必要时可以用模式进行修饰。如果使用文本字符串来声明权限类型，该文本字符串必须是SELECT、INSERT、UPDATE、DELETE、TRUNCATE、REFERENCES或TRIGGER之一的值。可以给权限类型添加WITH GRANT OPTION，用来测试权限是否拥有授权选项。也可以用逗号分隔列出的多个权限类型，如果拥有任何所列出的权限，则结果便为true。

示例：

```
SELECT has_table_privilege('tpcds.web_site', 'select');
 has_table_privilege
---------------------
 t
(1 row)

SELECT has_table_privilege('dbadmin', 'tpcds.web_site', 'select,INSERT WITH GRANT OPTION ');
```

 版权所有 © 华为云计算技术有限公司

```
 has_table_privilege
---------------------
 t
(1 row)
```

pg_has_role(user, role, privilege)

描述：指定用户是否有角色的权限。

返回类型：boolean

pg_has_role(role, privilege)

描述：当前用户是否有角色的权限。

返回类型：boolean

备注：pg_has_role检查用户是否能以特定方式访问角色。其参数类似has_table_privilege，除了public不能用做用户名。访问权限类型必须是MEMBER或USAGE的一些组合。 MEMBER表示的是角色中的直接或间接成员关系（也就是SET ROLE的权限），而USAGE表示无需通过SET ROLE也直接拥有角色的使用权限。

7.26.3 模式可见性查询函数

每个函数对数据库对象执行可见性检查。对于函数和操作符，如果在前面的搜索路径中没有相同的对象名称和参数的数据类型，则此对象是可见的。对于操作符类，则要同时考虑名字和相关索引的访问方法。

所有这些函数都需使用OID来标识要检查的对象。如果用户想通过名字测试对象，则可使用OID别名类型（regclass、regtype、regprocedure、regoperator、regconfig或regdictionary）。

比如，如果一个表所在的模式在搜索路径中，并且在前面的搜索路径中没有同名的表，那么这个表是可见的。它等效于表可以不带明确模式修饰进行引用。比如，要列出所有可见表的名字：

```
SELECT relname FROM pg_class WHERE pg_table_is_visible(oid);
```

pg_collation_is_visible(collation_oid)

描述：排序是否在搜索路径中可见。

返回类型：boolean

pg_conversion_is_visible(conversion_oid)

描述：转换是否在搜索路径中可见。

返回类型：boolean

pg_function_is_visible(function_oid)

描述：函数是否在搜索路径中可见。

返回类型：boolean

 版权所有 © 华为云计算技术有限公司

pg_opclass_is_visible(opclass_oid)

描述：操作符类是否在搜索路径中可见。

返回类型：boolean

pg_operator_is_visible(operator_oid)

描述：操作符是否在搜索路径中可见。

返回类型：boolean

pg_opfamily_is_visible(opclass_oid)

描述：操作符族是否在搜索路径中可见。

返回类型：boolean

pg_table_is_visible(table_oid)

描述：表是否在搜索路径中可见。

返回类型：boolean

pg_ts_config_is_visible(config_oid)

描述：文本检索配置是否在搜索路径中可见。

返回类型：boolean

pg_ts_dict_is_visible(dict_oid)

描述：文本检索词典是否在搜索路径中可见。

返回类型：boolean

pg_ts_parser_is_visible(parser_oid)

描述：文本搜索解析是否在搜索路径中可见。

返回类型：boolean

pg_ts_template_is_visible(template_oid)

描述：文本检索模板是否在搜索路径中可见。

返回类型：boolean

pg_type_is_visible(type_oid)

描述：类型（或域）是否在搜索路径中可见。

返回类型：boolean

 版权所有 © 华为云计算技术有限公司

7.26.4 系统表信息函数

format_type(type_oid, typemod)

描述：获取数据类型的SQL名称。

返回类型：text

备注：

format_type通过数据类型的类型OID以及可能的类型修饰词，返回其SQL名称。如果不知道具体的修饰词，则在类型修饰词的位置传入NULL。类型修饰词一般只对有长度限制的数据类型有意义。format_type所返回的SQL名称中包含数据类型的长度值，其大小是：实际存储长度len - sizeof(int32)，单位字节。数据存储时需要32位的空间来存储用户对数据类型的自定义长度信息，即实际存储长度要比用户定义长度多4个字节。在下例中，format_type返回的SQL名称为“character varying(6)”，6表示varchar类型的长度值是6字节，因此该类型的实际存储长度为10字节。

```
SELECT format_type((SELECT oid FROM pg_type WHERE typname='varchar'), 10);
    format_type
----------------------
 character varying(6)
(1 row)
```

pg_check_authid(role_oid)

描述：检查是否存在给定oid的角色名。

返回类型：bool

pg_describe_object(catalog_id, object_id, object_sub_id)

描述：获取数据库对象的描述。

返回类型：text

备注：pg_describe_object返回由目录OID，对象OID和一个（或许0个）子对象ID指定的数据库对象的描述。这有助于确认存储在pg_depend系统表中对象的身份。

pg_get_constraintdef(constraint_oid)

描述：获取约束的定义。

返回类型：text

pg_get_constraintdef(constraint_oid, pretty_bool)

描述：获取约束的定义。

返回类型：text

备注：pg_get_constraintdef和pg_get_indexdef分别从约束或索引上使用创建命令进行重构。

pg_get_expr(pg_node_tree, relation_oid)

描述：反编译表达式的内部形式，假设其中的任何Vars都引用第二个参数指定的关系。

 版权所有 © 华为云计算技术有限公司

返回类型：text

pg_get_expr(pg_node_tree, relation_oid, pretty_bool)

描述：反编译表达式的内部形式，假设其中的任何Vars都引用第二个参数指定的关系。

返回类型：text

备注：pg_get_expr反编译一个独立表达式的内部形式，比如一个字段的缺省值。在检查系统表的内容的时候很有用。如果表达式可能包含关键字，则指定它们引用相关的OID作为第二个参数；如果没有关键字，设置为零即可。

pg_get_indexdef(index_oid)

描述：获取索引的CREATE INDEX命令。

返回类型：text

index_oid为索引的OID，可以通过PG_STATIO_ALL_INDEXES系统视图查询。

示例：查询索引ds_ship_mode_t1_index1的OID及其创建命令。

```
SELECT indexrelid FROM PG_STATIO_ALL_INDEXES WHERE indexrelname = 'ds_ship_mode_t1_index1';
 indexrelid
------------
     136035
(1 row)
SELECT * FROM pg_get_indexdef(136035);
                                        pg_get_indexdef
-------------------------------------------------------------------------------------------------------------
 CREATE INDEX ds_ship_mode_t1_index1 ON tpcds.ship_mode_t1 USING psort (sm_ship_mode_sk)
TABLESPACE pg_default
(1 row)
```

pg_get_indexdef(index_oid, column_no, pretty_bool)

描述：获取索引的CREATE INDEX命令，或者如果column_no不为零，则只获取一个索引字段的定义。

返回类型：text

```
SELECT * FROM pg_get_indexdef(136035,0,false);
                                        pg_get_indexdef
-------------------------------------------------------------------------------------------------------------
 CREATE INDEX ds_ship_mode_t1_index1 ON tpcds.ship_mode_t1 USING psort (sm_ship_mode_sk)
TABLESPACE pg_default
(1 row)
SELECT * FROM pg_get_indexdef(136035,1,false);
 pg_get_indexdef
-----------------
 sm_ship_mode_sk
(1 row)
```

pg_get_keywords()

描述：获取SQL关键字和类别列表。

返回类型：setof record

备注：pg_get_keywords返回一组关于描述服务器识别SQL关键字的记录。

 版权所有 © 华为云计算技术有限公司

- word列为具体的关键字。
- catcode列为关键字类别代码：U表示非保留关键字，C表示列名关键字，T表示数据类型关键字，R表示完全保留关键字。 U类关键字可以用作其他数据库对象名，C类关键字不能用作函数和数据类型名，T类关键字不能用作表名和表的列名。
- barelabel列为bool类型，表示是否允许在不加AS的情况下表示表或者列的别名。该字段仅9.1.1.200及以上集群版本支持。
- catdesc列为catcode列的描述信息。
- baredesc列为barelabel列的描述信息。该字段仅9.1.1.200及以上集群版本支持。

```
SELECT * FROM (SELECT *, rank() OVER (PARTITION BY catcode, barelabel ORDER BY word) FROM
pg_get_keywords()) WHERE rank = 1;
    word      | catcode | barelabel |                 catdesc                  |     baredesc      | rank
--------------+---------+-----------+------------------------------------------+-------------------+------
 between      | C       | f         | unreserved (cannot be function or type name) | requires AS       |    1
 all          | R       | f         | reserved                                 | requires AS       |    1
 hot          | R       | t         | reserved                                 | can be bare label |    1
 authorization | T      | f         | reserved (can be function or type name)  | requires AS       |    1
 csn_num      | T       | t         | reserved (can be function or type name)  | can be bare label |    1
 abort        | U       | f         | unreserved                               | requires AS       |    1
 auto         | U       | t         | unreserved                               | can be bare label |    1
(7 rows)
```

barelabel字段查询是否允许在不加AS的情况下表示表或者列的别名。

非保留关键字auto的barelabel属性为true，可以不加AS用作表或者列的别名。

```
SELECT 1 auto FROM dual;
 auto
------
    1
(1 row)
```

非保留关键字abort的barelabel属性为false，用作表或者列的别名时必须加AS。

```
SELECT 1 abort FROM dual;
ERROR:  syntax error at or near "abort"

SELECT 1 AS abort FROM dual;
 abort
-------
     1
(1 row)
```

pg_get_ruledef(rule_oid)

描述：获取规则的CREATE RULE命令。

返回类型：text

pg_get_ruledef(rule_oid, pretty_bool)

描述：获取规则的CREATE RULE命令。

返回类型：text

pg_get_userbyid(role_oid)

描述：获取给定OID的角色名。

返回类型：name

备注：pg_get_userbyid通过角色的OID抽取对应的用户名。

 版权所有 © 华为云计算技术有限公司

pg_get_viewdef(viewname text [, pretty bool [, fullflag bool]])

描述：为视图获取底层的SELECT命令。

返回类型：text

备注：

- pg_get_viewdef重构定义视图的SELECT查询。pretty bool参数为true时，显示格式“适合打印”，且该格式易读。pretty bool参数缺省值为false，显示格式不易读。如果用于转储，那么尽可能使用缺省格式。pretty bool参数只对有效视图生效。
- fullflag bool参数为true时，显示视图的完整定义。其缺省值为false。

pg_get_viewdef(viewoid oid [, pretty bool [, fullflag bool]])

描述：为视图获取底层的SELECT命令。

返回类型：text

pg_get_viewdef(view_oid, wrap_column_int)

描述：为视图获取底层的SELECT命令；行字段被换到指定的列数，打印是隐含的。

返回类型：text

pg_get_raw_viewdef(viewname text)

描述：为视图获取完整的原始视图创建语句，包含/**/格式的注释。该函数仅9.1.1.100及以上集群版本支持。

返回类型：text

须知

- 不支持返回--格式注释。
- 针对视图名包含特殊字符场景，无法直接通过viewname查询的情况，可以先通过pg_class查询视图OID，再通过OID转成viewname进行查询。

pg_get_tabledef(table_oid)

描述：根据table_oid获取表定义。

返回类型：text

示例：先通过系统表pg_class获取表customer_t2的OID，再使用此函数查询表customer_t2的定义，可获取创建表customer_t2时的表字段，表的存储方式（行存或列存）及表的分布方式等信息。

```
select oid from pg_class where relname ='customer_t2';
  oid
-------
 17353
(1 row)
```

 版权所有 © 华为云计算技术有限公司

```
select * from pg_get_tabledef(17353);
             pg_get_tabledef
---------------------------------------------
 SET search_path = dbadmin;                +
 CREATE  TABLE customer_t2 (               +
        state_id character(2),             +
        state_name character varying(40), +
        area_id numeric                    +
 )                                         +
 WITH (orientation=column, compression=low)+
 DISTRIBUTE BY HASH(state_id)              +
 TO GROUP group_version1;
(1 row)
```

pg_get_tabledef(table_name)

描述：根据table_name获取表定义。

返回类型：text

备注：pg_get_tabledef重构出表定义的CREATE语句，包含了表定义本身、索引信息、comments信息。对于表对象依赖的group、schema、tablespace、server等信息，需要用户自己去创建，表定义里不会有这些对象的创建语句。

pg_get_tabledef(table_name/table_oid, forCreate boolean)

描述：根据table_name获取表定义。该函数仅9.1.0.100及以上集群版本支持。

返回类型：text

- forCreate参数为true时获得的表定义中不带TO GROUP
- forCreate参数为false时获得的表定义中含有TO GROUP。

pg_get_tabledef(table_name/table_oid, forCreate boolean,withColComm boolean)

描述：根据table_name获取表定义。该函数仅9.1.0.200及以上集群版本支持。

返回类型：text

- withColComment参数为true时，列字段的comment紧随对应字段显示。
- withColComment参数为false时，列字段的comment显示在表定义内容的末尾。

示例：

```
SELECT pg_get_tabledef('person',false,true);
                 pg_get_tabledef
-----------------------------------------------------
 SET search_path = public;                         +
 CREATE  TABLE person (                            +
        id integer COMMENT '学号',                  +
        name character varying(25) COMMENT '姓名',+
        address text COMMENT '地址'                 +
 )                                                 +
 WITH (orientation=row, compression=no)            +
 DISTRIBUTE BY ROUNDROBIN                          +
 TO GROUP group_version1;
(1 row)

SELECT pg_get_tabledef('person',false,false);
              pg_get_tabledef
-----------------------------------------------
```

 版权所有 © 华为云计算技术有限公司

```
SET search_path = public;                +
CREATE  TABLE person (                   +
        id integer,                      +
        name character varying(25),      +
        address text                     +
)                                        +
WITH (orientation=row, compression=no)     +
DISTRIBUTE BY ROUNDROBIN                   +
TO GROUP group_version1;                   +
COMMENT ON COLUMN person.id IS '学号';     +
COMMENT ON COLUMN person.name IS '姓名';   +
COMMENT ON COLUMN person.address IS '地址';
(1 row)
```

pg_get_stat_expressions(oid)

描述：获取扩展统计信息对象中的表达式。仅8.2.1.200及以上集群版本支持。

返回类型：text []

pg_options_to_table(reloptions)

描述：获取存储选项名称/值对的集合。

返回类型：setof record

备注：pg_options_to_table当通过pg_class.reloptions或pg_attribute.attoptions时返回存储选项名字/值对（ option_name/option_value ）的集合。

示例：

```
CREATE TABLE customer_test
(
  state_ID   CHAR(2),
  state_NAME VARCHAR2(40),
  area_ID    NUMBER
)
WITH (ORIENTATION = COLUMN,COMPRESSION=middle);
SELECT pg_options_to_table(reloptions) FROM pg_class WHERE relname='customer_test';
pg_options_to_table
----------------------
 (orientation,column)
 (compression,middle)
 (bucketnums,16384)
 (colversion,2.0)
 (enable_delta,false)
(5 rows)
```

pg_typeof(any)

描述：获取任何值的数据类型。

返回类型：regtype

备注：

pg_typeof返回传递给他的值的数据类型OID。这可能有助于故障排除或动态构造SQL查询。声明此函数返回regtype，这是一个OID别名类型（请参考**对象标识符类型**）；这意味着它是一个为了比较而显示类型名字的OID。

示例：

```
SELECT pg_typeof(33);
 pg_typeof
```

 版权所有 © 华为云计算技术有限公司

```
-----------
 integer
(1 row)

SELECT typlen FROM pg_type WHERE oid = pg_typeof(33);
 typlen
--------
      4
(1 row)
```

collation for (any)

描述：获取参数的排序。

返回类型：text

备注：

表达式collation for返回传递给他的值的排序。示例：

```
SELECT collation for (description) FROM pg_description LIMIT 1;
 pg_collation_for
------------------
 "default"
(1 row)
```

值可能是引号括起来的并且模式限制的。如果没有为参数表达式排序，则返回一个null值。如果参数不是排序的类型，则抛出一个错误。

getdistributekey(table_name)

描述：获取一个hash表的分布列。

返回类型：text

示例：

```
SELECT getdistributekey('item');
 getdistributekey
------------------
 i_item_sk
(1 row)
```

7.26.5 系统函数信息函数

pv_builtin_functions()

描述：查询系统内置函数的信息。

返回类型：record

pg_get_functiondef(func_oid)

描述：获取函数的定义。

返回类型：text

func_oid为函数的OID，可以通过PG_PROC系统表查询。

示例：查询函数justify_days的OID及其函数定义。

```
SELECT oid FROM pg_proc WHERE proname ='justify_days';
 oid
```

 版权所有 © 华为云计算技术有限公司

```
------
 1295
(1 row)

SELECT * FROM pg_get_functiondef(1295);
 headerlines |                          definition
-------------+------------------------------------------------------------
           4 | CREATE OR REPLACE FUNCTION pg_catalog.justify_days(interval)+
             |  RETURNS interval                                           +
             |  LANGUAGE internal                                          +
             |  IMMUTABLE STRICT NOT FENCED NOT SHIPPABLE                  +
             | AS $function$interval_justify_days$function$                +
             |
(1 row)
```

备注：pg_get_functiondef返回函数查询结果为存储过程原始文本格式，自带转义符'\'，便于后续应用的正确解析。

pg_get_function_arguments(func_oid)

描述：获取函数定义的参数列表（带默认值）。

返回类型：text

备注：pg_get_function_arguments返回一个函数的参数列表，需要在CREATE FUNCTION中使用这种格式。

pg_get_function_identity_arguments(func_oid)

描述：获取参数列表来确定一个函数 (不带默认值)。

返回类型：text

备注：pg_get_function_identity_arguments返回需要的参数列表用来标识函数，这种形式需要在ALTER FUNCTION中使用，并且这种形式省略了默认值。

pg_get_function_result(func_oid)

描述：获取函数的RETURNS子句。

返回类型：text

备注：pg_get_function_result为函数返回适当的RETURNS子句。

7.26.6 注释信息函数

col_description(table_oid, column_number)

描述：获取一个表字段的注释。

返回类型：text

备注：col_description返回一个表中字段的注释，通过表OID和字段号来声明。

示例：查询pg_class系统表获取表OID，查询INFORMATION_SCHEMA.COLUMNS系统视图获取column_number。

```
SELECT COLUMN_NAME,ORDINAL_POSITION FROM INFORMATION_SCHEMA.COLUMNS WHERE
TABLE_NAME = 't' AND COLUMN_NAME = 'a';
column_name | ordinal_position
-------------+------------------
```

 版权所有 © 华为云计算技术有限公司

```
a          |               1
(1 row)

SELECT oid FROM pg_class WHERE relname='t';
  oid
-------
 44020
(1 row)

SELECT col_description(44020,1);
    col_description
-----------------------
 This is a test table.
(1 row)
```

obj_description(object_oid, catalog_name)

描述：获取一个数据库对象的注释。

返回类型：text

备注：带有两个参数的obj_description返回一个数据库对象的注释，该对象是通过其OID和其所属的系统表名字声明。比如，obj_description(123456,'pg_class')将返回OID为123456的表的注释。只带一个参数的obj_description只要求对象OID。

obj_description不能用于表字段，因为字段没有自己的OID。

obj_description(object_oid)

描述：获取一个数据库对象的注释。

返回类型：text

shobj_description(object_oid, catalog_name)

描述：获取一个共享数据库对象的注释。

返回类型：text

备注：shobj_description和obj_description功能相似，不同之处仅在于shobj_description用于共享对象。一些系统表是通用于集群中所有数据库的全局表，因此这些表的注释也是全局存储的。

7.26.7 事务 ID 和快照

以下的函数在输出形式中提供服务器事务信息。这些函数的主要用途是为了确定在两个快照之间有哪个事务提交。

pgxc_is_committed(transaction_id)

描述：如果提交或忽略给定的XID（ gxid ）。NULL表示的状态是未知的（ 运行，准备，冻结等 ）。

返回类型：bool

txid_current()

描述：获取当前事务ID。

返回类型：bigint

 版权所有 © 华为云计算技术有限公司

txid_current_snapshot()

描述：获取当前快照。

返回类型：txid_snapshot

txid_snapshot_xip(txid_snapshot)

描述：在快照中获取正在进行的事务ID。

返回类型：setof bigint

txid_snapshot_xmax(txid_snapshot)

描述：获取快照的xmax。

返回类型：bigint

txid_snapshot_xmin(txid_snapshot)

描述：获取快照的xmin。

返回类型：bigint

txid_visible_in_snapshot(bigint, txid_snapshot)

描述：在快照中事务ID是否可见（不使用子事务ID）。

返回类型：boolean

内部事务ID类型（xid）是32位，每40亿事务一次循环。这些函数使用的数据类型txid_snapshot，存储在特定时刻事务ID可见性的信息。其组件描述在**表7-30**。

txid_snapshot的文本表示为：xmin:xmax:xip_list。

例如，10:20:10,14,15表示：xmin=10, xmax=20, xip_list=10, 14, 15。

表 7-30 快照组件

名字	描述
xmin	最早的事务ID（txid）仍然活动。所有较早事务将是已经提交可见的，或者是直接回滚。
xmax	作为尚未分配的txid。所有大于或等于此txids的都是尚未开始的快照时间，因此不可见。
xip_list	当前快照中活动的txids。这个列表只包含在xmin和xmax之间活动的txids；有可能活动的txids高于xmax。 介于大于等于xmin、小于xmax，并且不在这个列表中的txid，在这个时间快照已经完成的，因此按照提交状态查看他是可见还是回滚。这个列表不包含子事务的txids。

pgxc_gtm_txn_status()

描述：获取GTM上当前所有事务的信息。该函数仅9.1.1.300及以上集群版本支持。

 版权所有 © 华为云计算技术有限公司

表 7-31 pgxc_gtm_txn_status()返回字段

名称	类型	描述
slotnum	integer	GTM上的槽位号。
threadid	bigint	GTM上的线程号。
gxid	xid	全局事务ID。
state	text	事务状态。
xmin	xid	最小事务ID。
xmax	xid	最大事务ID。
isolevel	text	事务的隔离级别。 ● read committed：读已提交隔离级别。 ● repeatable read：可重复读隔离级别。
isreadonly	boolean	是否是只读事务。
isvacuum	boolean	是否是Vacuum事务。
nodename	text	节点名称。该字段仅9.1.2及以上集群版本支持。
remotepid	bigint	业务线程号。该字段仅9.1.2及以上集群版本支持。

示例：

```
SELECT * FROM pgxc_gtm_txn_status();
 slotnum |    threadid     | gxid  |  state   | xmin  | xmax  |    isolevel     | isreadonly | isvacuum |
nodename  |    remotepid
---------+-----------------+-------+----------+-------+-------+-----------------+------------+----------
+--------------+-----------------
       5 | 139836075788032 |     0 | STARTING | 1410028 |     0 | read committed  | f          | f        |
coordinator1 | 139956894525952
       4 | 139836073686784 |     0 | STARTING | 1410028 |     0 | repeatable read | f          | f        |
coordinator2 | 140466892542744
       2 | 139836077889280 | 1410028 | STARTING | 1410028 | 1410028 | read committed  | f          | f        |
coordinator1 | 139956894535808
(3 rows)
```

7.26.8 计算子集群函数

pv_compute_pool_workload()

描述：返回计算子集群当前的负载状态。

返回类型：void

示例：

```
SELECT * from pv_compute_pool_workload();
nodename  | rpinuse | maxrp | nodestate
-----------+---------+-------+-----------
datanode1 |       0 |  1000 | normal
datanode2 |       0 |  1000 | normal
(2 rows)
```

 版权所有 © 华为云计算技术有限公司

7.26.9 锁信息函数

pgxc_get_lock_conflicts()

描述：返回集群中有冲突的锁信息。当某个锁正在等待另一个锁，或正被另一个锁等待，则认为该锁是冲突的。

返回类型：setof record

示例：

```
SELECT * from pgxc_get_lock_conflicts();
                                        pgxc_get_lock_conflicts

-------------------------------------------------------------------------------------------------------------------
 (relation,cn_5001,gaussdb,u1,test,,,,,omm,648406,,,"<insufficient
privilege>",140651661784832,ExclusiveLock,f,2024-10-29 17:24:13.425672+08,)
(1 row
```

7.26.10 缓存信息函数

9.1.1.100集群版本开始DWS支持缓存信息函数，缓存信息函数提供共享缓冲区（shared buffer）使用情况的查询接口，帮助用户了解缓存的使用情况，优化缓存策略，从而提高系统的性能和稳定性。

表 7-32 DWS 支持的缓存信息函数

查询维度	功能描述	缓存占用的统计信息	缓存占用的详细信息
实例级	查询实例的整个共享缓冲区的汇总信息，以及具体每个缓冲区的使用情况（不关联到具体表）。	• pg_buffercache_summary() • pgxc_buffercache_summary() • pg_buffercache_usage_counts() • pgxc_buffercache_usage_counts()	• pg_buffercache_pages() • pgxc_buffercache_pages()
表级	将各缓冲区以表为粒度进行聚合，显示每个表占用的共享缓冲区情况。	• pg_buffercache_rel() • pgxc_buffercache_rel()	-

版权所有 © 华为云计算技术有限公司

查询维度	功能描述	缓存占用的统计信息	缓存占用的详细信息
单表查询	快速查询指定表的缓冲区占用情况。	• pg_buffercache_rel(nspname, relname, partname, tempnspname) • pgxc_buffercache_rel(nspname, relname, partname, tempnspname)	• pg_buffercache_pages(nspname, relname, partname, tempnspname) • pgxc_buffercache_pages(nspname, relname, partname, tempnspname)

其中，对于查询单表缓存信息的函数，根据不同的表类型，需要传入不同的参数值，具体请参考表7-33。使用示例可参考pg_buffercache_rel(nspname, relname, partname, tempnspname)。

表 7-33 单表查询入参差异

表类型	nspname	relname	partname	tempnspname
普通表	指定表模式名。	指定表名。	必须为NULL。	必须为NULL。
分区表	指定表模式名。	指定表名。	指定分区名或者指定为NULL查询所有分区。	必须为NULL。
全局临时表	指定表模式名。	指定表名。	必须为NULL。	指定临时模式名或者指定为NULL查询所有临时模式。

pg_buffercache_summary()

描述：汇总当前实例共享缓冲区的使用情况。

返回信息如下：

表 7-34 pg_buffercache_summary()返回字段

名称	类型	描述
buffers_used	integer	已使用的缓冲页数。
buffers_unused	integer	未使用的缓冲页数。

版权所有 © 华为云计算技术有限公司

名称	类型	描述
buffers_dirty	integer	脏缓冲页数。
buffers_pinned	integer	被pin住的缓冲页数。
usagecount_avg	real	缓冲页平均访问计数。

示例：

```
SELECT * FROM pg_catalog.pg_buffercache_summary();
 buffers_used | buffers_unused | buffers_dirty | buffers_pinned | usagecount_avg
--------------+----------------+---------------+----------------+----------------
         1761 |         129311 |           725 |              0 |        4.40829
(1 row)
```

pgxc_buffercache_summary()

描述：汇总所有实例共享缓冲区的使用情况。

返回字段说明：除新增一列nodename标识实例名称外，其他字段与**pg_buffercache_summary()**相同。

示例：

```
SELECT * FROM pg_catalog.pgxc_buffercache_summary();
   nodename    | buffers_used | buffers_unused | buffers_dirty | buffers_pinned | usagecount_avg
---------------+--------------+----------------+---------------+----------------+----------------
 cn_5001       |         1237 |          31531 |            13 |              0 |        3.89248
 cn_5002       |         1047 |          31721 |             9 |              0 |        4.01146
 cn_5003       |         1473 |          31295 |            15 |              0 |        3.93143
 dn_6001_6002 |         1768 |         129304 |           371 |              0 |        4.41572
 dn_6003_6004 |         1549 |         129523 |           180 |              0 |        3.74758
 dn_6005_6006 |         1488 |         129584 |           176 |              0 |         4.0578
(6 rows)
```

pg_buffercache_usage_counts()

描述：在当前实例上，按访问计数汇总共享缓冲区的使用情况。

返回信息如下：

表 7-35 pg_buffercache_usage_counts()返回字段

名称	类型	描述
usage_count	integer	缓冲页访问计数。
buffers	integer	对应访问计数的缓冲页总数。
dirty	integer	对应访问计数的所有缓冲页中的脏缓冲页数。
pinned	integer	对应访问计数的所有缓冲页中被pin住的缓冲页数。

版权所有 © 华为云计算技术有限公司

示例：

```
SELECT * FROM pg_catalog.pg_buffercache_usage_counts();
 usage_count | buffers | dirty | pinned
-------------+---------+-------+--------
           0 |  129302 |     0 |      0
           1 |     141 |     0 |      0
           2 |      82 |     2 |      0
           3 |      59 |     2 |      0
           4 |     109 |     0 |      0
           5 |    1379 |   169 |      0
(6 rows)
```

pgxc_buffercache_usage_counts()

描述：在所有实例上，按访问计数汇总共享缓冲区的使用情况。

返回字段说明：除新增一列nodename标识实例名称外，其他字段与**pg_buffercache_usage_counts()**相同。

示例：

```
SELECT * FROM pg_catalog.pgxc_buffercache_usage_counts() WHERE usage_count = 5;
   nodename    | usage_count | buffers | dirty | pinned
---------------+-------------+---------+-------+--------
 cn_5001       |           5 |     804 |    55 |      0
 cn_5002       |           5 |     694 |    32 |      0
 cn_5003       |           5 |     987 |    61 |      0
 dn_6001_6002 |           5 |    1380 |   124 |      0
 dn_6003_6004 |           5 |     871 |    87 |      0
 dn_6005_6006 |           5 |     975 |    93 |      0
(6 rows)
```

pg_buffercache_rel()

描述：查询当前实例中每张表的共享缓冲区占用统计信息。

返回信息如下：

表 7-36 pg_buffercache_rel 返回字段

名称	类型	描述
nspname	name	模式名。
relname	name	表名。
partname	name	分区名。如果不是分区表，则为NULL。
tempnspname	name	全局临时表的临时模式名。如果不是全局临时表，则为NULL。
buffers	bigint	缓冲页数。
dirty	bigint	脏缓冲页数。
usage_count	bigint	缓冲页访问计数总和。
pinned	bigint	被pin住的缓冲页数。

示例：

版权所有 © 华为云计算技术有限公司

```
SELECT * FROM pg_catalog.pg_buffercache_rel() ORDER BY usage_count DESC LIMIT 10;
  nspname   |              relname               | partname | tempnspname | buffers | dirty | usage_count | pinned
------------+------------------------------------+----------+-------------+---------+-------+-------------+--------
 cstore     | pg_hstore_delta_part_17696         |          |             |     539 |   142 |        2694 |      0
 pg_catalog | pg_attribute                       |          |             |     114 |     1 |         563 |      0
 pg_catalog | pg_class                           |          |             |      53 |     6 |         265 |      1
 pg_catalog | pg_statistic                       |          |             |      46 |    28 |         230 |      0
 cstore     | pg_hstore_delta_part_17696_tid_index |        |             |      44 |    11 |         216 |      0
 pg_catalog | pg_attribute_relid_attnum_index    |          |             |      29 |     1 |         139 |      0
 pg_toast   | pg_toast_2619                      |          |             |      23 |     3 |         109 |      0
 scheduler  | workload_collection                |          |             |      34 |     2 |          94 |      0
 pg_catalog | pg_partition                       |          |             |      18 |     0 |          84 |      0
 pg_catalog | pg_type                            |          |             |      33 |     1 |          77 |      0
(10 rows)
```

pgxc_buffercache_rel()

描述：查询所有实例中每张表的共享缓冲区占用统计信息。

返回字段说明：除新增一列nodename标识实例名称外，其他字段与**pg_buffercache_rel()**相同。

示例：

```
SELECT * FROM pg_catalog.pgxc_buffercache_rel() ORDER BY usage_count DESC LIMIT 10;
   nodename    |  nspname   |            relname            | partname | tempnspname | buffers | dirty | usage_count
| pinned
---------------+------------+-------------------------------+----------+-------------+---------+-------+-------------+--------
 dn_6001_6002 | cstore     | pg_hstore_delta_part_17696 |          |             |     539 |   187 |        2694 |      0
 dn_6003_6004 | cstore     | pg_hstore_delta_part_17704 |          |             |     486 |   266 |        1702 |      0
 dn_6005_6006 | cstore     | pg_hstore_delta_part_17696 |          |             |     445 |   181 |        1559 |      0
 cn_5003      | pg_catalog | pg_attribute               |          |             |     143 |     3 |         711 |      0
 cn_5001      | pg_catalog | pg_attribute               |          |             |     150 |     1 |         671 |      0
 dn_6005_6006 | pg_catalog | pg_attribute               |          |             |     125 |     1 |         574 |      0
 dn_6003_6004 | pg_catalog | pg_attribute               |          |             |     114 |     1 |         563 |      0
 dn_6001_6002 | pg_catalog | pg_attribute               |          |             |     114 |     1 |         563 |      0
 cn_5002      | pg_catalog | pg_attribute               |          |             |      99 |     1 |         483 |      0
 dn_6001_6002 | pg_catalog | pg_class                   |          |             |      53 |     6 |         265 |      1
(10 rows)
```

pg_buffercache_rel(nspname, relname, partname, tempnspname)

描述：查询指定表在当前实例上共享缓冲区占用的统计信息。

参数说明：根据不同的表类型，需要传入不同的参数值。具体参见**表7-33**。

返回字段说明：与**pg_buffercache_rel()**相同。

示例1：查询普通表。

```
SELECT * FROM pg_catalog.pg_buffercache_rel('public', 't1', NULL, NULL);
 nspname | relname | partname | tempnspname | buffers | dirty | usage_count | pinned
---------+---------+----------+-------------+---------+-------+-------------+--------
 public  | t1      |          |             |       4 |     2 |          12 |      0
(1 row)
```

示例2：查询分区表的指定分区。

```
SELECT * FROM pg_catalog.pg_buffercache_rel('public', 'part1', 'p1', NULL);
 nspname | relname | partname | tempnspname | buffers | dirty | usage_count | pinned
---------+---------+----------+-------------+---------+-------+-------------+--------
 public  | part1   | p1       |             |       1 |     1 |           5 |      0
(1 row)
```

示例3：查询分区表的所有分区。

```
SELECT * FROM pg_catalog.pg_buffercache_rel('public', 'part1', NULL, NULL) ORDER BY partname;
 nspname | relname | partname | tempnspname | buffers | dirty | usage_count | pinned
```


版权所有 © 华为云计算技术有限公司

```
---------+---------+----------+-------------+---------+-------+-------------+--------
 public  | part1   | p1       |             |       1 |     1 |           5 |      0
 public  | part1   | p2       |             |       1 |     1 |           5 |      0
(2 rows)
```

示例4：查询全局临时表的指定会话。

```
SELECT * FROM pg_catalog.pg_buffercache_rel('public', 'gtt1', NULL,
'pg_temp_cn_5001_4_1_139815154281712') ORDER BY tempnspname;
 nspname | relname | partname |              tempnspname              | buffers | dirty | usage_count | pinned
---------+---------+----------+---------------------------------------+---------+-------+-------------+--------
 public  | gtt1    |          | pg_temp_cn_5001_4_1_139815154281712 |       4 |     2 |          12 |      0
(1 row)
```

示例5：查询全局临时表的所有会话。

```
SELECT * FROM pg_catalog.pg_buffercache_rel('public', 'gtt1', NULL, NULL) ORDER BY tempnspname;
 nspname | relname | partname |              tempnspname              | buffers | dirty | usage_count | pinned
---------+---------+----------+---------------------------------------+---------+-------+-------------+--------
 public  | gtt1    |          | pg_temp_cn_5001_4_1_139815154281712 |       4 |     2 |          12 |      0
 public  | gtt1    |          | pg_temp_cn_5001_4_2_139815154283080 |       4 |     2 |          12 |      0
(2 rows)
```

pgxc_buffercache_rel(nspname, relname, partname, tempnspname)

描述：查询指定表在所有实例上共享缓冲区占用的统计信息。

参数说明：根据不同的表类型，需要传入不同的参数值。具体参见**表7-33**。

返回字段说明：与**pgxc_buffercache_rel()**相同。

示例：

```
SELECT * FROM pg_catalog.pgxc_buffercache_rel('public', 't1', NULL, NULL) ORDER BY nodename;
   nodename    | nspname | relname | partname | tempnspname | buffers | dirty | usage_count | pinned
---------------+---------+---------+----------+-------------+---------+-------+-------------+--------
 dn_6001_6002 | public  | t1      |          |             |       4 |     2 |          12 |      0
 dn_6003_6004 | public  | t1      |          |             |       4 |     2 |          12 |      0
 dn_6005_6006 | public  | t1      |          |             |       4 |     2 |          12 |      0
(3 rows)
```

pg_buffercache_pages()

描述：查询当前实例上共享缓冲区的详细信息。

返回信息如下：

表 7-37 pg_buffercache_pages 返回字段

名称	类型	描述
bufferid	integer	缓冲页ID，范围为[1, shared_buffers]。
relfilenode	oid	缓冲页对应的物理文件的编号。
reltablespace	oid	缓冲页对应的物理文件所属表空间对应的OID。
reldatabase	oid	缓冲页对应的物理文件所属数据库对应的OID。 如果是跨库共享系统表，该值为0。
relforknumber	smallint	缓冲页对应的物理文件的文件类型。

版权所有 © 华为云计算技术有限公司

名称	类型	描述
relblocknumber	bigint	缓冲页在对应的物理文件中的页面编号。
isdirty	boolean	缓冲页是否为脏页。
usage_count	smallint	缓冲页访问计数。
pinning_backends	integer	pin住缓冲页的后端线程数。

⚠ 注意

- 共享缓存中的每一个缓冲区对应一条记录。未使用的缓冲区中仅bufferid字段有值，其他字段为空。
- 该函数会列出所有缓冲页的使用情况，当shared_buffers较大时会导致结果过多。

示例：

```
SELECT * FROM pg_catalog.pg_buffercache_pages() LIMIT 10;
 bufferid | relfilenode | reltablespace | reldatabase | relforknumber | relblocknumber | isdirty | usage_count |
pinning_backends
----------+-------------+---------------+-------------+---------------+----------------+---------+-------------
+------------------
       1 |       16637 |          1664 |           0 |             0 |              0 | f       |           5 |
0
       2 |       17183 |          1664 |           0 |             0 |              0 | f       |           5 |
0
       3 |       16750 |          1663 |       17187 |             0 |              0 | f       |           5 |
0
       4 |       16621 |          1663 |       17187 |             0 |              0 | f       |           5 |
0
       5 |       16785 |          1663 |       17187 |             0 |              0 | f       |           5 |
0
       6 |       16801 |          1663 |       17187 |             0 |              0 | f       |           5 |
0
       7 |       16797 |          1663 |       17187 |             0 |              0 | f       |           5 |
0
       8 |       16810 |          1663 |       17187 |             0 |              0 | f       |           5 |
0
       9 |       16810 |          1663 |       17187 |             0 |              1 | f       |           2 |
0
      10 |       16810 |          1663 |       17187 |             0 |              2 | f       |           2 |
0
(10 rows)
```

pgxc_buffercache_pages()

描述：查询所有实例上共享缓冲区的详细信息。

返回字段说明：除新增一列nodename标识实例名称外，其他字段与**pg_buffercache_pages()**相同。

⚠ 注意

- 共享缓存中的每一个缓冲区对应一条记录。未使用的缓冲区中仅bufferid字段有值，其他字段为空。
- 该函数会列出所有实例所有缓冲页的使用情况，当集群规模较大或者shared_buffers较大时会导致结果过多。

示例：

```
SELECT * FROM pg_catalog.pgxc_buffercache_pages() WHERE isdirty LIMIT 10;
 nodename | bufferid | relfilenode | reltablespace | reldatabase | relforknumber | relblocknumber | isdirty |
```


版权所有 © 华为云计算技术有限公司

```
usage_count | pinning_backends
----------+----------+-------------+---------------+-------------+---------------+----------------+---------+-------------
+------------------
 cn_5001  |       7 |      17047 |        1664 |         0 |          0 |          1 | t      |         5 |           0
 cn_5001  |      15 |      16926 |        1663 |      17187 |          0 |          0 | t      |         5 |           0
 cn_5001  |      16 |      16931 |        1663 |      17187 |          0 |          1 | t      |         5 |           0
 cn_5001  |      19 |      16865 |        1663 |      17187 |          0 |         53 | t      |         5 |
0
 cn_5001  |      21 |      17022 |        1663 |      17187 |          0 |          0 | t      |         5 |           0
 cn_5001  |      22 |      17024 |        1663 |      17187 |          0 |          1 | t      |         5 |           0
 cn_5001  |      32 |      17050 |        1664 |         0 |          0 |          2 | t      |         5 |           0
 cn_5001  |      40 |      16617 |        1663 |      17187 |          1 |          2 | t      |         5 |           0
 cn_5001  |      41 |      16620 |        1663 |      17187 |          0 |          1 | t      |         5 |           0
 cn_5001  |      44 |      16750 |        1663 |      17187 |          0 |          1 | t      |         5 |           0
(10 rows)
```

pg_buffercache_pages(nspname, relname, partname, tempnspname)

描述：查询指定表在当前实例上占用的共享缓冲区的详细信息。

参数说明：根据不同的表类型，需要传入不同的参数值。具体参见**表7-33**。

返回字段说明：和**pg_buffercache_pages()**相同。

示例：

```
SELECT * FROM pg_catalog.pg_buffercache_pages('public', 't1', NULL, NULL);
 nspname | relname | partname | tempnspname | bufferid | relforknumber | relblocknumber | isdirty |
usage_count | pinning_backends
---------+---------+----------+-------------+----------+---------------+----------------+---------+-------------
+------------------
 public  | t1      |          |             |     1786 |             0 |              0 | t       |           5 |                0
 public  | t1      |          |             |     1787 |             1 |              2 | f       |           1 |                0
 public  | t1      |          |             |     1788 |             1 |              0 | f       |           1 |                0
 public  | t1      |          |             |     1789 |             0 |              1 | t       |           5 |                0
(4 rows)
```

pgxc_buffercache_pages(nspname, relname, partname, tempnspname)

描述：查询指定表在所有实例上占用的共享缓冲区的详细信息。

参数说明：根据不同的表类型，需要传入不同的参数值。具体参见**表7-33**。

返回字段说明：和**pgxc_buffercache_pages()**相同。

示例：

```
SELECT * FROM pg_catalog.pgxc_buffercache_pages('public', 't1', NULL, NULL) ORDER BY nodename,
bufferid;
   nodename    | nspname | relname | partname | tempnspname | bufferid | relforknumber | relblocknumber |
isdirty | usage_count | pinning_backends
---------------+---------+---------+----------+-------------+----------+---------------+----------------+---------
+-------------+------------------
 dn_6001_6002 | public  | t1      |          |             |     1786 |             0 |              0 | t       |           5
|                0
 dn_6001_6002 | public  | t1      |          |             |     1787 |             1 |              2 | f       |           1
|                0
 dn_6001_6002 | public  | t1      |          |             |     1788 |             1 |              0 | f       |           1
|                0
 dn_6001_6002 | public  | t1      |          |             |     1789 |             0 |              1 | t       |           5
|                0
 dn_6003_6004 | public  | t1      |          |             |     1740 |             0 |              0 | t       |           5
|                0
 dn_6003_6004 | public  | t1      |          |             |     1741 |             1 |              2 | f       |           1
|                0
 dn_6003_6004 | public  | t1      |          |             |     1742 |             1 |              0 | f       |           1
|                0
```


版权所有 © 华为云计算技术有限公司

```
 dn_6003_6004 | public  | t1      |        |        | 1743 |          0 |             1 | t     |          5
|              0
 dn_6005_6006 | public  | t1      |        |        | 1196 |          0 |             1 | t     |          5
|              0
 dn_6005_6006 | public  | t1      |        |        | 1197 |          1 |             0 | f     |          1
|              0
 dn_6005_6006 | public  | t1      |        |        | 1198 |          1 |             2 | f     |          1
|              0
 dn_6005_6006 | public  | t1      |        |        | 1199 |          0 |             0 | t     |          5
|              0
(12 rows)
```

7.27 系统管理函数

7.27.1 配置设置函数

配置设置函数是用于查询及修改运行时配置参数的函数。

current_setting(setting_name)

描述：当前的设置值。

返回值类型：text

备注：current_setting用于以查询形式获取setting_name的当前值。和SQL语句SHOW是等效的。比如：

```
SELECT current_setting('datestyle');

 current_setting
-----------------
 ISO, MDY
(1 row)
```

read_global_var(global_setting_name)

描述：用于读取全局变量当前的设置值。

返回值类型：text

备注：read_global_var用于以查询形式获取global_setting_name的当前值。和SQL语句SHOW是等效的。

示例：

```
SET my.var = 1000;
SELECT read_global_var('my.var');
 read_global_var
-----------------
 1000
(1 row)
```

说明

global_setting_name表示类似于my.var类型的变量，其特征为名称中存在小数点且小数点左右均不为空。这种类型的变量叫做全局变量。

全局变量只能通过SET命令设置，不支持gs_guc、ALTER DATABASE dbname SET parameter_name TO value、ALTER USER username SET parameter_name TO value等设置方法，对所有用户可见。

 版权所有 © 华为云计算技术有限公司

set_config(setting_name, new_value, is_local)

描述：设置参数并返回新值。

返回值类型：text

备注：set_config将参数setting_name设置为new_value，如果is_local为true，则新值将只应用于当前事务。如果希望新值应用于当前会话，可以使用false，和SQL语句SET是等效的。比如：

```
SELECT set_config('log_statement_stats', 'off', false);

 set_config
------------
 off
(1 row)
```

注意

如果GUC参数behavior_compat_options设置配置项为DISABLE_SET_GLOBAL_VAR_ON_DATANODE，则不可以通过该函数直接在DN上设置全局变量。

7.27.2 通用文件访问函数

通用文件访问函数提供了对数据库服务器上的文件的本地访问接口。只有数据库集群目录和log_directory目录里面的文件可以访问。使用相对路径访问集群目录里面的文件，以及匹配log_directory配置而设置的路径访问日志文件。只有数据库系统管理员才能使用这些函数。

pg_ls_dir(dirname text)

描述：列出目录中的文件。

返回值类型：setof text

备注：pg_ls_dir返回指定目录里面的除了特殊项“.”和“..”之外所有名字。

示例：

```
SELECT pg_ls_dir('./');
      pg_ls_dir
----------------------
 .postgresql.conf.swp
 postgresql.conf
 pg_tblspc
 PG_VERSION
 pg_ident.conf
 core
 server.crt
 pg_serial
 pg_twophase
 postgresql.conf.lock
 pg_stat_tmp
 pg_notify
 pg_subtrans
 pg_ctl.lock
 pg_xlog
 pg_clog
 base
```


版权所有 © 华为云计算技术有限公司

```
 pg_snapshots
 postmaster.opts
 postmaster.pid
 server.key.rand
 server.key.cipher
 pg_multixact
 pg_errorinfo
 server.key
 pg_hba.conf
 pg_replslot
 .pg_hba.conf.swp
 cacert.pem
 pg_hba.conf.lock
 global
 gaussdb.state
(32 rows)
```

pg_read_file(filename text, offset bigint, length bigint)

描述：返回一个文本文件的内容。

返回值类型：text

备注：pg_read_file返回一个文本文件的一部分，从offset开始，最多返回length字节（如果先达到文件结尾，则小于这个数值）。如果offset是负数，则它是相对于文件结尾回退的长度。如果省略了offset和length，则返回整个文件。

示例：

```
SELECT pg_read_file('postmaster.pid',0,100);
             pg_read_file
-----------------------------------------
 53078                                  +
/srv/BigData/hadoop/data1/coordinator+
 1500022474                             +
 8000                                  +
 /var/run/dws                           +
 localhost                             +
  2
(1 row)
```

pg_read_binary_file(filename text [, offset bigint, length bigint,missing_ok boolean])

描述：返回一个二进制文件的内容。

返回值类型：bytea

备注：pg_read_binary_file的功能与pg_read_file类似，除了结果的返回值为bytea类型不一致，相应地不会执行编码检查。与convert_from函数结合，这个函数可以用来读取用指定编码的一个文件。

```
SELECT convert_from(pg_read_binary_file('filename'), 'UTF8');
```

pg_stat_file(filename text)

描述：返回一个文本文件的状态信息。

返回值类型：record

备注：pg_stat_file返回一条记录，其中包含：文件大小、最后访问时间戳、最后更改时间戳、最后文件状态修改时间戳以及标识传入参数是否为目录的boolean值。典型的用法：

 版权所有 © 华为云计算技术有限公司

```
SELECT * FROM pg_stat_file('filename');
SELECT (pg_stat_file('filename')).modification;
```

示例：

```
SELECT * FROM pg_stat_file('postmaster.pid');

 size |         access         |      modification      |         change
| creation | isdir
------+------------------------+------------------------+------------------------
+----------+-------
  117 | 2017-06-05 11:06:34+08 | 2017-06-01 17:18:08+08 | 2017-06-01 17:18:08+08
|          | f
(1 row)
SELECT (pg_stat_file('postmaster.pid')).modification;
      modification
------------------------
 2017-06-01 17:18:08+08
(1 row)
```

7.27.3 服务器信号函数

服务器信号函数向其他服务器进程发送控制信号。只有系统管理员才能使用这些函数。

pg_cancel_backend(pid int)

描述：取消指定后端线程的当前查询。

返回值类型：boolean

备注：pg_cancel_backend函数向指定的后端进程（由pid标识）发送一个查询取消信号（SIGINT）。要获取活动后端进程的PID，可以通过pg_stat_activity视图的pid字段查询，或者在服务器上使用ps命令列出数据库进程。

示例：

```
SELECT pid FROM pg_stat_activity WHERE stmt_type ='RESET';
       pid
-----------------
 281471222065200
(1 row)

SELECT pg_cancel_backend(281471222065200);
 pg_cancel_backend
-------------------
 t
(1 row)
```

pg_cancel_backend(pid int, reason text)

描述：取消指定后端线程的当前查询，并附带取消的原因。该函数仅8.3.0及以上集群版本支持。

- pid，要取消的后端线程ID。
- reason，用于说明取消的原因的文本字符串。

返回值类型：boolean

备注：

- pg_cancel_backend函数向指定的后端进程（由pid标识）发送一个查询取消信号（SIGINT），并附带一个取消的原因。

 版权所有 © 华为云计算技术有限公司

- 9.1.0.226及以上9.1.0.xxx集群版本、9.1.1.200及以上集群版本中支持说明取消原因为“fault injection:malloc”，表示会对目标线程注入一个内存分配故障，可以取消不响应信号且进行内存分配的线程。
- 9.1.1.300及以上集群版本中支持说明取消原因为“fault injection:hash_seq_search”，表示会对目标线程注入一个hash_seq_search报错，可以取消不响应信号且进行hash_seq_search的线程。
- 9.1.1.300及以上集群版本中支持说明取消原因为“fault injection:sync_quit”，表示会对目标线程注入一个sync_quit退出标记，可以取消不响应信号且状态为sync_quit的线程，此场景必须先终止CN线程，再终止DN上tlevel>0的stream线程，最后终止DN上tlevel等于0的top consumer线程。

示例：

```
SELECT pid FROM pg_stat_activity WHERE stmt_type ='RESET';
       pid
-----------------
 281471222065200
(1 row)

SELECT pg_cancel_backend(281471222065200, 'fault injection:malloc');
 pg_cancel_backend
-------------------
 t
(1 row)
```

pg_reload_conf()

描述：服务器进程重新装载配置文件。

返回值类型：boolean

备注：pg_reload_conf给服务器发送一个SIGHUP信号，导致所有服务器进程重新装载配置文件。

示例：

```
SELECT pg_reload_conf();
 pg_reload_conf
----------------
 t
(1 row)
```

pg_rotate_logfile()

描述：滚动服务器的日志文件。

返回值类型：boolean

备注：pg_rotate_logfile通知日志文件管理器立即切换到一个新的输出文件。该函数仅在内置日志收集器运行时有效。

示例：

```
SELECT pg_rotate_logfile();
 pg_rotate_logfile
-------------------
 t
(1 row)
```

 版权所有 © 华为云计算技术有限公司

pg_terminate_backend(pid int)

描述：终止一个后台线程。

返回值类型：boolean

备注：如果成功，函数返回true，否则返回false。

示例：

```
SELECT pid FROM pg_stat_activity;
       pid
-----------------
 140657876268816
 140433774061312
 140433587902208
 140433656592128
 140433723717376
 140433637189376
 140433552770816
 140433481983744
 140433349310208
(9 rows)

SELECT pg_terminate_backend(140657876268816);
 pg_terminate_backend
----------------------
 t
(1 row)
```

pg_wlm_jump_queue(pid int)

描述：调整任务到CN队列的最前端。

返回值类型：boolean

备注：如果成功，函数返回true，否则返回false。

示例：

```
SELECT pid FROM pg_stat_activity WHERE stmt_type ='RESET';
       pid
-----------------
 281471222065200
(1 row)

SELECT pg_wlm_jump_queue(281471222065200);
 pg_wlm_jump_queue
-------------------
 t
(1 row)
```

gs_wlm_switch_cgroup(pid int, cgroup text)

描述：调整作业的优先级到新控制组。

返回值类型：boolean

备注：如果成功，函数返回true，否则返回false。

pg_cancel_query(queryId int)

描述：取消一个后端的当前查询。该函数8.1.2及以上版本支持。

返回值类型：boolean

 版权所有 © 华为云计算技术有限公司

备注：pg_cancel_query向由query_id标识的后端进程发送一个查询取消（SIGINT）信号。一个活动的后端进程的query_id可以从pg_stat_activity视图的query_id字段找到。

示例：

```
SELECT query_id FROM pgxc_stat_activity WHERE stmt_type ='RESET';
 query_id
----------
        0
        0
(2 rows)

SELECT pg_cancel_query(0);
 pg_cancel_query
-----------------
 f
(1 row)
```

pgxc_cancel_query(queryId int)

描述：取消当前集群下正在执行的查询。该函数8.1.2及以上版本支持。

返回值类型：boolean

备注：如果所有节点的查询均已取消，函数返回true，否则返回false。

示例：

```
SELECT query_id FROM pgxc_stat_activity WHERE stmt_type ='RESET';
 query_id
----------
        0
        0
(2 rows)

SELECT pgxc_cancel_query(0);
 pgxc_cancel_query
-------------------
 f
(1 row)
```

pg_terminate_query(queryId int)

描述：终止一个后端的当前查询。该函数8.1.2及以上版本支持。

返回值类型：boolean

示例：

```
SELECT query_id FROM pgxc_stat_activity WHERE stmt_type ='RESET';
 query_id
----------
        0
        0
(2 rows)

SELECT pg_terminate_query(0);
 pg_terminate_query
--------------------
 f
(1 row)
```

pgxc_terminate_query(queryId int)

描述：终止当前集群下正在执行的查询。该函数8.1.2及以上版本支持。

 版权所有 © 华为云计算技术有限公司

返回值类型：boolean

示例：

```
SELECT query_id FROM pgxc_stat_activity;
    query_id
-----------------
 72339069014638631
(1 row)

SELECT pgxc_terminate_query(72339069014638631);
 pgxc_terminate_query
----------------------
 t
(1 row)
```

7.27.4 快照同步函数

快照同步函数是导出当前快照的标识符。

create_wdr_snapshot()

描述：创建性能数据快照。

返回值类型：text

说明

- 该函数只有数据库管理员SYSADMIN才可以执行，非管理员执行会提示无权限。
- 该函数只能在CN上执行，在DN上执行会返回：“WDR snapshot can only be created on coordinator.”。
- 执行该函数前需确认enable_wdr_snapshot参数处于开启状态。如果enable_wdr_snapshot为off，执行该函数会返回：“WDR snapshot request can't be executed, because GUC parameter 'enable_wdr_snapshot' is off.”。
- 如果执行该函数时，快照线程由于节点重启等原因尚未启动，会提示错误：“WDR snapshot request can not be accepted, please retry later.”。
- 如果执行该函数失败，会提示：“Cannot respond to WDR snapshot request.”。
- 如果执行成功，会返回：“WDR snapshot request has been submitted.”。该提示表明创建快照请求已发送至后台快照线程，但不代表创建快照成功。

kill_snapshot(scope cstring)

描述：中止后台快照线程。该函数向后台快照线程发送中止信号并等待线程结束。

输入参数scope：表示操作范围。该参数取值范围为local和global。

- local表示中止当前CN上的快照线程。
- global表示不仅会中止当前CN上的快照线程，还会向集群中所有其他CN发送中止快照线程的请求，即中止集群中所有CN上的快照线程。
- 如果输入其他值，则报错“Scope is invalid, use "local" or "global".”。
- 输入参数可为空，表示默认取值为local。

返回值类型：无

 版权所有 © 华为云计算技术有限公司

说明

- 该函数只有数据库管理员SYSADMIN才有权执行，非管理员执行会提示无权限。
- 该函数只能在CN上执行，在DN上执行会提示：“kill_snapshot can only be executed on coordinator.”；
- 执行该函数会向后台快照线程发送中止信号并等待其结束。如果100s内快照线程仍未中止则会报错：“Kill snapshot thread failed”；

pg_export_snapshot()

描述：保存当前的快照并返回它的标识符。

返回值类型：text

备注：函数pg_export_snapshot保存当前的快照并返回一个文本字符串标识此快照。这个字符串必须传递给想要导入快照的客户端。可用在set transaction snapshot snapshot_id时导入snapshot，但是应用的前提是该事务设置了REPEATABLE READ隔离级别。该函数的输出不可用做set transaction snapshot的输入。

示例：

```
SELECT pg_export_snapshot();
 pg_export_snapshot
--------------------
 00000000000A7128-1
(1 row)
```

7.27.5 咨询锁函数

咨询锁函数用于管理咨询锁（Advisory Lock），此类函数目前仅内部使用。

pg_advisory_lock(key bigint)

描述：获取会话级别的排它咨询锁。

返回值类型：void

备注：pg_advisory_lock锁定应用程序定义的资源，该资源可以用一个64位或两个不重叠的32位键值标识。如果已经有另外的会话锁定了该资源，则该函数将阻塞到该资源可用为止。这个锁是排它的。多个锁定请求将会被压入栈中，因此，如果同一个资源被锁定了三次，它必须被解锁三次以将资源释放给其他会话使用。

pg_advisory_lock(key1 int, key2 int)

描述：获取会话级别的排它咨询锁。

返回值类型：void

pg_advisory_lock_shared(key bigint)

描述：获取会话级别的共享咨询锁。

返回值类型：void

pg_advisory_lock_shared(key1 int, key2 int)

描述：获取会话级别的共享咨询锁。

 版权所有 © 华为云计算技术有限公司

返回值类型：void

备注：pg_advisory_lock_shared类似于pg_advisory_lock，不同之处仅在于共享锁会话可以和其他请求共享锁的会话共享资源，但排它锁除外。

pg_advisory_unlock(key bigint)

描述：释放会话级别的排它咨询锁。

返回值类型：boolean

pg_advisory_unlock(key1 int, key2 int)

描述：释放会话级别的排它咨询锁。

返回值类型：boolean

备注：pg_advisory_unlock释放先前取得的排它咨询锁。如果释放成功则返回true。如果实际上并未持有指定的锁，将返回false并在服务器中产生一条SQL警告信息。

pg_advisory_unlock_shared(key bigint)

描述：释放会话级别的共享咨询锁。

返回值类型：boolean

pg_advisory_unlock_shared(key1 int, key2 int)

描述：释放会话级别的共享咨询锁。

返回值类型：boolean

备注：pg_advisory_unlock_shared类似于pg_advisory_unlock，不同之处在于该函数释放的是共享咨询锁。

pg_advisory_unlock_all()

描述：释放当前会话持有的所有咨询锁。

返回值类型：void

备注：pg_advisory_unlock_all将会释放当前会话持有的所有咨询锁，该函数在会话结束的时候被隐含调用，即使客户端异常地断开连接也是一样。

pg_advisory_xact_lock(key bigint)

描述：获取事务级别的排它咨询锁。

返回值类型：void

pg_advisory_xact_lock(key1 int, key2 int)

描述：获取事务级别的排它咨询锁。

返回值类型：void

备注：pg_advisory_xact_lock类似于pg_advisory_lock，不同之处在于锁是自动在当前事务结束时释放，而且不能被显式的释放。

 版权所有 © 华为云计算技术有限公司

pg_advisory_xact_lock_shared(key bigint)

描述：获取事务级别的共享咨询锁。

返回值类型：void

pg_advisory_xact_lock_shared(key1 int, key2 int)

描述：获取事务级别的共享咨询锁。

返回值类型：void

备注：pg_advisory_xact_lock_shared类似于pg_advisory_lock_shared，不同之处在于锁是在当前事务结束时自动释放，而且不能被显式的释放。

pg_try_advisory_lock(key bigint)

描述：尝试获取会话级排它咨询锁。

返回值类型：boolean

备注：pg_try_advisory_lock类似于pg_advisory_lock，不同之处在于该函数不会阻塞以等待资源的释放。它要么立即获得锁并返回true，要么返回false表示目前不能锁定。

pg_try_advisory_lock(key1 int, key2 int)

描述：尝试获取会话级排它咨询锁。

返回值类型：boolean

pg_try_advisory_lock_shared(key bigint)

描述：尝试获取会话级共享咨询锁。

返回值类型：boolean

pg_try_advisory_lock_shared(key1 int, key2 int)

描述：尝试获取会话级共享咨询锁。

返回值类型：boolean

备注：pg_try_advisory_lock_shared类似于pg_try_advisory_lock，不同之处在于该函数尝试获得共享锁而不是排它锁。

pg_try_advisory_xact_lock(key bigint)

描述：尝试获取事务级别的排它咨询锁。

返回值类型：boolean

pg_try_advisory_xact_lock(key1 int, key2 int)

描述：尝试获取事务级别的排它咨询锁。

返回值类型：boolean

 版权所有 © 华为云计算技术有限公司

备注：pg_try_advisory_xact_lock类似于pg_try_advisory_lock，不同之处在于如果得到锁，在当前事务的结束时自动释放，而且不能被显式的释放。

pg_try_advisory_xact_lock_shared(key bigint)

描述：尝试获取事务级别的共享咨询锁。

返回值类型：boolean

pg_try_advisory_xact_lock_shared(key1 int, key2 int)

描述：尝试获取事务级别的共享咨询锁。

返回值类型：boolean

备注：pg_try_advisory_xact_lock_shared类似于pg_try_advisory_lock_shared，不同之处在于如果得到锁，在当前事务结束时自动释放，而且不能被显式的释放。

7.27.6 复制函数

复制函数是系统为实现高可用在各个实例间进行日志同步或数据同步所提供的统计或操作方法。

说明

除统计查询外的复制函数为内部调用函数，不建议用户直接使用。

pg_create_logical_replication_slot('slot_name', 'plugin_name')

描述：创建逻辑复制槽。该函数在9.1.0.200及之后版本已废弃，为兼容历史版本功能保留该函数。

参数说明：

- slot_name
  流复制槽名称。
  取值范围：字符串，不支持除字母，数字，以及（ _?-. ）以外的字符。
- plugin_name
  插件名称。
  取值范围：字符串，当前只支持“mppdb_decoding”。

返回值类型：name, text

备注：第一个返回值表示slot_name，第二个返回值表示该逻辑复制槽解码的起始LSN位置。

pg_create_physical_replication_slot ('slot_name', isDummyStandby)

描述：创建物理复制槽。

参数说明：

- slot_name
  流复制槽名称。
  取值范围：字符串，不支持除字母，数字，以及（ _?-. ）以外的字符。

 版权所有 © 华为云计算技术有限公司

- isDummyStandby
  复制槽是否为从备。
  取值范围：布尔值，支持true或false。

返回值类型：name, text

备注：第一个返回值表示slot_name，第二个返回值表示该物理复制槽解码的起始LSN位置。

pg_get_replication_slots()

描述：显示当前DN上所有的复制槽信息。

返回值类型：record

函数返回信息如下：

表 7-38 pg_get_replication_slots()字段

名称	类型	描述
slot_name	text	复制槽的名称
plugin	name	逻辑复制槽对应的输出插件名
slot_type	text	复制槽的类型
datoid	oid	复制槽的数据库OID
active	boolean	复制槽是否为激活状态
xmin	xid	复制槽事务标识
catalog_xmin	text	逻辑复制槽对应的最早解码事务标识
restart_lsn	text	复制槽的Xlog文件信息
dummy_standby	boolean	复制槽是否为从备

示例：

```
SELECT * FROM pg_get_replication_slots();
    slot_name    | plugin | slot_type | datoid | active | xmin | catalog_xmin |    restart_lsn    | dummy_standby
-----------------+--------+-----------+--------+--------+------+--------------+-------------------+---------------
 gs_roach_common |        | physical  |      0 | f      |      |    602861775 | FFFFFFFF/FFFFFFFF | f
(1 row)
```

pg_drop_replication_slot('slot_name')

描述：删除流复制槽。

参数说明：

- slot_name
  流复制槽名称。
  取值范围：字符串，不支持除字母，数字，以及（ _?-. ）以外的字符。

 版权所有 © 华为云计算技术有限公司

返回值类型：void

pg_logical_slot_peek_changes('slot_name', 'LSN', upto_nchanges, 'options_name', 'options_value')

描述：解码并不推进流复制槽（下次解码可以再次获取本次解出的数据）。该函数在9.1.0.200及之后版本已废弃，为兼容历史版本功能保留该函数。

参数说明：

- slot_name

  流复制槽名称。

  取值范围：字符串，不支持除字母，数字，以及（_?-.）以外的字符。

- LSN

  日志的LSN，表示只解码小于等于此LSN的日志。

  取值范围：字符串（LSN，格式为xlogid/xrecoff），如'1/2AAFC60'。为NULL时表示不对解码截止的日志位置做限制。

- upto_nchanges

  解码条数（包含begin和commit）。假设一共有三条事务，分别包含3、5、7条记录，如果upto_nchanges为4，那么会解码出前两个事务共8条记录。解码完第二条事务时发现解码条数记录大于等于upto_nchanges，会停止解码。

  取值范围：非负整数。

  说明

  LSN和upto_nchanges中任一参数达到限制，解码都会结束。

- options：此项为可选参数。
  - include-xids

    解码出的data列是否包含xid信息。

    取值范围：0或1，默认值为1。

    - 0：设为0时，解码出的data列不包含xid信息。
    - 1：设为1时，解码出的data列包含xid信息。
  - skip-empty-xacts

    解码时是否忽略空事务信息。

    取值范围：0或1，默认值为0。

    - 0：设为0时，解码时不忽略空事务信息。
    - 1：设为1时，解码时会忽略空事务信息。
  - include-timestamp

    解码信息是否包含commit时间戳。

    取值范围：0或1，默认值为0。

    - 0：设为0时，解码信息不包含commit时间戳。
    - 1：设为1时，解码信息包含commit时间戳。

返回值类型：text, uint, text

 版权所有 © 华为云计算技术有限公司

备注：函数返回解码结果，每一条解码结果包含三列，对应上述返回值类型，分别表示LSN位置、xid和解码内容。

pg_logical_slot_get_changes('slot_name', 'LSN', upto_nchanges, 'options_name', 'options_value')

描述：解码并推进流复制槽。该函数在9.1.0.200及之后版本已废弃，为兼容历史版本功能保留该函数。

参数说明： 与pg_logical_slot_peek_changes一致，详细内容请参见**pg_logical_slot_peek_changes('slot_name', 'LSN', upto_nchanges, 'options_name', 'options_value')**。

pg_replication_slot_advance ('slot_name', 'LSN')

描述：直接推进流复制槽到指定LSN，不输出解码结果。

参数说明：

- slot_name

  流复制槽名称。

  取值范围：字符串，不支持除字母，数字，以及（ _?-. ）以外的字符。

- LSN

  推进到的日志LSN位置，下次解码时只会输出提交位置比该LSN大的事务结果。如果输入的LSN比当前流复制槽记录的推进位置还要小，则直接返回；如果输入的LSN比当前最新物理日志LSN还要大，则推进到当前最新物理日志LSN。

  取值范围：字符串（ LSN，格式为xlogid/xrecoff ）。

返回值类型：name, text

备注：返回值分别对应slot_name和实际推进至的LSN。

pg_stat_get_data_senders()

描述：显示当前DN上所有的数据页复制发送线程的统计信息。

返回值类型：record

函数返回信息如下：

表 7-39 pg_stat_get_data_senders()字段

名称	类型	描述
pid	bigint	线程pid。
sender_pid	integer	当前sender的pid。
local_role	text	本地的角色。
peer_role	text	对端的角色。
state	text	当前sender的复制状态。

 版权所有 © 华为云计算技术有限公司

名称	类型	描述
catchup_start	timestamp with time zone	catchup启动的时间。
catchup_end	timestamp with time zone	catchup结束的时间。
queue_size	text	数据队列大小。
queue_lower_tail	text	数据队列尾1位置。
queue_header	text	数据队列头位置。
queue_upper_tail	text	数据队列尾2位置。
send_position	text	发送端发送的位置。
receive_position	text	接收端接收的位置。
catchup_type	text	catchup方式为全量还是增量。
catchup_bcm_filename	text	catchup当前执行的bcm文件。
catchup_bcm_finished	integer	catchup已操作完成的bcm文件数量。
catchup_bcm_total	integer	catchup总共需要操作的bcm文件数量。
catchup_percent	text	catchup已经操作完成的百分比。
catchup_remaining_time	text	catchup预估剩余时间。

pg_stat_get_wal_senders()

描述：显示当前DN上所有的WAL复制发送线程的统计信息。

返回值类型：record

函数返回信息如下：

表 7-40 pg_stat_get_wal_senders()字段

名称	类型	描述
pid	bigint	线程pid
sender_pid	integer	当前sender的pid
local_role	text	本地的角色
peer_role	text	对端的角色
peer_state	text	对端的状态

 版权所有 © 华为云计算技术有限公司

名称	类型	描述
state	text	当前sender的复制状态
catchup_start	timestamp with time zone	catchup启动的时间
catchup_end	timestamp with time zone	catchup结束的时间
sender_sent_location	text	发送端发送的LSN位置
sender_write_location	text	发送端write的LSN位置
sender_flush_location	text	发送端flush的LSN位置
sender_replay_location	text	发送端replay的LSN位置
receiver_received_location	text	接收端received的LSN位置
receiver_write_location	text	接收端write的LSN位置
receiver_flush_location	text	接收端flush的LSN位置
receiver_replay_location	text	接收端replay的LSN位置
sync_percent	text	同步百分比
sync_state	text	同步状态（异步复制，同步复制，还是潜在同步者）
sync_priority	integer	同步复制的优先级（0表示异步）
sync_most_available	text	在备机同步失败时，是否阻塞主机
channel	text	WALSender的信道信息

pg_stat_get_wal_receiver()

描述：显示当前DN上所有的WAL复制接收线程的统计信息。

返回值类型：record

函数返回信息如下：

 版权所有 © 华为云计算技术有限公司

表 7-41 pg_stat_get_wal_receiver()

名称	类型	描述
receiver_pid	integer	当前receiver的pid
local_role	text	本地的角色
peer_role	text	远端的角色
peer_state	text	远端的状态
state	text	当前receiver的复制状态
sender_sent_location	text	发送端发送的LSN位置
sender_write_location	text	发送端write的LSN位置
sender_flush_location	text	发送端flush的LSN位置
sender_replay_location	text	发送端replay的LSN位置
receiver_received_location	text	接收端received的LSN位置
receiver_write_location	text	接收端write的LSN位置
receiver_flush_location	text	接收端flush的LSN位置
receiver_replay_location	text	接收端replay的LSN位置
sync_percent	text	同步百分比
channel	text	WALReceiver的信道信息

pg_stat_get_stream_replications()

描述：显示当前DN上所有的复制统计信息。

返回值类型：record

函数返回信息如下：

表 7-42 pg_stat_get_stream_replications()

名称	类型	描述
local_role	text	本地的角色
static_connections	integer	连接统计

 版权所有 © 华为云计算技术有限公司

名称	类型	描述
db_state	text	数据库状态
detail_information	text	详细信息

示例：

```
SELECT * FROM pg_stat_get_stream_replications();
 local_role | static_connections | db_state | detail_information
------------+--------------------+----------+--------------------
 Normal     |                  0 | Normal   | Normal
(1 row)
```

pg_stat_xlog_space()

描述：显示当前DN上Xlog空间使用信息。

返回值类型：record

函数返回信息如下：

表 7-43 pg_stat_xlog_space()

名称	类型	描述
xlog_files	bigint	pg_xlog目录下，去除backup、archive_status等子目录，所有识别为xlog文件的数目。
xlog_size	bigint	pg_xlog目录下，去除backup、archive_status等子目录，所有识别为xlog文件的文件大小之和，单位为MB。
other_size	bigint	pg_xlog目录下backup、archive_status等子目录文件的大小之和，单位为MB。

示例：

```
SELECT * FROM pg_stat_xlog_space();
 xlog_files | xlog_size | other_size
------------+-----------+------------
         79 |      1264 |          0
(1 row)
```

pgxc_stat_xlog_space()

描述：显示所有主DN上Xlog空间使用信息。

返回值类型：record

函数返回信息如下：

 版权所有 © 华为云计算技术有限公司

表 7-44 pgxc_stat_xlog_space()

名称	类型	描述
node_name	name	节点名称
xlog_files	bigint	pg_xlog目录下，去除backup、archive_status等子目录，所有识别为xlog文件的数目。
xlog_size	bigint	pg_xlog目录下，去除backup、archive_status等子目录，所有识别为xlog文件的文件大小之和，单位为MB。
other_size	bigint	pg_xlog目录下backup、archive_status等子目录文件的大小之和，单位为MB。

示例：

```
SELECT * FROM  pgxc_stat_xlog_space();
 node_name   | xlog_files | xlog_size | other_size
-------------+------------+-----------+------------
dn_6001_6002 |         73 |      1168 |          0
dn_6003_6004 |         73 |      1168 |          0
dn_6005_6006 |         73 |      1168 |          0
cn_5003      |         79 |      1264 |          0
cn_5001      |         72 |      1152 |          0
cn_5002      |         73 |      1168 |          0
(6 rows)
```

7.27.7 资源管理函数

资源管理模块相关函数介绍。

get_query_exec_info(queryid bigint)

描述：输入一个正在执行SQL的query_id，返回该SQL当前各计划节点的实时执行信息，包括算子名、已处理行数、预估行数、进度百分比。

该函数仅9.1.2及以上集群版本支持。

返回值类型：text

示例：

```
SELECT * FROM pg_catalog.get_query_exec_info(77405618595431677);
 NOTICE:  query duration: 33.295 s
  plan_node_id |              plan_node_name               |     A-time      | estimated_rows | tuple_processed |
progress
 --------------+-------------------------------------------+-----------------+----------------+-----------------+----------
             1 | ->  HashAggregate                         | 33.295          |       10000000 |               0 | 0.00%
             2 |    ->  Streaming (type: GATHER)           | 14.687          |       10000000 |         6451535 | 64.52%
             3 |       ->  HashAggregate                   | [5.849, 6.250] |       10000000 |         6851318 | 68.51%
             4 |          ->  Row Adapter                  | [0.557, 0.586] |       10000000 |        10000000 | 100.00%
             5 |             ->  CStore Scan on public.tb | [0.213, 0.234] |       10000000 |        10000000 | 100.00%
 (5 rows)
```

gs_switch_respool(query_id bigint, resource_pool_name name)

描述：作业切换资源池函数，9.1.0及以上集群版本支持。

 版权所有 © 华为云计算技术有限公司

入参要求resource_pool_name必须小于64字符。调用该函数后，如果作业处于排队阶段，作业切到新资源池会进行判断是否需要启动运行；如果作业处于运行阶段，作业切到新资源池可以启动原资源池的排队作业，同时受到新资源池的CPU管控限制。

该函数同时支持静态管控和动态管控模式下的资源池切换，也支持短查询切换资源池。

返回值类型：bool

⚠ 注意

- 该函数仅支持管理员或者管理员权限用户使用，禁止普通用户使用，该函数的使用前提是集群的CPU管控功能正常，即Cgroup功能正常。可使用pgxc_cgroup_reload_conf()函数查询cgroup是否正常，如果返回结果有非success内容，则需要进行修复。
- 该函数切换资源池后，会按照新资源池的所属异常规则进行管控。
- 当作业处于全局并发排队，即waiting in global queue时，切换资源池后并不会立即生效，此时还未进入资源池排队队列。
- 作业CPU管控切换，支持CN和DN，比如：资源池A占用40个核，资源池B占用5个核，将作业从资源池A切换到资源池B后，作业会在资源池B占用的5个核上执行，此时作业运行会更慢，但是隔离了该作业，不会再影响资源池A原有的作业运行。即实现语句的降级处理。
- 切换资源池时，用户可能会使用如下类似语句进行批量切换，在使用此类语句操作时，需要确保修改的是同一类状态的语句，如切换的都是排队作业或者切换的都是运行作业。如果切换不同类的语句，建议用户分两次进行，如果混合起来一起切换会导致语句运行并发数超限，作业异常唤醒。
  ```
  SELECT gs_switch_respool (queryid,'xxx') FROM xxx where xxx;
  ```
- 在切换语句资源池时，有以下约束：
  - 切换资源池A运行的作业到资源池B，资源池A的排队队首作业会被唤醒，如果资源池B并发作业数达到上限，作业A1不会直接运行。
  - 切换资源池A排队的作业A2到资源池B，如果资源池B并发作业数达到上限，作业A2会继续排队，如果未达到上限，则会被唤醒，直接运行。
  - 资源池切换时仅支持同车道切换，不支持不同车道切换，如资源池A的快车道切换后会到资源池B的快车道中，资源池A的慢车道切换后会到资源池B的慢车道中，无论资源池B是否开启了短查询加速功能，如资源池B未开启则作业会直接运行。
- 该函数仅支持在所有节点正常的情况下使用，如果节点故障，该函数会执行失败。

示例：

查询ID为74590868828368101的语句在执行过程中，管理员将其切换至poolg1资源池，切换成功：

```
CALL gs_switch_respool(74590868828368101,'poolg1');
 gs_switch_respool
-------------------
 t
(1 row)
```


版权所有 © 华为云计算技术有限公司

gs_increase_except_num(unique_sql_id int8)

描述：作业异常信息记录函数。入参要求必须大于0。调用该函数后会将作业异常次数+1，同时将作业最新异常时间更新为当前时间。

返回值类型：bool

示例：

```
SELECT gs_increase_except_num(111);
gs_increase_except_num
------------------------
t
(1 row)
```

gs_increase_except_num(unique_sql_id int8, except_num int4)

描述：作业异常信息记录函数。入参要求必须大于0。调用该函数后会将作业异常次数加except_num，同时将作业最新异常时间更新为当前时间。

返回值类型：bool

示例：

```
SELECT gs_increase_except_num(111, 4);
gs_increase_except_num
------------------------
t
(1 row)
```

gs_increase_except_num(unique_sql_id int8, except_num int4, except_time int8)

描述：作业异常信息记录函数。入参要求必须大于0。调用该函数后会将作业异常次数加except_num，同时将作业最新异常时间更新为except_time，except_time为时间戳，该函数主要用于内部调用。

返回值类型：bool

示例：

```
SELECT gs_increase_except_num(111, 4, 714623414421256);
gs_increase_except_num
------------------------
t
(1 row)
```

gs_append_blocklist(unique_sql_id int8)

描述：用于新增作业黑名单。即在系统表GS_BLOCKLIST_QUERY中新增一个黑名单，并更新GS_BLOCKLIST_QUERY中黑名单信息。

返回值类型：bool

示例：

```
SELECT gs_append_blocklist(111);
gs_append_blocklist
---------------------
t
(1 row)
```

 版权所有 © 华为云计算技术有限公司

gs_remove_blocklist(unique_sql_id int8)

描述：用于从作业黑名单中移除一个作业。即从系统表GS_BLOCKLIST_QUERY中移除一个黑名单，并更新GS_BLOCKLIST_QUERY中黑名单信息。

返回值类型：bool

示例：

```
SELECT gs_remove_blocklist(111);
gs_append_blocklist
--------------------
t
(1 row)
```

gs_increase_sql_except_num(sql_hash text)

描述：作业异常信息记录函数。调用该函数后会将作业异常次数+1，同时将作业最新异常时间更新为当前时间。该函数仅9.1.0.200及以上集群版本支持，主要用于内部调用。

返回值类型：bool

示例：

```
SELECT gs_increase_sql_except_num('sql_7bf6381d2a5a456a1936027198ad8b12');
gs_increase_sql_except_num
-------------------------
t
(1 row)
```

gs_increase_sql_except_num(sql_hash text, except_num int4)

描述：作业异常信息记录函数。调用该函数后会将作业异常次数加except_num，同时将作业最新异常时间更新为当前时间。该函数仅9.1.0.200及以上集群版本支持，主要用于内部调用。

返回值类型：bool

示例：

```
SELECT gs_increase_sql_except_num('sql_7bf6381d2a5a456a1936027198ad8b12', 4);
gs_increase_sql_except_num
-------------------------
t
(1 row)
```

gs_increase_sql_except_num(sql_hash text, except_num int4, except_time int8)

描述：作业异常信息记录函数。调用该函数后会将作业异常次数加except_num，同时将作业最新异常时间更新为except_time。其中，except_time为时间戳。该函数仅9.1.0.200及以上集群版本支持，主要用于内部调用。

返回值类型：bool

示例：

```
SELECT gs_increase_sql_except_num('sql_7bf6381d2a5a456a1936027198ad8b12', 4, 714623414421256);
gs_increase_sql_except_num
-------------------------
```

 版权所有 © 华为云计算技术有限公司

```
t
(1 row)
```

gs_append_blocklist(sql_hash text)

描述：用于新增作业黑名单。即在系统表GS_BLOCKLIST_SQL中新增一个黑名单，并更新GS_BLOCKLIST_SQL中黑名单信息。该函数仅9.1.0.200及以上集群版本支持。

返回值类型：bool

示例：

```
SELECT gs_append_blocklist('sql_7bf6381d2a5a456a1936027198ad8b12');
gs_append_blocklist
--------------------
t
(1 row)
```

gs_remove_blocklist(sql_hash text)

描述：用于从作业黑名单中移除一个作业。即从系统表GS_BLOCKLIST_SQL中移除一个黑名单，并更新GS_BLOCKLIST_SQL中黑名单信息。该函数仅9.1.0.200及以上集群版本支持。

返回值类型：bool

示例：

```
SELECT gs_remove_blocklist('sql_7bf6381d2a5a456a1936027198ad8b12');
gs_remove_blocklist
--------------------
t
(1 row)
```

gs_wlm_rebuild_except_rule_hash()

描述：用于重建当前节点的异常规则内存hash表，异常规则的生效阈值由监控线程实时从内存hash获取，hash异常时可以使用该函数进行rebuild恢复。

返回值类型：bool

示例：

```
SELECT gs_wlm_rebuild_except_rule_hash();
gs_wlm_rebuild_except_rule_hash
---------------------
t
(1 row)
```

gs_wlm_readjust_user_space(oid)

描述：用户永久存储空间校准函数。入参为用户oid，入参为0的情况下，会校准所有用户的永久存储空间。

返回值类型：text

示例：

```
SELECT gs_wlm_readjust_user_space(0);
gs_wlm_readjust_user_space
----------------------------
Exec Success
(1 row)
```

 版权所有 © 华为云计算技术有限公司

pgxc_wlm_readjust_schema_space()

描述：Schema永久存储空间校准函数。

返回值类型：text

示例：

```
SELECT pgxc_wlm_readjust_schema_space();
pgxc_wlm_readjust_schema_space
--------------------------------
Exec Success
(1 row)
```

pgxc_wlm_readjust_relfilenode_size_table()

描述：空间统计校准函数，不重建PG_RELFILENODE_SIZE系统表，重新校准用户和schema空间。

说明

由于校准函数执行时与其余业务之间的事务隔离性，校准函数对其他正在执行的业务不可见，导致校准函数会漏掉此类业务的空间变化，为避免校准之后出现空间误差，建议用户使用空间校准函数时，应选择空间大小稳定无变化的时候执行校准。

返回值类型：text

示例：

```
SELECT pgxc_wlm_readjust_schema_space();
pgxc_wlm_readjust_relfilenode_size_table
-----------------------------------------
Exec Success
(1 row)
```

pgxc_wlm_readjust_relfilenode_size_table(integer)

描述：空间统计校准函数。

说明

由于校准函数执行时与其余业务之间的事务隔离性，校准函数对其他的正在执行的业务不可见，导致校准函数会漏掉此类业务的空间变化，为避免校准之后出现空间误差，建议用户使用空间校准函数时，应选择空间大小稳定无变化的时候执行校准。

参数：取值范围为0~4，默认入参为0，不同的入参值代表不同的校准粒度。

- 入参为0：不重建PG_RELFILENODE_SIZE系统表，重新校准用户和schema空间。
- 入参为1：重建PG_RELFILENODE_SIZE系统表，并且重新校准用户和schema空间。
- 入参为2：重建PG_RELFILENODE_SIZE系统表。
- 入参为3：重新校准schema空间。
- 入参为4：重新校准用户空间。

返回值类型：text

示例：

```
SELECT * FROM pgxc_wlm_readjust_relfilenode_size_table(1);
  result
```

 版权所有 © 华为云计算技术有限公司

```
--------------
 Exec success
(1 row)
```

pgxc_wlm_get_schema_space(cstring)

描述：在CN上查询某个逻辑集群下各实例的Schema空间信息，入参为逻辑集群名称。

返回值类型：record

函数返回字段如下：

名称	类型	描述
schemaname	text	模式名
schemaid	oid	模式OID
databasename	text	数据库名
databaseid	oid	数据库OID
nodename	text	实例名称
nodegroup	text	节点组名称
usedspace	bigint	已使用的空间大小
permspace	bigint	空间上限值

示例：

```
SELECT * FROM pgxc_wlm_get_schema_space('group1');
    schemaname     | schemaid | databasename | databaseid |   nodename   |  nodegroup   | usedspace |
permspace
--------------------+----------+--------------+------------+--------------+--------------+-----------+-----------
 pg_catalog         |       11 | test1        |      16384 | datanode1    | installation |   9469952 |        -1
 public             |     2200 | postgres     |      15253 | datanode1    | installation |  25280512 |        -1
 pg_toast           |       99 | test1        |      16384 | datanode1    | installation |   1859584 |        -1
 cstore             |      100 | test1        |      16384 | datanode1    | installation |         0 |        -1
 data_redis         |    18106 | postgres     |      15253 | datanode1    | installation |    655360 |        -1
 data_redis         |    18116 | test1        |      16384 | datanode1    | installation |         0 |        -1
 public             |     2200 | test1        |      16384 | datanode1    | installation |     16384 |        -1
 dbms_om            |     3987 | postgres     |      15253 | datanode1    | installation |         0 |        -1
 dbms_job           |     3988 | postgres     |      15253 | datanode1    | installation |         0 |        -1
 dbms_om            |     3987 | test1        |      16384 | datanode1    | installation |         0 |        -1
 dbms_job           |     3988 | test1        |      16384 | datanode1    | installation |         0 |        -1
 sys                |    11693 | postgres     |      15253 | datanode1    | installation |         0 |        -1
 sys                |    11693 | test1        |      16384 | datanode1    | installation |         0 |        -1
 utl_file           |    14644 | postgres     |      15253 | datanode1    | installation |         0 |        -1
 utl_raw            |    14669 | postgres     |      15253 | datanode1    | installation |         0 |        -1
 dbms_sql           |    14674 | postgres     |      15253 | datanode1    | installation |         0 |        -1
 dbms_output        |    14662 | postgres     |      15253 | datanode1    | installation |         0 |        -1
 dbms_random        |    14666 | postgres     |      15253 | datanode1    | installation |         0 |        -1
 dbms_lob           |    14701 | postgres     |      15253 | datanode1    | installation |         0 |        -1
 information_schema |    14300 | postgres     |      15253 | datanode1    | installation |    294912 |        -1
 information_schema |    14300 | test1        |      16384 | datanode1    | installation |    294912 |        -1
 utl_file           |    14644 | test1        |      16384 | datanode1    | installation |         0 |        -1
 dbms_output        |    14662 | test1        |      16384 | datanode1    | installation |         0 |        -1
 dbms_random        |    14666 | test1        |      16384 | datanode1    | installation |         0 |        -1
 utl_raw            |    14669 | test1        |      16384 | datanode1    | installation |         0 |        -1
```


版权所有 © 华为云计算技术有限公司

```
dbms_sql          |    14674 | test1        |      16384 | datanode1     | installation |          0 |        -1
dbms_lob          |    14701 | test1        |      16384 | datanode1     | installation |          0 |        -1
pg_catalog        |       11 | postgres     |      15253 | datanode1     | installation |   13049856 |        -1
redisuser         |    16387 | postgres     |      15253 | datanode1     | installation |     630784 |        -1
pg_toast          |       99 | postgres     |      15253 | datanode1     | installation |    3080192 |        -1
cstore            |      100 | postgres     |      15253 | datanode1     | installation |    2408448 |        -1
pg_catalog        |       11 | test1        |      16384 | datanode2     | installation |    9469952 |        -1
public            |     2200 | postgres     |      15253 | datanode2     | installation |   25214976 |        -1
pg_toast          |       99 | test1        |      16384 | datanode2     | installation |    1859584 |        -1
cstore            |      100 | test1        |      16384 | datanode2     | installation |          0 |        -1
data_redis        |    18106 | postgres     |      15253 | datanode2     | installation |     655360 |        -1
data_redis        |    18116 | test1        |      16384 | datanode2     | installation |          0 |        -1
public            |     2200 | test1        |      16384 | datanode2     | installation |      16384 |        -1
dbms_om           |     3987 | postgres     |      15253 | datanode2     | installation |          0 |        -1
dbms_job          |     3988 | postgres     |      15253 | datanode2     | installation |          0 |        -1
dbms_om           |     3987 | test1        |      16384 | datanode2     | installation |          0 |        -1
dbms_job          |     3988 | test1        |      16384 | datanode2     | installation |          0 |        -1
```

pgxc_wlm_analyze_schema_space(cstring)

描述：在CN上查询某个逻辑集群下集群整体的Schema空间信息，入参为逻辑集群名称。

返回值类型：record

函数返回字段如下：

名称	类型	描述
schemaname	text	模式名。
databasename	text	数据库名。
nodegroup	text	节点组名称。
total_value	bigint	该模式的集群空间总值。
avg_value	bigint	该模式的各实例空间平均值。
skew_percent	integer	倾斜率。
extend_info	text	提供信息包括：单实例空间最大值、单实例空间最小值以及最大最小空间所在的实例名。

示例：

```
SELECT * FROM pgxc_wlm_analyze_schema_space('group1');
    schemaname     | databasename |  nodegroup   | total_value | avg_value | skew_percent |
extend_info
--------------------+--------------+--------------+-------------+-----------+--------------
+-----------------------------------------------
 pg_catalog         | test1        | installation |    56819712 |   9469952 |            0 | min:9469952
datanode1,max:9469952 datanode1
 public             | postgres     | installation |   150495232 |  25082538 |            0 | min:24903680
datanode6,max:25280512 datanode1
 pg_toast           | test1        | installation |    11157504 |   1859584 |            0 | min:1859584
datanode1,max:1859584 datanode1
 cstore             | test1        | installation |           0 |         0 |            0 | min:0 datanode1,max:0 datanode1
 data_redis         | postgres     | installation |     1966080 |    327680 |           50 | min:0
datanode4,max:655360 datanode1
 data_redis         | test1        | installation |           0 |         0 |            0 | min:0 datanode1,max:0 datanode1
```


版权所有 © 华为云计算技术有限公司

```
 public             | test1        | installation |       98304 |     16384 |           0 | min:16384
datanode1,max:16384 datanode1
 dbms_om            | postgres     | installation |           0 |         0 |           0 | min:0 datanode1,max:0
datanode1
 dbms_job           | postgres     | installation |           0 |         0 |           0 | min:0 datanode1,max:0
datanode1
 dbms_om            | test1        | installation |           0 |         0 |           0 | min:0 datanode1,max:0 datanode1
 dbms_job           | test1        | installation |           0 |         0 |           0 | min:0 datanode1,max:0 datanode1
 sys                | postgres     | installation |           0 |         0 |           0 | min:0 datanode1,max:0 datanode1
 sys                | test1        | installation |           0 |         0 |           0 | min:0 datanode1,max:0 datanode1
 utl_file           | postgres     | installation |           0 |         0 |           0 | min:0 datanode1,max:0 datanode1
 utl_raw            | postgres     | installation |           0 |         0 |           0 | min:0 datanode1,max:0 datanode1
 dbms_sql           | postgres     | installation |           0 |         0 |           0 | min:0 datanode1,max:0 datanode1
 dbms_output        | postgres     | installation |           0 |         0 |           0 | min:0 datanode1,max:0
datanode1
 dbms_random        | postgres     | installation |           0 |         0 |           0 | min:0 datanode1,max:0
datanode1
 dbms_lob           | postgres     | installation |           0 |         0 |           0 | min:0 datanode1,max:0
datanode1
 information_schema | postgres     | installation |     1769472 |    294912 |           0 | min:294912
datanode1,max:294912 datanode1
 information_schema | test1        | installation |     1769472 |    294912 |           0 | min:294912
datanode1,max:294912 datanode1
 utl_file           | test1        | installation |           0 |         0 |           0 | min:0 datanode1,max:0 datanode1
 dbms_output        | test1        | installation |           0 |         0 |           0 | min:0 datanode1,max:0
datanode1
 dbms_random        | test1        | installation |           0 |         0 |           0 | min:0 datanode1,max:0
datanode1
 utl_raw            | test1        | installation |           0 |         0 |           0 | min:0 datanode1,max:0 datanode1
 dbms_sql           | test1        | installation |           0 |         0 |           0 | min:0 datanode1,max:0 datanode1
 dbms_lob           | test1        | installation |           0 |         0 |           0 | min:0 datanode1,max:0 datanode1
 pg_catalog         | postgres     | installation |    75431936 |  12571989 |           3 | min:12124160
datanode4,max:13049856 datanode1
 redisuser          | postgres     | installation |     1884160 |    314026 |          50 | min:16384
datanode4,max:630784 datanode1
 pg_toast           | postgres     | installation |    17154048 |   2859008 |           7 | min:2637824
datanode4,max:3080192 datanode1
 cstore             | postgres     | installation |    15294464 |   2549077 |           5 | min:2408448
datanode1,max:2703360 datanode6
(31 rows)
```

gs_wlm_set_queryband_action(cstring,cstring,int4)

描述：设置query_band关联行为和次序。

返回值类型：boolean

函数入参字段如下：

名称	类型	描述
qband	cstring	query_band键值对，长度上限为63个字符。
action	cstring	query_band关联行为。
order	int4	query_band搜索次序，缺省参数，默认为-1。

示例：

```
SELECT * FROM gs_wlm_set_queryband_action('a=1','respool=p1');
 gs_wlm_set_queryband_action
-----------------------------
 t
(1 row)
```


版权所有 © 华为云计算技术有限公司

```
SELECT * FROM gs_wlm_set_queryband_action('a=3','respool=p1;priority=rush',1);
 gs_wlm_set_queryband_action
-----------------------------
 t
(1 row)
```

gs_wlm_set_queryband_order(cstring,int4)

描述：设置query_band次序。

返回值类型：boolean

函数入参字段如下：

名称	类型	描述
qband	cstring	query_band键值对。
order	int4	query_band搜索次序，缺省参数，默认为-1。

示例：

```
SELECT * FROM gs_wlm_set_queryband_order('a=1',2);
 gs_wlm_set_queryband_action
-----------------------------
 t
(1 row)
```

gs_wlm_get_queryband_action(cstring)

描述：查询query_band关联行为和次序。

返回值类型：record

函数返回字段如下：

名称	类型	描述
qband	cstring	query_band键值对。
respool_id	Oid	query band关联资源池OID。
respool	text	query band关联资源池名。
priority	text	query band关联队列内优先级。
qborder	int4	query_band搜索次序。

示例：

```
SELECT * FROM gs_wlm_get_queryband_action('a=1');
qband | respool_id | respool | priority | qborder
-------+------------+---------+----------+---------
 a=1   |      16388 | p1      | Medium   |      -1
(1 row)
```

 版权所有 © 华为云计算技术有限公司

gs_cgroup_reload_conf()

描述：在当前实例上进行cgroup配置文件在线加载。

返回值类型：record

函数返回字段如下：

名称	类型	描述
node_name	text	实例名称。
node_host	text	实例所在节点的IP地址。
result	text	cgroup在线加载是否成功。

示例：

```
SELECT * FROM gs_cgroup_reload_conf();
 node_name |   node_host    | result
-----------+----------------+---------
 cn_5001   | 192.168.178.35 | success
```

pgxc_cgroup_reload_conf()

描述：在系统所有实例上进行cgroup配置文件在线加载。

返回值类型：record

函数返回字段如下：

名称	类型	描述
node_name	text	实例名称。
node_host	text	实例所在节点的IP地址。
result	text	cgroup在线加载是否成功。

示例：

```
SELECT * FROM pgxc_cgroup_reload_conf();
  node_name   |    node_host    | result
--------------+-----------------+---------
 dn_6025_6026 | 192.168.178.177 | success
 dn_6049_6050 | 192.168.179.79  | success
 dn_6051_6052 | 192.168.179.79  | success
 dn_6055_6056 | 192.168.179.79  | success
 dn_6067_6068 | 192.168.181.57  | success
 dn_6023_6024 | 192.168.178.39  | success
 dn_6009_6010 | 192.168.181.21  | success
 dn_6011_6012 | 192.168.181.21  | success
 dn_6015_6016 | 192.168.181.21  | success
 dn_6029_6030 | 192.168.178.177 | success
 dn_6031_6032 | 192.168.178.177 | success
 dn_6045_6046 | 192.168.179.45  | success
 cn_5001      | 192.168.178.35  | success
 cn_5003      | 192.168.178.39  | success
```


版权所有 © 华为云计算技术有限公司

```
 dn_6061_6062 | 192.168.181.179 | success
 cn_5006      | 192.168.179.45  | success
 cn_5004      | 192.168.178.177 | success
 cn_5002      | 192.168.181.21  | success
 cn_5005      | 192.168.178.187 | success
 dn_6019_6020 | 192.168.178.39  | success
 dn_6007_6008 | 192.168.178.35  | success
 dn_6071_6072 | 192.168.181.57  | success
 dn_6003_6004 | 192.168.178.35  | success
 dn_6013_6014 | 192.168.181.21  | success
 dn_6035_6036 | 192.168.178.187 | success
 dn_6037_6038 | 192.168.178.187 | success
 dn_6001_6002 | 192.168.178.35  | success
 dn_6063_6064 | 192.168.181.179 | success
 dn_6005_6006 | 192.168.178.35  | success
 dn_6057_6058 | 192.168.181.179 | success
 dn_6069_6070 | 192.168.181.57  | success
 dn_6027_6028 | 192.168.178.177 | success
 dn_6059_6060 | 192.168.181.179 | success
 dn_6041_6042 | 192.168.179.45  | success
 dn_6043_6044 | 192.168.179.45  | success
 dn_6047_6048 | 192.168.179.45  | success
 dn_6033_6034 | 192.168.178.187 | success
 dn_6065_6066 | 192.168.181.57  | success
 dn_6021_6022 | 192.168.178.39  | success
 dn_6017_6018 | 192.168.178.39  | success
 dn_6039_6040 | 192.168.178.187 | success
 dn_6053_6054 | 192.168.179.79  | success
(42 rows)
```

pgxc_cgroup_reload_conf(text)

描述：在某个节点上进行cgroup配置文件在线加载。入参为节点的IP地址。

返回值类型：record

函数返回字段如下：

名称	类型	描述
node_name	text	实例名称。
node_host	text	实例所在节点的IP地址。
result	text	cgroup在线加载是否成功。

示例：

```
SELECT * FROM pgxc_cgroup_reload_conf('192.168.178.35');
  node_name   |   node_host    | result
--------------+----------------+---------
 cn_5001      | 192.168.178.35 | success
 dn_6007_6008 | 192.168.178.35 | success
 dn_6003_6004 | 192.168.178.35 | success
 dn_6001_6002 | 192.168.178.35 | success
 dn_6005_6006 | 192.168.178.35 | success
(5 rows)
```

gs_wlm_node_recover(boolean isForce)

描述：动态资源管理模式下，对CCN管控计数和作业信息进行更新恢复。该函数仅支持管理员执行，通常用于CN实例故障重启后的实例恢复，由集群管理组件（CM）调用，不建议用户直接调用。具体功能如下：

 版权所有 © 华为云计算技术有限公司

- CN执行：通知CCN清理该CN执行的作业信息和作业对应的管控计数信息。
- CCN执行：重置管控计数信息，并从CN上获取最新的慢车道作业信息。

返回值类型：bool

gs_wlm_node_clean(cstring nodename)

描述：动态资源管理模式下，对CCN上入参指定的CN执行的作业信息和对应的管控计数信息进行清理。该函数仅支持管理员执行，通常用于CN实例故障重启后的实例恢复，由集群管理组件（CM）调用，正常情况下不建议用户直接调用。

返回值类型：bool

pg_stat_get_wlm_node_resource_info(int4)

描述：显示所有DN资源的汇总信息。

返回值类型：record

函数返回字段如下：

名称	类型	描述
min_mem_util	integer	DN最小内存使用率。
max_mem_util	integer	DN最大内存使用率。
min_cpu_util	integer	DN最小CPU使用率。
max_cpu_util	integer	DN最大CPU使用率。
min_io_util	integer	DN最小IO使用率。
max_io_util	integer	DN最大IO使用率。
phy_usemem_rate	integer	物理节点最大内存使用率。

pg_stat_get_workload_struct_info()

描述：定位CCN排队问题的负载管理函数。此函数为内部函数，如需使用请联系技术支持工程师。

返回值类型：record

pgxc_query_resource_info(query_id bigint)

描述：显示指定query_id对应的语句在所有DN上的资源监控信息。该函数仅9.1.0.100及以上集群版本支持。

返回值类型：record

函数返回字段如下：

 版权所有 © 华为云计算技术有限公司

名称	类型	描述
node_name	text	实例名称，只包含DN。
user_id	oid	用户ID。
queryid	bigint	语句执行使用的内部query id。
used_mem	int	语句在当前DN使用的内存大小，单位MB。
cpu_time	bigint	语句在当前DN上的CPU时间，单位ms。
used_cpu	double	语句在当前DN使用的CPU个数。
spill_size	bigint	若发生下盘，语句在当前DN下盘数据量，默认为0。单位MB。
read_bytes	bigint	语句在当前DN使用的逻辑读字节数，单位KB。
write_bytes	bigint	语句在当前DN使用的逻辑写字节数，单位KB。
read_count	bigint	语句在当前DN使用的逻辑读次数。
write_count	bigint	语句在当前DN使用的逻辑写次数。
read_speed	int	语句在当前DN使用的逻辑读速率，单位KB/s。
write_speed	int	语句在当前DN使用的逻辑写速率，单位KB/s。
curr_iops	int	语句在当前DN上的每秒IO数值（列存单位是次/s，行存单位是万次/s）。
send_pkg	bigint	语句在当前DN上的通信包发送总量，单位packages。
recv_pkg	bigint	语句在当前DN上的通信包接收总量，单位packages
send_bytes	bigint	语句在当前DN上的通信流发送数据总量，单位Byte。
recv_bytes	bigint	语句在当前DN上的通信流接收数据总量，单位Byte。
send_speed	int	语句在当前DN使用的网络发送速率，单位KB/s。
recv_speed	int	语句在当前DN使用的网络接收速率，单位KB/s。

7.27.8 内存管理函数

内存管理函数仅9.1.0及以上集群版本支持。

pg_shared_chunk_detail(contextname char(64))

描述：查询指定共享内存下内存上下文申请的所有chunk信息。

参数contextname，表示内存上下文名称。

 版权所有 © 华为云计算技术有限公司

使用该函数需先使用pg_shared_chunk_dump(contextname char(64))函数将相关信息打印成文件。

返回值类型：record

示例：

```
SELECT * FROM pg_shared_chunk_detail('pgstat file hash table');
       parent        | level |  file_name   | line_number | chunk_size | requested_number | total_size
---------------------+-------+--------------+-------------+------------+------------------+------------
 pgstat file context |     2 | dynahash.cpp |         158 |       2048 |                1 |       2048
 pgstat file context |     2 | dynahash.cpp |         158 |        256 |                1 |        256
 pgstat file context |     2 | dynahash.cpp |         158 |       4096 |                9 |      36864
 pgstat file context |     2 | dynahash.cpp |         158 |       8192 |                4 |      32768
(4 rows)
```

pv_session_chunk_detail(tid int, contextname char(64))

描述：查询指定线程创建的某个内存上下文申请的所有chunk信息。

参数tid，表示线程ID；参数contextname，表示内存上下文名称。

使用该函数需使用pv_session_chunk_dump(tid int, contextname char(64))函数将相关信息打印成文件。

返回值类型：record

示例：

```
SELECT * FROM pv_session_chunk_detail(140178810990936, 'Timezones');
      parent       | level |   file_name   | line_number | chunk_size | requested_number | total_size
-------------------+-------+---------------+-------------+------------+------------------+------------
 TopMemoryContext |     1 | dynahash.cpp  |         158 |       1280 |                2 |       2560
 TopMemoryContext |     1 | dynahash.cpp  |         158 |        160 |                1 |        160
 TopMemoryContext |     1 | dynahash.cpp  |         158 |       2560 |                2 |       5120
 TopMemoryContext |     1 | localtime.cpp |        1965 |        128 |                1 |        128
 TopMemoryContext |     1 | localtime.cpp |        1965 |        448 |                1 |        448
(5 rows)
```

pg_shared_chunk_dump(contextname char(64))

描述：将指定共享内存下内存上下文申请的所有chunk信息打印成文件。

参数contextname，表示内存上下文名称。

返回值类型：boolean

示例：

```
SELECT * FROM pg_shared_chunk_dump('pgstat file hash table');
 pg_shared_chunk_dump
----------------------
 t
(1 row)
```

pv_session_chunk_dump(tid int, contextname char(64))

描述：将指定线程创建的某个内存上下文申请的所有chunk信息打印成文件。

参数tid，表示线程ID；参数contextname，表示内存上下文名称。

返回值类型：boolean

示例：

 版权所有 © 华为云计算技术有限公司

```
SELECT * FROM pv_session_chunk_dump(140472797325280, 'Timezones');
 pv_session_chunk_dump
-----------------------
 t
(1 row)
```

gs_print_memory_profiling()

描述：打印内存采集详情。该函数仅9.1.1.100及以上集群版本支持。

返回值类型：boolean

示例：

```
SELECT * FROM gs_print_memory_profiling();
 gs_print_memory_profiling
---------------------------
 t
(1 row)
```

gs_gsc_dbstat_info(database_id bigint DEFAULT NULL::bigint)

描述：查询当前节点全局缓存内存（Global SysCache）的统计信息，例如SysCache缓存命中情况及LRU淘汰情况。该函数仅9.1.1.100及以上集群版本支持。

参数：database_id，表示数据库OID。

- 当database_id为空时，查询所有数据库的统计信息。
- 当database_id为0时，查询共享数据库的统计信息。
- 当database_id为有效数据库OID时，查询共享数据库和该数据库的统计信息。

返回值类型：record

函数返回字段如下：

名称	类型	描述
database_id	BIGINT	数据库OID。
database_name	TEXT	数据库名称。
tup_searches	BIGINT	系统表缓存查找次数。
tup_hits	BIGINT	系统表缓存命中次数。
tup_miss	BIGINT	系统表缓存未命中触发扫描表的次数。
tup_swapout_count	BIGINT	被LRU淘汰的元组的数量。
tup_count	BIGINT	缓存中元组的数量。
tup_dead	BIGINT	垃圾队列中元组的数量。
tup_memory	BIGINT	缓存中元组的总内存。
total_memory	BIGINT	Global SysCache的总内存。
refcount	BIGINT	Global SysCache被线程引用的次数。

 版权所有 © 华为云计算技术有限公司

示例：

```
SELECT * FROM gs_gsc_dbstat_info();
 database_id | database_name | tup_searches | tup_hits | tup_miss | tup_swapout_count | tup_count |
tup_dead | tup_memory | total_memory | refcount
-------------+---------------+--------------+----------+----------+-------------------+-----------+----------+------------
+--------------+----------
           0 |               |        12474 |     9876 |       14 |                 0 |        16 |        0 |      12560 |       697664
|        0
       17185 | postgres      |        10156 |     5521 |     2223 |                 0 |      1398 |        0 |     752296 |
4417856 |       11
(2 rows)
```

gs_gsc_clean(database_id bigint DEFAULT NULL::bigint)

描述：清理指定数据库的全局缓存内存（Global SysCache）。该函数仅9.1.1.100及以上集群版本支持。

参数：database_id，表示数据库OID。

- 当database_id为空时，清理所有数据库的内存。
- 当database_id为0时，清理共享数据库的内存。
- 当database_id为有效数据库OID时，清理共享数据库和指定数据库的内存。

返回值类型：bool

示例：

```
SELECT * FROM gs_gsc_clean();
 gs_gsc_clean
--------------
 t
(1 row)
```

gs_table_oscache(regclass)

描述：查询表的数据文件在操作系统文件缓存中的缓存情况。该函数仅9.1.1.200及以上集群版本支持。

⚠ 注意

该函数涉及mmap锁操作，不建议频繁使用。

返回值类型：record

函数返回字段如下：

名称	类型	描述
schema_name	TEXT	模式名称。
table_name	TEXT	表名。
file_name	TEXT	数据文件名称。
file_type	TEXT	文件类型。
seg_count	INTEGER	文件段号。

版权所有 © 华为云计算技术有限公司

名称	类型	描述
file_size	BIGINT	文件大小。
cache_size	BIGINT	该文件被缓存的大小。
con_cache_count	BIGINT	连续缓存页的数量。
cache_dirty_size	BIGINT	缓存中脏页的大小。
con_cache_dirty_count	BIGINT	连续缓存脏页的数量。
in_fd_cache	BOOLEAN	文件是否在fd缓存中，即文件是否处于打开状态。

示例：

```
SELECT * FROM gs_table_oscache('employee');
 schema_name | table_name |     file_name      | file_type | seg_count | file_size | cache_size | con_cache_count
| cache_dirty_size | con_cache_dirty_count | in_fd_cache
-------------+------------+--------------------+-----------+-----------+-----------+------------+-----------------
+------------------+-----------------------+-------------
 public      | employee   | base/17253/17103 | RowRel    |         1 |         0 |          0 |                0
|                0 | f
(1 row)
```

gs_table_freecache(regclass)

描述：清理指定表的数据文件的操作系统文件缓存。该函数无法清理仍在打开状态的数据文件。该函数仅9.1.1.200及以上集群版本支持。

返回值类型：bool

示例：

```
SELECT * FROM gs_table_freecache('employee');
 gs_table_freecache
--------------------
 t
(1 row)
```

gs_file_freecache(IN file_path TEXT)

描述：清理指定文件的操作系统文件缓存。该函数可以清理仍在打开状态的文件。该函数仅9.1.1.200及以上集群版本支持。

参数：file_path，表示文件路径。

返回值类型：bool

示例：

```
SELECT * FROM gs_file_freecache('base/17253/17103');
 gs_file_freecache
-------------------
 t
(1 row)
```

 版权所有 © 华为云计算技术有限公司

7.27.9 缓存管理函数

9.1.1.100集群版本开始DWS支持缓存管理函数，缓存管理函数将缓存从共享缓冲区（shared buffer）中释放，可有效加速查询，优化数据库性能。

只有系统管理员可以执行缓存管理函数。

缓存管理函数提供了三种粒度的接口用于释放缓存：

- 释放指定缓冲页，对应函数**pg_buffercache_evict(bufferid integer)**。
- 释放指定表的所有缓冲页，对应函数**pg_buffercache_evict(nsp_name name, rel_name name, partition_name name, temp_nsp_name name)**、**pgxc_buffercache_evict(nsp_name name, rel_name name, partition_name name, temp_nsp_name name)**。
- 释放所有缓冲页，对应函数**pg_buffercache_evict()**和**pgxc_buffercache_evict()**。

  说明

  被pin的缓冲页不会被缓存管理函数释放。

pg_buffercache_evict(bufferid integer)

描述：释放指定ID的缓冲页。

返回值类型：bool

返回值说明：缓冲页是否被释放并成功写入磁盘。

示例：

```
SELECT pg_catalog.pg_buffercache_evict(1);
 pg_buffercache_evict
----------------------
 t
(1 row)
```

pg_buffercache_evict(nsp_name name, rel_name name, partition_name name, temp_nsp_name name)

描述：在当前实例上，释放指定表的所有未被pin住的缓冲页。

参数说明：根据不同的表类型，需要传入不同的参数值。具体参见**表7-33**。

返回值类型：integer

返回值说明：被成功释放的缓冲页数。

示例：

```
SELECT pg_catalog.pg_buffercache_evict('public', 't1', NULL, NULL);
 pg_buffercache_evict
----------------------
                    4
(1 row)
```

pgxc_buffercache_evict(nsp_name name, rel_name name, partition_name name, temp_nsp_name name)

描述：在所有实例上，释放指定表的所有未被pin住的缓冲页。

版权所有 © 华为云计算技术有限公司

参数说明：根据不同的表类型，需要传入不同的参数值。具体参见表7-33。

返回值类型：integer

返回值说明：被成功释放的缓冲页数。

示例：

```
SELECT pg_catalog.pgxc_buffercache_evict('public', 't1', NULL, NULL);
 pgxc_buffercache_evict
------------------------
                     10
(1 row)
```

pg_buffercache_evict()

描述：在当前实例上，释放所有未被pin住的缓冲页。

返回值类型：integer

返回值说明：被成功释放的缓冲页数。

示例：

```
SELECT pg_catalog.pg_buffercache_evict();
 pg_buffercache_evict
----------------------
                 2654
(1 row)
```

pgxc_buffercache_evict()

描述：在所有实例上，释放所有未被pin住的缓冲页。

返回值类型：integer

返回值说明：被成功释放的缓冲页数。

示例：

```
SELECT pg_catalog.pgxc_buffercache_evict();
 pgxc_buffercache_evict
------------------------
                  10881
(1 row)
```

7.27.10 plan management 管理函数

plan management模块相关函数介绍。

pgxc_bind_plan(sql_hash text, outline_name text, apply_skew_hint boolean DEFAULT false, status text DEFAULT 'approved'::text)

描述：为sql_hash对应的语句绑定outline_name对应的outline，即把SQL_OUTLINE系统表中主键outline_name和sql_hash对应的outline绑定给sql_hash对应的语句。

- apply_skew_hint，表示是否应用outline中的skew hint，默认为false，即不应用outline中的skew hint。
- status，表示绑定状态。（ 该参数仅9.1.1.200及以上集群版本支持 ）

 版权所有 © 华为云计算技术有限公司

status有四种取值：preferred（优先的）、approved（可用的）、unapproved（不推荐的）和rejected（拒绝的），计划选择时会依据status的优先级进行选择。

该函数需要数据库管理员用户，或授予了gs_role_sql_management角色权限的用户才能执行。

返回值类型：boolean

示例：

```
SELECT pgxc_bind_plan(sql_hash, outline_name) FROM outlines limit 1;
 pgxc_bind_plan
----------------
 t
(1 row)

SELECT pgxc_bind_plan('sql_be2995e824abb8b712b69fca4825b532',
'outline_e5214c8e3901da4d7d34166cb9cb84ee');
 pgxc_bind_plan
----------------
 t
(1 row)

SELECT pgxc_bind_plan(sql_hash, outline_name, true) FROM outlines WHERE outline ILIKE '%skew%' limit 1;
 pgxc_bind_plan
----------------
 t
(1 row)

SELECT pgxc_bind_plan('sql_be2995e824abb8b712b69fca4825b532',
'outline_e5214c8e3901da4d7d34166cb9cb84ee', true, 'preferred');
 pgxc_bind_plan
----------------
 t
(1 row)
```

pgxc_unbind_plan(sql_hash text, outline_name text DEFAULT NULL::text)

描述：为sql_hash对应的语句解除绑定的outline。

outline_name参数仅9.1.1.200及以上集群版本支持。

- 不指定outline_name参数时，解除该sql_hash对应语句所有绑定的outline。
- 指定outline_name参数时，仅解除该sql_hash对应语句与指定outline_name绑定的outline。

需要数据库管理员用户或授予了gs_role_sql_management角色权限的用户才能执行。

返回值类型：boolean

示例：

```
SELECT pgxc_unbind_plan(sql_hash) FROM sql_plan_baseline;
 pgxc_unbind_plan
------------------
 t
(1 row)

SELECT pgxc_unbind_plan('sql_be2995e824abb8b712b69fca4825b532');
 pgxc_unbind_plan
------------------
 t
(1 row)
```

 版权所有 © 华为云计算技术有限公司

```
SELECT pgxc_unbind_plan('sql_be2995e824abb8b712b69fca4825b532',
'outline_e5214c8e3901da4d7d34166cb9cb84ee');
 pgxc_unbind_plan
------------------
 t
(1 row)
```

pgxc_clean_plan(clean_period integer)

描述：手动执行一次计划清理任务，清理超出指定时间范围（clean_period，单位为天）的如下计划。该函数仅9.1.1.200及以上集群版本支持。

- pg_plan_baseline中超过clean_period指定时间范围，且status为“unapproved”的baseline。
- dbms_om.sql_outline中超过clean_period指定时间范围，且未绑定outline（即不在pg_plan_baseline中绑定的outline）。

需要数据库管理员用户，或授予了gs_role_sql_management角色权限的用户才能执行。

返回值类型：boolean

说明

计划清理由plan management后台线程执行。由于计划清理需等待排在其前面的其他任务（例如，要保存outline等）执行完成后才会开始执行计划清理，因此该清理操作可能会存在一定延时。

示例：

清理pg_plan_baseline中last_execute_time为两天前，且状态为“unapproved”的绑定outline；清理dbms_om.sql_outline中create_time为两天前，且未绑定的outline。

```
SELECT pgxc_clean_plan(2);
 pgxc_clean_plan
------------------
 t
(1 row)
```

pgxc_create_auto_clean_plan_task(clean_period integer, clean_time timestamp without time zone, clean_interval interval)

描述：创建自动计划清理任务。该任务会根据设定的规则定期清理不在clean_period指定时间内的pg_plan_baseline和dbms_om.sql_outline数据，首次执行时间（clean_time）由用户指定，后续任务按固定时间间隔（clean_interval）进行清理执行，同**pgxc_clean_plan(clean_period integer)**。该函数仅9.1.1.200及以上集群版本支持。

需数据库管理员用户，或授予了gs_role_sql_management角色权限的用户才能执行。

参数说明：

- clean_period，任务的清理时间范围，单位为天。
- clean_time，任务的首次执行时间。
- clean_interval，表示两次任务执行之间的时间间隔。

返回值类型：boolean

 版权所有 © 华为云计算技术有限公司

说明

- 同pgxc_clean_plan()函数，自动计划清理操作可能存在一定延时。
- 自动计划清理任务只负责对执行创建自动清理任务语句的数据库进行清理。如需对多个数据库中的计划进行清理，需分别连接到需要执行计划清理的数据库，并创建自动计划清理任务。
- 如果数据库中已存在一个自动计划清理任务，再次创建自动计划清理任务将报错。如需修改自动计划清理任务的执行时间、执行间隔或清理范围，可调用**pgxc_alter_auto_clean_plan_task()**修改对应的参数配置。

示例：

创建自动计划清理任务，第一次执行时间为2025-11-1 00:00:00，执行间隔为1天，清理范围为32天前的数据。

```
SELECT pgxc_create_auto_clean_plan_task(32, '2025-11-1 00:00:00', '1 day');
 pgxc_create_auto_clean_plan_task
----------------------------------
 t
(1 row)
```

pgxc_alter_auto_clean_plan_task(clean_period integer, clean_time timestamp without time zone, clean_interval interval)

描述：修改自动计划清理任务的参数配置。需修改的参数配置输入新的参数值，无需修改的参数配置输入NULL以保持原设置。需数据库管理员用户或授予了gs_role_sql_management角色权限的用户才能执行。

该函数仅9.1.1.200及以上集群版本支持。

返回值类型：boolean

示例：

对自动计划清理任务的参数配置不进行任何修改。

```
SELECT pgxc_alter_auto_clean_plan_task(NULL, NULL, NULL);
WARNING:  Nothing need to change
CONTEXT:  referenced column: pgxc_alter_auto_clean_plan_task
 pgxc_alter_auto_clean_plan_task
---------------------------------
 t
(1 row)
```

修改自动计划清理任务的下次执行时间为2025-12-01，执行间隔为1天零1秒。

```
SELECT pgxc_alter_auto_clean_plan_task(NULL, '2025-12-01', '1 day 1 second');
 pgxc_alter_auto_clean_plan_task
---------------------------------
 t
(1 row)
```

pgxc_drop_auto_clean_plan_task()

描述：删除不再使用的自动计划清理任务。该函数仅9.1.1.200及以上集群版本支持。

需数据库管理员用户或授予了gs_role_sql_management角色权限的用户才能执行。

返回值类型：boolean

示例：

 版权所有 © 华为云计算技术有限公司

```
SELECT pgxc_drop_auto_clean_plan_task();
 pgxc_drop_auto_clean_plan_task
--------------------------------
 t
(1 row)
```

drop_outline(sql_hash text, outline_name text)

描述：从dbms_om.sql_outline系统表中删除指定的outline。只有未被绑定outline才可以删除，且只能由数据库管理员用户或授予了gs_role_sql_management角色权限的用户才可以删除。

返回值类型：boolean

示例：

```
SELECT drop_outline(sql_hash, outline_name) FROM outlines;
WARNING:  The outline(specified by sql_hash[sql_3d937542c0d24c9c995f31cba327fa56] and
outline_name[outline_413ac0e566ff20be320f5ef0774bf1eb]) is bind
CONTEXT:  referenced column: drop_outline
 drop_outline
--------------
 t
 t
 t
 t
 t
 t
 t
 t
 t
 t
 t
 f
 t
 t
 t
 t
 t
 t
 t
 t
 t
 t
 t
(23 rows)
```

pgxc_reset_planmgmt_hashtable()

描述：释放所有CN上plan management用于缓存outline所占用的内存。只能由数据库管理员用户或授予了gs_role_sql_management角色权限的用户才可以执行。

返回值类型：boolean

示例：

查询当前各CN上plan management保存outline占用的内存情况。

```
SELECT pg_catalog.pgxc_node_str(), contextname, level, parent, pg_size_pretty(totalsize) totalsize,
pg_size_pretty(freesize) freesize, pg_size_pretty(usedsize) usedsize FROM PG_SHARED_MEMORY_DETAIL
WHERE contextname ilike 'plan%' ORDER BY contextname;
 pgxc_node_str |        contextname         | level |      parent      | totalsize | freesize |  usedsize
---------------+----------------------------+-------+------------------+-----------+----------+------------
 coordinator1  | plan management hash table |     2 | PlanmgmtContext | 80 KB     | 45 KB    | 35 KB
 coordinator1  | PlanmgmtContext            |     1 | ProcessMemory   | 96 KB     | 94 KB    | 2432 bytes
(2 rows)
```


版权所有 © 华为云计算技术有限公司

```
SELECT pgxc_reset_planmgmt_hashtable();
 pgxc_reset_planmgmt_hashtable
-------------------------------
 t
(1 row)
```

执行pgxc_reset_planmgmt_hashtable()函数后，查询当前各CN上plan management保存outline占用的内存情况。

```
SELECT pg_catalog.pgxc_node_str(), contextname, level, parent, pg_size_pretty(totalsize) totalsize,
pg_size_pretty(freesize) freesize, pg_size_pretty(usedsize) usedsize FROM PG_SHARED_MEMORY_DETAIL
WHERE contextname ilike 'plan%' ORDER BY contextname;
 pgxc_node_str |         contextname          | level |      parent      | totalsize | freesize | usedsize
---------------+------------------------------+-------+------------------+-----------+----------+----------
 coordinator1  | plan management hash table |     2 | PlanmgmtContext | 56 KB     | 33 KB    | 23 KB
 coordinator1  | PlanmgmtContext              |     1 | ProcessMemory   | 0 bytes   | 0 bytes  | 0 bytes
(2 rows)
```

7.27.11 其他函数

pgxc_pool_check()

描述：检查连接池中缓存的连接数据是否与pgxc_node一致。8.3.0及以上集群版本中该参数已废弃。

返回值类型：boolean

pgxc_pool_reload()

描述：更新连接池中缓存的连接信息。

返回值类型：boolean

pg_pool_validate(clear boolean, co_node_name cstring)

描述：清理CN上无效的后台线程（这些后台线程持有无效的pooler连接，这里无效的pooler连接指的是连接到当前DN备实例的连接）。

返回值类型：record

pg_nodes_memory()

描述：查看所有节点的内存占用。

返回值类型：record

table_skewness(text)

描述：查看表数据在所有节点的占比。

参数：表示待查询表的表名，为text类型。

返回值类型：record

table_skewness(table_name text, column_name text[, row_num text])

描述：查看表里某列数据，按hash分布规则，在各节点的占比情况。结果以数据节点上的数据量排序。

 版权所有 © 华为云计算技术有限公司

参数：table_name为表名；column_name为列名；row_num表示查看当前列所有数据，可缺省，默认为0。非0时，表示抽取指定条数的数据查看占比情况（每次采样结果可能不相同，只保证采集row_num条数据进行占比计算）。

返回值类型：record

示例：

根据tx表中的a列，按hash分布，则在1，2，0编号的DN上分布的数据量分别为7条，2条，1条。

```
SELECT * FROM table_skewness('tx','a');
 seqnum | num |  ratio
--------+-----+----------
 1      | 7   | 70.000%
 2      | 2   | 20.000%
 0      | 1   | 10.000%
(3 rows)
```

table_data_skewness(data_row record, locatorType "char")

描述：计算指定表中列拼接出的record，对应的桶分布索引。

参数：data_row表示指定表中列拼接出的record，locatorType表示分布规则，当前建议指定'H'，按hash分布计算。

返回值类型：smallint

示例：

计算tx表中a列拼接的record，按照hash分布规则对应的桶分布索引。

```
select a, table_data_skewness(row(a), 'H') from tx;
 a | table_data_skewness
---+---------------------
 3 |                   0
 6 |                   2
 7 |                   2
 4 |                   1
 5 |                   1
(5 rows)
```

table_distribution(schemaname text, tablename text)

描述：查看指定表在各个节点上占用的存储空间。

参数：表示待查询表的模式名和表名，均为text类型。

返回值类型：record

说明

- 使用本函数查询指定表存储分布信息，需要具备管理员权限或预置角色gs_role_read_all_stats权限。
- table_distribution性能比table_skewness更优，尤其是在大集群大数据量场景下，请优先考虑使用table_distribution函数。
- 当使用table_distribution并希望直观的看到空间占比时，可使用dnsize/(sum(dnsize) over ())的方式查看出具体的占比情况。

table_distribution(regclass)

描述：查看指定表在各个节点上占用的存储空间。

 版权所有 © 华为云计算技术有限公司

参数：表示待查询表的表名或OID，表名可以有模式名限定。为regclass类型。

返回值类型：record

说明

- 使用本函数查询指定表存储分布信息，需要具备管理员权限或预置角色gs_role_read_all_stats权限。
- table_distribution性能比table_skewness更优，尤其是在大集群大数据量场景下，请优先考虑使用table_distribution函数。
- 当使用table_distribution并希望直观的看到空间占比时，可使用dnsize/(sum(dnsize) over ())的方式查看出具体的占比情况。

table_distribution()

描述：查看当前库中所有表在各节点的存储空间分布情况。

返回值类型：record

说明

- 使用本函数涉及全库表信息查询，需要具备管理员权限或预置角色gs_role_read_all_stats权限。
- 当前基于table_distribution()函数，DWS提供视图PGXC_GET_TABLE_SKEWNESS进行数据倾斜查询，建议在数据库中表数量（小于10000）较少的场景直接使用。

gs_table_distribution(schemaname text, tablename text)

描述：快速查看指定表在各个节点上占用的存储空间。

返回值类型：record

表 7-45 gs_table_distribution(schemaname text, tablename text)函数返回字段

名称	类型	描述
schemaname	name	模式名称。
tablename	name	表名。
relkind	character	类型。 ● i：索引 ● r：表
relpersistence	character	类型。 ● t：本地临时表 ● g：全局临时表 ● u：unlogged表 ● p：普通表
nodename	name	节点名称。
dnsize	bigint	表在该节点上的存储空间大小，单位：字节。

版权所有 © 华为云计算技术有限公司

名称	类型	描述
sessionid	bigint	对于全局临时表显示每个会话的session id。

说明

- 使用本函数查询指定表存储分布信息，需要具备管理员权限或预置角色gs_role_read_all_stats权限。
- 该函数基于PG_RELFILENODE_SIZE系统表上的物理文件存储空间记录，需确保GUC参数use_workload_manager和enable_perm_space必须开启。
- 性能上，单表查询时，gs_table_distribution函数低于table_distribution函数；在全库表查询时，gs_table_distribution函数大幅度优于table_distribution函数；在大集群大数据量场景下，如果进行全库表查询，建议优先使用gs_table_distribution函数。
- 全局临时表可以显示每个会话在存储空间分布情况。通过sessionid列区分不同会话。

gs_table_distribution()

描述：快速查看当前库中所有表在各节点的存储空间分布情况。

返回值类型：record

表 7-46 gs_table_distribution()函数返回字段

名称	类型	描述
schemaname	name	模式名称。
tablename	name	表名。
relkind	character	类型。 • i：索引 • r：表
relpersistence	character	类型。 • t：本地临时表 • g：全局临时表 • u：unlogged表 • p：普通表
nodename	name	节点名称。
dnsize	bigint	表在该节点上的存储空间大小，单位：字节。
sessionid	bigint	对于全局临时表显示每个会话的session id。

版权所有 © 华为云计算技术有限公司

说明

- 使用本函数查询指定表存储分布信息，需要具备管理员权限或预置角色gs_role_read_all_stats权限。
- 该函数基于PG_RELFILENODE_SIZE系统表上的物理文件存储空间记录，需确保GUC参数use_workload_manager和enable_perm_space必须开启。
- 性能上，单表查询时，gs_table_distribution函数低于table_distribution函数；在全库表查询时，gs_table_distribution函数大幅度优于table_distribution函数；在大集群大数据量场景下，如果进行全库表查询，建议优先使用gs_table_distribution函数。
- 基于gs_table_distribution()函数，DWS自8.2.1集群版本提供了视图PGXC_WLM_TABLE_DISTRIBUTION_SKEWNESS进行数据倾斜查询，建议在数据库中表数量（小于10000）较少的场景直接使用。

check_foreign_key_constraint(schema.table)

描述：检测外键表中的数据是否满足外键约束。该参数仅9.1.0.200及以上集群版本支持。

返回值类型：text

示例：

```
set info_constraint_options = 'foreign_key';

CREATE TABLE ptt1 (a int,b int,e varchar(50),f varchar(50),PRIMARY KEY (a,b)) DISTRIBUTE BY HASH(a,b);
CREATE TABLE tt2 (a int , b int, c int, d int) DISTRIBUTE BY HASH(a,b);

select check_foreign_key_constraint('table_constraints.tt2');
check_foreign_key_constraint
------------------------------------------
tt2 table has no foreign key constraint

ALTER TABLE tt2 ADD CONSTRAINT fk_tt1_a FOREIGN KEY (a,b) REFERENCES ptt1(a,b);

select check_foreign_key_constraint('table_constraints.tt2');
check_foreign_key_constraint
------------------------------------------
tt2 satisfied all Foreign key constraints
```

plan_seed()

描述：获取前一次查询语句的seed值（内部使用）。

返回值类型：integer

pg_stat_get_env()

描述：提供获取当前节点的环境变量信息。

返回值类型：record

pg_stat_get_thread()

描述：提供当前节点下所有线程的状态信息。

返回值类型：record

pgxc_get_os_threads()

描述：提供当前集群中所有正常节点下的线程状态信息。

 版权所有 © 华为云计算技术有限公司

返回值类型：record

pg_stat_get_sql_count()

描述：提供当前节点中所有用户执行的SELECT/UPDATE/INSERT/DELETE/MERGE INTO语句的计数和响应时间的统计结果以及DDL、DML、DCL语句的计数结果。

返回值类型：record

pgxc_get_sql_count()

描述：提供当前集群所有节点中所有用户执行的SELECT/UPDATE/INSERT/DELETE/MERGE INTO语句的计数和响应时间的统计结果以及DDL、DML、DCL语句的计数结果。

返回值类型：record

pgxc_get_workload_sql_count()

描述：提供当前集群所有CN上所有Workload控制组内执行的SELECT/UPDATE/INSERT/DELETE语句的计数统计结果以及DDL、DML、DCL类型语句的计数统计结果。

返回值类型：record

pgxc_get_workload_sql_elapse_time()

描述：提供当前集群所有CN上所有Workload控制组内执行的SELECT/UPDATE/INSERT/DELETE语句的响应时间统计结果。

返回值类型：record

get_instr_unique_sql()

描述：提供当前节点中收集的Unique SQL的信息。如果是CN节点，将返回该CN上收集的Unique SQL的完整信息，即会收集其他CN和DN上对应Unique SQL的信息并进行汇总展示；如果是DN节点，将返回本DN节点上的Unique SQL信息。详见视图GS_INSTR_UNIQUE_SQL。

返回值类型：record

reset_instr_unique_sql(cstring, cstring, INT8)

描述：清理已收集的Unique SQL信息。输入参数含义如下：

- GLOBAL/LOCAL：清理范围为所有节点或当前节点。
- ALL/BY_USERID/BY_CNID/BY_GUC：ALL表示清理所有，BY_USERID/BY_CNID表示按照USERID或CNID进行清理，BY_GUC表示清理操作是由GUC参数instr_unique_sql_count设置值变小引起的。
- 第三个参数值对应第二个参数设置，ALL/BY_GUC的情况下该值无意义。

返回值类型：bool

 版权所有 © 华为云计算技术有限公司

pgxc_get_instr_unique_sql()

描述：提供集群中所有CN上收集的Unique SQL的完整信息。该函数只能在CN上执行。

返回值类型：record

get_instr_unique_sql_remote_cns()

描述：提供集群中除正在执行此函数的CN之外的所有CN上收集的Unique SQL的完整信息。该函数只能在CN上执行。

返回值类型：record

pgxc_get_node_env()

描述：提供获取集群中所有节点的环境变量信息。

返回值类型：record

gs_switch_relfilenode()

描述：交换两个表或分区的元信息（重分布工具内部使用，用户直接使用会有错误信息提示）。

返回值类型：integer

copy_error_log_create()

描述：创建COPY FROM容错机制所需要的错误表（public.pgxc_copy_error_log）。

返回值类型：boolean

说明

- 此函数会尝试创建public.pgxc_copy_error_log表，表的详细信息请参见**表7-47**。
- 在relname列上创建B-tree索引，并REVOKE ALL on public.pgxc_copy_error_log FROM public对错误表进行权限控制（与COPY语句权限一致）。
- 由于尝试创建的public.pgxc_copy_error_log定义是一张行存表，因此集群上必须支持行存表的创建才能够正常运行此函数，并使用后续的COPY容错功能。需要特别注意的是，enable_hadoop_env这个GUC参数开启后会禁止在集群内创建行存表（DWS默认为off）。
- 此函数自身权限为Sysadmin及以上（与错误表、COPY权限一致）。
- 若创建前public.pgxc_copy_error_log表已存在或者copy_error_log_relname_idx索引已存在，则此函数会报错回滚。

表 7-47 错误表 public.pgxc_copy_error_log 信息

列名称	类型	描述
relname	varchar	表名称。以模式名.表名形式显示。
begintime	timestamp with time zone	出现数据格式错误的时间。
filename	varchar	出现数据格式错误的数据源文件名。

版权所有 © 华为云计算技术有限公司

列名称	类型	描述
rownum	bigint	在数据源文件中，出现数据格式错误的行号。
rawrecord	text	在数据源文件中，出现数据格式错误的原始记录。为了防止字段长度过大，限制字段的长度不超过1024 byte。
detail	text	详细错误信息。
columnname	varchar	在数据源文件中，出现数据格式错误的字段名称。仅8.2.1.100及以上版本支持。
errcode	varchar	出错信息对应的错误码，使用sqlstate错误码。仅8.2.1.100及以上版本支持。
queryid	bigint	执行copy的sql语句id，用于唯一标识一条sql语句。仅8.2.1.100及以上版本支持。

pg_stat_get_status(tid, num_node_display)

描述：查询当前实例中工作线程（backend thread）以及辅助线程（auxiliary thread）的阻塞等待情况，其返回结果的详细含义参见PG_THREAD_WAIT_STATUS视图。输入参数含义如下：

- tid：表示线程ID，bigint类型。如果为NULL，则返回所有工作线程和辅助线程的等待情况；否则只返回指定ID线程的等待情况。
- num_node_display：integer类型。对于等待状态为“wait node”的记录，指定其wait_status列中显示的被等待节点的最大数量。
  - 如果为空或者小于等于0，则只显示一个被等待节点。
  - 如果大于20，则最多只显示20个节点。
  - 如果大于0且小于等于20，则显示数量为num_node_display和实际被等待节点数量的最小者。例如查询“SELECT * from pg_stat_get_status(NULL, 10)”，如果实际被等待节点数量大于10，则只随机显示其中10个节点名称，如果实际被等待节点数量小于等于10，则显示全部被等待节点名称。当实际被等待节点数量大于显示数量时，被显示的节点名称为随机挑选。

返回值类型：record

pgxc_get_thread_wait_status(num_node_display)

描述：查询集群各个节点上所有SQL语句产生的线程之间的调用层次关系，以及各个线程的阻塞等待状态。其返回结果的详细含义参见PGXC_THREAD_WAIT_STATUS视图。输入参数num_node_display的类型和含义与上述函数pg_stat_get_status相同。

返回值类型：record

pgxc_os_run_info()

描述：查询集群中各节点上操作系统运行的状态信息。

 版权所有 © 华为云计算技术有限公司

返回值类型：record

表 7-48 pgxc_os_run_info()函数返回字段

名称	类型	描述
id	integer	编号。
name	text	操作系统运行状态名称。
value	numeric	操作系统运行状态值。
comments	text	操作系统运行状态注释。
cumulative	boolean	操作系统运行状态的值是否为累加值。

get_instr_wait_event()

描述：查询当前实例上各类等待状态和事件的统计信息。如果GUC参数enable_track_wait_event为off，则返回0行。

返回值类型：record

表 7-49 get_instr_wait_event()函数返回字段

名称	类型	描述
nodename	name	节点名称。
type	text	事件的类型，包括STATUS，LOCK_EVENT，LWLOCK_EVENT和IO_EVENT四种类型。
event	text	事件名称。
wait	bigint	事件发生次数。该字段及以下字段均为进程运行中的累计值。
failed_wait	bigint	等待失败次数。当前版本中只有LOCK和LWLOCK等锁超时或失败才会使用该字段。
total_wait_time	bigint	该事件总持续时间。
avg_wait_time	bigint	该事件平均持续时间。
max_wait_time	bigint	该事件最大等待时间。
min_wait_time	bigint	该事件最小等待时间。

pgxc_wait_events()

描述：查询集群中各节点上各类等待状态和事件的统计信息。如果GUC参数enable_track_wait_event为off，则返回0行。

 版权所有 © 华为云计算技术有限公司

返回值类型：record

表 7-50 pgxc_wait_events()函数返回字段

名称	类型	描述
nodename	name	节点名称。
type	text	事件的类型，包括STATUS，LOCK_EVENT，LWLOCK_EVENT和IO_EVENT四种类型。
event	text	事件名称。
wait	bigint	事件发生次数。该字段及以下字段均为进程运行中的累计值。
failed_wait	bigint	等待失败次数。当前版本中只有LOCK和LWLOCK等锁超时或失败才会使用该字段。
total_wait_time	bigint	该事件总持续时间。
avg_wait_time	bigint	该事件平均持续时间。
max_wait_time	bigint	该事件最大等待时间。
min_wait_time	bigint	该事件最小等待时间。

pgxc_stat_bgwriter()

描述：查询集群中各节点上后端写进程活动的统计信息。

返回值类型：record

表 7-51 pgxc_stat_bgwriter()函数返回字段

名称	类型	描述
checkpoints_timed	bigint	定期执行的检查点数量。
checkpoints_req	bigint	请求执行的检查点数量。
checkpoint_write_time	double precision	检查点期间将文件写入磁盘花费的时间，以毫秒为单位。
checkpoint_sync_time	double precision	检查点期间数据同步到磁盘花费的时间，以毫秒为单位。
buffers_checkpoint	bigint	检查点期间写入缓冲区的数量。
buffers_clean	bigint	后端写进程写的缓冲区数量。

 版权所有 © 华为云计算技术有限公司

名称	类型	描述
maxwritten_clean	bigint	由于写入缓冲区太多，后端写进程停止清理扫描的次数。
buffers_backend	bigint	后端直接写入的缓冲区数量。
buffers_backend_fsync	bigint	后端需要fsync的次数。
buffers_alloc	bigint	分配的缓冲区数量。
stats_reset	timestamp with time zone	统计重置的时间。

pgxc_stat_replication()

描述：查询集群中各节点上日志同步的状态信息。

返回值类型：record

表 7-52 pgxc_stat_replication()函数返回字段

名称	类型	描述
pid	bigint	线程的PID。
usesysid	oid	用户系统ID。
usename	name	用户名。
application_name	text	程序名称。
client_addr	inet	客户端地址。
client_hostname	text	客户端名。
client_port	integer	客户端端口号。
backend_start	timestamp with time zone	程序启动时间。
state	text	日志复制的状态（追赶状态或一致的流状态）。
sender_sent_location	text	发送端发送日志位置。
receiver_write_location	text	接收端write日志位置。
receiver_flush_location	text	接收端flush日志位置。

 版权所有 © 华为云计算技术有限公司

名称	类型	描述
receiver_replay_location	text	接收端replay日志位置。
sync_priority	integer	同步复制的优先级（0表示异步）。
sync_state	text	同步状态（异步复制、同步复制或潜在同步）。

pgxc_replication_slots()

描述：查询集群中各DN上复制的状态信息。

返回值类型：record

表 7-53 pgxc_replication_slots()函数返回字段

名称	类型	描述
slot_name	text	复制节点的名称。
plugin	name	逻辑复制槽对应的输出插件名。
slot_type	text	复制节点的类型。
datoid	oid	复制节点的数据库OID。
database	name	复制节点的数据库名称。
active	boolean	复制节点是否为激活状态。
xmin	xid	复制节点事务标识。
catalog_xmin	text	逻辑复制槽对应的最早解码事务标识。
restart_lsn	text	复制节点的Xlog文件信息。
dummy_standby	boolean	复制节点是否为假备。

pgxc_settings()

描述：查询集群中各节点上运行时参数的相关信息。

返回值类型：record

表 7-54 pgxc_settings()函数返回字段

名称	类型	描述
name	text	参数名称。

 版权所有 © 华为云计算技术有限公司

名称	类型	描述
setting	text	参数当前值。
unit	text	参数的隐式结构。
category	text	参数的逻辑组。
short_desc	text	参数的简单描述。
extra_desc	text	参数的详细描述。
context	text	设置参数值的上下文，包括internal，postmaster，sighup，backend，superuser和user。
vartype	text	参数类型，包括bool，enum，integer，real和string。
source	text	参数的赋值方式。
min_val	text	参数最小值。如果参数类型不是数值型，那么该字段值为null。
max_val	text	参数最大值。如果参数类型不是数值型，那么该字段值为null。
enumvals	text[]	enum类型参数合法值。如果参数类型不是enum型，那么该字段值为null。
boot_val	text	数据库启动时参数默认值。
reset_val	text	数据库重置时参数默认值。
sourcefile	text	设置参数值的配置文件。如果参数不是通过配置文件赋值，那么该字段值为null。
sourceline	integer	设置参数值的配置文件的行号。如果参数不是通过配置文件赋值，那么该字段值为null。
sync_to_stream	boolean	标记是否在Stream线程同步设置参数的变更情况。该字段仅9.1.1.300及以上集群版本支持。

pgxc_instance_time()

描述：查询集群中各节点的运行时间统计信息及各执行阶段所消耗时间。

返回值类型：record

表 7-55 pgxc_instance_time()函数返回字段

名称	类型	描述
stat_id	integer	类型编号。
stat_name	text	运行时间类型名称。

 版权所有 © 华为云计算技术有限公司

名称	类型	描述
value	bigint	运行时间值。

pg_stat_get_redo_stat()

描述：查询当前节点上的XLOG重做统计信息。

返回值类型：record

表 7-56 pg_stat_get_redo_stat()函数返回字段

名称	类型	描述
phywrts	bigint	物理写次数。
phyblkwrt	bigint	物理写块数。
writetim	bigint	物理写消耗时间。
avgiotim	bigint	平均每次写入时间。
lstiotim	bigint	上一次写入时间。
miniotim	bigint	最小写入时间。
maxiowtm	bigint	最大写入时间。

pgxc_redo_stat()

描述：查询集群中各节点上的XLOG重做统计信息。

返回值类型：record

表 7-57 pgxc_redo_stat()函数返回字段

名称	类型	描述
phywrts	bigint	物理写次数。
phyblkwrt	bigint	物理写块数。
writetim	bigint	物理写消耗时间。
avgiotim	bigint	平均每次写入时间。
lstiotim	bigint	上一次写入时间。
miniotim	bigint	最小写入时间。
maxiowtm	bigint	最大写入时间。

版权所有 © 华为云计算技术有限公司

get_local_rel_iostat()

描述：查询当前实例上磁盘读写的统计信息。

返回值类型：record

表 7-58 get_local_rel_iostat()函数返回字段

名称	类型	描述
phyrds	bigint	读磁盘次数。
phywrts	bigint	写磁盘次数。
phyblkrd	bigint	读磁盘页数。
phyblkwrt	bigint	写磁盘页数。

pgxc_rel_iostat()

描述：查询集群中各节点上磁盘读写的统计信息。

返回值类型：record

表 7-59 pgxc_rel_iostat()函数返回字段

名称	类型	描述
phyrds	bigint	读磁盘次数。
phywrts	bigint	写磁盘次数。
phyblkrd	bigint	读磁盘页数。
phyblkwrt	bigint	写磁盘页数。

get_node_stat_reset_time()

描述：返回当前实例统计信息被重置的时间。

返回值类型：timestamptz

pgxc_node_stat_reset_time()

描述：查询集群中各节点统计信息被重置的时间。

返回值类型：record

 版权所有 © 华为云计算技术有限公司

说明

实例正常运行过程中，内存中的各类统计数值会逐渐累加，以下情况会导致内存中的统计数值被重置为0：

- 实例重启或集群发生了切换；
- 数据库Database被删除（drop）；
- 用户执行了重置操作，如执行pgstat_recv_resetcounter函数会将数据库中的统计计数器清零，执行reset_instr_unique_sql函数会将Unique SQL数据清零。

如果发生了以上事件，DWS会记录统计信息被重置的时间，可通过**get_node_stat_reset_time()**函数查询。

pgxc_parallel_query(text, text, bool)

描述：在指定类型的数据实例上执行指定的SQL查询语句，并把查询语句的结果返回给当前CN。该函数8.1.2及以上版本支持。

第一个参数text，SQL语句在哪些实例上执行。当前支持的有效入参是 'dn'，'datanode'，'cn'， 'coordinator'， 'all'， 其他值会导致函数执行报错。其中 'dn'，'datanode'表示在所有DN上执行，'cn'， 'coordinator'表示在所有CN上执行， all表示在所有的CN和所有的DN上执行。

第二个参数text，需要发往远程节点执行的SQL语句，函数内部会对SQL语句中查询的对象进行校验，不支持用户表、分布式表和自定义的多结果集函数。

返回值类型：record

第三个参数bool，控制CN间是否并行，默认为false。true表示开启，目前仅LSM离线重分布场景支持开启该参数。

说明

- 该函数功能仅为便于开发人员高效的收集集群内实例执行信息或者状态的视图，不建议用户直接使用。
- 该函数为多结果集函数，返回的数据类型为record，所以需要在函数调用后面增加AS语句显式的指出输出的列名和数据类型，如下所示：
  ```
  SELECT * FROM pgxc_parallel_query('all', 'select node_name, db_name, thread_name, query_id, tid, lwtid, ptid, tlevel, smpid, wait_status, wait_event from pg_thread_wait_status') AS (node_name text, db_name text, thread_name text, query_id bigint, tid bigint, lwtid integer , ptid integer, tlevel integer , smpid integer, wait_status text, wait_event text);
  ```
- 函数第二个参数指定的SQL语句输出结果的数据类型必须跟AS后面指明的数据类型一致，否则执行时可能会因为类型不匹配而报错。
- 函数第二个参数指定的SQL语句中不能触发跨节点的查询动作，否则会触发执行报错。
- 函数第二个参数指定的SQL语句只能是SELECT/UPDATE/DELETE/INSERT语句中的一种，且存在如下限制：
  - 不支持语句中有returning语句。
  - 调用函数用户需具有SQL语句中对象相应的操作权限。
  - 如果是INSERT语句，不支持INSERT OVERWRITE、UPSERT和INSERT INTO。
  - 对于UPDATE/DELETE/INSERT语句，只允许初始化用户在就地升级模式下或者管理员用户在重分布模式下执行；要求语句在每个实例上修改的记录数相同，否则执行会报错；函数会输出一条一列bigint类型的数值，此数值表示语句在每个实例上操作的记录数。
    ```
    SELECT * FROM pgxc_parallel_query('cn', 'UPDATE pg_partition SET relpages = 0') AS (updated bigint);
    ```

 版权所有 © 华为云计算技术有限公司

generate_wdr_report(begin_snap_id bigint, end_snap_id bigint, report_type cstring, report_scope cstring, node_name cstring)

描述：创建负荷分析报告。

输入参数含义如下：

- begin_snap_id，end_snap_id：生成报告的起止快照ID，bigint型，要求begin_snap_id < end_snap_id，并且起止快照的时间没有交集。判断快照时间是否有交集可通过查询dbms_om.snapshot表格select s1.end_ts < s2.start_ts from (select * from dbms_om.snapshot where snapshot_id=end_snap_id) as s2;，如果返回true则没有交集，反之有交集。
- report_type：报告类型，cstring型，包括“summary”，“detail”，“all”三种类型。
- report_scope：报告范围，cstring型，包括“cluster”和“node”两种。
- node_name：节点名称，cstring型，如果report_scope="node"，该参数必须是pg_catalog. pgxc_node表格中node_name字段里的CN或DN节点名称。

返回值类型：text

说明

- 该函数只有数据库管理员SYSADMIN才有权执行，非管理员执行会提示无权限。
- 该函数只能在CN上执行，在DN上执行会提示错误：“WDR report can only be created on coordinator.”。
- 如果生成报告成功，会返回：“Report %s has been generated”。
- 生成报告的两个快照期间不能发生统计信息重置事件，否则会提示错误：“Instance reset time is different”。引起统计信息重置的事件参见**pgxc_node_stat_reset_time**函数。

wdr_xdb_query(db_name text, snapshot_id bigint, view_name text)

描述：查询指定数据库下的指定视图。有的视图在不同数据库中查询结果不同，例如global_table_stat视图用于查询表格的统计信息，由于不同数据库下表格不同，在不同数据库中查询该视图得到的结果也不同。wdr_xdb_query函数可以在当前连接中访问db_name指定的数据库，并在该数据库中查询view_name指定的视图。输入参数含义如下：

- db_name：指定的数据库名称，text型。
- snapshot_id：快照ID，bigint型，参见“性能视图快照”。
- view_name：指定视图名称，text型。视图名称必须在如下白名单中：
  - global_table_stat
  - global_table_change_stat
  - global_column_table_io_stat
  - global_row_table_io_stat

    返回值类型：record，其第一列为snapshot_id bigint，第二列为db_name text，其他列的名称、类型和顺序与view_name指定的视图相同。

    示例：

```
select snapshot_id, db_name, schemaname, relname, distribute_mode,
seq_scan ,seq_tuple_read ,index_scan ,index_tuple_read ,tuple_inserted
,tuple_updated ,tuple_deleted ,tuple_hot_updated ,live_tuples ,dead_tuples from
wdr_xdb_query('postgres'::text, 1, 'global_table_stat'::text) as i(snapshot_id bigint, db_name text,
schemaname name, relname name, distribute_mode char, seq_scan bigint, seq_tuple_read
```

 版权所有 © 华为云计算技术有限公司

```
bigint, index_scan bigint, index_tuple_read bigint, tuple_inserted bigint, tuple_updated bigint,
tuple_deleted bigint, tuple_hot_updated bigint, live_tuples bigint, dead_tuples bigint);
```

说明

- 该函数仅8.1.2及以上版本支持。
- 该函数只有数据库管理员SYSADMIN才有权执行，非管理员执行会提示无权限。
- 该函数只能查询白名单中的视图，如果查询其他视图，会报错：“Input view name is invalid.”。

vac_fileclear_relation(oid)

描述：用于强制清理指定列存表中被VACUUM重写的文件，完成空间回收。

参数：列存表oid。

返回值类型：integer

说明

- 使用本函数前需要设置参数colvacuum_threshold_scale_factor，并确保VACUUM对指定列存表的文件完成重写后，才会清理文件并回收空间。
- 本函数将对指定的列存表施加排他锁。
- 延迟DDL开启后，此函数无法使用。

vac_fileclear_all_relation()

描述：用于强制清理所有列存表中被VACUUM重写的文件，完成空间回收。

返回值类型：record

说明

延迟DDL开启后，此函数无法使用。

get_col_file_info(table_name)

描述：查询指定列存表的空CU文件数和总CU文件数，该函数仅8.2.0及以上集群版本支持。

参数：列存表名。

返回值类型：int

返回值字段：

- total_file_num int：总CU文件数，范围为-1~INT_MAX，返回-1时表示表类型不支持等原因的失败，0~INT_MAX为总文件数。
- empty_file_num int：空CU文件数，范围为-1~INT_MAX，返回-1时表示表类型不支持等原因的失败，0~INT_MAX为空文件数。

示例：

```
call get_col_file_info('t4');
 total_file_num | empty_file_num
----------------+----------------
             10 |              7
(1 row)
```

 版权所有 © 华为云计算技术有限公司

get_all_col_file_info()

描述：查询所有列存表的空CU文件数和总CU文件数，该函数仅8.2.0及以上集群版本支持。

返回值类型：record

返回值字段：

- space_name text：列存表所属schema
- table_name text：列存表名
- total_file_num int：总CU文件数，范围为-1~INT_MAX，返回-1时表示表类型不支持等原因的失败，0~INT_MAX为总文件数。
- empty_file_num int：空CU文件数，范围为-1~INT_MAX，返回-1时表示表类型不支持等原因的失败，0~INT_MAX为空文件数。

示例：

```
call get_all_col_file_info();
 space_name | table_name | total_file_num | empty_file_num
------------+------------+----------------+----------------
 public     | t4         |             10 |              7
 public     | t2         |              1 |              1
 public     | t1         |              3 |              0
(3 rows)
```

pg_get_bucket_epoch(relid)

描述：查询v3表在当前节点上的epoch值。对于分区表，需要输入父表oid，将打印所有子分区的epoch值。该函数只支持在写VW上执行。该函数仅9.1.0.200及以上集群版本支持。

返回值类型：record

返回值字段：

- relfilenode oid：列存非分区表的relfilenode。对于分区表，此值为0。
- partfilenode oid：列存分区表的relfilenode。对于非分区表，此值为0。
- epoch xid：当前节点上该表对应的epoch值。

示例：

```
select * from pg_get_bucket_epoch(2147483732);
 relfilenode | partfilenode |  epoch
-------------+--------------+--------
  2137383742 |            0 | 1000001
(1 row)
```

pg_get_bucket_epoch(tablespaceid, databaseid, relfilenode)

描述：查询v3表在当前节点上的epoch值。输入v3表所在的表空间oid，数据库oid和表对应的relfilenode。该函数在写VW或读VW上均可执行。该函数仅9.1.0.200及以上集群版本支持。

返回值类型：xid

返回值字段：

epoch xid：当前节点上该表对应的epoch值。

 版权所有 © 华为云计算技术有限公司

示例：

```
select * from pg_get_bucket_epoch(2147483729, 16792, 2147483732);
 epoch
-------
1000001
(1 row)
```

pgxc_get_bucket_epoch(tablespaceid, databaseid, relfilenode)

描述：查询v3表在所有dn节点上的epoch值。输入v3表所在的表空间oid，数据库oid和表对应的relfilenode。该函数在写VW或读VW上均可执行。该函数仅9.1.0.200及以上集群版本支持。

返回值类型：record

返回值字段：

- nodename text：节点名称。
- epoch xid：当前节点上该表对应的epoch值。

示例：

```
select * from pgxc_get_bucket_epoch(2147483729, 16792, 2147483732);
  nodename |  epoch
-----------+---------
 datanode3 | 1000001
 datanode1 |       0
 datanode2 |       0
(1 row)
```

pgxc_check_bucket_info(tablename regclass, stoponerror)

描述：检查指定v3表所在的所有DN上的cuid、fileid映射信息是否正确，需在CN上执行。该函数仅9.1.1.200及以上集群版本支持。

参数：

tablename，表示表名。

stoponerror为可选参数，默认为false。

- 设置为true时，有一个DN上有错误信息时就报ERROR，不再检查其他DN。
- 设置为false时，会检测所有DN上cuid、fileid。

返回值类型：SETOF record

表 7-60 返回值字段

名称	类型	描述
nodename	text	节点名称。
is_valid	bool	当前节点检查是否通过。
error_msg	text	显示检查结果的错误信息。

示例：查看v3表所在的所有DN上的cuid、fileid映射信息是否正确。

 版权所有 © 华为云计算技术有限公司

```
SELECT * FROM pgxc_check_bucket_info('t1');
 nodename  | is_valid | error_msg
-----------+----------+-----------
 datanode2 | t        |
 datanode1 | t        |
(2 rows)
```

current_temp_schema()

描述：返回当前会话的临时schema。如果当前会话没有临时schema，则返回空。该函数仅8.2.1.220及以上集群版本支持。

返回值类型：name

示例：

```
select current_temp_schema();
 current_temp_schema
---------------------


(1 row)
create temp table tmp1(a int,b int) distribute by hash(a);
CREATE TABLE
select current_temp_schema();
           current_temp_schema
-----------------------------------------
 pg_temp_coordinator1_2_2_140718823485088
(1 row)
```

get_volatile_pg_class()

描述：获取当前会话中所有volatile临时表相关的pg_class元数据。该函数仅8.2.0及以上集群版本支持。

返回值类型：record

返回值字段：

- oid：返回volatile临时表的oid。
- 其他字段：同pg_class系统表中的基础字段(不包括系统隐藏字段)。

get_volatile_pg_class(relname text)

描述：获取当前会话中指定volatile临时表相关的pg_class元数据。该函数仅8.2.0及以上集群版本支持。

参数：当前会话中volatile临时表的表名。

返回值类型：record

返回值字段：

- oid：返回volatile临时表的oid。
- 其他字段：同pg_class系统表中的基础字段(不包括系统隐藏字段)。

示例：

```
SELECT * FROM get_volatile_pg_class('tx1');
  oid  | relname | relnamespace | reltype | reloftype | relowner | relam | relfilenode | reltablespace | relpages |
reltuples | relallvisible | reltoastrelid | reltoastidxid | reldeltarelid |
 reldeltaidx | relcudescrelid | relcudescidx | relhasindex | relisshared | relpersistence | relkind | relnatts |
relchecks | relhasoids | relhaspkey | relhasrules | relhastriggers | relhassub
```


版权所有 © 华为云计算技术有限公司

```
class | relcmprs | relhasclusterkey | relrowmovement | parttype | relfrozenxid | relacl |
reloptions             | relreplident | relfrozenxid64
-------+---------+--------------+---------+----------+----------+-------+------------+---------------+----------
+-----------+--------------+--------------+--------------+--------------+
-------------+---------------+-------------+-------------+-------------+----------------+---------+----------
+-----------+------------+------------+------------+---------------+----------
------+----------+-----------------+---------------+----------+--------------+--------
+----------------------------------+-------------+---------------
 16772 | tx1     |       16770 |   16774 |        0 |       10 |     0 |       16772 |         1665 |        0 |        0
|            0 |        16775 |            0 |            0 |
          0 |            0 |           0 | f          | f          | v            | r       |       2 |        0 | f         | f         |
f          | f            | f
     | 1       | f             | f             | n       | 11815       |       | {orientation=row,compression=no} |
d          |         11815
(1 row)
```

get_volatile_pg_attribute()

描述：获取当前会话中所有volatile临时表相关的pg_attribute元数据。该函数仅8.2.0及以上集群版本支持。

返回值类型：record

返回值字段：

- oid：返回该列信息的oid。
- 其他字段：同pg_attribute系统表中的基础字段（不包括系统隐藏字段）。

get_volatile_pg_attribute(relname text, attrname text)

描述：获取当前会话中指定volatile临时表相关的pg_attribute元数据。该函数仅8.2.0及以上集群版本支持。

参数：

- relname：表名(必须在当前会话中)。
- attrname：列名。

返回值类型：record

返回值字段：

- oid：返回该列信息的oid。
- 其他字段：同pg_attribute系统表中的基础字段（不包括系统隐藏字段）。

示例：

```
SELECT * FROM get_volatile_pg_attribute('tx1', 'b');
 attrelid | attname | atttypid | attstattarget | attlen | attnum | attndims | attcacheoff | atttypmod | attbyval |
attstorage | attalign | attnotnull | atthasdef | attisdropped | attislocal |
 attcmprmode | attinhcount | attcollation | attacl | attoptions | attfdwoptions | attinitdefval | attkvtype
----------+---------+----------+---------------+--------+--------+----------+-------------+-----------+----------
+------------+----------+------------+-----------+--------------+------------+
-------------+-------------+--------------+--------+------------+---------------+---------------+-----------
   16772 | b       |       25 |            -1 |     -1 |      2 |        0 |          -1 |        -1 | f        | x          | i        |
f          | f         | f           | t         |
 127        |          0 |         100 |       |           |             |             | 0
(1 row)
```

 版权所有 © 华为云计算技术有限公司

pg_get_publication_tables(pubname text)

描述：根据发布的名称，返回对应发布要发布的表的relid列表。该函数仅8.2.0.100及以上集群版本支持。

参数：pubname

返回值类型：set of oid

示例：

```
SELECT * FROM pg_get_publication_tables('mypub');
 relid
-------
 16757
 16776
(2 rows)
```

pg_relation_is_publishable(relname regclass)

描述：检查一个表是否可以发布。该函数仅8.2.0.100及以上集群版本支持。

参数：relname

返回值类型：boolean

示例：

```
SELECT * FROM pg_relation_is_publishable('t1');
 pg_relation_is_publishable
----------------------------
 t
(1 row)
```

get_col_cu_info(schema_name text, table_name text, row_count int8, dirty_percent int8)

描述：查询某张列存表的CU信息，每个分区单独统计。该函数仅8.2.0.100及以上集群版本支持。

参数：模式名（必选），表名（必选），小CU的行数阈值（可选，默认200，取值范围为1~60000），脏CU的删除百分比阈值（可选，默认70，取值范围1~100）。

返回值类型：record

返回值字段：

- node_name：DN节点名。
- part_name：分区名，普通表此列为空。
- zero_size_cu_count：cuSize=0，且行数小于等于row_count的CU数。
- small_cu_count：cuSize=ALIGNOF_CUSIZE(8192)，且行数小于等于row_count的CU数。
- dirty_cu_count: 因delete，deadtuple占比超过dirty_percent的CU数。
- total_cu_count：总CU数。
- small_cu_size：8kCU的总大小。
- total_cu_size：总CU大小。

示例：

 版权所有 © 华为云计算技术有限公司

```
 SELECT * FROM get_col_cu_info('public','hs_part');
 node_name | part_name | zero_size_cu_count | small_cu_count | dirty_cu_count | total_cu_count |
small_cu_size | total_cu_size
-----------+-----------+--------------------+----------------+----------------+----------------+---------------
+---------------
 dn_1      | p1        |                  3 |              0 |              0 |              3 | 0 bytes       | 0 bytes
 dn_1      | p2        |                  3 |              0 |              0 |              3 | 0 bytes       | 0 bytes
 dn_1      | p3        |                  3 |              0 |              0 |              3 | 0 bytes       | 0 bytes
(3 rows)

 SELECT * FROM get_col_cu_info('public','hs_part', 200, 90);
 node_name | part_name | zero_size_cu_count | small_cu_count | dirty_cu_count | total_cu_count |
small_cu_size | total_cu_size
-----------+-----------+--------------------+----------------+----------------+----------------+---------------
+---------------
 dn_1      | p1        |                  3 |              0 |              0 |              3 | 0 bytes       | 0 bytes
 dn_1      | p2        |                  3 |              0 |              0 |              3 | 0 bytes       | 0 bytes
 dn_1      | p3        |                  3 |              0 |              0 |              3 | 0 bytes       | 0 bytes
(3 rows)
```

get_col_file_vacuum_info(schema_name text, table_name text, force_get_rewritten_file_num bool)

描述：查询某张列存表的文件vacuum相关信息，每个分区单独统计。该函数仅8.2.0.100及以上集群版本支持。

参数：模式名（必选），表名（必选），是否强制获取准确的可清理文件数（必选，默认为false）

返回值类型：record

返回值字段：

- node_name：DN节点名。
- part_name：分区名，普通表此列为空。
- total_file_num：总CU文件数。
- rewritable_file_num：可重写，但还未重写的文件数。
- rewritten_file_num：已重写未清理文件数，此值来源于内存态ClearCache，若由于重启等原因ClearCache丢失，可通过设置force_get_rewritten_file_num=true，消耗额外的时间，强制获取准确的可清理文件数。
- empty_file_num：已清理文件数。

示例：

```
SELECT * FROM get_col_file_vacuum_info('public','pa',false);
 node_name | part_name | total_file_num | rewritable_file_num | rewritten_file_num | empty_file_num
-----------+-----------+----------------+---------------------+--------------------+----------------
 datanode1 | pa1       |              1 |                   0 |                  0 |              0
 datanode1 | pa2       |              1 |                   0 |                  0 |              0
 datanode2 | pa1       |              1 |                   0 |                  0 |              0
 datanode2 | pa2       |              1 |                   0 |                  0 |              0
 datanode3 | pa1       |              1 |                   0 |                  0 |              0
 datanode3 | pa2       |              1 |                   0 |                  0 |              0
(6 rows)
```

get_col_file_vacuum_info(schema_name text, table_name text, colvacuum_threshold_scale_factor int)

描述：查询某张列存表的文件vacuum相关信息，每个分区单独统计。该函数仅8.2.0.100及以上集群版本支持。

 版权所有 © 华为云计算技术有限公司

参数：模式名（必选），表名（必选），colvacuum_threshold_scale_factor（必选，0-100为dead tuple的比例值）

返回值类型：record

返回值：

- node_name：DN节点名。
- part_name：分区名，普通表此列为空。
- total_file_num：总CU文件数。
- rewritable_file_num：可重写，但还未重写的文件数。
- rewritten_file_num：已重写未清理文件数，此值来源于内存态ClearCache，若由于重启等原因ClearCache丢失，可通过设置force_get_rewritten_file_num=true，消耗额外的时间，强制获取准确的可清理文件数。
- empty_file_num：已清理文件数。

示例：

```
SELECT * FROM get_col_file_vacuum_info('public','pa',10);
 node_name | part_name | total_file_num | rewritable_file_num | rewritten_file_num | empty_file_num
-----------+-----------+----------------+---------------------+--------------------+----------------
 datanode1 | pa1       |              1 |                   0 |                  0 |              0
 datanode1 | pa2       |              1 |                   0 |                  0 |              0
 datanode2 | pa1       |              1 |                   0 |                  0 |              0
 datanode2 | pa2       |              1 |                   0 |                  0 |              0
 datanode3 | pa1       |              1 |                   0 |                  0 |              0
 datanode3 | pa2       |              1 |                   0 |                  0 |              0
(6 rows)
```

gs_scan_for_online_vacuum_full()

描述：扫描全库中存在Online Vacuum Full残留的表。该函数仅9.1.2及以上集群版本支持。

扫描全库中满足以下条件的表：

- relkind='r'（普通表）且relpersistence='p'（持久表）。
- reloptions包含append_mode=on。
- 对应的辅助schema data_redis_ovf_<rel_cn_oid> 仍然存在。
- 排除正在被Online Vacuum Full处理的表（通过检查advisory lock判断）。

返回值类型：TABLE

返回值字段：

表 7-61 gs_scan_for_online_vacuum_full()返回字段

名称	类型	描述
schema_name	text	表所在的schema名称。
table_name	text	表名称。
full_name	text	带引号标识符的完整限定名，格式为schema.table，可直接用于清理函数的入参。

 版权所有 © 华为云计算技术有限公司

名称	类型	描述
rel_cn_oid	bigint	表在第一个CN上的OID，用于构造辅助schema名data_redis_ovf_<rel_cn_oid>。
aux_schema	text	辅助schema名，格式为data_redis_ovf_<rel_cn_oid>。

示例：

```
SELECT * FROM pg_catalog.gs_scan_for_online_vacuum_full();
 schema_name | table_name |     full_name     | rel_cn_oid |        aux_schema
-------------+------------+-------------------+------------+---------------------------
 public      | t_sales    | public.t_sales    |  16384     | data_redis_ovf_16384
 public      | t_orders   | public.t_orders   |  16392     | data_redis_ovf_16392
(2 rows)
```

gs_cleanup_for_online_vacuum_full(target_tables text[])

描述：批量清理指定表的Online Vacuum Full残留。目标表的过滤条件与**gs_scan_for_online_vacuum_full()**一致。该函数仅9.1.2及以上集群版本支持。

需注意，整个清理在一个事务内完成，无中间提交。如果表数量过多可能导致表被锁定的时间较长。

返回值类型：void

返回值字段：

表 7-62 gs_cleanup_for_online_vacuum_full(target_tables text[])返回字段

消息级别	消息内容	触发条件
INFO	successfully cleaned up table "<表名>": append_mode turned off, auxiliary schema "<schema名>" dropped	单表清理成功。
INFO	Cleanup completed: tables cleaned=<N>, tables skipped=<N>, tables failed=<N>	全部清理完成后的统计。
NOTICE	table "<表名>" does not have append_mode=on, not an OVF residual, skipping	表不是append_mode，跳过。
NOTICE	table "<表名>" has append_mode=on but OVF auxiliary schema "<schema名>" does not exist, not an OVF residual, skipping	有append_mode但无辅助schema，跳过。
NOTICE	table "<表名>" is currently being processed by OVF, skipping	表正在被Online Vacuum Full处理（advisory lock被持有），跳过。
NOTICE	parameter target_tables is NULL or empty, nothing to cleanup	传入空数组或 NULL参数。

版权所有 © 华为云计算技术有限公司

消息级别	消息内容	触发条件
WARNING	table "<表名>" has append_mode=on but no rel_cn_oid, cannot determine OVF auxiliary schema, skipping	有append_mode但无rel_cn_oid，跳过。
WARNING	failed to cleanup table "<表名>": <错误信息>, skipping	单表清理失败。
WARNING	failed to release advisory lock for table "<表名>": <错误信息>	释放advisory lock失败。
EXCEPTION	cannot execute gs_cleanup_for_online_vacuum_full during recovery	在恢复模式（备机）下执行。

示例：

```
SELECT pg_catalog.gs_cleanup_for_online_vacuum_full(ARRAY['public.t_sales', 'public.t_orders']);

INFO:  successfully cleaned up table "public.t_sales": append_mode turned off, auxiliary schema "data_redis_ovf_16384" dropped
INFO:  successfully cleaned up table "public.t_orders": append_mode turned off, auxiliary schema "data_redis_ovf_16392" dropped
INFO:  Cleanup completed: tables cleaned=2, tables skipped=0, tables failed=0
```

```
SELECT pg_catalog.gs_cleanup_for_online_vacuum_full(
    ARRAY(SELECT full_name FROM pg_catalog.gs_scan_for_online_vacuum_full())
);
```

gs_cleanup_for_online_vacuum_full(target_table text)

描述：清理指定单个表的Online Vacuum Full残留。该函数仅9.1.2及以上集群版本支持。

返回值类型：void

返回值字段：

示例：

```
SELECT pg_catalog.gs_cleanup_for_online_vacuum_full('public.t_sales');
```

get_all_col_cu_info(row_count int8)

描述：查询全库所有列存表的CU信息。该函数仅8.2.0.100及以上集群版本支持。

参数：小CU的行数阈值（可选，默认值200，取值范围为1~60000。）

返回值类型：record

返回值字段：

- node_name：DN节点名。
- schema_name：模式名。
- table_name：表名。
- zero_size_cu_count：cuSize=0，且行数小于等于row_count的CU数。
- small_cu_count：cuSize=ALIGNOF_CUSIZE(8192)，且行数小于等于row_count的CU数。

 版权所有 © 华为云计算技术有限公司

- total_cu_count：总CU数。
- small_cu_size：8kCU的总大小。
- total_cu_size：总CU大小。

示例：

```
SELECT * FROM get_all_col_cu_info(200);
 node_name | schema_name |      table_name      | zero_size_cu_count | small_cu_count | total_cu_count |
small_cu_size | total_cu_size
-----------+-------------+----------------------+--------------------+----------------+----------------+---------------
+---------------
 datanode1 | public      | udi_48076            |                  5 |              1 |              6 | 8192 bytes    | 8192 bytes
 datanode1 | public      | udi_48077            |                  5 |              1 |              6 | 8192 bytes    | 8192 bytes
 datanode2 | public      | udi_48076            |                  5 |              1 |              6 | 8192 bytes    | 8192 bytes
 datanode2 | public      | udi_48077            |                  5 |              1 |              6 | 8192 bytes    | 8192 bytes
 datanode3 | public      | udi_48076            |                  5 |              1 |              6 | 8192 bytes    | 8192 bytes
 datanode3 | public      | udi_48077            |                  5 |              1 |              6 | 8192 bytes    | 8192 bytes
(6 rows)
```

get_all_col_file_vacuum_info(force_get_rewritten_file_num bool)

描述：查询全库所有列存表的文件vacuum相关信息。该函数仅8.2.0.100及以上集群版本支持。

参数：是否强制获取准确的可清理文件数（必选，true或false）

返回值类型：record

返回值字段：

- node_name：DN节点名。
- schema_name：模式名。
- table_name：表名。
- total_file_num：总CU文件数。
- rewritable_file_num：可重写，但还未重写的文件数。
- rewritten_file_num：已重写未清理文件数，此值来源于内存态ClearCache，若由于重启等原因ClearCache丢失，可通过设置force_get_rewritten_file_num=true，消耗额外的时间，强制获取准确的可清理文件数。
- empty_file_num：已清理文件数。

示例：

```
SELECT * FROM get_all_col_file_vacuum_info(false);
 node_name | schema_name |      table_name      | total_file_num | rewritable_file_num | rewritten_file_num
| empty_file_num
-----------+-------------+----------------------+----------------+---------------------+--------------------
+----------------
 datanode1 | public      | udi_57373            |              2 |                   0 |                  0 |              1
 datanode1 | public      | udi_57374            |              2 |                   0 |                  0 |              1
 datanode2 | public      | udi_57373            |              2 |                   0 |                  0 |              1
 datanode2 | public      | udi_57374            |              2 |                   0 |                  0 |              1
 datanode3 | public      | udi_57373            |              2 |                   0 |                  0 |              1
 datanode3 | public      | udi_57374            |              2 |                   0 |                  0 |              1
```

show_tsc_info()

描述：查询当前数据库节点已获取的TSC（TimeStamp-Counter）信息。该函数仅8.2.1及以上集群版本支持。

 版权所有 © 华为云计算技术有限公司

返回值类型：record

表 7-63 返回值字段

名称	类型	描述
node_name	text	节点名称。
tsc_mult	bigint	TSC换算乘数。
tsc_shift	bigint	TSC换算位移数。
tsc_frequency	float8	TSC频率。
tsc_use_frequency	boolean	是否使用TSC频率进行时间换算。
tsc_ready	boolean	是否可以使用TSC频率进行时间换算。
tsc_scalar_error_info	text	获取TSC换算信息的错误信息。
tsc_freq_error_info	text	获取TSC频率的错误信息。

示例：

```
SELECT * FROM show_tsc_info();
  node_name   | tsc_mult | tsc_shift | tsc_frequency | tsc_use_frequency | tsc_ready |
tsc_scalar_error_info     | tsc_freq_error_info
--------------+----------+-----------+---------------+-------------------+-----------+-------------------------------
+---------------------
 coordinator1 |          |           |          2400 | t                 | t         | TSC scalar is not initialized |
```

get_tsc_info()

描述：重新获取当前数据库节点的TSC（TimeStamp-Counter）信息。该函数仅8.2.1及以上集群版本支持。

返回值类型：record

表 7-64 返回值字段

名称	类型	描述
node_name	text	节点名称。
tsc_mult	bigint	TSC换算乘数。
tsc_shift	bigint	TSC换算位移数。
tsc_frequency	float8	TSC频率。
tsc_use_frequency	boolean	是否使用TSC频率进行时间换算。
tsc_ready	boolean	是否可以使用TSC频率进行时间换算。

 版权所有 © 华为云计算技术有限公司

名称	类型	描述
tsc_scalar_error_info	text	获取TSC换算信息的错误信息。
tsc_freq_error_info	text	获取TSC频率的错误信息。

示例：

```
SELECT * FROM get_tsc_info();
  node_name   | tsc_mult | tsc_shift | tsc_frequency | tsc_use_frequency | tsc_ready |
tsc_scalar_error_info     | tsc_freq_error_info
--------------+----------+-----------+---------------+-------------------+-----------+-------------------------------
+---------------------
 coordinator1 |          |           |          2400 | t                 | t         | TSC scalar is not initialized |
```

test_tsc_info(time float8, loops int)

描述：测试当前节点通过TSC（TimeStamp-Counter）换算时间的准确性。该函数仅8.2.1及以上集群版本支持。

输入参数含义如下：

- time：测试的时间差（单位s），测试时长需要小于或者等于60s
- loops：测试次数，范围值[1，10]

返回值类型：record

返回值字段：

- id：循环次数。
- real_time_diff：通过gettimeofday获取的时间差（单位us）。
- est_time_scalar：通过TSC换算信息转换得到的时间差（单位us）。
- est_time_frequency：通过TSC频率转换得到的时间差（单位us）。

示例：

```
SELECT * FROM test_tsc_info(0.01,10);
 id | real_time_diff | est_time_scalar | est_time_frequency
----+----------------+-----------------+--------------------
  1 |          10057 |                 |            10056.9
  2 |          10057 |                 |   10057.4816666667
  3 |          10056 |                 |   10055.2841666667
  4 |          10054 |                 |   10054.4908333333
  5 |          10055 |                 |         10054.2875
  6 |          10055 |                 |   10054.7483333333
  7 |          10055 |                 |         10054.4725
  8 |          10054 |                 |   10054.0766666667
  9 |          10058 |                 |   10058.1016666667
 10 |          10057 |                 |   10056.3733333333
(10 rows)
```

gs_set_hold_table_level_oldestxmin()

描述：当开启表级oldestxmin特性后，可以通过该函数将事务标记为长事务，计算任何表的oldestxmin时，会将此事务作为活跃事务计算到oldestxmin中。该函数仅9.1.1.100及以上集群版本支持。

 版权所有 © 华为云计算技术有限公司

返回值类型：void

以下两种场景可以使用该函数：

1. 该事务是一个潜在的长事务，即运行时间大于old_txn_threshold，并且涉及多张表。
2. 执行语句可能因锁等待等原因，在CN上的执行时间长。如COPY、GDS导入导出场景等。

说明

- 长事务指的是运行大于某一时间后还未提交的事务，由GUC参数old_txn_threshold设置，默认为600秒。
- 执行该函数会影响表级oldestxmin的计算规则：无论事务运行时间是否大于old_txn_threshold设置的值或计算的表是否与该事务相关，都会将该事务当作活跃事务，影响所有表的oldestxmin的计算。
- 语句执行完成后建议及时调用gs_unset_hold_table_level_oldestxmin()函数取消标记，否则会影响VACUUM的清理效果。

gs_unset_hold_table_level_oldestxmin()

描述：当开启表级oldestxmin特性后，可以通过该函数取消标记为长事务的事务。该函数仅9.1.1.100及以上集群版本支持。

返回值类型：void

说明

长事务指的是运行大于某一时间后还未提交的事务，由GUC参数old_txn_threshold设置，默认为600秒。

7.28 备份恢复控制函数

7.28.1 备份控制函数

备份控制函数可帮助进行在线备份。

pg_create_restore_point(name text)

描述：为执行恢复创建一个命名点。（需要管理员角色）

返回值类型：text

备注：pg_create_restore_point创建了一个可以用作恢复目的、有命名的事务日志记录，并返回相应的事务日志位置。在恢复过程中，recovery_target_name可以通过这个名称定位对应的日志恢复点，并从此处开始执行恢复操作。避免使用相同的名称创建多个恢复点，因为恢复操作将在第一个匹配（恢复目标）的名称上停止。

pg_current_xlog_location()

描述：获取当前事务日志的写入位置。

返回值类型：text

 版权所有 © 华为云计算技术有限公司

备注：pg_current_xlog_location使用与前面那些函数相同的格式显示当前事务日志的写入位置。如果是只读操作，不需要系统管理员权限。

示例：

```
SELECT pg_current_xlog_location();
 pg_current_xlog_location
--------------------------
 0/4310D5D8
(1 row)
```

pg_current_xlog_insert_location()

描述：获取当前事务日志的插入位置。

返回值类型：text

备注：pg_current_xlog_insert_location显示当前事务日志的插入位置。插入点是事务日志在某个瞬间的“逻辑终点”，而实际的写入位置则是从服务器内部缓冲区写出时的终点。写入位置是可以从服务器外部检测到的终点，如果要归档部分完成事务日志文件，则该操作即可实现。插入点主要用于服务器调试目的。如果是只读操作，不需要系统管理员权限。

示例：

```
SELECT pg_current_xlog_insert_location();
 pg_current_xlog_insert_location
---------------------------------
 0/4316AD50
(1 row)
```

pg_start_backup(label text [, fast boolean])

描述：开始执行在线备份。（需要管理员角色或复制的角色）

返回值类型：text

备注：pg_start_backup接受一个用户定义的备份标签（通常这是备份转储文件存放地点的名字）。这个函数向数据库集群的数据目录写入一个备份标签文件，然后以文本方式返回备份的事务日志起始位置。

```
SELECT pg_start_backup('label_goes_here');
 pg_start_backup
-----------------
 0/3000020
(1 row)
```

pg_stop_backup()

描述：完成执行在线备份。（需要管理员角色或复制的角色）

返回值类型：text

备注：pg_stop_backup删除pg_start_backup创建的标签文件，并且在事务日志归档区里创建一个备份历史文件。这个历史文件包含给予pg_start_backup的标签、备份的事务日志起始与终止位置、备份的起始和终止时间。返回值是备份的事务日志终止位置。计算出中止位置后，当前事务日志的插入点将自动前进到下一个事务日志文件，这样，结束的事务日志文件可以被立即归档从而完成备份。

示例：

```
SELECT pg_stop_backup();
 pg_stop_backup
```

 版权所有 © 华为云计算技术有限公司

```
----------------
 0/4401E610
(1 row)
```

pg_switch_xlog()

描述：切换到一个新的事务日志文件。（需要管理员角色）

返回值类型：text

备注：pg_switch_xlog移动到下一个事务日志文件，以允许将当前日志文件归档（假定使用连续归档）。返回值是刚完成的事务日志文件的事务日志结束位置+1。如果从最后一次事务日志切换以来没有活动的事务日志，则pg_switch_xlog什么事也不做，直接返回当前事务日志文件的开始位置。

示例:

```
SELECT pg_switch_xlog();
 pg_switch_xlog
----------------
 0/45017978
(1 row)
```

pg_xlogfile_name(location text)

描述：将事务日志的位置字符串转换为文件名。

返回值类型：text

备注：pg_xlogfile_name仅抽取事务日志文件名称。如果给定的事务日志位置恰好位于事务日志文件的交界上，这两个函数都返回前一个事务日志文件的名字。这对于管理事务日志归档来说是非常有利的，因为前一个文件是当前最后一个需要归档的文件。

pg_xlogfile_name_offset(location text)

描述：将事务日志的位置字符串转换为文件名并返回在文件中的字节偏移量。

返回值类型：text,integer

备注：可以使用pg_xlogfile_name_offset从前述函数的返回结果中抽取相应的事务日志文件名称和字节偏移量。例如：

```
SELECT * FROM pg_xlogfile_name_offset(pg_stop_backup());
NOTICE:  pg_stop_backup cleanup done, waiting for required WAL segments to be archived
NOTICE:  pg_stop_backup complete, all required WAL segments have been archived
        file_name         | file_offset
--------------------------+-------------
000000010000000000000003 |         272
(1 row)
```

pg_xlog_location_diff(location text, location text)

描述：计算两个事务日志位置之间在字节上的区别。

返回值类型：numeric

pg_cbm_tracked_location()

描述：用于查询cbm解析到的lsn位置。

返回值类型：text

 版权所有 © 华为云计算技术有限公司

pg_cbm_get_merged_file(startLSNArg text, endLSNArg text)

描述：用于将指定lsn范围之内的cbm文件合并成一个cbm文件，并返回合并完的cbm文件名。

返回值类型：text

pg_cbm_get_changed_block(startLSNArg text, endLSNArg text)

描述：用于将指定lsn范围之内的cbm文件合并成一个表，并返回表的各行记录。

返回值类型：record

备注：pg_cbm_get_changed_block返回的表字段包含：合并起始的lsn，合并截止的lsn， 表空间oid，库oid，表的relfilenode，表的fork number，表是否被删除，表是否被创建，表是否被截断，表被截断后的页面数，有多少页被修改以及被修改的页号的列表。

pg_cbm_recycle_file(slotName name, targetLSNArg text)

描述：删除不再使用的cbm文件，并返回删除后的第一条lsn。其中，slotName为空字符串时，将直接使用targetLSNArg作为回收点；在备份和容灾时，因为并行因素，需要指定相应的slotName，首先记录该任务的targetLSNArg到该slot中，然后遍历所有的backup slots，找到最小的lsn作为回收点。

返回值类型：text

pg_cbm_force_track(targetLSNArg text,timeOut int)

描述：强制执行一次cbm追踪到指定的xlog位置，并返回实际追踪结束点的xlog位置。

返回值类型：text

pg_enable_delay_ddl_recycle()

描述：开启延迟DDL功能，并返回开启点的xlog位置。

返回值类型：text

示例：

```
SELECT pg_enable_delay_ddl_recycle();
 pg_enable_delay_ddl_recycle
-----------------------------
 00000000/4607B250
(1 row)
```

pg_disable_delay_ddl_recycle(barrierLSNArg text, isForce bool)

描述：关闭延迟DDL功能，并返回本次延迟DDL生效的xlog范围。

返回值类型：record

pg_enable_delay_xlog_recycle()

描述：开启延迟xlog回收功能。

返回值类型：void

 版权所有 © 华为云计算技术有限公司

示例：

```
SELECT pg_enable_delay_xlog_recycle();
 pg_enable_delay_xlog_recycle
------------------------------

(1 row)
```

pg_disable_delay_xlog_recycle()

描述：关闭延迟xlog回收功能。

返回值类型：void

示例：

```
SELECT pg_disable_delay_xlog_recycle();
 pg_disable_delay_xlog_recycle
-------------------------------

(1 row)
```

pgxc_get_senders_catchup_time()

描述：显示所有DN上当前活跃的主备发送线程的追赶信息。

返回值类型：record

函数返回信息如下：

表 7-65 pgxc_get_senders_catchup_time()字段

名称	类型	描述
node_name	text	节点名称。
lwpid	integer	当前sender的lwpid。
local_role	text	本地的角色。
peer_role	text	对端的角色。
state	text	当前sender的复制状态。
sender	text	当前sender的类型。
catchup_start	timestamp with time zone	catchup启动的时间。
catchup_end	timestamp with time zone	catchup结束的时间。
catchup_type	text	catchup方式为全量还是增量。
catchup_bcm_filename	text	catchup当前执行到的bcm文件
catchup_bcm_finished	integer	catchup已经操作完成的bcm文件数量。

版权所有 © 华为云计算技术有限公司

名称	类型	描述
catchup_bcm_total	integer	catchup总共需要操作的bcm文件数量。
catchup_percent	text	catchup已经操作完成的百分比。
catchup_remaining_time	text	catchup预估剩余时间。

pgxc_lock_for_backup()

描述：备份前给集群加锁，备份是为了在新增节点上执行恢复操作。

返回值类型：boolean

说明

pgxc_lock_for_backup是在使用gs_dump或gs_dumpall工具备份集群前，用来给集群加锁的。当给集群加锁后，不允许有改变系统结构的操作。该函数不影响DML语句。

示例：

```
SELECT pgxc_lock_for_backup();
 pgxc_lock_for_backup
----------------------
 t
(1 row)
```

7.28.2 恢复控制函数

恢复控制函数提供了当前备机状态的信息。此类函数可在恢复期间或数据库正常运行时执行。

pg_is_in_recovery()

描述：如果恢复仍然在进行中则返回true。

返回值类型：bool

示例：

```
SELECT pg_is_in_recovery();
 pg_is_in_recovery
-------------------
 f
(1 row)
```

pg_last_xlog_receive_location()

描述：获取最后接收事务日志的位置并通过流媒体复制同步到磁盘。当流复制仍在进行，事务日志将持续递增。如果恢复已完成，则这个值将保持最后接收事务记录的状态并在恢复期间同步到磁盘。如果不能用流复制，或还没有开始，该函数返回NULL。

返回值类型：text

示例：

 版权所有 © 华为云计算技术有限公司

```
SELECT pg_last_xlog_receive_location();
 pg_last_xlog_receive_location
-------------------------------

(1 row)
```

pg_last_xlog_replay_location()

描述：获取最后一个事务日志在恢复时重放的位置。当恢复仍在进行，事务日志将持续递增。如果已经完成恢复，则将保持在恢复期间最后接收WAL记录的值。如果未进行恢复但服务器正常启动时，则这个函数返回NULL。

返回值类型：text

示例：

```
SELECT pg_last_xlog_replay_location();
 pg_last_xlog_replay_location
-------------------------------
 0/2B16530
(1 row)
```

pg_last_xact_replay_timestamp()

描述：获取最后一个事务在恢复时重放的时间戳。这是为在主节点上生成事务提交或终止WAL记录的时间。如果在恢复时没有事务重放，则这个函数返回NULL。否则，如果恢复仍在进行，则事务日志将持续递增。如果恢复已经完成，则将保持在恢复期间最后接收WAL记录的值。如果服务器无需恢复就已正常启动，则这个函数返回NULL。

返回值类型：timestamp with time zone

恢复控制函数控制恢复的进程。这些函数可能只在恢复时被执行。

示例：

```
SELECT pg_last_xact_replay_timestamp();
 pg_last_xact_replay_timestamp
-------------------------------
 2023-01-04 07:03:08.098024+00
(1 row)
```

pg_is_xlog_replay_paused()

描述：如果恢复暂停则返回true。

返回值类型：bool

pg_xlog_replay_pause()

描述：立即暂停恢复。

返回值类型：void

pg_xlog_replay_resume()

描述：如果恢复处于暂停状态，则重新启动。

当恢复暂停时，没有发生数据库更改。如果是在热备里，所有新的查询将看到一致的数据库快照，并且不会有进一步的查询冲突产生，直到恢复继续。

 版权所有 © 华为云计算技术有限公司

如果不能使用流复制，则暂停状态将无限的延续。当流复制正在进行时，将连续接收WAL记录，最终将填满可用磁盘空间，这个进度取决于暂停的持续时间，WAL生成的速度和可用的磁盘空间。

返回值类型：void

pg_xlog_replay_completion()

描述：显示当前DN上xlog redo的进度信息。

返回值类型：record

示例：

```
SELECT * FROM pg_xlog_replay_completion();
 replay_start | replay_current | replay_end | replay_percent
--------------+----------------+------------+----------------
 0/2ACAB80    | 0/2B16530      | 0/4F62B090 | 0%
(1 row)
```

函数返回信息如下：

表 7-66 pg_xlog_replay_completion()字段

名称	类型	描述
replay_start	integer	Xlog Redo的起始LSN
replay_current	integer	Xlog Redo的当前replay的LSN
replay_end	integer	最大的需要Xlog Redo的LSN
replay_percent	integer	Xlog Redo的当前完成的百分比

pg_data_sync_from_dummy_completion()

描述：显示当前DN上Failover过程中数据页文件同步的进度信息。

返回值类型：record

示例：

```
SELECT * FROM pg_data_sync_from_dummy_completion();
 start_index | current_index | total_index | sync_percent
-------------+---------------+-------------+--------------
           0 |             0 |           0 | 100%
(1 row)
```

函数返回信息如下：

表 7-67 pg_data_sync_from_dummy_completion()字段

名称	类型	描述
start_index	integer	数据页文件同步的起始编号

 版权所有 © 华为云计算技术有限公司

名称	类型	描述
current_index	integer	数据页文件同步的当前编号
total_index	integer	数据页文件同步的最大编号
sync_percent	integer	数据页文件当前完成的百分比

gs_roach_stop_backup(backupid text)

描述：停止一个内部备份工具GaussRoach开启的备份，返回当前日志的插入位置。与pg_stop_backup系统函数类似，但更轻量。

返回值类型：text

gs_roach_enable_delay_ddl_recycle(backupid name)

描述：开启延迟DDL功能，并返回开启点的日志位置。与pg_enable_delay_ddl_recycle系统函数类似，但更轻量。且通过传入不同的backupid，可支持并发打开延迟DDL。

返回值类型：text

gs_roach_disable_delay_ddl_recycle(backupid text)

描述：关闭延迟DDL功能，并返回本次延迟DDL生效的日志范围，并删除该范围内被用户删除的列存表物理文件。与pg_enable_delay_ddl_recycle系统函数类似，但更轻量。且通过传入不同的backupid，可支持并发关闭延迟DDL功能。

返回值类型：record

gs_roach_switch_xlog(request_ckpt bool)

描述：切换当前使用的日志段文件，返回切段日志的位置。且如果request_ckpt为true，则触发一个全量检查点。

返回值类型：text

pg_resume_bkp_flag(backupid name)

描述：用于备份恢复获取delay xlog标志，并返回start_backup_flag boolean，to_delay boolean，ddl_delay_recycle_ptr text，rewind_time text。

返回值类型：record

pg_update_reloptions_in_newcluster

描述：用于V3表集群级全量恢复到新集群时，将obs-store-path-list参数指定文件中存储的路径写入pg_class表的reloptions字段中。该函数仅9.1.0及以上集群版本支持。

返回值类型：bool

版权所有 © 华为云计算技术有限公司

7.29 数据库对象函数

7.29.1 数据库对象尺寸函数

数据库对象尺寸函数用于计算数据库对象使用的实际磁盘空间，其计算结果都返回以字节为单位的大小。

pg_column_size(any)

描述：存储一个指定的数值需要的字节数（可能压缩过）。

返回值类型：integer

备注：pg_column_size显示用于存储某个独立数据值的空间。

```
SELECT pg_column_size(1);
 pg_column_size
----------------
              4
(1 row)
```

pg_database_size(oid)

描述：指定OID代表的数据库使用的磁盘空间。

返回值类型：bigint

pg_database_size(name)

描述：指定名称的数据库使用的磁盘空间。

返回值类型：bigint

备注：pg_database_size接收一个数据库的OID或者名字，然后返回该对象使用的全部磁盘空间。

示例：

```
SELECT pg_database_size('gaussdb');
 pg_database_size
------------------
         51590112
(1 row)
```

pg_relation_size(oid)

描述：指定OID代表的表或者索引所使用的磁盘空间。

返回值类型：bigint

get_db_source_datasize()

描述：估算当前数据库非压缩态的数据总容量。

返回值类型：bigint

 版权所有 © 华为云计算技术有限公司

备注：1. 调用该函数前需要执行ANALYZE；2. 通过估算列存的压缩率计算非压缩态的数据总容量。

示例：

```
analyze;
ANALYZE
SELECT get_db_source_datasize();
 get_db_source_datasize
------------------------
            35384925667
(1 row)
```

pg_relation_size(text)

描述：指定名称的表或者索引使用的磁盘空间。表名字可以用模式名修饰。

返回值类型：bigint

pg_relation_size(relation regclass, fork text)

描述：指定表或索引的指定分叉树（ 'main'，'fsm'或'vm' ）使用的磁盘空间。

返回值类型：bigint

pg_relation_size(relation regclass)

描述：pg_relation_size(..., 'main')的简写。

返回值类型：bigint

备注：pg_relation_size接受一个表、索引、压缩表的OID或者名字，然后返回它们的字节大小。

pg_partition_size(oid,oid)

描述：指定OID代表的分区使用的磁盘空间。其中，第一个oid为表的OID，第二个oid为分区的OID。

返回值类型：bigint

pg_partition_size(text, text)

描述：指定名称的分区使用的磁盘空间。其中，第一个text为表名，第二个text为分区名。

返回值类型：bigint

pg_partitions_size(table, text[])

描述：获取指定分区数组使用的磁盘空间，返回各分区大小之和。该函数仅9.1.2及以上集群版本支持。

- table，表示表名，支持schema.table格式。
- text，表示分区名数组。

返回值类型：bigint

示例：

 版权所有 © 华为云计算技术有限公司

```
SELECT pg_partitions_size('public.tps_col', ARRAY['p1','p2','p3']::text[]);
```

pg_partition_indexes_size(oid,oid)

描述：指定OID代表的分区的索引使用的磁盘空间。其中，第一个oid为表的OID，第二个oid为分区的OID。

返回值类型：bigint

pg_partition_indexes_size(text,text)

描述：指定名称的分区的索引使用的磁盘空间。其中，第一个text为表名，第二个text为分区名。

返回值类型：bigint

pg_indexes_size(regclass)

描述：附加到指定表的索引使用的总磁盘空间。

返回值类型：bigint

pg_size_pretty(bigint)

描述：把字节计算的尺寸转换成一个易读的尺寸。

返回值类型：text

pg_size_pretty(numeric)

描述：把用数值表示的字节计算的尺寸转换成一个易读的尺寸。

返回值类型：text

备注：pg_size_pretty用于把其他函数的结果格式化成一种易读的格式，可以根据情况使用KB 、MB 、GB 、TB。

pg_table_size(regclass)

描述：指定的表使用的磁盘空间，不计索引（但是包含TOAST，自由空间映射和可见性映射）。

返回值类型：bigint

pg_total_relation_size(oid)

描述：指定OID代表的表使用的磁盘空间，包括索引和压缩数据。

返回值类型：bigint

pg_total_relation_size(regclass)

描述：指定的表使用的总磁盘空间，包括所有的索引和TOAST数据。

返回值类型：bigint

 版权所有 © 华为云计算技术有限公司

pg_total_relation_size(text)

描述：指定名字的表所使用的全部磁盘空间，包括索引和压缩数据。表名字可以用模式名修饰。

返回值类型：bigint

备注：pg_total_relation_size接受一个表或者一个压缩表的OID或者名称，然后返回以字节计的数据和所有相关的索引和压缩表的尺寸。

pg_obs_file_size(regclass)

描述：指定列存V3版本表OID或者表名字，获取其在OBS上存储的CU文件大小、文件名、bucket号。该函数仅9.1.0及以上集群版本支持。

9.1.1.210及以上集群版本中，查询分区表时返回字段会增加partname（分区名）和relfilenode（物理文件编号）列。

参数：入参可以是表的OID或者表名。

返回值类型：record

表 7-68 返回字段

名称	类型	描述
bucketid	integer	CU文件存储的bucket号。
filename	text	CU文件名。
size	bigint	CU文件大小，单位为字节。

示例：

```
SELECT * FROM pg_obs_file_size('t3_col_part_dif_partition');
 bucketid |      filename      | size
----------+--------------------+------
        9 | 23488119324673.0 |  512
       22 | 23488102760449.0 |  512
       21 | 23488119521281.0 |  512
       16 | 23488102662145.0 |  512
(4 rows)
```

pg_obs_file_size(text, text)

描述：指定列存V3版本表(主表名,分区表)的某个分区，返回其在OBS上存储的CU文件大小、文件名、bucket号。该函数仅9.1.0及以上集群版本支持。

9.1.1.210及以上集群版本中，查询分区表时返回字段会增加partname（分区名）和relfilenode（物理文件编号）列。

参数：入参可以是表的OID或者表名，分区名则直接使用分区名。

返回值类型：record

 版权所有 © 华为云计算技术有限公司

名称	类型	描述
bucketid	integer	CU文件存储的bucket号。
filename	text	CU文件名。
size	bigint	CU文件大小，单位为字节。

pg_foreign_table_size(text, text)

描述：查询指定外表对应文件的总大小，单位为字节（Byte）。其中，第一个text为schema，第二个text为表名。该函数仅9.1.1.100及以上集群版本支持。

返回值类型：bigint

备注：仅支持dfs_fdw或hdfs_fdw外表。

```
SELECT * FROM pg_foreign_table_size('public','customer_address2');
 pg_foreign_table_size
-----------------------
               1439489
(1 row)
```

7.29.2 数据库对象位置函数

pg_relation_filenode(relation regclass)

描述：指定关系的文件节点数。

返回值类型：oid

备注：pg_relation_filenode接收一个表、索引、序列或压缩表的OID或者名字，并且返回当前分配给它的“filenode”数。文件节点是关系使用的文件名字的基本组件。对大多数表来说，结果和pg_class.relfilenode相同，但对确定的系统目录来说，relfilenode为0而且这个函数必须用来获取正确的值。如果传递一个没有存储的关系，比如一个视图，那么这个函数返回NULL。

pg_relation_filepath(relation regclass)

描述：指定关系的文件路径名。

返回值类型：text

备注：pg_relation_filepath类似于pg_relation_filenode，但是它返回关系的整个文件路径名（相对于数据库集群的数据目录PGDATA）。

7.29.3 分区管理函数

proc_add_partition (relname regclass, boundaries_interval interval)

描述：用于给开启自动创建分区功能的表添加分区。

返回值类型：void

备注：该函数运行时，会在现有分区boundary的基础上，创建多个时间范围为boundaries_interval的新分区，直到new_part_boundary – now_time >= 29 *

 版权所有 © 华为云计算技术有限公司

boundaries_interval，之后再额外多创建一个分区，保证该函数运行时，一定会创建一个新分区 。

示例：

```
call proc_add_partition('my_schema.my_table', interval '1 day');
 proc_add_partition
--------------------

(1 row)
```

proc_drop_partition (relname regclass, older_than interval)

描述：用于给开启自动删除分区功能的表删除分区。

返回值类型：void

备注：该函数运行时，遍历表所有分区，并删除其中boundary小于(now_time - older_than)的分区；如果所有分区都满足删除条件，则保留一个分区，并truncate该表。

示例：

```
call proc_drop_partition('my_schema.my_table', interval '7 day');
 proc_drop_partition
--------------------

(1 row)
```

pg_partition_management_time(table_name, time)

描述：用来修改自增自减分区调度任务的执行时间，默认为创建自增分区表的时间。该函数仅8.3.0及以上版本支持。

返回值：SUCCESS，任务时间修改成功。

参数说明：

- table_name为自增分区表的表名，类型为name。
- time为数据切换任务调度时间，类型为Time。

示例：

```
call pg_partition_management_time('partition_management_time','3:00');
 pg_partition_management_time
------------------------------
 SUCCESS
(1 row)
```

7.29.4 排序规则版本函数

pg_collation_actual_version (oid)

描述：返回当前安装在操作系统中的该排序规则对象的实际版本，目前仅对case_insensitive有效。

返回值类型：text

示例：

```
SELECT oid FROM pg_collation WHERE collname ='case_insensitive';
 oid
```

 版权所有 © 华为云计算技术有限公司

```
------
 3300
(1 row)

SELECT pg_collation_actual_version(3300);
 pg_collation_actual_version
-----------------------------
 153.14
(1 row)
```

7.29.5 冷热表用户函数

pg_obs_cold_refresh_time(table_name, time)

描述：用来修改冷热表的冷数据切换至OBS上的时间，默认为每日0点。

table_name为冷热表的表名，类型为name，time为数据切换任务调度时间，类型为Time。

返回值：SUCCESS，任务时间修改成功。

示例：

```
SELECT * FROM pg_obs_cold_refresh_time('lifecycle_table', '06:30:00');
pg_obs_cold_refresh_time
--------------------------
 SUCCESS
(1 row)
```

pg_refresh_storage()

描述：切换所有冷热表，将符合冷热切换规则的数据由热数据切换至冷数据（OBS中）。

返回值类型：int

返回值字段：

- success_count int：切换成功的表个数
- failed_count int：切换失败的表个数

示例：

```
SELECT * FROM pg_refresh_storage();
 success_count | failed_count
---------------+--------------
             1 |            0
(1 row)
```

pg_lifecycle_table_data_distribute(table_name)

描述：查看某个冷热表的数据分布情况。

table_name为表名，不可缺省。

返回值：record

示例：根据节点数量形成多条记录，如下示例为只有一个dn节点时w1表数据分布情况。

```
SELECT * FROM pg_catalog.pg_lifecycle_table_data_distribute('w1');
 schemaname | tablename | nodename | hotpartition | coldpartition | switchablepartition | hotdatasize |
```

 版权所有 © 华为云计算技术有限公司

```
colddatasize | switchabledatasize
------------+-----------+----------+--------------+---------------+--------------------+-------------+--------------
+--------------------
 public     | w1        | dn_1     | p2           | p1            |                    | 80 KB       | 0 bytes      | 0 bytes
(1 row)
```

pg_lifecycle_node_data_distribute()

描述：查看所有冷热表数据分布情况。

返回值：record

示例：数据库中当前存在两个冷热表，其数据分布情况如下。

```
SELECT * FROM pg_catalog.pg_lifecycle_node_data_distribute();
 schemaname | tablename | nodename | hotpartition | coldpartition | switchablepartition | hotdatasize |
colddatasize | switchabledatasize
------------+-----------+----------+--------------+---------------+---------------------+-------------+--------------
+--------------------
 public     | w1        | dn_1     | p2           | p1            |                     |       81920 |            0 |                0
 public     | w2        | dn_1     | p2           | p1            |                     |       81920 |            0 |                0
(2 rows)
```

refresh_hot_storage(relname text)

描述：用于将指定冷热表的所有分区数据全部刷到OBS上。返回的数值为执行该函数切换分区的个数。该函数仅8.2.1.100及以上版本支持。

参数：

relname：表名(指定冷热表的表名，使用其他非冷热表的表名，并不会报错，返回值为0)

返回值类型：integer

```
SELECT refresh_hot_storage('multi_temper_table');
 refresh_hot_storage
---------------------
                   4
(1 row)
```

refresh_hot_storage(relname text, partname text)

描述：用于将指定冷热表的分区数据刷到OBS上。返回的数值为执行该函数切换分区的个数。该函数仅8.2.1.100及以上版本支持。

参数：

- relname：表名（指定冷热表的表名，使用其他非冷热表的表名，并不会报错，返回值为0）。
- partname：分区名（指定冷热表下的分区名）。

返回值类型：integer

```
SELECT refresh_hot_storage('multi_temper_table','p1');
 refresh_hot_storage
---------------------
                   1
(1 row)
```

 版权所有 © 华为云计算技术有限公司

reload_cold_partition(relname text)

描述：用于将指定冷热表的所有冷分区转换成热分区。返回的数值为执行该函数切换分区的个数。该函数仅8.3.0及以上版本支持。

参数：

relname：表名（指定冷热表的表名，使用其他非冷热表的表名，并不会报错，返回值为0）。

返回值类型：integer

```
SELECT reload_cold_partition('multi_temper_table');
 reload_cold_partition
----------------------
                    4
(1 row)
```

reload_cold_partition(relname text, partname text)

描述：用于将指定冷热表的特定冷分区转换成热分区。返回的数值为执行该函数切换分区的个数。该函数仅8.3.0及以上版本支持。

参数：

- relname：表名（指定冷热表的表名，使用其他非冷热表的表名，并不会报错，返回值为0）。
- partname：分区名（指定冷热表下的分区名）。

返回值类型：integer

```
SELECT reload_cold_partition('multi_temper_table','p1');
 reload_cold_partition
----------------------
                    1
(1 row)
```

7.30 残留文件管理函数

7.30.1 获取残留文件列表函数

pg_get_residualfiles()

描述：用于获取当前节点的所有残留文件记录。该函数为实例级函数，与当前所在的数据库无关，可以在任意实例上运行。

返回值类型：record

返回信息如下：

表 7-69 pg_get_residualfiles()返回字段

名称	类型	描述
isverified	bool	是否已经验证。

 版权所有 © 华为云计算技术有限公司

名称	类型	描述
isdeleted	bool	是否已经被删除。
dbname	text	所属数据库名称。
residualfile	text	数据文件路径。
filepath	text	残留文件记录路径。
notes	text	注释。

示例：

```
SELECT * FROM pg_get_residualfiles();
 isverified | isdeleted | dbname |   residualfile    |         filepath          | notes
------------+-----------+--------+-------------------+---------------------------+-------
 f          | f         | db2    | base/49155/114691 | pgrf_20200908160211441546 |
 f          | f         | db2    | base/49155/114694 | pgrf_20200908160211441546 |
 f          | f         | db2    | base/49155/114696 | pgrf_20200908160211441546 |
(3 rows)
```

pgxc_get_residualfiles()

描述：pg_get_residualfiles()的CN统一查询函数。该函数为集群级函数，与当前所在的数据库无关，在CN实例上运行。

返回值类型：record

返回信息如下：

表 7-70 pgxc_get_residualfiles()返回字段

名称	类型	描述
nodename	text	节点名称。
isverified	bool	是否已经验证。
isdeleted	bool	是否已经被删除。
dbname	text	所属数据库名称。
residualfile	text	数据文件路径。
filepath	text	残留文件记录路径。
notes	text	注释。

示例：

```
SELECT * FROM pgxc_get_residualfiles();
   nodename   | isverified | isdeleted |  dbname  |   residualfile    |         filepath          | notes
--------------+------------+-----------+----------+-------------------+---------------------------+-------
 cn_5001      | f          | f         | postgres | base/15092/32803  | pgrf_20200910170129360401 |
 dn_6001_6002 | f          | f         | db2      | base/49155/114691 | pgrf_20200908160211441546 |
 dn_6001_6002 | f          | f         | db2      | base/49155/114694 | pgrf_20200908160211441546 |
```


版权所有 © 华为云计算技术有限公司

```
 dn_6001_6002 | f          | f          | db2      | base/49155/114696 | pgrf_20200908160211441546 |
(4 rows)
```

7.30.2 扫描残留文件函数

扫描残留文件函数仅8.3.0及以上集群版本支持，本章节所有函数仅限sysadmin角色调用。

pg_scan_residualfiles()

描述：用于扫描当前节点所连接数据库的所有残留文件记录。该函数为库级函数，与当前所在的数据库相关，可以在任意实例上运行。

- 连接到CN执行时，扫描当前CN节点所在数据库的本地残留文件和OBS全部残留文件。
- 连接到DN执行时，扫描当前DN节点所在数据库的本地残留文件。

返回类型：record

返回信息如下：

表 7-71 返回字段

名称	类型	描述
pgscrf	text	记录残留文件信息的元文件本地路径。

示例：

```
SELECT * FROM pg_scan_residualfiles();
                                pgscrf
-----------------------------------------------------------------------
 pg_residualfiles/pgscrf_meta_1663_16323_20231027143716005244
 pg_residualfiles/pgscrf_meta_2147484281_16323_20231027143716005713
(2 rows)
```

说明

该函数的使用约束如下：

- 只能扫描当前登录的数据库中是否存在残留文件，当前只扫描默认表空间，自定义表空间和OBS表空间。
- 该函数为重载操作，不支持单节点并发使用，也不建议在业务繁忙和资源负载较高时使用。
- 只支持冷热表热分区，不支持冷分区。
- 禁止在升级观察期内使用。
- 无法对已删除数据库的残留文件进行扫描。
- 如果集群元数据信息存在不一致情况，该函数可能会错误扫描到有效的数据文件。

pgxc_scan_residualfiles(query_flag)

描述：pg_scan_residualfiles()的CN统一执行函数。该函数为集群级函数，与当前所在的数据库相关，在CN实例上运行。

 版权所有 © 华为云计算技术有限公司

参数说明：query_flag。参数类型int，表示执行范围。1表示CN，2表示主DN，4表示备DN，通过或运算可以得到查询并集，如1|2=3表示CN+主DN，1|2|4=7表示CN+主DN+备DN，默认值是7。

返回类型：record

返回信息如下：

表 7-72 返回字段

名称	类型	描述
nodename	name	节点名称。
instance_id	text	实例名称。
pgscrf	text	记录残留文件信息的元文件本地路径。

示例：

```
SELECT * FROM pgxc_scan_residualfiles();
  node_name   | instance_id |                                pgscrf
--------------+-------------+-----------------------------------------------------------------------
 dn_6001_6002 | dn_6001     | pg_residualfiles/pgscrf_meta_1663_16323_20231027144103839354
 dn_6001_6002 | dn_6001     | pg_residualfiles/pgscrf_meta_2147484281_16323_20231027144103839826
 cn_5001      | cn_5001     | pg_residualfiles/pgscrf_meta_1663_16323_20231027144103946217
 dn_6007_6008 | dn_6008     | pg_residualfiles/pgscrf_meta_1663_16323_20231027144104171311
(4 rows)

SELECT * FROM pgxc_scan_residualfiles(1);
 node_name | instance_id |                            pgscrf
-----------+-------------+-------------------------------------------------------------------
 cn_5001   | cn_5001     | pg_residualfiles/pgscrf_meta_1663_16323_20231027144421861628
(1 row)

SELECT * FROM pgxc_scan_residualfiles(2);
  node_name   | instance_id |                              pgscrf
--------------+-------------+-----------------------------------------------------------------------
 dn_6001_6002 | dn_6001     | pg_residualfiles/pgscrf_meta_1663_16323_20231027144424210395
 dn_6001_6002 | dn_6001     | pg_residualfiles/pgscrf_meta_2147484281_16323_20231027144424210855
(2 rows)

SELECT * FROM pgxc_scan_residualfiles(4);
  node_name   | instance_id |                            pgscrf
--------------+-------------+-------------------------------------------------------------------
 dn_6007_6008 | dn_6008     | pg_residualfiles/pgscrf_meta_1663_16323_20231027144427492060
(1 row)
```

说明

该函数的使用约束如下：

- 只能扫描当前登录的数据库中是否存在残留文件，当前只扫描默认表空间，自定义表空间和OBS表空间。
- 该函数为重载操作，不支持单节点并发使用，也不建议在业务繁忙和资源负载较高时使用。
- 只支持冷热表热分区，不支持冷分区。
- 禁止在升级观察期内使用。
- 无法对已删除数据库的残留文件进行扫描。
- 如果集群元数据信息存在不一致情况，该函数可能会错误扫描到有效的数据文件。

 版权所有 © 华为云计算技术有限公司

pg_get_scan_residualfiles()

描述：用于获取当前节点所有扫描到的残留文件记录。该函数为实例级函数，与当前所在的数据库无关，可以在任意实例上运行。

返回类型：record

返回信息如下：

表 7-73 返回字段

名称	类型	描述
handled	bool	残留文件是否已经被处理，被移动或者被更改。
dbname	text	数据库名称。
residualfile	text	残留文件路径。
size	bigint	文件大小，单位为byte。OBS路径的残留文件该项为0。
inode	bigint	残留文件stat信息中的Inode。OBS路径的残留文件该项为0。
atime	timestamptz	残留文件stat信息中的Access time。OBS路径的残留文件该项为空。
mtime	timestamptz	残留文件stat信息中的Modify time。OBS路径的残留文件该项为空。
ctime	timestamptz	残留文件stat信息中的Change time。OBS路径的残留文件该项为空。
filepath	text	记录残留文件信息的元文件本地路径，对应pgscrf_meta文件。
notes	text	注释。

pgxc_get_scan_residualfiles(query_flag)

描述：pg_get_scan_residualfiles()的CN统一执行函数。该函数为集群级函数，与当前所在的数据库无关，在CN实例上运行。

参数query_flag，参数类型int，用于表示执行范围。1表示CN，2表示主DN，4表示备DN，通过或运算可以得到查询并集，如1|2=3表示CN+主DN，1|2|4=7表示CN+主DN+备DN。默认值是7。

返回类型：record

返回信息如下：

 版权所有 © 华为云计算技术有限公司

表 7-74 返回字段

名称	类型	描述
nodename	name	节点名称。
instance_id	text	实例名称。
handled	bool	残留文件是否已经被处理，被移动或者被更改。
dbname	text	数据库名称。
residualfile	text	残留文件路径。
size	bigint	文件大小，单位为byte。OBS路径的残留文件该项为0。
inode	bigint	残留文件stat信息中的Inode。OBS路径的残留文件该项为0。
atime	timestamptz	残留文件stat信息中的Access time。OBS路径的残留文件该项为空。
mtime	timestamptz	残留文件stat信息中的Modify time。OBS路径的残留文件该项为空。
ctime	timestamptz	残留文件stat信息中的Change time。OBS路径的残留文件该项为空。
filepath	text	记录残留文件信息的元文件本地路径，对应pgscrf_meta文件。
notes	text	注释。

pg_archive_scan_residualfiles()

描述：用于归档当前节点所连接数据库上所有扫描到的残留文件记录。该函数为实例级函数，与当前所在的数据库相关，可以在任意实例上运行。

返回类型：record

返回信息如下：

表 7-75 返回字段

名称	类型	描述
archive	text	归档后的本地文件夹路径。OBS路径的残留文件归档在对应OBS数据库目录下。
count	bigint	归档的残留文件数量。
size	bigint	归档的残留文件总大小，单位为byte。

示例：

 版权所有 © 华为云计算技术有限公司

```
SELECT * FROM pg_archive_scan_residualfiles();
                          archive                          | count | size
-----------------------------------------------------------+-------+-------
 pg_residualfiles/archive/pgscrf_archive_20231027144613791801 |     4 | 81920
(1 row)
```

说明

- 此函数只能归档当前登录的数据库中记录过的残留文件，归档时会记录中的残留文件进行相应的校验，校验结果分为以下三种情况：
  - 校验通过：校验通过后会对残留文件进行归档，并标记为已处理。
  - 校验失败：校验失败会跳过归档，并标记为已处理。
  - 延迟校验：延迟校验会跳过归档，通常表示当前未满足校验时机，延迟校验的情况通常和事务完成情况及备机REDO进度有关。
- 实际归档目录与对应表空间在同一个文件系统下，删除表空间也会删除对应的残留文件归档。
- 延迟DDL开启后，此函数无法使用。
- 此函数为重载操作，不支持单节点并发使用，也不建议在业务繁忙和资源负载较高时使用。

pgxc_archive_scan_residualfiles(query_flag)

描述：pg_archive_scan_residualfiles()的CN统一执行函数。该函数为集群级函数，与当前所在的数据库相关，在CN实例上运行。

参数query_flag，参数类型int，用于表示执行范围。1表示CN，2表示主DN，4表示备DN，通过或运算可以得到查询并集，如1|2=3表示CN+主DN，1|2|4=7表示CN+主DN+备DN。默认值是7。

返回类型：record

返回信息如下：

表 7-76 返回字段

名称	类型	描述
nodename	name	节点名称。
instance_id	text	实例名称。
archive	text	归档后的本地文件夹路径。OBS路径的残留文件归档在对应OBS数据库目录下。
count	bigint	归档的残留文件数量。
size	bigint	归档的残留文件总大小，单位为byte。

示例：

```
SELECT * FROM pgxc_archive_scan_residualfiles();
  node_name   | instance_id |                          archive                          | count | size
--------------+-------------+-----------------------------------------------------------+-------+-------
 cn_5001      | cn_5001     | pg_residualfiles/archive/pgscrf_archive_20231027145050896440 |     1 | 40960
 dn_6007_6008 | dn_6008     | pg_residualfiles/archive/pgscrf_archive_20231027145051018138 |     1 | 24576
(2 rows)
```


版权所有 © 华为云计算技术有限公司

说明

- 此函数只能归档当前登录的数据库中记录过的残留文件，归档时会记录中的残留文件进行相应的校验，校验结果分为以下三种情况：
  - 校验通过：校验通过后会对残留文件进行归档，并标记为已处理。
  - 校验失败：校验失败会跳过归档，并标记为已处理。
  - 延迟校验：延迟校验会跳过归档，通常表示当前未满足校验时机，延迟校验的情况通常和事务完成情况及备机REDO进度有关。
- 实际归档目录与对应表空间在同一个文件系统下，删除表空间也会删除对应的残留文件归档。
- 延迟DDL开启后，此函数无法使用。
- 此函数为重载操作，不支持单节点并发使用，也不建议在业务繁忙和资源负载较高时使用。

pg_rm_scan_residualfiles_archive()

描述：用于删除当前节点所有归档的残留文件记录。该函数为实例级函数，与当前所在的数据库无关，可以在任意实例上运行。

返回类型：record

返回信息如下：

表 7-77 返回字段

名称	类型	描述
count	bigint	归档中已删除的残留文件数量。本地路径的残留文件统计所删除的文件数，OBS路径的残留文件统计所删除的表目录数量。
size	bigint	归档中已删除的残留文件总大小，单位为byte。OBS路径的残留文件该项均为0。

示例：

```
SELECT * FROM pg_rm_scan_residualfiles_archive();
 count | size
-------+-------
     4 | 81920
(1 row)
```

pgxc_rm_scan_residualfiles_archive(query_flag)

描述：pg_rm_scan_residualfiles_archive()的CN统一执行函数。该函数为集群级函数，与当前所在的数据库无关，在CN实例上运行。

参数query_flag，参数类型int，用于表示执行范围。1表示CN，2表示主DN，4表示备DN，通过或运算可以得到查询并集，如1|2=3表示CN+主DN，1|2|4=7表示CN+主DN+备DN。默认值是7。

返回类型：record

返回信息如下：

 版权所有 © 华为云计算技术有限公司

表 7-78 返回字段

名称	类型	描述
nodename	name	节点名称。
instance_id	text	实例名称。
count	bigint	归档中已删除的残留文件数量。本地路径的残留文件统计所删除的文件数，OBS路径的残留文件统计所删除的表目录数量。
size	bigint	归档中已删除的残留文件总大小，单位为byte。OBS路径的残留文件该项均为0。

示例：

```
SELECT * FROM pgxc_rm_scan_residualfiles_archive();
  node_name   | instance_id | count | size
--------------+-------------+-------+-------
 dn_6001_6002 | dn_6001     |     4 | 81920
 cn_5001      | cn_5001     |     1 | 40960
 dn_6007_6008 | dn_6008     |     1 | 24576
 coordinator1 | coordinator1|     1 |     0
(4 rows)
```

7.30.3 验证残留文件函数

pg_verify_residualfiles(filepath)

描述：用于验证参数指定文件中记录的文件是否为残留文件。该函数为实例级函数，与当前所在的数据库相关，可以在任意实例上运行。

参数类型：text

返回值类型：bool

函数返回字段如下：

表 7-79 pg_verify_residualfiles(filepath)返回字段

名称	类型	描述
isverified	bool	是否完成验证。

示例：

```
SELECT * FROM pg_verify_residualfiles('pgrf_20200908160211441546');
 isverified
------------
 t
(1 row)
```

说明

本函数只能验证记录的文件在当前登录的数据库中是否是残留文件。如果记录的文件不属于当前登录的数据库，则不会进行校验行为。

版权所有 © 华为云计算技术有限公司

pg_verify_residualfiles()

描述：用于验证当前实例上所有残留文件列表中记录的文件是否为残留文件。该函数为实例级函数，与当前所在的数据库相关，可以在任意实例上运行。

参数类型：无

返回值类型：record

函数返回字段如下：

表 7-80 pg_verify_residualfiles()返回字段

名称	类型	描述
result	bool	是否完成验证。
filepath	text	残留文件记录路径。
notes	text	注释。

示例：

```
SELECT * FROM pg_verify_residualfiles();
 result |          filepath           | notes
--------+-----------------------------+-------
 t      | pgrf_20200908160211441546 |
(1 row)
```

说明

本函数只能验证记录的文件在当前登录的数据库中是否是残留文件。如果记录的文件不属于当前登录的数据库，则不会进行校验行为。

pgxc_verify_residualfiles()

描述：pg_verify_residualfiles()的CN统一查询函数。该函数为集群级函数，与当前所在的数据库相关，在CN实例上运行。

参数类型：无

返回值类型：record

函数返回字段如下：

表 7-81 pgxc_verify_residualfiles()返回字段

名称	类型	描述
nodename	text	节点名称。
result	bool	是否完成验证。
filepath	text	残留文件记录路径。
notes	text	注释。

 版权所有 © 华为云计算技术有限公司

示例：

```
SELECT * FROM pgxc_verify_residualfiles();
   nodename   | result |          filepath           | notes
--------------+--------+-----------------------------+-------
 cn_5001      | t      | pgrf_20200910170129360401 |
 dn_6001_6002 | t      | pgrf_20200908160211441546 |
(2 rows)
```

说明

本函数只能验证记录的文件在当前登录的数据库中是否是残留文件。如果记录的文件不属于当前登录的数据库，则不会进行校验行为。

pg_is_residualfiles(residualfile)

描述：用于查询当前库中指定的relfilenode是否为残留文件。该函数为实例级函数，与当前所在的数据库相关，可以在任意实例上运行。

参数类型：text

返回值类型：bool

函数返回字段如下：

表 7-82 pg_is_residualfiles(residualfile)返回字段

名称	类型	描述
result	bool	是否是残留文件。

示例：

```
SELECT * FROM pg_is_residualfiles('base/49155/114691');
 result
--------
 t
(1 row)
```

说明

本函数只能验证记录的文件在当前登录的database中是否为残留文件。如果记录的文件不属于当前登录的数据库，则会被检测为是残留文件。

例如：针对gaussdb数据库中的非残留文件base/15092/14790，如果在gaussdb库中查询，则认为是非残留文件；在其他数据库中查询，则认为是残留文件。

```
SELECT * FROM pg_is_residualfiles('base/15092/14790');
result
--------
f
(1 row)

\c db2
db2=# SELECT * FROM pg_is_residualfiles('base/15092/14790');
result
--------
t
(1 row)
```

 版权所有 © 华为云计算技术有限公司

7.30.4 删除残留文件函数

pg_rm_residualfiles(filepath)

描述：用于删除当前实例中指定残留文件列表中的文件。该函数为实例级函数，与当前所在的数据库无关，可以在任意实例上运行。

参数类型：text

返回值类型：record

函数返回字段如下：

表 7-83 pg_rm_residualfiles(filepath)返回字段

名称	类型	描述
result	bool	是否已经完成删除。

示例：

```
SELECT * FROM pg_rm_residualfiles('pgrf_20200908160211441599');
 result
--------
 t
(1 row)
```

说明

- 残留文件只有在调用pg_verify_residualfiles()进行验证后才能被真正删除。
- 删除动作不区分数据库，指定文件中所有已经验证的文件都会被删除。
- 如果指定文件中记录的所有文件都已经被删除，指定文件会被移除并备份到$PGDATA/pg_residualfile/backup目录下。

pg_rm_residualfiles()

描述：用于删除当前实例中所有的残留文件列表中的文件。该函数为实例级函数，与当前所在的数据库无关，可以在任意实例上运行。

参数类型：无

返回值类型：record

函数返回字段如下：

表 7-84 pg_rm_residualfiles()返回字段

名称	类型	描述
result	bool	是否已经完成删除。
filepath	text	残留文件记录路径。
notes	text	注释。

 版权所有 © 华为云计算技术有限公司

示例：

```
SELECT * FROM pg_rm_residualfiles();
 result |          filepath          | notes
--------+----------------------------+-------
 t      | pgrf_20200908160211441546 |
(1 row)
```

说明

- 残留文件只有在调用pg_verify_residualfiles()进行验证后才能被真正删除。
- 删除动作不区分数据库，指定文件中所有已经验证的文件都会被删除。
- 如果指定文件中记录的所有文件都已经被删除，指定文件会被移除并备份到$PGDATA/pg_residualfile/backup目录下。

pgxc_rm_residualfiles()

描述：pgxc_rm_residualfiles的CN统一查询函数。该函数为集群级函数，与当前所在的数据库无关，在CN实例上运行。

参数类型：无

返回值类型：record

函数返回字段如下：

表 7-85 pgxc_rm_residualfiles()返回字段

名称	类型	描述
nodename	text	节点名。
result	bool	是否已经完成删除。
filepath	text	残留文件记录路径。
notes	text	注释。

示例：

```
SELECT * FROM pgxc_rm_residualfiles();
   nodename   | result |          filepath          | notes
--------------+--------+----------------------------+-------
 cn_5001      | t      | pgrf_20200910170129360401 |
 dn_6001_6002 | t      | pgrf_20200908160211441546 |
(2 rows)
```

pgxc_clear_disk_cache()

描述：用于完全清理磁盘缓存文件。该函数仅9.0.2及以上集群版本支持。

返回值类型：void

示例：

```
SELECT pgxc_clear_disk_cache();
 pgxc_clear_disk_cache
-----------------------


(1 row)
```


版权所有 © 华为云计算技术有限公司

pgxc_clear_ttl_cache(clearAll)

描述：用于清理TTL队列使用的磁盘缓存文件。该函数仅9.1.1.100及以上集群版本支持。

- 入参为true时，清理整个TTL队列占用的磁盘缓存文件。
- 入参为false时，仅清理TTL当前已经失效的磁盘缓存文件。

返回值类型：void

示例：

```
SELECT pgxc_clear_ttl_cache(true);
 pgxc_clear_ttl_cache
----------------------

(1 row)

SELECT pgxc_clear_ttl_cache(false);
 pgxc_clear_ttl_cache
----------------------

(1 row)
```

7.30.5 残留文件管理函数应用

使用步骤

步骤1 调用pgxc_get_residualfiles()函数，获取存在残留文件的数据库名称。

步骤2 分别进入确认有残留文件的数据库，调用pgxc_verify_residualfiles()函数，对当前数据库中记录的残留文件进行验证。

步骤3 调用pgxc_rm_residualfiles()函数，删除所有已经验证过的残留文件。

----结束

说明

pgxc类残留文件管理函数只对CN和当前主DN进行操作，不会验证和清理备DN上的残留文件。所以主DN完成清理后，应在备DN上及时执行残留文件清理操作或对备机进行build，防止主备切换后由于增量build导致备机残留文件被重新复制回主DN，导致未成功清理的假象。

使用示例

以当前两个用户自建的数据库db1、db2为例：

```
postgres=# \l
                                 List of databases
   Name    | Owner  | Encoding  | Collate | Ctype | Access privileges
-----------+--------+-----------+---------+-------+-------------------
 db1       | fpi810 | SQL_ASCII | C       | C     |
 db2       | fpi810 | SQL_ASCII | C       | C     |
 postgres  | fpi810 | SQL_ASCII | C       | C     |
 template0 | fpi810 | SQL_ASCII | C       | C     | =c/fpi810        +
           |        |           |         |       | fpi810=CTc/fpi810
 template1 | fpi810 | SQL_ASCII | C       | C     | =c/fpi810        +
           |        |           |         |       | fpi810=CTc/fpi810
(5 rows)
```

 版权所有 © 华为云计算技术有限公司

1. 在CN上获取集群的所有残留文件记录：

```
SELECT * FROM pgxc_get_residualfiles() order by 4, 6; -- order by不是必须的
```

```
db1=# select * from pgxc_get_residualfiles() order by 4, 6;
   nodename    | isverified | isdeleted | dbname |   residualfile   |         filepath          | notes
---------------+------------+-----------+--------+------------------+---------------------------+-------
 dn_6001_6002 | f          | f         | db1    | base/16390/16395 | pgrf_20200921153355979438 |
 dn_6001_6002 | f          | f         | db1    | base/16390/24592 | pgrf_20200921153612088342 |
 dn_6001_6002 | f          | f         | db2    | base/16391/24579 | pgrf_20200921153612088342 |
 dn_6001_6002 | f          | f         | db2    | base/16391/24582 | pgrf_20200921153612088342 |
 dn_6001_6002 | f          | f         | db2    | base/16391/24584 | pgrf_20200921153612088342 |
 dn_6001_6002 | f          | f         | db2    | base/16391/24585 | pgrf_20200921153612088342 |
 dn_6001_6002 | f          | f         | db2    | base/16391/24588 | pgrf_20200921153612088342 |
 dn_6001_6002 | f          | f         | db2    | base/16391/24589 | pgrf_20200921153612088342 |
 dn_6001_6002 | f          | f         | db2    | base/16391/24591 | pgrf_20200921153612088342 |
 dn_6001_6002 | f          | f         | db2    | base/16391/24576 | pgrf_20200921153612088342 |
(10 rows)
```

当前集群中：

- dn_6001_6002 节点（当前的主节点实例）的db1和db2数据库中都存在残留文件记录。
- 残留文件在residualfile列展示。
- filepath列为记录残留文件的记录文件，保存在实例数据目录下pg_residualfiles目录中。

2. 调用pgxc_verify_residualfiles()函数对db1库进行验证：

```
SELECT * FROM pgxc_verify_residualfiles();
```

```
postgres=# \c db1
Non-SSL connection (SSL connection is recommended when requiring high-security)
You are now connected to database "db1" as user "fpi810".
db1=# select * from pgxc_verify_residualfiles();
   nodename    | result |         filepath          | notes
---------------+--------+---------------------------+-------
 dn_6001_6002 | t      | pgrf_20200921153355979438 |
 dn_6001_6002 | t      | pgrf_20200921153612088342 |
(2 rows)
```

因为**verify类函数都是database级别**，所以当前在db1库中调用verify函数时，只对属于db1的残留文件进行验证。

可以再次调用get函数查看是否验证完成：

```
SELECT * FROM pgxc_get_residualfiles() order by 4, 6;
```

```
db1=# select * from pgxc_get_residualfiles() order by 4, 6;
   nodename    | isverified | isdeleted | dbname |   residualfile   |         filepath          | notes
---------------+------------+-----------+--------+------------------+---------------------------+-------
 dn_6001_6002 | t          | f         | db1    | base/16390/16395 | pgrf_20200921153355979438 |
 dn_6001_6002 | t          | f         | db1    | base/16390/24592 | pgrf_20200921153612088342 |
 dn_6001_6002 | f          | f         | db2    | base/16391/24579 | pgrf_20200921153612088342 |
 dn_6001_6002 | f          | f         | db2    | base/16391/24582 | pgrf_20200921153612088342 |
 dn_6001_6002 | f          | f         | db2    | base/16391/24584 | pgrf_20200921153612088342 |
 dn_6001_6002 | f          | f         | db2    | base/16391/24585 | pgrf_20200921153612088342 |
 dn_6001_6002 | f          | f         | db2    | base/16391/24588 | pgrf_20200921153612088342 |
 dn_6001_6002 | f          | f         | db2    | base/16391/24589 | pgrf_20200921153612088342 |
 dn_6001_6002 | f          | f         | db2    | base/16391/24591 | pgrf_20200921153612088342 |
 dn_6001_6002 | f          | f         | db2    | base/16391/24576 | pgrf_20200921153612088342 |
(10 rows)
```

如上图所示，已确认db1数据库中的残留文件都已经验证，db2数据库中的残留文件都未进行验证。

3. 调用 pgxc_rm_residualfiles()函数删除残留文件。

```
SELECT * FROM pgxc_rm_residualfiles();
```

```
db1=# select * from pgxc_rm_residualfiles();
   nodename    | result |         filepath          | notes
---------------+--------+---------------------------+-------
 dn_6001_6002 | t      | pgrf_20200921153355979438 |
 dn_6001_6002 | t      | pgrf_20200921153612088342 |
(2 rows)
```


版权所有 © 华为云计算技术有限公司

4. 再次调用pgxc_get_residualfiles()函数检查删除后的结果。

```
SELECT * FROM pgxc_get_residualfiles() order by 4, 6;
```

```
db1=# select * from pgxc_get_residualfiles() order by 4, 6;
   nodename    | isverified | isdeleted | dbname |   residualfile   |         filepath          | notes
---------------+------------+-----------+--------+------------------+---------------------------+-------
 dn_6001_6002 | t          | t         | db1    | base/16390/24592 | pgrf_20200921153612088342 |
 dn_6001_6002 | f          | f         | db2    | base/16391/24579 | pgrf_20200921153612088342 |
 dn_6001_6002 | f          | f         | db2    | base/16391/24582 | pgrf_20200921153612088342 |
 dn_6001_6002 | f          | f         | db2    | base/16391/24584 | pgrf_20200921153612088342 |
 dn_6001_6002 | f          | f         | db2    | base/16391/24576 | pgrf_20200921153612088342 |
 dn_6001_6002 | f          | f         | db2    | base/16391/24588 | pgrf_20200921153612088342 |
 dn_6001_6002 | f          | f         | db2    | base/16391/24589 | pgrf_20200921153612088342 |
 dn_6001_6002 | f          | f         | db2    | base/16391/24591 | pgrf_20200921153612088342 |
 dn_6001_6002 | f          | f         | db2    | base/16391/24585 | pgrf_20200921153612088342 |
(9 rows)
```

结果显示db1数据库中的残留文件已经被删除（isdeleted标记为t），db2中的残留文件都未被删除。

同时可以看到查询出9条结果，与之前查询出的结果相比，缺少一条以9438结尾的残留文件记录文件。这是因为以9438结尾的残留文件记录文件中只有一条残留文件记录，这条记录在步骤3中被删除，当记录文件中的所有残留文件都被删除后，记录文件本身也会被删除，并备份到pg_residualfiles/backup目录中：

```
[fpi810@host-192-168-244-162 pg_residualfiles]$ ls
backup  pendingdeletesfile  pgrf_20200921153612088342
[fpi810@host-192-168-244-162 pg_residualfiles]$ ls backup/
pgrf_20200921153355979438_bak
```

5. 如果需要删除db2数据库中的文件，需要在db2中调用verify函数后再调用rm函数。

   a. 进入db2数据库，并调用验证函数：

   ```
   SELECT * FROM pgxc_verify_residualfiles();
   ```

   ```
   db1=# \c db2
   Non-SSL connection (SSL connection is recommended when requiring high-security)
   You are now connected to database "db2" as user "fpi810".
   db2=# select * from pgxc_verify_residualfiles();
      nodename    | result |         filepath          | notes
   ---------------+--------+---------------------------+-------
    dn_6001_6002 | t      | pgrf_20200921153612088342 |
   (1 row)
   ```

   此时可以查询验证的结果：

   ```
   SELECT * FROM pgxc_get_residualfiles() order by 4, 6;
   ```

   ```
   db2=# select * from pgxc_get_residualfiles() order by 4, 6;
      nodename    | isverified | isdeleted | dbname |   residualfile   |         filepath          | notes
   ---------------+------------+-----------+--------+------------------+---------------------------+-------
    dn_6001_6002 | t          | t         | db1    | base/16390/24592 | pgrf_20200921153612088342 |
    dn_6001_6002 | t          | f         | db2    | base/16391/24579 | pgrf_20200921153612088342 |
    dn_6001_6002 | t          | f         | db2    | base/16391/24582 | pgrf_20200921153612088342 |
    dn_6001_6002 | t          | f         | db2    | base/16391/24584 | pgrf_20200921153612088342 |
    dn_6001_6002 | t          | f         | db2    | base/16391/24576 | pgrf_20200921153612088342 |
    dn_6001_6002 | t          | f         | db2    | base/16391/24588 | pgrf_20200921153612088342 |
    dn_6001_6002 | t          | f         | db2    | base/16391/24589 | pgrf_20200921153612088342 |
    dn_6001_6002 | t          | f         | db2    | base/16391/24591 | pgrf_20200921153612088342 |
    dn_6001_6002 | t          | f         | db2    | base/16391/24585 | pgrf_20200921153612088342 |
   (9 rows)
   ```

   b. 调用删除函数：

   ```
   SELECT * FROM pgxc_rm_residualfiles();
   ```

   ```
   db2=# select * from pgxc_rm_residualfiles();
      nodename    | result |         filepath          | notes
   ---------------+--------+---------------------------+-------
    dn_6001_6002 | t      | pgrf_20200921153612088342 |
   (1 row)
   ```


版权所有 © 华为云计算技术有限公司

c. 再查询删除的结果：

```
SELECT * FROM pgxc_get_residualfiles() order by 4, 6;
```

```
db2=# select * from pgxc_get_residualfiles() order by 4, 6;
 nodename | isverified | isdeleted | dbname | residualfile | filepath | notes
----------+------------+-----------+--------+--------------+----------+-------
(0 rows)
```

此时因为 8342 结尾的记录文件中残留文件已经全部删除，所以整个记录文件也被删除并备份到backup目录下，所以查询到0条记录。

```
[fpi810@host-192-168-244-162 pg_residualfiles]$ ls
backup  pendingdeletesfile
[fpi810@host-192-168-244-162 pg_residualfiles]$ ls backup/
pgrf_20200921153355979438_bak  pgrf_20200921153612088342_bak
```

7.31 统计信息函数

统计信息函数根据访问对象分为两种类型：针对某个数据库进行访问的函数，以数据库中每个表或索引的OID作为参数，标识需要报告的数据库；针对某个服务器进行访问的函数，以一个服务器进程号为参数，其范围从1到当前活跃服务器的数目。

对于全局临时表，函数入参是表在pg_class中的OID，函数返回当前会话数据的统计信息，所以同一张全局临时表在不同会话查询下列统计信息函数得到的结果可能不同。

pg_stat_get_db_numbackends(oid)

描述：查询当前实例上指定数据库活跃的服务器线程数目。

返回值类型：integer

pg_stat_get_db_total_numbackends(oid)

描述：在CN上执行该函数，返回集群中所有CN上指定数据库活跃的服务器线程总数。在DN上执行该函数，返回当前实例上指定数据库活跃的服务器线程数目。

返回值类型：integer

pg_stat_get_db_xact_commit(oid)

描述：返回当前实例上指定数据库中已提交事务的数量。

返回值类型：bigint

pg_stat_get_db_total_xact_commit(oid)

描述：在CN上执行该函数，返回集群中所有CN上指定数据库中已提交事务的总数。在DN上执行该函数，返回当前实例上指定数据库中已提交事务的数量。

返回值类型：bigint

pg_stat_get_db_xact_rollback(oid)

描述：返回当前实例上指定数据库中回滚事务的数量。

返回值类型：bigint

 版权所有 © 华为云计算技术有限公司

pg_stat_get_db_total_xact_rollback(oid)

描述：在CN上执行该函数，返回集群中所有CN上指定数据库中回滚事务的总数。在DN上执行该函数，返回当前实例上指定数据库中回滚事务的数量。

返回值类型：bigint

pg_stat_get_db_blocks_fetched(oid)

描述：返回当前实例上指定数据库中磁盘块抓取请求的数量。

返回值类型：bigint

pg_stat_get_db_total_blocks_fetched(oid)

描述：在CN上执行该函数，返回集群中所有DN上指定数据库中磁盘块抓取请求的总数。在DN上执行该函数，返回当前实例上指定数据库中磁盘块抓取请求的数量。

返回值类型：bigint

pg_stat_get_db_blocks_hit(oid)

描述：返回当前实例上指定数据库在缓冲区中找到的请求磁盘块的数量。

返回值类型：bigint

pg_stat_get_db_total_blocks_hit(oid)

描述：在CN上执行该函数，返回集群中所有DN上指定数据库在缓冲区中找到的请求磁盘块的总数。在DN上执行该函数，返回当前实例上指定数据库在缓冲区中找到的请求磁盘块的数量。

返回值类型：bigint

pg_stat_get_db_tuples_returned(oid)

描述：返回当前实例上指定数据库返回的元组数量。

返回值类型：bigint

pg_stat_get_db_total_tuples_returned(oid)

描述：在CN上执行该函数，返回集群中所有DN上指定数据库返回的元组总数。在DN上执行该函数，返回当前实例上指定数据库返回的元组数量。

返回值类型：bigint

pg_stat_get_db_tuples_fetched(oid)

描述：返回当前实例上指定数据库中读取的元组数量。

返回值类型：bigint

 版权所有 © 华为云计算技术有限公司

pg_stat_get_db_total_tuples_fetched(oid)

描述：在CN上执行该函数，返回集群中所有DN上指定数据库读取的元组总数。在DN上执行该函数，返回当前实例上指定数据库读取的元组数量。

返回值类型：bigint

pg_stat_get_db_tuples_inserted(oid)

描述：返回当前实例上指定数据库中插入的元组数量。

返回值类型：bigint

pg_stat_get_db_total_tuples_inserted(oid)

描述：在CN上执行该函数，返回集群中所有DN上指定数据库插入的元组总数。在DN上执行该函数，返回当前实例上指定数据库插入的元组数量。

返回值类型：bigint

pg_stat_get_db_tuples_updated(oid)

描述：返回当前实例上指定数据库中更新的元组数量。

返回值类型：bigint

pg_stat_get_db_total_tuples_updated(oid)

描述：在CN上执行该函数，返回集群中所有DN上指定数据库更新的元组总数。在DN上执行该函数，返回当前实例上指定数据库更新的元组数量。

返回值类型：bigint

pg_stat_get_db_tuples_deleted(oid)

描述：返回当前实例上指定数据库中删除的元组数量。

返回值类型：bigint

pg_stat_get_db_total_tuples_deleted(oid)

描述：在CN上执行该函数，返回集群中所有DN上指定数据库删除的元组总数。在DN上执行该函数，返回当前实例上指定数据库删除的元组数量。

返回值类型：bigint

pg_stat_get_db_conflict_lock(oid)

描述：在CN上执行该函数，返回集群中所有CN和DN上指定数据库锁冲突的总数。在DN上执行该函数，返回当前实例上指定数据库中锁冲突数量。

返回值类型：bigint

pg_stat_get_db_deadlocks(oid)

描述：返回当前实例上指定数据库中死锁的数量。

 版权所有 © 华为云计算技术有限公司

返回值类型：bigint

pg_stat_get_db_total_deadlocks(oid)

描述：在CN上执行该函数，返回集群中所有CN和DN上指定数据库死锁的总数。在DN上执行该函数，返回当前实例上指定数据库中死锁的数量。

返回值类型：bigint

pg_stat_get_db_conflict_all(oid)

描述：返回当前实例上指定数据库中发生冲突恢复的次数。

返回值类型：bigint

pg_stat_get_db_total_conflict_all(oid)

描述：在CN上执行该函数，返回集群中所有CN和DN上指定数据库发生冲突恢复的总次数。在DN上执行该函数，返回当前实例上指定数据库中发生冲突恢复的次数。

返回值类型：bigint

pg_stat_get_db_temp_files(oid)

描述：返回当前实例上指定数据库中创建临时文件的个数。

返回值类型：bigint

pg_stat_get_db_total_temp_files(oid)

描述：在CN上执行该函数，返回集群中所有DN上指定数据库中创建临时文件的总个数。在DN上执行该函数，返回当前实例上指定数据库中创建临时文件的个数。

返回值类型：bigint

pg_stat_get_db_temp_bytes(oid)

描述：返回当前实例上指定数据库中创建临时文件的字节数。

返回值类型：bigint

pg_stat_get_db_total_temp_bytes(oid)

描述：在CN上执行该函数，返回集群中所有DN上指定数据库中创建临时文件的总字节数。在DN上执行该函数，返回当前实例上指定数据库中创建临时文件的字节数。

返回值类型：bigint

pg_stat_get_db_blk_read_time(oid)

描述：返回当前实例上指定数据库中读数据块所用的时间。

返回值类型：double

 版权所有 © 华为云计算技术有限公司

pg_stat_get_db_total_blk_read_time(oid)

描述：在CN上执行该函数，返回集群中所有DN上指定数据库中读数据块所用的总时间。在DN上执行该函数，返回当前实例上指定数据库中读数据块所用的时间。

返回值类型：double

pg_stat_get_db_blk_write_time(oid)

描述：返回当前实例上指定数据库中写数据块所用的时间。

返回值类型：double

pg_stat_get_db_total_blk_write_time(oid)

描述：在CN上执行该函数，返回集群中所有DN上指定数据库中写数据块所用的总时间。在DN上执行该函数，返回当前实例上指定数据库中写数据块所用的时间。

返回值类型：double

pg_stat_get_numscans(oid)

描述：如果参数是一个表，则顺序扫描读取的行数目。

如果参数是一个索引，则返回索引行的数目。

返回值类型：bigint

pg_stat_get_tuple()

描述：在CN和DN上均可以执行，该函数仅8.1.3及以上集群版本支持。

函数无参时，查询CN上所有系统表的统计信息及表在每个CN上的脏页率，查询DN上所有系统表和用户表的统计信息和表在每个DN上的脏页率；

函数带入参时，入参是schema和表名，带入参的函数执行时查询单张表的统计信息和脏页率。

说明

该函数的统计信息依赖于ANALYZE，为获取最准确的信息请先对表进行ANALYZE。

返回值类型：record

函数返回字段如下：

名称	类型	描述
nodename	text	节点名。
tableid	oid	表的OID。
partid	oid	分区表的分区OID。
last_vacuum	timestamp with time zone	最后一次手动vacuum时间。

 版权所有 © 华为云计算技术有限公司

名称	类型	描述
last_autovacuum	timestamp with time zone	最后一次autovacuum时间。
last_analyze	timestamp with time zone	最后一次手动analyze时间。
last_autoanalyze	timestamp with time zone	最后一次autoanalyze时间。
vacuum_count	bigint	vacuum次数。
autovacuum_count	bigint	autovacuum次数。
analyze_count	bigint	analyze次数。
autoanalyze_count	bigint	自动分析的次数。
n_tup_ins	bigint	插入的行数。
n_tup_upd	bigint	更新的行数。
n_tup_del	bigint	删除的行数。
n_tup_hot_upd	bigint	HOT更新的行数。
n_tup_change	bigint	analyze之后改变的行数。
n_live_tup	bigint	live行估计数。
n_dead_tup	bigint	dead行估计数。
dirty_rate	bigint	单节点的脏页率（单CN或单DN级）。
last_data_changed	timestamp with time zone	记录表最后一次数据变化的时间。该字段在8.2.0及以上版本固定为空，建议通过**pg_stat_get_last_data_changed_time(oid)**查询。

pg_stat_get_tuples_returned(oid)

描述：如果参数是一个表，则顺序扫描读取的行数目。

如果参数是一个索引，则返回的索引行的数目。

返回值类型：bigint

pg_stat_get_tuples_fetched(oid)

描述：如果参数是一个表，则位图扫描抓取的行数目。

如果参数是一个索引，则用简单索引扫描抓取的行数目。

返回值类型：bigint

版权所有 © 华为云计算技术有限公司

pg_stat_get_tuples_inserted(oid)

描述：插入表中行的数量。

返回值类型：bigint

pg_stat_get_local_tuples_inserted(oid)

描述：当前节点上插入表中行的数量，该函数仅8.1.2及以上版本支持。

返回值类型：bigint

pg_stat_get_tuples_updated(oid)

描述：在表中已更新行的数量。

返回值类型：bigint

pg_stat_get_local_tuples_updated(oid)

描述：当前节点上在表中已更新行的数量，该函数仅8.1.2及以上版本支持。

返回值类型：bigint

pg_stat_get_tuples_deleted(oid)

描述：从表中删除行的数量。

返回值类型：bigint

pg_stat_get_local_tuples_deleted(oid)

描述：当前节点上从表中删除行的数量，该函数仅8.1.2及以上版本支持。

返回值类型：bigint

pg_stat_get_tuples_changed(oid)

描述：在CN上查询该函数，返回该表自上一次analyze或autoanalyze之后插入、更新、删除行的总数量。在DN上查询该函数，返回该表在当前节点上自上一次analyze或autoanalyze之后插入、更新、删除行的总数量。

返回值类型：bigint

pg_stat_get_local_tuples_changed(oid)

描述：该表在当前节点上自从上一次analyze或autoanalyze之后插入、更新、删除行的数量。

返回值类型：bigint

pg_stat_get_tuples_hot_updated(oid)

描述：表热更新的行数。

返回值类型：bigint

 版权所有 © 华为云计算技术有限公司

pg_stat_get_local_tuples_hot_updated(oid)

描述：当前节点上表热更新的行数，该函数仅8.1.2及以上版本支持。

返回值类型：bigint

pg_stat_get_live_tuples(oid)

描述：表格的活元组数。

返回值类型：bigint

pg_stat_get_local_live_tuples(oid)

描述：当前节点上表格的活元组数，该函数仅8.1.2及以上版本支持。

返回值类型：bigint

pg_stat_get_dead_tuples(oid)

描述：表格的死元组数。

返回值类型：bigint

pg_stat_get_local_dead_tuples(oid)

描述：当前节点上表格的死元组数，该函数仅8.1.2及以上版本支持。

返回值类型：bigint

pg_stat_get_blocks_fetched(oid)

描述：表或者索引的磁盘块抓取请求的数量。

返回值类型：bigint

pg_stat_get_blocks_hit(oid)

描述：在缓冲区中找到的表或者索引的磁盘块请求数目。

返回值类型：bigint

pg_stat_get_partition_tuples_inserted(oid)

描述：插入相应表分区中行的数量。

返回值类型：bigint

pg_stat_get_partition_tuples_updated(oid)

描述：在相应表分区中已更新行的数量。

返回值类型：bigint

pg_stat_get_partition_tuples_deleted(oid)

描述：从相应表分区中删除行的数量。

 版权所有 © 华为云计算技术有限公司

返回值类型：bigint

pg_stat_get_partition_tuples_changed(oid)

描述：该表分区上一次analyze或autoanalyze之后插入、更新、删除行的总数量。

返回值类型：bigint

pg_stat_get_partition_live_tuples(oid)

描述：表分区的活行数。

返回值类型：bigint

pg_stat_get_partition_dead_tuples(oid)

描述：表分区的死行数。

返回值类型：bigint

pg_stat_get_xact_tuples_inserted(oid)

描述：表相关的活跃子事务中插入的tuple数。

返回值类型：bigint

pg_stat_get_xact_tuples_deleted(oid)

描述：表相关的活跃子事务中删除的tuple数。

返回值类型：bigint

pg_stat_get_xact_tuples_hot_updated(oid)

描述：表相关的活跃子事务中热更新的tuple数。

返回值类型：bigint

pg_stat_get_xact_tuples_updated(oid)

描述：表相关的活跃子事务中更新的tuple数。

返回值类型：bigint

pg_stat_get_xact_partition_tuples_inserted(oid)

描述：表分区相关的活跃子事务中插入的tuple数。

返回值类型：bigint

pg_stat_get_xact_partition_tuples_deleted(oid)

描述：表分区相关的活跃子事务中删除的tuple数。

返回值类型：bigint

 版权所有 © 华为云计算技术有限公司

pg_stat_get_xact_partition_tuples_hot_updated(oid)

描述：表分区相关的活跃子事务中热更新的tuple数。

返回值类型：bigint

pg_stat_get_xact_partition_tuples_updated(oid)

描述：表分区相关的活跃子事务中更新的tuple数。

返回值类型：bigint

pg_stat_get_last_vacuum_time(oid)

描述：用户在该表上最后一次手动启动清理或者autovacuum线程启动清理的时间。

返回值类型：timestamptz

pg_stat_get_last_autovacuum_time(oid)

描述：autovacuum守护线程在该表上最后一次启动清理的时间。

返回值类型：timestamptz

pg_stat_get_local_last_autovacuum_time(oid)

描述：当前节点的autovacuum守护线程在该表上最后一次启动清理的时间，该函数仅8.1.2及以上版本支持。

返回值类型：timestamptz

pg_stat_get_vacuum_count(oid)

描述：用户在该表上手动启动清理的次数。

返回值类型：bigint

pg_stat_get_autovacuum_count(oid)

描述：autovacuum守护线程在该表上启动清理的次数。

返回值类型：bigint

pg_stat_get_local_autovacuum_count(oid)

描述：当前节点上的autovacuum守护线程在该表上启动清理的次数，该函数仅8.1.2及以上版本支持。

返回值类型：bigint

pg_stat_get_last_analyze_time(oid)

描述：用户在该表上最后一次手动启动分析或者autovacuum线程启动分析的时间。

返回值类型：timestamptz

 版权所有 © 华为云计算技术有限公司

pg_stat_get_last_autoanalyze_time(oid)

描述：autovacuum守护线程在该表上最后一次启动分析的时间。

返回值类型：timestamptz

pg_stat_get_local_last_autoanalyze_time(oid)

描述：当前节点的autovacuum守护线程在该表上最后一次启动分析的时间，该函数仅8.1.2及以上版本支持。

返回值类型：timestamptz

pg_stat_get_analyze_count(oid)

描述：用户在该表上手动启动分析的次数。

返回值类型：bigint

pg_stat_get_autoanalyze_count(oid)

描述：autovacuum守护线程在该表上启动分析的次数。

返回值类型：bigint

pg_stat_get_local_autoanalyze_count(oid)

描述：当前节点的autovacuum守护线程在该表上启动分析的次数，该函数仅8.1.2及以上版本支持。

返回值类型：bigint

pg_stat_get_local_analyze_status(oid)

描述：指定表在当前节点上的是否需要analyze的状态，仅在CN端有意义。该函数仅8.1.2及以上版本支持。

- 如果该表的修改行数超过analyze的阈值（根据autovacuum_analyze_threshold + autovacuum_analyze_scale_factor * reltuples计算，其中reltuples是pg_class中记录的表的估算行数），则返回“Analyze needed”。
- 如果该表的修改行数不超过analyze的阈值，则返回“Analyze not needed”。

返回值类型：text

pg_stat_get_runtime_relstats(oid)

描述：返回轻量化autoanalyze生成的内存表级统计信息。

返回值类型：setof record

pg_stat_get_runtime_attstats(oid)

描述：返回轻量化autoanalyze生成的内存表列级统计信息。

返回值类型：setof record

 版权所有 © 华为云计算技术有限公司

pg_stat_get_runtime_partstats(oid)

描述：返回轻量化autoanalyze生成的内存分区级统计信息。仅8.2.1.200及以上集群版本支持。

返回值类型：setof record

pg_stat_get_runtime_part_attstats(oid)

描述：返回轻量化autoanalyze生成的内存分区列级统计信息。仅8.2.1.200及以上集群版本支持。

返回值类型：setof record

pg_stat_get_runtime_exprstats(oid)

描述：返回轻量化autoanalyze生成的内存表达式统计信息。仅8.2.1.200及以上集群版本支持。

返回值类型：setof record

pg_stat_get_predicate_columns(oid)

描述：返回表收集到的谓词列信息。仅9.1.0.100及以上集群版本支持。

返回值类型：setof smallint

pg_stat_delete_predicate_columns(oid)

描述：清空表的谓词列信息。仅9.1.0.100及以上集群版本支持。

返回值类型：void

pg_total_autovac_tuples(bool)

描述：返回total autovacuum相关的tuple记录，如nodename、nspname、relname以及各类tuple的IUD信息。

返回值类型：setof record

pg_autovac_status(oid)

描述：返回和autovacuum状态相关的参数信息。例如，nodename、nspname、relname、analyze及vacuum设置，analyze/vacuum阈值，analyze/vacuum tuple数等。

返回值类型：setof record

pg_autovac_timeout(oid)

描述：返回某个表执行autovacuum连续超时的次数，表信息非法或node信息异常返回NULL。

返回值类型：bigint

 版权所有 © 华为云计算技术有限公司

pg_autovac_coordinator(oid)

描述：返回对某个表执行autovacuum的coordinator名称，表信息非法或node信息异常返回NULL。

返回值类型：text

pg_stat_get_all(regclass)

描述：返回表在当前CN上pg_stat_object中对应的tuple记录，包括从databaseid到read_sql_count_1d字段。入参为表的oid。8.2.1.210及以上集群版本支持。

返回值类型：setof record

示例：查询当前节点public模式下t1表的最近一次访问时间。

```
SELECT data_access_timestamp FROM pg_stat_get_all('public.t1'::regclass);
data_access_timestamp
-------------------------------
2024-06-27 16:16:26.643132+08
(1 row)
```

pgxc_stat_get_all(relnamespace text, relname text)

描述：返回表在所有CN上pg_stat_object中对应的tuple记录，包括从databaseid到read_sql_count_1d字段。8.2.1.210及以上集群版本支持。

返回值类型：setof record

示例：查询所有CN上public模式下t1表的最近一次访问时间。

```
SELECT relid,data_access_timestamp FROM pgxc_stat_get_all('public','t1');
 relid      | data_access_timestamp
------------+-------------------------------
2147516676 | 2000-01-01 08:00:00+08
2147516674 | 2000-01-01 08:00:00+08
2147516655 | 2024-06-27 16:16:26.643132+08
(3 rows)
```

pgxc_stat_get_last_data_access_timestamp(relnamespace text, relname text)

描述：返回指定表在所有节点中最近一次的访问时间。9.1.0.100及以上集群版本支持。

返回值类型：timestamptz

示例：查询所有节点上public模式下t1表的最近一次访问时间。

```
SELECT * FROM pgxc_stat_get_last_data_access_timestamp('public','t1');
pgxc_stat_get_last_data_access_timestamp
-----------------------------------------
2024-06-27 16:16:26.643132+08
(1 row)
```

pgxc_stat_flush_object_data()

描述：手动执行持久化，每个节点将本地所有库中的所有表的统计信息写入pg_stat_object系统表。统计信息包括pg_stat_object中从databaseid到autovac_vacuum_count共29个字段。只有系统管理员可以执行此函数。9.1.0.100及以上集群版本支持。

返回值类型：boolean

 版权所有 © 华为云计算技术有限公司

示例：系统管理员用户执行持久化。

```
SELECT pgxc_stat_flush_object_data();
pgxc_stat_flush_object_data
-----------------------------
t
(1 row)
```

说明

此函数不支持并发执行。

pgxc_get_wlm_session_info_bytime(text, timestamp without time zone, timestamp without time zone, int)

描述：PGXC_WLM_SESSION_INFO视图在统计数据量很大的场景中性能较差，建议使用该函数进行筛选查询。入参分别为：筛选时间列（'start_time', 'finish_time'），起始区间时间，结束区间时间，每个CN返回的最大数量。返回值为GS_WLM_SESSION_HISTORY。

返回值类型：setof record

pgxc_get_wlm_current_instance_info(text, int default null)

描述：在CN节点上查询集群各节点当前的资源使用情况，读取内存中还未存到GS_WLM_INSTANCE_HISTORY系统表的数据。入参分别为：节点名称（可以输入ALL、C、D、实例名称），每个节点返回的最大数量。返回值为GS_WLM_INSTANCE_HISTORY。

返回值类型：setof record

pgxc_get_wlm_history_instance_info(text, TIMESTAMP, TIMESTAMP, int default null)

描述：在CN节点上查询集群各节点历史资源使用情况，读取GS_WLM_INSTANCE_HISTORY系统表的数据。入参分别为：节点名称（可以输入ALL、C、D、实例名称），起始区间时间，结束区间时间，每个实例返回的最大数量。返回值为GS_WLM_INSTANCE_HISTORY。

返回值类型：setof record

pg_stat_get_last_data_changed_time(oid)

描述：insert/update/delete, exchange/truncate/drop partition在该表上最后一次操作的时间，PG_STAT_ALL_TABLES视图last_data_changed列的数据是通过该函数求值，在表数量很大的场景中，通过视图获取表数据最后修改时间的性能较差，建议直接使用该函数获取表数据的最后修改时间。

返回值类型：timestamptz

pg_stat_set_last_data_changed_time(oid)

描述：手动设置该表上最后一次insert/update/delete, exchange/truncate/drop partition操作的时间。

返回值类型：void

 版权所有 © 华为云计算技术有限公司

pg_stat_set_last_data_changed_num(oid)

描述：设置该表上当前节点的历史累计修改计数。

返回值类型：void

pv_session_time()

描述：统计当前节点各会话线程的运行时间信息及各执行阶段所消耗时间。

返回值类型：record

pv_instance_time()

描述：统计当前节点的运行时间信息及各执行阶段所消耗时间。

返回值类型：record

pg_stat_get_activity(integer)

描述：返回一个关于带有特殊PID的后台进程的记录信息，当参数为NULL时，则返回每个活动的后台进程的记录。返回结果是PG_STAT_ACTIVITY视图中的一个子集，不包含connection_info列。

返回值类型：setof record

pg_stat_get_activity_with_conninfo(integer)

描述：返回一个关于带有特殊PID的后台进程的记录信息，当参数为NULL时，则返回每个活动的后台进程的记录。返回结果是PG_STAT_ACTIVITY视图中的一个子集。

返回值类型：setof record

pg_user_iostat(text)

描述：该函数8.1.2版本中已废弃，当前版本查询无效。8.2.0及以上集群版本中可使用PGXC_TOTAL_USER_RESOURCE_INFO视图查询实例上用户实时资源消耗信息。

返回值类型：record

名称	类型	描述
userid	oid	用户ID。
min_curr_iops	int4	当前该用户io在各DN中的最小值。
max_curr_iops	int4	当前该用户io在各DN中的最大值。
min_peak_iops	int4	该用户io峰值中，各DN的最小值。
max_peak_iops	int4	该用户io峰值中，各DN的最大值。
io_limits	int4	用户指定的资源池所设置的每秒可触发IO次数上限。
io_priority	text	该用户所设置的I/O优先级。

 版权所有 © 华为云计算技术有限公司

pg_stat_get_function_calls(oid)

描述：函数已被调用次数。

返回值类型：bigint

pg_stat_get_function_total_time(oid)

描述：该函数花费的总挂钟时间，以微秒为单位。包括花费在此函数调用上的时间。

返回值类型：double precision

pg_stat_get_function_self_time(oid)

描述：在当前事务中仅花费在此函数上的时间。不包括花费在调用函数上的时间。

返回值类型：double precision

pg_stat_get_backend_idset()

描述：设置当前活动的服务器进程数（从1到活动服务器进程的数量）。

返回值类型：setofinteger

pg_stat_get_backend_pid(integer)

描述：给定的服务器线程的线程ID。

返回值类型：bigint

```
SELECT pg_stat_get_backend_pid(1);
 pg_stat_get_backend_pid
-------------------------
        139706243217168
(1 row)
```

pg_stat_get_backend_dbid(integer)

描述：连接到给定服务器进程的数据库ID。

返回值类型：oid

pg_stat_get_backend_userid(integer)

描述：给定的服务器进程的用户ID。

返回值类型：oid

pg_stat_get_backend_activity(integer)

描述：给定服务器进程的当前活动查询，仅在调用者是系统管理员或被查询会话的用户，并且打开track_activities的时候才能获得结果。

返回值类型：text

 版权所有 © 华为云计算技术有限公司

pg_stat_get_backend_waiting(integer)

描述：如果给定服务器进程在等待某个锁，并且调用者是系统管理员或被查询会话的用户，并且打开track_activities的时候才返回真。

返回值类型：boolean

pg_stat_get_backend_activity_start(integer)

描述：给定服务器进程当前正在执行的查询的起始时间，仅在调用者是系统管理员或被查询会话的用户，并且打开track_activities的时候才能获得结果。

返回值类型：timestampwithtimezone

pg_stat_get_backend_xact_start(integer)

描述：给定服务器进程当前正在执行的事务的开始时间，但只有当前用户是系统管理员或被查询会话的用户，并且打开track_activities的时候才能获得结果。

返回值类型：timestampwithtimezone

pg_stat_get_backend_start(integer)

描述：给定服务器进程启动的时间，如果当前用户不是系统管理员或被查询的后端的用户，则返回NULL。

返回值类型：timestampwithtimezone

pg_stat_get_backend_client_addr(integer)

描述：连接到给定客户端后端的IP地址。

如果是通过Unix域套接字连接的则返回NULL；如果当前用户不是系统管理员或被查询会话的用户，返回NULL。

返回值类型：inet

备注：该函数中IP地址作为入参时，需写为不带圆点的格式，例如，192.168.100.128写为192168100128。

pg_stat_get_backend_client_port(integer)

描述：连接到给定客户端后端的TCP端口。

如果是通过Unix域套接字连接的则返回-1；如果当前用户不是系统管理员或被查询会话的用户，返回NULL。

返回值类型：integer

pg_stat_get_bgwriter_timed_checkpoints()

描述：后台写进程开启定时检查点的时间（因为checkpoint_timeout时间已经过期了）。

返回值类型：bigint

 版权所有 © 华为云计算技术有限公司

pg_stat_get_bgwriter_requested_checkpoints()

描述：后台写进程开启基于后端请求的检查点的时间，因为已经超过了checkpoint_segments或因为已经执行了CHECKPOINT。

返回值类型：bigint

pg_stat_get_bgwriter_buf_written_checkpoints()

描述：在检查点期间后台写进程写入的缓冲区数目。

返回值类型：bigint

pg_stat_get_bgwriter_buf_written_clean()

描述：为日常清理脏块，后台写进程写入的缓冲区数目。

返回值类型：bigint

pg_stat_get_bgwriter_maxwritten_clean()

描述：后台写进程停止清理扫描的时间，因为已经写入了更多的缓冲区（相比bgwriter_lru_maxpages参数声明的缓冲区数）。

返回值类型：bigint

pg_stat_get_buf_written_backend()

描述：后端进程写入的缓冲区数，因为它们需要分配一个新的缓冲区。

返回值类型：bigint

pg_stat_get_buf_alloc()

描述：分配的总缓冲区数。

返回值类型：bigint

pg_stat_clear_snapshot()

描述：清理当前的统计快照。

返回值类型：void

pg_stat_reset()

描述：为当前数据库重置统计计数器为0（需要系统管理员权限）。

返回值类型：void

pg_stat_reset_shared(text)

描述：重置shared cluster每个节点当前数据统计计数器为0（需要系统管理员权限）。

返回值类型：void

 版权所有 © 华为云计算技术有限公司

pg_stat_reset_single_table_counters(oid)

描述：为当前数据库中的一个表或索引重置统计为0（需要系统管理员权限）。

返回值类型：void

pg_stat_reset_single_function_counters(oid)

描述：为当前数据库中的一个函数重置统计为0（需要系统管理员权限）。

返回值类型：void

pg_stat_session_cu(int, int, int)

描述：获取当前节点所运行session的CU命中统计信息。

返回值类型：record

gs_get_stat_session_cu(text, int, int, int)

描述：获取集群所有运行session的CU命中统计信息。

返回值类型：record

gs_get_stat_db_cu(text, text, bigint, bigint, bigint)

描述：获取集群一个数据库的CU命中统计信息。

返回值类型：record

pg_stat_get_cu_mem_hit(oid)

描述：获取当前节点当前数据库中一个列存表的CU内存命中次数。

返回值类型：bigint

pg_stat_get_cu_hdd_sync(oid)

描述：获取当前节点当前数据库中一个列存表从磁盘同步读取CU次数。

返回值类型：bigint

pg_stat_get_cu_hdd_asyn(oid)

描述：获取当前节点当前数据库中一个列存表从磁盘异步读取CU次数。

返回值类型：bigint

pg_stat_get_db_cu_mem_hit(oid)

描述：获取当前节点一个数据库CU内存命中次数。

返回值类型：bigint

pg_stat_get_db_cu_hdd_sync(oid)

描述：获取当前节点一个数据库从磁盘同步读取CU次数。

 版权所有 © 华为云计算技术有限公司

返回值类型：bigint

pg_stat_get_db_cu_hdd_asyn(oid)

描述：获取当前节点一个数据库从磁盘异步读取CU次数。

返回值类型：bigint

pgxc_fenced_udf_process()

描述：查看UDF Master和Work进程数。

返回值类型：record

pgxc_terminate_all_fenced_udf_process()

描述：终止所有的UDF Work进程。

返回值类型：bool

gs_all_nodegroup_control_group_info(text)

描述：提供了所有逻辑集群的控制组信息。该函数在调用时需指定要查询的逻辑集群的名称。例如，要查询逻辑集群“installation”的控制组信息：

```
SELECT * FROM gs_all_nodegroup_control_group_info('installation');
```

返回值类型：record

函数返回字段如下：

名称	类型	描述
name	text	控制组的名称。
type	text	控制组的类型。
gid	bigint	控制组ID。
classgid	bigint	Workload所属Class的控制组ID。
class	text	Class控制组。
workload	text	Workload控制组。
shares	bigint	控制组分配的CPU资源配额。
limits	bigint	控制组分配的CPU资源限额。
wdlevel	bigint	Workload控制组层级。
cpucores	text	控制组使用的CPU核的信息。

gs_get_nodegroup_tablecount(name)

描述：得到一个逻辑集群中所有数据库包含的用户表数目。

返回值类型：integer

 版权所有 © 华为云计算技术有限公司

pgxc_max_datanode_size(name)

描述：得到一个逻辑集群的所有DN节点中数据库文件占用磁盘空间的最大值，单位为字节。

返回值类型：bigint

gs_check_logic_cluster_consistency()

描述：检查当前系统中所有逻辑集群是否存在系统信息不一致的情况，如果返回空记录，表示不存在不一致情况；否则，逻辑集群中CN和DN上的NodeGroup信息存在不一致。该函数应该在非扩缩容重分布时调用。

返回值类型：record

gs_check_tables_distribution()

描述： 检查当前系统中用户表的分布是否存在不一致，如果返回空记录，表示不存在不一致。该函数应该在非扩缩容重分布时调用。

返回值类型：record

pg_stat_bad_block()

描述：获取当前节点自启动后，读取出现Page/CU的损坏信息。

返回值类型：record

返回信息如下：

返回值字段	类型	描述
nodename	text	节点名称。
databaseid	integer	数据库OID。
tablespaceid	integer	表空间OID。
relfilenode	integer	文件对象ID。
forknum	integer	文件类型。
error_count	integer	出现校验失败的次数。
first_time	timestamp with time zone	第一次出现时间。
last_time	timestamp with time zone	最近一次出现时间。

pgxc_stat_bad_block()

描述：获取集群所有节点自启动后，读取出现Page/CU的损坏信息。

返回值类型：record

返回信息如下：

 版权所有 © 华为云计算技术有限公司

返回值字段	类型	描述
nodename	text	节点名称。
databaseid	oid	数据库OID。
tablespaceid	oid	表空间OID。
relfilenode	oid	文件对象ID。
forknum	integer	文件类型。
error_count	integer	出现校验失败的次数。
first_time	timestamp with time zone	第一次出现的时间。
last_time	timestamp with time zone	最近一次出现的时间。

pg_stat_bad_block_clear()

描述：清理节点记录的读取出现的Page/CU损坏信息（需要系统管理员权限）。

返回值类型：void

pgxc_stat_bad_block_clear()

描述：清理集群所有节点记录的读取出现的Page/CU损坏信息（需要系统管理员权限）。

返回值类型：void

gs_respool_exception_info(pool text)

描述：查看某个资源池关联的查询规则信息。

返回值类型：record

gs_control_group_info(pool text)

描述：查看资源池关联的控制组信息

返回值类型：record

返回信息如下：

属性	属性值	描述
name	class_a:workload_a1	class和workload名称。
class	class_a	Class控制组名称。
workload	workload_a1	Workload控制组名称。

 版权所有 © 华为云计算技术有限公司

属性	属性值	描述
type	DEFWD	控制组类型（Top、CLASS、BAKWD、DEFWD、TSWD）。
gid	87	控制组ID。
shares	30	占父节点CPU资源的百分比。
limits	0	占父节点CPU核数的百分比。
rate	0	Timeshare中的分配比例。
cpucores	0-3	CPU核心数。

gs_wlm_user_resource_info(name text)

描述：查询具体某个用户的资源限额和资源使用情况。

返回值类型：record

pgxc_stat_single_table(schema, tablename)

描述：在CN上执行，入参为schema和表名。查询单张表在全库中的统计信息及该表在每个DN上的脏页率。

该函数仅8.1.3及以上集群版本支持。

说明

该函数的统计信息依赖于ANALYZE，为获取该表最准确的信息请先对表进行ANALYZE。

返回值类型：record

返回值字段与函数**pg_stat_get_tuple()**相同。

```
SELECT * FROM pgxc_stat_single_table('public','t1');
 nodename  | tableid | partid |      last_vacuum       |    last_autovacuum     |        last_analyze         |
last_autoanalyze    | vacuum_count | autovacuum_count | analyze_count | autoanalyze_count | n_tup_ins | n_
tup_upd | n_tup_del | n_tup_hot_upd | n_tup_change | n_live_tup | n_dead_tup | dirty_rate |
last_data_changed
-----------+---------+--------+------------------------+------------------------+-----------------------------
+------------------------+--------------+------------------+---------------+-------------------+-----------+---
--------+-----------+---------------+--------------+------------+------------+------------------
 datanode1 | 1270075 |        | 2000-01-01 08:00:00+08 | 2000-01-01 08:00:00+08 | 2023-01-09
09:38:43.220876+08 | 2000-01-01 08:00:00+08 |            0 |                0 |             1 |                 0 |         0 |
       0 |         0 |             0 |             0 |          0 |          0 |
(1 row)
```

pgxc_stat_wal_write()

描述：在所有CN上执行，记录各个DN上与CN交互的线程信息，并统计wal日志和数据页导入量及速率。该函数仅8.2.0及以上集群版本支持。

返回值类型：bigint

返回值字段：

- node_name：实例名。

版权所有 © 华为云计算技术有限公司

- application_name：应用名。
- query_start：当前正在执行语句的起始时间。
- datapage_write：当前query导入的数据页总量，单位为bytes。
- datapage_write_speed：数据页导入速率，单位为bytes/s。
- wal_write：当前query导入的wal日志总量，单位为bytes。
- wal_write_speed：wal导入速率，单位为bytes/s。
- total_datapage_write：当前线程产生数据页总量，单位为bytes。
- total_wal_write：wal总量，单位为bytes。

```
SELECT * FROM pgxc_stat_wal_write();
 node_name | datid |       pid       | lwtid | usesysid |    application_name    | state  |          query_start
|         backend_start         |     query_id      | datapage_write | datapage_write_speed | wal_write |
wal_write_speed | total_datapage_write | total_wal_write
-----------+-------+-----------------+-------+----------+------------------------+--------+-------------------------------
+-------------------------------+-------------------+----------------+----------------------+-----------+-----------------
+----------------------+-----------------
 datanode1 | 15979 | 140328473442048 | 12408 |       10 | CalculateSpaceInfo     | active | 2022-11-15
10:39:45.00219+08  | 2022-11-15 10:39:45.000918+08 |                 0 |              0 |                   0 |     70697
|              0 |                   0 |              0
 datanode1 |     0 | 140328075503360 | 12510 |       10 | WalSender to Secondary | idle
|                               | 2022-11-15 10:39:46.708557+08 |                 0 |              0 |                     |       210
|              0 |                   0 |              0
 datanode1 | 15979 | 140327896741632 | 13612 |       10 | coordinator1           | active | 2022-11-15
10:46:32.832548+08 | 2022-11-15 10:40:20.117516+08 | 72620543991349940 |       586579968 |
17425000 |   11743056 |          348000 |          5337505792 |       107245825
 datanode1 | 15979 | 140327583217408 | 13614 |       10 | coordinator1           | active | 2022-11-15
10:46:32.832548+08 |                               | 72620543991349940 |              0 |                   0 |         0
|              0 |                   0 |              0
 datanode1 | 15979 | 140327485175552 | 27914 |       10 | coordinator1           | active | 2022-11-15
10:47:06.493584+08 | 2022-11-15 10:47:06.489062+08 | 72620543991350020 |              0 |                   0
|         0 |              0 |                   0 |           8675
(5 rows)
```

说明

行存不带索引批量导入时，数据页copy导入会产生logical newpage相关的xlog日志，在xlog量大于默认值时同样会触发流控。

pg_stat_wal_write()

描述：在CN或者DN上执行，记录当前实例的线程信息，并统计wal日志和数据页导入量及速率，这里以CN为例。

该函数仅8.2.0及以上集群版本支持。

返回值类型：bigint

返回值字段：

- node_name：实例名。
- application_name：应用名。
- query_start：当前正在执行语句的起始时间。
- datapage_write：当前query导入的数据页总量，单位为bytes。
- datapage_write_speed：数据页导入速率，单位为bytes/s。
- wal_write：当前query导入的wal日志总量，单位为bytes。
- wal_write_speed：wal导入速率，单位为bytes/s。

 版权所有 © 华为云计算技术有限公司

- total_datapage_write：当前线程产生数据页总量，单位为bytes。
- total_wal_write：wal总量，单位为bytes。

```
SELECT * FROM pg_stat_wal_write();
  node_name   | datid |       pid       | lwtid | usesysid |   application_name   | state  |          query_start
|         backend_start         |     query_id      | datapage_write | datapage_write_speed | wal_write |
wal_write_speed | total_datapage_write | total_wal_write
--------------+-------+-----------------+-------+----------+----------------------+--------+-------------------------------
+-------------------------------+-------------------+----------------+----------------------+-----------+-----------------
+----------------------+-----------------
 coordinator1 | 15979 | 140234153498368 | 25353 |       10 | JobScheduler         | active |                               |
2022-11-15 11:57:54.396347+08 |                 0 |              0 |                      |         0 |              0
|                    0 |              0
 coordinator1 | 15979 | 140234132027136 | 25354 |       10 | StatCollector        | idle   |
|                               |                 0 |              0 |                      |         0 |              0 |                  0
|              0
 coordinator1 | 15979 | 140234090084096 | 25356 |       10 | WDRSnapshot          | idle   | 2022-11-15
11:57:54.388948+08 | 2022-11-15 11:57:54.403055+08 |                 0 |              0 |                  0 |         0
|              0 |                    0 |              0
 coordinator1 | 15979 | 140234027169536 | 25359 |       10 | workload             | active | 2022-11-15
11:57:54.404836+08 | 2022-11-15 11:57:54.389259+08 | 72620543991404472 |              0 |                  0 |
1722112 |              0 |                    0 |              0
 coordinator1 | 15979 | 140234004621056 | 25360 |       10 | CalculateSpaceInfo | active | 2022-11-15
11:57:54.405677+08 | 2022-11-15 11:57:54.389847+08 | 72620543991349266 |              0 |                  0
|    191326 |              0 |                    0 |              0
 coordinator1 | 15979 | 140233987839744 | 25361 |       10 | WorkloadMonitor      | active | 2022-11-15
11:57:54.406263+08 | 2022-11-15 11:57:54.390373+08 | 72620543991404418 |              0 |                  0 |
1183741 |              0 |                    0 |              0
 coordinator1 | 15979 | 140233971058432 | 25362 |       10 | WLMArbiter           | active | 2022-11-15
11:57:54.406367+08 | 2022-11-15 11:57:54.390947+08 |                 0 |              0 |                  0 |         0
|              0 |                    0 |              0
 coordinator1 | 15979 | 140233228666624 | 14415 |       10 | gsql                 | active | 2022-11-15
21:19:32.200305+08 | 2022-11-15 21:09:38.916931+08 | 72620543991404422 |              0 |                  0
|     30872 |           1000 |                    0 |              0
 coordinator1 | 15979 | 140233172035328 | 15516 |       10 | gsql                 | active | 2022-11-15
21:19:49.7877+08   | 2022-11-15 21:10:13.447312+08 | 72620543991404485 |              0 |                  0
|         0 |              0 |                    0 |              0
 coordinator1 |     0 | 140234245269248 | 25348 |        0 | Background writer    | idle   |                               |
2022-11-15 11:57:54.383651+08 |                 0 |              0 |                      |         0 |              0
|                    0 |              0
 coordinator1 |     0 | 140234291410688 | 25347 |        0 | CheckPointer         | idle   |                               |
2022-11-15 11:57:54.383231+08 |                 0 |              0 |                      |         0 |              0
|                    0 |              0
 coordinator1 |     0 | 140234228487936 | 25349 |        0 | Wal Writer           | idle   |                               |
2022-11-15 11:57:54.384069+08 |                 0 |              0 |                      |         0 |              0
|                    0 |              0
 coordinator1 |     0 | 140234068584192 | 25357 |        0 | TwoPhase Cleaner     | idle   |                               |
2022-11-15 11:57:54.388332+08 |                 0 |              0 |                      |         0 |              0
|                    0 |              0
 coordinator1 |     0 | 140234048657152 | 25358 |        0 | LWLock Monitor       | idle   |                               |
2022-11-15 11:57:54.389239+08 |                 0 |              0 |                      |         0 |              0
|                    0 |              0
 coordinator1 |     0 | 140234194925312 | 25351 |        0 | CBM Writer           | idle   |                               |
2022-11-15 11:57:54.38506+08  |                 0 |              0 |                      |         0 |              0
|                    0 |              0
 coordinator1 |     0 | 140234211706624 | 25350 |        0 | Tick service         | idle   |                               |
2022-11-15 11:57:54.384136+08 |                 0 |              0 |                      |         0 |              0
|                    0 |              0
(16 rows)
```

说明

行存不带索引批量导入时，数据页copy导入会产生logical newpage相关的xlog日志，在xlog量大于默认值时同样会触发流控。

版权所有 © 华为云计算技术有限公司

gs_stack()

描述：获取CN或DN进程的堆栈信息，函数入参tid均需要通过视图进行获取，不支持lwtid。该函数仅8.2.1及以上集群版本支持。

gs_stack()提供了比gdb、gstack等命令更灵活的、更快速的堆栈获得手段。仅建议管理员用户在问题分析定位过程中使用，不建议作为日常的监控工具使用。

返回值字段：

- tid：线程id。
- lwtid：轻量级线程id。
- stack：tid或lwtid线程对应的堆栈信息。

gs_stack()函数有四种使用方式：

- 方式一：gs_stack(tid)函数入参值为0，表示打印gsql所连接的CN或DN进程内所有线程的堆栈。

  说明

  - 可通过gsql连接CN或DN执行。
  - 获取CN或DN进程内所有线程的运行堆栈，运行时间长，对正常业务运行有影响，不建议频繁使用。

```
SELECT * FROM gs_stack(0);
      tid       | lwtid |
stack
-----------------+-------
+-----------------------------------------------------------------------------------------------------------------------
---
 140186894757888 | 95275 | __poll +
0x2d                                                                                                                  +
                |       | ServerLoop() +
0x457                                                                                                                 +
                |       | PmStartupThreads() +
0x151                                                                                                                 +
                |       | PostmasterMain(int, char**) +
0x22a                                                                                                                 +
                |       | main +
0x250                                                                                                                 +
                |       | __libc_start_main +
0xf5                                                                                                                  +
                |       |
0xb8d2d7                                                                                                              +
                |       |
 140185903825152 | 95316 | __poll +
0x2d                                                                                                                  +
                |       | GsWaiter::wait(void*, long) +
0x11a                                                                                                                 +
                |       | GsWaiter::wait(void*, int, int, long) +
0x7e                                                                                                                  +
                |       | SysLoggerMain() +
0x109                                                                                                                 +
                |       | SubPostmasterMain(tag_gs_thread_args*) +
0xc9f                                                                                                                 +
                |       | MainStarterThreadFunc(void*) +
0x6a                                                                                                                  +
                |       | ThreadStarterFunc(void*) +
0x66                                                                                                                  +
                |       | start_thread +
0xc5                                                                                                                  +
                |       | clone +
0x6d                                                                                                                  +
                |       |
 140185903825528 | 95317 | __poll +
```

 版权所有 © 华为云计算技术有限公司

```
0x2d                                                                +
        |       | GsWaiter::wait(void*, long) +
0x11a                                                           +
        |       | AlarmCheckerMain() +
0x21b                                                       +
        |       | SubPostmasterMain(tag_gs_thread_args*) +
0xcc6                                                   +
        |       | MainStarterThreadFunc(void*) +
0x6a                                                +
        |       | ThreadStarterFunc(void*) +
0x66                                                    +
        |       | start_thread +
0xc5                                                        +
        |       | clone +
0x6d                                                            +
        |       |
 140185903825904 | 95320 | __poll +
0x2d                                                            +
        |       | GsWaiter::wait(void*, long) +
0x11a                                                       +
        |       | ReaperBackendMain() +
0x1ee                                                           +
        |       | SubPostmasterMain(tag_gs_thread_args*) +
0xd6f                                                   +
        |       | MainStarterThreadFunc(void*) +
0x6a                                                +
        |       | ThreadStarterFunc(void*) +
0x66                                                    +
        |       | start_thread +
0xc5                                                        +
        |       | clone +
0x6d                                                            +
        |       |
 140185903826656 | 95324 | __poll +
0x2d                                                            +
        |       | GsWaiter::wait(void*, long) +
0x11a                                                       +
        |       | CheckpointerMain() +
0x8bb                                                           +
        |       | AuxiliaryThreadUnderPm(AuxProcType) +
0x210                                                       +
        |       | SubPostmasterMain(tag_gs_thread_args*) +
0x7e0                                                   +
        |       | MainStarterThreadFunc(void*) +
0x6a                                                +
        |       | ThreadStarterFunc(void*) +
0x66                                                    +
        |       | start_thread +
0xc5                                                        +
        |       | clone +
0x6d                                                            +
        |       |
 140185903827032 | 95325 | __poll +
0x2d                                                            +
        |       | GsWaiter::wait(void*, long) +
0x11a                                                       +
        |       | BackgroundWriterMain() +
0x541                                                           +
        |       | AuxiliaryThreadUnderPm(AuxProcType) +
0x210                                                       +
        |       | SubPostmasterMain(tag_gs_thread_args*) +
0x7e0                                                   +
        |       | MainStarterThreadFunc(void*) +
0x6a                                                +
        |       | ThreadStarterFunc(void*) +
0x66                                                    +
```

- 方式二：gs_stack(tid)函数入参值为指定的tid，表示打印gsql所连接的CN或DN进程内tid线程的堆栈。

 版权所有 © 华为云计算技术有限公司

说明

- 可通过gsql连接CN或DN执行。
- tid必须为所连CN或DN进程上存在的线程tid，否则会报错“invalid thread id”。
- tid需通过其他视图获取，不支持lwtid。
- 获取指定CN或DN进程内指定线程的运行堆栈，运行时间短，对正常业务运行无影响，推荐使用。

```
SELECT * FROM gs_stack(140185903826656);
      tid       | lwtid |                    stack
----------------+-------+-----------------------------------------------
 140185903826656 | 95275 | __poll + 0x2d                                +
                 |       | GsWaiter::wait(void*, long) + 0x11a          +
                 |       | CheckpointerMain() + 0x8bb                  +
                 |       | AuxiliaryThreadUnderPm(AuxProcType) + 0x210   +
                 |       | SubPostmasterMain(tag_gs_thread_args*) + 0x7e0+
                 |       | MainStarterThreadFunc(void*) + 0x6a          +
                 |       | ThreadStarterFunc(void*) + 0x66             +
                 |       | start_thread + 0xc5                        +
                 |       | clone + 0x6d                              +
                 |       |
(1 row)
```

- 方式三：gs_stack('nodename', 0)函数第一个参数值为nodename，第二个参数值为0，表示打印由nodename指定的进程内所有线程的堆栈。

说明

- 底层实现使用了“execute direct on”，因此必须使用gsql连接CN上执行。
- 第一个参数nodename需用单引号包括。
- 获取指定CN或DN进程内所有线程的运行堆栈，运行时间长，对正常业务运行有影响，不建议频繁使用。

```
SELECT * FROM gs_stack('datanode2', 0);
      tid       | lwtid |
stack
----------------+-------
+-----------------------------------------------------------------------------------------------------------------------------------------
---
 140634541242112 | 95442 | do_futex_wait +
0x5f                                                                +
                |       | __new_sem_wait_slow +
0x57                                                              +
                |       | sem_timedwait +
0x35                                                               +
                |       | tagBinarySemaphore::timed_wait(int) +
0x4f                                                             +
                |       | AuxiliaryThreadMain(void*) +
0x6d                                                              +
                |       | LibcommThreadRoutine(void*) +
0x41                                                             +
                |       | start_thread +
0xc5                                                              +
                |       | clone +
0x6d                                                                   +
                |       |
 140634549634816 | 95438 | epoll_wait +
0x33                                                                 +
                |       | LibcommEpollWait(int, int) +
0x38                                                              +
                |       | RecvDataThreadMain(void*) +
0x8e                                                              +
                |       | LibcommThreadRoutine(void*) +
0x41                                                             +
                |       | start_thread +
0xc5                                                              +
```

 版权所有 © 华为云计算技术有限公司

```
           |      | clone +
0x6d                                                                +
           |      |
 140634558027520 | 95432 | epoll_wait +
0x33                                                              +
           |      | LibcommEpollWait(int, int) +
0x38                                                           +
           |      | RecvDataThreadMain(void*) +
0x8e                                                           +
           |      | LibcommThreadRoutine(void*) +
0x41                                                          +
           |      | start_thread +
0xc5                                                             +
           |      | clone +
0x6d                                                                +
           |      |
 140634566420224 | 95430 | epoll_wait +
0x33                                                              +
           |      | LibcommEpollWait(int, int) +
0x38                                                           +
           |      | RecvDataThreadMain(void*) +
0x8e                                                           +
           |      | LibcommThreadRoutine(void*) +
0x41                                                          +
           |      | start_thread +
0xc5                                                             +
           |      | clone +
0x6d                                                                +
           |      |
 140634589476608 | 95426 | epoll_wait +
0x33                                                              +
           |      | LibcommEpollWait(int, int) +
0x38                                                           +
           |      | RecvDataThreadMain(void*) +
0x8e                                                           +
           |      | LibcommThreadRoutine(void*) +
0x41                                                          +
           |      | start_thread +
0xc5                                                             +
           |      | clone +
0x6d                                                                +
           |      |
 140634597869312 | 95424 | do_futex_wait +
0x5f                                                           +
           |      | __new_sem_wait_slow +
0x57                                                        +
           |      | sem_timedwait +
0x35                                                           +
           |      | tagBinarySemaphore::timed_wait(int) +
0x4f                                                        +
           |      | SendDataWait() +
0x58                                                           +
           |      | SendDataThreadMain(void*) +
0xe4                                                           +
           |      | LibcommThreadRoutine(void*) +
0x41                                                          +
           |      | start_thread +
0xc5                                                             +
           |      | clone +
0x6d                                                                +
           |      |
 140634606262016 | 95387 | do_futex_wait +
0x5f                                                           +
           |      | __new_sem_wait_slow +
0x57                                                        +
           |      | sem_timedwait +
0x35                                                           +
           |      | tagBinarySemaphore::timed_wait(int) +
0x4f                                                        +
```


版权所有 © 华为云计算技术有限公司

```
            |      | SendDataWait() +
0x58                                                                +
            |      | SendDataThreadMain(void*) +
0xe4                                                          +
```

- 方式四：gs_stack('nodename', 0)函数第一个参数值为nodename，第二个参数值为tid，表示打印由nodename指定的进程内tid线程的堆栈。

  说明

  - 底层实现使用了“execute direct on”，因此必须gsql连接CN上执行。
  - 第一个参数nodename需用单引号包括。
  - 第二个参数必须是nodename指定进程内存在的线程，否则会报错“invalid thread id”。
  - tid需通过其他视图获取，不支持lwtid。
  - 仅获取指定CN或DN进程内指定线程的运行堆栈，运行时间短，对正常业务运行无影响，推荐使用。

```
SELECT * FROM gs_stack('datanode2', 140634549634816);
      tid       | lwtid |               stack
----------------+-------+-----------------------------------
 140634549634816 | 95438 | epoll_wait + 0x33               +
            |      | LibcommEpollWait(int, int) + 0x38 +
            |      | RecvDataThreadMain(void*) + 0x8e  +
            |      | LibcommThreadRoutine(void*) + 0x41+
            |      | start_thread + 0xc5              +
            |      | clone + 0x6d                    +
            |      |
(1 row)
```

pgxc_get_xlog_stats()

描述：在CN上执行，统计各个DN节点自开机到当前不同xlog类型数量。该函数仅8.3.0及以上集群版本支持。

返回值类型：record

返回信息如下：

名称	类型	描述
node_name	text	节点名称。
rmgr_name	text	xlog主类型。
record_info	text	xlog info信息。
record_count	bigint	数量信息。
count_pct	text	数量占比。
record_size	bigint	record大小， 单位byte。
record_pct	text	record大小占比。
fpi_size	bigint	fpi(full page image)大小，单位byte。
fpi_pct	text	fpi大小占比。
combined_size	bigint	xlog总大小，单位byte。

版权所有 © 华为云计算技术有限公司

名称	类型	描述
combined_pct	text	xlog总大小占比。

说明

- 该函数只统计DN节点最近一次启动以来，不同类型的xlog数量与大小。
- 该函数统计的xlog的大小未考虑到其写盘时的对齐问题，与实际xlog段文件大小存在部分误差。

```
SELECT * FROM pgxc_get_xlog_stats();
 node_name |  rmgr_name  |          record_info          | record_count | count_pct | record_size | record_pct | fpi_size | fpi_pct | combined_size | combined
_pct
-----------+-------------+-------------------------------+--------------+-----------+-------------+------------+----------+---------+---------------+---------
-----
 datanode1 | XLOG        | checkpoint: shutdown          |            0 | 0.00%     |           0 | 0.00%      |        0 | 0.00%   |             0 | 0.00%
 datanode1 | XLOG        | checkpoint: online            |            0 | 0.00%     |           0 | 0.00%      |        0 | 0.00%   |             0 | 0.00%
 datanode1 | XLOG        | xlog no-op                    |            0 | 0.00%     |           0 | 0.00%      |        0 | 0.00%   |             0 | 0.00%
 datanode1 | XLOG        | nextOid                       |            0 | 0.00%     |           0 | 0.00%      |        0 | 0.00%   |             0 | 0.00%
 datanode1 | XLOG        | xlog switch                   |            0 | 0.00%     |           0 | 0.00%      |        0 | 0.00%   |             0 | 0.00%
 datanode1 | XLOG        | backup end                    |            0 | 0.00%     |           0 | 0.00%      |        0 | 0.00%   |             0 | 0.00%
 datanode1 | XLOG        | parameter change              |            0 | 0.00%     |           0 | 0.00%      |        0 | 0.00%   |             0 | 0.00%
 datanode1 | XLOG        | restore point                 |            0 | 0.00%     |           0 | 0.00%      |        0 | 0.00%   |             0 | 0.00%
 datanode1 | XLOG        | full page writes              |            0 | 0.00%     |           0 | 0.00%      |        0 | 0.00%   |             0 | 0.00%
 datanode1 | XLOG        | xlog recovery end             |            0 | 0.00%     |           0 | 0.00%      |        0 | 0.00%   |             0 | 0.00%
 datanode1 | XLOG        | page hint                     |            1 | 0.12%     |          56 | 0.11%      |     8192 | 17.92%  |          8248 | 8.71%
 datanode1 | XLOG        | xlog fpi                      |            0 | 0.00%     |           0 | 0.00%      |        0 | 0.00%   |             0 | 0.00%
 datanode1 | Transaction | commit                        |            3 | 0.35%     |         246 | 0.50%      |        0 | 0.00%   |           246 | 0.26%
 datanode1 | Transaction | prepare                       |            1 | 0.12%     |         469 | 0.96%      |        0 | 0.00%   |           469 | 0.50%
 datanode1 | Transaction | abort                         |            0 | 0.00%     |           0 | 0.00%      |        0 | 0.00%   |             0 | 0.00%
 datanode1 | Transaction | commit prepared               |            1 | 0.12%     |         106 | 0.22%      |        0 | 0.00%   |           106 | 0.11%
 datanode1 | Transaction | abort prepared                |            0 | 0.00%     |           0 | 0.00%      |        0 | 0.00%   |             0 | 0.00%
 datanode1 | Transaction | xid assignment xtop           |            0 | 0.00%     |           0 | 0.00%      |        0 | 0.00%   |             0 | 0.00%
 datanode1 | Transaction | commit compact                |            0 | 0.00%     |           0 | 0.00%      |        0 | 0.00%   |             0 | 0.00%
 datanode1 | Storage     | file create                   |            0 | 0.00%     |           0 | 0.00%      |        0 | 0.00%   |             0 | 0.00%
 datanode1 | Storage     | file truncate                 |            0 | 0.00%     |           0 | 0.00%      |        0 | 0.00%   |             0 | 0.00%
 datanode1 | Storage     | vacuum clear file             |            0 | 0.00%     |           0 | 0.00%      |        0 | 0.00%   |             0 | 0.00%
 datanode1 | CLOG        | zero page                     |            0 | 0.00%     |           0 | 0.00%      |        0 | 0.00%   |             0 | 0.00%
 datanode1 | CLOG        | truncate before               |            0 | 0.00%     |           0 | 0.00%      |        0 | 0.00%   |             0 | 0.00%
```

 版权所有 © 华为云计算技术有限公司

```
 datanode1 | Database    | create db                    |           1 | 0.12%     |          50 | 0.10%      |        0 |
0.00%   |           50 | 0.05%
 datanode1 | Database    | drop db                      |           0 | 0.00%     |           0 | 0.00%      |        0 |
0.00%   |            0 | 0.00%
 datanode1 | Tablespace  | create tablespace            |           0 | 0.00%     |           0 | 0.00%      |        0 |
0.00%   |            0 | 0.00%
 datanode1 | Tablespace  | drop tablespace              |           0 | 0.00%     |           0 | 0.00%      |        0 |
0.00%   |            0 | 0.00%
 datanode1 | Tablespace  | relative create              |           0 | 0.00%     |           0 | 0.00%      |        0 |
0.00%   |            0 | 0.00%
 datanode1 | MultiXact   | zero offsets page            |           0 | 0.00%     |           0 | 0.00%      |        0 |
0.00%   |            0 | 0.00%
 datanode1 | MultiXact   | zero members page            |           0 | 0.00%     |           0 | 0.00%      |        0 |
0.00%   |            0 | 0.00%
 datanode1 | MultiXact   | create multixact             |           0 | 0.00%     |           0 | 0.00%      |        0 |
0.00%   |            0 | 0.00%
 datanode1 | RelMap      | update relmap                |           0 | 0.00%     |           0 | 0.00%      |        0 |
0.00%   |            0 | 0.00%
 datanode1 | Standby     | AccessExclusive locks        |           0 | 0.00%     |           0 | 0.00%      |        0 |
0.00%   |            0 | 0.00%
 datanode1 | Standby     | xlog running xacts           |           0 | 0.00%     |           0 | 0.00%      |        0 |
0.00%   |            0 | 0.00%
 datanode1 | Standby     | release AccessExclusive locks |           0 | 0.00%     |           0 | 0.00%      |        0 |
0.00%   |            0 | 0.00%
 datanode1 | Standby     | xlog standby csn             |           0 | 0.00%     |           0 | 0.00%      |        0 |
0.00%   |            0 | 0.00%
 datanode1 | Heap2       | freeze                       |           0 | 0.00%     |           0 | 0.00%      |        0 | 0.00%
|            0 | 0.00%
 datanode1 | Heap2       | clean                        |           0 | 0.00%     |           0 | 0.00%      |        0 | 0.00%
|            0 | 0.00%
 datanode1 | Heap2       | page upgrade                 |           0 | 0.00%     |           0 | 0.00%      |        0 |
0.00%   |            0 | 0.00%
 datanode1 | Heap2       | cleanup info                 |           0 | 0.00%     |           0 | 0.00%      |        0 |
0.00%   |            0 | 0.00%
 datanode1 | Heap2       | visible                      |           0 | 0.00%     |           0 | 0.00%      |        0 | 0.00%
|            0 | 0.00%
 datanode1 | Heap2       | multi-insert                 |           0 | 0.00%     |           0 | 0.00%      |        0 |
0.00%   |            0 | 0.00%
 datanode1 | Heap2       | bcm                          |           0 | 0.00%     |           0 | 0.00%      |        0 |
0.00%   |            0 | 0.00%
 datanode1 | Heap2       | logical newpage rel          |           0 | 0.00%     |           0 | 0.00%      |        0 |
0.00%   |            0 | 0.00%
 datanode1 | Heap2       | no repair page               |           0 | 0.00%     |           0 | 0.00%      |        0 |
0.00%   |            0 | 0.00%
 datanode1 | Heap2       | multi-insert(init) base xid  |           0 | 0.00%     |           0 | 0.00%      |        0 |
0.00%   |            0 | 0.00%
 datanode1 | Heap        | insert                       |         280 | 32.67%    |       15964 | 32.63%     |    14314 |
31.31%  |        30278 | 31.99%
 datanode1 | Heap        | delete                       |           0 | 0.00%     |           0 | 0.00%      |        0 | 0.00%
|            0 | 0.00%
 datanode1 | Heap        | update                       |           0 | 0.00%     |           0 | 0.00%      |        0 |
0.00%   |            0 | 0.00%
 datanode1 | Heap        | base shift                   |           0 | 0.00%     |           0 | 0.00%      |        0 | 0.00%
|            0 | 0.00%
 datanode1 | Heap        | hot update                   |           0 | 0.00%     |           0 | 0.00%      |        0 |
0.00%   |            0 | 0.00%
 datanode1 | Heap        | new page                     |           0 | 0.00%     |           0 | 0.00%      |        0 |
0.00%   |            0 | 0.00%
 datanode1 | Heap        | lock type                    |           0 | 0.00%     |           0 | 0.00%      |        0 |
0.00%   |            0 | 0.00%
 datanode1 | Heap        | inplace                      |           2 | 0.23%     |         116 | 0.24%      |     8164 |
17.86%  |         8280 | 8.75%
 datanode1 | Heap        | init page                    |           2 | 0.23%     |         130 | 0.27%      |       92 |
0.20%   |          222 | 0.23%
 datanode1 | Heap        | insert(init) base xid        |           0 | 0.00%     |           0 | 0.00%      |        0 |
0.00%   |            0 | 0.00%
 datanode1 | Btree       | insert leaf                  |         562 | 65.58%    |       31484 | 64.35%     |    13352 |
29.21%  |        44836 | 47.37%
 datanode1 | Btree       | insert upper                 |           1 | 0.12%     |          72 | 0.15%      |      148 |
```


版权所有 © 华为云计算技术有限公司

```
0.32%  |          220 | 0.23%
 datanode1 | Btree      | insert meta               |          0 | 0.00%    |          0 | 0.00%      |        0 |
0.00%  |            0 | 0.00%
 datanode1 | Btree      | split left                |          1 | 0.12%    |         78 | 0.16%      |      744 | 1.63%
|          822 | 0.87%
 datanode1 | Btree      | split right               |          0 | 0.00%    |          0 | 0.00%      |        0 | 0.00%
|            0 | 0.00%
 datanode1 | Btree      | split left root           |          0 | 0.00%    |          0 | 0.00%      |        0 | 0.00%
|            0 | 0.00%
 datanode1 | Btree      | split right root          |          1 | 0.12%    |         78 | 0.16%      |      672 |
1.47%  |          750 | 0.79%
 datanode1 | Btree      | delete                    |          0 | 0.00%    |          0 | 0.00%      |        0 | 0.00%
|            0 | 0.00%
 datanode1 | Btree      | delete page               |          0 | 0.00%    |          0 | 0.00%      |        0 |
0.00%  |            0 | 0.00%
 datanode1 | Btree      | delete page meta          |          0 | 0.00%    |          0 | 0.00%      |        0 |
0.00%  |            0 | 0.00%
 datanode1 | Btree      | lev                       |          1 | 0.12%    |         78 | 0.16%      |       40 | 0.09%
|          118 | 0.12%
 datanode1 | Btree      | delete page half          |          0 | 0.00%    |          0 | 0.00%      |        0 |
0.00%  |            0 | 0.00%
 datanode1 | Btree      | vacuum                    |          0 | 0.00%    |          0 | 0.00%      |        0 |
0.00%  |            0 | 0.00%
 datanode1 | Btree      | reuse page                |          0 | 0.00%    |          0 | 0.00%      |        0 |
0.00%  |            0 | 0.00%
 datanode1 | Gin        | Create index              |          0 | 0.00%    |          0 | 0.00%      |        0 |
0.00%  |            0 | 0.00%
 datanode1 | Gin        | Create ptree              |          0 | 0.00%    |          0 | 0.00%      |        0 |
0.00%  |            0 | 0.00%
 datanode1 | Gin        | Insert item               |          0 | 0.00%    |          0 | 0.00%      |        0 | 0.00%
|            0 | 0.00%
 datanode1 | Gin        | Page split                |          0 | 0.00%    |          0 | 0.00%      |        0 | 0.00%
|            0 | 0.00%
 datanode1 | Gin        | Vacuum page               |          0 | 0.00%    |          0 | 0.00%      |        0 |
0.00%  |            0 | 0.00%
 datanode1 | Gin        | Delete page               |          0 | 0.00%    |          0 | 0.00%      |        0 |
0.00%  |            0 | 0.00%
 datanode1 | Gin        | Update metapage           |          0 | 0.00%    |          0 | 0.00%      |        0 |
0.00%  |            0 | 0.00%
 datanode1 | Gin        | Insert new list page      |          0 | 0.00%    |          0 | 0.00%      |        0 |
0.00%  |            0 | 0.00%
 datanode1 | Gin        | Delete list pages         |          0 | 0.00%    |          0 | 0.00%      |        0 |
0.00%  |            0 | 0.00%
 datanode1 | Gin        | Vacuum data leaf page     |          0 | 0.00%    |          0 | 0.00%      |        0 |
0.00%  |            0 | 0.00%
 datanode1 | Gist       | page update               |          0 | 0.00%    |          0 | 0.00%      |        0 |
0.00%  |            0 | 0.00%
 datanode1 | Gist       | page split                |          0 | 0.00%    |          0 | 0.00%      |        0 | 0.00%
|            0 | 0.00%
 datanode1 | Gist       | Create index              |          0 | 0.00%    |          0 | 0.00%      |        0 |
0.00%  |            0 | 0.00%
 datanode1 | Gist       | page delete               |          0 | 0.00%    |          0 | 0.00%      |        0 |
0.00%  |            0 | 0.00%
 datanode1 | Sequence   | log                       |          0 | 0.00%    |          0 | 0.00%      |        0 | 0.00%
|            0 | 0.00%
 datanode1 | SPGist     | create index              |          0 | 0.00%    |          0 | 0.00%      |        0 |
0.00%  |            0 | 0.00%
 datanode1 | SPGist     | add leaf to page          |          0 | 0.00%    |          0 | 0.00%      |        0 |
0.00%  |            0 | 0.00%
 datanode1 | SPGist     | move leafs                |          0 | 0.00%    |          0 | 0.00%      |        0 |
0.00%  |            0 | 0.00%
 datanode1 | SPGist     | add node                  |          0 | 0.00%    |          0 | 0.00%      |        0 |
0.00%  |            0 | 0.00%
 datanode1 | SPGist     | prefix off                |          0 | 0.00%    |          0 | 0.00%      |        0 | 0.00%
|            0 | 0.00%
 datanode1 | SPGist     | picksplit                 |          0 | 0.00%    |          0 | 0.00%      |        0 | 0.00%
|            0 | 0.00%
 datanode1 | SPGist     | vacuum leaf               |          0 | 0.00%    |          0 | 0.00%      |        0 |
0.00%  |            0 | 0.00%
```

 版权所有 © 华为云计算技术有限公司

```
 datanode1 | SPGist      | vacuum root               |          0 | 0.00%     |          0 | 0.00%      |        0 |
0.00%   |            0 | 0.00%
 datanode1 | SPGist      | newest XID                |          0 | 0.00%     |          0 | 0.00%      |        0 |
0.00%   |            0 | 0.00%
 datanode1 | Slot        | create slot               |          0 | 0.00%     |          0 | 0.00%      |        0 | 0.00%
|            0 | 0.00%
 datanode1 | Slot        | advance slot              |          0 | 0.00%     |          0 | 0.00%      |        0 |
0.00%   |            0 | 0.00%
 datanode1 | Slot        | drop slot                 |          0 | 0.00%     |          0 | 0.00%      |        0 | 0.00%
|            0 | 0.00%
 datanode1 | Slot        | check slot                |          0 | 0.00%     |          0 | 0.00%      |        0 | 0.00%
|            0 | 0.00%
 datanode1 | Heap3       | new cid                   |          0 | 0.00%     |          0 | 0.00%      |        0 |
0.00%   |            0 | 0.00%
 datanode1 | Heap3       | heap rewrite              |          0 | 0.00%     |          0 | 0.00%      |        0 |
0.00%   |            0 | 0.00%
 datanode1 | Barrier     | barrier                   |          0 | 0.00%     |          0 | 0.00%      |        0 | 0.00%
|            0 | 0.00%
(102 rows)
```

pgxc_get_wal_speed()

描述：在CN上执行，获取每一个DN实例的wal生成速率以及备DN receive、write、flush、redo速率。该函数仅8.3.0及以上集群版本支持。

返回值类型：record

返回信息如下：

名称	类型	描述
node_name	text	节点名称。
pid	bigint	线程PID号。
local_role	text	当前节点role。
peer_role	text	接收方role。
peer_state	text	接收方状态。
state	text	传输状态。
generate_speed	bigint	主DNwal生成速率，单位为byte/s。
receive_speed	bigint	备DN接收速率，单位为byte/s。
write_speed	bigint	备DN写速率，单位为byte/s。
flush_speed	bigint	备DN刷盘速率，单位为byte/s。
redo_speed	bigint	备DN redo速率，单位为byte/s。

```
SELECT * FROM pgxc_get_wal_speed();
node_name |       pid       | local_role | peer_role | peer_state |    state    | generate_speed | receive_speed |
```

 版权所有 © 华为云计算技术有限公司

```
write_speed | flush_speed | apply_speed
-----------+-----------------+------------+-----------+------------+-----------+----------------+--------------
+-------------+-------------+-------------
 datanode1 | 140240389260824 | Primary    | Secondary | Normal     | Streaming |              0 |            0
|           0 |           0 |           0
 datanode1 | 140240389262192 | Primary    | Standby   | Normal     | Streaming |              0 |            0
|           0 |           0 |           0
(2 rows)
```

pg_xlog_display_one_lsn(start_lsn)

描述：可以在CN或者DN上执行，根据起始LSN号完整解析当前位置的xlog。该函数仅8.3.0及以上集群版本支持。

入参：start_lsn，表示起始LSN。需要保证输入的起始LSN号是一条xlog的开始处。

返回值类型：record

返回信息如下：

名称	类型	描述
node_name	text	当前实例名。
start_lsn	text	设定的起始LSN。
end_lsn	text	设定的终止LSN。
startlsn	text	xlog起始LSN。
endlsn	text	xlog终止LSN。
prelsn	text	前一条xlog起始LSN。
xid	xid	xlog事务ID号。
datalen	int4	xlog数据长度，单位为byte。
totallen	int4	xlog长度，单位为byte。
type	text	xlog类型。
desc	text	xlog内容。
blkref	text	xlog关联的relfilenode。

示例：

```
SELECT * FROM pg_xlog_display_one_lsn('0/101CA00');
node_name | startlsn  |  endlsn   |  prelsn   | xlog_tid | datalen | totallen |  xlogtype   |
xlogdescribe
          | blkref
-----------+-----------+-----------+-----------+----------+---------+----------+-------------
+-------------------------------------------------------------------------
----------+--------
datanode1 | 0/101CA00 | 0/101CA78 | 0/101C3F8 |        5 |      80 |      114 | Transaction | commit:
2023-10-19 22:21:38.617092+08; csn:0; inval msgs: catcache 11 ca
tcache 10 |
(1 row)
```


版权所有 © 华为云计算技术有限公司

pg_xlogdump (tablename)

描述：在CN或者DN上执行，根据表名解析xlog文件并过滤。该函数仅8.3.0及以上集群版本支持。

返回值类型：record

返回信息如下：

名称	类型	描述
node_name	text	节点名称。
start_lsn	text	设定的起始LSN。
end_lsn	text	设定的终止LSN。
startlsn	text	xlog起始LSN。
endlsn	text	xlog终止LSN。
prelsn	text	前一条xlog起始LSN。
xid	xid	xlog事务ID号。
datalen	int4	xlog数据长度，单位为byte。
totallen	int4	xlog长度，单位为byte。
type	text	xlog类型。
desc	text	xlog内容。
blkref	text	xlog关联的relfilenode。

说明

- 由于不同xlog类型xlogdescribe字段的长度不一致，pg_xlogdump()函数会对该字段进行裁剪，仅保留前64个字节。
- 可以找到目标xlog后结合pg_xlog_display_one_lsn()函数查看完整xlog内容。

示例：

```
SELECT * FROM pg_xlogdump(pg_class);
 node_name | startlsn  |  endlsn   |  prelsn   | xlog_tid | datalen | totallen | xlogtype |  xlogdescribe
|                    blkref
-----------+-----------+-----------+-----------+----------+---------+----------+----------+-----------------
+-------------------------------------------------
 datanode1 | 0/2DFC660 | 0/2DFE1D0 | 0/2DFA610 |        0 |       2 |     7000 | Heap     | inplace: off 17 |
blkrel #0: rel 1663/16324/15920, fork main, blk 2
 datanode1 | 0/2E02270 | 0/2E02D88 | 0/2E00220 |        0 |       2 |     2840 | Heap     | inplace: off 2  | blkrel
#0: rel 1663/16324/15920, fork main, blk 0
 datanode1 | 0/2E06E60 | 0/2E06F68 | 0/2E06E28 |        0 |       2 |      264 | Heap     | inplace: off 10 | blkrel
#0: rel 1663/16324/15920, fork main, blk 0
 datanode1 | 0/2E0FC08 | 0/2E0FD10 | 0/2E0EBB8 |        0 |       2 |      264 | Heap     | inplace: off 13 | blkrel
#0: rel 1663/16324/15920, fork main, blk 0
 datanode1 | 0/2E12960 | 0/2E12A68 | 0/2E11960 |        0 |       2 |      264 | Heap     | inplace: off 15 | blkrel
#0: rel 1663/16324/15920, fork main, blk 0
```

 版权所有 © 华为云计算技术有限公司

```
 datanode1 | 0/2E12C88 | 0/2E12D90 | 0/2E12B78 |   0 |   2 |   264 | Heap     | inplace: off 17 | blkrel
#0: rel 1663/16324/15920, fork main, blk 0
 datanode1 | 0/2E13FF8 | 0/2E14118 | 0/2E13110 |   0 |   2 |   264 | Heap     | inplace: off 18 | blkrel
#0: rel 1663/16324/15920, fork main, blk 0
 datanode1 | 0/2E169F8 | 0/2E185E0 | 0/2E169B0 |   0 |   2 |  7120 | Heap     | inplace: off 1  | blkrel
#0: rel 1663/16324/15920, fork main, blk 1
 datanode1 | 0/2E188B8 | 0/2E189D0 | 0/2E185E0 |   0 |   2 |   280 | Heap     | inplace: off 2  | blkrel
#0: rel 1663/16324/15920, fork main, blk 1
 datanode1 | 0/2E18BF8 | 0/2E18CE0 | 0/2E189D0 |   0 |   2 |   232 | Heap     | inplace: off 28 | blkrel
#0: rel 1663/16324/15920, fork main, blk 1
 datanode1 | 0/2E18E80 | 0/2E18F88 | 0/2E18DC0 |   0 |   2 |   264 | Heap     | inplace: off 4  | blkrel
#0: rel 1663/16324/15920, fork main, blk 1
 datanode1 | 0/2E19160 | 0/2E19268 | 0/2E19088 |   0 |   2 |   264 | Heap     | inplace: off 5  | blkrel
#0: rel 1663/16324/15920, fork main, blk 1
 datanode1 | 0/2E19460 | 0/2E1B1D0 | 0/2E19388 |   0 |   2 |  7508 | Heap     | inplace: off 16 | blkrel
#0: rel 1663/16324/15920, fork main, blk 6
 datanode1 | 0/2E1B1D0 | 0/2E1B2E8 | 0/2E19460 |   0 |   2 |   280 | Heap     | inplace: off 18 | blkrel
#0: rel 1663/16324/15920, fork main, blk 6
 datanode1 | 0/2E1B2E8 | 0/2E1B3D0 | 0/2E1B1D0 |   0 |   2 |   232 | Heap     | inplace: off 30 | blkrel
#0: rel 1663/16324/15920, fork main, blk 1
(15 rows)
```

pg_xlogdump (xid)

描述：可以在CN或者DN上执行，根据事务id号解析xlog文件并过滤。该函数仅8.3.0及以上集群版本支持。

返回值类型：record

返回信息如下：

名称	类型	描述
node_name	text	节点名称。
start_lsn	text	设定的起始LSN。
end_lsn	text	设定的终止LSN。
startlsn	text	xlog起始LSN。
endlsn	text	xlog终止LSN。
prelsn	text	前一条xlog起始LSN。
xid	xid	xlog事务ID号。
datalen	int4	xlog数据长度，单位为byte。
totallen	int4	xlog长度，单位为byte。
type	text	xlog类型。
desc	text	xlog内容。
blkref	text	xlog关联的relfilenode。

 版权所有 © 华为云计算技术有限公司

说明

- 由于不同xlog类型xlogdescribe字段的长度不一致，pg_xlogdump()函数会对该字段进行裁剪，仅保留前64个字节。
- 可以找到目标xlog后结合pg_xlog_display_one_lsn()函数查看完整xlog内容。

示例：

```
SELECT * FROM pg_xlogdump(1);
 node_name | startlsn  |  endlsn   |  prelsn   | xlog_tid | datalen | totallen | xlogtype |  xlogdescribe
|                 blkref
-----------+-----------+-----------+-----------+----------+---------+----------+----------+----------------
+-------------------------------------------
 datanode1 | 0/10000A0 | 0/10020F0 | 0/1000028 |        1 |       0 |     8248 | XLOG     | page hint      | blkrel
#0: rel 1663/1/1255, fork fsm, blk 2
 datanode1 | 0/10020F0 | 0/1004140 | 0/10000A0 |        1 |       0 |     8248 | XLOG     | page hint      | blkrel
#0: rel 1663/1/1247, fork fsm, blk 2
 datanode1 | 0/1004140 | 0/1004168 | 0/10020F0 |        1 |       4 |       38 | XLOG     | nextOid: 18192 |
 datanode1 | 0/1004168 | 0/10061B8 | 0/1004140 |        1 |       0 |     8248 | XLOG     | page hint      | blkrel
#0: rel 1663/1/2616, fork fsm, blk 2
 datanode1 | 0/10061B8 | 0/1008208 | 0/1004168 |        1 |       0 |     8248 | XLOG     | page hint      | blkrel
#0: rel 1663/1/2602, fork fsm, blk 2
 datanode1 | 0/1008208 | 0/100A258 | 0/10061B8 |        1 |       0 |     8248 | XLOG     | page hint      | blkrel
#0: rel 1663/1/2603, fork fsm, blk 2
 datanode1 | 0/100A258 | 0/100C2A8 | 0/1008208 |        1 |       0 |     8248 | XLOG     | page hint      | blkrel
#0: rel 1663/1/2609, fork fsm, blk 2
 datanode1 | 0/100C2A8 | 0/100E2F8 | 0/100A258 |        1 |       0 |     8248 | XLOG     | page hint      | blkrel
#0: rel 1663/1/2605, fork fsm, blk 2
 datanode1 | 0/100E2F8 | 0/1010348 | 0/100C2A8 |        1 |       0 |     8248 | XLOG     | page hint      | blkrel
#0: rel 1663/1/1259, fork fsm, blk 2
 datanode1 | 0/1010348 | 0/1012398 | 0/100E2F8 |        1 |       0 |     8248 | XLOG     | page hint      | blkrel
#0: rel 1663/1/2610, fork fsm, blk 2
(10 rows)
```

pg_xlogdump (start_lsn, end_lsn)

描述：可以在CN或者DN上执行，根据起始和终止lsn解析xlog文件。该函数仅8.3.0及以上集群版本支持。

入参：start_lsn，表示设定的起始LSN号；end_lsn，表示设定的终止LSN号。无需保证起始LSN号是一条xlog的起始位置，会从起始LSN号后第一条有效的xlog开始解析。

返回值类型：record

返回信息如下：

字段名称	类型	含义
node_name	text	节点名称。
start_lsn	text	设定的起始LSN。
end_lsn	text	设定的终止LSN。
startlsn	text	xlog起始LSN。
endlsn	text	xlog终止LSN。
prelsn	text	前一条xlog起始LSN。
xid	xid	xlog事务ID号。

 版权所有 © 华为云计算技术有限公司

字段名称	类型	含义
datalen	int4	xlog数据长度，单位为byte。
totallen	int4	xlog长度，单位为byte。
type	text	xlog类型。
desc	text	xlog内容。
blkref	text	xlog关联的relfilenode。

说明

- 由于不同xlog类型xlogdescribe字段的长度不一致，pg_xlogdump()函数会对该字段进行裁剪，仅保留前64个字节。
- 可以找到目标xlog后结合pg_xlog_display_one_lsn()函数查看完整xlog内容。

示例：

```
SELECT * FROM pg_xlogdump('0/101EF70','0/101F590');
 node_name | startlsn  |  endlsn   |  prelsn   | xlog_tid | datalen | totallen | xlogtype |     xlogdescribe
|                 blkref
-----------+-----------+-----------+-----------+----------+---------+----------+----------+----------------------
+-----------------------------------------------
 datanode1 | 0/101EF70 | 0/101EFC0 | 0/101EF18 |        6 |       2 |       80 | Btree    | insert leaf: off 1    |
blkrel #0: rel 1663/1/2674, fork main, blk 1
 datanode1 | 0/101EFC0 | 0/101F018 | 0/101EF70 |        6 |       3 |       88 | Heap     | insert: off 2 flags 0 |
blkrel #0: rel 1663/1/2608, fork main, blk 0
 datanode1 | 0/101F018 | 0/101F068 | 0/101EFC0 |        6 |       2 |       80 | Btree    | insert leaf: off 1    |
blkrel #0: rel 1663/1/2673, fork main, blk 1
 datanode1 | 0/101F068 | 0/101F0B8 | 0/101F018 |        6 |       2 |       80 | Btree    | insert leaf: off 2    |
blkrel #0: rel 1663/1/2674, fork main, blk 1
 datanode1 | 0/101F0B8 | 0/101F110 | 0/101F068 |        6 |       3 |       88 | Heap     | insert: off 3 flags 0 |
blkrel #0: rel 1663/1/2608, fork main, blk 0
 datanode1 | 0/101F110 | 0/101F160 | 0/101F0B8 |        6 |       2 |       80 | Btree    | insert leaf: off 1    |
blkrel #0: rel 1663/1/2673, fork main, blk 1
 datanode1 | 0/101F160 | 0/101F1B0 | 0/101F110 |        6 |       2 |       80 | Btree    | insert leaf: off 3    |
blkrel #0: rel 1663/1/2674, fork main, blk 1
 datanode1 | 0/101F1B0 | 0/101F208 | 0/101F160 |        6 |       3 |       88 | Heap     | insert: off 4 flags 0 |
blkrel #0: rel 1663/1/2608, fork main, blk 0
 datanode1 | 0/101F208 | 0/101F258 | 0/101F1B0 |        6 |       2 |       80 | Btree    | insert leaf: off 1    |
blkrel #0: rel 1663/1/2673, fork main, blk 1
 datanode1 | 0/101F258 | 0/101F2A8 | 0/101F208 |        6 |       2 |       80 | Btree    | insert leaf: off 4    |
blkrel #0: rel 1663/1/2674, fork main, blk 1
 datanode1 | 0/101F2A8 | 0/101F300 | 0/101F258 |        6 |       3 |       88 | Heap     | insert: off 5 flags 0 |
blkrel #0: rel 1663/1/2608, fork main, blk 0
 datanode1 | 0/101F300 | 0/101F350 | 0/101F2A8 |        6 |       2 |       80 | Btree    | insert leaf: off 1    |
blkrel #0: rel 1663/1/2673, fork main, blk 1
 datanode1 | 0/101F350 | 0/101F3A0 | 0/101F300 |        6 |       2 |       80 | Btree    | insert leaf: off 5    |
blkrel #0: rel 1663/1/2674, fork main, blk 1
 datanode1 | 0/101F3A0 | 0/101F3F8 | 0/101F350 |        6 |       3 |       88 | Heap     | insert: off 6 flags 0 |
blkrel #0: rel 1663/1/2608, fork main, blk 0
 datanode1 | 0/101F3F8 | 0/101F448 | 0/101F3A0 |        6 |       2 |       80 | Btree    | insert leaf: off 1    |
blkrel #0: rel 1663/1/2673, fork main, blk 1
 datanode1 | 0/101F448 | 0/101F498 | 0/101F3F8 |        6 |       2 |       80 | Btree    | insert leaf: off 5    |
blkrel #0: rel 1663/1/2674, fork main, blk 1
 datanode1 | 0/101F498 | 0/101F4F0 | 0/101F448 |        6 |       3 |       88 | Heap     | insert: off 7 flags 0 |
blkrel #0: rel 1663/1/2608, fork main, blk 0
 datanode1 | 0/101F4F0 | 0/101F540 | 0/101F498 |        6 |       2 |       80 | Btree    | insert leaf: off 1    |
blkrel #0: rel 1663/1/2673, fork main, blk 1
 datanode1 | 0/101F540 | 0/101F590 | 0/101F4F0 |        6 |       2 |       80 | Btree    | insert leaf: off 6    |
```

 版权所有 © 华为云计算技术有限公司

```
blkrel #0: rel 1663/1/2674, fork main, blk 1
(19 rows)
```

7.32 触发器函数

pg_get_triggerdef(oid)

描述：获取触发器的定义信息。

参数：待查触发器的OID。

返回值类型：text

示例：

```
SELECT pg_get_triggerdef(oid) FROM pg_trigger;
                                        pg_get_triggerdef
-----------------------------------------------------------------------------------------------------------------------
 CREATE TRIGGER insert_trigger BEFORE INSERT ON test_trigger_src_tbl FOR EACH ROW EXECUTE
PROCEDURE tri_insert_func()
(1 row)
```

pg_get_triggerdef(oid, boolean)

描述：获取触发器的定义信息。

参数：待查触发器的OID及是否以pretty方式展示。

返回值类型：text

说明

仅在创建trigger时指定WHEN条件的情况下，布尔类型参数才生效。

示例：

```
SELECT pg_get_triggerdef(oid,true) FROM pg_trigger;
                                        pg_get_triggerdef
-----------------------------------------------------------------------------------------------------------------------
 CREATE TRIGGER insert_trigger BEFORE INSERT ON test_trigger_src_tbl FOR EACH ROW EXECUTE
PROCEDURE tri_insert_func()
(1 row)

SELECT pg_get_triggerdef(oid,false) FROM pg_trigger;
                                        pg_get_triggerdef
-----------------------------------------------------------------------------------------------------------------------
 CREATE TRIGGER insert_trigger BEFORE INSERT ON test_trigger_src_tbl FOR EACH ROW EXECUTE
PROCEDURE tri_insert_func()
(1 row)
```

7.33 XML 函数

7.33.1 产生 XML 内容

本节函数和类函数的表达式可以用来从SQL数据产生XML内容。适用于将查询结果格式化成XML文档以便于在客户端应用中处理。

 版权所有 © 华为云计算技术有限公司

XMLPARSE ({ DOCUMENT | CONTENT } *value*)

描述：从字符数据中生成一个XML类型的值。

返回值类型：xml

示例：

```
SELECT xmlparse(document '<foo>bar</foo>');
   xmlparse
----------------
 <foo>bar</foo>
(1 row)
```

XMLSERIALIZE ({ DOCUMENT | CONTENT } *value* AS *type*

描述：从XML类型的值生成一个字符串。

返回值类型：type，可以是character，character varying或text（或其别名）

示例：

```
SELECT xmlserialize(content 'good' AS CHAR(10));
 xmlserialize
--------------
 good
(1 row)
```

xmlcomment(text)

描述：用于创建一个XML值，它含有一个以指定文本作为内容的XML注释。该文本中不能包含“--”或以一个“-”结尾，这样的文本才是有效的XML注释。当参数为空时，结果也为空。

返回值类型：xml

示例：

```
SELECT xmlcomment('hello');
  xmlcomment
--------------
 <!--hello-->
(1 row)
```

xmlconcat(xml[, ...])

描述：用于将XML值组成的列表串接成一个单独的值。其中空值会被忽略，当所有参数都为空时，结果都为空。

返回值类型：xml

示例：

```
SELECT xmlconcat('<abc/>', '<bar>foo</bar>');
     xmlconcat
----------------------
 <abc/><bar>foo</bar>
(1 row)
```

说明：XML声明如果存在，结果如下：如果所有参数值都使用相同的XML版本声明，则在结果中使用版本，否则不用版本。如果所有参数值都有standalone属性，且值都为yes， 则结果中standalone值为yes，如果至少有一个是no，则结果中standalone值为no。否则结果中将不带standalone属性。

 版权所有 © 华为云计算技术有限公司

示例：

```
SELECT xmlconcat('<?xml version="1.1"?><foo/>', '<?xml version="1.1" standalone="no"?><bar/>');
           xmlconcat
-----------------------------------
 <?xml version="1.1"?><foo/><bar/>
(1 row)
```

xmlelement(name name [, xmlattributes(value [AS attname] [, ...])] [, content, ...])

描述：使用给定名称、属性和内容产生一个XML元素。

返回值类型：xml

示例：

```
SELECT xmlelement(name foo);
 xmlelement
------------
 <foo/>
(1 row)

SELECT xmlelement(name foo, xmlattributes('xyz' as bar));
   xmlelement
------------------
 <foo bar="xyz"/>
(1 row)

SELECT xmlelement(name foo, xmlattributes(current_date as bar), 'cont', 'ent');
            xmlelement
---------------------------------------
 <foo bar="2023-08-16">content</foo>
(1 row)
```

没有合法XML名称的元素和属性名会被转义，转义的方式是将不合法的字符用序列_xHHHH_替换，其中HHHH是该字符以16进制表达的Unicode代码点。例如：

```
SELECT xmlelement(name "foo$bar", xmlattributes('xyz' as "a&b"));

           xmlelement
---------------------------------
 <foo_x0024_bar a_x0026_b="xyz"/>
```

如果属性值是一个列引用，则不需要指定显式的属性名，这种情况下默认将把列名用作该属性的名字。在其他情况下，必须给该属性一个显式的命名。因此这个例子是合法的：

```
CREATE TABLE test (a xml, b xml);
SELECT xmlelement(name test, xmlattributes(a, b)) FROM test;
```

但这些就不合法：

```
SELECT xmlelement(name test, xmlattributes('constant'), a, b) FROM test;
SELECT xmlelement(name test, xmlattributes(func(a, b))) FROM test;
```

元素内容（如果指定）将被根据其数据类型进行格式化。如果内容本身是类型xml，则会构建复杂的XML文档。例如：

```
SELECT xmlelement(name foo, xmlattributes('xyz' as bar),xmlelement(name
abc),xmlcomment('test'),xmlelement(name xyz));
                  xmlelement
----------------------------------------------
 <foo bar="xyz"><abc/><!--test--><xyz/></foo>
```

 版权所有 © 华为云计算技术有限公司

其他类型的内容将被格式化为合法的XML字符数据。这意味着特别的字符<、>以及&将会被转换成实体。二进制数据（数据类型bytea）将被表示为base64或者十六进制编码，具体取决于配置参数xmlbinary。

xmlforest(content [AS name] [, ...])

描述：使用给定名称和内容产生一个元素的XML森林（序列）。

返回值类型：xml

示例：

```
SELECT xmlforest('abc' AS foo, 123 AS bar);
          xmlforest
------------------------------
 <foo>abc</foo><bar>123</bar>
(1 row)

SELECT xmlforest(table_name, column_name) FROM ALL_TAB_COLUMNS WHERE schema = 'pg_catalog';
                                                xmlforest
-----------------------------------------------------------------------------------------------------------------------------------
 <table_name>pg_authid</table_name><column_name>rolsuper</column_name>
 <table_name>pg_authid</table_name><column_name>rolinherit</column_name>
 <table_name>pg_authid</table_name><column_name>rolcreaterole</column_name>
```

xmlpi(name target [, content])

描述：创建一个XML处理指令。内容不能包含字符序列?>

返回值类型：xml

示例：

```
SELECT xmlpi(name php, 'echo "hello world";');
            xmlpi
-----------------------------
 <?php echo "hello world";?>
(1 row)
```

xmlroot(xml, version text | no value [, standalone yes|no|no value])

描述：修改一个XML值的根节点的属性。如果指定了一个版本，它会替换根节点的版本声明中的值；如果指定了一个standalone值，它会替换根节点standalone中的值。

返回值类型：xml

示例：

```
SELECT xmlroot(xmlparse(document '<?xml version="1.0" standalone="no"?><content>abc</content>'),
version '1.1', standalone yes);
                            xmlroot
----------------------------------------------------------------
 <?xml version="1.1" standalone="yes"?><content>abc</content>
(1 row)
```

xmlagg(xml)

描述：函数xmlagg是一个聚集函数，将输入值串接起来。

返回值类型：xml

示例：

 版权所有 © 华为云计算技术有限公司

```
CREATE TABLE test (y int, x xml);
INSERT INTO test VALUES (1, '<foo>abc</foo>');
INSERT INTO test VALUES (2, '<bar/>');

SELECT xmlagg(x) FROM test;
       xmlagg
----------------------
 <bar/><foo>abc</foo>
(1 row)
```

为聚集调用增加一个ORDER BY子句即可决定串接的顺序，例如：

```
SELECT xmlagg(x ORDER BY y DESC) FROM test;
       xmlagg
----------------------
 <bar/><foo>abc</foo>
(1 row)
```

7.33.2 XML 谓词

本节的函数用于检查xml值的属性。

xml IS DOCUMENT

描述：如果参数XML值是一个正确的XML文档，则IS DOCUMENT返回真；如果非正确XML文档，则返回假；参数为空时返回空。

返回值类型：bool

```
SELECT '<abc/>' is document;
 ?column?
----------
 t
(1 row)
```

xml IS NOT DOCUMENT

描述：如果参数XML值不是一个正确的XML文档，则IS NOT DOCUMENT返回真；如果是正确XML文档，则返回假；参数为空时返回空。

返回值类型：bool

```
SELECT 'abc' is document;
 ?column?
----------
 f
(1 row)
```

XMLEXISTS(text PASSING [BY REF] xml [BY REF])

描述：如果第一个参数中的XPath表达式返回任何节点，则函数XMLEXISTS返回真，否则返回假（如果哪一个参数为空，则结果就为空）。BY REF子句没有作用，用于保持SQL兼容性。

返回值类型：bool

示例：

```
SELECT xmlexists('//town[text() = ''TScity'']' PASSING BY REF '<towns><town>TScity</town><town>TOcity</
town></towns>');
 xmlexists
------------
 t
(1 row)
```

 版权所有 © 华为云计算技术有限公司

xml_is_well_formed(text)

描述：检查text字符串是不是格式良好的XML值，返回布尔结果。如果xmloption参数设置为DOCUMENT则检查文档，如果设置为CONTENT则检查内容。

返回值类型：bool

示例：

```
SET xmloption TO DOCUMENT;
SET

SELECT xml_is_well_formed('<>');
 xml_is_well_formed
--------------------
 f
(1 row)

SELECT xml_is_well_formed('<abc/>');
 xml_is_well_formed
--------------------
 t
(1 row)

SET xmloption TO CONTENT;
SELECT xml_is_well_formed('abc');
 xml_is_well_formed
--------------------
 t
(1 row)
```

xml_is_well_formed_document(text)

描述：检查text字符串是不是格式良好的文档，返回布尔结果。

返回值类型：bool

示例：

```
SELECT xml_is_well_formed_document('<test:foo xmlns:test="http://test.com/test">bar</test:foo>');
 xml_is_well_formed
--------------------
 t
(1 row)

SELECT xml_is_well_formed_document('<test:foo xmlns:test="http://test.com/test">bar</my:foo>');
 xml_is_well_formed_document
-----------------------------
 f
(1 row)
```

xml_is_well_formed_content(text)

描述：检查text字符串是不是格式良好的内容，返回布尔结果。

返回值类型：bool

示例：

```
SELECT xml_is_well_formed_content('content');
 xml_is_well_formed_content
----------------------------
 t
(1 row)

SELECT xml_is_well_formed_content('<content');
```

 版权所有 © 华为云计算技术有限公司

```
 xml_is_well_formed_content
----------------------------
 f
(1 row)
```

7.33.3 处理 XML

为了处理数据类型XML的值，DWS提供了函数xpath和xpath_exists计算XPath表达式以及XMLTABLE表函数。

xpath(xpath, xml [, nsarray])

描述：它返回一个XML值的数组对应xpath表达式所产生的节点集。如果xpath表达式返回一个标量值而不是节点集，将返回一个单个元素的数组。

该函数的第二个参数xml必须是一个完整的XML文档，它必须有一个根节点元素。

第三个参数为可选参数，是一个命名空间的数组映射。这个数组应该是一个二维text数组，第二个维的长度等于2（它应该是一个数组的数组，其中每一个正好包含2个元素）。每个数组项的第一个元素是命名空间名称的别名，第二个元素是命名空间URI。这个数组中提供的别名不必与XML文档本身中使用的别名相同。换句话说，在XML文档和xpath函数上下文中，别名都是本地的。

返回值类型：xml值的数组

示例：

```
SELECT xpath('/my:a/text()', '<my:a xmlns:my="http://example.com">test</my:a>', ARRAY[ARRAY['my',
'http://example.com']]);
 xpath
--------
 {test}
(1 row)
```

要处理默认（匿名）命名空间：

```
SELECT xpath('//mydefns:b/text()', '<a xmlns="http://example.com"><b>test</b></
a>',ARRAY[ARRAY['mydefns', 'http://example.com']]);
 xpath
--------
 {test}
(1 row)
```

xpath_exists(xpath, xml [, nsarray])

描述：函数xpath_exists是xpath函数的一种特殊形式。这个函数不是返回满足XPath的XML值，它返回一个布尔值表示查询是否被满足。这个函数等价于标准的XMLEXISTS谓词，不过它还提供了对一个名字空间映射参数的支持。

返回值：bool

示例：

```
SELECT xpath_exists('/my:a/text()', '<my:a xmlns:my="http://example.com">test</my:a>',
ARRAY[ARRAY['my', 'http://example.com']]);
 xpath_exists
--------------
 t
(1 row)
```

 版权所有 © 华为云计算技术有限公司

xmltable

描述：xmltable函数根据输入的XML数据、XPath路径表达式、列的定义信息生成一个表。xmltable在语法上类似于一个函数，但它只能以表的形式出现在查询的FROM子句里。

返回值：setof record

语法：

```
XMLTABLE ( [ XMLNAMESPACES ( namespace_uri AS namespace_name [,  ...] ), ]
            row_expression PASSING [ BY  { REF | VALUE } ]
document_expression [ BY  { REF | VALUE } ]
COLUMNS name  { type  [ PATH column_expression  ] [ DEFAULT default_expression ] [ NOT NULL |
NULL ] | FOR ORDINALITY }
[, ...]
)
```

参数说明：

- 可选的XMLNAMESPACES子句是一个逗号分隔的命名空间定义列表，其中namespace_uri都是text类型表达式，namespace_name都是简单标识符。XMLNAMESPACES指定了在文档中使用的XML命名空间及它们的别名。当前不支持默认的命名空间声明。
- 必选的row_expression参数是一个XPath 1.0表达式，该表达式根据提供的XML文档document_expression进行计算获取XML节点序列，该序列是xmltable转换成输出行的顺序。如果document_expression值为NULL，或row_expression产生空的节点集，结果将不返回任何行。
- document_expression参数用于输入XML文档，输入的文档必须是符合XML格式的文档，不支持XML片段数据或格式错误的XML文档。BY REF和BY VALUE子句不起作用，仅用于实现SQL标准兼容性。
- COLUMNS子句指定输出表中字段列表定义，列名和列的数据类型是必选的，路径、默认值和是否为空子句是可选的。
  - 列的column_expression是一个XPath 1.0表达式，用于从row_expression计算出当前行中提取出列的值。如果没有给出column_expression，那么字段名将用作隐式路径。
  - 列可以被标记为NOT NULL，如果NOT NULL列的column_expression不返回任何数据，并且没有DEFAULT子句或者default_expression的计算结果为NULL，就会报错。
  - 标记为FOR ORDINALITY的字段将从1开始填充行号，其顺序为从row_expression结果集中检索到的节点顺序，最多只有一列可以标记为FOR ORDINALITY。

  须知

  XPath 1.0没有为节点指定顺序，因此返回结果的顺序取决于数据获取顺序。

示例：

```
SELECT * FROM XMLTABLE('/ROWS/ROW'
PASSING '<ROWS><ROW id="1"><CITY_ID>1002a</CITY_ID><CITY_NAME>snowcity</CITY_NAME></
ROW><ROW id="2"><CITY_ID>1003b</CITY_ID><CITY_NAME>icecity</CITY_NAME></ROW><ROW
id="3"><CITY_ID>1004c</CITY_ID><CITY_NAME>windcity</CITY_NAME></ROW></ROWS>'
COLUMNS id INT PATH '@id',
ordinality FOR ORDINALITY,
```

 版权所有 © 华为云计算技术有限公司

```
CITY_id TEXT PATH 'CITY_ID',CITY_name TEXT PATH 'CITY_NAME' NOT NULL);
 id | ordinality | city_id | city_name
----+------------+---------+-----------
  1 |          1 | 1002a   | snowcity
  2 |          2 | 1003b   | icecity
  3 |          3 | 1004c   | windcity
(3 rows)
```

7.33.4 将表映射到 XML

table_to_xml(tbl regclass, nulls boolean, tableforest boolean, targetns text)

描述：把表的内容映射成XML值。

返回值类型：xml

table_to_xmlschema(tbl regclass, nulls boolean, tableforest boolean, targetns text)

描述：把关系表的模式映射成XML模式文档。

返回值类型：xml

table_to_xml_and_xmlschema(tbl regclass, nulls boolean, tableforest boolean, targetns text)

描述：把关系表映射成XML值和模式文档。

返回值类型：xml

query_to_xml(query text, nulls boolean, tableforest boolean, targetns text)

描述：把SQL查询的内容映射成XML值。

返回值类型：xml

query_to_xmlschema(query text, nulls boolean, tableforest boolean, targetns text)

描述：把SQL查询映射成XML模式文档。

返回值类型：xml

query_to_xml_and_xmlschema(query text, nulls boolean, tableforest boolean, targetns text)

描述：把SQL查询映射成XML值和模式文档。

返回值类型：xml

cursor_to_xml(cursor refcursor, count int, nulls boolean, tableforest boolean, targetns text)

描述：把游标查询映射成XML值。

返回值类型：xml

 版权所有 © 华为云计算技术有限公司

cursor_to_xmlschema(cursor refcursor, nulls boolean, tableforest boolean, targetns text)

描述：把游标查询映射成XML模式文档。

返回值类型：xml

schema_to_xml(schema name, nulls boolean, tableforest boolean, targetns text)

描述：把模式中的表映射成XML值。

返回值类型：xml

schema_to_xmlschema(schema name, nulls boolean, tableforest boolean, targetns text)

描述：把模式中的表映射成XML模式文档

返回值类型：xml

schema_to_xml_and_xmlschema(schema name, nulls boolean, tableforest boolean, targetns text)

描述：把模式中的表映射成XML值和模式文档。

返回值类型：xml

database_to_xml(nulls boolean, tableforest boolean, targetns text)

描述：把数据库的表映射成XML值。

返回值类型：xml

database_to_xmlschema(nulls boolean, tableforest boolean, targetns text)

描述：把数据库的表映射成XML模式文档

返回值类型：xml

database_to_xml_and_xmlschema(nulls boolean, tableforest boolean, targetns text)

描述：把数据库的表映射成XML值和模式文档。

返回值类型：xml

 版权所有 © 华为云计算技术有限公司

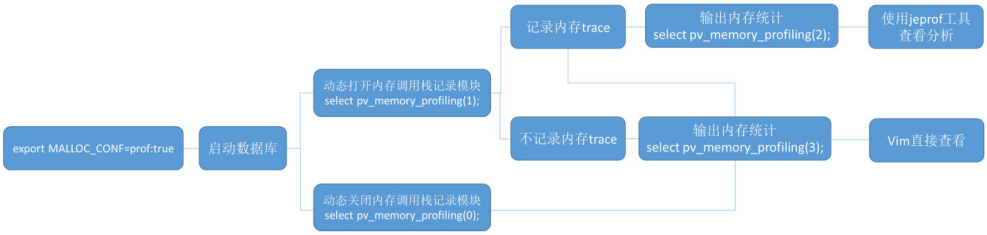
说明

将表映射到XML有关函数的参数说明如下：

- tbl：表名。
- nulls：在输出中是否包含空值，若为true，列中的空值表示为：<columnname xsi:nil="true"/>，若为false，包含空值的列会从输出中省略。
- tableforest：若为true，则输出xml片段，false，输出xml文档。
- targetns：指定想要结果的XML命名空间。若不指定，应传递一个空字符串。
- query：SQL查询语句。
- cursor：游标名。
- count：从游标中获取的数据量。
- schema：模式名

7.34 调用栈记录函数

DWS通过pv_memory_profiling(type int)和环境变量MALLOC_CONF，控制gaussdb进程中malloc等内存分配调用栈记录模块开启关闭、以及进程级内存调用栈输出等，使用流程如下图所示：



MALLOC_CONF

环境变量MALLOC_CONF用于控制监控模块是否开启，位于${BIGDATA_HOME}/mppdb/.mppdbgs_profile文件内，默认开启。需注意：

- 修改此环境变量变化，需要重启数据库。
- 如果在集群中启用了om_monitor，完成环境变量设置后，先重启om_monitor进程后，然后重启数据库，使得开关生效。
- 该环境变量可以设置在集群所有服务器中，也可以仅设置在需要开启模块的个别服务器中，对GaussDB进程而言，各进程是根据各自MALLOC_CONF环境变量，控制模块是否打开。

MALLOC_CONF开启和关闭命令：

- 开启监控模块：
  ```
  export MALLOC_CONF=prof:true
  ```
- 关闭监控模块：
  ```
  export MALLOC_CONF=prof:false
  ```

pv_memory_profiling (type int)

参数说明：控制gaussdb进程中malloc等内存分配函数的backtrace记录及输出。

取值范围：正整数，0~3

 版权所有 © 华为云计算技术有限公司

表 7-86 pv_memory_profiling 取值含义

pv_memory_profiling 取值	含义
0	关闭内存trace，不记录malloc等调用栈信息。
1	开启内存trace，记录malloc等调用栈信息。
2	输出malloc等调用栈的trace日志。 ● 输出路径为：/proc/pid/cwd指向的目录，pid为gaussdb进程ID。 ● 输出日志命名规则为：jeprof.<pid>.*.heap，其中pid表示GaussDB进程ID，*表示trace日志输出的序号，序号唯一。如：jeprof.195473.0.u0.heap
3	输出内存统计信息。 ● 输出路径为：/proc/pid/cwd指向的目录，pid为gaussdb进程ID。 ● 命名规则为节点名称+进程id+时间+“heap_stats”+.out。该文件可用vim直接打开。

返回值类型：boolean

备注：

- 如果成功，函数返回true，否则返回false，返回相关提示信息。
- 如果失败提示信息为“Memory profiling failed, check if $MALLOC_CONF contain 'prof:true'.”，说明未设置MALLOC_CONF=prof:true的情况下，使用了本模块，需要设置环境变量即可解决。
- 如果失败提示信息为“Type %d is not supported. The valid range is 0-3.”，说明用户输入参数错误，正确数值为0,1,2,3。
- 如果失败提示信息为“Memory profiling failed, input type is %d, failed number is %d.”请联系技术支持工程师提供技术支持。

输出内存调用栈信息

操作步骤：

步骤1 输出内存调用栈信息，在GaussDB进程所在目录输出trace文件：

```
SELECT * FROM pv_memory_profiling(2);
```

步骤2 使用jemalloc中提供的jeprof工具，解析日志信息：

方式1：以text格式输出。

```
jeprof --text --show_bytes $GAUSSHOME/bin/gaussdb trace文件1 >prof.txt
```

方式2：以pdf格式输出。

```
jeprof --pdf --show_bytes $GAUSSHOME/bin/gaussdb trace文件1 > prof.pdf
```

 版权所有 © 华为云计算技术有限公司

说明

- 解析内存调用栈信息，需要依靠GaussDB源码进行分析，需要将trace文件返回给研发工程师，进行分析。
- 分析trace文件，需要使用jeprof工具，该工具由Jemalloc生成。在常规使用中，需要依赖perl环境，如果需要生成pdf调用图，需要安装与操作系统匹配的GraphViz工具。

----结束

示例

系统管理员用户登录，设置环境变量，启动数据库：

```
export MALLOC_CONF=prof:true
```

在数据库运行期间，关闭内存trace记录功能：

```
SELECT pv_memory_profiling(0);
pv_memory_profiling
----------------------------
t
(1 row)
```

在数据库运行期间，打开内存trace记录功能：

```
SELECT pv_memory_profiling(1);
pv_memory_profiling
----------------------------
t
(1 row)
```

输出内存trace记录功能：

```
SELECT pv_memory_profiling(2);
pv_memory_profiling
----------------------------
t
(1 row)
```

在GaussDB进程所在目录输出trace文件，以text格式或PDF格式输出：

```
jeprof --text --show_bytes $GAUSSHOME/bin/gaussdb trace文件1 >prof.txt
jeprof --pdf --show_bytes $GAUSSHOME/bin/gaussdb trace文件1 > prof.pdf
```

输出内存统计信息。执行以下命令，在GaussDB进程所在目录输出内存统计文件，可直接读取：

```
SELECT pv_memory_profiling(3);
pv_memory_profiling
----------------------------
t
(1 row)
```

7.35 Hudi 系统函数

Hudi系统函数仅9.1.0.100及以上集群版本支持。

pg_show_custom_settings()

描述：查询Hudi外表参数设置详情。

返回值类型：setof record

 版权所有 © 华为云计算技术有限公司

示例：

```
SELECT * FROM pg_show_custom_settings();
                        name                         |      setting       | unit |      category      |        short_desc        |
extra_desc | context | vartype | source  | min_val | max_val |
 enumvals | boot_val | reset_val | sourcefile | sourceline
-----------------------------------------------------+--------------------+------+--------------------
+--------------------------+------------+---------+---------+---------+---------+
----------+----------+-----------+------------+------------
 hoodie.public.hudi_mor_ft.consume.ending.timestamp | 20230404172329544 |      | Customized Options |
GUC placeholder variable |            | user    | string  | session |         |         |
          |          |           |            |
 hoodie.public.hudi_mor_ft.consume.mode             | incremental       |      | Customized Options | GUC
placeholder variable |            | user    | string  | session |         |         |
          |          |           |            |
 hoodie.public.hudi_mor_ft.consume.start.timestamp  | 20230404172329543 |      | Customized Options |
GUC placeholder variable |            | user    | string  | session |         |         |
          |          |           |            |
(3 rows)
```

hudi_get_options(regclass)

描述：查询Hudi外表的属性信息（hoodie.properties）。以key-value键值对表示。

返回值类型：setof record

示例：

```
select * from hudi_get_options('public.hudi_mor_ft');
                        key                         |
                                                
```

hudi_get_max_commit(regclass)

描述：获取当前Hudi外表最新commit的时间戳和数据写入时间。

返回值类型：record

示例：

```
SELECT * FROM hudi_get_max_commit('public.hudi_mor_ft');
   max_commit   |       write_time
----------------+------------------------
 20221207141822 | 2022-12-07 14:18:30+08
(1 row)
```

hudi_get_commit(regclass, cstring, int)

描述：获取当前Hudi外表从指定commit开始，到第N个commit的时间戳和对应commit数据写入时间。当第N个commit不存在，则返回最新的commit和对应数据写入时间。

返回值类型：record

示例：

```
SELECT * FROM hudi_get_commit('public.hudi_mor_ft', '20230329174744657', 3);
    end_commit     |       write_time
-------------------+------------------------
 20230329174808908 | 2023-08-31 15:43:08+08
(1 row)
```

hudi_sync_task_submit(regclass, regclass)

描述：提交Hudi自动同步任务。第一个入参为同步目标表，第二个入参为HUDI外表。提交任务成功会返回任务的task-id。

返回值类型：text

说明

- 同步目标表必须包含主键，且主键需要与hudi recordkey一致。
- 如果hudi表包含precombine字段，那么同步目标表也必须包含与之对应的字段。
- 如果同步目标表只包含主键（除了主键外没有其他字段），则无法正常提交同步任务。
- 用户需要有同步目标表的insert和update权限、HUDI外表的select权限，否则无法正常提交同步任务。

示例：

```
SELECT hudi_sync_task_submit('public.hudi_sync_i','public.hudi_mor_ft');
       hudi_sync_task_submit
--------------------------------------
 6465efe2-3ea1-0b00-dde5-b57dfb30fffe
(1 row)
```

hudi_sync_task_submit(regclass, regclass, interval)

描述：功能与hudi_sync_task_submit(regclass, regclass)一致。不同之处在于，用户可以额外指定一个interval类型入参，用于指定任务的调度周期，取值范围为5秒至24小时。提交任务成功会返回任务的task-id。

返回值类型：text

 版权所有 © 华为云计算技术有限公司

说明

- 同步目标表必须包含主键，且主键需要与hudi recordkey一致。
- 如果hudi表包含precombine字段，那么同步目标表也必须包含与之对应的字段。
- 如果同步目标表只包含主键（除了主键外没有其他字段），则无法正常提交同步任务。
- 用户需要有同步目标表的insert和update权限、HUDI外表的select权限，否则无法正常提交同步任务。

示例：

```
SELECT hudi_sync_task_submit('public.hudi_sync_i','public.hudi_mor_ft','1 hour');
       hudi_sync_task_submit
--------------------------------------
 6465efe2-3ea1-0b00-dde5-b57dfb30fffe
(1 row)
```

hudi_sync_task_submit(regclass, regclass, text, text)

描述：功能与hudi_sync_task_submit(regclass, regclass)一致，不同之处在于，用户可以额外指定两个text入参，表示用户期望哪些字段被同步。字段之间使用','分隔，支持引号和转义字符的解析。两个text参数的字段数量和顺序应当一致，表示同步字段之间的对应关系。提交任务成功会返回任务的task-id。

返回值类型：text

说明

- 同步目标表必须包含主键，且主键需要与hudi recordkey一致。
- 如果hudi表包含precombine字段，那么同步目标表也必须包含与之对应的字段。
- 如果同步目标表只包含主键（除了主键外没有其他字段），则无法正常提交同步任务。
- 用户需要有同步目标表的insert和update权限、HUDI外表的select权限，否则无法正常提交同步任务。

示例：

```
SELECT hudi_sync_task_submit('public.hudi_sync_i','public.hudi_mor_ft','_hoodie_commit_time, col_bigint,
col_text', '_hoodie_commit_time, col_bigint, col_text');
       hudi_sync_task_submit
--------------------------------------
 646610bc-cdd1-0d00-d07d-b57e89a0fffe
(1 row)
```

hudi_sync_task_submit(regclass, regclass, text, text, interval)

描述：功能与hudi_sync_task_submit(regclass, regclass, text, text)一致，不同之处在于，用户可以额外指定一个interval类型入参，用于指定任务的调度周期，取值范围为5秒至24小时。

返回值类型：text

说明

- 同步目标表必须包含主键，且主键需要与hudi recordkey一致。
- 如果hudi表包含precombine字段，那么同步目标表也必须包含与之对应的字段。
- 如果同步目标表只包含主键（除了主键外没有其他字段），则无法正常提交同步任务。
- 用户需要有同步目标表的insert和update权限、HUDI外表的select权限，否则无法正常提交同步任务。

 版权所有 © 华为云计算技术有限公司

示例：

```
SELECT hudi_sync_task_submit('public.hudi_sync_i','public.hudi_mor_ft','_hoodie_commit_time, col_bigint,
col_text', '_hoodie_commit_time, col_bigint, col_text', '10 minute 30second');
       hudi_sync_task_submit
--------------------------------------
 646610bc-cdd1-0d00-d07d-b57e89a0fffe
(1 row)
```

hudi_show_sync_state()

描述：获取Hudi自动同步任务的同步状态。

返回值类型：setof record

示例：

```
SELECT * FROM hudi_show_sync_state();
    target_tbl     |    source_ftbl     |                      payload_type                       | precombine_key |
latest_commit
--------------------+--------------------+---------------------------------------------------------+----------------
+-------------------
 public.hudi_sync_i | public.hudi_mor_ft | org.apache.hudi.common.model.OverwriteWithLatestAvroPayload |
col_int        | 20230511114021573
(1 row)
```

hudi_sync(regclass, regclass)

描述：存储过程，Hudi自动同步任务调用入口。使用pg_catalog.hudi_sync_task_submit(regclass, regclass)提交的任务会执行该存储过程。执行成功会提示同步的行数和时间戳。

返回值类型：text

示例：

```
CALL hudi_sync('public.hudi_sync_i', 'public.hudi_mor_ft');
NOTICE:  execute full sync
CONTEXT:  PL/pgSQL function hudi_sync(regclass,regclass) line 11 at RETURN
            hudi_sync
--------------------------------------
 sync 1 row up to 20230511114021573.
(1 row)
```

hudi_sync_custom(regclass, regclass, text)

描述：存储过程，Hudi自动同步任务调用入口，支持用户自定义目标表和数据源表的字段同步对应关系。使用pg_catalog.hudi_sync_task_submit(regclass, regclass, text, text)提交的任务会执行该存储过程。其中text为json风格字符串，表示两张表字段间的同步对应关系。执行成功会提示同步的行数和时间戳。

返回值类型：text

示例：

```
CALL hudi_sync_custom('public.hudi_sync_i', 'public.hudi_mor_ft', '{"_hoodie_commit_time" :
"_hoodie_commit_time", "col_bigint" : "col_bigint", "col_text" : "col_text"}');
NOTICE:  execute full sync
CONTEXT:  PL/pgSQL function hudi_sync_custom(regclass,regclass,text) line 14 at RETURN
          hudi_sync_custom
--------------------------------------
 sync 1 row up to 20230511114021573.
(1 row)
```

 版权所有 © 华为云计算技术有限公司

hudi_set_sync_commit(regclass, regclass, text)

描述：设置Hudi自动同步任务首次同步的起点时间戳，避免在已经同步了部分数据的情况下，重新同步已有数据。第一个参数为同步目标表，第二个参数为hudi外表，第三个参数为用户期望的同步起点。该函数需要在提交同步任务之前使用。

返回值类型：text

示例：

```
select hudi_set_sync_commit('public.hudi_sync_i', 'public.hudi_mor_ft', '20230511114021573');
NOTICE:  set sync commit successfully, the next synchronization will start from 20230511114021573
CONTEXT:  referenced column: hudi_set_sync_commit
 hudi_set_sync_commit
----------------------
 20230511114021573
(1 row)
```

说明

用户需要有同步目标表的insert和update权限、HUDI外表的select权限，否则无法正常设置同步进度。

hudi_set_sync_commit(text, text)

描述：设置Hudi自动同步任务下一次同步的起点时间戳，可以用于重复同步历史数据或者跳过某些数据。第一个参数为任务id，第二个参数为用户期望的下一次同步起点。该函数需要在提交同步任务后使用，并且使用前需要暂停任务。

返回值类型：text

示例：

```
select hudi_set_sync_commit('6524c8e3-aae9-0000-5a14-be8ec000fffe', '20230511114021573');
NOTICE:  set sync commit successfully, the next synchronization will start from 20230511114021573
CONTEXT:  referenced column: hudi_set_sync_commit
 hudi_set_sync_commit
----------------------
 20230511114021573
(1 row)
```

说明

- 只有拥有目标任务权限的用户才能成功调用该函数。
- 调用该函数前，目标任务需要处于暂停状态，且至少成功执行过一次。

pg_task_show(text)

描述：查询当前自动调度任务信息，对于hudi同步任务，入参应该是'SQLonHudi'。

返回值类型：setof record

示例：

```
SELECT * FROM pg_task_show('SQLonHudi');
              task_id                |
what                                                
```

```
 64d257e9-1e9b-0d00-3ce3-7e61b5e0fffe | call pg_catalog.hudi_sync_custom('public.hudi_read_target',
'public.hudi_read101', '{"_hoodie_commit_seqno" : "_hoodie_commit_seqno", "id" : "id", "ts" : "ts",
"long_field" : "ts"}'); | SQLonHudi   |     10 | f         | '00:00:10' |           | 2023-08-08 22:58:15.846903+08
|          |          |         5 | 2023-08-08 22:58:15.846903+08 | 2023-08-08 22:58:24.846903+08 |
|           0
(1 row)
```

说明

last_log和failure_times字段用于记录上一次任务的状态。

- last_log于任务结束时刷新，若任务成功，则清空内容；若任务失败则记录任务的失败日志。
- failure_times于时间窗结束时刷新，若任务成功，置为0，失败则累加1，本轮未拉起则不变，可用于推断第一次失败出现的时间。

pg_task_remove(text)

描述：删除某个自动调度任务，入参为该任务的task-id，函数返回被删除任务的个数。

返回值类型：integer

示例：

```
SELECT pg_task_remove('64661705-8ada-0100-d07f-b57e89a0fffe');
 pg_task_remove
----------------
              1
(1 row)
```

pg_task_pause(text)

描述：暂停某个自动调度任务，入参为该任务的task-id，函数返回被暂停任务的个数。

返回值类型：integer

示例：

```
SELECT pg_task_pause('64661705-8ada-0100-d07f-b57e89a0fffe');
 pg_task_pause
---------------
             1
(1 row)
```

pg_task_resume(text)

描述：恢复某个自动调度任务执行，入参为该任务的task-id。函数返回被恢复执行任务的个数。

返回值类型：integer

示例：

```
SELECT pg_task_resume('64661705-8ada-0100-d07f-b57e89a0fffe');
 pg_task_resume
----------------
              1
(1 row)
```

 版权所有 © 华为云计算技术有限公司

pg_task_reset_interval(text, interval)

描述：修改某个同步任务的调度周期，第一个入参为任务的task_id，第二个入用于指定任务的调度周期，取值范围为5秒至24小时。函数返回被修改周期的任务的个数。

返回值类型：integer

示例：

```
select pg_task_reset_interval('64bfd69c-a016-0000-120e-1e802978fffe', '10 hours 30 minutes');
pg_task_reset_interval
------------------------
1
(1 row)
```

7.36 漏斗函数和留存函数

漏斗和留存相关函数仅8.3.0及以上集群版本支持。

背景信息

漏斗函数与**留存函数**都是常见的用户行为分析工具，广泛应用于产品分析和数据分析领域，尤其是在产品经理、数据科学家和市场营销人员的工作中。这些函数的目标是帮助分析用户的行为路径、理解用户的流失和转化问题，以及评估产品的长期吸引力和用户忠诚度。

- 漏斗分析：一种用于分析用户在执行某一目标过程中所经历的各个步骤的分析方法。通常，漏斗分析可以帮助产品团队或营销人员了解用户在使用产品或服务时的行为轨迹，以及每个步骤中的转化率和流失情况。漏斗分析的关键在于定义一个“漏斗”，即用户从进入某个流程（如访问网站、注册账户、选购商品等）到完成最终目标（如完成支付、完成注册等）的整个路径。
- 留存分析：留存分析关注的是在某一特定时间段后，仍然活跃使用产品的用户比例。具体来说，留存分析计算的是从某一时间点（例如，用户注册、首次下单、首次使用等）开始，经过一段时间后，仍然继续使用产品的用户数量。常见的留存分析函数通常涉及的是“日留存率”、“周留存率”、“月留存率”等，它们帮助分析用户在首次接触后是否会继续使用该产品。

漏斗函数 windowfunnel

函数描述：windowfunnel函数用于在滑动的时间窗口中搜索事件列表并计算条件匹配的事件列表的最大长度。

DWS根据用户定义的事件列表，从第一个事件开始匹配，依次做有序最长的匹配，返回匹配的最大长度。一旦匹配失败，结束整个匹配。具体介绍如下：

假设在窗口足够大的条件下：

- 条件事件为c1，c2，c3，而用户数据为c1，c2，c3，c4，最终匹配到c1，c2，c3，函数返回值为3。
- 条件事件为c1，c2，c3，而用户数据为c4，c3，c2，c1，最终匹配到c1，函数返回值为1。
- 条件事件为c1，c2，c3，而用户数据为c4，c3，最终没有匹配到事件，函数返回值为0。

函数语法

 版权所有 © 华为云计算技术有限公司

```
windowFunnel(window, mode, timestamp, cond1, cond2, ..., condN)
```

入参说明

- window：bigint类型。滑动的时间窗口的大小，指从第一个事件开始，往后推移的时间窗口大小，单位为秒。
- mode：text类型。目前仅支持default模式，其他模式报错处理。Default模式是指在同一个窗口期内，从第一个事件开始匹配，尽量匹配多的事件。
- timestamp：事件发生的时间范围，支持timestamp without time zone、timestamp with time zone、date、int、bigint类型。
- cond：变长boolean数组。指当前Tuple的数据满足事件的哪个步骤。DWS只支持1~32个condition，不在此范围报错处理。

返回值

level：int类型。条件匹配的事件列表的最大长度。

留存函数 retention

函数描述：retention函数可以将一组条件作为参数，分析事件是否满足该条件。

函数语法

```
retention(cond1, cond2, ..., cond32);
```

入参说明

cond：变长boolean数组，最大长度32，用来表示事件是否满足特定条件。DWS只支持1~32个condition，不在此范围报错处理。

返回值

retention condition：tinyint数组类型。返回结果的表达式，与入参cond长度一致，若第cond1和condi条件满足，则返回值第i个值为1，否则为0。

留存扩展函数

DWS留存扩展函数支持：range_retention_count和range_retention_sum函数。

- **range_retention_count**

  函数描述：记录每个用户的留存情况，该函数返回数组，可以作为range_retention_sum函数的入参。

  函数语法

  ```
  range_retention_count(is_first, is_active, dt, retention_interval, retention_granularity, output_format)
  ```

  入参说明

  - is_first：bool类型，是否符合初始行为，true表示符合， false表示不符合 。
  - is_active：bool类型，是否符合留存行为，true表示符合， false表示不符合 。
  - dt：date类型，发生行为的日期。
  - retention_interval：数组类型，表示留存间隔，最多支持15个留存间隔。例如ARRAY[1,3,5,7,15,30]。
  - retention_granularity：text类型，表示留存分析粒度，执行日（day）、周（week）、月（month）三种。

 版权所有 © 华为云计算技术有限公司

  - output_format：text类型，表示输出格式，支持normal（默认）和expand（可取得每日留存明细）两种。

  返回值：用户留存情况BIGINT数组。

- **range_retention_sum**

  函数描述：汇总计算所有用户每天(周/月)的留存率。

  函数语法

  ```
  range_retention_sum(range_retention_count_result)
  ```

  入参说明：range_retention_count函数返回值。

  返回值：用户留存统计情况text数组。

示例

创建表funnel_test：

```
CREATE TABLE IF NOT EXISTS funnel_test
(
    user_id INT ,
    event_type TEXT,
    event_time TIMESTAMP,
    event_timez TIMESTAMP WITH TIME ZONE,
    event_time_int BIGINT
);
```

插入数据：

```
INSERT INTO funnel_test VALUES
(1,'Browse','2021-01-31 11:00:00', '2021-01-31 11:00:00+08', 10),
(1,'Click','2021-01-31 11:10:00', '2021-01-31 11:10:00+07', 20),
(1,'Order','2021-01-31 11:20:00', '2021-01-31 11:20:00+06', 30),
(1,'Pay','2021-01-31 11:30:00', '2021-01-31 11:30:00+05', 40),
(2,'Order','2021-01-31 11:00:00', '2021-01-31 11:00:00+08', 11),
(2,'Pay','2021-01-31 11:10:00', '2021-01-31 11:10:00+08', 12),
(1,'Browse','2021-01-31 11:00:00', '2021-01-31 11:00:00+01', 50),
(3,'Browse','2021-01-31 11:20:00', '2021-01-31 11:20:00-04', 30),
(3,'Click','2021-01-31 12:00:00', '2021-01-31 12:00:00-04', 80),
(4,'Browse','2021-01-31 11:50:00', '2021-01-31 11:50:00-01', 1000),
(4,'Pay','2021-01-31 12:00:00', '2021-01-31 12:00:00-02', 900),
(4,'Order','2021-01-31 12:00:00', '2021-01-31 12:00:00-03', 1001),
(4,'Click','2021-01-31 12:00:00', '2021-01-31 12:00:00-04', 1001),
(5,'Browse','2021-01-31 11:50:00', '2021-01-31 11:50:00+08', NULL),
(5,'Click','2021-01-31 12:00:00', '2021-01-31 12:00:00+08', 776),
(5,'Order','2021-01-31 11:10:00', '2021-01-31 11:10:00+08', 999),
(6,'Browse','2021-01-31 11:50:00', '2021-01-31 11:50:00+01', -1),
(6,'Click','2021-01-31 12:00:00', '2021-01-31 12:00:00+02', -2),
(6,'Order','2021-01-31 12:10:00', '2021-01-31 12:00:00+03', -3);
```

计算每个用户的漏斗情况。返回结果如下，其中level=0表示用户在窗口期内匹配最大事件深度为0，level=1表示用户在窗口期内匹配最大事件深度为1：

```
SELECT
    user_id,
  windowFunnel(
    0, 'default', event_timez,
    event_type = 'Browse', event_type = 'Click', event_type = 'Order', event_type = 'Pay'
  ) AS level
FROM funnel_test
GROUP BY user_id
ORDER by user_id;

 user_id | level
---------+-------
       1 |     1
       2 |     0
```

 版权所有 © 华为云计算技术有限公司

```
      3 |     1
      4 |     1
      5 |     1
      6 |     1
(6 rows)
```

计算每个用户的漏斗情况，指定滑动的时间窗口的大小为NULL，返回报错：

```
SELECT
    user_id,
  windowFunnel(
    NULL, 'default', event_time,
    event_type = 'Browse', event_type = 'Click', event_type = 'Order', event_type = 'Pay'
  ) AS level
FROM funnel_test
GROUP BY user_id
ORDER by user_id;
ERROR:  Invalid parameter : window length or mode is null.
```

计算每个用户的漏斗情况，指定多个条件：

```
SELECT
    user_id,
  windowFunnel(
    40, 'default', date(event_time),
    true, true, false, true
  ) AS level
FROM funnel_test
GROUP BY user_id
ORDER by user_id;
 user_id | level
---------+-------
       1 |     2
       2 |     2
       3 |     2
       4 |     2
       5 |     2
       6 |     2
(6 rows)
```

分析用户的留存情况：

```
SELECT
    user_id,
  retention(
    event_type = 'Browse', event_type = 'Click', event_type = 'Order', event_type = 'Pay'
    ) AS r
FROM funnel_test
GROUP BY user_id
ORDER BY user_id ASC;
 user_id |     r
---------+-----------
       1 | {1,1,1,1}
       2 | {0,0,0,0}
       3 | {1,1,0,0}
       4 | {1,1,1,1}
       5 | {1,1,1,0}
       6 | {1,1,1,0}
(6 rows)
```

分析用户的留存情况，指定第一个时间为false：

```
SELECT
    user_id,
  retention(
    false, event_type = 'Browse', event_type = 'Click', event_type = 'Order', event_type = 'Pay'
    ) AS r
FROM funnel_test
GROUP BY user_id
ORDER BY user_id ASC;
```

 版权所有 © 华为云计算技术有限公司

```
 user_id |      r
---------+-------------
       1 | {0,0,0,0,0}
       2 | {0,0,0,0,0}
       3 | {0,0,0,0,0}
       4 | {0,0,0,0,0}
       5 | {0,0,0,0,0}
       6 | {0,0,0,0,0}
(6 rows)
```

分析用户的留存情况总和：

```
SELECT sum(r[1]), sum(r[2]), sum(r[3]), sum(r[4])
FROM
(
    SELECT
    retention(event_type = 'Browse', event_type = 'Click', event_type = 'Order', event_type = 'Pay') AS r
    FROM funnel_test
    GROUP BY user_id
);
 sum | sum | sum | sum
-----+-----+-----+-----
   5 |   5 |   4 |   2
(1 row)
```

创建表retention_test：

```
CREATE TABLE retention_test(
uid INT,
event TEXT,
event_time TIMESTAMP
);
```

插入数据：

```
INSERT INTO retention_test VALUES
(1, 'pay', '2024-05-01'),
(1, 'login', '2024-05-01'),
(1, 'pay', '2024-05-02'),
(1, 'login', '2024-05-02'),
(2, 'login', '2024-05-01'),
(3, 'login', '2024-05-02'),
(3, 'pay', '2024-05-03'),
(3, 'pay', '2024-05-04');
```

统计每个用户在1，2天后的付费留存率：

```
WITH retention_count_info AS (
    SELECT
        uid,
        range_retention_count(event = 'login', event = 'pay',
                              DATE(event_time), array[1, 2], 'day') AS info
    FROM
        retention_test
    GROUP BY
        uid
), retention_sum AS (
         SELECT regexp_split_to_array(unnest(range_retention_sum(info)), ',') AS s
         FROM retention_count_info
     ) SELECT to_date(s[1]::int) AS login_date,
              s[3]::numeric / s[2]::numeric AS retention_rate1,
              s[4]::numeric / s[2]::numeric AS retention_rate2
FROM retention_sum
ORDER BY login_date;
     login_date      |     retention_rate1     |     retention_rate2
---------------------+-------------------------+-------------------------
 2024-05-01 00:00:00 | .50000000000000000000 | 0.00000000000000000000
 2024-05-02 00:00:00 | .50000000000000000000 |  .50000000000000000000
(2 rows)
```


版权所有 © 华为云计算技术有限公司

7.37 EXTERNAL SCHEMA 系统函数

EXTERNAL SCHEMA相关系统函数仅8.3.0及以上集群版本支持。

pg_get_external_schema_table_list(schema_name text)

描述：获取external schema的全量表名信息，仅支持lakeformation。

入参: schema_name为external schema名称。

返回值类型：setof record

示例：

```
SELECT * FROM pg_get_external_schema_table_list('ex_lf');
 database_name |           table_name
---------------+--------------------------------
 dws-ci-hlt-db | lf_hms_lineitem_parq
 dws-ci-hlt-db | lf_hms_lineitem_parq
 dws-ci-hlt-db | lf_hms_part_parq
 dws-ci-hlt-db | lf_hms_part_parq
 dws-ci-hlt-db | lf_hms_supplier_parq
 dws-ci-hlt-db | lf_hms_partsupp_parq
 dws-ci-hlt-db | lf_hms_partsupp_parq
 dws-ci-hlt-db | lf_hms_nation_parq
 dws-ci-hlt-db | lf_hms_region_parq
 dws-ci-hlt-db | lf_hms_customer_parq
(10 rows)
```

pg_get_external_schema_table_options(schema_name text, table_name text)

描述：获取external schema表的属性信息。

入参: schema_name为external schema名称，table_name为表名称。

返回值类型：setof record

示例：

```
SELECT * FROM pg_get_external_schema_table_options('ex_lf', 'test_lf');
 option_name |            option_value
-------------+-------------------------------------
 encoding    | utf8
 format      | parquet
 foldername  | /***/***/***
(3 rows)
```

pg_get_external_schema_table_col(schema_name text, table_name text)

描述：获取external schema表的列信息。

入参: schema_name为external schema名称，table_name为表名称。

返回值类型：setof record

示例：

```
SELECT * FROM pg_get_external_schema_table_col('ex_lf', 'test_lf');
       col_name        |   col_type    | part_col
-----------------------+---------------+----------
```

 版权所有 © 华为云计算技术有限公司

```
field_smallint          | smallint      | f
field_int               | int           | f
field_integer           | int           | f
field_bigint            | bigint        | f
field_float             | float         | f
field_double            | double        | f
field_double_precision | double         | f
field_decimal           | decimal(10,0) | f
field_numeric           | decimal(10,0) | f
field_timestamp         | timestamp     | f
field_date              | date          | f
field_varchar           | varchar(5)    | f
field_char              | char(5)       | f
field_boolean           | boolean       | f
field_string            | string        | f
(15 rows)
```

7.38 存算分离函数

7.38.1 数据共享函数

pgxc_group_add_subscription(src_vw_name, target_vw_name)

描述：存算分离架构下，计算组（Virtual Warehouse，以下简称VW）之间建立KV订阅关系，建立了订阅关系之后，消费者VW的KVcache会定期的从生产者的obs cudesc快照目录增量同步sst文件到本地，后续就可以支持在消费者VW中跨VW查询生产者VW中的表。 创建的订阅关系可以在pgxc_group_subscription表中查询到记录。该函数仅9.0.3及以上集群版本支持。

返回值类型：void

返回信息如下：

表 7-87 pgxc_group_add_subscription(src_vw_name, target_vw_name)返回字段

名称	类型	描述
src_vw_name	text	生产者VW名称，通常作为写入数据的VW。
target_vw_name	text	消费者VW名称，通常作为读取数据的VW。

示例：

```
SELECT pgxc_group_add_subscription('write_group', 'single_read_group');
pgxc_group_add_subscription
-----------------

(1 row)
```

pgxc_group_drop_subscription(src_vw_name, target_vw_name)

描述：存算分离架构下，删除VW之间已经建立的KV订阅关系，删除操作是删除pgxc_group_subscription表的记录， 删除KV订阅关系后，消费者VW无法继续跨VW查询生产者VW中的表。该函数仅9.0.3及以上集群版本支持。

 版权所有 © 华为云计算技术有限公司

返回值类型：void

返回信息如下：

表 7-88 pgxc_group_drop_subscription(src_vw_name, target_vw_name)返回字段

名称	类型	描述
src_vw_name	text	生产者VW名称，通常作为写入数据的VW。
target_vw_name	text	消费者VW名称，通常作为读取数据的VW。

示例：

```
SELECT  pgxc_group_drop_subscription('write_group', 'single_read_group');
pgxc_group_drop_subscription
------------------------------

(1 row)
```

7.38.2 异步 IO 限流

异步IO限流功能仅9.1.1.200及以上集群版本支持。

obs_add_limit_config(access_key, bucket, metric, speed_ratio, expires_s)

描述：存算分离架构下通过添加限流配置来限制OBS请求，达到降低OBS指标的目的。

函数输入参数如下：

名称	类型	描述
access_key	text	用于访问OBS服务的access_key，添加时需要确保系统表PG_TABLESPACE中存在对应的OBS服务配置或者PG_FOREIGN_SERVER中存在对应的OBS服务配置。
bucket	text	访问的OBS桶，该字段可为空字符串，为空字符串时表示需要限制整个OBS租户。
metric	text	被限制的指标，指标包括：conn，read_tps，write_tps，list_tps，total_tps，read_bps，write_bps，total_bps。
speed_ratio	bigint	限速比例。取值范围为1~100，默认值为80，100表示不限速。
expires_s	bigint	限速配置有效期。取值范围 1~3600 ，单位为秒， 默认值为60s。

返回值类型：record

 版权所有 © 华为云计算技术有限公司

函数返回字段如下：

名称	类型	描述
nodename	text	执行语句的节点名称。
result	boolean	限流配置是否添加成功。 ● t：添加成功。 ● f：添加失败。

示例：

添加限流配置来限制OBS请求。

XXXXXXXXXXXXX表示访问OBS服务的access_key，访问的OBS桶名为test， 被限制的指标conn， 限速比为80，限速配置的有效时间为100秒。使用时请根据业务需求替换为实际值。

```
SELECT * FROM obs_add_limit_config('XXXXXXXXXXXXX','test','conn','80','100');
   nodename   | result
--------------+--------
 coordinator1 | t
(1 row)
```

expires_s参数缺省时，表示限速配置的有效时间为默认值60秒。

```
SELECT * FROM obs_add_limit_config('XXXXXXXXXXXXX','test','conn','80');
   nodename   | result
--------------+--------
 coordinator1 | t
(1 row)
```

speed_ratio与expires_s参数缺省时，分别表示限速比为80，限速配置的有效时间为60秒。

```
SELECT * FROM obs_add_limit_config('XXXXXXXXXXXXX','test','conn');
   nodename   | result
--------------+--------
 coordinator1 | t
(1 row)
```

obs_delete_limit_config(access_key, bucket, metric)

描述：删除当前节点中的OBS限流配置。

函数输入参数如下：

名称	类型	描述
access_key	text	已添加的限流配置的access_key。
bucket	text	已添加的限流配置中的OBS桶名称。
metric	text	已添加的限流配置中的指标。

返回值类型：record

 版权所有 © 华为云计算技术有限公司

函数返回字段如下：

名称	类型	描述
nodename	text	执行语句的节点名称。
result	boolean	OBS限流配置是否删除成功。 ● t：删除成功。 ● f：删除失败。

示例：

删除当前节点中的OBS限流配置。

XXXXXXXXXXXXX表示访问OBS服务的access_key，访问的OBS桶名为test， 被限制的指标conn。使用时请根据业务需求替换为实际值。

```
SELECT * FROM obs_delete_limit_config('XXXXXXXXXXXXX','test','conn');
   nodename   | result
--------------+--------
 coordinator1 | t
(1 row)
```

obs_clear_limit_configs()

描述：清除当前节点中的所有OBS限流配置。

返回值类型：record

函数返回字段如下：

名称	类型	描述
nodename	text	执行语句的节点名称。
result	boolean	限流配置是否清理成功。 ● t：清理成功。 ● f：清理失败。

示例：

清除当前节点中的所有OBS限流配置。

```
SELECT * FROM obs_clear_limit_configs();
   nodename   | result
--------------+--------
 coordinator1 | t
(1 row)
```

obs_query_limit_configs()

描述：查询当前节点中的所有OBS限流配置。

返回值类型：record

 版权所有 © 华为云计算技术有限公司

函数返回字段如下：

名称	类型	描述
nodename	text	执行语句的节点名称。
access_key	bool	已添加的限流配置的access_key。
bucket	text	OBS桶名称。
metric	text	被限制的指标。
speed_ratio	bigint	限速比例。
expires_s	bigint	限速配置的有效期。
is_expired	boolean	限速配置是否过期。 ● t：已过期 ● f：未过期
expiration_time	timestamp	限速配置的过期时间。
create_time	timestamp	限速配置首次添加时间。
update_time	timestamp	限速配置更新时间，如果当前配置已经存在，后续直接更新update_time。

示例：

查询当前节点中的所有OBS限流配置。

查询显示名称为test的OBS桶，被限制的指标conn，其限速比为80，限速配置的过期时间为100秒，当前配置已过期，限速配置首次添加时间为2025-10-11 09:09:43+08等信息。

```
SELECT * FROM obs_query_limit_configs();
   nodename   |      access_key      | bucket | metric | speed_ratio | expires_s | is_expired |     expiration_time
|      create_time       |       update_time
--------------+----------------------+--------+--------+-------------+-----------+------------+-------------------------
+-------------------------+-------------------------
 coordinator1 | xxxxxxxxxxxxxxxxxxxx|  test  | conn   |          80 |       100 | t          | 2025-10-11 09:11:23+08
| 2025-10-11 09:09:43+08 | 2025-10-11 09:09:43+08
```

pgxc_obs_add_limit_config(target, access_key, bucket, metric, speed_ratio, expires_s)

描述：存算分离架构下通过添加限流配置来限制OBS请求，达到降低OBS指标的目的。

函数输入参数如下：

 版权所有 © 华为云计算技术有限公司

名称	类型	描述
target	text	目标节点，默认为all。取值如下： ● all，表示所有节点。 ● cn，表示全部CN节点。 ● coordinator，表示全部CN节点。 ● dn，表示全部DN节点。 ● datanode，表示全部DN节点。
access_key	text	用于访问OBS服务的access_key，添加时需要确保系统表PG_TABLESPACE中存在对应的OBS服务配置或者PG_FOREIGN_SERVER中存在对应的OBS服务配置。
bucket	text	访问的OBS桶，该字段可为空字符串，为空字符串时表示需要限制整个OBS租户。
metric	text	被限制的指标，指标包括：conn，read_tps，write_tps，list_tps，total_tps，read_bps，write_bps，total_bps。
speed_ratio	bigint	限速比例，该比例是相对不限速情况下的比值。取值范围为 1-100，默认值为 80。
expires_s	bigint	限速配置有效期。取值范围为1~3600，单位为秒， 默认值为60s。

返回值类型：record

函数返回字段如下：

名称	类型	描述
total_instances	int	执行添加配置的节点总数。
success_instances	int	添加限流配置成功的节点数量。
failed_instances	int	添加限流配置失败的节点数量。
failed_nodes	text	添加限流配置失败的节点，会显示添加限流配置失败的具体节点信息。

示例：

```
SELECT * FROM pgxc_obs_add_limit_config('all','XXXXXXXXXXX','test','conn', 70, 1200);
 total_instances | success_instances | failed_instances | failed_nodes
-----------------+-------------------+------------------+--------------
               4 |                 4 |                0 |
(1 row)
```


版权所有 © 华为云计算技术有限公司

expires_s参数缺省时，表示限速配置的有效时间为默认值60秒。

```
select * from pgxc_obs_add_limit_config('all','XXXXXXXXXXXX','test','conn', 70);
 total_instances | success_instances | failed_instances | failed_nodes
-----------------+-------------------+------------------+--------------
               4 |                 4 |                0 |
(1 row)
```

speed_ratio与expires_s参数缺省时，分别表示限速比为80，限速配置的有效时间为60秒。

```
SELECT * FROM pgxc_obs_add_limit_config('all','XXXXXXXXXXXX','test','conn');
 total_instances | success_instances | failed_instances | failed_nodes
-----------------+-------------------+------------------+--------------
               4 |                 4 |                0 |
(1 row)
```

pgxc_obs_delete_limit_config(target,access_key, bucket, metric)

描述：批量删除目标节点中的限流配置。

函数输入参数如下：

名称	类型	描述
target	text	目标节点，默认为all。取值如下： ● all，表示所有节点。 ● cn，表示全部CN节点。 ● coordinator，表示全部CN节点。 ● dn，表示全部DN节点。 ● datanode，表示全部DN节点。
access_key	text	已添加的限流配置的access_key。
bucket	text	已添加的限流配置中的bucket。
metric	text	已添加的限流配置中的指标。

返回值类型：record

函数返回字段如下：

名称	类型	描述
total_instances	int	执行删除配置的节点总数。
success_instances	int	删除成功的节点数量。
failed_instances	int	删除失败的节点数量。
failed_nodes	text	删除失败的节点，显示删除失败的具体节点信息。

 版权所有 © 华为云计算技术有限公司

示例：

```
SELECT * FROM pgxc_obs_delete_limit_config('all','XXXXXXXXXXX','test','conn');
 total_instances | success_instances | failed_instances | failed_nodes
-----------------+-------------------+------------------+--------------
               4 |                 4 |                0 |
(1 row)
```

pgxc_obs_clear_limit_configs(target)

描述：批量清除目标节点的所有OBS限流配置。

函数输入参数：target，类型：text，表示目标节点，默认为all。取值如下：

- all，表示所有节点。
- cn，表示全部CN节点。
- coordinator，表示全部CN节点。
- dn，表示全部DN节点。
- datanode，表示全部DN节点。

返回值类型：record

函数返回字段如下：

名称	类型	描述
total_instances	int	执行清理配置的节点总数。
success_instances	int	清除成功的节点数量。
failed_instances	int	清除失败的节点数量。
failed_nodes	text	清除失败的节点，显示清除失败的具体节点信息。

示例：

清除所有节点的所有OBS限流配置。

```
SELECT * FROM pgxc_obs_clear_limit_configs('all');
 total_instances | success_instances | failed_instances | failed_nodes
-----------------+-------------------+------------------+--------------
               4 |                 4 |                0 |
(1 row)
```

target参数缺省时，默认清除所有节点的所有OBS限流配置。

```
SELECT * FROM pgxc_obs_clear_limit_configs();
 total_instances | success_instances | failed_instances | failed_nodes
-----------------+-------------------+------------------+--------------
               4 |                 4 |                0 |
(1 row)
```

 版权所有 © 华为云计算技术有限公司

pgxc_obs_query_limit_configs(target)

描述：查询目标节点中的所有限流配置。

函数输入参数：target，类型：text，表示目标节点，默认为all。取值如下：

- all，表示所有节点。
- cn，表示全部CN节点。
- coordinator，表示全部CN节点。
- dn，表示全部DN节点。
- datanode，表示全部DN节点。

返回值类型：record

函数返回字段如下：

名称	类型	描述
nodename	text	执行语句的节点名称。
access_key	bool	限流配置是否删除成功。 ● t：删除成功。 ● f：删除失败。
bucket	text	OBS桶名称。
metric	text	被限制的指标。
speed_ratio	bigint	限速比例。
expires_s	bigint	限速配置的有效期。
is_expired	boolean	限速配置是否过期。 ● t : 已过期。 ● f : 未过期。
expiration_time	timestamp	限速配置的过期时间。
create_time	timestamp	限速配置首次添加时间。
update_time	timestamp	限速配置更新时间，如果当前配置已经存在，后续直接更新update_time。

示例：

查询所有节点中的所有限流配置。

```
SELECT * FROM pgxc_obs_query_limit_configs('all');
   nodename   |      access_key      |     bucket     | metric | speed_ratio | expires_s | is_expired |
expiration_time     |      create_time       |      update_time
--------------+----------------------+----------------+--------+-------------+-----------+------------
+------------------------+------------------------+------------------------
 datanode3    | XXXXXXXXXXXXX        | test           | conn   |          70 |      1200 | f          | 2025-10-11
10:07:37+08 | 2025-10-11 09:47:37+08 | 2025-10-11 09:47:37+08
```

 版权所有 © 华为云计算技术有限公司

```
 datanode2    | XXXXXXXXXXXXX        | test          | conn   |          70 |      1200 | f          | 2025-10-11
10:07:37+08 | 2025-10-11 09:47:37+08 | 2025-10-11 09:47:37+08
 datanode1    | XXXXXXXXXXXXX        | test          | conn   |          70 |      1200 | f          | 2025-10-11
10:07:37+08 | 2025-10-11 09:47:37+08 | 2025-10-11 09:47:37+08
 coordinator1 | XXXXXXXXXXXXX        | test          | conn   |          70 |      1200 | f          | 2025-10-11
10:07:37+08 | 2025-10-11 09:47:37+08 | 2025-10-11 09:47:37+08
(4 rows)
```

未指定函数入参target，即使用默认值查询节点中的所有限流配置。

```
SELECT * FROM pgxc_obs_query_limit_configs();
   nodename   |      access_key      |     bucket     | metric | speed_ratio | expires_s | is_expired |
expiration_time     |      create_time       |      update_time
--------------+----------------------+----------------+--------+-------------+-----------+------------
+------------------------+------------------------+------------------------
 datanode3    | XXXXXXXXXXXXX        | test          | conn   |          70 |      1200 | f          | 2025-10-11
10:07:37+08 | 2025-10-11 09:47:37+08 | 2025-10-11 09:47:37+08
 datanode2    | XXXXXXXXXXXXX        | test          | conn   |          70 |      1200 | f          | 2025-10-11
10:07:37+08 | 2025-10-11 09:47:37+08 | 2025-10-11 09:47:37+08
 datanode1    | XXXXXXXXXXXXX        | test          | conn   |          70 |      1200 | f          | 2025-10-11
10:07:37+08 | 2025-10-11 09:47:37+08 | 2025-10-11 09:47:37+08
 coordinator1 | XXXXXXXXXXXXX        | test          | conn   |          70 |      1200 | f          | 2025-10-11
10:07:37+08 | 2025-10-11 09:47:37+08 | 2025-10-11 09:47:37+08
(4 rows)
```

查询所有DN节点中的所有限流配置。

```
SELECT * FROM pgxc_obs_query_limit_configs('dn');
   nodename   |      access_key      |     bucket     | metric | speed_ratio | expires_s | is_expired |
expiration_time     |      create_time       |      update_time
--------------+----------------------+----------------+--------+-------------+-----------+------------
+------------------------+------------------------+------------------------
 datanode3    | XXXXXXXXXXXXX        | test          | conn   |          70 |      1200 | f          | 2025-10-11
10:07:37+08 | 2025-10-11 09:47:37+08 | 2025-10-11 09:47:37+08
 datanode2    | XXXXXXXXXXXXX        | test          | conn   |          70 |      1200 | f          | 2025-10-11
10:07:37+08 | 2025-10-11 09:47:37+08 | 2025-10-11 09:47:37+08
 datanode1    | XXXXXXXXXXXXX        | test          | conn   |          70 |      1200 | f          | 2025-10-11
10:07:37+08 | 2025-10-11 09:47:37+08 | 2025-10-11 09:47:37+08
(3 rows)
```

查询所有CN节点中的所有限流配置。

```
SELECT * FROM pgxc_obs_query_limit_configs('cn');
   nodename   |      access_key      |     bucket     | metric | speed_ratio | expires_s | is_expired |
expiration_time     |      create_time       |      update_time
--------------+----------------------+----------------+--------+-------------+-----------+------------
+------------------------+------------------------+------------------------
 coordinator1 | XXXXXXXXXXXXX        | test          | conn   |          70 |      1200 | f          | 2025-10-11
10:07:37+08 | 2025-10-11 09:47:37+08 | 2025-10-11 09:47:37+08
(1 row)
```

7.38.3 LStore 存储引擎函数

本节所有函数仅适用LStore存储引擎，需开启LSM引擎（enable_lsm_engine参数设置为on）后使用。

LStore存储引擎函数仅9.1.2及以上集群版本支持。

gs_kvstore_checkpoint(group_name text)

描述：执行VW级的checkpoint将数据落盘。

返回值类型：bool

函数返回字段：

 版权所有 © 华为云计算技术有限公司

表 7-89 gs_kvstore_checkpoint(write_group)返回字段

名称	类型	描述
result	bool	是否完成checkpoint。

示例：

```
SELECT gs_kvstore_checkpoint('write_group');
```

kvstore_checkpoint(vgroupId uint64)

描述：VW中每个DN都会执行checkpoint进行数据落盘。

返回值类型：bool

函数返回字段：

表 7-90 kvstore_checkpoint(vgroupId)返回字段

名称	类型	描述
result	bool	各DN是否完成checkpoint。

示例：

```
SELECT kvstore_checkpoint(10875761377281);
```

gs_startup_kvstore(group_name text)

描述：启动对应VW的kvstore存储。

返回值类型：bool

函数返回字段：

表 7-91 gs_startup_kvstore(group_name)返回字段

名称	类型	描述
result	bool	指定VW的kvstore是否启动完成。

示例：

```
SELECT gs_startup_kvstore('write_group');
```

startup_kvstore(group_name text, vgroupId uint64)

描述：指定VW的每个DN启动kvstore。

返回值类型：bool

函数返回字段：

 版权所有 © 华为云计算技术有限公司

表 7-92 startup_kvstore(group_name)返回字段

名称	类型	描述
result	bool	指定VW的各DN的kvstore是否启动完成。

示例：

```
SELECT startup_kvstore('write_group',10875761377281);
```

gs_shutdown_kvstore(group_name text)

描述：关闭对应VW的kvstore存储。

返回值类型：bool

函数返回字段：

表 7-93 gs_shutdown_kvstore(group_name)返回字段

名称	类型	描述
result	bool	指定VW的kvstore是否关闭完成。

示例：

```
SELECT gs_shutdown_kvstore('write_group');
```

shutdown_kvstore(vgroupId uint64)

描述：指定VW的每个DN上关闭kvstore。

返回值类型：bool

函数返回字段：

表 7-94 shutdown_kvstore(vgroupId)返回字段

名称	类型	描述
result	bool	指定VW的DN上kvstore是否关闭完成。

示例：

```
SELECT shutdown_kvstore(10875761377281);
```

gs_write_kvstore_meta_to_obs(group_name1 text, group_name2 text)

描述：LStore存储引擎扩缩容场景下，将当前VW的元信息存储到OBS。

返回值类型：bool

函数返回字段：

 版权所有 © 华为云计算技术有限公司

表 7-95 gs_write_kvstore_meta_to_obs(group_name1, group_name2)返回字段

名称	类型	描述
result	bool	元信息转储到OBS是否完成。

示例：

```
SELECT gs_write_kvstore_meta_to_obs('write_group','dst_resize_group');
```

write_kvstore_meta_to_obs(src_vgroupId uint64, dst_vgroupId uint64)

描述：LStore存储引擎扩缩容场景下，将当前VW中每个DN执行元信息转储到OBS。

返回值类型：bool

函数返回字段：

表 7-96 write_kvstore_meta_to_obs(srcVgroupId, dstVgroupId)返回字段

名称	类型	描述
result	bool	指定VW的DN元信息转储到OBS是否完成。

示例：

```
SELECT write_kvstore_meta_to_obs(10875761377281, 10875761377282);
```

gs_get_kvstore_meta_from_obs(src_group_name text, dst_group_name text)

描述：LStore存储引擎扩缩容场景下，从OBS获取对应VW的元信息。

返回值类型：bool

函数返回字段：

表 7-97 gs_get_kvstore_meta_from_obs(src_group_name, dst_group_name)返回字段

名称	类型	描述
result	bool	从OBS获取元信息是否完成。

示例：

```
SELECT gs_get_kvstore_meta_from_obs('write_group','dst_resize_group');
```

get_kvstore_meta_from_obs(src_vgroupId uint64, dst_vgroupId uint64)

描述：用于LStore存储引擎扩缩容场景下，当前VW中的DN从OBS获取对应VW的元信息。

返回值类型：bool

 版权所有 © 华为云计算技术有限公司

函数返回字段：

表 7-98 get_kvstore_meta_from_obs(src_vgroupId, dst_vgroupId)返回字段

名称	类型	描述
result	bool	指定VW的DN从OBS获取元信息是否完成。

示例：

```
SELECT get_kvstore_meta_from_obs(10875761377281, 10875761377282);
```

gs_lstore_lsmbackupstate()

描述：LStore存储引擎备份场景下，检查当前备份是否处于可用状态。

返回值类型：integer

函数返回字段：

表 7-99 gs_lstore_lsmbackupstate()返回字段

名称	类型	描述
result	integer	可用状态返回为0，否则返回为-1。

示例：

```
SELECT gs_lstore_lsmbackupstate();
```

gs_lstore_beginbackup()

描述：LStore存储引擎备份场景下，在开始备份前执行准备操作，包括确认环境状态和备份中禁止删除、合并sst文件。

返回值类型：integer

函数返回字段：

表 7-100 gs_lstore_beginbackup()返回字段

名称	类型	描述
result	integer	执行成功返回为0，否则返回为-1。

示例：

```
execute direct on(datanode1) 'select gs_lstore_beginbackup()';
 gs_lstore_beginbackup
-----------------------
                     0
(1 row)
```

 版权所有 © 华为云计算技术有限公司

gs_lstore_endbackup()

描述：LStore存储引擎备份场景下，在备份操作后执行恢复操作，包括确认环境状态和允许删除、合并sst文件。

返回值类型：integer

函数返回字段：

表 7-101 gs_lstore_endbackup()返回字段

名称	类型	描述
result	integer	执行成功返回为0，否则返回为-1。

示例：

```
execute direct on(datanode1) 'select gs_lstore_endbackup()';
 gs_lstore_endbackup
---------------------
                   0
(1 row)
```

gs_lstore_getlivefiles()

描述：LStore存储引擎备份场景下，执行备份操作并返回备份内容。

返回值类型：record

函数返回字段：

表 7-102 gs_lstore_getlivefiles()返回字段

名称	类型	描述
barrierlsn	text	备份barrier点的lsn号。
ftypebyte	bigint	备份的文件集合类型。
metafilepath	text	备份meta文件的路径。
op	text	备份操作的类型。
file_name	text	文件名称。
file_type	text	文件类型。
file_size	bigint	文件大小。
file_checksum	text	文件校验。
in_obs	bool	文件是否在OBS上。
path	text	文件的路径。
mv_dstpath	text	文件要保存到的目标路径。

 版权所有 © 华为云计算技术有限公司

示例：

```
execute direct on(datanode1) 'select op, file_name, file_type, in_obs, path, mv_dstpath from
gs_lstore_getlivefiles(''FFFFFFFF/FFFFFFFF'',1,'''') order by path, file_name';
 op |     file_name     | file_type | in_obs |
path                                                
```

表 7-104 clean_kvstore_meta(group_name, src_vgroupId)返回字段

名称	类型	描述
result	bool	指定VW的DN清除对应kvstore元信息是否成功。

示例：

```
SELECT clean_kvstore_meta('write_group', 10875761377281);
```

pg_lock_for_lsm_offline_dilatation()

描述：LStore存储引擎离线扩缩容期间对集群加锁。

返回值类型：bool

函数返回字段：

表 7-105 pg_lock_for_lsm_offline_dilatation()返回字段

名称	类型	描述
result	bool	离线扩缩容期间对集群加锁。

示例：

```
SELECT pg_lock_for_lsm_offline_dilatation();
```

pg_unlock_for_lsm_offline_dilatation()

描述：LStore存储引擎离线扩缩容期间对集群释放锁。

返回值类型：bool

函数返回字段：

表 7-106 pg_unlock_for_lsm_offline_dilatation()返回字段

名称	类型	描述
result	bool	离线扩缩容期间对集群释放锁。

示例：

```
SELECT pg_unlock_for_lsm_offline_dilatation();
```

gs_destroy_kvstore(group_name text)

描述：关闭对应VW的kvstore。

返回值类型：bool

函数返回字段：

 版权所有 © 华为云计算技术有限公司

表 7-107 gs_destroy_kvstore(group_name)返回字段

名称	类型	描述
result	bool	关闭对应VW的kvstore是否成功。

示例：

```
SELECT gs_destroy_kvstore('write_group');
```

destroy_kvstore(vgroupId uint64, groupOid uint64)

描述：在VW对应的DN上执行关闭kvstore。

返回值类型：bool

函数返回字段：

表 7-108 destroy_kvstore(vgroupId, groupOid)返回字段

名称	类型	描述
result	bool	对应VW的DN关闭kvstore是否成功。

示例：

```
SELECT destroy_kvstore(10875761377281,2147483649);
```

gs_lsm_compact_db(dbName text, cfName text)

描述：对指定的数据库和列族（Column Family）执行compaction操作。

- cfName：目标列族名称，取值范围为'index'（主键索引列族）、'delta'（delta列族）、'cudesc'（列存元数据的列族）和'all'（全部三个列族）。为null或缺省时，默认对全部三个列族执行。
- dbName：目标数据库名称。为null或缺省时，默认对当前数据库执行。

返回值类型：bool

函数返回字段：

表 7-109 gs_lsm_compact_db(dbName text, cfName text)返回字段

名称	类型	描述
result	bool	对应数据库和列族执行的compaction是否成功。

示例：

对数据库mydb的index列族执行compaction。

```
gs_lsm_compact_db('mydb','index');
```


版权所有 © 华为云计算技术有限公司

对当前数据库的全部三个列族执行compaction。

```
gs_lsm_compact_db();
gs_lsm_compact_db(null, null);
```

lsm_compact_db(dbOid int8, cfName name)

描述：对指定的数据库和列族（Column Family）执行compaction操作，需要连接DN执行。

- dbOid：目标数据库的OID。
- cfName：目标列族名称，取值范围为'index'（主键索引列族）、'delta'（delta列族）、'cudesc'（列存元数据的列族）和'all'（全部三个列族）。

返回值类型：bool

函数返回字段：

表 7-110 lsm_compact_db(dbOid int8, cfName name)返回字段

名称	类型	描述
result	bool	对应数据库和列族执行的compaction是否成功。

示例：

```
SELECT lsm_compact_db(17379, 'index');
```

gs_lsm_compact_relation(tableName text, cfName text)

描述：对指定的LStore表和列族（Column Family）执行compaction操作。

- tableName：目标LStore表名称，支持schema限定的格式（如public.t1）。
- cfName：目标列族名称，取值范围为'index'（主键索引列族）、'delta'（delta列族）、'cudesc'（列存元数据的列族）和'all'（全部三个列族）。为null或缺省时，默认对全部三个列族执行。

返回值类型：bool

函数返回字段：

表 7-111 gs_lsm_compact_relation(tableName text, cfName text)返回字段

名称	类型	描述
result	bool	对应LStore表和列族执行的compaction是否成功。

示例：

对public.t1表的index列族执行compaction。

```
SELECT gs_lsm_compact_relation('public.t1', 'index');
```

对当前schema下t1表的全部列族执行compaction。

 版权所有 © 华为云计算技术有限公司

```
SELECT gs_lsm_compact_relation('t1');
```

lsm_compact_relation(relOid int8, cfName name)

描述：对指定的LStore表和列族（Column Family）执行compaction操作，需要连接DN执行。

- relOid：目标LStore表的OID。
- cfName：目标列族名称，取值范围为'index'（主键索引列族）、'delta'（delta列族）、'cudesc'（列存元数据的列族）和'all'（全部三个列族）。

返回值类型：bool

函数返回字段：

表 7-112 lsm_compact_relation(relOid int8, cfName name)返回字段

名称	类型	描述
result	bool	对应LStore表和列族执行的compaction是否成功。

示例：

```
SELECT lsm_compact_relation(2147483650, 'index');
```

pgxc_clean_delay_file()

描述：用于清理集群中所有节点上PG_DELAY_DELETE_FILE系统表延迟删除的文件记录。DN同步删除对应的OBS文件，随后删除记录。需要管理员权限才可执行该函数。

返回值类型：bigint

函数返回字段：

表 7-113 pgxc_clean_delay_file()返回字段

名称	类型	描述
result	bigint	已清理的记录数量。

示例：

```
SELECT pgxc_clean_delay_file();
```

pgxc_clean_delay_file(spcNode oid, dbId bigint, relNode oid, fileId bigint)

描述：用于清理指定表空间、数据库、relfilenode、文件ID对应的记录。

返回值类型：bigint

函数返回字段：

 版权所有 © 华为云计算技术有限公司

表 7-114 pgxc_clean_delay_file(spcNode oid, dbId bigint, relNode oid, fileId bigint) 返回字段

名称	类型	描述
result	bigint	已清理的记录数量。

示例：

```
SELECT pgxc_clean_delay_file(16385::oid, 16388::bigint, 32768::oid, 10::bigint);
```

pgxc_get_delay_file_info()

描述：查询集群中所有节点上PG_DELAY_DELETE_FILE系统表记录的延迟删除文件信息。需要管理员权限才可执行该函数。

返回值类型：setof record

函数返回字段：

表 7-115 pgxc_get_delay_file_info()返回字段

名称	类型	描述
node_name	text	记录所属节点名称。
databaseid	bigint	数据库ID。
tablespaceid	oid	表空间ID。
relfilenode	oid	表文件的relfilenode。
fileid	bigint	文件ID。
filesize	bigint	文件大小（PG_DELAY_DELETE_FILE表中未记录，固定为NULL）。
xid	bigint	产生该延迟删除记录的事务ID。
delete_time	timestamptz	记录删除时间。

示例：

```
SELECT * FROM pgxc_get_delay_file_info();
```

7.39 Iceberg 表服务函数

Iceberg表服务函数仅9.1.2及以上集群版本支持。

gs_iceberg_snapshots(foreigntable regclass)

描述：查看Iceberg表快照历史记录。

返回值类型：record

 版权所有 © 华为云计算技术有限公司

函数返回字段：

表 7-116 gs_iceberg_snapshots(foreigntable regclass)返回字段

名称	类型	描述
sequence_number	bigint	快照序列号。
snapshot_id	bigint	快照ID。
create_time	timestamp with time zone	快照创建时间。
parent_id	bigint	父快照ID。
operation	text	操作类型（APPEND/OVERWRITE/DELETE）。
manifest_list	text	manifest list文件路径。
summary	text	快照摘要（JSON）。

gs_iceberg_schemas(foreigntable regclass)

描述：查看Iceberg表DDL信息。

返回值类型：record

函数返回字段：

表 7-117 gs_iceberg_schemas(foreigntable regclass)返回字段

名称	类型	描述
schema_id	integer	schema版本ID。
field_id	integer	字段ID。
field_name	text	字段名。
field_type	text	字段类型。

gs_iceberg_partition_specs(foreigntable regclass)

描述：查看Iceberg表分区键。

返回值类型：record

函数返回字段：

 版权所有 © 华为云计算技术有限公司

表 7-118 gs_iceberg_partition_specs(foreigntable regclass)返回字段

名称	类型	描述
spec_id	integer	分区规格ID。
field_id	integer	分区字段ID。
field_name	text	分区字段名。
transform	text	变换函数（identity/hour/day等）。
source_field_id	integer	源字段ID。

gs_iceberg_refs(foreigntable regclass)

描述：查看Iceberg表快照引用信息。

返回值类型：record

函数返回字段：

表 7-119 gs_iceberg_refs(foreigntable regclass)返回字段

名称	类型	描述
name	text	引用名（BRANCH/TAG名）。
type	text	引用类型（BRANCH/TAG）。
snapshot_id	bigint	指向的快照ID。
max_ref_age_ms	bigint	引用最大保留时间（毫秒）。
min_snapshots_to_keep	integer	最少保留快照数。
max_snapshot_age_ms	bigint	快照最大保留时间（毫秒）。

gs_iceberg_manifests(foreigntable regclass)

描述：查看Iceberg表清单文件。

返回值类型：record

函数返回字段：

表 7-120 gs_iceberg_manifests(foreigntable regclass)返回字段

名称	类型	描述
path	text	manifest文件路径。

 版权所有 © 华为云计算技术有限公司

名称	类型	描述
length	bigint	文件大小（字节）。
partition_spec_id	integer	分区规格ID。
added_snapshot_id	bigint	添加该manifest的快照ID。
added_data_files_count	integer	新增数据文件数。
existing_data_files_count	integer	已有数据文件数。
deleted_data_files_count	integer	删除数据文件数。
added_delete_files_count	integer	新增删除文件数。
existing_delete_files_count	integer	已有删除文件数。
deleted_delete_files_count	integer	已删除的删除文件数。
content	integer	内容类型（0=data, 1=deletes）。
partition_summaries	text	分区摘要（JSON）。

gs_iceberg_partitions(foreigntable regclass)

描述：查看Iceberg表分区统计信息。

返回值类型：record

函数返回字段：

表 7-121 gs_iceberg_partitions(foreigntable regclass)返回字段

名称	类型	描述
partition	text	分区值。
spec_id	integer	分区规格ID。
record_count	bigint	记录数。
file_count	integer	文件数。
total_data_file_size_in_bytes	bigint	数据文件总大小（字节）。

 版权所有 © 华为云计算技术有限公司

名称	类型	描述
position_delete_record_count	bigint	位置删除记录数。
position_delete_file_count	integer	位置删除文件数。
equality_delete_record_count	bigint	等值删除记录数。
equality_delete_file_count	integer	等值删除文件数。
last_updated_at	timestamp with time zone	最后更新时间。
last_updated_snapshot_id	bigint	最后更新快照ID。
small_file_count	bigint	小文件数。

gs_iceberg_manifest_entries(foreigntable regclass)

描述：查看Iceberg表清单文件状态条目。

返回值类型：record

函数返回字段：

表 7-122 gs_iceberg_manifest_entries(foreigntable regclass)返回字段

名称	类型	描述
status	integer	状态（1=EXISTING, 2=ADDED, 3=DELETED）。
snapshot_id	bigint	快照ID。
sequence_number	bigint	序列号。
file_sequence_number	bigint	文件序列号。
data_file	text	数据文件描述（JSON）。
readable_metrics	text	可读指标（JSON）。

gs_iceberg_files(foreigntable regclass)

描述：查看Iceberg表数据文件信息。

返回值类型：record

函数返回字段：

 版权所有 © 华为云计算技术有限公司

表 7-123 gs_iceberg_files(foreigntable regclass)返回字段

名称	类型	描述
content	integer	内容类型（0=data, 1=position-deletes, 2=equality-deletes）。
file_path	text	文件路径。
format	text	文件格式（PARQUET/ORC/AVRO）。
spec_id	integer	分区规格ID。
record_count	bigint	记录数。
file_size	bigint	文件大小（字节）。
column_sizes	text	各列大小（JSON）。
value_counts	text	各列值的计数（JSON）。
null_value_counts	text	各列NULL计数（JSON）。
nan_value_counts	text	各列NaN计数（JSON）。
lower_bounds	text	各列下界（JSON）。
upper_bounds	text	各列上界（JSON）。

gs_iceberg_history(foreigntable regclass)

描述：查看Iceberg表时间线历史。

返回值类型：record

函数返回字段：

表 7-124 gs_iceberg_history(foreigntable regclass)返回字段

名称	类型	描述
made_current_at	timestamp with time zone	成为当前快照时间。
snapshot_id	bigint	快照ID。
parent_id	bigint	父快照ID。
is_current_ancestor	boolean	是否为当前快照的祖先。

gs_iceberg_metadata_log_entries(foreigntable regclass)

描述：查看Iceberg表元数据日志条目。

返回值类型：record

 版权所有 © 华为云计算技术有限公司

函数返回字段：

表 7-125 gs_iceberg_metadata_log_entries(foreigntable regclass)返回字段

名称	类型	描述
time_stamp	timestamp with time zone	时间戳。
file	text	metadata JSON文件路径。
latest_snapshot_id	bigint	最新快照ID。
latest_schema_id	bigint	最新Schema ID。
latest_sequence_num	bigint	最新序列号。

gs_iceberg_pos_delete_files(foreigntable regclass)

描述：查看Iceberg表删除记录。

返回值类型：record

函数返回字段：

表 7-126 gs_iceberg_pos_delete_files(foreigntable regclass)返回字段

名称	类型	描述
file_path	text	数据文件路径。
pos	bigint	删除位置（行号）。
partition	text	分区值。
spec_id	bigint	分区规格ID。
del_file_path	text	删除文件路径。

7.40 时序函数

DWS支持的时序计算函数参见表7-127。

表 7-127 时序计算函数列表

时序计算函数	函数描述
delta(field numeric)	计算按照时间排序后两行之间的差值。
spread(field numeric)	计算某段时间内最大和最小值的差值。

版权所有 © 华为云计算技术有限公司

时序计算函数	函数描述
mode() within group (order by value anyelement)	对给定的列，返回出现频率最高的值，如果多个值频率相同，返回这些值中最小的那个值。
value_of_percentile(column float, percentile float, compression float)	计算百分位。是percentile_cont的近似算法。
percentile_of_value(column float, percentilevalue float, compression float)	给定百分位，计算对应的值。是value_of_percentile的逆运算。
first(column1, column2)	通过比较分组内column2列的值，找到其中的最小值，输出对应column1列的值。
last(column1, column2)	通过比较分组内column2列的值，找到其中的最大值，输出对应column1列的值。

delta(field numeric)

描述：计算按照时间排序后两行之间的差值。该函数通常用于时序场景计算按照时间排序后相邻两行差值，用于流量，速度等指标监控。

参数说明：field，需要计算的列。

⚠ 注意

delta是一个窗口函数，需要与over窗口函数使用。且over中rows语句不会改变delta函数结果，比如，delta(value) over(order by time rows 1 preceding) 和 delta(value) over(order by time rows 3 preceding) 返回的结果是一致的。

示例：

```
SELECT
    delta(value) over (rows 1 preceding)
FROM
    (VALUES ('2019-07-12 00:00:00'::timestamptz, 1),
            ('2019-07-12 00:01:00'::timestamptz, 3),
            ('2019-07-12 00:02:00'::timestamptz, 7))
    v(time,value) order by 1;
 delta
-------
     0
     2
     4
(3 rows)
```

spread(field numeric)

描述：计算每个指标在分组内的增量，通常按照时间排序后计算。

参数说明：field，需要计算的列。

版权所有 © 华为云计算技术有限公司

⚠ 注意

每个分组内如果少于2个元组，返回结果为0，不要和over窗口函数混用。

示例：

```
SELECT
    time, SPREAD(value)
 FROM
    (VALUES ('2019-07-12 00:00:00'::timestamptz, 1), ('2019-07-12 00:00:00'::timestamptz, 3),
            ('2019-07-12 00:01:00'::timestamptz, 3), ('2019-07-12 00:01:00'::timestamptz, 6),
            ('2019-07-12 00:02:00'::timestamptz, 7), ('2019-07-12 00:02:00'::timestamptz, 9))
    v(time,value) group by time order by time;
         time          | spread
-----------------------+--------
 2019-07-12 00:00:00+08 |      2
 2019-07-12 00:01:00+08 |      3
 2019-07-12 00:02:00+08 |      2
(3 rows)
```

mode() within group (order by value anyelement)

描述：对给定的列，返回出现频率最高的值，如果多个值频率相同，返回这些值中最小的那个值。

参数说明：value，查询列。

⚠ 注意

- 该函数需要与within group一起使用，无within group语句，会报错。该函数参数放在group的order by后面。
- 不能和over子句一起使用。

示例：

```
SELECT
    mode() within group (order by value)
 FROM
    (VALUES ('2019-07-12 00:00:00'::timestamptz, 1),
            ('2019-07-12 00:01:00'::timestamptz, 1),
            ('2019-07-12 00:02:00'::timestamptz, 7))
    v(time,value);
 mode
------
    1
(1 row)
```

value_of_percentile(column float, percentile float, compression float)

描述：对于给定的列按照从小到大的顺序返回百分位的近似值。是percentile_cont的近似结果，但是性能比percentile_cont好。

表 7-128 参数说明

参数名	类型	描述
column	float	要计算百分位的列。

版权所有 © 华为云计算技术有限公司

参数名	类型	描述
percentile	float	指定的百分位的值，取值范围[0，1]。
compression	float	指定的压缩系数，取值范围为[0，500]，默认值是300，这个值取值越大，函数计算过程占用内存越大，同时结果的精度也相对高。如果指定的值不在取值范围内，会按照300求解。

示例：

```
SELECT value_of_percentile(values, 0.8, 0) from TABLE;
```

percentile_of_value(column float, percentilevalue float, compression float)

描述：对于给定的列按照从小到大的顺序返回百分位。是value_of_percentile的逆过程。

表 7-129 参数说明

参数名	类型	描述
column	float	要计算百分位的列。
percentilevalue	float	指定该值，用于计算它所在的百分位。
compression	float	指定的压缩系数，取值范围为[0，500]，默认值是300，这个值取值越大，函数计算过程占用内存越大，同时结果的精度也相对高。如果指定的值不在取值范围内，会按照300求解。

示例：

```
SELECT percentile_of_value(values, 80, 0) from TABLE;
```

first(column1, column2)

描述：聚合函数。通过比较分组内column2列的值，找到其中的最小值，输出对应column1列的值。

表 7-130 参数说明

参数名	类型	描述
column1	bigint/text/double/numeric	最终的输出列。
column2	timestamp/timestamptz/numeric	比较列。

 版权所有 © 华为云计算技术有限公司

示例：样例数据参考**time_fill()**表达式中的表定义和数据。

求按照scope_name分组，每个分组内按照时间排序最靠前的idle的值：

```
select first(idle, time_string) from dcs_cpu group by scope_name;
 first
-------
     1
     3
(2 rows)
```

last(column1, column2)

描述：聚合函数。通过比较分组内column2列的值，找到其中的最大值，输出对应column1列的值。

表 7-131 参数说明

参数名	类型	描述
column1	bigint/text/double/numeric	最终的输出列。
column2	timestamp/timestamptz/ numeric	比较列。

示例：样例数据参考**time_fill()**表达式中的表定义和数据。

求按照scope_name分组，每个分组内按照时间排序最靠后的idle的值：

```
select last(idle, time_string) from dcs_cpu group by scope_name;
 last
------
    3
    2
(2 rows)
```

时序填充函数

DWS支持的时序填充函数参见**表7-132**。

表 7-132 时序填充函数列表

时序填充表达式	描述
time_fill(interval, time_column, start_time, end_time), **fill_last**(agg_function(agg_column))	按照时间列排序后，补充缺失的时间数据信息（聚合以后的结果） 用按照时间排序后的**前值填充后值**。
time_fill(interval, time_column, start_time, end_time), **fill_first**(agg_function(agg_column))	按照时间列排序后，补充缺失的时间数据信息（聚合以后的结果） 用按照时间排序后的**后值填充前值**。

版权所有 © 华为云计算技术有限公司

时序填充表达式	描述
time_fill(interval, time_column, start_time, end_time), **fill_avg**(agg_function(agg_column))	按照时间列排序后，补充缺失的时间数据信息（聚合以后的结果） 用按照时间排序后**前后两者的值填充当前值**。

⚠ 注意

- time_fill函数需要作为聚合函数使用，group by需要引用自身计算结果，不支持与自身嵌套使用，不支持单条查询内多次调用，不支持作为下层计算节点使用，不支持与within group使用。
- time_fill在使用时，必须与fill_avg、fill_first或者fill_last配合使用；或者与Agg函数组合使用。
- time_fill必须出现在group by中，并且group by中只能出现这一列。
- time_fill不支持出现在处理select后面的其他位置，比如where后面或其他关联条件中。
- start时间戳的值必须小于finish时间戳的值， 且两者间距需要大于window_width的值。
- 所有参数不支持空值，start和finish需为确定值。

表 7-133 时序填充表达式的参数说明

参数名	类型	描述
interval	时间间隔类型：INTERVAL，最小单位是1秒。	按照时间分组的时间间隔。
time_column	时间类型，timestamp/ timestamptz。	按照指定列做时间分组。
start_time	时间类型，timestamp/ timestamptz。	分组的起始时间。
end_time	时间类型，timestamp/ timestamptz。	分组的结束时间。
agg_function(agg_column)	指定Agg函数对指定列做聚合。比如max(col)。	对Agg的结果按照指定的填充方法填充。

time_fill(interval, time_column, start_time, end_time), fill_last(agg_function(agg_column))

描述：按照时间列排序后，用前值补充后值。

创建示例表dcs_cpu，并且插入数据。

```
create table dcs_cpu(
idle real,
```

 版权所有 © 华为云计算技术有限公司

```
vcpu_num int,
node text,
scope_name text,
server_ip text,
iowait numeric,
time_string timestamp
)with (orientation=column) distribute by hash(node);
insert  into dcs_cpu VALUES(1.0,1,'node_a','scope_a','1.1.1.1',1.0,'2019-07-12 00:10:10');
insert  into dcs_cpu VALUES(2.0,2,'node_b','scope_a','1.1.1.2',2.0,'2019-07-12 00:12:10');
insert  into dcs_cpu VALUES(3.0,3,'node_c','scope_b','1.1.1.3',3.0,'2019-07-12 00:13:10');
```

以1min为单位，求分组内的均值。用前一时间段的值，填充后一时间段的值。

```
select time_fill(interval '1 min',time_string,'2019-07-12 00:09:00','2019-07-12 00:14:00'), fill_last(avg(idle))
from dcs_cpu group by time_fill order by time_fill;
     time_fill      | fill_last
--------------------+-----------
 2019-07-12 00:09:00 |
 2019-07-12 00:10:00 |         1
 2019-07-12 00:11:00 |         1
 2019-07-12 00:12:00 |         2
 2019-07-12 00:13:00 |         3
 2019-07-12 00:14:00 |         3
(6 rows)
```

time_fill(interval, time_column, start_time, end_time), fill_first(agg_function(agg_column))

描述：按照时间列排序后，用后值补充前值。

示例：

以1min为单位，求分组内的均值。用后一时间段的值，填充前一时间段的值（样例数据参考**time_fill(interval, time_column, start_time, end_time), fill_last(agg_function(agg_column))**中的表定义和数据）。

```
select time_fill(interval '1 min',time_string,'2019-07-12 00:09:00','2019-07-12 00:14:00'), fill_first(avg(idle))
from dcs_cpu group by time_fill order by time_fill;
     time_fill      | fill_first
--------------------+------------
 2019-07-12 00:09:00 |          1
 2019-07-12 00:10:00 |          1
 2019-07-12 00:11:00 |          2
 2019-07-12 00:12:00 |          2
 2019-07-12 00:13:00 |          3
 2019-07-12 00:14:00 |
(6 rows)
```

time_fill(interval, time_column, start_time, end_time), fill_avg(agg_function(agg_column))

描述：按照时间排序后，前后两者的值填充当前值。

示例：

以1min为单位，求分组内的平均值，用前后两个时间的加权平均值，填充当前值（样例数据参考**time_fill(interval, time_column, start_time, end_time), fill_last(agg_function(agg_column))**中的表定义和数据）。

```
select time_fill(interval '1 min',time_string,'2019-07-12 00:09:00','2019-07-12 00:14:00'), fill_avg(avg(idle))
from dcs_cpu group by time_fill order by time_fill;
     time_fill      | fill_avg
--------------------+----------
 2019-07-12 00:09:00 |        1
 2019-07-12 00:10:00 |        1
```


版权所有 © 华为云计算技术有限公司

```
 2019-07-12 00:11:00 |     1.5
 2019-07-12 00:12:00 |       2
 2019-07-12 00:13:00 |       3
 2019-07-12 00:14:00 |       3
(6 rows)
```

7.41 实时数仓函数

hstore_light_merge(rel_name text)

描述：用于手动对HStore表进行轻量化清理操作，持有目标表的三级锁。

返回值类型：int

示例：

```
SELECT hstore_light_merge('reason_select');
```

hstore_full_merge(rel_name text, partitionName text)

描述：用于手动对HStore表进行全量清理操作。

partitionName为可选参数，用于指定单分区进行操作。

返回值类型：int

注意

- 执行该操作会强制将DELTA表上的所有可见操作Merge到主表，然后建一张新的空Delta表， 期间持有该表的八级锁。
- 该操作的耗时长度与DELTA表上的数据量有关，务必打开HStore的清理线程，确保HStore表的及时清理。
- 仅在8.3.0.100及以上集群版本支持第二个可选参数partitionName，因此8.3.0.100及以上集群版本不支持使用call调用此函数（ call不支持重载 ）

示例：

```
SELECT hstore_full_merge('reason_select', 'part1');
```

pgxc_get_small_cu_info(rel_name text, row_count int)

描述：用于获取目标表的小CU信息。该函数仅8.2.1.200及以上集群版本支持。

row_count为可选参数，表示小CU的阈值，存活元组数小于该参数值会被认为是小CU，默认值200。

返回值类型：record

返回值：

- node_name：DN节点名。
- part_name：分区名，非分区表此列为空。
- zero_cu_count：0CU的数量。当一个CU中的所有数据都被删除时，称之为0CU。

版权所有 © 华为云计算技术有限公司

- small_cu_count：小CU数量。当一个CU中**有存活数据**且**存活数量小于阈值**时，称之为小CU。
- total_cu_count：总的CU的数量。
- sec_part_cu_num：每个二级分区的CU数量。当secondary_part_column被指定时，该列才会显示。该字段仅8.3.0及以上集群版本支持。
- bucket_cu_num：每个bucket的CU数量。当表是V3表时，该列才会显示。该字段仅9.1.1.100及以上集群版本支持。

说明

此处CU是跨列的概念，并非一列一个的CU。

示例：

二级分区表：

```
SELECT * FROM pgxc_get_small_cu_info('hs');
node_name | part_name | zero_cu_count | small_cu_count | total_cu_count |
sec_part_cu_num
-----------+-----------+---------------+----------------+----------------+-----------------------------------------
 datanode1 |           |             0 |              4 |              4 | p1:1 p2:0 p3:1 p4:0 p5:1 p6:0 p7:1 p8:0
 datanode2 |           |             0 |              4 |              4 | p1:0 p2:1 p3:0 p4:1 p5:0 p6:1 p7:0 p8:1
(2 rows)
```

V3表：

```
SELECT * FROM pgxc_get_small_cu_info('hs_v3');
 node_name | part_name | zero_cu_count | small_cu_count | total_cu_count | sec_part_cu_num |
bucket_cu_num
-----------+-----------+---------------+----------------+----------------+-----------------+----------------------
 datanode1 |           |             1 |              1 |              4 |                 | b0:2 b1:0 b2:2 b3:0
 datanode2 |           |             0 |              0 |              2 |                 | b0:0 b1:1 b2:0 b3:1
(2 rows)
```

pgxc_get_small_cu_info_with_percentage(rel_name text, row_count int)

描述：用于获取目标表的小CU信息。该函数仅9.1.1.200及以上集群版本支持。

row_count为可选参数，表示小CU的阈值，存活元组数小于该参数值会被认为是小CU，默认值200。

返回值类型：record

返回值：

- node_name：DN节点名。
- part_name：分区名，非分区表此列为空。
- zero_cu_count：0CU的数量。当一个CU中的所有数据都被删除时，称之为0 CU。
- small_cu_count：小CU数量。当一个CU中**有存活数据**且**存活数量小于阈值**时，称之为小CU。
- total_cu_count：总的CU的数量。
- sec_part_cu_num：每个二级分区的CU数量。当secondary_part_column被指定时，该列才会显示。
- bucket_cu_num：每个bucket的CU数量。当表是V3表时，该列才会显示。
- total_cu_percentage：当前节点在整个表或分区的CU占比。

 版权所有 © 华为云计算技术有限公司

说明

此处CU是跨列的概念，并非一列一个的CU。

示例：

```
SELECT * FROM pgxc_get_small_cu_info_with_percentage('tb') order by node_name, part_name;
node_name | part_name | zero_cu_count | small_cu_count | total_cu_count | sec_part_cu_num
|                    bucket_cu_num                     | total_cu_percentage
--------------+------------+----------------+----------------+----------------+-------------------+-----------
-----------------------------------+---------------------
datanode1  | part_1       |           0        |           8        |           8        |                     | b0:1
b2:1 b4:1 b6:1 b8:1 b10:1 b12:1 b14:1  |           50.000
datanode2  | part_1       |           0        |           8        |           8        |                     | b1:1
b3:1 b5:1 b7:1 b9:1 b11:1 b13:1 b15:1  |           50.000
(2 rows)
```

pgxc_get_hstore_delta_info(rel_name text)

描述：用于获取目标表的delta表信息，包括delta表的大小，insert/delete/update各种类型记录的数量等。该函数仅8.2.1.100及以上集群版本支持。

返回值类型：record

返回值：

- node_name：DN节点名。
- part_name：分区名，非分区表此列为non partition table。
- live_tup：存活的元组数量。
- n_ui_type：type是ui（小cu合并和upsert走update插入）的记录的数量。一条ui记录表示一次插入，可以是单插或者批插。该参数仅8.3.0.100及以上版本支持。
- n_i_type：type是i（insert）的记录的数量。一条i记录表示一次插入，可以是单插或者批插。
- n_d_type：type是d（delete）的记录的数量。一条d记录表示一次删除，可以是单条删或者批量删除。
- n_x_type：type是x（由update产生的delete）记录的数量。
- n_u_type：type是u（轻量化update）的记录数量。
- n_m_type：type是m（merge）的记录数量。
- data_size：delta表的总大小，单位为字节（包括delta上的索引与toast数据的大小）。

示例：

```
SELECT * FROM pgxc_get_hstore_delta_info('hs_part');
 node_name | part_name | live_tup | n_ui_type | n_i_type | n_d_type | n_x_type | n_u_type | n_m_type |
data_size
-----------+-----------+----------+-----------+----------+----------+----------+----------+----------+-----------
 dn_1      | p1        |        2 |         0 |        2 |        0 |        0 |        0 |        0 |      8192
 dn_1      | p2        |        2 |         0 |        2 |        0 |        0 |        0 |        0 |      8192
 dn_1      | p3        |        2 |         0 |        2 |        0 |        0 |        0 |        0 |      8192
(3 rows)
```

pgxc_get_binlog_sync_point(rel_name text, slot_name text, checkpoint bool, node_id int)

描述：用于从pg_binlog_slots系统表上获取槽位对应的同步点信息，只能对开启binlog或binlog时间戳的表使用。该函数仅9.1.0.200及以上集群版本支持。

 版权所有 © 华为云计算技术有限公司

返回值类型：record

返回值：

- node_name：节点名称。
- node_id：节点ID。
- last_sync_point：上次同步点。
- latest_sync_point：当前最新同步点。
- xmin：同步点对应xmin。

示例：

```
SELECT * FROM pg_catalog.pgxc_get_binlog_sync_point('hstore_binlog_source', 'slot1', false, 0);
 node_name |   node_id   | last_sync_point | latest_sync_point | xmin
-----------+-------------+-----------------+-------------------+-------
 dn_2      | -1051926843 |               0 |             10512 | 10507
 dn_1      | -1300059100 |               0 |             10512 | 10508
(2 rows)
```

pgxc_get_binlog_changes(rel_name text, node_id int, start_csn bigint, end_csn bigint)

描述：用于获取目标表在指定DN上指定同步点区间的增量数据（node_id给0表示指定所有DN），只能对开启binlog或binlog时间戳的表使用。该函数仅9.1.0.200及以上集群版本支持。

返回值类型：record

返回值：

- gs_binlog_sync_point：同步点。
- gs_binlog_event_sequence：用于表示同一事务内的先后顺序。
- gs_binlog_event_type：binlog类型。
- gs_binlog_timestamp_us：binlog记录的时间戳，对于enable_binlog_timestamp为false的binlog表，该列返回空。
- value columns：目标表上各个用户字段的数据。

示例：

```
SELECT * FROM pgxc_get_binlog_changes('hstore_binlog_source', 0, 0 , 9999999999);
 gs_binlog_sync_point | gs_binlog_event_sequence | gs_binlog_event_type | gs_binlog_timestamp_us | c1  | c2
| c3
----------------------+--------------------------+----------------------+------------------------+-----+----+----
               10516 |                        2 | I                    |      1731570520900211 | 100 |  1 |  1
               10517 |                        3 | d                    |      1731570520904425 | 100 |  1 |  1
               10518 |                        2 | I                    |      1731570520909055 | 200 |  1 |  1
               10519 |                        3 | B                    |      1731570520914102 | 200 |  1 |  1
               10519 |                        4 | U                    |      1731570520914154 | 200 |  2 |  1
(5 rows)
```

pgxc_register_binlog_sync_point(rel_name text, slot_name text, node_id int, end_csn bigint, checkpoint bool, xmin bigint)

描述：用于登记同步点，只能对开启binlog或binlog时间戳的表使用。该函数仅9.1.0.200及以上集群版本支持。

返回值类型：int

 版权所有 © 华为云计算技术有限公司

返回值：登记成功的节点数量。

示例：

```
SELECT pgxc_register_binlog_sync_point('hstore_binlog_source', 'slot1', 0, 9999999999, false, 100);
 pgxc_register_binlog_sync_point
---------------------------------
                               2
(1 row)
```

pgxc_consumed_binlog_records(rel_name text, node_id int)

描述：用于获取目标表在指定DN上的消费状态，只能对开启binlog或binlog时间戳的表使用。该函数仅9.1.0.200及以上集群版本支持。

其中，node_id为0，表示查询所有DN。

返回值类型：int

返回值：

- 0，表示目标表的binlog仍没有被消费完毕（包含所有槽位以及checkpoint同步点）。
- 1，表示目标表的binlog被全部消费完毕。

示例：

```
SELECT * FROM pgxc_consumed_binlog_records('hstore_binlog_source',0);
 pgxc_consumed_binlog_records
------------------------------
                            1
(1 row)
```

pgxc_get_binlog_slots(rel_name text, node_id int)

描述：用于获取目标表所有槽位的进度信息。该函数仅9.1.0.223及以上集群版本支持。

返回值类型：record

其中，node_id为0，表示查询所有DN。

返回值：

- node_name：节点名。
- slot_name：槽位名。
- sync_point： 该槽位最近一次注册的同步点。
- sync_timestamp：该槽位最近一次注册的时间。
- oldest_xmin：该槽位最近一次注册时对应的oldest_xmin。
- is_checkpoint：是否属于checkpoint点位。

示例：

```
SELECT * FROM pgxc_get_binlog_slots('binlog_view_part_t1', 0);
 node_name | slot_name | sync_point |        sync_timestamp         | oldest_xmin |  is_checkpoint
-----------+-----------+------------+-------------------------------+-------------+----------------
 dn_1      |   slot1   |    210059  | 2025-09-19 10:31:38.894357+08 |       10054 |  f
(1 row)
```

 版权所有 © 华为云计算技术有限公司

pgxc_binlog_clear_target_slot(rel_name text, slot_name text)

描述：用于清理目标表的指定槽位。该函数仅9.1.0.223及以上集群版本支持。

返回值类型：int

返回值：返回所有节点被清理的槽位数量之和。

示例：

```
SELECT * FROM pgxc_binlog_clear_target_slot('binlog_view_part_t1',  'slot1');
 pgxc_binlog_clear_target_slot
-------------------------------
                             4
(1 row)
```

pgxc_get_binlog_cursor_by_timestamp(rel_name text, timestamp timestampTz, node_id int)

描述：用于获取目标表上指定时间点后的第一个binlog记录的信息，只能对开启binlog时间戳的表使用。该函数仅9.1.0.200及以上集群版本支持。

返回值类型：record

返回值：

- node_name：节点名称。
- node_id：节点ID。
- latest_sync_point：当前最新同步点。
- binlog_sync_point：时间点后的第一个binlog记录的同步点。
- binlog_timestamp_us：时间点后的第一个binlog记录的时间戳。
- binlog_xmin：时间点后的第一个binlog记录的xmin。

示例：

```
SELECT * FROM pgxc_get_binlog_cursor_by_timestamp('hstore_binlog_source','2024-11-14
15:48:40.900211+08', 0);
 node_name |   node_id   | latest_sync_point | binlog_sync_point | binlog_timestamp_us | binlog_xmin
-----------+-------------+-------------------+-------------------+---------------------+-------------
 dn_2      | -1051926843 |             10532 |             10516 |    1731570520900211 |       10510
 dn_1      | -1300059100 |             10532 |             10518 |    1731570520909055 |       10510
(2 rows)
```

pgxc_get_binlog_consume_progress(rel_name text, node_id int)

描述：用于获取目标表的消费进度信息，只能对开启binlog时间戳的表使用。该函数仅9.1.0.200及以上集群版本支持。

返回值类型：record

返回值：

- node_name：节点名称。
- node_id：节点ID。
- slot_name：槽位名。
- checkpoint：是否属于checkpoint点位。

 版权所有 © 华为云计算技术有限公司

- latest_consumed_timestamp：已消费的最新的binlog的时间戳。
- latest_timestamp：所有binlog里最新的时间戳。
- latest_consumed_csn：已消费的最新的binlog的csn点位。
- latest_csn：所有binlog里最新的csn点位。
- unconsumed_binlog_count：还未消费的binlog的数量。

示例：

```
SELECT * FROM pgxc_get_binlog_consume_progress('hstore_binlog_source', 0);
 node_name |   node_id   | slot_name | checkpoint | latest_consumed_timestamp |    latest_timestamp    |
latest_consumed_csn | latest_csn | unconsumed_binlog_count
-----------+-------------+-----------+------------+---------------------------+------------------------
+--------------------+------------+-------------------------
 dn_1      | -1300059100 | slot1     | f          | 2025-04-15 15:55:55+08    | 2025-04-15 15:55:55+08 |
417725 |     417725 |                     0
 dn_1      | -1300059100 | slot1     | t          | 2025-04-15 15:55:55+08    | 2025-04-15 15:55:55+08 |
417725 |     417725 |                     0
(2 rows)
```

pgxc_get_binlog_cursor_by_syncpoint(rel_name text, csn int8, node_id int)

描述：用于获取目标表上指定同步点后的第一个binlog记录的信息，只能对开启binlog时间戳的表使用。该函数仅9.1.0.200及以上集群版本支持。

返回值类型：record

返回值：

- node_name：节点名称。
- node_id：节点ID。
- latest_sync_point：当前最新同步点。
- binlog_sync_point：同步点后的第一个binlog记录的同步点。
- binlog_timestamp_us：同步点后的第一个binlog记录的时间戳。
- binlog_xmin：同步点后的第一个binlog记录的xmin。

示例：

```
SELECT * FROM pgxc_get_binlog_cursor_by_syncpoint('hstore_binlog_source',10516,0);
 node_name |   node_id   | latest_sync_point | binlog_sync_point | binlog_timestamp_us | binlog_xmin
-----------+-------------+-------------------+-------------------+---------------------+-------------
 dn_1      | -1300059100 |             11187 |             10518 |    1731570520909055 |       10510
 dn_2      | -1051926843 |             11187 |             10516 |    1731570520900211 |       10510
(2 rows)
```

pgxc_get_cstore_dirty_ratio(rel_name text, partition_name)

描述：用于获取各个DN上目标表的cu、delta以及cudesc的脏页率以及大小，仅支持HStore_opt表。该函数仅9.1.0.100及以上集群版本支持。

partition_name为可选参数，如果传入分区名则只返回该分区对应的信息；如果没传入分区名且是主表时分区表返回所有分区对应的信息。

返回值类型：record

返回值：

- node_name：DN节点名称。

 版权所有 © 华为云计算技术有限公司

- database_name：表所在数据库名称。
- rel_name：主表名。
- part_name：分区名。
- cu_dirty_ratio：cu文件的脏页率。
- cu_size：cu文件大小。
- delta_dirty_ratio：delta表脏页率。
- delta_size：delta表大小。
- cudesc_dirty_ratio：cudesc表脏页率。
- cudesc_size：cudesc表大小。
- toast_size：toast表大小。该字段仅9.1.2及以上集群版本支持。
- index_size：索引大小。该字段仅9.1.2及以上集群版本支持。

示例：

```
SELECT * FROM pgxc_get_cstore_dirty_ratio('hs_opt');
 node_name | database_name |      rel_name | partition_name | cu_dirty_ratio | cu_size | delta_dirty_ratio |
delta_size | cudesc_dirty_ratio | cudesc_size | toast_size | index_size
-----------+---------------+---------------+----------------+----------------+---------+-------------------+------------
+--------------------+-------------+------------+------------
 dn_1      | postgres      | public.hs_opt | p1             |              0 |       0 |                 0 |      16384 |
0 |       24576 |          0 |          0
 dn_1      | postgres      | public.hs_opt | p2             |              0 |       0 |                 0 |      16384 |
0 |       24576 |          0 |          0
 dn_1      | postgres      | public.hs_opt | p3             |              0 |       0 |                 0 |      16384 |
0 |       24576 |          0 |          0
 dn_1      | postgres      | public.hs_opt | p4             |              0 |       0 |                 0 |      16384 |
0 |       24576 |          0 |          0
 dn_1      | postgres      | public.hs_opt | other          |              0 | 1105920 |                 0 |     524288
|                0 |       40960 |          0 |          0
(5 rows)
```

get_bucket_cuid_fileid_info(bucketid integer)

描述：根据指定的bucketid查询对应的cuid、fileid等信息。该函数仅9.1.1.300及以上集群版本支持。

返回值类型：record

返回值：

- min_globalcuid：指定bucketid对应的最小globalcuid。
- max_globalcuid：指定bucketid对应的最大globalcuid。
- min_localcuid：指定bucketid对应的最小localcuid。
- max_localcuid：指定bucketid对应的最大localcuid。
- localcuid_num：指定bucketid对应的localcuid的最大数量。
- min_globalfileid：指定bucketid对应的globalfileid的最小值。
- max_globalfileid：指定bucketid对应的globalfileid的最大值。
- min_seqno：指定bucketid对应的localfileid的最小值。
- max_seqno：指定bucketid对应的localfileid的最大值。
- seqno_num：指定bucketid对应的localfileid的最大数量。

示例：

 版权所有 © 华为云计算技术有限公司

```
SELECT * FROM get_bucket_cuid_fileid_info(1);
 min_globalcuid | max_globalcuid | min_localcuid | max_localcuid | localcuid_num | min_globalfileid |
max_globalfileid | min_seqno | max_seqno | seqno_num
----------------+----------------+---------------+---------------+---------------+------------------+------------------
+-----------+-----------+-----------
          1025 |     4294950913 |             1 |       4194288 |       4194288 |            16385 |            32767 |
1 |     16383 |     16383
(1 row)
```

get_table_cuid_fileid_info(tableName regclass)

描述：查询指定表中所有bucketid的存储信息，如cuid、fileid等。该函数仅9.1.1.300及以上集群版本支持。

返回值类型：record

返回值：

- nodename：节点名称。
- partname：分区表名，非分区表显示空。
- bucketid：对应的bucketid。
- min_globalcuid：指定表中globalcuid的最小值。
- max_globalcuid：指定表中globalcuid的最大值。
- globalcuid_num：指定表中globalcuid的数量。
- rowcount：指定表对应bucket中的数据行数。
- bucket_skew_ratio：bucket倾斜率，每个bucket数据行数占总行数的比率。
- min_globalfileid：指定表中globalfileid的最小值。
- max_globalfileid：指定表中globalfileid的最大值。
- globalfileid_num：指定表中globalfileid的数量。
- max_seqno：指定表中localfileid的最大值。
- seqno_num：指定表中localfileid的数量。

示例：

```
SELECT * FROM get_table_cuid_fileid_info('t1');
 nodename  | partname | bucketid | min_globalcuid | max_globalcuid | globalcuid_num | rowcount |
bucket_skew_ratio | min_globalfileid | max_globalfileid | globalfileid_num | max_seqno | seqno_num
-----------+----------+----------+----------------+----------------+----------------+----------+-------------------
+------------------+------------------+------------------+-----------+-----------
 datanode1 |          |        0 |           4096 |           4096 |              1 |    37615 |             .1254 |
97307852800002 |   97307852800002 |                1 |         2 |         1
 datanode1 |          |        2 |           4098 |           4098 |              1 |    37490 |              .125 |
97307852832770 |   97307852832770 |                1 |         2 |         1
 datanode1 |          |        4 |           4100 |           4100 |              1 |    37632 |             .1254 |
97307852865538 |   97307852865538 |                1 |         2 |         1
 datanode1 |          |        6 |           4102 |           4102 |              1 |    37578 |             .1253 |
97307852898306 |   97307852898306 |                1 |         2 |         1
 datanode2 |          |        1 |           4097 |           4097 |              1 |    37678 |             .1256 |
97307869593602 |   97307869593602 |                1 |         2 |         1
 datanode2 |          |        3 |           4099 |           4099 |              1 |    37339 |             .1245 |
97307869626370 |   97307869626370 |                1 |         2 |         1
 datanode2 |          |        5 |           4101 |           4101 |              1 |    37380 |             .1246 |
97307869659138 |   97307869659138 |                1 |         2 |         1
 datanode2 |          |        7 |           4103 |           4103 |              1 |    37288 |             .1243 |
97307869691906 |   97307869691906 |                1 |         2 |         1
(8 rows)
```


版权所有 © 华为云计算技术有限公司

8 表达式

8.1 简单表达式

逻辑表达式

逻辑表达式的操作符和运算规则，请参见**逻辑操作符**。

比较表达式

常用的比较操作符，请参见**比较操作符**。

除比较操作符外，还可以使用以下句式结构：

- BETWEEN操作符
  a BETWEEN x AND y等效于a >= x AND a <= y
  a NOT BETWEEN x AND y等效于a < x OR a > y
- 检查一个值是不是null，可使用：
  expression IS NULL
  expression IS NOT NULL
  或者与之等价的句式结构，但不是标准的：
  expression ISNULL
  expression NOTNULL

须知

不要写expression=NULL或expression<>(!=)NULL，因为NULL代表一个未知的值，不能通过该表达式判断两个未知值是否相等。

示例

```
SELECT 2 BETWEEN 1 AND 3 AS RESULT;
 result
----------
```

 版权所有 © 华为云计算技术有限公司

```
 t
(1 row)

SELECT 2 >= 1 AND 2 <= 3 AS RESULT;
 result
----------
 t
(1 row)

SELECT 2 NOT BETWEEN 1 AND 3 AS RESULT;
 result
----------
 f
(1 row)

SELECT 2 < 1 OR 2 > 3 AS RESULT;
 result
----------
 f
(1 row)

SELECT 2+2 IS NULL AS RESULT;
 result
----------
 f
(1 row)

SELECT 2+2 IS NOT NULL AS RESULT;
 result
----------
 t
(1 row)

SELECT 2+2 ISNULL AS RESULT;
 result
----------
 f
(1 row)

SELECT 2+2 NOTNULL AS RESULT;
 result
----------
 t
(1 row)

SELECT 2+2 IS DISTINCT FROM NULL AS RESULT;
 result
----------
 t
(1 row)

SELECT 2+2 IS NOT DISTINCT FROM NULL AS RESULT;
 result
----------
 f
(1 row)
```

8.2 条件表达式

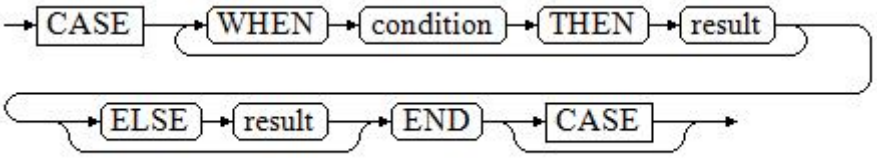
在执行SQL语句时，可通过条件表达式筛选出符合条件的数据。

条件表达式主要有以下几种：

- CASE
  CASE表达式是条件表达式，类似于其他编程语言中的CASE语句。
  CASE表达式的语法图请参考**图8-1**。

 版权所有 © 华为云计算技术有限公司

图 8-1 case::=



CASE子句可以用于合法的表达式中。condition是一个返回BOOLEAN数据类型的表达式：

- 如果结果为真，CASE表达式的结果就是符合该条件所对应的result。
- 如果结果为假，则以相同方式处理随后的WHEN或ELSE子句。
- 如果各WHEN condition都不为真，表达式的结果就是在ELSE子句执行的result。如果省略了ELSE子句且没有匹配的条件，结果为NULL。

⚠ 注意

- case when嵌套使用时不建议超过3层，解析器在解析时需要将所有涉及字段递归展开，递归过深可能导致内存占用过大。
- 如果case when很复杂可以直接使用子查询的方式替换，参考**子查询表达式**。

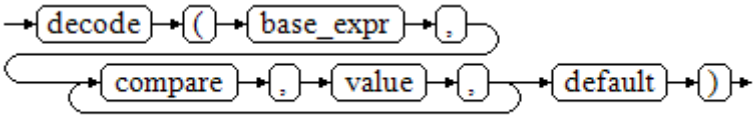
示例：

```
CREATE TABLE tpcds.case_when_t1(CW_COL1 INT)  DISTRIBUTE BY HASH (CW_COL1);

INSERT INTO tpcds.case_when_t1 VALUES (1), (2), (3);

SELECT * FROM tpcds.case_when_t1;
 cw_col1
---------
       3
       1
       2
(3 rows)

SELECT CW_COL1, CASE WHEN CW_COL1=1 THEN 'one' WHEN CW_COL1=2 THEN 'two' ELSE 'other'
END FROM tpcds.case_when_t1;
 cw_col1 | case
---------+-------
       3 | other
       1 | one
       2 | two
(3 rows)

DROP TABLE tpcds.case_when_t1;
```

- DECODE

  DECODE的语法图请参见**图8-2**。

图 8-2 decode::=



版权所有 © 华为云计算技术有限公司

将表达式base_expr与后面的每个compare(n) 进行比较，如果匹配返回相应的value(n)。如果没有发生匹配，则返回default。

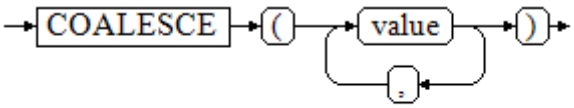
示例请参见**条件表达式函数**。

```
SELECT DECODE('A','A',1,'B',2,0);
 case
------
    1
(1 row)
```

- COALESCE

  COALESCE的语法图请参见**图8-3**。

  图 8-3 coalesce::=

  COALESCE返回它的第一个非NULL的参数值。如果参数都为NULL，则返回NULL。它常用于在显示数据时用缺省值替换NULL。和CASE表达式一样，COALESCE只计算用来判断结果的参数，即在第一个非空参数右边的参数不会被计算。

  示例：

  ```
  CREATE TABLE tpcds.c_tabl(description varchar(10), short_description varchar(10), last_value
  varchar(10))
  DISTRIBUTE BY HASH (last_value);

  INSERT INTO tpcds.c_tabl VALUES('abc', 'efg', '123');
  INSERT INTO tpcds.c_tabl VALUES(NULL, 'efg', '123');

  INSERT INTO tpcds.c_tabl VALUES(NULL, NULL, '123');

  SELECT description, short_description, last_value, COALESCE(description, short_description, last_value)
  FROM tpcds.c_tabl ORDER BY 1, 2, 3, 4;
   description | short_description | last_value | coalesce
  -------------+-------------------+------------+----------
   abc         | efg               | 123        | abc
               | efg               | 123        | efg
               |                   | 123        | 123
  (3 rows)

  DROP TABLE tpcds.c_tabl;
  ```

  如果description不为NULL，则返回description的值，否则计算下一个参数short_description；如果short_description不为NULL，则返回short_description的值，否则计算下一个参数last_value；如果last_value不为NULL，则返回last_value的值，否则返回（none）。

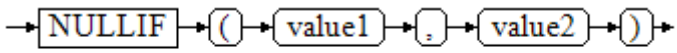
  ```
  SELECT COALESCE(NULL,'Hello World');
    coalesce
  ---------------
   Hello World
  (1 row)
  ```

- NULLIF

  NULLIF的语法图请参见**图8-4**。

 版权所有 © 华为云计算技术有限公司

图 8-4 nullif::=



只有当value1和value2相等时，NULLIF才返回NULL。否则它返回value1。

示例：

```
CREATE TABLE tpcds.null_if_t1 (
    NI_VALUE1 VARCHAR(10),
    NI_VALUE2 VARCHAR(10)
)  DISTRIBUTE BY HASH (NI_VALUE1);

INSERT INTO tpcds.null_if_t1 VALUES('abc', 'abc');
INSERT INTO tpcds.null_if_t1 VALUES('abc', 'efg');

SELECT NI_VALUE1, NI_VALUE2, NULLIF(NI_VALUE1, NI_VALUE2) FROM tpcds.null_if_t1 ORDER BY 1,
2, 3;

 ni_value1 | ni_value2 | nullif
-----------+-----------+--------
 abc       | abc       |
 abc       | efg       | abc
(2 rows)
DROP TABLE tpcds.null_if_t1;
```

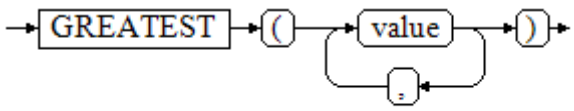
如果value1等于value2则返回NULL，否则返回value1。

```
SELECT NULLIF('Hello','Hello World');
 nullif
--------
 Hello
(1 row)
```

- GREATEST（最大值），LEAST（最小值）

  GREATEST的语法图请参见**图8-5**。

  图 8-5 greatest::=

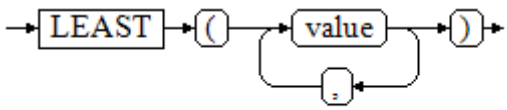
  

  从一个任意数字表达式的列表里选取最大的数值。

  ```
  SELECT greatest(9000,155555,2.01);
   greatest
  ----------
    155555
  (1 row)
  ```

  LEAST的语法图请参见**图8-6**。

  图 8-6 least::=

版权所有 © 华为云计算技术有限公司

从一个任意数字表达式的列表里选取最小的数值。

以上的数字表达式必须都可以转换成一个普通的数据类型，该数据类型将是结果类型。

列表中的NULL值将被忽略。只有所有表达式的结果都是NULL的时候，结果才是NULL。

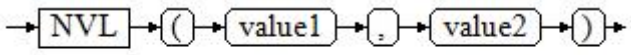
```
SELECT least(9000,2);
 least
--------
    2
(1 row)
```

示例请参见**条件表达式函数**。

- NVL

  NVL的语法图请参见**图8-7**。

  图 8-7 nvl::=

  如果value1为NULL则返回value2，如果value1非NULL，则返回value1。

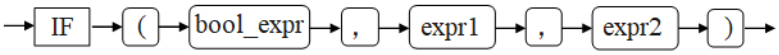
  示例：

```
SELECT nvl(null,1);
nvl
-----
 1
(1 row)
SELECT nvl ('Hello World' ,1);
      nvl
---------------
 Hello World
(1 row)
```

- IF

  IF的语法图请参见**图8-8**。

  图 8-8 if::=

  当bool_expr为true时，返回expr1，否则返回expr2。

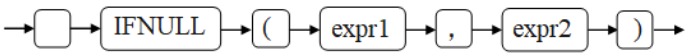
  示例请参见**条件表达式函数**。

- IFNULL

  IFNULL的语法图请参见**图8-9**。

 版权所有 © 华为云计算技术有限公司

图 8-9 ifnull::=



当expr1不为NULL时，返回expr1，否则返回expr2。

示例请参见**条件表达式函数**。

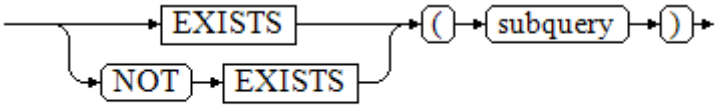
8.3 子查询表达式

子查询表达式主要有以下几种：

- EXISTS/NOT EXISTS

  EXISTS/NOT EXISTS的语法图请参见**图8-10**。

  图 8-10 EXISTS/NOT EXISTS::=

  EXISTS的参数是一个任意的SELECT语句，或者说子查询。系统对子查询进行运算以判断它是否返回行。如果它至少返回一行，则EXISTS结果就为"真"；如果子查询没有返回任何行， EXISTS的结果是"假"。

  这个子查询通常只是运行到能判断它是否可以生成至少一行为止，而不是等到全部结束。

  示例：

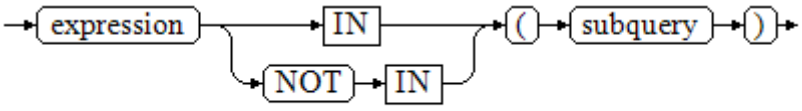
  ```
  SELECT sr_reason_sk,sr_customer_sk FROM tpcds.store_returns WHERE EXISTS (SELECT d_dom FROM tpcds.date_dim WHERE d_dom = store_returns.sr_reason_sk and sr_customer_sk <10);
  sr_reason_sk | sr_customer_sk
  --------------+----------------
            13 |              2
            22 |              5
            17 |              7
            25 |              7
             3 |              7
            31 |              5
             7 |              7
            14 |              6
            20 |              4
             5 |              6
            10 |              3
             1 |              5
            15 |              2
             4 |              1
            26 |              3
  (15 rows)
  ```

- IN/NOT IN

  IN/NOT IN的语法请参见**图8-11**。

 版权所有 © 华为云计算技术有限公司

图 8-11 IN/NOT IN::=



右边是一个圆括号括起来的子查询，它必须只返回一个字段。左边表达式对子查询结果的每一行进行一次计算和比较。如果找到任何相等的子查询行，则IN结果为"真"。如果没有找到任何相等行，则结果为"假"（包括子查询没有返回任何行的情况）。

表达式或子查询行里的NULL遵照SQL处理布尔值和NULL组合时的规则。如果两个行对应的字段都相等且非空，则这两行相等；如果任意对应字段不等且非空，则这两行不等；否则结果是未知（NULL）。如果每一行的结果都是不等或NULL，并且至少有一个NULL，则IN的结果是NULL。

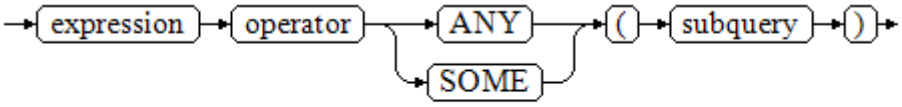
示例：

```
 SELECT sr_reason_sk,sr_customer_sk  FROM tpcds.store_returns WHERE sr_customer_sk IN (SELECT
d_dom FROM tpcds.date_dim WHERE d_dom < 10);
sr_reason_sk | sr_customer_sk
--------------+----------------
          10 |              3
          26 |              3
          22 |              5
          31 |              5
           1 |              5
          32 |              5
          32 |              5
           4 |              1
          15 |              2
          13 |              2
          33 |              4
          20 |              4
          33 |              8
           5 |              6
          14 |              6
          17 |              7
           3 |              7
          25 |              7
           7 |              7
(19 rows)
```

- ANY/SOME

  ANY/SOME的语法图请参见图8-12。

图 8-12 any/some::=



右边是一个圆括号括起来的子查询，它必须只返回一个字段。左边表达式使用operator对子查询结果的每一行进行一次计算和比较，其结果必须是布尔值。如果至少获得一个真值，则ANY结果为“真”。如果全部获得假值，则结果是“假”（包括子查询没有返回任何行的情况）。SOME是ANY的同义词。IN与ANY可以等效替换。

 版权所有 © 华为云计算技术有限公司

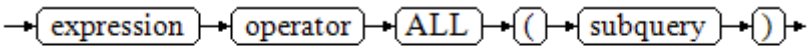
示例：

```
SELECT sr_reason_sk,sr_customer_sk  FROM tpcds.store_returns WHERE sr_customer_sk < ANY
(SELECT d_dom FROM tpcds.date_dim WHERE d_dom < 10);
sr_reason_sk | sr_customer_sk
--------------+----------------
          26 |              3
          17 |              7
          32 |              5
          32 |              5
          13 |              2
          31 |              5
          25 |              7
           5 |              6
           7 |              7
          10 |              3
           1 |              5
          14 |              6
           4 |              1
           3 |              7
          22 |              5
          33 |              4
          20 |              4
          33 |              8
          15 |              2
(19 rows)
```

- ALL

  ALL的语法请参见图8-13。

  图 8-13 all::=

  右边是一个圆括号括起来的子查询，它必须只返回一个字段。左边表达式使用operator对子查询结果的每一行进行一次计算和比较，其结果必须是布尔值。如果全部获得真值，ALL结果为"真"（包括子查询没有返回任何行的情况）。如果至少获得一个假值，则结果是"假"。

  示例：

```
SELECT sr_reason_sk,sr_customer_sk  FROM tpcds.store_returns WHERE sr_customer_sk < all(SELECT
d_dom FROM tpcds.date_dim WHERE d_dom < 10);
 sr_reason_sk | sr_customer_sk
--------------+----------------
(0 rows)
```

8.4 数组表达式

IN

expression **IN** *(value [, ...])*

右侧括号中的是一个表达式列表。左侧表达式的结果与表达式列表的内容进行比较。如果列表中的内容符合左侧表达式的结果，则IN的结果为true。如果没有相符的结果，则IN的结果为false。

示例如下：

```
SELECT 8000+500 IN (10000, 9000) AS RESULT;
 result
```


版权所有 © 华为云计算技术有限公司

```
----------
 f
(1 row)
```

说明

如果表达式结果为null，或者表达式列表不符合表达式的条件且右侧表达式列表返回结果至少一处为空，则IN的返回结果为null，而不是false。这样的处理方式和SQL返回空值的布尔组合规则是一致的。

NOT IN

expression **NOT IN** *(value [, ...])*

右侧括号中的是一个表达式列表。左侧表达式的结果与表达式列表的内容进行比较。如果在列表中的内容没有符合左侧表达式结果的内容，则NOT IN的结果为true。如果有符合的内容，则NOT IN的结果为false。

示例如下：

```
SELECT 8000+500 NOT IN (10000, 9000) AS RESULT;
 result
----------
 t
(1 row)
```

说明

如果查询语句返回结果为空，或者表达式列表不符合表达式的条件且右侧表达式列表返回结果至少一处为空，则NOT IN的返回结果为null，而不是false。这样的处理方式和SQL返回空值的布尔组合规则是一致的。

提示：在所有情况下X NOT IN Y等价于NOT(X IN Y)。

ANY/SOME (array)

expression operator **ANY** *(array expression)*

expression operator **SOME** *(array expression)*

```
SELECT 8000+500 < SOME (array[10000,9000]) AS RESULT;
 result
----------
 t
(1 row)
SELECT 8000+500 < ANY (array[10000,9000]) AS RESULT;
 result
----------
 t
(1 row)
```

右侧括号中的是一个数组表达式，它必须产生一个数组值。左侧表达式的结果使用操作符对数组表达式的每一行结果都进行计算和比较，比较结果必须是布尔值。

- 如果对比结果至少获取一个真值，则ANY的结果为true。
- 如果对比结果没有真值，则ANY的结果为false。

说明

如果结果没有真值，并且数组表达式生成至少一个值为null，则ANY的值为NULL，而不是false。这样的处理方式和SQL返回空值的布尔组合规则是一致的。

SOME是ANY的同义词。

 版权所有 © 华为云计算技术有限公司

ALL (array)

expression operator **ALL** *(array expression)*

右侧括号中的是一个数组表达式，它必须产生一个数组值。左侧表达式的结果使用操作符对数组表达式的每一行结果都进行计算和比较，比较结果必须是布尔值。

- 如果所有的比较结果都为真值（包括数组不含任何元素的情况），则ALL的结果为true。
- 如果结果有任一结果是false，则ALL的结果为false。

如果数组表达式产生一个NULL数组，则ALL的结果为NULL。如果左边表达式的值为NULL，则ALL的结果通常也为NULL(某些不严格的比较操作符可能得到不同的结果)。另外，如果右边的数组表达式中包含null元素并且比较结果没有假值，则ALL的结果将是NULL(某些不严格的比较操作符可能得到不同的结果)， 而不是真。这样的处理方式和SQL返回空值的布尔组合规则是一致的。

```
SELECT 8000+500 < ALL (array[10000,9000]) AS RESULT;
 result
----------
 t
(1 row)
```

8.5 行表达式

语法：

row_constructor operator row_constructor

两边都是一个行构造器，两行值必须具有相同数目的字段，每一行都进行比较，行比较允许使用=，<>，<，<=，>=等操作符，或其中一个相似的语义符。

=<>和别的操作符使用略有不同。如果两行值的所有字段都是非空并且相等，则认为两行是相等的；如果两行值的任意字段为非空并且不相等，则认为两行是不相等的；否则比较结果是未知的（null）。

对于<，<=，>，> =的情况下，行中元素从左到右依次比较，直到遇到一对不相等的元素或者一对为空的元素。如果这对元素中存在至少一个null值，则比较结果是未知的（null），否则这对元素的比较结果为最终的结果。

示例：

```
SELECT ROW(1,2,NULL) < ROW(1,3,0) AS RESULT;
 result
----------
 t
(1 row)
```

 版权所有 © 华为云计算技术有限公司

9 类型转换

9.1 概述

背景信息

在SQL语言中，每个数据都与一个决定其行为和用法的数据类型相关。DWS提供一个可扩展的数据类型系统，该系统比其他SQL实现更具通用性和灵活性。因而，DWS中大多数类型转换是由通用规则来管理的，这种做法允许使用混合类型的表达式。

DWS扫描/分析器只将词法元素分解成五个基本种类：整数、浮点数、字符串、标识符和关键字。大多数非数字类型首先表现为字符串。SQL语言的定义允许将常量字符串声明为具体的类型。例，下面查询：

```
SELECT text 'Origin' AS "label", point '(0,0)' AS "value";
 label  | value
--------+-------
 Origin | (0,0)
(1 row)
```

示例中有两个文本常量，类型分别为text和point。如果没有为字符串文本声明类型，则该文本首先被定义成一个unknown类型。

在DWS分析器里，有四种基本的SQL结构需要独立的类型转换规则：

- 函数调用

  多数SQL类型系统是建筑在一套丰富的函数上的。函数调用可以有一个或多个参数。因为SQL允许函数重载，所以不能通过函数名直接找到要调用的函数，分析器必须根据函数提供的参数类型选择正确的函数。

- 操作符

  SQL允许在表达式上使用前缀或后缀（单目）操作符，也允许表达式内部使用双目操作符（两个参数）。像函数一样，操作符也可以被重载，因此操作符的选择也和函数一样取决于参数类型。

- 值存储

  INSERT和UPDATE语句将表达式结果存入表中。语句中的表达式类型必须和目标字段的类型一致或者可以转换为一致。

- UNION，CASE和相关构造

  因为联合SELECT语句中的所有查询结果必须在一列里显示出来，所以每个SELECT子句中的元素类型必须相互匹配并转换成一个统一类型。类似地，一个CASE构造

 版权所有 © 华为云计算技术有限公司

的结果表达式必须转换成统一的类型，这样整个case表达式会有一个统一的输出类型。同样的要求也存在于ARRAY构造以及GREATEST和LEAST函数中。

系统表pg_cast存储了有关数据类型之间的转换关系以及如何执行这些转换的信息。详细信息请参见PG_CAST。

语义分析阶段会决定表达式的返回值类型并选择适当的转换行为。数据类型的基本类型分类，包括：boolean、numeric、string、bitstring、datetime、timespan、geometric和network。每种类型都有一种或多种首选类型用于解决类型选择的问题。根据首选类型和可用的隐含转换，就可能保证有歧义的表达式（那些有多个候选解析方案的）得到有效的方式解决。

所有类型转换规则都是建立在下面几个基本原则上的：

- 隐含转换绝不能有奇怪的或不可预见的输出。
- 如果一个查询不需要隐含的类型转换，分析器和执行器不应该进行更多的额外操作。这就是说，任何一个类型匹配、格式清晰的查询不应该在分析器里耗费更多的时间，也不应该向查询中引入任何不必要的隐含类型转换调用。
- 另外，如果一个查询在调用某个函数时需要进行隐式转换，当用户定义了一个有正确参数的函数后，解释器应该选择使用新函数。

TD 兼容模式下，空串转换为数值类型的处理

- TD数据库不同于Oracle，Oracle将空串当做NULL进行处理，TD在将空串转换为数值类型的时候，默认将空串转换为0进行处理，因此查询空串会查询到数值为0的数据。同样地，在TD兼容模式下，字符串转换数值的过程中，也会将空串默认转换为相应数值类型的0值进行处理。除此之外，' - '、' + '、' '这些字符串也都会在TD兼容模式下默认转换为0进行处理，但是小数点字符串' . '会报错。例如：

```
create table t1(no int,col varchar);
insert into t1 values(1,'');
insert into t1 values(2,null);
select * from t1 where col is null;
 no | col
----+-----
  2 |
(1 row)
select * from t1 where col='';
 no | col
----+-----
  1 |
(1 row)
```

- MySQL兼容模式下对空串转换为数值类型的处理和TD兼容模式相同。

9.2 操作符

操作符类型解析

1. 从系统表pg_operator中选出要考虑的操作符。如果可以找到一个参数类型以及参数个数都一致的操作符，那么这个操作符就是最终使用的操作符。如果找到了多个备选的操作符，将从中选择一个最合适的。
2. 寻找最优匹配。
   a. 丢弃输入类型不匹配以及无法隐式转换成匹配的候选操作符。unknown文本在这种情况下可以转换成任何类型。如果只剩下一个候选操作符，则使用，否则继续下一步。

 版权所有 © 华为云计算技术有限公司

b. 查看所有候选操作符，并保留输入类型最匹配的操作符。此时，域被看作和其基本类型相同。如果没有完全匹配的操作符，则保留所有候选。如果只剩下一个候选操作符，则使用，否则继续下一步。

c. 查看所有候选操作符，保留需要类型转换时接受(属于输入数据类型的类型范畴的)首选类型位置最多的操作符。如果没有接受首选类型的操作符，则保留所有候选。如果只剩下一个候选操作符，则使用，否则继续下一步。

d. 如果有任何输入参数是unknown类型，请检查其余候选操作符对应参数位置的类型范畴。在每一个能够接受string类型范畴的位置使用string类型（这种偏向字符串的做法合理，因为unknown文本跟字符串相似）。另外，如果所有剩下的候选操作符都接受相同的类型范畴，则选择该类型范畴，否则会报错（因为在没有更多线索的条件下无法作出正确的选择）。现在丢弃不接受选定类型范畴的候选操作符。此外，如果有任意候选操作符接受该范畴中的首选类型，则丢弃该参数接受非首选类型的候选操作符。如果没有一个操作符能被保留，则保留所有候选。如果只剩下一个候选操作符，则使用，否则继续下一步。

e. 如果同时有unknown和已知参数，并且所有已知参数都是相同的类型，那么假设unknown参数也属于该类型，并检查哪些候选操作符在unknown参数位置接受该类型。如果只有一个操作符符合，则使用。否则，报错。

示例

示例1：阶乘操作符类型解析。在系统表中只有一个阶乘操作符（后缀!），它以bigint作为参数。扫描器给下面查询表达式的参数赋予bigint的初始类型：

```
SELECT 40 ! AS "40 factorial";
```

```
gaussdb=> SELECT 40 ! AS "40 factorial";
                  40 factorial
-----------------------------------------------
 815915283247897734345611269596115894272000000000
(1 row)
```

分析器对参数做类型转换，查询等效于：

```
SELECT CAST(40 AS bigint) ! AS "40 factorial";
```

示例2：字符串连接操作符类型分析。一种字符串风格的语法既可以用于字符串也可以用于复杂的扩展类型。未声明类型的字符串将被所有可能的候选操作符匹配。有一个未声明的参数的例子：

```
SELECT text 'abc' || 'def' AS "text and unknown";
```

```
gaussdb=> SELECT text 'abc' || 'def' AS "text and unknown";
 text and unknown
------------------
 abcdef
(1 row)
```

本例中分析器寻找两个参数都是text的操作符。确实有这样的操作符，两个参数都是text类型。

下面是连接两个未声明类型的值：

```
SELECT 'abc' || 'def' AS "unspecified";
```

 版权所有 © 华为云计算技术有限公司

```
gaussdb=> SELECT 'abc' || 'def' AS "unspecified";
 unspecified
-------------
 abcdef
(1 row)
```

说明

因为查询中没有声明任何类型，所以本例中对类型没有任何初始提示。因此，分析器查找所有候选操作符，发现既存在接受字符串类型范畴的操作符也存在接受位串类型范畴的操作符。因为字符串类型范畴是首选，所以选择字符串类型范畴的首选类型text作为解析未知类型文本的声明类型。

示例3：绝对值和取反操作符类型分析。DWS操作符表里面有几条记录对应于前缀操作符@，它们都用于为各种数值类型实现绝对值操作。其中之一用于float8类型，它是数值类型范畴中的首选类型。因此，在面对unknown输入的时候， DWS会使用该类型：

```
SELECT @ '-4.5' AS "abs";
```

```
gaussdb=> SELECT @ '-4.5' AS "abs";
 abs
-----
 4.5
(1 row)
```

此处，系统在应用选定的操作符之前隐式地转换unknown类型的文字为float8类型。

示例4：数组包含操作符类型分析。以操作符两侧带有一个已知类型和一个未知类型的情况为例：

```
SELECT array[1,2] <@ '{1,2,3}' as "is subset";
```

```
gaussdb=> SELECT array[1,2] <@ '{1,2,3}' as "is subset";
 is subset
-----------
 t
(1 row)
```

说明

DWS的pg_operator表中有多条记录对应于中缀操作符<@，但是只有“数组包含(anyarray <@ anyarray)”和“范围包含(anyelement <@ anyrange)”可在左侧接受一个整数数组。因为多态伪类型(参阅**伪类型**)不被认为是首选操作符类型，因此解析器无法在此基础上解决歧义。不过，**寻找最优匹配**的最后一条解析规则又指出，假定未知类型的文字和另一个已知的输入类型相同，也就是只有两个运算符之一可以匹配，所以选择数组包含。（如果选择了范围包含，就会报错，因为字符串格式不正确不能作为范围。）

9.3 函数

函数类型解析

1. 从系统表pg_proc中选择所有可能被选到的函数。如果使用了一个不带模式修饰的函数名称，那么认为该函数是那些在当前搜索路径中的函数。如果给出一个带修饰的函数名，那么只考虑指定模式中的函数。

 版权所有 © 华为云计算技术有限公司

如果搜索路径中找到了多个不同参数类型的函数。将从中选择一个合适的函数。

2. 查找和输入参数类型完全匹配的函数。如果找到一个，则用之。如果输入的实参类型都是unknown类型，则不会找到匹配的函数。
3. 如果未找到完全匹配，请查看该函数是否为一个特殊的类型转换函数。
4. 寻找最优匹配。
   a. 抛弃那些输入类型不匹配并且也不能隐式转换成匹配的候选函数。unknown文本在这种情况下可以转换成任何东西。如果只剩下一个候选项，则用之，否则继续下一步。
   b. 遍历所有候选函数，保留那些输入类型匹配最准确的。此时，域被看作和它们的基本类型相同。如果没有一个函数能准确匹配，则保留所有候选。如果只剩下一个候选项，则用之，否则继续下一步。
   c. 遍历所有候选函数，保留那些需要类型转换时接受首选类型位置最多的函数。如果没有接受首选类型的函数，则保留所有候选。如果只剩下一个候选项，则用之，否则继续下一步。
   d. 如果有任何输入参数是unknown类型，检查剩余的候选函数对应参数位置的类型范畴。在每一个能够接受字符串类型范畴的位置使用string类型（这种对字符串的偏爱是合适的，因为unknown文本确实像字符串）。另外，如果所有剩下的候选函数都接受相同的类型范畴，则选择该类型范畴，否则抛出一个错误（因为在没有更多线索的条件下无法作出正确的选择）。现在抛弃不接受选定的类型范畴的候选函数，然后，如果任意候选函数在那个范畴接受一个首选类型，则抛弃那些在该参数位置接受非首选类型的候选函数。如果没有一个候选符合这些测试则保留所有候选。如果只有一个候选函数符合，则使用它；否则，继续下一步。
   e. 如果同时有unknown和已知类型的参数，并且所有已知类型的参数有相同的类型，假设unknown参数也是这种类型，检查哪个候选函数可以在unknown参数位置接受这种类型。如果正好一个候选符合，那么使用它。否则，产生一个错误。

示例

示例1：圆整函数参数类型解析。只有一个round函数有两个参数（第一个是numeric，第二个是integer）。所以下面的查询自动把第一个类型为integer的参数转换成numeric类型。

```
SELECT round(4, 4);
 round
--------
 4.0000
(1 row)
```

实际上它被分析器转换成：

```
SELECT round(CAST (4 AS numeric), 4);
```

因为带小数点的数值常量初始时被赋予numeric类型，因此下面的查询将不需要类型转换，并且可能会略微高效一些：

```
SELECT round(4.0, 4);
```

示例2：子字符串函数类型解析。有好几个substr函数，其中一个接受text和integer类型。如果用一个未声明类型的字符串常量调用它，系统将选择接受string类型范畴的首选类型（也就是text类型）的候选函数。

```
SELECT substr('1234', 3);
 substr
```

 版权所有 © 华为云计算技术有限公司

```
--------
    34
(1 row)
```

如果该字符串声明为varchar类型，就像从表中取出来的数据一样，分析器将试着将其转换成text类型：

```
SELECT substr(varchar '1234', 3);
 substr
--------
    34
(1 row)
```

被分析器转换后实际上变成：

```
SELECT substr(CAST (varchar '1234' AS text), 3);
```

说明

分析器从pg_cast表中了解到text和varchar是二进制兼容的，意思是说一个可以传递给接受另一个的函数而不需要做任何物理转换。因此，在这种情况下，实际上没有做任何类型转换。

而且，如果以integer为参数调用函数，分析器将试图将其转换成text类型：

```
SELECT substr(1234, 3);
substr
--------
 34
(1 row)
```

被分析器转换后实际上变成：

```
SELECT substr(CAST (1234 AS text), 3);
 substr
--------
    34
(1 row)
```

9.4 值存储

值存储数据类型解析

1. 查找与目标字段准确的匹配。
2. 试着将表达式直接转换成目标类型。如果已知这两种类型之间存在一个已登记的转换函数，那么直接调用该转换函数即可。如果表达式是一个未知类型文本，该文本字符串的内容将交给目标类型的输入转换过程。
3. 检查目标类型是否有长度转换。长度转换是一个从某类型到自身的转换。如果在pg_cast表里面找到一个，那么在存储到目标字段之前先在表达式上应用。这样的转换函数总是接受一个额外的类型为integer的参数，它接收目标字段的atttypmod值（实际上是其声明长度，atttypmod的解释随不同的数据类型而不同），并且它可能接受一个boolean类型的第三个参数，表示转换是显式的还是隐式的。转换函数负责施加那些长度相关的语义，比如长度检查或者截断。

示例

character存储类型转换。

本示例演示向character(20) 类型列插入数据时的处理逻辑。尽管插入的字符串长度小于定义长度，但查询结果显示存储长度为20字节，说明系统自动进行了空白填充。

```
CREATE TABLE x1
(
```

 版权所有 © 华为云计算技术有限公司

```
    customer_sk             integer,
    customer_id             char(20),
    first_name              char(6),
    last_name               char(8)
)
with (orientation = column,compression=middle)
distribute by hash (last_name);

INSERT INTO x1(customer_sk, customer_id, first_name) VALUES (3769, 'abcdef', 'Grace');

SELECT customer_id, octet_length(customer_id) FROM x1;
     customer_id      | octet_length
----------------------+--------------
 abcdef               |           20
(1 row)
```

说明

character存储类型转换流程如下：

1. 两个unknown文本系统默认将其解析成text，这样就允许||操作符解析成text连接。
2. 操作符的text结果转换成bpchar（"空白填充的字符型"， character类型内部名称）以匹配目标字段类型。从text到bpchar的转换是二进制兼容的，此转换过程是隐含的，且实际上不调用任何函数。
3. 在系统表里找到长度转换函数bpchar(bpchar, integer, boolean)并应用于该操作符的结果和存储的字段长度。该函数执行必要的长度检查，并在数据尾部填充空白以满足定长要求。

9.5 UNION，CASE 和相关构造

SQL UNION构造把不相同的数据类型进行匹配输出为统一的数据类型结果集。因为SELECT UNION语句中的所有查询结果必须在一列里显示出来，所以每个SELECT子句中的元素类型必须相互匹配并转换成一个统一的数据类型。类似地，一个CASE构造的结果表达式必须转换成统一的类型，这样整个case表达式会有一个统一的输出类型。同样的要求也存在于ARRAY构造以及GREATEST和LEAST函数中。

UNION，CASE 和相关构造解析

- 如果所有输入都是相同的数据类型，不包括unknown类型（即输入的字符串文本未声明类型，该文本首先被定义成一个未知类型），那么解析成所输入的相同数据类型。
- 如果所有输入都是unknown类型则解析成text类型（字符串类型范畴的首选类型）。否则，忽略unknown输入。
- 如果输入不属于同一个类型范畴，查询失败（unknown类型除外）。
- 如果输入类型是同一个类型范畴，则选择该类型范畴的首选类型（union操作会选择第一个分支的类型作为所选类型的情况除外）。

  说明

  系统表pg_type中typcategory表示数据类型范畴， typispreferred表示是否是typcategory分类中的首选类型。
- 把所有输入转换为所选的类型（对于字符串保持原有长度）。如果从给定的输入到所选的类型没有隐式转换则失败。
- 若输入中含json、txid_snapshot、sys_refcursor或几何类型，则不能进行union。

 版权所有 © 华为云计算技术有限公司

对于 CASE、COALESCE、IF 和 IFNULL，在 TD 兼容模式下的处理

- 如果所有输入都是相同的数据类型，不包括unknown类型，那么解析成所输入的相同数据类型。
- 如果所有输入都是unknown类型则解析成text类型。
- 如果输入字符串（包括unknown，unknown当text来处理）和数字类型，那么解析成字符串类型，如果是其他不同的类型范畴，则报错。
- 如果输入类型是同一个类型范畴，则选择该类型的优先级较高的类型。
- 把所有输入转换为所选的类型。如果从给定的输入到所选的类型没有隐式转换则失败。

对于 CASE、COALESCE、IF 和 IFNULL，在 MySQL 兼容模式下的处理

- 如果所有输入都是相同的类型，不包括unknown类型，那么解析成所输入的相同数据类型。
- 如果所有输入都是unknown类型则解析成text类型。
- 如果输入是unknown类型和某一非unknown类型，则解析成该非unknown类型。
- 如果存在多种非unknown类型，将enum类型当做text类型，再进行比较。
- 如果输入类型是同一个类型范畴，则选择该类型的优先级较高的类型。如果是不同的类型范畴，则解析成text类型。
- 把所有输入转换为所选的类型。如果从给定的输入到所选的类型没有隐式转换则失败。

示例

示例1：Union中的未知类型解析。示例中未知类型文本'b'将被解析成text类型。

```
SELECT text 'a' AS "text" UNION SELECT 'b';
 text
------
 a
 b
(2 rows)
```

示例2：简单Union中的类型解析。文本1.2是numeric类型，且integer类型值1可以隐式地转换为numeric，因此使用numeric类型。

```
SELECT 1.2 AS "numeric" UNION SELECT 1;
 numeric
---------
       1
     1.2
(2 rows)
```

示例3：转换Union中的类型解析。示例中由于类型real不能被隐式地转换为integer，而integer可以隐式地转换成real类型，那么联合结果类型被系统决定为real。

```
SELECT 1 AS "real" UNION SELECT CAST('2.2' AS REAL);
 real
------
    1
  2.2
(2 rows)
```

示例4：COALESCE函数输入int和varchar类型，ORA模式下会报错，TD模式下解析为varchar类型，MySQL模式下解析为text类型。

 版权所有 © 华为云计算技术有限公司

```
--指定兼容模式创建数据库ora_db、td_db、mysql_db。
CREATE DATABASE ora_db dbcompatibility = 'ORA';
CREATE DATABASE td_db dbcompatibility = 'TD';
CREATE DATABASE mysql_db dbcompatibility = 'MySQL';

--切换数据库为ora_db。
\c ora_db

--创建表t1。
ora_db=# CREATE TABLE t1(a int, b varchar(10));

--查看coalesce参数输入int和varchar类型的查询语句的执行计划。
ora_db=# EXPLAIN SELECT coalesce(a, b) FROM t1;
ERROR:  COALESCE types integer and character varying cannot be matched
CONTEXT:  referenced column: coalesce

--删除表。
ora_db=# DROP TABLE t1;

--切换数据库为td_db。
ora_db=# \c td_db

--创建表t2。
td_db=# CREATE TABLE t2(a int, b varchar(10));

--查看coalesce参数输入int和varchar类型的查询语句的执行计划。
td_db=# EXPLAIN VERBOSE select coalesce(a, b) from t2;
                                  QUERY PLAN
------------------------------------------------------------------------------------------------
 id |                operation                 | E-rows | E-distinct | E-width | E-costs
----+------------------------------------------+--------+------------+---------+---------
  1 | ->  Data Node Scan on "__REMOTE_FQS_QUERY__" |      0 |            |       0 | 0.00

                    Targetlist Information (identified by plan id)
------------------------------------------------------------------------------------------------
  1 --Data Node Scan on "__REMOTE_FQS_QUERY__"
        Output: (COALESCE((t2.a)::character varying, t2.b))
        Node/s: All datanodes
        Remote query: SELECT COALESCE(a::character varying, b) AS "coalesce" FROM public.t2
(10 rows)

--删除表。
td_db=# DROP TABLE t2;

--切换数据库为mysql_db。
td_db=# \c mysql_db

--创建表t3。
mysql_db=# CREATE TABLE t3(a int, b varchar(10));

--查看coalesce参数输入int和varchar类型的查询语句的执行计划。
mysql_db=# EXPLAIN VERBOSE select coalesce(a, b) from t3;
                                  QUERY PLAN
------------------------------------------------------------------------------------------------
 id |                operation                 | E-rows | E-distinct | E-width | E-costs
----+------------------------------------------+--------+------------+---------+---------
  1 | ->  Data Node Scan on "__REMOTE_FQS_QUERY__" |      0 |            |       0 | 0.00

                    Targetlist Information (identified by plan id)
------------------------------------------------------------------------------------------------
  1 --Data Node Scan on "__REMOTE_FQS_QUERY__"
        Output: (COALESCE((t3.a)::text, (t3.b)::text))
        Node/s: All datanodes
        Remote query: SELECT COALESCE(a::text, b::text) AS "coalesce" FROM public.t3
(10 rows)

--删除表。
mysql_db=# DROP TABLE t3;
```

 版权所有 © 华为云计算技术有限公司

```
--切换数据库为gaussdb
mysql_db=# \c gaussdb

--删除数据库。
DROP DATABASE ora_db;
DROP DATABASE td_db;
DROP DATABASE mysql_db;
```


版权所有 © 华为云计算技术有限公司

10 全文检索

10.1 介绍

10.1.1 全文检索概述

文本搜索操作符在数据库中已存在多年。DWS为文本数据类型提供~、~*、LIKE和ILIKE操作符，但这些操作符缺乏现代信息系统所要求的许多必要属性，不过这一问题可以通过使用索引及词典进行解决。

文本检索缺乏信息系统所要求的必要属性：

- 没有语义支持，即使是英语。

  由于要识别派生词并不是那么容易，因此正则表达式也不能满足要求。如，satisfies和satisfy，当使用正则表达式寻找satisfy时，并不会查询到包含satisfies的文档。用户可以使用OR搜索多种派生形式，但过程非常繁琐。并且有些词会有上千的派生词，因此容易出错。

- 没有对搜索结果的分类（排序）。当搜索出成千的文档时，查找效率很低。
- 由于没有索引的支持，每一次的搜索需要遍历所有的文档，整体搜索比较缓慢。

使用全文索引可以对文档进行预处理，并且可以使后续的搜索更快速。预处理过程包括：

- 将文档解析成token。

  为每个文档标记不同类别的token是非常有必要的，例如：数字、文字、复合词、电子邮件地址，这样就可以做不同的处理。原则上token的类别依赖于具体的应用，但对于大多数的应用来说，可以使用一组预定义的token类。

- 将token转换为词素。

  词素像token一样是一个字符串，但它已经标准化处理，这样同一个词的不同形式是一样的。例如，标准化通常包括：将大写字母折成小写字母、删除后缀（如英语中的s或者es）。这将允许通过搜索找到同一个词的不同形式，不需要繁琐地输入所有可能的变形样式。同时，这一步通常会删除停用词。这些停用词通常因为太常见而对搜索无用。（总之，token是文档文本的原片段，而词素被认为是有用的索引和搜索词。）DWS使用词典执行这一步，且提供了各种标准的词典。

- 保存搜索优化后的预处理文档。

版权所有 © 华为云计算技术有限公司

比如，每个文档可以呈现为标准化词素的有序组合。伴随词素，通常还需要存储词素位置信息以用于邻近排序。因此文档包含的查询词越密集其排序越高。

词典能够对token如何标准化做到细粒度控制。使用合适的词典，可以定义不被索引的停用词。

数据类型tsvector用于存储预处理文档，tsquery用于存储查询条件，详细内容可参见**文本搜索类型**。为数据类型tsvector提供的函数和操作符可参见**文本检索函数和操作符**，其中最重要的是匹配运算符@@，请参见**基本文本匹配**。

10.1.2 文档概念

文档是全文搜索系统的搜索单元，例如：杂志上的一篇文章或电子邮件消息。文本搜索引擎必须能够解析文档，而且可以存储父文档的关联词素（关键词）。后续，这些关联词素用来搜索包含查询词的文档。

在DWS中，文档通常是一个数据库表中一行的文本字段，或者这些字段的可能组合（级联）。文档可能存储在多个表中或者需动态获取。换句话说，一个文档由被索引化的不同部分构成，因此无法存储为一个整体。比如：

```
SELECT d_dow || '-' || d_dom || '-' || d_fy_week_seq  AS identify_serials FROM tpcds.date_dim WHERE
d_fy_week_seq = 1;
identify_serials
------------------
 5-6-1
 0-8-1
 2-3-1
 3-4-1
 4-5-1
 1-2-1
 6-7-1
(7 rows)
```

须知

实际上，在这些示例查询中，应该使用coalesce防止一个独立的NULL属性导致整个文档的NULL结果。

另外一种可能是：文档在文件系统中作为简单的文本文件存储。在这种情况下，数据库可以用于存储全文索引并且执行搜索，同时可以使用一些唯一标识从文件系统中检索文档。然而，从数据库外部检索文件需要拥有系统管理员权限或者特殊函数支持。因此，还是将所有数据保存在数据库中比较方便。同时，将所有数据保存在数据库中可以方便地访问文档元数据以便于索引和显示。

为实现文本搜索，需先将每个文档预处理为tsvector格式。搜索与相关性排序都是在tsvector格式的文档上执行。原始文档仅在呈现给用户时才会被检索。因此tsvector虽常被称为“文档”，实际只是完整文档的一种紧凑表示。

10.1.3 基本文本匹配

DWS的全文检索基于匹配算子@@，当一个tsvector(document)匹配到一个tsquery(query)时，则返回true。其中，tsvector(document)和tsquery(query)两种数据类型可以任意排序。

```
SELECT 'a fat cat sat on a mat and ate a fat rat'::tsvector @@ 'cat & rat'::tsquery AS RESULT;
 result
----------
```

 版权所有 © 华为云计算技术有限公司

```
 t
(1 row)
SELECT 'fat & cow'::tsquery @@ 'a fat cat sat on a mat and ate a fat rat'::tsvector AS RESULT;
 result
----------
 f
(1 row)
```

正如上面例子表明，tsquery不仅是文本，且比tsvector包含的要多。tsquery包含已经标注化为词条的搜索词，同时可能是使用AND、OR、或NOT操作符连接的多个术语。详细请参见**文本搜索类型**。函数to_tsquery和plainto_tsquery对于将用户书写文本转换成适合的tsquery是非常有用的，比如将文本中的词标准化。类似地，to_tsvector用于解析和标准化文档字符串。因此，实际中文本搜索匹配看起来更像这样：

```
SELECT to_tsvector('fat cats ate fat rats') @@ to_tsquery('fat & rat') AS RESULT;
result
----------
 t
(1 row)
```

需要注意的是，下面这种方式是不可行的：

```
SELECT 'fat cats ate fat rats'::tsvector @@ to_tsquery('fat & rat')AS RESULT;
result
----------
 f
(1 row)
```

由于tsvector没有对rats进行标准化，所以rats不匹配rat。

@@操作符也支持text输入，允许一个文本字符串的显示转换为tsvector或者在简单情况下忽略tsquery。可用形式是：

```
tsvector @@ tsquery
tsquery  @@ tsvector
text @@ tsquery
text @@ text
```

形式text @@ tsquery等价于to_tsvector(text) @@ tsquery，而text @@ text等价于to_tsvector(text) @@ plainto_tsquery(text)。

10.1.4 分词器

全文检索功能还可以做更多事情：忽略索引某个词（停用词），处理同义词和使用复杂解析，例如，不仅基于空格的解析。这些功能通过文本搜索分词器控制。DWS支持多语言的预定义的分词器，并且可以创建分词器（gsql的\dF命令显示了所有可用分词器）。

在安装期间选择一个合适的分词器，并且在postgresql.conf中相应的设置default_text_search_config。如果为了整个集群使用同一个文本搜索分词器可以使用postgresql.conf中的值。如果需要在集群中使用不同分词器，可以使用ALTER DATABASE ... SET在任一数据库进行配置。用户也可以在每个会话中设置default_text_search_config。

每个依赖于分词器的文本搜索函数有一个可选的配置参数，用于明确声明所使用的分词器。仅当忽略这个参数的时候，才使用default_text_search_config。

为了更方便的建立自定义文本搜索分词器，可以通过简单的数据库对象建立分词器。DWS文本搜索功能提供了四种类型与分词器相关的数据库对象：

- 文本搜索解析器将文档分解为token，并且分类每个token（例如：词和数字）。
- 文本搜索词典将token转换成规范格式并且丢弃停用词。

 版权所有 © 华为云计算技术有限公司

- 文本搜索模板提供潜在的词典功能：一个词典指定一个模板，并且为模板设置参数。
- 文本搜索分词器选择一个解析器，并且使用一系列词典规范化语法分析器产生的token。

10.2 表和索引

10.2.1 搜索表

在不使用索引的情况下也可以进行全文检索。

- 一个简单查询：将body字段中包含science的每一行打印出来。

```
DROP SCHEMA IF EXISTS tsearch CASCADE;

CREATE SCHEMA tsearch;

CREATE TABLE tsearch.pgweb(id int, body text, title text, last_mod_date date);

INSERT INTO tsearch.pgweb VALUES(1, 'Philology is the study of words, especially the history and development of the words in a particular language or group of languages.', 'Philology', '2010-1-1');

INSERT INTO tsearch.pgweb VALUES(2, 'Mathematics is the science that deals with the logic of shape, quantity and arrangement.', 'Mathematics', '2010-1-1');

INSERT INTO tsearch.pgweb VALUES(3, 'Computer science is the study of processes that interact with data and that can be represented as data in the form of programs.', 'Computer science', '2010-1-1');

INSERT INTO tsearch.pgweb VALUES(4, 'Chemistry is the scientific discipline involved with elements and compounds composed of atoms, molecules and ions.', 'Chemistry', '2010-1-1');

INSERT INTO tsearch.pgweb VALUES(5, 'Geography is a field of science devoted to the study of the lands, features, inhabitants, and phenomena of the Earth and planets.', 'Geography', '2010-1-1');

INSERT INTO tsearch.pgweb VALUES(6, 'History is a subject studied in schools, colleges, and universities that deals with events that have happened in the past.', 'History', '2010-1-1');

INSERT INTO tsearch.pgweb VALUES(7, 'Medical science is the science of dealing with the maintenance of health and the prevention and treatment of disease.', 'Medical science', '2010-1-1');

INSERT INTO tsearch.pgweb VALUES(8, 'Physics is one of the most fundamental scientific disciplines, and its main goal is to understand how the universe behaves.', 'Physics', '2010-1-1');


SELECT id, body, title FROM tsearch.pgweb WHERE to_tsvector('english', body) @@ to_tsquery('english', 'science');
 id |                                                 body                                                 |  title
----
+----------------------------------------------------------------------------------------------------------------------------------------
----+---------

  2 | Mathematics is the science that deals with the logic of shape, quantity and
arrangement.                                      | Mathematics
  3 | Computer science is the study of processes that interact with data and that can be represented as
data in the form of programs. | Computer science
  5 | Geography is a field of science devoted to the study of the lands, features, inhabitants, and
phenomena of the Earth and planets.   | Geography
  7 | Medical science is the science of dealing with the maintenance of health and the prevention and
treatment of disease.           | Medical science
(4 rows)
```

像science这样的相关词也会被找到，因为这些词都被处理成了相同标准的词条。

上面的查询指定english配置来解析和规范化字符串。也可以省略此配置，通过default_text_search_config进行配置设置：

 版权所有 © 华为云计算技术有限公司

```
SHOW default_text_search_config;
 default_text_search_config
----------------------------
 pg_catalog.english
(1 row)

SELECT id, body, title FROM tsearch.pgweb WHERE to_tsvector(body) @@ to_tsquery('science');
 id |                                                  body                                                  |  title
----
+-------------------------------------------------------------------------------------------------------------------------------------
----+---------

 2 | Mathematics is the science that deals with the logic of shape, quantity and
arrangement.                                        | Mathematics
 3 | Computer science is the study of processes that interact with data and that can be represented as
data in the form of programs. | Computer science
 5 | Geography is a field of science devoted to the study of the lands, features, inhabitants, and
phenomena of the Earth and planets.   | Geography
 7 | Medical science is the science of dealing with the maintenance of health and the prevention and
treatment of disease.            | Medical science

(4 rows)
```

- 一个复杂查询：检索出在title或者body字段中包含treatment和science的最近10篇文档：

  ```
  SELECT title FROM tsearch.pgweb WHERE to_tsvector(title || ' ' || body) @@ to_tsquery('treatment &
  science') ORDER BY last_mod_date DESC LIMIT 10;
   title
  --------
  Medical science
  (1 row)
  ```

  为了清晰，举例中没有调用coalesce函数在两个字段中查找包含NULL的行。

  以上例子均在没有索引的情况下进行查询。对于大多数应用程序来说，这个方法很慢。因此除了偶尔的特定搜索，文本搜索在实际使用中通常需要创建索引。

10.2.2 创建索引

为了加速文本搜索，可以创建GIN索引。

```
CREATE INDEX pgweb_idx_1 ON tsearch.pgweb USING gin(to_tsvector('english', body));
```

to_tsvector()函数有两个版本，只输一个参数的版本和输两个参数的版本。

只输一个参数时，系统默认采用default_text_search_config所指定的分词器。

创建索引时必须使用to_tsvector的两参数版本，否则索引内容可能不一致。只有指定了分词器名称的全文检索函数才可以在索引表达式中使用。因为索引的内容不受default_text_search_config的影响。由于default_text_search_config的值可以随时调整，从而导致不同条目生成的tsvector采用了不同的分词器，并且无法区分究竟使用了哪个分词器。正确地转储和恢复这样的索引也是不支持的。

在上述创建索引中to_tsvector使用了两个参数，只有当查询时也使用了两个参数，且参数值与索引中相同时，才会使用该索引。例如，WHERE to_tsvector('english', body) @@ 'a & b' 可以使用索引，但WHERE to_tsvector(body) @@ 'a & b'不能使用索引。这确保索引各条目是使用相同的分词器创建的。

索引中的分词器名称由另一列指定时可以建立更复杂的表达式索引。例如：

```
CREATE INDEX pgweb_idx_2 ON tsearch.pgweb USING gin(to_tsvector('zhparser', body));
```

说明

本示例中zhparser仅支持UTF8/GBK的数据库编码格式，在Encoding为SQL_ASCII下会报错。

 版权所有 © 华为云计算技术有限公司

其中body是pgweb表中的一列。当对索引的各条目使用了哪个分词器进行记录时，允许在同一索引中存在混合分词器。在某些场景下这将是有用的。例如，文档集合中包含不同语言的文档时。再次强调，打算使用索引的查询必须措辞匹配，例如，WHERE to_tsvector(config_name, body) @@ 'a & b'与索引中的to_tsvector措辞匹配。

索引甚至可以连接列：

```
CREATE INDEX pgweb_idx_3 ON tsearch.pgweb USING gin(to_tsvector('english', title || ' ' || body));
```

另一个方法是创建一个单独的tsvector列控制to_tsvector的输出。下面的例子是title和body的连接， 当其他是NULL的时候，使用coalesce确保一个字段仍然会被索引：

```
ALTER TABLE tsearch.pgweb ADD COLUMN textsearchable_index_col tsvector;
UPDATE tsearch.pgweb SET textsearchable_index_col = to_tsvector('english', coalesce(title,'') || ' ' ||
coalesce(body,''));
```

然后为加速搜索创建一个GIN索引：

```
CREATE INDEX textsearch_idx_4 ON tsearch.pgweb USING gin(textsearchable_index_col);
```

现在，就可以执行一个快速全文搜索了：

```
SELECT title
FROM tsearch.pgweb
WHERE textsearchable_index_col @@ to_tsquery('science & Computer')
ORDER BY last_mod_date DESC
LIMIT 10;
 title
------------------
 Computer science
(1 row)
```

相比于一个表达式索引，单独列方法的一个优势是：它没有必要在查询时明确指定分词器以便能使用索引。正如上面例子所示，查询可以依赖于default_text_search_config。另一个优势是搜索比较快速，因为它没有必要重新利用to_tsvector调用来验证索引匹配。表达式索引方法更容易建立，且它需要较少的磁盘空间，因为tsvector形式没有明确存储。

10.2.3 索引使用约束

下面是一个使用索引的例子，由于SQL_ASCII的数据库编码格式不支持中文字符，请在Encoding为UTF8/GBK的数据库中执行以下示例：

```
create table table1 (c_int int,c_bigint bigint,c_varchar varchar,c_text text) with(orientation=row);

create text search configuration ts_conf_1(parser=POUND);
create text search configuration ts_conf_2(parser=POUND) with(split_flag='%');

set default_text_search_config='ts_conf_1';
create index idx1 on table1 using gin(to_tsvector(c_text));

set default_text_search_config='ts_conf_2';
create index idx2 on table1 using gin(to_tsvector(c_text));

select c_varchar,to_tsvector(c_varchar) from table1 where to_tsvector(c_text) @@ plainto_tsquery('￥#@……&**')  and to_tsvector(c_text) @@ plainto_tsquery('某公司 ')  and c_varchar is not null order by 1 desc limit 3;
```

该例子的关键点是表table1的同一个列c_text上建立了两个GIN索引：idx1和idx2，但这两个索引是在不同default_text_search_config的设置下建立的。该例子和同一张表的同一个列上建立普通索引的不同之处在于：

- GIN索引使用了不同的parser（即分隔符不同），那么idx1和idx2的索引数据是不同的；

 版权所有 © 华为云计算技术有限公司

- 在同一张表的同一个列上建立的多个普通索引的索引数据是相同的；

因此当执行同一个查询时，使用idx1和idx2查询出的结果是不同的。

使用约束

通过上面的例子，索引使用满足如下条件时：

- 在同一个表的同一个列上建立了多个GIN索引；
- 这些GIN索引使用了不同的parser（即分隔符不同）；
- 在查询中使用了该列，且执行计划中使用索引进行扫描；
  为了避免使用不同GIN索引导致查询结果不同的问题，需要保证在物理表的一列上只有一个gin索引可用。

10.3 控制文本搜索

10.3.1 解析文档

DWS中提供了to_tsvector函数把文档处理成tsvector数据类型。

to_tsvector([*config* **regconfig**,] *document* **text**) **returns tsvector**

to_tsvector将文本文档解析为token，再将token简化到词素，并返回一个tsvector。其中tsvector中列出了词素及它们在文档中的位置。文档是根据指定的或默认的文本搜索分词器进行处理的。这里有一个简单的例子：

```
SELECT to_tsvector('english', 'a fat  cat sat on a mat - it ate a fat rats');
                  to_tsvector
-----------------------------------------------------
 'ate':9 'cat':3 'fat':2,11 'mat':7 'rat':12 'sat':4
```

通过以上例子可发现结果tsvector不包含词a、on或者it，rats变成rat，并且忽略标点符号-。

to_tsvector函数内部调用一个解析器，将文档的文本分解成token并给每个token指定一个类型。对于每个token，有一系列词典可供查询。词典系列因token类型的不同而不同。识别token的第一本词典将发出一个或多个标准词素来表示token。例如：

- rats变成rat因为词典认为词rats是rat的复数形式。
- 有些词被作为停用词（请参考**停用词**），这样它们就会被忽略，因为它们出现得太过频繁以致于搜索中没有用处。比如例子中的a、on和it。
- 如果没有词典识别token，那么它也被忽略。在这个例子中，符号“-”被忽略，因为词典没有给它分配token类型（空间符号），即空间符号永远不会被索引。

语法解析器、词典和要索引的token类型由选定的文本搜索分词器决定。可以在同一个数据库中有多种不同的分词器，以及提供各种语言的预定义分词器。在以上例子中，使用缺省分词器english。

函数setweight可以给tsvector的记录加权重，权重是字母A、B、C、D之一。这通常用于标记来自文档不同部分的记录，比如标题、正文。之后，这些信息可以用于排序搜索结果。

因为to_tsvector(NULL)会返回空，当字段可能是空的时候，建议使用coalesce。以下是推荐的为结构化文档创建tsvector的方法：

 版权所有 © 华为云计算技术有限公司

```
CREATE TABLE tsearch.tt (id int, title text, keyword text, abstract text, body text, ti tsvector);

INSERT INTO tsearch.tt(id, title, keyword, abstract, body) VALUES (1, 'book', 'literature', 'Ancient
poetry','Tang poem Song jambic verse');

UPDATE tsearch.tt SET ti =
    setweight(to_tsvector(coalesce(title,'')), 'A')    ||
    setweight(to_tsvector(coalesce(keyword,'')), 'B')  ||
    setweight(to_tsvector(coalesce(abstract,'')), 'C') ||
    setweight(to_tsvector(coalesce(body,'')), 'D');
DROP TABLE tsearch.tt;
```

上例使用setweight标记已完成的tsvector中的每个词的来源，并且使用tsvector连接操作符||合并标记过的tsvector值，**处理tsvector**一节详细介绍了这些操作。

10.3.2 解析查询

DWS提供了函数to_tsquery和plainto_tsquery将查询转换为tsquery数据类型，to_tsquery提供比plainto_tsquery更多的功能，但对其输入要求更严格。

to_tsquery([config regconfig,] querytext text) returns tsquery

to_tsquery从querytext中创建一个tsquery，querytext必须由布尔运算符& (AND)，| (OR)和! (NOT)分割的单个token组成。这些运算符可以用圆括号分组。换句话说，to_tsquery输入必须遵循tsquery输入的通用规则，具体请参见**文本搜索类型**。不同的是基本tsquery以token表面值作为输入，而to_tsquery使用指定或默认分词器将每个token标准化成词素，并依据分词器丢弃属于停用词的token。例如：

```
SELECT to_tsquery('english', 'The & Fat & Rats');
  to_tsquery
---------------
 'fat' & 'rat'
(1 row)
```

像在基本tsquery中的输入一样，weight(s)可以附加到每个词素来限制它只匹配那些有相同weight(s)的tsvector词素。比如：

```
SELECT to_tsquery('english', 'Fat | Rats:AB');
    to_tsquery
------------------
 'fat' | 'rat':AB
(1 row)
```

同时，*也可以附加到词素来指定前缀匹配：

```
SELECT to_tsquery('supern:*A & star:A*B');
       to_tsquery
--------------------------
 'supern':*A & 'star':*AB
(1 row)
```

这样的词素将匹配tsquery中指定字符串和权重的项。

plainto_tsquery([config regconfig,] querytext text) returns tsquery

plainto_tsquery将未格式化的文本querytext变换为tsquery。类似于to_tsvector，文本被解析并且标准化，然后在存在的词之间插入&(AND)布尔算子。

比如：

```
SELECT plainto_tsquery('english', 'The Fat Rats');
 plainto_tsquery
-----------------
 'fat' & 'rat'
(1 row)
```


版权所有 © 华为云计算技术有限公司

请注意，plainto_tsquery无法识别布尔运算符、权重标签，或在其输入中的前缀匹配标签：

```
SELECT plainto_tsquery('english', 'The Fat & Rats:C');
  plainto_tsquery
---------------------
 'fat' & 'rat' & 'c'
(1 row)
```

在这里，所有输入的标点符号作为空格符号丢弃。

10.3.3 排序查询结果

排序会根据查询内容评估文档的相关性，将最匹配的结果优先展示。DWS提供了两个预置的排序函数。函数考虑了词法，距离，和结构信息；也就是，考虑查询词在文档中出现的频率、紧密程度、以及出现的地方在文档中的重要性。然而，相关性的概念是模糊的，并且是跟应用强相关的。不同的应用程序可能需要额外的信息来排序，比如，文档的修改时间，内置的排序函数等。也可以开发自己的排序函数或者采用附加因素组合这些排序函数的结果来满足特定需求。

两个预置的排序函数：

```
ts_rank([ weights float4[], ] vector tsvector, query tsquery [, normalization integer ]) returns float4
```

基于词素匹配率对vector进行排序：

```
ts_rank_cd([ weights float4[], ] vector tsvector, query tsquery [, normalization integer ]) returns float4
```

该函数需要位置信息的输入。因此它不能在"剥离"tsvector值的情况下运行—它将总是返回零。

对于这两个函数，可选的weights参数提供给词加权重的能力，词的权重大小取决于所加的权值。权重阵列指定在排序时为每类词汇加多大的权重。

```
{D-weight, C-weight, B-weight, A-weight}
```

如果没有提供weights，则使用缺省值：{0.1, 0.2, 0.4, 1.0}

通常的权重是用来标记文档特殊领域的词，如标题或最初的摘要，所以相对于文章主体中的词它们有着更高或更低的重要性。

由于较长的文档有更多的机会包含查询词，因此有必要考虑文档的大小。例如，包含有5个搜索词的一百字文档比包含有5个搜索词的一千字文档相关性更高。两个预置的排序函数都采用了一个整型的标准化选项来定义文档长度是否影响排序及如何影响。这个整型选项控制多个行为，所以它是一个屏蔽字：可以使用|指定一个或多个行为（例如，2|4）。

- 0（缺省）表示：跟长度大小没有关系
- 1 表示：排名（rank）除以(文档长度的对数+1)
- 2表示：排名除以文档的长度
- 4表示：排名除以两个扩展词间的调和平均距离。只能使用ts_rank_cd实现
- 8表示：排名除以文档中单独词的数量
- 16表示：排名除以单独词数量的对数+1
- 32表示：排名除以排名本身+1

当指定多个标志位时，会按照所列的顺序依次进行转换。

需要特别注意的是，排序函数不使用任何全局信息，所以不可能产生一个某些情况下需要的1%或100%的理想标准值。标准化选项32 (rank/(rank+1))可用于所有规模的从

 版权所有 © 华为云计算技术有限公司

零到一之间的排序。需要注意的是：这只是一个表面变化，并不会影响搜索结果的排序。

下面是一个例子，仅选择排名前十的匹配：

由于SQL_ASCII的数据库编码格式不支持中文字符，请在Encoding为UTF8/GBK的数据库中执行以下示例：

```
SELECT id, title, ts_rank_cd(to_tsvector(body), query) AS rank
FROM tsearch.pgweb, to_tsquery('science') query
WHERE query @@ to_tsvector(body)
ORDER BY rank DESC
LIMIT 10;
 id |  title  | rank
----+---------+------
 11 | Philology |   .2
  2 | Mathematics |   .1
 12 | Geography |   .1
 13 | Computer science |   .1
(4 rows)
```

使用标准化排序的示例：

```
SELECT id, title, ts_rank_cd(to_tsvector(body), query, 32 /* rank/(rank+1) */ ) AS rank
FROM tsearch.pgweb, to_tsquery('science') query
WHERE  query @@ to_tsvector(body)
ORDER BY rank DESC
LIMIT 10;
 id |  title  |   rank
----+---------+----------
 11 | Philology  | .166667
  2 | Mathematics | .0909091
 12 | Geography  | .0909091
 13 | Computer science  | .0909091
(4 rows)
```

使用中文分词法排序查询的示例：

```
CREATE TABLE tsearch.ts_zhparser(id int, body text);
INSERT INTO tsearch.ts_zhparser VALUES(1, '排序');
INSERT INTO tsearch.ts_zhparser VALUES(2, '排序查询');
INSERT INTO tsearch.ts_zhparser VALUES(3, '查询排序');
--精确匹配
SELECT id, body, ts_rank_cd(to_tsvector('zhparser',body), query) AS rank FROM tsearch.ts_zhparser,
to_tsquery('排序') query WHERE query @@ to_tsvector(body);
 id | body | rank
----+------+------
  1 | 排序 |   .1
(1 row)

--模糊匹配
SELECT id, body, ts_rank_cd(to_tsvector('zhparser',body), query) AS rank FROM tsearch.ts_zhparser,
to_tsquery('排序') query WHERE query @@ to_tsvector('zhparser',body);
 id |   body   | rank
----+----------+------
  3 | 查询排序 |   .1
  1 | 排序     |   .1
  2 | 排序查询 |   .1
(3 rows)
```

排序要遍历每个匹配的tsvector，因此资源消耗多，可能会因为I/O限制导致排序慢。可是这是很难避免的，因为实际查询中通常会有大量的匹配。

10.3.4 高亮搜索结果

搜索结果的理想显示是：列出每篇文档中与搜索相关的部分，并标识为什么与查询相关。搜索引擎能够显示标识了搜索词的文档片段。DWS提供了函数ts_headline支持这部分功能。

版权所有 © 华为云计算技术有限公司

ts_headline([config regconfig,] document text, query tsquery [, options text]) returns text

ts_headline的输入是带有查询条件的文档，其返回文档中的摘录，在摘录中查询词是高亮显示的。用来解析文档的分词器由config参数指定。如果省略config，则使用default_text_search_config的值所指定的分词器。

指定options字符串时，需由一个或多个option=value对组成，且必须用逗号分隔。options可以是下面的选项：

- StartSel，StopSel：分隔文档中出现的查询词，以区别于其他摘录词。当包含有空格或逗号时，必须用双引号将字符串引起来。
- MaxWords，MinWords：定义摘录的最长和最短值。
- ShortWord：在摘录的开始和结束会丢弃此长度或更短的词。默认值3会消除常见的英语冠词。
- HighlightAll：布尔标志。如果为真，整个文档将作为摘录。忽略前面三个参数的值。
- MaxFragments：要显示的文本摘录或片段的最大数量。默认值0表示选择非片段的摘录生成方法。 大于0的值表示选择基于片段的摘录生成。此方法查找带有尽可能多查询词的文本片段，并显示查询词周围的上下文片段。因此，查询词临近每个片段的中间，且查询词两边都有词。每个片段至多有MaxWords，并且长度为ShortWord或更短的词在每一个片段开始和结束被丢弃。如果在文档中没有找到所有的查询词，则文档中开头将显示MinWords单片段。
- FragmentDelimiter：当有一个以上的片段时，通过该字符串分隔这些片段。

不声明选项时，采用下面的缺省值：

```
StartSel=<b>, StopSel=</b>,
MaxWords=35, MinWords=15, ShortWord=3, HighlightAll=FALSE,
MaxFragments=0, FragmentDelimiter=" ... "
```

比如：

```
SELECT ts_headline('english',
'The most common type of search
is to find all documents containing given query terms
and return them in order of their similarity to the
query.',
to_tsquery('english', 'query & similarity'));
                        ts_headline
------------------------------------------------------------
 containing given <b>query</b> terms
 and return them in order of their <b>similarity</b> to the
 <b>query</b>.
(1 row)

SELECT ts_headline('english',
'The most common type of search
is to find all documents containing given query terms
and return them in order of their similarity to the
query.',
to_tsquery('english', 'query & similarity'),
'StartSel = <, StopSel = >');
                        ts_headline
-------------------------------------------------------
 containing given <query> terms
 and return them in order of their <similarity> to the
 <query>.
(1 row)
```

ts_headline使用原始文档，而不是tsvector摘录，因此使用起来会慢，应慎重使用。

 版权所有 © 华为云计算技术有限公司

10.4 附加功能

10.4.1 处理 tsvector

DWS提供了用来操作tsvector类型的函数和操作符。

- tsvector || tsvector

  tsvector连接操作符返回一个新的tsvector类型，它综合了两个tsvector中词素和位置信息，并保留词素的位置信息和权重标签。右侧的tsvector的起始位置位于左侧tsvector的最后位置，因此，新生成的tsvector几乎等同于将两个原始文档字符串连接后进行to_tsvector操作。（这个等价是不准确的，因为任何从左边tsvector中删除的停用词都不会影响结果，但是，在使用文本连接时，则会影响词素在右侧tsvector中的位置。）

  相较于对文本进行连接后再执行to_tsvector操作，使用tsvector类型进行连接操作的优势在于，可以对文档的不同部分使用不同配置进行解析。因为setweight函数会对给定的tsvector中的语素进行统一设置，如果想要对文档的不同部分设置不同的权重，需要在连接之前对文本进行解析和权重设置。

- setweight(vector tsvector, weight "char") returns tsvector

  setweight返回一个输入tsvector的副本，其中每一个位置都使用给定的权重做了标记。权值可以为A、B、C或D（D是tsvector副本的默认权重，并且不在输出中呈现）。当对tsvector进行连接操作时，这些权重标签将会被保留，文档不同部分以不同的权重进行排序。

  须知

  权重标签作用于位置，而不是词素。如果传入的tsvector已经被剥离了位置信息，那么setweight函数将什么都不做。

- length(vector tsvector) returns integer

  返回vector中的词素的数量。

- strip(vector tsvector) returns tsvector

  返回一个tsvector类型，其中包含输入的tsvector的同义词，但不包含任何位置和权重信息。虽然在相关性排序中，这里返回的tsvector要比未拆分的tsvector的作用小很多，但它通常都比未拆分的tsvector小的多。

10.4.2 处理查询

DWS提供了函数和操作符用来操作tsquery类型的查询。

- tsquery && tsquery

  返回两个给定查询tsquery的与结果。

- tsquery || tsquery

  返回两个给定查询tsquery的或结果。

- !! tsquery

  返回给定查询tsquery的非结果。

 版权所有 © 华为云计算技术有限公司

- numnode(query tsquery) returns integer

  返回tsquery中的节点数目（词素加操作符），这个函数在检查查询是否有效（返回值大于0），或者只包含停用词（返回值等于0）时，是有用的。例如：

```
SELECT numnode(plainto_tsquery('the any'));
NOTICE:  text-search query contains only stop words or doesn't contain lexemes, ignored
CONTEXT:  referenced column: numnode
 numnode
---------
       0

SELECT numnode('foo & bar'::tsquery);
 numnode
---------
       3
```

- querytree(query tsquery) returns text

  返回可用于索引搜索的tsquery部分，该函数对于检测非索引查询是有用的（例如只包含停用词或否定项）。例如:

```
SELECT querytree(to_tsquery('!defined'));
 querytree
-----------
 T
(1 row)
```

10.4.3 查询重写

ts_rewrite函数族可以从tsquery中搜索一个特定的目标子查询，并在该子查询每次出现的地方都替换为另一个子查询。 实际上这只是通过字符串替换而得到的一个特定tsquery版本。目标子查询和替换查询组合起来可以被认为是一个重写规则。一组类似的重写规则可以为搜索提供强大的帮助。例如，可以使用同义词扩大搜索范围（例如，new york, big apple, nyc, gotham）或限制搜索范围在用户直接感兴趣的热点话题上。

- ts_rewrite (query tsquery, target tsquery, substitute tsquery) returns tsquery

  ts_rewrite的这种形式只适用于一个单一的重写规则：任何出现目标子查询的地方都被无条件替换。例如：

```
SELECT ts_rewrite('a & b'::tsquery, 'a'::tsquery, 'c'::tsquery);
 ts_rewrite
------------
 'b' & 'c'
```

- ts_rewrite (query tsquery, select text) returns tsquery

  ts_rewrite的这种形式接受一个起始查询和SQL查询命令。 这里的查询命令是文本字符串形式，必须产生两个tsquery列。查询结果的每一行，第一个字段的值（目标子查询） 都会被第二个字段（替代子查询）替换。

  说明

  当多个规则需要重写时，重写顺序非常重要； 因此在实践中需要使用ORDER BY将源查询按照某些字段进行排序。

  例如：举一个现实生活中天文学上的例子。将使用表驱动的重写规则扩大supernovae的查询范围：

```
CREATE TABLE tsearch.aliases (id int, t tsquery, s tsquery);

INSERT INTO tsearch.aliases VALUES(1, to_tsquery('supernovae'), to_tsquery('supernovae|sn'));

SELECT ts_rewrite(to_tsquery('supernovae & crab'), 'SELECT t, s FROM tsearch.aliases');

          ts_rewrite
```

 版权所有 © 华为云计算技术有限公司

```
---------------------------------
 'crab' & ( 'supernova' | 'sn' )
```

可以通过更新表修改重写规则：

```
UPDATE tsearch.aliases
SET s = to_tsquery('supernovae|sn & !nebulae')
WHERE t = to_tsquery('supernovae');

SELECT ts_rewrite(to_tsquery('supernovae & crab'), 'SELECT t, s FROM tsearch.aliases');

                ts_rewrite
-----------------------------------------------
 'crab' & ( 'supernova' | 'sn' & !'nebula' )
```

需要重写的规则越多，重写操作就越慢。因为它要检查每一个可能匹配的规则。为了过滤明显的非候选规则，可以使用tsquery类型的操作符来实现。在下面的例子中， 只选择那些可能与原始查询匹配的规则：

```
SELECT ts_rewrite('a & b'::tsquery, 'SELECT t,s FROM tsearch.aliases WHERE ''a & b''::tsquery @> t');

 ts_rewrite
------------
 'b' & 'a'
(1 row)
DROP TABLE ts_rewrite;
```

10.4.4 收集文献统计

函数ts_stat可用于检查配置和查找候选停用词。

```
ts_stat(sqlquery text, [ weights text, ]
    OUT word text, OUT ndoc integer,
    OUT nentry integer) returns setof record
```

sqlquery是一个包含SQL查询语句的文本，该SQL查询将返回一个tsvector。ts_stat执行SQL查询语句并返回一个包含tsvector中每一个不同的语素（词）的统计信息。返回信息包括：

- word text：词素。
- ndoc integer：词素在文档（tsvector）中的编号。
- nentry integer：词素出现的频率。

如果设置了权重条件，只有标记了对应权重的词素才会统计频率。例如，在一个文档集中检索使用频率最高的十个单词：

```
SELECT * FROM ts_stat('SELECT to_tsvector(''english'', sr_reason_sk) FROM tpcds.store_returns WHERE
sr_customer_sk < 10') ORDER BY nentry DESC, ndoc DESC, word LIMIT 10;;
  word | ndoc | nentry
------+------+--------
 32   |    2 |      2
 33   |    2 |      2
 1    |    1 |      1
 10   |    1 |      1
 13   |    1 |      1
 14   |    1 |      1
 15   |    1 |      1
 17   |    1 |      1
 20   |    1 |      1
 22   |    1 |      1
(10 rows)
```

同样的情况，但是只计算权重为A或者B的单词使用频率：

```
SELECT * FROM ts_stat('SELECT to_tsvector(''english'', sr_reason_sk) FROM tpcds.store_returns WHERE
sr_customer_sk < 10', 'a') ORDER BY nentry DESC, ndoc DESC, word LIMIT 10;
 word | ndoc | nentry
```


版权所有 © 华为云计算技术有限公司

```
------+------+--------
(0 rows)
```

10.5 解析器

文本搜索解析器负责将原文档文本分解为多个token，并标识每个token的类型。这里的类型集由解析器本身定义。注意，解析器并不修改文本，它只是确定合理的单词边界。由于这一限制，人们更需要定制词典，而不是为每个应用程序定制解析器。

目前DWS提供了四个内置的解析器，分别为pg_catalog.default/pg_catalog.ngram/pg_catalog.zhparser/pg_catalog.pound，其中pg_catalog.default适用于英文分词场景，pg_catalog.ngram/pg_catalog.zhparser/pg_catalog.pound是为了支持中文全文检索功能新增的三种解析器，适用于中文及中英混合分词场景。

内置解析器pg_catalog.default，它能识别23种token类型，显示在**表10-1**中。

表 10-1 默认解析器类型

别名	描述	示例
asciiword	Word, all ASCII letters	elephant
word	Word, all letters	mañana
numword	Word, letters and digits	beta1
asciihword	Hyphenated word, all ASCII	up-to-date
hword	Hyphenated word, all letters	lógico-matemática
numhword	Hyphenated word, letters and digits	postgresql-beta1
hword_asciipart	Hyphenated word part, all ASCII	postgresql in the context postgresql-beta1
hword_part	Hyphenated word part, all letters	lógico or matemática in the context lógico-matemática
hword_numpart	Hyphenated word part, letters and digits	beta1 in the context postgresql-beta1
email	Email address	foo@example.com
protocol	Protocol head	http://
url	URL	example.com/stuff/index.html
host	Host	example.com
url_path	URL path	/stuff/index.html, in the context of a URL
file	File or path name	/usr/local/foo.txt, if not within a URL
sfloat	Scientific notation	-1.23E+56

版权所有 © 华为云计算技术有限公司

别名	描述	示例
float	Decimal notation	-1.234
int	Signed integer	-1234
uint	Unsigned integer	1234
version	Version number	8.3.0
tag	XML tag	<a href="dictionaries.html">
entity	XML entity	&
blank	Space symbols	(any whitespace or punctuation not otherwise recognized)

注意：对于解析器来说，一个“字母”的概念是由数据库的语言区域设置，即lc_ctype设置决定的。只包含基本ASCII字母的词被报告为一个单独的token类型，因为这类词有时需要被区分出来。大多数欧洲语言中，对token类型word和asciiword的处理方法是类似的。

email不支持某些由RFC 5322定义的有效电子邮件字符。具体来说，可用于email用户名的非字母数字字符仅包含句号、破折号和下划线。

解析器可能对同一内容进行重叠token。例如，包含连字符的单词将作为一个整体进行报告，其组件也会分别被报告：

```
SELECT alias, description, token FROM ts_debug('english','foo-bar-beta1');
      alias      |               description                |     token
-----------------+------------------------------------------+---------------
 numhword        | Hyphenated word, letters and digits      | foo-bar-beta1
 hword_asciipart | Hyphenated word part, all ASCII          | foo
 blank           | Space symbols                            | -
 hword_asciipart | Hyphenated word part, all ASCII          | bar
 blank           | Space symbols                            | -
 hword_numpart   | Hyphenated word part, letters and digits | beta1
```

这种行为是有必要的，因为它支持搜索整个复合词和各组件。这里是另一个例子：

```
SELECT alias, description, token FROM ts_debug('english','http://example.com/stuff/index.html');
  alias   |  description  |            token
----------+---------------+------------------------------
 protocol | Protocol head | http://
 url      | URL           | example.com/stuff/index.html
 host     | Host          | example.com
 url_path | URL path      | /stuff/index.html
```

N-gram是一种机械分词方法，适用于无语义中文分词场景。N-gram分词法可以保证分词的完备性，但是为了照顾所有可能，把很多不必要的词也加入到索引中，导致索引项增加。N-gram支持中文编码包括GBK、UTF-8。内置6种token类型，如**表10-2** 所示。

 版权所有 © 华为云计算技术有限公司

表 10-2 token 类型

Alias	Description
zh_words	chinese words
en_word	english word
numeric	numeric data
alnum	alnum string
grapsymbol	graphic symbol
multisymbol	multiple symbol

Zhparser是基于词典的语义分词方法，底层调用SCWS(https://github.com/hightman/scws)分词算法，适用于有语义的中文分词场景。SCWS是一套基于词频词典的机械式中文分词引擎，可以将一整段的中文文本正确地切分成词。支持GBK、UTF-8两种中文编码格式。内置26种token类型如表10-3所示：

表 10-3 token 类型

Alias	Description
A	形容词
B	区别词
C	连词
D	副词
E	叹词
F	方位词
G	语素
H	前接成分
I	成语
J	简称略语
K	后接成分
L	习用语
M	数词
N	名词
O	拟声词
P	介词
Q	量词

 版权所有 © 华为云计算技术有限公司

Alias	Description
R	代词
S	处所词
T	时间词
U	助词
V	动词
W	标点符号
X	未知词
Y	语气词
Z	状态词

Pound是一种固定格式分词方法，适用于无语义但待解析文本以固定分隔符分割开来的中英文分词场景。支持中文编码包括GBK、UTF8，支持英文编码包括ASCII。内置6种token类型，如**表4 token类型**所示；支持5种分隔符，如**表10-5**所示，在用户不进行自定义设置的情况下分隔符默认为“#”。Pound限制单个token长度不能超过256个字符。

表 10-4 token 类型

Alias	Description
zh_words	chinese words
en_word	english word
numeric	numeric data
alnum	alnum string
grapsymbol	graphic symbol
multisymbol	multiple symbol

表 10-5 分隔符类型

分隔符	描述
@	Special character
#	Special character
$	Special character
%	Special character

版权所有 © 华为云计算技术有限公司

分隔符	描述
/	Special character

10.6 词典

10.6.1 词典概述

词典用于定义停用词（stop words），即全文检索时不搜索哪些词。

词典还可以用于对同一词的不同形式进行规范化，这样同一个词的不同派生形式都可以进行匹配。规范化后的词称为词位（lexeme）。

除了提高检索质量外，词的规范化和删除停用词可以减少文档tsvector格式的大小，从而提高性能。词的规范化和删除停用词并不总是具有语言学意义，用户可以根据应用环境在词典定义文件中自定义规范化和删除规则。

一个词典是一个程序，接收标记（token）作为输入，并返回：

- 如果token在词典中已知，返回对应lexeme数组（注意，一个标记可能对应多个lexeme）。
- 一个lexeme。一个新token会代替输入token被传递给后继词典（当前词典可被称为过滤词典）。
- 如果token在词典中已知，但它是一个停用词，返回空数组。
- 如果词典不能识别输入的token，返回NULL。

DWS提供了多种语言的预定义词典，同时提供了五种预定义的词典模板，分别是Simple，Synonym，Thesaurus，Ispell，和Snowball，可用于创建自定义参数的新词典。

在使用全文检索时，建议用户：

- 可以在文本搜索配置中定义一个解析器，以及一组用于处理该解析器的输出标记词典。对于解析器返回的每个标记类型，可以在配置中指定不同的词典列表进行处理。当解析器输出一种类型的标记后，在对应列表的每个词典中会查阅该标记，直到某个词典识别它。如果它被识别为一个停用词， 或者没有任何词典识别，该token将被丢弃，即不被索引或检索到。通常情况下，第一个返回非空结果的词典决定了最终结果，后继词典将不会继续处理。但是一个过滤类型的词典可以依据规则替换输入token，然后将替换后的token传递给后继词典进行处理。
- 配置词典列表的一般规则是，第一个位置放置一个应用范围最小的、最具体化定义的词典，其次是更一般化定义的词典， 最后是一个普适定义的词典，比如Snowball词干词典或Simple词典。在下面例子中，对于一个针对天文学的文本搜索配置astro_en，可以定义标记类型asciiword（ASCII词）对应的词典列表为：天文术语的Synonym同义词词典， Ispell英语词典和Snowball英语词干词典。

  ```
  ALTER TEXT SEARCH CONFIGURATION astro_en
    ADD MAPPING FOR asciiword WITH astro_syn, english_ispell, english_stem;
  ```

  过滤类型的词典可以放置在词典列表中除去末尾的任何地方，放置在末尾时是无效的。使用这些词典对标记进行部分规范化，可以有效简化后继词典的处理。

 版权所有 © 华为云计算技术有限公司

10.6.2 停用词

停用词是很常见的词，几乎出现在每一个文档中，并且没有区分值。因此，在全文搜索的语境下可忽视它们。停用词处理逻辑和词典类型相关。例如，Ispell词典会先对标记进行规范化，然后再查看停用词表，而Snowball词典会最先检查输入标记是否为停用词。

例如，每个英文文本包含像a和the的单词，因此没必要将它们存储在索引中。然而，停用词影响tsvector中的位置，同时位置也会影响相关度：

```
SELECT to_tsvector('english','in the list of stop words');
        to_tsvector
----------------------------
 'list':3 'stop':5 'word':6
```

位置1、2、4是停用词，所以不显示。为包含和不包含停用词的文档计算出的排序是完全不同的：

```
SELECT ts_rank_cd (to_tsvector('english','in the list of stop words'), to_tsquery('list & stop'));
 ts_rank_cd
------------
        .05

SELECT ts_rank_cd (to_tsvector('english','list stop words'), to_tsquery('list & stop'));
 ts_rank_cd
------------
         .1
```

10.6.3 Simple 词典

Simple词典首先将输入标记转换为小写字母，然后检查停用词表。如果识别为停用词则返回空数组，即表示该标记会被丢弃。否则，输入标记的小写形式作为规范化后的lexeme返回。此外，Simple词典可通过设置参数Accept为false（默认值true），将非停用词报告为未识别，传递给后继词典继续处理。

注意事项

- 大多数词典的功能依赖于词典定义文件，词典定义文件名仅支持小写字母、数字、下划线组合。
- 临时模式pg_temp下不允许创建词典。
- 词典定义文件的字符集编码必须为UTF-8格式。实际应用时，如果与数据库的字符编码格式不一致，在读入词典定义文件时会进行编码转换。
- 通常情况下，每个session仅读取词典定义文件一次，当且仅当在第一次使用该词典时。需要修改词典文件时，可通过ALTER TEXT SEARCH DICTIONARY命令进行词典定义文件的更新和重新加载。

操作步骤

步骤1 创建Simple词典。

```
CREATE TEXT SEARCH DICTIONARY public.simple_dict (
    TEMPLATE = pg_catalog.simple,
    STOPWORDS = english
);
```

其中，停用词表文件全名为english.stop。关于创建simple词典的语法和更多参数，请参见**CREATE TEXT SEARCH DICTIONARY**。

步骤2 使用Simple词典。

 版权所有 © 华为云计算技术有限公司

```
SELECT ts_lexize('public.simple_dict','Yes');
 ts_lexize
-----------
 {yes}
(1 row)

SELECT ts_lexize('public.simple_dict','The');
 ts_lexize
-----------
 {}
(1 row)
```

步骤3 设置参数ACCEPT=false，使Simple词典返回NULL，而不是返回非停用词的小写形式。

```
ALTER TEXT SEARCH DICTIONARY public.simple_dict ( Accept = false );
SELECT ts_lexize('public.simple_dict','Yes');
 ts_lexize
-----------

(1 row)

SELECT ts_lexize('public.simple_dict','The');
 ts_lexize
-----------
 {}
(1 row)
```

----结束

10.6.4 Synonym 词典

Synonym词典用于定义、识别token的同义词并转化，不支持词组（词组形式的同义词可用Thesaurus词典定义，详细请参见**Thesaurus词典**）。

示例

- Synonym词典可用于解决语言学相关问题，例如，为避免使单词"Paris"变成"pari"，可在Synonym词典文件中定义一行"Paris paris"，并将该词典放置在预定义的english_stem词典之前。

```
SELECT * FROM ts_debug('english', 'Paris');
   alias   |   description   | token |  dictionaries  |  dictionary  | lexemes
-----------+-----------------+-------+----------------+--------------+---------
 asciiword | Word, all ASCII | Paris | {english_stem} | english_stem | {pari}
(1 row)

CREATE TEXT SEARCH DICTIONARY my_synonym (
    TEMPLATE = synonym,
    SYNONYMS = my_synonyms,
    FILEPATH =  'obs://bucket01/obs.example.com accesskey=xxxxx secretkey=xxxxx region=ap-southeast-1'
);

ALTER TEXT SEARCH CONFIGURATION english
    ALTER MAPPING FOR asciiword
    WITH my_synonym, english_stem;

SELECT * FROM ts_debug('english', 'Paris');
   alias   |   description   | token |       dictionaries        | dictionary | lexemes
-----------+-----------------+-------+---------------------------+------------+---------
 asciiword | Word, all ASCII | Paris | {my_synonym,english_stem} | my_synonym | {paris}
(1 row)

SELECT * FROM ts_debug('english', 'paris');
   alias   |   description   | token |       dictionaries        | dictionary | lexemes
-----------+-----------------+-------+---------------------------+------------+---------
```

 版权所有 © 华为云计算技术有限公司

```
 asciiword | Word, all ASCII | Paris | {my_synonym,english_stem} | my_synonym | {paris}
(1 row)

ALTER TEXT SEARCH DICTIONARY my_synonym ( CASESENSITIVE=true);

SELECT * FROM ts_debug('english', 'Paris');
   alias   |   description   | token |       dictionaries        | dictionary | lexemes
-----------+-----------------+-------+---------------------------+------------+---------
 asciiword | Word, all ASCII | Paris | {my_synonym,english_stem} | my_synonym | {paris}
(1 row)

SELECT * FROM ts_debug('english', 'paris');
   alias   |   description   | token |       dictionaries        | dictionary | lexemes
-----------+-----------------+-------+---------------------------+------------+---------
 asciiword | Word, all ASCII | Paris | {my_synonym,english_stem} | my_synonym | {pari}
(1 row)
```

其中，同义词词典文件全名为my_synonyms.syn，所在目录为 'obs://bucket01/obs.example.com accesskey=xxxxx secretkey=xxxxx region=*ap-southeast-1*'。关于创建词典的语法和更多参数，请参见**CREATE TEXT SEARCH DICTIONARY**。

- 星号（*）可用于词典文件中的同义词结尾，表示该同义词是一个前缀。在to_tsvector()中该星号将被忽略，但在to_tsquery()中会匹配该前缀并对应输出结果（参照**处理查询**一节）。

  假设词典文件synonym_sample.syn内容如下：

  ```
  postgres        pgsql
  postgresql      pgsql
  postgre pgsql
  gogle   googl
  indices index*
  ```

  创建并使用词典：

  ```
  CREATE TEXT SEARCH DICTIONARY syn (
      TEMPLATE = synonym,
      SYNONYMS = synonym_sample
  );

  SELECT ts_lexize('syn','indices');
   ts_lexize
  -----------
   {index}
  (1 row)

  CREATE TEXT SEARCH CONFIGURATION tst (copy=simple);

  ALTER TEXT SEARCH CONFIGURATION tst ALTER MAPPING FOR asciiword WITH syn;

  SELECT to_tsvector('tst','indices');
   to_tsvector
  -------------
   'index':1
  (1 row)

  SELECT to_tsquery('tst','indices');
   to_tsquery
  ------------
   'index':*
  (1 row)

  SELECT 'indexes are very useful'::tsvector;
              tsvector
  ---------------------------------
   'are' 'indexes' 'useful' 'very'
  (1 row)

  SELECT 'indexes are very useful'::tsvector @@ to_tsquery('tst','indices');
   ?column?
  ```

 版权所有 © 华为云计算技术有限公司

```
----------
 t
(1 row)
```

10.6.5 Thesaurus 词典

Thesaurus词典，也叫做分类词典（缩写为TZ），是一组定义了词以及词组间关系的集合，包括广义词（BT）、狭义词（NT）、首选词、非首选词、相关词等。根据词典文件中的定义，TZ词典用一个指定的短语替换对应匹配的所有短语，并且可选择保留原始短语进行索引。TZ词典实际上是Synonym词典的一个扩展，增加了短语支持。

注意事项

- 由于TZ词典需要识别短语，所以在处理过程中必须保存当前状态并与解析器进行交互，以决定是否处理下一个token或是结束当前识别。此外，TZ词典配置时需谨慎，如果设置TZ词典仅处理asciiword类型的token，则类似one 7的分类词典定义将不会生效，因为uint类型的token不会传给TZ词典处理。
- 在索引期间要用到分类词典，因此分类词典参数中的任何变化都要求重新索引。对于其他大多数类型的词典来说，类似添加或删除停用词这种修改并不需要强制重新索引。

操作步骤

步骤1 创建一个名为thesaurus_astro的TZ词典。

以一个简单的天文学词典thesaurus_astro为例，其中定义了两组天文短语及其同义词如下：

```
supernovae stars : sn
crab nebulae : crab
```

执行如下语句创建TZ词典：

```
CREATE TEXT SEARCH DICTIONARY thesaurus_astro (
    TEMPLATE = thesaurus,
    DictFile = thesaurus_astro,
    Dictionary = pg_catalog.english_stem,
    FILEPATH = 'obs://bucket_name/path accesskey=ak secretkey=sk region=rg'
);
```

其中，词典定义文件全名为thesaurus_astro.ths，所在目录为 "obs://bucket_name/path accesskey=ak secretkey=sk region=rg"。子词典pg_catalog.english_stem是预定义的Snowball类型的英语词干词典，用于规范化输入词，子词典自身相关配置（例如停用词等）不在此处显示。关于创建词典的语法和更多参数，请参见**CREATE TEXT SEARCH DICTIONARY**。

步骤2 创建词典后，将其绑定到对应文本搜索配置中需要处理的token类型上：

```
ALTER TEXT SEARCH CONFIGURATION english
    ALTER MAPPING FOR asciiword, asciihword, hword_asciipart
    WITH thesaurus_astro, english_stem;
```

步骤3 使用TZ词典。

- 测试TZ词典。

  ts_lexize函数对于测试TZ词典作用不大，因为该函数是按照单个token处理输入。可以使用plainto_tsquery、to_tsvector、to_tsquery函数测试TZ词典，这些函数能够将输入分解成多个token（to_tsquery函数需要将输入加上引号）。

  ```
  SELECT plainto_tsquery('english','supernova star');
   plainto_tsquery
  -----------------
  ```

 版权所有 © 华为云计算技术有限公司

```
 'sn'
(1 row)

SELECT to_tsvector('english','supernova star');
 to_tsvector
-------------
 'sn':1
(1 row)

SELECT to_tsquery('english','''supernova star''');
 to_tsquery
------------
 'sn'
(1 row)
```

其中，supernova star匹配了词典thesaurus_astro定义中的supernovae stars，这是因为在thesaurus_astro词典定义中指定了Snowball类型的子词典english_stem，该词典移除了e和s。

- 如果同时需要索引原始短语，只要将其同时放置在词典定义文件中对应定义的右侧即可，如下：

```
supernovae stars : sn supernovae stars

ALTER TEXT SEARCH DICTIONARY thesaurus_astro (
    DictFile = thesaurus_astro,
    FILEPATH = 'file:///home/dicts/');

SELECT plainto_tsquery('english','supernova star');
       plainto_tsquery
-----------------------------
 'sn' & 'supernova' & 'star'
(1 row)
```

----结束

10.6.6 Ispell 词典

Ispell词典模板支持词法词典，它可以把一个词的各种语言学形式规范化成相同的词位。比如，一个Ispell英语词典可以匹配搜索词bank的词尾变化和词形变化，如banking、banked、banks、banks'和bank's等。

DWS不提供任何预定义的Ispell类型词典或词典文件。dict文件和affix文件支持多种开源词典格式，包括Ispell、MySpell和Hunspell等。

操作步骤

步骤1 获取词典定义文件和词缀文件。

用户可以使用开源词典，直接获取的开源词典后缀名可能为.aff和.dic，此时需要将扩展名改为.affix和.dict。此外，对于某些词典文件，还需要使用下面的命令把字符转换成UTF-8编码，比如挪威语词典：

```
iconv -f ISO_8859-1 -t UTF-8 -o nn_no.affix nn_NO.aff
iconv -f ISO_8859-1 -t UTF-8 -o nn_no.dict nn_NO.dic
```

步骤2 创建Ispell词典。

```
CREATE TEXT SEARCH DICTIONARY norwegian_ispell (
    TEMPLATE = ispell,
    DictFile = nn_no,
    AffFile = nn_no,
    FilePath =  'obs://bucket01/obs.example.com accesskey=xxxxx secretkey=xxxxx region=ap-southeast-1'
);
```

 版权所有 © 华为云计算技术有限公司

其中，词典文件全名为nn_no.dict和nn_no.affix，所在目录为 'obs://bucket01/ obs.example.com accesskey=xxxxx secretkey=xxxxx region=*ap-southeast-1*'。关于创建词典的语法和更多参数，请参见**CREATE TEXT SEARCH DICTIONARY**。

步骤3 使用Ispell词典进行复合词拆分。

```
SELECT ts_lexize('norwegian_ispell', 'sjokoladefabrikk');
     ts_lexize
---------------------
 {sjokolade,fabrikk}
(1 row)
```

MySpell不支持复合词，Hunspell对复合词有较好的支持。DWS仅支持Hunspell中基本的复合词操作。通常情况下，Ispell词典能够识别的词是一个有限集合，其后应该配置一个更广义的词典，例如一个可以识别所有词的Snowball词典。

----结束

10.6.7 Snowball 词典

Snowball词典模板支持词干分析词典，基于Martin Porter的Snowball项目，内置有许多语言的词干分析算法。DWS中预定义有多种语言的Snowball词典，可通过系统表PG_TS_DICT查看预定义的词干分析词典以及支持的语言词干分析算法。

无论是否可以简化，Snowball词典将标识所有输入为已识别，因此它应当被放置在词典列表的最后。把Snowball词典放在任何其他词典前面会导致后继词典失效，因为输入token不会通过Snowball词典进入到下一个词典。

关于Snowball词典的语法，请参见**CREATE TEXT SEARCH DICTIONARY**。

10.7 配置示例

文本搜索配置（Text Search Configuration），指定了将文档转换成tsvector过程中所必需的组件：

- 解析器，用于把文本分解成标记token；
- 词典列表，用于将每个token转换成词位lexeme。

每次to_tsvector或to_tsquery函数调用时，都需要指定一个文本搜索配置来指定具体的处理过程。GUC参数default_text_search_config指定了默认的文本搜索配置，当文本搜索函数中没有显式指定文本搜索配置参数时，将会使用该默认值进行处理。

DWS中预定义有一些可用的文本搜索配置，用户也可创建自定义的文本搜索配置。此外，为了便于管理文本搜索对象，还提供有多个gsql元命令，可以显示有关文本搜索对象的信息（详细请参见《数据仓库服务工具指南》中“元命令参考”章节）。

操作步骤

步骤1 创建一个文本搜索配置ts_conf，复制预定义的文本搜索配置english。

```
CREATE TEXT SEARCH CONFIGURATION ts_conf ( COPY = pg_catalog.english );
CREATE TEXT SEARCH CONFIGURATION
```

步骤2 创建Synonym词典。

假设同义词词典定义文件pg_dict.syn内容如下：

```
postgres    pg
pgsql       pg
postgresql  pg
```

 版权所有 © 华为云计算技术有限公司

执行如下语句创建Synonym词典：

```
CREATE TEXT SEARCH DICTIONARY pg_dict (
    TEMPLATE = synonym,
    SYNONYMS = pg_dict,
    FILEPATH =  'obs://bucket01/obs.example.com accesskey=xxxxx secretkey=xxxxx region=ap-southeast-1'
);
```

步骤3 创建一个Ispell词典english_ispell（词典定义文件来自开源词典）。

```
CREATE TEXT SEARCH DICTIONARY english_ispell (
   TEMPLATE = ispell,
   DictFile = english,
   AffFile = english,
   StopWords = english,
   FILEPATH =  'obs://bucket01/obs.example.com accesskey=xxxxx secretkey=xxxxx region=ap-southeast-1'
);
```

步骤4 设置文本搜索配置ts_conf，修改某些类型的token对应的词典列表。关于token类型的详细信息，请参见**解析器**。

```
ALTER TEXT SEARCH CONFIGURATION ts_conf
   ALTER MAPPING FOR asciiword, asciihword, hword_asciipart,
                     word, hword, hword_part
   WITH pg_dict, english_ispell, english_stem;
```

步骤5 在文本搜索配置中，选择设置不索引或搜索某些token类型。

```
ALTER TEXT SEARCH CONFIGURATION ts_conf
   DROP MAPPING FOR email, url, url_path, sfloat, float;
```

步骤6 使用文本检索调测函数ts_debug()对所创建的词典配置ts_conf进行测试。

```
SELECT * FROM ts_debug('ts_conf', '
PostgreSQL, the highly scalable, SQL compliant, open source object-relational
database management system, is now undergoing beta testing of the next
version of our software.
');
```

步骤7 可以设置当前session使用ts_conf作为默认的文本搜索配置。此设置仅在当前session有效。

```
\dF+ ts_conf
     Text search configuration "public.ts_conf"
Parser: "pg_catalog.default"
     Token      |            Dictionaries
-----------------+-----------------------------------------
 asciihword      | pg_dict,english_ispell,english_stem
 asciiword       | pg_dict,english_ispell,english_stem
 file            | simple
 host            | simple
 hword           | pg_dict,english_ispell,english_stem
 hword_asciipart | pg_dict,english_ispell,english_stem
 hword_numpart   | simple
 hword_part      | pg_dict,english_ispell,english_stem
 int             | simple
 numhword        | simple
 numword         | simple
 uint            | simple
 version         | simple
 word            | pg_dict,english_ispell,english_stem

SET default_text_search_config = 'public.ts_conf';
SET
SHOW default_text_search_config;
 default_text_search_config
----------------------------
 public.ts_conf
(1 row)
```

----结束

 版权所有 © 华为云计算技术有限公司

10.8 测试和调试文本搜索

10.8.1 分词器测试

函数ts_debug允许简单测试文本搜索分词器。

```
ts_debug([ config regconfig, ] document text,
        OUT alias text,
        OUT description text,
        OUT token text,
        OUT dictionaries regdictionary[],
        OUT dictionary regdictionary,
        OUT lexemes text[])
        returns setof record
```

ts_debug显示document的每个token信息，token是由解析器生成，由指定的词典进行处理。如果忽略对应参数，则使用config指定的分词器或者default_text_search_config指定的分词器。

ts_debug为文本解析器标识的每个token返回一行记录。记录中的列分别是：

- alias：text类型，token的别名。
- description：text类型，token的描述。
- token：text类型，token的文本内容。
- dictionaries：regdictionary数组类型，是分词器为token选定的词典。
- dictionary：regdictionary类型，用来识别token的词典。如果为空，则不做识别。
- lexemes：text数组类型，词典识别token时生成的词素。如果为空，则不生成词素。空数组（{}）意味着token将被识别成停用词。

一个简单的例子：

```
SELECT * FROM ts_debug('english','a fat  cat sat on a mat - it ate a fat rats');
   alias   |   description   | token |  dictionaries  |  dictionary  | lexemes
-----------+-----------------+-------+----------------+--------------+---------
 asciiword | Word, all ASCII | a     | {english_stem} | english_stem | {}
 blank     | Space symbols   |       | {}             |              |
 asciiword | Word, all ASCII | fat   | {english_stem} | english_stem | {fat}
 blank     | Space symbols   |       | {}             |              |
 asciiword | Word, all ASCII | cat   | {english_stem} | english_stem | {cat}
 blank     | Space symbols   |       | {}             |              |
 asciiword | Word, all ASCII | sat   | {english_stem} | english_stem | {sat}
 blank     | Space symbols   |       | {}             |              |
 asciiword | Word, all ASCII | on    | {english_stem} | english_stem | {}
 blank     | Space symbols   |       | {}             |              |
 asciiword | Word, all ASCII | a     | {english_stem} | english_stem | {}
 blank     | Space symbols   |       | {}             |              |
 asciiword | Word, all ASCII | mat   | {english_stem} | english_stem | {mat}
 blank     | Space symbols   |       | {}             |              |
 blank     | Space symbols   | -     | {}             |              |
 asciiword | Word, all ASCII | it    | {english_stem} | english_stem | {}
 blank     | Space symbols   |       | {}             |              |
 asciiword | Word, all ASCII | ate   | {english_stem} | english_stem | {ate}
 blank     | Space symbols   |       | {}             |              |
 asciiword | Word, all ASCII | a     | {english_stem} | english_stem | {}
 blank     | Space symbols   |       | {}             |              |
 asciiword | Word, all ASCII | fat   | {english_stem} | english_stem | {fat}
 blank     | Space symbols   |       | {}             |              |
 asciiword | Word, all ASCII | rats  | {english_stem} | english_stem | {rat}
(24 rows)
```


版权所有 © 华为云计算技术有限公司

10.8.2 解析器测试

函数ts_parse可以直接测试文本搜索解析器。

```
ts_parse(parser_name text, document text,
        OUT tokid integer, OUT token text) returns setof record
```

ts_parse解析指定的document并返回一系列的记录，一条记录代表一个解析生成的token。每条记录包括标识token类型的tokid，及token文本。比如：

```
SELECT * FROM ts_parse('default', '123 - a number');
 tokid | token
-------+--------
    22 | 123
    12 |
    12 | -
     1 | a
    12 |
     1 | number
(6 rows)
ts_token_type(parser_name text, OUT tokid integer,
             OUT alias text, OUT description text) returns setof record
```

ts_token_type返回一个表，这个表描述了指定解析器可以识别的每种token类型。对于每个token类型，表中给出了整数类型的tokid--用于解析器标记对应的token类型；alias——命名分词器命令中的token类型；及简单描述。比如：

```
SELECT * FROM ts_token_type('default');
 tokid |      alias      |               description
-------+-----------------+------------------------------------------
     1 | asciiword       | Word, all ASCII
     2 | word            | Word, all letters
     3 | numword         | Word, letters and digits
     4 | email           | Email address
     5 | url             | URL
     6 | host            | Host
     7 | sfloat          | Scientific notation
     8 | version         | Version number
     9 | hword_numpart   | Hyphenated word part, letters and digits
    10 | hword_part      | Hyphenated word part, all letters
    11 | hword_asciipart | Hyphenated word part, all ASCII
    12 | blank           | Space symbols
    13 | tag             | XML tag
    14 | protocol        | Protocol head
    15 | numhword        | Hyphenated word, letters and digits
    16 | asciihword      | Hyphenated word, all ASCII
    17 | hword           | Hyphenated word, all letters
    18 | url_path        | URL path
    19 | file            | File or path name
    20 | float           | Decimal notation
    21 | int             | Signed integer
    22 | uint            | Unsigned integer
    23 | entity          | XML entity
(23 rows)
```

10.8.3 词典测试

函数ts_lexize用于进行词典测试。

ts_lexize(dict regdictionary, token text) returns text[]如果输入的token可以被词典识别，那么ts_lexize返回词素的数组；如果token可以被词典识别但它是一个停用词，则返回空数组；如果是一个不可识别的词则返回NULL。

比如：

```
SELECT ts_lexize('english_stem', 'stars');
 ts_lexize
```

 版权所有 © 华为云计算技术有限公司

```
-----------
 {star}

SELECT ts_lexize('english_stem', 'a');
 ts_lexize
-----------
 {}
```

须知

ts_lexize函数支持单一token，不支持文本。

10.9 限制约束

DWS的全文检索功能当前限制约束是：

- 每个分词长度必须小于2K字节。
- tsvector结构（分词+位置）的长度必须小于1兆字节。
- tsvector的位置值必须大于0，小于等于16,383。
- 每个分词在文档中位置数必须小于256，若超过将舍弃后面的位置信息。
- tsquery中的关键字及对应运算符最大支持到32,768。

 版权所有 © 华为云计算技术有限公司

11 系统操作

DWS通过SQL语句执行不同的系统操作，比如：设置变量，显示执行计划和垃圾收集等操作。

设置变量

设置会话或事务中需要使用的各种参数，请参考**SET**。

显示执行计划

显示DWS为SQL语句规划的执行计划，请参考**EXPLAIN**。

事务日志检查点

预写式日志（WAL）缺省时在事务日志中每隔一段时间放置一个检查点。CHECKPOINT强制立即进行检查，而不是等到下一次调度时的检查点。请参考**CHECKPOINT**。

垃圾收集

进行垃圾收集以及可选择的对数据库进行分析。请参考**VACUUM**。

收集统计信息

收集与数据库中表内容相关的统计信息。请参考**ANALYZE | ANALYSE**。

设置当前事务的约束检查模式

设置当前事务里的约束检查的特性。请参考**SET CONSTRAINTS**。

版权所有 © 华为云计算技术有限公司

12 事务管理

DWS支持数据库事务ACID属性，提供了事务的读已提交隔离级别和可重复读隔离级别。

警告

- 客户端显式开启事务或手动关闭Autocommit场景，业务最后必须手动执行Commit提交事务。
- 针对执行时长超过30分钟的语句，建议进行优化。
- **尽量避免**出现执行时长超过2小时的业务，避免长事务、长持锁等影响。
- 更多开发设计规范参见**总体开发设计规范**。

事务的概念

- **事务**指一个操作，由多个步骤组成，要么全部成功，要么全部失败。
- **数据库事务**（TRANSACTION）是数据库管理系统执行过程中的一个逻辑单位，由一个有限的数据库操作序列构成（通常由事务开始与事务结束之间执行的全部数据库操作组成），这些操作要么全部执行，要么全部不执行，是一个不可分割的执行单位。

事务的作用

数据库事务的目的主要是：

- 为数据库操作序列提供了一个从失败中恢复到正常状态的方法，同时提供了数据库即使在异常状态下仍能保持一致性的方法。
- 当多个应用程序在并发访问数据库时，可以在这些应用程序之间提供一个隔离方法，以防止彼此的操作互相干扰。

事务的执行过程

当事务被提交给数据库管理系统（DBMS）后，DBMS需要确保该事务中的所有操作都成功完成，并且其结果被永久保存在数据库中。如果事务中有操作没有成功完成，则事务中的所有操作都需要回滚，回到事务执行前的状态；同时，该事务对数据库或者其他事务的执行无影响，所有的事务互不干扰和影响，好像在独立的环境中运行。

版权所有 © 华为云计算技术有限公司

事务的属性

事务具有以下四个标准属性，通常根据首字母缩写为ACID。

- Atomicity（原子性）：事务中的全部操作在数据库中是不可分割的，整个事务中的所有操作要么全部完成，要么全部失败，对于一个事务来说，不能只执行其中的一部分操作。

  比如： A给B转账，A扣除500 USD，B增加500 USD。整个事务的操作要么全部成功，要么全部失败，不能出现A扣钱，但是B不增加的情况。如果原子性不能保证，就会出现一致性问题。

- Consistency（一致性）：在事务开始之前和事务结束以后，数据库的完整性没有被破坏。这表示写入的数据必须完全符合所有的预设规则，这包含数据的精确度、串联性以及后续数据库可以自发性地完成预定的工作。

  比如：A给B转账，A扣除500 USD ，B增加500 USD，扣除的钱-500与增加的钱+500，相加应该为0。如从A账户转账500 USD到B账户，不管操作成功与否，A和B的存款总额是不变的。

- Isolation（隔离性）：一个事务的执行不能被其他事务干扰。即一个事务内部的操作及使用的数据对并发的其他事务是隔离的，并发执行的各个事务之间不能互相干扰。

  数据库允许多个并发事务同时对其数据进行读写和修改的能力，隔离性可以防止多个事务并发执行时由于交叉执行而导致数据的不一致。事务隔离分为不同级别，包括读未提交（read uncommitted）、读提交（read committed）、可重复读（repeatable read）和串行化（serializable）。

- Durability（持久性）：一旦事务提交，则其所做的修改就会永久保存到数据库中。即使系统故障，已经提交的修改数据也不会丢失。

表 12-1 ACID 用途

ACID	属性	用途
Atomicity	原子性	并发控制，故障恢复。
Consistency	一致性	SQL的完整性约束（主键约束、外键约束）。
Isolation	隔离性	并发控制。
Durability	持久性	故障恢复。

常用的并发控制技术有基于锁的并发控制和基于时间戳的并发控制，DWS数据库针对DDL语句采用两阶段锁技术，而针对DML语句则采用多版本控制技术（Multi-Version Concurrency Control，MVCC）。DWS数据库的故障恢复采用WAL日志的方式来实现，目前主要支持Redo日志，通过Redo日志和MVCC可以保证事务读写的一致性。

Redo日志（重做日志）是数据库系统中用于保证数据持久性和故障恢复的核心机制之一，是数据库中用于故障恢复的关键日志文件，它记录了数据库事务对数据所做的“修改操作本身”（而非修改前后的数据），目的是在数据库发生故障（如宕机、崩溃）后，能够重新执行这些操作，将数据恢复到故障发生前的一致状态。通过Redo日志恢复过程可参见Redo恢复过程。

Redo日志的恢复过程：数据库崩溃重启后，通过读取预写日志记录(WAL日志)，从最近一个检查点(Checkpoint)（类似“存档点”）开始，扫描所有已提交但未持久化的

 版权所有 © 华为云计算技术有限公司

操作日志，重新执行这些操作（即“前滚（Redo）”），将内存丢失的数据变更重新写入磁盘；同时，对于未提交的事务，则通过逆向操作回滚其修改（即“后滚（Undo）”），最终确保数据恢复到崩溃前一致且完整的状态，既保证已提交操作的持久性，又维持未提交操作的原子性。这一机制是数据库实现高可靠性的核心基础。

隔离级别

Isolation（隔离性）可以防止多个事务并发执行时由于交叉执行而导致数据的不一致。事务隔离级别，决定多个事务并发操作同一个对象时的处理方式。

DWS的事务隔离级别，由GUC参数transaction_isolation或**SET TRANSACTION**语法设置，支持以下隔离级别，默认为读已提交（read committed）。

- read committed：读已提交隔离级别，只能读到已经提交的数据，而不会读到未提交的数据。
- read uncommitted：读未提交隔离级别，DWS不支持read uncommitted，如果设置了read uncommitted，实际上使用的是read committed。
- repeatable read：可重复读隔离级别，仅仅能看到事务开始之前提交的数据，不能看到未提交的数据，以及在事务执行期间由其他并发事务提交的修改。
- serializable：事务可序列化，DWS不支持serializable，如果设置了serializable，实际上使用的是repeatable read。

事务控制语法

- 启动事务

  DWS通过START TRANSACTION和BEGIN语法启动事务，请参考**START TRANSACTION**和**BEGIN**。

- 设置事务

  DWS通过SET TRANSACTION或者SET LOCAL TRANSACTION语法设置事务，请参考**SET TRANSACTION**。

- 提交事务

  DWS通过COMMIT或者END可完成提交事务的功能，即提交事务的所有操作，请参考**COMMIT | END**。

- 回滚事务

  回滚是在事务运行的过程中发生了某种故障，事务不能继续执行，系统将事务中对数据库的所有已完成的操作全部撤销。请参考**ROLLBACK**。

  说明

  数据库中收到的一次执行请求（不在事务块中），如果含有多条语句，将会被打包成一个事务，如果其中有一个语句失败，那么整个请求都将会被回滚。

- 其他事务操作
  - SAVEPOINT用于在当前事务里建立一个新的保存点。即在一个事务中标记一个位置并且允许做部分回滚。用户可以回滚在一个保存点之后执行的命令但保留该保存点之前执行的命令。请参考**SAVEPOINT**。
  - ROLLBACK TO SAVEPOINT回滚事务到一个保存点。隐含地删除所有在该保存点之后建立的保存点。请参考**ROLLBACK TO SAVEPOINT**。
  - RELEASE SAVEPOINT删除一个事务内的保存点。请参考**RELEASE SAVEPOINT**。

 版权所有 © 华为云计算技术有限公司

事务场景示例

某顾客在商店使用电子支付购买100 USD的物品，当中至少包括两个操作：1. 该顾客的账户减少100 USD。2. 商店账户（商户）增加100 USD。支持事务的数据库管理系统就是要确保以上两个操作（整个“事务”）都能完成，或一起取消。

1. 创建样例数据：

   创建一个简单的用户金额表并向表中插入数据（假设商户和顾客的账户上各有500 USD）。

   ```
   CREATE TABLE customer_info (
      NAME VARCHAR(32) PRIMARY KEY,
      MONEY INTEGER
   );
   INSERT INTO customer_info (name, money) VALUES ('buyer', 500), ('shop', 500);
   ```

   查看表数据显示商户和顾客各有500 USD。

   ```
   SELECT * FROM customer_info;
   ```

   ```
    name  | money
   -------+-------
    buyer |   500
    shop  |   500
   (2 rows)
   ```

2. 普通操作（正常模式）。

   模拟正常购买过程，顾客先扣款100 USD，商户再增加款额100 USD。

   ```
   UPDATE customer_info SET money = money-100 WHERE name IN (SELECT name FROM
   customer_info WHERE name = 'buyer');
   UPDATE customer_info SET money = money+100 WHERE name IN (SELECT name FROM
   customer_info WHERE name = 'shop');

   SELECT * FROM customer_info;
   ```

   ```
    name  | money
   -------+-------
    buyer |   400
    shop  |   600
   (2 rows)
   ```

3. 恢复初始值。

   ```
   UPDATE customer_info SET money=500;
   select * from customer_info;
   ```

   ```
    name  | money
   -------+-------
    shop  |   500
    buyer |   500
   (2 rows)
   ```

4. 普通操作（异常模式）。

   模拟购买过程出现状况，顾客发生扣款100 USD，商户没有增加款额。

   a. 顾客先扣款100 USD。

      ```
      UPDATE customer_info SET money = money-100 WHERE name IN (SELECT name FROM
      customer_info WHERE name = 'buyer');
      ```


版权所有 © 华为云计算技术有限公司

b. 商户发现支付有问题，终止了后续交易。商户增加款操作直接报错，终止执行下面的语句。（仅商户觉得支付有问题）

```
UPDATE customer_info SET money = money+100 WHERE name IN (SELECT name FROM customer_info WHERE name = 'shop');
```

c. 查询结果发现：消费者已经扣款，但商户没增加款额，这里顾客的金额少了100 USD。

```
SELECT * FROM customer_info;
```

```
 name  | money
-------+-------
 buyer |   400
 shop  |   500
(2 rows)
```

因此，如果没有事务，一旦SQL语句中间出现异常，整个账户系统的收支就不平衡了。

使用数据库事务，模拟出现异常操作时，进行事务回滚。

1. 恢复初始值：

```
UPDATE customer_info SET money=500;
```

2. 开启事务后，顾客先扣款100 USD。

```
BEGIN TRANSACTION;
UPDATE customer_info SET money = money-100 WHERE name IN (SELECT name FROM customer_info WHERE name = 'buyer');
```

3. 商户增加款额操作直接报错，终止执行下面的语句。

```
UPDATE customer_info SET money = money+100 WHERE name IN (SELECT name FROM customer_info WHERE name = 'shop');
```

4. 回滚事务，在事务运行的过程中发生了某种故障，事务不能继续执行，系统将事务中对数据库的所有已完成的操作全部撤销。

```
ERROR:  syntax error at or near "shop"
LINE 1: ...e IN (SELECT name FROM customer_info WHERE name = ''shop'');
```

```
END TRANSACTION;
ROLLBACK
```

5. 查询显示顾客和商户的账户金额仍旧完整一致。即数据库在事务执行过程中发生错误，会被恢复（Rollback）到事务开始前的状态，数据库的完整性没有被破坏。

```
SELECT * FROM customer_info;
```

```
 name  | money
-------+-------
 buyer |   500
 shop  |   500
(2 rows)
```

两阶段事务

DWS属于分布式share-nothing架构，表的数据分布在不同的节点上。客户端的一条或多条语句可能会同时修改多个节点上的数据，这种情况下，会产生分布式事务。DWS采用两阶段提交事务来保证分布式事务中数据的一致性和事务的原子性。顾名思义，两阶段提交就是将事务提交划分为两个阶段，通常针对的是包含写操作的事务。当写操作将数据写入不同的节点时，需要满足事务的原子性要求，要么全部提交，要么全部回滚。

 版权所有 © 华为云计算技术有限公司

不支持两阶段的场景如下：

- 不支持显式的两阶段提交语法PREPARE TRANSACTION。
```
BEGIN;
PREPARE TRANSACTION 'p1';
```
```
ERROR: Explicit prepare transaction is not supported.
```
- 不支持在两阶段事务中修改系统表的文件映射关系。
```
REINDEX TABLE pg_class;
```
```
ERROR: cannot PREPARE a transaction that modified relation mapping.
```
- 不支持在跨节点的事务中提交导出事务快照。
```
BEGIN;
CREATE TABLE t1(a int);
SELECT pg_export_snapshot();
END;
```
```
ERROR: cannot PREPARE a transaction that has exported snapshots.
```

 版权所有 © 华为云计算技术有限公司

13 DDL 语法

13.1 DDL 语法一览表

DDL（Data Definition Language数据定义语言），用于定义或修改数据库中的对象。如：表、索引、视图等。

说明

DWS不支持CN不完整的情况下执行DDL操作。例如：集群中存在1个CN故障时，执行新建数据库、表等操作都会失败。

定义数据库

数据库是组织、存储和管理数据的仓库，而数据库定义主要包括：创建数据库、修改数据库属性，以及删除数据库。所涉及的SQL语句，请参考下表。

表 13-1 数据库定义相关 SQL

功能	相关SQL
创建数据库	CREATE DATABASE
修改数据库属性	ALTER DATABASE
删除数据库	DROP DATABASE

定义模式

模式是一组数据库对象的集合，主要用于控制对数据库对象的访问。所涉及的SQL语句，请参考下表。

表 13-2 模式定义相关 SQL

功能	相关SQL
创建模式	CREATE SCHEMA

版权所有 © 华为云计算技术有限公司

功能	相关SQL
修改模式属性	ALTER SCHEMA
删除模式	DROP SCHEMA

定义表

表是数据库中的一种特殊数据结构，用于存储数据对象以及对象之间的关系。所涉及的SQL语句，请参考下表。

表 13-3 表定义相关 SQL

功能	相关SQL
创建表	CREATE TABLE
修改表属性	ALTER TABLE
修改表名	RENAME TABLE
删除表	DROP TABLE
删除表的所有数据	TRUNCATE

定义分区表

分区表是一种逻辑表，数据是由普通表存储的，主要用于提升查询性能。所涉及的SQL语句，请参考下表。

表 13-4 分区表定义相关 SQL

功能	相关SQL
创建分区表	CREATE TABLE PARTITION
修改分区	ALTER TABLE PARTITION
修改分区表属性	ALTER TABLE PARTITION
删除分区	ALTER TABLE PARTITION
删除分区表	DROP TABLE

定义索引

索引是对数据库表中一列或多列的值进行排序的一种结构，使用索引可快速访问数据库表中的特定信息。所涉及的SQL语句，请参考下表。

版权所有 © 华为云计算技术有限公司

表 13-5 索引定义相关 SQL

功能	相关SQL
创建索引	CREATE INDEX
修改索引属性	ALTER INDEX
删除索引	DROP INDEX
重建索引	REINDEX

定义角色

角色是用来管理权限的，从数据库安全的角度考虑，可以把所有的管理和操作权限划分到不同的角色上。所涉及的SQL语句，请参考下表。

表 13-6 角色定义相关 SQL

功能	相关SQL
创建角色	CREATE ROLE
修改角色属性	ALTER ROLE
删除角色	DROP ROLE

定义用户

用户是用来登录数据库的，通过对用户赋予不同的权限，可以方便地管理用户对数据库的访问及操作。所涉及的SQL语句，请参考下表。

表 13-7 用户定义相关 SQL

功能	相关SQL
创建用户	CREATE USER
修改用户属性	ALTER USER
删除用户	DROP USER

定义脱敏策略

数据脱敏策略是指对某些敏感信息通过脱敏规则进行数据的变形，实现敏感隐私数据的可靠保护。用户可以在指定表对象创建脱敏策略，并限定策略生效范围，也可以新增、修改、删除脱敏列信息。所涉及的SQL语句，请参考下表。

版权所有 © 华为云计算技术有限公司

表 13-8 脱敏策略定义相关 SQL

功能	相关SQL
对表创建数据脱敏策略	CREATE REDACTION POLICY
修改应用在指定表的脱敏策略	ALTER REDACTION POLICY
删除应用在指定表的脱敏策略	DROP REDACTION POLICY

定义行级访问控制

行级访问控制策略控制数据库表中行级数据可见性。不同用户执行相同的SQL查询操作，读取到的结果不同。所涉及的SQL语句，请参考下表。

表 13-9 行级访问控制定义相关 SQL

功能	相关SQL
创建行访问控制策略	CREATE ROW LEVEL SECURITY POLICY
修改已存在的行访问控制策略	ALTER ROW LEVEL SECURITY POLICY
删除表上某个行访问控制策略	DROP ROW LEVEL SECURITY POLICY

定义存储过程

存储过程是一组为了完成特定功能的SQL语句集，经编译后存储在数据库中，用户通过指定存储过程的名字并给出参数（如果该存储过程带有参数）来执行它。所涉及的SQL语句，请参考下表。

表 13-10 存储过程定义相关 SQL

功能	相关SQL
创建存储过程	CREATE PROCEDURE
删除存储过程	DROP PROCEDURE

定义函数

DWS的函数是一段可重用的代码块，它接受输入参数，执行特定操作，并返回一个值或结果集。所涉及的SQL语句，请参考下表。

版权所有 © 华为云计算技术有限公司

表 13-11 函数定义相关 SQL

功能	相关SQL
创建函数	CREATE FUNCTION
修改函数属性	ALTER FUNCTION
删除函数	DROP FUNCTION

定义视图

视图是基于SQL查询定义的逻辑虚拟表，仅保存视图定义而不物理存储数据。当基表数据发生变化时，视图查询结果自动同步更新，视图提供动态数据窗口，帮助用户高效聚焦目标数据。所涉及的SQL语句，请参考下表。

表 13-12 视图定义相关 SQL

功能	相关SQL
创建视图	CREATE VIEW
修改视图	ALTER VIEW
删除视图	DROP VIEW

定义游标

为了处理SQL语句，存储过程进程分配一段内存区域来保存上下文联系。游标是指向上下文区域的句柄或指针。借助游标，存储过程可以控制上下文区域的变化，请参考下表。

表 13-13 游标定义相关 SQL

功能	相关SQL
创建游标	CURSOR
移动游标	MOVE
从游标中提取数据	FETCH
关闭游标	CLOSE

操作会话

用户与数据库之间建立的连接称为会话，请参考下表。

 版权所有 © 华为云计算技术有限公司

表 13-14 会话相关 SQL

功能	相关SQL
修改会话	ALTER SESSION
结束会话	ALTER SYSTEM KILL SESSION

定义资源池

资源池是负载管理模块使用的系统表，主要用于指定资源管理相关的属性，如控制组。所涉及的SQL语句，请参考下表。

表 13-15 资源池定义相关 SQL

功能	相关SQL
创建资源池	CREATE RESOURCE POOL
修改资源池属性	ALTER RESOURCE POOL
删除资源池	DROP RESOURCE POOL

定义同义词

同义词是兼容Oracle的一种特殊的数据库对象，用于存储与另一个数据库对象名间的映射关系，目前仅支持使用同义词关联以下数据库对象：表、视图、函数和存储过程。所涉及的SQL语句，请参考下表。

表 13-16 同义词定义相关 SQL

功能	相关SQL
创建同义词	CREATE SYNONYM
修改同义词	ALTER SYNONYM
删除同义词	DROP SYNONYM

定义文本搜索配置

文本搜索配置指定了文本搜索解析器，该文本搜索解析器可以将字符串划分为标记，外加一些词典（可被用来决定哪些标记是搜索感兴趣的）。所涉及的SQL语句，请参考下表。

版权所有 © 华为云计算技术有限公司

表 13-17 文本搜索配置相关 SQL

功能	相关SQL
创建文本搜索配置	CREATE TEXT SEARCH CONFIGURATION
修改文本搜索配置	ALTER TEXT SEARCH CONFIGURATION
删除文本搜索配置	DROP TEXT SEARCH CONFIGURATION

定义全文检索词典

词典是在全文检索时识别特定词并进行处理。词典的创建依赖于预定义模板（在系统表PG_TS_TEMPLATE中定义），支持创建五种类型的词典，分别是Simple、Ispell、Synonym、Thesaurus、以及Snowball，每种类型的词典可以完成不同的任务。所涉及的SQL语句，请参考下表。

表 13-18 全文检索词典相关 SQL

功能	相关SQL
创建全文检索词典	CREATE TEXT SEARCH DICTIONARY
修改全文检索词典	ALTER TEXT SEARCH DICTIONARY
删除全文检索词典	DROP TEXT SEARCH DICTIONARY

定义查询过滤规则

对于提前识别到的慢SQL语句，用户可以通过创建查询过滤器规则来进行拦截。所涉及的SQL语句，请参考下表。

表 13-19 查询过滤规则相关 SQL

功能	相关SQL
创建查询过滤规则	CREATE BLOCK RULE
修改查询过滤规则	ALTER BLOCK RULE
删除查询过滤规则	DROP BLOCK RULE

13.2 数据库&模式&数据类型

版权所有 © 华为云计算技术有限公司

13.2.1 CREATE DATABASE

功能描述

创建新的数据库。

DWS采用从模板数据库复制的方式来创建新的数据库。DWS系统提供了两个模板数据库（template0和template1）以及一个默认的用户数据库（gaussdb）。默认情况下，新的数据库基于template1创建。template1不允许修改字符编码，因此若要变更编码，用户可以通过指定**template**参数选择template0模板。

注意事项

- 只有拥有CREATEDB权限的用户才可以创建新数据库，系统管理员默认拥有此权限。
- CREATE DATABASE不能在事务块中执行。
- 在创建数据库过程中，若出现类似“could not initialize database directory”的错误提示，可能是由于文件系统上数据目录的权限不足或磁盘满等原因引起。

语法格式

```
CREATE DATABASE database_name
   [ [ WITH ] { [ OWNER [=] user_name ] |
            [ TEMPLATE [=] template ] |
            [ ENCODING [=] encoding ] |
            [ LC_COLLATE [=] lc_collate ] |
            [ LC_CTYPE [=] lc_ctype ] |
            [ DBCOMPATIBILITY [=] compatibility_type ] |
            [ DBCOMPATIBILITY_BEHAVIOR [=] opt_compat_behavior] |

            [ CONNECTION LIMIT [=] connlimit ]}[...] ];
```

参数说明

表 13-20 CREATE DATABASE 参数说明

参数	描述	取值范围
database_name	要创建的数据库名称。	字符串，需符合**标识符命名规范**。
user_name	指定新数据库所有者的名称。缺省时，新数据库的所有者是当前用户。	已存在的用户名。

版权所有 © 华为云计算技术有限公司

参数	描述	取值范围
template	创建新数据库时指定的模板名称，即使用哪个模板创建新数据库。不指定时，系统默认使用template1作为模板。	template0和template1。 • template0，如果需要指定字符集编码（encoding），需指定为template0。 • template1，缺省值。模板template1不允许指定字符集编码（encoding）。
encoding	指定数据库使用的字符集编码，可以是字符串（如'SQL_ASCII'）、整数编号。不指定时，默认使用模板数据库的编码。 • 模板数据库template0和template1的编码默认与操作系统环境相关。 • template1不允许指定字符编码。 • 创建数据库时如果需要指定encoding，需同template参数搭配使用，且template取值为template0。 **说明** SQL_ASCII编码不校验入库数据编码，若数据库存入非ASCII字符数据时，可能会导致GaussRoach.py等服务端工具异常。	常用取值为GBK、UTF8、Latin1、SQL_ASCII或GB18030等。 如需支持**GB18030、GB18030-2022字符集**，encoding参数需指定为UTF8或者GB18030，且client_encoding为**UTF8**字符编码。
lc_collate	指定新数据库使用的字符集。例如，通过lc_collate = 'zh_CN.gbk'设定该参数。 该参数的使用会影响到对字符串的排序顺序（例如使用ORDER BY的查询中，以及在文本列上使用索引的顺序）。默认使用模板数据库的排序顺序。有关其他限制请参考**字符集编码的说明**。	有效的排序类型。
lc_ctype	指定新数据库使用的字符分类。例如，通过lc_ctype = 'zh_CN.gbk'设定该参数。 该参数的使用会影响到字符的分类，如大写、小写和数字。默认是使用模板数据库的字符分类。	有效的字符分类。

版权所有 © 华为云计算技术有限公司

参数	描述	取值范围
compatibility_type	指定兼容的数据库的类型。若不指定该参数，默认为ORA。	ORA、TD或MySQL。 分别表示兼容Oracle、Teradata或MySQL数据库。
opt_compat_behavior	指定数据库的兼容性行为。该参数仅9.1.0及以上集群版本支持。若不指定该参数，默认为NO_BEHAVIOR，表示无特殊行为。	td_rtrim或pg_char。 ● td_rtrim，表示TD兼容模式下变长字符串去尾空。 ● pg_char，表示PG兼容模式下varchar类型转为nvarchar2类型。
connlimit	指定数据库的并发连接数。默认值为-1，表示没有限制。 ● 系统管理员不受此参数的限制。 ● 为保证集群正常使用，connection limit的最小值是集群中CN的数目。在集群执行ANALYZE时，其他CN节点会连接当前执行ANALYZE的CN节点来同步元数据。例如集群中有3个CN节点，那么connection limit应该设置为>=3。	大于等于-1的整数。

字符集编码的说明

- 可使用“show server_encoding;”命令查看当前数据库的字符集编码。
- 为了适应全球化的需求，使数据库编码能够存储与表示绝大多数的字符，建议创建数据库时使用UTF8编码。
- 指定新的数据库字符集编码必须与所选择的语言环境中（LC_COLLATE和LC_CTYPE）的设置兼容。
- 若语言环境设置为C（或POSIX），则允许所有的编码类型，但是对于其他的语言环境设置，字符集编码必须和区域设置相同。
- 编码和区域设置必须匹配模板数据库，除了将template0当作模板。 因为其他数据库可能会包含不匹配指定编码的数据，或者可能包含排序顺序受LC_COLLATE和LC_CTYPE影响的索引。复制这些数据会导致在新数据库中的索引失效。template0是不包含任何会受到影响的数据或者索引。
- 支持的有效编码类型与当前所处的环境有关。若出现“invalid locale name”的字可通过locale -a命令检查环境所支持的字符集编码。
- 当指定的字符集编码为GBK时，部分中文生僻字无法直接作为对象名。这是因为GBK第二个字节的编码范围在0x40-0x7E之间时，字节编码与ASCII字符@A-Z[\]^_`a-z{|}重叠。其中@[\]^_'{|}是数据库中的操作符，直接作为对象名时，会语法报错。例如“俦”字，GBK16进制编码为0x8240，第二个字节为0x40，与

 版权所有 © 华为云计算技术有限公司

ASCII“@”符号编码相同，因此无法直接作为对象名使用。如果确实要使用，可以在创建和访问对象时，通过增加双引号来规避这个问题。

- 当前版本GBK字符集支持了欧元符'€'，十六进制表示为'0x80'，用户可以在GBK库中操作欧元符，也使得DWS的GBK字符集可以兼容CP936字符集。需注意GBK字符集约等于CP936字符集，但是GBK字符集中不包含欧元符的定义。
- 9.1.1.100及以上集群版本支持使用GB18030、GB18030-2022字符集创建数据库，需要注意的是只支持GB18030/GB18030-2022到UTF8之间的映射，所以必须设置client_encoding参数为UTF8。

示例

创建一个GBK编码的数据库music（本地环境的编码格式必须也为GBK）。

```
CREATE DATABASE music ENCODING 'GBK' template = template0;
```

创建数据库music2，并指定所有者为jim。

```
CREATE DATABASE music2 OWNER jim;
```

用模板template0创建数据库music3，并指定所有者为jim。

```
CREATE DATABASE music3 OWNER jim TEMPLATE template0;
```

创建兼容ORA格式的数据库，连接限制为10。

```
CREATE DATABASE ora_compatible_db DBCOMPATIBILITY 'ORA'CONNECTION LIMIT = 10;
```

创建兼容GB18030字符集的数据库db_gb18030（本地环境的编码格式必须也是GB18030）。

```
CREATE DATABASE db_gb18030 encoding='gb18030' LC_COLLATE='zh_CN.GB18030' LC_CTYPE
='zh_CN.GB18030' TEMPLATE=template0;
```

创建兼容GB18030-2022字符集的数据库db_gb180302022（本地环境的编码格式必须也是GB18030）。

```
CREATE DATABASE db_gb180302022 encoding='gb18030_2022' LC_COLLATE='zh_CN.GB18030' LC_CTYPE
='zh_CN.GB18030' TEMPLATE=template0;
```

清理环境。

```
DROP DATABASE music;
DROP DATABASE music2;
DROP DATABASE music3;
DROP DATABASE ora_compatible_db;
DROP DATABASE db_gb18030;
DROP DATABASE db_gb180302022;
```

相关链接

ALTER DATABASE，**DROP DATABASE**

常见问题

如何查看和设置DWS数据库的字符集编码格式

Oracle/TD/MySQL兼容模式下查询结果不一致

 版权所有 © 华为云计算技术有限公司

13.2.2 ALTER DATABASE

功能描述

修改数据库的属性，包括它的名称、所有者、连接数限制、对象隔离属性等。

注意事项

- 只有拥有数据库所有者权限的用户才能执行ALTER DATABASE命令，系统管理员默认拥有此权限。如果是非系统管理员，针对所要修改属性的不同，对其还有以下权限约束：
  - 修改数据库名称，必须拥有CREATEDB权限。
  - 修改数据库所有者，当前用户必须是该数据库的所有者，且拥有CREATEDB权限，并确保该用户是新所有者角色的成员。
  - 修改数据库默认表空间，该用户必须是该数据库的所有者或系统管理员，且拥有新表空间的CREATE权限。该语法会从物理上将一个数据库原来缺省表空间上的表和索引移至新的表空间。注意不在缺省表空间的表和索引不受此影响。
  - 修改某个按数据库设置的相关参数，只有数据库所有者或者系统管理员可以改变这些设置。
  - 修改某个数据库对象隔离属性，只有数据库所有者或者系统管理员可以执行此操作。
- 不能重命名当前使用的数据库，如果需要重新命名，必须连接至其他数据库上。
- 不支持修改现有数据库的兼容模式，只能在创建数据库时指定兼容模式，详情请参见**CREATE DATABASE**。

语法格式

- 对指定的数据库连接是否加锁，锁定后会阻止其他用户连接该数据库。该语法仅9.1.1.200及以上集群版本支持。

```
ALTER_DATABASE database_name
    [ [ WITH ] ALLOW_CONNECTIONS allow_connections ]
```

- 修改数据库的最大连接数。

```
ALTER DATABASE database_name
    [ [ WITH ] CONNECTION LIMIT connlimit ];
```

- 修改数据库名称。

```
ALTER DATABASE database_name
    RENAME TO new_name;
```

  说明

  若该数据库中有OBS冷热表，则不支持修改数据库名。

- 修改数据库所属者。

```
ALTER DATABASE database_name
    OWNER TO new_owner;
```

- 修改数据库指定会话参数值。

```
ALTER DATABASE database_name
    SET configuration_parameter { { TO | = } { value | DEFAULT } | FROM CURRENT };
```

- 数据库配置参数重置。

```
ALTER DATABASE database_name RESET
    { configuration_parameter | ALL };
```


版权所有 © 华为云计算技术有限公司

- 修改数据库对象隔离属性。
```
ALTER DATABASE database_name [ WITH ] { ENABLE | DISABLE } PRIVATE OBJECT;
```

  说明

  - 修改数据库的对象隔离属性时须连接至该数据库，否则无法更改。
  - 新创建的数据库，对象隔离属性默认是关闭的。当开启数据库对象隔离属性后，普通用户只能查看有权访问的对象（表、函数、视图、字段等）。对象隔离特性对管理员用户不生效，当开启对象隔离特性后，管理员也可以查看到全量的数据库对象。

参数说明

表 13-21 ALTER DATABASE 参数说明

参数	描述	取值范围
database_name	需要修改属性的数据库名称。	字符串，需符合**标识符命名规范**。
connlimit	数据库可以接收的最大并发连接数（管理员用户连接除外）。 生产环境中建议设置合理上限，防止过多连接耗尽资源。	整数，建议填写1~50。 -1（缺省）表示没有限制。
allow_connections	对指定的数据库连接是否加锁。锁定后不允许用户连接该数据库。	取值为true或false。 缺省值为true，表示不对数据库连接加锁。
new_name	数据库的新名称。	字符串，需符合**标识符命名规范**。
new_owner	数据库的新所有者。	字符串，有效的用户名。
configuration_parameter value	将指定的数据库会话参数值设置为给定值。	字符串： • DEFAULT，表示在新的会话中使用系统的缺省设置。 • RESET，表示在新的会话中使用系统的缺省设置。 • OFF，表示关闭设置。
FROM CURRENT	根据当前会话连接的数据库设置该参数的值。	-
RESET configuration_parameter	重置指定的数据库会话参数值。	-
RESET ALL	重置全部的数据库会话参数值。 修改的数据库会话参数值，将在下一次会话中生效。	-

版权所有 © 华为云计算技术有限公司

示例

创建数据库testdb和用户user10。

```
CREATE DATABASE testdb ENCODING 'UTF8' template = template0;
DROP USER IF EXISTS user10 CASCADE;
CREATE USER user10 PASSWORD '{Password}';
```

设置数据库testdb的连接数为10。

```
ALTER DATABASE testdb CONNECTION LIMIT= 10;
```

将testdb名称改为testdb_1。

```
ALTER DATABASE testdb RENAME TO testdb_1;
```

将数据库testdb_1的所属者修改为user10。

```
ALTER DATABASE testdb_1 OWNER TO user10;
```

关闭在数据库testdb_1上缺省的索引扫描。

```
ALTER DATABASE testdb_1 SET enable_indexscan TO off;
```

重置enable_indexscan参数。

```
ALTER DATABASE testdb_1 RESET enable_indexscan;
```

对数据库testdb_1连接进行加锁，不允许用户连接。

```
ALTER DATABASE testdb_1 allow_connections false;
```

清理环境。

```
CLEAN CONNECTION TO ALL FORCE FOR DATABASE testdb_1;
DROP DATABASE testdb_1;
DROP USER IF EXISTS user10 CASCADE;
```

相关链接

CREATE DATABASE，**DROP DATABASE**

13.2.3 DROP DATABASE

功能描述

删除数据库。

注意事项

- 只有数据库所有者有权限执行DROP DATABASE命令，系统管理员默认拥有此权限。
- 不能对系统默认安装的三个数据库（gaussdb、TEMPLATE0和TEMPLATE1）执行删除操作，系统做了保护。如果想查看当前服务中有哪几个数据库，可以用gsql的\l命令查看。
- 如果有用户正在与要删除的数据库连接，则删除操作失败。如果要查看当前存在哪些数据库连接，可以通过视图v$session查看。
- 不能在事务块中执行DROP DATABASE命令。
- 如果执行DROP DATABASE失败，事务回滚，需要再次执行一次DROP DATABASE IF EXISTS。

版权所有 © 华为云计算技术有限公司

- DROP DATABASE若提示database is being accessed by other users类错误，可能原因为CLEAN CONNECTION过程存在线程无法及时响应信号，出现连接清理不完全的情况，需要再次执行CLEAN CONNECTION。

危险

- **DROP对象**操作（如DATABASE、USER/ROLE、SCHEMA、TABLE、VIEW等对象）存在数据丢失风险，尤其含带CASCADE级联删除场景，会将关联的对象一并删除，操作需谨慎，操作前需考虑数据备份。
- 更多开发设计规范参见**总体开发设计规范**。

语法格式

```
DROP DATABASE [ IF EXISTS ] database_name ;
```

参数说明

表 13-22 参数说明

参数	描述	取值范围
IF EXISTS	如果指定的数据库不存在时返回一个提示而不是错误。	-
database_name	需要删除的数据库名称。	已存在的数据库名称。

示例

删除名称为music的数据库。如删除数据库失败，请参见**常见问题**处理。

```
DROP DATABASE IF EXISTS music;
```

相关链接

CREATE DATABASE，**ALTER DATABASE**

常见问题

当执行DROP DATABASE语法时出现以下类似报错，表示待删除数据库被其他用户连接。

```
database "mysql_compatible_db" is being accessed by other users Detail: There are 2 other sessions using the database.
```

可执行以下语法，先断开所有用户对该数据库建立的连接。

```
CLEAN CONNECTION TO ALL FORCE FOR DATABASE mysql_compatible_db;
```

再重新执行DROP DATABASE，可删除成功（如删除失败，可以重复执行以上CLEAN CONNECTION再重试）。

```
DROP DATABASE mysql_compatible_db;
```


版权所有 © 华为云计算技术有限公司

13.2.4 CREATE SCHEMA

功能描述

创建Schema。

Schema又称作模式，从逻辑上组织一个数据库中的对象和数据。通过管理Schema，允许多个用户使用同一数据库而不相互干扰，同时便于将第三方应用添加到相应的Schema下而不引起冲突。

Schema的核心作用是：

- **逻辑隔离**：实现数据、功能、租户的隔离。
- **权限控制**：提供细粒度的安全边界。
- **组织管理**：使数据库结构清晰可维护。
- **命名空间**：避免对象命名冲突。

相同的数据库对象名称可以应用在同一数据库的不同Schema中，而没有冲突。例如，a_schema和b_schema都可以包含名为mytable的表。具有所需权限的用户可以访问数据库的多个Schema中的对象。

在当前数据库中创建用户时，系统会在当前数据库中为新用户创建一个同名Schema，例如创建用户user1，默认会创建一个名称为user1的Schema。

注意事项

- 只要用户对当前数据库有CREATE权限，就可以创建模式。
- 系统管理员在普通用户同名Schema下创建的对象，所有者为Schema的同名用户（非系统管理员）。
- 访问命名对象时可以使用模式名作为前缀进行访问，如果无模式名前缀，则访问当前模式下的命名对象。创建命名对象时也可用模式名作为前缀修饰。
- CREATE SCHEMA可以包括在新模式中创建对象的子命令，这些子命令和那些在创建完模式后发出的命令没有任何区别。如果使用了AUTHORIZATION子句，则所有创建的对象都将被该用户所拥有。

语法格式

- 根据指定的名字创建模式：

```
CREATE SCHEMA schema_name
   [ AUTHORIZATION user_name ] [ WITH PERM SPACE 'space_limit'] [ schema_element [ ... ] ];
CREATE SCHEMA IF NOT EXISTS schema_name
   [ AUTHORIZATION user_name ] [ WITH PERM SPACE 'space_limit'] ;
```

- 根据用户名创建模式：

```
CREATE SCHEMA AUTHORIZATION user_name [ WITH PERM SPACE 'space_limit'] [ schema_element
[ ... ] ];
CREATE SCHEMA IF NOT EXISTS AUTHORIZATION user_name [ WITH PERM SPACE 'space_limit'] ;
```

 版权所有 © 华为云计算技术有限公司

参数说明

表 13-23 CREATE SCHEMA 参数说明

参数	描述	取值范围
schema_name	模式名称。	字符串，要符合标识符的命名规范。参见**标识符命名规范**。 **须知** • 模式名不能和当前数据库里其他的模式重名。 • 模式的名字不可以“pg_”开头。
AUTHORIZATION user_name	指定模式的所有者。当不指定schema_name时，把user_name当作模式名，此时user_name只能是角色名。	取值范围：已存在的用户名/角色名。
WITH PERM SPACE 'space_limit'	指定模式的永久表存储空间上限。当不指定space_limit时，则不限制。	取值范围：字符串格式为正整数+单位，单位当前支持K/M/G/T/P。解析后的数值以K为单位，且范围不能够超过64比特表示的有符号整数，即1KB~9007199254740991KB。 **示例**请参见**示例：创建一个设置了永久存储空间的Schema**。
schema_element	在模式里创建对象的SQL语句。目前仅支持CREATE TABLE、CREATE VIEW、CREATE INDEX、CREATE PARTITION、GRANT子句。 子命令所创建的对象都被AUTHORIZATION子句指定的用户所拥有。	-

版权所有 © 华为云计算技术有限公司

参数	描述	取值范围
IF NOT EXISTS	指定IF NOT EXISTS时，若不存在同名SCHEMA，则可以成功创建SCHEMA。若已存在同名SCHEMA，创建时不会报错，仅会提示该SCHEMA已存在，且不执行任何操作。当使用该参数时，不能包括Schema_element子命令。 **IF NOT EXISTS参数仅9.1.0及以上版本支持。** 如果当前搜索路径上的模式中存在同名对象时，需要明确指定引用对象所在的模式。可以通过命令SHOW SEARCH_PATH来查看当前搜索路径上的模式。	-

示例：基础示例

创建一个角色role1：

```
DROP ROLE IF EXISTS role1;
CREATE ROLE role1 IDENTIFIED BY '{password}';
```

为用户role1创建一个同名schema，子命令创建的表films和winners的拥有者为role1：

```
DROP TABLE IF EXISTS role1.films;
CREATE SCHEMA AUTHORIZATION role1
    CREATE TABLE films (title text, release date, awards text[])
    CREATE VIEW winners AS
    SELECT title, release FROM films WHERE awards IS NOT NULL;
```

示例：创建一个设置了永久存储空间的 Schema

以下模拟一个示例，创建一个名为testsche的schema，并设置永久存储空间为10MB，当数据超过该设置值，系统报错。

除了通过SQL语句设置永久存储空间之外，您也可以通过控制台界面进行Schema永久存储空间设置，详情参见**配置DWS数据库的Schema存储空间**章节。

步骤1 创建一个名为testsche的schema，并设置永久存储空间为10MB。

```
DROP SCHEMA IF EXISTS testsche;
CREATE SCHEMA testsche  WITH PERM SPACE '10M';
```

步骤2 使用系统表**PG_NAMESPACE**查询Schema的存储空间大小，其中permspace为永久存储空间限额，usedspace为已使用永久存储空间大小。

```
SELECT permspace,usedspace FROM PG_NAMESPACE WHERE nspname = 'testsche';
```

 版权所有 © 华为云计算技术有限公司

```
gaussdb=> CREATE SCHEMA testsche  WITH PERM SPACE '10M';
CREATE SCHEMA
gaussdb=> SELECT permspace,usedspace FROM PG_NAMESPACE WHERE nspname = 'testsche';
 permspace | usedspace
-----------+-----------
  10485760 |         0
(1 row)
```

步骤3 创建一张测试表，并导入数据。

```
DROP TABLE IF EXISTS testsche.src;
DROP TABLE IF EXISTS testsche.t1;
CREATE TABLE testsche.src AS SELECT 1;
CREATE TABLE testsche.t1(a int, b numeric(15,2)) WITH(orientation=column);
INSERT INTO testsche.t1 SELECT generate_series(1,20000000) % 1000,generate_series(1,20000000) FROM testsche.src;
```

回显报错信息，超过该schema的永久存储空间大小。

```
gaussdb=> CREATE TABLE testsche.src AS SELECT 1;
NOTICE:  The 'DISTRIBUTE BY' clause is not specified. Using round-robin as the distribution mode by default.
HINT:  Please use 'DISTRIBUTE BY' clause to specify suitable data distribution column.
INSERT 0 1
gaussdb=> CREATE TABLE testsche.t1(a int, b numeric(15,2)) WITH(orientation=column);
NOTICE:  The 'DISTRIBUTE BY' clause is not specified. Using round-robin as the distribution mode by default.
HINT:  Please use 'DISTRIBUTE BY' clause to specify suitable data distribution column.
CREATE TABLE
gaussdb=> INSERT INTO testsche.t1 SELECT generate_series(1,20000000) % 1000,generate_series(1,20000000) FROM testsche.src;
ERROR:  dn_6005_6006: out of testsche schema's perm space limit, usedspace: 15204352, needspace: 0, spacelimit: 10485760
gaussdb=>
```

步骤4 调大Schema的永久存储空间大小为10GB，再次导入，显示成功。

```
ALTER SCHEMA testsche WITH PERM SPACE '10G';
INSERT INTO testsche.t1 SELECT generate_series(1,20000000) % 1000,generate_series(1,20000000) FROM testsche.src;
```

```
gaussdb=> ALTER SCHEMA testsche WITH PERM SPACE '10G';
ALTER SCHEMA
gaussdb=> INSERT INTO testsche.t1 SELECT generate_series(1,20000000) % 1000,generate_series(1,20000000) FROM testsche.src;
INSERT 0 20000000
gaussdb=>
```

步骤5 查看已用空间。

```
SELECT permspace,usedspace FROM PG_NAMESPACE WHERE nspname = 'testsche';
```

```
gaussdb=> SELECT permspace,usedspace FROM PG_NAMESPACE WHERE nspname = 'testsche';
  permspace  | usedspace
-------------+-----------
 10737418240 |     16384
(1 row)
```

----结束

相关链接

ALTER SCHEMA，**DROP SCHEMA**

创建和管理schema

配置DWS数据库的Schema存储空间

13.2.5 ALTER SCHEMA

功能描述

修改模式属性，包括模式名称、所有者、永久表存储空间上限。

版权所有 © 华为云计算技术有限公司

注意事项

- 只有模式的所有者或者被授予了模式ALTER权限的用户，有权限执行ALTER SCHEMA命令，系统管理员默认拥有此权限。
- 非系统管理员要修改模式的所有者或模式的永久表存储空间上限，该用户还必须是新的所有角色的直接或间接成员， 并且该成员必须在此数据库上有CREATE权限。

语法格式

- 修改模式的名称。
```
ALTER SCHEMA schema_name
    RENAME TO new_name;
```
- 修改模式的所有者。
```
ALTER SCHEMA schema_name
    OWNER TO new_owner;
```
- 修改模式的永久表存储空间限制。
```
ALTER SCHEMA schema_name
    WITH PERM SPACE 'space_limit';
```

参数说明

表 13-24 ALTER SCHEMA 参数说明

参数	描述	取值范围
schema_name	需要修改的模式名称。	已存在的模式名。
new_name	模式的新名称。	字符串，需符合**标识符命名规范**。
new_owner	模式的新所有者。	已存在的用户名/角色名。
WITH PERM SPACE 'space_limit'	模式的永久表存储空间上限。	字符串格式为正整数+单位，单位当前支持K/M/G/T/P。解析后的数值以K为单位，且范围不能够超过64比特表示的有符号整数，即1KB~9007199254740991KB。

示例

创建示例模式schema_test和用户user_a：

```
CREATE SCHEMA schema_test;
CREATE USER user_a PASSWORD '{Password}';
```

将当前模式schema_test更名为schema_test1：

```
ALTER SCHEMA schema_test RENAME TO schema_test1;
```

将schema_test1的所有者修改为user_a：

```
ALTER SCHEMA schema_test1 OWNER TO user_a;
```


版权所有 © 华为云计算技术有限公司

示例：创建一个设置了永久存储空间的 Schema

以下模拟一个示例，创建一个名为testsche的schema，并设置永久存储空间为10MB，当数据超过该设置值，系统报错。

除了通过SQL语句设置永久存储空间之外，您也可以通过控制台界面进行Schema永久存储空间设置，详情参见**配置DWS数据库的Schema存储空间**章节。

步骤1 创建一个名为testsche的schema，并设置永久存储空间为10MB。

```
DROP SCHEMA IF EXISTS testsche;
CREATE SCHEMA testsche  WITH PERM SPACE '10M';
```

步骤2 使用系统表**PG_NAMESPACE**查询Schema的存储空间大小，其中permspace为永久存储空间限额，usedspace为已使用永久存储空间大小。

```
SELECT permspace,usedspace FROM PG_NAMESPACE WHERE nspname = 'testsche';
```

```
gaussdb=> CREATE SCHEMA testsche  WITH PERM SPACE '10M';
CREATE SCHEMA
gaussdb=> SELECT permspace,usedspace FROM PG_NAMESPACE WHERE nspname = 'testsche';
 permspace | usedspace
-----------+-----------
  10485760 |         0
(1 row)
```

步骤3 创建一张测试表，并导入数据。

```
DROP TABLE IF EXISTS testsche.src;
DROP TABLE IF EXISTS testsche.t1;
CREATE TABLE testsche.src AS SELECT 1;
CREATE TABLE testsche.t1(a int, b numeric(15,2)) WITH(orientation=column);
INSERT INTO testsche.t1 SELECT generate_series(1,20000000) % 1000,generate_series(1,20000000) FROM testsche.src;
```

回显报错信息，超过该schema的永久存储空间大小。

```
gaussdb=> CREATE TABLE testsche.src AS SELECT 1;
NOTICE:  The 'DISTRIBUTE BY' clause is not specified. Using round-robin as the distribution mode by default.
HINT:  Please use 'DISTRIBUTE BY' clause to specify suitable data distribution column.
INSERT 0 1
gaussdb=> CREATE TABLE testsche.t1(a int, b numeric(15,2)) WITH(orientation=column);
NOTICE:  The 'DISTRIBUTE BY' clause is not specified. Using round-robin as the distribution mode by default.
HINT:  Please use 'DISTRIBUTE BY' clause to specify suitable data distribution column.
CREATE TABLE
gaussdb=> INSERT INTO testsche.t1 SELECT generate_series(1,20000000) % 1000,generate_series(1,20000000) FROM testsche.src;
ERROR:  dn_6005_6006: out of testsche schema's perm space limit, usedspace: 15204352, needspace: 0, spacelimit: 10485760
gaussdb=>
```

步骤4 调大Schema的永久存储空间大小为10GB，再次导入，显示成功。

```
ALTER SCHEMA testsche WITH PERM SPACE '10G';
INSERT INTO testsche.t1 SELECT generate_series(1,20000000) % 1000,generate_series(1,20000000) FROM testsche.src;
```

```
gaussdb=> ALTER SCHEMA testsche WITH PERM SPACE '10G';
ALTER SCHEMA
gaussdb=> INSERT INTO testsche.t1 SELECT generate_series(1,20000000) % 1000,generate_series(1,20000000) FROM testsche.src;
INSERT 0 20000000
gaussdb=>
```

步骤5 查看已用空间。

```
SELECT permspace,usedspace FROM PG_NAMESPACE WHERE nspname = 'testsche';
```

```
gaussdb=> SELECT permspace,usedspace FROM PG_NAMESPACE WHERE nspname = 'testsche';
  permspace  | usedspace
-------------+-----------
 10737418240 |     16384
(1 row)
```

----结束

 版权所有 © 华为云计算技术有限公司

相关链接

CREATE SCHEMA，DROP SCHEMA

配置DWS数据库的Schema存储空间

13.2.6 DROP SCHEMA

功能描述

从数据库中删除模式。

注意事项

- 只有模式的所有者或者被授予了模式DROP权限的用户有权限执行DROP SCHEMA命令，系统管理员默认拥有此权限。

危险

- **DROP对象**操作（如DATABASE、USER/ROLE、SCHEMA、TABLE、VIEW等对象）存在数据丢失风险，尤其含带CASCADE级联删除场景，会将关联的对象一并删除，操作需谨慎，操作前需考虑数据备份。
- 更多开发设计规范参见**总体开发设计规范**。

语法格式

```
DROP SCHEMA [ IF EXISTS ] schema_name [, ...] [ CASCADE | RESTRICT ];
```

参数说明

表 13-25 DROP SCHEMA 参数说明

参数	描述	取值范围
IF EXISTS	如果指定的模式不存在，发出一个notice而不是抛出一个错误。	-
schema_name	要删除的模式名。	已存在的模式名。
CASCADE | RESTRICT	当执行删除操作时，如何处理依赖对象中的相关数据。	• CASCADE：自动删除包含在模式中的对象。 • RESTRICT：如果模式包含任何对象，则删除失败（缺省行为）。

版权所有 © 华为云计算技术有限公司

须知

- 不要随意删除pg_temp或pg_toast_temp开头的模式，这些模式是系统内部使用的，如果删除，可能导致无法预知的结果。
- 无法删除当前模式。如果要删除当前模式，须切换到其他模式下。

示例

先创建模式ds_new。

```
CREATE SCHEMA ds_new;
```

删除模式ds_new：

```
DROP SCHEMA ds_new;
```

相关链接

ALTER SCHEMA，**CREATE SCHEMA**

13.2.7 ALTER EXTERNAL SCHEMA

功能描述

修改EXTERNAL SCHEMA（外部模式）的属性。该语法仅8.3.0及以上版本支持。

语法格式

根据指定的名称修改EXTERNAL SCHEMA。

```
ALTER EXTERNAL SCHEMA schema_name
   WITH [ SOURCE source_name ]
        [ DATABASE 'database_name' ]
        [ SERVER server_name ]
        [ CATALOG 'catalog_name' ]
        [ OPTIONS ( { option_name ' value ' } [, ...] ) ]
        [ METAADDRESS 'address']
        [ CONFIGURATION 'confpath'];
```

参数说明

表 13-26 ALTER EXTERNAL SCHEMA 参数说明

参数	描述	取值范围
schema_name	需要修改的外部模式的名称。	已有的外部模式名。
SOURCE	外部元数据存储引擎的类型。	当前source_type仅支持dli，lakeformation，hive。
DATABASE	指定外部SCHEMA所对应的要访问的数据库。	-
SERVER	通过external schema关联foreign server以达到访问外部数据的目的。	已存在的foreign server且type类型为lf、obs、dli或hdfs。

版权所有 © 华为云计算技术有限公司

参数	描述	取值范围
CATALOG	Lakeformation中对应的要访问的目录名称。	-
OPTIONS	指定外表数据的各类参数。该参数仅8.3.0及以上版本支持。	dli_project_id，DLI服务对应的项目编号，可在管理控制台上获取项目ID，该参数仅支持server类型为DLI时设置。
METAADDRESS	Hive Metastore通讯接口地址。该参数仅9.1.0及以上版本支持。	-
CONFIGURATION	Hive Metastore相关配置文件存放路径。该参数仅9.1.0及以上版本支持。 如果当前搜索路径上的模式中存在同名对象时，需要明确指定引用对象所在的模式。可以通过命令SHOW SEARCH_PATH来查看当前搜索路径上的模式。	-

示例

准备前置条件：

1. 参见《 数据仓库服务用户指南 》的“管理LakeFormation数据源”章节分别创建名称为lf_server和my_server的server。
2. 创建名称为ex1的EXTERNAL SCHEMA。

```
CREATE EXTERNAL SCHEMA ex1
    WITH SOURCE lakeformation
        DATABASE 'demo'
        SERVER lf_server
        CATALOG 'hive';
```

修改名称为ex1的EXTERNAL SCHEMA对应的数据库和FOREIGN SERVER。

```
ALTER EXTERNAL SCHEMA ex1
    WITH DATABASE 'demo'
        SERVER my_server;
```

13.2.8 CREATE EXTERNAL SCHEMA

功能描述

创建EXTERNAL SCHEMA，即外部模式。该语法仅8.3.0及以上版本支持。

使用CREATE EXTERNAL SCHEMA建立DWS集群与外部数据源（如Lakeformation、DLI、Hive Metastore）的元数据映射通道，实时高效访问外部数据源。用户通过“外部模式名.表名” 直接查询，无需手动创建外表或数据搬迁，显著跨源数据访问与分析。

 版权所有 © 华为云计算技术有限公司

注意事项

- 具有当前数据库CREATE权限的用户可创建外部SCHEMA。
- 创建命名对象时不可用EXTERNAL SCHEMA作为前缀修饰，即不支持在EXTERNAL SCHEMA下创建对象。
- CREATE EXTERNAL SCHEMA不支持在新模式中创建对象的子命令。

语法格式

创建EXTERNAL SCHEMA。

```
CREATE EXTERNAL SCHEMA schema_name
    WITH SOURCE source_name
        DATABASE 'database_name'
        SERVER server_name
        [ CATALOG 'catalog_name' ]
        [ OPTIONS ( { option_name ' value ' } [, ...] ) ]
        [METAADDRESS 'address']
        [CONFIGURATION 'confpath'];
```

参数说明

表 13-27 CREATE EXTERNAL SCHEMA 参数说明

参数	描述	取值范围
schema_name	外部模式名称。 • 模式名不能和当前数据库里其他的模式重名。 • 模式的名称不能以“pg_”开头。	字符串，需符合**标识符命名规范**。
SOURCE	外部元数据存储引擎的类型。	当前仅支持dli，lakeformation，hive。
DATABASE	指定外部SCHEMA所对应的要访问的数据库。	-
SERVER	外部SCHEMA关联的外部服务器名称，关联后可访问到外部数据。	支持关联type类型为lf、obs、dli或hdfs的Server。详情可参见**CREATE SERVER**。其中lf类型的Server，请参见《数据仓库服务用户指南》的“管理LakeFormation数据源”进行创建。
CATALOG	Lakeformation中对应的要访问的目录名称，当Server的type为lf时需要指定catalog，否则不需要指定。	-
OPTIONS	用于指定外表数据的各类参数，当前支持的参数类型如右列所示。该参数仅8.3.0及以上版本支持。	dli_project_id：DLI服务对应的项目编号，可在管理控制台上获取项目ID，该参数仅支持type类型为dli的Server。

版权所有 © 华为云计算技术有限公司

参数	描述	取值范围
METAADDRESS	表示Hive Metastore通讯接口。该参数仅9.1.0及以上版本支持。	-
CONFIGURATION	表示Hive Metastore相关配置文件存放路径。该参数仅9.1.0及以上版本支持。	如果当前搜索路径上的模式中存在同名对象时，需要明确指定引用对象所在的模式。可以通过命令SHOW SEARCH_PATH来查看当前搜索路径上的模式。

示例：通过 external schema 读取 Lakeformation 表

通过external schema读取Lakeformation表的方式如下：

1. 创建lf_server，对应的foreign data wrapper为DFS_FDW。

   说明

   如何创建lf_server请参见《数据仓库服务用户指南》中“管理LakeFormation数据源”章节。

2. 创建external schema。

```
CREATE EXTERNAL SCHEMA ex_lf
   WITH SOURCE lakeformation
        DATABASE 'demo'
        SERVER lf_server
        CATALOG 'hive';
```

   - SOURCE为lakeformation。
   - 表关联的DLI服务器为lf_server。
   - DATABASE为对应的Lakeformation数据库。
   - CATALOG为对应要访问的Lakeformation的目录名称，根据实际替换。

3. 角色授权。

   a. 查询当前用户。

```
SELECT current_user;
```

   b. Lakeformation管控面创建与当前同名角色并对要访问的表授权。

4. 通过external schema查询Lakeformation表的数据，test_lf为要访问的Lakeformation表。

```
SELECT COUNT(*) FROM ex_dli.test_lf;
 count
-------
    20
(1 row)
```

通过 external schema 读取 Hive Metastore 表

通过external schema读取Hive Metastore表。仅9.1.0及以上集群版本支持。

1. 创建obs/hdfs server，对应的foreign data wrapper为DFS_FDW。

```
CREATE SERVER hdfs_server
FOREIGN DATA WRAPPER HDFS_FDW
OPTIONS (
address '***.***.***.***:9000',
type'HDFS');
```

 版权所有 © 华为云计算技术有限公司

```
CREATE SERVER obs_server
FOREIGN DATA WRAPPER dfs_fdw
OPTIONS (
    address 'obs.ap-southeast-1.myhuaweicloud.com' ,     ACCESS_KEY
'access_key_value_to_be_replaced',
    SECRET_ACCESS_KEY 'secret_access_key_value_to_be_replaced',
    encrypt 'on',
    type 'obs' );
```

2. 手动创建external schema对接Hive Metastore时，用户需先准备好相关配置文件，具体如下：
   a. 从集群管理页面或hive安装目录获取hive-site.xml、hivemetastore-site.xml这两个配置文件，上传至CONFIGURATION目录。
   b. 新增elk-site.xml配置文件并添加配置项（或在原有的elk-site.xml配置文件中新增），需要根据实际替换用户名、域名、keytab文件名等信息。

```
    <property>
        <name>hive.kerberos.principal</name>
        <value>user/hadoop.hadoop.com@HADOOP.COM</value> //连接Hadoop的用户@域名
    </property>
    <property>
        <name>hive.kerberos.keytab</name>
        <value>{配置目录}/user.keytab</value> //连接Hadoop用户的keytab文件的绝对路径
    </property>
```

3. 创建external schema。

```
CREATE EXTERNAL SCHEMA ex_hms
    WITH SOURCE source_type
    DATABASE 'db_name'
    SERVER srv_name
    METAADDRESS 'address'
    CONFIGURATION 'confpath';
```

   - SOURCE为hive。
   - SERVER为表关联的服务器obs/hdfs server。
   - DATABASE为对应的HMS数据库。
   - METAADDRESS为Hive Metastore通讯接口。
   - CONFIGURATION是一个绝对路径，表示Hive Metastore相关配置文件存放路径，根据实际替换。

4. 通过external schema查询HMS表的数据，test_hms为要访问的HMS表，根据实际替换。

```
SELECT COUNT(*) FROM ex_hms.test_hms;
 count
-------
    20
(1 row)
```

相关链接

ALTER EXTERNAL SCHEMA

使用实践

使用EXTERNAL SCHEMA跨集群访问HiveMetaStore元数据

版权所有 © 华为云计算技术有限公司

13.2.9 CREATE TYPE

功能描述

该语法用于创建**用户自定义类型（User-defined type，简称UDT)**，本质是将一组相关字段或值封装成一个新的、可复用的数据类型，类似编程语言中的“结构体”，让数据库表设计更贴合业务逻辑，减少冗余、提升可读性。**该功能仅适用于行存表。**

当前支持创建的UDT包括：复合类型、基本类型、shell类型和枚举类型。

- **复合类型**

  复合类型通过“属性名 + 对应数据类型”的列表来定义；若属性的数据类型支持排序，还可专门为该属性指定排序规则。复合类型的本质与数据表的行类型完全一致，但如果仅需定义一套结构化的数据类型，通过使用CREATE TYPE语法可实现更轻量化的类型定义，而无需创建一个实际的数据表。

  独立的复合类型具备很高的实用价值，例如作为自定义函数的参数类型或返回值类型。典型场景如用户表需存储家庭、工作地址，地址包含省、市、详细地址、邮编等信息，需复用地址结构，避免表字段冗余，此时可创建复合类型。

  需要注意的是，创建复合类型时，用户必须拥有其所有属性对应数据类型的USAGE权限。

- **基本类型**

  创建一种全新的基本类型（非组合、枚举），需自定义输入、 输出函数实现类型的编解码，适用于业务专属的特殊类型（如手机号、身份证号、带币种的金额）。使用了该基本类型的自定义函数必须使用C语言或者另外一种底层语言所编写。

  典型场景，例如用户联系方式需强制为11位手机号，自动过滤非数字字符，避免非法格式录入。

- **shell类型**

  shell类型是一种用于后面要定义的类型的占位符，通过发出一个只带类型名参数的CREATE TYPE命令来创建该类型。在创建**基本类型**时，需要shell类型作为一种向前引用。

- **枚举类型**

  由若干个标签构成的列表，每一个标签值都是一个非空字符串，且字符串长度必须不超过64个字节。其核心作用是限定字段取值为固定的枚举值集合，替代VARCHAR等其他数据类型的检查约束，提升数据一致性和可读性，例如可以避免拼写错误。

  典型场景如，订单状态仅允许：待支付、已支付、已发货 、已送达、已取消，需保证所有表（订单表、退款表）的状态值一致，避免填入其他取值，导致拼写错误。

注意事项

- **该功能仅适用于行存表，列存表不支持。**
- 如果给定一个模式名，那么该类型将被创建在指定的模式中，否则它会被创建在当前模式中。类型名称必须与同一个模式中任何现有的类型或者域有所区别（因为表具有相关的数据类型，类型名称也必须与同一个模式中任何现有表的名字不同）。
- 自定义基本类型时，参数可以以任意顺序出现，input_function和output_function为必选参数，其他为可选参数。

 版权所有 © 华为云计算技术有限公司

- 输入和输出函数能被声明为具有新类型的结果或参数是因为：必须在创建新类型之前先创建这两个函数。而新类型应该首先被定义为一种shell type，它是一种占位符类型，除了名称和拥有者之外它没有其他属性，这可以通过不带额外参数的命令CREATE TYPE name做到，接着用C语言写的I/O函数可以被定义为引用这种shell type。最后，用带有完整定义的CREATE TYPE把该shell type替换为一个完全的、合法的类型定义，之后新类型就可以正常使用了。

语法格式

复合类型

```
CREATE TYPE name AS
    ( [ attribute_name data_type [ COLLATE collation ] [, ... ] ] )
```

基本类型

```
CREATE TYPE name (
    INPUT = input_function,
    OUTPUT = output_function
    [ , RECEIVE = receive_function ]
    [ , SEND = send_function ]
    [ , TYPMOD_IN = type_modifier_input_function ]
    [ , TYPMOD_OUT = type_modifier_output_function ]
    [ , ANALYZE = analyze_function ]
    [ , INTERNALLENGTH = { internallength | VARIABLE } ]
    [ , PASSEDBYVALUE ]
    [ , ALIGNMENT = alignment ]
    [ , STORAGE = storage ]
    [ , LIKE = like_type ]
    [ , CATEGORY = category ]
    [ , PREFERRED = preferred ]
    [ , DEFAULT = default ]
    [ , ELEMENT = element ]
    [ , DELIMITER = delimiter ]
    [ , COLLATABLE = collatable ]
)
```

shell类型

```
CREATE TYPE name
```

枚举类型

```
CREATE TYPE name AS ENUM
    ( [ 'label' [, ... ] ] )
```

参数说明

表 13-28 复合类型参数说明

参数	描述	取值范围
name	要创建的类型的名称，可以用模式修饰。	需符合**标识符命名规范**。
attribute_name	复合类型的一个属性（列）的名称。	-
data_type	复合类型属性对应的数据类型，必须是DWS已存在的类型（比如 int、varchar (50)、date 等）。	-

版权所有 © 华为云计算技术有限公司

参数	描述	取值范围
collation	复合类型属性的 “排序规则”，控制字符串类数据的排序、比较逻辑（比如中文拼音排序、大小写不敏感排序）。	-

表 13-29 基本类型参数说明

参数	描述	取值范围
name	要创建的类型的名称，可以用模式修饰。	需符合**标识符命名规范**。
input_function	将数据从类型的**外部文本形式**转换为内部形式的函数名。 输入函数可以被声明为有一个cstring类型的参数，或者有三个类型分别为cstring、 oid、integer的参数。 ● cstring参数是以C字符串存在的输入文本。 ● oid参数是该类型自身的OID（对于数组类型则是其元素类型的OID）。 ● integer参数是目标列的typmod，如果typmod不明确，则将传递 -1。 输入函数必须返回一个该数据类型本身的值。通常，一个输入函数应该被声明为STRICT。 如果不是这样，在读到一个NULL输入值时，调用输入函数时第一个参数会是NULL。在这种情况下，该函数必须仍然返回NULL，除非调用函数发生了错误（这种情况主要是想支持域输入函数，域输入函数可能需要拒绝NULL输入）。	-
output_function	将数据从类型的内部形式转换为**外部文本形式**的函数名。 输出函数必须被声明为有一个新数据类型的参数。输出函数必须返回类型cstring。对于NULL值不会调用输出函数。	-

 版权所有 © 华为云计算技术有限公司

参数	描述	取值范围
receive_function（可选参数）	将数据从类型的**外部二进制形式**转换成内部形式的函数名。 如果没有该函数，该类型不能参与到二进制输入中。二进制表达转换成内部形式代价更低，然而却更容易移植（例如，标准的整数数据类型使用网络字节序作为外部二进制表达，而内部表达是机器本地的字节序）。receive_function应该执行足够的检查以确保该值是有效的。 接收函数可以被声明为有一个internal类型的参数，或者有三个类型分别为internal、oid、integer的参数。 ● internal参数是一个指向StringInfo缓冲区的指针，其中保存着接收到的字节串。 ● oid和integer参数和文本输入函数的相同。 接收函数必须返回一个该数据类型本身的值。通常，一个接收函数应该被声明为STRICT。如果不是这样，在读到一个NULL输入值时调用接收函数时第一个参数会是NULL。在这种情况下，该函数必须仍然返回NULL，除非接收函数发生了错误（这种情况主要是想支持域接收函数，域接收函数可能需要拒绝NULL输入）。	-
send_function（可选参数）	将数据从类型的内部形式转换为**外部二进制形式**的函数名。 如果没有该函数，该类型将不能参与到二进制输出中。发送函数必须被声明为有一个新数据类型的参数。发送函数必须返回类型bytea。对于NULL值不会调用发送函数。	-

 版权所有 © 华为云计算技术有限公司

参数	描述	取值范围
type_modifier_input_function（可选参数）参数。	将类型的**修饰符数组**转换为内部形式的函数名。 如果类型支持修饰符（附加在类型声明上的可选约束，例如，char(5)或numeric(30,2)），则需要可选的type_modifier_input_function和type_modifier_output_function。DWS允许用户定义的类型有一个或者多个简单常量或者标识符作为修饰符。不过，为了存储在系统目录中，该信息必须能被打包到一个非负整数值中。所声明的修饰符会被以cstring数组的形式传递给type_modifier_input_function。type_modifier_input_function必须检查该值的合法性（如果值错误就抛出一个错误），如果值正确，要返回一个非负integer值，该值将被存储在“typmod”列中。如果类型没有type_modifier_input_function则类型修饰符将被拒绝。type_modifier_output_function把内部的整数typmod值转换回正确的形式用于用户显示。type_modifier_output_function必须返回一个cstring值，该值就是追加到类型名称后的字符串。例如，numeric的函数可能会返回(30,2)。如果默认的显示格式就是只把存储的typmod整数值放在圆括号内，则允许省略type_modifier_output_function。	-
type_modifier_output_function（可选参数）	将类型的修饰符的内部形式转换为外部文本形式的函数名。	-
analyze_function（可选参数）	为该数据类型执行统计分析的函数名的可选参数。 默认情况下，如果该类型有一个默认的B-Tree操作符类，ANALYZE将尝试用类型的“equals”和“less-than”操作符来收集统计信息。这种行为对于非标量类型并不合适，因此可以通过指定一个自定义分析函数来覆盖这种行为。分析函数必须被声明为有一个类型为internal的参数，并且返回一个boolean结果。	-
internallength（可选参数）	一个数字常量，用于指定新类型的内部表达的字节长度。默认为变长。 虽然只有I/O函数和其他为该类型创建的函数才知道新类型的内部表达的细节， 但是内部表达的一些属性必须被向DWS声明。其中最重要的是internallength。基本数据类型可以是定长的（这种情况下internallength是一个正整数）或者是变长的（把internallength设置为VARIABLE，在内部通过把typlen设置为-1表示）。所有变长类型的内部表达都必须以一个4字节整数开始，internallength定义了总长度。	-

版权所有 © 华为云计算技术有限公司

参数	描述	取值范围
PASSEDBYVALUE（可选参数）	表示此数据类型的值需要被传值而不是传引用。传值的类型必须是定长的，并且它们的内部表达不能超过Datum类型（某些机器上是4字节，其他机器上是8字节）的尺寸。	-
alignment（可选参数）	该参数指定数据类型的存储对齐需求。如果被指定，必须是char、int2、int4或者double。默认是int4。 允许的值等同于以1、2、4或8字节边界对齐。要注意变长类型的alignment参数必须至少为4，因为它们需要包含一个int4作为它们的第一个组成部分。	-
storage（可选参数）	该数据类型的存储策略。 如果被指定，必须是plain、external、extended或者main。 默认是plain。 ● plain指定该类型的数据将总是被存储在线内并且不会被压缩。（对定长类型只允许plain） ● extended指定系统将首先尝试压缩一个长的数据值，并且将在数据仍然太长的情况下把值移出主表行。 ● external允许值被移出主表， 但是系统将不会尝试对它进行压缩。 ● main允许压缩，但是不鼓励把值移出主表（如果没有其他方法让行的大小变得合适，具有这种存储策略的数据项仍将被移出主表，但比起extended以及external项来，这种存储策略的数据项会被优先考虑保留在主表中）。 除plain之外所有的storage值都暗示该数据类型的函数能处理被TOAST过的值。指定的值仅决定一种可TOAST数据类型的列的默认TOAST存储策略，用户可以使用ALTER TABLE SET STORAGE为列选取其他策略。	-
like_type（可选参数）	与新类型具有相同表达的现有数据类型的名称。会从这个类型中复制internallength、passedbyvalue、 alignment以及storage的值（ 除非在这个CREATE TYPE命令的其他地方用显式说明覆盖）。 当新类型的底层实现是以一种现有的类型为参考时，用这种方式指定表达特别有用。	-

 版权所有 © 华为云计算技术有限公司

参数	描述	取值范围
category（可选参数）	此类型的分类码（单个ASCII字符）。 默认为'U'，表示用户定义类型。为了创建自定义分类，也可以选择其他ASCII字符。 category和preferred参数可以被用来帮助控制在混淆的情况下应用哪一种隐式造型。每一种数据类型都属于一个用单个ASCII字符命名的分类，并且每一种类型可以是其所属分类中的“首选”。当有助于解决重载函数或操作符时，解析器将优先造型到首选类型（但是只能从同类的其他类型造型）。对于没有隐式转换到或来自任意其他类型的类型，让这些设置保持默认即可。不过，对于有隐式转换的相关类型的组，把它们都标记为属于同一个类别并且选择一种或两种“最常用”的类型作为该类别的首选通常是很有用的。在把一种用户定义的类型增加到一个现有的内建类别（例如，数字或者字符串类型）中时，category参数特别有用。不过，也可以创建新的全部是用户定义类型的类别。对这样的类别，可选择除大写字母之外的任何ASCII字符。	-
preferred（可选参数）	如果此类型是其类型分类中的优先类型则为TRUE，否则为FALSE。默认为FALSE。在现有类型分类中创建新的优先类型要非常谨慎， 因为这可能会导致很大的改变。	-
default（可选参数）	数据类型的默认值。如果被省略，默认值是空。 如果用户希望该数据类型的列被默认为某种非空值，可以指定一个默认值。默认值可以用DEFAULT关键词指定（这样一个默认值可以被附加到一个特定列的显式DEFAULT子句覆盖）。	-
element（可选参数）	被创建的类型是一个数组，element指定了数组元素的类型。例如，要定义一个4字节整数的数组（int4）， 应指定ELEMENT = int4。	-
delimiter（可选参数）	指定此类型组成的数组中分隔值的定界符。 可以把delimiter设置为一个特定字符，默认的定界符是逗号（,）。注意定界符是与数组元素类型相关的，而不是数组类型本身相关。	-
collatable（可选参数）	如果此类型的操作可以使用排序规则信息，则为TRUE。默认为FALSE。 如果collatable为TRUE，此类型的列定义和表达式可能通过使用COLLATE子句携带有排序规则信息。在该类型上操作的函数的实现负责真正利用这些信息，仅把类型标记为可排序的并不会让它们自动地去使用这类信息。	-

版权所有 © 华为云计算技术有限公司

表 13-30 shell 类型参数说明

参数	描述	取值范围
name	要创建的类型的名称，可以用模式修饰。	需符合**标识符命名规范**。

表 13-31 枚举类型参数说明

参数	描述	取值范围
name	要创建的类型的名称，可以用模式修饰。	需符合**标识符命名规范**。
label（可选参数）	与枚举类型的一个值相关的文本标签，其值为长度不超过64个字符的非空字符串。	-

示例：创建复合类型表示地址信息

步骤1 创建一个复合类型用于表示地址信息，包括所在省、城市、街道、门牌号。

```
DROP TYPE IF EXISTS address_type;
CREATE TYPE address_type AS (
    province VARCHAR(20),
    city VARCHAR(50),
    street VARCHAR(100),
    number VARCHAR(10)
);
```

步骤2 使用该类型创建表。

```
DROP TABLE IF EXISTS customers;
CREATE TABLE customers (
    customer_id SERIAL PRIMARY KEY,
    name VARCHAR(100),
    billing_address address_type,
    shipping_address address_type
);
```

步骤3 插入数据。

```
INSERT INTO customers (name, billing_address, shipping_address)
VALUES (
    'John Doe',
    ('SDong', 'QD', '123 Main St', '10001')::address_type,
    ('SDong', 'QD', '456 Park Ave', '10022')::address_type
);
```

步骤4 查询数据。

```
SELECT
    name,
    (billing_address).street as bill_street,
    (shipping_address).city as ship_city
FROM customers;
```

```
   name   | bill_street | ship_city
----------+-------------+-----------
 John Doe | 123 Main St | QD
(1 row)
```


版权所有 © 华为云计算技术有限公司

步骤5 更新数据。

```
UPDATE customers
SET billing_address = ('SDong', 'QD','789 Broadway', '10003')::address_type
WHERE customer_id = 1;
```

----结束

示例：创建枚举类型用于电商订单管理

步骤1 创建枚举类型，里面的枚举值分别对应“待支付”、“已支付”、“已发货”、“已送达”、“已取消”。

```
DROP TYPE IF EXISTS order_status;
CREATE TYPE order_status AS ENUM (
    'pending', 'paid', 'shipped', 'delivered', 'cancelled'
);
```

步骤2 创建订单表，状态列status使用枚举类型order_status。

```
DROP TABLE IF EXISTS orders;
CREATE TABLE orders (
    order_id varchar(32) PRIMARY KEY,
    user_id int,
    goods_id int,
    amount decimal(10,2),
    status order_status DEFAULT 'pending', -- 默认待支付
    create_time timestamp DEFAULT now(),
    update_time timestamp DEFAULT now()
);
```

步骤3 插入测试数据。

```
INSERT INTO orders (order_id, user_id, goods_id, amount)
VALUES ('ORD20260127001', 1001, 5001, 299.99); -- 状态默认pending
```

步骤4 更新订单状态，仅允许更新为限定的枚举值。

```
UPDATE orders SET status = 'paid', update_time = now() WHERE order_id = 'ORD20260127001';
```

步骤5 查询数据。

```
SELECT * FROM orders WHERE order_id = 'ORD20260127001';
```

```
gaussdb=> SELECT * FROM orders WHERE order_id = 'ORD20260127001';
    order_id    | user_id | goods_id | amount | status |        create_time         |        update_time
----------------+---------+----------+--------+--------+----------------------------+----------------------------
 ORD20260127001 |    1001 |     5001 | 299.99 | paid   | 2026-01-29 16:21:17.813356 | 2026-01-29 16:21:25.098978
(1 row)
```

步骤6 更新为非枚举值中的其他值时，系统会报错，以此来保证数据一致性。

```
UPDATE orders SET status = 'wait_pay' WHERE order_id = 'ORD20260127001';
```

```
gaussdb=> UPDATE orders SET status = 'wait_pay' WHERE order_id = 'ORD20260127001';
ERROR:  invalid input value for enum order_status: "wait_pay"
LINE 1: UPDATE orders SET status = 'wait_pay' WHERE order_id = 'ORD2...
                                   ^
CONTEXT:  referenced column: status
```

----结束

示例：创建基本类型，用于 C 语言函数调用

步骤1 在创建基本类型前，先创建shell类型（占位符），用于后续定义I/O函数时引用该类型。

```
DROP TYPE IF EXISTS complex;
CREATE TYPE complex;
```

 版权所有 © 华为云计算技术有限公司

这个语句的作用是为要定义的类型创建了一个占位符，这样允许在定义其I/O函数时引用该类型。

步骤2 定义I/O函数，这些函数都引用到前面的shell类型complex，需要注意的是在创建函数时function必须声明为NOT FENCED模式，即约束该函数的执行在CN或者DN进程中进行，减少fork进程以及通信的开销。了解更多请参见**CREATE FUNCTION**。

```
--创建文本输入函数
CREATE OR REPLACE FUNCTION
complex_in(cstring)
   RETURNS complex
   AS 'filename'   --替换为实际的.so文件路径，例如/opt/dws/complex.so
   LANGUAGE C IMMUTABLE STRICT not fenced;

--创建文本输出函数
CREATE FUNCTION
complex_out(complex)
   RETURNS cstring
   AS 'filename'    --同上，替换为实际的.so文件路径。
   LANGUAGE C IMMUTABLE STRICT not fenced;

--创建二进制输入函数
CREATE FUNCTION
complex_recv(internal)
   RETURNS complex
   AS 'filename'      --同上，替换为实际的.so文件路径。
   LANGUAGE C IMMUTABLE STRICT not fenced;

--创建二进制输出函数
CREATE FUNCTION
complex_send(complex)
   RETURNS bytea
   AS 'filename'       --同上，替换为实际的.so文件路径。
   LANGUAGE C IMMUTABLE STRICT not fenced;
```

步骤3 创建定义完整的基本类型，替换之前的shell类型。

```
CREATE TYPE complex (
internallength = 16,      -- 定长16字节（2个double各8字节）
input = complex_in,       -- 绑定文本输入函数
output = complex_out,     -- 绑定文本输出函数
receive = complex_recv,   -- 绑定二进制输入函数
send = complex_send,      -- 绑定二进制输出函数
alignment = double        -- 指定8字节的存储对齐需求，匹配double类型存储
);
```

步骤4 下面是input、output、receive及send函数对应的C语言函数代码，供进一步开发参考。

```
--定义结构体Complex如下：
typedef struct Complex {
   double      x;
   double      y;
} Complex;

--定义input函数：
PG_FUNCTION_INFO_V1(complex_in);

Datum
complex_in(PG_FUNCTION_ARGS)
{
       char       *str = PG_GETARG_CSTRING(0);
   double      x,
               y;
   Complex    *result;

   if (sscanf(str, " ( %lf , %lf )", &x, &y) != 2)
      ereport(ERROR,
              (errcode(ERRCODE_INVALID_TEXT_REPRESENTATION),
               errmsg("invalid input syntax for complex: \"%s\"",
```

 版权所有 © 华为云计算技术有限公司

```
                    str)));

    result = (Complex *) palloc(sizeof(Complex));
    result->x = x;
    result->y = y;
    PG_RETURN_POINTER(result);
}

--定义output函数：
PG_FUNCTION_INFO_V1(complex_out);

Datum
complex_out(PG_FUNCTION_ARGS)
{
        Complex    *complex = (Complex *) PG_GETARG_POINTER(0);
        char       *result;

        result = (char *) palloc(100);
        snprintf(result, 100, "(%g,%g)", complex->x, complex->y);
        PG_RETURN_CSTRING(result);
}

--定义receive函数：
PG_FUNCTION_INFO_V1(complex_recv);

Datum
complex_recv(PG_FUNCTION_ARGS)
{
    StringInfo  buf = (StringInfo) PG_GETARG_POINTER(0);
    Complex    *result;

    result = (Complex *) palloc(sizeof(Complex));
    result->x = pq_getmsgfloat8(buf);
    result->y = pq_getmsgfloat8(buf);
    PG_RETURN_POINTER(result);
}

--定义send函数：
PG_FUNCTION_INFO_V1(complex_send);

Datum
complex_send(PG_FUNCTION_ARGS)
{
    Complex    *complex = (Complex *) PG_GETARG_POINTER(0);
    StringInfoData buf;

    pq_begintypsend(&buf);
    pq_sendfloat8(&buf, complex->x);
    pq_sendfloat8(&buf, complex->y);
    PG_RETURN_BYTEA_P(pq_endtypsend(&buf));
}
```

----结束

相关链接

ALTER TYPE，**DROP TYPE**

13.2.10 ALTER TYPE

功能描述

修改一个类型的定义。

 版权所有 © 华为云计算技术有限公司

注意事项

- 要使用ALTER TYPE，必须是该类型的所有者。
- 要修改类型的模式，还必须在新模式上拥有CREATE权限。
- 要修改所有者，必须是新的所有角色的直接或间接成员， 并且该角色必须在此类型的模式上有CREATE权限。 这些限制强制要求修改所有者不能执行任何通过删除和重建该类型无法实现的操作。
- 系统管理员拥有以任何方式修改任意类型的所有权。
- 要增加一个属性或是修改一个属性的类型，也必须有该类型的USAGE权限。

语法格式

- 修改类型。

```
ALTER TYPE name action [, ... ];
ALTER TYPE name OWNER TO { new_owner | CURRENT_USER | SESSION_USER };
ALTER TYPE name RENAME ATTRIBUTE attribute_name TO new_attribute_name [ CASCADE |
RESTRICT ];
ALTER TYPE name RENAME TO new_name;
ALTER TYPE name SET SCHEMA new_schema;
ALTER TYPE name ADD VALUE [ IF NOT EXISTS ] new_enum_value [ { BEFORE | AFTER }
neighbor_enum_value ];
ALTER TYPE name RENAME VALUE existing_enum_value TO new_enum_value;

where action is one of:
    ADD ATTRIBUTE attribute_name data_type [ COLLATE collation ] [ CASCADE | RESTRICT ];
    DROP ATTRIBUTE [ IF EXISTS ] attribute_name [ CASCADE | RESTRICT ];
    ALTER ATTRIBUTE attribute_name [ SET DATA ] TYPE data_type [ COLLATE collation ] [ CASCADE |
RESTRICT ];
```

- 给复合类型增加新的属性。

```
ALTER TYPE name ADD ATTRIBUTE attribute_name data_type [ COLLATE collation ] [ CASCADE |
RESTRICT ];
```

- 从复合类型删除一个属性。

```
ALTER TYPE name DROP ATTRIBUTE [ IF EXISTS ] attribute_name [ CASCADE | RESTRICT ];
```

- 改变一种复合类型中某个属性的类型。

```
ALTER TYPE name ALTER ATTRIBUTE attribute_name [ SET DATA ] TYPE data_type [ COLLATE
collation ] [ CASCADE | RESTRICT ];
```

说明

ADD ATTRIBUTE、DROP ATTRIBUTE和ALTER ATTRIBUTE选项可以组合成一个列表同时执行。例如，在一条命令中同时增加多个属性或是更改多个属性的类型。

- 改变类型的所有者。

```
ALTER TYPE name OWNER TO { new_owner | CURRENT_USER | SESSION_USER };
```

- 改变类型的名称或是一个复合类型中的一个属性的名称。

```
ALTER TYPE name RENAME TO new_name;
ALTER TYPE name RENAME ATTRIBUTE attribute_name TO new_attribute_name [ CASCADE |
RESTRICT ];
```

- 将类型移至一个新的模式中。

```
ALTER TYPE name SET SCHEMA new_schema;
```

- 为枚举类型增加一个新值。

```
ALTER TYPE name ADD VALUE [ IF NOT EXISTS ] new_enum_value [ { BEFORE | AFTER }
neighbor_enum_value ];
```

- 重命名枚举类型的一个标签值。

```
ALTER TYPE name RENAME VALUE existing_enum_value TO new_enum_value;
```

 版权所有 © 华为云计算技术有限公司

参数说明

表 13-32 ALTER TYPE 参数说明

参数	描述	取值范围
name	需要修改的现有类型的名称，可以用模式修饰。	有效的类型名称。
new_name	类型的新名称。	字符串。
new_owner	类型新所有者的用户名。	有效的用户名。
new_schema	类型的新模式。	有效的模式名。
attribute_name	要增加、更改或删除的属性的名称。	-
new_attribute_name	要重命名的属性的新名称。	-
data_type	要新增属性的数据类型，或是要更改的属性的新类型名。	-
new_enum_value	枚举类型新增加的标签值。	字符串。非空且长度不超过64个字节。
neighbor_enum_value	一个已有枚举标签值，新值应该被增加在紧接着该枚举值之前或者之后的位置上。	-
existing_enum_value	现有的要重命名的枚举值。	字符串。非空且长度不超过64个字节。
CASCADE	自动级联更新需更新类型以及相关联的记录和继承它们的子表。	-
RESTRICT	如果要级联更新类型是已更新类型的关联记录，则拒绝更新。RESTRICT为缺省选项。	-

示例

创建复合类型test，枚举类型testdata和用户user_t。

```
CREATE TYPE test AS (col1 int, col text);
CREATE TYPE testdata AS ENUM ('create', 'modify', 'closed');
CREATE USER user_t PASSWORD '{Password}';
```

重命名数据类型test为test1。

```
ALTER TYPE test RENAME TO test1;
```

修改用户定义类型test1的所有者为user_t。

```
ALTER TYPE test1 OWNER TO user_t;
```

把用户定义类型test1的模式改为user_t。

```
ALTER TYPE test1 SET SCHEMA user_t;
```

 版权所有 © 华为云计算技术有限公司

给用户定义类型test1添加一个新的属性col3。

```
ALTER TYPE user_t.test1 ADD ATTRIBUTE col3 int;
```

给枚举类型testdata添加一个标签值。

```
ALTER TYPE testdata ADD VALUE IF NOT EXISTS 'regress' BEFORE 'closed';
```

重命名枚举类型testdata的一个标签值。

```
ALTER TYPE testdata RENAME VALUE 'create' TO 'new';
```

相关链接

CREATE TYPE，**DROP TYPE**

13.2.11 DROP TYPE

功能描述

删除一个用户定义的数据类型。只有类型所有者才有删除权限。

语法格式

```
DROP TYPE [ IF EXISTS ] name [, ...] [ CASCADE | RESTRICT ]
```

参数说明

表 13-33 DROP TYPE 参数说明

参数	描述	取值范围
IF EXISTS	如果指定的类型不存在，那么发出一个notice而不是抛出一个错误。	-
name	要删除的类型名（可以有模式修饰）。	已存在的类型名。
CASCADE | RESTRICT	当执行删除操作时，如何处理依赖对象中的相关数据。	• CASCADE：级联删除依赖该类型的对象（比如字段、函数、操作符等）。 • RESTRICT：如果有依赖对象，则拒绝删除该类型（缺省行为）。

示例

先创建compfoo类型。

```
CREATE TYPE compfoo AS (f1 int, f2 text);
```

删除compfoo类型：

```
DROP TYPE IF EXISTS compfoo cascade;
```

 版权所有 © 华为云计算技术有限公司

相关链接

ALTER TYPE，CREATE TYPE

13.3 用户和角色

13.3.1 CREATE USER

功能描述

创建新的数据库用户。

注意事项

- 通过CREATE USER创建的用户，默认具有LOGIN权限。
- 通过CREATE USER创建用户的同时系统会在执行该命令的数据库中，为该用户创建一个同名的SCHEMA；其他数据库中，则不自动创建同名的SCHEMA；用户可使用CREATE SCHEMA命令，分别在其他数据库中，为该用户创建同名SCHEMA。
- 系统管理员在普通用户同名SCHEMA下创建的对象，所有者为SCHEMA的同名用户（非系统管理员）。
- 除系统管理员之外，其他用户即使被授权了SCHEMA的所有权限也无法在普通用户的同名SCHEMA下创建对象，除非把同名SCHEMA相关的角色权限赋予其他用户。具体操作可参考**赋予用户schema的all权限后建表依然报错**章节。

⚠ 注意

- **尽量避免**所有业务使用同一个数据库用户运行，请按业务模块规划不同数据库用户。
- **不建议**使用系统管理员用户跑业务，不同模块业务请通过多用户和权限进行访问控制。
- 更多开发设计规范参见**总体开发设计规范**。

语法格式

```
CREATE USER user_name [ [ WITH ] option [ ... ] ] [ ENCRYPTED | UNENCRYPTED ] { PASSWORD |
IDENTIFIED BY } { 'password' | DISABLE };
```

其中option子句用于设置权限及属性等信息。

```
{SYSADMIN | NOSYSADMIN}
   | {AUDITADMIN | NOAUDITADMIN}
   | {CREATEDB | NOCREATEDB}
   | {USEFT | NOUSEFT}
   | {CREATEROLE | NOCREATEROLE}
   | {INHERIT | NOINHERIT}
   | {LOGIN | NOLOGIN}
   | {REPLICATION | NOREPLICATION}
   | {INDEPENDENT | NOINDEPENDENT}
   | {VCADMIN | NOVCADMIN}
   | CONNECTION LIMIT connlimit
   | VALID BEGIN 'timestamp'
```


版权所有 © 华为云计算技术有限公司

```
| VALID UNTIL 'timestamp'
| RESOURCE POOL 'respool'
| USER GROUP 'groupuser'
| PERM SPACE 'spacelimit'
| TEMP SPACE 'tmpspacelimit'
| SPILL SPACE 'spillspacelimit'
| NODE GROUP logic_cluster_name
| IN ROLE role_name [, ...]
| IN GROUP role_name [, ...]
| ROLE role_name [, ...]
| ADMIN role_name [, ...]
| USER role_name [, ...]
| SYSID uid
| DEFAULT TABLESPACE tablespace_name
| PROFILE DEFAULT
| PROFILE profile_name
| PGUSER
| AUTHINFO 'authinfo'
| PASSWORD EXPIRATION period
```

参数说明

表 13-34 CREATE USER 参数说明

参数	描述	取值范围或默认值
user_name	要创建的用户名。	字符串，需符合标识符的命名规范且最多为63个字符。参见**标识符命名规范**。
password	登录密码。 密码规则如下： • 密码默认不少于8个字符。 • 不能与用户名及用户名倒序相同。 • 至少包含大写字母（A-Z），小写字母（a-z），数字（0-9），非字母数字字符（~!@#$%^&*()-_=+|[{}];:,<.>/?）四类字符中的三类字符。使用范围外的字符会收到告警，但依然允许创建。	字符串。
DISABLE	默认情况下，用户可以更改自己的密码，除非密码被禁用。要禁用用户的密码，请指定DISABLE。禁用某个用户的密码后，将从系统中删除该密码，此类用户只能通过外部认证来连接数据库，例如：IAM认证、kerberos或ldap认证。只有管理员才能启用或禁用密码。普通用户不能禁用初始用户的密码。要启用密码，请运行ALTER USER并指定密码。	-

版权所有 © 华为云计算技术有限公司

参数	描述	取值范围或默认值
ENCRYPTED | UNENCRYPTED	控制密码存储在系统表里的密码是否加密。（如果未显式指定，默认行为由配置参数password_encryption_type控制。） 按照产品安全要求，密码必须加密存储，因此UNENCRYPTED在DWS中禁止使用。由于系统无法对指定的加密密码字符串进行解密，如果当前的密码字符串已使用SHA256加密的格式，则无论是否声明ENCRYPTED或UNENCRYPTED，都将按原格式继续存储。这样就允许在dump/restore时重新加载已加密的密码。	-
SYSADMIN | NOSYSADMIN	决定新用户是否为“系统管理员”，具有SYSADMIN属性的用户拥有系统最高权限。	缺省为NOSYSADMIN。
AUDITADMIN | NOAUDITADMIN	定义用户是否有审计管理属性。	缺省为NOAUDITADMIN。
CREATEDB | NOCREATEDB	控制新用户是否能创建数据库。 新用户没有创建数据库的权限。	缺省为NOCREATEDB。
USEFT | NOUSEFT	决定一个新用户是否能操作外表，包括：新建外表、删除外表、修改外表、读写外表。 缺省为NOUSEFT，表示新建用户默认不能操作外表。 新用户没有操作外表的权限。	-
CREATEROLE | NOCREATEROLE	控制用户是否可以创建新角色或用户（也就是执行CREATE ROLE和CREATE USER）。拥有CREATEROLE权限的用户也可以修改和删除其他用户或角色。	缺省为NOCREATEROLE。
INHERIT | NOINHERIT	控制用户是否可以“继承”它所在组的用户的权限。不推荐使用。	-
LOGIN | NOLOGIN	控制用户是否可以登录数据库。具有LOGIN属性的用户才可以登录数据库。	缺省为LOGIN。
REPLICATION | NOREPLICATION	控制用户是否允许流复制或设置系统为备份模式。REPLICATION属性是特定的用户，仅用于复制。	缺省为NOREPLICATION。

 版权所有 © 华为云计算技术有限公司

参数	描述	取值范围或默认值
INDEPENDENT | NOINDEPENDENT	定义私有、独立的用户。具有INDEPENDENT属性的用户，管理员对其进行的控制、访问的权限被分离，具体规则如下： ● 未经INDEPENDENT用户授权，管理员无权对其表对象进行增、删、查、改、拷贝、授权操作。 ● 未经INDEPENDENT用户授权，管理员无权修改INDEPENDENT用户的继承关系。 ● 管理员无权修改INDEPENDENT用户的表对象的属主。 ● 管理员无权去除INDEPENDENT用户的INDEPENDENT属性。 ● 管理员无权修改INDEPENDENT用户的数据库密码，INDEPENDENT用户需管理好自身密码，密码丢失无法重置。 ● 管理员属性用户不允许定义修改为INDEPENDENT属性。	-
VCADMIN | NOVCADMIN	定义逻辑集群管理员用户。具有逻辑集群管理员属性的用户，和普通用户相比，有如下额外权限： ● 在所关联逻辑集群中创建、修改和删除资源池的权限。 ● 将所关联的逻辑集群的访问权限授予其他用户或角色，或回收其他用户或角色对关联逻辑集群的访问权限。	-
CONNECTION LIMIT	定义用户在单个CN上可以使用的最大并发连接数量。 系统管理员不受此参数的限制。	整数，大于等于-1。 缺省值为-1，表示没有限制。
VALID BEGIN	设置用户生效的时间戳。如果省略了该子句，用户无有效开始时间限制。	-
VALID UNTIL	设置用户失效的时间戳，即过了这个时间后，账户不可使用。如果省略了该子句，用户无有效结束时间限制。	-
RESOURCE POOL	设置用户使用的resource pool名称，该名称属于系统表：pg_resource_pool	-
USER GROUP 'groupuser'	创建用户的子用户。	-

版权所有 © 华为云计算技术有限公司

参数	描述	取值范围或默认值
PERM SPACE	设置用户永久表存储空间限额。 space_limit，表示永久表存储空间上限。	space_limit，字符串格式为正整数+单位，单位当前支持K/M/G/T/P。 0表示不限制。
TEMP SPACE	设置用户临时表存储空间限额。 tmpspacelimit，表示临时表存储空间限额。	tmpspacelimit，字符串格式为正整数+单位，单位当前支持K/M/G/T/P。 0表示不限制。
SPILL SPACE	设置用户算子落盘空间限额。 spillspacelimit，表示算子落盘空间限额。	spillspacelimit，字符串格式为正整数+单位，单位当前支持K/M/G/T/P。 0表示不限制。
NODE GROUP	设置用户关联的逻辑集群名称。如果需要关联的逻辑集群名称包含大写字符或特殊字符，指定逻辑集群名称时需要加双引号。	-
IN ROLE	新用户立即拥有IN ROLE子句中列出的一个或多个现有用户拥有的权限。不推荐使用。	-
IN GROUP	IN GROUP是IN ROLE废弃的拼写方式。不推荐使用。	-
ROLE	ROLE子句列出一个或多个现有的用户，它们将自动添加为这个新用户的成员，拥有新用户所有的权限。	-
ADMIN	ADMIN子句类似ROLE子句，不同的是ADMIN后的用户可以把新用户的权限赋予其他用户。	-
USER	USER子句是ROLE子句废弃的拼写方式。	-
SYSID	SYSID子句将被忽略，无实际意义。	-
DEFAULT TABLESPACE	DEFAULT TABLESPACE子句将被忽略，无实际意义。	-
PROFILE	PROFILE子句将被忽略，无实际意义。	-

版权所有 © 华为云计算技术有限公司

参数	描述	取值范围或默认值
PGUSER	该属性用于兼容开源Postgres的连接通讯，开源的Postgres客户端接口（推荐使用Postgres 9.2.19版本的相关客户端接口）可以使用具有该属性的数据库用户连接数据库。 该属性只用于兼容连接过程，而由于本产品与Postgres的内核差异导致的不兼容，不在此属性控制范围内。	-
AUTHINFO 'authinfo'	该属性用于指定用户认证类型，authinfo为类型说明字符串，大小写敏感。 当前仅支持LDAP类型，对应的类型说明字符串为'ldap'。LDAP属于外部认证，故需要同时指定PASSWORD DISABLE。 ● authinfo可以加上LDAP认证的额外信息，比如LDAP认证中的fulluser，fulluser等同于ldapprefix+username+ldapsuffix。当authinfo为'ldap'，表示用户认证类型为LDAP，此时ldapprefix和ldapsuffix的信息由pg_hba.conf中匹配的记录提供。 ● ALTER ROLE时不允许用户切换认证类型，仅允许LDAP用户修改LDAP属性。	-
PASSWORD EXPIRATION period	声明该用户的登录密码过期天数，登录密码过期之前用户需要及时修改密码。登录密码过期后用户无法登录，需要管理员设置新的登录密码后登录。	整数，-1~999。 ● -1（缺省值），表示没有过期限制。 ● 0，表示用户登录密码立即过期。

示例

创建用户new_u1。

```
CREATE USER new_u1 PASSWORD '{password}';
```

下面语句与上面的等价。

```
CREATE USER new_u1 IDENTIFIED BY '{password}';
```

如果创建有“创建数据库”权限的用户，则需要指定CREATEDB关键字。

```
CREATE USER new_u1 CREATEDB PASSWORD '{password}';
```

相关链接

ALTER USER，**CREATE ROLE**，**DROP USER**

 版权所有 © 华为云计算技术有限公司

常见问题

如何查看DWS数据库用户的创建时间?

13.3.2 ALTER USER

功能描述

修改数据库用户的属性。

注意事项

ALTER USER修改的会话参数只针对指定的用户，且在下一次会话中有效。

语法格式

- 修改用户的权限等信息。

```
ALTER USER user_name [ [ WITH ] option [ ... ] ];
```

其中option子句为：

```
{ CREATEDB | NOCREATEDB }
    | { CREATEROLE | NOCREATEROLE }
    | { INHERIT | NOINHERIT }
    | { AUDITADMIN | NOAUDITADMIN }
    | { SYSADMIN | NOSYSADMIN }
    | { USEFT | NOUSEFT }
    | { LOGIN | NOLOGIN }
    | { REPLICATION | NOREPLICATION }
    | {INDEPENDENT | NOINDEPENDENT}
    | {VCADMIN | NOVCADMIN}
    | CONNECTION LIMIT connlimit
    | [ ENCRYPTED | UNENCRYPTED ] PASSWORD { 'password' | DISABLE }
    | [ ENCRYPTED | UNENCRYPTED ] IDENTIFIED BY { 'password' [ REPLACE 'old_password' ] |
DISABLE }
    | VALID BEGIN 'timestamp'
    | VALID UNTIL 'timestamp'
    | RESOURCE POOL 'respool'
    | USER GROUP 'groupuser'
    | PERM SPACE 'spacelimit'
    | TEMP SPACE 'tmpspacelimit'
    | SPILL SPACE 'spillspacelimit'
    | NODE GROUP logic_cluster_name
    | ACCOUNT { LOCK | UNLOCK }
    | PGUSER
    | AUTHINFO 'authinfo'
    | PASSWORD EXPIRATION period
```

- 修改用户名。

```
ALTER USER user_name
    RENAME TO new_name;
```

- 修改与用户关联的指定会话参数值。

```
ALTER USER user_name
    SET configuration_parameter { { TO | = } { value | DEFAULT } | FROM CURRENT };
```

- 重置与用户关联的指定会话参数值。

```
ALTER USER user_name
    RESET { configuration_parameter | ALL };
```

 版权所有 © 华为云计算技术有限公司

参数说明

表 13-35 ALTER USER 参数说明

参数	描述	取值范围及默认值
user_name	需要修改的用户名。	有效的用户名。
new_name	用户的新名称。	字符串，需符合**标识符命名规范**。
CREATEDB | NOCREATEDB	控制用户是否能创建数据库。 新用户没有创建数据库的权限。	缺省为NOCREATEDB。
CREATEROLE | NOCREATEROLE	控制用户是否可以创建新用户或角色（也就是执行CREATE ROLE和CREATE USER）。拥有CREATEROLE权限的用户也可以修改和删除其他用户。	缺省为NOCREATEROLE。
INHERIT | NOINHERIT	控制用户是否可以“继承”它所在组的用户的权限。不推荐使用。	-
AUDITADMIN | NOAUDITADMIN	定义用户是否有审计管理属性。	缺省为NOAUDITADMIN。
SYSADMIN | NOSYSADMIN	控制用户是否为“系统管理员”，具有SYSADMIN属性的用户拥有系统最高权限。	缺省为NOSYSADMIN。
USEFT | NOUSEFT	控制用户是否有权限操作外表，包括：新建外表、删除外表、修改外表、读写外表。 缺省为NOUSEFT，表示新建用户默认不能操作外表。 新用户没有操作外表的权限。	-
LOGIN | NOLOGIN	控制用户是否可以登录数据库。具有LOGIN属性的用户才可以登录数据库。	缺省为LOGIN。
REPLICATION | NOREPLICATION	控制用户是否允许流复制或设置系统为备份模式。REPLICATION属性是特定的用户，仅用于复制。	缺省为NOREPLICATION。

版权所有 © 华为云计算技术有限公司

参数	描述	取值范围及默认值
INDEPENDENT | NOINDEPENDENT	定义私有、独立的用户。具有INDEPENDENT属性的用户，管理员对其进行的控制、访问的权限被分离，具体规则如下： ● 未经INDEPENDENT用户授权，管理员无权对其表对象进行增、删、查、改、拷贝、授权操作。 ● 未经INDEPENDENT用户授权，管理员无权修改INDEPENDENT用户的继承关系。 ● 管理员无权修改INDEPENDENT用户的表对象的属主。 ● 管理员无权去除INDEPENDENT用户的INDEPENDENT属性。 ● 管理员无权修改INDEPENDENT用户的数据库密码，INDEPENDENT用户需管理好自身密码，密码丢失无法重置。 ● 管理员属性用户不允许定义修改为INDEPENDENT属性。	-
VCADMIN | NOVCADMIN	定义逻辑集群管理员用户。具有逻辑集群管理员属性的用户，和普通用户相比，有如下额外权限： ● 在所关联逻辑集群中创建、修改和删除资源池的权限。 ● 将所关联的逻辑集群的访问权限授予其他用户或角色，或回收其他用户或角色对关联逻辑集群的访问权限。	-
CONNECTION LIMIT	定义用户在单个CN上可以使用的最大并发连接数量。 系统管理员不受此参数的限制。	取值为整数，大于等于-1。 缺省值为-1，表示没有限制。

版权所有 © 华为云计算技术有限公司

参数	描述	取值范围及默认值
ENCRYPTED | UNENCRYPTED	控制密码存储在系统表里的密码是否加密。（如果未显式指定，默认行为由配置参数password_encryption_type控制。） 按照产品安全要求，密码必须加密存储，因此UNENCRYPTED在DWS中禁止使用。由于系统无法对指定的加密密码字符串进行解密，如果当前的密码字符串已使用SHA256加密的格式，则无论是否声明ENCRYPTED或UNENCRYPTED，都将按原格式继续存储。这样就允许在dump/restore时重新加载已加密的密码。 ● password，登录密码。 ● DISABLE，默认情况下，用户可以更改自己的密码，除非密码被禁用。要禁用用户的密码，请指定DISABLE。禁用某个用户的密码后，将从系统中删除该密码，此类用户只能通过外部认证来连接数据库，例如：IAM认证、Kerberos或LDAP认证。只有管理员才能启用或禁用密码。普通用户不能禁用初始用户的密码。要启用密码，请运行ALTER USER并指定密码。	密码规则： ● 密码默认不少于8个字符。 ● 不能与用户名及用户名倒序相同。 ● 至少包含大写字母（A-Z）、小写字母（a-z）、数字（0-9）、非字母数字字符（~!@#$ %^&*()-_=+| [{}];:,<.>/?）四类字符中的三类字符。使用范围外的字符会收到告警，但依然允许创建。
VALID BEGIN	设置用户生效的时间戳。如果省略了该子句，用户无有效开始时间限制。	-
VALID UNTIL	设置用户失效的时间戳，即过了这个时间后，账户不可使用。如果省略了该子句，用户无有效结束时间限制。	-
RESOURCE POOL	设置用户使用的资源池名称，该名称属于系统表：pg_resource_pool	-
USER GROUP 'groupuser'	创建用户的子用户。	-
PERM SPACE	设置用户永久表存储空间限额。 space_limit，表示永久表存储空间上限。	space_limit，字符串格式为正整数+单位，单位当前支持K/M/G/T/P。 0表示不限制。
TEMP SPACE	设置用户临时表存储空间限额。 tmpspacelimit，表示临时表存储空间限额。	tmpspacelimit，字符串格式为正整数+单位，单位当前支持K/M/G/T/P。 0表示不限制。

版权所有 © 华为云计算技术有限公司

参数	描述	取值范围及默认值
SPILL SPACE	设置用户算子落盘空间限额。 spillspacelimit，表示算子落盘空间限额。	spillspacelimit，字符串格式为正整数+单位，单位当前支持K/M/G/T/P。 0表示不限制。
NODE GROUP	设置用户关联的逻辑集群名称。如果需要关联的逻辑集群名称包含大写字符或特殊字符，指定逻辑集群名称时需要加双引号。	-
ACCOUNT LOCK | ACCOUNT UNLOCK	锁定或解锁用户账号： • ACCOUNT LOCK：锁定账户，禁止登录数据库。 • ACCOUNT UNLOCK：解锁账户，允许登录数据库。	-
PGUSER	该属性用于兼容开源Postgres的连接通讯，开源的Postgres客户端接口（推荐使用Postgres 9.2.19版本的相关客户端接口）可以使用具有该属性的数据库用户连接数据库。 当前版本不允许修改用户的PGUSER属性。更多说明内容请参考**PGUSER参数建议**。	-
AUTHINFO 'authinfo'	该属性用于指定用户认证类型，authinfo为类型说明字符串，大小写敏感。 当前仅支持LDAP类型，对应的类型说明字符串为'ldap'。LDAP属于外部认证，故需要同时指定PASSWORD DISABLE。 • authinfo可以加上LDAP认证的额外信息，比如LDAP认证中的fulluser，fulluser等同于ldapprefix+username+ldapsuffix。当authinfo为'ldap'，表示用户认证类型为LDAP，此时ldapprefix和ldapsuffix的信息由pg_hba.conf中匹配的记录提供。 • ALTER ROLE时不允许用户切换认证类型，仅允许LDAP用户修改LDAP属性。	-
PASSWORD EXPIRATION period	设置用户的登录密码过期天数，登录密码过期之前用户需要及时修改密码。登录密码过期后用户无法登录，需要管理员设置新的登录密码后登录。	整数，-1~999。 • -1，表示没有过期限制，缺省值。 • 0，表示用户登录密码立即过期。

版权所有 © 华为云计算技术有限公司

参数	描述	取值范围及默认值
SET configuration_parameter	设置用户的会话参数。ALTER USER中修改的会话参数只针对指定的用户，且在下一次该用户启动的会话中有效。 • DEFAULT，表示清除configuration_parameter参数的值，configuration_parameter参数的值将继承本角色新产生的SESSION的默认值。 • FROM CURRENT，取当前会话中的值设置为configuration_parameter参数的值。	configuration_parameter和value的取值请参见**SET**。
RESET configuration_parameter/ALL	清除configuration_parameter参数的值。与SET configuration_parameter TO DEFAULT的效果相同。 ALL表示清除所有参数的值。	-

PGUSER 参数建议

PGUSER属性只用于兼容连接过程，而由于本产品与Postgres的内核差异导致的不兼容，不在此属性控制范围内。

由于具有PGUSER属性的用户的认证方式与其他用户不同，开源客户端的报错信息可能导致数据库用户PGUSER属性被枚举，建议使用本产品自有的客户端。例如：

```
#normaluser是不具有PGUSER属性的用户，psql是Postgres的客户端工具
pg@dws04:~> psql -d postgres -p 8000 -h 10.11.12.13 -U normaluser
psql: authentication method 10 not supported

#pguser用户是具有PGUSER属性的用户
pg@dws04:~> psql -d postgres -p 8000 -h 10.11.12.13 -U pguser
Password for user pguser:
```

示例

创建示例用户u1。

```
DROP USER IF EXISTS u1 CASCADE;
CREATE USER u1 PASSWORD '{Password}';
```

修改用户u1的登录密码。

```
ALTER USER u1 IDENTIFIED BY '{new_Password}' REPLACE '{Password}';
```

为用户u1追加CREATEROLE权限。

```
ALTER USER u1 CREATEROLE;
```

设置用户u1的最大数据库连接数为10。

```
ALTER USER u1 with  CONNECTION LIMIT 10;
```

将与用户u1关联的会话参数enable_seqscan值设置为on，设置成功后，在下一次启动的会话中生效。

```
ALTER USER u1 SET enable_seqscan TO on;
```


版权所有 © 华为云计算技术有限公司

重置u1的enable_seqscan参数。

```
ALTER USER u1 RESET enable_seqscan;
```

锁定u1账户。

```
ALTER USER u1 ACCOUNT LOCK;
```

相关链接

CREATE ROLE，**CREATE USER**，**DROP USER**

常见问题

账号被锁住了，如何解锁?

普通用户创建或删除GDS/OBS外表语句时报错，提示没有权限或权限不足

13.3.3 DROP USER

功能描述

删除用户，同时会删除同名的schema。

注意事项

- 须使用CASCADE级联删除依赖用户的对象（除数据库外）。当删除用户的级联对象时，如果级联对象处于锁定状态，则此级联对象无法被删除，直到对象被解锁或锁定级联对象的进程被终止。
- 在数据库中删除用户时，如果依赖用户的对象在其他数据库中或者依赖用户的对象是其他数据库，请用户先手动删除其他数据库中的依赖对象或直接删除依赖数据库，再删除用户。即drop user不支持跨数据库进行级联删除。
- 在多租户场景下，删除组用户时，业务用户也会同时被删除，如果指定CASCADE级联删除，那么删除业务用户时同时也指定CASCADE。如果在删除某个用户失败时，会报错，同时其他用户也无法成功删除。
- 如果用户A创建的GDS外表指定的错误表在用户B的schema下，则无法通过drop user指定CASCADE关键字直接删除用户B。
- DROP USER若提示role is being used by other users错误，可能原因为CLEAN CONNECTION过程存在线程无法及时响应信号，出现连接清理不完全的情况，需要再次执行CLEAN CONNECTION。

危险

- **DROP对象**操作（如DATABASE、USER/ROLE、SCHEMA、TABLE、VIEW等对象）存在数据丢失风险，尤其含带CASCADE级联删除场景，会将关联的对象一并删除，操作需谨慎，操作前需考虑数据备份。
- 更多开发设计规范参见**总体开发设计规范**。

语法格式

```
DROP USER [ IF EXISTS ] user_name [, ...] [ CASCADE | RESTRICT ];
```


版权所有 © 华为云计算技术有限公司

参数说明

表 13-36 DROP USER 参数说明

参数	描述	取值范围
IF EXISTS	如果指定的用户不存在，发出一个notice而不是抛出一个错误。	-
user_name	待删除的用户名。	已存在的用户名。
CASCADE | RESTRICT	当执行删除操作时，如何处理依赖对象中的相关数据。 **说明** 在DWS中，存在一个配置参数enable_kill_query，此参数在配置文件postgresql.conf中。此参数影响级联删除用户对象的行为： • 当参数enable_kill_query为on ，且使用CASCADE模式删除用户时，会自动kill锁定用户级联对象的进程，并删除用户。 • 当参数enable_kill_query为off，且使用CASCADE模式删除用户时，会等待锁定级联对象的进程结束之后再删除用户。	• CASCADE：级联删除依赖用户的表等对象。级联删除用户的时候，会删除掉owner是这个用户的对象，并清理掉其他对象对这个用户的授权信息。 • RESTRICT：如果用户还有任何依赖的对象，则拒绝删除该用户（缺省行为）。

示例

删除用户jim：

```
DROP USER IF EXISTS jim CASCADE;
```

相关链接

ALTER USER，**CREATE USER**

13.3.4 CREATE ROLE

功能描述

创建角色。

角色是拥有数据库对象和权限的实体。在不同的环境中角色可以认为是一个用户，一个组或者兼顾两者。

注意事项

- 在数据库中添加一个新角色，角色无登录权限。
- 创建角色的用户必须具备CREATE ROLE的权限或者是系统管理员。
- 使用CREATE ROLE创建的角色必须使用ALTER ROLE语法进行修改，不能使用ALTER USER。

版权所有 © 华为云计算技术有限公司

语法格式

```
CREATE ROLE role_name [ [ WITH ] option [ ... ] ] [ ENCRYPTED | UNENCRYPTED ] { PASSWORD |
IDENTIFIED BY } { 'password' | DISABLE };
```

其中角色信息设置子句option语法为：

```
{SYSADMIN | NOSYSADMIN}
    | {AUDITADMIN | NOAUDITADMIN}
    | {CREATEDB | NOCREATEDB}
    | {USEFT | NOUSEFT}
    | {CREATEROLE | NOCREATEROLE}
    | {INHERIT | NOINHERIT}
    | {LOGIN | NOLOGIN}
    | {REPLICATION | NOREPLICATION}
    | {INDEPENDENT | NOINDEPENDENT}
    | {VCADMIN | NOVCADMIN}
    | CONNECTION LIMIT connlimit
    | VALID BEGIN 'timestamp'
    | VALID UNTIL 'timestamp'
    | RESOURCE POOL 'respool'
    | USER GROUP 'groupuser'
    | PERM SPACE 'spacelimit'
    | TEMP SPACE 'tmpspacelimit'
    | SPILL SPACE 'spillspacelimit'
    | NODE GROUP logic_cluster_name
    | IN ROLE role_name [, ...]
    | IN GROUP role_name [, ...]
    | ROLE role_name [, ...]
    | ADMIN role_name [, ...]
    | USER role_name [, ...]
    | SYSID uid
    | DEFAULT TABLESPACE tablespace_name
    | PROFILE DEFAULT
    | PROFILE profile_name
    | PGUSER
    | AUTHINFO 'authinfo'
    | PASSWORD EXPIRATION period
```

参数说明

表 13-37 CREATE ROLE 参数说明

参数	描述	取值范围
role_name	角色名称。	字符串，要符合标识符的命名规范。且最多为63个字节。参见**标识符命名规范**。

 版权所有 © 华为云计算技术有限公司

参数	描述	取值范围
password	登录密码。	密码规则如下： ● 密码默认不少于8个字符。 ● 不能与用户名及用户名倒序相同。 ● 至少包含大写字母（A-Z），小写字母（a-z），数字（0-9），非字母数字字符（~!@#$%^&*()-_=+|[{}];:,<.>/?）四类字符中的三类字符。使用范围外的字符会收到告警，但依然允许创建。 取值范围：字符串。
DISABLE	默认情况下，用户可以更改自己的密码，除非密码被禁用。要禁用用户的密码，请指定DISABLE。 禁用某个用户的密码后，将从系统中删除该密码，此类用户只能通过外部认证来连接数据库，例如：IAM认证、Kerberos、LDAP认证或OneAccess认证。只有管理员才能启用或禁用密码。普通用户不能禁用初始用户的密码。要启用密码，请运行ALTER USER并指定密码。	-
ENCRYPTED | UNENCRYPTED	控制密码存储在系统表里的密码是否加密。（如果未显示指定，默认行为由配置参数password_encryption_type控制。） 按照产品安全要求，密码必须加密存储，因此UNENCRYPTED在DWS中禁止使用。 由于系统无法对指定的加密密码字符串进行解密，如果当前的密码字符串已使用SHA256加密的格式，则无论是否声明ENCRYPTED或UNENCRYPTED，都将按原格式继续存储。这样就允许在dump/restore时重新加载已加密的密码。	-
SYSADMIN | NOSYSADMIN	决定一个新角色是否为“系统管理员”，具有SYSADMIN属性的角色拥有系统最高权限。	默认为NOSYSADMIN。

版权所有 © 华为云计算技术有限公司

参数	描述	取值范围
AUDITADMIN | NOAUDITADMIN	定义角色是否有审计管理属性。	默认为NOAUDITADMIN。
CREATEDB | NOCREATEDB	决定一个新角色是否能创建数据库。新角色没有创建数据库的权限。	默认为NOCREATEDB。
USEFT | NOUSEFT	决定一个新角色是否能操作外表，包括：新建外表、删除外表、修改外表、读写外表。 新角色没有操作外表的权限。	缺省为NOUSEFT。 **说明** 8.2.0及以上集群版本中，若安全模式下GUC参数security_enable_options的foreign_table_options选项打开，即表示允许使用外表操作功能，不需要再赋予用户useft权限。
CREATEROLE | NOCREATEROLE	决定一个角色是否可以创建新角色（也就是执行CREATE ROLE和CREATE USER）。 一个拥有CREATEROLE权限的角色也可以修改和删除其他角色。	默认为NOCREATEROLE。
INHERIT | NOINHERIT	这些子句决定一个角色是否“继承”它所在组的角色的权限。不推荐使用。	-
LOGIN | NOLOGIN	具有LOGIN属性的角色才可以登录数据库。一个拥有LOGIN属性的角色可以认为是一个用户。	默认为NOLOGIN。
REPLICATION | NOREPLICATION	定义角色是否允许流复制或设置系统为备份模式。REPLICATION属性是特定的角色，仅用于复制。	默认为NOREPLICATION。

版权所有 © 华为云计算技术有限公司

参数	描述	取值范围
INDEPENDENT | NOINDEPENDENT	定义私有、独立的角色。具有INDEPENDENT属性的角色，管理员对其进行的控制、访问的权限被分离。 具体规则如下： ● 未经INDEPENDENT角色授权，管理员无权对其表对象进行增、删、查、改、拷贝、授权操作。 ● 未经INDEPENDENT角色授权，管理员无权修改INDEPENDENT角色的继承关系。 ● 管理员无权修改INDEPENDENT角色的表对象的属主。 ● 管理员无权去除INDEPENDENT角色的INDEPENDENT属性。 ● 管理员无权修改INDEPENDENT角色的数据库密码，INDEPENDENT角色需管理好自身密码，密码丢失无法重置。 ● 管理员属性用户不允许定义修改为INDEPENDENT属性。	-
VCADMIN | NOVCADMIN	定义逻辑集群管理员角色。具有逻辑集群管理员属性的角色，和普通用户相比，有如下额外权限： ● 在所关联逻辑集群中创建、修改和删除资源池的权限。 ● 将所关联的逻辑集群的访问权限授予其他用户或角色，或回收其他用户或角色对关联逻辑集群的访问权限。	-
CONNECTION LIMIT	声明该角色在单个CN上可以使用的并发连接数量。 为保证集群正常使用，connection limit的最小值是集群中CN的数目。在集群执行ANALYZE时，其他CN节点会连接当前执行ANALYZE的CN节点来同步元数据。例如集群中有3个CN节点，那么connection limit应该设置为>=3。	整数，>=-1，缺省值为-1，表示没有限制。
VALID BEGIN	设置角色生效的时间戳。如果省略了该子句，角色无有效开始时间限制。	-

版权所有 © 华为云计算技术有限公司

参数	描述	取值范围
VALID UNTIL	设置角色失效的时间戳。如果省略了该子句，角色无有效结束时间限制。	-
RESOURCE POOL	设置角色使用的resource pool名字，该名字属于系统表：pg_resource_pool。	-
USER GROUP 'groupuser'	创建一个user的子用户。	-
PERM SPACE	设置用户永久表存储空间限额。	space_limit：永久表存储空间上限。取值范围：字符串格式为正整数+单位，单位当前支持K/M/G/T/P。0表示不限制。
TEMP SPACE	设置用户临时表存储空间限额。	tmpspacelimit：临时表存储空间限额。取值范围：字符串格式为正整数+单位，单位当前支持K/M/G/T/P。0表示不限制。
SPILL SPACE	设置用户算子落盘空间限额。	spillspacelimit：算子落盘空间限额。取值范围：字符串格式为正整数+单位，单位当前支持K/M/G/T/P。0表示不限制。
NODE GROUP	设置用户关联的逻辑集群名称。如果需要关联的逻辑集群名称包含大写字符或特殊字符，指定逻辑集群名称时需要加双引号。	-
IN ROLE	新角色立即拥有IN ROLE子句中列出的一个或多个现有角色拥有的权限。不推荐使用。	-
IN GROUP	IN GROUP是IN ROLE废弃的拼写方式。不推荐使用。	-
ROLE	ROLE子句列出一个或多个现有的角色，它们将自动添加为这个新角色的成员，拥有新角色所有的权限。	-
ADMIN	ADMIN子句类似ROLE子句，不同的是ADMIN后的角色可以把新角色的权限赋予其他角色。	-
USER	USER子句是ROLE子句废弃的拼写方式。	-

版权所有 © 华为云计算技术有限公司

参数	描述	取值范围
SYSID	SYSID子句将被忽略，无实际意义。	-
DEFAULT TABLESPACE	DEFAULT TABLESPACE子句将被忽略，无实际意义。	-
PROFILE	PROFILE子句将被忽略，无实际意义。	-
PGUSER	该属性用于兼容开源Postgres的连接通讯，开源的Postgres客户端接口（推荐使用Postgres 9.2.19版本的相关客户端接口）可以使用具有该属性的数据库用户连接数据库。 该属性只用于兼容连接过程，而由于本产品与Postgres的内核差异导致的不兼容，不在此属性控制范围内。	由于具有PGUSER属性的用户的认证方式与其他用户不同，开源客户端的报错信息可能导致数据库用户PGUSER属性被枚举，建议使用本产品自有的客户端。 #normaluser是不具有PGUSER属性的用户，psql是Postgres的客户端工具 pg@dws04:~> psql -d postgres -p 8000 -h 10.11.12.13 -U normaluser psql: authentication method 10 not supported #pguser用户是具有PGUSER属性的用户 pg@dws04:~> psql -d postgres -p 8000 -h 10.11.12.13 -U pguser Password for user pguser:

 版权所有 © 华为云计算技术有限公司

参数	描述	取值范围
AUTHINFO 'authinfo'	该属性用于指定角色认证类型，authinfo为类型说明字符串，大小写敏感。当前仅支持LDAP和OneAccess类型，对应的类型说明字符串为“ldap”和“oneaccess”。LDAP和OneAccess属于外部认证，故需要同时指定PASSWORD DISABLE。	• authinfo可以加上LDAP认证的额外信息，比如LDAP认证中的fulluser，fulluser等同于ldapprefix +username+ldapsuffix。当authinfo为“ldap”，表示角色认证类型为LDAP，此时ldapprefix和ldapsuffix的信息由pg_hba.conf中匹配的记录提供。 • 在支持OneAccess认证方式时，authinfo必须有oneaccessClientId和domain信息，满足'oneaccessClientId=xxxx,domain=xxxx' 格式。oneaccessClientId是oneaccess用户的ID，最大长度为256字符；domain是oneaccess的访问域名，最大长度为64字符。 • ALTER ROLE时不允许用户切换认证类型，仅允许LDAP用户修改LDAP属性或OneAccess用户修改OneAccess属性。 • OneAccess是通过内部代码逻辑识别的，无需配置pg_hba.conf，认证信息由authinfo给出。
PASSWORD EXPIRATION period	声明该角色的登录密码过期天数，登录密码过期之前用户需要及时修改密码。登录密码过期后用户无法登录，需要管理员设置新的登录密码后登录。	整数，-1~999。缺省值为-1，表示没有过期限制；0表示用户登录密码立即过期。

示例

创建一个角色，名为manager：

```
CREATE ROLE manager IDENTIFIED BY '{password}';
```

创建一个角色，从2015年1月1日开始生效，到2026年1月1日失效：

```
CREATE ROLE miriam WITH LOGIN PASSWORD '{password}' VALID BEGIN '2015-01-01' VALID UNTIL '2026-01-01';
```

创建一个角色，认证类型是LDAP，LDAP认证的其他信息由pg_hba.conf提供：

```
CREATE ROLE role1 WITH LOGIN AUTHINFO 'ldap' PASSWORD DISABLE;
```

 版权所有 © 华为云计算技术有限公司

创建一个角色，认证类型是LDAP，LDAP认证的fulluser信息在创建时指定，此时ldap大小写敏感，需要用单引号包含：

```
CREATE ROLE role2 WITH LOGIN AUTHINFO 'ldapcn=role2,cn=user,dc=lework,dc=com' PASSWORD DISABLE;
```

创建一个角色，认证类型是OneAccess，OneAccess认证的authinfo信息在创建时指定，此时oneaccess大小写敏感，需要用单引号包括：

```
CREATE ROLE role3 WITH LOGIN AUTHINFO
'oneaccessClientId=AbCd123,domain=example.exampleoneaccess.com' PASSWORD DISABLE;
```

创建一个角色，登录密码有效期是30天：

```
CREATE ROLE role4 WITH LOGIN PASSWORD '{password}' PASSWORD EXPIRATION 30;
```

相关链接

SET ROLE，**ALTER ROLE**，**DROP ROLE**，**GRANT**，**REVOKE**

13.3.5 ALTER ROLE

功能描述

修改角色属性。

注意事项

无。

语法格式

- 修改角色的权限。

  ```
  ALTER ROLE role_name [ [ WITH ] option [ ... ] ];
  ```

  其中权限项子句option为：

  ```
  {CREATEDB | NOCREATEDB}
    | {CREATEROLE | NOCREATEROLE}
    | {INHERIT | NOINHERIT}
    | {AUDITADMIN | NOAUDITADMIN}
    | {SYSADMIN | NOSYSADMIN}
    | {USEFT | NOUSEFT}
    | {LOGIN | NOLOGIN}
    | {REPLICATION | NOREPLICATION}
    | {INDEPENDENT | NOINDEPENDENT}
    | {VCADMIN | NOVCADMIN}
    | CONNECTION LIMIT connlimit
    | [ ENCRYPTED | UNENCRYPTED ] PASSWORD 'password'
    | [ ENCRYPTED | UNENCRYPTED ] IDENTIFIED BY 'password' [ REPLACE 'old_password' ]
    | [ ENCRYPTED | UNENCRYPTED ] PASSWORD { 'password' | DISABLE }
    | [ ENCRYPTED | UNENCRYPTED ] IDENTIFIED BY { 'password' [ REPLACE 'old_password' ] |
  DISABLE }
    | VALID BEGIN 'timestamp'
    | VALID UNTIL 'timestamp'
    | RESOURCE POOL 'respool'
    | USER GROUP 'groupuser'
    | PERM SPACE 'spacelimit'
    | TEMP SPACE 'tmpspacelimit'
    | SPILL SPACE 'spillspacelimit'
    | NODE GROUP logic_cluster_name
    | ACCOUNT { LOCK | UNLOCK }
    | PGUSER
    | AUTHINFO 'authinfo'
    | PASSWORD EXPIRATION period
  ```

 版权所有 © 华为云计算技术有限公司

- 修改角色的名称。
```
ALTER ROLE role_name
    RENAME TO new_name;
```
- 设置角色的配置参数。
```
ALTER ROLE role_name [ IN DATABASE database_name ]
    SET configuration_parameter {{ TO | = } { value | DEFAULT } | FROM CURRENT};
```
- 重置角色的配置参数。
```
ALTER ROLE role_name
    [ IN DATABASE database_name ] RESET {configuration_parameter|ALL};
```

参数说明

表 13-38 ALTER ROLE 参数说明

参数	描述	取值范围
role_name	需要修改的角色名。	已存在的角色名。
new_name	角色的新名称。	字符串，需符合**标识符命名规范**。
CREATEDB | NOCREATEDB	决定一个新角色是否能创建数据库。缺省为NOCREATEDB。 新角色没有创建数据库的权限。	-
CREATEROLE | NOCREATEROLE	决定一个角色是否可以创建新角色（也就是执行CREATE ROLE和CREATE USER）。 缺省为NOCREATEROLE。 拥有CREATEROLE权限的角色也可以修改和删除其他角色。	-
INHERIT | NOINHERIT	决定一个角色是否“继承”它所在组的角色的权限。不推荐使用。	-
AUDITADMIN | NOAUDITADMIN	定义角色是否有审计管理属性。 缺省为NOAUDITADMIN。	-
SYSADMIN | NOSYSADMIN	决定一个新角色是否为“系统管理员”，具有SYSADMIN属性的角色拥有系统最高权限。 缺省为NOSYSADMIN。	-
USEFT | NOUSEFT	决定一个新角色是否能操作外表，包括：新建外表、删除外表、修改外表、读写外表。 新角色没有操作外表的权限。 缺省为NOUSEFT，表示新建角色默认不能操作外表。	-

版权所有 © 华为云计算技术有限公司

参数	描述	取值范围
LOGIN | NOLOGIN	定义用户是否可以登录数据库。具有LOGIN属性的角色才可以登录数据库。一个拥有LOGIN属性的角色可以认为是一个用户。 缺省为NOLOGIN。	-
REPLICATION | NOREPLICATION	定义角色是否允许流复制或设置系统为备份模式。REPLICATION属性是特定的角色，仅用于复制。 缺省为NOREPLICATION。	-
INDEPENDENT | NOINDEPENDENT	定义私有、独立的角色。具有INDEPENDENT属性的角色，管理员对其进行的控制、访问的权限被分离，具体规则如下： ● 未经INDEPENDENT角色授权，管理员无权对其表对象进行增、删、查、改、拷贝、授权操作。 ● 未经INDEPENDENT角色授权，管理员无权修改INDEPENDENT角色的继承关系。 ● 管理员无权修改INDEPENDENT角色的表对象的属主。 ● 管理员无权去除INDEPENDENT角色的INDEPENDENT属性。 ● 管理员无权修改INDEPENDENT角色的数据库密码，INDEPENDENT角色需管理好自身密码，密码丢失无法重置。 ● 管理员属性用户不允许定义修改为INDEPENDENT属性。	-
VCADMIN | NOVCADMIN	定义逻辑集群管理员角色。具有逻辑集群管理员属性的角色，和普通用户相比，有如下额外权限： ● 在所关联逻辑集群中创建、修改和删除资源池的权限。 ● 将所关联的逻辑集群的访问权限授予其他用户或角色，或回收其他用户或角色对关联逻辑集群的访问权限。	-
CONNECTION LIMIT	定义角色在单个CN上可以使用的最大并发连接数量。	整数，大于等于-1。 缺省值为-1，表示没有限制。

 版权所有 © 华为云计算技术有限公司

参数	描述	取值范围
ENCRYPTED | UNENCRYPTED	控制密码存储在系统表里的密码是否加密。（如果未显式指定，默认行为由配置参数password_encryption_type控制。） 按照产品安全要求，密码必须加密存储，因此UNENCRYPTED在DWS中禁止使用。由于系统无法对指定的加密密码字符串进行解密，如果当前的密码字符串已使用SHA256加密的格式，则无论是否声明ENCRYPTED或UNENCRYPTED，都将按原格式继续存储。这样就允许在dump/restore时重新加载已加密的密码。 ● password 登录密码。 ● DISABLE 默认情况下，用户可以更改自己的密码，除非密码被禁用。要禁用用户的密码，请指定DISABLE。禁用某个用户的密码后，将从系统中删除该密码，此类用户只能通过外部认证来连接数据库，例如：IAM认证、kerberos或ldap认证。只有管理员才能启用或禁用密码。普通用户不能禁用初始用户的密码。要启用密码，请运行ALTER USER并指定密码。	密码规则： ● 密码默认不少于8个字符。 ● 不能与用户名及用户名倒序相同。 ● 至少包含大写字母（A-Z），小写字母（a-z），数字（0-9），非字母数字字符（~!@#$%^&*()-_=+\\|[{}];:,<.>/?）四类字符中的三类字符。使用范围外的字符会收到告警，但依然允许创建。
VALID BEGIN	设置角色生效的时间戳。如果省略了该子句，角色无有效开始时间限制。	-
VALID UNTIL	设置角色失效的时间戳。如果省略了该子句，角色无有效结束时间限制。	-
RESOURCE POOL	设置角色使用的resource pool名称，该名称属于系统表：pg_resource_pool	-
USER GROUP 'groupuser'	创建一个user的子用户。	-
PERM SPACE	设置用户永久表存储空间限额。 space_limit表示永久表存储空间上限。	space_limit，字符串格式为正整数+单位，单位当前支持K/M/G/T/P。0表示不限制。
TEMP SPACE	设置用户临时表存储空间限额。 tmpspacelimit表示临时表存储空间限额。	tmpspacelimit，字符串格式为正整数+单位，单位当前支持K/M/G/T/P。0表示不限制。

版权所有 © 华为云计算技术有限公司

参数	描述	取值范围
SPILL SPACE	设置用户算子落盘空间限额。 spillspacelimit：算子落盘空间限额。	spillspacelimit，字符串格式为正整数+单位，单位当前支持K/M/G/T/P。0表示不限制。
NODE GROUP	设置用户关联的逻辑集群名称。如果需要关联的逻辑集群名称包含大写字符或特殊字符，指定逻辑集群名称时需要加双引号。	-
ACCOUNT LOCK | ACCOUNT UNLOCK	• ACCOUNT LOCK：锁定账户，禁止登录数据库。 • ACCOUNT UNLOCK：解锁账户，允许登录数据库。	-
PGUSER	该属性用于兼容开源Postgres的连接通讯，开源的Postgres客户端接口（推荐使用Postgres 9.2.19版本的相关客户端接口）可以使用具有该属性的数据库用户连接数据库。 当前版本不允许修改角色的PGUSER属性。更多说明内容请参考**PGUSER参数建议**。	-
AUTHINFO 'authinfo'	该属性用于指定角色认证类型，authinfo为类型说明字符串，大小写敏感。当前仅支持LDAP类型，对应的类型说明字符串为'ldap'。LDAP属于外部认证，故需要同时指定PASSWORD DISABLE。 • authinfo可以加上LDAP认证的额外信息，比如LDAP认证中的fulluser，fulluser等同于ldapprefix+username+ldapsuffix。当authinfo为'ldap'，表示角色认证类型为LDAP，此时ldapprefix和ldapsuffix的信息由pg_hba.conf中匹配的记录提供。 • ALTER ROLE时不允许用户切换认证类型，仅允许LDAP用户修改LDAP属性。	-
PASSWORD EXPIRATION period	声明该角色的登录密码过期天数，登录密码过期之前用户需要及时修改密码。登录密码过期后用户无法登录，需要管理员设置新的登录密码后登录。	整数，-1~999。 • -1，表示没有过期限制，缺省值。 • 0，表示用户登录密码立即过期。

版权所有 © 华为云计算技术有限公司

参数	描述	取值范围
IN DATABASE database_name	表示修改角色在指定数据库上的参数。	-
SET configuration_parameter	设置角色的参数。ALTER ROLE中修改的会话参数只针对指定的角色，且在下一次该角色启动的会话中有效。 • DEFAULT，表示清除configuration_parameter参数的值，configuration_parameter参数的值将继承本角色新产生的SESSION的默认值。 • FROM CURRENT，表示取当前会话中的值设置为configuration_parameter参数的值。	configuration_parameter和value的取值请参见**SET**。
RESET configuration_parameter/ALL	清除configuration_parameter参数的值。与SET configuration_parameter TO DEFAULT的效果相同。 ALL表示清除所有参数的值。	-

PGUSER 参数建议

PGUSER属性只用于兼容连接过程，而由于本产品与Postgres的内核差异导致的不兼容，不在此属性控制范围内。

由于具有PGUSER属性的用户的认证方式与其他用户不同，开源客户端的报错信息可能导致数据库用户PGUSER属性被枚举，建议使用本产品自有的客户端。例如：

```
#normaluser是不具有PGUSER属性的用户，psql是Postgres的客户端工具
pg@dws04:~> psql -d postgres -p 8000 -h 10.11.12.13 -U normaluser
psql: authentication method 10 not supported

#pguser用户是具有PGUSER属性的用户
pg@dws04:~> psql -d postgres -p 8000 -h 10.11.12.13 -U pguser
Password for user pguser:
```

示例

创建示例角色r1，r2和r3。

```
CREATE ROLE r1 IDENTIFIED BY '{Password}';
CREATE ROLE r2 WITH LOGIN AUTHINFO 'ldapcn=r2,cn=user,dc=lework,dc=com' PASSWORD DISABLE;
CREATE ROLE r3 WITH LOGIN PASSWORD '{Password}' PASSWORD EXPIRATION 30;
```

修改角色r1的登录权限：

```
ALTER ROLE r1 login;
```

修改角色r1的密码：

```
ALTER ROLE r1 IDENTIFIED BY '{new_Password}' REPLACE '{Password}';
```

修改角色manager为系统管理员：

版权所有 © 华为云计算技术有限公司

```
ALTER ROLE r1 SYSADMIN;
```

修改LDAP认证角色的fulluser信息：

```
ALTER ROLE r2 WITH LOGIN AUTHINFO 'ldapcn=role2,cn=user2,dc=func,dc=com' PASSWORD DISABLE;
```

修改角色的登录密码有效期为90天：

```
ALTER ROLE r3 PASSWORD EXPIRATION 90;
```

相关链接

CREATE ROLE，**DROP ROLE**，**SET**

13.3.6 SET ROLE

功能描述

设置当前会话的当前用户标识符，即在当前数据库会话中切换或启用指定角色。

注意事项

- 当前会话的用户必须是指定的role_name角色的成员，系统管理员可以切换至任意角色。
- 可以使用SET ROLE添加权限或限制某个用户的权限。如果会话用户的角色具有INHERITS属性，则它将自动拥有SET ROLE所设置角色的所有权限；在这种情况下，SET ROLE实际上是删除了所有直接赋予会话用户的权限，以及它的所属角色的权限，仅保留指定角色的权限。另一方面，如果会话用户的角色具有NOINHERITS属性，则SET ROLE将删除直接赋予会话用户的权限，而获取指定角色的权限。

语法格式

- 在当前数据库会话中切换或启用指定角色。
  ```
  SET [ SESSION | LOCAL ] ROLE role_name PASSWORD '{password}';
  ```
- 重置当前会话角色。
  ```
  RESET ROLE;
  ```

参数说明

表 13-39 SET ROLE 参数说明

参数	描述	取值范围
SESSION	声明这个命令只对当前会话生效（缺省值）。	-
LOCAL	声明该命令仅对当前事务生效。	-
role_name	角色名称。	已有的角色名称。
password	角色的密码。	-
RESET ROLE	用于重置当前用户标识。	-

版权所有 © 华为云计算技术有限公司

示例

查看当前会话用户。

```
SELECT CURRENT_USER;
 current_user
--------------
 dbadmin
(1 row)
```

创建名称为role_new的角色，并切换当前会话用户为role_new。

```
CREATE ROLE role_new IDENTIFIED BY '{Password}';
SET ROLE role_new PASSWORD '{password}';
```

查看当前会话用户，显示为role_new。

```
SELECT CURRENT_USER;
 current_user
--------------
 role_new
(1 row)
```

将当前用户重置为原会话用户。

```
RESET role;
```

13.3.7 DROP ROLE

功能描述

删除指定的角色。

注意事项

DROP ROLE若提示role is being used by other users错误，可能原因为CLEAN CONNECTION过程存在线程无法及时响应信号，出现连接清理不完全的情况，需要再次执行CLEAN CONNECTION。

危险

- **DROP对象**操作（如DATABASE、USER/ROLE、SCHEMA、TABLE、VIEW等对象）存在数据丢失风险，尤其含带CASCADE级联删除场景，会将关联的对象一并删除，操作需谨慎，操作前需考虑数据备份。
- 更多开发设计规范参见**总体开发设计规范**。

语法格式

```
DROP ROLE [ IF EXISTS ] role_name [, ...];
```


版权所有 © 华为云计算技术有限公司

参数说明

表 13-40 DROP ROLE 参数说明

参数	描述	取值范围
IF EXISTS	如果指定的角色不存在，则发出一个notice而不是抛出一个错误。	-
role_name	要删除的角色名称。	已存在的角色。

示例

删除角色manager：

```
DROP ROLE IF EXISTS manager;
```

相关链接

CREATE ROLE，**ALTER ROLE**，**SET ROLE**

13.3.8 CREATE GROUP

功能描述

创建一个新用户组。

注意事项

CREATE GROUP是CREATE ROLE的别名，非SQL标准语法，不推荐使用，建议用户直接使用CREATE ROLE替代。

语法格式

```
CREATE GROUP group_name [ [ WITH ] option [ ... ] ]
    [ ENCRYPTED | UNENCRYPTED ] { PASSWORD | IDENTIFIED BY } { 'password' | DISABLE };
```

其中可选项action子句语法为：

```
where option can be:
{SYSADMIN | NOSYSADMIN}
    | {AUDITADMIN | NOAUDITADMIN}
    | {CREATEDB | NOCREATEDB}
    | {USEFT | NOUSEFT}
    | {CREATEROLE | NOCREATEROLE}
    | {INHERIT | NOINHERIT}
    | {LOGIN | NOLOGIN}
    | {REPLICATION | NOREPLICATION}
    | {INDEPENDENT | NOINDEPENDENT}
    | {VCADMIN | NOVCADMIN}
    | CONNECTION LIMIT connlimit
    | VALID BEGIN 'timestamp'
    | VALID UNTIL 'timestamp'
    | RESOURCE POOL 'respool'
    | USER GROUP 'groupuser'
    | PERM SPACE 'spacelimit'
    | NODE GROUP logic_group_name
```

 版权所有 © 华为云计算技术有限公司

```
| IN ROLE role_name [, ...]
| IN GROUP role_name [, ...]
| ROLE role_name [, ...]
| ADMIN role_name [, ...]
| USER role_name [, ...]
| SYSID uid
| DEFAULT TABLESPACE tablespace_name
| PROFILE DEFAULT
| PROFILE profile_name
| PGUSER
```

参数说明

请参考CREATE ROLE的**参数说明**。

相关链接

ALTER GROUP，**DROP GROUP**，**CREATE ROLE**

13.3.9 ALTER GROUP

功能描述

修改用户组的属性。

注意事项

ALTER GROUP是ALTER ROLE的别名，非SQL标准语法，不推荐使用，建议用户直接使用ALTER ROLE替代。

语法格式

- 向用户组中添加用户。
  ```
  ALTER GROUP group_name
    ADD USER user_name [, ... ];
  ```
- 从用户组中删除用户。
  ```
  ALTER GROUP group_name
    DROP USER user_name [, ... ];
  ```
- 修改用户组的名称。
  ```
  ALTER GROUP group_name
    RENAME TO new_name;
  ```

参数说明

更多参数信息，请参考ALTER ROLE的**参数说明**。

表 13-41 ALTER GROUP 参数说明

参数	描述	取值范围
group_name	需要修改的用户组名称。	字符串，已存在的组名称。
user_name	要添加到组或从组中删除的用户。	已存在的用户名。
new_name	组的新名称。	字符串，需符合**标识符命名规范**。

版权所有 © 华为云计算技术有限公司

相关链接

CREATE GROUP，DROP GROUP，ALTER ROLE

13.3.10 DROP GROUP

功能描述

在数据库运维管理中，用户组用于将具有相同权限属性的用户归类管理，便于批量授权和权限维护。随着业务调整和组织架构变化，某些用户组可能不再使用——例如项目解散、权限模型重构，或者组内用户已全部迁移至其他用户组。当需要删除不再使用的用户组以清理权限资源时，可使用DROP GROUP命令。

DROP GROUP是DROP ROLE的别名。两者功能和效果完全一致。

注意事项

DROP GROUP是集群管理工具封装的内部接口，用来实现集群管理。该接口不建议用户直接使用，以免对集群状态造成影响。

语法格式

```
DROP GROUP [ IF EXISTS ] group_name [, ...];
```

参数说明

请参见DROP ROLE的参数说明。

示例

以下示例演示如何删除用户组，并验证删除结果。

步骤1 使用CREATE ROLE语法创建一个测试用户组dev_group。

```
CREATE ROLE dev_group IDENTIFIED BY '{password}';
CREATE GROUP dev_group IDENTIFIED BY 'Gauss_202611';  --与CREATE ROLE等价
```

步骤2 查询当前存在的用户组，确认要删除的用户组名称。

```
SELECT rolname FROM pg_roles WHERE rolname = 'dev_group';
```

步骤3 删除用户组dev_group（使用IF EXISTS避免用户组不存在时报错）。

```
DROP GROUP IF EXISTS dev_group;
```

步骤4 验证用户组是否已成功删除。

```
SELECT rolname FROM pg_roles WHERE rolname = 'dev_group';
```

预期输出（查询返回空结果，表示用户组已删除）：

```
 rolname
---------
(0 rows)
```

如需同时删除多个用户组，可使用逗号分隔：

```
DROP GROUP IF EXISTS dev_group,test_group;
```

----结束

版权所有 © 华为云计算技术有限公司

相关链接

CREATE GROUP，ALTER GROUP，DROP ROLE

常见问题

- **Q：删除用户组会删除组内的用户吗?**

  A：不会。DROP GROUP仅删除用户组本身，组内的用户账号和独立权限不受影响，用户仍可正常登录数据库。

- **Q：DROP GROUP和DROP ROLE有什么区别?**

  A：没有区别。DROP GROUP是DROP ROLE的别名，两者功能和效果完全一致。DROP GROUP主要用于兼容PostgreSQL语法。

13.3.11 DROP OWNED

功能描述

在数据库的日常运维中，管理员经常需要进行角色清理或迁移操作。例如，当某个应用下线、员工离职或系统迁移时，需要删除特定角色拥有的所有数据库对象（如表、视图、序列等）并撤销其在共享对象上的权限。此时，可使用DROP OWNED命令一次性完成这些操作，避免逐个对象手动删除的繁琐过程。

DROP OWNED命令用于删除一个数据库角色所拥有的数据库对象，并撤销该角色在当前数据库和共享对象（数据库、表空间）上的所有权限。

DROP OWNED与DROP ROLE是角色管理中密切相关的命令，其适用场景对比如下：

命令	作用	对象处理方式	角色本身	典型场景
DROP OWNED	删除角色拥有的所有对象并撤销其权限。	删除对象。	仍然存在。	角色清理：需要删除对象且不再需要这些对象。
DROP ROLE	直接删除角色本身。	如果角色仍拥有对象则报错。	删除角色。	角色删除：需先通过DROP OWNED清理对象后才能执行。

注意事项

- 所有该角色在当前数据库里和共享对象（数据库，表空间）上的所有对象上的权限都将被撤销。
- DROP OWNED仅删除角色拥有的对象和权限，**角色本身不会被删除**，仍保留在系统中。如需删除角色本身，需在执行DROP OWNED后配合使用DROP ROLE命令。
- 使用CASCADE选项时会级联删除所有依赖对象，可能影响其他角色的对象，请谨慎使用。

版权所有 © 华为云计算技术有限公司

语法格式

```
DROP OWNED BY name [, ...] [ CASCADE | RESTRICT ];
```

参数说明

表 13-42 DROP OWNED 参数说明

参数	描述	取值范围
name	将要删除所拥有对象并且撤销其权限的角色名。	已存在的角色名。
CASCADE | RESTRICT	当执行删除操作时，如何处理依赖对象中的相关数据。	• CASCADE：级联删除所有依赖于被删除对象的对象。 • RESTRICT：缺省值，拒绝删除那些有任何依赖对象存在的对象。

示例

创建角色u1并创建一个由u1拥有的表。

```
CREATE ROLE u1 IDENTIFIED BY '{password}';
CREATE TABLE u1_table (id int);
ALTER TABLE u1_table OWNER TO u1;
```

删除名称为u1的角色拥有的所有数据库对象。

```
DROP OWNED BY u1;
```

相关文档

DROP ROLE

常见问题

执行 DROP ROLE 报错“role "xx" cannot be dropped because some objects depend on it”

原因：角色仍拥有数据库对象，直接执行DROP ROLE会被拒绝。

解决方法：先执行DROP OWNED BY 角色名 CASCADE;清除所有对象和权限，然后再执行DROP ROLE。

执行 DROP OWNED 后发现其他数据库中的对象未被删除

原因：DROP OWNED仅作用于当前连接的数据库，不会跨数据库删除对象。

解决方法：切换到其他数据库，分别执行DROP OWNED BY name;命令。

13.4 表和外部数据源

版权所有 © 华为云计算技术有限公司

13.4.1 CREATE TABLE

功能描述

在当前数据库中创建一个新的空白表。

该表由命令执行者所有，如果系统管理员在某个与普通用户同名的schema下创建表，则该表的所有者为该schema对应的普通用户，而非系统管理员。

注意事项

表 13-43 CREATE TABLE 注意事项

约束项	详细内容
创建列存和HDFS分区表的数量	建议不超过1000个。
列存表支持的数据类型	请参考**列存表支持的数据类型**。
主键与唯一约束要求	表中的主键约束（PRIMARY KEY）和唯一约束（UNIQUE）必须包含分布列。
分布列修改限制	不支持修改已有表的分布列数据类型。
行存REPLICATION表限制	行存REPLICATION分布表不支持将系统列设置为主键。
列存表约束支持范围	列存表仅支持PARTIAL CLUSTER KEY、主键和唯一表级约束，不支持外键（FOREIGN KEY）表级约束。
列存表的字段约束	仅支持NULL、NOT NULL和DEFAULT常量值。
冷热表	• 冷热表仅支持列存分区表，依赖于可用的OBS服务。 • 冷热表仅支持默认表空间为default_obs_tbs，如需新增OBS表空间可联系技术支持。
列存表支持delta表	列存表支持delta表，由表级参数enable_delta控制是否开启，由参数deltarow_threshold控制进入delta表的阈值。不推荐使用列存带Delta表，因为会出现MERGE处理不及时而导致的磁盘膨胀以及性能劣化等问题。列存delta表由于上述问题在9.1.0.200后续版本不再支持新创建，统一使用HStore表替代。

版权所有 © 华为云计算技术有限公司

约束项	详细内容
HStore表	● 创建HStore表时，必须确保数据库GUC参数设置满足以下条件： – autovacuum设置为on。 – 设置hstore表automerge工作线程的数量，autovacuum_max_workers的取值必须大于autovacuum_max_workers_hstore取值，建议autovacuum_max_workers=6，autovacuum_max_workers_hstore=3。 ● 创建HStore表的用户需要拥有schema cstore的USAGE权限。 ● 表级参数enable_delta与enable_hstore_opt无法同时开启，因为enable_delta用于控制普通列存表的delta开启，会与enable_hstore_opt冲突。 ● 每一个HStore表绑定一张delta表，delta表的oid记录在pg_class中reldeltaidx字段（reldelta字段被列存表的delta表使用）。
V3表	● 创建表后，不支持通过ALTER TABLE语法将非V3表切换为V3表（即colversion为2.0不支持切为3.0）。 ● V3表（即colversion=3.0，存算分离表，以下简称V3表）不支持设置delta表和列存二级分区。 ● V3表不支持设置为冷热表、时序表。 ● V3表不支持创建全局临时表和临时表，创建的临时表会自动转化为colversion=2.0的临时表。

 版权所有 © 华为云计算技术有限公司

约束项	详细内容
加密表	• 加密表语法仅9.1.1.200及以上集群版本支持。 • 加密表语法仅行存表、列存2.0表支持。 • **集群级透明加密和细粒度透明加密功能互斥，无法同时开启**。已经开启了集群级透明加密，则无法开启细粒度透明加密功能。 • 加密表分为表级加密和列级加密。 表级加密：设置encrypt_option=table时，表示全表进行加密。 列级加密：设置encrypt_option=column时，表示对表进行列级加密。若没有明确指定加密列，表示表中没有加密列。 • 行存表不支持列级的透明加密配置，即encrypt_option=column。 • 升级观察期内，不支持创建透明加密表。 • 索引不支持修改透明加密配置。 • 如果已经创建加密表并向加密表中写入数据，关闭细粒度透明加密开关后，无法对加密表进行读写操作。 • 普通列存+delta表、HStore非opt表、行列混存、LSM表及临时表不支持表级加密。 • PG_STATISTIC系统表保存的列统计信息不支持加密。
其他	如果在建表过程中数据库系统发生故障，系统恢复后可能无法自动清除之前已创建的、大小为0的磁盘文件。此种情况出现概率小，不影响数据库系统的正常运行。

⚠ 警告

- 不建议创建普通表时指定自定义TABLESPACE。
- 创建行存表时应避免指定COMPRESS压缩属性。
- 创建HASH分布的表对象时，要确保数据分布均匀（10G以上数据量的表，倾斜率控制在10%以内）。
- 创建REPLICATION分布的表对象，要确保表数据量控制在100万行以内。
- 针对存在时间字段的大表（数据量5000万行以上），必须设计成分区表，根据查询特征合理设计分区间隔。
- 针对有大批量数据增删改的表，索引个数建议控制在3个以内，最多不超过5个。
- 更多开发设计规范参见**总体开发设计规范**。

语法格式

```
CREATE [ [ GLOBAL | LOCAL | VOLATILE ] { TEMPORARY | TEMP } | UNLOGGED ] TABLE [ IF NOT EXISTS ]
table_name
```


版权所有 © 华为云计算技术有限公司

```
{ ({ column_name data_type [ compress_mode ] [ COLLATE collation ] [ column_constraint [ ... ] ]
    | table_constraint
    | LIKE source_table [ like_option [...] ] }
    [, ... ])|
    LIKE source_table [ like_option [...] ] }
[ WITH ( {storage_parameter = value} [, ... ] ) ]
[ ON COMMIT { PRESERVE ROWS | DELETE ROWS } ]
[ COMPRESS | NOCOMPRESS ]
[ DISTRIBUTE BY { REPLICATION | ROUNDROBIN | { HASH ( column_name [,...] ) } } ]
[ TO { GROUP groupname | NODE ( nodename [, ... ] ) } ]
[ COMMENT [=] 'text' ];
```

- 其中列约束column_constraint为：

```
[ CONSTRAINT constraint_name ]
{ NOT NULL |
  NULL |
  CHECK ( expression ) |
  DEFAULT default_expr |
  ON UPDATE on_update_expr |
  COMMENT 'text' |
  UNIQUE [ NULLS [NOT] DISTINCT | NULLS IGNORE ] index_parameters [NOT ENFORCED [ENABLE QUERY OPTIMIZATION | DISABLE QUERY OPTIMIZATION]]|
  PRIMARY KEY index_parameters [NOT ENFORCED [ENABLE QUERY OPTIMIZATION | DISABLE QUERY OPTIMIZATION]] |
  REFERENCES reftable [ ( refcolumn ) ] }
[ DEFERRABLE | NOT DEFERRABLE | INITIALLY DEFERRED | INITIALLY IMMEDIATE ]
```

- 其中列的压缩可选项compress_mode为：

```
{ DELTA | PREFIX | DICTIONARY | NUMSTR | NOCOMPRESS }
```

- 其中表约束table_constraint为：

```
[ CONSTRAINT constraint_name ]
{ CHECK ( expression ) |
  UNIQUE [ NULLS [NOT] DISTINCT | NULLS IGNORE ] ( column_name [, ... ] ) index_parameters [NOT ENFORCED [ENABLE QUERY OPTIMIZATION | DISABLE QUERY OPTIMIZATION]]|
  PRIMARY KEY ( column_name [, ... ] ) index_parameters [NOT ENFORCED [ENABLE QUERY OPTIMIZATION | DISABLE QUERY OPTIMIZATION]] |
  PARTIAL CLUSTER KEY ( column_name [, ... ] ) }
[ DEFERRABLE | NOT DEFERRABLE | INITIALLY DEFERRED | INITIALLY IMMEDIATE ]
```

- 其中like选项like_option为：

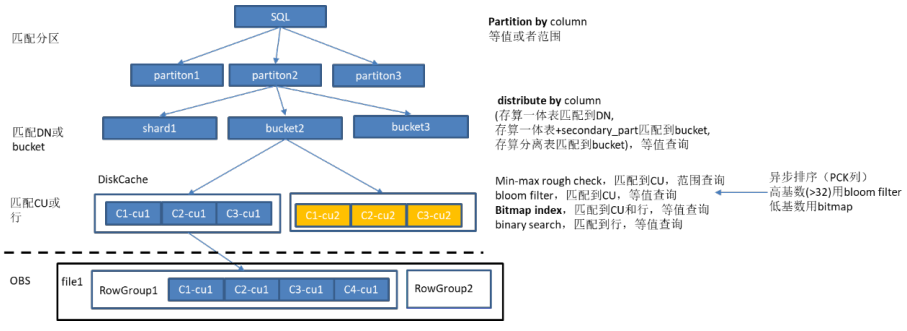
```
{ INCLUDING | EXCLUDING } { DEFAULTS | CONSTRAINTS | INDEXES | STORAGE | COMMENTS | PARTITION | RELOPTIONS | DISTRIBUTION | DROPCOLUMNS | ALL }
```

- 其中索引参数index_parameters为：

```
[ WITH ( {storage_parameter = value} [, ... ] ) ]
```

表设计参考

DWS兼容PostgreSQL生态，行存及其B-Tree索引和PostgreSQL类似，列存及其索引为自研。建表时，选择合适的存储方式、分布列、分区键、索引，能够使SQL在执行时快速命中数据，减少IO消耗。下图是一个SQL从发起到获取数据的执行流程，通过下图，可以理解每个技术手段的作用，用于指导性能调优。



 版权所有 © 华为云计算技术有限公司

1. SQL执行时，分区表会通过Partition Column分区列进行分区裁剪，定位到所在分区。
2. Hash分布表会通过Distribute Column快速定位到数据所在的数据分片（存算一体定位到DN，存算分离定位到bucket）。
3. 行存通过B-Tree快速定位到数据所在的Page；列存通过min-max索引快速定位到可能存在数据的CU数据块，PCK(Partial Cluster Key)列上的min-max索引过滤效果最好。
4. 系统自动为列存的所有列维护min-max索引，无需用户定义。min-max索引是粗过滤，满足min-max条件的CU（Column Unit）数据块不一定真正存在满足filter条件的数据行，如果定义了bitmap column，可以通过bitmap index快速定位到CU内符合条件的数据所在的行号；对于有序的CU，也会通过二分查找快速定位数据的行号。
5. 列存也支持B-Tree和GIN索引，通过B-Tree/GIN索引也可以快速定位到满足条件的数据所在的CU及行号，但是，索引的维护代价比较大，除非对点查有极高的性能诉求，推荐使用bitmap index替代B-Tree/GIN。

DWS已有的优化手段可参考下表。

表 13-44 优化手段

编号	优化手段	使用建议	SQL示例	建表后是否可修改
1	字符串类型	1. 变长字符串类型（如VARCHAR）比定长类型（如CHAR）性能稍差，如果能使用定长类型时不建议使用变长字符串类型。 2. 尽量指定长度在16以内，性能会显著提高。	-	是，已有数据会重写。
2	Numeric类型	Numeric类型要求都指定精度，性能会显著提高，不建议使用无精度的numeric。	-	是，已有数据会重写。
3	Partition by Column（分区键）	1. 需用户定义，适用分区表，通过分区键进行分区裁剪（Partition Pruning），减少扫描范围，支持partition wise join，适用等值和范围查询。 2. 分区数目不建议超过1000，分区列不建议超过两列。	SELECT * FROM t1 WHERE t1.c1='p1';	否，如需修改请重新建表。
4	secondary_part_column（二级分区列）	1. 需用户定义，只适用于列存表，适用等值查询。 2. 可在最常用的一个等值过滤列上指定二级分区。	SELECT * FROM t1 WHERE t1.c1='p1';	否，如需修改请重新建表。

版权所有 © 华为云计算技术有限公司

编号	优化手段	使用建议	SQL示例	建表后是否可修改
5	Distribute by Column（分布列）	需用户定义，适用于频繁进行GROUP BY或者多表JOIN的关键字段，通过本地连接（Local Join）减少数据重分布（Shuffle），适用等值查询。	SELECT * FROM t1 join t2 on t1.c3 = t2.c1;	否，如需修改请重新建表。
6	Bitmap column（位图列）	需用户定义，根据CU内的重复值，自适应创建bitmap index（基数<=32），或者bloom filter（基数>32），适用VARCHAR/TEXT类型列的等值查询，建议对WHERE条件中频繁过滤的列创建。	SELECT * FROM t1 WHERE t1.c4='hello';	可以修改，已有数据不重写，只影响新数据。
7	Min-max索引（稀疏索引）	1. min-max索引无需用户定义，适用等值和范围查询。 2. min-max的过滤效果取决于数据的有序性，指定PCK（Partial Cluster Key）的列，过滤效果会更好。	SELECT * FROM t1 WHERE c3 > 100 and c3 < 200;	PCK可以修改，已有数据不重写，只影响新数据。
8	主键（B-Tree索引）	1. UPSERT入库强依赖主键，需用户定义，适用等值和范围查询，不建议超过5列。 2. 满足业务前提下，能使用定长类型的列更好，定义时尽可能将distinct值多的列放前面。	SELECT * FROM t1 WHERE c3 > 100 and c3 < 200;	可以修改，重建索引。
9	Gin索引	1. 需用户定义，适用不固定的多条件等值查询，Distinct值超过100万的列不建议使用。 2. 过滤后数据量小于1000的适用，过滤后数据量依然较大的，不推荐使用。	SELECT * FROM t1 WHERE c1 = 100 and c3 = 200 and c2 = 105;	可以修改，重建索引。
10	Orientation=column/row	指定表为行存或者列存，行存不支持压缩，适合点查询和频繁更新场景；列存支持压缩，适合分析型（OLAP）场景。	-	否，如需修改请重新建表。

版权所有 © 华为云计算技术有限公司

参数说明

表 13-45 CREATE TABLE 参数说明

参数	描述	取值范围及示例
UNLOGGED	如果指定UNLOGGED关键字，则创建的表为非日志表（Unlogged Table）。写入非日志表的数据不会被写入预写日志（WAL），因此写入速度比普通表快很多。但非日志表在冲突、执行操作系统重启、强制重启、切断电源操作或异常关机后会被自动截断（TRUNCATE），造成数据丢失的风险。非日志表中的内容也不会被复制到备服务器中，在非日志表中创建的索引也不会被自动记录。 **须知** 1. UNLOGGED表无主备机制，在系统故障或异常断点等情况下，会有数据丢失风险，不可用来存储基础数据。 2. 9.1.0版本，UNLOGGED表默认存储在pg_unlogged表空间下，且不可迁移或指定到其他表空间。 3. 从低版本升级到9.1.0版本后，旧版本创建的UNLOGGED表还存储在原表空间下。 4. 由于实例异常重启时会对UNLOGGED表进行重置，所以可能影响实例RTO。9.1.0版本提供UNLOGGED表迁移脚本switch_unlogged_tablespace.py，配合GUC参数enable_unlogged_tablespace_compat，可以优化RTO。	使用场景：非日志表不能保证数据的安全性，用户应该在确保数据已经做好备份的前提下使用，例如系统升级时进行数据的备份。 故障处理：当异常关机等操作导致非日志表上的索引发生数据丢失时，用户应该对发生错误的索引进行重建。 示例： `CREATE UNLOGGED TABLE table_unlog ( col1 VARCHAR(50) , col2 TEXT );`

 版权所有 © 华为云计算技术有限公司

参数	描述	取值范围及示例
GLOBAL | LOCAL | VOLATILE	创建临时表时在TEMP或TEMPORARY前指定GLOBAL、LOCAL、VOLATILE关键字，以创建不同属性的临时表。仅8.2.1.220及以上集群版本支持全局临时表。 • 指定关键字LOCAL，则创建本地临时表（LOCAL临时表）。 • 指定关键字GLOBAL，则创建全局临时表（GLOBAL临时表），具体行为取决于GUC参数**enable_global_temp_table**的设置（默认值为ON）。 当enable_global_temp_table=on时，表示创建GLOBAL临时表； 当enable_global_temp_table=off时，表示创建LOCAL临时表，等同于指定关键字LOCAL。 • 指定关键字VOLATILE，则创建VOLATILE临时表。 当**default_temptable_type**=local时，不指定关键字创建的临时表为LOCAL类型临时表，当default_temptable_type=volatile时，不指定关键字创建的临时表为VOLATILE类型临时表。	示例： CREATE **LOCAL** TEMP TABLE temp_test (col1 INT, col2 DECIMAL(10,2));
TEMPORARY | TEMP	如果指定TEMP或TEMPORARY关键字，则创建的表为临时表。 临时表只在当前会话可见，本会话结束后会自动删除。因此，在除当前会话连接的CN以外的其他CN故障时，仍然可以在当前会话上创建和使用临时表。 由于临时表只在当前会话创建，对于涉及对临时表操作的DDL语句，可能会产生DDL失败的报错。因此，建议DDL语句中不要对临时表进行操作。TEMP和TEMPORARY等价。	该参数其他使用约束请参见**TEMPORARY | TEMP参数使用说明**。

版权所有 © 华为云计算技术有限公司

参数	描述	取值范围及示例
IF NOT EXISTS	指定IF NOT EXISTS时，若不存在同名表，则可以成功创建表。若已存在同名表，创建时不会报错，仅会提示该表已存在并跳过创建。	示例：使用IF NOT EXISTS避免表重复创建错误。 CREATE TABLE **IF NOT EXISTS** test_t1(col1 INT,col2 VARCHAR(30));
table_name	要创建的表名。	• 表名长度不超过63个字节，以字母或下划线开头，可包含字母、数字、下划线、$、#。 • 支持中文表名，支持的中文字数与数据库字符集编码有关，请参见**数据库对象命名规则**。 • 应避免使用**关键字**作为表名。 • 使用双引号括起来的表名可以包含空格和特殊字符，但不建议在表名中使用这些字符，因为这样可能会使表名难以引用和使用，而且不同的数据库兼容模式下可能对于这种表名的处理方式也有所不同。
column_name	新表中要创建的列名。	列名长度不超过63个字节，以字母或下划线开头，可包含字母、数字、下划线、$、#。
data_type	列的数据类型。	在兼容Teradata或MySQL语法的数据库中，列数据类型指定为DATE时同样返回为DATE类型，否则返回TIMESTAMP类型。
compress_mode	列的压缩选项，当前仅对行存表有效。该选项指定表字段优先使用的压缩算法。	DELTA、PREFIX、DICTIONARY、NUMSTR、NOCOMPRESS。
COLLATE collation	COLLATE子句指定列的排序规则（该列必须是可排序的数据类型）。 如果没有指定，则使用默认的排序规则。 了解排序规则，参见**DWS排序规则**章节。	示例：创建表t1，使用case_insensitive的排序规则，表示列不区分大小写。 CREATE TABLE t1 (a text collate case_insensitive);

版权所有 © 华为云计算技术有限公司

参数	描述	取值范围及示例
LIKE source_table [like_option ...]	LIKE子句声明一个表，新表自动从这个表中继承所有列名及其数据类型和非空约束。	LIKE相关使用约束参见**LIKE参数使用说明**。
WITH	这个子句为表或索引指定可选的存储参数。 **说明** 使用任意精度类型Numeric定义列时，建议指定精度p以及刻度s。在不指定精度和刻度时，会按输入的显示出来。	WITH包含的子参数较多，按场景分类如下，开发者仅需关注特定场景下的参数即可： • 存储类相关参数：参见**表13-46**。 • 分区表相关参数：参见**表13-47**。 • delta表相关参数：参见**表13-48**。 • HStore表相关参数：参见**表13-49** • 性能相关参数：参见**表13-51**。 • 加密表相关参数：参见**表8 WITH参数说明(透明加密相关)**。
ON COMMIT { PRESERVE ROWS | DELETE ROWS }	ON COMMIT选项决定在事务中创建临时表后，当事务提交时该临时表的后续行为。 • PRESERVE ROWS（默认值）：提交时不对临时表做任何操作，临时表及其表数据保持不变。 • DELETE ROWS：提交时删除临时表中数据。 全局临时表只支持PRESERVE ROWS选项。	-
COMPRESS | NOCOMPRESS	创建新表时，需要在CREATE TABLE语句中指定关键字COMPRESS，这样，当对该表进行批量插入时就会触发压缩特性。该特性会在页范围内扫描所有元组数据，生成字典、压缩元组数据并进行存储。指定关键字NOCOMPRESS则不对表进行压缩。	缺省值：NOCOMPRESS，即不对元组数据进行压缩。

版权所有 © 华为云计算技术有限公司

参数	描述	取值范围及示例
DISTRIBUTE BY	指定表数据如何在节点之间的分布或者复制方式。 取值范围： • REPLICATION：表的每一行存储在所有数据节点（DN）中，即每个数据节点都有完整的表数据。 • ROUNDROBIN：表的每一行被轮询发送给各个DN，因此数据会被均匀地分布在各个DN中。（ROUNDROBIN仅8.1.2及以上版本支持） • HASH（column_name）：对指定的列进行Hash，通过映射，把数据分布到指定DN。 其他详细使用约束参见**DISTRIBUTE BY参数使用说明**。	示例： 指定表的数据分布方式为HASH。 CREATE TABLE test_t2(col1 INT,col2 VARCHAR(30))**DISTRIBUTE BY** HASH(col1);

版权所有 © 华为云计算技术有限公司

参数	描述	取值范围及示例
TO { GROUP groupname | NODE (nodename [, ...]) }	TO GROUP指定创建表所在的Node Group，目前不支持hdfs表使用。TO NODE主要供内部扩容工具使用，一般用户不应使用。	在逻辑集群模式下，如果不指定TO GROUP，表默认会创建在逻辑集群用户关联的节点组中；如果用户没有管理逻辑集群（例如管理员用户或其他普通用户），表默认会创建在第一个逻辑集群中（pgxc_group中oid最小的逻辑集群是第一个逻辑集群）。 如果TO GROUP指定的节点组是复制表节点组，表将创建在所有CN和DN节点上，但复制表数据将只分布在复制表节点组包含的DN节点上。 存算分离3.0版本支持只读计算组，如果未绑定到只读计算组的用户，建表语句中TO GROUP指定的是只读计算组，则创建表会报错；但绑定到只读计算组的用户，创建表会遵循未绑定计算组用户的建表规则，将表创建到由GUC参数default_storage_nodegroup指定的计算组上。如果default_storage_nodegroup是installation, 表会创建在第一个计算组中。
COMMENT [=] 'text'	COMMENT子句可在创建表时指定表注释。	示例： 添加表注释。 CREATE TABLE table_user(user_id INT,username VARCHAR(50)) **COMMENT** '*user Information*';
COMMENT 'text'	COMMENT子句可以指定列的注释。	示例： 添加指定列的注释。 CREATE TABLE table_users(user_id INT **COMMENT** '*employee id*',username VARCHAR(50));

版权所有 © 华为云计算技术有限公司

参数	描述	取值范围及示例
CONSTRAINT constraint_name	列约束或表约束的名字。可选的约束子句用于声明约束，新行或者更新的行必须满足这些约束才能成功插入或更新。 定义约束有两种方法： ● 列约束：作为一个列定义的一部分，仅影响该列。 ● 表约束：不和某个列绑在一起，可以作用于多个列。	示例： 创建表table_c_products，定义了主键约束：表的主键由C_CUSTKEY和C_NAME组成。 CREATE TABLE table_c_products (C_CUSTKEY BIGINT, C_NAME VARCHAR(25) , C_ADDRESS VARCHAR(40), C_NATIONKEY INT, C_PHONE CHAR(15) , C_ACCTBAL DECIMAL(15,2), **CONSTRAINT** C_CUSTKEY_KEY PRIMARY KEY(C_CUSTKEY,C_NAME)) DISTRIBUTE BY HASH(C_CUSTKEY,C_NAME);
NOT NULL	字段值不允许为NULL。	-
NULL	字段值允许为NULL，这是默认值。 这个子句只是为了和非标准SQL数据库兼容。不建议使用。	-
CHECK (expression)	CHECK约束声明一个布尔表达式，每次要插入的新行或者要更新的行的新值必须使表达式结果为真或未知才能成功，否则会抛出一个异常并且不会修改数据库。 声明为字段约束的检查约束应该只引用该字段的数值，而在表约束里出现的表达式可以引用多个字段。	expression表达式中，如果存在“<>NULL”或“!=NULL”，这种写法是无效的，需要写成“is NOT NULL”。
DEFAULT default_expr	DEFAULT子句给字段指定缺省值。	该数值可以是任何不含变量的表达式(不允许使用子查询和对本表中的其他字段的交叉引用)。缺省表达式的数据类型必须和字段类型匹配。 缺省表达式将被用于任何未声明该字段数值的插入操作。若未显式指定DEFAULT，则默认值为NULL。

版权所有 © 华为云计算技术有限公司

参数	描述	取值范围及示例
ON UPDATE on_update_expr	ON UPDATE子句给字段指定时间戳函数。ON UPDATE子句支持指定的字段类型必须为timestamp类型或者timestamptz类型。 当执行含有update操作的SQL语句时，自动更新此列为时间戳函数所代表的时间。 on_update_expr时间戳函数仅支持CURRENT_TIMESTAMP, CURRENT_TIME, CURRENT_DATE, LOCALTIME, LOCALTIMESTAMP。	示例：当更新t1表的b列数据时，第三列会显示对应的更新时间。 CREATE TABLE t1 (a INT PRIMARY KEY, b INT, col_update_time TIMESTAMP ON UPDATE CURRENT_TIMESTAMP); INSERT INTO t1(a,b) VALUES (1,2); UPDATE t1 SET b = 3 WHERE a = 1; SELECT * FROM t1; gaussdb=> SELECT * FROM t1; a | b | col_update_time ---+---+--- 1 | 3 | 2026-06-25 11:35:41.764667 (1 row)
UNIQUE [NULLS [NOT] DISTINCT | NULLS IGNORE] index_parameters UNIQUE [NULLS [NOT] DISTINCT | NULLS IGNORE] (column_name [, ...]) index_parameters	UNIQUE约束表示表里的一个字段或多个字段的组合必须在全表范围内唯一。 其中[NULLS [NOT] DISTINCT | NULLS IGNORE]字段用来指定Unique唯一索引中索引列NULL值的处理方式。	默认值：该参数默认缺省，即NULL值可重复插入。 在对插入的新数据和表中原始数据进行列的等值比较时，对于NULL值有以下三种处理方式： • NULLS DISTINCT：NULL值互不相等，即NULL值可重复插入。 • NULLS NOT DISTINCT：NULL值相等。若索引列全为NULL，则NULL值不可重复插入；部分索引列为NULL，只有非NULL值不相等，才可成功插入数据。 • NULLS IGNORE：在等值比较时跳过NULL值。若索引列全为NULL，则NULL值可重复插入；部分索引列为NULL，只有非NULL值不相等，才可成功插入数据。 三种处理方式具体的行为如表13-53所示。 **说明** 如果没有声明DISTRIBUTE BY REPLICATION，则唯一约束的列集合中必须包含分布列。

版权所有 © 华为云计算技术有限公司

参数	描述	取值范围及示例
PRIMARY KEY index_parameters PRIMARY KEY (column_name [, ...]) index_parameters	主键约束声明表中的一个或者多个列只能包含唯一的非NULL值。 一个表只能声明一个主键。 **注意** 如果没有声明DISTRIBUTE BY REPLICATION，则主键约束的列集合中必须包含分布列。	示例： 创建表products，定义product_no为整数类型，并将其设为主键。 CREATE TABLE products (product_no integer PRIMARY KEY, name text, price numeric) DISTRIBUTE BY HASH(product_no);
NOT ENFORCED	指定所建立的约束为信息约束，数据库对该约束不进行数据校验，由用户来保证约束要求。	-
ENFORCED	预留参数，目前不支持ENFORCED。	-
ENABLE QUERY OPTIMIZATION	利用信息约束对执行计划进行优化。	-
DISABLE QUERY OPTIMIZATION	禁止利用信息约束对执行计划优化。	-
REFERENCES reftable [(refcolumn)]	外键约束要求新表中一列或多列构成的组应该只包含、匹配被参考表中被参考字段值。被参考列应该是被参考表中的唯一字段或主键。 由于DWS不会进行外键约束检查，使用外键约束时需通过函数check_foreign_key_constraint来检测外键表中的数据是否满足外键约束。	-
DEFERRABLE | NOT DEFERRABLE	这两个关键字设置该约束是否可延迟。一个不可延迟的约束将在每条命令之后马上检查。可延迟约束可以延迟到事务结尾使用**SET CONSTRAINTS**命令检查。 缺省是NOT DEFERRABLE。目前，只有行存的UNIQUE约束和主键约束可以接受这个子句。所有其他约束类型都是不可延迟的。	-

版权所有 © 华为云计算技术有限公司

参数	描述	取值范围及示例
PARTIAL CLUSTER KEY	局部聚簇存储，列存表导入数据时按照指定的列(单列或多列)，进行局部排序。 列存/hstore实时入库场景中使用PARTIAL CLUSTER KEY时，GUC参数autovacuum_asyncsort_time_limit不能设置为0。	示例： 创建表products_new，并在product_no列上定义了局部聚簇。 CREATE TABLE products_new (product_no integer, name text, price numeric, PARTIAL CLUSTER KEY(product_no)) WITH (ORIENTATION = COLUMN);
INITIALLY IMMEDIATE | INITIALLY DEFERRED	如果约束是可延迟的，则这个子句声明检查约束的缺省时间。	● 如果约束是INITIALLY IMMEDIATE（缺省），则在每条语句执行之后就立即检查它。 ● 如果约束是INITIALLY DEFERRED ，则只有在事务结尾才检查它。 ● 约束检查的时间可以用SET CONSTRAINTS命令修改。

表 13-46 WITH 参数说明(存储相关)

参数	描述	取值范围及默认值
ORIENTATION	指定表数据的存储方式，即行存方式、列存方式，该参数设置成功后就不再支持修改。 ● 行存储适合于OLTP业务，此类型的表上交互事务比较多，一次交互会涉及表中的多个列，用行存查询效率较高。 ● 列存储适合于数据仓库业务，此类型的表上会做大量的聚合计算，且涉及的列操作较少。	● ROW，表示表的数据将以行式存储。 ● COLUMN，表示表的数据将以列式存储。 默认值：ROW，即行存方式。 **说明** ● 如需创建列存表，需显式设置orientation属性为column，存算分离版本如需创建本地表（数据全部存储在EVS盘），需显式指定colversion=2.0。 ● 对于HStore表，应当使用列存方式，同时设置enable_hstore_opt为true。

版权所有 © 华为云计算技术有限公司

参数	描述	取值范围及默认值
COMPRESSION	指定表数据的压缩级别，它决定了表数据的压缩比以及压缩时间。一般来讲，压缩级别越高，压缩比也越大，压缩时间也越长；反之亦然。实际压缩比取决于加载的表数据的分布特征。	该参数列存表和HStore表支持，有效值为：LOW、MIDDLE或HIGH。默认值为LOW。 DWS内部提供的压缩算法参见**表13-52**。 **说明** • 暂不支持行存表压缩功能。 • ORC格式迁移到DWS列存表，如果使用low级别压缩，单副本大小大约是ORC的1.5~2倍；如果使用high级别压缩，单副本大小基本与ORC持平，在新建DWS集群规模时，需考虑该转换关系。 • 列存middle压缩固定使用字典压缩，对于数据特征不适合字典压缩的数据，使用middle压缩的结果可能比low压缩更大。
COMPRESSLEVEL	指定表数据同一压缩级别下的不同压缩水平，它决定了同一压缩级别下表数据的压缩比以及压缩时间。 对同一压缩级别进行了更加详细的划分，为用户选择压缩比和压缩时间提供了更多的空间。总体来讲，此值越大，表示同一压缩级别下压缩比越大，压缩时间越长；反之亦然。该参数只对列存表有效。	0~3 默认值：0
MAX_BATCHROW	指定了在数据加载过程中一个存储单元可以容纳记录的最大数目。该参数只对列存表有效。	10000~60000 默认值：60,000
PARTIAL_CLUSTER_ROWS	指定了在数据加载过程中进行局部聚簇存储的记录数目。该参数只对列存表有效。	600000~2147483647 默认值：4,200,000

版权所有 © 华为云计算技术有限公司

参数	描述	取值范围及默认值
COLVERSION	指定列存存储格式的版本，支持不同存储格式版本之间的切换。	• 1.0：列存表的每列以一个单独的文件进行存储，文件名以relfilenode.C1.0、relfilenode.C2.0、relfilenode.C3.0等命名。 • 2.0：列存表的每列合并存储在一个文件中，文件名以relfilenode.C1.0命名。 • 3.0：列存表的每列合并存储在一个文件中，文件存储在OBS文件系统上，文件名以C1_fileid.0命名。 默认值：存算一体版本默认值2.0，存算分离版本默认值3.0。

版权所有 © 华为云计算技术有限公司

参数	描述	取值范围及默认值

		说明
		● HStore表及OBS冷热表只支持2.0版本。
		● 8.1.0集群版本该参数默认值为1.0，8.1.1及以上集群版本该参数默认值为2.0，若集群版本由8.1.0升级至8.1.1或以上版本，该参数默认值也会由1.0变为2.0。
		● 在建列存表时，选择COLVERSION=2.0，相比于1.0存储格式，在以下场景中性能有明显提升：
		  ● 创建列存宽表场景下，建表时间显著减少。
		  ● roach备份数据场景下，备份时间显著减少。
		  ● build、catch up耗时显著减少。
		● 存算分离版本兼容列存所有版本，建表时需显式指定colversion=1.0/2.0/3.0。当colversion=3.0时建立表为存算分离表，不显式指定colversion时默认创建3.0版本列存表。需注意，创建存算分离表时指定colversion为3.0的同时需要将orientation属性设置为column。
		● V3存算分离表不支持使用ALTER TABLE切换colversion，比如从2.0升级到3.0。

 版权所有 © 华为云计算技术有限公司

参数	描述	取值范围及默认值
cache_policy（该参数仅9.0.2及以上的存算分离版本支持）	控制了表或者分区表（磁盘）缓存的方式，如果在缓存策略中指定相应参数范围，则会进行热缓存，否则进行冷缓存。热缓存相比冷缓存占用的空间更大，技术上使用更加复杂的替换策略。 • ALL： 对整个表进行热缓存。 • NONE：对整个表进行冷缓存。 • HPN(HOT PARTITION NUMBER)：分区表中，更晚创建的N个分区会被热缓存，其余分区进行冷缓存。 • HPL(HOT PARTITION LIST)：P1, P2, ... 分区表中在缓存策略中被指定名称的分区会被热缓存，其余分区进行冷缓存。 默认值：ALL **说明** • 对于外表和非分区内表只支持ALL和NONE两种缓存策略。 • 仅range和list分区内表支持HPN和HPL缓存策略。	• HPN的使用示例请参见**示例：创建HPN磁盘缓存策略的分区表**。 • HPL的示例请参见**示例：创建HPL磁盘缓存策略的分区表**
disk_cache_ttl	设置新导入的数据期望在（磁盘）缓存中保留的时间，确保新导入的数据在（磁盘）缓存中保留一段时间不被淘汰。该参数仅适用于COLVERSION为3.0的列存表。 该参数仅9.1.1.100及以上的存算分离集群版本支持。 • 支持alter table set的方式修改disk_cache_ttl属性。 • 支持alter table reset的方式取消disk_cache_ttl属性。	1 second ~ 100 years。 使用示例请参见**示例：设置disk_cache_ttl参数，确保新导入数据在磁盘缓存中保留一定时间不被淘汰**。
clsm_mode	指定列存表是否开启LSM，仅9.1.2及以上集群版本支持。该参数仅适用于COLVERSION为3.0的hstore_opt表。使用该参数时，保证已开启GUC参数enable_lsm_engine。	• pk，表示主键索引为LSM的表。 • all，表示主键索引及cudesc/delta为LSM的表。 • off，表示非LSM表，默认值。

版权所有 © 华为云计算技术有限公司

参数	描述	取值范围及默认值
enable_disaster_cstore	指定了列存表是否开启细粒度容灾功能。该参数仅适用于COLVERSION为2.0的普通列存表，不支持hstore表、行存表、临时表、UNLOGGED表、外表、时序表、冷热表，物化视图以及开启delta表功能，并且不能和enable_hstore_opt，enable_delta同时打开。	默认值：off
fine_disaster_table_role	该参数8.2.1版本中已废弃，为兼容历史版本功能保留该参数，当前版本设置无效。 指定了细粒度容灾表的表角色为主表还是备表。使用该参数时，必须要保证已经开启了enable_disaster_cstore参数。	● primary，表示细粒度容灾主表。 ● standby，表示细粒度容灾备表。
on_commit_preserve_rows	与**ON COMMIT { PRESERVE ROWS | DELETE ROWS } 类似**。两者不能同时指定。该参数仅用于全局临时表。	默认值：true。
disable_vacuum_module	对于特定业务场景（例如包含大量只读分区表，或频繁进行批量导入的中间表），可以通过配置此参数来跳过列存表上的后台维护任务（如VACUUM、索引维护、异步排序等）。这有助于提升系统性能，并减少不必要的资源消耗。 该参数仅9.1.1.300及以上集群版本支持。	默认为空，表示不跳过任何操作。多个值可使用英文逗号分隔，如“index_only,async_sort”。 ● all，表示跳过全部操作，包括：列存VACUUM、索引VACUUM、异步排序。 ● index_only，表示跳过索引VACUUM。 ● async_sort，表示跳过异步排序 (仅对hstore_opt表生效）。 ● vacuum_only，表示跳过列存CU VACUUM。

表 13-47 WITH 参数说明(分区表相关)

参数	描述	取值范围及默认值
TTL	设置分区表定时删除分区task任务。默认不创建删除分区task任务。	1 hour ~ 100 years

版权所有 © 华为云计算技术有限公司

参数	描述	取值范围及默认值
PERIOD	设置分区表定时创建分区task任务。如果设置TTL，PERIOD不能大于TTL。	1 hour ~ 100 years 默认值：1 day
time_format	自增减分区支持INT4/INT8/VARCHAR/TEXT列，存储内容为时间时使用。time_format选项仅当分区键为INT4/INT8/VARCHAR/TEXT时，同时指定period才生效。该参数仅9.1.0.200及以上集群版本支持。	1. time_format写法遵从DWS规范，例如：time_format = 'yyyymmdd'。 2. 分区键的类型**VARCHAR/TEXT**对time_format的规格有如下限制： ● 精度支持到秒级。 ● 输入内容不包括字母类型，例如month，am，pm等。 ● 时间必须从大到小排列，例如年月日小时分钟秒。 **INT4/INT8**对time_format的规格有如下限制： ● 精度支持到小时级。 ● 输入内容仅包括Y,M,D,HH24。 ● 时间必须从大到小排列，例如年月日小时。 3. ALTER相关限制 ● 不支持set。 ● 当period被reset时（表示已经关闭自动分区，会返回提示信息），此时可以reset此选项。

版权所有 © 华为云计算技术有限公司

表 13-48 WITH 参数说明(delta 表相关)

参数	描述	取值范围及默认值
enable_delta	• 指定了列存表是否开启delta表。该参数只对列存表有效。V3表暂不支持该参数，即COLVERSION为3.0时不支持设置enable_delta为on。 • 不推荐使用列存带Delta表，因为会出现MERGE处理不及时而导致的磁盘膨胀以及性能劣化等问题。 • 对HStore表不能开启该参数。	默认值：off
DELTAROW_THRESHOLD	• 指定列存表导入时小于多少行的数据进入delta表，该参数只对列存表和HStore表有效。 实时数仓场景下：指定HStore表导入时小于多少行(SUB_PARTITION_COUNT * DELTAROW_THRESHOLD)的数据进入delta表。 • 9.1.1.300及以上集群版本的标准数仓场景下：DELTAROW_THRESHOLD参数值默认为**0**，表示创建为**批量模式hstore_opt表**（批量模式，即批量数据更新，无并发实时入库能力，查询性能与实时数仓场景持平，适用于高效处理大批量数据同步更新的场景）。DELTAROW_THRESHOLD参数值不为0时，表示创建为hstore_opt标准模式。 hstore_opt批量模式的约束和限制： • 标准数仓场景下默认创建hstore_opt批量模式表，要求必须指定orientation=column，如果不指定orientation默认创建的表为行存表。 • 支持ALTER TABLE修改DELTAROW_THRESHOLD参数设置值。 • hstore_opt批量模式不支持UPSERT操作。 • hstore_opt批量模式不支持binlog和增量物化视图。 • hstore_opt批量模式强绑定turbo存储和turbo引擎。 • hstore_opt只有批量模式支持细粒度容灾，实时数仓场景下不支持细粒度容灾。 • hstore_opt批量模式下，当并发更新同一个CU时，将触发锁等待或者死锁。	0～60000 默认值：1000

版权所有 © 华为云计算技术有限公司

表 13-49 WITH 参数说明(HStore 表相关)

参数	描述	取值范围及默认值
enable_hstore	指定了是否创建为HStore表（基于列存表实现）。该参数只对列存表有效。 ● 该参数仅8.2.0.100及以上集群版本支持。 ● 9.1.0.100及以上集群版本中，为保持前向兼容保留该参数，开启HStore表请指定enable_hstore_opt参数。	取值为on/off 默认值：off
enable_hstore_opt	enable_hstore_opt表级参数打开时默认会同时打开表级参数enable_hstore和enable_turbo_store，该参数仅9.1.0.100及以上集群版本支持。 新安装的9.1.1.300及以上集群版本中，建表时若未显式指定enable_hstore_opt参数，系统将根据GUC参数column_default_format控制是否开启hstore_opt表。该参数默认值为column_hstore_opt，即在创建列存表（显式指定orientation=column或隐式指定）时根据集群类型适配： ● 实时数仓集群：创建标准模式hstore_opt表。 ● 标准数仓集群：创建批量模式hstore_opt表。	取值为true/false 默认值：false **说明** 打开该参数时必须设置以下GUC参数用于保证HStore表的清理，推荐值如下： autovacuum=on，autovacuum_max_workers=6，autovacuum_max_workers_hstore=3。
enable_binlog	用于控制HStore表是否开启binlog功能。该参数仅9.1.0及以上集群版本支持。	取值为on/off 默认值：off
enable_binlog_timestamp	用于控制HStore表是否开启带时间戳的binlog功能。该参数与enable_binlog无法同时打开，且仅9.1.0.200及以上集群版本支持。	取值为on/off 默认值：off

版权所有 © 华为云计算技术有限公司

参数	描述	取值范围及默认值
enable_light_update	指定hstore_opt表是否**开启轻量化更新**，开启后，在对hstore_opt表的列进行update更新入库时，不再对非更新列进行反查操作，以此减少入库反查带来的性能和资源开销，进一步提高Hstore_opt表的入库性能。该参数仅9.1.1.100及以上集群版本支持。 • 开启此功能，请确保控制全局的GUC参数**enable_hstore_lightupdate_table**（即是否支持创建轻量化更新的hstore_opt表）已打开。 • 仅支持hstore_opt表开启此参数。 • 该参数不支持ALTER操作，仅支持建表时指定。 • 不支持REPLICATION分布表开启该参数。 • 建议创建轻量化更新表时指定主键。 • 建表时若不指定PCK列，会默认使用主键列（若存在）第一列作为PCK列。 • 创建PCK约束时表必须包含主键，且PCK列必须为主键列。 • 轻量化update的表只支持创建一个索引，该索引只能为主键或唯一索引。	取值为on/off 默认值：off

版权所有 © 华为云计算技术有限公司

参数	描述	取值范围及默认值
storage_mode	指定hstore_opt表建立**行列混合存储表（简称行列混存表，也称行列共存表）**的方式。该参数仅9.1.1.100及以上版本支持。 • 了解行列混存表，请参见**使用DWS行列共存表**。 • **轻量化更新**（light update）表、binlog表仅在9.1.1.200及以上集群版本中支持行列混存表。 • **该参数依赖enable_hstore_opt = on**，即Hstore_opt表才支持。	取值为col/row/mix 默认值：不指定，仍为hstore_opt表 • col模式，表示行列混存的列模式，同hstore_opt相同。 • row模式，表示行列混存的行模式。支持压缩，该模式为全新的行模式，与行存表不同。 • mix模式，表示同时存储col模式和row模式两份数据。 **示例：创建行列混存表**
secondary_part_column	指定列存表二级分区列的列名，仅能指定一列作为二级分区列且只适用于HStore列存表。该参数仅8.3.0及以上集群版本支持。	• 被指定为二级分区的列不能被删除或者修改。 • 只能在建表时指定二级分区列，建表后不支持修改。 • 不建议指定分布列作为二级分区列。 • 指定二级分区列后，该表将会在DN上按照二级分区列进行逻辑上的HASH分区存储，从而有效提升该列等值查询的性能。
secondary_part_num	指定列存表二级分区的数量，仅适用于HStore列存表。该参数仅8.3.0及以上集群版本支持。	1~32 默认值：8 **须知** • 只有secondary_part_column被指定时，该参数才可以指定。 • 只能在建表时指定二级分区的数量，建表后不支持修改。 • 不建议用户随意设置该默认值，否则会影响导入和查询的性能。

版权所有 © 华为云计算技术有限公司

参数	描述	取值范围及默认值
bitmap_columns	bitmap index仅适用于hstore_opt表，开启表级参数enable_hstore_opt且开启bitmap_columns="指定列"，才能生成bitmap index映射关系。该参数仅8.3.0及以上集群版本支持。	-
enable_wide_table	用于控制是否开启大宽表功能。该参数仅9.1.1.300及以上集群版本支持。 仅适用于hstore_opt表，开启后最多支持5000列。 大宽表功能主要约束场景： ● 不支持多VW场景。 ● 不支持外表导出功能。 ● 不支持UPDATE语句。 ● 不支持游标等计划中包含行存执行引擎算子的场景。 ● 不支持pck、异步排序、binlog、行列混存表等。	取值为on/off 默认值：off

表 13-50 WITH 参数说明(透明加密相关)

参数	描述	取值范围
encrypt_option	指定表的加密级别。	● table，表示全表加密。 ● column，表示列级加密，仅支持列存表。 ● off，表示非加密。

 版权所有 © 华为云计算技术有限公司

参数	描述	取值范围
encrypt_algo	指定加密算法。默认值为none。 若未显式设置该参数时，由GUC参数transparent_encrypt_algorithm控制加密算法类型（默认值为aes_ctr_128）；若显式设置，则以表级配置为准。 **开启国密SM4-CTR-128算法**时，建议开启KAE（鲲鹏加速引擎，需联系技术支持开启GUC参数enable_kae_accelerate），以降低对集群性能的影响。（未开启KAE，集群性能开销将大幅增加；开启KAE后，集群性能开销显著降低，因SM4加密本身会带来一定的性能开销，在TPC-DS复杂查询场景下的仍会出现：列存整体性能劣化通常在5%至15%之间及行存整体性能劣化通常在50%以上的情况，具体的劣化幅度取决于硬件是否提供KAE鲲鹏加速引擎）。	● AES_CTR_128 ● SM4_CTR_128 ● AES_128_CTR_SHA256_HMAC（9.1.2及以上集群版本支持） ● SM4_CTR_SM3_HMAC（9.1.2及以上集群版本支持）

 版权所有 © 华为云计算技术有限公司

表 13-51 WITH 参数说明(其他性能相关参数)

参数	描述	取值范围及默认值
FILLFACTOR	一个表的填充因子（fillfactor）是一个介于10和100之间的百分数。100（完全填充）是默认值。如果指定了较小的填充因子，INSERT操作仅按照填充因子指定的百分率填充表页。每个页上的剩余空间将用于在该页上更新行，这就使得UPDATE有机会在同一页上放置同一条记录的新版本，这比把新版本放置在其他页上更有效。对于一个从不更新的表将填充因子设为100是合适的选择，但是对于频繁更新的表，选择较小的填充因子则更加合适。该参数对于列存表没有意义。	10~100
analyze_mode	控制表级自动analyze的方式。	● frozen：禁止所有形式的analyze（例外：无统计信息时，还能触发动态采样）。 ● backend：仅允许autovacuum轮询触发的analyze。 ● runtime：仅允许优化器触发的runtime analyze。 ● all：backend和runtime两种自动analyze都允许触发。 默认值：all
incremental_analyze	控制分区表是否启用增量analyze模式。该参数只对分区表有效，复制表不支持设置。该参数仅9.1.0.100及以上集群版本支持。	默认值：false

 版权所有 © 华为云计算技术有限公司

参数	描述	取值范围及默认值
enable_turbo_store	用于控制是否创建为turbo表（基于列存表实现）。该参数只对列存表有效。该选项仅9.1.0.100及以上集群版本支持。 **说明** • 3.0普通列存表不支持设置turbo存储格式，仅3.0 hstore_opt表支持turbo存储格式。 • 2.0版本列存表无限制。 • turbo表当前不支持时序表、delta表以及轻量化update。	默认值：off **说明** turbo表对部分numeric和varchar数据类型做了整数格式存储优化，以提升对于numeric和varchar数据类型的处理速度。 • 为获取最佳性能，推荐定义numeric类型时指定precision和scale(p,s)。具体来说，当p小于等于9时，使用4字节整数存储；p小于等于18时，使用8字节整数存储；p小于等于37时，使用16字节整数存储；p大于37或未指定时，则依然使用变长格式存储，其占用较多存储空间，且性能较差。 • 如果定义varchar类型时指定其最大字节长度n小于等于2字节，使用2字节整数存储；n小于等于4字节时，使用4字节整数存储；依次类推，最大当n小于等于16字节时，使用16字节整数存储。因此对于varchar类型的GROUP BY列或者HashJoin关联列，如果可使用整数存储字符串，建议定义时指定其最大字节长度n小于等于16，可以成倍提升计算性能。
SKIP_FPI_HINT	顺序扫描过程中，若需要写FPW(full page writes)日志时，该参数控制是否跳过设置HintBits操作。	默认值：false **说明** 设置SKIP_FPI_HINT=true时，在对某表执行checkpoint操作后，若对该表进行顺序扫描，将不再产生Xlog。适用于查询次数较少的中间表，有效减少Xlog的大小，提升查询性能。
enable_column_autovacuum_garbage	用于控制是否开启列存表AUTOVACUUM重写CU的逻辑。该参数仅8.2.1.100及以上集群版本支持。 当前列存轻量化UPDATE与后台列存AUTOVACUUM并发会小概率报错， 可以通过设置该表级参数为off来避免。	默认值：true

版权所有 © 华为云计算技术有限公司

表 13-52 列存压缩算法

COMPRESSION	NUMERIC	STRING	INT
LOW	delta压缩+RLE压缩	lz4压缩	delta压缩（RLE可选）
MIDDLE	delta压缩+RLE压缩+lz4压缩	dict压缩或lz4压缩	delta压缩或lz4压缩（RLE可选）
HIGH	delta压缩+RLE压缩+zlib压缩	dict压缩或zlib压缩	delta压缩或zlib压缩（RLE可选）

表 13-53 唯一索引中索引列 NULL 值的处理方式

字段控制	索引列全为NULL	部分索引列为NULL
NULLS DISTINCT	可重复插入	可重复插入
NULLS NOT DISTINCT	不可重复插入	非NULL值相等，不可插入；非NULL值不相等，则插入成功
NULLS IGNORE	可重复插入	非NULL值相等，不可插入；非NULL值不相等，则插入成功

TEMPORARY | TEMP 参数使用说明

- LOCAL/VOLATILE临时表通过每个会话独立的以pg_temp开头的schema来保证只对当前会话可见，因此，不建议用户在日常操作中手动删除以pg_temp，pg_toast_temp开头的schema。
- 如果建表时不指定TEMPORARY/TEMP关键字，而指定表的schema为当前会话的pg_temp_开头的schema，则此表会被创建为临时表。
- LOCAL临时表的所有相关元数据同普通表类似，都存储在系统表内，而VOLATILE临时表会将除schema外的相关表结构元数据直接存储在内存中。所以相对本地临时表而言，VOLATILE临时表有更多约束：
  - 当前CN或DN重启之后，对应实例上的内存数据丢失，相关volatile临时表会失效。
  - VOLATILE临时表当前不支持ALTER/GRANT等修改表结构相关操作。
  - VOLATILE临时表和LOCAL临时表共用一种临时schema，所以在同一session中，VOLATILE临时表和LOCAL临时表不能存在同名表。
  - VOLATILE临时表信息不存储在系统表内，所以无法通过对系统表执行DML语句查询到VOLATILE相关元数据。
  - VOLATILE临时表仅支持普通的行存、列存表，不支持delta表、时序表、冷热表。
  - 不支持基于VOLATILE临时表创建视图。
  - 不支持创建临时表时指定tablespace（VOLATILE临时表默认tablespace均为pg_volatile）。

版权所有 © 华为云计算技术有限公司

  - 创建VOLATILE临时表时不支持指定约束：CHECK约束、UNIQUE约束、主键约束、触发器约束、EXCLUDE约束、PARTIAL CLUSTER约束。
- GLOBAL临时表的所有相关元数据同普通表类似，都存储在系统表内。
  - GLOBAL临时表与LOCAL临时表不同的是，会话退出时元数据不会删除，但会话的数据会删除。不同会话的数据独立，但共享同一份GLOBAL临时表的元数据。
  - 全局临时表与普通表一样必须定义在常规Schema中（如public或用户自定义Schema），不能使用以pg_temp开头的临时Schema。全局临时表可以与LOCAL/VOLATILE临时表同名，因为两者存储在不同的Schema中。
  - 全局临时表仅支持普通的行存、列存表，不支持delta表、时序表、冷热表。
  - 不支持操作其他逻辑集群的全局临时表。

LIKE 参数使用说明

新表与源表之间在创建动作完毕之后是完全无关的。在源表做的任何修改都不会传播到新表中，并且也不可能在扫描源表的时候包含新表的数据。

被复制的列和约束并不使用相同的名字进行融合。如果明确的指定了相同的名字或者在另外一个LIKE子句中，将会报错。

- 源表上的字段缺省表达式或者ON UPDATE表达式只有在指定INCLUDING DEFAULTS时，才会复制到新表中。缺省是不包含缺省表达式的，即新表中的所有字段的缺省值都是NULL。
- 源表上的CHECK约束仅在指定INCLUDING CONSTRAINTS时，会复制到新表中，而其他类型的约束永远不会复制到新表中。非空约束总是复制到新表中。此规则同时适用于表约束和列约束。
- 如果指定了INCLUDING INDEXES，则源表上的索引也将在新表上创建，默认不建立索引。
- 如果指定了INCLUDING STORAGE，则复制列的STORAGE设置会复制到新表中，默认情况下不包含STORAGE设置。
- 如果指定了INCLUDING COMMENTS，则源表列、约束和索引的注释会复制到新表中。默认情况下，不复制源表的注释。
- 如果指定了INCLUDING PARTITION，则源表的分区定义会复制到新表中，同时新表将不能再使用PARTITION BY子句。默认情况下，不复制源表的分区定义。
- 如果指定了INCLUDING RELOPTIONS，则源表的存储参数（即源表的WITH子句）会复制到新表中。默认情况下，不复制源表的存储参数。

  说明

  WITH中的'PERIOD'，'TTL'为partition相关参数，LIKE INCLUDING RELOPTIONS不会复制到新表中，若要复制需INCLUDING PARTITION。
- 如果指定了INCLUDING DISTRIBUTION，则源表的分布信息会复制到新表中，包括分布类型和分布列，同时新表将不能再使用DISTRIBUTE BY子句。默认情况下，不复制源表的分布信息。
- 如果指定了INCLUDING DROPCOLUMNS，则源表被删除的列信息会被复制到新表中。默认情况下，不复制源表的删除列信息。
- INCLUDING ALL包含了INCLUDING DEFAULTS、INCLUDING CONSTRAINTS、INCLUDING INDEXES、INCLUDING STORAGE、INCLUDING COMMENTS、INCLUDING PARTITION、INCLUDING RELOPTIONS、INCLUDING DISTRIBUTION和INCLUDING DROPCOLUMNS的内容。

 版权所有 © 华为云计算技术有限公司

- 如果指定了EXCLUDING，则表示不包括指定的参数。
- 如果是OBS冷热表，INCLUDING PARTITION后新表所有分区均为本地热分区。
- 对源表同时执行ALTER TABLE...ADD PARTITION与CREATE TABLE语句，偶发可能会出现不同节点的分区数量不一致的情况，从而导致CREATE TABLE语句报错。出现此情况时，建议重新执行CREATE TABLE语句，即可解决。

须知

- 如果源表包含serial、bigserial、smallserial类型，或者源表字段的默认值是sequence，且sequence属于源表（通过CREATE SEQUENCE ... OWNED BY创建），这些Sequence不会关联到新表中，新表中会重新创建属于自己的sequence。这和之前版本的处理逻辑不同。如果用户希望源表和新表共享Sequence，需要首先创建一个共享的Sequence（避免使用OWNED BY），并配置为源表字段默认值，这样创建的新表会和源表共享该Sequence。
- 不建议将其他表私有的Sequence配置为源表字段的默认值，尤其是其他表只分布在特定的NodeGroup上，这可能导致CREATE TABLE ... LIKE执行失败。另外，如果源表配置其他表私有的Sequence，当该表删除时Sequence也会连带删除，这样源表的Sequence将不可用。如果用户希望多个表共享Sequence，建议创建共享的Sequence。

DISTRIBUTE BY 参数使用说明

指定表如何在节点之间分布或者复制。

取值范围：

- REPLICATION：表的每一行存储在所有数据节点（DN）中，即每个数据节点都有完整的表数据。
- ROUNDROBIN：表的每一行被轮询发送给各个DN，因此数据会被均匀地分布在各个DN中。（ROUNDROBIN仅8.1.2及以上版本支持）
- HASH（column_name）：对指定的列进行Hash，通过映射，把数据分布到指定DN。

  说明

  - 当指定DISTRIBUTE BY HASH (column_name)参数时，创建主键和唯一索引必须包含“column_name”列。
  - 当被参照表指定DISTRIBUTE BY HASH (column_name)参数时，参照表的外键必须包含“column_name”列。
  - 如果TO GROUP指定为复制表节点组（8.1.2及以上版本支持），DISTRIBUTE BY必须指定为REPLICATION。如果没有指定DISTRIBUTE BY，创建的表会自动设置为复制表。

默认值：由GUC参数default_distribution_mode控制。

- 当default_distribution_mode=roundrobin时，DISTRIBUTE BY的默认值按如下规则选取：
  a. 若建表时包含主键/唯一约束，则选取HASH分布，分布列为主键/唯一约束对应的列。
  b. 若建表时不包含主键/唯一约束，则选取ROUNDROBIN分布。
- 当default_distribution_mode=hash时，DISTRIBUTE BY的默认值按如下规则选取：

 版权所有 © 华为云计算技术有限公司

a. 若建表时包含主键/唯一约束，则选取HASH分布，分布列为主键/唯一约束对应的列。
b. 若建表时不包含主键/唯一约束，但存在数据类型支持作分布列的列，则选取HASH分布，分布列为第一个数据类型支持作分布列的列。
c. 若建表时不包含主键/唯一约束，也不存在数据类型支持作分布列的列，选取ROUNDROBIN分布。

以下数据类型支持作为分布列：

- INTEGER TYPES：TINYINT，SMALLINT，INT，BIGINT，NUMERIC/DECIMAL
- CHARACTER TYPES：CHAR，BPCHAR，VARCHAR，VARCHAR2，NVARCHAR2，TEXT
- DATE/TIME TYPES：DATE，TIME，TIMETZ，TIMESTAMP，TIMESTAMPTZ，INTERVAL，SMALLDATETIME

说明

在建表时，选择分布列和分区键可对SQL查询性能产生重大影响。因此，需要根据一定策略选择合适的分布列和分区键。

- 选择合适的分布列

  对于采用散列（Hash）方式的数据分布表，一个合适的分布列应将一个表内的数据，均匀分散存储在多个DN内，避免出现数据倾斜现象（即多个DN内数据分布不均）。请按照如下原则判定合适的分布列：

  1. 判断是否已发生数据倾斜现象。

     连接数据库，执行如下语句，查看各DN内元组数目。命令中的斜体部分tablename，请填入待分析的表名。

     SELECT a.count,b.node_name FROM (SELECT count(*) AS count,xc_node_id FROM *tablename* GROUP BY xc_node_id) a, pgxc_node b WHERE a.xc_node_id=b.node_id ORDER BY a.count DESC;

     如果各DN内元组数目相差较大（如相差数倍、数十倍），则表明已发生数据倾斜现象，请按照下面原则调整分布列。

  2. 重新选择分布列，可通过ALTER TABLE语句调整分布列，选择原则如下：

     分布列的列值应比较离散，以便数据能够均匀分布到各个DN。例如，考虑选择表的主键为分布列，如在人员信息表中选择身份证号码为分布列。

     在满足上面原则的情况下，考虑选择查询中的连接条件为分布列，以便Join任务能够下推到DN中执行，且减少DN之间的通信数据量。

  3. 如果找不到一个合适的分布列，使数据能够均匀分布到各个DN，那么可以考虑使用REPLICATION或ROUNDROBIN的数据分布方式。由于REPLICATION的数据分布方式会在每个DN中存放完整的数据，因此在表较大且找不到合适的分布列时，推荐使用ROUNDROBIN的数据分布方式。（ROUNDROBIN分布方式8.1.2及以上版本支持）

- 选择合适的分区键

  数据分区功能，可根据表的一列或者多列，将要插入表的记录分为若干个范围（这些范围在不同的分区里没有重叠）。然后为每个范围创建一个分区，用来存储相应的数据。

  调整分区键，使每次查询结果尽可能存储在相同或者最少的分区内（称为“分区剪枝”），通过获取连续I/O大幅度提升查询性能。

  实际业务中，经常将时间作为查询对象的过滤条件，因此，可考虑选择时间列为分区键，键值范围可根据总数据量、一次查询数据量调整。

示例：创建存算一体表

创建表test_t4。

```
DROP TABLE IF EXISTS test_t4;
CREATE TABLE test_t4 (  ---当数据库中没有同名表（test_t4）时创建表
```

 版权所有 © 华为云计算技术有限公司

```
    col_id INT PRIMARY KEY, ---定义一个整数类型的列，并将其设为主键
    col_number INT NOT NULL, ---定义一个整数类型的列，并设置为非空
    col_date DATE NOT NULL，---定义一个日期类型的列，并设置为非空
    col_price NUMERIC(10,2)，--- 定义一个数值类型的列，允许的最大整数位数为10，小数位数为2
    col_status TEXT    ---定义一个文本类型的列
)
WITH (ORIENTATION = COLUMN,COLVERSION=2.0,COMPRESSION=middle) ---表的存储方式为列存，表的压缩级别为MIDDLE
DISTRIBUTE BY HASH (col_id)                    ---使用哈希分布
COMMENT ' order information for test_table4';   ---表注释
```

示例：创建存算分离 V3 表

创建一个存算分离V3表（仅存算分离3.0版本支持）。

```
DROP TABLE IF EXISTS test_t5;
CREATE TABLE test_t5
(
id integer not null,
data integer,
age integer
)
WITH (ORIENTATION =COLUMN, COLVERSION =3.0)
DISTRIBUTE BY ROUNDROBIN;
```

创建表的时候指定缓存策略（仅存算分离3.0版本集群支持）。

```
DROP TABLE IF EXISTS sports;
CREATE TABLE sports
(
    N_NATIONKEY  INT NOT NULL
  , N_NAME       CHAR(25) NOT NULL
  , N_REGIONKEY  INT NOT NULL
  , N_COMMENT    VARCHAR(152)
) WITH (orientation = column, colversion = 3.0, cache_policy = 'HPL: Balls, Basketball')
tablespace cu_obs_tbs
DISTRIBUTE BY ROUNDROBIN
partition by list(N_NAME)
(
  partition Balls values ('Basketball', 'football', 'badminton'),
  partition Athletics values ('High jump', 'long jump', 'javelin'),
  partition Water_Sports values ('Surfing', 'diving', 'swimming'),
  partition Shooting values ('air guns', 'Rifles', 'archery'),
  partition rest values (DEFAULT)
);
```

示例：创建 HStore 表

```
DROP TABLE IF EXISTS public.hstore_opt_table_demo;
CREATE table public.hstore_opt_table_demo(
t_code character varying(20),
t_gisid character varying(800),
t_datetime timestamp(6) without time zone,
t_gmid character varying(64)
)
WITH (orientation=column, enable_hstore_opt=on)    -- 创建表时，默认使用该配置。
DISTRIBUTE BY hash (t_gmid)   --分布键，选择主键或者关联字段
PARTITION BY range(t_datetime)    -- 分区键
(
partition p2024_1 start('2024-01-01') end ('2024-06-01') every (interval '1 month'),
partition p2024_7 start('2024-06-01') end ('2024-12-31') every (interval '1 month')
);
```

示例：创建大宽表

```
DROP TABLE IF EXISTS wide_table;
CREATE TABLE wide_table(a int,b int,c int) WITH
(orientation=column,enable_hstore_opt=on,enable_wide_table=on);
```

 版权所有 © 华为云计算技术有限公司

示例：创建表时定义约束

为表定义主键表约束，可以在表的一列或多列上定义主键表约束。

```
DROP TABLE IF EXISTS CUSTOMER;
CREATE TABLE CUSTOMER
(
   C_CUSTKEY     BIGINT       ,
   C_NAME        VARCHAR(25)  ,
   C_ADDRESS     VARCHAR(40)  ,
   C_NATIONKEY   INT          ,
   C_PHONE       CHAR(15)     ,
   C_ACCTBAL     DECIMAL(15,2)   ,
   CONSTRAINT C_CUSTKEY_KEY PRIMARY KEY(C_CUSTKEY,C_NAME)
)
DISTRIBUTE BY HASH(C_CUSTKEY,C_NAME);
```

为表定义唯一列约束。

```
DROP TABLE IF EXISTS CUSTOMER;
CREATE TABLE CUSTOMER
(
   C_CUSTKEY     BIGINT NOT NULL CONSTRAINT C_CUSTKEY_pk PRIMARY KEY,
   C_NAME        VARCHAR(25)  ,
   C_ADDRESS     VARCHAR(40)  ,
   C_NATIONKEY   INT          ,
   C_PHONE       CHAR(15)     ,
   C_ACCTBAL     DECIMAL(15,2)
)
DISTRIBUTE BY HASH(C_CUSTKEY);
```

定义CHECK表约束。

```
DROP TABLE IF EXISTS CUSTOMER;
CREATE TABLE CUSTOMER
(
   C_CUSTKEY     BIGINT NOT NULL CONSTRAINT C_CUSTKEY_pk PRIMARY KEY  ,
   C_NAME        VARCHAR(25)      ,
   C_ADDRESS     VARCHAR(40)      ,
   C_NATIONKEY   INT              ,
   CONSTRAINT C_CUSTKEY_KEY2 CHECK(C_CUSTKEY > 0 AND C_NAME <> '')
)
DISTRIBUTE BY HASH(C_CUSTKEY);
```

定义CHECK列约束。

```
DROP TABLE IF EXISTS CUSTOMER;
CREATE TABLE CUSTOMER
(
   C_CUSTKEY     BIGINT NOT NULL CONSTRAINT C_CUSTKEY_pk PRIMARY KEY  ,
   C_NAME        VARCHAR(25)  ,
   C_ADDRESS     VARCHAR(40)  ,
   C_NATIONKEY   INT NOT NULL  CHECK (C_NATIONKEY > 0)
)
DISTRIBUTE BY HASH(C_CUSTKEY);
```

使用DEFAULT为列col_STATE声明默认值为“GA”。

```
DROP TABLE IF EXISTS test_t5;
CREATE TABLE test_t5
(
   col_SK INTEGER NOT NULL,
   col_ID CHAR(16) NOT NULL,
   col_NAME VARCHAR(20) UNIQUE DEFERRABLE,
   col_WAREHOUSE INTEGER,
   col_COUNTY VARCHAR(30),
   col_STATE CHAR(2) DEFAULT 'GA',
   col_ZIP CHAR(10)
);
```

 版权所有 © 华为云计算技术有限公司

示例：创建列存表并指定压缩方式

创建列存表customer_address并指定该表的数据压缩级别为高压缩。

```
DROP TABLE IF EXISTS customer_address;
CREATE TABLE customer_address
(
    ca_address_sk       INTEGER                 NOT NULL   ,
    ca_address_id       CHARACTER(16)           NOT NULL   ,
    ca_street_number    CHARACTER(10)                      ,
    ca_street_name      CHARACTER varying(60)              ,
    ca_street_type      CHARACTER(15)                      ,
    ca_suite_number     CHARACTER(10)
)
WITH (ORIENTATION = COLUMN, COMPRESSION=HIGH,COLVERSION=2.0)
DISTRIBUTE BY HASH (ca_address_sk);
```

示例：使用 CREATE TABLE LIKE 语句创建表

使用CREATE TABLE LIKE方式快速复制现有表test_t6结构创建表test_t7。

```
DROP TABLE IF EXISTS test_t6;
CREATE TABLE  test_t6
(
    col_id INT PRIMARY KEY,
    col_number INT NOT NULL,
    col_date DATE NOT NULL,
    col_price NUMERIC(10,2),
    col_status TEXT
)
WITH (ORIENTATION = COLUMN, COMPRESSION=middle)
DISTRIBUTE BY HASH (col_id);

CREATE TABLE test_t7 (LIKE test_t6 INCLUDING ALL);
```

示例：创建加密表

创建表级加密表。

```
DROP TABLE IF EXISTS test_encrypt;
CREATE TABLE test_encrypt
(
    col_id INT PRIMARY KEY,
    col_number INT NOT NULL,
    col_date DATE NOT NULL,
    col_price NUMERIC(10,2),
    col_status TEXT
)
WITH (ORIENTATION = COLUMN, ENCRYPT_OPTION=TABLE, ENCRYPT_ALGO=aes_ctr_128)
DISTRIBUTE BY HASH (col_id);
```

创建列级加密表。

```
DROP TABLE IF EXISTS test_encrypt;
CREATE TABLE test_encrypt
(
    col_id INT PRIMARY KEY encrypt,
    col_number INT NOT NULL,
    col_date DATE NOT NULL,
    col_price NUMERIC(10,2),
    col_status TEXT
)
WITH (ORIENTATION = COLUMN, ENCRYPT_OPTION=column, ENCRYPT_ALGO=aes_ctr_128)
DISTRIBUTE BY HASH (col_id);
```

 版权所有 © 华为云计算技术有限公司

示例：创建行列混存表

以下演示如何创建一个行列混存表，注意需创建成Hstore_opt表（设置enable_hstore_opt = on）。

了解行列混存表更多知识，请参见**使用DWS行列共存表**。

步骤1 建立一个简单的数据表。

```
DROP TABLE IF EXISTS data;
CREATE TABLE data(a INT, b BIGINT, c VARCHAR(10), d VARCHAR(10));
INSERT INTO data values(generate_series(1,100),1,'asdfasdf','gergqer');
```

步骤2 扩充表数据至20W。

```
INSERT INTO data SELECT * FROM data;
INSERT INTO data SELECT * FROM data;
INSERT INTO data SELECT * FROM data;
INSERT INTO data SELECT * FROM data;
INSERT INTO data SELECT * FROM data;
INSERT INTO data SELECT * FROM data;
INSERT INTO data SELECT * FROM data;
INSERT INTO data SELECT * FROM data;
INSERT INTO data SELECT * FROM data;
INSERT INTO data SELECT * FROM data;
SELECT COUNT(*) FROM data;
```

```
gaussdb=> SELECT COUNT(*) FROM data;
 count
--------
 204800
(1 row)
```

步骤3 构造一个简单的行列共存表。

```
DROP TABLE IF EXISTS rowmode_test1;
CREATE TABLE rowmode_test1 (
    a INT,
    b BIGINT,
    c VARCHAR(10),
    d VARCHAR(10)
)
WITH (
    orientation = column,
    enable_hstore_opt = on,
    storage_mode = 'mix'
);
```

步骤4 将数据导入行列共存表。

```
INSERT INTO rowmode_test1 SELECT * FROM data;
```

----结束

示例：设置 disk_cache_ttl 参数，确保新导入数据在磁盘缓存中保留一定时间不被淘汰

该示例仅9.1.1.100及以上的存算分离集群版本支持。

在建表时，设置相应的表级参数disk_cache_ttl，即可将该表的数据使用TTL策略进行缓存。disk_cache_ttl表示，新导入的数据期望在缓存中保留的时间，参数的类型为interval，最小值为"1 second"、最大值为 "100 years"。

版权所有 © 华为云计算技术有限公司

```
DROP TABLE IF EXISTS ProductDimension;
CREATE TABLE ProductDimension
(
    ProductID INT,
    ProductName VARCHAR(100),
    Category VARCHAR(50),
    Price DECIMAL(10, 2),
    Description TEXT
)
WITH (
 orientation = column,
 colversion = 3.0,
 disk_cache_ttl = '1 day'
)
DISTRIBUTE BY ROUNDROBIN;  -- 热缓存
```

上表中，所有新导入的数据将在缓存中被保留 1天。系统当前支持修改表的 TTL 时间，用户可以根据实际需求将 TTL 的时间延长或减短。

```
ALTER TABLE ProductDimension SET (disk_cache_ttl = '1000 seconds');
```

⚠ 注意

- 修改后的 TTL 值不会对已经在缓存中的数据生效，仅对后续进入缓存的数据有效。
- 如果在建表时没有设置 TTL，用户同样可以通过执行 ALTER 语句来修改表的 TTL 属性。
- 了解更多磁盘缓存，请参见**DWS磁盘缓存最佳实践**。

示例：创建 HPN 磁盘缓存策略的分区表

cache policy选项可以在执行CREATE TABLE或者ALTER TABLE语句时设置成HPN : N，用户需要提供一个数字N(N >= -1600 && N <= 1600)，之后获取表中的数据，**更晚创建的N个分区中的数据会被热缓存**，其余分区会被冷缓存。当N等于0时，所有分区中的数据都会被冷缓存。HPN只对V3表中的range分区表和list分区表生效。

了解更多磁盘缓存，请参见**DWS磁盘缓存最佳实践**。

```
DROP TABLE IF EXISTS lineitem;
CREATE TABLE lineitem
(
 L_ORDERKEY BIGINT NOT NULL
 , L_PARTKEY BIGINT NOT NULL
 , L_SUPPKEY BIGINT NOT NULL
 , L_LINENUMBER BIGINT NOT NULL
 , L_QUANTITY DECIMAL(15,2) NOT NULL
 , L_EXTENDEDPRICE DECIMAL(15,2) NOT NULL
 , L_DISCOUNT DECIMAL(15,2) NOT NULL
 , L_TAX DECIMAL(15,2) NOT NULL
 , L_RETURNFLAG CHAR(1) NOT NULL
 , L_LINESTATUS CHAR(1) NOT NULL
 , L_SHIPDATE DATE NOT NULL
 , L_COMMITDATE DATE NOT NULL
 , L_RECEIPTDATE DATE NOT NULL
 , L_SHIPINSTRUCT CHAR(25) NOT NULL
 , L_SHIPMODE CHAR(10) NOT NULL
 , L_COMMENT VARCHAR(44) NOT NULL
) with (
 orientation = column,
 colversion = 3.0,
 cache_policy = 'HPN : 3'
)
```

 版权所有 © 华为云计算技术有限公司

```
TABLESPACE cu_obs_tbs
DISTRIBUTE BY HASH(L_ORDERKEY)
PARTITION BY RANGE(L_SHIPDATE)
(
PARTITION L_SHIPDATE_4 VALUES LESS THAN('1993-01-01 00:00:00'), -- 冷缓存
PARTITION L_SHIPDATE_5 VALUES LESS THAN('1994-01-01 00:00:00'), -- 冷缓存
PARTITION L_SHIPDATE_3 VALUES LESS THAN('1995-01-01 00:00:00'), -- 冷缓存
PARTITION L_SHIPDATE_1 VALUES LESS THAN('1996-01-01 00:00:00'), -- 冷缓存
PARTITION L_SHIPDATE_2 VALUES LESS THAN('1997-01-01 00:00:00'), -- 热缓存
PARTITION L_SHIPDATE_7 VALUES LESS THAN('1998-01-01 00:00:00'), -- 热缓存
PARTITION L_SHIPDATE_6 VALUES LESS THAN('1999-01-01 00:00:00')  -- 热缓存
);
```

示例：创建 HPL 磁盘缓存策略的分区表

cache policy选项可以在执行CREATE TABLE或者ALTER TABLE语句时设置成HPL : , , ...，用户需要提供指定为热分区的分区名，之后获取表中的数据，热分区中的数据会被热缓存，而其余分区中的数据则会被冷缓存。HPL只对V3表中的range分区表和list分区表生效。

了解更多磁盘缓存，请参见**DWS磁盘缓存最佳实践**。

```
DROP TABLE IF EXISTS nation;
CREATE TABLE nation
(
 N_NATIONKEY INT NOT NULL,
 N_NAME CHAR(25) NOT NULL,
 N_REGIONKEY INT NOT NULL,
 N_COMMENT VARCHAR(152)
) WITH (
 orientation = column,
 colversion = 3.0,
 cache_policy = 'HPL : ASIA, AMERICA, REST'
)
TABLESPACE cu_obs_tbs
DISTRIBUTE BY ROUNDROBIN
PARTITION BY LIST(N_NAME)
(
PARTITION ASIA VALUES ('INDIA', 'INDONESIA', 'JAPAN'), -- 热缓存
PARTITION EUROPE VALUES ('FRANCE', 'GERMANY', 'UNITED KINGDOM'), -- 冷缓存
PARTITION AFRICA VALUES ('ALGERIA', 'ETHIOPIA', 'KENYA'), -- 冷缓存
PARTITION AMERICA VALUES ('ARGENTINA', 'BRAZIL', 'CANADA'), -- 热缓存
PARTITION REST VALUES (DEFAULT) -- 热缓存
);
```

相关链接

ALTER TABLE，**12.101-RENAME TABLE**，**DROP TABLE**

常见问题

如何调整DWS分布列?

如果DWS建表时没有指定分布列，数据会怎么存储?

13.4.2 CREATE TABLE PARTITION

功能描述

创建分区表。逻辑上的一张表根据某种方案分成几张物理块进行存储，这张逻辑上的表称之为分区表，物理块称之为分区。分区表是一张逻辑表，不存储数据，数据实际是存储在分区上的。

 版权所有 © 华为云计算技术有限公司

常见的分区策略都是按照某一列或者某几列定义一些数据分布范围，然后每个分区承载一个范围的数据，这些列称之为**分区键**。

目前DWS行存表、列存表支持的分区类型包括**范围分区**和**列表分区**。

- **范围分区（Range Partitioning）**

  根据表的一列或者多列，将要插入表的记录分为若干个范围，这些范围在不同的分区里没有重叠。为每个范围创建一个分区，用来存储相应的数据。
  - **核心特征**：数据按连续区间分布，常用于有序且按范围检索的场景。
  - **适用场景**：分区键具有线性增长或连续区间特征，数据查询常围绕 “某一范围” 展开，且新数据会自然落入新的区间。例如按时间范围进行分区的订单数据，按数值范围分区的用户等级数据等。
  - **范围分区策略**是指记录插入分区的方式，根据分区键值将记录映射到已创建的某个分区上，如果可以映射到已创建的某一分区上，则把记录插入到对应的分区上，否则给出报错和提示信息，这是最常用的分区策略。目前范围分区仅支持范围分区策略。
- **列表分区（List Partitioning）**

  根据表的一列，将要插入表的记录通过每一个分区中出现的键值划分到对应的分区中，这些键值在不同的分区里没有重叠。为每组键值创建一个分区，用来存储相应的数据。**列表分区仅8.1.3及以上集群版本支持**。
  - **核心特征**：数据按离散的枚举值划分，常用于固定分类或高频过滤的维度。
  - **适用场景**：分区键值是固定且离散的类别，数据查询常围绕 “某一类别” 展开，键值不会无规律新增。例如按省份分区的用户信息表，每个省份对应一个分区。
  - **列表分区策略**是根据分区键值将记录映射到已创建的某个分区上，如果可以映射到已创建的某一分区上，则把记录插入到对应的分区上，否则给出报错和提示信息。

范围分区和列表分区的核心差异

表 13-54 范围分区和列表分区

维度	范围分区（Range）	列表分区（List）
数据分布	连续区间（时间、数值）	离散枚举（分类、状态）
查询优化	范围过滤（BETWEEN, >, <）	精确匹配（=, IN）
管理场景	定期增删区间（如按月份滚动）	固定分类维护（如地区、类型）
典型用例	日志、时间序列、数值归档	地域划分、状态隔离、多租户数据

分区的优势

- 某些类型的查询性能可以得到极大提升，特别是表中访问率较高的行位于一个单独分区或少数几个分区上的情况下。分区可以减少数据的搜索空间，提高数据访问效率。

 版权所有 © 华为云计算技术有限公司

- 当查询或更新一个分区的大部分记录时，连续扫描对应分区而不是访问整个表可以获得巨大的性能提升。
- 如果需要大量加载或者删除的记录位于单独的分区上，则可以通过直接读取或删除那个分区以获得巨大的性能提升，同时还可以避免由于大量DELETE导致的VACUUM超载（仅范围分区）。
- 通常情况下，单个DN数据量**超过1000W行**的场景，建议将表规划为分区表，建立完分区后，保持单DN上，单个分区的数据量控制在100W~1000W之间。

注意事项

- 单表的分区数最大为32767个，所有分区的边界值个数最大为32767个。建议不超过1000个。
- 有限地支持唯一约束和主键约束，即唯一约束和主键约束的约束键必须包含所有分区键。

语法格式

```
CREATE TABLE [ IF NOT EXISTS ] partition_table_name
( [
   { column_name data_type [ COLLATE collation ] [ column_constraint [ ... ] ]
   | table_constraint
   | LIKE source_table [ like_option [...] ] }[, ... ]
] )
   [ WITH ( {storage_parameter = value} [, ... ] ) ]
   [ COMPRESS | NOCOMPRESS ]
   [ DISTRIBUTE BY { REPLICATION | ROUNDROBIN | { [ HASH ] ( column_name ) } } ]
   [ TO { GROUP groupname | NODE ( nodename [, ... ] ) } ]
   PARTITION BY {
      {VALUES (partition_key)} |
      {RANGE (partition_key) ( partition_less_than_item [, ... ] )} |
      {RANGE (partition_key) ( partition_start_end_item [, ... ] )} |
      {LIST (partition_key) (list_partition_item [, ...])}
   } [ { ENABLE | DISABLE } ROW MOVEMENT ];
```

CREATE TABLE PARTITION 主句参数说明

表 13-55 CREATE TABLE PARTITION 主句参数说明

参数	描述	取值范围或示例
IF NOT EXISTS	如果已经存在相同名称的表，不会抛出一个错误，而会发出一个通知，告知表关系已存在。	-
partition_table_name	分区表的名称。	字符串，要符合标识符的命名规范。参见**标识符命名规范**。
column_name	新表中要创建的列名。	字符串，要符合标识符的命名规范。参见**标识符命名规范**。
data_type	列的数据类型。	-

版权所有 © 华为云计算技术有限公司

参数	描述	取值范围或示例

COLLATE collation	COLLATE子句指定列的排序规则，仅支持char、varchar、text、nchar、nvarchar这些数据类型的排序规则，排序规则主要影响查询结果的排列顺序，例如SELECT ... ORDER BY后展示的优先级。如果没有指定，则使用默认的排序规则C（即严格区分大小写）。
	常见的排序规则有：
	• C：基于二进制编码的字节顺序进行排序，完全区分大小写、重音和特殊字符（严格按ASCII 码值比较）。
	• case_insensitive：不区分大小写，8.1.3.x及以上版本支持。
	• en_US.utf8：美式英语UTF8编码，不区分大小写。
	DWS支持的排序规则可通过系统表 **PG_COLLATION** 查询：
	`SELECT * FROM pg_collation;`
	查看当前数据库默认的排序规则可使用如下语句：
	`SELECT pg_database.datcollate AS current_collation FROM pg_catalog.pg_database WHERE pg_database.datname = pg_catalog.current_database();`
		例如，对某列的取值采用不区分大小写排序规则。
		`CREATE TABLE t1 (a text collate case_insensitive);`
		`INSERT INTO t1 values('a'),('A'),('b'),('B');`
		`SELECT a, a='a' FROM t1;`
		查询时结果如下，表示A和a等价。
		a | ?column?
		---+----------
		b | f
		A | t
		a | t
		B | f
		(4 rows)

CONSTRAINT constraint_name	列约束或表约束的名字。可选的约束子句用于声明约束，新行或者更新的行必须满足这些约束才能成功插入或更新。 定义约束有两种方法： • 列约束：作为一个列定义的一部分，仅影响该列。 • 表约束：不和某个列绑在一起，可以作用于多个列。	支持的约束参见**表13-56**。
LIKE source_table [like_option ...]	LIKE子句声明一个表，新表自动从声明的表中继承所有列名及其数据类型和非空约束。	使用约束参见**LIKE参数使用说明**。
WITH	这个子句为表或索引指定一个可选的存储参数。	支持的WITH参数参见**表13-57**。

版权所有 © 华为云计算技术有限公司

参数	描述	取值范围或示例
COMPRESS / NOCOMPRESS	创建一个新表时，需要在创建表语句中指定关键字COMPRESS，这样，当对该表进行批量插入时就会触发压缩特性。该特性会在页范围内扫描所有元组数据，生成字典、压缩元组数据并进行存储。指定关键字NOCOMPRESS则不对表进行压缩。	缺省值为NOCOMPRESS，即不对元组数据进行压缩。
DISTRIBUTE BY	指定表如何在节点之间分布或者复制。	取值范围： ● REPLICATION：表的每一行存储在所有数据节点(DN)中，即每个数据节点都有完整的表数据。 ● ROUNDROBIN：表的每一行被轮询发送给各个DN，因此数据会被均匀地分布在各个DN中。（ROUNDROBIN仅8.1.2及以上版本支持） ● HASH (column_name)：对指定的列进行Hash，通过映射，把数据分布到指定DN。 其他详细使用约束参见**DISTRIBUTE BY参数使用说明**。 **注意** ● 当指定DISTRIBUTE BY HASH (column_name)参数时，创建主键和唯一索引必须包含“column_name”列。 ● 当被参照表指定DISTRIBUTE BY HASH (column_name)参数时，参照表的外键必须包含“column_name”列。
TO { GROUP groupname | NODE (nodename [, ...]) }	TO GROUP指定创建表所在的Node Group，目前不支持hdfs表使用。TO NODE主要供内部扩容工具使用，一般用户不建议使用。	-

版权所有 © 华为云计算技术有限公司

参数	描述	取值范围或示例
{ ENABLE | DISABLE } ROW MOVEMENT	行迁移开关，当更新表中某行分区键值时，如果新值属于其他分区，则会根据开关给出报错信息或者将该行数据迁移到对应的分区。 • ENABLE：打开行迁移功能，适用于频繁更新分区键的场景。 • DISABLE：关闭行迁移功能，适用于严格禁止数据跨分区更新的场景，更新分区键会报错。 默认为DISABLE，关闭。 **使用建议**： • **启用条件**：当业务需要通过更新操作改变分区键值，且性能影响可接受时使用。 • **性能考量**：ENABLE ROW MOVEMENT会增加更新操作的开销，特别是在高并发场景下需谨慎使用。 • **替代方案**：若需频繁调整分区归属，可考虑先删除原行再插入新行，或使用分区交换（**Exchange Partition**）操作。 **注意** 开启ROW MOVEMENT则允许跨分区更新，但此时如果有SELECT FOR UPDATE查询该分区表并发执行，存在查询结果瞬时不一致的可能性，需要谨慎使用。	举例，某分区表的日期列（分区键），按季度进行分区，分为分区p_2023q1和p_2023q2，某一行的“2023-02-15”原属于第一季度的分区p_2023q1，当更新该值为“2023-05-15”后，对应的这行数据，会根据ROW MOVEMENT开关是否打开而决定是否行数据迁移： • ENABLE：打开行迁移功能，则该数据迁移到新分区p_2023q2。 • DISABLE：关闭行迁移功能，则系统会报错。 详细示例参见**示例：开启和关闭行迁移功能ROW MOVEMENT**

列约束 CONSTRAINT constraint_name 参数说明

```
{ NOT NULL |
  NULL |
  CHECK ( expression ) |
  DEFAULT default_expr |
  UNIQUE [ NULLS [NOT] DISTINCT | NULLS IGNORE ] index_parameters |
  PRIMARY KEY index_parameters }
[ DEFERRABLE | NOT DEFERRABLE | INITIALLY DEFERRED | INITIALLY IMMEDIATE ]
```


版权所有 © 华为云计算技术有限公司

表 13-56 CONSTRAINT constraint_name 参数说明

参数	描述	取值范围或示例
NOT NULL	字段值不允许为NULL。 ENABLE用于语法兼容，可省略。	-
NULL	字段值允许NULL ，这是默认值，该子句只是为和非标准SQL数据库兼容。不建议使用。	-

CHECK (condition) [NO INHERIT]	CHECK约束声明一个布尔表达式，每次要插入的新行或者要更新的行的新值必须使表达式结果为真或未知才能成功，否则会报错并且执行失败。
	声明为字段约束的检查约束应该只引用该字段的数值，而在表约束里出现的表达式可以引用多个字段。
	用NO INHERIT标记的约束将不会传递到子表中去。
		单个字段约束:
		价格必须为正数：
		`CREATE TABLE products ( product_id INT, price DECIMAL(10,2) CHECK (price > 0), discount DECIMAL(3,2) );`
		插入一条价格为负数的数据：
		`INSERT INTO products (product_id, price, discount) VALUES (2, -10.00, 1.1);`
		违反约束，则报错，插入不成功。
		`ERROR: dn_6003_6004: new row for relation "products" violates check constraint "products_price_check"`
		表约束（约束多个字段）：
		创建一个库存表，分两个字段，可用量和预定量，且可用量必须大于等于预定量：
		`CREATE TABLE inventory ( item_id INT PRIMARY KEY, quantity_on_hand INT, quantity_reserved INT, CHECK (quantity_on_hand >= quantity_reserved));`

DEFAULT default_expr	DEFAULT子句给字段指定缺省值。该数值可以是任何不含变量的表达式(不允许使用子查询和对本表中的其他字段的交叉引用)。缺省表达式的数据类型必须和字段类型匹配。 缺省表达式将被用于任何未声明该字段数值的插入操作。若未显式指定DEFAULT，则默认值为NULL。	-

 版权所有 © 华为云计算技术有限公司

参数	描述	取值范围或示例
UNIQUE [NULLS [NOT] DISTINCT | NULLS IGNORE] index_parameters UNIQUE [NULLS [NOT] DISTINCT | NULLS IGNORE] (column_name [, ...]) index_parameters	UNIQUE约束表示表里的一个字段或多个字段的组合必须在全表范围内唯一。 其中[NULLS [NOT] DISTINCT | NULLS IGNORE]字段用来指定Unique唯一索引中索引列NULL值的处理方式。	默认取值：该参数默认缺省，即NULL值可重复插入。 在对插入的新数据和表中原始数据进行列的等值比较时，对于NULL值有以下三种处理方式： • NULLS DISTINCT：将每个NULL值视为不同的值，即允许NULL值重复插入。 • NULLS NOT DISTINCT：将所有NULL值视为相同的值，即索引中只保留一个NULL值条目，不允许NULL值重复插入，只有非NULL值且与原索引列中已有数据不相等的数据可插入。 • NULLS IGNORE：在等值比较时跳过NULL值。若索引列全为NULL，则NULL值可重复插入；部分索引列为NULL，只有非NULL值不相等，才可成功插入数据。 三种处理方式具体的行为如**表13-61**所示。 更多使用方法可以参见**关于NULLS DISTINCT、NULLS NOT DISTINCT、NULLS IGNORE使用上的场景差异**。 **注意** 如果没有声明DISTRIBUTE BY REPLICATION，则唯一约束的列集合中必须包含分布列。
PRIMARY KEY index_parameters PRIMARY KEY (column_name [, ...]) index_parameters	主键约束声明表中的一个或者多个字段只能包含唯一的非NULL值。 一个表只能声明一个主键。 **注意** 如果没有声明DISTRIBUTE BY REPLICATION，则主键约束的列集合中必须包含分布列。	-

版权所有 © 华为云计算技术有限公司

参数	描述	取值范围或示例
DEFERRABLE | NOT DEFERRABLE	设置该约束是否可延迟。**分区表不支持该参数，DEFERRABLE仅用于非分区表。**一个不可延迟的约束将在每条命令之后马上检查。可延迟约束可以延迟到事务结尾使用**SET CONSTRAINTS**命令检查。缺省是NOT DEFERRABLE。目前，行存的UNIQUE约束和主键约束可以接受这个子句。所有其他约束类型都是不可延迟的。	-
INITIALLY IMMEDIATE | INITIALLY DEFERRED	如果约束是可延迟的，则这个子句声明检查约束的缺省时间。 ● 如果约束是INITIALLY IMMEDIATE（缺省），则在每条语句执行之后就立即检查它； ● 如果约束是INITIALLY DEFERRED ，则只有在事务结尾才检查它。 约束检查的时间可以用SET CONSTRAINTS命令修改。	-

表约束 table_constraint 参数说明

表约束table_constraint，参数解释同列约束相似不再赘述，参见**列约束CONSTRAINT constraint_name参数说明**。

```
{ CHECK ( expression ) |
  UNIQUE [ NULLS [NOT] DISTINCT | NULLS IGNORE ] ( column_name [, ... ] ) index_parameters |
  PRIMARY KEY ( column_name [, ... ] ) index_parameters}
[ DEFERRABLE | NOT DEFERRABLE | INITIALLY DEFERRED | INITIALLY IMMEDIATE ]
```

LIKE 参数使用说明

like选项like_option：

```
{ INCLUDING | EXCLUDING } { DEFAULTS | CONSTRAINTS | INDEXES | STORAGE | COMMENTS |
RELOPTIONS | DISTRIBUTION | ALL }
```

新表与原来的表之间在创建动作完毕之后是完全无关的。在源表做的任何修改都不会传播到新表中，并且也不可能在扫描源表的时候包含新表的数据。

字段缺省表达式只有在声明了INCLUDING DEFAULTS之后才会包含进来。缺省是不包含缺省表达式的，即新表中所有字段的缺省值都是NULL。

非空约束将总是复制到新表中，CHECK约束则仅在指定了INCLUDING CONSTRAINTS的时候才复制，而其他类型的约束则永远也不会被复制。此规则同时适用于表约束和列约束。

被复制的列和约束并不使用相同的名字进行融合。如果明确指定了相同的名字或者在另外一个LIKE子句中，将会报错。

 版权所有 © 华为云计算技术有限公司

- 如果指定了INCLUDING INDEXES，则源表上的索引也将在新表上创建，默认不建立索引。
- 如果指定了INCLUDING STORAGE，则拷贝列的STORAGE设置也将被拷贝，默认情况下不包含STORAGE设置。
- 如果指定了INCLUDING COMMENTS，则源表列、约束和索引的注释也会被拷贝过来。默认情况下，不拷贝源表的注释。
- 如果指定了INCLUDING RELOPTIONS，则源表的存储参数（即源表的WITH子句）也将拷贝至新表。默认情况下，不拷贝源表的存储参数。
- 如果指定了INCLUDING DISTRIBUTION，则新表将拷贝源表的分布信息，包括分布类型和分布列，同时新表将不能再使用DISTRIBUTE BY子句。默认情况下，不拷贝源表的分布信息。
- INCLUDING ALL是INCLUDING DEFAULTS INCLUDING CONSTRAINTS INCLUDING INDEXES INCLUDING STORAGE INCLUDING COMMENTS INCLUDING RELOPTIONS INCLUDING DISTRIBUTION的简写形式。

WITH 子句参数说明

表 13-57 WITH 参数说明

参数	描述	取值范围
FILLFACTOR	一个表的填充因子（fillfactor）是一个介于10和100之间的百分数。100（完全填充）是默认值。如果指定了较小的填充因子，INSERT操作仅按照填充因子指定的百分率填充表页。每个页上的剩余空间将用于在该页上更新行，这就使得UPDATE有机会在同一页上放置同一条记录的新版本，这比把新版本放置在其他页上更有效。对于一个从不更新的表将填充因子设为100是优良选择，但是对于频繁更新的表，选择较小的填充因子则更加合适。该参数对于列存表没有意义。	10~100
ORIENTATION	决定了数据的存储方式。	• COLUMN：表的数据将以列式存储。 • ROW（缺省值）：表的数据将以行式存储。 • ORC：表的数据将以ORC格式存储（仅HDFS表）。 **注意** orientation不支持修改。

 版权所有 © 华为云计算技术有限公司

参数	描述	取值范围
COMPRESSION	该参数仅列存表支持，有效值为：LOW、MIDDLE或HIGH。默认值为LOW。 **说明** 暂不支持行存表压缩功能。	-
MAX_BATCHROW	指定了在数据加载过程中一个存储单元可以容纳记录的最大数目。该参数只对列存表有效。	10000~60000 默认值：60000
PARTIAL_CLUSTER_ROWS	指定了在数据加载过程中进行局部聚簇存储的记录数目。该参数只对列存表有效。	其有效值为大于等于10万。此值是MAX_BATCHROW的倍数。
enable_delta	指定列存表是否开启delta表。该参数只对列存表有效。不推荐使用列存带Delta表，因为会出现MERGE处理不及时而导致的磁盘膨胀以及性能劣化等问题。	默认值：off
DELTAROW_THRESHOLD	预留参数。该参数只对列存表有效。	0～60000，默认值为6000
COLD_TABLESPACE	指定冷分区保存的OBS表空间，仅冷热表支持。该参数仅支持列存分区表，且该参数不支持修改，需与STORAGE_POLICY同时使用。在指定STORAGE_POLICY时，可不设置该参数，默认为default_obs_tbs。	有效的OBS TABLESPACE名。

版权所有 © 华为云计算技术有限公司

参数	描述	取值范围
STORAGE_POLICY	指定冷热分区切换规则，仅冷热表支持。该参数需与COLD_TABLESPACE同时使用。	"冷热切换策略名称:冷热切换的阈值"，目前冷热切换的策略名称只支持LMT和HPN，LMT指按分区的最后更新时间切换，HPN指保留热分区的个数切换。 ● LMT：[day]：表示切换[day]时间前修改的热分区数据为冷分区，将该数据迁至OBS表空间中。其中[day]为整型，范围[0, 36500]，单位为天。 ● HPN:[hot_partition_num]：表示保留[hot_partition_num]个有数据的分区为热分区。保留规则为查找出有数据的分区的最大的Sequence ID，大于Sequence ID的无数据分区为热分区，并按这个Sequence ID从大到小保留[hot_partition_num]个分区为热分区；分区Sequence ID小于保留的最小热分区的Sequence ID的分区为冷分区，在冷热切换时，需要将数据迁移至OBS表空间中。其中[hot_partition_num]为整型，范围为[0,1600]。 **注意** ● 对于LIST分区，建议谨慎使用HPN策略，否则可能出现新增分区不是热分区的情况。
ENABLE_MULTI_TEMP_DISKCACHE	冷热表性能优化参数，仅支持冷热表设置，该参数仅9.1.0及以上版本支持。	on/true，off/false 默认值：on/true

版权所有 © 华为云计算技术有限公司

参数	描述	取值范围
PERIOD	指定分区管理中自动创建分区的周期，并开启自动创建分区功能。仅支持行存、列存范围分区表、时序表以及冷热表；分区键唯一并且类型仅支持TIMESTAMP[(p)] [WITHOUT TIME ZONE]、TIMESTAMP[(p)] [WITH TIME ZONE]、DATE；不支持存在maxvalue分区；(nowTime - boundaryTime) / PERIOD需要小于分区个数上限，其中nowTime为当前时间，boundaryTime为现有分区中最早的分区边界时间；不支持在小型机、加速集群上使用。	1 hour ~ 100 years **注意** • 在兼容Teradata或MySQL的数据库中，分区键类型为DATE时，PERIOD不能小于1 day。 • 建分区表时，如果设置了PERIOD，则可以只指定分区键不指定分区。建表时将创建两个默认分区，这两个默认分区的分区时间范围均为PERIOD。其中，第一个默认分区的边界时间是大于当前时间的第一个整时/整天/整周/整月/整年的时间，具体选择哪种整点时间取决于PERIOD的最大单位；第二个默认分区的边界时间是第一个分区边界时间加PERIOD。假设当前时间是2022-02-17 16:32:45，各种情况的第一个默认分区的分区边界选择如**表13-60**。 有关默认分区的更多内容，请参见**示例：创建不指定分区的自动分区管理分区表，指定分区管理中自动创建分区的周期PERIOD为1 day，分区键为time**。

版权所有 © 华为云计算技术有限公司

参数	描述	取值范围
TTL	指定分区管理中分区过期的时间，并开启自动删除分区功能。不支持单独设置，必须要提前或同时设置PERIOD，并且要大于或等于PERIOD。 **TTL（Time To Live）**指明该表的数据保存周期，超过TTL周期的数据将被清理，具体做法是以period周期，自动调用**proc_drop_partition (relname regclass, older_than interval)**函数。PERIOD和TTL的值为Interval类型，例如：“1 hour”，“1 day”，“1 week”，“1 month”，“1 year”，“1 month 2 day 3 hour”。 PERIOD指明按照时间划分的周期对数据进行分区，分区的大小可能对查询性能有影响，同时每隔周期时间会创建一个新的周期大小的分区，具体做法是以period周期，自动调用**proc_add_partition (relname regclass, boundaries_interval interval)**函数。	1 hour ~ 100 years **注意** 自动创建与自动删除分区功能是在创建自动分区任务时开始执行，例如：2026-3-12 15:30:00创建自动分区管理的表(TTL=7 days, PERIOD = 1 day)，从2026-3-12 15:30:00开始，每天15:30:00都会触发自动创建与自动删除分区。

版权所有 © 华为云计算技术有限公司

参数	描述	取值范围
time_format	自增减分区支持INT4/INT8/VARCHAR/TEXT列，存储内容为时间时使用。time_format选项仅当分区键为INT4/INT8/VARCHAR/TEXT时，同时指定period才生效。该参数仅9.1.0.200及以上集群版本支持。	1. time_format写法遵从DWS规范，例如：time_format = 'yyyymmdd'。 2. 分区键的类型**VARCHAR/TEXT**对time_format的规格有如下限制： ● 精度支持到秒级。 ● 输入内容不包括字母类型，例如month，am，pm等。 ● 时间必须从大到小排列，例如年月日小时分钟秒。 **INT4/INT8**对time_format的规格有如下限制： ● 精度支持到小时级。 ● 输入内容仅包括Y,M,D,HH24。 ● 时间必须从大到小排列，例如年月日小时。 3. ALTER相关限制 ● 不支持set。 ● 当period被reset时（表示已经关闭自动分区，会报出提示），此时可以reset此选项。

 版权所有 © 华为云计算技术有限公司

参数	描述	取值范围
COLVERSION	指定列存存储格式的版本，支持不同存储格式版本之间的切换，但分区表不支持存储格式版本切换。	• 1.0：列存表的每列以一个单独的文件进行存储，文件名以relfilenode.C1.0、relfilenode.C2.0、relfilenode.C3.0等命名。 • 2.0：列存表的每列合并存储在一个文件中，文件名以relfilenode.C1.0命名。 默认值：2.0 需注意，OBS冷热表仅支持colversion 2.0格式。 **注意** 在建列存表时选择COLVERSION=2.0，相比于1.0存储格式，在以下场景中性能有明显提升： 1. 创建列存宽表场景下，建表时间显著减少。 2. roach备份数据场景下，备份时间显著减少。 3. build、catch up耗时显著减少。
SKIP_FPI_HINT	顺序扫描过程中，若需要写FPW(full page writes)日志时，该参数控制是否跳过设置HintBits操作。	默认值：false **注意** 设置SKIP_FPI_HINT=true时，在对某表执行checkpoint操作后，若对该表进行顺序扫描，将不再产生Xlog。适用于查询次数较少的中间表，有效减少Xlog的大小，提升查询性能。

PARTITION BY 子句参数说明

- partition_less_than_item：
  ```
  PARTITION partition_name VALUES LESS THAN ( { partition_value | MAXVALUE } )
  ```
- partition_start_end_item：
  ```
  PARTITION partition_name {
        {START(partition_value) END (partition_value) EVERY (interval_value)} |
        {START(partition_value) END ({partition_value | MAXVALUE})} |
        {START(partition_value)} |
        {END({partition_value | MAXVALUE})}
  }
  ```
- list_partition_item：
  ```
  PARTITION partition_name VALUES ( { (partition_value) [, ...] | DEFAULT } )
  ```

 版权所有 © 华为云计算技术有限公司

表 13-58 PARTITION BY 子句参数说明

参数	描述	取值范围
PARTITION BY RANGE(partition_key)	指定范围分区策略语法，partition_key为分区键的名称。	• 对于从句是VALUES LESS THAN的语法格式： **注意** 对于从句是VALUES LESS THAN的语法格式，范围分区策略的分区键最多支持4列，且分区键只能是列名。当存在多个分区键时，一个列名只能出现一次，且相邻的两个分区键要使用逗号隔开。 该情形下，分区键支持的数据类型为：SMALLINT、INTEGER、BIGINT、DECIMAL、NUMERIC、REAL、DOUBLE PRECISION、CHARACTER VARYING(n)、VARCHAR(n)、CHARACTER(n)、CHAR(n)、CHARACTER、CHAR、TEXT、NVARCHAR2、NAME、TIMESTAMP[(p)] [WITHOUT TIME ZONE]、TIMESTAMP[(p)] [WITH TIME ZONE]、DATE。 • 对于从句是START END的语法格式： **注意** 对于从句是START END的语法格式，范围分区策略的分区键仅支持1列。 该情形下，分区键支持的数据类型为：SMALLINT、INTEGER、BIGINT、DECIMAL、NUMERIC、REAL、DOUBLE PRECISION、TIMESTAMP[(p)] [WITHOUT TIME ZONE]、TIMESTAMP[(p)]

 版权所有 © 华为云计算技术有限公司

参数	描述	取值范围
		[WITH TIME ZONE]、DATE。
PARTITION BY LIST (partition_key,[...])	指定列表分区策略语法，partition_key为分区键的名称。	列表分区策略分区键支持的数据类型为：TINYINT、SMALLINT、INTEGER、BIGINT、NUMERIC/DECIMAL、TEXT、NVARCHAR2、VARCHAR(n)、CHAR、BPCHAR、TIME、TIME WITH TIMEZONE、TIMESTAMP、TIMESTAMP WITH TIME ZONE、DATE、INTERVAL、SMALLDATETIME。 **注意** 列表分区策略的分区键最多支持4列。
partition_less_than_item	范围分区策略下分区（简称为范围分区）的定义语法。partition_name为范围分区的名称。partition_value为范围分区的上边界，取值依赖于partition_key的类型。MAXVALUE表示分区的上边界，它通常用于设置最后一个范围分区的上边界。 `PARTITION partition_name VALUES LESS THAN ( { partition_value | MAXVALUE } )` **注意** • 每个分区都需要指定一个上边界。 • 分区上边界的类型应当和分区键的类型一致。 • 分区列表是按照分区上边界升序排列的，值较小的分区位于值较大的分区之前。 • 如果分区键由多个字段组成，比较大小时，先比较第一个字段，当第一个字段相等时比较第二个字段，以此类推。	-

 版权所有 © 华为云计算技术有限公司

参数	描述	取值范围
partition_start_end_item	使用起始值以及间隔值定义范围分区的语法。 PARTITION partition_name {START (partition_value) END (partition_value) EVERY (interval_value)} | {START (partition_value) END (partition_value| MAXVALUE)} | {START(partition_value)} | {END (partition_value| MAXVALUE)}	各参数含义如下： ● partition_name：范围分区的名称或名称前缀，除以下情形外（假定其中的partition_name是p1），均为分区的名称。 – 若该定义是START+END+EVERY从句，则语义上定义的分区的名称依次为p1_1, p1_2, ...。例如对于定义“PARTITION p1 START(1) END(4) EVERY(1)”，则生成的分区是：[1, 2), [2, 3) 和 [3, 4)，名称依次为p1_1, p1_2和p1_3，即此处的p1是名称前缀。 – 若该定义是第一个分区定义，且该定义有START值，则范围（MINVALUE, START）将自动作为第一个实际分区，其名称为p1_0，然后该定义语义描述的分区名称依次为p1_1，p1_2，……。例如，对于完整定义“PARTITION p1 START(1), PARTITION p2 START(2)”，则生成的分区是：(MINVALUE, 1)，[1, 2) 和[2, MAXVALUE)，其名称依次为p1_0，p1_1和p2，即此处p1是名称前缀，p2是分区名称。这里MINVALUE表示最小值。 ● partition_value：范围分区的端点值（起始或终点），取值依赖于

 版权所有 © 华为云计算技术有限公司

参数	描述	取值范围
		partition_key的类型，不可是MAXVALUE。 ● interval_value：对[START，END) 表示的范围进行切分，interval_value是指定切分后每个分区的宽度，不可是MAXVALUE；如果（END-START）值不能整除以EVERY值，则仅最后一个分区的宽度小于EVERY值。 ● MAXVALUE：表示最大值，它通常用于设置最后一个范围分区的上边界。

 版权所有 © 华为云计算技术有限公司

参数	描述	取值范围
		注意 1. 在创建分区表若第一个分区定义含START值，则范围（MINVALUE，START）将自动作为实际的第一个分区。 2. START END语法需要遵循以下限制： • 每个partition_start_end_item中的START值（若存在，下同）必须小于其END值； • 相邻的两个partition_start_end_item，第一个的END值必须等于第二个的START值； • 每个partition_start_end_item中的EVERY值必须是正向递增的，且必须小于（END-START）值； • 每个分区包含起始值，不包含终点值，即形如：[起始值，终点值)，起始值是MINVALUE时则不包含； • 一个partition_start_end_item创建的每个分区所属的TABLESPACE一样； • partition_name作为分区名称前缀时，其长度不要超过57字节，超过时自动截断； • 在创建、修改分区表时请注意分区表的分区总数不可超过最大限制（32767）； 3. 在创建分区表时START END与LESS THAN语法不可混合使用。 4. 即使创建分区表时使用START END语法，备份（gs_dump）出的SQL语句也是VALUES LESS THAN语法格式。

版权所有 © 华为云计算技术有限公司

参数	描述	取值范围
list_partition_item	列表分区策略下分区（简称为列表分区）的定义语法。partition_name为分区的名称。partition_value为列表分区边界的一个枚举值，取值依赖于partition_key的类型。DEFAULT表示默认分区的边界。 PARTITION partition_name VALUES ({ (partition_value) [, ...] | DEFAULT })	对于列表分区表，存在以下约定和约束： • 边界值为DEFAULT的分区，称之为默认分区。 • 每个列表分区表只能有一个DEFAULT分区。 • 分区表的所有分区数不超过32767个，所有分区的边界值个数不大于32767个。 • 不管分区键的个数，DEFAULT分区的边界只能是一个DEFAULT。 • 如果分区键由多个字段组成，每个partition_value需要包含所有分区键的值，当分区键只有一列时，partition_value两侧的括号可以省略，参见**示例：创建多个分区键的list分区表**。 • 如果分区键由多个字段组成，比较大小时，先逐个字段比较大小，任何一个字段值不一样即可认为是不一样的键值。 • 边界中不同的partition_value值不能重复。 • 数据插入时，如果数据的分区键值能匹配任何非DEFAULT分区的边界，那么数据会写入对应的分区；否则数据会写入DEFAULT分区。

表 13-59 唯一索引中索引列 NULL 值的处理方式

字段控制	索引列全为NULL	部分索引列为NULL
NULLS DISTINCT	可重复插入	可重复插入
NULLS NOT DISTINCT	不可重复插入	非NULL值相等，不可插入；非NULL值不相等，则插入成功

版权所有 © 华为云计算技术有限公司

字段控制	索引列全为NULL	部分索引列为NULL
NULLS IGNORE	可重复插入	非NULL值相等，不可插入；非NULL值不相等，则插入成功

表 13-60 分区边界选择

period	period最大单位	第一个默认分区的分区边界
1hour	Hour	2022-02-17 17:00:00
1day	Day	2022-02-18 00:00:00
1month	Month	2022-03-01 00:00:00
13month	Year	2023-01-01 00:00:00

表 13-61 唯一索引中索引列 NULL 值的处理方式

字段控制	索引列全为NULL	部分索引列为NULL
NULLS DISTINCT	可重复插入	可重复插入
NULLS NOT DISTINCT	不可重复插入	非NULL值相等，不可插入；非NULL值不相等，则插入成功
NULLS IGNORE	可重复插入	非NULL值相等，不可插入；非NULL值不相等，则插入成功

表 13-62 分区边界选择

period	period最大单位	第一个默认分区的分区边界
1hour	Hour	2022-02-17 17:00:00
1day	Day	2022-02-18 00:00:00
1month	Month	2022-03-01 00:00:00
13month	Year	2023-01-01 00:00:00

DISTRIBUTE BY 参数使用说明

默认值：由GUC参数default_distribution_mode控制。

- 当default_distribution_mode=roundrobin时，DISTRIBUTE BY的默认值按如下规则选取：

 版权所有 © 华为云计算技术有限公司

a. 若建表时包含主键/唯一约束，则选取HASH分布，分布列为主键/唯一约束对应的列。

b. 若建表时不包含主键/唯一约束，则选取ROUNDROBIN分布。

- 当default_distribution_mode=hash时，DISTRIBUTE BY的默认值按如下规则选取：

  a. 若建表时包含主键/唯一约束，则选取HASH分布，分布列为主键/唯一约束对应的列。

  b. 若建表时不包含主键/唯一约束，但存在数据类型支持作分布列的列，则选取HASH分布，分布列为第一个数据类型支持作分布列的列。

  c. 若建表时不包含主键/唯一约束，也不存在数据类型支持作分布列的列，选取ROUNDROBIN分布。

以下数据类型支持作为分布列：

- INTEGER TYPES：TINYINT，SMALLINT，INT，BIGINT，NUMERIC/DECIMAL
- CHARACTER TYPES：CHAR，BPCHAR，VARCHAR，VARCHAR2，NVARCHAR2，TEXT
- DATE/TIME TYPES：DATE，TIME，TIMETZ，TIMESTAMP，TIMESTAMPTZ，INTERVAL，SMALLDATETIME

示例：使用 LESS THAN 语法创建 range 分区表

Range分区表customer_address含有4个分区，分区键为integer类型。分区的范围分别为：ca_address_sk<2450815，2450815<= ca_address_sk< 2451179，2451179<= ca_address_sk< 2451544，2451544<=ca_address_sk。

```
DROP TABLE IF EXISTS customer_address;
CREATE TABLE customer_address
(
    ca_address_sk       INTEGER                 NOT NULL   ,
    ca_address_id       CHARACTER(16)           NOT NULL   ,
    ca_street_number    CHARACTER(10)                      ,
    ca_street_name      CHARACTER varying(60)              ,
    ca_street_type      CHARACTER(15)                      ,
    ca_suite_number     CHARACTER(10)
)
DISTRIBUTE BY HASH (ca_address_sk)
PARTITION BY RANGE(ca_address_sk)
(
        PARTITION P1 VALUES LESS THAN(2450815),
        PARTITION P2 VALUES LESS THAN(2451179),
        PARTITION P3 VALUES LESS THAN(2451544),
        PARTITION P4 VALUES LESS THAN(MAXVALUE)
);
```

查看分区表信息：

```
SELECT relname, boundaries FROM pg_partition p where p.parentid='customer_address'::regclass ORDER BY 1;
```

```
gaussdb=> SELECT relname, boundaries FROM pg_partition p where p.parentid='customer_address'::regclass ORDER BY 1;
     relname      | boundaries
------------------+------------
 customer_address |
 p1               | {2450815}
 p2               | {2451179}
 p3               | {2451544}
 p4               | {NULL}
(5 rows)
```

查询分区P1的行数：

 版权所有 © 华为云计算技术有限公司

```
SELECT count(*) FROM customer_address PARTITION (P1);
SELECT count(*) FROM customer_address PARTITION FOR (2450815);
```

示例：开启和关闭行迁移功能 ROW MOVEMENT

在更新某行的分区键为新值且新值属于其他分区时，如果打开ROW MOVEMENT，则该行数据对应迁移到新的分区中。

频繁更新分区键时，ENABLE ROW MOVEMENT 可能导致额外I/O开销，需评估业务需求。

创建按日期范围分区的表，并启用行迁移：

```
DROP TABLE IF EXISTS sale_data;
CREATE TABLE  sales_data (
    sale_id  INT,
    product_name VARCHAR(100),
    sale_date DATE,
    amount DECIMAL(10, 2)
)
PARTITION BY RANGE (sale_date) (
    PARTITION p_2023q1 VALUES LESS THAN ('2023-04-01'),
    PARTITION p_2023q2 VALUES LESS THAN ('2023-07-01'),
    PARTITION p_2023q3 VALUES LESS THAN ('2023-10-01'),
    PARTITION p_2023q4 VALUES LESS THAN ('2024-01-01')
)
ENABLE ROW MOVEMENT;
```

插入测试数据：

```
INSERT INTO sales_data (sale_id, product_name, sale_date, amount)
VALUES
(1, 'Product A', '2023-02-15', 1000.00),  -- 属于 p_2023q1
(2, 'Product B', '2023-05-20', 1500.00);  -- 属于 p_2023q2
```

查询各分区数据分布：

```
SELECT tableoid::regclass AS partition, * FROM sales_data ORDER BY sale_id;
```

```
 partition  | sale_id | product_name |      sale_date      | amount
------------+---------+--------------+---------------------+---------
 2147617615 |       1 | Product A    | 2023-02-15 00:00:00 | 1000.00
 2147617616 |       2 | Product B    | 2023-05-20 00:00:00 | 1500.00
(2 rows)
```

尝试更新分区键值（从 p_2023q1 移动到 p_2023q2）：

```
UPDATE sales_data SET sale_date = '2023-05-01'  WHERE sale_id = 1;
```

再次查询各分区数据分布，确认行已移动

```
SELECT tableoid::regclass AS partition, *  FROM sales_data  ORDER BY sale_id;
```

```
 partition  | sale_id | product_name |      sale_date      | amount
------------+---------+--------------+---------------------+---------
 2147617616 |       1 | Product A    | 2023-05-01 00:00:00 | 1000.00
 2147617616 |       2 | Product B    | 2023-05-20 00:00:00 | 1500.00
(2 rows)
```

关闭行迁移功能：

```
ALTER TABLE sales_data DISABLE ROW MOVEMENT;
```


版权所有 © 华为云计算技术有限公司

再次修改分区键（从 p_2023q2 移动回到 p_2023q2）：

```
UPDATE sales_data SET sale_date = '2023-02-15'  WHERE sale_id = 1;
```

回显提示，行迁移关闭，不允许更新分区键。

```
gaussdb=> UPDATE sales_data SET sale_date = '2023-02-15'  WHERE sale_id = 1;
ERROR:  dn_6003_6004: fail to update partitioned table "sales_data"
DETAIL:  disable row movement
```

示例：使用 START END 语法创建列存 range 分区表

```
DROP TABLE IF EXISTS customer_address_SE;
CREATE TABLE customer_address_SE
(
    ca_address_sk       INTEGER               NOT NULL   ,
    ca_address_id       CHARACTER(16)           NOT NULL   ,
    ca_street_number    CHARACTER(10)                      ,
    ca_street_name      CHARACTER varying(60)              ,
    ca_street_type      CHARACTER(15)                     ,
    ca_suite_number     CHARACTER(10)
)
WITH (ORIENTATION = COLUMN)
DISTRIBUTE BY HASH (ca_address_sk)
PARTITION BY RANGE(ca_address_sk)
(
    PARTITION p1 START(1) END(1000) EVERY(200),
    PARTITION p2 END(2000),
    PARTITION p3 START(2000) END(5000)
);
```

查看分区表信息：

```
SELECT relname, boundaries FROM pg_partition p where p.parentid='customer_address_SE'::regclass ORDER BY 1;
```

```
gaussdb=> SELECT relname, boundaries FROM pg_partition p where p.parentid='customer_address_SE'::regclass ORDER BY 1;
       relname       | boundaries
---------------------+------------
customer_address_se |
p1_0                | {1}
p1_1                | {201}
p1_2                | {401}
p1_3                | {601}
p1_4                | {801}
p1_5                | {1000}
p2                  | {2000}
p3                  | {5000}
(9 rows)
```

示例：START END 语法创建、修改 Range 分区表

创建临时schema：

```
DROP SCHEMA IF EXISTS tpcds;
CREATE SCHEMA tpcds;
SET CURRENT_SCHEMA TO tpcds;
```

创建分区表，分区键是integer类型：

```
CREATE TABLE tpcds.startend_pt (c1 INT, c2 INT)
DISTRIBUTE BY HASH (c1)
PARTITION BY RANGE (c2) (
    PARTITION p1 START(1) END(1000) EVERY(200),
    PARTITION p2 END(2000),
    PARTITION p3 START(2000) END(2500),
    PARTITION p4 START(2500),
    PARTITION p5 START(3000) END(5000) EVERY(1000)
)
ENABLE ROW MOVEMENT;
```

 版权所有 © 华为云计算技术有限公司

查看分区表信息：

```
SELECT relname, boundaries FROM pg_partition p where p.parentid='tpcds.startend_pt'::regclass ORDER BY 1;
```

```
gaussdb=> SELECT relname, boundaries FROM pg_partition p where p.parentid='tpcds.startend_pt'::regclass ORDER BY 1;
   relname   | boundaries
-------------+------------
 p1_0        | {1}
 p1_1        | {201}
 p1_2        | {401}
 p1_3        | {601}
 p1_4        | {801}
 p1_5        | {1000}
 p2          | {2000}
 p3          | {2500}
 p4          | {3000}
 p5_1        | {4000}
 p5_2        | {5000}
 startend_pt |
(12 rows)
```

导入数据，查看分区数据量：

```
INSERT INTO tpcds.startend_pt VALUES (GENERATE_SERIES(0, 4999), GENERATE_SERIES(0, 4999));
SELECT COUNT(*) FROM tpcds.startend_pt PARTITION FOR (0);
SELECT COUNT(*) FROM tpcds.startend_pt PARTITION (p3);
```

```
gaussdb=> SELECT COUNT(*) FROM tpcds.startend_pt PARTITION FOR (0);
 count
-------
     1
(1 row)
```

```
gaussdb=> SELECT COUNT(*) FROM tpcds.startend_pt PARTITION (p3);
 count
-------
   500
(1 row)
```

查看分区表信息：

```
SELECT relname, boundaries FROM pg_partition p where p.parentid='tpcds.startend_pt'::regclass ORDER BY 1;
```

```
gaussdb=> SELECT relname, boundaries FROM pg_partition p where p.parentid='tpcds.startend_pt'::regclass ORDER BY 1;
   relname   | boundaries
-------------+------------
 p1_0        | {1}
 p1_1        | {201}
 p1_2        | {401}
 p1_3        | {601}
 p1_4        | {801}
 p1_5        | {1000}
 p2          | {2000}
 p3          | {2500}
 p4          | {3000}
 p5_1        | {4000}
 p5_2        | {5000}
 startend_pt |
(12 rows)
```

示例：创建一个分区键的 list 分区表

```
DROP TABLE IF EXISTS data_list;
CREATE TABLE data_list
(
    id int,
    time int,
    salary decimal(12,2)
)
PARTITION BY LIST (time)
(
        PARTITION P1 VALUES (202209),
        PARTITION P2 VALUES (202210,202208),
```

 版权所有 © 华为云计算技术有限公司

```
    PARTITION P3 VALUES (202211),
    PARTITION P4 VALUES (202212),
    PARTITION P5 VALUES (202301)
);
```

示例：创建多个分区键的 list 分区表

分区表有两个分区键：period, city。

```
DROP TABLE IF EXISTS sales_info;
CREATE TABLE sales_info
(
sale_time  timestamptz,
period     int,
city       text,
price      numeric(10,2),
remark     varchar2(100)
)
DISTRIBUTE BY HASH(sale_time)
PARTITION BY LIST (period, city)
(
PARTITION north_2022 VALUES (('202201', 'north1'), ('202202', 'north2')),
PARTITION south_2022 VALUES (('202201', 'south1'), ('202202', 'south2'), ('202203', 'south2')),
PARTITION rest VALUES (DEFAULT)
);
```

示例：创建不指定分区的自动分区管理分区表，指定分区管理中自动创建分区的周期 PERIOD 为 1 day，分区键为 time

```
DROP TABLE IF EXISTS time_part;
CREATE TABLE time_part (id integer,time timestamp) with (PERIOD='1 day') partition by range(time);
```

建表时将创建两个默认分区，第一个默认分区的边界时间是大于当前时间的第一个整天的时间，即2022-12-13 00:00:00；第二个默认分区的边界时间是第一个分区边界时间加PERIOD，即2022-12-13 00:00:00+1day=2022-12-14 00:00:00。

```
SELECT now();
SELECT relname, boundaries FROM pg_partition p where p.parentid='time_part'::regclass ORDER BY 1;
```

```
gaussdb=> SELECT now();
              now
-------------------------------
 2026-05-06 20:38:44.76038+08
(1 row)
```

```
gaussdb=> SELECT relname, boundaries FROM pg_partition p where p.parentid='time_part'::regclass ORDER BY 1;
    relname     |       boundaries
----------------+------------------------
 default_part_1 | {"2026-05-07 00:00:00"}
 default_part_2 | {"2026-05-08 00:00:00"}
 time_part      |
(3 rows)
```

示例：创建指定分区的自动分区管理分区表

```
DROP TABLE IF EXISTS CPU;
CREATE TABLE CPU(
    id integer,
    idle numeric,
    IO numeric,
    scope text,
    IP text,
    time timestamp
) with (TTL='7 days',PERIOD='1 day')
partition by range(time)
(
    PARTITION P1 VALUES LESS THAN('2022-01-05 16:32:45'),
```

 版权所有 © 华为云计算技术有限公司

```
    PARTITION P2 VALUES LESS THAN('2022-01-06 16:56:12')
);
```

示例：创建指定分区的自动分区管理分区表(非标准 timestamp 时间类型的分区列)

当分区列的数据类型非标准timestamp，而为VARCHAR、TEXT、INT或BIGINT类型时，需要通过设置表级参数**time_format**指定分区列中的时间格式。该参数用于指示系统如何解析分区列中的时间值，以便进行自动分区管理。time_format选项仅当分区键为INT4/INT8/VARCHAR/TEXT，同时也指定period时才能成功设置。

创建分区管理表CPU2和CPU3并指定分区，分区列分别为INT、VARCHAR类型。

```
-- INT类型
DROP TABLE IF EXISTS CPU2;
CREATE TABLE CPU2(
    id integer,
    IP text,
    time integer
) with (TTL='7 days',PERIOD='1 day', TIME_FORMAT='YYYYMMDD')
partition by range(time)
(
    PARTITION P1 VALUES LESS THAN('20230213'),
    PARTITION P2 VALUES LESS THAN('20230215')
);

-- VARCHAR类型
DROP TABLE IF EXISTS CPU3;
CREATE TABLE CPU3(
    id integer,
    IP text,
    time varchar
) with (TTL='7 days',PERIOD='1 day', TIME_FORMAT='YYYY-MM-DD HH24:MI:SS')
partition by range(time)
(
    PARTITION P1 VALUES LESS THAN('2023-02-13 16:32:45'),
    PARTITION P2 VALUES LESS THAN('2023-02-15 16:48:12')
);
```

示例：按照月份创建分区表 customer_address，含有 13 个分区，分区键为 date 类型

创建分区表customer_address：

```
DROP TABLE IF EXISTS customer_address;
CREATE TABLE customer_address
(
    ca_address_sk       integer           NOT NULL,
    ca_address_date       date            NOT NULL
)
DISTRIBUTE BY HASH (ca_address_sk)
PARTITION BY RANGE (ca_address_date)
(
        PARTITION p202001 VALUES LESS THAN('20200101'),
        PARTITION p202002 VALUES LESS THAN('20200201'),
        PARTITION p202003 VALUES LESS THAN('20200301'),
        PARTITION p202004 VALUES LESS THAN('20200401'),
        PARTITION p202005 VALUES LESS THAN('20200501'),
        PARTITION p202006 VALUES LESS THAN('20200601'),
        PARTITION p202007 VALUES LESS THAN('20200701'),
        PARTITION p202008 VALUES LESS THAN('20200801'),
        PARTITION p202009 VALUES LESS THAN('20200901'),
        PARTITION p202010 VALUES LESS THAN('20201001'),
        PARTITION p202011 VALUES LESS THAN('20201101'),
        PARTITION p202012 VALUES LESS THAN('20201201'),
        PARTITION p202013 VALUES LESS THAN(MAXVALUE)
);
```


版权所有 © 华为云计算技术有限公司

插入数据：

```
INSERT INTO customer_address values('1','20200215');
INSERT INTO customer_address values('7','20200805');
INSERT INTO customer_address values('9','20201111');
INSERT INTO customer_address values('4','20201231');
```

查询分区：

```
SELECT * FROM customer_address PARTITION(p202009);
```

```
gaussdb=> SELECT * FROM customer_address PARTITION(p202009);
 ca_address_sk |   ca_address_date
---------------+---------------------
             7 | 2020-08-05 00:00:00
(1 row)
```

示例：使用 START END 语法一次创建含有多个分区的分区表

- 创建分区表day_part，每一天为一个分区，分区键为date类型。

  ```
  DROP TABLE IF EXISTS day_part;
  CREATE TABLE day_part(id int,d_time date)
  DISTRIBUTE BY HASH (id)
  PARTITION BY RANGE (d_time)
  (PARTITION p1 START('2022-01-01') END('2022-01-31') EVERY(interval '1 day'));
  ALTER TABLE  day_part ADD PARTITION pmax VALUES LESS THAN (maxvalue);
  ```

- 创建分区表week_part，每7天为一个分区，分区键为date类型。

  ```
  DROP TABLE IF EXISTS week_part;
  CREATE TABLE week_part(id int,w_time date)
  DISTRIBUTE BY HASH (id)
  PARTITION BY RANGE (w_time)
  (PARTITION p1 START('2021-01-01') END('2022-01-01') EVERY(interval '7 day'));
  ALTER TABLE  week_part ADD PARTITION pmax VALUES LESS THAN (maxvalue);
  ```

- 创建分区表month_part，每一个月为一个分区，分区键为date类型。

  ```
  DROP TABLE IF EXISTS month_part;
  CREATE TABLE month_part(id int,m_time date)
  DISTRIBUTE BY HASH (id)
  PARTITION BY RANGE (m_time)
  (PARTITION p1 START('2021-01-01') END('2022-01-01') EVERY(interval '1 month'));
  ALTER TABLE  month_part ADD PARTITION pmax VALUES LESS THAN (maxvalue);
  ```

示例：创建冷热表

仅支持列存分区表，使用obs默认表空间，冷热切换规则设置LMT为30。

```
DROP TABLE IF EXISTS cold_hot_table;
CREATE TABLE cold_hot_table
(
    W_WAREHOUSE_ID          CHAR(16)            NOT NULL,
    W_WAREHOUSE_NAME          VARCHAR(20)                ,
    W_STREET_NUMBER          CHAR(10)                   ,
    W_STREET_NAME           VARCHAR(60)                ,
    W_STREET_ID             CHAR(15)                   ,
    W_SUITE_NUMBER           CHAR(10)
)
WITH (ORIENTATION = COLUMN, storage_policy = 'LMT:30')
DISTRIBUTE BY HASH (W_WAREHOUSE_ID)
PARTITION BY RANGE(W_STREET_ID)
(
    PARTITION P1 VALUES LESS THAN(100000),
    PARTITION P2 VALUES LESS THAN(200000),
    PARTITION P3 VALUES LESS THAN(300000),
    PARTITION P4 VALUES LESS THAN(MAXVALUE)
)ENABLE ROW MOVEMENT;
```


版权所有 © 华为云计算技术有限公司

相关链接

ALTER TABLE PARTITION，**DROP TABLE**

常见问题

- **使用实践：使用DWS分区自动管理功能降低电商和物联网行业数据分区维护成本**
- 使用实践：**创建和管理DWS分区表**
- 使用实践：**案例：改建分区表**
- **分区表插入数据报错：inserted partition key does not map to any table partition**
- **范围分区表添加新分区报错upper boundary of adding partition MUST overtop last existing partition**

13.4.3 CREATE TABLE AS

功能描述

根据查询结果创建表。适用于数据迁移或复制场景。

CREATE TABLE AS创建一个表，并使用SELECT查询结果填充该表。该表的字段与SELECT输出字段的名称及数据类型关联。不过用户可以通过明确地给出一个字段名称列表来覆盖SELECT输出字段的名称。

CREATE TABLE AS对源表进行一次查询，然后将数据写入新表中，而查询视图结果会根据源表的变化而有所改变。相比之下，每次做查询的时候，视图都重新计算定义它的SELECT语句。

注意事项

- 分区表不能采用此方式进行创建。
- 如果在建表过程中数据库系统发生故障，系统恢复后可能无法自动清除之前已创建的、大小非0的磁盘文件。此种情况出现概率小，不影响数据库系统的正常运行。
- UNLOGGED表无主备机制，在系统故障或异常断点等情况下，会有数据丢失风险，不可用来存储基础数据。
- UNLOGGED表不能保证数据的安全性，用户应该在确保数据已经做好备份的前提下使用，例如系统升级时进行数据的备份。

语法格式

```
CREATE [ [ GLOBAL | LOCAL | VOLATILE ] { TEMPORARY | TEMP } | UNLOGGED ] TABLE table_name
   [ (column_name [, ...] ) ]
   [ WITH ( {storage_parameter = value} [, ... ] ) ]
   [ COMPRESS | NOCOMPRESS ]

   [ DISTRIBUTE BY { REPLICATION | ROUNDROBIN | { [HASH ] ( column_name ) } } ]

   [ COMMENT [=] 'text' ]
   AS query
   [ WITH [ NO ] DATA ];
```


版权所有 © 华为云计算技术有限公司

参数说明

表 13-63 CREATE TABLE AS 参数说明

参数	描述	取值范围
GLOBAL | LOCAL | VOLATILE	指定临时表类型。	具体参考**参数说明**。
TEMPORARY | TEMP	指定TEMP或TEMPORARY关键字时创建的表为临时表。	-
UNLOGGED	指定表为非日志表。在非日志表中写入的数据不会被写入到预写日志中，这样就会比普通表快很多。但是，它也是不安全的，非日志表在冲突或异常关机后会被自动删截。非日志表中的内容也不会被复制到备用服务器中。在该类表中创建的索引也不会被自动记录。 当异常关机等操作导致非日志表上的索引发生数据丢失时，用户应该对发生错误的索引进行重建。	自9.1.0版本，UNLOGGED表默认存储在pg_unlogged表空间下，且不可迁移或指定到其他表空间；从低版本升级到9.1.0版本后，旧版本创建的UNLOGGED表还存储在原表空间下。 由于实例异常重启时会对UNLOGGED表进行重置，所以可能影响实例RTO。自9.1.0版本提供UNLOGGED表迁移脚本switch_unlogged_tablespace.py，配合GUC参数enable_unlogged_tablespace_compat，可以优化RTO。
table_name	要创建的表名。	字符串，需符合**标识符命名规范**。
column_name	新表中要创建的字段名。	字符串，需符合**标识符命名规范**。
WITH (storage_parameter [= value] [, ...])	WITH子句设置表或其索引的存储选项。	参数的详细说明如**表13-64**所示。
COMPRESS | NOCOMPRESS	• 指定关键字COMPRESS，对表进行批量插入时就会触发压缩特性。该特性会在页范围内扫描所有元组数据，生成字典、压缩元组数据并进行存储。 • 指定关键字NOCOMPRESS则不对表进行压缩。 缺省值：NOCOMPRESS，即不对元组数据进行压缩	-

版权所有 © 华为云计算技术有限公司

参数	描述	取值范围
DISTRIBUTE BY	指定表如何在节点之间分布或者复制。 ● REPLICATION：表的每一行存储在所有数据节点(DN)中，即每个数据节点都有完整的表数据。 ● ROUNDROBIN：表的每一行被依次发送给各个DN，在这种分布策略下可以保证数据分布不会存在倾斜，但是因为数据分布节点是随机的，导致这类表在计算时会更大概率的触发此表的重分布。各列倾斜都比较严重的大表推荐使用此种分布策略。（ ROUNDROBIN仅8.1.2及以上版本支持 ） ● HASH (column_name)：对指定的列进行Hash，通过映射，把数据分布到指定DN。 默认值：由GUC参数default_distribution_mode控制。DISTRIBUTE BY的默认值选取规则请参考此表下方的须知内容。	以下数据类型支持作为分布列： ● INTEGER TYPES：TINYINT，SMALLINT，INT，BIGINT，NUMERIC/DECIMAL ● CHARACTER TYPES：CHAR，BPCHAR，VARCHAR，VARCHAR2，NVARCHAR2，TEXT ● DATE/TIME TYPES：DATE，TIME，TIMETZ，TIMESTAMP，TIMESTAMPTZ，INTERVAL，SMALLDATETIME
COMMENT [=] 'text'	COMMENT子句可以在创建表时指定表注释。	-
AS query	一个SELECT VALUES命令或者一个运行预备好的SELECT或VALUES查询的EXECUTE命令。	-
[WITH [NO] DATA]	创建表时，是否也插入查询到的数据。默认是要数据，选择“NO”参数时，则不要数据。	-

 版权所有 © 华为云计算技术有限公司

须知

DISTRIBUTE BY参数说明：

- 当指定DISTRIBUTE BY HASH (column_name)参数时，创建主键和唯一索引必须包含“column_name”列。
- 当被参照表指定DISTRIBUTE BY HASH (column_name)参数时，参照表的外键必须包含“column_name”列。
- 当default_distribution_mode=roundrobin时，DISTRIBUTE BY的默认值按如下规则选取：
  1. 若建表时包含主键/唯一约束，则选取HASH分布，分布列为主键/唯一约束对应的列。
  2. 若建表时不包含主键/唯一约束，则选取ROUNDROBIN分布。
- 当default_distribution_mode=hash时，DISTRIBUTE BY的默认值按如下规则选取：
  1. 若建表时包含主键/唯一约束，则选取HASH分布，分布列为主键/唯一约束对应的列。
  2. 若建表时不包含主键/唯一约束，但存在数据类型支持作分布列的列，则选取HASH分布，分布列为第一个数据类型支持作分布列的列。
  3. 若建表时不包含主键/唯一约束，也不存在数据类型支持作分布列的列，选取ROUNDROBIN分布。

表 13-64 storage_parameter 参数说明

storage_parameter 参数	描述	取值范围
FILLFACTOR	一个表的填充因子（fillfactor）是一个介于10和100之间的百分数。如果指定了较小的填充因子，INSERT操作仅按照填充因子指定的百分率填充表页。每个页上的剩余空间将用于在该页上更新行，这就使得UPDATE有机会在同一页上放置同一条记录的新版本，这比把新版本放置在其他页上更有效。对于一个从不更新的表将填充因子设为100是合适的选择，但是对于频繁更新的表，选择较小的填充因子则更加合适。 默认值为100，表示完全填充。 该参数只对行存表有效。	10~100

版权所有 © 华为云计算技术有限公司

storage_parameter 参数	描述	取值范围
ORIENTATION	指定表数据的存储方式，即行存方式、列存方式，该参数设置成功后就不再支持修改。 默认值：ROW，即行存方式。	• ROW，表示表的数据将以行式存储。 行存储适合于OLTP业务，此类型的表上交互事务比较多，一次交互会涉及表中的多个列，用行存查询效率较高。 • COLUMN，表示表的数据将以列式存储。 列存储适合于数据仓库业务，此类型的表上会做大量的汇聚计算，且涉及的列操作较少。
COMPRESSION	指定表数据的压缩级别，它决定了表数据的压缩比以及压缩时间。一般来讲，压缩级别越高，压缩比也越大，压缩时间也越长；反之亦然。实际压缩比取决于加载的表数据的分布特征。 **说明** 暂不支持行存表压缩功能。	该参数仅列存表支持，有效值为：LOW、MIDDLE或HIGH。默认值为LOW。
MAX_BATCHROW	指定了在数据加载过程中一个存储单元可以容纳记录的最大数目。该参数只对列存表有效。 默认值：60000	10000~60000
PARTIAL_CLUSTER_ROWS	指定了在数据加载过程中进行局部聚簇存储的记录数目。该参数只对列存表有效。 默认值：4,200,000	600000~2147483647
enable_delta	指定了在列存表是否开启delta表。该参数只对列存表有效。 不推荐使用列存带Delta表，因为会出现MERGE处理不及时而导致的磁盘膨胀以及性能劣化等问题。 默认值：off	on或off

 版权所有 © 华为云计算技术有限公司

storage_parameter 参数	描述	取值范围
COLVERSION	指定列存存储格式的版本，支持不同存储格式版本之间的切换。 默认值：2.0 **说明** 在建列存表时选择COLVERSION=2.0，相比于1.0存储格式，在以下场景中性能有明显提升： 1. 创建列存宽表场景下，建表时间显著减少。 2. roach备份数据场景下，备份时间显著减少。 3. build、catch up耗时显著减少。	取值范围： ● 1.0：列存表的每列以一个单独的文件进行存储，文件名以relfilenode.C1.0、relfilenode.C2.0、relfilenode.C3.0等命名。 ● 2.0：列存表的每列合并存储在一个文件中，文件名以relfilenode.C1.0命名
SKIP_FPI_HINT	顺序扫描过程中，若需要写FPW（full page writes）日志时，该参数控制是否跳过设置HintBits操作。 默认值：false **说明** 设置SKIP_FPI_HINT=true时，在对某表执行checkpoint操作后，若对该表进行顺序扫描，将不再产生Xlog。适用于查询次数较少的中间表，有效减少Xlog的大小，提升查询性能。	true或false

示例

创建表customer。

```
DROP TABLE IF EXISTS customer CASCADE;
CREATE TABLE customer
(
   C_CUSTKEY     BIGINT NOT NULL CONSTRAINT C_CUSTKEY_pk PRIMARY KEY ,
   C_NAME        VARCHAR(25) ,
   C_ADDRESS     VARCHAR(40) ,
   C_NATIONKEY   INT NOT NULL  CHECK (C_NATIONKEY > 0)
)
DISTRIBUTE BY HASH(C_CUSTKEY);
```

创建表store_returns_t1并插入CUSTOMER表中C_CUSTKEY字段中大于4795的数值。

```
DROP TABLE IF EXISTS store_returns_t1;
CREATE TABLE store_returns_t1 AS SELECT * FROM CUSTOMER WHERE C_CUSTKEY > 4795;
```

使用store_returns_t1复制一个新表store_returns_t2。

```
DROP TABLE IF EXISTS store_returns_t2;
CREATE TABLE store_returns_t2 AS table store_returns_t1;
```

 版权所有 © 华为云计算技术有限公司

相关链接

CREATE TABLE，SELECT

13.4.4 ALTER TABLE

功能描述

修改表，包括修改表的定义、重命名表、重命名表中指定的列、重命名表的约束、设置表的所属模式、添加/更新多个列、打开/关闭行访问控制开关。

只有表的所有者或者被授予了表ALTER权限的用户有权限执行ALTER TABLE命令，系统管理员默认拥有此权限。但要修改表的所有者或者修改表的模式，当前用户必须是该表的所有者或者系统管理员，且该用户是新的所有角色的直接或间接成员。

注意事项

表 13-65 ALTER TABLE 注意事项

约束项	详细内容
不支持修改表数据的存储方式	建表后不支持修改存储参数ORIENTATION（该参数指定表为行存或列存）。
列存表约束支持范围	列存表支持PARTIAL CLUSTER KEY，不支持外键表级约束。列存表自8.1.1集群版本开始建表时支持创建主键和唯一表级约束。
列存表约束修改限制	• 支持通过ALTER方式对已有列存表添加非空约束NOT NULL以及主键约束。该约束仅8.2.0及以上集群版本支持。 • 通过约束名删除PRIMARY KEY约束时，不会删除NOT NULL约束。如有需要，可手动删除NOT NULL约束。
列存表的字段约束	列存表支持的字段约束包括NULL、NOT NULL和DEFAULT常量值；对字段约束的修改，当前支持对DEFAULT值的修改（SET DEFAULT）、删除（DROP DEFAULT）和NOT NULL约束的删除。
列存表支持操作	• 添加字段ADD COLUMN，对于添加的字段和修改的字段类型要求是列存支持的数据类型。 • 修改字段的数据类型ALTER TYPE，ALTER TYPE的USING选项只支持常量表达式和涉及本字段的表达式，暂不支持涉及其他字段的表达式。 • 设置单个字段的收集目标SET STATISTICS。 • 支持更改表名称。 • 支持删除字段DROP COLUMN。
分布列修改限制	不支持修改已有表的分布列数据类型。
行存REPLICATION表限制	行存REPLICATION分布表不支持将系统列设置为主键。

版权所有 © 华为云计算技术有限公司

约束项	详细内容
不支持增加自增列	不支持增加自增列，或者增加DEFAULT值中包含nextval()表达式的列。
OBS冷热表修改约束	● OBS冷热表不支持ALTER RESET的cold_tablespace，storage_policy参数，不支持修改COLVERSION为1.0。 ● 可将原列存COLVERSION为2.0表ALTER为OBS冷热表，需同时增加cold_tablespace和storage_policy参数。 ● ALTER TABLE支持针对RELOPTIONS的storage_policy的选项更改。冷热切换策略发生更改后，不改变现有冷数据的冷热属性，只改变下一次执行冷热切换的规则，下一次执行冷热切换命令时将按新规则进行冷热切换。
V3表	V3表（COLVERSION为3.0）不支持修改COLVERSION，也不支持将非V3表切换为V3表（即COLVERSION为2.0不支持切为3.0）。
触发表重建的操作	对表执行以下ALTER TABLE操作时会触发表重建（表重建过程中会先把数据转储到一个新的数据文件中，重建完成之后会删除原始文件），当表比较大时，重建会消耗较多的磁盘空间。当磁盘空间不足时，要谨慎对待大表ALTER TABLE操作，防止触发集群只读。 ● 修改列数据类型。 ● 行存表增加列（包括oid列）。 ● 列存表修改COLVERSION。 ● 列存表增加列指定DEFAULT常量值，DEFAULT值中包含volatile函数以及DEFAULT值非NULL且不属于特定数据类型中的任何一个。
加密表	● 加密表语法仅9.1.1.200及以上集群版本支持。 ● 加密表语法仅行存表、列存2.0表支持。 ● **集群级透明加密和细粒度透明加密功能互斥，无法同时开启**。已经开启了集群级透明加密，则无法开启细粒度透明加密功能。 ● 加密表分为表级加密和列级加密。 表级加密：设置encrypt_option=table时，表示全表进行加密。 列级加密：设置encrypt_option=column时，表示对表进行列级加密。若没有明确指定加密列，表示表中没有加密列。 ● 行存表不支持列级的透明加密配置，即encrypt_option=column。 ● 索引不支持修改透明加密配置。 ● 支持非加密表转加密表，加密表转非加密表。 ● 支持新增加密列。

版权所有 © 华为云计算技术有限公司

约束项	详细内容
其他	• 修改列存表存储参数COLVERSION或者enable_delta时，不能与其他ALTER操作同时进行。 • 对unlogged表执行ALTER TABLE时，不允许指定表空间。对非unlogged表执行ALTER TABLE时，不允许指定表空间为pg_unlogged。 • 全局临时表不支持修改分布列。

警告

- 避免在业务高峰期执行ALTER TABLE/ALTER TABLE PARTITION（增删改查、DROP PARTITION）、TRUNCATE操作，避免有长SQL阻塞ALTER、TRUNCATE操作或SQL业务被ALTER、TRUNCATE阻塞。
- 更多开发设计规范参见**总体开发设计规范**。

ALTER TABLE 语法格式

```
ALTER TABLE [ IF EXISTS ] { table_name [*] | ONLY table_name | ONLY ( table_name ) }
    action [, ... ];
```

- 修改表已存在字段的数据类型，可同时修改空约束，default值和注释信息，只允许相同大类的类型转换（数值之间，字符串之间，时间之间等）。

  ```
  ALTER TABLE [ IF EXISTS ] table_name
      MODIFY ( { column_name data_type | [ CONSTRAINT constraint_name ] NOT NULL [ ENABLE ] |
          [ CONSTRAINT constraint_name ] NULL | DEFAULT default_expr | COMMENT 'text' } [, ...] );
  ```

- 重命名表。对名称的修改不会影响所存储的数据；不支持跨schema对表进行重命名操作；支持新表名前带有原表的schema名；如果new_table_name不指定schema，则表示new_table_name与原表处于同一schema中。

  ```
  ALTER TABLE [ IF EXISTS ] table_name
      RENAME TO new_table_name;
  ALTER TABLE [ IF EXISTS ] table_name
      RENAME TO schema.new_table_name;
  ```

- 重命名表中指定的列。

  ```
  ALTER TABLE [ IF EXISTS ] { table_name [*] | ONLY table_name | ONLY ( table_name )}
      RENAME [ COLUMN ] column_name TO new_column_name;
  ```

- 重命名表的约束。

  ```
  ALTER TABLE [ IF EXISTS ] { table_name  [*] | ONLY table_name | ONLY ( table_name  )}
      RENAME CONSTRAINT constraint_name TO new_constraint_name;
  ```

- 设置表的所属模式。

  ```
  ALTER TABLE [ IF EXISTS ] table_name
      SET SCHEMA new_schema;
  ```


版权所有 © 华为云计算技术有限公司

说明

- SET SCHEMA操作不支持修改为系统内部模式，当前仅支持用户模式之间的修改。
- 这种形式把表移动到另外一个模式。相关的索引、约束都跟着移动。目前序列不支持改变schema。 若该表拥有序列，需要将序列删除，重建，或者取消拥有关系， 才能将表schema更改成功。
- 要修改一个表的模式，用户必须在新模式上拥有CREATE权限。要把该表添加为一个父表的新子表，用户必须同时又是父表的所有者。要修改所有者，用户还必须是新的所有角色的直接或间接成员，并且该成员必须在此表的模式上有CREATE权限。这些限制规定了该用户不能执行重建和删除表以外的操作，但系统管理员能以任何方式修改任意表的所有权限。
- 除了RENAME和SET SCHEMA之外所有动作都可以捆绑在一个经过多次修改的列表中并行使用。比如，可以在一个命令里增加几个字段或修改几个字段的类型。对于大表，此种操作带来的效率提升更明显，原因在于只需要对该大表做一次处理。
- 增加一个CHECK或NOT NULL约束将会扫描该表，以保证现有的行符合约束要求。
- 用一个非空缺省值增加一个字段或者改变一个字段的现有类型会重写整个表。对于大表来说，这个操作可能会花很长时间，并且它还临时需要两倍的磁盘空间。

- 设置表已存在字段的加密属性。

```
ALTER TABLE [ IF EXISTS ] table_name
    MODIFY ( { column_name to encrypt[decrypt] } [, ...] );
```

- 修改表的加密属性。

```
ALTER TABLE [ IF EXISTS ] table_name
    SET ( encrypt_option=table[column|off] } );
```

- 修改表的加密算法。

```
ALTER TABLE [ IF EXISTS ] table_name
    SET ( encrypt_algo=aes_ctr_128[sm4_ctr_128] } );
```

- 向列存表中添加加密列。

```
ALTER TABLE [ IF EXISTS ] table_name
    ADD ( { column_name data_type [ compress_mode ] [ COLLATE collation ] [ column_constraint
[ ... ] ] [encrypt | decrypt]} [, ...] );
```

- 添加多个列。

```
ALTER TABLE [ IF EXISTS ] table_name
    ADD ( { column_name data_type [ compress_mode ] [ COLLATE collation ] [ column_constraint
[ ... ] ]} [, ...] );
```

- 更新多个列。

```
ALTER TABLE [ IF EXISTS ] table_name
    MODIFY ( { column_name data_type | column_name [ CONSTRAINT constraint_name ] NOT NULL
[ ENABLE ] | column_name [ CONSTRAINT constraint_name ] NULL } [, ...] );
```

- 删除指定表上的索引。

```
ALTER TABLE table_name DROP INDEX index_name;
```

表操作相关的 action 子句

具体表操作action可以是以下子句之一：

```
column_clause
    | ADD table_constraint [ NOT VALID ]
    | ADD table_constraint_using_index
    | VALIDATE CONSTRAINT constraint_name
    | DROP CONSTRAINT [ IF EXISTS ]  constraint_name [ RESTRICT | CASCADE ]
    | CLUSTER ON index_name
    | SET WITHOUT CLUSTER
    | SET ( {storage_parameter = value} [, ... ] )
    | RESET ( storage_parameter [, ... ] )
    | OWNER TO new_owner
    | SET TABLESPACE new_tablespace
    | SET {COMPRESS|NOCOMPRESS}
    | DISTRIBUTE BY { REPLICATION | ROUNDROBIN | { HASH ( column_name [,...] ) } }
    | TO { GROUP groupname | NODE ( nodename [, ... ] ) }
```

 版权所有 © 华为云计算技术有限公司

```
| ADD NODE ( nodename [, ... ] )
| DELETE NODE ( nodename [, ... ] )
| DISABLE TRIGGER [ trigger_name | ALL | USER ]
| ENABLE TRIGGER [ trigger_name | ALL | USER ]
| ENABLE REPLICA TRIGGER trigger_name
| ENABLE ALWAYS TRIGGER trigger_name
| DISABLE ROW LEVEL SECURITY
| ENABLE ROW LEVEL SECURITY
| FORCE ROW LEVEL SECURITY
| NO FORCE ROW LEVEL SECURITY
| REFRESH STORAGE
| REFRESH DISTRIBUTION
```

- **ADD table_constraint [NOT VALID]**

  给表增加一个新的约束。与NOT VALID选项组合时，这种约束仅在对外键和CHECK类约束有效。 如果约束条件增加了NOT VALID选项后，表中已有记录是否满足初始约束检查会被跳过。

- **ADD table_constraint_using_index**

  根据已有唯一索引为表增加主键约束或唯一约束。

- **VALIDATE CONSTRAINT constraint_name**

  验证一个外键或是一个使用NOT VALID选项创建的检查类约束，通过扫描全表来保证所有记录都符合约束条件。如果约束已标记为有效时，什么操作也不会发生。

- **DROP CONSTRAINT [IF EXISTS] constraint_name [RESTRICT | CASCADE]**

  删除一个表上的约束。

- **CLUSTER ON index_name**

  为将来的CLUSTER操作选择默认索引。实际上并没有重新盘簇化处理该表。

- **SET WITHOUT CLUSTER**

  从表中删除最新使用的CLUSTER索引。这样会影响将来那些没有声明索引的集群操作。

- **SET ({storage_parameter = value} [, ...])**

  修改表的一个或多个存储参数。

- **RESET (storage_parameter [, ...])**

  重置表的一个或多个存储参数。与SET一样，根据参数的不同可能需要重写表才能获得想要的效果。

- **OWNER TO new_owner**

  将表、序列、视图的属主改变成指定的用户。

- **SET {COMPRESS|NOCOMPRESS}**

  修改表的压缩特性。表压缩特性的改变只会影响后续批量插入的数据的存储方式，对已有数据的存储毫无影响。也就是说，表压缩特性的修改会导致该表中同时存在着已压缩和未压缩的数据。

- **DISTRIBUTE BY { REPLICATION | ROUNDROBIN | { HASH (column_name [,...]) } }**

  修改表的分布方式，在修改表分布信息的同时会将表数据在物理上按新分布方式重新分布，修改完成后建议对被修改表执行ANALYZE，以便收集全新的统计信息。

 版权所有 © 华为云计算技术有限公司

说明

- 本操作属于重大变更操作，涉及表分布信息的修改以及数据的物理重分布，修改过程中会阻塞业务，修改完成后原有业务的执行计划会发生变化，请按照正规变更流程进行。
- 本操作属于资源密集操作，针对大表的分布方式修改，建议在计算和存储资源充裕情况下进行，保证整个集群和原表所在表空间有足够的剩余空间能存储一张与原表同等大小且按照新分布方式进行分布的表。

- **TO { GROUP groupname | NODE (nodename [, ...]) }**

  此语法仅在扩展模式（GUC参数enable_cluster_resize为on时）下可用。该模式谨慎打开，主要供内部扩容工具使用，一般用户不应使用该模式。

- **ADD NODE (nodename [, ...])**

  此语法主要供内部扩容工具使用，一般用户不建议使用。

- **DELETE NODE (nodename [, ...])**

  此语法主要供内部缩容工具使用，一般用户不建议使用。

- **DISABLE TRIGGER [trigger_name | ALL | USER]**

  禁用trigger_name所表示的单个触发器，或禁用所有触发器，或仅禁用用户触发器（此选项不包括内部生成的约束触发器，例如，可延迟唯一性和排除约束的约束触发器）。

  说明

  应谨慎使用此功能，因为如果不执行触发器，则无法保证原先期望的约束的完整性。

- **ENABLE TRIGGER [trigger_name | ALL | USER]**

  启用trigger_name所表示的单个触发器，或启用所有触发器，或仅启用用户触发器。

- **ENABLE REPLICA TRIGGER trigger_name**

  触发器触发机制受配置变量session_replication_role的影响，当复制角色为“origin”（默认值）或“local”时，将触发简单启用的触发器。

  配置为ENABLE REPLICA的触发器仅在会话处于“replica”模式时触发。

- **ENABLE ALWAYS TRIGGER trigger_name**

  无论当前复制模式如何，配置为ENABLE ALWAYS的触发器都将触发。

- **DISABLE/ENABLE ROW LEVEL SECURITY**

  开启或关闭表的行访问控制开关。

  当开启行访问控制开关时，如果未在该数据表定义相关行访问控制策略，数据表的行级访问将不受影响；如果关闭表的行访问控制开关，即使定义了行访问控制策略，数据表的行访问也不受影响。详细信息参见**CREATE ROW LEVEL SECURITY POLICY**章节。

  说明

  不支持对HDFS表、外表、临时表开启行访问控制开关。

- **NO FORCE/FORCE ROW LEVEL SECURITY**

  强制开启或关闭表的行访问控制开关。

  默认情况，表所有者不受行访问控制特性影响，但当强制开启表的行访问控制开关时，表的所有者（不包含系统管理员用户）会受影响。系统管理员可以绕过所有的行访问控制策略，不受影响。

- **REFRESH STORAGE**

 版权所有 © 华为云计算技术有限公司

根据OBS冷热表storage_policy所定义的规则，将符合条件的本地热分区切换为存储在OBS上的冷分区。

例如创建OBS冷热表时，设置storage_policy为 'LMT:10'，则在执行该操作时可将10日前无修改的分区切为冷存储，存至OBS中。

- **REFRESH DISTRIBUTION**

  根据表的分布列信息和DN节点信息，刷新pgxc_class系统表里的记录信息，用于数据分布的一致性校验，会同步刷新CN和DN的信息。该参数仅9.1.1.200及以上集群版本支持。

  说明

  低版本升级到9.1.1.200及以上集群版本时，需要使用ALTER TABLE REFRESH DISTRIBUTION刷新生成，因为低版本上创建的表在DN上没有pgxc_class系统表的记录。

列相关的操作 column_clause 子句

其中列相关的操作column_clause可以是以下子句之一：

```
ADD [ COLUMN ] column_name data_type [ compress_mode ] [ COLLATE collation ] [ column_constraint [ ... ] ]
    | MODIFY [ COLUMN ] column_name data_type
    | MODIFY [ COLUMN ] column_name [ CONSTRAINT constraint_name ] NOT NULL [ ENABLE ]
    | MODIFY [ COLUMN ] column_name [ CONSTRAINT constraint_name ] NULL
    | MODIFY [ COLUMN ] column_name DEFAULT default_expr
    | MODIFY [ COLUMN ] column_name ON UPDATE on_update_expr
    | MODIFY [ COLUMN ] column_name COMMENT comment_text
    | DROP [ COLUMN ] [ IF EXISTS ] column_name [ RESTRICT | CASCADE ]
    | ALTER [ COLUMN ] column_name [ SET DATA ] TYPE data_type [ COLLATE collation ] [ USING expression ]
    | ALTER [ COLUMN ] column_name { SET DEFAULT expression | DROP DEFAULT }
    | ALTER [ COLUMN ] column_name { SET | DROP } NOT NULL
    | ALTER [ COLUMN ] column_name SET STATISTICS [PERCENT] integer
    | ADD STATISTICS (( column_1_name, column_2_name [, ...] ))
    | ADD { INDEX | UNIQUE [ INDEX ] } [ index_name ] ( { { column_name | ( expression ) } [ COLLATE collation ] [ opclass ] [ ASC | DESC ] [ NULLS LAST ] } [, ...] ) [ USING method ] [ NULLS [ NOT ] DISTINCT | NULLS IGNORE ] [ COMMENT 'text' ] LOCAL [ ( { PARTITION index_partition_name } [, ...] ) ] [ WITH ( { storage_parameter = value } [, ...] ) ]
    | ADD { INDEX | UNIQUE [ INDEX ] } [ index_name ] ({ { column_name | ( expression ) } [ COLLATE collation ] [ opclass ] [ ASC | DESC ] [ NULLS { FIRST | LAST } ] }[, ...] ) [ USING method ] [ NULLS [ NOT ] DISTINCT | NULLS IGNORE ] [ COMMENT 'text' ] [ WITH ( {storage_parameter = value} [, ... ] ) ] [ WHERE predicate ]
    | DROP { INDEX | KEY } index_name
    | CHANGE [ COLUMN ] old_column_name new_column_name data_type [ [ CONSTRAINT constraint_name ] NOT NULL [ ENABLE ] |
        [ CONSTRAINT constraint_name ] NULL | DEFAULT default_expr | COMMENT 'text' ]
    | DELETE STATISTICS (( column_1_name, column_2_name [, ...] ))
    | ALTER [ COLUMN ] column_name SET ( {attribute_option = value} [, ... ] )
    | ALTER [ COLUMN ] column_name RESET ( attribute_option [, ... ] )
    | ALTER [ COLUMN ] column_name SET STORAGE { PLAIN | EXTERNAL | EXTENDED | MAIN }
```

- **ADD [COLUMN] column_name data_type [compress_mode] [COLLATE collation] [column_constraint [...]]**

  向表中增加一个新的字段。用ADD COLUMN增加一个字段，所有表中现有行都初始化为该字段的缺省值（如果没有声明DEFAULT子句，值为NULL）。

- **ADD ({ column_name data_type [compress_mode] } [, ...])**

  向表中增加多列。

- **MODIFY [COLUMN] column_name data_type**

  修改表已存在字段的数据类型。需注意：不支持修改分布列数据类型。

- **MODIFY [COLUMN] column_name [CONSTRAINT constraint_name] NOT NULL [ENABLE]**

 版权所有 © 华为云计算技术有限公司

为表的某列添加not null约束，列存表暂不支持。

- **MODIFY [COLUMN] column_name [CONSTRAINT constraint_name] NULL**

  为表的某列移除not null约束。

- **MODIFY [COLUMN] column_name DEFAULT default_expr**

  修改表的default值。

- **MODIFY [COLUMN] column_name ON UPDATE on_update_expr**

  修改表中指定列的on update表达式，该列必须为timestamp类型或者timestamptz类型，当on_update_expr为NULL时，则删除ON UPDATE子句。

- **MODIFY [COLUMN] column_name COMMENT comment_text**

  修改表的注释信息。

- **DROP [COLUMN] [IF EXISTS] column_name [RESTRICT | CASCADE]**

  从表中删除一个字段，和这个字段相关的索引和表约束也会被自动删除。如果任何表之外的对象依赖于这个字段，必须声明CASCADE ，比如外键参考、视图等。

  DROP COLUMN命令并不是物理上把字段删除，而只是简单地把它标记为对SQL操作不可见。随后对该表的插入和更新将在该字段存储一个NULL。因此，删除一个字段是很快的，但是它不会立即释放表在磁盘上的空间，因为被删除了的字段占据的空间还没有回收。这些空间将在执行VACUUM时而得到回收。

- **ALTER [COLUMN] column_name [SET DATA] TYPE data_type [COLLATE collation] [USING expression]**

  改变表字段的数据类型，只允许相同大类的类型转换（数值之间，字符串之间，时间之间等）。该字段涉及的索引和简单的表约束将被自动地转换为使用新的字段类型，方法是重新分析最初提供的表达式。

  ALTER TYPE要求重写整个表的特性有时候是一个优点，因为重写的过程消除了表中没用的空间。比如，要想立刻回收被一个已经删除的字段占据的空间，最快的方法是：

  ```
  ALTER TABLE table ALTER COLUMN anycol TYPE anytype;
  ```

  这里的anycol是任何在表中还存在的字段，而anytype是和该字段的原类型一样的类型。这样的结果是在表上没有任何可见的语义的变化，但是这个命令强制重写，这样就删除了不再使用的数据。

  ALTER TYPE的USING选项实际上可以声明涉及该行旧值的任何表达式，即可引用正在被转换的字段之外的其他字段。因此可以用ALTER TYPE语法实现非常通用的类型转换。由于这种灵活性，USING表达式不会作用于该字段的缺省值（若存在），因此其结果可能不是缺省表达式要求的常量表达式。这意味着，如果从旧类型到新类型没有隐式或赋值转换，即使存在USING子句，ALTER TYPE也可能无法将缺省值转换为新的类型。这种情况下，应先用DROP DEFAULT删除缺省，再执行ALTER TYPE，最后使用SET DEFAULT设置新的缺省值。类似的问题也适用于涉及该字段的索引和约束。

- **ALTER [COLUMN] column_name { SET DEFAULT expression | DROP DEFAULT }**

  为一个字段设置或者删除缺省值。缺省值只应用于后续执行的INSERT命令，不会修改表中已经存在的行。也可以为视图创建缺省值，在视图的ON INSERT规则应用之前，默认值会插入到INSERT语句中。

- **ALTER [COLUMN] column_name { SET | DROP } NOT NULL**

  修改一个字段是否允许NULL值或者拒绝NULL值。如果表在字段中包含非NULL，则只能使用SET NOT NULL。

 版权所有 © 华为云计算技术有限公司

- **ALTER [COLUMN] column_name SET STATISTICS [PERCENT] integer**

  为随后的ANALYZE操作设置针对每个字段的统计收集目标。目标的范围可以在0到10000之内设置。设置为-1时表示重新恢复到使用系统缺省的统计目标。

- **ADD { INDEX | UNIQUE [INDEX] } [index_name] ({ { column_name | (expression) } [COLLATE collation] [opclass] [ASC | DESC] [NULLS LAST] } [, ...]) [USING method] [NULLS [NOT] DISTINCT | NULLS IGNORE] [COMMENT 'text'] LOCAL [({ PARTITION index_partition_name } [, ...])] [WITH ({ storage_parameter = value } [, ...])]**

  为表的分区表创建索引，具体参数可参考**CREATE INDEX**。

- **ADD { INDEX | UNIQUE [INDEX] } [index_name] ({ { column_name | (expression) } [COLLATE collation] [opclass] [ASC | DESC] [NULLS { FIRST | LAST }] }[, ...]) [USING method] [NULLS [NOT] DISTINCT | NULLS IGNORE] [COMMENT 'text'] [WITH ({storage_parameter = value} [, ...])] [WHERE predicate]**

  在表上创建索引，具体参数可参考**CREATE INDEX**。

- **DROP { INDEX | KEY } index_name**

  删除一个表上的索引。

- **CHANGE [COLUMN] old_column_name new_column_name data_type [[CONSTRAINT constraint_name] NOT NULL [ENABLE] |**

  [CONSTRAINT constraint_name] NULL | DEFAULT default_expr | COMMENT 'text']

  修改表中列信息，可将旧列名修改成新列名，以及修改列字段信息。

- **{ADD | DELETE} STATISTICS ((column_1_name, column_2_name [, ...]))**

  用于添加和删除多列统计信息声明（不实际进行多列统计信息收集），以便在后续进行全表或全库analyze时进行多列统计信息收集。每组多列统计信息最多支持32列。不支持添加/删除多列统计信息声明的表：系统表、外表。

- **ALTER [COLUMN] column_name SET ({attribute_option = value} [, ...])**

  ALTER [COLUMN] column_name RESET (attribute_option [, ...])

  设置/重置属性选项。

  属性选项定义的参数有：n_distinct、n_distinct_inherited和cstore_cu_sample_ratio。n_distinct设置并固定表的distinct值统计信息，n_distinct_inherited设置并固定继承表的distinct值统计信息，cstore_cu_sample_ratio设置对cstore列存表进行analyze时所选CU的比例。目前禁止SET/RESET n_distinct_inherited参数。

  - n_distinct

    手动设置该列的distinct值统计信息。

    取值范围：-1.0 ~ INT_MAX

    默认值：0，表示不设置。

  - n_distinct_inherited

    手动设置继承表的该列的distinct值统计信息。

    取值范围：-1.0 ~ INT_MAX

    默认值：0，表示不设置。

  - cstore_cu_sample_ratio

    设置列存表执行analyze，计算需要采样的CU个数时，需要扩大的倍数。

 版权所有 © 华为云计算技术有限公司

取值范围：1.0 ~ 10000.0

默认值：1.0

- **ALTER [COLUMN] column_name SET STORAGE { PLAIN | EXTERNAL | EXTENDED | MAIN }**

  为一个字段设置存储模式。这个设置控制这个字段是内联保存还是保存在一个附属的表里，以及数据是否要压缩。仅支持对行存表的设置；对列存表没有意义，执行时报错。SET STORAGE本身并不改变表上的任何东西，只是设置将来的表操作时，建议使用的策略。

其他参数的选项

- 列约束column_constraint为：

```
[ CONSTRAINT constraint_name ]
   { NOT NULL |
     NULL |
     CHECK ( expression ) |
     DEFAULT default_expr  |
     UNIQUE [ NULLS [ NOT ] DISTINCT | NULLS IGNORE ] index_parameters [NOT ENFORCED [ENABLE QUERY OPTIMIZATION | DISABLE QUERY OPTIMIZATION]] |
     PRIMARY KEY index_parameters [NOT ENFORCED [ENABLE QUERY OPTIMIZATION | DISABLE QUERY OPTIMIZATION]] }
   [ DEFERRABLE | NOT DEFERRABLE | INITIALLY DEFERRED | INITIALLY IMMEDIATE ]
```

- 列的压缩可选项compress_mode为：

```
[ DELTA | PREFIX | DICTIONARY | NUMSTR | NOCOMPRESS ]
```

- 根据已有唯一索引为表增加主键约束或唯一约束table_constraint_using_index为：

```
[ CONSTRAINT constraint_name ]
   { UNIQUE | PRIMARY KEY } USING INDEX index_name
   [ DEFERRABLE | NOT DEFERRABLE | INITIALLY DEFERRED | INITIALLY IMMEDIATE ] [NOT ENFORCED [ENABLE QUERY OPTIMIZATION | DISABLE QUERY OPTIMIZATION]]
```

- 增加外键约束REFERENCES为：

```
[ CONSTRAINT constraint_name ]
 FOREIGN KEY ( column_name [, ... ] ) REFERENCES reftable [ ( refcolumn [, ... ] ) ] }
```

- 表约束table_constraint为：

```
[ CONSTRAINT constraint_name ]
   { CHECK ( expression ) |
     UNIQUE [ NULLS [ NOT ] DISTINCT | NULLS IGNORE ] ( column_name [, ... ] ) index_parameters [NOT ENFORCED [ENABLE QUERY OPTIMIZATION | DISABLE QUERY OPTIMIZATION]] |
     PRIMARY KEY ( column_name [, ... ] ) index_parameters [NOT ENFORCED [ENABLE QUERY OPTIMIZATION | DISABLE QUERY OPTIMIZATION]] }|
     PARTIAL CLUSTER KEY ( column_name [, ... ] ) }
   [ DEFERRABLE | NOT DEFERRABLE | INITIALLY DEFERRED | INITIALLY IMMEDIATE ]
```

- 索引参数index_parameters为：

```
[ WITH ( {storage_parameter = value} [, ... ] ) ]
   [ USING INDEX TABLESPACE tablespace_name ]
```

参数说明

表 13-66 ALTER TABLE 参数说明

参数	描述	取值范围
IF EXISTS	如果表名不存在时返回一个提示而不是错误。	-

 版权所有 © 华为云计算技术有限公司

参数	描述	取值范围
ONLY	仅对指定的表名执行操作。 如果不使用ONLY关键字， 该表及其所有子表都将会被更改。 需注意：只允许在父表或子表中添加或删除列，或更改列的类型。 父表及其后代表必须始终具有相同的列和类型。	-
table_name	需要修改的表名称。 可以在表名称后面显式地增加*选项来指定包括子表，即表示所有后代表都被扫描，这是默认行为	有效的表名。
new_table_name	表的新名称。	字符串，需符合**标识符命名规范**。
column_name， column_1_name， column_2_name	现有的或新字段的名称。	有效的字段名。
new_column_name	指定列的新名称。	字符串，需符合**标识符命名规范**。
constraint_name	约束的名称。	有效的约束名。
new_constraint_name	表约束的新名称。	字符串，需符合**标识符命名规范**。
schema_name	表所在的模式名称。	有效的模式名
new_schema	修改后新的模式名称。	有效的模式名。
index_name	索引名称。	有效的索引名。
new_owner	表的新所有者的名称。	有效的用户名或角色名。
new_tablespace	表所属的新表空间的名称。	有效的表空间名。
data_type	新字段的类型，或者现有字段的新类型。	字符串，需符合**标识符命名规范**。
storage_parameter	表的存储参数。	参见**表13-67**。
compress_mode	表字段的压缩可选项，当前仅对行存表有效。该子句指定该字段优先使用的压缩算法	-
collation	字段排序规则名称。可选字段COLLATE指定了新字段的排序规则，如果省略，排序规则为新字段的默认类型。	-

版权所有 © 华为云计算技术有限公司

参数	描述	取值范围
USING expression	USING子句声明如何从旧的字段值里计算新的字段值；如果省略，缺省从旧类型向新类型的赋值转换。如果从旧数据类型到新类型没有隐含或者赋值的转换，则必须提供一个USING子句。	-
NOT NULL | NULL	设置列是否允许空值。	-
integer	带符号的整数常量。当使用PERCENT时表示按照表数据的百分比收集统计信息。	0~100
attribute_option	属性选项。	-
PLAIN | EXTERNAL | EXTENDED | MAIN	字段存储模式。 ● PLAIN必须用于定长的数值（比如integer）并且是内联的、不压缩的。 ● MAIN用于内联、可压缩的数据。 ● EXTERNAL用于外部保存、不压缩的数据。使用EXTERNAL将令在text和bytea字段上的子字符串操作更快，但付出的代价是增加了存储空间。 ● EXTENDED用于外部的压缩数据，EXTENDED是大多数支持非PLAIN存储的数据的缺省。	-
CHECK (expression)	每次将要插入的新行或者将要被更新的行必须使表达式结果为真才能成功，否则会抛出一个异常并且不会修改数据库。 声明为字段约束的检查约束应该只引用该字段的数值，而在表约束里出现的表达式可以引用多个字段。 目前，CHECK表达式不能包含子查询也不能引用除当前行字段之外的变量。	-
DEFAULT default_expr	给字段指定缺省值。 缺省表达式的数据类型必须和字段类型匹配。 缺省表达式将被用于任何未声明该字段数值的插入操作。若未显式指定DEFAULT，则默认值为NULL。 default_expr中若使用后缀操作符（如!），需使用括号括起来。	-

版权所有 © 华为云计算技术有限公司

参数	描述	取值范围
UNIQUE [NULLS [NOT] DISTINCT | NULLS IGNORE] index_parameters UNIQUE (column_name [, ...]) [NULLS [NOT] DISTINCT | NULLS IGNORE] index_parameters	UNIQUE约束表示表里的一个或多个字段的组合必须在全表范围内唯一。 其中[NULLS [NOT] DISTINCT | NULLS IGNORE]字段用来指定Unique唯一索引中索引列NULL值的处理方式。具体可参考**表13-68**。 默认取值：该参数默认取值为空，即NULL值可重复插入。	-
PRIMARY KEY index_parameters PRIMARY KEY (column_name [, ...]) index_parameters	主键约束表明表中的一个或者一些字段只能包含唯一（不重复）的非NULL值。	-
DEFERRABLE | NOT DEFERRABLE | INITIALLY DEFERRED | INITIALLY IMMEDIATE	设置该约束是否可推迟，列存暂不支持。 • DEFERRABLE：可以推迟到事务结尾使用SET CONSTRAINTS命令检查。 • NOT DEFERRABLE：在每条命令之后马上检查。 • INITIALLY IMMEDIATE：那么每条语句之后就立即检查它。 • INITIALLY DEFERRED：只有在事务结尾才检查它。	-
NOT ENFORCED	指定所建立的约束为信息约束，数据库对该约束不进行数据校验，由用户来保证约束要求。	-
ENFORCED	预留参数，目前不支持ENFORCED。	-
ENABLE QUERY OPTIMIZATION	利用信息约束对执行计划进行优化。	-
DISABLE QUERY OPTIMIZATION	禁止利用信息约束对执行计划优化。	-
WITH ({storage_parameter = value} [, ...])	为表或索引指定一个可选的存储参数。	-

版权所有 © 华为云计算技术有限公司

参数	描述	取值范围
COMPRESS| NOCOMPRESS	● NOCOMPRESS：如果指定关键字NOCOMPRESS则不会修改表的现有压缩特性。 ● COMPRESS：如果指定COMPRESS关键字，则对该表进行批量插入元组时触发该特性。	-
CASCADE	级联删除依赖于被依赖字段或者约束的对象（比如引用该字段的视图）。	-
RESTRICT	如果字段或者约束还有任何依赖的对象，则拒绝删除该字段。这是缺省行为。	-
cache_policy	表缓存策略，仅存算分离3.0版本支持。 默认值为ALL。	● ALL： 对整个表进行热缓存。 ● NONE：对整个表进行冷缓存。 ● HPN：N分区表中前N个分区会被热缓存，其余分区进行冷缓存。 ● HPL：P1, P2, ... 分区表中在缓存策略中被指定名称的分区会被热缓存，其余分区进行冷缓存。 **说明** ● 对于外表和非分区内表只支持ALL和NONE两种缓存策略。 ● 仅range和list分区内表支持HPN和HPL缓存策略。

 版权所有 © 华为云计算技术有限公司

表 13-67 storage_parameter 参数说明

storage_parameter 参数	参数项	描述	取值范围
分区管理新增的两个选项	PERIOD	设置分区管理中自动创建分区的周期。 • 在建表时，如果没有设置该参数，可以通过set的方式添加该参数，并开启自动创建分区功能；如果之前已经设置该参数，则通过set的方式修改该参数。 • 用户可以通过reset该参数的方式关闭自动创建分区功能，但是在自动删除分区功能存在的情况下，不支持关闭自动创建分区功能。	interval类型。 PERIOD的范围要求以及开启该功能的约束请参考**WITH子句参数说明**中的PERIOD参数。
	TTL	设置分区管理中自动删除分区的分区过期时间。 • 在建表时，如果没有设置该参数，可以通过set的方式添加该参数，并开启自动删除分区功能；如果之前已经设置该参数，则通过set的方式修改该参数。 • 用户可以通过reset该参数的方式关闭自动删除分区功能。	interval类型。 TTL的范围要求以及开启该功能的约束请参考**WITH子句参数说明**中的TTL参数。
turbo存储格式列存表新增的选项	enable_turbo_store	设置列存表是否为turbo存储格式。该选项仅9.1.0.100及以上集群版本支持。 • 3.0普通列存表不支持设置turbo存储格式，仅3.0 hstore_opt表支持turbo存储格式。 • 2.0版本列存表无限制。	-

版权所有 © 华为云计算技术有限公司

storage_parameter 参数	参数项	描述	取值范围
hstore_opt表新增的选项	enable_hstore_opt	设置列存表是否为hstore_opt表。该选项仅9.1.2及以上集群版本支持。 ● 标准数仓集群：创建批量模式hstore_opt表。DELTAROW_THRESHOLD参数值默认为0，表示创建为批量模式hstore_opt表（批量模式，即批量数据更新，无并发实时入库能力，查询性能与实时数仓场景持平，适用于高效处理大批量数据同步更新的场景）。 DELTAROW_THRESHOLD参数值不为0时，表示创建为hstore_opt标准模式。 ● 实时数仓集群：创建标准模式hstore_opt表。	取值为true/false。
行列混存表选项	storage_mode	设置hstore_opt表为行列混存表。该选项仅9.1.2及以上集群版本支持。 非hstore_opt表可以通过同时指定enable_hstore_opt开启。	取值为mix，表示行列混存表。
hstore_opt开启轻量化更新功能	enable_light_update	设置hstore_opt表开启轻量化更新表。该选项仅9.1.2及以上集群版本支持。 非hstore_opt表可以通过同时指定enable_hstore_opt开启。	取值为true/false。

唯一索引中索引列NULL值的处理方式的说明如下：

[NULLS [NOT] DISTINCT | NULLS IGNORE]字段用来指定Unique唯一索引中索引列NULL值的处理方式。

在对插入的新数据和表中原始数据进行列的等值比较时，对于NULL值有以下三种处理方式：

 版权所有 © 华为云计算技术有限公司

- NULLS DISTINCT：NULL值互不相等，即NULL值可重复插入。
- NULLS NOT DISTINCT：NULL值相等。若索引列全为NULL，则NULL值不可重复插入；部分索引列为NULL，只有非NULL值不相等，才可成功插入数据。
- NULLS IGNORE：在等值比较时跳过NULL值。若索引列全为NULL，则NULL值可重复插入；部分索引列为NULL，只有非NULL值不相等，才可成功插入数据。

表 13-68 唯一索引中索引列 NULL 值的处理方式

字段控制	索引列全为NULL	部分索引列为NULL
NULLS DISTINCT	可重复插入	可重复插入
NULLS NOT DISTINCT	不可重复插入	非NULL值相等，不可插入；非NULL值不相等，则插入成功
NULLS IGNORE	可重复插入	非NULL值相等，不可插入；非NULL值不相等，则插入成功

表操作示例

创建示例表customer。

```
DROP TABLE IF EXISTS customer;
CREATE TABLE customer
(
   C_CUSTKEY     BIGINT       ,
   C_NAME        VARCHAR(25)  ,
   C_ADDRESS     VARCHAR(40)  ,
   C_NATIONKEY   INT          ,
   C_PHONE       CHAR(15)     ,
   C_ACCTBAL     DECIMAL(15,2)
)
DISTRIBUTE BY HASH(C_CUSTKEY);
```

创建示例列存表customer_address。

```
DROP TABLE IF EXISTS customer_address;
CREATE TABLE customer_address
(
   ca_address_sk       INTEGER                 NOT NULL   ,
   ca_address_id       CHARACTER(16)           NOT NULL   ,
   ca_street_number    CHARACTER(10)                      ,
   ca_street_name      CHARACTER varying(60)              ,
   ca_street_type      CHARACTER(15)                      ,
   ca_suite_number     CHARACTER(10)
)
WITH (ORIENTATION = COLUMN, COMPRESSION=HIGH,COLVERSION=2.0)
DISTRIBUTE BY HASH (ca_address_sk);
```

- 修改表的模式。

  设置表customer的所属模式为myschema。

  ```
  CREATE SCHEMA myschema;
  ALTER TABLE customer SET SCHEMA myschema;
  ```

- 根据已有唯一索引为表增加主键约束或唯一约束。

  为表customer创建唯一索引customer_constraint1，然后根据已有唯一索引增加主键约束，并对前面创建的索引进行重命名。

  ```
  CREATE UNIQUE INDEX customer_constraint1 ON myschema.customer(C_CUSTKEY);
  ALTER TABLE myschema.customer ADD CONSTRAINT customer_constraint2 PRIMARY KEY USING
  INDEX customer_constraint1;
  ```

 版权所有 © 华为云计算技术有限公司

- 重命名表约束。
```
ALTER TABLE myschema.customer RENAME CONSTRAINT customer_constraint2 TO
customer_constraint;
```
- 删除表约束。
```
ALTER TABLE myschema.customer DROP CONSTRAINT customer_constraint;
```
- 给表新增主键约束。
```
ALTER TABLE myschema.customer ADD PRIMARY KEY(C_CUSTKEY);
```
- 向列存表customer_address中添加局部聚簇列。
```
ALTER TABLE customer_address ADD CONSTRAINT customer_address_cluster PARTIAL CLUSTER
KEY(ca_address_sk);
```
- 删除列存表customer_address中的局部聚簇列。
```
ALTER TABLE customer_address DROP CONSTRAINT customer_address_cluster;
```
- 修改列存表的存储格式。**注意**：实时数仓HStore表不支持该示例。
```
ALTER TABLE customer_address SET (COLVERSION = 1.0);
```
- 修改表的分布方式。
```
ALTER TABLE customer_address DISTRIBUTE BY REPLICATION;
```
- 重命名表。
```
ALTER TABLE myschema.customer RENAME TO customer_t;
```
- 修改表缓存策略（仅存算分离3.0版本集群支持）。
```
DROP TABLE IF EXISTS orders;
CREATE TABLE orders(a int, b int);
ALTER TABLE orders SET (cache_policy = 'NONE');
```
- 关闭表自动删除分区功能。创建表的示例可参考**示例：创建指定分区的自动分区管理分区表**。
```
DROP TABLE IF EXISTS CPU;
CREATE TABLE CPU(
    id integer,
    idle numeric,
    IO numeric,
    scope text,
    IP text,
    time timestamp
) with (TTL='7 days',PERIOD='1 day')
partition by range(time)
(
    PARTITION P1 VALUES LESS THAN('2022-01-05 16:32:45'),
    PARTITION P2 VALUES LESS THAN('2022-01-06 16:56:12')
);
ALTER TABLE CPU RESET (ttl);
```

索引操作示例

- 给表增加一个索引。
```
ALTER TABLE myschema.customer_t ADD INDEX customer_index(C_CUSTKEY);
```
- 删除表索引。
```
ALTER TABLE myschema.customer_t DROP INDEX customer_index;
ALTER TABLE myschema.customer_t ADD INDEX customer_index(C_CUSTKEY);
ALTER TABLE myschema.customer_t DROP KEY customer_index;
```
- 给表中某一列添加索引。
```
ALTER TABLE myschema.customer_t ADD c_address_id varchar(20) CONSTRAINT ca_address_index
CHECK (c_address_id > 0);
```

冷热表操作示例

- 单表冷热切换。
```
DROP TABLE IF EXISTS lifecycle_table;

CREATE TABLE lifecycle_table(i int, val text) WITH (ORIENTATION = COLUMN, storage_policy =
```

 版权所有 © 华为云计算技术有限公司

```
'LMT:100')
PARTITION BY RANGE (i)
(
PARTITION P1 VALUES LESS THAN(5),
PARTITION P2 VALUES LESS THAN(10),
PARTITION P3 VALUES LESS THAN(15),
PARTITION P8 VALUES LESS THAN(MAXVALUE)
)
ENABLE ROW MOVEMENT;
ALTER TABLE lifecycle_table REFRESH STORAGE;
```

- 列存分区表修改为冷热表。

```
DROP TABLE IF EXISTS test_1;
CREATE TABLE test_1(id int,d_time date)
WITH(ORIENTATION=COLUMN)
DISTRIBUTE BY HASH (id)
PARTITION BY RANGE (d_time)
(PARTITION p1 START('2022-01-01') END('2022-01-31') EVERY(interval '1 day'))

ALTER TABLE test_1 SET (storage_policy = 'LMT:100');
```

列操作示例

创建示例表warehouse_t。

```
DROP TABLE IF EXISTS warehouse_t;
CREATE TABLE warehouse_t
(
   W_WAREHOUSE_SK            INTEGER               NOT NULL,
   W_WAREHOUSE_ID            CHAR(16)              NOT NULL,
   W_WAREHOUSE_NAME            VARCHAR(20)   UNIQUE DEFERRABLE,
   W_WAREHOUSE_SQ_FT         INTEGER                      ,
   W_COUNTY                VARCHAR(30)                  ,
   W_STATE                 CHAR(2)           DEFAULT 'GA',
   W_ZIP                   CHAR(10)
);
```

- 增加列。

  向表warehouse_t中增加单列。

  ```
  ALTER TABLE warehouse_t ADD W_GOODS_CATEGORY int;
  ```

  向表warehouse_t中增加多列。

  ```
  ALTER TABLE warehouse_t ADD (W_COMMENT VARCHAR(117) NOT NULL, W_COUNT int);
  ```

- 删除列。

  从表warehouse_t中删除一列。

  ```
  ALTER TABLE warehouse_t DROP COLUMN W_STATE;
  ```

- 重命名列。

  表warehouse_t的列名W_ZIP修改为new_W_ZIP。

  ```
  ALTER TABLE warehouse_t RENAME W_ZIP TO new_W_ZIP;
  ```

- 变更列名同时修改列字段信息。

  表warehouse_t的列名W_GOODS_CATEGORY修改为W_GOODS_CATEGORY2，同时修改W_GOODS_CATEGORY2字段数据类型为DECIMAL，字段约束为非空。

  ```
  ALTER TABLE warehouse_t CHANGE W_GOODS_CATEGORY W_GOODS_CATEGORY2 DECIMAL NOT
  NULL COMMENT 'W_GOODS_CATEGORY';
  ```

- 变更列的注释信息（COMMENT）。

  ```
  ALTER TABLE warehouse_t MODIFY W_GOODS_CATEGORY2 COMMENT 'W_GOODS_CATEGORY';
  ```

- 变更列的数据类型。

  修改表字段warehouse_t的数据类型，并将字段约束设置为非空。

  ```
  ALTER TABLE warehouse_t MODIFY W_WAREHOUSE_SQ_FT varchar(20) NOT NULL;
  ```

- 指定列添加非空约束（not null）。

  ```
  ALTER TABLE warehouse_t ALTER COLUMN W_COUNTY SET NOT NULL;
  ```

 版权所有 © 华为云计算技术有限公司

- 为表warehouse_t增加主键。
  ```
  ALTER TABLE warehouse_t ADD PRIMARY KEY(w_warehouse_name);
  ```
- 修改表中指定列的on update表达式。

  为warehouse_t表增加一个带有on update的timestamp列，在表执行UPDATE操作时自动更新该列的时间戳。
  ```
  ALTER TABLE warehouse_t ADD COLUMN C_TIME timestamp on update current_timestamp;
  ```
  将warehouse_t表中带有on update的timestamp列删除。
  ```
  ALTER TABLE warehouse_t MODIFY COLUMN C_TIME timestamp on update NULL;
  ```

加密表操作示例

- 指定列添加加密属性。
  ```
  ALTER TABLE encrypt_table MODIFY (i_int to encrypt);
  ```
- 修改表的加密属性。
  ```
  ALTER TABLE encrypt_table SET (encrypt_option = table);
  ```
- 修改表的加密算法。
  ```
  ALTER TABLE encrypt_table SET (encrypt_algo=aes_ctr_128);
  ```
- 添加加密列。
  ```
  ALTER TABLE encrypt_table add column i_int int encrypt;
  ```

相关链接

CREATE TABLE，**RENAME TABLE**，**DROP TABLE**

常见问题

如何调整DWS分布列?

数据类型转换出现报错：invalid input syntax for xxx

向表中插入数据报错：duplicate key value violates unique constraint "%s

13.4.5 ALTER TABLE PARTITION

功能描述

本章节介绍修改表分区的语法功能，包括增删分区、切割分区、合成分区、从普通表数据迁移到分区表，以及修改分区属性等。

逻辑上的一张表根据某种方案分成几张物理块进行存储，这张逻辑上的表称之为**分区表**，物理块称之为**分区**。分区表是一张逻辑表，不存储数据，数据实际是存储在分区上的。

常见的分区策略都是按照某一列或者某几列定义一些数据分布范围，然后每个分区承载一个范围的数据，这些列称之为**分区键**。

目前DWS行存表、列存表支持的分区类型包括**范围分区**和**列表分区**。

- **范围分区（Range Partitioning）**

  根据表的一列或者多列，将要插入表的记录分为若干个范围，这些范围在不同的分区里没有重叠。为每个范围创建一个分区，用来存储相应的数据。
  - **核心特征**：数据按连续区间分布，常用于有序且按范围检索的场景。

 版权所有 © 华为云计算技术有限公司

  - **适用场景**：分区键具有线性增长或连续区间特征，数据查询常围绕 “某一范围” 展开，且新数据会自然落入新的区间。例如按时间范围进行分区的订单数据，按数值范围分区的用户等级数据等。
  - **范围分区策略**是指记录插入分区的方式，根据分区键值将记录映射到已创建的某个分区上，如果可以映射到已创建的某一分区上，则把记录插入到对应的分区上，否则给出报错和提示信息，这是最常用的分区策略。目前范围分区仅支持范围分区策略。

- **列表分区（List Partitioning）**

  根据表的一列，将要插入表的记录通过每一个分区中出现的键值划分到对应的分区中，这些键值在不同的分区里没有重叠。为每组键值创建一个分区，用来存储相应的数据。**列表分区仅8.1.3及以上集群版本支持**。

  - **核心特征**：数据按离散的枚举值划分，常用于固定分类或高频过滤的维度。
  - **适用场景**：分区键值是固定且离散的类别，数据查询常围绕 “某一类别” 展开，键值不会无规律新增。例如按省份分区的用户信息表，每个省份对应一个分区。
  - **列表分区策略**是根据分区键值将记录映射到已创建的某个分区上，如果可以映射到已创建的某一分区上，则把记录插入到对应的分区上，否则给出报错和提示信息。

注意事项

- 添加分区的名称不能与该分区表已有分区的名称相同。
- 对于范围分区表，要添加的分区的边界值要和分区表的分区键的类型一致，且要大于分区表的最后一个分区的上边界。
- 对于列表分区表，如果已经定义DEFAULT分区，则不能添加新分区。
- 若文档中未特殊注明，则表明范围分区表和列存分区的语法使用相同。
- 如果目标分区表中已有分区数达到了最大值（32767），则不能继续添加分区。
- 当分区表只有一个分区时，不能删除该分区。
- 删除分区（DROP PARTITION）时会连同分区内数据一起删除。
- 选择分区使用PARTITION FOR()，括号里指定值个数应该与定义分区时使用的列个数相同，并且一一对应。
- Value分区表不支持相应的Alter Partition操作。
- OBS冷热表对于move，exchange，merge，split操作，不支持指定分区表的表空间为OBS表空间；执行ALTER语法时，需保持分区数据冷热属性不变（即冷分区操作后为冷分区，热分区操作后为热分区），不支持将冷分区数据切至本地表空间；对于冷分区仅支持默认表空间；merge操作不支持将冷分区与热分区进行合并，exchange操作不支持冷分区交换。

⚠ 警告

- 避免在业务高峰期执行ALTER TABLE/ALTER TABLE PARTITION（增删改查、DROP PARTITION）、TRUNCATE操作，避免有长SQL阻塞ALTER、TRUNCATE操作或SQL业务被ALTER、TRUNCATE阻塞。
- 更多开发设计规范参见**总体开发设计规范**。

版权所有 © 华为云计算技术有限公司

语法格式

修改表分区主语法：

```
ALTER TABLE [ IF EXISTS ] { table_name  [*] | ONLY table_name | ONLY ( table_name  )}
    action [, ... ];
```

其中action统指如下分区维护子语法。当存在多个分区维护子句时，保证了分区的连续性，无论这些子句的排序如何，DWS总会先执行DROP PARTITION再执行ADD PARTITION操作，最后顺序执行其他分区维护操作。

- **MODIFY PARTITION子语法：设置分区索引是否可用**
- **REBUILD PARTITION子语法：重建分区索引**
- **EXCHANGE PARTITION（交换分区）子语法：把普通表的数据迁移到指定的分区**
- **ROW MOVEMENT子语法：设置分区表的行迁移开关**
- **MERGE PARTITIONS子语法：把多个分区合并成一个分区**
- **SPLIT PARTITION子语法：把一个分区切割成多个分区**
- **增加分区ADD PARTITION子语法：为指定的分区表添加一个或多个分区**
- **删除分区DROP PARTITION子语法：删除分区表中的指定分区或多个分区**
- **TRUNCATE PARTITION子语法：清理表分区的数据**

ALTER TABLE PARTITION 主句参数说明

表 13-69 ALTER TABLE PARTITION 主句参数说明

参数	描述	取值范围
table_name	需要修改的分区表的名称。	已存在的分区表名。
action [, ...]	action统指分区维护子语法。当存在多个分区维护子句时，保证了分区的连续性，无论这些子句的排序如何，DWS总会先执行DROP PARTITION再执行ADD PARTITION操作，最后顺序执行其他分区维护操作。	参见以下各个子语法使用。
partition_name	需要修改的分区名。	已存在的分区名。
partition_value	分区键值。 通过PARTITION FOR (partition_value [, ...])子句指定的这一组值，可以唯一确定一个分区。	-
PURGE	指定Purge参数时，无论回收站特性是否开启，直接删除，不放入回收站。	-

版权所有 © 华为云计算技术有限公司

参数	描述	取值范围
KEEP ttl HOUR	对于V3表，回收站特性开启时，设置TTL，物理文件会放入回收站中，后续可以从回收站中恢复，TTL到期之后，后台自动清理回收站中的对象。	-

MODIFY PARTITION 子语法：设置分区索引是否可用

```
ALTER TABLE [ IF EXISTS ] { table_name  [*] | ONLY table_name | ONLY ( table_name  )}
MODIFY PARTITION partition_name { UNUSABLE LOCAL INDEXES | REBUILD UNUSABLE LOCAL INDEXES }
```

表 13-70 modify_clause 子语法参数说明

参数	描述	取值范围
partition_name	需要修改的分区名。	已存在的分区名。
UNUSABLE LOCAL INDEXES	设置该分区上的所有索引不可用。	-
REBUILD UNUSABLE LOCAL INDEXES	重建该分区上的所有索引。	-

REBUILD PARTITION 子语法：重建分区索引

REBUILD PARTITION是DWS中用于重建分区表索引的操作，主要用于修复因异常操作（如分区数据大量变更、索引损坏等）导致的分区索引不可用状态，或优化分区索引性能。与重建整个表的索引相比，它仅针对指定分区操作，能显著减少资源消耗和执行时间。

典型使用场景:

- **分区索引失效后修复**：当分区执行了大量DML操作（如批量删除、更新）或异常中断后，分区索引可能变为 UNUSABLE 状态，需重建恢复可用性。
- **性能优化**：长期运行的分区表，索引可能产生碎片，重建可优化索引结构，提升查询效率。
- **分区数据迁移后**：通过EXCHANGE PARTITION交换分区数据后，索引可能需要重建以保持一致性。

该语法**仅8.3.0.100及以上集群版本**支持。

```
ALTER TABLE [ IF EXISTS ] { table_name  [*] | ONLY table_name | ONLY ( table_name  )}
REBUILD PARTITION partition_name [ WITHOUT UNUSABLE ]
```

 版权所有 © 华为云计算技术有限公司

表 13-71 rebuild_clause 子语法参数说明

参数	描述	取值范围
WITHOUT UNUSABLE	重建该分区上的索引时，忽略UNUSABLE状态的索引。	-

EXCHANGE PARTITION（交换分区）子语法：把普通表的数据迁移到指定的分区

EXCHANGE PARTITION（交换分区，或称分区交换）是数据仓库中用于高效管理分区表的关键操作。其核心功能是将**分区表中的一个分区与普通表进行元数据层交换，而非物理移动数据**。这意味着数据块的所有权快速切换，避免大规模I/O操作，极大提升性能。

使用场景：

1. **数据批量加载：**将预处理好的数据先加载到临时表，再替换到目标分区，减少业务高峰期对生产表的写入压力。
2. **历史数据归档：**将不再频繁访问的分区交换到独立表中，便于迁移到低成本存储。
3. **分区数据修复**：将问题分区的数据交换到临时表进行修正，再重新交换回分区表，降低修复期间的锁竞争。
4. **数据迁移：**跨表或跨库迁移特定分区数据时，直接交换避免逐条插入。

语法格式：

```
ALTER TABLE [ IF EXISTS ] { table_name  [*] | ONLY table_name | ONLY ( table_name  )}
EXCHANGE PARTITION { ( partition_name ) | FOR ( partition_value [, ...] ) }
   WITH TABLE {[ ONLY ] ordinary_table_name | ordinary_table_name * | ONLY ( ordinary_table_name )}
   [ { WITH | WITHOUT } VALIDATION ] [ VERBOSE ]
```

表 13-72 exchange_clause 子语法参数说明

参数	描述	取值范围
partition_value	分区键值。 通过PARTITION FOR (partition_value [, ...])子句指定的这一组值，可以唯一确定一个分区。	-
ordinary_table_name	进行迁移的普通表的名称。	已存在的普通表名。

版权所有 © 华为云计算技术有限公司

参数	描述	取值范围
{ WITH | WITHOUT } VALIDATION	在进行数据迁移时，是否检查普通表中的数据满足指定分区的分区键范围。默认为WITH。 由于检查比较耗时，特别是当数据量很大的情况下更甚。所以在保证当前普通表中的数据满足分区的分区键范围时，可以加上WITHOUT来指明不进行检查。	• WITH：对于普通表中的数据要检查是否满足分区的分区键范围，如果有数据不满足，则报错。 • WITHOUT：对于普通表中的数据不检查是否满足分区的分区键范围。
VERBOSE	在VALIDATION是WITH状态时，如果检查出普通表有不满足要交换分区的分区键范围的数据，那么把这些数据插入到正确的分区，如果路由不到任何分区，再报错。 **注意** 只有在VALIDATION是WITH状态时，才可以指定VERBOSE。	-

进行交换的普通表和分区表必须满足如下条件：

- 普通表和分区表的列数目相同，对应列的信息严格一致，包括：列名、列的数据类型、列约束、列的Collation信息、列的存储参数、列的压缩信息、已删除字段的数据类型等。
- 普通表和分区表的表压缩信息严格一致。
- 普通表和分区表的分布列信息严格一致。
- 普通表和分区表的索引个数相同，且对应索引的信息严格一致。
- 普通表和分区表的表约束个数相同，且对应表约束的信息严格一致。
- 普通表不可以是临时表和unlogged表。
- 普通表和分区表应该在同一个逻辑集群或节点组（NodeGroup）中，如果不在同一个逻辑集群或节点组，将会采用将数据插入对方表内方式来实现交换分区，这样交换分区的时间与表数据量有关，对于数据量非常大的表和分区表，交换分区将会非常耗时。
- 在线扩容重分布场景中，如果普通表和分区表正在重分布，交换分区语句有可能中断正在重分布的普通表或分区表（取决于交换分区和重分布语句是否产生锁冲突），通常重分布的普通表或分区表被中断后会重试2次，但同一个表交换分区执行过于频繁可能导致普通表或分区表多次重试重分布都失败。如果普通表重分布过程被交换分区操作打断，在重试重分布时，数据已经被替换为原分区表中的数据，会重新进行全量重分布。
- 如果行存分区表中最后一个有效字段后的其他字段全部被删除，在不考虑这些删除字段的情况下，分区表与普通表字段信息一致时，分区表和普通表可以进行交换。
- 列存普通表和列存分区表的表级参数colversion必须一致：禁止colversion2.0与colversion1.0执行交换分区操作。

 版权所有 © 华为云计算技术有限公司

完成交换后，普通表和分区表的数据被置换，同时普通表和分区表的表空间信息被置换。此时，普通表和分区表的统计信息变得不可靠，需要对普通表和分区表重新执行analyze。

ROW MOVEMENT 子语法：设置分区表的行迁移开关

```
ALTER TABLE [ IF EXISTS ] { table_name  [*] | ONLY table_name | ONLY ( table_name  )}
{ ENABLE | DISABLE } ROW MOVEMENT
```

表 13-73 row_clause 子语法参数说明

参数	描述	取值范围
ENABLE/DISABLE ROW MOVEMENT	行迁移开关，当更新表中某行分区键值时，如果新值属于其他分区，则会根据开关给出报错信息或者将该行数据迁移到对应的分区。 • ENABLE：打开行迁移功能，适用于频繁更新分区键的场景。 • DISABLE：关闭行迁移功能，适用于严格禁止数据跨分区更新的场景，更新分区键会报错。 默认为DISABLE，关闭。 **注意** 开启ROW MOVEMENT则允许跨分区更新，但此时如果有SELECT FOR UPDATE查询该分区表并发执行，存在查询结果瞬时不一致的可能性，需要谨慎使用。	举例，某分区表的日期列（分区键），按季度进行分区，分为分区p_2023q1和p_2023q2，某一行的“2023-02-15”原属于第一季度的分区p_2023q1，当更新该值为“2023-05-15”后，对应的这行数据，会根据ROW MOVEMENT开关是否打开而决定是否行数据迁移： • ENABLE：打开行迁移功能，则该数据迁移到新分区p_2023q2。 • DISABLE：关闭行迁移功能，则系统会报错。 详细示例参见**示例：开启和关闭行迁移功能ROW MOVEMENT**。

MERGE PARTITIONS 子语法：把多个分区合并成一个分区

```
ALTER TABLE [ IF EXISTS ] { table_name  [*] | ONLY table_name | ONLY ( table_name  )}
MERGE PARTITIONS { partition_name } [, ...] INTO PARTITION partition_name
```

- INTO关键字前的分区称为源分区，INTO关键字后的分区称为目标分区。
- 源分区个数不能小于2个。
- 源分区名称不能重复。
- 源分区不能存在unusable的索引，否则执行会报错。
- 目标分区名只能跟最后一个源分区的名称相同，或者跟表的所有分区名都不相同。
- 目标分区的边界是所有源分区边界的并集。

版权所有 © 华为云计算技术有限公司

- 对于范围分区表，所有的源分区必须是边界连续的分区。
- 对于列表分区，如果源分区中包含DEFAULT分区，那么目标分区的边界也是DEFAULT。

SPLIT PARTITION 子语法：把一个分区切割成多个分区

范围分区的split_clause语法如下：

```
ALTER TABLE [ IF EXISTS ] { table_name  [*] | ONLY table_name | ONLY ( table_name  )}
SPLIT PARTITION { partition_name | FOR ( partition_value [, ...] ) } { split_point_clause |
no_split_point_clause }
```

- 指定切割点split_point_clause的语法为：
  ```
  AT ( partition_value ) INTO ( PARTITION partition_name  , PARTITION partition_name  )
  ```

  ⚠ 注意

  切割点的大小要位于正在被切割分区的分区键范围内，指定切割点的方式只能把一个分区切割成两个新分区。

- 不指定切割点no_split_point_clause的语法为。
  ```
  INTO { ( partition_less_than_item [, ...] ) | ( partition_start_end_item [, ...] ) }
  ```

  ⚠ 注意

  - 不指定切割点的方式，partition_less_than_item指定的第一个新分区的分区键要大于正在被切割的分区的前一个分区（若存在）的分区键，partition_less_than_item指定的最后一个分区的分区键要等于正在被切割分区的分区键大小。
  - 不指定切割点的方式，partition_start_end_item指定的第一个新分区的起始点（若存在）必须等于正在被切割的分区的前一个分区（若存在）的分区键，partition_start_end_item指定的最后一个分区的终止点（若存在）必须等于正在被切割分区的分区键。
  - partition_less_than_item支持的分区键个数最多为4，而partition_start_end_item仅支持1个分区键，其支持的数据类型参见**表13-55**中的“PARTITION BY RANGE(partition_key)”参数。
  - 在同一语句中partition_less_than_item和partition_start_end_item两者不可同时使用；不同split语句之间没有限制。

- 分区项partition_less_than_item的语法为：
  ```
  PARTITION partition_name VALUES LESS THAN ( { partition_value | MAXVALUE }  [, ...] )
  ```

- 分区项partition_start_end_item的语法为，其约束参见**表13-55**中的“partition_start_end_item”参数。
  ```
  PARTITION partition_name {
        {START(partition_value) END (partition_value) EVERY (interval_value)} |
        {START(partition_value) END ({partition_value | MAXVALUE})} |
        {START(partition_value)} |
        {END({partition_value | MAXVALUE})}
  }
  ```

列表分区的split_clause语法如下：

 版权所有 © 华为云计算技术有限公司

```
ALTER TABLE [ IF EXISTS ] { table_name  [*] | ONLY table_name | ONLY ( table_name  )}
SPLIT PARTITION { partition_name | FOR ( partition_value [, ...] ) } { split_values_clause |
split_no_values_clause }
```

- 指定切割点的split_values_clause的语法为：
  ```
  VALUES ( { (partition_value) [, ...]  | DEFAULT } ) INTO ( PARTITION partition_name  , PARTITION
  partition_name  )
  ```

  注意

  - 如果源分区不是**DEFAULT分区**，那么切割点所指定的边界是源分区边界的一个非空真子集；如果源分区是DEFAULT分区，那么切割点所指定的边界不能和其他非DEFAULT分区的边界存在重叠。
  - 切割点的指定的边界是INTO关键字后面的第一个分区的边界，源分区边界与切割点的指定的边界的差集是第二个分区的边界。
  - 当源分区是DEFAULT分区时，第二个分区的边界还是DEFAULT。

- 不指定切割点的split_no_values_clause的语法为：
  ```
  INTO ( list_partition_item [, ....], PARTITION partition_name )
  ```

  注意

  - 此处的**list_partition_item**和创建列表分区表的时候指定分区的语法一样，除了此处的分区定义中边界值不能为DEFAULT。
  - 除了最后一个分区，其他分区需要显式定义边界，定义的边界不能是DEFAULT，并且必须是源分区边界的非空真子集。最后一个分区的边界是源分区边界与其他分区边界的差集，且最后一个分区的边界为空(即差集不能为空集)。
  - 如果源分区是DEFAULT分区，则最后一个分区的边界为DEFAULT。

增加分区 ADD PARTITION 子语法：为指定的分区表添加一个或多个分区

范围分区的add_clause语法如下：

```
ALTER TABLE [ IF EXISTS ] { table_name  [*] | ONLY table_name | ONLY ( table_name  )}
ADD { partition_less_than_item... | partition_start_end_item }
```

注意

- 使用**partition_less_than_item**语法时，分区表必须是范围分区表，否则执行会报错。
- 此处**partition_less_than_item**和创建范围分区表的时候指定分区的语法一样。
- 当前分区表的最后一个分区的边界为MAXVALUE，不允许添加新的分区，否则执行会报错。

列表分区的add_clause语法如下：

```
ALTER TABLE [ IF EXISTS ] { table_name  [*] | ONLY table_name | ONLY ( table_name  )}
ADD list_partition_item
```


版权所有 © 华为云计算技术有限公司

⚠ 注意

- 使用**list_partition_item**语法时，分区表必须是列表分区表，否则执行会报错。
- 此处的**list_partition_item**和创建列表分区表的时候指定分区的语法一样。
- 当前分区表存在DEFAULT分区时，不允许添加新的分区动作，否则执行会报错。

删除分区 DROP PARTITION 子语法：删除分区表中的指定分区或多个分区

删除指定分区。

```
ALTER TABLE [ IF EXISTS ] { table_name  [*] | ONLY table_name | ONLY ( table_name  )}
DROP PARTITION { [ IF EXISTS ] partition_name | FOR (  partition_value [, ...] )  }
```

删除多个分区。（8.1.3.100及以上集群版本支持。）

```
ALTER TABLE [ IF EXISTS ] { table_name  [*] | ONLY table_name | ONLY ( table_name  )}
DROP PARTITION { [ IF EXISTS ] partition_name [, ... ] }
```

TRUNCATE PARTITION 子语法：清理表分区的数据

```
ALTER TABLE [ IF EXISTS ] { table_name  [*] | ONLY table_name | ONLY ( table_name  )}
TRUNCATE PARTITION { [ IF EXISTS ] partition_name | FOR (  partition_value  [, ...] )  } [ PURGE ] [KEEP ttl HOUR];
```

- partition_value为分区键值。支持指定多个分区键值，多个分区键值以英文逗号分隔。
- 使用PARTITION FOR子句时，partition_value所在的整个分区会被清空。

修改表分区名称语法

```
ALTER TABLE [ IF EXISTS ] { table_name [*] | ONLY table_name | ONLY ( table_name  )}
    RENAME PARTITION { partition_name | FOR ( partition_value [, ...] ) } TO partition_new_name;
```

表 13-74 修改分区名称参数说明

参数	描述	取值范围
partition_new_name	分区的新名称。	字符串，需符合**标识符命名规范**。

示例：创建范围分区表 customer_address

```
DROP TABLE IF EXISTS customer_address;
CREATE TABLE customer_address
(
    ca_address_sk       INTEGER                 NOT NULL   ,
    ca_address_id       CHARACTER(16)           NOT NULL   ,
    ca_street_number    CHARACTER(10)                      ,
    ca_street_name      CHARACTER varying(60)              ,
    ca_street_type      CHARACTER(15)                      ,
    ca_suite_number     CHARACTER(10)
)
DISTRIBUTE BY HASH (ca_address_sk)
PARTITION BY RANGE(ca_address_sk)
(
        PARTITION P1 VALUES LESS THAN(100),
        PARTITION P2 VALUES LESS THAN(200),
        PARTITION P3 VALUES LESS THAN(300)
);
```


版权所有 © 华为云计算技术有限公司

示例：创建 list 分区表

```
DROP TABLE IF EXISTS data_list;
CREATE TABLE data_list(
    id int,
    time int,
    salary decimal(12,2)
)PARTITION BY LIST (time)(
        PARTITION P1 VALUES (202209),
        PARTITION P2 VALUES (202210,202208),
        PARTITION P3 VALUES (202211),
        PARTITION P4 VALUES (202212),
        PARTITION P5 VALUES (202301)
);
```

示例：使用 MODIFY PARTITION 子句设置分区索引是否可用

给分区表customer_address创建LOCAL索引student_grade_index，并指定分区的索引名称。

```
CREATE INDEX customer_address_index ON customer_address(ca_address_id) LOCAL
(
        PARTITION P1_index,
        PARTITION P2_index,
        PARTITION P3_index
);
```

重建分区表customer_address中分区P1上的所有索引。

```
ALTER TABLE customer_address MODIFY PARTITION P1 REBUILD UNUSABLE LOCAL INDEXES;
```

设置分区表customer_address的分区P3上的所有索引不可用。

```
ALTER TABLE customer_address MODIFY PARTITION P3 UNUSABLE LOCAL INDEXES;
```

示例：使用 REBUILD PARTITION 子语法重建插入过数据的分区索引

步骤1 创建分区表。

```
DROP TABLE IF EXISTS sales;
CREATE TABLE sales (
    id          INT,
    sale_date   DATE,
    amount      DECIMAL(10,2)
)
PARTITION BY RANGE (sale_date)
(
    PARTITION p202310 VALUES LESS THAN ('2023-11-01'),
    PARTITION p202311 VALUES LESS THAN ('2023-12-01')
);
```

步骤2 创建索引。

```
CREATE INDEX idx_sale_date ON sales (sale_date)  LOCAL;
```

步骤3 插入10月、11月分区数据。

```
INSERT INTO sales PARTITION (p202310)
VALUES
(1, '2023-10-05', 100.50),
(2, '2023-10-10', 200.75);
INSERT INTO sales PARTITION (p202311)
VALUES
(3, '2023-11-15', 300.00);
```

步骤4 重建特定分区p202310的索引。

 版权所有 © 华为云计算技术有限公司

```
ALTER TABLE sales  REBUILD PARTITION p202310  WITHOUT UNUSABLE;
```

----结束

示例：使用 ADD PARTITION 子句为指定的分区表添加一个或多个分区

为范围分区表customer_address增加分区。

```
ALTER TABLE customer_address ADD PARTITION P5 VALUES LESS THAN (500);
```

为范围分区表customer_address增加分区: [500, 600), [600, 700)。

```
ALTER TABLE customer_address ADD PARTITION p6 START(500) END(700) EVERY(100);
```

为范围分区表customer_address增加MAXVALUE分区p7。

```
ALTER TABLE customer_address ADD PARTITION p7 END(MAXVALUE);
```

为列表分区表增加分区P6。

```
ALTER TABLE data_list ADD PARTITION P6 VALUES (202302,202303);
```

示例：使用 SPLIT PARTITION 子句将一个分区切割成多个分区

将范围分区表customer_address的P7分区以800为分割点切分。

```
ALTER TABLE customer_address SPLIT PARTITION P7 AT(800) INTO (PARTITION P6a,PARTITION P6b);
```

将范围分区表customer_address中400所在的分区分割成多个分区。

```
ALTER TABLE customer_address SPLIT PARTITION FOR(400) INTO(PARTITION p_part START(300) END(500)
EVERY(100));
```

将列表分区表data_list的分区P2分割成p2a和p2b两个分区。

```
ALTER TABLE data_list SPLIT PARTITION P2 VALUES(202210) INTO (PARTITION p2a,PARTITION p2b);
```

示例：使用 EXCHANGE PARTITION（交换分区）子句将普通表的数据迁移到指定分区

如下示例将演示一个普通表math_grade数据迁移到分区表student_grade中分区（math）的操作。

1. 创建分区表student_grade。

```
DROP TABLE IF EXISTS student_grade;
CREATE TABLE student_grade (
        stu_name     char(5),
        stu_no       integer,
        grade        integer,
        subject      varchar(30)
)
PARTITION BY LIST(subject)
(
        PARTITION gym VALUES('gymnastics'),
        PARTITION phys VALUES('physics'),
        PARTITION history VALUES('history'),
        PARTITION math VALUES('math')
);
```

2. 插入数据到分区表student_grade中。

```
INSERT INTO student_grade VALUES
        ('Ann', 20220101, 75, 'gymnastics'),
        ('Jeck', 20220103, 60, 'math'),
        ('Anna', 20220108, 56, 'history'),
```

 版权所有 © 华为云计算技术有限公司

```
    ('Jann', 20220107, 82, 'physics'),
    ('Molly', 20220104, 91, 'physics'),
    ('Sam', 20220105, 72, 'math');
```

3. 查询分区表student_grade的math分区记录。

```
SELECT * FROM student_grade PARTITION (math);
```

查询结果如下：

```
gaussdb=> SELECT * FROM student_grade PARTITION (math);
 stu_name |  stu_no  | grade | subject
----------+----------+-------+---------
 Sam      | 20220105 |    72 | math
 Jack     | 20220103 |    60 | math
(2 rows)
```

4. 创建一个与分区表student_grade定义匹配的普通表math_grade。

```
DROP TABLE IF EXISTS math_grade;
CREATE TABLE math_grade
(
    stu_name    char(5),
    stu_no      integer,
    grade       integer,
    subject     varchar(30)
);
```

5. 插入数据到表math_grade中。数据与分区表student_grade的math分区的分区规则一致。

```
INSERT INTO math_grade VALUES
    ('Ann', 20220101, 75, 'math'),
    ('Jeck', 20220103, 60, 'math'),
    ('Anna', 20220108, 56, 'math'),
    ('Jann', 20220107, 82, 'math');
```

6. 将普通表math_grade数据迁移到分区表student_grade的分区（math）。

```
ALTER TABLE student_grade EXCHANGE PARTITION (math) WITH TABLE math_grade;
```

7. 查询分区表student_grade，结果显示表math_grade中的数据已和分区表student_grade的分区math中的数据交换。

```
SELECT * FROM student_grade PARTITION (math);
```

```
gaussdb=> SELECT * FROM student_grade PARTITION (math);
 stu_name |  stu_no  | grade | subject
----------+----------+-------+---------
 Jack     | 20220103 |    60 | math
 Anna     | 20220108 |    56 | math
 Ann      | 20220101 |    75 | math
 John     | 20220107 |    82 | math
(4 rows)
```

8. 查询表math_grade，显示之前存储在分区表student_grade的分区math中的数据已交换到表math_grade中。

```
SELECT * FROM math_grade;
```

```
gaussdb=> SELECT * FROM math_grade;
 stu_name |  stu_no  | grade | subject
----------+----------+-------+---------
 Jack     | 20220103 |    60 | math
 Sam      | 20220105 |    72 | math
(2 rows)
```

示例：使用 TRUNCATE PARTITION 子语法清理表分区的数据

清空表customer_address分区p1：

```
ALTER TABLE customer_address TRUNCATE PARTITION IF EXISTS p1;
```


版权所有 © 华为云计算技术有限公司

示例：开启和关闭行迁移功能 ROW MOVEMENT

在更新某行的分区键为新值且新值属于其他分区时，如果打开ROW MOVEMENT，则该行数据对应迁移到新的分区中。

频繁更新分区键时，ENABLE ROW MOVEMENT 可能导致额外I/O开销，需评估业务需求。

创建按日期范围分区的表，并启用行迁移。

```
DROP TABLE IF EXISTS sale_data;
CREATE TABLE  sales_data (
    sale_id  INT,
    product_name VARCHAR(100),
    sale_date DATE,
    amount DECIMAL(10, 2)
)
PARTITION BY RANGE (sale_date) (
    PARTITION p_2023q1 VALUES LESS THAN ('2023-04-01'),
    PARTITION p_2023q2 VALUES LESS THAN ('2023-07-01'),
    PARTITION p_2023q3 VALUES LESS THAN ('2023-10-01'),
    PARTITION p_2023q4 VALUES LESS THAN ('2024-01-01')
)
ENABLE ROW MOVEMENT;
```

插入测试数据。

```
INSERT INTO sales_data (sale_id, product_name, sale_date, amount)
VALUES
(1, 'Product A', '2023-02-15', 1000.00),  -- 属于 p_2023q1
(2, 'Product B', '2023-05-20', 1500.00);  -- 属于 p_2023q2
```

查询各分区数据分布。

```
SELECT tableoid::regclass AS partition, * FROM sales_data ORDER BY sale_id;
```

```
 partition  | sale_id | product_name |      sale_date      | amount
------------+---------+--------------+---------------------+---------
 2147617615 |       1 | Product A    | 2023-02-15 00:00:00 | 1000.00
 2147617616 |       2 | Product B    | 2023-05-20 00:00:00 | 1500.00
(2 rows)
```

尝试更新分区键值（从 p_2023q1 移动到 p_2023q2）。

```
UPDATE sales_data SET sale_date = '2023-05-01'  WHERE sale_id = 1;
```

再次查询各分区数据分布，确认行已移动。

```
SELECT tableoid::regclass AS partition, *  FROM sales_data  ORDER BY sale_id;
```

```
 partition  | sale_id | product_name |      sale_date      | amount
------------+---------+--------------+---------------------+---------
 2147617616 |       1 | Product A    | 2023-05-01 00:00:00 | 1000.00
 2147617616 |       2 | Product B    | 2023-05-20 00:00:00 | 1500.00
(2 rows)
```

关闭行迁移功能。

```
ALTER TABLE sales_data DISABLE ROW MOVEMENT;
```

再次修改分区键（从 p_2023q2 移动回到 p_2023q2）。

版权所有 © 华为云计算技术有限公司

```
UPDATE sales_data SET sale_date = '2023-02-15'  WHERE sale_id = 1;
```

回显提示，行迁移关闭，不允许更新分区键。

```
gaussdb=> UPDATE sales_data SET sale_date = '2023-02-15'  WHERE sale_id = 1;
ERROR:  dn_6003_6004: fail to update partitioned table "sales_data"
DETAIL:  disable row movement
```

示例：使用 MERGE PARTITIONS 子句把多个分区合并成一个分区

将范围分区表customer_address的P2，P3两个分区合并为一个分区。

```
ALTER TABLE customer_address MODIFY PARTITION P3 REBUILD UNUSABLE LOCAL INDEXES;  --上面的示例设置成P3索引不可用，先执行该语句重建索引再做MERGE，否则下行示例会报错
ALTER TABLE customer_address MERGE PARTITIONS P2, P3 INTO PARTITION P_M;
```

示例：使用 DROP PARTITION 子句删除分区表中的指定分区

删除分区表customer_address的多个分区P6a，P6b。

```
ALTER TABLE customer_address DROP PARTITION P6a, P6b;
```

相关链接

CREATE TABLE PARTITION，**DROP TABLE**

常见问题

- **使用实践：使用DWS分区自动管理功能降低电商和物联网行业数据分区维护成本**
- 使用实践：**创建和管理DWS分区表**
- 使用实践：**案例：改建分区表**
- **分区表插入数据报错：inserted partition key does not map to any table partition**
- **范围分区表添加新分区报错upper boundary of adding partition MUST overtop last existing partition**

13.4.6 RENAME TABLE

功能描述

重命名指定表。

RENAME TABLE命令与如下命令的作用相同：

```
ALTER TABLE table_name RENAME to new_table_name
```

注意事项

不支持跨schema对表进行重命名操作。

语法格式

```
RENAME TABLE
{[schema.]table_name TO new_table_name} [, ...];
```


版权所有 © 华为云计算技术有限公司

参数说明

表 13-75 RENAME TABLE 参数说明

参数	描述	取值范围
schema	模式名称。	-
table_name	需要修改的表名称。	-
new_table_name	修改后新的表名称。	-

示例

修改表名：

```
RENAME TABLE customer_address TO new_customer_address;
```

相关链接

CREATE TABLE，**ALTER TABLE**，**DROP TABLE**

13.4.7 DROP TABLE

功能描述

删除指定的表。

注意事项

- DROP TABLE会强制删除指定的表，删除表后，依赖该表的索引会被删除，而使用到该表的函数和存储过程将无法执行。删除分区表，会同时删除分区表中的所有分区。
- 只有表的所有者、模式所有者或者被授予了表的DROP权限的用户才能执行DROP TABLE，系统管理员默认拥有该权限。要清空指定表中的行但是不删除该表定义，可以使用TRUNCATE或者DELETE。

危险

- **DROP对象**操作（如DATABASE、USER/ROLE、SCHEMA、TABLE、VIEW等对象）存在数据丢失风险，尤其含带CASCADE级联删除场景，会将关联的对象一并删除，操作需谨慎，操作前需考虑数据备份。
- 更多开发设计规范参见**总体开发设计规范**。

语法格式

```
DROP TABLE [ IF EXISTS ]
  { [schema.]table_name } [, ...] [ CASCADE | RESTRICT ] [ PURGE ] [KEEP ttl HOUR];
```


版权所有 © 华为云计算技术有限公司

参数说明

表 13-76 DROP TABLE 参数说明

参数	描述	取值范围
IF EXISTS	如果指定的表不存在，则发出一个notice而不是抛出一个错误。	-
schema	要删除的表所属的模式名称。	-
table_name	要删除的表名称。	已存在的表名称。
CASCADE | RESTRICT	当执行删除操作时，如何处理依赖对象中的相关数据。	• CASCADE：级联删除依赖于表的对象（比如视图）。 • RESTRICT（缺省项）：如果存在依赖对象，则拒绝删除该表。这个是缺省。
PURGE	指定PURGE参数时，无论回收站特性是否开启，直接删除，不放入回收站。该参数仅9.1.0.200及以上集群版本支持。 PURGE与KEEP ttl HOUR参数不能同时指定。	-
KEEP ttl HOUR	V3表开启回收站特性（GUC参数enable_recyclebin设置为on）并设置TTL后，物理文件会放入回收站中，在TTL到期前可从回收站中恢复；TTL到期后，后台自动清理回收站中的对象。该参数仅9.1.0.200及以上集群版本支持。	• ttl取值需大于0且最大为360小时（15天）。 • ttl为0时，与不设置该参数效果一致。

示例

删除表warehouse_t1：

```
DROP TABLE IF EXISTS tpcds.warehouse_t1;
```

相关链接

ALTER TABLE，**12.101-RENAME TABLE**，**CREATE TABLE**

13.4.8 CREATE FOREIGN TABLE (SQL on OBS or Hadoop)

功能描述

在当前数据库创建一个HDFS或OBS外表，用来访问存储在HDFS或者OBS分布式集群文件系统上的结构化数据，也可以通过HDFS或OBS外表导出ORC和PARQUET格式数据到HDFS或者OBS上。

版权所有 © 华为云计算技术有限公司

外表（Foreign Table）是数据库或大数据平台中一种虚拟表结构，它不实际存储数据，而是通过元数据信息（如表结构、存储位置、文件格式等）将外部存储系统（如HDFS、OBS等）中的数据映射到数据库内部，使用户能够以标准SQL操作直接访问或操作外部数据。外表的本质是一种“数据代理”，将外部文件系统的数据逻辑映射成数据库可识别的表结构。

外表的核心价值是 “打破数据孤岛”： 通过元数据映射，让数据库用户以熟悉的方式访问外部系统数据，同时支持数据向外部系统的导出，实现跨系统数据的高效协同、分析与共享，尤其在大数据场景中能显著降低数据移动成本、提升协作效率。

DWS通过外表访问外部数据源时，对于不同的数据源，访问性能有所不同：

数据存储在OBS：数据存储和计算分离，集群存储成本低，存储量不受限制，并且集群可以随时删除，但计算性能取决于OBS访问性能，相对HDFS有所下降，建议在数据计算不频繁场景下使用。

数据存储在HDFS：数据存储和计算不分离，集群成本较高，计算性能高，但存储量受磁盘空间限制，删除集群前需将数据导出保存，建议在数据计算频繁场景下使用。

说明

以下描述的读写外表，仅存算分离3.0版本支持。

注意事项

- 使用OBS外表访问OBS桶数据时，应确保DWS与OBS桶在同一个区域(Region)下，例如都在“中国-香港”下。
- 仅系统管理员dbadmin角色或已授权了USEFT权限的普通用户可执行外表操作。
  ```
  ALTER USER user_name USEFT;
  ```
- HDFS外表与OBS外表分为只读外表、只写外表和读写外表，只读外表用于查询操作，只写外表可以将DWS中的数据导出到分布式文件系统中，读写外表可以查询和数据导出，其中**读写外表仅9.1.0.100及以上版本支持**。
- 格式支持：
  - **导入查询**支持ORC、PARQUET、HUDI、ICEBERG、TEXT、CSV、CARBONDATA和JSON格式。
  - **导出数据**中，OBS外表支持ORC、PARQUET（仅存算分离3.0版本支持）格式，HDFS外表支持ORC、PARQUET格式。
- 该方式使用外表前需要用户先手动创建**外部服务器**，**外部服务器**是数据库中配置的一个虚拟对象，用于定义连接到外部数据源的参数（如地址、协议、认证信息等）。通过外部服务器，本地数据库可以直接查询、插入外部数据，就像操作本地表一样（语法类似，但可能受性能或功能限制）。具体请参见**CREATE SERVER**。
- 若手动创建外部服务器Server时指定foreign data wrapper为HDFS_FDW或者DFS_FDW，创建只读外表时需DISTRIBUTE BY子句指定分布方式。

表 13-77 OBS 外表支持读写格式说明

数据类型	DFS_FDW/HDFS_FDW		
-	READ ONLY	WRITE ONLY	READ WRITE
ORC	√	√	√

版权所有 © 华为云计算技术有限公司

数据类型	DFS_FDW/HDFS_FDW		
PARQUET	√	√	√
CARBONDATA	√	×	×
TEXT	√	×	×
CSV	√	×	×
JSON	√	×	×
HUDI	√	×	×
ICEBERG	√	×	×

语法格式

创建外表。

```
CREATE FOREIGN TABLE [ IF NOT EXISTS ] table_name
( [ { column_name type_name
   [ { [CONSTRAINT constraint_name] NULL |
   [CONSTRAINT constraint_name] NOT NULL |
     column_constraint [...]} ] |
     table_constraint [, ...]} [, ...] ] )
   SERVER server_name
   OPTIONS ( { option_name ' value ' } [, ...] )
   [ {WRITE ONLY | READ ONLY | READ WRITE} ]
   DISTRIBUTE BY {ROUNDROBIN | REPLICATION}

   [ PARTITION BY ( column_name ) [ AUTOMAPPED ] ] ;
```

- 其中column_constraint为：
  ```
  [CONSTRAINT constraint_name]
  {PRIMARY KEY | UNIQUE}
  [NOT ENFORCED [ENABLE QUERY OPTIMIZATION | DISABLE QUERY OPTIMIZATION] | ENFORCED]
  ```
- 其中table_constraint为：
  ```
  [CONSTRAINT constraint_name]
  {PRIMARY KEY | UNIQUE} (column_name)
  [NOT ENFORCED [ENABLE QUERY OPTIMIZATION | DISABLE QUERY OPTIMIZATION] | ENFORCED]
  ```

参数说明

表 13-78 CREATE FOREIGN TABLE (SQL on OBS or Hadoop)参数说明

参数	描述	取值范围
IF NOT EXISTS	如果已经存在相同名称的表，系统不会报错，而会发出一个通知，告知表关系已存在。	-
table_name	外表的表名。	字符串，要符合标识符的命名规范。参见**标识符命名规范**。

版权所有 © 华为云计算技术有限公司

参数	描述	取值范围
column_name	外表中的字段名。可以选择多个字段名，中间用“,”隔开。	字符串，要符合标识符的命名规范。参见**标识符命名规范**。 JSON对象由嵌套或并列的name-value对组成，具有顺序无关性，当导入JSON格式数据时，需要通过字段名与name的自动对应来确定字段与value的对应关系。用户需要定义恰当的字段名，否则可能导致导入结果不符合预期。字段名与name的自动对应规则如下： • 无嵌套无数组的情况下，字段名应当与name一致，不区分大小写。 • 字段名使用‘_’字符拼接两个name，标识嵌套关系。 • 字段名使用‘#’字符加十进制非负整数‘n’标识数组的第n个元素（从0开始）。 例如，要导入JSON对象{"A" : "simple", "B" : {"C" : "nesting"}, "D" : ["array", 2, {"E" : "complicated"}]}中的每个元素，外表字段名应当分别定义为a、b、b_c、d、d#0、d#1、d#2、d#2_e，字段的定义顺序不会影响导入结果的正确性。
type_name	字段的数据类型。	-
constraint_name	外表的表约束名。	-

版权所有 © 华为云计算技术有限公司

参数	描述	取值范围
{ NULL | NOT NULL }	标识此列是否允许NULL值。 在创建表时，对于列的约束NULL/NOT NULL，并不能保证该表在HDFS系统中的数据为NULL或者NOT NULL，数据的一致性由用户保证。所以需要由用户判断该列是否一定不为空或者一定为空，在建表的时候选用NULL或NOT NULL。(优化器对列为NULL/NOT NULL做了优化处理，会产生更优的计划。)	-
SERVER server_name	外表的server名字。允许用户自定义名字。	字符串，要符合标识符的命名规范，并且这个server必须存在。
OPTIONS	用于指定外表数据的各类参数。	不同场景下，需要指定的OPTIONS参数有所不同，说明如下。 • 数据格式参数：参见**表13-79**。 • 容错性参数：参见**表13-80**。 • 仅Server类型为DLI时需要指定的参数：参见**表13-81**。 • 仅format为hudi时需要指定的参数：参见**表6 OPTIONS参数说明（仅format为hudi时设置）**。 • OBS外表，各数据源格式支持的OPTIONS参数有所不同，具体参见**表13-84**。 • HDFS外表，各数据源格式支持的OPTIONS参数有所不同，具体参见**表13-85**。

版权所有 © 华为云计算技术有限公司

参数	描述	取值范围
WRITE ONLY | READ ONLY | READ WRITE	指定创建的外表类型。	• WRITE ONLY指定创建HDFS/OBS的只写外表。 • READ ONLY指定创建HDFS/OBS的只读外表。 • READ WRITE指定创建HDFS/OBS的读写外表。 如果不指定创建的外表的类型，**默认为只读外表。**
DISTRIBUTE BY ROUNDROBIN	指定HDFS/OBS外表为ROUNDROBIN分布方式。表的每一行被轮番地发送给各个数据节点（DN），因此数据会被均匀地分布在各个DN中。	-
DISTRIBUTE BY REPLICATION	指定HDFS外表为REPLICATION分布方式。 表的每一行存储在所有数据节点（DN）中，即每个数据节点都有完整的表数据。	-

版权所有 © 华为云计算技术有限公司

参数	描述	取值范围
PARTITION BY (column_name) AUTOMAPPED	column_name指定分区列。对于分区表，AUTOMAPPED表示HDFS分区外表指定的分区列会和HDFS数据中的分区目录信息自动对应，前提是必须保证HDFS分区外表指定分区列的顺序和HDFS数据中分区目录定义的顺序一致，该功能只适用于只读外表，只写外表不支持。	• 只写分区外表最多支持四级分区（即最多指定四个列为分区列）。且至少需要保留一列为非分区列。 • HDFS只读外表支持text，csv，carbondata，orc，parquet格式分区表。 • HDFS只写外表支持orc，parquet格式分区表。 • HDFS读写外表支持orc，parquet格式分区表。 • OBS只读外表/只写外表/读写外表均支持orc，parquet格式分区表。 • 不支持浮点类型和布尔类型的列作为分区列。 • 分区字段长度限制可通过guc参数**dfs_partition_directory_length**调整。 • 分区目录名称由 "分区列名=分区列值" 组成。此名称内如果包含特殊字符还会经过转义，所以转义前推荐长度不要超过(dfs_partition_directory_length + 1) / 3，以此保证其转义后总长度不会因为超过dfs_partition_directory_length而出现报错。 • 不推荐包含过长中文字符的列作为分区列。因为一个中文字符和一个英文字符所占的空间大小并不一致，用户不易计算最终的分区目录名称长度，更容易触发超过dfs_partition_director

 版权所有 © 华为云计算技术有限公司

参数	描述	取值范围
		y_length长度限制的报错。
CONSTRAINT constraint_name	用于指定外表所建立的信息约束（Informational Constraint）的名字。	字符串，要符合标识符的命名规范。
PRIMARY KEY	主键约束，表示表里的一个或者一些字段只能包含唯一（不重复）的非NULL值。一个表只能声明一个主键。	-
UNIQUE	唯一约束，表示表里的一个或者多个字段的组合必须在全表范围内唯一。对于唯一约束，NULL被认为是互相不等的。	-
NOT ENFORCED	指定所建立的约束为信息约束，即数据库不会强制验证该约束，这个选项与后续的ENABLE QUERY OPTIMIZATION搭配使用，主要有以下作用： • **仅为信息约束**：通过PRIMARY KEY 或 UNIQUE 声明列的唯一性约束，帮助优化查询执行计划。 • **不强制验证**：数据库不会检查外部数据源中的数据是否实际满足约束条件（如是否存在重复值）。 • **性能优化**：允许查询优化器利用约束信息生成更高效的执行计划，但需用户自行确保数据符合约束，否则可能导致查询结果不准确。 • **风险**：若实际数据违反约束（如存在重复值），可能导致查询结果错误（如聚合结果异常、连接产生多余记录）。 更多信息参见**信息约束（Informational Constraint）**。	例如声明primary key，但是不强制验证。 CREATE FOREIGN TABLE hdfs_users (user_id INT PRIMARY KEY NOT ENFORCED, ...

版权所有 © 华为云计算技术有限公司

参数	描述	取值范围
ENFORCED	数据库强制验证约束，预留参数，目前DWS不支持设置为ENFORCED。	-
PRIMARY KEY (column_name)	指定所建立的信息约束位于column_name列上。	字符串，要符合标识符的命名规范，并且这个column_name必须存在。
ENABLE QUERY OPTIMIZATION	允许查询优化器利用信息约束生成更高效的执行计划，与前面的NOT ENFORCED搭配使用。	创建外表，为字段名p_partkey int增加唯一约束，但是不强制验证该约束，且允许优化器利用该信息约束生成更优执行计划： CREATE FOREIGN TABLE ft_part (p_partkey int UNIQUE NOT ENFORCED ENABLE QUERY OPTIMIZATION, ...
DISABLE QUERY OPTIMIZATION	禁止查询优化器利用信息约束生成高效的执行计划。	-

表 13-79 OPTIONS 数据格式参数说明

参数	描述	取值范围
header	指定导入数据文件是否包含标题行，标题行一般用来描述表中每个字段的信息。header只能用于CSV格式的文件中。	● true/on，则数据文件第一行会被识别为标题行，导出时会忽略此行。 ● false/off，则数据文件中第一行会被识别为数据。 默认为false/off。
quote	CSV格式文件下的引号字符，缺省值为双引号。	quote参数不能和分隔符、null参数相同。 quote参数只能是单字节的字符。 推荐不可见字符作为quote，例如0x07，0x08，0x1b等。

 版权所有 © 华为云计算技术有限公司

参数	描述	取值范围
escape	CSV格式下，用来指定逃逸字符，逃逸字符只能指定为单字节字符。 在数据库中，逃逸字符（Escape Character）是一种特殊字符，用于转义其他字符，它的主要作用是让数据库系统将原本具有特殊含义的字符当作普通字符处理，让数据库识别并正确处理那些可能与SQL语法冲突的字符（如引号、特殊符号等）。常见的逃逸字符有反斜杠\或双引号''。 缺省值为双引号。当与quote值相同时，会被替换为'\0'。	-
location	OBS外表参数，指定存储在OBS上的文件路径（不建议路径中使用空格，避免引发找不到文件或语法错误的问题），多个桶的数据源数据之间使用分隔符‘|’进行分割，例如：LOCATION 'obs://bucket1/folder/ | obs://bucket2/'，数据库将会扫描指定路径文件夹下面的所有对象。 当访问DLI多版本表时，无需指定location参数。	-

 版权所有 © 华为云计算技术有限公司

参数	描述	取值范围
format	外表中数据源文件的格式。	HDFS外表READ ONLY外表支持ORC、TEXT、JSON、CSV、PARQUET、HUDI、ICEBERG文件格式，WRITE ONLY/READ WRITE外表支持ORC和PARQUET文件格式。 OBS外表READ ONLY外表支持ORC、TEXT、JSON、CSV、CARBONDATA、PARQUET、HUDI、ICEBERG文件格式，WRITE ONLY外表支持ORC和PARQUET文件格式。 **注意** • 对于JSON格式数据，仅支持JSON对象（object，最外层由{}构造）导入，不支持JSON数组（array，最外层由[]构造）导入，但支持JSON对象内部数组的导入。 • Hudi即Apache Hudi，是一个事务性数据湖平台。当前版本支持将OBS外表的format参数指定为Hudi来访问相应的Hudi表数据集，并提供访问Hudi表增量/实时视图的方法。

foldername	外表中数据源文件目录（不建议目录路径中使用空格，避免引发找不到文件或语法错误的问题），即表数据目录在HDFS文件系统和OBS上对应的文件目录。此选项对WRITE ONLY和READ WRITE可写外表为必选项，对READ ONLY外表为可选项。
	当访问DLI多版本表时，无需指定foldername参数。
	• OBS外表，格式为：'/OBS桶名/文件夹名/'。应确保OBS桶与DWS集群在同一个区域，不支持跨区访问OBS桶数据。
	• HDFS外表，格式为远端HDFS中的数据文件所在目录，如MRS集群，可登录**MRS控制台**，进入集群基本信息页面，单击“文件管理”，选择“HDFS文件列表”，找到数据存储的目录。
		OBS外表，如下举例：
		`OPTIONS ( foldername '/obs-demo01/obs-dws/', format 'parquet', encoding 'utf8' )`
		HDFS外表：如下举例：
		`OPTIONS ( foldername '/user/hive/warehouse/demo.db/product_info_orc/', format 'orc', encoding 'utf8')`

版权所有 © 华为云计算技术有限公司

参数	描述	取值范围
encoding	外表中数据源文件的编码格式名称，缺省为utf8。此选项为可选参数。	-
totalrows	可选参数，估计表的行数，仅OBS外表使用。由于OBS上文件可能很多，执行ANALYZE可能会很慢，通过此参数让用户设置一个预估的值，使优化器能通过这个值做大小表的估计。一般预估值和实际值相近时，查询效率较高。	-
filenames	外表中数据源文件，以","间隔。	• 推荐通过使用foldername参数指定数据源的位置，对于只读外表filenames参数与foldername参数两者必有其一，而只写外表和读写外表只能通过foldername指定。 • foldername为绝对目录时，前后必须有'/', 多个路径用', '分隔。 • 查询分区表时，会先根据分区信息进行剪枝，然后查询满足条件的数据文件。由于剪枝操作会涉及多次扫描HDFS分区目录内容，不建议使用重复度非常小的列作为分区列，因为这可能导致分区目录非常的多，增加对HDFS的查询压力。 • OBS只读外表和读写外表不支持。

 版权所有 © 华为云计算技术有限公司

参数	描述	取值范围
delimiter	指定数据文件行数据的字段分隔符，不指定则使用默认分隔符，TEXT格式的默认分隔符是水平制表符（tab）。	支持多字符分隔符，但分隔符不能超过10个字节。 ● 分隔符不能是\r和\n。 ● 分隔符不能和null参数相同。 ● 分隔符不能包含“\”、“.”、数字和字母。 ● 数据文件中单行数据长度需<1GB，如果分隔符较长且数据列较多的情况下，会影响导出有效数据的长度。 ● 分隔符推荐使用多字符（例如'$^&'）和不可见字符（例如0x07、0x08、0x1b等）。 ● delimiter参数只在TEXT和CSV格式下有效。
eol	指定导入数据文件换行符样式。	支持多字符换行符，但换行符不能超过10个字节。常见的换行符，如\r、\n、\r\n（设成0x0D、0x0A、0x0D0A效果是相同的），其他字符或字符串，如$、#。 ● eol参数只能用于TEXT格式的导入。 ● eol参数不能和分隔符、null参数相同。 ● eol参数不能包含：数字，字母和符号“.”。
null	用来指定数据文件中空值的表示。	在TEXT格式下缺省值是\N。 ● null值不能是\r和\n，最大为100个字符。 ● null值不能是分隔符。 ● null参数只在TEXT和CSV格式下有效。
noescaping	TEXT格式下，不对'\'和后面的字符进行转义。	该参数只在TEXT格式下有效。true/on，false/off。缺省值为false/off。

 版权所有 © 华为云计算技术有限公司

参数	描述	取值范围
date_format	导入对于DATE类型指定格式。此语法仅对READ ONLY的外表有效。	合法DATE格式。可参考**时间日期函数和操作符**。 • 对于指定为ORACLE兼容类型的数据库，则DATE类型内建为TIMESTAMP类型。在导入的时候，若需指定格式，可以参考下面的timestamp_format参数。 • date_format参数只在TEXT和CSV格式下有效。
time_format	导入对于TIME类型指定格式。此语法仅对READ ONLY的外表有效。	该参数只在TEXT和CSV格式下有效。 取值范围：合法TIME格式，不支持时区。可参考**时间日期函数和操作符**。
timestamp_format	导入对于TIMESTAMP类型指定格式。此语法仅对READ ONLY的外表有效。	该参数只在TEXT和CSV格式下有效。 取值范围：合法TIMESTAMP格式，不支持时区。可参考**时间日期函数和操作符**。
smalldatetime_format	导入对于SMALLDATETIME类型指定格式。此语法仅对READ ONLY的外表有效。	该参数只在TEXT和CSV格式下有效。 取值范围：合法SMALLDATETIME格式。可参考**时间日期函数和操作符**。
dataencoding	在数据库编码与数据表的数据编码不一致时，该参数用于指定导出数据表的数据编码。比如数据库编码为Latin-1，而导出的数据表中的数据为UTF-8编码。此选项为可选项，如果不指定该选项，默认采用数据库编码。此语法仅对HDFS的WRITE ONLY/READ WRITE外表有效。	该参数只对ORC格式的WRITE ONLY/READ WRITE的HDFS外表有效。 取值范围：该数据库编码支持转换的数据编码。
filesize	指定WRITE ONLY外表的文件大小。此选项为可选项，不指定该选项默认分布式文件系统配置中文件大小的配置值。此语法仅对WRITE ONLY/READ WRITE可写外表有效。	该参数只对ORC格式的WRITE ONLY/READ WRITE的HDFS外表有效。 取值范围：[1, 1024]的整数。

版权所有 © 华为云计算技术有限公司

参数	描述	取值范围
compression	指定文件的压缩方式，此选项为可选项，该参数仅8.2.0及以上集群版本支持。 对WRITE ONLY/READ WRITE的外表：指定ORC格式文件的压缩方式。 对READ ONLY的外表：指定TEXT、CSV或JSON格式文件的压缩方式。	• 当format为ORC时：zlib，zstd，snappy，lz4，缺省值为zlib。 • 当format为PARQUET时：zlib，zstd，snappy，lz4，lz4_hadoop，缺省值为snappy。 • 当format为TEXT、CSV或JSON时：gzip。 **注意** • 8.2.0以下版本不支持READ ONLY外表指定compression参数。 • lz4_hadoop压缩选项，该选项适配hive的lz4，当parquet外表设置该压缩格式时，可以与hive端相互读取。
version	指定ORC格式的版本号，此选项为可选项。此语法仅对WRITE ONLY/READ WRITE的外表有效。	目前仅支持0.12。缺省值为0.12。
julian_adjust	指定是否矫正数据中的儒略日，此选项为可选项，该参数仅8.3.0及以上集群版本支持。 spark2.X版本使用了julian历，3.0版本后使用Proleptic Gregorian公历，导致了parquet 1582年前的部分日期有差别，根据实际情况选择是否需要矫正。	取值为true或false。 缺省值为true。
cache_policy	指定外表磁盘缓存(disk cache)策略。该参数仅存算分离3.0版本支持。	外表只支持ALL和NONE两种缓存策略，ALL表示使用disk cache中的热缓存，NONE则使用冷缓存。热缓存相比冷缓存占用的空间更大，技术上使用更加复杂的替换算法。
column_index_access	控制外表字段与实际文件字段的匹配方式。仅支持ORC和PARQUET格式只读外表。该参数仅9.1.1.200及以上集群版本支持。 • 参数打开时，表示按顺序匹配，即外表第i个字段对应文件第i个字段。 • 参数关闭时表示按字段名进行匹配，外表字段仅匹配文件中的同名字段，对不存在字段使用null填充。	取值为true或false。 缺省值为true。

版权所有 © 华为云计算技术有限公司

参数	描述	取值范围
metadata	针对Iceberg外表，手动指定metadata中所指定的元数据文件。该参数仅9.1.2及以上集群版本支持。	缺省时，通过metadata命名获取最新的metadata文件。 读取metadata中所指定的元数据文件不存在时，会报错，需检视元数据文件是否存在或路径是否正确。

表 13-80 OPTIONS 容错性参数说明

参数	描述	取值范围
fill_missing_fields	当数据导入时，若数据源文件中某一行的末尾字段（一个或多个）缺失的处理方式。	true/on，false/off。缺省值为false/off。 ● true/on，表示末尾字段缺失，则将末尾字段的值设置为NULL，不报错。 ● false/off，表示末尾字段缺失会直接报错，报错信息类似如下： missing data for column "tt" **注意** ● TEXT格式下执行SELECT COUNT(*) 不会去解析具体字段，因此不会对字段缺失情况报错。 ● fill_missing_fields参数只在TEXT和CSV格式下有效。
ignore_extra_data	若数据源文件比外表定义列数多，是否会忽略多出的列。该参数只在数据导入过程中使用。	true/on，false/off。缺省值为false/off。 ● true/on，表示若数据源文件比外表定义列数多，则忽略行尾多出来的列。 ● false/off，表示若数据源文件比外表定义列数多，会显示如下错误信息。 extra data after last expected column **注意** ● 如果行尾换行符丢失，使两行变成一行时，设置此参数为true将导致后一行数据被忽略掉。 ● TEXT格式下执行SELECT COUNT(*) 不会去解析具体字段，因此不会对多余的情况报错。 ● ignore_extra_data参数只在TEXT和CSV格式下有效。

 版权所有 © 华为云计算技术有限公司

参数	描述	取值范围
checkencoding	是否检查字符编码。	no、low、high 。缺省值为low。 TEXT格式下，导入非法字符容错规则如下： • 对于'\0'，容错后转换为空格； • 对于其他非法字符，容错后转换为问号； • 若checkencoding为low，表示导入时对于非法字符进行容错处理，若NULL、DELIMITER设置为空格或问号则会通过如“illegal chars conversion may confuse null 0x20”等报错信息提示用户修改可能引起混淆的参数，以避免导入错误。 ORC格式下，导入非法字符容错规则如下： • checkencoding为no，表示导入时不检查非法字符。在用户确认无非法字符的情况下，此选项可以缩短导入时间。 • checkencoding为low，表示导入时检查到某个字段中包含非法字符，则自动将当前列当前行的字段整体替换为同样长度的‘？’字符； • checkencoding为high，表示导入时检查到某个字段中包含非法字符，则报错退出。
force_mapping	JSON格式下，外表列无法匹配到正确的name-value键值对时的处理方式。	true，false。缺省值为true。 • force_mapping为true，相应的列填null，该null与JSON定义中的null含义相同。 • force_mapping为false，查询报错，提示不存在这样的列。 **注意** 由于对JSON对象没有限制，但外表字段定义需要符合DWS的标识符规范（例如长度、字符等限制），因此这种导入方式可能导致异常：例如，字段无法正确标识JSON name、字段需重复定义等。建议使用容错性选项force_mapping或json操作符（可参考**JSON/JSONB函数和操作符**）来规避。 JSON格式下执行SELECT COUNT(*) 不会去解析具体字段，因此不会对字段缺失、格式错误等情况报错。

表 13-81 OPTIONS 参数说明（仅 Server 类型为 DLI 时设置）

参数	描述	取值范围
dli_project_id	DLI服务对应的项目编号，可在管理控制台上获取项目ID，该参数仅支持server类型为DLI时设置。该参数仅8.1.1及以上集群版本支持。	-
dli_database_name	待访问的DLI多版本表所在的数据库名称，该参数仅支持server类型为DLI时设置。该参数仅8.1.1及以上集群版本支持。	-

版权所有 © 华为云计算技术有限公司

参数	描述	取值范围
dli_table_name	待访问的DLI多版本表的名称，该参数仅支持server类型为DLI时设置。该参数仅8.1.1及以上集群版本支持。	-

表 13-82 OPTIONS 参数说明（仅 format 为 hudi 时设置，以下参数 9.1.0.100 及以上集群版本支持。）

参数	描述	取值范围
auth_server	指定一个server，用于支持hudi外表数据使用KMS加密时的用户身份认证，仅支持format为hudi时指定。	可用的hdfs server名。 MRS服务提供了hudi数据列级加密的能力，通过RangerKMS对用户以及keyId进行管理以及加解密权限控制。用户通过在创建外表时指定auth_server参数，绑定RangerKMS用户信息，进一步进行安全认证、获取密钥，从而实现对读取加密数据的支持。
kms_url	指定KMS服务的ip和端口。仅支持format为hudi、且指定了auth_server时指定。	合法的ip:port风格字符串。 一般情况下，仅指定auth_server即可自动解析KMS服务的ip和端口。如果指定了该参数，则会强制使用地址访问KMS服务。
cow_improve	针对COPY_ON_WRITE表的优化参数，该参数打开时，DN会替代dws-bigdata组件直接从OBS上读取数据。仅支持format为hudi时指定，且不能与auth_server同时指定。	true/on，false/off。缺省值为true/on。 **说明** MERGE_ON_READ表指定该参数无效。

表 13-83 OPTIONS 参数说明（仅 format 为 ICEBERG 时设置，仅 9.1.2 及以上集群版本支持。）

参数	描述	取值范围
cow_improve	针对COPY_ON_WRITE表的优化参数，该参数打开时，DN会替代dws-bigdata组件直接从OBS上读取数据。 ICEBERG格式READ ONLY外表支持该参数。	true/on，false/off。缺省值为true/on。

版权所有 © 华为云计算技术有限公司

表 13-84 OBS 外表支持的 OPTIONS 说明

参数名称	OBS											
-	TEXT	CSV	JSON	ORC			CARBONDATA	PARQUET			HUDI	ICEBERG
	READONLY	READONLY	READONLY	READONLY	WRITEONLY	READWRITE	READONLY	READONLY	WRITEONLY	READWRITE	READONLY	READONLY
location	√	√	√	√	×	×	√	√	×	×	×	×
format	√	√	√	√	√	√	√	√	√	√	√	√
header	×	√	×	×	×	×	×	×	×	×	×	×
delimiter	√	√	×	×	×	×	×	×	×	×	×	×
quote	×	√	×	×	×	×	×	×	×	×	×	×
escape	×	√	×	×	×	×	×	×	×	×	×	×
null	√	√	×	×	×	×	×	×	×	×	×	×
noescaping	√	×	×	×	×	×	×	×	×	×	×	×
encoding	√	√	√	√	√	√	√	√	√	√	√	√
fill_missing_fields	√	√	×	×	×	×	×	×	×	×	×	×

版权所有 © 华为云计算技术有限公司

参数名称	OBS											
ignore_extra_data	√	√	×	×	×	×	×	×	×	×	×	×
date_format	√	√	√	×	×	×	×	×	×	×	×	×
time_format	√	√	√	×	×	×	×	×	×	×	×	×
timestamp_format	√	√	√	×	×	×	×	×	×	×	×	×
smalldatetime_format	√	√	√	×	×	×	×	×	×	×	×	×
chunksize	√	√	√	×	×	×	×	×	×	×	×	×
filenames	×	×	×	×	×	×	×	×	×	×	×	×
foldername	√	√	√	√	√	√	√	√	√	√	√	√
dataencoding	×	×	×	×	×	×	×	×	×	×	×	×
filesize	×	×	×	×	×	×	×	×	×	×	×	×

版权所有 © 华为云计算技术有限公司

参数名称	OBS											
compression	√	√	√	×	√	√	×	×	√	√	×	×
version	×	×	×	×	√	√	×	×	×	×	×	×
checkencoding	√	√	√	√	×	√	√	√	√	√	√	√
totalrows	√	√	√	√	×	√	×	×	×	×	×	×
force_mapping	×	×	√	×	×	×	×	×	×	×	×	×
auth_server	×	×	×	×	×	×	×	×	×	×	√	×
kms_url	×	×	×	×	×	×	×	×	×	×	√	×
cow_improve	×	×	×	×	×	×	×	×	×	×	√	√
julian_adjust	×	×	×	×	×	×	×	√	×	×	√	√
column_index_access	×	×	×	√	×	√	×	√	×	√	×	×

版权所有 © 华为云计算技术有限公司

表 13-85 HDFS 外表支持的 OPTIONS 说明

参数名称	HDFS										
-	TEXT	CSV	JSON	ORC			PARQUET			HUDI	ICEBERG
	READ ONLY	READ ONLY	READ ONLY	READ ONLY	WRITE ONLY	READ WRITE	READ ONLY	WRITE ONLY	READ WRITE	READ ONLY	READ ONLY
location	×	×	×	×	×	×	×	×	×	×	×
format	√	√	√	√	√	√	√	√	√	√	√
header	×	√	×	×	×	×	×	×	×	×	×
delimiter	√	√	×	×	×	×	×	×	×	×	×
quote	×	√	×	×	×	×	×	×	×	×	×
escape	×	√	×	×	×	×	×	×	×	×	×
null	√	√	×	×	×	×	×	×	×	×	×
noescaping	√	×	×	×	×	×	×	×	×	×	×
encoding	√	√	√	√	√	√	√	√	√	√	√
fill_missing_fields	√	√	×	×	×	×	×	×	×	×	×
ignore_extra_data	√	√	×	×	×	×	×	×	×	×	×
date_format	√	√	√	×	×	×	×	×	×	×	×

 版权所有 © 华为云计算技术有限公司

参数名称	HDFS										
time_format	√	√	√	×	×	×	×	×	×	×	×
timestamp_format	√	√	√	×	×	×	×	×	×	×	×
smalldatetime_format	√	√	√	×	×	×	×	×	×	×	×
chunksize	√	√	√	×	×	×	×	×	×	×	×
filenames	√	√	√	√	×	×	√	×	×	×	×
foldername	√	√	√	√	√	√	√	√	√	√	√
dataencoding	×	×	×	×	√	√	×	×	×	×	×
filesize	×	×	×	×	√	√	×	×	×	×	×
compression	√	√	√	×	√	√	×	√	√	×	×
version	×	×	×	×	√	√	×	×	×	×	×
checkencoding	√	√	√	√	√	√	√	√	√	√	√
totalrows	×	×	×	×	×	×	×	×	×	×	×
force_mapping	×	×	√	×	×	×	×	×	×	×	×

版权所有 © 华为云计算技术有限公司

参数名称	HDFS										
auth_server	×	×	×	×	×	×	×	×	×	√	×
kms_url	×	×	×	×	×	×	×	×	×	√	×
cow_improve	×	×	×	×	×	×	×	×	×	√	√
julian_adjust	×	×	×	×	×	×	√	×	×	√	√
column_index_access	×	×	×	√	×	√	√	×	√	×	×

说明

chunksize表示在DN中每个HDFS读取线程的缓存大小，支持text，csv，json格式，默认大小为4MB。

信息约束（Informational Constraint）

在DWS中，数据的约束完全由使用者保证，数据源数据能够严格遵守某种信息约束条件，能够加速对已经具有这种约束特征数据的查询。目前外表不支持索引，所以采取使用Informational Constraint信息优化Plan，提高查询性能。

建立外表信息约束的约束条件：

- 只有用户保证表中的其中一列的非空值具有唯一性时才可以建立Informational Constraint，否则查询结果将与期望值不同。
- DWS的Informational Constraint只支持PRIMARY KEY和UNIQUE两种约束。
- DWS的Informational Constraint支持NOT ENFORCED属性，不支持ENFORCED属性。
- 一个表上的多列可以分别建立UNIQUE类型的Informational Constraint，但是PRIMARY KEY一个表中只能建立一个。
- 一个表的一列上可以建立多个Informational Constraint（由于一个列上有多个约束和一个的作用一致，所以不建议一个列上建立多个Informational Constraint），但是Primary Key类型只能建立一个。
- 不支持表级COMMENT。
- 不支持多列组合约束。
- ORC格式只写外表不支持同一个集群不同CN向同一外表并发导出。

 版权所有 © 华为云计算技术有限公司

- ORC格式只写外表的目录，只能用于DWS的单个外表的导出目录，不能用于多个外表，并且其他组件不能向此目录写入其他文件。

示例：创建 HDFS 外表

在HDFS通过HIVE导入TPC-H benchmark测试数据表part表及region表。part表的文件路径为/user/hive/warehouse/partition.db/part_4，region表的文件路径为/user/hive/warehouse/mppdb.db/region_orc11_64stripe/。

1. 创建HDFS_Server，对应的foreign data wrapper为HDFS_FDW或者DFS_FDW。

```
DROP SERVER IF EXISTS hdfs_server;
CREATE SERVER hdfs_server FOREIGN DATA WRAPPER HDFS_FDW OPTIONS
(address '10.10.0.100:25000,10.10.0.101:25000',
hdfscfgpath '/opt/hadoop_client/HDFS/hadoop/etc/hadoop',
type'HDFS');
```

说明

- 在可选项options里面写入了HDFS集群对应的NameNode的IP地址及端口号。具体端口号请在MRS-HDFS服务配置中搜索参数“dfs.namenode.rpc.port”查看。本示例假设端口号为25000。
- ‘10.10.0.100:25000,10.10.0.101:25000’中列出了两组NameNode的地址及端口号，分别表示HDFS的主NameNode及备NameNode，这里推荐使用该种主备方式填写。两组信息中间使用“,”进行分割。

2. 创建HDFS外表。表关联的HDFS server为hdfs_server，表ft_region对应的HDFS服务器上的文件格式为‘orc’，在HDFS文件系统上对应的文件目录为'/user/hive/warehouse/mppdb.db/region_orc11_64stripe/'。
   - 创建不包含分区列的HDFS外表：

```
DROP FOREIGN TABLE IF EXISTS ft_region;
CREATE FOREIGN TABLE ft_region
(
    R_REGIONKEY INT4,
    R_NAME TEXT,
    R_COMMENT TEXT
)
SERVER
    hdfs_server
OPTIONS
(
    FORMAT 'orc',
    encoding 'utf8',
    FOLDERNAME '/user/hive/warehouse/mppdb.db/region_orc11_64stripe/'
)
DISTRIBUTE BY
     roundrobin;
```

   - 创建包含分区列的HDFS外表：

```
DROP FOREIGN TABLE IF EXISTS ft_part;
CREATE FOREIGN TABLE ft_part
(
     p_partkey int,
     p_name text,
     p_mfgr text,
     p_brand text,
     p_type text,
     p_size int,
     p_container text,
     p_retailprice float8,
     p_comment text
)
SERVER
     hdfs_server
OPTIONS
(
```

 版权所有 © 华为云计算技术有限公司

```
    FORMAT 'orc',
    encoding 'utf8',
    FOLDERNAME '/user/hive/warehouse/partition.db/part_4'
)
DISTRIBUTE BY
    roundrobin
PARTITION BY
    (p_mfgr) AUTOMAPPED;
```

说明

DWS支持2种文件指定方式：通过关键字filenames指定和通过foldername指定。推荐通过使用foldername进行指定。关键字distribute指定了表ft_region的存储分布方式。

3. 查看创建的外表：

```
SELECT * FROM pg_foreign_table WHERE ftrelid='ft_region'::regclass;
```

```
 ftrelid    | ftserver | ftwriteonly |                                                                                                ftoptions                                                                                                | ftreadwriteopts
------------+----------+-------------+---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------+-----------------
 2147560041 |    35023 | f           | {format=orc,encoding=utf8,foldername=/user/hive/warehouse/mppdb.db/region_orc11_64stripe/} |               1
(1 row)
```

```
SELECT * FROM pg_foreign_table where ftrelid='ft_part'::regclass;
```

```
gaussdb=> SELECT * FROM pg_foreign_table where ftrelid='ft_part'::regclass;
 ftrelid    | ftserver | ftwriteonly |                                                                                                ftoptions                                                                                                | ftreadwriteopts
------------+----------+-------------+---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------+-----------------
 2147560044 |    35023 | f           | {format=orc,encoding=utf8,foldername=/user/hive/warehouse/partition.db/part_4} |               1
(1 row)
```

示例：创建 HDFS 只写外表

通过HDFS只写外表，将TPC-H benchmark测试数据表region中的数据导出至HDFS文件系统的/user/hive/warehouse/mppdb.db/region_orc/目录下。

1. 创建HDFS SERVER，对应的foreign data wrapper为HDFS_FDW或者DFS_FDW，同**示例一**。
2. 创建HDFS只写外表。

```
DROP FOREIGN TABLE IF EXISTS ft_wo_region;
CREATE FOREIGN TABLE ft_wo_region
(
    R_REGIONKEY INT4,
    R_NAME TEXT,
    R_COMMENT TEXT
)
SERVER
    hdfs_server
OPTIONS
(
    FORMAT 'orc',
    encoding 'utf8',
    FOLDERNAME '/user/hive/warehouse/mppdb.db/region_orc/'
)
WRITE ONLY;
```

3. 通过只写外表向HDFS文件系统写入数据。

```
INSERT INTO ft_wo_region SELECT * FROM region;
```

示例：创建 HDFS 只写分区外表（仅存算分离 3.0 版本支持）

通过HDFS多级分区只写外表将TPC-H benchmark测试数据表region表中的数据导出HDFS文件系统的/user/hive/warehouse/mppdb.db/region_orc/目录下。

1. 创建HDFS外表，对应的foreign data wrapper为HDFS_FDW或者DFS_FDW，同**示例：创建HDFS外表**。
2. 创建HDFS多级分区只写外表。

```
DROP FOREIGN TABLE IF EXISTS ft_wo_region_partition;
CREATE FOREIGN TABLE ft_wo_region_partition
```


版权所有 © 华为云计算技术有限公司

```
(
    R_REGIONKEY INT4,
    R_NAME TEXT,
    R_COMMENT TEXT
)
SERVER
    hdfs_server
OPTIONS
(
    FORMAT 'orc',
    encoding 'utf8',
    FOLDERNAME '/user/hive/warehouse/mppdb.db/region_orc/'
)
WRITE ONLY
PARTITION BY (R_REGIONKEY, R_NAME);
```

3. 通过只写外表向HDFS文件系统写入数据。

```
INSERT INTO ft_wo_region_partition SELECT * FROM region;
```

示例：创建 OBS 只读外表（仅存算分离 3.0 版本支持）

通过partsupp外表从obs文件系统读取数据。

1. 创建obs server。

须知

- 认证用的AK和SK硬编码到代码中或者明文存储都有很大的安全风险，建议在配置文件或者环境变量中密文存放，使用时解密，确保安全。
- 了解更多创建Server语法请参见**CREATE SERVER**。

```
DROP SERVER IF EXISTS obs_srv;
CREATE SERVER obs_srv FOREIGN DATA WRAPPER dfs_fdw OPTIONS (
    address 'obs.ap-southeast-1.myhuaweicloud.com',
    type 'obs',
    access_key 'xxx',
    secret_access_key 'xxx'
);
```

2. 创建obs外表partsupp，使用disk cache中的冷缓存。

```
DROP FOREIGN TABLE IF EXISTS PARTSUPP;
CREATE FOREIGN TABLE PARTSUPP
(
   PS_PARTKEY     BIGINT NOT NULL,
   PS_SUPPKEY     BIGINT NOT NULL,
   PS_AVAILQTY    BIGINT NOT NULL,
   PS_SUPPLYCOST  DECIMAL(15,2)  NOT NULL,
   PS_COMMENT     VARCHAR(199) NOT NULL
) SERVER obs_srv options (
    encoding     'utf-8',
    format       'parquet',
    foldername   'xxx',
    cache_policy 'NONE'
) READ ONLY DISTRIBUTE BY ROUNDROBIN;
```

3. 通过外表partsupp从obs上读取数据。

```
SELECT * FROM PARTSUPP;
```

示例：创建包含信息约束（指定 not enforced）的 HDFS 外表

关于包含信息约束（Informational Constraint）HDFS外表的相关操作。

- 创建含有信息约束（Informational Constraint）的HDFS外表。

 版权所有 © 华为云计算技术有限公司

```
DROP FOREIGN TABLE IF EXISTS ft_region;
CREATE FOREIGN TABLE ft_region   (
 R_REGIONKEY  int,
 R_NAME TEXT,
 R_COMMENT TEXT
  , primary key (R_REGIONKEY) not enforced)
SERVER hdfs_server
OPTIONS(format 'orc',
    encoding 'utf8',
 foldername '/user/hive/warehouse/mppdb.db/region_orc11_64stripe')
DISTRIBUTE BY roundrobin;
```

- 查看region表是否有信息约束索引：

```
SELECT relname,relhasindex FROM pg_class WHERE oid='ft_region'::regclass;
```

```
gaussdb=> SELECT relname,relhasindex FROM pg_class WHERE oid='ft_region'::regclass;
  relname  | relhasindex
-----------+-------------
 ft_region | f
(1 row)
```

```
SELECT conname, contype, consoft, conopt, conindid, conkey FROM pg_constraint WHERE conname
='ft_region_pkey';
```

```
gaussdb=> SELECT conname, contype, consoft, conopt, conindid, conkey FROM pg_constraint WHERE conname ='ft_region_pkey';
    conname     | contype | consoft | conopt | conindid | conkey
----------------+---------+---------+--------+----------+--------
 ft_region_pkey | p       | t       | t      |        0 | {1}
(1 row)
```

- 删除信息约束：

```
ALTER FOREIGN TABLE ft_region DROP CONSTRAINT ft_region_pkey RESTRICT;

SELECT conname, contype, consoft, conindid, conkey FROM pg_constraint WHERE conname
='ft_region_pkey';
```

```
gaussdb=> SELECT conname, contype, consoft, conindid, conkey FROM pg_constraint WHERE conname ='ft_region_pkey';
 conname | contype | consoft | conindid | conkey
---------+---------+---------+----------+--------
(0 rows)
```

- 添加一个唯一信息约束：

```
ALTER FOREIGN TABLE ft_region ADD CONSTRAINT constr_unique UNIQUE(R_REGIONKEY) NOT
ENFORCED;
```

删除唯一信息约束：

```
ALTER FOREIGN TABLE ft_region DROP CONSTRAINT constr_unique RESTRICT;

SELECT conname, contype, consoft, conindid, conkey FROM pg_constraint WHERE conname
='constr_unique';
```

```
gaussdb=> SELECT conname, contype, consoft, conindid, conkey FROM pg_constraint WHERE conname ='constr_unique';
 conname | contype | consoft | conindid | conkey
---------+---------+---------+----------+--------
(0 rows)
```

- 添加一个唯一信息约束：

```
ALTER FOREIGN TABLE ft_region ADD CONSTRAINT constr_unique UNIQUE(R_REGIONKEY) NOT
ENFORCED disable query optimization;

SELECT relname,relhasindex FROM pg_class WHERE oid='ft_region'::regclass;
```

```
gaussdb=> SELECT relname,relhasindex FROM pg_class WHERE oid='ft_region'::regclass;
  relname  | relhasindex
-----------+-------------
 ft_region | f
(1 row)
```

删除唯一信息约束：

```
ALTER FOREIGN TABLE ft_region DROP CONSTRAINT constr_unique CASCADE;
```

示例：通过 OBS 外表读取 OBS 的 json 数据

通过外表读取OBS上的json数据。

 版权所有 © 华为云计算技术有限公司

1. OBS上有如下json文件，json对象中存在嵌套、数组，部分对象的某些字段缺失，部分对象name重复。

```
{"A" : "simple1", "B" : {"C" : "nesting1"}, "D" : ["array", 2, {"E" : "complicated"}]}
{"A" : "simple2", "D" : ["array", 2, {"E" : "complicated"}]}
{"A" : "simple3", "B" : {"C" : "nesting3"}, "D" : ["array", 2, {"E" : "complicated3"}]}
{"B" : {"C" : "nesting4"},"A" : "simple4",  "D" : ["array", 2, {"E" : "complicated4"}]}
{"A" : "simple5", "B" : {"C" : "nesting5"}, "D" : ["array", 2, {"E" : "complicated5"}]}
```

2. 创建obs_server，对应的foreign data wrapper为DFS_FDW。

```
DROP SERVER IF EXISTS obs_server;
CREATE SERVER obs_server FOREIGN DATA WRAPPER DFS_FDW OPTIONS (
  ADDRESS 'obs.ap-southeast-1.myhuaweicloud.com',
  ACCESS_KEY 'xxxxxxxxx',
  SECRET_ACCESS_KEY 'yyyyyyyyyyyyy',
  TYPE 'OBS'
);
```

说明

- ADDRESS是OBS的终端节点（Endpoint），请根据实际替换。也是使用region参数，通过指定regionCode在region_map文件中查找对应的域名。
- ACCESS_KEY和SECRET_ACCESS_KEY是云账号体系访问密钥。请根据实际替换。
- TYPE表示创建的Server为OBS Server。请保持OBS取值不变。

3. 创建OBS外表json_f，定义字段名，以d#2_e为例，从命名可以看出该字段是数组d的第二个元素里嵌套的e对象。表关联的OBS服务器为obs_server。foldername为外表中数据源文件目录，即表数据目录在OBS上对应的文件目录。

须知

认证用的AK和SK硬编码到代码中或者明文存储都有很大的安全风险，建议在配置文件或者环境变量中密文存放，使用时解密，确保安全。

```
DROP FOREIGN TABLE IF EXISTS json_f;
CREATE FOREIGN TABLE json_f (
  a VARCHAR(10),
  b_c TEXT,
  d#1 INTEGER,
  d#2_e VARCHAR(30)
)SERVER obs_server OPTIONS (
    foldername '/xxx/xxx/',
    format 'json',
    encoding 'utf8',
    force_mapping 'true'
)distribute by roundrobin;
```

4. 查询外表json_f。由于容错性参数force_mapping默认打开，json对象缺失的字段会填NULL；json对象name重复的以最后一次出现的name为准。

```
SELECT * FROM json_f;
```

示例：通过 DLI 外表访问 DLI 上的数据

通过外表读取DLI多版本外表。DLI多版本外表示例仅8.1.1及以上版本支持。

1. 创建dli_server，对应的foreign data wrapper为DFS_FDW。

```
DROP SERVER IF EXISTS dli_server;
CREATE SERVER dli_server FOREIGN DATA WRAPPER DFS_FDW OPTIONS (
  ADDRESS 'obs.ap-southeast-1.myhuaweicloud.com',
  ACCESS_KEY 'xxxxxxxxx',
  SECRET_ACCESS_KEY 'yyyyyyyyyyyyy',
  TYPE 'DLI',
  DLI_ADDRESS 'dli.example.com',
```

 版权所有 © 华为云计算技术有限公司

```
 DLI_ACCESS_KEY 'xxxxxxxxx',
 DLI_SECRET_ACCESS_KEY 'yyyyyyyyyyyyy'
);
```

说明

- ADDRESS是OBS的终端节点（Endpoint）。DLI_ADDRESS是DLI的终端节点（Endpoint），请根据实际替换。
- ACCESS_KEY和SECRET_ACCESS_KEY是云账号体系访问OBS服务的密钥。请根据实际替换。
- DLI_ACCESS_KEY和DLI_SECRET_ACCESS_KEY是云账号体系访问DLI服务的密钥。请根据实际替换。
- TYPE表示创建的Server为DLI Server。请保持DLI取值不变。

2. 创建访问DLI多版本的OBS外表customer_address_ft，不包含分区列，表关联的DLI服务器为dli_server。其中project_id为xxxxxxxxxxxxxxx，dli上的database_name为database123，需要访问的table_name为table456，根据实际替换。

须知

认证用的AK和SK硬编码到代码中或者明文存储都有很大的安全风险，建议在配置文件或者环境变量中密文存放，使用时解密，确保安全。

```
DROP FOREIGN TABLE IF EXISTS customer_address_ft;
CREATE FOREIGN TABLE customer_address_ft
(
    ca_address_sk             integer               not null,
    ca_address_id             char(16)              not null,
    ca_street_number          char(10)                      ,
    ca_street_name            varchar(60)                   ,
    ca_street_type            char(15)                      ,
    ca_suite_number           char(10)                      ,
    ca_city                   varchar(60)                   ,
    ca_county                 varchar(30)                   ,
    ca_state                  char(2)                       ,
    ca_zip                    char(10)                      ,
    ca_country                varchar(20)                   ,
    ca_gmt_offset             decimal(36,33)                ,
    ca_location_type          char(20)
)
SERVER dli_server OPTIONS (
    FORMAT 'ORC',
    ENCODING 'utf8',
    DLI_PROJECT_ID 'xxxxxxxxxxxxxxx',
    DLI_DATABASE_NAME 'database123',
    DLI_TABLE_NAME 'table456'
)
DISTRIBUTE BY roundrobin;
```

3. 通过外表查询DLI多版本表的数据。

```
SELECT COUNT(*) FROM customer_address_ft;
```

示例：创建 Iceberg 只读外表

1. 创建Iceberg外表（不指定metadata）。

```
CREATE foreign TABLE customer
(
    C_CUSTKEY     INT
  , C_NAME        VARCHAR(25)
  , C_ADDRESS     VARCHAR(40)
  , C_NATIONKEY   INT
```


版权所有 © 华为云计算技术有限公司

```
 , C_PHONE       CHAR(15)
 , C_ACCTBAL     DECIMAL(15,2)
 , C_MKTSEGMENT  CHAR(10)
 , C_COMMENT     VARCHAR(117)
)
SERVER obs_server
OPTIONS(encoding 'utf-8', format 'iceberg', foldername '/bucket/dws_iceberg/customer/') distribute
by roundrobin;
```

2. 创建Iceberg外表（指定metadata）。

```
CREATE foreign TABLE customer_01
(
   C_CUSTKEY     INT
 , C_NAME        VARCHAR(25)
 , C_ADDRESS     VARCHAR(40)
 , C_NATIONKEY   INT
 , C_PHONE       CHAR(15)
 , C_ACCTBAL     DECIMAL(15,2)
 , C_MKTSEGMENT  CHAR(10)
 , C_COMMENT     VARCHAR(117)
)
SERVER obs_server
OPTIONS(encoding 'utf-8', format 'iceberg', foldername '/bucket/dws_iceberg/customer/',metadata
'00001-a64c9059-ea93-482e-847c-d136fc6e281a.metadata.json') distribute by roundrobin;
```

3. 查询Iceberg外表。

```
SELECT * FROM customer;
SELECT * FROM customer_01 ORDER BY C_CUSTKEY;
```

相关链接

- 修改和删除外表语法请参见**ALTER FOREIGN TABLE (For HDFS or OBS)**和**DROP FOREIGN TABLE**。
- **更多最佳实践**：
  - 使用HDFS外表访问MRS远端数据，请参见**从MRS Hive导入表数据到DWS集群**。
  - 使用OBS外表，请参见**导入OBS桶数据到DWS集群及从DWS导出数据到OBS桶**。
  - 使用DLI外表，请参见**从DLI导入表数据到DWS集群**。

13.4.9 CREATE FOREIGN TABLE (OBS 导入导出)

功能描述

在当前数据库创建一个OBS外表，用于OBS数据并行导入导出。该方式不需要用户再额外创建外部服务器server，使用数据库默认创建的gsmpp_server即可。

本语法功能仅支持OBS桶上的**TEXT和CSV格式**数据，如果需要处理**ORC、CARBONDATA、PARQUET**等格式，请参见**CREATE FOREIGN TABLE (SQL on OBS or Hadoop)**，且需要额外创建外部server。

外表（Foreign Table）是数据库或大数据平台中一种虚拟表结构，它不实际存储数据，而是通过元数据信息（如表结构、存储位置、文件格式等）将外部存储系统（如HDFS、OBS等）中的数据映射到数据库内部，使用户能够以标准SQL操作直接访问或操作外部数据。外表的本质是一种“数据代理”，将外部文件系统的数据逻辑映射成数据库可识别的表结构。

版权所有 © 华为云计算技术有限公司

注意事项

- 这种方式仅支持TEXT和CSV格式，并且需要额外指定OBS连接信息。
- 使用OBS外表访问OBS桶数据时，应确保DWS与OBS桶在同一个区域(Region)下，例如都在“中国-香港”下。
- 创建的外表分为只读外表（READ ONLY）和只写外表（WRITE ONLY），缺省为只读外表。导入OBS数据时，请将外表设为READ ONLY；将数据导出到OBS时，请设为WRITE ONLY。
- 仅系统管理员dbadmin角色或已授权了USEFT权限的普通用户可执行外表操作。
  ```
  ALTER USER user_name USEFT;
  ```
- OBS外表不需要显式指定分布方式，默认支持ROUNDROBIN分布方式；
- 所创建外表只对**信息约束（Informational Constraint）**约束生效。
- 不支持OBS桶中带中文路径的场景。

表 13-86 OBS 外表支持读写格式说明

数据类型	DIST_FDW	
-	READ ONLY	WRITE ONLY
ORC	×	×
PARQUET	×	×
CARBONDATA	×	×
TEXT	√	√
CSV	√	√
JSON	×	×

语法格式

```
CREATE FOREIGN TABLE [ IF NOT EXISTS ] table_name
( { column_name type_name [column_constraint ]
    | LIKE source_table | table_constraint [, ...]} [, ...] )
SERVER server_name
OPTIONS ( { option_name ' value ' } [, ...] )
[ { WRITE ONLY | READ ONLY }]
[ WITH error_table_name | LOG INTO error_table_name]
[PER NODE REJECT LIMIT 'value'] ;
```

- 其中column_constraint为：
  ```
  [CONSTRAINT constraint_name]
  {PRIMARY KEY | UNIQUE}
  [NOT ENFORCED [ENABLE QUERY OPTIMIZATION | DISABLE QUERY OPTIMIZATION] | ENFORCED]
  ```
- 其中table_constraint为：
  ```
  [CONSTRAINT constraint_name]
  {PRIMARY KEY | UNIQUE} (column_name)
  [NOT ENFORCED [ENABLE QUERY OPTIMIZATION | DISABLE QUERY OPTIMIZATION] | ENFORCED]
  ```


版权所有 © 华为云计算技术有限公司

参数说明

表 13-87 CREATE FOREIGN TABLE (OBS 导入导出)参数说明

参数	描述	取值范围或示例
IF NOT EXISTS	如果已经存在相同名称的表，系统不会报错，而会发出一个通知，告知表关系已存在。	-
table_name	外表的表名。	字符串，要符合标识符的命名规范。参见**标识符命名规范**。
column_name	外表中的字段名。	字符串，要符合标识符的命名规范。参见**标识符命名规范**。
type_name	字段的数据类型。	-
SERVER server_name	外表的server名字。对于导入导出的OBS外表，可以使用初始数据库默认创建的gsmpp_server，也可以使用自定义的server。	• 如果使用自定义server，FOREIGN DATA WRAPPER应该为dist_fdw。 • 8.2.0及以上集群版本支持在自定义dist_fdw server中指定OBS访问相关参数：access_key、secret_access_key和security_token。当server中指定以上参数时，外表中不需要再重复指定。
OPTIONS	用于指定外表数据的各类参数。	• 数据格式参数，参见**表13-88**。 • 容错性参数，参见**表13-89**。 • 性能参数，参见**表13-90**。
READ ONLY	外表只读，该参数只供数据导入使用。	-
WRITE ONLY	外表只写。该参数只供数据导出使用。	-

版权所有 © 华为云计算技术有限公司

参数	描述	取值范围或示例
WITH error_table_name	数据导入过程中出现的数据格式错误信息将被写入error_table_name指定的错误信息表中，可以在并行导入结束后查询此错误信息表，获取详细的错误信息。此参数只在设置了reject_limit参数时有效。	取值范围：字符串，要符合标识符的命名规范。 **注意** 如果为了兼容postgres开源接口，此语法建议用LOG INTO代替。该参数指定时错误表自动创建。
LOG INTO error_table_name	数据导入过程中出现的数据格式错误信息将被写入error_table_name指定的错误信息表中，可以在并行导入结束后查询此错误信息表，获取详细的错误信息。	取值范围：字符串，要符合标识符的命名规范。 ● 若没有指定PER NODE REJECT LIMIT参数，则此参数不起作用。 ● 该参数指定时，错误表自动创建。
PER NODE REJECT LIMIT 'value'	指定本次数据导入过程中每个DN实例上允许出现的数据格式错误的数量，如果有一个DN实例上的错误数量大于设定值，本次导入失败，报错退出。	取值范围：整型值，unlimited（无限），不指定该参数时，有错误信息立即返回。 **注意** 此语法指定的是单个节点的错误容忍度。 数据格式错误是指缺少或者多出字段值，数据类型错误或者编码错误。对于非数据格式错误，一旦发生就将导致整个数据扫描失败。

 版权所有 © 华为云计算技术有限公司

参数	描述	取值范围或示例
NOT ENFORCED	指定所建立的约束为信息约束，即数据库不会强制验证该约束，这个选项与后续的ENABLE QUERY OPTIMIZATION搭配使用，主要有以下作用： • **仅为信息约束**：通过 PRIMARY KEY 或 UNIQUE 声明列的唯一性约束，帮助优化查询执行计划。 • **不强制验证**：数据库不会检查外部数据源中的数据是否实际满足约束条件（如是否存在重复值）。 • **性能优化**：允许查询优化器利用约束信息生成更高效地执行计划，但需用户自行确保数据符合约束，否则可能导致查询结果不准确。 • **风险**：若实际数据违反约束（如存在重复值），可能导致查询结果错误（如聚合结果异常、连接产生多余记录）。 更多信息参见**location参数说明**。	例如声明primary key，但是不强制验证。 `CREATE FOREIGN TABLE hdfs_users ( user_id INT PRIMARY KEY NOT ENFORCED, ...`
ENFORCED	数据库强制验证约束，预留参数，目前DWS不支持设置为ENFORCED。	-
PRIMARY KEY (column_name)	指定所建立的信息约束位于column_name列上。	取值范围：字符串，要符合标识符的命名规范，并且这个column_name必须存在。
ENABLE QUERY OPTIMIZATION	允许查询优化器利用信息约束生成更高效的执行计划，与前面的NOT ENFORCED搭配使用。	创建外表，为字段名p_partkey int增加唯一约束，但是不强制验证该约束，且允许优化器利用该信息约束生成更优执行计划： `CREATE FOREIGN TABLE ft_part ( p_partkey int UNIQUE NOT ENFORCED ENABLE QUERY OPTIMIZATION, ...`

 版权所有 © 华为云计算技术有限公司

参数	描述	取值范围或示例
DISABLE QUERY OPTIMIZATION	禁止查询优化器利用信息约束生成高效的执行计划。	-

表 13-88 OPTIONS 数据格式参数说明

参数	描述	取值范围或示例
encrypt	数据传输过程中使用HTTPS，否则使用HTTP，默认off。	-
access_key	OBS访问协议对应的AK值（由用户从服务界面上用户信息里获取），创建外表时AK值不会加密保存到数据库的元数据表中。创建外表时不校验该正确性。 AK可参见**访问密钥**获取。	-
secret_access_key	OBS访问协议对应的SK值（由用户从服务界面上用户信息里获取），创建外表时SK值会加密保存到数据库的元数据表中。创建外表时不校验该正确性。 SK可参见**访问密钥**获取。	-
security_token	对应统一身份认证服务IAM中临时安全凭证的SecurityToken值，与临时AK、SK共同组成临时安全凭证，临时安全凭证有效期不超过24小时。该参数仅8.2.0及以上集群版本支持。 **注意** • 该参数仅8.2.0及以上集群版本支持。 • 使用该参数时，access_key和secret_access_key参数分别对应临时AK、SK。	-
chunksize	在DN中每个OBS读取线程的缓存大小，可指定范围8~512，默认大小为64，单位为MB。	-

版权所有 © 华为云计算技术有限公司

参数	描述	取值范围或示例
location	OBS外表的数据源位置，即OBS数据的存放目录，目前支持URL描述，格式为'obs://OBS桶名/文件夹名'，多个URL使用‘|’分割。应确保OBS桶与DWS集群在同一个区域，不支持跨区访问OBS桶数据。 使用该参数请参见**location参数说明**。	举例如下： OPTIONS (encoding 'utf8', location 'obs://<obs_bucket_name>/traffic-data/gcxx', format 'text', ...
region	可选参数，region参数指定regionCode，regionCode为云上的region信息。 若显式指定此参数，域名信息将会读取指定的region参数；若此参数不指定，则读取defaultRegion的值，即安装集群时指定的region。	TEXT、CSV格式的OBS导入导出外表格式参数使用说明如下： • location参数必选，其中前缀gsobs、obs均支持，都识别为OBS的信息，若为gsobs时，其中包含obs url、bucket、prefix，若为obs时则表示bucket、prefix。 • 多个桶的数据源数据之间使用分隔符‘|’进行分割，LOCATION 'obs://bucket1/folder/ | obs://bucket2/'，数据库将会扫描指定路径文件夹下面的所有对象。
format	外表中数据源文件的格式。	取值范围：CSV、TEXT，缺省值为TEXT。DWS只支持CSV和TEXT格式。 • CSV（逗号分隔文件格式）： – 格式的文件，可以有效处理数据列中的换行符，但对一些特殊字符处理有欠缺。 – 由记录组成，每条记录被分隔符分隔为字段，且每条记录都有同样的字段序列。 • TEXT（文本格式）： – 由换行符区分每条记录，由分隔符区分每个字段。可以有效处理一些特殊字符，但无法正确处理数据列中的换行符。

版权所有 © 华为云计算技术有限公司

参数	描述	取值范围或示例
header	指定导出数据文件是否包含标题行，标题行一般用来描述表中每个字段的信息。	OBS导出数据时不支持该参数为true/on；在导入数据时，如果header选项为true/on，则数据文件中第一行会被识别为标题行，会忽略此行。如果header为false/off，数据文件中第一行会被识别为数据。 取值范围：true/on，false/off。 缺省值为false/off。
delimiter	指定数据文件行数据的字段分隔符，不指定则使用默认分隔符，TEXT格式的默认分隔符是水平制表符（tab），CSV格式的默认分隔符为“,”。	支持多字符分隔符，但分隔符不能超过10个字节。 ● TEXT格式，分隔符不能是\r和\n。 ● 分隔符不能和null参数相同，CSV格式数据的分隔符不能和quote参数相同。 ● TEXT格式数据的分隔符不能包含：字母、数字以及特殊字符“\”和“.”。 ● 数据文件中单行数据长度需<1GB，如果分隔符较长且数据列较多的情况下，会影响导出有效数据的长度。 ● 分隔符推荐使用多字符（例如'$^&'）和不可见字符（例如0x07、0x08、0x1b等）。
quote	CSV格式文件下的引号字符，缺省值为双引号。	● quote参数不能和分隔符、null参数相同。 ● quote参数只能是单字节的字符。 ● 推荐不可见字符作为quote，例如0x07，0x08，0x1b等。
escape	CSV格式下，用来指定逃逸字符，逃逸字符只能指定为单字节字符。	缺省值为双引号。当与quote值相同时，会被替换为'\0'。

版权所有 © 华为云计算技术有限公司

参数	描述	取值范围或示例
null	用来指定数据文件中空值的表示方式。	• 在TEXT格式下缺省值是\N。 • CSV格式下缺省值是一个没有引号的空字符串。 • null的值不能是\r和\n，最大为100个字符。 • null值不能和分隔符、quote参数相同。
noescaping	TEXT格式下，开启后不对'\'和后面的字符进行转义。	noescaping参数只在TEXT格式下有效。 true/on，false/off。缺省值为false/off。
encoding	指定数据文件的编码格式名称，即需要以何编码格式对数据文件进行解析和校验/输出文件为何种编码格式。缺省值为当前数据库的默认客户端编码格式，即client_encoding。	导入外表此处强烈建议指定为文件的编码格式，或根据文件的字符集在导入前对client_encoding进行设置。否则可能会导致不必要的解析、校验错误以及其导致的导入报错回滚，甚至非法数据入库。导出外表同样希望指定此选项，以避免导出采用默认字符集设置时与预期不符。 在创建外表时此选项未指定，会在客户端给出对应Warning信息。 **注意** • 目前OBS导入外表不支持解析带有多种字符集编码格式混合的文件。 • 目前OBS导出外表不支持写出带有多种字符集编码格式混合的文件。

版权所有 © 华为云计算技术有限公司

参数	描述	取值范围或示例
eol	指定导入导出数据文件换行符样式。	取值范围：支持多字符换行符，但换行符不能超过10个字节。常见的换行符，如\r、\n、\r\n（设成0x0D、0x0A、0x0D0A效果是相同的），其他字符或字符串，如$、#。 ● eol参数只能用于TEXT格式的导入导出，不支持CSV格式。 ● eol参数不能和分隔符、null参数相同。 ● eol参数不能包含：数字，字母和符号“.”。
date_format	导入对于DATE类型指定格式。此语法仅对READ ONLY的外表有效。	取值范围：合法DATE格式。可参考**时间日期函数和操作符**。 **说明** 对于指定为ORACLE兼容类型的数据库，则DATE类型内建为TIMESTAMP类型。在导入的时候，若需指定格式，可以参考下面的timestamp_format参数。
time_format	导入对于TIME类型指定格式。此语法仅对READ ONLY的外表有效。	取值范围：合法TIME格式，不支持时区。
timestamp_format	导入对于TIMESTAMP类型指定格式。此语法仅对READ ONLY的外表有效。	取值范围：合法TIMESTAMP格式，不支持时区。
smalldatetime_format	导入对于SMALLDATETIME类型指定格式。此语法仅对READ ONLY的外表有效。	取值范围：合法SMALLDATETIME格式。
bom	标识CSV文件是否包含utf8 BOM字段。	仅在只读外表且文件编码为utf8时生效。 取值范围：true/on，false/off。 缺省值：false

版权所有 © 华为云计算技术有限公司

表 13-89 OPTIONS 容错性参数说明

参数	描述	取值范围
fill_missing_fields	当数据导入时，若数据源文件中某一行的末尾字段（一个或多个）缺失的处理方式。	true/on，false/off。缺省值为false/off。 ● true/on，表示末尾字段缺失，则将末尾字段的值设置为NULL，不报错。 ● false/off，表示末尾字段缺失会直接报错，报错信息类似如下： `missing data for column "tt"`
ignore_extra_data	数据源文件中的字段比外表定义列数多时，是否忽略多出的列。该参数只在数据导入过程中使用。	true/on，false/off。缺省值为false/off。 ● true/on，表示若数据源文件比外表定义列数多，则忽略行尾多出来的列。 ● false/off，表示若数据源文件比外表定义列数多，会显示如下错误信息。 `extra data after last expected column` **注意** 如果行尾换行符丢失，使两行变成一行时，设置此参数为true将导致后一行数据被忽略掉。
reject_limit	指定本次数据导入允许出现的数据格式错误个数，当导入过程中出现的数据格式错误未达到限定值时，本次数据导入可以成功。	整型值、unlimited（无限制）。 不指定该参数时，有错误信息立即返回。 **注意** 此语法建议用PER NODE REJECT LIMIT 'value'替代。 数据格式错误是指缺少或者多出字段值，数据类型错误或者编码错误。对于非数据格式错误，一旦发生就将导致整个数据导入失败。
force_save_err	表示是否在导入报错退出后，保存错误信息至错误表中。	true/on，false/off。缺省值为false/off。 与reject_limit配合使用，若该参数开启： ● reject_limit不指定时，保留一条错误记录至错误表中。 ● reject_limit设置为N，保留N+1条错误记录至错误表中。

版权所有 © 华为云计算技术有限公司

参数	描述	取值范围
obs_null_file	用于dws和obs之间空文件导入导出的处理。 **注意** • 该参数仅8.2.1及以上版本支持。 • obs_null_file为true/on时，导出时若导出目录只有“_SUCCESS”空文件时，可以重复导出，obs_null_file为false/off时不能。 • obs_null_file为true/on时，多桶导入，对于第一个文件不存在的路径进行报错。	true/on，false/off。缺省值为false/off。 当obs_null_file为true/on时： • dws导出空表时，生成名字为“_SUCCESS”的空文件，表示导出成功；导出非空表时，生成本身的文件+“_SUCCESS”空文件。 • dws导入文件时，如果文件不存在或者路径错误，会显示如下错误信息。 `No such file or directory: 'XXX'`
compatible_illegal_chars	导入非法字符容错参数。此语法仅对READ ONLY的外表有效。	true/on，false/off。缺省值为false/off。 • true/on，表示则导入时遇到非法字符进行容错处理，非法字符转换后入库，不报错，不中断导入。 • false/off，表示导入时遇到非法字符进行报错，中断导入。 • 导入非法字符容错规则如下： 1. 对于'\0'，容错后转换为空格； 2. 对于其他非法字符，容错后转换为问号； 3. 若compatible_illegal_chars为true/on，表示导入时对于非法字符进行容错处理，若NULL、DELIMITER、QUOTE、ESCAPE设置为空格或问号则会通过如“illegal chars conversion may confuse COPY escape 0x20”等报错信息提示用户修改可能引起混淆的参数以避免导入错误。 **注意** Windows平台下OBS若按照文本格式读取数据文件，遇到0x1A会作为EOF符号结束数据读入造成解析错误，这是Windows平台的实现约束。由于OBS不支持BINARY形式读取，可将相应数据文件交由Linux平台下的OBS读取。

 版权所有 © 华为云计算技术有限公司

表 13-90 OPTIONS 性能参数说明

参数	描述	取值范围
file_split_threshold	TEXT格式导入性能优化参数，含义是指定文件逻辑分块的大小下限。当指定该参数时，会根据实际文件和DN状况，通过对大文件进行拆分来提高导入并发度，最终目的是让各个DN的任务均衡分布。因此适合文件数少于DN数或文件大小不均衡的导入场景。	0~2147483647，单位MB，默认为0，不会拆分文件。 ● 该参数仅8.2.0及以上版本支持。 ● 该参数仅支持TEXT格式READ ONLY外表。 ● 该参数的含义是指定文件逻辑分块的大小下限，因此文件不一定按照该参数切块，但切块大小一定不会小于file_split_threshold。 例如：当前文件大小为1024MB，DN数为4。当指定file_split_threshold小于256时，文件会被均匀的分成4块，每个DN分配一个256MB的文件导入任务；当指定file_split_threshold为500时，由于用户明确告知不希望分块小于500MB，因此最终文件被拆分成500MB和524MB，分配给2个DN。多文件时该参数仍适用。 ● 除非有明确的分块大小需求，一般建议设置为较小的值（例如10），否则可能影响并发效果。

location 参数说明

- 对于只读外表（默认为只读）的URL末尾可以指定到对象路径的前缀或直接指定到对象全路径。指定方式为obs://bucket/prefix。（其中，prefix是指对象路径的前缀。）例如：obs://mybucket/tpch/nation/
- 对于obs://bucket/prefix格式，若显式指定region参数时，域名信息将会读取指定的region参数；若region参数不指定，则读取defaultRegion的值，即安装集群时指定的region。
- 对于可写外表，URL不需要指定文件名。外部表数据源位置只可指定一个，并且要预先创建好对应目录。
- 对于只读外表不能使用多个相同的URL地址。
- 向外表中插入数据需要指定location。
- location参数中前缀gsobs、obs均支持，都识别为OBS的信息，若为gsobs时，其中包含obs url、bucket、prefix，若为obs时则表示bucket、prefix。

在实际导入导出数据时，location参数使用建议如下：

 版权所有 © 华为云计算技术有限公司

- 导入时“location”建议指定到具体文件名。如果仅指定到OBS桶或目录，则会导入其中的所有文本文件。当数据格式不正确时，则会报错。如果设置了容错，则容错表可能导入大量数据。
- 支持OBS单桶多文件导入，根据文件名前缀进行匹配，匹配到的文件都会被导入。

  例如，有以下两个数据文件，只要在“location”中指定前缀mybucket/input_data/product_info就能识别并导入这两个文件。

  ```
  mybucket/input_data/product_info.0
  mybucket/input_data/product_info.1
  ```
- 导入时如果指定到文件名，例如“1.csv”，那么在此文件的桶或目录存在此名称为前缀的其他文件，也会被导入。即“1.csv1”、“1.csv22”等，都会被自动导入。
- 导入时，“location”中如果使用obs方式，支持多个url，并且用‘|’分隔；如果使用gsobs方式，则不支持多个路径。
- 导出时“location”默认按目录处理。如果仅指定到自定义名称的文件，则导出时会以该文件为名称创建目录，然后再生成导出文件。文件名由DWS自动生成。
- 导出时“location”只支持一个路径。

示例

创建外表OBS_ft，用来以txt格式导入OBS上指定的对象数据到row_tbl表中：

⚠ 注意

认证用的AK和SK硬编码到代码中或者明文存储都有很大的安全风险，建议在配置文件或者环境变量中密文存放，使用时解密，确保安全。

```
DROP FOREIGN TABLE IF EXISTS OBS_ft;

CREATE FOREIGN TABLE OBS_ft
( a int, b int)
SERVER gsmpp_server
OPTIONS
(location 'obs://gaussdbcheck/obs_ddl/test_case_data/txt_obs_informational_test001',
format 'text',
encoding 'utf8',
chunksize '32',
 encrypt 'on',
ACCESS_KEY 'access_key_value_to_be_replaced',
SECRET_ACCESS_KEY 'secret_access_key_value_to_be_replaced',
delimiter E'\x08')
read only;

DROP TABLE IF EXISTS row_tbl;


CREATE TABLE row_tbl( a int, b int);

INSERT INTO row_tbl SELECT * FROM OBS_ft;
```

相关链接

ALTER FOREIGN TABLE (For HDFS or OBS)，**DROP FOREIGN TABLE**

最佳实践：**使用DWS秒级查询交通卡口通行车辆行驶路线**

版权所有 © 华为云计算技术有限公司

优化建议

- delimiter
  - 分隔符不能是\r和\n，分隔符不能和null参数相同，CSV格式数据的分隔符不能和quote参数相同。
  - 数据文件中单行数据长度需<1GB，如果分隔符较长且数据列较多的情况下，会影响导出有效数据的长度。
  - 分隔符推荐使用多字符（例如'$^&'）和不可见字符（例如0x07、0x08、0x1b等）。
- quote
  - quote参数只能是单字节的字符，quote参数不能和分隔符、null参数相同。
  - 推荐不可见字符作为quote，例如0x07，0x08，0x1b等。
- mode Normal
  - 支持所有文件格式(包括CSV、TEXT、FIXED)，数据导入需要在数据服务器上启动Gauss data service协助完成。
- mode Shared
  - 支持TEXT文件格式，此时导入无需Gauss data service的协助，但是需要将用户数据通过NFS或挂载到所有节点的相同路径下。
- mode Private
  - 适用于用户数据已经存放到Datanode节点的本地目录下，并且位于相同的路径。

13.4.10 CREATE FOREIGN TABLE (GDS 导入导出)

功能描述

在当前数据库创建一个GDS外表，用于数据并行导入导出。GDS外表分为只读外表和只写外表，分别用于数据并行导入和并行导出，缺省为只读外表。

GDS是一款运行在Linux操作系统的命令行工具，通过和外表机制的配合，实现数据的高速导入导出。GDS工具包需要安装在数据源文件所在的服务器上，数据源文件所在的服务器称为数据服务器，也叫GDS服务器。

安装和配置GDS，请参见**安装配置和启动GDS**。

注意事项

- 仅系统管理员dbadmin角色或已授权了USEFT权限的普通用户可执行外表操作。
  ```
  ALTER USER user_name USEFT;
  ```
- GDS外表不需要显式指定分布方式，默认支持ROUNDROBIN分布方式。
- 对于GDS外表指定任何约束（列约束、表约束等）均不生效。
- GDS导入导出支持的文件格式：TEXT、CSV和FIXED。

语法格式

```
CREATE FOREIGN TABLE [ IF NOT EXISTS ] table_name
   ( [ { column_name type_name POSITION(offset,length) | LIKE source_table } [, ...] ] )
   SERVER gsmpp_server
   OPTIONS ( { option_name ' value ' } [, ...] )
   [ { WRITE ONLY | READ ONLY }]
   [ WITH error_table_name | LOG INTO error_table_name]
```

 版权所有 © 华为云计算技术有限公司

```
[REMOTE LOG 'name']
[PER NODE REJECT LIMIT 'value']
[ TO { GROUP groupname | NODE ( nodename [, ... ] ) } ];
```

参数说明

表 13-91 CREATE FOREIGN TABLE (GDS 导入导出)参数说明

参数	描述	取值范围
IF NOT EXISTS	如果已经存在相同名称的表，系统不会报错，而会发出一个通知，告知表关系已存在。	-
table_name	外表的表名。	字符串，要符合标识符的命名规范。参见**标识符命名规范**。
column_name	外表中的字段名。	
type_name	字段的数据类型。	-
POSITION(offset,length)	在固定长度模式中，定义每一个字段在数据文件中的位置。 offset为该字段在数据源文件中的起始位置，length为该字段的长度。	offset取值不能小于0字节，单位为字节。 每条记录的长度不能大于1GB，文件中没有出现的列默认以空值代替。
SERVER gsmpp_server	外表的server名字。对于GDS外表，其server是初始数据库默认创建的，即gsmpp_server。	-
OPTIONS	用于指定外表数据的各类参数。	• 数据格式参数：参见**表13-92**。 • 容错性参数：参见**表13-93**。 • 性能参数：参见**表13-94**。
READ ONLY	只读外表，该参数只供数据导入使用。	-

版权所有 © 华为云计算技术有限公司

表 13-92 OPTIONS 数据格式参数说明

参数	描述	取值范围
location	外表的数据源位置，目前支持URL方式的描述。多个URL使用‘|’分隔。 gds目前可以支持导出的时候自动创建外表定义的目录。 如外表location定义“gsfs://192.168.0.91:5000/2019/09”执行导出任务的时候，如果gds数据目录下的子目录“2019/09”不存在则会自动的创建该子目录，不需要用户手动创建外表中指定的目录。	● 对于使用GDS从远端服务器并行导入时的只读外表（默认为只读）的URL末尾必须指定文件的匹配模式或者文件名。 例如: gsfs://192.168.0.90:5000/*或者file:///data/data.txt或者gsfs:// 192.168.0.90:5000/* | gsfs:// 192.168.0.91:5000/* ● 对于使用GDS并行导出到远端服务器时的只写外表，URL不需要指定文件名。当导出数据文件存储位置为远端URL时，例如gsfs://192.168.0.90:5000/，则数据源位置可指定多个，此时：若导出数据文件存储位置数量小于等于数据节点数量时，使用此外部表执行导出任务，数据将被平均分配至各数据源位置；若导出数据存储位置数量大于数据节点数量时，执行导出任务，数据将被平均分配给此位置列表中从前端开始等于数据节点数量的数据源位置下，剩余数据源位置仍会创建数据文件，但文件中不会有任何数据。 ● 对于使用GDS从远端服务器并行导入时的只读外表，URL个数应小于DN个数，且不能使用多个location相同的URL。 ● 当使用gsfss协议，即当URL为“gsfss://”开头，进行加密导入导出时，并发数量不能超过10。 ● gds导出时location指定的文件路径“gsfs://127.0.0.1:7789/2019/09/”中的2019/09子目录会在执行导出任务的时候自动创建。 ● 设置file_type为“pipe”时，GDS会根据URL中最后一个字符是否为“/”来判断导入导出的目标是管道文件还是目录。如： – gsfs://192.168.0.90:5000/a/b，GDS会将b识别成一个管道文件。 – gsfs://192.168.0.90:5000/a/b/，GDS将b识别成一个目录，并在b目录下创建管道文件。

 版权所有 © 华为云计算技术有限公司

参数	描述	取值范围
format	外表中数据源文件的格式。	取值范围：CSV、TEXT、FIXED，缺省值为TEXT。 ● CSV格式的文件，对一些转义序列按照普通字符串进行处理，可以有效处理数据列中的换行符。 ● TEXT格式的文件，仅支持部分转义序列，无法正确处理数据列中的换行符。 ● FIXED格式的文件，适用于每条数据的数据列都比较固定的数据，长度不足的列会添加空格补齐。 **说明** ● 转义序列指的是反斜杠开头的字符串，包括：\b（退格）、\f（换页）、\n（换行）、\r（回车）、\t（横向制表）、\v（纵向制表）、\数字（八进制编码）、\x数字（十六进制编码）。 TEXT格式支持按转义序列本身含义进行处理，CSV和FIXED格式只能按照普通字符串进行处理。 TEXT格式导出时，若字段值存在反斜杠和分隔符，系统会自动在反斜杠和分隔符前添加反斜杠（例如，\变为 \\）进行转义，避免解析时发生混淆或错位。 ● 定长格式（FIXED）定义如下。当为FIXED时，必须为每一列指定POSITION。 1. 每条记录的每个字段长度相同。 2. 长度不足的字段以空格填充，数字类型字段左对齐，字符字段右对齐。 3. 字段和字段之间没有分隔符。

 版权所有 © 华为云计算技术有限公司

参数	描述	取值范围
header	指定导出数据文件是否包含标题行，标题行一般用来描述表中每个字段的信息。header只能用于CSV，FIXED格式的文件中。 在导入数据时，如果header选项为on，则数据文件中第一行会被识别为标题行，会忽略此行。如果header为off，数据文件中第一行会被识别为数据。 在导出数据时，如果header选项为on，则需要指定**fileheader**。fileheader用来指定导出头文件的格式。如果header为off，则导出数据文件不包含标题行。	true/on，false/off。缺省值为false/off。
headernum	指定导入文件header的行数，即指定具体多少行作为标题行。该参数仅9.1.1.300及以上集群版本支持。	取值范围：1-10
fileheader	指定导出数据要包含的标题行定义的文件，文件一般只包含一行用来描述每一列数据信息的字符串。 例如：在包含商品信息的数据前加标题行，定义文件如下 The information of products.\n	• 标题行定义文件仅在header为on或true的情况下有效，且需要提前写好备用。 • 在Remote导出模式下，定义文件必须放在GDS的工作目录（即启动gds时指定的-d路径）下。 • 定义文件只能包含一行标题信息，并以换行符结尾，多余的行将被丢弃（标题信息不能包含换行符）。 • 定义文件包括换行符在内长度不超过1MB。

 版权所有 © 华为云计算技术有限公司

参数	描述	取值范围
out_filename_prefix	指定write only外表导出时，GDS端生成导出数据文件的文件名前缀。 file_type设置为pipe时，会生成“dbName_schemaName_foreignTableName.pipe”的管道文件。 如果out_filename_prefix和location中都指定了管道名，则以location中指定的管道文件名为准。	● 指定文件名前缀需合法，符合GDS部署物理环境使用的文件系统的约束，否则出现文件创建失败： – 指定的导出文件名前缀中不含有非法字符（/、?、*、:、|、\\、<、>、@、#、$、&、(、)、+以及-），其中非法字符包含但不限于以上列举，允许的字符范围为小写字母（a~z）、大写字母（A~Z）、数字（0~9）和下划线（_）。 – 指定的导出文件名前缀中不可以是一些Windows和linux预留的特性字段，其中包括但不限于：con、aux、nul、prn、com0、com1、com2、com3、com4、com5、com6、com7、com8、com9、lpt0、lpt1、lpt2、lpt3、lpt4、lpt5、lpt6、lpt7、lpt8或lpt9。 – 指定的导出文件名前缀，与GDS -d目录和“.dat”或者“.pipe”拼接为绝对路径后必须符合GDS所在部署文件系统的文件名长度要求。 – 指定的导出文件名前缀，需要可以被数据文件的最终接收方正确解析识别（包括但不限于GDS再次导入库中），对于造成文件名解析问题的指定选项，需要用户识别。 ● 在多文件同时导出的高并发导出场景下，请确认并发的导出任务不要使用同一个文件名前缀设定，否则从操作系统/文件系统层面可能会出现导出文件的覆盖和丢失。
delimiter	指定数据文件行数据的字段分隔符，不指定则使用默认分隔符，TEXT格式的默认分隔符是水平制表符（tab），CSV格式的默认分隔符为“,”，FIXED格式没有分隔符。	支持多字符分隔符，但分隔符不能超过10个字节。 ● 分隔符不能是\r和\n。 ● 分隔符不能和null参数相同，CSV格式数据的分隔符不能和quote参数相同。 ● TEXT格式数据的分隔符不能包含：\.abcdefghijklmnopqrstuvwxyz0123456789。 ● 数据文件中单行数据长度需<1GB，如果分隔符较长且数据列较多的情况下，会影响导出有效数据的长度。 ● 分隔符推荐使用多字符（例如'$^&'）和不可见字符（例如0x07、0x08、0x1b等）。 ● 建议TEXT格式下多字符分隔符中的字符不要完全相同，例如不建议使用delimiter '---'。

版权所有 © 华为云计算技术有限公司

参数	描述	取值范围
quote	用于设置将CSV格式数据源文件中的什么字符识别为引号字符。缺省值为双引号。	• quote参数不能和分隔符、null参数相同。 • quote参数只能是单字节的字符。 • 推荐不可见字符作为quote，例如0x07，0x08，0x1b等。 • 导出时，如果字段值中含有分隔符（delimiter）、换行符（eol）、quote字符本身时，会自动使用quote包裹，使字段内容不会被误解析为语法字符。
escape	用来指定CSV格式的数据源文件中，什么字符为逃逸字符。逃逸字符只能指定为单字节字符。 缺省值和quote相同。	-
null	用来指定数据文件中空值的表示。	• 在TEXT格式下缺省值是\N。 • CSV格式下缺省值是一个没有引号的空字符串。 • null值不能是\r和\n，最大为100个字符。 • null值不能和分隔符、quote参数相同。
blank_number_str_to_null	用来指定数据文件中数值类型（INT1OID、INT2OID、INT4OID、INT8OID、FLOAT4OID、FLOAT8OID）不定数量的空格转换为NULL值。该参数8.2.1.236及以上集群版本支持。 例如，设置该选项为true时，定义null ''，null ' '（一个空格）、null ' '（两个空格）时都返回NULL值。	true/false，默认值为false，表示不做转换。
noescaping	TEXT格式下，不对'\'和后面的字符进行转义。	noescaping参数只在TEXT格式下有效。true/on，false/off。缺省值为false/off。

版权所有 © 华为云计算技术有限公司

参数	描述	取值范围
encoding	指定数据文件的编码格式名称，即需要以何编码格式对数据文件进行解析和校验/输出文件为何种编码格式。缺省值为当前数据库的默认客户端编码格式，即client_encoding。 导入外表此处强烈建议指定为文件的编码格式，或根据文件的字符集在导入前对client_encoding进行设置。否则可能会导致不必要的解析、校验错误以及其导致的导入报错回滚，甚至非法数据入库。导出外表同样希望指定此选项，以避免导出采用默认字符集设置时与预期不符。 在创建外表时此选项未指定，会在客户端给出对应Warning信息。	● 目前GDS导入外表不支持解析带有多种字符集编码格式混合的文件，GDS导出外表不支持写出带有多种字符集编码格式混合的文件。 ● 常用的编码格式有UTF8、GBK、GB18030，其中GB18030有两个版本，GB18030和GB18030_2022。GB18030_2022，用于支持汉字最新的国家标准**GB 18030-2022**。
dataencoding	用于指定数据在latin1数据库表中的实际编码，实现GDS导入导出时数据从dataencoding与encoding之间的转换。该参数仅8.2.0及以上集群版本支持。	GBK、GB18030、GB18030_2022、UTF8。 典型场景： 1. 将GB编码的文本数据导入到latin1库中并转换为UTF8编码，设置dataencoding为 'UTF8'，encoding为 'GB'。 2. 将latin1库中GB编码的数据导出为UTF8的数据文件，设置dataencoding为 'GB'，encoding为 'UTF8'。 **注意** ● 该参数仅支持在latin1的库中使用。 ● 使用该参数时，encoding参数的取值范围也被限定在GBK、GB18030、GB18030_2022和UTF8之内。
mode	指定数据导入过程中，数据导入策略。DWS只支持Normal策略。	取值范围： ● Normal（缺省值）：支持所有文件格式(包括CSV、TEXT、FIXED)，数据导入需要在数据服务器上启动Gauss data service协助完成。

 版权所有 © 华为云计算技术有限公司

参数	描述	取值范围
eol	指定导入导出数据文件换行符样式。	取值范围：支持多字符换行符，但换行符不能超过10个字节。常见的换行符，如\r、\n、\r\n（设成0x0D、0x0A、0x0D0A效果是相同的），其他字符或字符串，如$、#。 ● eol参数一般用于TEXT格式的导入导出，不支持CSV格式和FIXED格式导入。为了兼容原有eol参数，仍然支持导出CSV格式和FIXED格式时指定eol参数为0x0D或0x0D0A。 ● 导出FIXED格式时可指定eol参数为空串''，这样会将FIXED格式数据导出为一行，但是这种方式导出的数据不支持再通过同样的方式导入到DWS数据库中。 ● eol参数不能和分隔符、null参数相同。 ● eol参数不能包含：数字，字母和符号“.”。

 版权所有 © 华为云计算技术有限公司

参数	描述	取值范围
conflict_delimiter	此参数一般配合**compatible_illegal_chars**参数一起使用，当用户的数据文件中包含半个汉字字符的时候，并且这半个字符和分隔符会由于外表的编码和数据库编码不一致被编码成一个其他的汉字，导致分隔符被掩盖从而报错缺少字段。 如果用户不希望让这半个字符和分隔符编码成一个其他字符则可以使用此参数。	取值范围：true/on，false/off。缺省值为false/off。 ● 参数为true/on，允许这半个字符和分隔符编码成一个其他字符。 ● 参数为false/off，不允许这半个字符和分隔符编码成一个其他字符。 此参数默认关闭。由于场景较少出现，不建议打开。如果未能识别场景而打开此参数则会有入表信息错乱的风险。 例如：假设有一行GBK数据"3|+|腠+|20191212"要被导入到UTF8的数据库，用户自定义的字段分隔符为"|+|"。 这行GBK数据和十六进制对应关系如下 33 7C 2B 7C C4 7C 2B 7C 323031393132313 ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ 3 | + | ** | + | 20191212 当导入UTF8数据库时，数据库会发生如下转换。 GBK -> UTF8 33->33 (3) 7C->7C (|) 2B->2B (+) 7C->7C(|) C47C-> E886A2(腠) 2B->2B(+) 7C->7C(|) 323031393132313-> 20191212 由于用户定义的字段分隔符号为"|+|"，因此对该行数据分隔后获取的数据是3，腠+|20191212, 实际上用户可能要获得的是3，？,20191212。 "腠" 这个字符其实是一个用户不期望的半个字符，和用户定义的分隔符“|+|”中的“|”产生了冲突，导致数据错乱或者导入失败。 如果用户想要忽略这种字符可以使用conflict_delimiter配合**compatible_illegal_chars**参数将“C4”这个半个GBK字符转换成“？”导入UTF8数据库。

 版权所有 © 华为云计算技术有限公司

参数	描述	取值范围
file_type	指定导入或者导出的文件类型。	取值范围：normal，pipe，interconn。缺省为normal。 ● 参数为normal，表示导入或者导出的文件类型为普通文件。 ● 参数为pipe，表示导入或者导出的文件类型为命名管道文件。 ● 参数为interconn，表示当前的GDS外表用于互联互通压缩导入导出，该参数必须配合gds_compress参数使用，仅8.2.0及以上集群版本支持。 **注意** 当file_type的参数为interconn时，gds需要升级到8.2.0及以上版本，否则会出现ERROR: un-support format报错。
auto_create_pipe	用于设置GDS进程是否自动创建命名管道文件。 设置auto_create_pipe参数时，file_type必须设置为pipe或者interconn，否则不能成功创建外表。	取值范围：true/on，false/off。缺省值为true/on。 ● 参数为true/on，表示允许GDS进程自动创建命名管道文件。 ● 参数为false/off，表示用户需手动创建命名管道文件。 ● auto_create_pipe设置为false，执行导入导出时，若未指定管道文件，会打开“数据库名_模式名_外表名.pipe”文件；若已指定管道文件，会打开location参数中指定的管道文件。该命名管道文件在pipe-timeout参数设置的时间内没有被其他程序写入或者以写的方式打开，则导入导出任务报错超时。若发现该文件不是管道文件则导入导出任务直接报错。 ● auto_create_pipe设置为true，执行导入导出时，若未指定管道文件，会打开“数据库名_模式名_外表名.pipe”文件；若该文件为普通文件类型，则导入导出任务报错。若为管道文件会自动删除该文件，并重新创建该命名管道文件。 ● 管道文件导出时**location**参数可以指定导出的管道文件，如：location‘gsfs://127.0.0.1:7789/aa.pipe'，当auto_create_pipe设置为true，GDS会自动在数据目录下创建“aa.pipe”管道文件。

版权所有 © 华为云计算技术有限公司

参数	描述	取值范围
del_pipe	用于设置不落地导入/导出业务结束后是否自动删除管道文件。 设置del_pipe时，file_type必须设置为pipe或者interconn，否则不能成功创建外表。	取值范围：true/on，false/off。缺省值为true/on。 ● 参数为true/on，表示允许GDS进程自动删除命名管道文件。 ● 参数为false/off，表示GDS进程不会删除命名管道文件。
gds_compress	该参数用于降低用户使用GDS管道文件时集群之间互联互通的网络带宽，当前仅支持snappy算法。该参数仅8.2.0及以上集群版本支持。	取值范围：snappy。 当前仅支持snappy算法，表示GDS管道导入导出的数据是以snappy算法压缩和解压缩的。 ● 设置该参数时，file_type参数必须为interconn，format必须为CSV， 默认不压缩。当前仅支持设置为snappy，使用该功能时需要保证本地集群，远端集群，GDS的版本至少在同一版本。 ● 当file_type的参数为interconn时，gds需要升级到8.2.0及以上版本，否则会出现ERROR: un-support format报错。 ● 该参数会降低集群间的网络带宽，但是会增加集群的CPU负载，用户可以根据实际情况选择是否开启该参数。 ● 为了实现GDS互联互通外表迁移数据时，数据在网络层压缩传输，GDS外表必须支持该参数，用户不仅可以使用互联互通外表进行数据压缩迁移，也可以手动使用GDS外表实现数据在网络层压缩迁移，由于手动实现时需要用户设置参数较多，推荐使用互联互通外表的方案。 ● 导入和导出端的外表必须都指定该参数。
preserve_blanks	用于定长导入时，控制每列数据拖尾的空白字符（包括空格，\t，\v，\f）是否保留。该参数仅8.2.0.100及以上集群版本支持。	取值范围：true/on，false/off。缺省值为false/off。 ● 参数为true/on，表示针对BYTEAOID、CHAROID、NAMEOID、TEXTOID、BPCHAROID、VARCHAROID、NVARCHAR2OID、CSTRINGOID对应类型的列数据，保留拖尾 \t，不进行修剪。 ● 参数为false/off，表示针对所有数据类型的数据拖尾的空白字符（包括空格，\t，\v，\f）进行修剪。

版权所有 © 华为云计算技术有限公司

参数	描述	取值范围
fix	指定每一行定长格式数据的长度。按字节计算。此语法仅对READ ONLY的外表有效。	取值范围：N >= POSITION指定的总长度（总长度即为表定义最后一个字段的offset与length的和）AND N < 1GB
out_fix_alignment	定长导出中，指定BYTEAOID、CHAROID、NAMEOID、TEXTOID、BPCHAROID、VARCHAROID、NVARCHAR2OID、CSTRINGOID对应类型所在列的对齐方式。	取值范围：align_left、align_right 默认值：align_right **注意** 由于bytea数据类型要求十六进制格式（如"\XXXX"）或八进制格式（如"\XXX\XXX\XXX"），导入时需要左对齐（即列数据以两种格式开头，而非空格）。因此若导出文件需要重新以GDS外表入库且数据长度小于外表formatter指定长度，导出时需要指定左对齐，否则会在入库的过程中报错。
out_fix_num_alignment	定长导出中，指定数值类型TINYINT、SMALLINT、INTEGER、BINARY_INTEGER、BIGINT、NUMERIC[(p[,s])]、DECIMAL[(p[,s])]、NUMBER[(p[,s])]、REAL、FLOAT4、DOUBLE PRECISION、FLOAT8、FLOAT[(p)]、BINARY_DOUBLE、DEC[(p[,s])]、INTEGER[(p[,s])]、SMALLSERIAL、SERIAL、BIGSERIAL对应类型所在列的对齐方式。该选项仅9.1.0及以上集群版本支持。	取值范围：align_left、align_right 默认值：align_left
date_format	导入对于DATE类型指定格式。此语法仅对READ ONLY的外表有效。	取值范围：合法DATE格式。可参考**时间日期函数和操作符**。 **注意** 对于指定为ORACLE兼容类型的数据库，则DATE类型内建为TIMESTAMP类型。在导入的时候，若需指定格式，可以参考下面的timestamp_format参数。

版权所有 © 华为云计算技术有限公司

参数	描述	取值范围
time_format	导入对于TIME类型指定格式。此语法仅对READ ONLY的外表有效。	取值范围：合法TIME格式，不支持时区。可参考**时间日期函数和操作符**。
timestamp_format	导入对于TIMESTAMP类型指定格式。此语法仅对READ ONLY的外表有效。	取值范围：合法TIMESTAMP格式，不支持时区。可参考**时间日期函数和操作符**。
smalldatetime_format	导入对于SMALLDATETIME类型指定格式。此语法仅对READ ONLY的外表有效。	取值范围：合法SMALLDATETIME格式。可参考**时间日期函数和操作符**。

表 13-93 OPTIONS 容错性参数说明

参数	描述	取值范围
fill_missing_fields	当数据导入时，若数据源文件中某一行的末尾字段（一个或多个）缺失的处理方式。	true/on，false/off。缺省值为false/off。 • true/on，表示末尾字段缺失，则将末尾字段的值设置为NULL，不报错。 • false/off，表示末尾字段缺失会直接报错，报错信息类似如下： `missing data for column "tt"`
ignore_extra_data	若数据源文件比外表定义列数多，是否会忽略对多出的列。该参数只在数据导入过程中使用。	true/on，false/off。缺省值为false/off。 • 参数为true/on，若数据源文件比外表定义列数多，则忽略行尾多出来的列。 • 参数为false/off，若数据源文件比外表定义列数多，会显示如下错误信息。 `extra data after last expected column` **注意** 如果行尾换行符丢失，使两行变成一行时，设置此参数为true将导致后一行数据被忽略掉。

版权所有 © 华为云计算技术有限公司

参数	描述	取值范围
reject_limit	指定本次数据导入允许出现的数据格式错误个数，当导入过程中出现的数据格式错误未达到限定值时，本次数据导入可以成功。 **说明** 此语法建议用PER NODE REJECT LIMIT 'value'替代。 数据格式错误是指缺少或者多出字段值，数据类型错误或者编码错误。对于非数据格式错误，一旦发生就将导致整个数据导入失败。	正整型值、unlimited（无限制）。指定正整型参数时需要添加单引号。 不指定该参数时，有错误信息立即返回。
compatible_illegal_chars	导入导出非法字符容错参数。此语法对READ ONLY和WRITE ONLY的外表有效。 导出容错仅8.1.3.331及以上版本集群支持。	true/on，false/off。缺省值为false/off。 ● true/on，表示导入导出时遇到非法字符进行容错处理，非法字符转换后入库或出库，不报错，不中断导入导出。 ● false/off，表示导入导出时遇到非法字符进行报错，中断导入导出。 导入非法字符容错规则如下： ● 对于'\0'，容错后转换为空格。 ● 对于其他非法字符，容错后转换为问号。 ● 如果compatible_illegal_chars为true/on标识导入时对于非法字符进行容错处理，如果NULL、DELIMITER、QUOTE、ESCAPE设置为空格或问号则会通过如"illegal chars conversion may confuse COPY escape 0x20"等报错信息提示用户修改可能引起混淆的参数，以避免导入错误。 ● 需注意：导出外表容错参数开启后，非法字符数据导出被容错为问号，再次导入回DWS库中就会与原数据不一致。
replace_illegal_chars	导入非法字符容错参数，用于将'\0'替换为指定字符串。此语法仅对READ ONLY的外表有效。	指定用于替换'\0'的字符串有如下约束： ● 仅支持compatible_illegal_chars参数为true/on时设置。 ● 仅支持format参数为TEXT或CSV时设置。 ● 字符串长度不能超过10。 ● 字符串不能与delimiter相同或者包含delimiter，不能包含以下字符：数字，字母以及特殊字符“\”和“.”。

版权所有 © 华为云计算技术有限公司

参数	描述	取值范围
WITH error_table_name	数据导入过程中出现的数据格式错误信息将被写入error_table_name指定的错误信息表中，可以在并行导入结束后查询此错误信息表，获取详细的错误信息。此参数只在设置了reject_limit参数时有效。 **说明** 如果为了兼容postgres开源接口，此语法建议用LOG INTO代替。	字符串，要符合标识符的命名规范。
LOG INTO error_table_name	数据导入过程中出现的数据格式错误信息将被写入error_table_name指定的错误信息表中，可以在并行导入结束后查询此错误信息表，获取详细的错误信息。 **说明** 若没有指定PER NODE REJECT LIMIT参数，则此参数不起作用。	字符串，要符合标识符的命名规范。
REMOTE LOG 'name'	数据导入过程中出现的数据格式错误信息将被写到GDS端以文件方式保存。name为错误数据文件的文件名前缀。	-

版权所有 © 华为云计算技术有限公司

参数	描述	取值范围
PER NODE REJECT LIMIT 'value'	指定本次数据导入过程中每个DN实例上允许出现的数据格式错误的数量，如果有一个DN实例上的错误数量大于设定值，本次导入失败，报错退出。 **注意** 此语法指定的是单个节点的错误容忍度。 数据格式错误是指缺少或者多出字段值，数据类型错误或者编码错误。对于非数据格式错误，一旦发生就将导致整个数据扫描失败。	整型值，unlimited（无限），不指定该参数时，有错误信息立即返回。

 版权所有 © 华为云计算技术有限公司

表 13-94 OPTIONS 性能参数说明

参数	描述	取值范围
file_sequence	用于多任务GDS外表并行导入，提升单个文件的导入性能。该参数仅供数据导入使用。 格式为file_sequence '文件被拆分的总数-当前分片' 。例如： file_sequence '3-1' 表示导入的文件在逻辑上被拆分成3份，当前外表导入的数据为第一个分片上的数据。 file_sequence '3-2' 表示导入的文件在逻辑上被拆分成3份，当前外表导入的数据为第二个分片上的数据。 file_sequence '3-3' 表示导入的文件在逻辑上被拆分成3份，当前外表导入的数据为第三个分片上的数据。	使用该参数需遵循以下约束： • 文件被拆分的总数小于等于8。 • 当前分片小于等于文件被拆分的总数。 • 导入的文件仅支持CSV和text格式。 使用CSV格式进行并行导入时，在某些场景中会因为CSV本身的规则和GDS拆分逻辑冲突而导致其中的某些分片导入失败，详情参见**使用file_sequence参数异常场景说明**。

示例

创建外表customer_ft，用来以TEXT格式导入GDS服务器10.10.123.234上的数据：

```
DROP FOREIGN TABLE IF EXISTS customer_ft;
CREATE FOREIGN TABLE customer_ft
(
    c_customer_sk             integer               ,
    c_customer_id             char(16)              ,
    c_current_cdemo_sk        integer               ,
    c_current_hdemo_sk        integer               ,
    c_current_addr_sk         integer               ,
    c_first_shipto_date_sk    integer               ,
    c_first_sales_date_sk     integer               ,
    c_salutation              char(10)              ,
    c_first_name              char(20)              ,
    c_last_name               char(30)              ,
    c_preferred_cust_flag     char(1)               ,
    c_birth_day               integer               ,
    c_birth_month             integer               ,
    c_birth_year              integer               ,
    c_birth_country           varchar(20)           ,
    c_login                   char(13)              ,
    c_email_address           char(50)              ,
    c_last_review_date        char(10)
```


版权所有 © 华为云计算技术有限公司

```
)
   SERVER gsmpp_server
   OPTIONS
(
   location 'gsfs://10.10.123.234:5000/customer1*.dat',
   FORMAT 'TEXT' ,
   DELIMITER '|',
   encoding 'utf8',
   mode 'Normal')
READ ONLY;
```

创建外表foreign_HR_staffs_ft，用来以TEXT格式导入GDS服务器192.168.0.90和192.168.0.91上的数据，导入过程错误信息将记录到err_HR_staffs中。本次数据导入允许出现的数据格式错误个数为2。

```
DROP FOREIGN TABLE IF EXISTS foreign_HR_staffs_ft;
CREATE FOREIGN TABLE foreign_HR_staffs_ft
(
  staff_ID       NUMBER(6) ,
  FIRST_NAME     VARCHAR2(20),
  LAST_NAME      VARCHAR2(25),
  EMAIL          VARCHAR2(25),
  PHONE_NUMBER   VARCHAR2(20),
  HIRE_DATE      DATE,
  employment_ID  VARCHAR2(10),
  SALARY         NUMBER(8,2),
  COMMISSION_PCT NUMBER(2,2),
  MANAGER_ID     NUMBER(6),
  section_ID  NUMBER(4)
) SERVER gsmpp_server OPTIONS (location 'gsfs://192.168.0.90:5000/* | gsfs://192.168.0.91:5000/*', format
'TEXT', delimiter E'\x08',  null '',reject_limit '2') WITH err_HR_staffs_ft;
```

建立外表，用来以CSV格式导入input_data目录下存放在各个节点名文件下的所有文件。

```
DROP FOREIGN TABLE IF EXISTS foreign_HR_staffs_ft1;
CREATE FOREIGN TABLE foreign_HR_staffs_ft1
(
  staff_ID       NUMBER(6) ,
  FIRST_NAME     VARCHAR2(20),
  LAST_NAME      VARCHAR2(25),
  EMAIL          VARCHAR2(25),
  PHONE_NUMBER   VARCHAR2(20),
  HIRE_DATE      DATE,
  employment_ID  VARCHAR2(10),
  SALARY         NUMBER(8,2),
  COMMISSION_PCT NUMBER(2,2),
  MANAGER_ID     NUMBER(6),
  section_ID     NUMBER(4)
) SERVER gsmpp_server OPTIONS (location 'file:///input_data/*', format 'csv', quote E'\x08', mode 'private',
delimiter ',') WITH err_HR_staffs_ft1;
```

建立外表，用来以CSV格式导出数据到output_data目录下。

```
DROP FOREIGN TABLE IF EXISTS foreign_HR_staffs_ft2;
CREATE FOREIGN TABLE foreign_HR_staffs_ft2
(
  staff_ID       NUMBER(6) ,
  FIRST_NAME     VARCHAR2(20),
  LAST_NAME      VARCHAR2(25),
  EMAIL          VARCHAR2(25),
  PHONE_NUMBER   VARCHAR2(20),
  HIRE_DATE      DATE,
  employment_ID  VARCHAR2(10),
  SALARY         NUMBER(8,2),
  COMMISSION_PCT NUMBER(2,2),
  MANAGER_ID     NUMBER(6),
  section_ID  NUMBER(4)
) SERVER gsmpp_server OPTIONS (location 'file:///output_data/', format 'csv', quote E'\x08', delimiter '|',
header 'on') WRITE ONLY;
```

 版权所有 © 华为云计算技术有限公司

使用 file_sequence 参数异常场景说明

使用CSV格式进行并行导入时，在如下示例场景中会因为CSV本身的规则和GDS拆分逻辑冲突而导致其中的某些分片导入失败。

场景：csv文件中包含未转义的换行符，且该换行符被quote指定的字符所包含，并且该行数据处于逻辑分片的第一行。

示例：并行导入一个文件big.csv ，正确导入显示内容如下：

```
--id, username, address
10001,"customer1 name","Rose District"
10002,"customer2 name","
23 Road Rose
District NewCity"
10003,"customer3 name","NewCity"
```

文件被拆分成两份后，第一个分片显示内容如下：

```
10001,"customer1 name","Rose District"
10002,"customer2 name","
23
```

第二个分片显示内容如下：

```
Road Rose
District NewCity"
10003,"customer3 name","NewCity"
```

因为第二个分片第一行后面的换行符包含在一个双引号之间，导致GDS无法分辨该换行符是字段中的换行符还是行中的分隔符，因此第一个分片会成功导入两条数据，第二个分片导入失败。

相关链接

ALTER FOREIGN TABLE (GDS导入导出)，**DROP FOREIGN TABLE**

安装和启动GDS

最佳实践：**使用GDS从远端服务器上导入表数据到DWS集群**

常见问题：**GDS导入/导出类问题**

13.4.11 ALTER FOREIGN TABLE (For HDFS or OBS)

功能描述

对HDFS外表和OBS外表进行修改。

外表（Foreign Table）是数据库或大数据平台中一种虚拟表结构，它不实际存储数据，而是通过元数据信息（如表结构、存储位置、文件格式等）将外部存储系统（如HDFS、OBS等）中的数据映射到数据库内部，使用户能够以标准SQL操作直接访问或操作外部数据。外表的本质是一种“数据代理”，将外部文件系统的数据逻辑映射成数据库可识别的表结构。

外表的核心价值是 “打破数据孤岛”： 通过元数据映射，让数据库用户以熟悉的方式访问外部系统数据，同时支持数据向外部系统的导出，实现跨系统数据的高效协同、分析与共享，尤其在大数据场景中能显著降低数据移动成本、提升协作效率。

版权所有 © 华为云计算技术有限公司

注意事项

无。

语法格式

- 设置外表属性。
```
ALTER FOREIGN TABLE [ IF EXISTS ]  table_name
    OPTIONS ( {[ ADD | SET | DROP ] option ['value']} [, ... ]);
```
- 设置外表的所有者。
```
ALTER FOREIGN TABLE [ IF EXISTS ] tablename
    OWNER TO new_owner;
```
- 更新外表列。
```
ALTER FOREIGN TABLE [ IF EXISTS ] table_name
    MODIFY ( { column_name data_type | column_name [ CONSTRAINT constraint_name ] NOT NULL
[ ENABLE ] | column_name [ CONSTRAINT constraint_name ] NULL } [, ...] );
```
- 修改外表的列。
```
ALTER FOREIGN TABLE [ IF EXISTS ] tablename
    action [, ... ];
```
其中action语法为：
```
ALTER [ COLUMN ] column_name [ SET DATA ] TYPE data_type
  | ALTER [ COLUMN ] column_name { SET | DROP } NOT NULL
  | ALTER [ COLUMN ] column_name SET STATISTICS [PERCENT] integer
  | ALTER [ COLUMN ] column_name OPTIONS ( {[ ADD | SET | DROP ] option ['value'] } [, ... ])
  | MODIFY column_name data_type
  | MODIFY column_name [ CONSTRAINT constraint_name ] NOT NULL [ ENABLE ]
  | MODIFY column_name [ CONSTRAINT constraint_name ] NULL
```
参考**ALTER TABLE**。
- 增加外表信息约束（Informational Constraint）。
```
ALTER FOREIGN TABLE [ IF EXISTS ] tablename
    ADD [ CONSTRAINT constraint_name ]
    { PRIMARY KEY | UNIQUE } ( column_name )
    [ NOT ENFORCED [ ENABLE QUERY OPTIMIZATION | DISABLE QUERY OPTIMIZATION ] |
ENFORCED ];
```
对于增加外表信息约束（Informational Constraint）相关参数请参考**信息约束（Informational Constraint）**。
- 删除外表的信息约束（Informational Constraint）。
```
ALTER FOREIGN TABLE [ IF EXISTS ] tablename
    DROP CONSTRAINT constraint_name ;
```
- 设置外表行级访问控制开关（9.1.0及以上集群版本支持）。
```
ALTER FOREIGN TABLE [ IF EXISTS ] tablename
      ENABLE | DISABLE | FORCE | NO FORCE ROW LEVEL SECURITY
```

参数说明

表 13-95 ALTER FOREIGN TABLE (For HDFS or OBS)参数说明

参数	描述	取值范围
IF EXISTS	如果指定的外表不存在时返回一个提示而不是错误。	-
tablename	需要修改的外表名称。	已存在的外表名。
new_owner	外表的新所有者。	字符串，有效的用户名。

版权所有 © 华为云计算技术有限公司

参数	描述	取值范围
data_type	已存在字段的新数据类型。	字符串。
constraint_name	要添加/删除的约束的名称。	字符串。
column_name	已存在字段的名称。	字符串。

修改外表语法中其他参数，请参见ALTER TABLE的**参数说明**。

示例

1. 创建HDFS_Server，对应的foreign data wrapper为HDFS_FDW或者DFS_FDW。
```
CREATE SERVER hdfs_server FOREIGN DATA WRAPPER HDFS_FDW OPTIONS (address
'10.10.0.100:25000,10.10.0.101:25000',hdfscfgpath '/opt/hadoop_client/HDFS/hadoop/etc/
hadoop',type'HDFS');
```
2. 创建不包含分区列的HDFS外表：
```
DROP FOREIGN TABLE IF EXISTS ft_region;
CREATE FOREIGN TABLE ft_region
(
    R_REGIONKEY INT4,
    R_NAME TEXT,
    R_COMMENT TEXT
)
SERVER
    hdfs_server
OPTIONS
(
    FORMAT 'orc',
    encoding 'utf8',
    FOLDERNAME '/user/hive/warehouse/gauss.db/region_orc11_64stripe/'
)
DISTRIBUTE BY
     roundrobin;
```
3. 修改外表ft_region的r_name字段类型为text。
```
ALTER FOREIGN TABLE ft_region ALTER r_name TYPE TEXT;
```
4. 标记外表ft_region的r_name列为not null。
```
ALTER FOREIGN TABLE ft_region ALTER r_name SET NOT NULL;
```

相关链接

CREATE FOREIGN TABLE (SQL on OBS or Hadoop)，**DROP FOREIGN TABLE**

13.4.12 ALTER FOREIGN TABLE (GDS 导入导出)

功能描述

对GDS外表进行修改。

注意事项

无。

版权所有 © 华为云计算技术有限公司

语法格式

- 设置外表属性。
```
ALTER FOREIGN TABLE [ IF EXISTS ]  table_name
   OPTIONS ( {[ ADD | SET | DROP ] option ['value']}[, ... ]);
```
- 设置外表新的所有者。
```
ALTER FOREIGN TABLE [ IF EXISTS ] tablename
   OWNER TO new_owner;
```
- 设置外表行级访问控制开关（9.1.0及以上集群版本支持）。
```
ALTER FOREIGN TABLE [ IF EXISTS ] tablename
      ENABLE | DISABLE | FORCE | NO FORCE ROW LEVEL SECURITY
```

参数说明

表 13-96 ALTER FOREIGN TABLE (GDS 导入导出)参数说明

参数	描述	取值范围
IF EXISTS	如果指定的索引不存在时返回一个提示而不是错误。	-
table_name	需要修改的外表名称。	已存在的外表名。
option	需要修改的option名称。	请参见CREATE FOREIGN TABLE的**表13-92**。
value	option的新值。	-
new_owner	外表的新所有者。	字符串，有效的用户名。

示例

创建外表customer_ft，用来以TEXT格式导入GDS服务器10.10.123.234上的数据：

```
DROP FOREIGN TABLE IF EXISTS customer_ft;
CREATE FOREIGN TABLE customer_ft
(
   c_customer_sk             integer               ,
   c_customer_id             char(16)              ,
   c_current_cdemo_sk        integer               ,
   c_current_hdemo_sk        integer               ,
   c_current_addr_sk         integer               ,
   c_first_shipto_date_sk    integer               ,
   c_first_sales_date_sk     integer               ,
   c_salutation              char(10)              ,
   c_first_name              char(20)              ,
   c_last_name               char(30)              ,
   c_preferred_cust_flag     char(1)               ,
   c_birth_day               integer               ,
   c_birth_month             integer               ,
   c_birth_year              integer               ,
   c_birth_country           varchar(20)           ,
   c_login                   char(13)              ,
   c_email_address           char(50)              ,
   c_last_review_date        char(10)
)
   SERVER gsmpp_server
   OPTIONS
(
   location 'gsfs://10.10.123.234:5000/customer1*.dat',
```

 版权所有 © 华为云计算技术有限公司

```
  FORMAT 'TEXT' ,
  DELIMITER '|',
  encoding 'utf8',
  mode 'Normal')READ ONLY;
```

修改外表customer_ft属性，删除mode选项：

```
ALTER FOREIGN TABLE customer_ft options(drop mode);
```

相关链接

CREATE FOREIGN TABLE (GDS导入导出)，DROP FOREIGN TABLE

13.4.13 DROP FOREIGN TABLE

功能描述

删除指定的外表。

注意事项

DROP FOREIGN TABLE会强制删除指定的外表，删除外表后，依赖该外表的索引会被删除，而使用到该外表的函数和存储过程将无法执行。

语法格式

```
DROP FOREIGN TABLE [ IF EXISTS ]
  table_name [, ...] [ CASCADE | RESTRICT ];
```

参数说明

表 13-97 DROP FOREIGN TABLE 参数说明

参数	描述	取值范围
IF EXISTS	如果指定的外表不存在时返回一个提示而不是错误。	-
table_name	要删除的外表名称。	已存在的表名。
CASCADE | RESTRICT	当执行删除操作时，如何处理依赖对象中的相关数据。	• CASCADE：级联删除依赖于外表的对象（比如视图）。 • RESTRICT：如果存在依赖对象，则拒绝删除该外表（缺省行为）。

示例

删除名称为customer_ft的外表：

```
DROP FOREIGN TABLE IF EXISTS customer_ft;
```


版权所有 © 华为云计算技术有限公司

相关链接

ALTER FOREIGN TABLE (GDS导入导出)，**ALTER FOREIGN TABLE (For HDFS or OBS)**，**CREATE FOREIGN TABLE (GDS导入导出)**，**CREATE FOREIGN TABLE (SQL on OBS or Hadoop)**

13.4.14 CREATE SERVER

功能描述

创建一个外部服务器。

外部服务器是存储HDFS集群信息、OBS服务器信息、DLI连接信息或其他同构集群信息的载体。

注意事项

默认只有系统管理员才可以创建外部服务器，否则需要对所使用的FOREIGN DATA WRAPPER授权，授权语法为：

```
GRANT USAGE ON FOREIGN DATA WRAPPER fdw_name TO username;
```

其中fdw_name为FOREIGN DATA WRAPPER的名字，username为创建SERVER的用户名。

语法格式

```
CREATE SERVER server_name
    FOREIGN DATA WRAPPER fdw_name
    OPTIONS ( { option_name ' value ' } [, ...] ) ;
```

参数说明

表 13-98 CREATE SERVER 参数说明

参数	描述	取值范围
server_name	要创建的外部服务器的名称。服务器名称在数据库中必须唯一。	取值范围：长度必须小于等于63。
FOREIGN DATA WRAPPER fdw_name	指定外部数据封装器的名字。	取值范围：fdw_name是数据库初始化时系统创建的数据封装器，目前对于HDFS集群，fdw_name的名字可以是hdfs_fdw或者dfs_fdw；对于其他同构集群，fdw_name为gc_fdw；对于数据导入导出场景，GDS外表使用gsmpp_server，fdw_name为dist_fdw。
OPTIONS	用于指定外部服务器的各类参数。	支持的各类参数参见**表13-99**。

版权所有 © 华为云计算技术有限公司

表 13-99 OPTIONS 参数说明

参数	描述	取值范围
address	指定的OBS服务终端节点或HDFS集群的IP地址。	OBS：OBS服务的终端节点（Endpoint）。 HDFS：HDFS集群的元数据节点（NameNode）所在的IP地址以及端口，或者同构其他集群的CN的IP地址以及端口。 为保证HA（High Availability），HDFS NameNode经常采用主备模式。主备NameNode的地址都需要加入到address值中。DWS访问HDFS服务时，会动态查找当前处于active状态的主NameNode。 若HDFS为联邦模式时，可将Router的地址都加入到address值中，DWS访问HDFS服务时，会随机选择router进行访问。 **说明** ● address option必须存在，若用于跨集群互联互通场景则只允许设置1个。 ● 当server类型为DLI时，address为DLI服务上数据所存储的OBS address。 ● 若HDFS为联邦模式时，即fed 'rbf'，address可设置为多组IP、port，对应为HDFS Router的address。
hdfscfgpath	该参数仅支持type为HDFS时设置。 用户通过配置hdfscfgpath参数来指定HDFS配置文件路径。DWS会根据配置文件路径下的HDFS配置文件指定的连接配置方式，以及安全模式，来访问HDFS集群。非安全模式连接HDFS集群时，不支持数据传输加密。 如果没有指定address选项，默认采用hdfscfgpath指定的配置文件中指定的address。	-

 版权所有 © 华为云计算技术有限公司

参数	描述	取值范围
fed	表示dfs_fdw连接的是HDFS为联邦模式。	取值rbf，表示HDFS为联邦rbf方式。 **说明** 该参数8.1.2及以上版本支持。
encrypt	是否对数据进行加密，该参数仅支持type为OBS时设置。默认值为off。	取值范围： ● on表示对数据进行加密。 ● off表示不对数据进行加密。
access_key	OBS访问协议对应的AK值（OBS云服务界面由用户获取），创建外表时AK值会保存到数据库的元数据表中。	● FOREIGN DATA WRAPPER为dfs_fdw时，仅支持type为OBS时设置。 ● 8.2.0及以上版本集群，支持FOREIGN DATA WRAPPER为dist_fdw时指定该参数。
secret_access_key	OBS访问协议对应的SK值（OBS云服务界面由用户获取），创建外表时SK值会加密保存到数据库的元数据表中。	● FOREIGN DATA WRAPPER为dfs_fdw时，仅支持type为OBS时设置。 ● 8.2.0及以上集群版本，支持FOREIGN DATA WRAPPER为dist_fdw时指定该参数。
security_token	对应统一身份认证服务IAM中临时安全凭证的SecurityToken的值，与临时AK、SK共同组成临时安全凭证，临时安全凭证有效期不超过24小时。该参数仅8.2.0及以上集群版本支持。	● FOREIGN DATA WRAPPER为dfs_fdw时，仅支持type为OBS时设置。 ● 8.2.0及以上集群版本，支持FOREIGN DATA WRAPPER为dist_fdw时指定该参数。 ● 使用该参数时，access_key和secret_access_key参数分别对应临时AK、SK。
type	表示dfs_fdw连接的类型。	取值范围： ● OBS表示连接的是OBS。 ● S3表示连接的是支持S3标准接口的对象存储。该取值仅9.1.1.300及以上集群版本支持。 ● HDFS表示连接的是HDFS。 ● DLI表示连接的是DLI。
dli_address	DLI服务的终端节点，即endpoint。该参数仅支持type为DLI时设置。	-

版权所有 © 华为云计算技术有限公司

参数	描述	取值范围
dli_access_key	DLI访问协议对应的AK值（DLI云服务界面由用户获取），创建外表时AK值会保存到数据库的元数据表中。该参数仅支持type为DLI时设置。	-
dli_secret_access_key	DLI访问协议对应的SK值（DLI云服务界面由用户获取），创建外表时SK值会加密保存到数据库的元数据表中。该参数仅支持type为DLI时设置。	-
dbname	用于协同分析、跨集群互联互通，表示将要连接的远端集群的数据库名字。	-
username	用于协同分析、跨集群互联互通，表示将要连接的远端集群的用户名。	-
password	用于协同分析、跨集群互联互通，表示将要连接的远端集群的用户密码。 **说明** 对于云下集群迁移到云上的场景，从云下集群导出的server配置中密码为密文，由于云上和云下集群加解密的密钥不同，如果直接在云上集群执行导出时的CREATE SERVER，会执行失败，报解密失败的错误。这种场景下需要将CREATE SERVER中的password手动修改成明文密码进行配置。	-
syncsrv	仅用于跨集群互联互通，表示数据同步过程中使用到的GDS服务，设置方式与GDS外表的location属性相同。	-

示例

建立一个hdfs_server，其中hdfs_fdw为数据库中存在的foreign data wrapper。（address、hdfscfgpath的值仅供参考，请根据实际配置情况进行修改。）

```
DROP SERVER IF EXISTS hdfs_server;
CREATE SERVER hdfs_server FOREIGN DATA WRAPPER HDFS_FDW OPTIONS
  (address '10.10.0.100:25000,10.10.0.101:25000',
   hdfscfgpath '/opt/hadoop_client/HDFS/hadoop/etc/hadoop',
```

 版权所有 © 华为云计算技术有限公司

```
    type 'HDFS'
) ;
```

建立一个obs_server，其中dfs_fdw为数据库中存在的foreign data wrapper：

```
DROP SERVER IF EXISTS obs_server;
CREATE SERVER obs_server FOREIGN DATA WRAPPER DFS_FDW OPTIONS (
  address 'obs.ap-southeast-1.myhuaweicloud.com',
   access_key 'xxxxxxxxx',
  secret_access_key 'yyyyyyyyyyyyy',
  type 'obs'
);
```

建立一个s3_server，其中dfs_fdw为数据库中存在的foreign data wrapper：

```
DROP SERVER IF EXISTS s3_server;
CREATE SERVER s3_server FOREIGN DATA WRAPPER DFS_FDW OPTIONS (
  address '10.10.0.100:25000,10.10.0.101:25000',
  access_key 'xxxxxxxxx',
  secret_access_key 'yyyyyyyyyyyyy',
  type 'S3'
);
```

建立一个dli_server，其中dfs_fdw为数据库中存在的foreign data wrapper：

```
DROP SERVER IF EXISTS dli_server;
CREATE SERVER dli_server FOREIGN DATA WRAPPER DFS_FDW OPTIONS (
  address 'obs.ap-southeast-1.myhuaweicloud.com',
  access_key 'xxxxxxxxx',
  secret_access_key 'yyyyyyyyyyyyy',
  type 'dli',
  dli_address 'dli.example.com',
  dli_access_key 'xxxxxxxxx',
  dli_secret_access_key 'yyyyyyyyyyyyy'
);
```

建立另外一个同构集群的server，其中gc_fdw为数据库中存在的foreign data wrapper：

```
DROP SERVER IF EXISTS server_remote;
CREATE SERVER server_remote FOREIGN DATA WRAPPER GC_FDW OPTIONS
   (address '10.10.0.100:25000,10.10.0.101:25000',
  dbname 'test',
  username 'test',
  password 'xxxxxxxx'
);
```

建立一个FOREIGN DATA WRAPPER为dist_fdw的server，用于obs上文本数据的导入导出：

```
DROP SERVER IF EXISTS import_server;
CREATE SERVER import_server FOREIGN DATA WRAPPER DIST_FDW OPTIONS
(
  access_key 'ak_string',
  secret_access_key 'sk_string'
);
```

相关链接

ALTER SERVER，**DROP SERVER**

13.4.15 ALTER SERVER

功能描述

增加、修改和删除外部服务器的定义。

 版权所有 © 华为云计算技术有限公司

现有外部服务器（server）可以使用pg_foreign_server系统表查询。

注意事项

- 只有服务器的所有者有权限执行ALTER SERVER命令，系统管理员默认拥有此权限。
- 要更改服务器的所有者，必须是外部服务器的OWNER并且也是新的所有者角色的直接或者间接成员，且必须对外部服务器的外部数据封装器有USAGE权限。

语法格式

- 修改外部服务的参数。
```
ALTER SERVER server_name [ VERSION 'new_version' ]
    [ OPTIONS ( {[ ADD | SET | DROP ] option ['value']} [, ... ] ) ];
```
- 修改外部服务的所有者。
```
ALTER SERVER server_name
    OWNER TO new_owner;
```
- 修改外部服务的名称。
```
ALTER SERVER server_name
    RENAME TO new_name;
```
- 刷新HDFS配置文件。
```
ALTER SERVER server_name REFRESH OPTIONS;
```

参数说明

表 13-100 ALTER SERVER 参数说明

参数	描述	取值范围
server_name	需要修改的外部服务的名称。	已存在的外部服务器名。
new_version	新的外部服务器版本。	-
OPTIONS	更改服务器的选项。ADD、SET和DROP指定要执行的操作。OPTIONS选项未指定时默认为ADD操作。 对于HDFS Server目前只支持SET操作，ADD/DROP操作现有版本不支持。语法中SET和DROP操作语法依然保留，以便后续扩展使用。	option和value为对应操作的参数可参考**表13-101**。
new_owner	外部服务器的新所有者。	有效的用户名。
new_name	外部服务器的新名称。	字符串，需符合**标识符命名规范**。
REFRESH OPTIONS	刷新HDFS的配置文件缓存信息，在配置文件有变动时执行，若不执行可能会访问报错。 该参数用于更新hdfs server配置缓存，不能与其相关的HDFS外表访问并发。	-

版权所有 © 华为云计算技术有限公司

表 13-101 修改外部服务器所支持的 **OPTIONS** 参数说明

参数	描述	取值范围
address	OBS服务的终端节点（Endpoint）。 HDFS集群的主备节点所在的IP地址以及端口。 **说明** ● 对于HDFS server，address必须存在，所以ADD和DROP操作不被允许。 ● address目前只支持点分十进制格式的IPv4格式，且address字符串中不能出现空格，多组address以逗号作为分隔符。ip和port之间使用“:”来区分。HDFS集群中ip、port组对推荐设置两组，分别对应HDFS NameNode主备节点的address。 ● 当server类型为DLI时，address为DLI服务上数据所存储的OBS address。	-
hdfscfgpath	HDFS集群的配置文件。 **说明** ● 若HDFS为安全模式时，hdfscfgpath是必选项，否则为可选项。 ● 若设置hdfscfgpath时，path仅能设置一个。	-
fed	表示dfs_fdw连接的是HDFS为联邦模式。 取值rbf，表示HDFS为联邦rbf方式。 **说明** 该参数8.1.2及以上版本支持；8.0.0基线版本下，仅8.0.0.10及以上版本支持。	-
encrypt	是否对数据进行加密，该参数仅支持在type为OBS时设置。默认值为off。	● on表示对数据进行加密。 ● off表示不对数据进行加密。
access_key	OBS访问协议对应的AK值（OBS云服务界面由用户获取），创建外表时AK值会加密保存到数据库的元数据表中。该参数仅支持type为OBS时设置。	-
secret_access_key	OBS访问协议对应的SK值（OBS云服务界面由用户获取），创建外表时SK值会加密保存到数据库的元数据表中。该参数仅支持type为OBS时设置。	-
dli_address	DLI服务的终端节点，即endpoint。该参数仅支持type为DLI时设置。	-
dli_access_key	DLI访问协议对应的AK值（DLI云服务界面由用户获取），创建外表时AK值会保存到数据库的元数据表中。该参数仅支持type为DLI时设置。	-

版权所有 © 华为云计算技术有限公司

参数	描述	取值范围
dli_secret_access_key	DLI访问协议对应的SK值（DLI云服务界面由用户获取），创建外表时SK值会加密保存到数据库的元数据表中。该参数仅支持type为DLI时设置。	-
region	此参数表示OBS服务的IP地址或者域名信息。该参数仅支持type为OBS时设置。	-
dbname	用于协同分析、跨集群互联互通，表示将要连接的远端集群的数据库名称。	-
username	用于协同分析、跨集群互联互通，表示将要连接的远端集群的用户名。	-
password	用于协同分析、跨集群互联互通，表示将要连接的远端集群的用户密码。	-
syncsrv	仅用于跨集群互联互通，表示数据同步过程中使用到的GDS服务，设置方式与GDS外表的location属性相同。该参数仅8.1.2及以上版本支持。	-

示例

创建外部服务器hdfs_server，其中hdfs_fdw为数据库中存在的foreign data wrapper。

```
DROP SERVER IF EXISTS hdfs_server;
CREATE SERVER hdfs_server FOREIGN DATA WRAPPER HDFS_FDW OPTIONS
   (address '10.10.0.100:25000,10.10.0.101:25000',
    hdfscfgpath '/opt/hadoop_client/HDFS/hadoop/etc/hadoop',
    type 'HDFS'
) ;
```

修改外部服务器hdfs_server的address选项。

```
ALTER SERVER hdfs_server OPTIONS ( SET address '10.10.0.110:25000,10.10.0.120:25000');
```

修改外部服务器hdfs_server的hdfscfgpath选项。

```
ALTER SERVER hdfs_server OPTIONS ( SET hdfscfgpath '/opt/bigdata/hadoop');
```

相关链接

CREATE SERVER，**DROP SERVER**

13.4.16 DROP SERVER

功能描述

删除现有的一个数据服务器。

 版权所有 © 华为云计算技术有限公司

注意事项

只有server的owner才可以删除。

语法格式

```
DROP SERVER [ IF EXISTS ] server_name [ {CASCADE | RESTRICT} ] ;
```

参数描述

表 13-102 DROP SERVER 参数说明

参数	描述	取值范围
IF EXISTS	如果指定的服务器不存在，则发出一个notice而不是抛出一个错误。	-
server_name	要删除的服务器名称。	已存在的服务器名称。
CASCADE | RESTRICT	当执行删除操作时，如何处理依赖对象中的相关数据。	• CASCADE：级联删除依赖于server的对象 。 • RESTRICT（缺省值）：如果存在依赖对象，则拒绝删除该server。

示例

删除hdfs_server：

```
DROP SERVER IF EXISTS hdfs_server;
```

相关链接

CREATE SERVER，**ALTER SERVER**

13.5 视图和物化视图

13.5.1 CREATE VIEW

功能描述

CREATE VIEW用于创建视图。

视图（View）是基于SQL查询定义的逻辑虚拟表，仅保存视图定义而不物理存储数据。当基表数据发生变化时，视图查询结果自动同步更新，视图提供动态数据窗口，帮助用户高效聚焦目标数据。

 版权所有 © 华为云计算技术有限公司

语法格式

```
CREATE [ OR REPLACE ] [ TEMP | TEMPORARY ] VIEW view_name [ ( column_name [, ...] ) ]
    [ WITH ( {view_option_name [= view_option_value]} [, ... ] ) ]
    AS query;
```

说明

- GUC参数view_independent打开时，支持普通视图删除列。但如果存在列级约束，不支持该列的删除。
- 视图解耦功能下不支持CTE重名。例如：

```
CREATE TABLE t1(a1 INT, b1 INT);
CREATE TABLE t2(a2 INT, b2 INT, c2 INT);
CREATE OR REPLACE VIEW v1 AS WITH tmp AS (SELECT * FROM t2) ,tmp1 AS (SELECT b2,c2
FROM tmp WHERE b2 = (WITH RECURSIVE tmp(aa, bb) AS (SELECT a1,b1 FROM t1) SELECT bb
FROM tmp WHERE aa = c2)) SELECT c2 FROM tmp1;
```

参数说明

表 13-103 CREATE VIEW 参数说明

参数	描述	取值范围
OR REPLACE	如果已存在相同名称的视图，则重新定义。	-
TEMP | TEMPORARY	创建临时视图。	-
view_name	要创建的视图名称。可以用模式修饰。	字符串，需符合**标识符命名规范**。
column_name	可选的名字列表，用作视图的字段名。如果未给出，从查询的字段名中获取。	字符串，需符合**标识符命名规范**。
WITH ({view_option_name [= view_option_value]} [, ...])	该子句指定视图的可选参数。 目前view_option_name仅支持security_barrier。 创建视图时使用WITH (security_barriers)可以创建一个相对安全的视图，避免攻击者利用低成本函数的RAISE语句打印出隐藏的基表数据。	security_barrier为boolean类型： • TRUE • FALSE
query	为视图提供行和列的SELECT或VALUES语句。	-

示例

创建基表customer并插入数据。

```
DROP TABLE IF EXISTS customer CASCADE;

CREATE TABLE customer
(id int, name varchar(20) , age int)
with (orientation = column,COMPRESSION=MIDDLE)
```


版权所有 © 华为云计算技术有限公司

```
distribute by hash(id);

INSERT INTO customer VALUES (1,'lily',10),(2, 'lucy',12),(3,'lilei',15);
```

基于表customer创建视图customer_details_view_v1，用于查看表字段age < 12的数据内容。

```
DROP VIEW IF EXISTS customer_details_view_v1;
CREATE VIEW customer_details_view_v1 AS
SELECT * FROM customer
WHERE age < 12;
```

可更新的视图

可更新的视图功能需开启GUC参数enable_view_update（用于设置是否开启视图更新功能，默认值为off），开启该参数后系统允许对简单视图使用INSERT，UPDATE、DELETE和MERGE INTO语句进行更新。

满足以下所有条件的视图可进行更新：

- 视图定义的FROM语句中只能有一个普通表，不能是系统表、外表、dfs表、delta表、toast表、错误表。
- 视图中包含可更新的列，这些列是对基础表可更新列的简单引用。
- 视图定义不能包含WITH、DISTINCT、GROUP BY、ORDER BY、FOR UPDATE、FOR SHARE、HAVING、TABLESAMPLE、LIMIT、OFFSET子句。
- 视图定义不能包含UNION、INTERSECT、EXCEPT集合操作。
- 视图定义的选择列表不能包含聚集函数、窗口函数、返回集合的函数。
- 对于INSERT、UPDATE和DELETE语句，视图上不能有触发时机为INSTEAD OF的触发器。对于MERGE INTO语句，视图和基础表上都不能有触发器。
- 视图定义不能包含子链接。
- 视图定义不能包含属性为VOLATILE的函数（函数值可以在一次表扫描内改变的函数）
- 视图定义不能对表的分布键所在列起别名，或将普通列起别名为分布键列名。
- 视图更新操作中包含RETURNING子句时，视图定义中的列只能来自于基础表。

如果可更新的视图定义包含WHERE条件，则该条件将会限制UPDATE和DELETE语句修改基础表上的行。如果UPDATE语句更改行后不再满足WHERE条件，更新后通过视图将无法查询到。类似地如果INSERT命令插入了不满足WHERE条件的数据，插入后通过视图将无法查询到。在视图上执行插入、更新或删除的用户必须在视图和表上具有相应的插入、更新或删除权限。

相关链接

ALTER VIEW，**DROP VIEW**

常见问题

视图查询时执行失败

通过DWS视图管控数据权限

 版权所有 © 华为云计算技术有限公司

13.5.2 ALTER VIEW

功能描述

ALTER VIEW更改视图的各种辅助属性。（如果用户是更改视图的查询定义，要使用CREATE OR REPLACE VIEW。）

注意事项

- 用户必须是视图的所有者才可以执行ALTER VIEW。
- 要修改视图的模式，用户必须要有新模式的CREATE权限。
- 要修改视图的所有者，用户必须是新所属角色的直接或者间接的成员，并且此角色必须有视图模式的CREATE权限。
- 管理员用户可以更改任何视图的所属关系。

语法格式

- 设置视图列的默认值。
```
ALTER VIEW [ IF EXISTS ] view_name
    ALTER [ COLUMN ] column_name SET DEFAULT expression;
```
- 取消列视图列的默认值。
```
ALTER VIEW [ IF EXISTS ] view_name
    ALTER [ COLUMN ] column_name DROP DEFAULT;
```
- 修改视图的所有者。
```
ALTER VIEW [ IF EXISTS ] view_name
    OWNER TO new_owner;
```
- 重命名视图，支持新视图名前带有原视图所在的schema名，不支持同时修改schema名。
```
ALTER VIEW [ IF EXISTS ] view_name
    RENAME TO new_name;
ALTER VIEW [ IF EXISTS ] view_name
    RENAME TO schema.new_name;
```
- 设置视图的所属模式。
```
ALTER VIEW [ IF EXISTS ] view_name
    SET SCHEMA new_schema;
```
- 设置视图的选项。
```
ALTER VIEW [ IF EXISTS ] view_name
    SET ( { view_option_name [ = view_option_value ] } [, ... ] );
```
- 重置视图的选项。
```
ALTER VIEW [ IF EXISTS ] view_name
    RESET ( view_option_name [, ... ] );
```
- 重建本视图及上层和下层依赖的无效视图。
```
ALTER VIEW [ IF EXISTS ] view_name
    REBUILD;
```
- 重建本视图及下层依赖的视图。
```
ALTER VIEW [ IF EXISTS ] ONLY view_name
    REBUILD;
```

 版权所有 © 华为云计算技术有限公司

参数说明

表 13-104 ALTER VIEW 参数说明

参数	描述	取值范围
IF EXISTS	如果视图不存在时返回一个提示而不是错误。	-
view_name	需要修改的视图的名称，可以用模式修饰。	有效的视图名。
column_name	可选的名称列表，视图的字段名。如果没有给出，字段名取自查询中的字段名。	有效的字段名。
SET/DROP DEFAULT	设置或删除一个列的缺省值，该参数暂无实际意义。	-
new_owner	视图的新所有者。	有效的用户名。
new_name	视图的新名称。	字符串，需符合**标识符命名规范**。
new_schema	视图的新模式。	有效的模式名。
SET ({ view_option_name [= view_option_value] } [, ...])	该子句为视图指定一个可选的参数。 目前view_option_name仅支持security_barrier。 创建视图时使用WITH (security_barriers)可以创建一个相对安全的视图，避免攻击者利用低成本函数的RAISE语句打印出隐藏的基表数据。	security_barrier为boolean类型： • TRUE • FALSE
REBUILD	该子句用于视图解耦，可使用已保存的原始语句重新创建视图，恢复依赖关系。9.1.0.200及以上集群版本中，视图解耦功能无需开启GUC参数view_independent，检测视图如果为无效状态，则可以进行重建。	REBUILD的详细使用说明请参见**REBUILD（视图解耦）使用说明**。
ONLY	控制视图重建的范围，只重建视图及其所依赖的视图。此功能无需开启GUC参数view_independent，检测视图如果为无效状态，则可以进行重建。	-

REBUILD（视图解耦）使用说明

- 重建视图会从当前视图开始，依次向后级联刷新与其关联的所有视图，如果其依赖的前向视图也为不可用状态，会触发自动重建。

版权所有 © 华为云计算技术有限公司

- 不支持对有依赖关系的临时表及临时视图的解耦DROP，可以对没有依赖关系的临时视图进行REBUILD操作。
- 支持视图模式名称及视图名称的修改，REBUILD按照最新的名称重建，但是query部分保留原始定义。
- 基表字段类型仅支持大类（字符型、数字型、时间类型等）范围内的修改；当基表添加字段时，视图不会置为无效，且定义不变。
- 无效视图备份时以注释形式导出，恢复时需要自行手动处理。
- 触发上层级联视图无效的场景：
  - DROP TABLE/VIEW
  - RENAME TABLE/VIEW
  - ALTER TABLE DROP COLUMN
  - ALTER TABLE CHANGE/ALTER COLUMN TYPE
  - ALTER TABLE CHANGE/ALTER COLUMN NAME
  - ALTER TABLE/VIEW NAMESPACE
  - ALTER TABLE/VIEW RENAME

示例

创建视图myview。

```
CREATE OR REPLACE VIEW myview AS
    SELECT * FROM pg_tablespace WHERE spcname = 'pg_default';
```

修改视图名称为product_view。

```
ALTER VIEW myview RENAME TO product_view;
```

修改视图product_view的schema为schetest。

```
CREATE SCHEMA schetest;
ALTER VIEW product_view SET schema schetest;
```

重建视图schetest.product_view。

```
ALTER VIEW schetest.product_view REBUILD;
```

重建依赖视图schetest.product_view。

```
ALTER VIEW ONLY schetest.product_view REBUILD;
```

相关链接

CREATE VIEW，**DROP VIEW**

使用实践

使用DWS视图重建功能实现视图解耦以提升开发效率

13.5.3 DROP VIEW

功能描述

数据库中强制删除已有的视图。

 版权所有 © 华为云计算技术有限公司

注意事项

只有视图的所有者有权限执行DROP VIEW的命令，系统管理员默认拥有此权限。

危险

- **DROP对象**操作（如DATABASE、USER/ROLE、SCHEMA、TABLE、VIEW等对象）存在数据丢失风险，尤其含带CASCADE级联删除场景，会将关联的对象一并删除，操作需谨慎，操作前需考虑数据备份。
- 更多开发设计规范参见**总体开发设计规范**。

语法格式

```
DROP VIEW [ IF EXISTS ] view_name [, ...] [ CASCADE | RESTRICT ];
```

参数说明

表 13-105 DROP VIEW 参数说明

参数	描述	取值范围
IF EXISTS	如果指定的视图不存在，则发出一个notice而不是抛出一个错误。	-
view_name	要删除的视图名字。	已存在的视图。
CASCADE | RESTRICT	当执行删除操作时，如何处理依赖对象中的相关数据。	• CASCADE：级联删除依赖此视图的对象（比如其他视图）。 • RESTRICT：如果有依赖对象存在，则拒绝删除此视图。此选项为缺省值。

示例

删除视图myView：

```
DROP IF EXISTS VIEW myView;
```

删除视图customer_details_view_v2：

```
DROP VIEW IF EXISTS public.customer_details_view_v2;
```

相关链接

ALTER VIEW，**CREATE VIEW**

版权所有 © 华为云计算技术有限公司

13.5.4 CREATE MATERIALIZED VIEW

功能描述

创建新的物化视图。物化视图语法仅8.2.1.220及以上集群支持。

物化视图（Materialized View）是数据库中一种特殊类型的数据库对象。它将复杂计算的查询结果预计算并存储在数据库中，实现加速查询的目的。

适用于频繁对大量数据进行复杂的查询和分析场景，使用物化视图可以预先计算和存储该场景的查询结果，显著提高查询性能。

物化视图权限要求

1. 创建权限：创建物化视图时需要schema的CREATE权限和基表或列的SELECT权限。
2. 查询权限：查询物化视图需要物化视图的SELECT权限。
3. 刷新权限：刷新需要物化视图的INSERT、UPDATE、DELETE和基表或列的SELECT权限。
4. 其他权限：

   物化视图支持ANALYZE/VACUUM/ALTER/DROP等细粒度权限。

   物化视图支持with grant option的权限传递操作。
5. 安全管理：

   物化视图不支持更高级别的安全控制，如果基表存在行级访问控制、脱敏策略或owner为私有用户等限制SELECT权限的场景，则禁止创建物化视图；

   如果已存在物化视图，基表增加RLS、脱敏策略或修改owner为私有用户，则物化视图可以执行查询，但无法刷新。

注意事项

- 使用CREATE MATERIALIZED VIEW语句需确保GUC参数**enable_matview**设置为on。
- 物化视图的基表可以是行存表、列存表、hstore表、冷热表（9.1.0.200及以上集群版本支持）、分区表或者指定某个分区、外表、其他物化视图。
- 不支持包含临时表（包括全局临时表、volatile临时表和普通临时表），不支持对自动分区表指定分区。
- 不支持物化视图定义中的CTE包含除SELECT语句以外的其他DML。
- 不支持基表上存在脱敏策略和行级访问控制或者基表属于私有用户。脱敏策略和行级访问控制可能会导致结果集错误，而私有用户可能导致刷新物化视图失败等问题。
- 物化视图不支持的存储类型：外表、时序表。
- 物化视图禁止使用INSERT/UPDATE/MERGE INTO/DELETE直接对数据进行修改，只能通过REFRESH命令刷新数据。
- 物化视图执行一次将结果并保存，每次查询结果是一致的。BUILD IMMEDIATE或REFRESH后，物化视图可以查询出正确结果。
- 物化视图不能通过语法指定Node Group。支持物化视图的基表指定Node Group创建，物化视图会继承基表NodeGroup信息创建，需要多个基表的NodeGroup相同。

 版权所有 © 华为云计算技术有限公司

语法格式

```
CREATE MATERIALIZED VIEW [ IF NOT EXISTS ] materialized_view_name [ ( column_name [ COMMENT
'string' ] [, ...] ) ]
[ BUILD { DEFERRED | IMMEDIATE } ]
[ REFRESH [ [DEFAULT] | [ RESTRICT ] | [ CASCADE FORWARD ] | [ CASCADE BACKWARD ] | [ CASCADE
ALL ] ]
        [ COMPLETE | FAST ]
        [ [ ON DEMAND ] | [ ON STATEMENT ] | [ ON COMMIT ] ]
        [ [ START WITH (timestamptz) ] | [ EVERY (interval) ] ] ]
[ { ENABLE | DISABLE } QUERY REWRITE ]
[ WITH ( {storage_parameter = value} [, ... ] ) ]
[ DISTRIBUTE BY { REPLICATION | ROUNDROBIN | { HASH ( column_name [,...] ) } } ]
[ PARTITION BY {(column_name) | date_trunc('time_unit', column_name)} ]
[ COMMENT 'string' ]
AS query;
CREATE MATERIALIZED VIEW [ IF NOT EXISTS ] materialized_view_name
[ ( column_name [ COMMENT 'string' ] [, ...] ) ]
[ BUILD IMMEDIATE ]
REFRESH AUTO
ENABLE QUERY REWRITE
[ WITH ( {storage_parameter = value} [, ... ] ) ]
[ COMMENT 'string' ]
AS query;
```

参数说明

表 13-106 CREATE MATERIALIZED VIEW 参数说明

参数	描述	取值范围
BUILD DEFERRED | IMMEDIATE	• IMMEDIATE，表示创建物化视图时即包含最新数据。 • DEFERRED，表示创建物化视图时需要等到第一次refresh时才会包含数据。	-

版权所有 © 华为云计算技术有限公司

参数	描述	取值范围
REFRESH	指定物化视图的刷新方式。 创建物化视图后，物化视图中的数据只反映创建时刻基表的状态。当基表中的数据发生变化时，需要通过刷新物化视图（REFRESH MATERIALIZED VIEW）更新物化视图中的数据。	• 强制模式：DEFAULT默认刷新，RESTRICT 强制刷新。仅创建后台异步刷新的物化视图可以指定。 • 刷新方向：级联嵌套使用物化视图时的刷新方向。FORWARD向上刷新，BACKWARD向下刷新，ALL 全部都刷。仅创建后台异步刷新的物化视图可以指定。具体参考**REFRESH MATERIALIZED VIEW**。 • 刷新方法： COMPLETE：全量刷新，执行物化视图定义的查询语句并更新物化视图。 FAST：增量刷新，抽取每张基表自上次刷新依赖的增量数据，进行增量计算后写入物化视图。 • 刷新触发方式。 – ON DEMAND：手动按需刷新。 – ON COMMIT：基表修改的事务提交时，会触发基表上所有物化视图刷新。保证了基表与物化视图的数据强一致性，但会影响基表的数据写入速度，请谨慎考虑。 – ON STATEMENT：基表修改的DML语句执行时，触发基表上所有物化视图刷新。保证了基表与物化视图的数据强一致性，但会影响基表的数据写入速度，请谨慎考虑。 ON COMMIT和ON STATEMENT都不支持在可重复读的事务隔离级别下使用，触发时会有报错拦截。 • START WITH (timestamptz) | EVERY (interval)：定时自动刷新。START WITH指定首次刷新时间，EVERY 指定刷新间隔，根据指定的时间定时自动刷新，支持MONTH、DAY、HOUR、MINUTE及SECOND。

版权所有 © 华为云计算技术有限公司

参数	描述	取值范围
ENABLE | DISABLE QUERY REWRITE	是否开启查询重写。 查询重写是指在对基表进行查询时， 如果基表上创建有物化视图，数据库系统自动判断是否可以使用物化视图中的预计算结果处理查询。 如果可以使用某个物化视图，会直接从该物化视图读取预计算结果，起到加速查询的作用。	默认关闭。 指定ENABLE QUERY REWRITE时，需要设置GUC参数**mv_rewrite_rule**才能启用物化视图查询重写功能。
WITH	创建物化视图指定的相关参数。	参见**表13-107**。
DISTRIBUTE BY	指定表如何在节点之间分布或者复制。	取值范围： ● REPLICATION：表的每一行存储在所有数据节点（DN）中，即每个数据节点都有完整的表数据。 ● ROUNDROBIN：表的每一行被轮番地发送给各个DN，因此数据会被均匀地分布在各个DN中。（ROUNDROBIN仅8.1.2及以上版本支持） ● HASH：对指定的列进行Hash，通过映射，把数据分布到指定DN。 默认值：由参数default_distribution_mode决定。 **说明** 当物化视图指定为hash分布时，可能存在数据倾斜。可使用普通表检查数据是否倾斜的方法对物化视图进行检查，详细内容请参考《数据仓库服务开发指南》的“查看数据倾斜状态”章节。 若物化视图的数据存储存在倾斜时，可参考《数据仓库服务开发指南》的“数据倾斜调优”章节进行存储层的倾斜调优。

 版权所有 © 华为云计算技术有限公司

参数	描述	取值范围
PARTITION BY	创建分区物化视图语法仅9.1.0.200及以上集群版本支持。 分区物化视图的约束与限制： ● 物化视图只支持range分区。 ● 不支持基表分区的merge和split。 ● 物化视图的分区键，必须是被查询表中某个分区表的分区键，用于建立按分区失效和刷新的对应关系。 ● 物化视图的分区键，必须包含在查询的最外层SQL中，以便将查询数据按分区键刷入物化视图表。 ● 不支持Outer join中会产生空值的列作为分区键的情况。 ● 被查询表可以是组合分区，如果被查询表有多个分区表且分区条件都一样，都会与物化视图自动建立映射关系。 ● 不支持外表的分区映射，因为无法感知外表的数据变化。	● 等比例关联，指定查询中某个分区表的分区键作为物化视图的分区键，并建立一比一的映射关系。 ● 上卷式关联，通过date_trunc函数对查询中某个分区表的分区键进行上卷，并建立上卷式的映射关系。要求数据类型必须是date_trunc函数可接受的日期类型。
COMMENT	为物化视图增加注释说明。该参数仅9.1.2及以上集群版本支持。	-
AS query	基于query的结果创建物化视图。	-

 版权所有 © 华为云计算技术有限公司

表 13-107 WITH 参数说明

参数	描述	取值范围
ORIENTATION	指定表数据的存储方式，即行存方式、列存方式，该参数设置成功后就不再支持修改。	ROW或COLUMN： ● ROW，表示表的数据将以行式存储。 行存储适合于OLTP业务，此类型的表上交互事务比较多，一次交互会涉及表中的多个列，用行存查询效率较高。 ● COLUMN，表示表的数据将以列式存储。 列存储适合于数据仓库业务，此类型的表上会做大量的汇聚计算，且涉及的列操作较少。 默认值：ROW。
bitmap_columns	bitmap index只适用于hstore_opt表，只有开启表级参数enable_hstore_opt且开启bitmap_columns='指定列'。该参数仅9.1.0.200及以上集群版本支持。	-
enable_foreign_table_query_rewrite	指定是否允许包含外表的物化视图进行查询重写，需要与ENABLE QUERY REWRITE一起使用。 外表数据有变化，物化视图无法感知。如果需要对包含外表的物化视图使用查询重写功能，需要指定此选项。	● on，表示允许包含外表的物化视图进行查询重写。 ● off，表示不允许包含外表的物化视图进行查询重写。 默认值：off
enable_hstore_opt	enable_hstore_opt表级参数打开时会默认同时打开enable_hstore表级参数。该参数仅9.1.0.200及以上集群版本支持。	true或者false。 默认值：false
enable_turbo_store	用于控制是否创建为turbo表（基于列存表实现）。该参数只对列存表有效。该参数仅9.1.0.200及以上集群版本支持。	on或者off。 默认值：off

版权所有 © 华为云计算技术有限公司

参数	描述	取值范围
excluded_inactive_tables	指定的基表发生数据变化不会失效该物化视图。该参数仅9.1.0.200及以上集群版本支持。 设置方式为：excluded_inactive_tables='schemaName1.tableName1,schemaName2.tableName2'	默认值：空
force_rewrite_timeout	在刷新后的指定时间间隔内，可以查询重写，无论此物化视图数据是否新鲜。该参数仅9.1.0.200及以上集群版本支持。	默认值：0，单位为秒。
mv_analyze_mode	控制物化视图自动analyze的方式。该参数仅9.1.0.200及以上集群版本支持。	• none，表示物化视图在刷新后不自动执行analyze。 • light，表示物化视图在刷新后执行light analyze。 默认值：light
mv_computing_resource	物化视图刷新指定的弹性逻辑集群，可减少刷新对当前集群资源的影响。该参数仅9.1.1.200及以上集群版本支持。	默认值：空
mv_incremental_mode	指定增量物化视图的格式。该参数仅9.1.1.200及以上集群版本支持。	• 空，表示要借助辅助表方式进行增量计算，性能优。 • v2，表示不借助辅助表方式进行增量计算。 默认值：空
mv_pck_column	物化视图的局部聚簇存储，列存表导入数据时按照指定的列（单列或多列），进行局部排序。该参数仅9.1.0.200及以上集群版本支持。 开启mv_pck_column='指定列'。	默认值：空

版权所有 © 华为云计算技术有限公司

参数	描述	取值范围
mv_support_function_type	开启物化视图创建时查询语句中可以使用的函数属性。该参数仅9.1.0.200及以上集群版本支持。	• stable，表示支持在查询语句中使用类型为STABLE和IMMUTABLE的函数。 • volatile，表示支持在查询语句中使用类型为VOLATILE，STABLE和IMMUTABLE的函数。 默认值：空
mv_refresh_guc	当前物化视图刷新时需要使用的GUC参数。该参数仅9.1.1.100及以上集群版本支持。	默认值：空
mv_direct_query_policy	物化视图直查策略参数。该参数仅9.1.2及以上集群版本支持。 物化视图直查策略参数也支持通过GUC参数mv_direct_query_policy（默认值为force_mv）进行设置。	• force_mv，表示直接查询物化视图，无论视图是否失效。 • prefer_mv，表示少量分区失效时，使用可用的分区物化视图与基表进行UNION ALL查询；大量分区失效时，仍使用物化视图进行查询，性能优先。 • prefer_base，表示少量分区失效时，使用可用的分区物化视图与基表进行UNION ALL查询；大量分区失效时，使用基表进行查询。 • force_base，表示强制使用基表进行查询，完全不使用物化视图。 默认值：force_mv
mv_partition_rewrite_threshold	分区物化视图重写阈值，该参数仅9.1.2及以上集群版本支持。	• 正值，表示最大失效分区数。 • 负值，表示最大失效分区比例。 默认值：-0.3， 表示30%。
secondary_part_num	指定列存表二级分区的数量，仅适用于HStore列存表。该参数仅9.1.0.200及以上集群版本支持。	1~32 默认值：8

版权所有 © 华为云计算技术有限公司

参数	描述	取值范围
TABLESPACE tablespace_name	声明一个表空间。用于V3存储格式的创建，如果default_tablespace为空，将使用数据库的缺省表空间。该参数仅9.1.0.200及以上集群版本支持。	-

示例

开启物化视图功能。

```
set enable_matview = on;
```

创建基表，并向基表插入数据。

```
DROP TABLE IF EXISTS t1;
CREATE TABLE t1 (a int, b int) DISTRIBUTE BY HASH(a);
INSERT INTO t1 SELECT x,x FROM generate_series(1,10) x;
```

创建默认BUILD IMMEDIATE方式的物化视图。

```
CREATE MATERIALIZED VIEW mv1 AS SELECT * FROM t1;
```

创建指定列存方式的物化视图。

```
CREATE MATERIALIZED VIEW mv2 WITH(orientation = column) AS SELECT * FROM t1;
```

创建手动按需刷新的物化视图。

```
CREATE MATERIALIZED VIEW mv3 BUILD DEFERRED REFRESH ON DEMAND AS SELECT * FROM t1;
```

创建指定刷新时间的物化视图。

```
CREATE MATERIALIZED VIEW mv4 BUILD DEFERRED REFRESH START WITH(trunc(sysdate)) EVERY
(interval '1 day') AS SELECT * FROM t1;
```

创建带有bitmap index的物化视图。

```
DROP TABLE IF EXISTS base_table;
CREATE TABLE base_table (col1 varchar(50), col2 int , col3 int) WITH (ORIENTATION=COLUMN);
CREATE MATERIALIZED VIEW mv1
with (ORIENTATION = COLUMN, enable_hstore=true, enable_hstore_opt=on, bitmap_columns='col1')  AS
SELECT * FROM base_table;
```

创建物化视图并指定列存表二级分区的数量。

```
CREATE MATERIALIZED VIEW mv
WITH (ORIENTATION=COLUMN, ENABLE_HSTORE=ON, enable_hstore_opt=on, mv_pck_column='col3',
secondary_part_column = 'col2', secondary_part_num = 8)  AS SELECT * FROM base_table;
```

创建物化视图并指定pck列进行排序。

```
CREATE MATERIALIZED VIEW mv2
WITH (ORIENTATION=COLUMN, ENABLE_HSTORE=ON, enable_hstore_opt=on, mv_pck_column='col3')  AS
SELECT * FROM base_table;
```

创建物化视图并指定其analyze的方式。

```
CREATE MATERIALIZED VIEW mv1 enable query rewrite
with(excluded_inactive_tables='matview_basic."T1",matview_basic."a=b"',mv_analyze_mode='none') as
SELECT * FROM base_table;
```


版权所有 © 华为云计算技术有限公司

创建V3物化视图。

```
DROP TABLE IF EXISTS dicttbl_low;
CREATE TABLE dicttbl_low (col1 int, col2 int ) WITH (ORIENTATION=COLUMN);
CREATE MATERIALIZED VIEW mv1
with (orientation=column, enable_hstore=true, compression=low, enable_hstore_opt=on, COLVERSION =
3.0) TABLESPACE cu_obs_tbs distribute by hash(col1)
AS SELECT * FROM dicttbl_low;
```

创建包含外表的物化视图进行查询重写。**注意**，该示例的base_table_ft为外表。

```
CREATE MATERIALIZED VIEW mv1 with (enable_foreign_table_query_rewrite = true) as SELECT * FROM
base_table_ft;
```

创建物化视图指定查询语句中可以使用volatile函数。

```
DROP TABLE IF EXISTS t_date;
CREATE TABLE t_date (a int, b int ) WITH (ORIENTATION=COLUMN);
CREATE MATERIALIZED VIEW mv_date with(mv_support_function_type = 'volatile') as select to_date(a)
from t_date;
```

创建物化视图时添加注释。

```
CREATE MATERIALIZED VIEW mv1 DISTRIBUTE BY HASH(id)
COMMENT 'table level comment'
AS SELECT * FROM base_table;
```

创建物化视图时添加列的注释。

```
CREATE MATERIALIZED VIEW mv2 (
  id COMMENT 'identifier',
  name COMMENT 'person name',
  age
) DISTRIBUTE BY HASH(id)
COMMENT 'table and column comments'
AS SELECT * FROM base_table;
```

相关链接

ALTER MATERIALIZED VIEW，**DROP MATERIALIZED VIEW**，**REFRESH MATERIALIZED VIEW**

13.5.5 ALTER MATERIALIZED VIEW

功能描述

修改物化视图属性。物化视图语法仅8.2.1.220及以上集群支持。

注意事项

使用ALTER MATERIALIZED VIEW语句需确保**enable_matview**设置为on。

语法格式

- 设置是否对物化视图启动查询重写。

  ```
  ALTER MATERIALIZED VIEW [ IF EXISTS ] { materialized_view_name }
      [ ENABLE | DISABLE ] QUERY REWRITE;
  ```

- 修改物化视图的刷新方式或更新物化视图。

  ```
  ALTER MATERIALIZED VIEW [ IF EXISTS ] { materialized_view_name }
      REFRESH [ [DEFAULT] | [ RESTRICT ] | [ CASCADE FORWARD ] | [ CASCADE BACKWARD ] |
  [ CASCADE ALL ] ] [ COMPLETE ] [ ON DEMAND ] [ [ START WITH (timestamptz) ] | [ EVERY
  (interval) ] ];
  ```

 版权所有 © 华为云计算技术有限公司

  - REFRESH ON DEMAND，表示手动按需刷新。
  - START WITH，指定首次刷新时间。
  - EVERY指定刷新间隔，根据指定的时间定时刷新，支持MONTH、DAY、HOUR、MINUTE及SECOND。
- 修改后台异步刷新物化视图的刷新方式，仅针对后台异步刷新的物化视图。该语法仅9.1.0.200及以上集群版本支持。具体内容参考**REFRESH MATERIALIZED VIEW**。

```
ALTER MATERIALIZED VIEW [ IF EXISTS ] { materialized_view_name }
    REFRESH [DEFAULT] | [ RESTRICT ] | [ CASCADE FORWARD ] | [ CASCADE BACKWARD ] |
[ CASCADE ALL ] ;
```

- 修改物化视图的所有者。

```
ALTER MATERIALIZED VIEW { materialized_view_name }
    OWNER TO new_owner;
```

- 设置物化视图的表属性。该语法仅9.1.0.200及以上集群版本支持。

```
ALTER MATERIALIZED VIEW { materialized_view_name }
    SET ( {storage_parameter = value} [, ... ] )
    | RESET ( storage_parameter [, ... ] )
```

- 重命名物化视图。不支持跨schema对物化视图进行重命名操作。重命名物化视图时原物化视图名materialized_view_name可指定模式名，new_materialized_view_name不能指定模式名。

```
ALTER MATERIALIZED VIEW [ IF EXISTS ] materialized_view_name
    RENAME TO new_materialized_view_name;
```

参数说明

表 13-108 ALTER MATERIALIZED VIEW 参数说明

参数	描述	取值范围
IF EXISTS	如果指定的物化视图不存在时，返回一个提示而不是错误。	-
materialized_view_name	需要修改的物化视图的名称。	已存在的物化视图名。
ENABLE | DISABLE QUERY REWRITE	设置是否开启查询重写。 指定ENABLE QUERY REWRITE时，需要设置GUC参数**mv_rewrite_rule**才能启用物化视图查询重写功能。	默认为DISABLE，即关闭查询重写。

版权所有 © 华为云计算技术有限公司

参数	描述	取值范围
REFRESH	修改物化视图的刷新方式。 当基表中的数据发生变化时，需要通过刷新物化视图（REFRESH MATERIALIZED VIEW）更新物化视图中的数据。 • 刷新方式：DEFAULT默认刷新，RESTRICT强制刷新。仅创建后台异步刷新的物化视图可以指定。 • 刷新方向：级联嵌套使用物化视图时的刷新方向，FORWARD向上刷新，BACKWARD向下刷新，ALL 全部都刷。仅创建后台异步刷新的物化视图可以指定。具体参考**REFRESH MATERIALIZED VIEW**。 • 目前只支持COMPLETE全量刷新这一种刷新方式。执行物化视图定义的查询语句并更新物化视图。 • 刷新触发方式。 ON DEMAND：手动按需刷新。 START WITH (timestamptz) | EVERY (interval)：定时刷新。START WITH指定首次刷新时间，EVERY 指定刷新间隔，根据指定的时间定时刷新，支持MONTH、DAY、HOUR、MINUTE及SECOND。	-
storage_parameter	设置物化视图的表属性。 支持设置参数：mv_pck_column，bitmap_columns，enable_foreign_table_query_rewrite，excluded_inactive_tables，force_rewrite_timeout，mv_analyze_mode等。	更多支持设置的参数可参考**参数说明**。
new_owner	物化视图的新所有者。	已存在的用户名。
new_materialized_view_name	新的物化视图名称。	字符串，需符合**标识符命名规范**。

示例

开启物化视图功能。

```
SET enable_matview = on;
```

创建基表，并向基表插入数据。

```
DROP TABLE IF EXISTS t1 CASCADE;
CREATE TABLE t1 (a int, b int)
WITH (ORIENTATION = COLUMN )
```

 版权所有 © 华为云计算技术有限公司

```
DISTRIBUTE BY HASH(a);
INSERT INTO t1 SELECT x,x FROM generate_series(1,10) x;
```

创建指定列存方式的物化视图：

```
CREATE MATERIALIZED  VIEW mv1 WITH(orientation = column, enable_foreign_table_query_rewrite = false ) AS SELECT * FROM t1;
```

启用物化视图的查询重写。

```
ALTER MATERIALIZED VIEW mv1 ENABLE QUERY REWRITE;
```

修改物化视图的表属性。

```
ALTER MATERIALIZED VIEW mv1 SET (force_rewrite_timeout=100);
ALTER MATERIALIZED VIEW mv1 SET (mv_pck_column='b');
ALTER MATERIALIZED VIEW mv1 SET (enable_foreign_table_query_rewrite = true);
```

修改物化视图的刷新时间。

```
ALTER MATERIALIZED VIEW mv1 REFRESH START WITH('2025-01-01 15:15:15'::timestamptz) EVERY (interval '60 s');
```

修改物化视图的名称。

```
ALTER MATERIALIZED VIEW my_schema.mv1 RENAME TO new_mv;
```

相关链接

CREATE MATERIALIZED VIEW，DROP MATERIALIZED VIEW，REFRESH MATERIALIZED VIEW

13.5.6 REFRESH MATERIALIZED VIEW

功能描述

刷新物化视图。该语法仅8.2.1.220及以上集群支持。

刷新方式由CREATE MATERIALIZED VIEW语法中的REFRESH参数指定。目前支持全量刷新或者定时刷新。

注意事项

- 刷新操作会阻塞基表的DML操作。
- 增量物化视图的REFRESH不支持explain performance。

语法格式

```
REFRESH MATERIALIZED VIEW
{[schema.]materialized_view_name} [ [DEFAULT] | [ RESTRICT ] | [ CASCADE FORWARD ] | [ CASCADE BACKWARD ] | [ CASCADE ALL ] ]
```

参数说明

表 13-109 REFRESH MATERIALIZED VIEW 参数说明

参数	描述
materialized_view_name	需要刷新的物化视图的名称。

 版权所有 © 华为云计算技术有限公司

参数	描述
DEFAULT	采用默认的刷新方式。即刷新单个物化视图，若基表未发生变化，不进行刷新操作。
RESTRICT	采用强制刷新。不关注基表是否变化，一定会执行刷新操作。
CASCADE FORWARD	级联向上刷新。先刷新当前物化视图，再依次刷新依赖此物化视图的嵌套物化视图。刷新方式为默认刷新方式。
CASCADE BACKWARD	级联向下刷新。先刷新当前物化视图所依赖的最底层物化视图，再依次刷新底层物化视图依赖的嵌套物化视图，最后刷新当前物化视图。刷新方式为默认刷新方式。
CASCADE ALL	级联刷新全部。先执行级联向下刷新，然后执行级联向上刷新。

示例

刷新物化视图。

```
SET enable_matview = ON;
DROP TABLE IF EXISTS t;
CREATE TABLE t (a int, b int);
DROP MATERIALIZED VIEW IF EXISTS mv1;
CREATE MATERIALIZED VIEW mv1 BUILD DEFERRED REFRESH ON DEMAND AS SELECT * FROM t;
REFRESH MATERIALIZED VIEW mv1;
```

级联向下刷新物化视图。

```
DROP TABLE IF EXISTS t1;
DROP MATERIALIZED VIEW IF EXISTS mv1;
DROP MATERIALIZED VIEW IF EXISTS mv2;
CREATE TABLE t1(a int,b int);
CREATE MATERIALIZED VIEW mv1 AS SELECT * FROM t1 where a > 0;
CREATE MATERIALIZED VIEW mv2 AS SELECT * FROM mv1;
INSERT INTO t1 values(1,1),(2,2);
REFRESH MATERIALIZED VIEW mv2 cascade backward; -- 刷新完mv2, mv1数据也进行了更新

SELECT * FROM mv1 ORDER BY 1,2;
```

```
gaussdb=> SELECT * FROM mv1 ORDER BY 1,2;
 a | b
---+---
 1 | 1
 2 | 2
(2 rows)
```

```
SELECT * FROM mv2 ORDER BY 1,2;
```

```
gaussdb=> SELECT * FROM mv2 ORDER BY 1,2;
 a | b
---+---
 1 | 1
 2 | 2
(2 rows)
```


版权所有 © 华为云计算技术有限公司

相关链接

CREATE MATERIALIZED VIEW，ALTER MATERIALIZED VIEW，DROP MATERIALIZED VIEW

13.5.7 DROP MATERIALIZED VIEW

功能描述

删除物化视图。物化视图语法仅8.2.1.220及以上集群支持。

注意事项

无

语法格式

```
DROP MATERIALIZED VIEW [ IF EXISTS ]
{[schema.]materialized_view_name} [, ...] [ CASCADE | RESTRICT ];
```

参数说明

表 13-110 DROP MATERIALIZED VIEW 参数说明

参数	描述	取值范围
IF EXISTS	如果指定的物化视图不存在，则发出一个notice而不是抛出一个错误。	-
materialized_view_name	需要删除的物化视图名称。	已存在的物化视图名称。
CASCADE | RESTRICT	当执行删除操作时，如何处理依赖对象中的相关数据。	• CASCADE：自动删除包含在模式中的对象。 • RESTRICT：如果模式包含任何对象，则删除失败（缺省行为）。

示例

删除物化视图：

```
DROP MATERIALIZED IF EXISTS VIEW mv1;
```

相关链接

ALTER MATERIALIZED VIEW，CREATE MATERIALIZED VIEW，REFRESH MATERIALIZED VIEW

版权所有 © 华为云计算技术有限公司

13.6 索引&序列&文件目录

13.6.1 CREATE INDEX

功能描述

在指定的表上创建索引。**索引（Index）**是一种通过优化数据检索效率的数据结构，其核心作用类似于书籍的目录——通过预先组织关键信息，帮助数据库快速定位所需数据，减少查询时需要扫描的数据量。

索引的核心作用是：

1. **加速查询速度：**

   在处理大量数据的数据库应用中，用户经常遇到查询速度慢的问题，尤其是在没有索引的情况下，数据库查询需要扫描整个表（全表扫描），效率极低。如何提高查询效率，减少数据处理时间？通过建立索引，可以显著提升查询速度。有索引时，可通过索引直接定位到符合条件的数据位置，大幅减少数据扫描量。例如：查询where id=100时，若id列有索引，能直接找到该记录，无需遍历所有行。

2. **优化排序和分组操作：**

   索引本身是有序的（如 Btree索引），当查询涉及ORDER BY（排序）或GROUP BY（分组）时，可直接利用索引的有序性，避免额外的排序操作，提升效率。

3. **约束数据完整性**：

   某些索引（如主键索引、唯一索引）可强制列的值唯一（如主键不能重复），辅助维护数据的完整性。

DWS 索引支持的类型

DWS支持创建的索引类型包括：

- **B-Tree**：**B-Tree**索引使用一种类似于B+树的结构来存储数据的键值，通过这种结构能够快速的查找索引。**B-Tree**适合范围查询、排序、等值查询。
- **GIN**：GIN索引是倒排索引，可以处理包含多个键的值（比如数组），适用场景：多值元素、模糊匹配、全文检索。
- **GiST**：**GiST**索引适用于几何和地理等多维数据类型和集合数据类型，适用场景：多维数据、空间搜索、复杂范围。
- **PSORT**：**PSORT**索引。针对列存表进行局部排序索引。

其中，

- **行存表**支持：**B-Tree**（行存表缺省值）、**GIN**、**GiST**。
- **列存表**支持：**PSORT**（列存表缺省值）、**B-Tree**、**GIN**。

从**分区表**的角度分析，又分为本地索引（LOCAL INDEX），全局索引（GLOBAL INDEX）。

- **本地索引**（LOCAL INDEX）：也叫分区索引，即在分区表的每个区上单独创建索引，本地索引与分区表的分区一一对应，具有与其表相同的分区数和相同的分区范围。当表分区发生变化时，只需维护对应的索引分区。

 版权所有 © 华为云计算技术有限公司

- **全局索引**（GLOBAL INDEX）：全局索引是跨分区的全局索引，索引结构独立于基表的分区方式，覆盖整个表的所有分区数据。目前DWS不支持全局索引。

DWS 支持的索引方式

表 13-111 索引方式

索引方式	描述
唯一索引	可用于约束索引属性值的唯一性，或者属性组合值的唯一性。如果一个表声明了唯一约束或者主键，则DWS自动在组成主键或唯一约束的字段上创建唯一索引（可能是多字段索引），以实现这些约束。目前，DWS只有B-Tree可以创建唯一索引。
多字段索引（或复合索引、联合索引）	一个索引可以定义在表中的多个属性上。目前，DWS中的B-Tree支持多字段索引，且最多可在32个字段上创建索引。
部分索引	建立在一个表的子集上的索引，这种索引方式只包含满足条件表达式的元组。
表达式索引	索引建立在一个函数或者从表中一个或多个属性计算出来的表达式上。表达式索引只有在查询时使用与创建时相同的表达式才会起作用。

创建索引的原则

索引可以用来提高数据库查询性能，但是不恰当的使用将导致数据库性能下降。建议仅在匹配如下某条原则时创建索引：

- 经常执行查询的字段。
- 在连接条件上创建索引，对于存在多字段连接的查询，建议在这些字段上建立组合索引。例如，select * from t1 join t2 on t1.a=t2.a and t1.b=t2.b，可以在t1表上的a，b字段上建立组合索引。
- where子句的过滤条件字段上（尤其是范围条件）。
- 在经常出现在order by、group by和distinct后的字段。
- 对于点查询场景，推荐建立B-Tree索引。

在分区表上创建索引与在普通表上创建索引的语法不太相同，使用时请注意，如分区表上不支持并行创建索引、不支持创建部分索引、不支持**NULL FIRST**特性。

注意事项

- 索引自身也占用存储空间、消耗计算资源，创建过多的索引将对数据库性能造成负面影响（尤其影响数据导入的性能，建议在数据导入后再建索引）。因此，仅在必要时创建索引。
- 索引定义里的所有函数和操作符都必须是immutable类型的，即结果必须依赖于其输入参数，而不受任何外部的影响（如另外一个表的内容或者当前时间），该限制可以确保该索引的行为是定义良好的。在一个索引上或WHERE语句中使用用户定义函数，请将其标记为immutable类型函数。

 版权所有 © 华为云计算技术有限公司

- 在分区表上创建唯一索引时，索引项中必须包含分布列和所有分区键。
- DWS在分区表上创建索引时只支持本地（LOCAL）索引，不支持全局（GLOBAL）索引。
- 列存表和HDFS表支持B-tree索引，不支持创建表达式索引、部分索引。
- 列存表支持通过B-tree索引建立唯一索引。
- 列存表和HDFS表支持的PSORT索引不支持创建表达式索引、部分索引和唯一索引。
- 列存表支持的GIN索引支持创建表达式索引，但表达式不能包含空分词、空列和多列，不支持创建部分索引和唯一索引。
- 列存索引不支持OR查询过滤条件及inlist场景。
- roundrobin表不支持创建主键/唯一索引。
- 对表执行CREATE INDEX或REINDEX操作时会触发索引重建（索引重建过程中会先把数据转储到一个新的数据文件中，重建完成之后会删除原始文件），当表比较大时，重建会消耗较多的磁盘空间。当磁盘空间不足时，要谨慎对待大表CREATE INDEX或REINDEX操作，防止触发集群只读。
- 在GUC参数enable_consistent_oid为on时，对同一张分区表进行CREATE INDEX与REINDEX TABLE并发操作时，低概率会导致不同节点间分区索引的数量不一致，进而导致REINDEX TABLE报错。出现此情况时，重试执行REINDEX TABLE语句即可成功。

警告

- 针对有大批量数据增删改的表，索引个数建议控制在3个以内，最多不超过5个。
- 避免在业务高峰期执行对大表执行CREATE INDEX和REINDEX操作。
- 更多开发设计规范参见**总体开发设计规范**。

语法格式

- 在表上创建索引。

```
CREATE [ UNIQUE ] INDEX [ [IF NOT EXISTS] [ schema_name. ] index_name ] ON table_name
[ USING method ]
    ({ { column_name | ( expression ) } [ COLLATE collation ] [ opclass ] [ ASC | DESC ] [ NULLS
{ FIRST | LAST } ] }[, ...] )
    [ NULLS [ NOT ] DISTINCT | NULLS IGNORE ]
    [ COMMENT 'text' ]
    [ WITH ( {storage_parameter = value} [, ... ] ) ]

    [ WHERE predicate ];
```

- 在分区表上创建索引。

```
CREATE [ UNIQUE ] INDEX [ [IF NOT EXISTS] [ schema_name. ] index_name ] ON table_name
[ USING method ]
    ( {{ column_name | ( expression ) } [ COLLATE collation ] [ opclass ] [ ASC | DESC ] [ NULLS
LAST ] }[, ...] )
    [ NULLS [ NOT ] DISTINCT | NULLS IGNORE ]
    [ COMMENT 'text' ]
    LOCAL [ ( { PARTITION index_partition_name  } [, ...] ) ]
    [ WITH ( { storage_parameter = value } [, ...] ) ]
    ;
```


版权所有 © 华为云计算技术有限公司

参数说明

表 13-112 CREATE INDEX 参数说明

参数	描述	取值范围
UNIQUE	创建唯一性索引，每次添加数据时检测表中是否有重复值。如果插入或更新的值会引起重复的记录时，系统会报错。 目前只有行存表B-tree索引和列存表的B-tree索引支持唯一索引。	-
IF NOT EXISTS	指定IF NOT EXISTS时，若不存在同名索引，则可以成功创建索引。若已存在同名索引，创建时不会报错，仅会提示该索引已存在，且不执行任何操作。当使用IF NOT EXISTS时，需要指定索引名。该参数仅9.1.0及以上集群版本支持。	-
schema_name	要创建的索引所在的模式名。指定的模式名需与表所在的模式相同。	-
index_name	要创建的索引名，索引的模式与表相同。	取值范围：字符串，要符合标识符的命名规范。
table_name	需要为其创建索引的表的名字，可以用模式修饰。	取值范围：已存在的表名。

版权所有 © 华为云计算技术有限公司

参数	描述	取值范围
USING method	指定创建索引的方法。	取值范围： ● **btree**：B-tree索引使用一种类似于B+树的结构来存储数据的键值，通过这种结构能够快速的查找索引。B-Tree适合支持比较查询以及查询范围。 ● **gin**：GIN索引是倒排索引，可以处理包含多个键的值（比如数组）。 ● **gist**：GiST索引适用于几何和地理等多维数据类型和集合数据类型。 ● **psort**：PSORT索引。针对列存表进行局部排序索引。 行存表支持的索引类型：btree（行存表缺省值）、gin、gist。 列存表支持的索引类型：psort（列存表缺省值）、btree、gin。
column_name	表中需要创建索引的列的名字（字段名）。	如果索引方式支持多字段索引，可以声明多个字段。最多可以声明32个字段。
expression	创建一个基于该表的一个或多个字段的表达式索引，通常必须写在圆括号中。	如果表达式有函数调用的形式，圆括号可以省略。 表达式索引可用于获取对基本数据的某种变形的快速访问。比如，一个在upper(col)上的函数索引将允许WHERE upper(col) = 'JIM'子句使用索引。 在创建表达式索引时，如果表达式中包含IS NULL子句，则这种索引是无效的。此时，建议用户尝试创建一个部分索引。
COLLATE collation	COLLATE子句指定列的排序规则（该列必须是可排列的数据类型）。如果没有指定，则使用默认的排序规则。	-

 版权所有 © 华为云计算技术有限公司

参数	描述	取值范围
opclass	操作符类的名字。	对于索引的每一列可以指定一个操作符类，操作符类标识了索引那一列的使用的操作符。例如一个B-tree索引在一个四字节整数上可以使用int4_ops；这个操作符类包括四字节整数的比较函数。实际上对于列上的数据类型默认的操作符类是足够用的。操作符类主要用于一些有多种排序的数据。例如，用户想按照绝对值或者实数部分排序一个复数。能通过定义两个操作符类然后在建立索引时选择合适的类。
ASC	指定按升序排序（默认）。本选项仅行存支持。	-
DESC	指定按降序排序。本选项仅行存支持。	-
NULLS FIRST	指定空值在排序中排在非空值之前，当指定DESC排序时，本选项为默认的。	-
NULLS LAST	指定空值在排序中排在非空值之后，未指定DESC排序时，本选项为默认的。	-

 版权所有 © 华为云计算技术有限公司

参数	描述	取值范围
NULLS DISTINCT NULLS NOT DISTINCT NULLS IGNORE	指定Unique唯一索引中索引列NULL值的处理方式。	默认取值：该参数默认取值为空，即NULL值可重复插入。 在对插入的新数据和表中原始数据进行列的等值比较时，对于NULL值有以下三种处理方式： • NULLS DISTINCT：将每个NULL值视为不同的值，即允许NULL值重复插入。 • NULLS NOT DISTINCT：将所有NULL值视为相同的值，即索引中只保留一个NULL值条目，不允许NULL值重复插入，只有非NULL值且与原索引列中已有数据不相等的数据可插入。 • NULLS IGNORE：在等值比较时跳过NULL值。若索引列全为NULL，则NULL值可重复插入；部分索引列为NULL，只有非NULL值且与原索引列中已有数据不相等的数据可插入。 三种处理方式具体的行为参见**表13-114**所示。
COMMENT 'text'	指定索引的注释信息。	-
WITH ({storage_parameter = value} [, ...])	指定索引方法的存储参数。包括：**FILLFACTOR**、**INVISIBLE**、**FASTUPDATE**、**GIN_PENDING_LIST_LIMIT**。	取值范围和详细用法参见**表13-113**。

版权所有 © 华为云计算技术有限公司

参数	描述	取值范围
WHERE predicate	创建一个部分索引。 部分索引是一个只包含表的一部分记录的索引，通常是该表中比其他部分数据更有用的部分。例如，有一个表，表里包含已记账和未记账的订单，未记账的订单只占表的一小部分而且这部分是最常用的部分，此时就可以通过只在未记账部分创建一个索引来改善性能。另外一个可能的用途是使用带有UNIQUE的WHERE强制一个表的某个子集的唯一性。	取值范围：predicate表达式只能引用表的字段，它可以使用所有字段，而不仅是被索引的字段。目前，子查询和聚集表达式不能出现在WHERE子句里。
PARTITION index_partition_name	索引分区的名称。	取值范围：字符串，要符合标识符的命名规范。

版权所有 © 华为云计算技术有限公司

表 13-113 WITH ({storage_parameter = value} [, ...])指定索引方法的存储参数

存储参数	描述	示例
FILLFACTOR	**GIN和PSORT索引不支持该参数**，表示一个索引的填充因子（fillfactor），取值范围是一个介于10和100之间的百分比值，用于指定索引页初始填充的百分比，剩余的空间可以用于后续的插入和更新操作，从而减少页面分裂和提高性能。 该参数的作用是控制索引页的填充程度，从而影响数据库的性能和存储空间的利用效率，通过调整FILLFACTOR的值，可以减少后续更新或插入操作导致的页分裂，优化索引的存储和性能。 **FILLFACTOR对索引性能的影响**： ● **性能**：当索引页填满后，新的数据插入会导致页面分裂，这会消耗更多的I/O资源和时间。较低的 fillfactor可以减少页面分裂，提高插入和更新性能，但会增加存储空间的使用。较高的fillfactor可以提高存储空间的利用率，但可能会导致更多的页面分裂，降低插入和更新性能。 ● **存储空间**：较低的fillfactor会增加每个数据页面的空闲空间，导致更多的磁盘空间使用。较高的fillfactor会更紧凑地存储数据，减少磁盘空间使用。 **如何选择合适的FILLFACTOR值**： ● 对于更新频繁的表，选择较小fillfactor值。例如60~80，预留空间，减少页分裂。 ● 对于更新较少的表，选择较大fillfactor值。例如90~100，以提高存储空间利用率。	在表tpcds.ship_mode_t1上的SM_SHIP_MODE_SK字段上创建指定B-tree索引，并指定该索引的填充因子**FILLFACTOR**为70%。 这意味着每个数据页面在初次填充时只会填满70%。剩余的30%空间留给将来的插入和更新操作。 `CREATE INDEX ds_ship_mode_t1_index6 ON tpcds.ship_mode_t1 USING btree(SM_SHIP_MODE_SK) WITH (**FILLFACTOR** = 70);`

 版权所有 © 华为云计算技术有限公司

存储参数	描述	示例
INVISIBLE	控制优化器是否生成索引扫描相关计划。 **适用场景**：维护期间临时禁用索引，或测试索引对性能的影响。 **取值范围**： • ON表示不生成索引扫描相关计划。 • OFF表示生成索引扫描相关计划。 **默认值**：OFF	在表tpcds.ship_mode_t1上的SM_SHIP_MODE_SK字段上创建指定B-tree索引，并设置**优化器不生成该索引扫描相关计划** CREATE INDEX ds_ship_mode_t1_index7 ON tpcds.ship_mode_t1 USING btree(SM_SHIP_MODE_SK) WITH (**INVISIBLE** = ON);
FASTUPDATE	**仅GIN索引支持该参数**，GIN索引是否使用快速更新，当启用后，数据变更会暂存到待处理列表（Pending List）。 **取值范围**： • ON，表示使用快速更新。 • OFF，表示关闭快速更新。 **默认值**：ON	在表tpcds.ship_mode_t1上的SM_SHIP_MODE_SK字段上创建指定GIN索引，并打开快速更新参数**FASTUPDATE**，设置该索引Pending List容量的最大值为8MB。 CREATE INDEX ds_ship_mode_t1_index8 ON tpcds.ship_mode_t1 USING gin(SM_SHIP_MODE_SK) WITH (FASTUPDATE = ON, GIN_PENDING_LIST_LIMIT = 8192);
GIN_PENDING_LIST_LIMIT	**仅GIN索引支持该参数**，当GIN索引启用fastupdate时，设置该索引Pending List容量的最大值。 **Pending List**是GIN索引特有的数据结构，用于临时存储索引更新操作。当GIN索引启用FASTUPDATE参数时，新的索引项不会直接写入主索引结构，而是先存入Pending List，待满足特定条件时再批量合并到主索引中。 **取值范围**：64~INT_MAX，单位KB。 **默认值**：由GUC参数**gin_pending_list_limit**的值决定（默认为4MB）。	

表 13-114 唯一索引中索引列 NULL 值的处理方式

字段控制	索引列全为NULL	部分索引列为NULL
NULLS DISTINCT	可重复插入	可重复插入。
NULLS NOT DISTINCT	不可重复插入	非NULL值相等，不可插入；非NULL值不相等，则插入成功。

版权所有 © 华为云计算技术有限公司

字段控制	索引列全为NULL	部分索引列为NULL
NULLS IGNORE	可重复插入	非NULL值相等，不可插入；非NULL值不相等，则插入成功。

示例

- 创建示例表tpcds.ship_mode_t1：

```
DROP SCHEMA IF EXISTS tpcds CASCADE;
CREATE SCHEMA tpcds;
DROP TABLE IF EXISTS tpcds.ship_mode_t1;
CREATE TABLE tpcds.ship_mode_t1
(
   SM_SHIP_MODE_SK          INTEGER               NOT NULL,
   SM_SHIP_MODE_ID          CHAR(16)              NOT NULL,
   SM_TYPE                  CHAR(30)                      ,
   SM_CODE                  CHAR(10)                      ,
   SM_CARRIER               CHAR(20)                      ,
   SM_CONTRACT              CHAR(20)
)
DISTRIBUTE BY HASH(SM_SHIP_MODE_SK);
```

  在表tpcds.ship_mode_t1上的SM_SHIP_MODE_SK字段上创建唯一索引：

```
CREATE UNIQUE INDEX ds_ship_mode_t1_index1 ON tpcds.ship_mode_t1(SM_SHIP_MODE_SK);
```

  在表tpcds.ship_mode_t1上的SM_SHIP_MODE_SK字段上创建UNIQUE唯一索引，并指定NULL值的处理方式：

```
CREATE UNIQUE INDEX ds_ship_mode_t1_index5 ON tpcds.ship_mode_t1(SM_SHIP_MODE_SK)
NULLS NOT DISTINCT;
```

  在表tpcds.ship_mode_t1上的SM_SHIP_MODE_SK字段上创建索引时添加索引的注释。

```
CREATE INDEX ds_ship_mode_t1_index_comment ON tpcds.ship_mode_t1(SM_SHIP_MODE_SK)
COMMENT 'index';
```

  在表tpcds.ship_mode_t1上的SM_SHIP_MODE_SK字段上创建指定B-tree索引。

```
CREATE INDEX ds_ship_mode_t1_index4 ON tpcds.ship_mode_t1 USING btree(SM_SHIP_MODE_SK);
```

  在表tpcds.ship_mode_t1上的SM_SHIP_MODE_SK字段上创建指定B-tree索引，并指定该索引的填充因子**FILLFACTOR**为70%。

```
CREATE INDEX ds_ship_mode_t1_index6 ON tpcds.ship_mode_t1 USING btree(SM_SHIP_MODE_SK)
WITH (FILLFACTOR = 70);
```

  在表tpcds.ship_mode_t1上的SM_SHIP_MODE_SK字段上创建指定B-tree索引，并设置**优化器不生成该索引扫描相关计划**(默认为OFF，表示生成，ON表示不生成)。

```
CREATE INDEX ds_ship_mode_t1_index7 ON tpcds.ship_mode_t1 USING btree(SM_SHIP_MODE_SK)
WITH (INVISIBLE = ON);
```

  在表tpcds.ship_mode_t1上的SM_SHIP_MODE_SK字段上创建指定GIN索引，并打开快速更新参数**FASTUPDATE**，设置该索引Pending List容量的最大值为8MB。

  在表tpcds.ship_mode_t1上SM_CODE字段上创建表达式索引。

```
CREATE INDEX ds_ship_mode_t1_index2 ON tpcds.ship_mode_t1(SUBSTR(SM_CODE,1 ,4));
```

  在表tpcds.ship_mode_t1上的SM_SHIP_MODE_SK字段上创建SM_SHIP_MODE_SK大于10的部分索引。

```
CREATE UNIQUE INDEX ds_ship_mode_t1_index3 ON tpcds.ship_mode_t1(SM_SHIP_MODE_SK)
WHERE SM_SHIP_MODE_SK>10;
```

- 创建示例表tpcds.customer_address_p1。

```
DROP TABLE IF EXISTS tpcds.customer_address_p1;
CREATE TABLE tpcds.customer_address_p1
```

 版权所有 © 华为云计算技术有限公司

```
(
    CA_ADDRESS_SK             INTEGER               NOT NULL,
    CA_ADDRESS_ID             CHAR(16)              NOT NULL,
    CA_STREET_NUMBER          CHAR(10)                      ,
    CA_STREET_NAME            VARCHAR(60)                   ,
    CA_STREET_TYPE            CHAR(15)                      ,
    CA_SUITE_NUMBER           CHAR(10)                      ,
    CA_CITY                   VARCHAR(60)                   ,
    CA_COUNTY                 VARCHAR(30)                   ,
    CA_STATE                  CHAR(2)                       ,
    CA_ZIP                    CHAR(10)                      ,
    CA_COUNTRY                VARCHAR(20)                   ,
    CA_GMT_OFFSET             DECIMAL(5,2)                  ,
    CA_LOCATION_TYPE          CHAR(20)
)
DISTRIBUTE BY HASH(CA_ADDRESS_SK)
PARTITION BY RANGE(CA_ADDRESS_SK)
(
   PARTITION p1 VALUES LESS THAN (3000),
   PARTITION p2 VALUES LESS THAN (5000) ,
   PARTITION p3 VALUES LESS THAN (MAXVALUE)
)
ENABLE ROW MOVEMENT;
```

创建分区表索引ds_customer_address_p1_index1，不指定索引分区的名字。

```
CREATE INDEX ds_customer_address_p1_index1 ON tpcds.customer_address_p1(CA_ADDRESS_SK)
LOCAL;
```

创建分区表索引ds_customer_address_p1_index2，并指定索引分区的名字。

```
CREATE INDEX ds_customer_address_p1_index2 ON tpcds.customer_address_p1(CA_ADDRESS_SK)
LOCAL
(
    PARTITION CA_ADDRESS_SK_index1,
    PARTITION CA_ADDRESS_SK_index2,
    PARTITION CA_ADDRESS_SK_index3
)
;
```

创建分区表索引ds_customer_address_p1_index_comment，并添加索引注释。

```
CREATE INDEX ds_customer_address_p1_index_comment ON
tpcds.customer_address_p1(CA_ADDRESS_SK) COMMENT 'index' LOCAL
(
    PARTITION CA_ADDRESS_SK_index1,
    PARTITION CA_ADDRESS_SK_index2,
    PARTITION CA_ADDRESS_SK_index3
)
;
```

关于 NULLS DISTINCT、NULLS NOT DISTINCT、NULLS IGNORE 使用上的场景差异

假设有一张表users，包含id和email（可能为NULL，表示未提供邮箱）。

场景 1：默认 NULLS DISTINCT（NULL可重复插入）

```
-- 创建表
DROP TABLE IF EXISTS users;
CREATE TABLE users (
  id INT,
  email TEXT
)
DISTRIBUTE BY HASH(email);
-- 创建唯一索引（默认 NULLS DISTINCT）
CREATE UNIQUE INDEX idx_email_unique ON users (email);
-- 插入数据
INSERT INTO users (id, email) VALUES (1, NULL); -- 成功
INSERT INTO users (id, email) VALUES (2, NULL); -- 成功（NULL 被视为不同）
```

 版权所有 © 华为云计算技术有限公司

```
INSERT INTO users (id, email) VALUES (3, 'a@example.com'); -- 成功
INSERT INTO users (id, email) VALUES (4, 'a@example.com'); -- 失败（非 NULL 重复）
```

```
error_code: DWS.S0010010
error_msg: Constraint Violation. STATE: 23505, message: ERROR: dn_6005_6006: duplicate key value violates unique constraint "idx_email_unique"
Detail: Key (email)=(a@example.com) already exists.
```

场景 2：NULLS NOT DISTINCT（NULL不可重复插入）

```
-- 重新创建唯一索引，指定NULLS NOT DISTINCT
DROP TABLE IF EXISTS users;
CREATE TABLE users (
  id INT,
  email TEXT
)
DISTRIBUTE BY HASH(email);
CREATE UNIQUE INDEX idx_email_unique ON users (email) NULLS NOT DISTINCT;
-- 插入数据
INSERT INTO users (id, email) VALUES (5, NULL); -- 成功（第一个 NULL）
INSERT INTO users (id, email) VALUES (6, NULL); -- 失败（第二个 NULL，违反唯一约束）
INSERT INTO users (id, email) VALUES (7, 'b@example.com'); -- 成功
INSERT INTO users (id, email) VALUES (8, 'b@example.com'); -- 失败（非 NULL 重复）
```

场景 3：NULLS IGNORE（忽略NULL值）

```
-- 创建索引，忽略NULL值
DROP TABLE IF EXISTS users;
CREATE TABLE users (
  id INT,
  email TEXT
)
DISTRIBUTE BY HASH(email);
CREATE UNIQUE INDEX idx_email_ignore ON users (email) NULLS IGNORE;
-- 插入数据
INSERT INTO users (id, email) VALUES (9, NULL); -- 成功（第一个 NULL）
INSERT INTO users (id, email) VALUES (10, NULL); -- 成功（NULL值被忽略）
INSERT INTO users (id, email) VALUES (11, 'c@example.com'); -- 成功
INSERT INTO users (id, email) VALUES (12, 'c@example.com'); -- 失败（非 NULL 重复）
```

相关链接

- 修改和删除索引语法参见**ALTER INDEX**和**DROP INDEX**。
- 更多索引案例请参见**创建和管理索引**章节。
- 有关GIN索引的更多示例，请参见**GIN索引使用实践**。

常见问题

- **DWS唯一约束和唯一索引有什么区别**
- **重建索引失败**
- **修改索引只调用索引名提示索引不存在**
- **DWS元数据不一致-分区索引异常**
- **DWS查询时索引失效场景解析**

13.6.2 ALTER INDEX

功能描述

ALTER INDEX用于修改索引的定义。

 版权所有 © 华为云计算技术有限公司

注意事项

只有索引的所有者有权限执行此命令，系统管理员默认拥有此权限。

语法格式

- 重命名表索引的名称，支持新的索引名称前带有原索引所在的schema名，但是不支持同时修改schema名。
```
ALTER INDEX [ IF EXISTS ] index_name
    RENAME TO new_name;
ALTER INDEX [ IF EXISTS ] index_name
    RENAME TO schema.new_name;
```
- 修改表索引的存储参数。
```
ALTER INDEX [ IF EXISTS ] index_name
    SET ( {storage_parameter = value} [, ... ] );
```
- 修改索引的状态标记。
```
ALTER INDEX [ IF EXISTS ] index_name
    SET ( {invisible = value} [, ... ] );
```
- 重置表索引的存储参数。
```
ALTER INDEX [ IF EXISTS ] index_name
    RESET ( storage_parameter [, ... ] ) ;
```
- 设置表索引或索引分区不可用。
```
ALTER INDEX [ IF EXISTS ] index_name
    [ MODIFY PARTITION index_partition_name ] UNUSABLE;
```
- 重建表索引或索引分区。
```
ALTER INDEX index_name
    REBUILD [ PARTITION index_partition_name ] [WITHOUT UNUSABLE];
```
- 重命名索引分区。**PG_OBJECT**系统表记录索引的最后修改时间时，**暂不支持对该语法的记录**。
```
ALTER INDEX [ IF EXISTS ] index_name
    RENAME PARTITION index_partition_name TO new_index_partition_name;
```
- 添加，修改索引的注释。
```
ALTER INDEX [ IF EXISTS ] index_name
    COMMENT 'text';
```
- 删除索引的注释。
```
ALTER INDEX [ IF EXISTS ] index_name
    COMMENT '';
ALTER INDEX [ IF EXISTS ] index_name
    COMMENT NULL;
```

参数说明

表 13-115 ALTER INDEX 参数说明

参数	描述	取值范围
IF EXISTS	如果指定的索引不存在时，返回一个提示而不是错误。	-
RENAME TO	修改索引的名称，对存储的数据本身没有任何影响。	-

版权所有 © 华为云计算技术有限公司

参数	描述	取值范围
SET ({ STORAGE_PARAMETER = value } [, ...])	修改索引的一个或多个索引方法特定的存储参数。 此命令不会立即修改索引内容，根据参数的不同，可能需要使用REINDEX重建索引来获得所需的效果。	取值范围和详细用法参见**表13-116**。
RESET ({ storage_parameter } [, ...])	重置索引的一个或多个索引方法特定的存储参数为默认值。与SET一样，可能需要使用REINDEX来完全更新索引。	-
[MODIFY PARTITION index_partition_name] UNUSABLE	用于设置表或者索引分区上的索引不可用。	-
REBUILD [PARTITION index_partition_name]	用于重建表或者索引分区上的索引。	-
WITHOUT UNUSABLE	重建表或者索引分区上的索引时不重建状态为UNUSABLE的索引。	-
RENAME PARTITION	用于重命名索引分区。	-
COMMENT comment_text	用于添加，修改或删除索引的注释。	-
index_name	需要修改的索引名。	已存在的索引名。
new_name	新的索引名。	字符串，需符合**标识符命名规范**。
storage_parameter	索引方法特定的参数名。	-
invisible	控制优化器是否生成索引扫描相关计划。默认值：OFF。	ON或OFF。 • ON表示不生成索引扫描相关计划。 • OFF表示生成索引扫描相关计划。
value	索引方法特定的存储参数的新值。根据参数的不同，这可能是一个数字或单词。	-
new_index_partition_name	新索引分区名。	字符串，需符合**标识符命名规范**。
index_partition_name	索引分区名。	字符串，需符合**标识符命名规范**。
comment_text	索引的注释信息。	-

版权所有 © 华为云计算技术有限公司

表 13-116 WITH ({storage_parameter = value} [, ...])指定索引方法的存储参数

存储参数	描述	示例
FILLFACTOR	**GIN和PSORT索引不支持该参数**，表示一个索引的填充因子（ fillfactor ），取值范围是一个介于10和100之间的百分比值，用于指定索引页初始填充的百分比，剩余的空间可以用于后续的插入和更新操作，从而减少页面分裂和提高性能。 该参数的作用是控制索引页的填充程度，从而影响数据库的性能和存储空间的利用效率，通过调整FILLFACTOR的值，可以减少后续更新或插入操作导致的页分裂，优化索引的存储和性能。 **FILLFACTOR对索引性能的影响**： ● **性能**：当索引页填满后，新的数据插入会导致页面分裂，这会消耗更多的I/O资源和时间。较低的 fillfactor可以减少页面分裂，提高插入和更新性能，但会增加存储空间的使用。较高的fillfactor可以提高存储空间的利用率，但可能会导致更多的页面分裂，降低插入和更新性能。 ● **存储空间**：较低的fillfactor会增加每个数据页面的空闲空间，导致更多的磁盘空间使用。较高的fillfactor会更紧凑地存储数据，减少磁盘空间使用。 **如何选择合适的FILLFACTOR值**： ● 对于更新频繁的表，选择较小fillfactor值如60~80，预留空间，减少页分裂。 ● 对于更新较少的表，选择较大fillfactor值如90~100，以提高存储空间利用率。	修改索引的填充因子**FILLFACTOR**为60%。 `ALTER INDEX tpcds.ds_ship_mode_t1_index7 SET (**FILLFACTOR** = 60);`

 版权所有 © 华为云计算技术有限公司

存储参数	描述	示例
INVISIBLE	控制优化器是否生成索引扫描相关计划。 **适用场景**：维护期间临时禁用索引，或测试索引对性能的影响。 **取值范围**： • ON表示不生成索引扫描相关计划。 • OFF表示生成索引扫描相关计划。 **默认值**：OFF	设置**优化器不生成索引的扫描相关计划。** ALTER INDEX tpcds.ds_ship_mode_t1_index9 SET (INVISIBLE = ON);
FASTUPDATE	**仅GIN索引支持该参数**，GIN索引是否使用快速更新，当启用后，数据变更会暂存到待处理列表（Pending List）。 **取值范围**：ON，OFF **默认值**：ON	打开快速更新参数**FASTUPDATE**，设置该索引Pending List容量的最大值为8MB。 ALTER INDEX tpcds.ds_ship_mode_t1_index10 SET (FASTUPDATE = ON, GIN_PENDING_LIST_LIMIT = 8192);
GIN_PENDING_LIST_LIMIT	**仅GIN索引支持该参数**，当GIN索引启用fastupdate时，设置该索引Pending List容量的最大值。 **Pending List**是GIN索引特有的数据结构，用于临时存储索引更新操作。当GIN索引启用FASTUPDATE参数时，新的索引项不会直接写入主索引结构，而是先存入Pending List，待满足特定条件时再批量合并到主索引中。 **取值范围**：64~INT_MAX，单位KB。 **默认值**：由GUC参数**gin_pending_list_limit**的值决定（默认为4MB）。	

示例

创建示例表tpcds.ship_mode_t1。

```
DROP SCHEMA IF EXISTS tpcds CASCADE;
CREATE SCHEMA tpcds;
DROP TABLE IF EXISTS tpcds.ship_mode_t1;
CREATE TABLE tpcds.ship_mode_t1
(
    SM_SHIP_MODE_SK          INTEGER               NOT NULL,
    SM_SHIP_MODE_ID          CHAR(16)              NOT NULL,
    SM_TYPE                  CHAR(30)                      ,
    SM_CODE                  CHAR(10)                      ,
    SM_CARRIER               CHAR(20)                      ,
    SM_CONTRACT              CHAR(20)
)
DISTRIBUTE BY HASH(SM_SHIP_MODE_SK);
```


版权所有 © 华为云计算技术有限公司

在表tpcds.ship_mode_t1上的SM_SHIP_MODE_SK字段上创建唯一索引。

```
DROP INDEX IF EXISTS ds_ship_mode_t1_index1;
CREATE UNIQUE INDEX ds_ship_mode_t1_index1 ON tpcds.ship_mode_t1(SM_SHIP_MODE_SK);
```

在表tpcds.ship_mode_t1上SM_CODE字段上创建表达式索引。

```
DROP INDEX IF EXISTS ds_ship_mode_t1_index2;
CREATE INDEX ds_ship_mode_t1_index2 ON tpcds.ship_mode_t1(SUBSTR(SM_CODE,1 ,4));
```

重命名现有的索引ds_ship_mode_t1_index1为ds_ship_mode_t1_index5。

```
ALTER INDEX tpcds.ds_ship_mode_t1_index1 RENAME TO ds_ship_mode_t1_index5;
```

设置索引ds_ship_mode_t1_index2不可用。

```
ALTER INDEX tpcds.ds_ship_mode_t1_index2 UNUSABLE;
```

重建索引ds_ship_mode_t1_index2。

```
ALTER INDEX tpcds.ds_ship_mode_t1_index2 REBUILD;
```

相关链接

CREATE INDEX，**DROP INDEX**，**REINDEX**

13.6.3 REINDEX

功能描述

为表中的数据重建索引。

在以下几种情况下需要使用REINDEX重建索引：

- 索引崩溃，并且不再包含有效的数据。
- 索引变得“臃肿”，包含大量的空页或接近空页。
- 为索引更改了存储参数（例如填充因子），并且希望这个更改完全生效。
  使用CONCURRENTLY选项创建索引失败，留下了一个“非法”索引。

注意事项

- REINDEX DATABASE和SYSTEM这种形式的重建索引不能在事务块中执行。
- 存算分离架构下，不支持REINDEX全库索引。
- 在GUC参数enable_consistent_oid为on时，对同一张分区表进行CREATE INDEX与REINDEX TABLE并发操作时，低概率会导致不同节点间分区索引的数量不一致，进而导致REINDEX TABLE报错。出现此情况时，重试执行REINDEX TABLE语句即可成功。
- 使用REINDEX INDEX PARTITION重建索引分区，会在对应的索引分区对象上增加排他锁。在重建过程中，阻止使用该索引分区的读写操作（如索引扫描），不使用该索引分区的读写操作不受影响。

⚠ 警告

- 避免在业务高峰期执行对大表执行CREATE INDEX和REINDEX操作。
- 更多开发设计规范参见**总体开发设计规范**。

版权所有 © 华为云计算技术有限公司

语法格式

- 重建全库索引或者系统表的索引。
  ```
  REINDEX { DATABASE | SYSTEM } name [ FORCE ];
  ```
- 重建普通索引或者普通表上的索引。
  ```
  REINDEX { INDEX |  TABLE  } name [ FORCE | WITHOUT UNUSABLE ]
  ```
- 重建索引分区或者分区上的索引。
  ```
  REINDEX  { |   TABLE} name
     PARTITION partition_name [ FORCE | WITHOUT UNUSABLE ];
  ```

参数说明

表 13-117 REINDEX 参数说明

参数	描述	取值范围
INDEX	重新建立指定的索引。	-
TABLE	重新建立指定表的所有索引，如果表有从属的“TOAST”表，则这个表也会重建索引。	-
DATABASE	重建当前数据库里的所有索引。	-
SYSTEM	在当前数据库上重建所有系统表上的索引。不会处理在用户表上的索引。	-
name	需要重建索引的索引、表、数据库的名称。表和索引可以有模式修饰。 **说明** REINDEX DATABASE和SYSTEM只能重建当前数据库的索引，所以name必须和当前数据库名称相同。	-
FORCE	无效选项，会被忽略。	-
WITHOUT UNUSABLE	指定该参数时，在REINDEX时会忽略对象上的UNUSABLE索引或者UNUSABLE索引分区。该参数仅9.1.0.100及以上版本支持。	-
partition_name	需要重建索引的分区的名字或者索引分区的名字。	• 如果前面是REINDEX INDEX，则这里应该指定索引分区的名字。 • 如果前面是REINDEX TABLE，则这里应该指定分区的名字。

须知

REINDEX DATABASE和SYSTEM这种形式的重建索引不能在事务块中执行。

版权所有 © 华为云计算技术有限公司

示例

重建一个单独索引：

```
DROP TABLE IF EXISTS customer;
CREATE TABLE customer (a int, b int) DISTRIBUTE BY HASH(a);
CREATE  UNIQUE INDEX customer_index1 ON customer (a);
REINDEX INDEX customer_index1;
```

重建表customer上的所有索引：

```
DROP TABLE IF EXISTS customer;
CREATE TABLE customer (a int, b int) DISTRIBUTE BY HASH(a);
CREATE UNIQUE INDEX customer_index1 ON customer (a);
CREATE INDEX customer_index2 ON customer (b);
REINDEX TABLE customer;
```

重建分区表customer_address分区P1的分区索引：

```
DROP TABLE IF EXISTS customer_address;
CREATE TABLE customer_address
(
    ca_address_sk       INTEGER                 NOT NULL   ,
    ca_address_id       CHARACTER(16)           NOT NULL   ,
    ca_street_number    CHARACTER(10)                      ,
    ca_street_name      CHARACTER varying(60)              ,
    ca_street_type      CHARACTER(15)                     ,
    ca_suite_number     CHARACTER(10)
)
DISTRIBUTE BY HASH (ca_address_sk)
PARTITION BY RANGE(ca_address_sk)
(
        PARTITION P1 VALUES LESS THAN(2450815),
        PARTITION P2 VALUES LESS THAN(2451179),
        PARTITION P3 VALUES LESS THAN(2451544),
        PARTITION P4 VALUES LESS THAN(MAXVALUE)
);

CREATE INDEX customer_address_index on customer_address(CA_ADDRESS_SK) LOCAL;

REINDEX TABLE customer_address PARTITION P1;
```

13.6.4 DROP INDEX

功能描述

删除索引。

注意事项

只有索引的所有者有权限执行DROP INDEX命令，系统管理员默认拥有此权限。

语法格式

```
DROP INDEX [ CONCURRENTLY ][ IF EXISTS ]
    index_name [, ...] [ CASCADE | RESTRICT ];
```

 版权所有 © 华为云计算技术有限公司

参数说明

表 13-118 DROP INDEX 参数说明

参数	描述	取值范围
CONCURRENTLY	删除索引而不锁定索引表上的并发选择、插入、更新和删除。 普通的DROP INDEX在表上获取排他锁，从而阻止其他访问，直到可以完成索引删除为止。 使用此选项，命令将一直等到冲突的事务完成。 **说明** 使用此选项时需要注意：只能指定一个索引名称，并且不支持CASCADE选项（因此，不能以这种方式删除支持UNIQUE或PRIMARY KEY约束的索引）。可以在事务块内执行常规的DROP INDEX命令，但不能以DROP INDEX CONCURRENTLY方式执行。	-
IF EXISTS	如果指定的索引不存在，则发出一个notice而不是抛出一个错误。	-
index_name	要删除的索引名。	已存在的索引。
CASCADE | RESTRICT	当执行删除操作时，如何处理依赖对象中的相关数据。	• CASCADE：表示允许级联删除依赖于该索引的对象。 • RESTRICT（缺省值）：表示有依赖此索引的对象存在，则该索引无法被删除。

示例

删除现有的索引ds_ship_mode_t1_index2：

```
DROP INDEX IF EXISTS tpcds.ds_ship_mode_t1_index2;
```

相关链接

ALTER INDEX，**CREATE INDEX**

13.6.5 CREATE SEQUENCE

功能描述

向当前数据库里增加一个新的序列。

序列是一个数据库对象，主要用于生成一系列唯一的、递增或递减的数值，通常用作表中的主键或唯一标识符。

版权所有 © 华为云计算技术有限公司

DWS的9.1.1.300及以上集群版本中新增GUC参数enable_multi_sequence_file（用于控制是否开启GTM中的gtm.sequence文件拆分）。开启该参数后稳定支持百万级别并发序列对象 (SEQUENCE) 的创建、管理与高性能访问，可满足超大规模应用需求。

注意事项

- 创建序列的用户为该序列的所有者。
- 如果指定了模式名称，则会在指定模式中创建序列，否则会在当前模式中创建。序列名必须和同一个模式中的其他序列、表、索引、视图或外表的名字不同。
- 创建序列后，在表中使用序列的nextval()函数和generate_series(1,N)函数对表插入数据，请保证nextval的可调用次数大于等于N+1次，否则会因为generate_series()函数会调用N+1次而导致报错。
- 不支持在template1数据库中创建SEQUENCE。
- 建议控制SEQUENCE的使用数量（含SERIAL类型），因分布式场景下SEQUENCE中序列值的产生依赖GTM全局管理，频繁的与GTM交互会严重影响性能。建议引用公共的SEQUENCE对象，或者针对需要唯一值的场景使用UUID进行替换。同时为尽量减少与GTM的交互，针对手动创建的SEQUENCE，建议指定CACHE值不小于10000。

语法格式

```
CREATE SEQUENCE [ IF NOT EXISTS ] name [ INCREMENT [ BY ] increment ]
    [ MINVALUE minvalue | NO MINVALUE | NOMINVALUE ] [ MAXVALUE maxvalue | NO MAXVALUE | NOMAXVALUE]
    [ START [ WITH ] start ] [ CACHE cache ] [ [ NO ] CYCLE | NOCYCLE ]
    [ OWNED BY { table_name.column_name | NONE } ];
```

参数说明

表 13-119 CREATE SEQUENCE 参数说明

参数	描述	取值范围
IF NOT EXISTS	指定IF NOT EXISTS时，若不存在同名SEQUENCE，则可以成功创建SEQUENCE。若已存在同名SEQUENCE，创建时不会报错，仅会提示该SEQUENCE已存在，且不执行任何操作。 该参数仅9.1.0及以上版本支持。	-
name	将要创建的序列名称。	字符串，需符合**标识符命名规范**。
increment	指定序列的步长。正值将生成一个递增的序列，负值将生成一个递减的序列。 默认值为1。	-

版权所有 © 华为云计算技术有限公司

参数	描述	取值范围
MINVALUE minvalue | NO MINVALUE| NOMINVALUE	确定序列可以生成的最小值。如果没有指定minvalue或者指定了NO MINVALUE，则递增序列的缺省值为1，递减序列的缺省值为-2^{63}-1。 NOMINVALUE等价于NO MINVALUE。	-
MAXVALUE maxvalue | NO MAXVALUE| NOMAXVALUE	确定序列可以生成的最大值。如果没有声明maxvalue或者声明了NO MAXVALUE，则递增序列的缺省值为2^{63}-1，递减序列的缺省值为-1。 NOMAXVALUE等价于NO MAXVALUE。	-
start	指定序列的起始值。 默认值：对于递增序列为minvalue，递减序列为maxvalue。	-
cache	指定数据库预先缓存（预分配）多少个序列号到内存中，从而加快访问速度。一个缓存周期内，CN不再向GTM索取序列号，而是使用本地预先申请的序列号。 9.1.0.100及以上集群版本中缺省值由GUC参数default_sequence_cache控制。 ● 新安装的9.1.0.100及以上集群版本中，默认值为20。 ● 从低版本升级到9.1.0.100及以上集群版本，默认值为1。表示一次只能生成一个值，即没有缓存。	● 不建议同时定义cache和maxvalue或minvalue。因为定义cache后不能保证序列的连续性，可能会产生空洞，造成序列号段浪费。 ● 建议cache值不要设置过大，否则会出现缓存序列号时（每个cache周期的第一个nextval）耗时过长的情况；同时建议cache值小于100000000。实际使用时应根据业务设置合理的cache值，既能保证快速访问，又不会浪费序列号。

版权所有 © 华为云计算技术有限公司

参数	描述	取值范围
CYCLE | NOCYCLE	当序列达到maxvalue或者minvalue时，允许序列可循环重新开始。 默认值为NO CYCLE，表示序列达到极大或极小值后将无法生成新值。NOCYCLE的作用等价于NO CYCLE。 ● 如果指定序列为CYCLE，则不能保证序列的唯一性。 ● 如果指定了NO CYCLE，则在序列达到其最大值后任何对nextval的调用都将返回错误。	-
OWNED BY	将序列与特定表的指定字段相关联。这样，如果该字段或其所在表被删除，该序列也将被自动删除。关联的表和序列的所有者必须是同一个用户，并且在同一个模式中。 缺省值为OWNED BY NONE，表示不存在这种关联。	● 通过指定OWNED BY，仅建立了表的对应列和SEQUENCE之间关联关系，并不会在插入数据时在该列上产生自增序列。 ● 通过指定OWNED BY创建的SEQUENCE不建议用于其他表，如果希望多个表共享SEQUENCE，该SEQUENCE不应该从属于特定表。

示例

创建一个名为serial的递增序列，从101开始。

```
DROP SEQUENCE IF EXISTS  serial;
CREATE SEQUENCE serial
 START 101
 CACHE 20;
```

从序列中选出下一个数字。

```
SELECT nextval('serial');
```

```
gaussdb=> SELECT nextval('serial');
 nextval
---------
     101
(1 row)
```

从序列中选出下一个数字。

```
SELECT nextval('serial');
```

 版权所有 © 华为云计算技术有限公司

```
gaussdb=> SELECT nextval('serial');
 nextval
---------
     102
(1 row)
```

创建与表关联的序列。

```
DROP TABLE IF EXISTS customer_address;
CREATE TABLE customer_address
(
    ca_address_sk             integer               not null,
    ca_address_id             char(16)              not null,
    ca_street_number          char(10)                      ,
    ca_street_name            varchar(60)                   ,
    ca_street_type            char(15)                      ,
    ca_suite_number           char(10)                      ,
    ca_city                   varchar(60)                   ,
    ca_county                 varchar(30)                   ,
    ca_state                  char(2)                       ,
    ca_zip                    char(10)                      ,
    ca_country                varchar(20)                   ,
    ca_gmt_offset             decimal(5,2)                  ,
    ca_location_type          char(20)
) ;

DROP SEQUENCE IF EXISTS serial1;
CREATE SEQUENCE serial1
 START 101
 CACHE 20
OWNED BY customer_address.ca_address_sk;
```

使用serial创建主键自增序列表serial_table。

```
DROP TABLE IF EXISTS serial_table;
CREATE TABLE serial_table(a int, b serial);
INSERT INTO serial_table (a) VALUES (1),(2),(3);
SELECT * FROM serial_table ORDER BY b;
```

```
gaussdb=> SELECT * FROM serial_table ORDER BY b;
 a | b
---+---
 1 | 1
 2 | 2
 3 | 3
(3 rows)
```

相关链接

ALTER SEQUENCE，**DROP SEQUENCE**

13.6.6 ALTER SEQUENCE

功能描述

修改序列定义。

注意事项

- 使用ALTER SEQUENCE的用户必须是该序列的所有者。

 版权所有 © 华为云计算技术有限公司

- 用户要修改序列的所有者，必须是新角色的直接或者间接成员，并且所有者角色必须有序列所在模式上的CREATE权限。
- 当前版本仅支持修改拥有者、归属列和最大值。若要修改其他参数，可以删除重建，并用Setval函数恢复当前值。
- ALTER SEQUENCE MAXVALUE不支持在事务、函数和存储过程中使用。
- 修改序列的最大值后，会清空该序列在所有会话的cache。
- ALTER SEQUENCE会阻塞nextval、setval、currval和lastval的调用。

语法格式

修改序列最大值或归属列。

```
ALTER SEQUENCE [ IF EXISTS ] name
    [ MAXVALUE maxvalue | NO MAXVALUE | NOMAXVALUE ]
    [ CACHE cache ]
    [ OWNED BY { table_name.column_name | NONE } ] ;
```

修改序列的所有者。

```
ALTER SEQUENCE [ IF EXISTS ] name OWNER TO new_owner;
```

参数说明

表 13-120 ALTER SEQUENCE 参数说明

参数	描述	取值范围
IF EXISTS	如果序列不存在时返回一个提示而不是错误。	-
name	需要修改的序列的名称。	已存在的序列名称。
MAXVALUE maxvalue | NO MAXVALUE	序列所能达到的最大值。如果指定了NO MAXVALUE，则递增序列的缺省值为2^{63}-1，递减序列的缺省值为-1。NOMAXVALUE等价于NO MAXVALUE。	-
CACHE cache	指定要预先分配多少序列号并存储在内存中，以加快访问速度。该参数仅8.2.1.100及以上集群版本支持。	• cachevalue最小值为 1（一次只能生成一个值，即NOCACHE）。 • 如果未指定，则保留旧缓存值。

 版权所有 © 华为云计算技术有限公司

参数	描述	取值范围
OWNED BY	将序列和表的指定字段进行关联。这样，在删除指定字段或其所在表的时候会自动删除已关联的序列。 如果序列已经和表有关联后，使用OWNED BY参数后新的关联关系会覆盖旧的关联。关联的表和序列的所有者必须是同一个用户，并且在同一个模式中。使用OWNED BY NONE将删除任何已经存在的关联。	-
new_owner	序列新所有者的用户名。	已存在的用户名。

示例

创建示例序列seq_test和示例表t1。

```
CREATE SEQUENCE seq_test START 101;
DROP TABLE IF EXISTS t1;
CREATE TABLE t1(c1 bigint default nextval('seq_test'));
```

将序列serial的最大值修改为200。

```
ALTER SEQUENCE seq_test MAXVALUE 200;
```

将序列seq_test的归属列变为t1.c1。

```
ALTER SEQUENCE seq_test OWNED BY t1.c1;
```

相关链接

CREATE SEQUENCE，**DROP SEQUENCE**

13.6.7 DROP SEQUENCE

功能描述

从当前数据库里删除序列。

注意事项

只有序列的所有者或者系统管理员才能删除。

语法格式

```
DROP SEQUENCE [ IF EXISTS ] {[schema.]sequence_name} [ , ... ] [ CASCADE | RESTRICT ];
```


版权所有 © 华为云计算技术有限公司

参数说明

表 13-121 DROP SEQUENCE 参数说明

参数	描述	取值范围
IF EXISTS	如果指定的序列不存在，则发出一个notice而不是抛出一个错误。	-
name	要删除的序列名称。	已存在的序列名称。
CASCADE | RESTRICT	当执行删除操作时，如何处理依赖对象中的相关数据。	• CASCADE：级联删除依赖序列的对象。 • RESTRICT：如果存在任何依赖的对象，则拒绝删除序列（缺省行为）。

示例

删除序列serial：

```
DROP SEQUENCE IF EXISTS serial;
```

相关链接

CREATE SEQUENCE，**ALTER SEQUENCE**

13.6.8 CREATE TEXT SEARCH CONFIGURATION

功能描述

创建新的文本搜索配置。

文本搜索配置通过文本搜索解析器（PARSER）将原始字符串解析为标记，再使用词典（DICTIONARY）决定哪些标记是搜索感兴趣的，从而满足搜索需求，提升全文搜索效率。

注意事项

- 如果仅指定了解析器，则新的文本搜索配置初始没有从标记类型到词典的映射，因此会忽略所有的单词。后面必须使用ALTER TEXT SEARCH CONFIGURATION命令创建映射使配置生效。如果指定了COPY选项，则会自动拷贝指定的文本搜索配置的解析器、映射、配置选项等信息。
- 如果指定了模式名称，则会在指定模式中创建文本搜索配置，否则会在当前模式中创建。
- 定义文本搜索配置的用户为其所有者。
- PARSER和COPY选项是互斥的，因为复制现有配置时，其解析器配置也会被复制。

版权所有 © 华为云计算技术有限公司

语法格式

```
CREATE TEXT SEARCH CONFIGURATION name
    ( PARSER = parser_name | COPY = source_config )
    [ WITH ( {configuration_option = value} [, ...] )];
```

参数说明

表 13-122 CREATE TEXT SEARCH CONFIGURATION 参数说明

参数	描述	取值范围
name	要创建的文本搜索配置的名称，可以用模式修饰。	字符串，需符合**标识符命名规范**。
parser_name	用于该配置的文本搜索解析器的名称。	目前支持default、ngram、zhparser三种类型的解析器。 ● default，默认解析器，可参考**解析器**。 ● ngram，基于n-gram分词算法，将文本按固定长度的字符或单词分割为标记。适用于灵活的分词粒度（如短文本、模糊搜索）或多语言或非结构化文本搜索（如日志分析、短文本匹配）的场景。 ● zhparser，基于词典的语义分词方法，将中文文本按语义或语言规则拆分为标记。底层调用SCWS（简易中文分词系统）分词算法，适用于有语义的中文分词场景。
source_config	要复制的现有文本搜索配置的名称。	-
configuration_option	文本搜索配置的配置参数。为指定解析器（parser_name解析器或source_config隐含解析器）配置参数，优化分词逻辑。	● default解析器无自定义配置参数configuration_option。 ● ngram和zhparser解析器对应的配置参数参考**表13-123**。

版权所有 © 华为云计算技术有限公司

表 13-123 ngram、zhparser 类型解析器对应的配置参数

解析器	配置参数	参数描述	取值范围
ngram	gram_size	分词长度。	正整数，1~4。 默认值：2。
	punctuation_ignore	是否忽略标点符号。	• true（默认值）：忽略标点符号。 • false：不忽略标点符号。
	grapsymbol_ignore	是否忽略图形化字符。	• true：忽略图形化字符。 • false（默认值）：不忽略图形化字符。
zhparser	punctuation_ignore	分词结果是否忽略所有的标点等特殊符号（不会忽略\r和\n）。	• true（默认值）：忽略所有的标点等特殊符号。 • false：不忽略所有的标点等特殊符号。
	seg_with_duality	是否将闲散文字自动以二字分词法聚合。	• true：将闲散文字自动以二字分词法聚合。 • false（默认值）：不将闲散文字自动以二字分词法聚合。
	multi_short	分词执行时是否执行针对长词复合切分。	• true（默认值）：执行针对长词复合切分。 • false：不执行针对长词复合切分。
	multi_duality	设定是否将长词内的文字自动以二字分词法聚合。	• true：将长词内的文字自动以二字分词法聚合。 • false（默认值）：不将长词内的文字自动以二字分词法聚合。
	multi_zmain	是否将重要单字单独显示。	• true：将重要单字单独显示。 • false（默认值）：不将重要单字单独显示。
	multi_zall	是否将全部单字单独显示。	• true：将全部单字单独显示。 • false（默认值）：不将全部单字单独显示。

版权所有 © 华为云计算技术有限公司

示例

创建名为ngram1的文本搜索配置，使用ngram解析器，分词长度为两个字符，保留文本中的图形化字符。

```
DROP TEXT SEARCH CONFIGURATION IF EXISTS  ngram1;
CREATE TEXT SEARCH CONFIGURATION ngram1 (parser=ngram) WITH (gram_size = 2, grapsymbol_ignore = false);
```

创建名为ngram2的文本搜索配置，复制现有配置ngram1的所有参数。

```
DROP TEXT SEARCH CONFIGURATION IF EXISTS  ngram2;
CREATE TEXT SEARCH CONFIGURATION ngram2 (copy=ngram1);
```

创建名为english_1的文本搜索配置，使用默认的文本解析器（parser为default）。

```
DROP TEXT SEARCH CONFIGURATION IF EXISTS english_1;
CREATE TEXT SEARCH CONFIGURATION english_1 (parser=default);
```

相关链接

ALTER TEXT SEARCH CONFIGURATION，**DROP TEXT SEARCH CONFIGURATION**，**全文检索**

13.6.9 ALTER TEXT SEARCH CONFIGURATION

功能描述

修改文本搜索配置的定义。

用户可以将映射从标记类型调整为字典，更改配置的名称或者所有者，或者修改搜索配置的配置参数。

注意事项

- 当一个搜索配置已经被引用（如被用来创建索引），则不允许用户修改此文本搜索配置。
- 要使用ALTER TEXT SEARCH CONFIGURATION，用户必须是配置的所有者。

语法格式

- 增加文本搜索配置标记类型映射。
  ```
  ALTER TEXT SEARCH CONFIGURATION name
      ADD MAPPING FOR token_type [, ... ] WITH dictionary_name [, ... ];
  ```
- 修改文本搜索配置中指定标记类型的字典。
  ```
  ALTER TEXT SEARCH CONFIGURATION name
      ALTER MAPPING FOR token_type [, ... ] REPLACE old_dictionary WITH new_dictionary;
  ```
- 修改文本搜索配置标记类型。
  ```
  ALTER TEXT SEARCH CONFIGURATION name
      ALTER MAPPING FOR token_type [, ... ] WITH dictionary_name [, ... ];
  ```
- 修改文本搜索配置字典。
  ```
  ALTER TEXT SEARCH CONFIGURATION name
      ALTER MAPPING REPLACE old_dictionary WITH new_dictionary;
  ```
- 删除文本搜索配置标记类型映射。
  ```
  ALTER TEXT SEARCH CONFIGURATION name
      DROP MAPPING [ IF EXISTS ] FOR token_type [, ... ];
  ```
- 重命名文本搜索配置所有者。

 版权所有 © 华为云计算技术有限公司

```
ALTER TEXT SEARCH CONFIGURATION name OWNER TO new_owner;
```

- 重命名文本搜索配置名称。
```
ALTER TEXT SEARCH CONFIGURATION name RENAME TO new_name;
```
- 重命名文本搜索配置命名空间。
```
ALTER TEXT SEARCH CONFIGURATION name SET SCHEMA new_schema;
```
- 修改文本搜索配置属性。
```
ALTER TEXT SEARCH CONFIGURATION name SET ( { configuration_option = value } [, ...] );
```
- 重置文本搜索配置属性。
```
ALTER TEXT SEARCH CONFIGURATION name RESET ( {configuration_option} [, ...] );
```

说明

- ADD MAPPING FOR选项为文本搜索配置增加标记类型映射；如果ADD MAPPING FOR后面任何一个标记类型的映射已经存在于此文本搜索配置中，那么系统将会报错。
- ALTER MAPPING FOR选项会首先清除已有的标记类型映射，然后添加指定的标记类型映射。
- ALTER MAPPING REPLACE ... WITH ... 与ALTER MAPPING FOR ... REPLACE ... WITH ...选项会直接使用new_dictionary替换old_dictionary。需要注意的是，只有pg_ts_config_map系统表中存在maptokentype与old_dictionary对应关系的元组时，才能更新成功，否则不会成功，也不会有任何提示信息返回。
- DROP MAPPING FOR选项会删除当前文本搜索配置中指定的标记类型映射。 如果没有指定IF EXISTS选项，当DROP MAPPING FOR选项指定的标记类型映射在文本搜索配置中不存在时，数据库会报错。

参数说明

表 13-124 ALTER TEXT SEARCH CONFIGURATION 参数说明

参数	描述	取值范围
name	需要修改的文本搜索配置的名称，可以用模式修饰。	-
token_type	与配置的语法解析器关联的标记类型的名称。	详细信息参见**解析器**。
dictionary_name	文本搜索字典名称。 如果有多个字典，则它们会按指定的顺序搜索。	-
old_dictionary	映射中要替换的文本搜索字典名称。	-
new_dictionary	替换old_dictionary的文本搜索字典的名称。	-
new_owner	文本搜索配置的新所有者。	有效的用户名。
new_name	文本搜索配置的新名称。	字符串，需符合**标识符命名规范**。
new_schema	文本搜索配置的新模式名。	有效的模式名。
configuration_option	文本搜索配置项。	详细信息参见**ALTER TEXT SEARCH CONFIGURATION**。

版权所有 © 华为云计算技术有限公司

参数	描述	取值范围
value	文本搜索配置项的值。	-

示例

创建文本搜索配置ngram1。

```
DROP TEXT SEARCH CONFIGURATION IF EXISTS ngram1;
CREATE TEXT SEARCH CONFIGURATION ngram1 (parser=ngram) WITH (gram_size = 2, grapsymbol_ignore = false);
```

给文本搜索类型ngram1添加类型映射。

```
ALTER TEXT SEARCH CONFIGURATION ngram1 ADD MAPPING FOR multisymbol WITH simple;
```

修改文本搜索配置的所有者。

```
CREATE USER joe password '{Password}';
ALTER TEXT SEARCH CONFIGURATION ngram1 OWNER TO joe;
```

修改文本搜索配置的schema。

```
ALTER TEXT SEARCH CONFIGURATION ngram1 SET SCHEMA joe;
```

重命名文本搜索配置。

```
ALTER TEXT SEARCH CONFIGURATION joe.ngram1 RENAME TO ngram_1;
```

删除类型映射。

```
ALTER TEXT SEARCH CONFIGURATION joe.ngram_1 DROP MAPPING IF EXISTS FOR multisymbol;
```

创建文本搜索配置english_1。

```
DROP TEXT SEARCH CONFIGURATION IF EXISTS english_1;
CREATE TEXT SEARCH CONFIGURATION english_1 (parser=default);
```

增加文本搜索配置标记类型映射。

```
ALTER TEXT SEARCH CONFIGURATION english_1 ADD MAPPING FOR word WITH simple,english_stem;
```

增加文本搜索配置标记类型映射。

```
ALTER TEXT SEARCH CONFIGURATION english_1 ADD MAPPING FOR email WITH english_stem, french_stem;
```

修改文本搜索配置标记类型映射。

```
ALTER TEXT SEARCH CONFIGURATION english_1 ALTER MAPPING REPLACE french_stem with german_stem;
```

查询文本搜索配置相关信息。

```
SELECT b.cfgname,a.maptokentype,a.mapseqno,a.mapdict,c.dictname FROM pg_ts_config_map a,pg_ts_config b, pg_ts_dict c WHERE a.mapcfg=b.oid AND a.mapdict=c.oid AND b.cfgname='english_1' ORDER BY 1,2,3,4,5;
```

```
  cfgname  | maptokentype | mapseqno | mapdict |   dictname
-----------+--------------+----------+---------+--------------
 english_1 |            2 |        1 |    3765 | simple
 english_1 |            2 |        2 |   14505 | english_stem
 english_1 |            4 |        1 |   14505 | english_stem
 english_1 |            4 |        2 |   14511 | german_stem
(4 rows)
```


版权所有 © 华为云计算技术有限公司

相关链接

CREATE TEXT SEARCH CONFIGURATION，DROP TEXT SEARCH CONFIGURATION，全文检索

查看数据库中的文本搜索配置信息

DWS提供以下系统表用于查看当前数据库中的文本搜索配置信息：

- PG_TS_CONFIG系统表记录当前数据库中创建的文本搜索配置信息，包括文本搜索配置的名称、所有者及配置选项等。
- PG_TS_CONFIG_MAP系统表记录文本搜索配置的解析器输出的符号类型，关联的文本搜索字典及搜索顺序。

13.6.10 DROP TEXT SEARCH CONFIGURATION

功能描述

删除已有文本搜索配置。

注意事项

要执行DROP TEXT SEARCH CONFIGURATION命令，用户必须是该文本搜索配置的所有者。

语法格式

```
DROP TEXT SEARCH CONFIGURATION [ IF EXISTS ] name [ CASCADE | RESTRICT ];
```

参数说明

表 13-125 DROP TEXT SEARCH CONFIGURATION 参数说明

参数	描述	取值范围
IF EXISTS	如果指定的文本搜索配置不存在，那么发出一个notice而不是抛出一个错误。	-
name	要删除的文本搜索配置名称（可有模式修饰）。	已存在的文本搜索配置名称。
CASCADE | RESTRICT	当执行删除操作时，如何处理依赖对象中的相关数据。	• CASCADE：级联删除依赖文本搜索配置的对象。 • RESTRICT：如果有任何对象依赖文本搜索配置则拒绝删除它。这是默认情况。此选项为缺省值。

示例

删除文本搜索配置ngram1:

 版权所有 © 华为云计算技术有限公司

```
DROP TEXT SEARCH CONFIGURATION IF EXISTS  ngram1;
```

相关链接

ALTER TEXT SEARCH CONFIGURATION，CREATE TEXT SEARCH CONFIGURATION

13.6.11 CREATE TEXT SEARCH DICTIONARY

功能描述

创建一个新的全文检索词典。全文检索词典指定在全文检索时识别特定词并处理的方法。

词典的创建依赖于预定义模板（在系统表PG_TS_TEMPLATE中定义），支持创建五种类型的词典，分别是Simple、Ispell、Synonym、Thesaurus、以及Snowball，每种类型的词典可以完成不同的任务。

注意事项

- 具有SYSADMIN权限的用户可以执行创建词典操作，创建该词典的用户自动成为其所有者。
- 临时模式（pg_temp）下不允许创建词典。
- 创建或修改词典之后，任何对于用户自定义的词典定义文件的修改，将不会影响到数据库中的词典。如果需要在数据库中使用这些修改，需使用ALTER语句更新对应词典的定义文件。
- 不支持修改词典的TEMPLATE参数值。
- 不支持仅修改FILEPATH参数而不修改对应的词典定义文件参数。
- 词典定义文件的文件名仅支持小写字母、数字、下划线混合。
- 预定义词典文件位于$GAUSSHOME/share/postgresql/tsearch_data目录下。

语法格式

```
CREATE TEXT SEARCH DICTIONARY name (
    TEMPLATE = template
    [, option = value [, ... ]]
);
```

参数说明

表 13-126 CREATE TEXT SEARCH DICTIONARY 参数说明

参数	描述	取值范围
name	要创建的全文检索词典的名称，可以用模式修饰。	字符串，需符合标识符命名规范。
template	模板名。	系统表PG_TS_TEMPLATE中定义的模板：Simple/Synonym/Thesaurus/Ispell/Snowball。

版权所有 © 华为云计算技术有限公司

参数	描述	取值范围
option	全文检索词典特定选项的参数名称。与template对应，不同的词典类型具有不同的参数列表，且与指定顺序无关。	详细参数说明参见表13-127。
value	参数值。如果不是简单的标识符或数字，则参数值必须加单引号（标识符和数字同样可以加上单引号）。	-

表 13-127 Simple、Ispell、Synonym、Thesaurus、以及 Snowball 的 option 参数

option	option参数选项的名称	描述
Simple词典对应的option	STOPWORDS	停用词表文件名，默认后缀名为stop。例如STOPWORDS = french，则实际文件名为french.stop。停用词文件格式为一组word列表，每行定义一个停用词。词典处理时，文件中的空行和空格会被忽略，并将stopword词组转换为小写形式。
	ACCEPT	是否将非停用词设置为已识别。默认值为true。 当Simple词典设置参数ACCEPT=true时，将不会传递任何token给后继词典，此时建议将其放置在词典列表的最后。反之，当ACCEPT=false时，建议将该Simple词典放置在列表中的至少一个词典之前。

 版权所有 © 华为云计算技术有限公司

option	option参数选项的名称	描述
	FILEPATH	停用词文件所在目录。停用词文件可以存放于本地，也可以存放于对象存储服务OBS服务器。如果存放在本地，目录格式为'file://absolute_path'。如果存放于OBS服务器，目录格式为'obs://bucket/path accesskey=ak secretkey=sk region=region_name'。注意目录要用单引号包含。默认值为预定义词典文件所在目录。FILEPATH参数必须和STOPWORDS参数同时指定，不允许单独指定。 使用OBS服务器上的停用词文件创建字典的过程如下： 1. 将停用词文件上传到OBS服务器。例如将french.stop文件上传到OBS服务器obsv3.sa-fb-1.externaldemo.com上名为gaussdb的桶中，其URL为https://gaussdb.obsv3.sa-fb-1.externaldemo.com/french.stop。上传文件及查询URL的方式请参考OBS服务的用户手册。 2. 修改$GAUSSHOME/etc/region_map文件，在其中加入一行"region_name": "obs domain"。region_name可以为任意由大小写字母、数字、斜杠（/）或下划线组成的字符串。obs domain为OBS服务器的域名。 示例，将region_name设为rg，region_map包含的内容如："rg": "obsv3.sa-fb-1.externaldemo.com"。 **须知** region_name和obs domain都用双引号，冒号的左边没有空格，右边有一个空格。 3. 执行CREATE TEXT SEARCH DICTIONARY命令创建字典。命令如下： `CREATE TEXT SEARCH DICTIONARY french_dict ( TEMPLATE = pg_catalog.simple, STOPWORDS = french, FILEPATH = 'obs://gaussdb accesskey=xxx secretkey=yyy region=rg' );` 由于french.stop文件放在gaussdb桶的根目录下，因此path为空。
Synonym词典对应的option	SYNONYM	同义词词典的定义文件名，默认后缀名为syn。 文件格式为一组同义词列表，每行格式为"token synonym"，即token和其对应的synonym，中间以空格相连。
	CASESENSITIVE	设置是否大小写敏感，默认值为false，此时词典文件中的token和synonym均会转为小写形式处理。如果设置为true，则不会进行小写转换。

版权所有 © 华为云计算技术有限公司

option	option参数选项的名称	描述
	FILEPATH	同义词词典文件所在目录。目录可以指定为本地目录和OBS目录两种形式。默认值为预定义词典文件所在目录。其中目录格式、以及使用OBS服务器上的文件创建Synonym字典的过程与Simple字典相同。
Thesaurus词典对应的option	DICTFILE	词典定义文件名，默认后缀名为ths。 文件格式为一组同义词列表，每行格式为"sample words : indexed words"，中间冒号（:）作为短语和其替换词间的分隔符。TZ词典处理时，如果有多个匹配的sample words，将选择最长匹配输出。
	DICTIONARY	用于词规范化的子词典名，必须且仅能定义一个。该词典必须是已经存在的，在检查短语匹配之前使用，用于识别和规范输入文本。 如果子词典无法识别输入词，将会报错。此时，需要移除该词或者更新子词典使其识别。此外，可在indexed words的开头放上一个星号（*）来跳过在其上应用子词典，但是所有sample words必须可以被子词典识别。 如果词典文件定义的sample words中，含有子词典中定义的停用词，需要用问号（？）替代停用词。假设a和the是子词典中所定义的停用词，如下： `? one ? two : swsw` 上述同义词组定义会匹配"a one the two"以及"the one a two"，这两个短语均会被swsw替代输出。
	FILEPATH	词典定义文件所在目录。目录可以指定为本地目录和OBS目录两种形式。默认值为预定义词典文件所在目录。其中目录格式、以及使用OBS服务器上的文件创建Synonym字典的过程与Simple字典相同。
Ispell词典	DICTFILE	词典定义文件名，默认后缀名为dict。
	AFFFILE	词缀文件名，默认后缀名为affix。
	STOPWORDS	停用词文件名，默认后缀名为stop，文件格式要求与Simple类型词典的停用词文件相同。
	FILEPATH	词典文件所在目录。可以指定为本地目录和OBS目录两种形式。默认值为预定义词典文件所在目录。其中目录格式、以及使用OBS服务器上的文件创建Synonym字典的过程与Simple字典相同。
Snowball词典	LANGUAGE	语言名，标识使用哪种语言的词干分析算法。算法按照对应语言中的拼写规则，缩减输入词的常见变体形式为一个基础词或词干。
	STOPWORDS	停用词表文件名，默认后缀名为stop，文件格式要求与Simple类型词典的停用词文件相同。

版权所有 © 华为云计算技术有限公司

option	option参数选项的名称	描述
	FILEPATH	词典定义文件所在目录。可以指定为本地目录或者OBS目录。默认值为预定义词典文件所在目录。FILEPATH参数必须和STOPWORDS参数同时指定，不允许单独指定。其中目录格式、以及用OBS服务器上的文件创建Snowball字典的过程与Simple字典相同。

示例

创建一个Ispell词典类型的词典english_ispell（词典定义文件来自开源词典）。

```
CREATE TEXT SEARCH DICTIONARY english_ispell (
    TEMPLATE = ispell,
    DictFile = english,
    AffFile = english,
    StopWords = english,
    FilePath = 'obs://bucket_name/path accesskey=ak secretkey=sk region=rg'
);
```

请参见**配置示例**一节的示例。

相关链接

ALTER TEXT SEARCH DICTIONARY，**CREATE TEXT SEARCH DICTIONARY**

13.6.12 ALTER TEXT SEARCH DICTIONARY

功能描述

修改全文检索词典的相关定义，包括参数、名称、所有者以及模式等。

注意事项

- 预定义词典不支持ALTER操作。
- 只有词典的所有者可以执行ALTER操作，系统管理员默认拥有此权限。
- 创建或修改词典之后，任何对于filepath路径下用户自定义的词典定义文件的修改，将不会影响到数据库中的词典。如果需要在数据库中使用这些修改，需使用ALTER TEXT SEARCH DICTIONARY语句更新对应词典的定义文件。

语法格式

- 修改词典定义。
  ```
  ALTER TEXT SEARCH DICTIONARY name ( option = value | option [, ...] );
  ```
- 重命名词典。
  ```
  ALTER TEXT SEARCH DICTIONARY name RENAME TO new_name;
  ```
- 设置词典的所属模式。
  ```
  ALTER TEXT SEARCH DICTIONARY name SET SCHEMA new_schema;
  ```
- 修改词典的所属者。
  ```
  ALTER TEXT SEARCH DICTIONARY name OWNER TO new_owner;
  ```


版权所有 © 华为云计算技术有限公司

参数说明

表 13-128 ALTER TEXT SEARCH DICTIONARY 参数说明

参数	描述	取值范围
name	需要修改的全文检索词典的名称，可以用模式修饰。	有效的全文检索词典名。
option	要修改的全文检索词典特定选项的参数名称。与template对应，不同的词典类型具有不同的参数列表，且与指定顺序无关。 • 不支持修改词典的TEMPLATE参数值。 • 不支持仅修改FILEPATH参数而不修改对应的词典定义文件参数。 • 词典定义文件的文件名仅支持小写字母、数字、下划线混合。	详细参数说明请见**参数说明**中option参数。
value	要修改的参数值。如果省略等号（=）和value，则表示删除该option的先前设置，使用默认值。	对应option定义。
new_name	全文检索词典的新名称。	字符串，需符合**标识符命名规范**。
new_owner	全文检索词典的新所有者。	有效的用户名。
new_schema	全文检索词典的新模式。	有效的模式名。

示例

创建全文检索词典my_dict。

```
CREATE TEXT SEARCH DICTIONARY my_dict (
    TEMPLATE = snowball,
    Language = english,
    StopWords = english
);
```

修改全文检索词典my_dict的Language参数，并删除停用词定义。

```
ALTER TEXT SEARCH DICTIONARY my_dict (Language = dutch,StopWords);
```

更新全文检索词典my_dict定义，实际不更改任何内容。

```
ALTER TEXT SEARCH DICTIONARY my_dict (dummy);
```

相关链接

CREATE TEXT SEARCH DICTIONARY，**DROP TEXT SEARCH DICTIONARY**

版权所有 © 华为云计算技术有限公司

13.6.13 DROP TEXT SEARCH DICTIONARY

功能描述

删除全文检索词典。

注意事项

- 预定义词典不支持DROP操作。
- 只有词典的所有者可以执行DROP操作，系统管理员默认拥有此权限。
- 谨慎执行DROP...CASCADE操作，该操作将级联删除使用该词典的文本搜索配置（TEXT SEARCH CONFIGURATION）。

语法格式

```
DROP TEXT SEARCH DICTIONARY [ IF EXISTS ] name [ CASCADE | RESTRICT ]
```

参数说明

表 13-129 DROP TEXT SEARCH DICTIONARY 参数说明

参数	描述	取值范围
IF EXISTS	如果指定的全文检索词典不存在，那么发出一个Notice而不是报错。	-
name	要删除的词典名称（可指定模式名，否则默认在当前模式下）。	已存在的词典名。
CASCADE | RESTRICT	当执行删除操作时，如何处理依赖对象中的相关数据。	• CASCADE：自动删除依赖于该词典的对象，并依次删除依赖于这些对象的所有对象。如果存在任何一个使用该词典的文本搜索配置，此DROP命令将不会成功。可添加CASCADE以删除引用该词典的所有文本搜索配置以及词典。 • RESTRICT：如果任何对象依赖词典，则拒绝删除该词典。此选项是缺省值。

示例

删除词典english：

```
DROP TEXT SEARCH DICTIONARY IF EXISTS english;
```

 版权所有 © 华为云计算技术有限公司

相关链接

ALTER TEXT SEARCH DICTIONARY，CREATE TEXT SEARCH DICTIONARY

13.6.14 CLUSTER

功能描述

根据索引对表进行聚簇排序。

- CLUSTER基于索引名指定的索引来聚簇由表名指定的表。索引名指定的索引必须已经定义在指定表上。
- 当对一个表聚簇后，该表将基于索引信息进行物理排序。聚簇是一次性操作：当表被更新之后， 更改的内容不会被聚簇。也就是说，系统不会试图按照索引顺序对新的存储内容及更新记录进行重新聚簇。
- 在对表聚簇时，DWS会记录在哪个索引上建立了聚簇。 形式CLUSTER table_name会使用前面所用的同一个索引对表重新聚簇。用户也可以用CLUSTER或ALTER TABLE的SET WITHOUT CLUSTER形式来设置索引用于后续的聚簇操作或清除任何之前的设置。
- 不带任何参数的CLUSTER会将当前用户所拥有的数据库中的已执行过聚簇的所有表重新聚簇，如果是系统管理员调用，则是所有已被聚簇过的表。
- 在对表进行聚簇时，会在其上请求ACCESS EXCLUSIVE锁。这样避免了在CLUSTER完成之前对此表执行其他的操作（包括读写）。

注意事项

- CLUSTER不允许在事务中执行。
- HDFS表不能进行CLUSTER操作。
- 全局临时表不能进行CLUSTER操作。
- 对表执行CLUSTER操作时会触发表重建（表重建过程中会先把数据转储到一个新的数据文件中，重建完成之后会删除原始文件），当表比较大时，重建会消耗较多的磁盘空间。当磁盘空间不足时，要谨慎对待大表CLUSTER操作，防止触发集群只读。
- 存算分离架构下，对CLUSTER全库的操作会有提示不支持。
- 只有行存B-Tree索引支持CLUSTER操作。
- 如果用户只是随机访问表中的行，那么表中数据的实际存储顺序是无关紧要的。但是， 如果对某些数据的访问多于其他数据，而且有一个索引将这些数据分组，那么使用CLUSTER会有所帮助。如果从一个表中请求一定索引范围的值， 或者是一个索引值对应多行，CLUSTER也会有所帮助，因为如果索引标识出第一匹配行所在的存储页，所有其他行也可能已经在同一个存储页里了，因此可以节省磁盘访问的时间并加快查询速度。
- 在聚簇过程中，系统先创建一个按照索引顺序建立的表的临时副本，同时也创建了表上每个索引的临时副本。因此，磁盘上需要有足够的可用空间， 至少等于表大小和索引大小的总和。
- 由于CLUSTER会记忆聚集信息，因此可以在第一次手动对要聚簇的表进行聚簇，然后设置一个定期维护脚本，这样就可以周期地自动对表进行聚簇操作。
- 因为优化器会记录有关表排序的统计信息，所以建议在新的聚簇表上运行ANALYZE。否则，优化器可能会选择错误的查询规划。

版权所有 © 华为云计算技术有限公司

语法格式

- 对指定表进行聚簇排序。
  ```
  CLUSTER [ VERBOSE ] table_name [ USING index_name ];
  ```
- 对指定分区进行聚簇排序。
  ```
  CLUSTER [ VERBOSE ] table_name PARTITION ( partition_name ) [ USING index_name ];
  ```
- 对已执行过聚簇的表重新进行聚簇。
  ```
  CLUSTER [ VERBOSE ];
  ```

参数说明

表 13-130 CLUSTER 参数说明

参数	描述	取值范围
VERBOSE	启用显示进度信息。	-
table_name	表名称。	已存在的表名称。
index_name	索引名称。	已存在的索引名称。
partition_name	分区名称。	已存在的分区名称。

示例

创建一个分区表。

```
DROP SCHEMA IF EXISTS tpcds CASCADE;
CREATE SCHEMA tpcds;
CREATE TABLE tpcds.inventory_p1
(
   INV_DATE_SK             INTEGER              NOT NULL,
   INV_ITEM_SK             INTEGER              NOT NULL,
   INV_WAREHOUSE_SK        INTEGER              NOT NULL,
   INV_QUANTITY_ON_HAND    INTEGER
)
DISTRIBUTE BY HASH(INV_ITEM_SK)
PARTITION BY RANGE(INV_DATE_SK)
(
      PARTITION P1 VALUES LESS THAN(2451179),
      PARTITION P2 VALUES LESS THAN(2451544),
      PARTITION P3 VALUES LESS THAN(2451910),
      PARTITION P4 VALUES LESS THAN(2452275),
      PARTITION P5 VALUES LESS THAN(2452640),
      PARTITION P6 VALUES LESS THAN(2453005),
      PARTITION P7 VALUES LESS THAN(MAXVALUE)
);
```

创建索引ds_inventory_p1_index1。

```
CREATE INDEX ds_inventory_p1_index1 ON tpcds.inventory_p1 (INV_ITEM_SK) LOCAL;
```

对表tpcds.inventory_p1进行聚簇。

```
CLUSTER tpcds.inventory_p1 USING ds_inventory_p1_index1;
```

对分区p3进行聚簇。

```
CLUSTER tpcds.inventory_p1 PARTITION (p3) USING ds_inventory_p1_index1;
```

对数据库中的表进行聚簇。

 版权所有 © 华为云计算技术有限公司

```
CLUSTER;
```

13.7 存储过程&函数&触发器&游标

13.7.1 CREATE FUNCTION

功能描述

FUNCTION（函数）是一段可重用的代码块，它接受输入参数，执行特定操作，并返回一个值或结果集。CREATE FUNCTION语句用于在数据库中创建**用户自定义函数（UDF）**，实现DWS当前内置的**系统函数**中还未支持的其他计算功能，覆盖用户特定的复杂计算场景。

函数的主要特点包括：

- 支持输入参数，可根据不同条件返回结果。
- 需声明返回值的数据类型，执行结果返回单一值（标量函数）或表（表值函数）。
- 将复杂逻辑封装为单一函数，可以在SQL语句中像内置函数一样调用，提高代码的复用性。

DWS 函数易变属性的三种行为

SQL中函数的行为特性通常分为三类，包括**immutable**（不可变函数）、**stable**（稳定函数）、**volatile**（易变函数），三者关系如**表13-131**所示：

表 13-131 函数的易变属性

易变属性	定义	典型场景
immutable	immutable函数的结果仅由输入参数决定，与任何外部状态（如数据库当前数据、时间、事务上下文等）无关。无论何时、何地调用，只要输入参数相同，返回结果永远一致。且函数不会修改数据库中的数据。	数学计算、哈希、字符串操作，例如ABS(x)，POWER(a, b)，MD5(str)。
stable	stable函数的结果在单次查询执行期间保持稳定。即在同一个查询的多次调用中，只要输入参数相同，结果一致。但结果可能依赖数据库的当前状态（如表数据、会话变量等），因此在不同查询或事务中可能返回不同结果。	基于当前数据的过滤、聚合，例如部分聚合函数：MAX(col), MIN(col)（在单次扫描中结果稳定）。
volatile	volatile函数是最“不稳定”的类型。其结果可能在每次调用时变化，或可能修改数据库状态（如写入数据）。优化器无法预测其行为，必须每次调用都重新计算。	随机数、时间戳、DML操作，例如随机数函数：RANDOM()（每次调用返回不同值）。

版权所有 © 华为云计算技术有限公司

查看函数属于哪种属性，可以查看PG_PROC系统表的provolatile字段（i表示immutable，s表示stable，v表示volatile）。

函数是否下推到 DN（数据节点）执行（SHIPPABLE 和 NOT SHIPPABLE 介绍）

CREATE FUNCTION语法中，通过SHIPPABLE（下推）和NOT SHIPPABLE（不下推）来控制该函数是否下推到数据节点（DN）执行。

- 对于IMMUTABLE类型的函数，函数始终可以下推到DN上执行。
- 对于STABLE/VOLATILE类型的函数，仅当函数的属性是SHIPPABLE的时候，函数可以下推到DN执行。

  用户在定义函数的SHIPPABLE属性时也需特别慎重，SHIPPABLE意味着整个函数会下推到DN上执行，如果设置不当，会导致结果错误等严重问题。

  与定义IMMUTABLE属性一样，SHIPPABLE属性的定义也有诸多约束，简单来说就是函数内不能有不可下推的因素，函数下推到单DN执行后，函数内部的计算逻辑仅依赖当前DN的数据集合。

不能定义为IMMUTABLE的情况，举例如下：

1. 如果自定义函数中引用了表、视图等对象，那么该函数就不能定义为IMMUTABLE，因为当表的数据发生变化的时候，函数的返回值可能发生变化。
2. 如果自定义函数中引用了STABLE/VOLATILE类型的函数，那么该函数不能定义为IMMUTABLE。
3. 如果自定义函数中有不下推的因素，则该函数不能定义成IMMUTABLE，因为IMMUTABLE意味着要下推到DN执行，与函数内部的不下推因素相互冲突。典型场景例如，包含不下推的函数、语法等。
4. 如果自定义函数中含有聚合运算，但聚合运算的运算需要生成STREAM计划才能完成计算的（部分结果在DN计算，部分结果在CN计算，例如listagg函数等）。

   说明

   为了防止错误定义IMMUTABLE可能导致严重问题，可以通过设置behavior_compat_options参数的check_function_conflicts选项，检查自定义plpgsql/SQL函数的IMMUTABLE属性。可识别上述1和2场景问题。

不能定义为SHIPPABLE的情况，举例如下：

1. 如果函数内部引用了表，并且表为HASH分布，那么该函数通常不能定义为SHIPPABLE。
2. 函数内部有不可下推的因素，例如不支持下推的函数、语法等，那么该函数不能定义为SHIPPABLE，可参考语句下推调优章节。
3. 函数内部的计算过程可能需要跨DN数据，这种情况该函数通常不能定义为SHIPPABLE，例如一些聚合运算等。

注意

用户在使用DWS时，应该正确指定函数属性，错误指定函数属性不仅会导致查询语句执行效率低，而且可能会导致结果集不稳定的情况。

注意事项

- 如果创建函数时参数或返回值带有精度，不进行精度检测。

版权所有 © 华为云计算技术有限公司

- 创建函数时，函数定义中对表对象的操作建议都显式指定模式（例如hr.table1，明确指定是hr模式下面的表），否则可能会导致函数执行异常。
- 在创建函数时，函数内部通过SET语句设置current_schema和search_path无效。函数执行完的search_path和current_schema与函数执行前的search_path和current_schema保持一致。
- 如果函数参数中带有出参，SELECT调用函数必须缺省出参，CALL调用函数适配Oracle必须指定出参，对于调用重载的带有PACKAGE属性的函数，CALL调用函数可以缺省出参，具体信息参见**CALL**的示例。
- 兼容PostgreSQL风格的函数或者带有PACKAGE属性的函数支持重载。在指定REPLACE的时候，如果参数个数、类型、返回值有变化，不会替换原有函数，而是会建立新的函数。
- SELECT调用可以指定不同参数来进行同名函数调用。由于语法CALL适配自Oracle，因此不支持调用不带有PACKAGE属性的同名函数。
- 在创建function时，不能在avg函数外面嵌套其他agg函数，或者其他系统函数。
- 在非逻辑集群模式下，暂不支持将返回值、参数以及变量设置为建在非系统默认安装Node Group的表，sql function内部语句暂不支持对建在非系统默认安装Node Group的表操作。
- 在逻辑集群模式下，如果函数返回值和参数是用户表类型，所有涉及表都必须在同一逻辑集群内；如果函数内部涉及对多个逻辑集群表操作，函数定义时不能为IMMUTABLE和SHIPPABLE类型，以避免函数被下推到DN执行。
- 在逻辑集群模式下，函数参数、返回值不能用%type引用表字段类型，否则会导致函数创建失败。
- 新创建的函数默认会给PUBLIC授予执行权限（详见**GRANT**）。用户可以选择收回PUBLIC默认执行权限，然后根据需要将执行权限授予其他用户，为了避免出现新函数能被所有人访问的时间窗口，应在一个事务中创建函数并且设置函数执行权限。
- 在多个CN的集群中，不能将函数的入参或出参设置为临时表类型，因为在非当前连接CN上执行创建函数过程中，无法单独根据表名获取到正确的临时schema，进而无法获取到准确的表类型。

语法格式

- 兼容PostgreSQL风格的创建自定义函数语法。

```
CREATE [ OR REPLACE ] FUNCTION function_name
    ( [ { argname [ argmode ] argtype [ { DEFAULT | := | = } expression ]} [, ...] ] )
    [ RETURNS rettype [ DETERMINISTIC ] | RETURNS TABLE ( { column_name column_type } 
[, ...] )]
    LANGUAGE lang_name
    [
      {IMMUTABLE | STABLE | VOLATILE }
       | {SHIPPABLE | NOT SHIPPABLE}
       | WINDOW
       | [ NOT ] LEAKPROOF
       | {CALLED ON NULL INPUT | RETURNS NULL ON NULL INPUT | STRICT }
       | {[ EXTERNAL ] SECURITY INVOKER | [ EXTERNAL ] SECURITY DEFINER | AUTHID DEFINER | 
AUTHID CURRENT_USER}
       | {FENCED | NOT FENCED}
       | {MASKED | NOT MASKED}
       | {PACKAGE}

       | COST execution_cost
       | ROWS result_rows
       | SET configuration_parameter { {TO | =} value | FROM CURRENT }}
     ][...]
    {
```

 版权所有 © 华为云计算技术有限公司

```
    AS 'definition'
    | AS 'obj_file', 'link_symbol'
  }
```

- Oracle风格的创建自定义函数的语法。

```
CREATE [ OR REPLACE  ] FUNCTION function_name
  ( [  { argname [ argmode  ] argtype [  { DEFAULT | := | =  } expression  ] }  [, ...]  ] )
  RETURN rettype [ DETERMINISTIC  ]
  [
    {IMMUTABLE  | STABLE  | VOLATILE }
    | {SHIPPABLE | NOT SHIPPABLE}
    | {PACKAGE}
    | {FENCED | NOT FENCED}
    | {MASKED | NOT MASKED}
    | [ NOT  ] LEAKPROOF
    | {CALLED ON NULL INPUT | RETURNS NULL ON NULL INPUT | STRICT }
    | {[ EXTERNAL  ] SECURITY INVOKER  | [ EXTERNAL  ] SECURITY DEFINER |
AUTHID DEFINER | AUTHID CURRENT_USER
}
    | COST execution_cost
    | ROWS result_rows
    | SET configuration_parameter { {TO | =} value  | FROM CURRENT

  ][...]

  {
    IS  | AS
} plsql_body
/
```

参数说明

表 13-132 CREATE FUNCTION 参数说明

参数	描述	取值范围
function_name	要创建的函数名称（可以用模式修饰）。	取值范围：字符串，要符合标识符的命名规范。参见**标识符命名规范**。 **注意** 如果创建的函数名与系统函数同名，建议指定schema。调用自定义函数时需指定schema，否则系统会优先调用系统函数。
argname	函数参数的名字。	字符串，要符合标识符的命名规范。
argmode	函数参数的模式。	IN，OUT，INOUT或VARIADIC。 • IN：输入参数，缺省值是IN。 • OUT：输出参数，只有OUT模式的参数后面能跟VARIADIC。 • INOUT：既是输入又是输出参数。 • OUT VARIADIC: 表示输出参数，且声明为数组类型。 **注意** OUT和INOUT模式的参数不能用在RETURNS TABLE的函数定义中。

 版权所有 © 华为云计算技术有限公司

参数	描述	取值范围
argtype	函数参数的数据类型。	常用的数据类型参见**数据类型**。
expression	函数参数的默认表达式。	-
rettype	函数返回值的数据类型。	如果存在OUT或INOUT参数，可以省略RETURNS子句。如果该子句存在，则必须和输出参数所表示的结果类型一致：如果有多个输出参数，则为RECORD，否则与单个输出参数的类型相同。 SETOF修饰词表示该函数将返回一个集合，而不是单独一项。
DETERMINISTIC	为适配Oracle SQL语法，未实现功能，不推荐使用。	-
column_name	字段名称。	-
column_type	字段类型。	-
definition	一个定义函数的字符串常量，含义取决于语言。它可以是一个内部函数名称、一个指向某个目标文件的路径、一个SQL查询、一个过程语言文本。	-
LANGUAGE lang_name	用于实现函数的语言的名字。可以是SQL，internal，或者是用户定义的过程语言名字。为了保证向下兼容，该名字可以用单引号（包围）。若采用单引号，则引号内必须为大写。	-
WINDOW	表示该函数是窗口函数，替换函数定义时不能改变WINDOW属性。 **注意** 自定义窗口函数只支持LANGUAGE是internal，并且引用的内部函数必须是窗口函数。	-

版权所有 © 华为云计算技术有限公司

参数	描述	取值范围
IMMUTABLE	表示该函数在给出同样的参数值时总是返回同样的结果。	如果函数的入参是常量，会在优化器阶段计算该函数的值。益处是可以尽早获取表达式的值，从而能更准确的进行代价估算，生成的执行计划也更优。 用户自定义的IMMUTABLE的函数是会被自动下推到DN执行的，但是这样可能有潜在的风险，即如果用户错误定义了函数的IMMUTABLE属性，但是函数执行的过程并不是IMMUTABLE的，那么可能会导致结果错误等严重问题。因此，用户在指定函数的属性为IMMUTABLE的时候，要特别慎重。 举例如下： 1. 如果自定义函数中引用了表，视图等对象，那么该函数就不能定义为IMMUTABLE，因为当表的数据发生变化的时候，函数的返回值可能发生变化。 2. 如果自定义函数中引用了STABLE/VOLATILE类型的函数，那么该函数不能定义为IMMUTABLE。 3. 如果自定义函数中有不下推的因素，则该函数不能定义成IMMUTABLE，因为IMMUTABLE意味着要下推到DN执行，与函数内部的不下推因素相互冲突。典型场景例如，包含不下推的函数、语法等。 4. 如果自定义函数中含有聚合运算，但聚合运算的运算需要生成STREAM计划才能完成计算的（部分结果在DN计算，部分结果在CN计算，例如listagg函数等）。 同时，为了防止这种情况下可能出现严重问题，数据库内部可以通过设置behavior_compat_options=‘check_function_conflicts’来开启对函数定义冲突的检查，目前可以识别出上述1和2场景。

版权所有 © 华为云计算技术有限公司

参数	描述	取值范围
STABLE	表示该函数不能修改数据库，对相同参数值，在同一次表扫描里，该函数的返回值不变，但是返回值可能在不同SQL语句之间变化。	-
VOLATILE	表示该函数值可以在一次表扫描内改变，因此不会做任何优化。	-
SHIPPABLE NOT SHIPPABLE	表示该函数是否可以下推到DN上执行。	• 对于IMMUTABLE类型的函数，函数始终可以下推到DN上执行。 • 对于STABLE/VOLATILE类型的函数，仅当函数的属性是SHIPPABLE的时候，函数可以下推到DN执行。 用户在定义函数的SHIPPABLE属性时也需特别慎重，SHIPPABLE意味着整个函数会下推到DN上执行，如果设置不当，会导致结果错误等严重问题。 与定义IMMUTABLE属性一样，SHIPPABLE属性的定义也有诸多约束，简单来说就是函数内不能有不可下推的因素，函数下推到单DN执行后，函数内部的计算逻辑仅依赖当前DN的数据集合。 举例如下： 1. 如果函数内部引用了表，并且表为HASH分布，那么该函数通常不能定义为SHIPPABLE。 2. 函数内部有不可下推的因素，例如不支持下推的函数、语法等，那么该函数不能定义为SHIPPABLE。 3. 函数内部的计算过程可能需要跨DN数据，这种情况该函数通常不能定义为SHIPPABLE，例如一些聚合运算等。

 版权所有 © 华为云计算技术有限公司

参数	描述	取值范围
PACKAGE	表示该函数是否支持重载。PostgreSQL风格的函数本身就支持重载，此参数主要是针对Oracle风格的函数。	• 不允许package函数和非package函数重载或者替换。 • package函数不支持VARIADIC类型的参数。 • 不允许修改函数的package属性。
LEAKPROOF	指出该函数的参数只包括返回值。LEAKPROOF只能由系统管理员设置。	-
CALLED ON NULL INPUT	表明该函数的某些参数是NULL时可以按照正常的方式调用。该参数可以省略。	-
RETURNS NULL ON NULL INPUT STRICT	STRICT用于指定如果函数的某个参数是NULL，此函数总是返回NULL。如果声明了这个参数，当有NULL值参数时该函数不会被执行；而只是自动返回一个NULL结果。 RETURNS NULL ON NULL INPUT和STRICT的功能相同。	-
EXTERNAL	目的是和SQL兼容，是可选的，这个特性适合于所有函数，而不仅是外部函数。	-
SECURITY INVOKER AUTHID CURRENT_USER	表明该函数将带着调用它的用户的权限执行。该参数可以省略。 SECURITY INVOKER和AUTHID CURRENT_USER的功能相同。	-
SECURITY DEFINER AUTHID DEFINER	声明该函数将以创建它的用户的权限执行。 AUTHID DEFINER和SECURITY DEFINER的功能相同。	-

版权所有 © 华为云计算技术有限公司

参数	描述	取值范围
FENCED NOT FENCED	该参数只对用户定义的C函数生效，声明函数是在保护模式还是非保护模式下执行。如果函数声明为NOT FENCED模式，则函数的执行在CN或者DN进程中进行。如果函数声明为FENCED模式，则函数在新fork的进程执行，这样函数的异常不会影响CN或者DN进程。	FENCED/NOT FENCED模式的选择： ● 正在开发或者调试的Function使用FENCED模式。开发测试完成，使用NOT FENCED模式执行，减少fork进程以及通信的开销。 ● 复杂的操作系统操作，例：打开文件，信号处理，线程处理等操作，使用FENCED模式。否则可能影响DWS数据库的执行。 ● 默认值为FENCED。
MASKED NOT MASKED	该参数在GUC参数enable_redactcol_computable开启的情况下，是否进行脱敏。系统函数支持修改MASKED、NOT MASKED选项。	-
COST execution_cost	用来估计函数的执行成本。execution_cost以cpu_operator_cost为单位。	取值范围：正数。
ROWS result_rows	估计函数返回的行数。用于函数返回的是一个集合。	正数，默认值是1000行。
configuration_parameter	该函数配置的参数值。	● **value** 把指定的数据库会话参数值设置为给定的值。如果value是DEFAULT或者RESET，则在新的会话中使用系统的缺省设置。OFF关闭设置。 取值范围：字符串 – DEFAULT – OFF – RESET 指定默认值。 ● **from current** 取当前会话中的值设置为configuration_parameter的值。

版权所有 © 华为云计算技术有限公司

参数	描述	取值范围
plsql_body	PL/SQL存储过程体。 当在函数中创建用户时，日志中会记录密码的明文。因此不建议用户在函数中创建用户。	-

示例：创建一个两个整数求和的 SQL 函数

语法如下：

```
CREATE  OR REPLACE FUNCTION func_add_sql(
integer,                --第一个整数参数，未命名，使用位置参数$1引用
integer                 --第二个整数参数，未命名，使用位置参数$2引用
)
RETURNS integer             --指定函数返回类型为整数
    AS 'select $1 + $2;'    --函数体：使用位置参数$1和$2相加
    LANGUAGE SQL            --使用SQL语言实现该函数
    IMMUTABLE               --函数类型为IMMUTABLE，表示相同输入总是返回相同结果
    RETURNS NULL ON NULL INPUT;      --等同于STRICT，任何参数为NULL时，函数返回NULL
```

执行函数，返回两个入参的求和值：

```
SELECT func_add_sql(8,3);
```

```
 func_add_sql
--------------
           11
```

执行函数，输入为Null，返回为Null。

```
SELECT func_add_sql(Null,3);
```

```
 func_add_sql
--------------
 (Null)
```

示例：创建一个处理整数自增的 PL/pgSQL 语言函数

语法如下：

```
CREATE OR REPLACE FUNCTION func_increment_plsql(
i integer                           --输入参数，整数类型
) RETURNS integer                       --函数返回值为整数类型
AS $$
    BEGIN                           --表示函数体开始
        RETURN i + 1;               --返回输入参数加1的结果
    END;
$$ LANGUAGE plpgsql;                    --指定函数使用PL/pgSQL语言编写
```

执行函数，返回整数自增结果。

```
SELECT func_increment_plsql(5);
```

```
 func_increment_plsql
----------------------
                    6
```

 版权所有 © 华为云计算技术有限公司

入参由一个表numbers的第一列指定，使用函数计算后，将计算结果返回到第二列：

```
DROP TABLE IF EXISTS numbers;
CREATE TABLE numbers (value integer);
INSERT INTO numbers VALUES (1), (2), (3), (4), (5);
SELECT value, func_increment_plsql(value) AS incremented_value FROM numbers ORDER BY 1;
```

value	incremented_value
1	2
2	3
3	4
4	5
5	6

示例：创建一个返回 RECORD 类型的 PL/pgSQL 语言函数：

```
CREATE OR REPLACE FUNCTION compute(
i int,                          --输入参数，整数类型
out result_1 bigint,            --输出结果1，大范围整数类型
out result_2 bigint             --输出结果2，大范围整数类型
)
RETURNS SETOF RECORD            --结果返回一组记录
AS $$
BEGIN
    result_1 = i + 1;           --计算第一个结果值，输入值+1
    result_2 = i * 10;          --计算第二个结果值，输入值乘以10
RETURN next;                    --返回当前计算结果并继续（允许返回多行）
END;
$$LANGUAGE plpgsql;             --指定函数使用PL/pgSQL语言编写
```

单次调用函数，返回结果：

```
SELECT * FROM compute(5);
```

result_1	result_2
6	50

生成一个包含1到3的序列，并对序列的每个值调用一次compute函数。

```
SELECT i, (compute(i)).*        --先调用compute函数，再展开函数返回的所有列
FROM generate_series(1, 3)      --生成一个包含1，2，3的临时序列，用于i的赋值
AS t(i);                        --将结果作为临时表，别名为t，列名为i
```

 版权所有 © 华为云计算技术有限公司

i	result_1	result_2
1	2	10
2	3	20
3	4	30

与一个表test_data的数据结合使用，表中的value列作为函数compute的入参值，计算结果：

```
DROP TABLE IF EXISTS test_data;
CREATE TABLE test_data (id serial, value int);
INSERT INTO test_data (value) VALUES (2), (4), (6);

SELECT id, value, (compute(value)).*
FROM test_data ORDER BY 1;
```

id	value	result_1	result_2
1	2	3	20
2	4	5	40
3	6	7	60

示例：创建一个返回结果包含多个输出参数的 SQL 语言函数

语法如下：

```
CREATE OR REPLACE FUNCTION func_dup_sql(
in int,                          --输入参数，整数类型
out f1 int,                       --输出参数1，整数类型
out f2 text)                       --输出参数2，文本类型
 AS $$
SELECT $1, CAST($1 AS text) || ' is text'        --函数体，查询返回两列，第一列是输入参数$1，第二列是将参数转为文本后拼接'is text'
 $$   LANGUAGE SQL;
```

执行函数

```
SELECT * FROM func_dup_sql(42);
```

f1	f2
42	42 is text

示例：创建 package 属性的重载函数：

定义了两个重载的函数，名为package_func_overload，它们位于名为package的包中。这两个函数根据参数类型的不同有不同的实现。

```
CREATE OR REPLACE FUNCTION package_func_overload(
col int,                                --输入参数1，整数类型
```

 版权所有 © 华为云计算技术有限公司

```
col2  int                                --输入参数2，整数类型
)
RETURN integer PACKAGE
AS
DECLARE
    col_type text;                  --声明局部变量
BEGIN
     col := 122;                    --将第一个参数的值固定为122
     dbms_output.put_line('two int parameters ' || col2);
     RETURN 0;                       --返回整数值0
END;
/

CREATE OR REPLACE FUNCTION package_func_overload(
col int,
col2 smallint
)
RETURN integer PACKAGE
AS
DECLARE
    col_type text;
BEGIN
     col := 122;
     dbms_output.put_line('two smallint parameters ' || col2);
     RETURN 0;
END;
/
```

函数下推逻辑及示例

函数下推逻辑分为两部分：

1. 函数体是否下推：通过函数属性（SHIPPABLE）判断。
2. 函数体内的语句是否下推：当函数体不下推的时候，函数体内的SQL语句进行独立的下推判断（优化器自动判断）。

调用函数的语句不包含用户表

当调用函数的语句不包含用户表的时候，这个语句总是在CN上执行，函数体不下推，此时执行逻辑如下：

1. 先执行函数：语句中没有用户表，这个语句总是在CN上执行。
2. 接着执行函数体的语句，如果不存在不下推的因素，那么总是可以下推的（走下推框架：FQS 或 stream计划）。

⚠ 注意

- 单独调用函数（调用函数的语句不包含用户表），函数中包含表关联查询的SQL，查询结果是准确的。
- 函数下推只有在函数调用语句中直接出现用户表的时候才有意义。
- 函数体里面的执行计划无法直接对外展示，可以设置enable_track_record_subsql=on记录函数体内部的语句。

示例：函数体为NOT SHIPPABLE，调用函数的语句不包含用户表。

```
CREATE TABLE student (
        id integer,
        name character varying(50),
        age integer
```

 版权所有 © 华为云计算技术有限公司

```
)
WITH (orientation=row, compression=no)
DISTRIBUTE BY HASH(id);

CREATE TABLE stuscore(
        id integer,
        score integer
)
WITH (orientation=row, compression=no)
DISTRIBUTE BY HASH(id);

INSERT INTO student VALUES(1,'bob',17),(3,'peter',18),(4,'david',15),(8,'alice',17);
INSERT INTO stuscore VALUES(1,88),(3,68),(8,100),(4,90);

--创建一个NOT SHIPPABLE的函数
CREATE OR REPLACE FUNCTION get_student_id(name character varying)
    RETURNS TABLE(id integer)
    LANGUAGE plpgsql
    VOLATILE NOT FENCED NOT SHIPPABLE
AS $function$
DECLARE
BEGIN
    RETURN QUERY (SELECT id FROM mytest.student WHERE name=$1);
END $function$;
```

查看单独调用函数的计划：

```
EXPLAIN VERBOSE SELECT * FROM get_student_id('alice');
```

从返回结果看，计划中没有函数是否下推的说明，因为函数下推只有在函数调用语句中直接出现用户表的时候才有意义。

 版权所有 © 华为云计算技术有限公司

```
QUERY PLAN
 id |            operation             | E-rows | E-distinct | E-width | E-costs
----+----------------------------------+--------+------------+---------+---------
  1 | ->  Function Scan on public.get_student_id |  1000 |            |       4 | 10.25


        Predicate Information (identified by plan id)
-------------------------------------------------------------
  1 --Function Scan on public.get_student_id
        Function Call: get_student_id('alice'::character varying)


      Targetlist Information (identified by plan id)
-------------------------------------------------------
  1 --Function Scan on public.get_student_id
        Output: id
```

调用函数的语句包含用户表

当调用函数的语句包含用户表的时候，语句是否下推由函数属性（SHIPPABLE）决定。

示例：函数属性为NOT SHIPPABLE，调用函数的语句包含用户表。

查看调用函数的计划，函数不可下推：

```
EXPLAIN VERBOSE SELECT get_student_id('alice') FROM stuscore;
```

 版权所有 © 华为云计算技术有限公司

```
 QUERY PLAN
 id |                operation                 | E-rows | E-distinct | E-width | E-costs
----+------------------------------------------+--------+------------+---------+---------
  1 | ->  Data Node Scan on stuscore "_REMOTE_TABLE_QUERY_" | 30000 |            |       0 | 157.35


               SQL Diagnostic Information

-------------------------------------------------
SQL is not plan-shipping
        reason: Function get_student_id() can not be shipped
```

示例：函数属性SHIPPABLE，调用函数的语句包含用户表。

```
--修改函数属性为SHIPPABLE。
CREATE OR REPLACE FUNCTION get_student_id(name character varying)
    RETURNS TABLE(id integer)
    LANGUAGE plpgsql
    VOLATILE NOT FENCED SHIPPABLE
AS $function$
DECLARE
BEGIN
    RETURN QUERY (SELECT id FROM mytest.student WHERE name=$1);
END $function$;
```

查看调用函数的计划，函数可下推。

```
EXPLAIN VERBOSE SELECT get_student_id('alice') FROM stuscore;
```

```
 QUERY PLAN
 id |              operation               | E-rows | E-distinct | E-width | E-costs
----+--------------------------------------+--------+------------+---------+---------
  1 | ->  Data Node Scan on "__REMOTE_FQS_QUERY__" |      0 |            |       0 | 0.00


                    Targetlist Information (identified by plan id)

------------------------------------------------------------------------------------------------
 1 --Data Node Scan on "__REMOTE_FQS_QUERY__"
       Output: (get_student_id('alice'::character varying))
       Node/s: All datanodes (v3_logical, bucket:32)
       Remote query: SELECT public.get_student_id('alice'::character varying) AS get_student_id FROM public.stuscore
```

 版权所有 © 华为云计算技术有限公司

相关链接

ALTER FUNCTION，DROP FUNCTION

13.7.2 ALTER FUNCTION

功能描述

修改自定义函数的属性。

注意事项

- 只有该函数的所有者或者授予了函数ALTER权限的用户才有权限执行ALTER FUNCTION，系统管理员默认拥有该权限。
- 要修改函数的所有者的用户必须是新拥有角色的直接或间接成员。
- 要修改函数的所有者，新所有者必须拥有该函数所属模式的CREATE权限。
- 要修改函数的所属模式，必须拥有新模式的CREATE权限。
- 如果函数中涉及对临时表相关的操作，则无法使用ALTER FUNCTION。

语法格式

- 修改自定义函数的附加参数。

```
ALTER FUNCTION function_name ( [ { [ argmode ] [ argname ] argtype} [, ...] ] )
    action [ ... ] [ RESTRICT ];
```

其中附加参数action子句语法为：

```
{CALLED ON NULL INPUT | RETURNS NULL ON NULL INPUT | STRICT}
| {IMMUTABLE | STABLE | VOLATILE}
| {SHIPPABLE | NOT SHIPPABLE}
| {NOT FENCED | FENCED}
| {MASKED | NOT MASKED}
| [ NOT ] LEAKPROOF
| { [ EXTERNAL ] SECURITY INVOKER | [ EXTERNAL ] SECURITY DEFINER }
| AUTHID { DEFINER | CURRENT_USER }
| COST execution_cost
| ROWS result_rows
| SET configuration_parameter { { TO | = } { value | DEFAULT }| FROM CURRENT}
| RESET {configuration_parameter | ALL}
```

- 修改自定义函数的名称，支持新函数名前带有原函数所在的schema名，不支持同时修改schema名。

```
ALTER FUNCTION funname ( [ { [ argmode ] [ argname ] argtype} [, ...] ] )
    RENAME TO new_name;
ALTER FUNCTION funname ( [ { [ argmode ] [ argname ] argtype} [, ...] ] )
    RENAME TO schema.new_name;
```

- 修改自定义函数的所属者。

```
ALTER FUNCTION funname ( [ { [ argmode ] [ argname ] argtype} [, ...] ] )
    OWNER TO new_owner;
```

- 修改自定义函数的模式。

```
ALTER FUNCTION funname ( [ { [ argmode ] [ argname ] argtype} [, ...] ] )
    SET SCHEMA new_schema;
```


版权所有 © 华为云计算技术有限公司

参数说明

表 13-133 ALTER FUNCTION 参数说明

参数	描述	取值范围
function_name	需要修改的函数名称。	已存在的函数名。
argmode	参数的模式。标识该参数是输入、输出还是输入输出参数。	IN、OUT或者IN OUT。 缺省值是IN。
argname	参数名称。	字符串，需符合**标识符命名规范**。
argtype	函数的参数类型。	有效的类型，请参考**数据类型**。
new_name	函数的新名称。	字符串，需符合**标识符命名规范**。
new_owner	函数的新所有者。	有效的用户名或角色名。
new_schema	函数的新模式。	已存在的模式。
CALLED ON NULL INPUT | RETURNS NULL ON NULL INPUT | STRICT	• CALLED ON NULL INPUT，表明该函数的某些参数是NULL时可以按照正常的方式调用。缺省时与指定此参数的作用相同。 • STRICT，如果函数的某个参数是NULL时，此函数总是返回NULL。如果指定了该参数，当参数为NULL值时不被调用而只是自动返回一个NULL结果。 • RETURNS NULL ON NULL INPUT和STRICT的功能相同。	-
IMMUTABLE | STABLE | VOLATILE	• IMMUTABLE，表示该函数在给出同样的参数值时总是返回同样的结果。 • STABLE，表示该函数不能修改数据库，对相同参数值，在同一次表扫描里，该函数的返回值不变，但是返回值可能在不同SQL语句之间变化。 • VOLATILE，表示该函数值可以在一次表扫描内改变，不会做任何优化。	IMMUTABLE、STABLE和VOLATILE的详细介绍可参考**DWS函数易变属性的三种行为**。

版权所有 © 华为云计算技术有限公司

参数	描述	取值范围
SHIPPABLE | NOT SHIPPABLE	表示该函数是否可以下推到DN上执行。 ● 对于IMMUTABLE类型的函数，函数始终可以下推到DN上执行。 ● 对于STABLE/VOLATILE类型的函数，仅当函数的属性是SHIPPABLE的时候，函数可以下推到DN执行。	-
MASKED | NOT MASKED	表示该函数在guc参数enable_redactcol_computable开启的情况下，是否进行脱敏。系统函数支持修改MASKED、NOT MASKED选项，不支持和ALTER FUNCTION中的其他action选项一起修改。	-
LEAKPROOF	表示该函数没有副作用，指出参数只包括返回值。LEAKPROOF只能由系统管理员设置。	-
EXTERNAL（可选）	允许使用EXTERNAL关键字实现SQL兼容，此功能适用于所有函数而不仅是外部函数。	-
SECURITY INVOKER/AUTHID CURRENT_USER	表明该函数将以调用它的用户的权限执行。缺省时与指定此参数的作用相同。 SECURITY INVOKER和AUTHID CURRENT_USER的功能相同。	-
SECURITY DEFINER/AUTHID DEFINER	声明该函数将以创建它的用户的权限执行。缺省时与指定此参数的作用相同。 AUTHID DEFINER和SECURITY DEFINER的功能相同。	-
COST execution_cost	用来估计函数的执行成本。 execution_cost以cpu_operator_cost为单位。	正数。
ROWS result_rows	估计函数返回的行数。用于函数返回的是一个集合。	正数，默认值为1000行。

 版权所有 © 华为云计算技术有限公司

参数	描述	取值范围
configuration_parameter	• value 把指定的数据库会话参数值设置为给定的值。 – DEFAULT或RESET，表示新的会话中使用系统的缺省设置。 – OFF，表示关闭设置。 指定默认值。 • from current 取当前会话中的值设置为configuration_parameter的值。	字符串。

示例

创建计算两个整数和的函数func_add_sql2并返回结果。如果输入为null，则返回null。

```
DROP FUNCTION IF EXISTS func_add_sql2;
CREATE FUNCTION func_add_sql2(num1 integer, num2 integer) RETURN integer
AS
BEGIN
RETURN num1 + num2;
END;
/
;
```

修改函数func_add_sql2的执行规则为IMMUTABLE，即参数不变时返回相同结果。

```
ALTER FUNCTION func_add_sql2(INTEGER, INTEGER) IMMUTABLE;
```

将函数func_add_sql2的名称修改为add_two_number。

```
ALTER FUNCTION func_add_sql2(INTEGER, INTEGER) RENAME TO add_two_number;
```

将函数add_two_number的所有者修改为dbadmin。

```
ALTER FUNCTION add_two_number(INTEGER, INTEGER) OWNER TO dbadmin;
```

相关链接

CREATE FUNCTION，**DROP FUNCTION**

13.7.3 DROP FUNCTION

功能描述

删除一个已存在的函数。

注意事项

- 如果所删除的函数为重载函数，则删除时需要指明该函数的参数类型，如为非重载函数，则可直接指定函数名进行删除。
- 如果函数中涉及对临时表相关操作，则无法使用DROP FUNCTION删除函数。

版权所有 © 华为云计算技术有限公司

语法格式

```
DROP FUNCTION [ IF EXISTS ] function_name
[ ( [ {[ argmode ] [ argname ] argtype} [, ...] ] ) [ CASCADE | RESTRICT ] ];
```

参数说明

表 13-134 DROP FUNCTION 参数说明

参数	描述	取值范围
IF EXISTS	如果指定的外表不存在，则发出一个notice而不是抛出一个错误。	-
function_name	要删除的函数名称。	已存在的函数名。
argmode	函数参数的模式。	-
argname	函数参数的名称。	-
argtype	函数参数的数据类型。	-
CASCADE | RESTRICT	当执行删除操作时，如何处理依赖对象中的相关数据。	• CASCADE：级联删除依赖于函数的对象（比如操作符）。 • RESTRICT：如果有任何依赖对象存在，则拒绝删除该函数（缺省行为）。

示例

删除名称为add_two_number的函数：

```
DROP FUNCTION IF EXISTS add_two_number;
```

相关链接

ALTER FUNCTION，**CREATE FUNCTION**

13.7.4 CREATE PROCEDURE

功能描述

创建一个新的存储过程。

存储过程（PROCEDURE）是由一组预编译的SQL语句组成的集合，可以在数据库中进行存储并反复调用。

注意事项

- 创建存储过程时，请显式指定表对象的所属模式（如schema.table），否则可能引发执行异常。
- 创建存储过程时，存储过程内部通过SET语句设置的current_schema/search_path无效，函数search_path和current_schema执行完成后与函数执行前保持一致。

版权所有 © 华为云计算技术有限公司

- 创建存储过程时，禁止在avg函数外面嵌套其他聚合函数或系统函数。
- 创建存储过程时，参数或返回值带有精度，不进行精度检测。
- 在多个CN的集群中，禁止将存储过程的入参/出参设置为临时表类型（因为在非当前连接CN上执行创建存储过程时，无法单独根据表名获取到正确的临时schema，进而无法获取到准确的表类型。）
- 存储过程可以有多个返回值，也可以没有返回值。无返回值的存储过程调用后会有空的回显。
- 如果存储过程参数中带有出参：SELECT调用存储过程必须缺省出参；CALL调用存储过程适配Oracle；调用非重载函数时必须指定出参；对于重载的PACKAGE函数，out参数可以缺省。具体信息参见**CALL**的示例。
- 存储过程指定PACKAGE属性时支持重载。
- 存储过程不支持分段提交。
- 函数创建也适用于存储过程，更多内容请参考**CREATE FUNCTION**。创建存储过程后查询定义时会返回CREATE FUNCTION的定义。

语法格式

```
CREATE [ OR REPLACE ] PROCEDURE procedure_name
    [ ( {[ argmode ] [ argname ] argtype [ { DEFAULT | := | = } expression ]}[,...]) ]
    [
      { IMMUTABLE | STABLE | VOLATILE }
      | { SHIPPABLE | NOT SHIPPABLE }
      | {PACKAGE}
      | [ NOT ] LEAKPROOF
      | { CALLED ON NULL INPUT | RETURNS NULL ON NULL INPUT | STRICT }
      | {[ EXTERNAL ] SECURITY INVOKER | [ EXTERNAL ] SECURITY DEFINER | AUTHID DEFINER | AUTHID
CURRENT_USER}
      | COST execution_cost
      | ROWS result_rows
      | SET configuration_parameter { [ TO | = ] value | FROM CURRENT }
    ][ ... ]
 { IS | AS }
plsql_body
/
```

参数说明

表 13-135 CREATE PROCEDURE 参数说明

参数	描述	取值范围
OR REPLACE	当存在同名的存储过程时，替换原来的定义。	-
procedure_name	要创建的存储过程名字，可以用模式修饰。	字符串，要符合标识符的命名规范。参见**标识符命名规范**。

版权所有 © 华为云计算技术有限公司

参数	描述	取值范围
argmode	参数的模式。	IN，OUT，INOUT或VARIADIC。 • IN：输入参数，缺省值是IN。 • OUT：输出参数，只有OUT模式的参数后面能跟VARIADIC。 • INOUT：既是输入又是输出参数。 • OUT VARIADIC: 表示输出参数，且声明为数组类型。 **注意** OUT和INOUT模式的参数不能用在RETURNS TABLE的存储过程定义中。
argname	参数的名字。	字符串，要符合标识符的命名规范。参见**标识符命名规范**。 **说明** argname和argmode的顺序没有严格要求，推荐按照argname、argmode、argtype的顺序使用。
argtype	参数的数据类型。	可用的数据类型。
IMMUTABLE | STABLE | VOLATILE	行为约束可选项。各参数的功能与CREATE FUNCTION类似，详细说明见**CREATE FUNCTION**。	-

版权所有 © 华为云计算技术有限公司

参数	描述	取值范围
SHIPPABLE | NOT SHIPPABLE	表示该存储过程是否可以下推到DN上执行。	• 对于IMMUTABLE类型的存储过程，存储过程始终可以下推到DN上执行。 • 对于STABLE/VOLATILE类型的存储过程，仅当存储过程的属性是SHIPPABLE的时候，存储过程可以下推到DN执行。用户在定义存储过程的SHIPPABLE属性时也需特别慎重，SHIPPABLE意味着整个存储过程会下推到DN上执行，如果设置不当，会导致结果错误等严重问题。 与定义IMMUTABLE属性一样，SHIPPABLE属性的定义也有诸多约束，简单来说就是存储过程内不能有不可下推的因素，存储过程下推到单DN执行后，存储过程内部的计算逻辑仅依赖当前DN的数据集合。 举例如下： 1. 如果存储过程内部引用了表，并且表为HASH分布，那么该存储过程通常不能定义为SHIPPABLE。 2. 存储过程内部有不可下推的因素，例如不支持下推的存储过程、语法等，那么该存储过程不能定义为SHIPPABLE。 3. 存储过程内部的计算过程可能需要跨DN数据，这种情况该存储过程通常不能定义为SHIPPABLE，例如一些聚合运算等。
PACKAGE	表示该存储过程是否支持重载。PostgreSQL风格的存储过程本身就支持重载，此参数主要是针对Oracle风格的存储过程。	• 不允许package存储过程和非package存储过程重载或者替换。 • package存储过程不支持VARIADIC类型的参数。 • 不允许修改存储过程的package属性。
LEAKPROOF	指出该存储过程的参数只包括返回值。LEAKPROOF只能由系统管理员设置。	-

 版权所有 © 华为云计算技术有限公司

参数	描述	取值范围
CALLED ON NULL INPUT	表明该存储过程的某些参数是NULL时可以按照正常的方式调用。该参数可以省略。	-
RETURNS NULL ON NULL INPUT | STRICT	STRICT用于指定如果存储过程的某个参数是NULL，此存储过程总是返回NULL。如果声明了这个参数，当有NULL值参数时该存储过程不会被执行；而只是自动返回一个NULL结果。 RETURNS NULL ON NULL INPUT和STRICT的功能相同。	-
EXTERNAL	目的是和SQL兼容，是可选的，这个特性适合于所有存储过程，而不仅是外部存储过程。	-
SECURITY INVOKER | AUTHID CURRENT_USER	表明该存储过程将带着调用它的用户的权限执行。该参数可以省略。 SECURITY INVOKER和AUTHID CURRENT_USER的功能相同。	-
SECURITY DEFINER | AUTHID DEFINER	声明该存储过程将以创建它的用户的权限执行。 AUTHID DEFINER和SECURITY DEFINER的功能相同。	-
COST execution_cost	用来估计存储过程的执行成本。execution_cost以cpu_operator_cost为单位。	正数。
ROWS result_rows	估计存储过程返回的行数。用于存储过程返回的是一个集合。	正数，默认值是1000行。

 版权所有 © 华为云计算技术有限公司

参数	描述	取值范围
configuration_parameter	该存储过程配置的参数值。	• **value** 把指定的数据库会话参数值设置为给定的值。如果value是DEFAULT或者RESET，则在新的会话中使用系统的缺省设置。OFF关闭设置。 取值范围：字符串 – DEFAULT – OFF – RESET 指定默认值。 • **from current** 取当前会话中的值设置为configuration_parameter的值。
plsql_body	PL/SQL存储过程体。	**须知** 当在存储过程体中进行创建用户等涉及用户密码相关操作时，系统表及csv日志中会记录密码的明文。因此不建议用户在存储过程体中进行涉及用户密码的相关操作。

示例

- 创建存储过程prc_add，param1为输入参数，param2为输入输出参数，计算param1+param2，并将结果赋值给param2，并输出结果。

```
CREATE OR REPLACE PROCEDURE prc_add
(
    param1 IN INTEGER,
    param2 IN OUT INTEGER
)
AS
BEGIN
   param2:= param1 + param2;
   dbms_output.put_line('result is: '||to_char(param2));
END;
/
```

  调用存储过程prc_add。

```
CALL prc_add(2,3);
```

- 创建存储过程pro_variadic，param1参数模式为VARIADIC，将传入的整数数组转换为TEXT格式，并通过param2输出。

```
CREATE OR REPLACE PROCEDURE pro_variadic (param1 VARIADIC int4[],param2 OUT TEXT)
AS
BEGIN
    param2:= param1::text;
END;
/
```

  执行存储过程pro_variadic。

```
SELECT pro_variadic(VARIADIC param1=> array[1,2,3,4]);
```

- 创建带有PACKAGE属性的存储过程package_proc_overload。

 版权所有 © 华为云计算技术有限公司

```
CREATE OR REPLACE PROCEDURE package_proc_overload(col int, col2 out varchar)
PACKAGE
AS
DECLARE
    col_type text;
BEGIN
     col2 := '122';
        dbms_output.put_line('two varchar parameters ' || col2);
END;
/
```

相关链接

删除存储过程参见**DROP PROCEDURE**，调用存储过程参见**CALL**。

13.7.5 DROP PROCEDURE

功能描述

删除已存在的存储过程。

注意事项

无。

语法格式

```
DROP PROCEDURE [ IF EXISTS  ] procedure_name ;
```

参数说明

表 13-136 DROP PROCEDURE 参数说明

参数	描述	取值范围
IF EXISTS	如果指定的存储过程不存在，发出一个notice而不是抛出一个错误。	-
procedure_name	要删除的存储过程名字。	已存在的存储过程名。

示例

删除存储过程：

```
DROP PROCEDURE IF EXISTS prc_add;
```

相关链接

CREATE PROCEDURE

 版权所有 © 华为云计算技术有限公司

13.7.6 CREATE TRIGGER

功能描述

创建一个新触发器。 触发器将与指定的表或视图关联，并在特定条件下执行指定的函数。

注意事项

- 当前仅支持在普通行存表上创建触发器，不支持在列存表、临时表、unlogged表等类型表上创建触发器。
- 如果为同一事件定义了多个相同类型的触发器，则按触发器的名称字母顺序触发它们。
- 一个触发器只能作用在一张表上，对创建的触发器数量无限制但一个表上的触发器越多，性能消耗越大。
- 触发器常用于多表间数据关联同步场景，对SQL执行性能影响较大，不建议在大数据量同步及对性能要求高的场景中使用。
- 当触发器满足如下条件时，触发语句能和触发器一起下推到DN执行并提升触发器执行性能：
  - 开关enable_trigger_shipping和enable_fast_query_shipping开启（默认均开启）。
  - 源表触发器使用的触发器函数为plpgsql类型（推荐类型）。
  - 源表与触发表分布键的类型、数量完全相同，均为行存表，且所属的nodegroup相同。
  - 原INSERT/UPDATE/DELETE语句条件中包含所有分布键与NEW/OLD等值比较表达式。
  - 原INSERT/UPDATE/DELETE语句在没有触发器的情况下原本就能query shipping。
  - 源表上只有INSERT/UPDATE/DELETE AFTER/BEFORE FOR EACH ROW六类触发器，且所有触发器都可下推。
- 关于触发器种类：
  - INSTEAD OF的触发器必须标记为FOR EACH ROW，并且只能在视图上定义。
  - BEFORE和AFTER触发器作用在视图上时，只能标记为FOR EACH STATEMENT。
  - TRUNCATE类型触发器仅限FOR EACH STATEMENT。

语法格式

```
CREATE [ CONSTRAINT ] TRIGGER trigger_name { BEFORE | AFTER | INSTEAD OF } { event [ OR ... ] }
   ON table_name
   [ FROM referenced_table_name ]
   { NOT DEFERRABLE | [ DEFERRABLE ] { INITIALLY IMMEDIATE | INITIALLY DEFERRED } }
   [ FOR [ EACH ] { ROW | STATEMENT } ]
   [ WHEN ( condition ) ]
   EXECUTE PROCEDURE function_name ( arguments );
```

其中event包含以下几种：

```
   INSERT
   UPDATE [ OF column_name [, ... ] ]
```

 版权所有 © 华为云计算技术有限公司

```
DELETE
TRUNCATE
```

参数说明

表 13-137 CREATE TRIGGER 参数说明

参数	描述	取值范围
CONSTRAINT	指定此参数将创建约束触发器，即触发器作为约束来使用。除了可以使用SET CONSTRAINTS调整触发器触发的时间之外，这与常规触发器相同。 约束触发器必须是AFTER ROW触发器。	-
trigger_name	要创建的触发器的名称。 该名称不能限定模式，因为触发器自动继承其所在表的模式，且同一个表的触发器不能重名。 对于约束触发器，使用**SET CONSTRAINTS**修改触发器行为时也使用此名称。	字符串，需符合**标识符命名规范**。
BEFORE	触发器函数是在触发事件发生前执行。	-
AFTER	触发器函数是在触发事件发生后执行，约束触发器只能指定为AFTER。	-
INSTEAD OF	触发器函数直接替代触发事件。	-
event	启动触发器的事件。	取值范围包括：INSERT、UPDATE、DELETE或TRUNCATE，也可以通过OR同时指定多个触发事件。 对于UPDATE事件类型，可以使用下面语法指定列： `UPDATE OF column_name1 [, column_name2 ... ]` 表示只有这些列作为UPDATE语句的目标列时，才会启动触发器，但是INSTEAD OF UPDATE类型不支持指定列信息
table_name	需要创建触发器的表名称。	数据库中已经存在的表名称。
referenced_table_name	约束引用的另一个表的名称。 只能为约束触发器指定，常见于外键约束。由于当前不支持外键，因此不建议使用。	数据库中已经存在的表名称。

版权所有 © 华为云计算技术有限公司

参数	描述	取值范围
DEFERRABLE | NOT DEFERRABLE	约束触发器的启动时机，仅作用于约束触发器。这两个关键字设置该约束是否可推迟。 详细介绍请参见**CREATE TABLE**。	-
INITIALLY IMMEDIATE | INITIALLY DEFERRED	如果约束是可推迟的，则这个子句声明检查约束的缺省时间，仅作用于约束触发器。 详细介绍请参见**CREATE TABLE**。	-
FOR EACH ROW | FOR EACH STATEMENT	触发器的触发频率。 ● FOR EACH ROW是指该触发器是受触发事件影响的每一行触发一次。 ● FOR EACH STATEMENT是指该触发器是每个SQL语句只触发一次。 未指定时默认值为FOR EACH STATEMENT。约束触发器只能指定为FOR EACH ROW。	-
condition	决定是否实际执行触发器函数的条件表达式。当指定WHEN时，只有在条件返回true时才会调用该函数。 ● 在FOR EACH ROW触发器中，WHEN条件可以通过分别写入OLD.column_name或NEW.column_name来引用旧行或新行值的列。 需要注意，INSERT触发器不能引用OLD和DELETE触发器不能引用NEW。 ● INSTEAD OF触发器不支持WHEN条件。 ● WHEN表达式不能包含子查询。 ● 对于约束触发器，WHEN条件的评估不会延迟，而是在执行更新操作后立即发生。 如果条件返回值不为true，则触发器不会排队等待延迟执行。	-
function_name	用户定义的函数，必须声明为不带参数并返回类型为触发器，在触发器触发时执行。	-
arguments	执行触发器时要提供给函数的可选的以逗号分隔的参数列表。参数是文字字符串常量，简单的名称和数字常量也可以写在这里，但它们都将被转换为字符串。 请检查触发器函数的实现语言的描述，以了解如何在函数内访问这些参数。	-

版权所有 © 华为云计算技术有限公司

表 13-138 表和视图上支持的触发器种类：

触发时机	触发事件	行级	语句级
BEFORE	INSERT/UPDATE/DELETE	表	表和视图
	TRUNCATE	不支持	表
AFTER	INSERT/UPDATE/DELETE	表	表和视图
	TRUNCATE	不支持	表
INSTEAD OF	INSERT/UPDATE/DELETE	视图	不支持
	TRUNCATE	不支持	不支持

表 13-139 PLPGSQL 类型触发器函数特殊变量：

变量名	变量含义
NEW	INSERT及UPDATE操作涉及tuple信息中的新值，对DELETE为空。
OLD	UPDATE及DELETE操作涉及tuple信息中的旧值，对INSERT为空。
TG_NAME	触发器名称。
TG_WHEN	触发器触发时机（BEFORE/AFTER/INSTEAD OF）。
TG_LEVEL	触发频率（ROW/STATEMENT）。
TG_OP	触发操作（INSERT/UPDATE/DELETE/TRUNCATE）。
TG_RELID	触发器所在表OID。
TG_RELNAME	触发器所在表名（已废弃，现用TG_TABLE_NAME替代）。
TG_TABLE_NAME	触发器所在表名。
TG_TABLE_SCHEMA	触发器所在表的SCHEMA信息。
TG_NARGS	触发器函数参数个数。
TG_ARGV[]	触发器函数参数列表。

示例

创建源表及触发表。

```
DROP TABLE IF EXISTS test_trigger_src_tbl;
DROP TABLE IF EXISTS test_trigger_des_tbl;
```

 版权所有 © 华为云计算技术有限公司

```
CREATE TABLE test_trigger_src_tbl(id1 INT, id2 INT, id3 INT);
CREATE TABLE test_trigger_des_tbl(id1 INT, id2 INT, id3 INT);
```

创建触发器函数tri_insert_func()。

```
CREATE OR REPLACE FUNCTION tri_insert_func() RETURNS TRIGGER AS
        $$
        DECLARE
        BEGIN
                INSERT INTO test_trigger_des_tbl VALUES(NEW.id1, NEW.id2, NEW.id3);
                RETURN NEW;
        END
        $$ LANGUAGE PLPGSQL;
```

创建触发器函数tri_update_func()。

```
CREATE OR REPLACE FUNCTION tri_update_func() RETURNS TRIGGER AS
        $$
        DECLARE
        BEGIN
                UPDATE test_trigger_des_tbl SET id3 = NEW.id3 WHERE id1=OLD.id1;
                RETURN OLD;
        END
        $$ LANGUAGE PLPGSQL;
```

创建触发器函数tri_delete_func()。

```
CREATE OR REPLACE FUNCTION tri_delete_func() RETURNS TRIGGER AS
        $$
        DECLARE
        BEGIN
                DELETE FROM test_trigger_des_tbl WHERE id1=OLD.id1;
                RETURN OLD;
        END
        $$ LANGUAGE PLPGSQL;
```

创建INSERT触发器。

```
DROP TRIGGER IF EXISTS insert_trigger ON test_trigger_src_tbl;
CREATE TRIGGER insert_trigger
        BEFORE INSERT ON test_trigger_src_tbl
        FOR EACH ROW
        EXECUTE PROCEDURE tri_insert_func();
```

创建UPDATE触发器。

```
DROP TRIGGER IF EXISTS update_trigger ON test_trigger_src_tbl;
CREATE TRIGGER update_trigger
        AFTER UPDATE ON test_trigger_src_tbl
        FOR EACH ROW
        EXECUTE PROCEDURE tri_update_func();
```

创建DELETE触发器。

```
DROP TRIGGER IF EXISTS delete_trigger ON test_trigger_src_tbl;
CREATE TRIGGER delete_trigger
        BEFORE DELETE ON test_trigger_src_tbl
        FOR EACH ROW
        EXECUTE PROCEDURE tri_delete_func();
```

相关链接

ALTER TRIGGER，**DROP TRIGGER**，**ALTER TABLE**

 版权所有 © 华为云计算技术有限公司

13.7.7 ALTER TRIGGER

功能描述

修改触发器定义。

注意事项

只有触发器所在表的所有者可以执行ALTER TRIGGER操作，系统管理员默认拥有此权限。

语法格式

```
ALTER TRIGGER trigger_name ON table_name RENAME TO new_name;
```

参数说明

表 13-140 ALTER TRIGGER 参数说明

参数	描述	取值范围
trigger_name	需要修改的触发器的名称。	有效的触发器名。
table_name	需要修改的触发器所在的表名称。	触发器所在的有效表名。
new_name	新的触发器名称。	字符串，需符合**标识符命名规范**。

示例

创建示例表test_trigger_src_tbl和test_trigger_des_tbl。

```
DROP TABLE IF EXISTS test_trigger_src_tbl;
DROP TABLE IF EXISTS test_trigger_des_tbl;

CREATE TABLE test_trigger_src_tbl(id1 INT, id2 INT, id3 INT);
CREATE TABLE test_trigger_des_tbl(id1 INT, id2 INT, id3 INT);
```

创建触发器函数tri_insert_func()。

```
DROP FUNCTION IF EXISTS tri_insert_func;
CREATE OR REPLACE FUNCTION tri_insert_func() RETURNS TRIGGER AS
        $$
        DECLARE
        BEGIN
                INSERT INTO test_trigger_des_tbl VALUES(NEW.id1, NEW.id2, NEW.id3);
                RETURN NEW;
        END
        $$ LANGUAGE PLPGSQL;
```

创建触发器insert_trigger。

```
CREATE TRIGGER insert_trigger
        BEFORE INSERT ON test_trigger_src_tbl
        FOR EACH ROW
        EXECUTE PROCEDURE tri_insert_func();
```

修改触发器insert_trigger。

版权所有 © 华为云计算技术有限公司

```
ALTER TRIGGER insert_trigger ON test_trigger_src_tbl RENAME TO delete_trigger_renamed;
```

禁用触发器delete_trigger_renamed。

```
ALTER TABLE test_trigger_src_tbl DISABLE TRIGGER delete_trigger_renamed;
```

禁用表test_trigger_src_tbl所有触发器。

```
ALTER TABLE test_trigger_src_tbl DISABLE TRIGGER ALL;
```

相关链接

DROP TRIGGER，**DROP TRIGGER**，**ALTER TABLE**

13.7.8 DROP TRIGGER

功能描述

删除触发器。

注意事项

只有触发器的所有者可以执行DROP TRIGGER操作，系统管理员默认拥有此权限。

语法格式

```
DROP TRIGGER [ IF EXISTS ] trigger_name ON table_name [ CASCADE | RESTRICT ];
```

参数说明

表 13-141 DROP TRIGGER 参数说明

参数	描述	取值范围
IF EXISTS	如果指定的触发器不存在，则发出一个notice而不是抛出一个错误。	-
trigger_name	要删除的触发器名字。	已存在的触发器。
table_name	要删除的触发器所在的表名称。	已存在的含触发器的表。
CASCADE | RESTRICT	当执行删除操作时，如何处理依赖对象中的相关数据。	• CASCADE：级联删除依赖此触发器的对象。 • RESTRICT：如果有依赖对象存在，则拒绝删除此触发器。此选项为缺省值。

示例

删除触发器insert_trigger：

```
DROP TRIGGER IF EXISTS insert_trigger ON test_trigger_src_tbl;
```


版权所有 © 华为云计算技术有限公司

相关链接

CREATE TRIGGER，ALTER TRIGGER，ALTER TABLE

13.7.9 FETCH

功能描述

FETCH通过已创建的游标来检索数据。

每个游标都有一个供FETCH使用的关联位置。游标的关联位置可以在查询结果的第一行之前，或者在结果中的任意行，或者在结果的最后一行之后：

- 游标刚创建完之后，关联位置在第一行之前的。
- 在抓取了一些移动行之后，关联位置在检索到的最后一行上。
- 如果FETCH抓取完了所有可用行，它就停在最后一行后面，或者在反向抓取的情况下是停在第一行前面。
- FETCH ALL或FETCH BACKWARD ALL将总是把游标的关联位置放在最后一行或者在第一行前面。

注意事项

- 如果游标定义了NO SCROLL，则不允许使用例如FETCH BACKWARD之类的反向抓取。
- NEXT，PRIOR，FIRST，LAST，ABSOLUTE，RELATIVE形式在恰当地移动游标之后抓取一条记录。如果后面没有数据行，就返回一个空的结果，此时游标就会停在查询结果的最后一行之后（向后查询时）或者第一行之前（向前查询时）。
- FORWARD和BACKWARD形式在向前或者向后移动的过程中抓取指定的行数，然后把游标定位在最后返回的行上；或者是，如果count大于可用的行数，则在所有行之后（向后查询时）或者之前（向前查询时）。
- RELATIVE 0，FORWARD 0，BACKWARD 0都要求在不移动游标的前提下抓取当前行，也就是重新抓取最近刚抓取过的行。除非游标定位在第一行之前或者最后一行之后，这个动作都应该成功，而在那两种情况下，不返回任何行。
- 当FETCH的游标上涉及列存表时，不支持BACKWARD、PRIOR、FIRST等涉及反向获取操作。

语法格式

```
FETCH [ direction { FROM | IN } ] cursor_name;
```

其中direction子句为可选参数。

```
NEXT
  | PRIOR
  | FIRST
  | LAST
  | ABSOLUTE count
  | RELATIVE count
  | count
  | ALL
  | FORWARD
  | FORWARD count
  | FORWARD ALL
  | BACKWARD
  | BACKWARD count
  | BACKWARD ALL
```


版权所有 © 华为云计算技术有限公司

参数说明

表 13-142 FETCH 参数说明

参数	描述	取值范围
direction_clause	定义抓取数据的方向。	请参见**表13-143**。
{ FROM | IN } cursor_name	使用关键字FROM或IN指定游标名称。	已创建的游标的名称。

表 13-143 direction_clause 取值范围

取值范围	描述	备注
NEXT（缺省值）	从当前关联位置开始，抓取下一行。	-
PRIOR	从当前关联位置开始，抓取上一行。	-
FIRST	抓取查询的第一行（和ABSOLUTE 1相同）。	-
LAST	抓取查询的最后一行（和ABSOLUTE -1相同）。	-
ABSOLUTE count	抓取查询中第count行。 ABSOLUTE抓取不会比用相对位移移动到需要的数据行更快，因为下层的实现必须遍历所有中间的行。	count取值范围：有符号的整数 • count为正数，就从查询结果的第一行开始，抓取第count行。当count小于当前游标位置时，涉及到rewind操作，暂不支持。 • count为负数或0，涉及到反向扫描操作，暂不支持。
RELATIVE count	从当前关联位置开始，抓取随后或前面的第count行。	count取值范围：有符号的整数 • count为正数就抓取当前关联位置之后的第count行。 • count为负数，涉及到反向扫描操作，暂不支持。 • 如果有数据，RELATIVE 0重新抓取当前行。
count	抓取随后的count行（和FORWARD count一样）。	-

版权所有 © 华为云计算技术有限公司

取值范围	描述	备注
ALL	从当前关联位置开始，抓取所有剩余的行（和FORWARD ALL一样）。	-
FORWARD	抓取下一行（和NEXT一样）。	-
FORWARD count	与RELATIVE count的效果相同，从当前关联位置开始，抓取随后或前面的第count行。	-
FORWARD ALL	从当前关联位置开始，抓取所有剩余行。	-
BACKWARD	从当前关联位置开始，抓取前面一行（和PRIOR一样）。	-
BACKWARD count	从当前关联位置开始，抓取前面的count行（向后扫描）。	count取值范围：有符号的整数 ● count为正数就抓取当前关联位置之前的第count行。 ● count为负数就抓取当前关联位置之后的第abs(count)行。 ● 如果有数据，BACKWARD 0重新抓取当前行。
BACKWARD ALL	从当前关联位置开始，抓取所有前面的行（向后扫描）。	-

示例

示例一：SELECT语句，用一个游标读取一个表。

准备表数据。

```
DROP TABLE IF EXISTS customer_address;
CREATE TABLE customer_address
(ca_address_sk INT,
ca_address_id VARCHAR(16),
ca_street_number VARCHAR(10),
ca_street_name VARCHAR(60),
ca_street_type VARCHAR(15),
ca_suite_number VARCHAR(10));

INSERT INTO customer_address VALUES
(1, 'ID1', '100', 'Main', 'St', 'A1'),
(2, 'ID2', '200', 'Oak', 'Ave', 'B2'),
(3, 'ID3', '300', 'Pine', 'Blvd', 'C3');
```

建立一个名为cursor1的游标：

```
START TRANSACTION;
CURSOR cursor1 FOR SELECT * FROM customer_address ORDER BY 1;
```

抓取头3行到游标cursor1里。

```
FETCH FORWARD 3 FROM cursor1;
```

 版权所有 © 华为云计算技术有限公司

```
gaussdb=> FETCH FORWARD 3 FROM cursor1;
 ca_address_sk | ca_address_id | ca_street_number | ca_street_name | ca_street_type | ca_suite_number
---------------+---------------+------------------+----------------+----------------+-----------------
             1 | ID1           | 100              | Main           | St             | A1
             2 | ID2           | 200              | Oak            | Ave            | B2
             3 | ID3           | 300              | Pine           | Blvd           | C3
(3 rows)
```

示例二：VALUES子句，用一个游标读取VALUES子句中的内容。

建立一个名为cursor2的游标：

```
CURSOR cursor2 FOR VALUES(1,2),(0,3) ORDER BY 1;
```

抓取头2行到游标cursor2里：

```
FETCH FORWARD 2 FROM cursor2;
```

```
gaussdb=> FETCH FORWARD 2 FROM cursor2;
 column1 | column2
---------+---------
       0 |       3
       1 |       2
(2 rows)
```

相关链接

CLOSE，**MOVE**

13.7.10 MOVE

功能描述

MOVE在不检索数据的情况下重新定位一个游标。MOVE的作用类似于**FETCH**命令，但只是重定位游标而不返回行。

注意事项

无。

语法格式

```
MOVE [ direction [ FROM | IN ] ] cursor_name;
```

其中direction子句为可选参数。

```
NEXT
  | PRIOR
  | FIRST
  | LAST
  | ABSOLUTE count
  | RELATIVE count
  | count
  | ALL
  | FORWARD
  | FORWARD count
  | FORWARD ALL
  | BACKWARD
  | BACKWARD count
  | BACKWARD ALL
```


版权所有 © 华为云计算技术有限公司

参数说明

MOVE命令的参数与FETCH的相同，详细请参见FETCH的**参数说明**。

说明

成功完成时，MOVE命令将返回一个“MOVE count”的标签，count是一个使用相同参数的FETCH命令会返回的行数（可能为零）。

示例

创建表reason，并向表中插入数据。

```
DROP TABLE IF EXISTS reason;
CREATE TABLE reason
(
  a    int primary key,
  b    int,
  c    int
);

INSERT INTO reason VALUES (1, 2, 3);
```

开始一个事务：

```
START TRANSACTION;
```

定义一个名为cursor1的游标：

```
CURSOR cursor1 FOR SELECT * FROM reason;
```

忽略游标cursor1的前3行：

```
MOVE FORWARD 3 FROM cursor1;
```

抓取游标cursor1的前4行：

```
FETCH 4 FROM cursor1;
```

关闭游标：

```
CLOSE cursor1;
```

结束一个事务：

```
END;
```

相关链接

CLOSE，**FETCH**

13.7.11 CURSOR

功能描述

CURSOR语句定义游标，用于从查询里面检索出指定的几行数据。

为了处理SQL语句，存储过程进程分配一段内存区域来保存上下文联系。游标是指向上下文区域的句柄或指针。借助游标，存储过程可以控制上下文区域的变化。

版权所有 © 华为云计算技术有限公司

注意事项

- 游标命令只能在事务块里使用。
- 通常游标和SELECT一样返回文本格式。因为数据在系统内部是用二进制格式存储的，系统必须对数据做一定转换以生成文本格式。一旦数据是以文本形式返回，客户端应用需要把它们转换成二进制进行操作。使用FETCH语句，游标可以返回文本或二进制格式。
- 应该谨慎使用二进制游标。文本格式一般都比对应的二进制格式占用的存储空间大。二进制游标返回内部二进制形态的数据，可能更易于操作。如果想以文本方式显示数据，则以文本方式检索会为用户节约很多客户端的工作。比如，如果查询从某个整数列返回1，在缺省的游标里将获得一个字符串1，但在二进制游标里将得到一个4字节的包含该数值内部形式的数值（大端顺序）。

语法格式

```
CURSOR cursor_name
   [ BINARY ]  [ NO SCROLL ]  [ { WITH | WITHOUT } HOLD ]
   FOR query ;
```

参数说明

表 13-144 CURSOR 参数说明

参数	描述	取值范围
cursor_name	要创建的游标名。	字符串，需符合**标识符命名规范**。
BINARY	指明游标以二进制而不是文本格式返回数据。	-
NO SCROLL	声明游标检索数据行的方式。 • NO SCROLL：声明该游标不能用于以倒序的方式检索数据行。 • 未声明：根据执行计划的不同，自动判断该游标是否可以用于以倒序的方式检索数据行。	-
WITH HOLD | WITHOUT HOLD	声明当创建游标的事务结束后，游标是否能继续使用。 • WITH HOLD：声明该游标在创建它的事务结束后仍可继续使用。 • WITHOUT HOLD：声明该游标在创建它的事务之外不能再继续使用，此游标将在事务结束时被自动关闭。（默认行为）	-
query	使用SELECT或VALUES子句指定游标返回的行。	SELECT或VALUES子句。

版权所有 © 华为云计算技术有限公司

示例

准备表数据。

```
DROP TABLE IF EXISTS customer_address;
CREATE TABLE customer_address
(ca_address_sk INT,
ca_address_id VARCHAR(16),
ca_street_number VARCHAR(10),
ca_street_name VARCHAR(60),
ca_street_type VARCHAR(15),
ca_suite_number VARCHAR(10));

INSERT INTO customer_address VALUES
(1, 'ID1', '100', 'Main', 'St', 'A1'),
(2, 'ID2', '200', 'Oak', 'Ave', 'B2'),
(3, 'ID3', '300', 'Pine', 'Blvd', 'C3');
```

开启事务。

```
START TRANSACTION;
```

创建名称为cursor1的游标。

```
CURSOR cursor1 FOR SELECT * FROM customer_address ORDER BY 1;
```

创建名称为cursor2的游标。

```
CURSOR cursor2 FOR VALUES(1,2),(0,3) ORDER BY 1;
```

WITH HOLD游标的使用：

1. 创建一个with hold游标。
   ```
   DECLARE cursor3 CURSOR WITH HOLD FOR SELECT * FROM customer_address ORDER BY 1;
   ```
2. 抓取头2行到游标cursor3中。
   ```
   FETCH FORWARD 2 FROM cursor3;
   ```
   ```
   gaussdb=> FETCH FORWARD 2 FROM cursor3;
    ca_address_sk | ca_address_id | ca_street_number | ca_street_name | ca_street_type | ca_suite_number
   ---------------+---------------+------------------+----------------+----------------+-----------------
                1 | ID1           | 100              | Main           | St             | A1
                2 | ID2           | 200              | Oak            | Ave            | B2
   (2 rows)
   ```
3. 结束事务：
   ```
   END;
   ```
4. 抓取下一行到游标cursor3中。
   ```
   FETCH FORWARD 1 FROM cursor3;
   ```
   ```
   gaussdb=> FETCH FORWARD 1 FROM cursor3;
    ca_address_sk | ca_address_id | ca_street_number | ca_street_name | ca_street_type | ca_suite_number
   ---------------+---------------+------------------+----------------+----------------+-----------------
                3 | ID3           | 300              | Pine           | Blvd           | C3
   (1 row)
   ```
5. 关闭游标。
   ```
   CLOSE cursor3;
   ```

相关链接

FETCH

13.7.12 CLOSE

功能描述

CLOSE用于释放与游标关联的所有资源。

版权所有 © 华为云计算技术有限公司

注意事项

- 不允许对一个已关闭的游标再进行任何操作。
- 不再使用的游标应将其关闭。
- 当创建游标的事务用COMMIT或ROLLBACK终止之后，每个非持有的已打开游标都隐式关闭。
- 当创建游标的事务通过ROLLBACK退出之后，则该游标将隐式关闭。
- 当创建游标的事务成功提交，则持有的游标将保持打开状态，直到执行了显式的CLOSE或者客户端断开连接为止。
- DWS没有显式地打开游标的OPEN语句，因为一个游标在使用CURSOR命令定义的时候就打开了。可以通过查询系统视图pg_cursors查看所有可用的游标。

语法格式

```
CLOSE { cursor_name | ALL };
```

参数说明

表 13-145 CLOSE 参数说明

参数	描述	取值范围
cursor_name	要关闭的游标的名称。	已存在的游标名称。
ALL	关闭所有已打开的游标。	-

示例

关闭游标。

```
DROP TABLE IF EXISTS customer;
CREATE TABLE customer (a int, b int);
START TRANSACTION;
CURSOR cursor1 FOR SELECT * FROM customer;
CLOSE cursor1;
```

相关链接

FETCH，**MOVE**

13.8 数据脱敏和安全相关

13.8.1 CREATE ROW LEVEL SECURITY POLICY

功能描述

对表创建行级访问控制策略。

版权所有 © 华为云计算技术有限公司

行级访问控制策略（ROW LEVEL SECURITY POLICY）表示将数据库访问控制权限精确到数据表行级别，使数据库达到行级访问控制的能力。不同用户执行相同的SQL查询操作，系统将自动根据创建的控制策略返回不同的读取结果。

对表创建行级访问控制策略时，需**打开该表的行级访问控制开关**（ALTER TABLE ... ENABLE ROW LEVEL SECURITY | ALTER FOREIGN TABLE ... ENABLE ROW LEVEL SECURITY）策略才能生效，否则不生效。

策略应用

- 行级访问控制策略名称是针对表的，同一个数据表上不能有同名的行级访问控制策略；对不同的数据表，可以有同名的行级访问控制策略。
- 行级访问控制策略可以应用到指定的用户（角色），也可应用到全部用户（PUBLIC）；定义行级访问控制策略时，若未指定受影响的用户，默认为PUBLIC。
- 行级访问控制策略可以应用到指定的操作（SELECT、UPDATE、DELETE、ALL），ALL表示会影响SELECT、UPDATE、DELETE三种操作；定义行级访问控制策略时，若未指定受影响的相关操作，默认为ALL。
- 当前行级访问控制会影响数据表的读取操作（SELECT、UPDATE、DELETE），暂不影响数据表的写入操作（INSERT、MERGE INTO）。表所有者或系统管理员可以在USING子句中创建表达式，并在客户端执行数据表读取操作时，数据库后台在查询重写阶段会将满足条件的表达式拼接并应用到执行计划中。针对数据表的每一条元组，当USING表达式返回TRUE时，元组对当前用户可见，当USING表达式返回FALSE或NULL时，元组对当前用户不可见。

注意事项

- 支持对行存表、行存分区表、列存表、列存分区表、复制表、unlogged表、hash表、非EXTERNAL SCHEMA的外表定义行级访问控制策略。
- 同一张表上可以创建多个行级访问控制策略，一张表最多创建100个行级访问控制策略。
- 具有管理员权限的用户、初始运维用户(Ruby)、表的owner和表的owner角色组成员不受行级访问控制影响，可以查看表的全量数据。
- 通过SQL语句、视图、函数、存储过程查询包含行级访问控制策略的表，都会受影响。
- 不支持HDFS表、EXTERNAL SCHEMA的外表、临时表定义行级访问控制策略。
- 不支持对视图定义行级访问控制策略。
- 不支持对行级访问控制策略依赖的列进行类型修改。例如，不支持如下修改：
  ```
  ALTER TABLE public.all_data ALTER COLUMN role TYPE text;
  ```

语法格式

```
CREATE [ ROW LEVEL SECURITY ] POLICY policy_name ON table_name
    [ AS { PERMISSIVE | RESTRICTIVE } ]
    [ FOR { ALL | SELECT | UPDATE | DELETE } ]
    [ TO { role_name | PUBLIC } [, ...] ]
    USING ( using_expression )
```

 版权所有 © 华为云计算技术有限公司

参数说明

表 13-146 CREATE ROW LEVEL SECURITY POLICY 参数说明

参数	描述	取值范围
policy_name	要创建的行级访问控制策略名称。	同一个数据表上不能有同名的行级访问控制策略。
table_name	要应用行级访问控制策略的表名。	有效的表名称。
PERMISSIVE	指定行级访问控制策略的类型为宽容性策略。 对于一个给定的查询，将使用“OR”操作符将所有的宽容性策略组合。行级访问控制策略的类型默认为宽容性策略。	-
RESTRICTIVE	指定行级访问控制策略的类型为限制性策略。 对于一个给定的查询，将使用“AND”操作符将所有的限制性策略组合。 至少需要一条宽容性策略允许对记录的访问。如果只有限制性策略存在，则不能访问任何记录。当宽容性和限制性策略共存时，只有当记录能通过至少一条宽容性策略以及所有的限制性策略时，该记录才能访问。	-

版权所有 © 华为云计算技术有限公司

参数	描述	取值范围
command	当前行级访问控制影响的SQL操作，可指定操作包括：ALL、SELECT、UPDATE、DELETE。 未指定时，ALL为默认值，涵盖SELECT、UPDATE、DELETE操作。	• 当command为SELECT时，SELECT类操作受行级访问控制的影响，只能查看到满足条件（using_expression返回值为TRUE）的元组数据，受影响的操作包括SELECT，UPDATE ... RETURNING，DELETE ... RETURNING。 • 当command为UPDATE时，UPDATE类操作受行级访问控制的影响，只能更新满足条件（using_expression返回值为TRUE）的元组数据，受影响的操作包括UPDATE，UPDATE ... RETURNING，SELECT ... FOR UPDATE/SHARE。 • 当command为DELETE时，DELETE类操作受行级访问控制的影响，只能删除满足条件（using_expression返回值为TRUE）的元组数据，受影响的操作包括DELETE，DELETE ... RETURNING。 行级访问控制策略与适配的SQL语法关系参见**表13-147**。
role_name	行级访问控制影响的数据库用户。系统管理员不受行级访问控制特性影响。 未指定时，PUBLIC为默认值，PUBLIC表示影响所有数据库用户，可以指定多个受影响的数据库用户。	有效的角色名称。
using_expression	行级访问控制的表达式（返回boolean值）。	条件表达式中不能包含AGG函数和窗口（WINDOW）函数。在查询重写阶段，如果数据表的行级访问控制开关打开，满足条件的表达式会添加到计划树中。针对数据表的每条元组，会进行表达式计算，只有表达式返回值为TRUE时，行数据对用户才可见（SELECT、UPDATE、DELETE）；当表达式返回FALSE时，该元组对当前用户不可见，用户无法通过SELECT语句查看此元组，无法通过UPDATE语句更新此元组，无法通过DELETE语句删除此元组。

版权所有 © 华为云计算技术有限公司

表 13-147 ROW LEVEL SECURITY 策略与适配 SQL 语法关系

Command	SELECT/ALL policy	UPDATE/ALL policy	DELETE/ALL policy
SELECT	Existing row	No	No
SELECT FOR UPDATE/SHARE	Existing row	Existing row	No
UPDATE	No	Existing row	No
UPDATE RETURNING	Existing row	Existing row	No
DELETE	No	No	Existing row
DELETE RETURNING	Existing row	No	Existing row

示例 1：创建行级访问控制策略，当前用户只能查看用户自身的数据

1. 创建用户alice和bob。
```
DROP ROLE IF EXISTS alice;
DROP ROLE IF EXISTS bob;
CREATE ROLE alice PASSWORD '{password}';
CREATE ROLE bob PASSWORD '{password}';
```
2. 创建数据表public.all_data。
```
DROP TABLE IF EXISTS public.all_data;
CREATE TABLE public.all_data(id int, role varchar(100), data varchar(100));
```
3. 向数据表插入数据。
```
INSERT INTO all_data VALUES(1, 'alice', 'alice data');
INSERT INTO all_data VALUES(2, 'bob', 'bob data');
INSERT INTO all_data VALUES(3, 'peter', 'peter data');
```
4. 将表all_data的读取权限赋予alice和bob用户。
```
GRANT SELECT ON all_data TO alice, bob;
```
5. 打开行级访问控制策略开关。
```
ALTER TABLE all_data ENABLE ROW LEVEL SECURITY;
```
6. 创建行级访问控制策略，当前用户只能查看用户自身的数据。
```
CREATE ROW LEVEL SECURITY POLICY all_data_rls ON all_data USING(role = CURRENT_USER);
```
7. 查看表all_data相关信息。
```
SELECT * FROM PG_GET_TABLEDEF('all_data');
```
```
gaussdb=> SELECT * FROM PG_GET_TABLEDEF('all_data');
                        pg_get_tabledef
------------------------------------------------------------------------
 SET search_path = public;                                             +
 CREATE  TABLE all_data (                                              +
         id integer,                                                   +
         role character varying(100),                                  +
         data character varying(100)                                   +
 )                                                                     +
 WITH (orientation=row, compression=no, enable_rowsecurity=true)+
 DISTRIBUTE BY ROUNDROBIN                                              +
 TO GROUP group_version1;
(1 row)
```

 版权所有 © 华为云计算技术有限公司

8. 当前用户执行SELECT操作。

```
SELECT * FROM all_data;
```

```
gaussdb=> SELECT * FROM all_data;
 id | role  |    data
----+-------+------------
  1 | alice | alice data
  2 | bob   | bob data
  3 | peter | peter data
(3 rows)
```

```
EXPLAIN(COSTS OFF) SELECT * FROM all_data;
```

```
gaussdb=> EXPLAIN(COSTS OFF) SELECT * FROM all_data;
             QUERY PLAN
-----------------------------------
  id |          operation
 ----+-------------------------------
   1 | ->  Streaming (type: GATHER)
   2 |    ->  Seq Scan on all_data

   ====== Query Summary =====
 -------------------------------
 System available mem: 5160960KB
 Query Max mem: 5160960KB
 Query estimated mem: 1024KB
(10 rows)
```

9. 切换至alice用户。

```
SET role alice password '{password}';
```

10. 执行SELECT操作。

```
SELECT * FROM all_data;
```

```
gaussdb=> SELECT * FROM all_data;
 id | role  |    data
----+-------+------------
  1 | alice | alice data
(1 row)
```

```
EXPLAIN(COSTS OFF) SELECT * FROM all_data;
```

 版权所有 © 华为云计算技术有限公司

```
gaussdb=> EXPLAIN(COSTS OFF) SELECT * FROM all_data;
                          QUERY PLAN
------------------------------------------------------------------
 id |          operation
----+-------------------------------
  1 | ->  Streaming (type: GATHER)
  2 |     ->  Seq Scan on all_data

        Predicate Information (identified by plan id)
------------------------------------------------------------------
  2 --Seq Scan on all_data
        Filter: ((role)::name = 'alice'::name)
Notice: This query is influenced by row level security feature

  ====== Query Summary =====
-------------------------------
System available mem: 5160960KB
Query Max mem: 5160960KB
Query estimated mem: 1024KB
(16 rows)
```

示例 2：通过行级控制实现分区权限管理

1. 创建用户alice。

   ```
   SET role dbadmin password '{password}';
   CREATE ROLE alice PASSWORD '{password1}';
   ```

2. 创建范围分区表web_returns_p1，并插入数据。

   ```
   DROP TABLE IF EXISTS web_returns_p1;
   CREATE TABLE web_returns_p1
   (
      wr_returned_date_sk       integer,
      wr_returned_time_sk       integer,
      wr_item_sk                integer NOT NULL,
      wr_refunded_customer_sk   integer
   )
   WITH (orientation = column)
   DISTRIBUTE BY HASH (wr_item_sk)
   PARTITION BY RANGE(wr_returned_date_sk)
   (
      PARTITION p2016 START(800) END(830) EVERY(1)
   );

   INSERT INTO web_returns_p1 values (801,17,11,102);
   INSERT INTO web_returns_p1 values (802,18,12,103);
   ```

3. 将表web_returns_p1的读取权限赋予alice用户。

   ```
   GRANT SELECT ON web_returns_p1 TO alice;
   ```

4. 打开行级访问控制策略开关。

   ```
   ALTER TABLE web_returns_p1 ENABLE ROW LEVEL SECURITY;
   ```

5. 创建行级访问控制策略web_returns_rsl。其中wr_returned_date_sk为web_returns_p1分区表的分区名，801为分区值。

   ```
   CREATE ROW LEVEL SECURITY POLICY web_returns_rsl ON web_returns_p1
   USING('wr_returned_date_sk' = '801');
   ```

6. 将行级访问控制策略web_returns_rsl的赋予用户alice。

   ```
   ALTER ROW LEVEL SECURITY POLICY web_returns_rsl ON web_returns_p1 TO alice;
   ```

7. 切换至alice用户。

   ```
   SET ROLE alice password '{password1}';
   ```

8. 查询表web_returns_p1。

   ```
   SELECT * FROM web_returns_p1;
   ```

 版权所有 © 华为云计算技术有限公司

相关链接

ALTER ROW LEVEL SECURITY POLICY，DROP ROW LEVEL SECURITY POLICY

常见问题

Q：如何查看表上当前有哪些行级访问控制策略?

A：可通过系统表PG_RLSPOLICY查询，语句如下，tablename替换成实际表名。

```
SELECT p.polname,
      c.relname AS table_name,
      CASE p.polcmd WHEN '*' THEN 'ALL' WHEN 'r' THEN 'SELECT' WHEN 'w' THEN 'UPDATE' WHEN 'd'
THEN 'DELETE' END AS cmd,
      CASE p.polpermissive WHEN true THEN 'PERMISSIVE' WHEN false THEN 'RESTRICTIVE' END AS
policy_type,
      pg_get_userbyid(p.polroles[1]) AS affected_role,
      p.polqual
FROM pg_rlspolicy p
JOIN pg_class c ON p.polrelid = c.oid
WHERE c.relname = 'tablename';
```

13.8.2 ALTER ROW LEVEL SECURITY POLICY

功能描述

修改已有行级访问控制策略的定义，包括行级访问控制策略的名称、策略应用的用户及策略表达式等。

注意事项

表的所有者或管理员用户才能执行ALTER ROW LEVEL SECURITY POLICY。

语法格式

```
ALTER [ ROW LEVEL SECURITY ] POLICY [ IF EXISTS ] policy_name ON table_name RENAME TO
new_policy_name

ALTER [ ROW LEVEL SECURITY ] POLICY policy_name ON table_name
    [ TO { role_name | PUBLIC } [, ...] ]
    [ USING ( using_expression ) ]
```

参数说明

表 13-148 ALTER ROW LEVEL SECURITY POLICY 参数说明

参数	描述	取值范围
policy_name	要修改的行级访问控制策略名称。	已有的行级访问控制策略。
table_name	应用行级访问控制策略的表名	已存在的表名。
new_policy_name	新的行级访问控制策略名称。	-

版权所有 © 华为云计算技术有限公司

参数	描述	取值范围
role_name	行级访问控制策略应用的数据库用户，可以指定多个用户，PUBLIC表示应用到所有用户。	已存在的角色名。
using_expression	行级访问控制的表达式，返回值为boolean类型。	-

示例

创建示例表all_data和用户alice、bob。

```
DROP TABLE IF EXISTS all_data;
CREATE TABLE all_data (a int, b int);
DROP USER IF EXISTS alice CASCADE;
DROP USER IF EXISTS bob CASCADE;
CREATE USER alice PASSWORD '{password}';
CREATE USER bob PASSWORD '{password}';
```

开启表all_data的行级访问控制策略功能。

```
ALTER TABLE all_data ENABLE ROW LEVEL SECURITY;
```

在表all_data上创建行级访问控制策略all_data_rls。

```
CREATE ROW LEVEL SECURITY POLICY all_data_rls ON all_data USING(b = CURRENT_USER);
```

修改行级访问控制策略的名称all_data_rls为新的名称all_data_new_rls。

```
ALTER ROW LEVEL SECURITY POLICY all_data_rls ON all_data RENAME TO all_data_new_rls;
```

修改行级访问控制策略all_data_new_rls应用的用户为alice和bob。

```
ALTER ROW LEVEL SECURITY POLICY all_data_new_rls ON all_data TO alice, bob;
```

修改行级访问控制策略all_data_new_rls的过滤条件（alice和bob只能访问a列值大于100且b列值等于当前登录用户名的数据行）。

```
ALTER ROW LEVEL SECURITY POLICY all_data_new_rls ON all_data USING (a > 100 AND b = current_user);
```

相关链接

CREATE ROW LEVEL SECURITY POLICY，**DROP ROW LEVEL SECURITY POLICY**

13.8.3 DROP ROW LEVEL SECURITY POLICY

功能描述

在数据库安全管理中，行级访问控制（Row Level Security）策略用于限制不同用户对表中行数据的访问范围。随着业务变化，某些策略可能不再适用——例如业务部门调整后旧的数据隔离规则需要移除，或策略规则需要重建以适配新的权限模型。当行级访问控制策略不再需要或需要重建时，可使用DROP ROW LEVEL SECURITY POLICY命令删除现有策略。

需要特别注意的是： 删除策略后，表的行级访问控制将失效，所有有权限访问该表的用户均可访问全部行数据。因此，在删除策略前，请确认该操作不会导致敏感数据泄露。

版权所有 © 华为云计算技术有限公司

注意事项

- 仅表的所有者或者管理员用户才能执行DROP ROW LEVEL SECURITY POLICY。
- **安全风险提示**：删除行级访问控制策略后，该策略所限制的数据访问范围将解除，原本被策略过滤的行数据将变为可访问。
- 如果需要修改现有策略而非删除重建，可使用**ALTER ROW LEVEL SECURITY POLICY**命令，避免删除策略导致的中间状态数据暴露风险。

语法格式

```
DROP [ ROW LEVEL SECURITY ] POLICY [ IF EXISTS ] policy_name ON table_name [ CASCADE | RESTRICT ]
```

参数说明

表 13-149 DROP ROW LEVEL SECURITY POLICY 参数说明

参数	描述	取值范围
IF EXISTS	如果指定的行级访问控制策略不存在，发出一个notice而不是抛出一个错误。	-
policy_name	要删除的行级访问控制策略的名称。	已存在的行级访问控制策略的名称。
table_name	要删除的行级访问控制策略所在的数据表名称。	已存在的行级访问控制策略所在的数据表名。
CASCADE/RESTRICT	仅适配此语法，无对象依赖于该行级访问控制策略，CASCADE和RESTRICT效果相同。	-

示例

步骤1 使用系统表**PG_RLSPOLICY**查询当前存在的行级访问控制策略。

```
SELECT polname, polrelid::regclass AS table_name FROM PG_RLSPOLICY;
```

```
gaussdb=> SELECT polname, polrelid::regclass AS table_name FROM PG_RLSPOLICY;
     polname     |    table_name
-----------------+------------------
 pg_task_control | scheduler.pg_task
 all_data_rls    | all_data
(2 rows)
```

步骤2 删除表all_data上的行级访问控制策略all_data_rls。

```
DROP ROW LEVEL SECURITY POLICY all_data_rls ON all_data;
```

步骤3 验证策略是否已成功删除。通过查询PG_RLSPOLICY系统表，确认目标策略不再存在。

```
SELECT polname, polrelid::regclass AS table_name FROM PG_RLSPOLICY WHERE polname = 'all_data_rls';
```

 版权所有 © 华为云计算技术有限公司

```
gaussdb=> SELECT polname, polrelid::regclass AS table_name FROM PG_RLSPOLICY WHERE polname = 'all_data_rls';
 polname | table_name
---------+------------
(0 rows)
```

----结束

相关链接

ALTER ROW LEVEL SECURITY POLICY，**CREATE ROW LEVEL SECURITY POLICY**

常见问题

Q：删除策略后，表的行级访问控制是否自动关闭?

A：不会自动关闭。删除策略仅移除指定的策略规则，表上行级访问控制的启用状态不变。如果表上还有其他策略，其他策略仍会生效。如果被删除的是表上唯一的策略，则虽然行级访问控制仍处于启用状态，但由于没有策略匹配，实际上不会对查询产生过滤效果，所有有权限的用户均可访问全部行数据。如需完全关闭行级访问控制，请使用ALTER TABLE … DISABLE ROW LEVEL SECURITY命令。

```
ALTER TABLE all_data DISABLE  ROW LEVEL SECURITY;
```

Q：如何查看表上当前有哪些行级访问控制策略?

A：可通过系统表**PG_RLSPOLICY**查询，语句如下，tablename替换成实际表名。

```
SELECT p.polname,
      c.relname AS table_name,
      CASE p.polcmd WHEN '*' THEN 'ALL' WHEN 'r' THEN 'SELECT' WHEN 'w' THEN 'UPDATE' WHEN 'd'
THEN 'DELETE' END AS cmd,
      CASE p.polpermissive WHEN true THEN 'PERMISSIVE' WHEN false THEN 'RESTRICTIVE' END AS
policy_type,
      pg_get_userbyid(p.polroles[1]) AS affected_role,
      p.polqual
FROM pg_rlspolicy p
JOIN pg_class c ON p.polrelid = c.oid
WHERE c.relname = 'tablename';
```

13.8.4 CREATE REDACTION POLICY

功能描述

在数据库表上创建数据脱敏策略。

DWS提供列级别的数据脱敏策略（REDACTION POLICY）功能。针对某些敏感信息（如身份证号、手机号、银行卡号等），通过应用脱敏函数进行原始数据的变形改写，实现敏感隐私数据的可靠保护，增强产品在数据安全和隐私保护方面的能力。

注意事项

- 数据脱敏特性使用GUC参数feature_support_options进行控制，当出现“REDACTION POLICY is not yet supported, please add enable_data_redaction into feature_support_options.“或者“Cannot use the feature of data redaction, please drop existed redaction policy or add enable_data_redaction into feature_support_options.“的报错信息时，说明没有开启数据脱敏特性，需联系技术支持人员设置feature_support_options参数，开启数据脱敏特性。

 版权所有 © 华为云计算技术有限公司

- 只有表对象的属主和被授予gs_role_redaction预置角色的用户具有创建脱敏策略的权限。
- 具有sysadmin权限的用户，可跳过脱敏策略检查，对脱敏列数据一直具有可见性，即脱敏策略不生效。
- 仅支持在普通表和非EXTERNAL SCHEMA的外表（OBS、HDFS、GDS、协调分析外表）创建数据脱敏策略，不支持为系统表、EXTERNAL SCHEMA的外表、临时表、UNLOGGED表以及视图和函数对象创建脱敏策略。
- 表对象与脱敏策略间一一对应。一个脱敏策略是表对象所有脱敏列的集合，可以给脱敏表对象的多个列字段指定脱敏函数，且不同脱敏列对象可以采用不同的脱敏函数。
- 创建脱敏策略时，默认策略生效，即属性enable值为true。
- 支持通过指定角色匹配脱敏策略。
- 不支持通过同义词向普通表对象创建脱敏策略。

语法格式

```
CREATE REDACTION POLICY policy_name ON table_name [ { AFTER | BEFORE } old_policy_name ]
   [INHERIT]
   [ WHEN (when_expression) ]
   [ ADD COLUMN column_name WITH redaction_function_name ( [ argument [, ...] ] )] [, ... ];
```

参数说明

表 13-150 CREATE REDACTION POLICY 参数说明

参数	描述	取值范围
policy_name	脱敏策略名称。	-
table_name	应用脱敏策略的表名。	有效的表名称。
BEFORE | AFTER	指定当前策略创建的相对位置，通常一张表使用一个脱敏策略即可。缺省时表示：默认当前策略创建在当前系统表中记录的目标表的候选策略的最后一个之后。	-

 版权所有 © 华为云计算技术有限公司

参数	描述	取值范围
WHEN (when_expression)	WHEN子句指定一个生效表达式。仅当此表达式为真时，脱敏策略才可能生效。	查询语句涉及脱敏表对象时，仅当脱敏策略的WHEN子句表达式为真时，查询对脱敏列数据才可能不可见，即脱敏策略生效。通常，采用WHEN子句来限定脱敏策略的生效用户范围，具有较严格的约束规格。 WHEN子句的规格约束如下： 1. 表达式可以是AND、OR连接的多个子表达式。 2. 每个子表达式仅支持=、<>、!=、>=、>、<=、<七种运算符，左右值只能取常量值或者下列系统常量值之一：SESSION_USER、CURRENT_USER、USER、CURRENT_ROLE、CURRENT_SCHEMA系统常量或者SYS_CONTEXT系统函数。 3. 每个子表达式可以是IN、NOT IN表达式，左值可以是上述2中所列系统常量值，右值数组的每个元素必须是常量值。 4. 每个子表达式可以是**pg_has_role(user, role, privilege)**系统函数。 5. 当脱敏策略永远成立时，即对所有用户（含表对象属主）均生效，建议使用表达式（1=1）创建脱敏策略。 6. WHEN子句缺省时，脱敏策略默认不生效，需用户手动指定WHEN子句表达式。
column_name	应用脱敏策略的表的列名。	-
redaction_function_name	脱敏函数名。	-

版权所有 © 华为云计算技术有限公司

参数	描述	取值范围
arguments	脱敏函数的参数列表。 DWS支持MASK_NONE、MASK_FULL、MASK_PARTIAL三种内置脱敏函数。同时支持使用C语言、PL/pgSQL语言创建的用户自定义脱敏函数，具体可参考**数据脱敏函数**。	• MASK_NONE，表示不进行任何脱敏处理。 • MASK_FULL，表示全脱敏成固定值（如数字转为0）。 • MASK_PARTIAL，表示按指定的字符类型，数值类型或时间类型进行部分脱敏处理（如手机号部分数字遮盖为*）。

示例 1：对指定用户创建脱敏策略

1. 打开数据脱敏功能。

   联系技术支持人员设置feature_support_options参数，开启数据脱敏特性。

   ```
   gs_guc set -Z coordinator -Z datanode -N all -I all -c
   "feature_support_options=enable_data_redaction"
   gs_om -t stop && gs_om -t start
   ```

2. 创建用户alice和matu。

   ```
   CREATE ROLE alice PASSWORD '{password}';
   CREATE ROLE matu PASSWORD '{password}';
   ```

3. 用户alice创建表对象emp并插入数据。

   ```
   CREATE TABLE emp(id int, name varchar(20), salary NUMERIC(10,2));
   INSERT INTO emp VALUES(1, 'July', 1230.10), (2, 'David', 999.99);
   ```

4. 查看表数据。

   ```
   SELECT * FROM emp;
   ```

   ```
   id | name  | salary
   ---+-------+--------
    2 | David |  999.99
    1 | July  | 1230.10
   (2 rows)
   ```

5. 用户alice为表对象emp创建脱敏策略mask_emp，字段salary对用户matu不可见。

   ```
   CREATE REDACTION POLICY mask_emp ON emp WHEN(current_user = 'matu') ADD COLUMN salary
   WITH mask_full(salary);
   ```

6. 用户alice授予用户matu表emp的SELECT权限。

   ```
   GRANT SELECT ON emp TO matu;
   ```

7. 切至用户matu。

   ```
   SET ROLE matu PASSWORD '{password}';
   ```

8. matu查询表emp，字段salary数据已脱敏。

   ```
   SELECT * FROM emp;
   ```

   ```
   id | name  | salary
   ---+-------+--------
    2 | David |      0
    1 | July  |      0
   (2 rows)
   ```


版权所有 © 华为云计算技术有限公司

示例 2：对指定角色创建脱敏策略

1. 创建角色redact_role。
```
CREATE ROLE redact_role PASSWORD '{password}';
```
2. 将用户matu，alice加入角色redact_role。
```
GRANT redact_role to matu,alice;
```
3. 管理员用户创建表对象emp1并插入数据。
```
CREATE TABLE emp1(id int, name varchar(20), salary NUMERIC(10,2));
INSERT INTO emp1 VALUES(3, 'Rose', 2230.20), (4, 'Jack', 899.88);
```
4. 查看表emp1数据。
```
SELECT * FROM emp1;
```
```
 id | name | salary
----+------+---------
  4 | Jack |  899.88
  3 | Rose | 2230.20
(2 rows)
```
5. 管理员用户为表对象emp1创建脱敏策略mask_emp1，使字段salary对角色redact_role不可见。
```
CREATE REDACTION POLICY mask_emp1 ON emp1 WHEN(pg_has_role(current_user, 'redact_role',
'member')) ADD COLUMN salary WITH mask_full(salary);
```
若不指定用户，默认为当前用户current_user。
```
CREATE REDACTION POLICY mask_emp1 ON emp1 WHEN (pg_has_role('redact_role', 'member'))
ADD COLUMN salary WITH mask_full(salary);
```
6. 管理员用户授予用户matu表emp1的SELECT权限。
```
GRANT SELECT ON emp1 TO matu;
```
7. 切换至用户matu。
```
SET ROLE matu PASSWORD '{password}';
```
8. matu查询表emp，字段salary数据已脱敏。
```
SELECT * FROM emp1;
```
```
 id | name | salary
----+------+--------
  3 | Rose |      0
  4 | Jack |      0
(2 rows)
```

相关链接

ALTER REDACTION POLICY，**DROP REDACTION POLICY**

使用实践

DWS数据脱敏

13.8.5 ALTER REDACTION POLICY

功能描述

修改数据库表上已有的数据脱敏策略，包括脱敏策略生效/失效、策略名称及脱敏列的增删改等内容。

版权所有 © 华为云计算技术有限公司

注意事项

只有待修改脱敏策略所应用表对象的属主和被授予gs_role_redaction预置角色的用户才有修改脱敏策略的权限。

语法格式

- 修改脱敏策略的生效表达式。
```
ALTER REDACTION POLICY policy_name ON table_name WHEN (when_expression);
```
- 使脱敏策略生效或失效。
```
ALTER REDACTION POLICY policy_name ON table_name ENABLE | DISABLE;
```
- 重命名脱敏策略。
```
ALTER REDACTION POLICY policy_name ON table_name RENAME TO new_policy_name;
```
- 修改脱敏列，包括新增、修改、删除脱敏列。
```
ALTER REDACTION POLICY policy_name ON table_name
  action;
```
其中，脱敏列操作action可以是以下子句之一：
```
[INHERIT] ADD COLUMN column_name WITH redaction_function_name ( [ argument [, ...] ] )
 | [INHERIT] MODIFY COLUMN column_name WITH redaction_function_name ( [ argument [, ...] ] )
 | DROP COLUMN column_name
```

参数说明

表 13-151 ALTER REDACTION POLICY 参数说明

参数	描述	取值范围
policy_name	要修改的脱敏策略名。	已存在的脱敏策略名。
table_name	要修改的脱敏策略应用的表名称。	已存在的表名。
when_expression	脱敏策略新的生效表达式。	-
ENABLE | DISABLE	当前脱敏策略是否生效。 • ENABLE，使之前失效的表对象的脱敏策略重新生效。 • DISABLE，使当前应用在表上的脱敏策略失效。	-
new_policy_name	新的脱敏策略名。	字符串，需符合**标识符命名规范**。
column_name	脱敏表的列字段名。 • 如果新增脱敏列，则指定列字段尚未绑定任何脱敏函数。 • 如果修改脱敏列，则指定列字段为已存在的脱敏列。 • 如果删除脱敏列，则指定列字段为已存在的脱敏列。	-
redaction_function_name	脱敏函数名。	支持的函数请参考**数据脱敏函数**。

版权所有 © 华为云计算技术有限公司

参数	描述	取值范围
arguments	脱敏函数的参数列表。 • MASK_NONE，表示不进行任何脱敏处理。 • MASK_FULL，表示全脱敏成固定值。 • MASK_PARTIAL，表示按指定的字符类型，数值类型或时间类型进行部分脱敏处理。	支持的函数请参考**数据脱敏函数**。

示例

打开数据脱敏功能。

联系技术支持人员设置feature_support_options参数，开启数据脱敏特性。

```
gs_guc set -Z coordinator -Z datanode -N all -I all -c "feature_support_options=enable_data_redaction"
gs_om -t stop && gs_om -t start
```

创建用户test_role和示例表emp并插入数据。

```
CREATE ROLE test_role PASSWORD '{Password}';

DROP TABLE IF EXISTS emp;
CREATE TABLE emp(id int, name varchar(20), salary NUMERIC(10,2));
INSERT INTO emp VALUES(1, 'July', 1230.10), (2, 'David', 999.99);
```

为表对象emp创建脱敏策略mask_emp，字段salary对用户test_role不可见。

```
CREATE REDACTION POLICY mask_emp ON emp WHEN(current_user = 'test_role') ADD COLUMN salary
WITH mask_full(salary);
```

修改脱敏策略生效表达式，使其对指定角色生效（若不指定用户，默认对当前用户生效）。

```
ALTER REDACTION POLICY mask_emp ON emp WHEN (pg_has_role(current_user, 'redact_role', 'member'));
ALTER REDACTION POLICY mask_emp ON emp WHEN (pg_has_role('redact_role', 'member'));
```

修改脱敏策略生效表达式，使其对所有用户均生效。

```
ALTER REDACTION POLICY mask_emp ON emp WHEN (1=1);
```

修改脱敏策略，使其失效。

```
ALTER REDACTION POLICY mask_emp ON emp DISABLE;
```

重新使脱敏策略生效。

```
ALTER REDACTION POLICY mask_emp ON emp ENABLE;
```

重命名脱敏策略为mask_emp_new。

```
ALTER REDACTION POLICY mask_emp ON emp RENAME TO mask_emp_new;
```

新增脱敏列。

```
ALTER REDACTION POLICY mask_emp_new ON emp ADD COLUMN name WITH mask_partial(name, '*', 1,
length(name));
```

使用脱敏函数MASK_FULL对name字段数据全脱敏。

```
ALTER REDACTION POLICY mask_emp_new ON emp MODIFY COLUMN name WITH mask_full(name);
```


版权所有 © 华为云计算技术有限公司

删除已存在的脱敏列。

```
ALTER REDACTION POLICY mask_emp_new ON emp DROP COLUMN name;
```

相关链接

CREATE REDACTION POLICY，**DROP REDACTION POLICY**

查看脱敏策略和脱敏列信息

DWS提供以下系统表和视图用于查看脱敏策略和脱敏列信息。

- **PG_REDACTION_POLICY**系统表，用于查看脱敏策略信息。
- **PG_REDACTION_COLUMN**系统表，用于查看脱敏列信息。
- **REDACTION_POLICIES**系统视图，用于查看当前数据库内所有脱敏对象信息
- **REDACTION_COLUMNS**系统视图，用于查看当前数据库内所有脱敏列信息。

13.8.6 DROP REDACTION POLICY

功能描述

删除应用在指定表上的脱敏策略。

注意事项

只有表的属主和被授予gs_role_redaction预置角色的用户才有权限删除脱敏策略。

语法格式

```
DROP REDACTION POLICY [ IF EXISTS ] policy_name ON table_name;
```

参数说明

表 13-152 DROP REDACTION POLICY 参数说明

参数	描述	取值范围
IF EXISTS	如果待删除的脱敏策略不存在，则发出一个notice而不是抛出一个错误。	-
policy_name	要删除的脱敏策略名称。	已存在的脱敏策略名称。
table_name	要删除的脱敏策略所应用的表名。	已存在的脱敏策略所应用的表名。

示例

删除表emp上的脱敏策略mask_emp：

```
DROP REDACTION POLICY mask_emp ON emp;
```


版权所有 © 华为云计算技术有限公司

相关链接

ALTER REDACTION POLICY，CREATE REDACTION POLICY

13.8.7 ALTER CLUSTER

功能描述

细粒度透明加密密钥轮转，创建细粒度透明加密集群时自动调用ALTER CLUSTER ENCRYPTION KEY ROTATION创建密钥。

DEK和CEK保存在系统表PG_TDE中，每次轮转时会对系统表每一行进行UPDATE。

密钥缓存和轮转

执行ALTER CLUSTER ENCRYPTION KEY ROTATION命令进行密钥轮转后，系统会生成新的数据加密密钥（DEK）。每次轮转产生的密钥都会在首次访问时从KMS解密并缓存到本地缓存中，以减少后续访问KMS的网络开销。

TDE缓存容量通过GUC参数tde_cache_config（该参数需联系技术支持设置）的cachesize字段设置，默认值为100000。当缓存中的密钥数量达到容量上限时，系统会按照最近最少使用策略淘汰旧密钥，为新密钥腾出空间。

若缓存容量设置过小，无法容纳所有历史轮转密钥，则被淘汰的密钥在下次访问时将产生缓存未命中，系统需重新从KMS获取并解密该密钥后再次写入缓存，这会增加I/O访问延迟。

缓存超时时间通过GUC参数tde_cache_config（该参数需联系技术支持设置）的cachetime字段设置，默认值-1表示不启用超时淘汰。当设置为正数时，系统会定期清理超过该时间未被访问的缓存密钥，同样会导致后续访问时需要重新连接KMS。

建议合理规划密钥轮转的频率，根据业务实际需要缓存的密钥数量设置cachesize，确保缓存能够容纳所有活跃密钥，避免因缓存淘汰产生额外的KMS访问开销。同时，根据密钥访问的时间局部性特征合理设置cachetime，在内存占用与访问性能之间取得平衡。

注意事项

只有拥有数据库所有者权限的用户才能执行ALTER CLUSTER。

语法格式

修改密钥。

```
ALTER CLUSTER ENCRYPTION KEY ROTATION;
```

13.9 资源池和过滤规则

13.9.1 CREATE RESOURCE POOL

功能描述

创建一个资源池，并指定此资源池相关联的控制组。数据库的系统资源（CPU资源、内存资源、IO资源和存储资源）是有限的，DWS在同时运行多种类型的业务（如数据

版权所有 © 华为云计算技术有限公司

加载、批量分析、实时查询等）时，各类型业务之间可能会竞争资源，从而出现资源性能瓶颈，导致吞吐量下降，造成整体性能低下。那么对系统的资源进行合理的分配，能避免产生资源不合理利用而导致的系统运行效率下降等问题，因此DWS提供了**资源管理功能**，用户可以根据自身业务将资源按需划分成不同的资源池，不同资源池之间资源互相隔离。

本语法为基础语法仅供参考，当前**推荐通过管理控制台进行资源池创建**，了解资源池更多内容，请参见**DWS资源负载管理**。

注意事项

只要用户对当前数据库有CREATE权限，就可以创建资源池。

语法格式

```
CREATE RESOURCE POOL pool_name
    [WITH ({MEM_PERCENT=pct | CONTROL_GROUP="group_name" | ACTIVE_STATEMENTS=stmt |
MAX_DOP = dop | MEMORY_LIMIT='memory_size' | io_limits=io_limits | io_priority='priority' |
nodegroup='nodegroup_name' | is_foreign = boolean | except_rule='except_rule' | weight=bandwidth_weight
| enable_concurrency_scaling=boolean}[, ... ])];
```

参数说明

表 13-153 CREATE RESOURCE POOL 参数说明

参数	描述	取值范围或示例
pool_name	资源池名称。资源池名称不能和当前数据库里已有的资源池重名。	字符串，要符合标识符的命名规范。

版权所有 © 华为云计算技术有限公司

参数	描述	取值范围或示例

group_name	控制组名称。
	控制组（Cgroups）是control groups的缩写，是Linux内核提供的一种可以限制、记录、隔离进程组所使用的物理资源（如：CPU、内存、IO等）的机制。如果一个进程加入了某一个控制组，该控制组对Linux的系统资源都有严格的限制，进程在使用这些资源时，不能超过其最大限制。更多Cgroups的原理介绍，请查看相关操作系统的产品手册。
	环境中已预置存在**DefaultClass控制组**，而DefaultClass控制组下层Timeshare控制组又分为以下4个层级：“Rush”、“High”、“Medium”或“Low”。
	● 设置控制组名称时，语法可以使用双引号，也可以使用单引号。
	● group_name对大小写敏感。
	● 不指定group_name时，默认指定的字符串为“Medium”，代表指定DefaultClass控制组的“Medium” Timeshare控制组。
		字符串，要符合说明中的规则，其指定已创建的控制组。
		创建一个默认资源池，其控制组为“DefaultClass”组下属的“Medium”Timeshare Workload控制组。
		`CREATE RESOURCE POOL pool1;`
		创建一个资源池，其控制组指定为“DefaultClass”组下属的“High”Timeshare Workload控制组。
		`CREATE RESOURCE POOL pool2 WITH (CONTROL_GROUP="High");`

ACTIVE_STATEMENTS	资源池**复杂语句**执行的最大并发数量。数值型，-1~INT_MAX，不指定时，默认是10。	创建资源池pool3，设定复杂语句执行的最大并发数量为5，内存使用比例50%，网络权重为5。 `CREATE RESOURCE POOL pool3 WITH (ACTIVE_STATEMENTS=5, MEM_PERCENT=50, weight=5);`
MAX_DOP	资源池**简单语句**执行的最大并发数量。数值型，-1~INT_MAX，-1和0代表不限制，不指定时，默认为-1，不限制。	创建资源池pool4，设定简单语句执行最大并发数量为10。 `CREATE RESOURCE POOL pool4 WITH (MAX_DOP=10);`

版权所有 © 华为云计算技术有限公司

参数	描述	取值范围或示例
MEMORY_LIMIT	使用该资源池的作业估算内存上限。	取值为字符串。 ● default，表示最大为资源池内存一半。 ● unlimited，表示不限制。 ● 数值范围为1KB~2047GB。 不指定时，默认为default。
MEM_PERCENT	资源池可用内存占全部内存或者组用户内存使用的比例。	取值范围为0-100，默认值为0。
io_limits	该参数8.1.2版本中已废弃，为兼容历史版本保留该参数。	-
io_priority	该参数8.1.2版本中已废弃，为兼容历史版本保留该参数。	-
nodegroup	在逻辑集群模式下，指定资源池所属的逻辑集群名称。必须是存在的逻辑集群。	如果逻辑集群名称包含大写字符、特殊符号或以数字开头，SQL语句中对逻辑集群名称需要加双引号。
is_foreign	在逻辑集群模式下，指定当前资源池用于控制没有关联本逻辑集群的普通用户的资源。这里的逻辑集群是由资源池nodegroup字段指定的。	● nodegroup必须是存在的逻辑集群，不能是elastic_group和安装的nodegroup (group_version1)。 ● 如果指定了is_foreign为true，则资源池不能再关联用户，即不允许通过CREATE USER ... RESOURCE POOL语句来将该资源池配置给用户。该资源池自动检查用户是否关联到资源池指定的逻辑集群，如果用户没有关联到该逻辑集群，则这些用户在逻辑集群所包含的DN上运行将受到该资源池的资源控制。

 版权所有 © 华为云计算技术有限公司

参数	描述	取值范围或示例
except_rule	异常规则集。 **异常规则集**是由一组定义了作业执行时间、作业排队时间、CPU使用率、最大落盘等各项规则阈值的集合，与资源池绑定搭配后使用，当用户执行作业时，触发了异常规则，则作业会选择终止或者降级。 了解更多，请参见**CREATE EXCEPT RULE**。	-
weight	资源池网络带宽权重。	取值为-1~2147483647的整数，默认配置为-1，不设置权重。
enable_concurrency_scaling	弹性并发扩展功能开关。该参数仅9.1.0.100及以上集群版本支持。	取值为true或false： • true表示开启该资源池作业的弹性并发扩展功能。 • false表示关闭该资源池作业的弹性并发扩展功能。 默认值：false

示例

本示例假定用户已预先成功创建控制组。

- 创建一个默认资源池，其控制组为“DefaultClass”组下属的“Medium” Timeshare Workload控制组：
```
DROP RESOURCE POOL IF EXISTS pool1;
CREATE RESOURCE POOL pool1;
```
- 创建一个资源池，其控制组指定为“DefaultClass”组下属的“High” Timeshare Workload控制组：
```
DROP RESOURCE POOL IF EXISTS pool2;
CREATE RESOURCE POOL pool2 WITH (CONTROL_GROUP="High");
```
- 创建资源池pool_p，设定复杂语句执行的最大并发数量为5，内存使用比例50%，网络权重为5。
```
DROP RESOURCE POOL IF EXISTS pool_p;
CREATE RESOURCE POOL pool_p WITH (ACTIVE_STATEMENTS=5, MEM_PERCENT=50, weight=5);
```
- 可通过**PG_RESOURCE_POOL**系统表查询资源池的配置参数。
```
SELECT * FROM PG_RESOURCE_POOL WHERE respool_name = 'pool_p';
```
```
gaussdb=> SELECT * FROM PG_RESOURCE_POOL WHERE respool_name = 'pool_p';                                                  M     at isa0060/serio1/
 respool_name | mem_percent | cpu_affinity |    control_group    | active_statements | max_dop | memory_limit | parentid | io_li
mits | io_priority |  nodegroup   | is_foreign | short_acc | except_rule | weight | enable_concurrency_scaling
-------------+-------------+--------------+---------------------+-------------------+---------+--------------+----------+---------+------
-----+-------------+--------------+------------+-----------+-------------+--------+---------------------------
 pool_p      |          50 |           -1 | DefaultClass:Medium |                 5 |      -1 | default      |        0 |
   0 | None        | installation | f          | t         | None        |      5 | f
(1 row)
```
- 创建一个异常规则集except_rule4，设置执行的最大时间elapsedtime为10秒，并创建一个资源池同时绑定异常规则。
```
DROP RESOURCE POOL IF EXISTS resource_pool_a4;
DROP EXCEPT RULE IF EXISTS except_rule4;
```


版权所有 © 华为云计算技术有限公司

```
CREATE EXCEPT RULE except_rule4 WITH (elapsedtime=10);
CREATE resource pool resource_pool_a4 WITH (except_rule='except_rule4');
```

- 创建一个资源池，并打开该资源池的弹性并发扩展功能：

```
DROP RESOURCE POOL IF EXISTS pool6;
CREATE RESOURCE POOL pool6 WITH (enable_concurrency_scaling=true);
```

示例：创建异常规则，当作业执行时间超过 10 秒，则终止该作业

在DWS资源负载管理中，异常规则需要绑定到资源池后生效，紧接着将执行作业的用户绑定到指定资源池后，总体上才能达到资源负载管理的效果，当用户执行某作业时，若触发异常规则，则作业将被终止，以下示例模拟了这一终止过程。

步骤1 先创建一个测试表并导入数据。

1. 以系统管理员连接数据库执行以下SQL，创建一个名为src的临时表，为后续的INSERT操作提供数据源。

```
DROP SCHEMA IF EXISTS test CASCADE;
CREATE SCHEMA test;
CREATE TABLE test.src AS SELECT 1;
```

2. 执行以下SQL，创建表test.test。

```
DROP TABLE IF EXISTS test.test;
CREATE TABLE test.test(a int, b numeric(15,2)) WITH(orientation=column);
```

3. 执行以下SQL，分别为两张表插入2000万行数据。执行时间约半分钟，请等待。

```
INSERT INTO test.test SELECT generate_series(1,20000000) % 1000,generate_series(1,20000000)
FROM test.src;
```

步骤2 创建一个异常规则集except_rule4，设置执行的最大时间elapsedtime为10秒

```
DROP EXCEPT RULE IF EXISTS except_rule4;
CREATE EXCEPT RULE except_rule4 WITH (elapsedtime=10);
```

步骤3 创建一个资源池resource_pool_a4，并将异常规则绑定到该资源池。

```
DROP resource pool IF EXISTS resource_pool_a4;
CREATE resource pool resource_pool_a4 WITH (except_rule='except_rule4');
```

步骤4 创建一个作业用户，并绑定到资源池resource_pool_a4中，密码请用户自定义。

```
CREATE USER user4 RESOURCE POOL 'resource_pool_a4' PASSWORD '{password}';
```

步骤5 授权作业用户user4可以访问表test.test。

```
GRANT USAGE ON SCHEMA test TO user4;
GRANT SELECT ON TABLE test.test TO user4;
```

步骤6 以作业用户user4连接数据库，执行以下SQL，超过10秒，语句被终止了。

```
SELECT * FROM test.test;
```

```
gaussdb=> SELECT * FROM test.test;
ERROR:  canceling statement due to workload manager exception.
DETAIL: except rule [except_rule4] is meet condition:rule [elapsedtime] is over limit, current value is:11.
gaussdb=>
```

步骤7 通过**PG_EXCEPT_RULE**系统表，可以查看异常规则定义。

```
SELECT * FROM PG_EXCEPT_RULE WHERE name = 'except_rule4';
```

```
gaussdb=> SELECT * FROM PG_EXCEPT_RULE WHERE name = 'except_rule4';
     name     |    rule     | value
--------------+-------------+-------
 except_rule4 | action      | abort
 except_rule4 | elapsedtime | 10
(2 rows)
```

----结束

 版权所有 © 华为云计算技术有限公司

相关链接

ALTER RESOURCE POOL，DROP RESOURCE POOL，CREATE EXCEPT RULE

使用实践：为两种作业绑定不同资源池以实现DWS资源负载能力

13.9.2 ALTER RESOURCE POOL

功能描述

修改一个资源池，指定其他控制组。

注意事项

只要用户对当前数据库有ALTER权限，就可以修改资源池。

语法格式

```
ALTER RESOURCE POOL pool_name
    WITH ({MEM_PERCENT=pct | CONTROL_GROUP='group_name' | ACTIVE_STATEMENTS=stmt | MAX_DOP
= dop | MEMORY_LIMIT='memory_size' | io_limits=io_limits | io_priority='priority' |
nodegroup='nodegroup_name' | except_rule='except_rule' | weight=bandwidth_weight |
enable_concurrency_scaling=boolean} [, ... ]);
```

参数说明

表 13-154 ALTER RESOURCE POOL 参数说明

参数	描述	取值范围
pool_name	需要修改的资源池名称。	已存在的资源池。
group_name	控制组名称。	已存在的控制组。
stmt	资源池语句执行的最大并发数量。	数值型，-1~INT_MAX。
dop	资源池简单语句执行的最大并发数量。	数值型，1~INT_MAX。
memory_size	使用该资源池的作业估算内存上限。	字符串 • default，表示最大为资源池内存一半。 • unlimited，表示不限制。 • 数值范围为1KB~2047GB。
mem_percent	资源池可用内存占全部内存或者组用户内存使用的比例。	普通用户的mem_percent范围为0-100的整数，默认值为0。
io_limits	该参数8.1.2集群版本中已废弃，为兼容历史版本保留该参数。	-

版权所有 © 华为云计算技术有限公司

参数	描述	取值范围
io_priority	该参数8.1.2集群版本中已废弃，为兼容历史版本保留该参数。	-
except_rule	异常规则集。 **异常规则集**是由一组定义了作业执行时间、作业排队时间、CPU使用率、最大落盘等各项规则阈值的集合，与资源池绑定搭配后使用，当用户执行作业时，触发了异常规则，则作业会选择终止或者降级。 了解更多，请参见**CREATE EXCEPT RULE**。	-
weight	资源池网络带宽权重。	-
enable_concurrency_scaling	弹性并发扩展功能开关。该参数仅9.1.0.100及以上集群版本支持。 默认值为false。	true或false。 • true，表示开启该资源池作业的弹性并发扩展功能。 • false，表示关闭该资源池作业的弹性并发扩展功能。

说明

- 设置控制组名称“**group_name**”时，语法可以使用双引号，也可以使用单引号。
- group_name对大小写敏感。
- 不指定group_name时，默认指定的字符串为 "Medium"，代表指定DefaultClass控制组的 "Medium" Timeshare控制组。
- 若数据库用户指定Timeshare控制组代表的字符串，即"Rush"、"High"、"Medium"或"Low"其中一种，如control_group的字符串为"High"；代表资源池指定到DefaultClass控制组下的 "High" Timeshare控制组。

示例

- 创建示例资源池pool_test，其控制组为“DefaultClass”组下属的“Medium” Timeshare Workload控制组。
```
DROP RESOURCE POOL IF EXISTS pool_test;
CREATE RESOURCE POOL pool_test;
```
- 修改资源池pool_test，其控制组指定为“DefaultClass”组下属的“High” Timeshare Workload控制组。
```
ALTER RESOURCE POOL pool_test WITH (CONTROL_GROUP="High");
```
- 关闭指定资源池作业的弹性并发扩展功能。
```
ALTER RESOURCE POOL pool_test WITH (enable_concurrency_scaling=false);
```

相关链接

CREATE RESOURCE POOL，**DROP RESOURCE POOL**

版权所有 © 华为云计算技术有限公司

13.9.3 DROP RESOURCE POOL

功能描述

删除一个资源池。

说明

如果某个角色已关联到该资源池，则无法删除该资源池。

注意事项

只要用户对当前数据库有DROP权限，就可以删除资源池。

语法格式

```
DROP RESOURCE POOL [ IF EXISTS ] pool_name;
```

参数说明

表 13-155 DROP RESOURCE POOL 参数说明

参数	描述	取值范围
IF EXISTS	如果指定的资源池不存在，系统给出提示而不是直接报错。	-
pool_name	已创建过的资源池名称。	字符串，要符合标识符的命名规范。

示例

删除资源池pool：

```
DROP RESOURCE POOL pool;
```

相关链接

ALTER RESOURCE POOL，**CREATE RESOURCE POOL**

13.9.4 CREATE EXCEPT RULE

功能描述

用于创建一个异常规则集合。创建时可以指定任意一个或者多个具体的规则阈值，也可以指定触发异常规则后所采取的操作，当前支持终止（abort）和降级（penalty）操作。

分布式数据库中，查询任务由多个节点并行执行，每个节点独立处理部分数据并汇总结果。但若任一节点出现性能瓶颈，即使其他节点快速完成，整个查询仍会被拖慢。在某些极端情况下，一个查询语句可能会占用大量资源，导致节点资源使用率过高，从而引发计算资源或存储资源不足的问题，不仅影响业务运行，还可能导致整个集群性能下降甚至不可用。

版权所有 © 华为云计算技术有限公司

DWS提供配置**异常规则**功能，允许用户设置查询的资源限制（如执行时间、内存、磁盘等），SQL语句在执行时触发异常规则会自动终止，避免异常查询拖垮集群，保障稳定性并减轻DBA运维负担。例如设置查询的最大执行时间，超时则自动终止；设置查询可以使用的磁盘空间大小，超过阈值则自动终止；或者设置查询可以使用的内存量，超过限制则终止。

在DWS资源负载管理中，异常规则需要绑定到资源池后生效，紧接着将执行作业的用户绑定到指定资源池后，总体上才能达到资源负载管理的效果，当用户执行某作业时，若触发异常规则，则作业将被终止。

您也可以**通过管理控制台**进行资源池和异常规则设置，了解**资源池、异常规则**介绍及配置，可参见**DWS资源管理**。

注意事项

同一组异常规则中的多个规则以“且”的关系生效，例如：同时设置执行时间elapsedtime=1000，内存大小memsize=500。表示一个作业，执行时间达到1000秒且内存使用达到500MB时，才会终止该作业。**仅触发其中一个不会生效。**

语法格式

```
CREATE EXCEPT RULE except_rule_name
    WITH (
        | BLOCKTIME = VALUE,
        | CPUTIME = VALUE,
        | ELAPSEDTIME = VALUE,
        | CPUSKEWPERCENT = VALUE,
        | SPILLSIZE = VALUE,
        | BROADCASTSIZE = VALUE,
        | MEMSIZE = VALUE,
        | CPUAVGPERCENT = VALUE,
        | BANDWIDTH = VALUE,
        | ACTION = ['abort' | 'penalty']
        );
```

参数说明

表 13-156 CREATE EXCEPT RULE 参数说明

参数	描述	取值范围
rule_name	异常规则集名称。	字符串，要符合标识符的命名规范。参见**标识符命名规范**。
blocktime	作业排队阻塞的最大时间，单位：秒。	数值型，-1，1~INT64_MAX。-1表示不设置，即不限制，下同。
elapsedtime	作业执行的最大时间，单位：秒。	数值型，-1，1~INT64_MAX。
allcputime	作业运行中使用的最大CPU时间，单位：秒。	数值型，-1，1~INT64_MAX。
cpuskewpercent	作业执行时的平均CPU使用率，单位：百分比。	数值型，-1，1~100。
cpuavgpercent	作业执行时的CPU使用倾斜率，单位：百分比。	数值型，-1，1~100。

版权所有 © 华为云计算技术有限公司

参数	描述	取值范围
spillsize	作业执行的最大下盘大小，单位：MB。	数值型，-1，1~INT64_MAX。
broadcastsize	作业执行的最大广播大小，单位：MB。	数值型，-1，1~INT64_MAX。
memsize	作业执行使用的最大内存大小，单位：MB。	数值型，-1，1~INT64_MAX。
bandwidth	作业执行可使用的最大带宽，单位：MB。	数值型，-1，1~INT64_MAX。

基础示例

创建一个异常规则集except_rule1并指定其排队阻塞时间blocktime为3000秒，下盘空间为4000MB，即某个语句排队阻塞超过3000秒且下盘空间超过4000MB，该语句会被终止。

```
DROP EXCEPT RULE IF EXISTS except_rule1;
CREATE EXCEPT RULE except_rule1 WITH (blocktime=3000, spillsize=4000, action=abort);
```

创建一个异常规则集except_rule2并指定其memsize规则阈值为5000MB，默认所采取的异常规则操作为abort。

```
DROP EXCEPT RULE IF EXISTS except_rule2;
CREATE EXCEPT RULE except_rule2 WITH (memsize=3000);
```

创建一个资源池并绑定异常规则集except_rule1。

```
DROP resource pool IF EXISTS resource_pool_a1;
CREATE resource pool resource_pool_a1 WITH (except_rule='except_rule1');
```

示例：创建异常规则，当作业执行时间超过 10 秒，则终止该作业

在DWS资源负载管理中，异常规则需要绑定到资源池后生效，紧接着将执行作业的用户绑定到指定资源池后，总体上才能达到资源负载管理的效果，当用户执行某作业时，若触发异常规则，则作业将被终止，以下示例模拟了这一终止过程。

步骤1 先创建一个测试表并导入数据。

1. 以系统管理员连接数据库执行以下SQL，创建一个名为src的临时表，为后续的INSERT操作提供数据源。
   ```
   DROP SCHEMA IF EXISTS test CASCADE;
   CREATE SCHEMA test;
   CREATE TABLE test.src AS SELECT 1;
   ```
2. 执行以下SQL，创建表test.test。
   ```
   DROP TABLE IF EXISTS test.test;
   CREATE TABLE test.test(a int, b numeric(15,2)) WITH(orientation=column);
   ```
3. 执行以下SQL，分别为两张表插入2000万行数据。执行时间约半分钟，请等待。
   ```
   INSERT INTO test.test SELECT generate_series(1,20000000) % 1000,generate_series(1,20000000)
   FROM test.src;
   ```

步骤2 创建一个异常规则集except_rule4，设置执行的最大时间elapsedtime为10秒

```
DROP EXCEPT RULE IF EXISTS except_rule4;
CREATE EXCEPT RULE except_rule4 WITH (elapsedtime=10);
```

 版权所有 © 华为云计算技术有限公司

步骤3 创建一个资源池resource_pool_a4，并将异常规则绑定到该资源池。

```
DROP resource pool IF EXISTS resource_pool_a4;
CREATE resource pool resource_pool_a4 WITH (except_rule='except_rule4');
```

步骤4 创建一个作业用户，并绑定到资源池resource_pool_a4中，密码请用户自定义。

```
CREATE USER user4 RESOURCE POOL 'resource_pool_a4' PASSWORD '{password}';
```

步骤5 授权作业用户user4可以访问表test.test。

```
GRANT USAGE ON SCHEMA test TO user4;
GRANT SELECT ON TABLE test.test TO user4;
```

步骤6 以作业用户user4连接数据库，执行以下SQL，超过10秒，语句被终止了。

```
SELECT * FROM test.test;
```

```
gaussdb=> SELECT * FROM test.test;
ERROR:  canceling statement due to workload manager exception.
DETAIL: except rule [except_rule4] is meet condition:rule [elapsedtime] is over limit, current value is:11.
gaussdb=>
```

步骤7 通过**PG_EXCEPT_RULE**系统表，可以查看异常规则定义。

```
SELECT * FROM PG_EXCEPT_RULE WHERE name = 'except_rule4';
```

```
gaussdb=> SELECT * FROM PG_EXCEPT_RULE WHERE name = 'except_rule4';
     name     |    rule     | value
--------------+-------------+-------
 except_rule4 | action      | abort
 except_rule4 | elapsedtime | 10
(2 rows)
```

----结束

相关链接

ALTER EXCEPT RULE，**DROP EXCEPT RULE**，**CREATE RESOURCE POOL**

使用实践：为两种作业绑定不同资源池以实现DWS资源负载能力

13.9.5 ALTER EXCEPT RULE

功能描述

修改一个异常规则集合，可以修改一组规则集中的任意一个或者多个具体的规则阈值。

注意事项

添加某个新规则阈值时，WITH参数中原有的其他规则会保留，并不是全部清空覆盖。例如原来一组规则中只有一条规则：作业排队阻塞的最大时间blocktime = 100秒。通过ALTER EXCEPT RULE语法新添加一条规则，设置作业执行最大时间elapsedtime为1000秒，则最终的规则包含两条。

语法格式

```
ALTER EXCEPT RULE except_rule_name
    WITH (
        | BLOCKTIME = VALUE,
        | CPUTIME = VALUE,
        | ELAPSEDTIME = VALUE,
```

 版权所有 © 华为云计算技术有限公司

```
| CPUSKEWPERCENT = VALUE,
| SPILLSIZE = VALUE,
| BROADCASTSIZE = VALUE,
| MEMSIZE = VALUE,
| CPUAVGPERCENT = VALUE,
| BANDWIDTH = VALUE,
| ACTION = ['abort' | 'penalty']
);
```

参数说明

表 13-157 ALTER EXCEPT RULE 参数说明

参数	描述	取值范围
rule_name	异常规则集名称。	字符串，需符合**标识符命名规范**。
blocktime	作业排队阻塞的最大时间。单位：秒。	数值型，-1，1~INT64_MAX。
elapsedtime	作业执行的最大时间。单位：秒。	数值型，-1，1~INT64_MAX。
allcputime	作业运行中使用的最大CPU时间。单位：秒。	数值型，-1，1~INT64_MAX。
cpuskewpercent	作业执行时的CPU使用倾斜率。单位：百分比。	数值型，-1，1~100。
cpuavgpercent	作业执行时的平均CPU使用率。单位：百分比。	数值型，-1，1~100。
spillsize	作业执行的最大下盘大小。单位：MB。	数值型，-1，1~INT64_MAX。
broadcastsize	作业执行的最大广播大小。单位：MB。	数值型，-1，1~INT64_MAX。
memsize	作业执行使用的最大内存大小。单位：MB。	数值型，-1，1~INT64_MAX。
bandwidth	作业执行可使用的最大网络带宽。单位：MB。	数值型，-1，1~INT64_MAX。

示例

创建异常规则except_rule1。

```
CREATE EXCEPT RULE except_rule_1 WITH (blocktime=2000, spillsize=3000, action=abort);
```

修改异常规则except_rule1的blocktime规则阈值为3000秒，下盘空间为4000MB。

```
ALTER EXCEPT RULE except_rule_1 WITH (blocktime=3000, spillsize=4000);
```

通过**PG_EXCEPT_RULE**系统表，可以查看异常规则已修改成功。

```
SELECT * FROM PG_EXCEPT_RULE WHERE name = 'except_rule_1';
```


版权所有 © 华为云计算技术有限公司

```
gaussdb=> SELECT * FROM PG_EXCEPT_RULE WHERE name = 'except_rule_1';
     name      |   rule    | value
---------------+-----------+-------
 except_rule_1 | action    | abort
 except_rule_1 | blocktime | 3000
 except_rule_1 | spillsize | 4000
(3 rows)
```

同时，需要再叠加一条新规则，作业执行的最大时间elapsedtime为2000秒。

```
ALTER EXCEPT RULE except_rule_1 WITH (elapsedtime=2000);
```

再次查看异常规则，已修改成功。**注意，原先的规则会保留，并不是清空。**

```
SELECT * FROM PG_EXCEPT_RULE WHERE name = 'except_rule_1';
```

```
gaussdb=> SELECT * FROM PG_EXCEPT_RULE WHERE name = 'except_rule_1';
     name      |    rule     | value
---------------+-------------+-------
 except_rule_1 | action      | abort
 except_rule_1 | blocktime   | 3000
 except_rule_1 | spillsize   | 4000
 except_rule_1 | elapsedtime | 2000
(4 rows)
```

创建一个资源池并绑定异常规则集except_rule_1。

```
DROP resource pool IF EXISTS resource_pool_a1;
CREATE resource pool resource_pool_a1 WITH (except_rule='except_rule_1');
```

解除资源池resource_pool_a1绑定的异常规则集。

```
ALTER resource pool resource_pool_a1 WITH (except_rule='None');
```

相关链接

CREATE EXCEPT RULE，DROP EXCEPT RULE

13.9.6 DROP EXCEPT RULE

功能描述

用于删除指定异常规则集合。

注意事项

只有系统管理员才能执行DROP EXCEPT RULE操作。

语法格式

```
DROP EXCEPT RULE [ IF EXISTS ]
  rule_name ;
```

 版权所有 © 华为云计算技术有限公司

参数说明

表 13-158 DROP EXCEPT RULE 参数说明

参数	描述	取值范围
IF EXISTS	如果指定的异常规则不存在时返回一个提示而不是错误。	-
rule_name	需要删除的异常规则集名称。	已存在的异常规则集名称。

示例

创建异常规则集合except_rule_1。

```
DROP EXCEPT RULE IF EXISTS except_rule_1;
CREATE EXCEPT RULE except_rule_1 WITH (blocktime=2000, spillsize=3000, action=abort);
```

删除名称为except_rule_1的规则集合。

```
DROP EXCEPT RULE except_rule_1;
```

如果存在名为except_rule_2的规则集合，则将其删除。

```
DROP EXCEPT RULE IF EXISTS except_rule_2;
```

相关链接

CREATE EXCEPT RULE，**ALTER EXCEPT RULE**

13.9.7 CREATE BLOCK RULE

功能描述

该语法用于创建**一条查询过滤规则**，包括查询过滤规则名称、绑定的客户端名称、客户端IP、用户以及匹配方式等。在日常数据库操作中，异常SQL的执行可能会导致资源过度消耗，影响数据库性能和稳定性，通过创建查询过滤规则，可以将此类SQL语句进行拦截。

当前支持**指定某个SQL语句的正则匹配**进行拦截，例如针对某用户对某张表的SELECT进行拦截；也支持**指定某SQL语句的unique_sql_id或sql_hash取值**进行拦截，例如已知某SQL语句的**unique_sql_id**值，将此SQL语句直接拦截。

本文档提供基础SQL语法，您也可以通过控制台进行配置，参见**使用DWS查询过滤器拦截慢SQL**。

注意事项

- 该语法仅9.1.0.100及以上集群版本支持。
- 只有拥有数据库所有者权限的用户或者授予gs_role_block角色权限的用户才能执行CREATE BLOCK RULE命令，系统管理员默认拥有此权限。
- 普通用户无法执行CREATE BLOCK RULE，可通过系统管理员进行授权，授权语法：
  ```
  GRANT gs_role_block TO user;
  ```


版权所有 © 华为云计算技术有限公司

- 建议**创建查询过滤规则时尽量缩小适用范围，避免误过滤，或者范围过大导致性能劣化**。

查询过滤规则的约束和限制

生效范围与隔离性

- 查询过滤器规则按**数据库级别**隔离，不支持跨数据库使用。
- 查询过滤器仅对**外部语句**生效，对内部语句不生效。
- 查询过滤器仅管控**主语句**，不管控子语句。

规则数量与参数限制

- 9.1.1.300及以上集群版本中，单个数据库最大支持创建**64条**查询过滤规则，超出后会报错无法创建。
- 单条规则可配置的最大并发数（max_active_num）为**1000**。

多规则命中时的处理

- 同一作业命中多条规则时，各规则的并发计数独立维护，作业被并发数上限最小的规则拦截。
- 同一作业命中多条规则，且同时存在报错与告警行为时，先打印所有告警规则相关信息，再报错退出。

CCN排队与并发管控交互场景

- CCN排队优先于查询过滤器并发控制。作业先经过CCN排队并发管控，再受查询过滤器并发控制。
- 作业在单个资源池内达到资源池并发上限时，进入CCN排队；达到查询过滤器并发控制上限时，作业执行失败并报错。
- CCN重启期间：非CCN节点与CCN通信失败，作业不经过查询过滤器并发管控，直接执行；CCN恢复后，查询过滤器管控重新生效。

语法格式

```
CREATE BLOCK RULE [ IF NOT EXISTS ] block_name
   [ [ TO user_name@'host' ] | [ TO user_name ] | [ TO 'host' ] ] |
   [ FOR UPDATE | SELECT | INSERT | DELETE | MERGE ] |
   FILTER BY
   { SQL ( 'text' ) | TEMPLATE ( template_parameter = value ) }
   [ WITH ( { with_parameter = value }, [, ... ] ) ];
```

参数说明

表 13-159 CREATE BLOCK RULE 参数说明

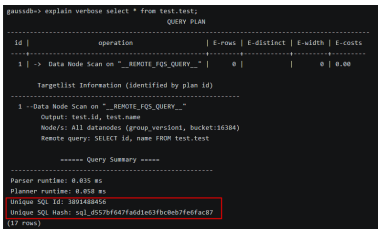
参数	描述	取值范围或示例
block_name	要创建的查询过滤规则名称。	字符串，需符合**标识符命名规范**。

版权所有 © 华为云计算技术有限公司

参数	描述	取值范围或示例

TO	对查询过滤规则生效的对象或客户端IP，支持以下方式：
	● **TO user_name@'host'**：用户名以及客户端IP，两者都要满足，查询过滤规则才能生效。
	● **TO user_name**：用户名，即只判定满足用户名条件，查询过滤规则就生效，不区分客户端IP。
	● **TO 'host'**：客户端IP，即只判定满足客户端IP条件，查询过滤规则就生效，不区分用户。
		示例：
		查询过滤规则对来自客户端IP为192.168.*.**且登录用户名为user1时生效。
		`CREATE BLOCK RULE query_block TO user1@'192.168.*.**' FOR SELECT FILTER BY SQL('test.test');`
		查询过滤规则对登录用户名为user1时生效（无论客户端IP是哪个）。
		`CREATE BLOCK RULE query_block TO user1 FOR SELECT  FILTER BY SQL('test.test');`
		查询过滤规则对来自客户端IP 192.168.*.**的访问时都生效（无论哪个用户登录）。
		`CREATE BLOCK RULE query_block TO '192.168.*.**' FOR SELECT  FILTER BY SQL('test.test');`

FOR	语句类型，支持对UPDATE/SELECT/INSERT/DELETE/MERGE INTO五种类型语句进行限制。	-

FILTER BY	过滤方法，支持两种形式：
	● **SQL**：根据SQL语句内容进行过滤，通过关键词对语句进行正则匹配，例如表名，其长度不能超过1024，建议尽量精简。
	● **TEMPLATE**：根据模板进行过滤，**此方法不区分用户，针对所有用户执行的特定语句都实施拦截，拦截效果更高效**，取值如下：
	– **unique_sql_id**：归一化的64位哈希值，重复概率较sql_hash大一些。
	– **sql_hash**：归一化的哈希值（MD5），一般不会重复，相较unique_sql_id更推荐使用。
		unique_sql_id和sql_hash取值，可以通过EXPLAIN VERBOSE查询，例如：
		`EXPLAIN VERBOSE SELECT * FROM test.test;`
		
		`gaussdb=> explain verbose select * from test.test;`
		`QUERY PLAN`
		`id | operation | E-rows | E-distinct | E-width | E-costs`
		`1 | -> Data Node Scan on "__REMOTE_FQS_QUERY__" | 0 | | 0 | 0.00`
		`Targetlist Information (identified by plan id)`
		`1 --Data Node Scan on "__REMOTE_FQS_QUERY__"`
		`Output: test.id, test.name`
		`Node/s: All datanodes (group_version1, bucket:16384)`
		`Remote query: SELECT id, name FROM test.test`
		`====== Query Summary =====`
		`Parser runtime: 0.035 ms`
		`Planner runtime: 0.058 ms`
		`Unique SQL Id: 3891488456`
		`Unique SQL Hash: sql_d557bf047fa6d1e63fbc0eb7fe6fac87`
		`(17 rows)`

版权所有 © 华为云计算技术有限公司

参数	描述	取值范围或示例
with_parameter	查询过滤规则选项参数。	以下选项可以一起设置，满足任何一个参数的限制查询将会被过滤。 ● application_name，客户端名称。 ● query_band，负载标识。 ● table_num，语句扫描的基表个数。 ● partition_num，算子预估扫描的最大分区数。 ● estimate_row，算子预估扫描的最大表行数。 ● resource_pool，需要切换到的资源池名称。 ● max_active_num，规则对应语句的最大并发数。 ● is_warning，控制拦截到语句的行为是告警还是报错。on表示告警，off表示报错。默认为off。

示例：基于匹配 SQL 语句进行 SQL 拦截

以系统管理员连接数据库执行以下SQL，创建一个名为src的临时表，为后续的INSERT操作提供数据源。

```
DROP SCHEMA IF EXISTS test CASCADE;
CREATE SCHEMA test;
CREATE TABLE test.src AS SELECT 1;
```

执行以下SQL，创建表test.test。

```
DROP TABLE IF EXISTS test.test;
CREATE TABLE test.test(a int, b numeric(15,2)) WITH(orientation=column);
```

执行以下SQL，分别为两张表插入2000万行数据。执行时间约半分钟，请等待。

```
INSERT INTO test.test SELECT generate_series(1,20000000) % 1000,generate_series(1,20000000) FROM test.src;
```

创建用户user1，并授权该用户可以访问表test.test。

```
DROP USER IF EXISTS user1 CASCADE;
CREATE USER user1 PASSWORD '{password}';
GRANT USAGE ON SCHEMA test TO user1;
GRANT SELECT ON TABLE test.test TO user1;
```

以用户user1连接数据库，查询表数据，结果表明在没有创建过滤规则的情况下，用户user1可以正常查询表test.test数据。

```
SELECT sum(b) FROM test.test GROUP BY a;
```

 版权所有 © 华为云计算技术有限公司

```
gaussdb=> SELECT sum(b) FROM test.test GROUP BY a;
        sum
------------------
 199990060000.00
 199990120000.00
 199990180000.00
 199990240000.00
 199990300000.00
 199990360000.00
 199990420000.00
 199990480000.00
 199990540000.00
 199990600000.00
 199990660000.00
 199990720000.00
 199990780000.00
```

以系统管理员连接数据库，创建名称为query_block的查询过滤规则，对于用户user1查询的SQL语句进行拦截，过滤方式为SQL语句，SQL语句中出现关键字test(即表名)会被过滤。

```
CREATE BLOCK RULE query_block
TO user1@'192.168.0.59'        --规则作用于用户user1，且客户端IP为192.168.0.59。
FOR SELECT                     --过滤SELECT的相关操作。
FILTER BY
SQL('test.test')               --过滤方式为SQL语句，SQL语句中出现关键字test(即表名)会被过滤。
WITH                           --满足以下条件即被过滤。
(
application_name='gsql',       --客户端名称为gsql
table_num='1',                 --扫描基表次数为1
estimate_row='10',             --扫描基表行数为10
is_warning='off'               --关闭告警，直接报错
)
;
```

通过视图**pg_get_blockruledef**查询已创建的过滤规则，从回显判断出规则已生效。

```
SELECT * FROM pg_get_blockruledef('query_block');
```

```
gaussdb=> SELECT * FROM pg_get_blockruledef('query_block');
                                                
```

从回显中记录unique_sql_id和sql_hash。

```
             ====== Query Summary =====
 ---------------------------------------------------
 System available mem: 1269760KB
 Query Max mem: 1269760KB
 Query estimated mem: 1919KB
 Turbo Engine: true
 Parser runtime: 0.071 ms
 Planner runtime: 0.337 ms
 Unique SQL Id: 3324222175
 Unique SQL Hash: sql_170d0c75c472043b0cd35e05903c382f
(34 rows)
```

删除原来的过滤规则，创建基于unique_sql_id的过滤规则。

```
DROP BLOCK RULE query_block;
CREATE BLOCK RULE query_block FILTER BY TEMPLATE(unique_sql_id='3324222175');
```

再次切到user1用户，查询报错，语句被拦截了。

```
SELECT sum(b) FROM test.test GROUP BY a;
```

```
gaussdb=> SELECT sum(b) FROM test.test GROUP BY a;
ERROR:  hit block rule query_block(unique_sql_id: 3324222175)
```

同样，以系统管理员去执行一遍，语句会被拦截，由此可知，**通过unique_sql_id进行过滤拦截，并不区分用户，对所有用户执行此类SQL语句都会进行拦截，拦截效果更优。**

```
SELECT sum(b) FROM test.test GROUP BY a;
```

```
gaussdb=> SELECT sum(b) FROM test.test GROUP BY a;
ERROR:  hit block rule query_block(unique_sql_id: 3324222175)
```

删除过滤规则，创建基于sql_hash的过滤规则。

```
DROP BLOCK RULE query_block;
CREATE BLOCK RULE query_block FILTER BY
TEMPLATE(sql_hash='sql_170d0c75c472043b0cd35e05903c382f');
```

再次切到user1用户，查询报错，语句被拦截了。

```
gaussdb=> SELECT sum(b) FROM test.test GROUP BY a;
ERROR:  hit block rule query_block(sql_hash: sql_170d0c75c472043b0cd35e05903c382f)
```

相关链接

ALTER BLOCK RULE，**DROP BLOCK RULE**

本文档提供基础SQL语法，您也可以通过控制台进行配置，参见**使用DWS查询过滤器拦截慢SQL**。

 版权所有 © 华为云计算技术有限公司

常见问题：创建了查询过滤规则后，并未成功拦截语句

不同的计划，可能部分字段无法按照预期进行拦截，例如：

```
CREATE BLOCK RULE test filter BY SQL('test')with(estimate_row='3');
SELECT * FROM test;
```

```
 c1 | c2
----+----
  1 |  2
  1 |  2
  1 |  2
  1 |  2
  1 |  2
(5 rows)
```

此时，语句关键字是匹配的，查询的行数也超过了3行的限制，但是没有成功拦截。

```
postgres=# explain verbose select * from test;
                                    QUERY PLAN
-----------------------------------------------------------------------------------------------
 id |                operation                 | E-rows | E-distinct | E-width | E-costs
----+------------------------------------------+--------+------------+---------+---------
  1 | ->  Data Node Scan on "__REMOTE_FQS_QUERY__" |      0 |            |       0 | 0.00

    Targetlist Information (identified by plan id)
-------------------------------------------------------------
  1 --Data Node Scan on "__REMOTE_FQS_QUERY__"
        Output: test.c1, test.c2
        Node/s: All datanodes (node_group, bucket:16384)
        Remote query: SELECT c1, c2 FROM public.test
```

通过计划可以看出，此时是**FQS（fast_query_shipping）**计划，即参数**enable_fast_query_shipping**设置为on，导致没有估算信息。因此无法进行拦截，对于CN轻量化的计划也是一样的，如果让语句强制使用stream计划，那么就可以拦截成功：

```
SET enable_stream_operator=on;
SET enable_fast_query_shipping=off;
```

 版权所有 © 华为云计算技术有限公司

```
postgres=# set enable_stream_operator=on;
SET
postgres=# set enable_fast_query_shipping=off;
SET
postgres=# select * from test;
ERROR:  hit block rule test(regexp_sql: test, estimate_row: 3(5))
postgres=#  explain verbose select * from test;
                                  QUERY PLAN
------------------------------------------------------------------------------------------
 id |               operation                | E-rows | E-distinct | E-width | E-costs
----+----------------------------------------+--------+------------+---------+---------
  1 | ->  Row Adapter                        |      5 |            |       8 | 69.00
  2 |    ->  Vector Streaming (type: GATHER) |      5 |            |       8 | 69.00
  3 |       ->  CStore Scan on public.test   |      5 |            |       8 | 59.01

     Targetlist Information (identified by plan id)
-----------------------------------------------------------
  1 --Row Adapter
        Output: c1, c2
  2 --Vector Streaming (type: GATHER)
        Output: c1, c2
        Node/s: All datanodes (node_group, bucket:16384)
  3 --CStore Scan on public.test
        Output: c1, c2
        Distribute Key: c1
```

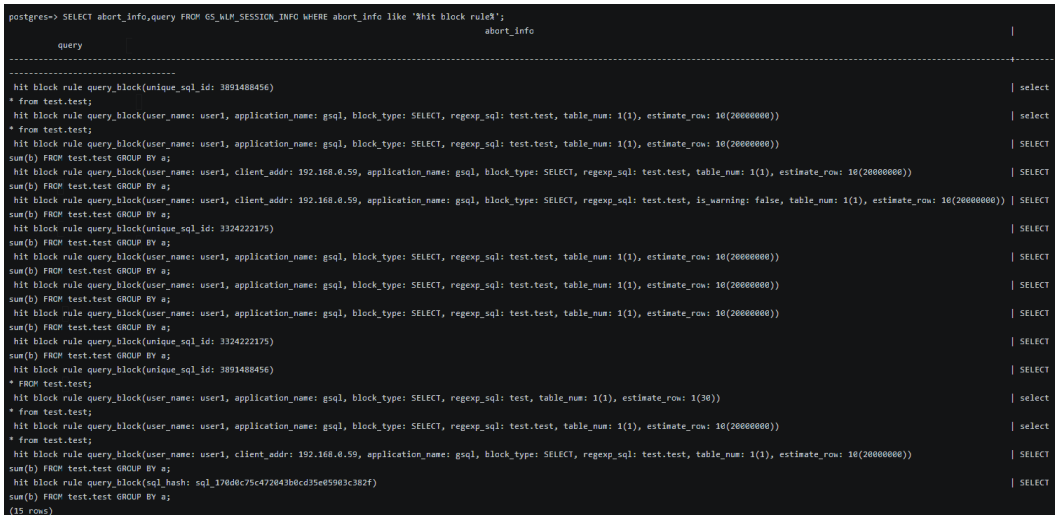
综上所述，如果估算信息不准确，也会导致误拦截或者漏拦截的情况，因为计划的信息是通过估算所得，通常这种情况无法避免。

常见问题：如何查看拦截记录

创建查询过滤规则后会拦截很多慢SQL，可以通过TopSQL系统表GS_WLM_SESSION_INFO查看已拦截的语句，abort_info会记录拦截信息，也就是查询的报错信息。

需连接postgres数据库执行：

```
SELECT abort_info,query FROM GS_WLM_SESSION_INFO WHERE abort_info like '%hit block rule%';    --查询所有拦截记录
SELECT abort_info,query FROM GS_WLM_SESSION_INFO WHERE abort_info like '%hit block rule query_block_new%'   --只查询规则名称为query_block_new的拦截记录
```



 版权所有 © 华为云计算技术有限公司

13.9.8 ALTER BLOCK RULE

功能描述

修改查询过滤规则属性，包括查询过滤规则名称、绑定的客户端名称、客户端IP、用户以及匹配方式等。

在日常数据库操作中，异常SQL的执行可能会导致资源过度消耗，影响数据库性能和稳定性，通过创建查询过滤规则，可以将此类SQL语句进行拦截。

注意事项

- 该语法仅9.1.0.100及以上集群版本支持。
- 只有拥有数据库所有者权限的用户或者授予gs_role_block角色权限的用户才能执行ALTER BLOCK RULE命令，系统管理员默认拥有此权限。
- 普通用户无法执行CREATE BLOCK RULE，可通过系统管理员进行授权，授权语法：
  ```
  GRANT gs_role_block TO user;
  ```
  建议**创建查询过滤规则时尽量缩小适用范围，避免误过滤，或者范围过大导致性能劣化**。
- DEFAULT语法用于对某一属性进行重置，但仅适用于非必选项。

语法格式

```
ALTER BLOCK RULE block_name
   [ [ TO user_name@'host' ] | [ TO user_name ] | [ TO 'host' ] | [ TO DEFAULT ] ] |
   [ FOR UPDATE | SELECT | INSERT | DELETE | MERGE | DEFAULT ] |
   FILTER BY
   { SQL ( 'text' ) | TEMPLATE ( template_parameter = value ) }
   [ WITH ( { with_parameter = value }, [, ... ] ) ];
ALTER BLOCK RULE block_name RENAME to new_block_name;
```

参数说明

表 13-160 ALTER BLOCK RULE 参数说明

参数	描述	取值范围
block_name	需要修改属性的查询过滤规则名称。	字符串，需符合**标识符命名规范**。

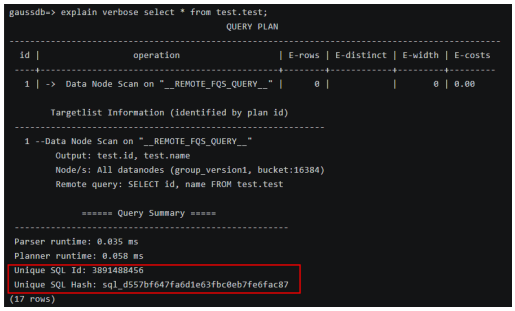
版权所有 © 华为云计算技术有限公司

参数	描述	取值范围

TO	对查询过滤规则生效的对象或客户端IP，支持以下方式：
	• **TO user_name@'host'**：用户名以及客户端IP，两者都要满足，查询过滤规则才能生效。
	• **TO user_name**：用户名，即只判定满足用户名条件，查询过滤规则就生效，不区分客户端IP。
	• **TO 'host'**：客户端IP，即只判定满足客户端IP条件，查询过滤规则就生效，不区分用户。
		举例：查询过滤规则对来自客户端IP为192.168.0.59且登录用户名为user1时生效。
		`ALTER BLOCK RULE query_block TO user1@'192.168.0.59' FOR SELECT  FILTER BY SQL('test.test');`
		查询过滤规则对登录用户名为user1时生效（无论客户端IP是哪个）。
		`ALTER BLOCK RULE query_block TO user1 FOR SELECT  FILTER BY SQL('test.test');`
		查询过滤规则对来自客户端IP 192.168.0.59的访问时都生效（无论哪个用户登录）。
		`ALTER BLOCK RULE query_block TO '192.168.0.59' FOR SELECT  FILTER BY SQL('test.test');`

FOR	语句类型，支持对UPDATE/SELECT/INSERT/DELETE/MERGE INTO五种类型语句进行限制。	-

版权所有 © 华为云计算技术有限公司

参数	描述	取值范围
FILTER BY	过滤方法，支持两种形式： • **SQL**：根据SQL语句内容进行过滤，通过关键词对语句进行正则匹配，例如表名，其长度不能超过1024，建议尽量精简。 • **TEMPLATE**：根据模板进行过滤，**此方法不区分用户，针对所有用户执行的特定语句都实施拦截，拦截效果更高效**，取值如下： – **unique_sql_id**：归一化的64位哈希值，重复概率较sql_hash大一些。 – **sql_hash**：归一化的哈希值（md5），一般不会重复，相较unique_sql_id更推荐使用。	unique_sql_id和sql_hash取值，可以通过EXPLAIN VERBOSE查询，例如： EXPLAIN VERBOSE SELECT * FROM test.test; 
with_parameter	查询过滤规则选项参数。	• application_name，客户端名称。 • query_band，负载标识。 • table_num，语句扫描的基表个数。 • partition_num，算子预估扫描的最大分区数。 • estimate_row，算子预估扫描的最大表行数。 • resource_pool，需要切换到的资源池名称。 • max_active_num，规则对应语句的最大并发数。 • is_warning，控制拦截到语句的行为是告警还是报错，on表示告警，off为报错，默认为off。

示例

创建查询规则query_block。

 版权所有 © 华为云计算技术有限公司

```
CREATE BLOCK RULE query_block FILTER BY SQL('update table_table set a=1');
```

创建一个测试用户。

```
DROP USER IF EXISTS user12 CASCADE;
CREATE USER user12 PASSWORD '{Password}';
```

修改查询规则query_block的属性。

```
ALTER BLOCK RULE query_block
TO user12@'192.168.0.59'
FOR SELECT
FILTER BY
SQL('select * from table_name')
WITH (application_name='gsql',
query_band='test1',
table_num='2',
partition_num='3',
estimate_row='1000',
resource_pool='rsp1',
max_active_num='3',
is_warning='off'
);
```

通过视图**pg_get_blockruledef**查询修改后的过滤规则。

```
SELECT * FROM pg_get_blockruledef('query_block');
```



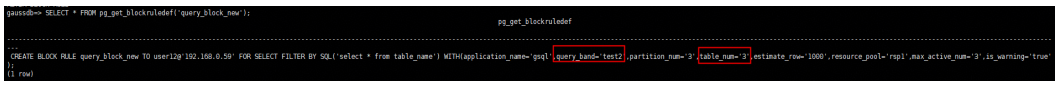
对查询规则query_block进行重命名。

```
ALTER BLOCK RULE query_block RENAME TO query_block_new;
```

注意，修改规则时，如果只重设置了部分参数，则原来的其他参数还保留原值，**即设置哪个参数，只修改哪个参数，不是覆盖清空全部参数。**

```
ALTER BLOCK RULE query_block_new
TO user12@'192.168.0.59'
FOR SELECT
FILTER BY
SQL('select * from table_name')
WITH (
query_band='test2',
table_num='3');

SELECT * FROM pg_get_blockruledef('query_block_new');
```

返回结果显示：只是将query_band、table_num的取值修改，其他参数项仍存在。



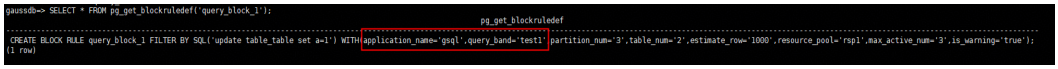
同样，**新增过滤条件，也不会将之前的条件清除**，例如：

```
CREATE BLOCK RULE query_block_1 FILTER BY SQL('update table_table set a=1')
WITH (application_name='gsql',
query_band='test1');

ALTER BLOCK RULE query_block_1 FILTER BY SQL('update table_table set a=1')
WITH (
table_num='2',
partition_num='3',
estimate_row='1000',
resource_pool='rsp1',
max_active_num='3',
is_warning='off'
```

 版权所有 © 华为云计算技术有限公司

```
);
SELECT * FROM pg_get_blockruledef('query_block_1');
```

返回结果显示：原来的application_name、query_band参数仍保留。



相关链接

CREATE BLOCK RULE，**DROP BLOCK RULE**

13.9.9 DROP BLOCK RULE

功能描述

该语法用于删除查询过滤规则。**查询过滤规则**，包括查询过滤规则名称、绑定的客户端名称、客户端IP、用户以及匹配方式等。在日常数据库操作中，异常SQL的执行可能会导致资源过度消耗，影响数据库性能和稳定性，通过创建查询过滤规则，可以将此类SQL语句进行拦截。

注意事项

- 该语法仅9.1.0.100及以上集群版本支持。
- 只有拥有数据库所有者权限的用户或者授予gs_role_block角色权限的用户才能执行DROP BLOCK RULE命令，系统管理员默认拥有此权限。
- 普通用户无法执行CREATE BLOCK RULE，可通过系统管理员进行授权，授权语法：
  ```
  GRANT gs_role_block TO user;
  ```
  建议**查询过滤规则操作尽量缩小适用范围，避免误过滤，或者范围过大导致性能劣化**。

语法格式

```
DROP BLOCK RULE [ IF EXISTS ] block_name;
```

参数说明

表 13-161 DROP BLOCK RULE 参数说明

参数	描述	取值范围
IF EXISTS	如果指定的查询过滤规则不存在时返回一个提示而不是错误。	-
block_name	需要删除的查询过滤规则名称。	已存在的查询过滤规则名称。

示例

创建名称为query_block的查询过滤规则。

 版权所有 © 华为云计算技术有限公司

```
DROP BLOCK RULE IF EXISTS query_block;
CREATE BLOCK RULE query_block FILTER BY SQL('update table_table set a=1');
```

删除名称为query_block的查询过滤规则。

```
DROP BLOCK RULE query_block;
DROP BLOCK RULE IF EXISTS query_block;
```

相关链接

CREATE BLOCK RULE，**ALTER BLOCK RULE**

13.10 性能和执行计划

13.10.1 CREATE STATISTICS

功能描述

在当前数据库中为表创建一个扩展统计信息对象，该对象由命令的执行者所有。该语法仅8.2.1.200及以上集群版本支持（8.3.0集群版本不支持）。

注意事项

- 目前该命令仅支持创建表达式统计信息。
- 不支持对临时表创建表达式统计信息。
- 不支持在表的分区上创建表达式统计信息。
- 不支持对包含聚集函数、子查询、窗口函数、返回集合类型的表达式创建统计信息。

语法格式

```
CREATE STATISTICS [ IF NOT EXISTS ] statistics_name
    ON ( expression ) [, ...]
    FROM table_name
```

参数说明

表 13-162 CREATE STATISTICS 参数说明

参数	描述	取值范围
statistics_name	新统计信息对象的名称。	取值范围：字符串，要符合标识符的命名规范。
expression	要收集统计信息的表达式，支持单个或多个表达式，多个表达式通过逗号分隔。	-
table_name	用于收集统计信息的表名。	-

版权所有 © 华为云计算技术有限公司

示例

创建示例表并插入数据。

```
CREATE TABLE t1_row(id int, dt timestamp) WITH (orientation = row) DISTRIBUTE BY HASH (id);
INSERT INTO t1_row values (generate_series(1, 721), generate_series('2023-01-01'::timestamp,
'2023-01-31'::timestamp, '1 hour'::interval));
```

为表创建一个表达式统计信息对象，使用ANALYZE命令收集统计信息。

```
CREATE STATISTICS s1_t1_row ON EXTRACT(HOUR FROM dt), EXTRACT(DAY FROM dt) FROM t1_row;
ANALYZE t1_row;
```

查看语句的查询计划。

```
SET enable_fast_query_shipping = OFF;
EXPLAIN VERBOSE SELECT * FROM t1_row WHERE EXTRACT(HOUR FROM dt) BETWEEN 0 AND 12;
```

相关链接

DROP STATISTICS

13.10.2 DROP STATISTICS

功能描述

删除一个扩展统计信息对象。该语法仅8.2.1.200及以上集群版本支持（8.3.0集群版本不支持）。

注意事项

无

语法格式

```
DROP STATISTICS [ IF EXISTS ] name [, ...] [ CASCADE | RESTRICT ]
```

参数说明

表 13-163 DROP STATISTICS 参数说明

参数	描述	取值范围
IF EXISTS	如果指定的扩展统计信息对象不存在，不会抛出一个错误，而会发出一个通知，告知对象不存在。	-
name	要删除的扩展统计信息对象名。	已存在的扩展统计信息对象名。
CASCADE | RESTRICT	该参数不起作用，没有对象依赖于扩展统计信息对象。	-

示例

删除一个扩展统计信息对象：

 版权所有 © 华为云计算技术有限公司

```
DROP STATISTICS s1_t1_row;
```

相关链接

CREATE STATISTICS

13.10.3 CREATE OUTLINE

功能描述

为指定的sql_hash创建一个新的outline。

注意事项

- 如果outline_text中的outline存在语法错误，创建报错syntax error。
- 执行成功后，会在SQL_OUTLINE系统表中插入一行。
- 该语法仅数据库管理员用户，或授予了gs_role_sql_management角色权限的用户可执行。

语法格式

```
CREATE OUTLINE outline_name FOR sql_hash USING outline_text;
```

参数说明

- **outline_name**
  指定新建outline的名字。
- **sql_hash**
  新建outline对应的sql的sql_hash。
- **outline_text**
  新建outline的具体内容。text类型，需要加单引号。

  说明

  只支持把所有的hint写在一个hint注释中。如果有多个hint注释块，需合并为一个。

示例

数据库管理员用户创建outline：

```
CREATE OUTLINE outline_test1 FOR sql_be2995e824abb8b712b69fca4825b532 USING '/*+ leading(@sel$1
public.t1@sel$1 s1.t2@sel$1) */';
CREATE OUTLINE outline_test2 FOR sql_be2995e824abb8b712b69fca4825b532 USING
'/*+
    begin_outline_data
     leading(@sel$1 public.t1@sel$1 s1.t2@sel$1)
    end_outline_data
 */';
```

数据库管理员用户创建outline时，outline_text指定多个hint注释块，报错syntax error；合并为一个hint注释块后，创建成功：

```
CREATE OUTLINE outline_test3 FOR sql_be2995e824abb8b712b69fca4825b532 USING '/*+ leading(@sel$1
public.t1@sel$1 s1.t2@sel$1) */ /*+ mergejoin(@sel$1 t1@sel$1 t1@sel$2)*/';
ERROR:  LINE 1: syntax error at '*'
```

 版权所有 © 华为云计算技术有限公司

```
CREATE OUTLINE outline_test3 FOR sql_be2995e824abb8b712b69fca4825b532 USING '/*+ leading(@sel$1 public.t1@sel$1 s1.t2@sel$1) */' '/*+ mergejoin(@sel$1 t1@sel$1 t1@sel$2)*/';
ERROR:  syntax error at or near "'/*+ mergejoin(@sel$1 t1@sel$1 t1@sel$2)*/'"
LINE 1: ...+ leading(@sel$1 public.t1@sel$1 s1.t2@sel$1) */' '/*+ merge...
                                                             ^
CREATE OUTLINE outline_test3 FOR sql_be2995e824abb8b712b69fca4825b532 USING '/*+ leading(@sel$1 public.t1@sel$1 s1.t2@sel$1) mergejoin(@sel$1 t1@sel$1 t1@sel$2)*/';
```

普通用户创建outline报错：

```
CREATE OUTLINE outline_test4 FOR sql_be2995e824abb8b712b69fca4825b532 USING '/*+ leading(@sel$1 public.t1@sel$1 s1.t2@sel$1) */';
ERROR:  Permission denied to create outline.
```

13.10.4 DROP OUTLINE

功能描述

删除一个outline。

注意事项

- 不能删除被绑定的outline。
- 该语法仅数据库管理员用户，或授予了gs_role_sql_management角色权限的用户可执行。

语法格式

```
DROP OUTLINE outline_name FOR sql_hash;
```

参数说明

表 13-164 DROP OUTLINE 参数说明

参数	描述	取值范围
outline_name	要删除的outline的名字。	-
sql_hash	删除该sql_hash对应的outline。	-

示例

删除一个未绑定的outline。

```
DROP OUTLINE outline_test3 FOR sql_be2995e824abb8b712b69fca4825b532;
```

删除一个绑定的outline，报错。

```
SELECT sql_hash, outline_name FROM sql_plan_baseline;
               sql_hash               |               outline_name
--------------------------------------+------------------------------------------
 sql_3d937542c0d24c9c995f31cba327fa56 | outline_413ac0e566ff20be320f5ef0774bf1eb
 sql_6e7be6e92878d67d30b602d16b779a99 | outline_62bcfca3d090e777d7a6e6a88a49e53f
(2 rows)

DROP OUTLINE outline_413ac0e566ff20be320f5ef0774bf1eb FOR sql_3d937542c0d24c9c995f31cba327fa56;
```

 版权所有 © 华为云计算技术有限公司

```
ERROR:  The outline(specified by sql_hash[sql_3d937542c0d24c9c995f31cba327fa56] and
outline_name[outline_413ac0e566ff20be320f5ef0774bf1eb]) is bind
```

13.11 连接&会话&资源清理

13.11.1 CLEAN CONNECTION

功能描述

用于清理数据库连接。允许在指定节点上清理指定数据库的指定用户的相关连接。

主要用于处理数据库异常、清理空闲/无效连接以及用于DROP DATABASE之前的连接访问检查，防止操作失败。

注意事项

参数**dbname和username**必须至少指定一个。

语法格式

```
CLEAN CONNECTION
   TO { COORDINATOR ( nodename [, ... ] ) | NODE ( nodename [, ... ] )| ALL [ CHECK ] [ FORCE ] }
   [ FOR DATABASE dbname ]
   [ TO USER username ];
```

参数说明

表 13-165 CLEAN CONNECTION 参数说明

参数	描述	取值范围
CHECK	仅在节点列表为TO ALL时可以指定。 如果指定该参数，会在清理连接之前检查数据库是否被其他会话连接访问。 此参数主要用于DROP DATABASE之前的连接访问检查，如果发现有其他会话连接，则将报错并停止删除数据库。	-
FORCE	仅在节点列表为TO ALL时可以指定。 如果指定该参数，所有和指定dbname和username相关的线程都会收到SIGTERM信号，然后被强制关闭。	-

 版权所有 © 华为云计算技术有限公司

参数	描述	取值范围
COORDINATOR (nodename [, ...]) | NODE (nodename [, ...]) | ALL	删除指定节点上的连接。有三种场景： 1. 删除指定CN上的连接。 2. 删除指定DN上的连接。 3. 删除所有节点上的连接，包括CN和DN。	可替换其中的nodename为已存在的节点名。
dbname	删除指定数据库上的连接。 如果不指定，则删除所有数据库的连接。	有效的数据库名。
username	删除指定用户上的连接。 如果不指定，则删除所有用户的连接。	有效的用户名。

示例

删除数据库template1在dn1和dn2节点上的连接。

```
CLEAN CONNECTION TO NODE (dn1,dn2) FOR DATABASE template1;
```

删除用户jack在dn1节点上的连接。

```
CLEAN CONNECTION TO NODE (dn1) TO USER jack;
```

删除在数据库gaussdb上的所有连接。

```
CLEAN CONNECTION TO ALL FORCE FOR DATABASE gaussdb;
```

常见问题

删除数据库失败，提示有session正在连接

13.11.2 ALTER SYSTEM KILL SESSION

功能描述

在DWS数据库的日常运维中，管理员可能会遇到某些会话长时间运行、占用大量资源或处于异常状态（如死锁、无限等待）的情况。这些异常会话会影响数据库的整体性能，甚至阻塞其他正常业务。此时，可使用ALTER SYSTEM KILL SESSION命令结束该会话以释放资源，恢复系统正常运行。

ALTER SYSTEM KILL SESSION命令用于结束一个指定的数据库会话。通过提供会话的SID和SERIAL标识，可以精确终止目标会话，释放其占用的系统资源。

注意事项

无。

版权所有 © 华为云计算技术有限公司

语法格式

```
ALTER SYSTEM KILL SESSION 'session_sid, serial' [ IMMEDIATE ];
```

参数说明

表 13-166 ALTER SYSTEM KILL SESSION 参数说明

参数	描述	取值范围
session_sid, serial	指定要结束的会话标识，由会话SID和SERIAL组成，以逗号分隔。	通过查看系统表V$SESSION可查看所有会话的SID和SERIAL。
IMMEDIATE	表示会话将在命令执行后立即结束。	-

示例

查询会话信息。

```
SELECT sid,serial#,username FROM V$SESSION;

       sid        | serial# | username
------------------+---------+----------
 140131075880720 |       0 |
 140131025549072 |       0 |
 140131073779472 |       0 |
 140131071678224 |       0 |
 140131125774096 |       0 |
 140131127875344 |       0 |
 140131113629456 |       0 |
 140131094742800 |       0 |
(8 rows)
```

结束SID为140131075880720的会话。

```
ALTER SYSTEM KILL SESSION '140131075880720,0' IMMEDIATE;
```

13.11.3 ALTER SESSION

功能描述

ALTER SESSION命令用于定义或修改那些对当前会话有影响的条件或参数。修改后的会话参数会一直保持，直到断开当前会话。

注意事项

- 如果执行SET TRANSACTION之前没有执行START TRANSACTION，则事务立即结束，命令无法显示效果。
- 可以用START TRANSACTION里面声明所需要的transaction_mode(s)的方法来避免使用SET TRANSACTION。

语法格式

- 设置会话的事务参数。

 版权所有 © 华为云计算技术有限公司

```
ALTER SESSION SET [ SESSION CHARACTERISTICS AS ] TRANSACTION
   { ISOLATION LEVEL { READ COMMITTED | READ UNCOMMITTED } | { READ ONLY  | READ
WRITE } } [, ...] ;
```

- 设置会话的其他运行时参数。

```
ALTER SESSION SET
   {{config_parameter { { TO  | =  }  { value | DEFAULT }
     | FROM CURRENT }} | CURRENT_SCHEMA [ TO | = ] { schema | DEFAULT }
     | TIME ZONE time_zone
     | SCHEMA schema
     | NAMES encoding_name
     | ROLE role_name PASSWORD 'password'
     | SESSION AUTHORIZATION { role_name PASSWORD 'password' | DEFAULT }
     | XML OPTION { DOCUMENT | CONTENT }
   } ;
```

参数说明

表 13-167 ALTER SESSION 参数说明

参数	描述	取值范围
SESSION	声明的参数只对当前会话起作用。如果SESSION和LOCAL都没出现，则SESSION为缺省值。 如果在事务中执行了此命令，命令的产生影响将在事务回滚之后消失。如果该事务已提交，影响将持续到会话的结束，除非被另外一个SET命令重置参数。	-
config_parameter	可设置的运行时参数的名称。可用的运行时参数可以使用SHOW ALL命令查看。 **说明** 部分通过SHOW ALL查看的参数不能通过SET设置。如max_datanodes。	-
value	config_parameter的新值。	可以声明为字符串常量、标识符、数字，或者逗号分隔的列表。DEFAULT用于把这些参数设置为它们的缺省值。
TIME ZONE timezone	用于指定当前会话的本地时区。	有效的本地时区。该选项对应的运行时参数名称为TimeZone，DEFAULT缺省值为PRC。
CURRENT_SCHEMA	CURRENT_SCHEMA用于指定当前的模式。	已存在的模式名称。
SCHEMA schema	同CURRENT_SCHEMA。此处的schema是个字符串。 例如：set schema 'public';	-

版权所有 © 华为云计算技术有限公司

参数	描述	取值范围
NAMES encoding_name	用于设置客户端的字符编码。等价于set client_encoding to encoding_name。	有效的字符编码。该选项对应的运行时参数名称为client_encoding，默认编码为UTF8。
XML OPTION option	用于设置XML的解析方式。	• CONTENT（缺省） • DOCUMENT

示例

创建模式ds：

```
CREATE SCHEMA ds;
```

设置模式搜索路径：

```
SET SEARCH_PATH TO ds, public;
```

设置日期时间风格为传统的POSTGRES风格（日在月前）：

```
SET DATESTYLE TO postgres, dmy;
```

设置当前会话的字符编码为UTF8：

```
ALTER SESSION SET NAMES 'UTF8';
```

设置时区为加州伯克利：

```
SET TIME ZONE 'PST8PDT';
```

设置时区为意大利：

```
SET TIME ZONE 'Europe/Rome';
```

设置当前模式：

```
ALTER SESSION SET CURRENT_SCHEMA TO tpcds;
```

设置XML OPTION为DOCUMENT：

```
ALTER SESSION SET XML OPTION DOCUMENT;
```

创建角色joe，并设置会话的角色为joe：

```
CREATE ROLE joe WITH PASSWORD '{password}';
ALTER SESSION SET SESSION AUTHORIZATION joe PASSWORD '{password}';
```

切换到默认用户：

```
ALTER SESSION SET SESSION AUTHORIZATION default;
```

相关链接

SET

 版权所有 © 华为云计算技术有限公司

13.11.4 DISCARD

功能描述

释放与数据库会话相关的内部资源。

DISCARD命令用于重设部分或全部的会话的状态，不同的DISCARD子命令释放不同类型的资源；DISCARD ALL释放所有与当前会话相关的临时资源，并重置到其初始状态。

注意事项

- 执行DISCARD VOLATILE { TEMPORARY | TEMP }操作后，当前session内所有volatile临时表资源都会被清理，不支持清理单个volatile临时表资源。
- 如果某张全局临时表在某个会话占用资源，需要DISCARD依次清理所有会话的资源后，才能执行DDL操作。
- 执行DISCARD ALL成功之后，以pg_temp和pg_toast_temp开头的schema也会被删除。
- DISCARD ALL不允许在事务中执行。

语法格式

```
DISCARD {{ GLOBAL { TEMPORARY | TEMP } [ TABLE table_name ] } |
      { VOLATILE { TEMPORARY | TEMP } }  |
      { ALL | TEMP | TEMPORARY | PLANS | SEQUENCES }}
```

参数说明

表 13-168 DISCARD 参数说明

参数	描述
VOLATILE { TEMPORARY | TEMP }	表示释放当前会话中VOLATILE临时表相关资源。
GLOBAL { TEMPORARY | TEMP } [TABLE table_name]	• DISCARD GLOBAL TEMP，释放当前会话中GLOBAL临时表相关资源。 • DISCARD GLOBAL TEMP TABLE table_name 释放当前会话中指定某张GLOBAL临时表的相关资源。
TEMP | TEMPORARY	释放当前会话中所有临时表的相关资源，包括VOLATILE临时表和GLOBAL临时表。
PLANS	释放当前会话中所有缓存的查询计划，并在下次使用相关prepare语句时强制进行重新计划。
SEQUENCES	丢弃缓存的所有序列相关的状态，包括currval()/lastval()信息和尚未由nextval()返回的任何预分配的序列值。

版权所有 © 华为云计算技术有限公司

参数	描述
ALL	释放所有与当前会话相关的临时资源，并将会话重置为其初始状态，这与执行以下语句序列具有几乎相同的效果： SET SESSION AUTHORIZATION DEFAULT; RESET ALL; DEALLOCATE ALL; CLOSE ALL; UNLISTEN *; SELECT pg_advisory_unlock_all(); DISCARD PLANS; DISCARD SEQUENCES; DISCARD TEMP;

示例

DISCARD全局临时表。

```
DROP TABLE IF EXISTS t_global_temp;
CREATE GLOBAL TEMP TABLE t_global_temp(a int,b int);
INSERT INTO t_global_temp VALUES(1,1),(2,2);
DROP TABLE t_global_temp;
```

```
gaussdb=> DROP TABLE t_global_temp;
ERROR:  can not DROP TABLE when global temp table "t_global_temp" is in use
```

```
DISCARD GLOBAL TEMP TABLE t_global_temp;
DROP TABLE t_global_temp;
```

```
gaussdb=> DISCARD GLOBAL TEMP TABLE t_global_temp;
DISCARD GLOBAL TEMP
gaussdb=> DROP TABLE t_global_temp;
DROP TABLE
```

DISCARD VOLATILE清理当前会话中所有volatile临时表相关资源。

```
DROP TABLE IF EXISTS TX1;
DROP TABLE IF EXISTS TX2;
CREATE VOLATILE TEMP TABLE TX1(A INT) DISTRIBUTE BY HASH(A);
CREATE VOLATILE TEMP TABLE TX2(A INT) DISTRIBUTE BY HASH(A);

SELECT * FROM TX1;
SELECT * FROM TX2;          ^
```

```
gaussdb=> SELECT * FROM TX1;
 a
---
(0 rows)

gaussdb=> SELECT * FROM TX2;
 a
---
(0 rows)
```

```
DISCARD VOLATILE TEMP;
SELECT * FROM TX1;
```

 版权所有 © 华为云计算技术有限公司

```
gaussdb=> DISCARD VOLATILE TEMP;
DISCARD VOLATILE TEMP
gaussdb=> SELECT * FROM TX1;
ERROR:  relation "tx1" does not exist
LINE 1: SELECT * FROM TX1;
                      ^
```

```
SELECT * FROM TX2;
```

```
gaussdb=> SELECT * FROM TX2;
ERROR:  relation "tx2" does not exist
LINE 1: SELECT * FROM TX2;
                      ^
```

DISCARD TEMP操作清理当前会话中所有临时表相关资源。

```
DROP TABLE IF EXISTS t_global_temp;
CREATE GLOBAL TEMP TABLE t_global_temp(a int,b int);
INSERT INTO t_global_temp VALUES(1,1),(2,2);
DROP TABLE IF EXISTS t_volatile_temp;
DROP TABLE IF EXISTS t_temp;
CREATE VOLATILE TEMP TABLE t_volatile_temp(a int,b int);
CREATE TEMP TABLE t_temp(a int,b int);

DISCARD TEMP;
SELECT * FROM t_global_temp;
```

```
gaussdb=> DISCARD TEMP;
DISCARD TEMP
gaussdb=> SELECT * FROM t_global_temp;
 a | b
---+---
(0 rows)
```

```
SELECT * FROM t_volatile_temp;
```

```
gaussdb=> SELECT * FROM t_volatile_temp;
ERROR:  relation "t_volatile_temp" does not exist
LINE 1: SELECT * FROM t_volatile_temp;
                      ^
```

```
SELECT * FROM t_temp;
```

```
gaussdb=> SELECT * FROM t_temp;
ERROR:  relation "t_temp" does not exist
LINE 1: SELECT * FROM t_temp;
                      ^
```

13.11.5 TRUNCATE

功能描述

TRUNCATE语法用于清理表数据，快速地从表中删除所有行。

它与在目标表上进行无条件的DELETE的效果相同，但由于TRUNCATE实际上并不扫描表，因此速度更快，尤其在大表上操作效果更明显。

 版权所有 © 华为云计算技术有限公司

TRUNCATE TABLE、DELETE TABLE 和 DROP TABLE 差异

表 13-169 TRUNCATE TABLE、DELETE TABLE 和 DROP TABLE 差异

差异类别	TRUNCATE TABLE	DELETE TABLE	DROP TABLE
语法	DDL	DML	DDL
删除内容	删除表中的所有数据，但不会删除表结构	删除内容，不删除定义	删除内容和定义
是否释放空间	释放空间	不释放空间	释放空间
执行速度	速度较快 释放存储表数据所用数据页来删除数据，并且只在事务日志中记录页的释放，使用系统和事务日志资源少。	速度慢 每删除一行，在事务日志中为所删除每行记录一项。	速度最快
适用场景	如果需要快速删除整个表中的数据，保留表结构，且是大批量数据的场景。	需要根据特定条件删除数据且数据量可控的情况。	需要删除整个表，包括表结构和数据。

注意事项

- 请谨慎执行TRUNCATE TABLE，确认表数据可以删除或者已有备份后，再执行删除操作。TRUNCATE TABLE命令删除表数据后，数据将无法恢复。
- TRUNCATE全局临时表只对当前会话的数据进行TRUNCATE。不影响其他会话的数据。
- 存算分离架构下，不支持同时TRUNCATE普通表和临时表。

警告

- 避免在业务高峰期执行ALTER TABLE/ALTER TABLE PARTITION（增删改查、DROP PARTITION）、TRUNCATE操作，避免有长SQL阻塞ALTER、TRUNCATE操作或SQL业务被ALTER、TRUNCATE阻塞。
- 更多开发设计规范参见**总体开发设计规范**。

语法格式

清理表数据。

```
TRUNCATE [ TABLE ] [ ONLY ] {[[database_name.]schema_name.]table_name [ * ]} [, ... ]
    [ CONTINUE IDENTITY ] [ CASCADE | RESTRICT ] [ PURGE ] [KEEP ttl HOUR];
```


版权所有 © 华为云计算技术有限公司

参数说明

表 13-170 TRUNCATE 参数说明

参数	描述	取值范围
ONLY	指定清空表的范围。 该参数仅为了兼容PG保留，DWS不支持继承表。	• 如果指定ONLY，则只有指定表会被清空。 • 如果没有指定ONLY，则指定表及其所有继承表（若有）都会被清空。
database_name	要清空的表所在的数据库名称。	已存在的数据库名。
schema_name	要清空的表的模式名称。	已存在的模式名。
table_name	指定要清空的表名（可以用模式修饰）。	已存在的表名。
CONTINUE IDENTITY	不更改序列的值。	该参数为缺省值。
CASCADE | RESTRICT	可选参数，清空有外键引用的表的方式。	• CASCADE：表示级联清空所有在该表上有外键引用的表，或者由于CASCADE而被添加到组中的表。 • RESTRICT：表示如果其他表在该表上有外键引用，则拒绝清空。 缺省值为RESTRICT。
PURGE	指定PURGE参数时，无论回收站特性是否开启，对象都会直接删除，不放入回收站。PURGE与KEEP ttl HOUR参数不能同时指定。 该参数仅9.1.0.200及以上集群版本支持。	-
KEEP ttl HOUR	V3表开启回收站特性（GUC参数enable_recyclebin设置为on）并设置TTL后，物理文件会放入回收站中，在TTL到期前可从回收站中恢复；TTL到期后，后台自动清理回收站中的对象。 该参数仅9.1.0.200及以上集群版本支持。	• ttl取值需大于0且最大为360小时（15天）。 • ttl为0时，与不设置该参数效果一致。

 版权所有 © 华为云计算技术有限公司

示例

1. 创建示例表test_t并插入数据。

```
DROP TABLE IF EXISTS test_t;
CREATE TABLE test_t
(
    col_id INT PRIMARY KEY,
    col_number INT NOT NULL,
    col_date DATE NOT NULL,
    col_price NUMERIC (10,2),
    col_status TEXT
)
WITH (ORIENTATION = COLUMN)
DISTRIBUTE BY HASH (col_id);

INSERT INTO test_t VALUES
(01,300,'2025-03-01',99.20,'sold'),
(02,400,'2025-04-01',95.50,'sold'),
(03,450,'2025-05-01',420.50,'sold'),
(04,100,'2025-06-01',100.85,'restock');
```

2. 查看表test_t的数据。

```
SELECT * FROM test_t;
```

```
gaussdb=> SELECT * FROM test_t;
 col_id | col_number |      col_date       | col_price | col_status
--------+------------+---------------------+-----------+------------
      3 |        450 | 2025-05-01 00:00:00 |    420.50 | sold
      4 |        100 | 2025-06-01 00:00:00 |    100.85 | restock
      2 |        400 | 2025-04-01 00:00:00 |     95.50 | sold
      1 |        300 | 2025-03-01 00:00:00 |     99.20 | sold
(4 rows)
```

3. 清理表test_t。

```
TRUNCATE test_t;
```

返回结果如下，表示已成功清理。

```
gaussdb=> TRUNCATE test_t;
TRUNCATE TABLE
```

4. 查看表定义。TRUNCATE TABLE操作清空表，保留表定义。

```
SELECT * FROM pg_get_tabledef('public.test_t');
```

```
gaussdb=> SELECT * FROM pg_get_tabledef('public.test_t');
                                                
```

VACUUM 与 VACUUM FULL 的区别

VACUUM命令存在两种形式，VACUUM和VACUUM FULL，VACUUM命令做的是LAZY VACUUM。从字面意思就可以看出来，LAZY VACUUM是VACUUM FULL的简化版，具体区别见下表。

表 13-171 VACUUM 与 VACUUM FULL 的区别

比较维度	VACUUM	VACUUM FULL
空间清理	如果删除记录位于表的末端，其所占用的空间将会被物理释放并归还操作系统。如果删除记录不是末端数据，会将表中或索引中dead tuple所占用的空间置为可用状态，从而复用这些空间。	不论被清理的数据处于何处，这些数据所占用的空间都将被物理释放并归还于操作系统。当再有数据插入后，系统会分配新的磁盘页面使用。
锁类型	4级锁，共享更新排他锁（ShareUpdateExclusiveLock），可以与其他操作并行。	8级锁，访问排他锁（AccessExclusiveLock），执行期间会同时锁表，基于该表的其他操作全部挂起。
物理空间	不会释放。	会释放。
事务ID	不回收。	回收
执行开销	开销较小，可以定期执行。	开销巨大，建议确认数据库所占磁盘页面空间接近临界值再执行操作，建议选择业务量较少的时段完成。
执行效果	执行后会有所提升。	执行完后，基于该表的操作效率大大提升。

关于自动清理（AUTOVACUUM）

由于VACUUM操作可释放空间，对提升查询性能至关重要，DWS提供了自动清理的能力，用户可根据自身业务需要，在指定条件下，让系统触发自动清理功能，通过管理控制台配置，参见**管理运维计划**。

在线 VACUUM FULL

DWS的9.1.1.300及以上集群版本中支持在线VACUUM FULL（对应语法**VACUUM FULL ONLINE**），该功能可以在不长时间加排他锁的情况下，允许对表的并发读写，从而减少对业务的影响。因此，当数据库管理员在业务低峰期发现业务表脏页率较高时，可以执行在线VACUUM FULL操作，以回收表空间并优化数据库性能，同时确保业务的连续性和数据库的可用性。

在线VACUUM FULL功能约束限制与注意事项如下：

 版权所有 © 华为云计算技术有限公司

- 目前支持行存表、V2列存表（含hstore/hstore_opt），V3列存表（含hstore/hstore_opt）以及LSM、行列混存表。不支持临时表、系统表、外表、复制表、物化视图、binlog、冷热表及大宽表。
- 用户需要同时具有目标表的VACUUM权限和ALTER权限，才能执行在线VACUUM FULL。
- 在线VACUUM FULL操作执行失败可能会导致数据残留。需手动执行以下命令并删除残留的辅助SCHEMA（以data_redis_ovf_<table_oid>命名）。
  ```
  ALTER TABLE [ table_name ] SET (append_mode=off);
  ```
- 处于重分布状态和扩缩容状态的集群，不能执行在线VACUUM FULL。
- 2.0集群升级3.0集群期间不能执行在线VACUUM FULL。
- 在线VACUUM FULL支持资源管控：
  - 在线VACUUM FULL会启动gs_redis子进程，gs_redis子进程以dbadmin连接数据库后，通过SET session_respool切换到调用用户绑定的资源池。若用户未绑定资源池，则使用dbadmin默认资源池。
  - 支持通过WITH子句中的参数控制I/O速率等，详见**表13-173**。
- 在线VACUUM FULL操作针对大表（例如含有索引和分区的2.0表）性能与VACUUM FULL持平。但对于小表性能劣化明显，主要限制为进程启动和状态检查，耗时数秒。
- 在线VACUUM FULL操作完成的最后阶段会阻塞写操作，在小时级的VACUUM FULL操作中，阻塞时间在分钟级。在最后瞬间会对读操作产生毫秒级的短暂阻塞。
- 建议业务低峰期执行，因为执行在线VACUUM FULL会降低DDL/DML操作的性能，且最后阻塞写操作的时长与业务数据量相关。
- 9.1.2及以上集群版本中支持分区级在线VACUUM FULL，即支持对指定分区表的单个或多个分区执行在线VACUUM FULL，仅对目标分区进行数据迁移和空间回收，对非目标分区不进行清理。
  - 单分区锁优化：在线VACUUM FULL指定单个分区时，主表锁降为3级（RowExclusiveLock），非目标分区可正常执行DML操作；指定多个分区时，主表保持7/8级锁以防止死锁。
  - 分区级在线VACUUM FULL仅创建包含目标分区的影子表，缓解因分区数过多导致的建表耗时长和长时间持锁问题。
  - 针对分区级在线VACUUM FULL，建议单次处理的分区数限制在100个以内。由于需要过滤分区，如果指定分区数过多，会导致性能可能与全表在线VACUUM FULL相比劣化。
- 表级在线VACUUM FULL支持使用bucket_nums参数来修改表的数据分布，分区级在线VACUUM FULL不支持修改表的数据分布。
- 在线VACUUM FULL参数enable_drop_column_optimize建议开启（参数默认开启）。如果手动关闭了该参数（设置为off），则每次对行存表执行在线VACUUM FULL时都会在原表上累积4个dropped列，导致pg_attribute膨胀；列存表不受影响。
- 在线Vacuum Full执行期间会为目标表设置append_mode=on并创建辅助schema data_redis_ovf_<rel_cn_oid>。如果在线Vacuum Full中途异常退出，对应两个残留不会被自动清理，导致表处于异常状态。可使用在线Vacuum Full辅助函数**gs_scan_for_online_vacuum_full()**、**gs_cleanup_for_online_vacuum_full(target_tables text[])**及**gs_cleanup_for_online_vacuum_full(target_table text)**扫描和清理这类残留。

 版权所有 © 华为云计算技术有限公司

注意事项

- 如果没有参数，VACUUM处理当前数据库里用户拥有相应权限的每个表。如果参数指定了一个表，VACUUM只处理指定的那个表。
- 要对一个表进行VACUUM操作，通常用户必须是表的所有者，被授予了指定表VACUUM权限的用户或者被授予了gs_role_vacuum_any角色的用户，系统管理员默认拥有此权限。数据库的所有者允许对数据库中除了共享目录以外的所有表进行VACUUM操作（该限制意味着只有系统管理员才能真正对一个数据库进行VACUUM操作）。VACUUM命令会跳过那些用户没有权限的表进行垃圾回收操作。
- VACUUM不能在事务块内执行。
- 建议生产数据库经常清理（至少每晚一次），以保证不断地删除失效的行。VACUUM ANALYZE先执行一个VACUUM操作，然后给每个选定的表执行一个ANALYZE。尤其是在增删了大量记录之后，对受影响的表执行VACUUM ANALYZE命令是一个很好的习惯。这样将更新系统目录为最近的更改，并且允许查询优化器在规划用户查询时有更好的选择。
- VACUUM会导致I/O流量的大幅增加，这可能会影响其他活动会话的性能。因此，有时候会建议使用基于开销的VACUUM延迟特性。
- 如果指定了VERBOSE选项，VACUUM将打印处理过程中的信息，以表明当前正在处理的表。各种有关当前表的统计信息也会打印出来。
- 语法格式中含有带括号的选项列表时，选项可以以任何顺序写入。如果没有括号，则选项必须按语法显示的顺序给出。
- VACUUM列存表内部执行的操作包括四个：VACUUM主表、VACUUM主表的desc表、VACUUM主表的delta表、迁移delta表中的数据到主表。该操作不会回收delta表的存储空间，如果要回收delta表的冗余存储空间，需要对该列存表执行VACUUM DELTAMERGE。VACUUM主表操作默认开启，可以通过调节参数colvacuum_threshold_scale_factor（该参数需联系技术支持修改）来关闭该功能。
- VACUUM列存主表功能不支持临时表、冷热表和时序表。
- VACUUM列存主表功能对空间的回收是延后的，若需要立即回收空间，在执行VACUUM后，运行vac_fileclear_relation函数，将会对指定表施加排他锁，完成空间回收。
- 简单的VACUUM（不带FULL选项）只是简单地回收空间并且令其可以再次使用。这种形式的命令可以和对表的普通读写并发操作，因为没有请求排他锁。VACUUM FULL执行更广泛的处理，包括跨块移动行，以便把表压缩到最少的磁盘块数目里。这种形式要慢许多并且在处理的时候需要在表上施加一个排他锁。
- 不建议日常使用FULL选项，但是可以在特殊情况下使用。例如在用户删除了一个表的大部分行之后，希望从物理上缩小该表以减少磁盘空间占用。VACUUM FULL通常要比单纯的VACUUM收缩更多的表尺寸。如果执行此命令后所占用物理空间无变化（未减少），请确认是否有其他活跃事务（删除数据事务开始之前开始的事务，并在VACUUM FULL执行前未结束）存在，如果有等其他活跃事务退出进行重试。
- VACUUM和VACUUM FULL时，会根据参数vacuum_defer_cleanup_age延迟清理行存表记录，即不会立即清理刚删除的元组。
- VACUUM FULL系统表只能离线操作，在线VACUUM FULL系统表除了会锁表，还可能导致一些异常情况并产生报错。
- 对列存分区表执行VACUUM FULL，会同时锁表和锁分区。对分区表指定的分区VACUUM FULL时，只锁定指定的分区，其他分区仍然可以被访问，但对整个分区表的操作可能受到影响。

 版权所有 © 华为云计算技术有限公司

- VACUUM FULL操作分区表时与用户DML语句在如下特定场景有并发时可能发生分布式死锁，请谨慎操作：
  - VACUUM FULL子分区与insert/update/delete主表。
  - VACUUM FULL全表与select全表/select子分区。
- 在存算分离架构下，不支持全库VACUUM/全库VACUUM FULL/全库VACUUM ANALYZE/全库VACUUM DELTAMERGE操作。
- 在存算分离架构下，正在重分布的列存3.0的表执行VACUUM操作时，系统不会进行实际的数据清理动作，但会返回如下提示信息：

  执行VACUUM操作时，会跳过VACUUM操作。

```
VACUUM tpcds.web_returns_p1;
NOTICE:  skip vacuum relation "web_returns_p1" since it is in redistribution state
```

  执行VACUUM ANALYZE操作时，会跳过VACUUM操作，但是会收集统计信息。

```
VACUUM ANALYZE tpcds.web_returns_p1;
NOTICE:  skip vacuum relation "web_returns_p1" since it is in redistribution state
```

警告

- 对表执行VACUUM FULL操作时会触发表重建和表上的索引重建（重建过程中会先把数据写到一个新的数据文件中，重建完成之后删除原始文件），当表比较大时，重建会消耗较多的磁盘空间。当磁盘空间不足时，要谨慎对待大表VACUUM FULL操作，防止触发集群只读。
- 定期对脏页率、小CU占比超过25%的表执行VACUUM FULL，普通表需在低峰期执行，系统表需离线执行。
- 更多开发设计规范参见**总体开发设计规范**。

语法格式

- 回收空间并更新统计信息，关键字顺序必须按语法显示的顺序给出。

```
VACUUM [ ( { FULL | FREEZE | VERBOSE | {ANALYZE | ANALYSE }} [,...] ) ]
    [ table_name [ (column_name [, ...] ) ] ] [ PARTITION ( partition_name ) ];
```

- 仅回收空间，不更新统计信息。

```
VACUUM [ FULL [COMPACT] ] [ FREEZE ] [ VERBOSE ] [ table_name ] [ PARTITION
( partition_name ) ];
```

- 回收空间并更新统计信息，且对关键字顺序有要求。

```
VACUUM [ FULL ] [ FREEZE ] [ VERBOSE ] { ANALYZE | ANALYSE } [ VERBOSE ]
    [ table_name [ (column_name [, ...] ) ] ] [ PARTITION ( partition_name ) ];
```

- 针对HDFS表和列存表，将delta table中的数据转移到主表存储。（partition_name参数仅8.2.1.200及以上集群版本支持）

```
VACUUM DELTAMERGE [ table_name ][partition_name];
```

- 针对HDFS表，删除HDFS表在HDFS存储上的空值分区目录。

```
VACUUM HDFSDIRECTORY [ table_name ];
```

- 在线VACUUM FULL，支持DDL/DML操作与VACUUM FULL操作同时执行，不阻塞DDL/DML操作（该语法仅9.1.1.300及以上集群版本的2.0表支持）。

```
VACUUM FULL [ table_name ] ONLINE;
```

- 在线VACUUM FULL，支持指定分区执行，可通过WITH子句控制并行度、IO限速、bucketnums等参数（该语法仅9.1.2及以上集群版本支持）。

```
VACUUM FULL table_name [ PARTITION ( partition_name [, ...] ) ] ONLINE [ WITH ( option = value
[, ...] ) ];
```


版权所有 © 华为云计算技术有限公司

参数说明

表 13-172 VACUUM 参数说明

参数	描述	取值范围
FULL	选择该参数，可以恢复更多的空间，但是会耗时更多，并且在表上施加了排他锁。 FULL选项还可以带有COMPACT参数，只针对HDFS表，指定该参数的VACUUM FULL操作性能要好于未指定该参数的VACUUM FULL操作。 **须知** • COMPACT和PARTITION参数不能同时使用。 • 使用FULL参数会导致统计信息丢失，如果需要收集统计信息，请在VACUUM FULL语句中加上analyze关键字。	-
FREEZE	用于解决事务ID回绕(Transaction ID Wraparound)问题，指定FREEZE相当于执行VACUUM时将vacuum_freeze_min_age参数设为0。 **须知** DWS使用64位xid，理论上不会存在绕回问题，所以不建议客户操作VACUUM FREEZE。	-
VERBOSE	为每个表打印一份详细的清理工作报告。	-
ANALYZE | ANALYSE	更新用于优化器的统计信息，以决定执行查询的最有效方法。	-
table_name	需要清理的表的名称（可以带有模式修饰）。	缺省时为当前数据库中的所有表。
column_name	需要分析的具体的字段名称。	缺省时为所有字段。
PARTITION	指定清理的表为分区表。HDFS表不支持PARTITION参数。 **说明** PARTITION参数和COMPACT同时使用会报错： COMPACT can not be used with PARTITION.	-
partition_name	要清理的表的分区名称。	缺省时为所有分区。

版权所有 © 华为云计算技术有限公司

参数	描述	取值范围
DELTAMERGE	只针对HDFS表和列存表，将HDFS表或列存表的delta表中的数据转移到主表上存储。对HDFS表而言，当delta表中数据量小于六万行，则不作迁移，只有在大于或等于六万行数据时，将delta表中所有数据迁移到HDFS上，并通过truncate清理delta表的存储空间。对列存表而言，此操作总是会将delta表中的数据全部转移到CU中。 **说明** 为了检查列存delta表中的信息（HDFS表的delta表数据可以通过explain analyze查到），提供下述DFX函数，用于获取某个列存表的delta表中数据存储情况： • pgxc_get_delta_info(TEXT)，传入参数为列存表名，搜集并显示各个节点上的对应delta表信息，包括当前存活tuple数量、表大小、使用的最大block ID。 • get_delta_info(TEXT)，传入参数为列存表名，汇总pgxc_get_delta_info得到的结果，返回其delta表整体的当前存活tuple数量、表大小、使用的最大block ID。查询临时表delta信息时，需要显式指定临时表的schema，否则会报错找不到表。	-
HDFSDIRECTORY	只针对HDFS表，清理HDFS表（存储在HDFS表的目录下）的空值分区目录。	-
ONLINE	在线执行VACUUM FULL操作，不会全程阻塞DDL/DML操作。该参数仅9.1.1.300及以上集群版本支持。 • **ONLINE参数不能与除FULL外的其他任何关键字同时使用**。 • 在线执行VACUUM FULL功能的约束和限制请参考**在线VACUUM FULL**。	-
PARTITION (partition_name [, ...])	指定在线VACUUM FULL的目标分区列表。该参数仅9.1.2及以上集群版本支持。 仅对分区表的目标分区进行数据迁移和空间回收。支持指定一个或多个分区。指定单分区时主表锁降为3级（RowExclusiveLock），非目标分区可正常执行DML；多分区时主表保持7/8级锁，防止死锁。**指定分区时不允许设置bucket_nums参数**。	缺省时为全表ONLINE VACUUM FULL。
WITH (option = value [, ...])	在线VACUUM FULL的扩展选项，用于控制并行度、IO限速、bucketnums等参数。各选项详见**表13-173**。 该子句仅9.1.2及以上集群版本支持。	缺省时使用各参数默认值。

版权所有 © 华为云计算技术有限公司

表 13-173 在线 VACUUM FULL WITH 子句参数说明

option	类型	取值范围	默认值	描述
parallel_threads	INTEGER	1 ~ 128	4	分区并行线程数，控制数据基线重分布和索引构建的分区并发线程数。
insert_delay_time_ms	INTEGER	0 ~ 1000	0	数据插入限速（毫秒）。限速公式：新速率 = 原速率 / (1 + insert_delay_time_ms / 100)。值为0表示不限速，业务高峰期建议设置100~500。
index_delay_time_ms	INTEGER	0 ~ 1000	0	索引创建限速（毫秒）。限速公式与insert_delay_time_ms相同。索引构建为I/O密集型操作，业务高峰期建议设置适当限速。
enable_batch_read_optimize	BOOLEAN	on / off	on	批量读取优化开关。开启后使用独立Ring Buffer进行批量读取，防止污染共享缓存，同时启用Disk Cache Bulk Read提升吞吐量。内存紧张时可关闭以降低内存压力。
bucket_nums	INTEGER	1~1024或16384，0表示保持原值	0	新表哈希桶数。借助ONLINE VACUUM FULL动态调整表的bucketnums。指定分区时不允许设置此参数。
enable_drop_column_optimize	BOOLEAN	on / off	on	DROP列优化开关。建议保持该参数开启。 • 开启时使用insert_desc_table机制追踪增量数据变更。 • 关闭时ONLINE VACUUM FULL通过在临时表上添加4个辅助列来追踪变更。对于行存表，这4个辅助列会导致pg_attribute中累积dropped column记录；列存表不受影响。

示例

清理当前数据库中的所有表。

```
VACUUM;
```

仅回收表tpcds.web_returns_p1分区P2的空间，不更新统计信息。

```
VACUUM FULL tpcds.web_returns_p1 PARTITION(P2);
```

回收表tpcds.web_returns_p1空间，并更新统计信息。

 版权所有 © 华为云计算技术有限公司

```
VACUUM FULL ANALYZE tpcds.web_returns_p1;
```

清理当前数据库中的所有表并收集查询优化器的统计信息。

```
VACUUM ANALYZE;
```

仅清理特定表reason。

```
VACUUM (VERBOSE, ANALYZE) tpcds.reason;
```

对列存表table_delta进行DELTAMERGE操作。

```
VACUUM DELTAMERGE tpcds.table_delta;
```

仅对列存表table_delta的分区p1进行DELTAMERGE操作。

```
VACUUM DELTAMERGE tpcds.table_delta partition(p1);
```

执行在线VACUUM FULL。

```
VACUUM FULL tpcds.web_returns_p1 ONLINE;
```

指定分区执行在线VACUUM FULL，仅对分区p1和p2进行空间回收。

```
VACUUM FULL tpcds.web_returns_p1 PARTITION(p1, p2) ONLINE;
```

执行在线VACUUM FULL，设置并行度和IO限速调整对业务的影响。

```
VACUUM FULL tpcds.web_returns_p1 ONLINE WITH(parallel_threads = 4, insert_delay_time_ms = 500,
index_delay_time_ms = 500);
```

执行在线VACUUM FULL，动态调整表的bucketnums。

```
VACUUM FULL tpcds.web_returns_p1 ONLINE WITH(bucket_nums = 16);
```

13.11.7 PURGE

功能描述

清理回收站对象，回收站对象被清理之后，将无法再恢复。该功能仅9.1.0.200及以上集群版本支持。

该功能9.1.2以下集群版本中为Beta特性，如需使用请联系技术支持；9.1.2及以上集群版本中已正式商用。

注意事项

- PURGE RECYCLEBIN清空当前数据库所有回收站对象，执行该操作时会修改ttl和回收站对象状态，实际的文件删除操作仍然由后台线程自动执行。
- 仅系统管理员有权限执行PURGE RECYCLEBIN操作。

语法格式

```
PURGE { TABLE [schema_name.]table_name [PARTITION partition_name]
      | RECYCLEBIN
    }
```

 版权所有 © 华为云计算技术有限公司

参数说明

表 13-174 PURGE 参数说明

参数	描述	取值范围
schema_name	目标表的模式名称。	有效的模式名。
table_name	table_name为原表名或回收站对象名。如果指定了schema_name，则table_name为原表名，否则为回收站对象名。	有效的表名或回收站对象名。
RECYCLEBIN	清理所有回收站中的对象，执行后会对回收站内的对象异步清理。	无

示例

清理指定表。

```
PURGE TABLE rb_purge.t3_col_obs_part;
```

根据回收站中的对象名称（示例中为tb_d_a3b62756_6b90_45f3_85ae_5570c9989961）进行清理，对象名称可通过pg_file_recyclbin系统表查询获取。

```
PURGE TABLE tb_d_a3b62756_6b90_45f3_85ae_5570c9989961;
```

清理指定分区。

```
PURGE TABLE rb_purge.t3_col_obs_part partition p2;
```

全部清除，执行该命令后，会创建清理任务，后台任务线程会执行清理。

```
PURGE recyclebin;
```

13.11.8 TIMECAPSULE

功能描述

从回收站中恢复V3表数据，回收站中的对象在TRUNCATE，DROP，ALTER TABLE TRUNCATE PARTITION时放入回收站，并设置了ttl，在有效期内支持从回收站中恢复数据至上述操作之前。该功能仅9.1.0.200及以上集群版本支持。

该功能9.1.2以下集群版本中为Beta特性，如需使用请联系技术支持；9.1.2及以上集群版本中已正式商用。

注意事项

- 闪回操作必须确保GUC参数enable_recyclebin设置为on。
- 回收站中的对象，在生命周期之内，不会实际删除物理空间，因此还会额外的占用本地存储或者OBS的存储空间并产生相关存储费用。
- 闪回点（TRUNCATE或者DROP操作的时间点）和当前点之间，执行过修改表结构或影响物理文件的结构（DDL、VACUUM FULL、add/drop/split/merge/

 版权所有 © 华为云计算技术有限公司

exchange等分区操作），闪回失败，报错：“ERROR：The table definition of %s has been changed.”。

- 回收站中不保留索引（主键索引会保留并重命名）。Truncate加入回收站未保留索引数据，闪回后未自动rebuild index，如果使用index可能存在数据不一致的风险，由用户自行决定是否需要reindex table。
- 恢复DROP表时，只恢复基表名，其他子对象名均保持回收站对象名。用户可根据需要，执行DDL命令手动调整子对象名。
- 集群重分布之后会导致所有回收站对象变为不可闪回。

语法格式

```
TIMECAPSULE TABLE
   [schema_name.]table_name [PARTITION part_name] TO { BEFORE { DROP [RENAME TO
[new_schema_name.]new_table_name] | TRUNCATE } }
```

参数说明

表 13-175 TIMECAPSULE 参数说明

参数	描述	取值范围
schema_name	目标表的模式名称。	字符串，有效的模式名称。
table_name	table_name可为原名或表回收站名。	如果指定了schema_name，则table_name为原表名，否则为回收站对象名。如果指定为原表名，且回收站存在多个同名对象时，优先恢复TRUNCATE/DROP时间最近的有效对象。
part_name	分区名称。	字符串，有效的分区名称。
new_schema_name	闪回到新的模式名称。	字符串，有效的模式名称。
new_table_name	闪回到新的表的名称。	字符串，有效的表名称。

示例

闪回TRUNCATE。

```
TIMECAPSULE TABLE rb_truncate.v3_rr TO BEFORE TRUNCATE;

--- tb_d_5cc4d328_418b_4b74_9943_7676beefba4c 为回收站对象名
TIMECAPSULE TABLE tb_d_5cc4d328_418b_4b74_9943_7676beefba4c TO BEFORE TRUNCATE;
```

闪回单分区TRUNCATE。

```
TIMECAPSULE TABLE rb_truncate_part.t4_col_obs_part PARTITION p2 TO BEFORE TRUNCATE;
TIMECAPSULE TABLE tb_d_5cc4d328_418b_4b74_9943_7676beefba4c PARTITION p2 TO BEFORE TRUNCATE;
```

闪回DROP。

```
TIMECAPSULE TABLE rb_drop.v3_rr_drop TO BEFORE DROP;
TIMECAPSULE TABLE tb_d_5cc4d328_418b_4b74_9943_7676beefba4c TO BEFORE DROP;
```

 版权所有 © 华为云计算技术有限公司

13.12 设置与注释

13.12.1 SHOW

功能描述

SHOW将显示当前运行时参数的数值。可以使用SET语句来设置这些参数。

注意事项

SHOW可以查看的某些参数是只读的，可以查看但不能设置它们的值。

语法格式

```
SHOW
 {
   configuration_parameter |
   CURRENT_SCHEMA |
   TIME ZONE |
   TRANSACTION ISOLATION LEVEL |
   SESSION AUTHORIZATION |
   ALL
 };
```

参数说明

显示变量的参数请参见RESET的**参数说明**。

示例

显示timezone参数值：

```
SHOW timezone;
```

显示参数DateStyle的当前设置：

```
SHOW DateStyle;
```

显示所有参数的当前设置：

```
SHOW ALL;
```

相关链接

SET，**RESET**

13.12.2 SET

功能描述

用于修改运行时配置参数。

 版权所有 © 华为云计算技术有限公司

注意事项

大多数运行时参数都可以用SET在运行时设置，但有些则在服务运行过程中或会话开始之后不能修改。

语法格式

- 设置所处的时区。
  ```
  SET [ SESSION | LOCAL ] TIME ZONE { timezone | LOCAL | DEFAULT };
  ```
- 设置所属的模式。
  ```
  SET [ SESSION | LOCAL ]
    {CURRENT_SCHEMA { TO | = } { schema | DEFAULT }
    | SCHEMA 'schema'};
  ```
- 设置客户端编码集。
  ```
  SET [ SESSION | LOCAL ] NAMES encoding_name;
  ```
- 设置XML的解析方式。
  ```
  SET [ SESSION | LOCAL ] XML OPTION { DOCUMENT | CONTENT };
  ```
- 设置其他运行时参数。
  ```
  SET [ LOCAL | SESSION ]
    { {config_parameter { { TO | = } { value | DEFAULT }
                        | FROM CURRENT }}};
  ```

参数说明

表 13-176 SET 参数说明

参数	描述	取值范围
SESSION	声明的参数只对当前会话起作用。如果SESSION和LOCAL都没出现，则SESSION为缺省值。 如果在事务中执行了此命令，命令的产生影响将在事务回滚之后消失。如果该事务已提交，影响将持续到会话的结束，除非被另外一个SET命令重置参数。	-
LOCAL	声明的参数只在当前事务中有效。在COMMIT或ROLLBACK之后，会话级别的设置将再次生效。 不论事务是否提交，此命令的影响只持续到当前事务结束。一个特例是：在一个事务里面，既有SET命令，又有SET LOCAL命令，且SET LOCAL在SET后面，则在事务结束之前，SET LOCAL命令会起作用，但事务提交之后，则是SET命令会生效。	-

 版权所有 © 华为云计算技术有限公司

参数	描述	取值范围
TIME ZONE timezone	用于指定当前会话的本地时区。	有效的本地时区。该选项对应的运行时参数名称为TimeZone，DEFAULT缺省值为PRC。
CURRENT_SCHEMA schema	CURRENT_SCHEMA用于指定当前的模式。	已存在模式名称。
SCHEMA schema	同CURRENT_SCHEMA。此处的schema是个字符串。 例如：set schema 'public';	-
NAMES encoding_name	用于设置客户端的字符编码。等价于set client_encoding to encoding_name。	有效的字符编码。该选项对应的运行时参数名称为client_encoding，默认编码为UTF8。
XML OPTION option	用于设置XML的解析方式。	CONTENT（缺省）、DOCUMENT
config_parameter	可设置的运行时参数的名称。可用的运行时参数可以使用SHOW ALL命令查看。 **说明** 部分通过SHOW ALL查看的参数不能通过SET设置。如max_datanodes。	-
value	config_parameter的新值。可以声明为字符串常量、标识符、数字，或者逗号分隔的列表。DEFAULT用于把这些参数设置为它们的缺省值。	-

示例

设置模式tpcds搜索路径：

```
SET search_path TO tpcds, public;
```

把日期时间风格设置为传统的POSTGRES风格(日在月前)：

```
SET datestyle TO postgres;
```

相关链接

RESET，**SHOW**

版权所有 © 华为云计算技术有限公司

13.12.3 RESET

功能描述

RESET将指定的运行时参数恢复为缺省值。这些参数的缺省值是指postgresql.conf配置文件中所描述的参数缺省值。

RESET命令与如下命令的作用相同：

SET configuration_parameter TO DEFAULT

注意事项

RESET的事务性行为和SET相同：它的影响将会被事务回滚撤销。

语法格式

```
RESET {configuration_parameter | CURRENT_SCHEMA | TIME ZONE | TRANSACTION ISOLATION LEVEL | SESSION AUTHORIZATION | ALL };
```

参数说明

表 13-177 RESET 参数说明

参数	描述	取值范围
configuration_parameter	运行时参数的名称。	可以使用SHOW ALL命令查看运行时参数。 **说明** 部分通过SHOW ALL查看的参数不能通过SET设置。如max_datanodes。
CURRENT_SCHEMA	当前模式。	-
TIME ZONE	时区。	-
TRANSACTION ISOLATION LEVEL	事务的隔离级别。	-
SESSION AUTHORIZATION	当前会话的用户标识符。	-
ALL	所有运行时参数。	-

示例

把timezone设为缺省值：

```
RESET timezone;
```

把所有参数设置为缺省值：

 版权所有 © 华为云计算技术有限公司

```
RESET ALL;
```

相关链接

SET，SHOW

13.12.4 SET CONSTRAINTS

功能描述

SET CONSTRAINTS设置当前事务检查行为的约束条件。

IMMEDIATE约束是在每条语句后面进行检查。DEFERRED约束一直到事务提交时才检查。每个约束都有自己的模式。

从创建约束条件开始，一个约束总是设定为DEFERRABLE INITIALLY DEFERRED，DEFERRABLE INITIALLY IMMEDIATE，NOT DEFERRABLE三个特性之一。第三种总是IMMEDIATE，并且不会受SET CONSTRAINTS影响。前两种以指定的方式启动每个事务，但是其行为可以在事务里用SET CONSTRAINTS改变。

带着一个约束名列表的SET CONSTRAINTS改变这些约束的模式（都必须是可推迟的）。如果有多个约束匹配某个名字，则所有都会被影响。SET CONSTRAINTS ALL改变所有可推迟约束的模式。

当SET CONSTRAINTS把一个约束从DEFERRED改成IMMEDIATE的时候，新模式反作用式地起作用：任何将在事务结束准备进行的数据修改都将在SET CONSTRAINTS的时候执行检查。如果违反了任何约束，SET CONSTRAINTS都会失败（并且不会修改约束模式）。因此，SET CONSTRAINTS可以用于强制在事务中某一点进行约束检查。

目前，只有外键约束被该设置影响。检查和唯一约束总是不可推迟的。

注意事项

SET CONSTRAINTS只在当前事务里设置约束的行为。因此，如果用户在事务块之外（START TRANSACTION/COMMIT对）执行这个命令，它将没有任何作用。

语法格式

```
SET CONSTRAINTS { ALL | { name } [, ...] } { DEFERRED | IMMEDIATE };
```

参数说明

表 13-178 SET CONSTRAINTS 参数说明

参数	描述	取值范围
name	约束名。	已存在的约束名。可以在系统表pg_constraint中查到。
ALL	所有约束。	-
DEFERRED	约束一直到事务提交时才检查。	-
IMMEDIATE	约束在每条语句后进行检查。	-

版权所有 © 华为云计算技术有限公司

示例

设置所有约束在事务提交时检查：

```
SET CONSTRAINTS ALL DEFERRED;
```

13.12.5 SET SESSION AUTHORIZATION

功能描述

在数据库运维管理中，管理员通常以自己的身份执行日常操作。但在某些场景下，管理员需要以其他用户身份执行特定操作或验证权限配置——例如排查某用户的查询权限是否正确、以目标用户身份执行一段存储过程、或在权限迁移后验证新权限是否生效。如果每次都退出当前会话并以目标用户重新登录，不仅操作繁琐，还会中断当前会话中的临时对象和上下文。

通过SET SESSION AUTHORIZATION命令，管理员可以在不重新登录的情况下，将当前会话的会话用户标识和当前用户标识同时切换为指定的角色，从而以目标用户身份继续执行操作。操作完成后，可通过RESET命令恢复为初始认证用户。

注意事项

只有在初始会话用户有系统管理员权限的时候，会话用户标识符才能改变。否则，只有在指定了被认证的用户名的情况下，系统才接受该命令。

语法格式

- 为当前会话设置会话用户标识符和当前用户标识符。
  ```
  SET [ SESSION | LOCAL ] SESSION AUTHORIZATION role_name PASSWORD '{password}';
  ```
- 重置会话和当前用户标识符为初始认证的用户名。
  ```
  {SET [ SESSION | LOCAL ] SESSION AUTHORIZATION DEFAULT
      | RESET SESSION AUTHORIZATION};
  ```

参数说明

表 13-179 SET SESSION AUTHORIZATION 参数说明

参数	描述	取值范围
SESSION	可选关键字。声明这个命令对**当前会话**起作用，效果在当前会话的整个生命周期内持续有效。如果省略SESSION和LOCAL，默认效果与SESSION相同。	-
LOCAL	可选关键字。声明该命令只在**当前事务**中有效，当事务提交或回滚后，用户标识恢复为切换前的状态。 与SESSION的区别在于：SESSION在整个会话期间生效，而LOCAL仅在当前事务内生效。	-
role_name	角色名。指定要切换到的目标角色名称。	已存在的角色名。

版权所有 © 华为云计算技术有限公司

参数	描述	取值范围
password	角色的密码。要求符合密码的命名规则。	-
DEFAULT	重置会话和当前用户标识符为初始认证的用户名。	-

示例

步骤1 设置当前用户为paul。

```
SET SESSION AUTHORIZATION paul password '{password}';
```

步骤2 查看当前会话用户和当前用户，验证身份切换是否成功。

```
SELECT SESSION_USER, CURRENT_USER;
```

```
gaussdb=> SELECT SESSION_USER, CURRENT_USER;
 session_user | current_user
--------------+--------------
 paul         | paul
(1 row)
```

步骤3 重置当前用户，恢复为初始认证用户。

```
RESET SESSION AUTHORIZATION;
```

步骤4 验证身份切换是否成功。

```
SELECT SESSION_USER, CURRENT_USER;
```

```
gaussdb=> SELECT SESSION_USER, CURRENT_USER;
 session_user | current_user
--------------+--------------
 dbadmin      | dbadmin
(1 row)
```

----结束

相关参考

SET ROLE

常见问题

执行 SET SESSION AUTHORIZATION 报错：“permission denied to set session authorization”

- **原因**：当前初始会话用户没有系统管理员权限，且未提供目标角色的密码。
- **排查方法**：确认当前用户是否具有系统管理员权限。如果不具备管理员权限，则必须在命令中指定PASSWORD参数，且密码需与目标角色的实际密码一致。

 版权所有 © 华为云计算技术有限公司

13.12.6 COMMENT

功能描述

定义或修改对象的注释。

注意事项

- 每个对象只存储一条注释，因此要修改对象的注释，对同一个对象发出一条新的COMMENT命令即可。要删除注释，在文本字符串的位置写入NULL。删除对象时，注释将自动删除。
- 目前查看注释没有相关安全机制：连接到数据库的任何用户都可以查看该数据库中对象的所有注释。共享对象（比如数据库、角色、表空间）的注释是全局存储的，连接到集群中任何数据库的任何用户都可以查看共享对象的所有注释。因此，不要在注释里添加与安全有关的敏感信息。
- 对大多数对象，只有对象的所有者才能设置注释。角色没有所有者，所以COMMENT ON ROLE命令仅可以由系统管理员对系统管理员角色执行，或者具有CREATE ROLE权限的角色也可以为非系统管理员角色添加注释。系统管理员可以对所有对象添加注释。

语法格式

```
COMMENT ON
{
  AGGREGATE agg_name (agg_type [, ...] ) |
  CAST (source_type AS target_type) |
  COLLATION object_name |
  COLUMN { table_name.column_name | view_name.column_name } |
  CONSTRAINT constraint_name ON table_name |
  CONVERSION object_name |
  DATABASE object_name |
  DOMAIN object_name |
  EXTENSION object_name |
  FOREIGN DATA WRAPPER object_name |
  FOREIGN TABLE object_name |
  FUNCTION function_name ( [ {[ argmode ] [ argname ] argtype} [, ...] ] ) |
  INDEX object_name |
  LARGE OBJECT large_object_oid |
  MATERIALIZED VIEW object_name |
  OPERATOR operator_name (left_type, right_type) |
  OPERATOR CLASS object_name USING index_method |
  OPERATOR FAMILY object_name USING index_method |
  [ PROCEDURAL ] LANGUAGE object_name |
  ROLE object_name |
  RULE rule_name ON table_name |
  SCHEMA object_name |
  SERVER object_name |
  TABLE object_name |
  TABLESPACE object_name |
  TEXT SEARCH CONFIGURATION object_name |
  TEXT SEARCH DICTIONARY object_name |
  TEXT SEARCH PARSER object_name |
  TEXT SEARCH TEMPLATE object_name |
  TYPE object_name |
  VIEW object_name
}
   IS 'text';
```

 版权所有 © 华为云计算技术有限公司

参数说明

表 13-180 COMMENT 参数说明

参数	描述	取值范围
agg_name	需要注释的聚集函数的名称。	-
agg_type	聚集函数参数的类型。	-
source_type	类型转换的源数据类型。	-
target_type	类型转换的目标数据类型。	-
object_name	对象的名称。	-
table_name.column_name/ view_name.column_name	定义/修改注释的列名称。前缀可加表名称或者视图名称。	-
constraint_name	定义/修改注释的表约束的名称。	已存在的表约束名称。
table_name	表名称。	已存在的表名称。
function_name	定义/修改注释的函数名称。	已存在的函数名称。
argmode/argname/ argtype	函数参数的模式、名称、类型。	-
large_object_oid	定义/修改注释的大对象的OID。	-
MATERIALIZED VIEW object_name	定义/修改注释的物化视图名称。该字段仅9.1.1.100及以上集群版本支持。	已存在的物化视图名称。
operator_name	操作符名称。	-
left_type/right_type	操作参数的数据类型（可以用模式指定）。当前置或者后置操作符不存在时，可以增加NONE选项。	-
text	新注释。	字符串。

示例

创建示例表CUSTOMER，并在表CUSTOMER的customer.c_customer_sk列上添加注释。

```
DROP TABLE IF EXISTS customer;
CREATE TABLE customer
(
    c_custkey     BIGINT NOT NULL CONSTRAINT c_custkey PRIMARY KEY  ,
    c_name        VARCHAR(25)  ,
    c_address     VARCHAR(40)  ,
    c_nationkey   INT          ,
```

 版权所有 © 华为云计算技术有限公司

```
    c_phone       CHAR(15)     ,
    c_acctbal     DECIMAL(15,2)
)
DISTRIBUTE BY HASH(c_custkey);

COMMENT ON COLUMN customer.c_custkey IS 'Primary key of customer demographics table.';
```

创建视图customer_details_view并给视图添加注释。

```
CREATE OR REPLACE VIEW customer_details_view AS SELECT * FROM customer;
COMMENT ON VIEW customer_details_view IS 'View of customer detail';
```

给表customer添加注释。

```
COMMENT ON TABLE customer IS 'This is my table';
```

给物化视图添加注释。

```
COMMENT ON materialized view matview_basic2.mv_date2 IS 'this is a matview.';
```

13.13 发布和订阅

13.13.1 CREATE PUBLICATION

功能描述

为当前数据库添加一个新的发布（PUBLICATION）。

发布本质上是通过逻辑复制将一组表的数据变更进行复制。数据库将自动跟踪并捕获这些源表上发生的数据操作（INSERT， UPDATE，DELETE和TRUNCATE），并通过逻辑复制机制，持续地将这些变更数据传输给所有配置并订阅（CREATE SUBSCRIPTION）该发布的目标数据库，从而实现源端到目标端的数据同步。

注意事项

- 该语法仅8.2.0.100及以上集群版本支持。
- 创建发布的用户必须拥有当前数据库的CREATE权限。
- 新创建的发布的名称不能与当前数据库中任何现有发布的名称相同。
- 要将表添加到发布中，操作用户必须拥有该表的所有权。FOR ALL TABLES和FOR ALL TABLES IN SCHEMA子句要求用户具有系统管理员权限。
- 如果创建发布时未指定FOR TABLE，也未指定FOR ALL TABLES， 则该发布是一组空表，后续可以通过**ALTER PUBLICATION**命令在该发布添加表。
- CREATE PUBLICATION不会启动复制过程，只为未来的订阅者定义一个分组和过滤逻辑。
- 不能同时通过FOR TABLE和FOR ALL TABLES IN SCHEMA方式添加同一个待发布表到同一个发布中。

语法格式

```
CREATE PUBLICATION name
    [ FOR ALL TABLES
      | FOR publication_object [, ... ] ]
    [ WITH ( publication_parameter [=value] [, ... ] ) ];
```

其中，发布对象publication_object为：

 版权所有 © 华为云计算技术有限公司

```
TABLE table_name [, ...]
| ALL TABLES IN SCHEMA schema_name [, ... ]
```

参数说明

表 13-181 CREATE PUBLICATION 参数说明

参数	描述	取值范围
name	新发布的名称。	字符串，要符合标识符的命名规范。参见**标识符命名规范**。
FOR ALL TABLES	将发布标记为复制数据库中所有细粒度容灾主表的更改，包括在将来创建的表。	-
FOR TABLE	指定要添加到发布的表的列表。只有细粒度容灾主表才能成为发布的一部分。	-
table_name	要添加到发布的表的名字，可以带模式名。	有效的表名称。
FOR ALL TABLES IN SCHEMA	将发布标记为复制指定模式列表中所有细粒度容灾主表的更改，包括在将来创建的表。	-
schema_name	要添加到发布的模式的名字。	有效的模式名称。
WITH (publication_parameter [=value] [, ...])	该子句指定发布的可选参数，当前支持publish参数。 publish参数决定了哪些DML操作将由新的发布给订阅者。 • 取值范围：字符串，用逗号分隔的操作列表。允许的操作为：INSERT、 UPDATE、DELETE和TRUNCATE。 • 默认值为：'insert, update, delete, truncate'，表示默认发布所有操作。	-

版权所有 © 华为云计算技术有限公司

示例

- 创建模式、示例表及样例数据。

```
CREATE SCHEMA pub_demo;

CREATE TABLE pub_demo.sales
 (
   sale_id    INTEGER,
   product    VARCHAR(50),
   amount     INTEGER NOT NULL,
   sale_date  DATE
) WITH (ORIENTATION = COLUMN,enable_disaster_cstore='on') DISTRIBUTE BY HASH ( sale_id);

CREATE TABLE pub_demo.customers
(
   cust_id   INTEGER,
   cust_name VARCHAR(50),
   region    VARCHAR(20)
) WITH (ORIENTATION = COLUMN,enable_disaster_cstore='on') DISTRIBUTE BY HASH ( cust_id);

INSERT INTO pub_demo.sales VALUES (1, 'Laptop', 500, '2026-01-15');
INSERT INTO pub_demo.sales VALUES (2, 'Mouse',  900,  '2026-01-16');
INSERT INTO pub_demo.customers VALUES (101, 'Alice', 'East');
INSERT INTO pub_demo.customers VALUES (102, 'Bob',   'West');
```

- 创建发布，发布所有表。

```
CREATE PUBLICATION pub_all_tables FOR ALL TABLES;
```

- 创建发布，发布单个表sales 。

```
CREATE PUBLICATION pub_sales_only FOR TABLE pub_demo.sales;
```

- 创建发布，发布单个表sales的所有操作（INSERT、 UPDATE、DELETE和TRUNCATE）。默认包含全部操作，也可显式声明。

```
CREATE PUBLICATION pub_sales_all_ops
FOR TABLE pub_demo.sales
WITH (publish = 'insert, update, delete, truncate');
```

- 创建发布，发布单个表customers的INSERT操作。

```
CREATE PUBLICATION pub_cust_insert
FOR TABLE pub_demo.customers
WITH (publish = 'insert');
```

相关链接

ALTER PUBLICATION，**DROP PUBLICATION**

13.13.2 ALTER PUBLICATION

功能描述

更改发布的属性。

注意事项

- 该语法仅8.2.0.100及以上集群版本支持。
- 发布的所有者和系统管理员才能执行ALTER PUBLICATION。
- 要更改发布的所有者，用户必须是新所有者角色的直接或者间接成员，并且新的所有者必须具有当前数据库的CREATE权限。
- FOR ALL TABLES发布的新所有者必须是系统管理员。
- 管理员用户可以更改任何发布的所属关系。

版权所有 © 华为云计算技术有限公司

语法格式

- 向发布中添加一个或多个发布对象。
  ```
  ALTER PUBLICATION name ADD publication_object [, ...]
  ```
- 从发布中删除一个或多个发布对象。
  ```
  ALTER PUBLICATION name DROP publication_object [, ...]
  ```
- 用指定的发布对象替换当前发布对象。
  ```
  ALTER PUBLICATION name SET publication_object [, ...]
  ```
- 设置发布的参数，未显式指定的参数保持原值。
  ```
  ALTER PUBLICATION name SET ( publication_parameter [= value] [, ... ] )
  ```
- 更改发布的所有者。
  ```
  ALTER PUBLICATION name OWNER TO new_owner
  ```
- 重命名发布。
  ```
  ALTER PUBLICATION name RENAME TO new_name
  ```

其中发布对象publication_object为：

```
TABLE table_name [, ...]
| ALL TABLES IN SCHEMA schema_name [, ... ]
```

参数说明

表 13-182 ALTER PUBLICATION 参数说明

参数	描述	取值范围
name	需要修改的发布的名称。	已存在的发布名。
table_name	表名称。	已存在的表名。
schema_name	模式的名称。	已存在的模式名。
SET (publication_parameter [= value] [, ...])	该子句修改最初由CREATE PUBLICATION设置的发布参数。	详细的参数说明请参考CREATE PUBLICATION的**参数说明**。
new_owner	发布的新所有者。	已存在的用户名。
new_name	发布的新名称。	字符串，需符合**标识符命名规范**。

示例

- 创建模式、示例表及样例数据。
  ```
  DROP SCHEMA IF EXISTS pub_demo;
  CREATE SCHEMA pub_demo;

  DROP TABLE IF EXISTS pub_demo.sales;
  CREATE TABLE pub_demo.sales
   (
     sale_id    INTEGER,
     product    VARCHAR(50),
     amount     INTEGER NOT NULL,
     sale_date  DATE
  ) WITH (ORIENTATION = COLUMN,enable_disaster_cstore='on') DISTRIBUTE BY HASH ( sale_id);

  DROP TABLE IF EXISTS pub_demo.sales;
  ```

 版权所有 © 华为云计算技术有限公司

```
CREATE TABLE pub_demo.customers
(
    cust_id   INTEGER,
    cust_name VARCHAR(50),
    region    VARCHAR(20)
) WITH (ORIENTATION = COLUMN,enable_disaster_cstore='on') DISTRIBUTE BY HASH ( cust_id);

INSERT INTO pub_demo.sales VALUES (1, 'Laptop', 500, '2026-01-15');
INSERT INTO pub_demo.sales VALUES (2, 'Mouse',  900,  '2026-01-16');
INSERT INTO pub_demo.customers VALUES (101, 'Alice', 'East');
INSERT INTO pub_demo.customers VALUES (102, 'Bob',   'West');
```

创建名称为pub_sales的发布。

```
CREATE PUBLICATION pub_sales FOR TABLE pub_demo.sales;
```

- 向发布中添加表。

```
ALTER PUBLICATION pub_sales ADD TABLE pub_demo.customers;
ALTER PUBLICATION pub_sales ADD TABLE pub_demo.customers;
```

- 从发布中删除表。

```
ALTER PUBLICATION pub_sales DROP ALL TABLES IN SCHEMA pub_demo;
ALTER PUBLICATION pub_sales DROP TABLE pub_demo.customers;
```

- 重新设置发布对象。

```
ALTER PUBLICATION pub_sales SET TABLE pub_demo.customers;
```

- 向发布中添加整个schema的所有表。

```
ALTER PUBLICATION pub_sales ADD ALL TABLES IN SCHEMA public;
```

- 修改发布参数，只发布INSERT操作。

```
ALTER PUBLICATION pub_sales SET (publish = 'insert');
```

- 修改发布的所有者。

```
CREATE ROLE new_publisher PASSWORD '********';
GRANT ALL PRIVILEGES ON ALL TABLES IN SCHEMA pub_demo TO new_publisher;
ALTER PUBLICATION pub_sales OWNER TO new_publisher;
```

- 修改发布名称。

```
ALTER PUBLICATION pub_sales RENAME TO pub_sales_new;
```

相关链接

CREATE PUBLICATION，**DROP PUBLICATION**

查看数据库中的发布信息

DWS提供以下系统表和视图用于查看当前数据库中的发布信息。

- **PG_PUBLICATION**系统表，用于查看当前数据库中创建的发布数据信息，包括发布名称、所有者、参数设置等。
- **PG_PUBLICATION_TABLES**系统视图，用于查看发布包含的表。

13.13.3 DROP PUBLICATION

功能描述

删除一个发布。

注意事项

- 该语法仅8.2.0.100及以上集群版本支持。
- 只有发布的所有者可以执行DROP PUBLICATION操作，系统管理员默认拥有此权限。

版权所有 © 华为云计算技术有限公司

语法格式

```
DROP PUBLICATION [ IF EXISTS ] name [, ...] [ CASCADE | RESTRICT ];
```

参数说明

表 13-183 DROP PUBLICATION 参数说明

参数	描述	取值范围
IF EXISTS	如果指定的发布不存在，不会抛出一个错误，而会发出一个通知，告知发布不存在。	-
name	要删除的发布名字。	已存在的发布。
CASCADE | RESTRICT	当前这些关键词没有任何作用，因为发布没有依赖关系。	-

示例

删除发布：

```
DROP PUBLICATION mypublication;
```

相关链接

ALTER PUBLICATION，**CREATE PUBLICATION**

13.13.4 CREATE SUBSCRIPTION

功能描述

为当前数据库添加一个新的订阅（SUBSCRIPTION）。

订阅（Subscription）是指一个数据库实例（称为订阅者）从另一个数据库实例（称为发布者）接收数据变更的过程。订阅者通过订阅发布者的数据变更，实现数据的同步和分发。

该语法在9.1.1.100及以上集群版本中已废弃，执行时会报错SUBSCRIPTION is not yet supported。

注意事项

- 只有系统管理员才可以创建订阅。
- 新创建的订阅名称不能与当前数据库中任何现有的订阅名称相同。

语法格式

```
CREATE SUBSCRIPTION name
    CONNECTION 'conninfo'
    PUBLICATION publication_name [, ...]
    [ WITH ( subscription_parameter [= value] [, ... ] ) ]
```


版权所有 © 华为云计算技术有限公司

参数说明

表 13-184 CREATE SUBSCRIPTION 参数说明

参数	描述	取值范围
name	新订阅的名称。	字符串，要符合标识符的命名规范。参见**标识符命名规范**。
conninfo	连接发布端的字符串。 • **host**，发布端IP地址，可以同时指定发布端主机和备机的IP地址，如果同时指定了多个IP，以英文逗号分隔。 • **port**，发布端端口，此处的端口不能使用主端口，而应该使用主端口+1端口，否则会与线程池冲突。可以同时指定发布端主机和备机的端口，如果同时指定了多个端口，以英文逗号分隔。**host和port的数量要一致，并且要相对应。** • **dbname**，发布所在的数据库。 • **user和password** 用于连接发布端且具有系统管理员权限(SYSADMIN)或者运维管理员权限(OPRADMIN)的用户名和密码。	例如：host=1.1.1.1,2.2.2.2 port=10000,20000 dbname=postgres user=repusr1 password=*password*。
publication_name	要订阅的发布端的发布名称，一个订阅可以对应多个发布。	已有的发布名称。

版权所有 © 华为云计算技术有限公司

参数	描述	取值范围
WITH (subscription_parameter [= value] [, ...])	该子句指定订阅的可选参数。 • **enabled**，指定订阅是否应该主动复制，或者只是先被创建而不立即运行。 取值范围：true，false。 默认值：true • **create_slot**，指定该命令是否要在发布者端创建复制槽。 取值范围：true，false。 默认值：true • **slot_name**，指定要在发布者端使用的复制槽名称。 取值范围：字符串。 默认值：默认使用订阅名称作为复制槽的名称。	-

示例

数据库发布的示例数据请参考**示例**。

- 创建订阅，订阅上游数据库的发布。
```
CREATE SUBSCRIPTION sub_demo
    CONNECTION 'host=10.***.*.** port=10000 user=u_admin dbname=testdb password=********'
    PUBLICATION pub_sales_only;
```
- 订阅多个发布。
```
CREATE SUBSCRIPTION sub_demo_1
    CONNECTION 'host=10.***.*.** port=10000 user=u_admin dbname=testdb password=********'
    PUBLICATION pub_sales_only,pub_sales_all_ops;
```
- 订阅所有表（需上游发布包含ALL TABLES IN SCHEMA）。
```
CREATE SUBSCRIPTION sub_demo_1
    CONNECTION 'host=10.***.*.** port=10000 user=u_admin dbname=testdb password=********'
    PUBLICATION pub_all_tables;
```

相关链接

ALTER SUBSCRIPTION，**DROP SUBSCRIPTION**

13.13.5 ALTER SUBSCRIPTION

功能描述

修改订阅的属性。

该语法在9.1.1.100及以上集群版本中已废弃，执行时会报错SUBSCRIPTION is not yet supported。

版权所有 © 华为云计算技术有限公司

注意事项

订阅的所有者才能执行ALTER SUBSCRIPTION，并且新的所有者必须是系统管理员。

语法格式

- 更新订阅的连接信息。
  ```
  ALTER SUBSCRIPTION name CONNECTION 'conninfo';
  ```
- 更新订阅的发布端的发布名称。
  ```
  ALTER SUBSCRIPTION name SET PUBLICATION publication_name [, ...];
  ```
- 启用订阅。
  ```
  ALTER SUBSCRIPTION name ENABLE;
  ```
- 禁用订阅。
  ```
  ALTER SUBSCRIPTION name DISABLE;
  ```
- 设置订阅的参数。
  ```
  ALTER SUBSCRIPTION name SET ( subscription_parameter [= value] [, ... ] );
  ```
- 更改订阅的所有者。
  ```
  ALTER SUBSCRIPTION name OWNER TO new_owner;
  ```
- 重命名订阅。
  ```
  ALTER SUBSCRIPTION name RENAME TO new_name;
  ```

参数说明

表 13-185 ALTER SUBSCRIPTION 参数说明

参数	描述	取值范围
name	需要修改的订阅名称。	已存在的订阅名。
CONNECTION 'conninfo'	该子句修改最初由CREATE SUBSCRIPTION设置的连接属性。	详细的参数说明请参考CREATE SUBSCRIPTION的**参数说明**。
publication_name	新发布的名称。	字符串，需符合**标识符命名规范**。
ENABLE	启用先前禁用的订阅，在事务结束时启动逻辑复制工作。	-
DISABLE	禁用正在运行的订阅，在事务结束时停止逻辑复制工作。	-
SET (publication_parameter [= value] [, ...])	该子句修改最初由CREATE PUBLICATION设置的发布参数。	详细的参数说明请参考CREATE SUBSCRIPTION的**参数说明**。
new_owner	订阅新所有者的用户名称。	已存在的用户名。
new_name	订阅的新名称。	字符串，需符合**标识符命名规范**。

版权所有 © 华为云计算技术有限公司

示例

- 创建订阅sub_demo。
```
DROP SUBSCRIPTION IF EXISTS sub_demo;
CREATE SUBSCRIPTION sub_demo
    CONNECTION 'host=10.***.*.** port=10000 user=u_admin dbname=testdb password=********'
    PUBLICATION pub_sales_only;
```
- 将订阅的发布更改为insert_only。
```
ALTER SUBSCRIPTION sub_demo SET PUBLICATION insert_only;
```
- 重命名订阅。
```
ALTER SUBSCRIPTION sub_demo RENAME TO new_sub_demo;
```
- 禁用订阅。
```
ALTER SUBSCRIPTION new_sub_demo DISABLE;
```

相关链接

CREATE SUBSCRIPTION，**DROP SUBSCRIPTION**

13.13.6 DROP SUBSCRIPTION

功能描述

删除一个订阅。

该语法在9.1.1.100及以上集群版本中已废弃，执行时会报错SUBSCRIPTION is not yet supported。

注意事项

只有系统管理员才可以删除订阅。

语法格式

```
DROP SUBSCRIPTION [ IF EXISTS ] name;
```

参数说明

表 13-186 DROP SUBSCRIPTION 参数说明

参数	描述	取值范围
IF EXISTS	如果指定的订阅不存在，不会抛出一个错误，而会发出一个通知，告知订阅不存在。	-
name	要删除的订阅名字。	已存在的订阅。

示例

删除订阅：
```
DROP SUBSCRIPTION mysub;
```


版权所有 © 华为云计算技术有限公司

相关链接

ALTER SUBSCRIPTION，CREATE SUBSCRIPTION

13.14 其他

13.14.1 CREATE SYNONYM

功能描述

创建一个同义词对象。同义词是数据库对象的别名，用于记录与其他数据库对象名间的映射关系，用户可以使用同义词访问关联的数据库对象。

注意事项

- 定义同义词的用户成为其所有者。
- 若指定模式名称，则同义词在指定模式中创建。否则，在当前模式创建。
- 支持通过同义词访问的数据库对象包括：表、视图、函数和存储过程。
- 使用同义词时，用户需要具有对关联对象的相应权限。
- 支持使用同义词的DML语句包括：SELECT、INSERT、UPDATE、DELETE、EXPLAIN、CALL。
- 不支持关联函数或存储过程的CREATE SYNONYM语句出现在存储过程中，建议存储过程中使用系统表pg_synonym中已存在的同义词对象。

语法格式

```
CREATE [ OR REPLACE ] SYNONYM synonym_name
    FOR object_name;
```

参数说明

表 13-187 CREATE SYNONYM 参数说明

参数	描述	取值范围
synonym	创建的同义词名字，可以带模式名。	取值范围：字符串，要符合标识符的命名规范。 参见标识符命名规范。
object_name	关联的对象名字，可以带模式名。	取值范围：字符串，要符合标识符的命名规范。 参见标识符命名规范。 **说明** object_name可以是不存在的对象名称。

版权所有 © 华为云计算技术有限公司

示例

创建模式ot：

```
DROP SCHEMA IF EXISTS ot;
CREATE SCHEMA ot;
```

创建表ot.t1及其同义词t1：

```
CREATE TABLE ot.t1(id int, name varchar2(10)) DISTRIBUTE BY hash(id);
DROP SYNONYM IF EXISTS t1;
CREATE OR REPLACE SYNONYM t1 FOR ot.t1;
```

使用同义词t1：

```
SELECT * FROM t1;
INSERT INTO t1 VALUES (1, 'ada'), (2, 'bob');
UPDATE t1 SET t1.name = 'cici' WHERE t1.id = 2;
```

创建同义词v1及其关联视图ot.v_t1：

```
DROP SYNONYM IF EXISTS v1;
CREATE SYNONYM v1 FOR ot.v_t1;
CREATE VIEW ot.v_t1 AS SELECT * FROM ot.t1;
```

使用同义词v1：

```
SELECT * FROM v1;
```

```
gaussdb=> SELECT * FROM v1;
 id | name
----+------
  1 | ada
  2 | cici
(2 rows)
```

创建重载函数ot.add及其同义词add：

```
CREATE OR REPLACE FUNCTION ot.add(a integer, b integer) RETURNS integer AS
$$
SELECT $1 + $2
$$
LANGUAGE sql;

CREATE OR REPLACE FUNCTION ot.add(a decimal(5,2), b decimal(5,2)) RETURNS decimal(5,2) AS
$$
SELECT $1 + $2
$$
LANGUAGE sql;

CREATE OR REPLACE SYNONYM add FOR ot.add;
```

使用同义词add：

```
SELECT add(1,2);
SELECT add(1.2,2.3);
```

 版权所有 © 华为云计算技术有限公司

```
gaussdb=> SELECT add(1,2);
 add
-----
   3
(1 row)

gaussdb=> SELECT add(1.2,2.3);
 add
-----
 3.5
(1 row)
```

创建存储过程ot.register及其同义词register：

```
DROP PROCEDURE IF EXISTS ot.register;
CREATE PROCEDURE ot.register(n_id integer, n_name varchar2(10))
SECURITY INVOKER
AS
BEGIN
    INSERT INTO ot.t1 VALUES(n_id, n_name);
END;
/

CREATE OR REPLACE SYNONYM register FOR ot.register;
```

使用同义词register，调用存储过程：

```
CALL register(3,'mia');
```

相关链接

ALTER SYNONYM，**DROP SYNONYM**

13.14.2 ALTER SYNONYM

功能描述

修改SYNONYM对象的属性。

注意事项

- 目前仅支持修改SYNONYM对象的所有者。
- 只有系统管理员和同义词的所有者有权限修改SYNONYM对象的所有者信息。
- 修改者必须是新所有者的直接或间接成员，并且新的所有者必须拥有该同义词所属模式的CREATE权限。

语法格式

```
ALTER SYNONYM synonym_name
    OWNER TO new_owner;
```

 版权所有 © 华为云计算技术有限公司

参数说明

表 13-188 ALTER SYNONYM 参数说明

参数	描述	取值范围
synonym_name	需要修改的同义词名称，可以带模式名。	已存在的同义词名称。
OWNER TO	用于修改同义词所有者的子句。	-
new_owner	同义词的新所有者。	有效的用户名。

示例

创建同义词t1。

```
CREATE OR REPLACE SYNONYM t1 FOR ot.t1;
```

创建新用户u1。

```
CREATE USER u1 PASSWORD '{password}';
```

修改同义词t1的所有者为u1。

```
ALTER SYNONYM t1 OWNER TO u1;
```

相关链接

CREATE SYNONYM，**DROP SYNONYM**

13.14.3 DROP SYNONYM

功能描述

删除指定的SYNONYM对象。

注意事项

只有SYNONYM的所有者有权限执行DROP SYNONYM命令，系统管理员默认拥有此权限。

语法格式

```
DROP SYNONYM [ IF EXISTS ] synonym_name [ CASCADE | RESTRICT ];
```

参数描述

表 13-189 DROP SYNONYM 参数说明

参数	描述	取值范围
IF EXISTS	如果指定的同义词不存在，则发出一个notice而不是抛出一个错误。	-

版权所有 © 华为云计算技术有限公司

参数	描述	取值范围
synonym_name	要删除的同义词名字，可以带模式名。	已存在的同义词名字。
CASCADE | RESTRICT	当执行删除操作时，如何处理依赖对象中的相关数据。	• CASCADE：级联删除依赖同义词的对象（比如视图）。 • RESTRICT：如果有依赖对象存在，则拒绝删除同义词。此选项为缺省值。

示例

删除同义词：

```
DROP SYNONYM IF EXISTS t1;
```

相关链接

ALTER SYNONYM，CREATE SYNONYM

13.14.4 ALTER LARGE OBJECT

功能描述

用于修改大对象的定义。ALTER LARGE OBJECT的唯一的功能是分配一个新的所有者。

注意事项

只有大对象的所有者有权限执行此命令，系统管理员默认拥有此权限。

语法格式

```
ALTER LARGE OBJECT large_object_oid
    OWNER TO new_owner;
```

参数说明

表 13-190 ALTER LARGE OBJECT 参数说明

参数	描述	取值范围
large_object_oid	需要修改的大对象的OID。	已存在的大对象名。
OWNER TO	用于修改大对象所有者的子句。	-
new_owner	大对象的新所有者。	已存在的用户名/角色名。

版权所有 © 华为云计算技术有限公司

示例

无。

13.14.5 CREATE BARRIER

功能描述

创建一个新集群节点间的同步点。该同步点可用于数据恢复。

注意事项

在创建同步点之前，首先要确认集群中CN和DN的gtm_backup_barrier和enable_cbm_tracking参数都已设置为on。

语法格式

```
CREATE BARRIER [ barrier_name  ] ;
```

参数说明

barrier_name

可选参数。同步点名称。

取值范围：字符串，需符合**标识符命名规范**。

示例

创建一个同步点，不指定名称。

```
CREATE BARRIER;
```

创建一个同步点并指定其名称为barrier1。

```
CREATE BARRIER 'barrier1';
```


版权所有 © 华为云计算技术有限公司

14 DML 语法

14.1 DML 语法一览表

DML（Data Manipulation Language数据操作语言），用于对数据库表中的数据进行操作。如：插入、更新、查询、删除。

插入数据

插入数据是往数据库表中添加一条或多条记录，请参考**INSERT**。

修改数据

修改数据是修改数据库表中的一条或多条记录，请参考**UPDATE**。

查询数据

数据库查询语句SELECT用于在数据库中检索符合条件的信息，请参考**SELECT**。

删除数据

删除表中指定条件的数据，请参考**DELETE**。

拷贝数据

DWS提供了在表和文件之间拷贝数据的语句，请参考**COPY**。

锁定表

DWS提供了多种锁模式用于控制对表中数据的并发访问，请参考**LOCK**。

调用函数

DWS提供了三个用于调用函数的语句，它们在语法结构上没有差别，请参考**CALL**。

版权所有 © 华为云计算技术有限公司

14.2 CALL

功能描述

使用CALL命令可以调用已定义的函数和存储过程。

注意事项

如果自定义函数的函数名与系统函数同名，调用自定义函数时需指定schema，否则系统会优先调用系统函数。

语法格式

```
CALL [schema.] {func_name| procedure_name} ( param_expr );
```

参数说明

表 14-1 CALL 参数说明

参数	描述	取值范围
schema	函数或存储过程所在的模式名称。	字符串，要符合标识符命名规范。
func_name	所调用函数或存储过程的名称。	已存在的函数名称。
param_expr	参数列表可以用符号“:=”或者“=>”将参数名和参数值隔开，优点是参数可以以任意顺序排列。若参数列表中仅出现参数值，则参数值的排列顺序必须和函数或存储过程定义时的相同。 **说明** 参数可以包含入参（参数名和类型之间指定“IN”关键字）和出参（参数名和类型之间指定“OUT”关键字），使用CALL命令调用函数或存储过程时，对于非重载的函数，参数列表必须包含出参，出参可以传入一个变量或者任一常量，详见**示例**。对于重载的package函数，参数列表里可以忽略出参，忽略出参时可能会导致函数找不到。包含出参时，出参只能是常量。	已存在的函数参数名称或存储过程参数名称。

示例

创建一个函数func_add_sql，计算两个整数的和，并返回结果：

```
CREATE FUNCTION func_add_sql(num1 integer, num2 integer) RETURN integer
AS
BEGIN
RETURN num1 + num2;
END;
/
```


版权所有 © 华为云计算技术有限公司

按参数值传递：

```
CALL func_add_sql(1, 3);
```

使用命名标记法传参：

```
CALL func_add_sql(num1 => 1,num2 => 3);
CALL func_add_sql(num2 := 2, num1 := 3);
```

删除函数：

```
DROP FUNCTION func_add_sql;
```

创建带出参的函数：

```
CREATE FUNCTION func_increment_sql(num1 IN integer, num2 IN integer, res OUT integer)
RETURN integer
AS
BEGIN
res := num1 + num2;
END;
/
```

出参传入常量：

```
CALL func_increment_sql(1,2,1);
```

出参传入变量：

```
DECLARE
res int;
BEGIN
func_increment_sql(1, 2, res);
dbms_output.put_line(res);
END;
/
```

创建重载的函数：

```
create or replace procedure package_func_overload(col int, col2 out int) package
as
declare
    col_type text;
begin
     col := 122;
         dbms_output.put_line('two out parameters ' || col2);
end;
/
create or replace procedure package_func_overload(col int, col2 out varchar)
package
as
declare
    col_type text;
begin
     col2 := '122';
         dbms_output.put_line('two varchar parameters ' || col2);
end;
/
```

函数调用：

```
call package_func_overload(1, 'test');
call package_func_overload(1, 1);
```

删除函数：

```
DROP FUNCTION func_increment_sql;
```

 版权所有 © 华为云计算技术有限公司

14.3 COPY

功能描述

通过COPY命令实现在表和文件之间拷贝数据。COPY FROM从一个文件拷贝数据到一个表，COPY TO把一个表的数据拷贝到一个文件。

COPY FROM/TO适用于低并发，本地小数据量导入导出的场景。

须知

- 由于云上DWS是全托管服务，用户无法登录后台集群节点，所以云上禁止使用COPY语法，可采用“\COPY”命令进行操作（仅gsql客户端支持，其他客户端包括SQL编辑器均不支持“\COPY”命令）。
- 云上推荐将数据文件放到OBS上，使用OBS外表进行入库。

注意事项

- 以安全模式(云上安全模式不支持关闭)启动CN、DN的开关，那么当前模式下禁止使用COPY FROM FILENAME或COPY TO FILENAME语法，可采用\copy的方式进行规避，请参考**如何使用\copy导入导出**中的示例。
- COPY只能用于表，不能用于视图。
- 对任何要插入数据的表必须有插入权限。
- 如果声明了一个字段列表，COPY将只在文件和表之间拷贝已声明字段的数据。如果表中有任何不在字段列表里的字段，COPY FROM将为这些字段插入缺省值。
- 如果声明了数据源文件，服务器必须可以访问该文件；如果指定了STDIN，数据将在客户前端和服务器之间流动，输入时，表的列与列之间使用TAB键分隔，在新的一行中以反斜杠和句点（\.）表示输入结束。
- 如果数据文件的任意行包含比预期多或者少的字段，COPY FROM将抛出一个错误。
- 数据的结束可以用一个只包含反斜杠和句点（\.）的行表示。如果从文件中读取数据，数据结束的标记是不必要的；如果在客户端应用之间拷贝数据，必须要有结束标记。
- COPY FROM中\N为空字符串，如果要输入实际数据值\N ，使用\\N。
- COPY FROM不支持在导入过程中对数据做预处理（比如说表达式运算，填充指定默认值等）。如果需要在导入过程中对数据做预处理，用户需先把数据导入到临时表中，然后执行SQL语句通过运算插入到表中，但此方法会导致I/O膨胀，降低导入性能。
- COPY FROM在遇到数据格式错误时会回滚事务，但没有足够的错误信息，不方便用户从大量的原始数据中定位错误数据。
- 支持使用COPY TO向OBS导出TEXT或CSV格式数据，但不支持使用COPY FROM从OBS导入数据。对于导出到OBS上的带utf8 BOM的CSV格式文件，如果有导入需求，建议使用OBS导入，否则可能无法正确识别BOM字段。

版权所有 © 华为云计算技术有限公司

语法格式

- 从一个文件拷贝数据到一个表：

```
COPY table_name [ ( column_name [, ...] ) ]
    FROM { 'filename' | STDIN }
    [ [ USING ] DELIMITERS 'delimiters' ]
    [ WITHOUT ESCAPING ]
    [ LOG ERRORS ]
    [ LOG ERRORS data ]
    [ REJECT LIMIT 'limit' ]
    [ [ WITH ] ( option [, ...] ) ]
    | copy_option
    | FIXED FORMATTER ( { column_name( offset, length ) } [, ...] ) [ ( option [, ...] ) | copy_option
[ ...] ] ];
```

说明

语法中的FIXED FORMATTER ({ column_name(offset, length) } [, ...])以及 [(option [, ...]) | copy_option [...]] 可以任意排列组合。

- 把一个表的数据拷贝到一个文件：

```
COPY table_name [ ( column_name [, ...] ) ]
    TO { 'filename' | STDOUT }
    [ [ USING ] DELIMITERS 'delimiters' ]
    [ WITHOUT ESCAPING ]
    [ [ WITH ] ( option [, ...] ) ]
    | copy_option
    | FIXED FORMATTER ( { column_name( offset, length ) } [, ...] ) [ ( option [, ...] ) | copy_option
[ ...] ] ];

COPY query
    TO { 'filename' | STDOUT }
    [ WITHOUT ESCAPING ]
    [ [ WITH ] ( option [, ...] ) ]
    | copy_option
    | FIXED FORMATTER ( { column_name( offset, length ) } [, ...] ) [ ( option [, ...] ) | copy_option
[ ...] ] ];
```

说明

- COPY TO语法形式约束如下：

  (query)与[USING] DELIMITER不兼容，即若COPY TO的数据来自于一个query的查询结果，那么COPY TO语法不能再指定[USING] DELIMITERS语法子句。
- 对于FIXED FORMATTER语法后面跟随的copy_option是以空格进行分隔的。
- copy_option是指COPY原生的参数形式，而option是兼容外表导入的参数形式。
- 语法中的FIXED FORMATTER ({ column_name(offset, length) } [, ...])以及 [(option [, ...]) | copy_option [...]] 可以任意排列组合。

参数说明

表 14-2 COPY 参数说明

参数	描述	取值范围
query	将查询结果进行拷贝。	一个必须用圆括号包围的SELECT或VALUES命令。
table_name	表的名字（可以有模式修饰）。	已存在的表名。

版权所有 © 华为云计算技术有限公司

参数	描述	取值范围
column_name	可选的待拷贝字段列表。	如果没有声明字段列表，将使用所有字段。
STDIN	声明输入是来自标准输入。	-
STDOUT	声明输出打印到标准输出。	-
FIXED	打开字段固定长度模式。 ● 在字段固定长度模式下，不能声明DELIMITER，NULL，CSV选项。 ● 指定FIXED类型后，不能再通过option或copy_option指定BINARY、CSV、TEXT等类型。	● 每条记录的每个字段长度相同。 ● 长度不足的字段以空格填充，数字类型字段左对齐，字符字段右对齐。 ● 字段和字段之间没有分隔符。
[USING] DELIMITER 'delimiters'	在文件中分隔各个字段的字符串，分隔符最大长度不超过10个字节。	不允许包含\.abcdefghijklmnopqrstuvwxyz0123456789中的任何一个字符。 在文本模式下，缺省是水平制表符，在CSV模式下是一个逗号。
WITHOUT ESCAPING	在TEXT格式中，不对'\'和后面的字符进行转义。	仅支持TEXT格式。

版权所有 © 华为云计算技术有限公司

参数	描述	取值范围
LOG ERRORS	若指定该参数，则开启对于COPY FROM语句中数据类型错误的容错机制，相关错误行的错误记录会记录到此库中的public.pgxc_copy_error_log表中，备后续查阅。 **须知** 此容错选项的使用限制如下： • 此容错机制仅捕捉COPY FROM过程中CN节点上数据解析过程中相关的数据类型错误（DATA_EXCEPTION），诸如CN与DN之间的网络交互错误或者是DN上的表达式转换错误等CN数据解析逻辑之外的过程无法涵盖在内。 • 在每个库第一次使用时COPY FROM容错时，请先行检查public.pgxc_copy_error_log（COPY错误表）是否存在，若不存在请调用copy_error_log_create() 函数创建；若存在，请转移此表数据并删除这张表后，调用copy_error_log_create() 函数创建。更多关于表public.pgxc_copy_error_log的字段信息，请参见**表7-47**。 • 若在指定了LOG ERRORS的COPY FROM运行时，public.pgxc_copy_error_log不存在（未创建或者已删除）或表定义不符合copy_error_log_create() 中的预设表定义，则会报错。因此请确定此COPY错误表是使用copy_error_log_create() 函数创建的，否则可能导致容错的COPY FROM语句无法正常执行。 • COPY已有的容错选项（如IGNORE_EXTRA_DATA）开启时，对应类型的错误会按照已有的方式处理而不会报出异常，因此错误表也不会有相应数据。 • 此容错机制的覆盖范围与GDS Foreign Table的容错范围相同。推荐用户通过表名列以及COPY FROM语句开始时间戳对查询结果进行过滤。错误数据处理可参考处理错误表章节。	仅支持导入（即COPY FROM）时指定。
LOG ERRORS DATA	LOG ERRORS DATA和LOG ERRORS的区别： • LOG ERRORS DATA会登记错误行数据信息（rawrecord）。 • 使用该参数时，若错误内容过于复杂可能存在写入容错表失败的风险，导致任务失败。	-

版权所有 © 华为云计算技术有限公司

参数	描述	取值范围
REJECT LIMIT 'limit'	与LOG ERROR选项共同使用，对COPY FROM的容错机制设置数值上限，一旦此COPY FROM语句错误数据超过选项指定条数，则会按照原有机制报错。 **说明** LOG ERRORS中描述的容错机制，REJECT LIMIT的计数也是按照执行COPY FROM的CN上遇到的解析错误数量计算，而不是每个DN上的错误数量，请与GDS容错机制区别开。	正整数（1-INTMAX），'unlimited'（无最大值限制） 缺省值： • 若未指定LOG ERRORS，则会报错； • 若指定LOG ERRORS，则默认为0。
FORMATTER	在固定长度模式中（即声明了FIXED参数），定义每一个字段在数据文件中的位置。按照column(offset,length)格式定义每一列在数据文件中的位置。	• offset取值不能小于0，以字节为单位。 • length取值不能小于0，以字节为单位。 • 所有列的总长度和不能大于1GB。 • 文件中没有出现的列默认以空值代替。
OPTION { option_name ' value ' }	用于指定兼容外表的各类参数。	具体请参见**表14-3**。
COPY_OPTION { option_name ' value ' }	用于指定COPY原生的各类参数。	具体请参见**表14-4**。

版权所有 © 华为云计算技术有限公司

表 14-3 OPTION 参数详情

参数	描述	取值范围
FORMAT	数据源文件的格式。	支持CSV、TEXT、FIXED、BINARY类型的文件。 ● CSV格式的文件，可以有效处理数据列中的换行符，但对一些特殊字符处理有欠缺。 ● TEXT格式的文件，可以有效处理一些特殊字符，但无法正确处理数据列中的换行符。 ● FIXED格式的文件，适用于每条数据的数据列都比较固定的数据，长度不足的列会添加空格补齐，过长的列则会自动截断。 ● BINARY形式的选项会使得所有的数据被存储/读作二进制格式而不是文本。 这比TEXT和CSV格式的要快一些，但是一个BINARY格式文件可移植性比较差。 缺省值：TEXT
OIDS	为每行拷贝内部对象标识（ oid ）。 **说明** 若COPY FROM对象为query或者对于没有oid的表，指定oids标识报错。	● true/on ● false/off 缺省值：false

版权所有 © 华为云计算技术有限公司

参数	描述	取值范围
DELIMITER	指定数据文件行数据的字段分隔符。 ● 分隔符不能是\r和\n。 ● 分隔符不能和null参数相同，CSV格式数据的分隔符不能和quote参数相同。 ● TEXT格式数据的分隔符不能包含： 小写字母、数字和特殊字符.\。 ● 数据文件中单行数据长度需<1GB，如果分隔符较长且数据列较多的情况下，会影响导出有效数据的长度。 ● 分隔符推荐使用多字符和不可见字符。多字符例如'$^&'；不可见字符例如0x07，0x08，0x1b等。 ● 建议TEXT格式下多字符分隔符中的字符不要完全相同，例如不建议使用delimiter '---' 。	支持多字符分隔符，但分隔符不能超过10个字节。 缺省值： ● TEXT格式的默认分隔符是水平制表符（ tab ）。 ● CSV格式的默认分隔符为“,”。 ● FIXED格式没有分隔符。
NULL	用来指定数据文件中空值的表示方式。	● null值不能是\r和\n，最大为100个字符。 ● null值不能和分隔符、quote参数相同。 缺省值： ● CSV格式下默认值是一个没有引号的空字符串。 ● 在TEXT格式下默认值是\N。

 版权所有 © 华为云计算技术有限公司

参数	描述	取值范围
HEADER	指定导出数据文件是否包含标题行，标题行一般用来描述表中每个字段的信息。header只能用于CSV，FIXED格式的文件中。 • 在导入数据时，如果header选项为on，则数据文件中第一行会被识别为标题行，会忽略此行。如果header为off，数据文件中第一行会被识别为数据。 • 在导出数据时，如果header选项为on，则需要指定fileheader。如果header为off，则导出数据文件不包含标题行。	• true/on • false/off 缺省值：false
QUOTE	CSV格式文件下，指定一个数据值被引用时使用的引用字符。	缺省值：双引号 • quote参数不能和分隔符、null参数相同。 • quote参数只能是单字节的字符。 • 推荐不可见字符作为quote，例如0x07，0x08，0x1b等。 • 导出时，如果字段值中含有分隔符（delimiter）、换行符（eol）、quote字符本身时，会自动使用quote包裹，使字段内容不会被误解析为语法字符。
ESCAPE	CSV格式下，用来指定**逃逸字符**，逃逸字符只能指定为单字节字符。	单字节字符。 缺省值：双引号

版权所有 © 华为云计算技术有限公司

参数	描述	取值范围
EOL 'newline_character'	指定导入导出数据文件换行符样式。	支持多字符换行符，但换行符不能超过10个字节。 常见的换行符，如\r、\n、\r\n（设成0x0D、0x0A、0x0D0A效果是相同的），其他字符或字符串，如$、#。 ● EOL参数只能用于TEXT格式的导入导出，不支持CSV格式和FIXED格式导入。为了兼容原有EOL参数，仍然支持导出CSV格式和FIXED格式时指定EOL参数为0x0D或0x0D0A。 ● EOL参数不能和分隔符、null参数相同。 ● EOL参数不能包含：.abcdefghijklmnopqrstuvwxyz0123456789。
FORCE_QUOTE { (column_name [, ...]) | * }	在CSV COPY TO模式下，强制对每个声明的字段的所有非NULL值都使用引号包围。NULL输出不会被引号包围。	已存在的字段。
FORCE_NOT_NULL (column_name [, ...])	在CSV COPY FROM模式下，指定的字段输入不能为空。	已存在的字段。
ENCODING	指定数据文件的编码格式名称。	缺省为当前数据库编码格式。 常用的编码格式有UTF8、GBK、GB18030，GB18030有两个版本，GB18030和GB18030_2022。GB18030_2022，用于支持汉字最新的国家标准**GB 18030-2022**。
IGNORE_EXTRA_DATA	若数据源文件比外表定义列数多，是否会忽略对多出的列。该参数只在数据导入过程中使用。	● 参数为true/on，若数据源文件比外表定义列数多，则忽略行尾多出来的列。 ● 参数为false/off，若数据源文件比外表定义列数多，会显示如下错误信息。 `extra data after last expected column` 缺省值：false/off **须知** 如果行尾换行符丢失，使两行变成一行时，设置此参数为true将导致后一行数据被忽略掉。

版权所有 © 华为云计算技术有限公司

参数	描述	取值范围
COMPATIBLE_ILLEGAL_CHARS	导入非法字符容错参数，仅对COPY FROM导入有效。 此参数不支持BINARY格式，会报“cannot specify bulkload compatibility options in BINARY mode”错误信息。 导入非法字符容错规则如下： • 对于'\0'，容错后转换为空格； • 对于其他非法字符，容错后转换为问号； • 如果compatible_illegal_chars为true/on，表示导入时对于非法字符进行容错处理，若NULL、DELIMITER、QUOTE、ESCAPE设置为空格或问号则会通过如“illegal chars conversion may confuse COPY escape 0x20”等报错信息提示用户修改可能引起混淆的参数以避免导入错误。	• 参数为true/on，则导入时遇到非法字符进行容错处理，非法字符转换后入库，不报错，不中断导入。 • 参数为false/off，导入时遇到非法字符进行报错，中断导入。 缺省值：false/off
FILL_MISSING_FIELDS	当数据导入时，若数据源文件中某一行的末尾字段（一个或多个）缺失的处理方式。	true/on，false/off。缺省值为false/off。 • true/on，表示末尾字段缺失，则将末尾字段的值设置为NULL，不报错。 • false/off，表示末尾字段缺失会直接报错，报错信息类似如下： `missing data for column "tt"`
DATE_FORMAT	指定导入DATE类型数据的格式。仅对COPY FROM导入有效。 此参数不支持BINARY格式，会报“cannot specify bulkload compatibility options in BINARY mode”错误信息。	有效的DATE格式。可参考**时间日期函数和操作符**。 **须知** 对于指定为ORACLE兼容类型的数据库（**建库时指定DBCOMPATIBILITY取值为“ORA”**），则DATE类型内建为TIMESTAMP类型。在导入的时候，若需指定格式，可以参考下面的timestamp_format参数。

版权所有 © 华为云计算技术有限公司

参数	描述	取值范围
TIME_FORMAT	指定导入TIME类型数据的格式。此参数仅对COPY FROM导入有效。 此参数不支持BINARY格式，会报“cannot specify bulkload compatibility options in BINARY mode”错误信息。	有效的TIME格式，不支持时区。可参考**时间日期函数和操作符**。
TIMESTAMP_FORMAT	指定导入TIMESTAMP类型数据的格式。此参数仅对COPY FROM导入有效。 此参数不支持BINARY格式，会报“cannot specify bulkload compatibility options in BINARY mode”错误信息。	有效的TIMESTAMP格式，不支持时区。可参考**时间日期函数和操作符**。
SMALLDATETIME_FORMAT	指定导入SMALLDATETIME类型数据的格式。此参数仅对COPY FROM导入有效。 此参数不支持BINARY格式，会报“cannot specify bulkload compatibility options in BINARY mode”错误信息。	有效的SMALLDATETIME格式。可参考**时间日期函数和操作符**。
SERVER	指定OBS SERVER，仅在COPY TO时生效。指定该参数时filename为OBS上的路径，表示导出到OBS。	已创建的OBS SERVER名。
BOM	标识是否对导出的CSV格式文件添加utf8 BOM字段。仅在COPY TO并且指定了有效SERVER参数时生效，仅支持导出文件为utf8编码。	• true/on • false/off 缺省值：false
MAXROW	标识导出文件的最大行数，超出该值会生成新的文件。仅在COPY TO并且指定了有效SERVER参数时生效。当HEADER取值为true时MAXROW需要大于1。该参数需要与FILEPREFIX同时指定。	1~ 2147483647。

版权所有 © 华为云计算技术有限公司

参数	描述	取值范围
FILEPREFIX	指定导出文件名的前缀。仅在COPY TO并且指定了有效SERVER参数时生效。该参数需要与MAXROW同时指定。	有效的字符串，且不能以/开始或结尾。
PRESERVE_BLANKS	用于定长导入时，控制每列数据拖尾的空白字符（包括空格，\t，\v，\f）是否保留。该参数仅8.2.0.100及以上集群版本支持。	• 参数为true/on，表示针对BYTEAOID、CHAROID、NAMEOID、TEXTOID、BPCHAROID、VARCHAROID、NVARCHAR2OID、CSTRINGOID对应类型的列数据，保留拖尾 \t，不进行修剪。 • 参数为false/off，表示针对所有数据类型的数据拖尾的空白字符（包括空格，\t，\v，\f）进行修剪。 缺省值为false/off。

表 14-4 COPY_OPTION 参数详情

参数	描述	取值范围
OIDS	为每行拷贝内部对象标识（oid）。 若COPY FROM对象为query或者对于没有oid的表，指定oids标识会报错。	-
NULL null_string	用来指定数据文件中空值的表示。 **须知** 在使用COPY FROM的时候，任何匹配这个字符串的字符串将被存储为NULL值，所以应该确保指定的字符串和COPY TO相同。	• null值不能是\r和\n，最大为100个字符。 • null值不能和分隔符、quote参数相同。 缺省值： • 在TEXT格式下默认值是\N。 • CSV格式下默认值是一个没有引号的空字符串。
HEADER	指定导出数据文件是否包含标题行，标题行一般用来描述表中每个字段的信息。header只能用于CSV，FIXED格式的文件中。	• 导入数据时，如果header选项为on，则数据文件中第一行会被识别为标题行，会忽略此行。如果header为off，数据文件中第一行会被识别为数据。 • 导出数据时，如果header选项为on，则需要指定fileheader。如果header为off，则导出数据文件不包含标题行。

版权所有 © 华为云计算技术有限公司

参数	描述	取值范围
FILEHEADER	导出数据时用于定义标题行的文件，一般用来描述每一列的数据信息。 **须知** • 仅在header为on或true的情况下有效。 • fileheader指定的是绝对路径。 • 该文件只能包含一行标题信息，并以换行符结尾，多余的行将被丢弃（标题信息不能包含换行符）。 • 该文件包括换行符在内长度不超过1M。	-
FREEZE	将COPY加载的数据行设置为已经被frozen，就等同于这些数据行执行过VACUUM FREEZE。 该参数为一个初始数据加载的性能选项。仅当以下三个条件同时满足时，数据行才会被frozen： • 在同一事务中create或truncate这张表之后执行COPY。 • 当前事务中没有打开的游标。 • 当前事务中没有原有的快照。 **须知** COPY完成后，所有其他会话将会立刻识别到这些数据。但这违反了MVCC可见性的一般原则，会导致潜在的风险。	-
FORCE NOT NULL column_name [, ...]	在CSV COPY FROM模式下，指定的字段输入不能为空。	已存在的字段。
FORCE QUOTE { column_name [, ...] | * }	在CSV COPY TO模式下，强制在每个声明的字段周围对所有非NULL值都使用引号包围。NULL输出不会被引号包围。	已存在的字段。

 版权所有 © 华为云计算技术有限公司

参数	描述	取值范围
BINARY	使用二进制格式存储和读取，而不是以文本的方式。在二进制模式下，不能声明DELIMITER，NULL，CSV选项。指定BINARY类型后，不能再通过option或copy_option指定CSV、FIXED、TEXT等类型。	-
CSV	打开逗号分隔变量（CSV）模式。指定CSV类型后，不能再通过option或copy_option指定BINARY、FIXED、TEXT等类型。	-
QUOTE [AS] 'quote_character'	CSV格式文件下的引号字符。 ● quote参数不能和分隔符、null参数相同。 ● quote参数只能是单字节的字符。 ● 推荐不可见字符作为quote，例如0x07，0x08，0x1b等。	缺省值：双引号。
ESCAPE [AS] 'escape_character'	CSV格式下，用来指定逃逸字符，逃逸字符只能指定为单字节字符。	缺省值为双引号。当与quote值相同时，会被替换为'\0'。
EOL 'newline_character'	指定导入导出数据文件换行符样式。	支持多字符换行符，但换行符不能超过10个字节。 常见的换行符，如\r、\n、\r\n（设成0x0D、0x0A、0x0D0A效果是相同的），其他字符或字符串，如$、#。 ● EOL参数只能用于TEXT格式的导入导出，不支持CSV格式和FIXED格式。为了兼容原有EOL参数，仍然支持导出CSV格式和FIXED格式时指定EOL参数为0x0D或0x0D0A。 ● EOL参数不能和分隔符、null参数相同。 ● EOL参数不能包含：.abcdefghijklmnopqrstuvwxyz0123456789。
ENCODING 'encoding_name'	指定文件编码格式名称。	有效的编码格式。 缺省值：当前编码格式。

版权所有 © 华为云计算技术有限公司

参数	描述	取值范围
IGNORE_EXTRA_DATA	指定当数据源文件比外表定义列数多时，忽略行尾多出来的列。该参数只在数据导入过程中使用。 若不使用该参数，在数据源文件比外表定义列数多，会显示如下错误信息。 `extra data after last expected column`	-
COMPATIBLE_ILLEGAL_CHARS	指定导入时对非法字符进行容错处理，非法字符转换后入库。不报错，不中断导入。该参数不支持BINARY格式，会报“cannot specify bulkload compatibility options in BINARY mode”错误信息。仅对COPY FROM导入有效。 若不使用该参数，导入时遇到非法字符进行报错，中断导入。 导入非法字符容错规则如下： 1. 对于'\0'，容错后转换为空格； 2. 对于其他非法字符，容错后转换为问号； 3. 若compatible_illegal_chars为true/on，表示导入时对于非法字符进行容错处理，若NULL、DELIMITER、QUOTE、ESCAPE设置为空格或问号则会通过如“illegal chars conversion may confuse COPY escape 0x20”等报错信息提示用户修改可能引起混淆的参数，以避免导入错误。	-

 版权所有 © 华为云计算技术有限公司

参数	描述	取值范围
FILL_MISSING_FIELDS	当数据导入时，若数据源文件中某一行的末尾字段（一个或多个）缺失的处理方式。 目前COPY指定此option实际不会生效，即不会有相应的容错处理效果。请不要指定此选项，原因如下： 打开此选项会导致解析器在CN数据解析阶段（即COPY错误表容错的涵盖范围）忽略此数据问题，而到DN重新报错，从而使得COPY错误表（打开LOG ERRORS REJECT LIMIT）在此选项打开的情况下无法成功捕获这类少列的数据异常。	• true/on • false/off 缺省值：false/off。
DATE_FORMAT 'date_format_string'	导入对于DATE类型指定格式。此参数不支持BINARY格式，会报“cannot specify bulkload compatibility options in BINARY mode”错误信息。此参数仅对COPY FROM导入有效。 **说明** 对于指定为ORACLE兼容类型的数据库（**建库时指定DBCOMPATIBILITY取值为“ORA”**），则DATE类型内建为TIMESTAMP类型。在导入的时候，若需指定格式，可以参考下面的timestamp_format参数。	有效的DATE格式。可参考**时间日期函数和操作符**。
TIME_FORMAT 'time_format_string'	导入对于TIME类型指定格式。此参数仅对COPY FROM导入有效。 此参数不支持BINARY格式，会报“cannot specify bulkload compatibility options in BINARY mode”错误信息。	有效的TIME格式，不支持时区。可参考**时间日期函数和操作符**。

版权所有 © 华为云计算技术有限公司

参数	描述	取值范围
TIMESTAMP_FORMAT 'timestamp_format_string'	导入对于TIMESTAMP类型指定格式。此参数仅对COPY FROM导入有效。 此参数不支持BINARY格式，会报“cannot specify bulkload compatibility options in BINARY mode”错误信息。	有效的TIMESTAMP格式，不支持时区。可参考**时间日期函数和操作符**。
SMALLDATETIME_FORMAT 'smalldatetime_format_string'	导入对于SMALLDATETIME类型指定格式。此参数仅对COPY FROM导入有效。 此参数不支持BINARY格式，会报“cannot specify bulkload compatibility options in BINARY mode”错误信息。	有效的SMALLDATETIME格式。可参考**时间日期函数和操作符**。
PRESERVE_BLANKS	用于定长导入时，控制每列数据拖尾的空白字符（包括空格，\t，\v，\f）是否保留。该参数仅8.2.0.100及以上集群版本支持。	• 参数为true/on，表示针对BYTEAOID、CHAROID、NAMEOID、TEXTOID、BPCHAROID、VARCHAROID、NVARCHAR2OID、CSTRINGOID对应类型的列数据，保留拖尾 \t，不进行修剪。 • 参数为false/off，表示针对所有数据类型的数据拖尾的空白字符（包括空格，\t，\v，\f）进行修剪。 缺省值为false/off。

COPY FROM能够识别的特殊反斜杠序列如下所示。

表 14-5 特殊反斜杠序列

反斜杠序列	含义
\b	反斜杠 （ ASCII 8 ）
\f	换页（ ASCII 12 ）
\n	换行符 （ ASCII 10 ）
\r	回车符 （ ASCII 13 ）
\t	水平制表符 （ ASCII 9 ）
\v	垂直制表符 （ ASCII 11 ）
\digits	反斜杠后面跟着一到三个八进制数，表示ASCII值为该数的字符。

版权所有 © 华为云计算技术有限公司

反斜杠序列	含义
\xdigits	反斜杠x后面跟着一个或两个十六进制位声明指定数值编码的字符。

示例（仅 gsql 客户端支持，其他客户端不支持）

准备数据：

```
DROP SCHEMA IF EXISTS tpcds CASCADE;
CREATE SCHEMA tpcds;
DROP TABLE IF EXISTS tpcds.ship_mode;
CREATE TABLE  tpcds.ship_mode
(
    SM_SHIP_MODE_SK           INTEGER               NOT NULL,
    SM_SHIP_MODE_ID           CHAR(16)              NOT NULL,
    SM_TYPE                   CHAR(30)                      ,
    SM_CODE                   CHAR(10)                      ,
    SM_CARRIER                CHAR(20)                      ,
    SM_CONTRACT               CHAR(20)
)
WITH (ORIENTATION = COLUMN,COMPRESSION=MIDDLE)

DISTRIBUTE BY HASH(SM_SHIP_MODE_SK );

INSERT INTO tpcds.ship_mode
VALUES (1, 'SHIP_MODE_001', 'AIR TRANSPORT', 'AIR001', 'FEDEX', 'FULL_CONTRACT');
```

将tpcds.ship_mode中的数据拷贝到/opt/ds_ship_mode.dat文件中：

```
\COPY tpcds.ship_mode TO '/opt/ds_ship_mode.dat';
```

将tpcds.ship_mode中的查询结果拷贝到/opt/ds_ship_mode.dat文件中：

```
\COPY (SELECT * FROM tpcds.ship_mode WHERE SM_SHIP_MODE_SK <> null) TO '/opt/ds_ship_mode.dat';
```

将tpcds.ship_mode输出到stdout：

```
\COPY tpcds.ship_mode TO stdout;
```

创建tpcds.ship_mode_t1表：

```
CREATE TABLE tpcds.ship_mode_t1
(
    SM_SHIP_MODE_SK           INTEGER               NOT NULL,
    SM_SHIP_MODE_ID           CHAR(16)              NOT NULL,
    SM_TYPE                   CHAR(30)                      ,
    SM_CODE                   CHAR(10)                      ,
    SM_CARRIER                CHAR(20)                      ,
    SM_CONTRACT               CHAR(20)
)
WITH (ORIENTATION = COLUMN,COMPRESSION=MIDDLE)
DISTRIBUTE BY HASH(SM_SHIP_MODE_SK );
```

从stdin拷贝数据到表tpcds.ship_mode_t1：

```
\COPY tpcds.ship_mode_t1 FROM stdin;
```

从/opt/ds_ship_mode.dat文件拷贝数据到表tpcds.ship_mode_t1：

```
\COPY tpcds.ship_mode_t1 FROM '/opt/ds_ship_mode.dat';
```

从/opt/ds_ship_mode.dat文件拷贝数据到表tpcds.ship_mode_t1，使用参数如下：导入格式为TEXT（format 'text'），分隔符为'\t'（delimiter E'\t'），忽略多余列（ignore_extra_data 'true'），不指定转义（noescaping 'true'）：

 版权所有 © 华为云计算技术有限公司

```
\COPY tpcds.ship_mode_t1 FROM '/opt/ds_ship_mode.dat' WITH(format 'text', delimiter E'\t',
ignore_extra_data 'true', noescaping 'true');
```

从/opt/ds_ship_mode.dat文件拷贝数据到表tpcds.ship_mode_t1，使用参数如下：导入格式为FIXED（FIXED），指定定长格式（FORMATTER(SM_SHIP_MODE_SK(0, 2), SM_SHIP_MODE_ID(2,16), SM_TYPE(18,30), SM_CODE(50,10), SM_CARRIER(61,20), SM_CONTRACT(82,20)）），忽略多余列（ignore_extra_data），有数据头（header）：

```
\COPY tpcds.ship_mode_t1 FROM '/opt/ds_ship_mode.dat' FIXED FORMATTER(SM_SHIP_MODE_SK(0, 2),
SM_SHIP_MODE_ID(2,16), SM_TYPE(18,30), SM_CODE(50,10), SM_CARRIER(61,20), SM_CONTRACT(82,20))
header ignore_extra_data;
```

从/opt/ds_ship_mode.dat文件拷贝数据到表tpcds.ship_mode_t1，使用参数如下：导入格式为FIXED（FIXED），指定定长格式（FORMATTER(SM_SHIP_MODE_SK(0, 2), SM_SHIP_MODE_ID(2,16), SM_TYPE(18,30), SM_CODE(50,10), SM_CARRIER(61,20), SM_CONTRACT(82,20)）），忽略多余列（ignore_extra_data），有数据头（header），保留末尾\t字符：

```
\COPY tpcds.ship_mode_t1 FROM '/opt/ds_ship_mode.dat' (FORMAT 'fixed', FORMATTER
(SM_SHIP_MODE_SK(0,2), SM_SHIP_MODE_ID(2,16), SM_TYPE(18,30), SM_CODE(50,10),
SM_CARRIER(61,20), SM_CONTRACT(82,20)), PRESERVE_BLANKS 'true', HEADER 'true',
IGNORE_EXTRA_DATA 'true');
```

将tpcds.ship_mode_t1导出为OBS的 '/bucket/path/'路径上的TEXT格式文件ds_ship_mode.dat。需要指定包含OBS访问信息的“server”option参数：

```
COPY tpcds.ship_mode_t1 TO '/bucket/path/ds_ship_mode.dat' WITH (format 'text', encoding 'utf8', server
'obs_server');
```

将tpcds.ship_mode_t1导出为OBS的 '/bucket/path/'路径上的CSV格式文件。需要指定包含OBS访问信息的“server”option参数。其中文件包含标题行，包含BOM头，单文件最大行数1000行（超出1000行生成新的文件），自定义文件名前缀为“justprefix”：

```
COPY (select * from tpcds.ship_mode_t1 where SM_SHIP_MODE_SK=1060) TO '/bucket/path/' WITH (format
'csv', header 'on', encoding 'utf8', server 'obs_server', bom 'on', maxrow '1000', fileprefix 'justprefix');
```

删除tpcds.ship_mode_t1：

```
DROP TABLE tpcds.ship_mode_t1;
```

常见问题

使用COPY FROM导入时报错“invalid byte sequence for encoding "UTF8": 0x00”

14.4 DELETE

功能描述

从指定的表里删除满足WHERE子句的行。如果WHERE子句不存在，将删除表中所有行，结果只保留表结构。

注意事项

- 要删除表中的数据，用户必须对它有DELETE权限。同样也必须有USING子句引用的表以及condition上读取的表的SELECT权限。

版权所有 © 华为云计算技术有限公司

- 如果需要删除表上的所有数据，建议使用TRUNCATE语法，可以有效提高性能同时减少空间膨胀。
- 对于复制表，仅支持两种场景下的DELETE操作：
  a. 有主键约束的场景。
  b. 执行计划能下推的场景。
- 对于列存表，暂时不支持RETURNING子句。
- HStore表上的单条DELETE操作，会往Delta中插入一条type是D的记录，同时再更新内存更新链用于管理并发。
- HStore表上的批量DELETE操作，对于每个CU上的连续DELETE，会插入一条type是D的记录。
- 对于并发DELETE场景， 传统列存储格式由于同时操作相同CU时会阻塞所以并发性能较差， 对于HStore表由于不需要阻塞等待，并发DELETE性能可达到列存的百倍以上。

⚠ 警告

- 避免使用UPDATE/DELETE大批量刷新和删除数据，考虑使用TRUNCATE PARTITION/DROP PARTITION代替。
- 更多开发设计规范参见**总体开发设计规范**。

语法格式

```
[ WITH [ RECURSIVE ] with_query [, ...] ]
    DELETE [/*+ plan_hint */] FROM [ ONLY ] table_name [ * ] [ [ AS ] alias ]
    [ PARTITION ( partition_name ) | PARTITION FOR ( partition_key_value [, ...] ) ]
    [ USING using_list ]
    [ WHERE condition | WHERE CURRENT OF cursor_name ]
    [ RETURNING { * | { output_expr [ [ AS ] output_name ] } [, ...] } ];
```


版权所有 © 华为云计算技术有限公司

参数说明

表 14-6 DELETE 参数说明

参数	描述	取值范围
WITH [RECURSIVE] with_query [, ...]	用于声明一个或多个可以在主查询中通过名字引用的子查询，相当于临时表。 如果声明了RECURSIVE，则允许SELECT子查询通过名字引用它自己。 其中with_query的详细格式为： `with_query_name [ ( column_name [, ...] ) ] AS ( {select | values | insert | update | delete} )` • with_query_name，指定子查询生成的结果集名字，在查询中可使用该名称访问子查询的结果集。 • column_name，指定子查询结果集中显示的列名。 • 每个子查询可以是SELECT、VALUES、INSERT、UPDATE或DELETE语句。	-
plan_hint子句	以/*+ */的形式在关键字后，用于对指定语句块生成的计划进行hint调优，详细用法请参见**使用Plan Hint进行调优**。	-
ONLY	如果指定ONLY，则只有指定表被删除；如果没有指定ONLY，则指定表及其所有继承表都会被删除。 该参数仅为了兼容PG保留，DWS不支持继承表。	-
table_name	目标表的名字（可以有模式修饰）。	已存在的表名。
alias	目标表的别名。	字符串，符合标识符命名规范。
partition_name	分区名。仅8.2.1及以上集群版本支持该选项。	已存在的分区名。
partition_key_value	分区键值。 通过PARTITION FOR (partition_key_value [, ...])子句指定的这一组值，可以唯一确定一个分区。	-
using_list	using子句。	-
condition	一个返回boolean值的表达式，用于判断哪些行需要被删除。	-
WHERE CURRENT OF cursor_name	当前不支持，仅保留语法接口。	-

版权所有 © 华为云计算技术有限公司

参数	描述	取值范围
output_expr	DELETE命令删除行之后计算输出结果的表达式。该表达式可以使用表的任意字段。可以使用*返回被删除行的所有字段。	-
output_name	一个字段的输出名称。	字符串，符合标识符命名规范。

示例

创建表tpcds.customer_address：

```
DROP SCHEMA IF EXISTS tpcds CASCADE;
CREATE SCHEMA tpcds;
CREATE TABLE tpcds.customer_address
(
    ca_address_sk             bigint               not null ,
    ca_address_id             char(16)               not null,
    ca_street_number            char(10)                       ,
    ca_street_name            varchar(60)                    ,
    ca_street_type            char(15)                     ,
    ca_suite_number             char(10)                     ,
    ca_city                   varchar(60)                  ,
    ca_county                   varchar(30)                  ,
    ca_state                  char(2)                      ,
    ca_zip                    char(10)                     ,
    ca_country                  varchar(20)                    ,
    ca_gmt_offset             decimal(5,2)                   ,
    ca_location_type          char(20)
)
 with (orientation = column)
distribute by hash (ca_address_sk);
```

删除tpcds.customer_address中ca_address_sk小于14888的职员：

```
DELETE FROM tpcds.customer_address WHERE ca_address_sk < 14888;
```

删除tpcds.customer_address中ca_address_sk为14891，14893和14895的职员：

```
DELETE FROM tpcds.customer_address WHERE ca_address_sk in (14891,14893,14895);
```

删除tpcds.customer_address中所有数据：

```
DELETE FROM tpcds.customer_address;
```

先通过子查询（删除行存表tpcds.warehouse_t30）得到一张临时表temp_t，然后查询临时表temp_t中的所有数据：

```
CREATE TABLE tpcds.warehouse_t30 (a int, b int);
WITH temp_t AS (DELETE FROM tpcds.warehouse_t30 RETURNING *) SELECT * FROM temp_t ORDER BY 1;
```

删除分区表test_range_row的指定分区p1：

```
DROP TABLE IF EXISTS test_range_row;
CREATE TABLE test_range_row(a int, d int)
DISTRIBUTE BY hash(a) PARTITION BY RANGE(d)
(
    PARTITION p1 values LESS THAN (60),
    PARTITION p2 values LESS THAN (75),
    PARTITION p3 values LESS THAN (90),
    PARTITION p4 VALUES LESS THAN (maxvalue)
);
INSERT OVERWRITE INTO test_range_row PARTITION(p1) VALUES(55,51);
```


版权所有 © 华为云计算技术有限公司

```
INSERT OVERWRITE INTO test_range_row PARTITION(p3) VALUES(85,80);

DELETE FROM test_range_row PARTITION(p1);
```

14.5 EXPLAIN

功能描述

EXPLAIN命令输出SQL语句的执行计划，通过分析执行计划来定位慢SQL，优化SQL语句。

执行计划将显示SQL语句所引用的表采用的扫描方式，如：简单的顺序扫描、索引扫描等。如果引用了多个表，执行计划还会显示使用的JOIN算法。执行计划的最关键部分是语句的预计执行开销，即计划生成器估算执行该语句将花费多长的时间。

指定ANALYZE选项（ EXPLAIN ANALYZE ），不仅显示执行计划，还会实际执行查询，并提供每个步骤的实际执行时间。这有助于验证执行计划的准确性，并评估实际性能。

详细的SQL调优介绍及指导请参考**SQL调优**。

注意事项

在指定ANALYZE选项时，语句会被执行。如果用户使用EXPLAIN分析INSERT，UPDATE，DELETE，CREATE TABLE AS或EXECUTE语句，但不改动数据（ 执行这些语句会影响数据 ），可使用以下所示的方法：

```
START TRANSACTION;
EXPLAIN ANALYZE ...;
ROLLBACK;
```

语法格式

- 显示SQL语句的执行计划，支持多种选项，对选项顺序无要求：
  ```
  EXPLAIN [ (  option  [, ...] ) ] statement;
  ```
  其中选项option子句的语法为：
  ```
  ANALYZE [ boolean ] |
     ANALYSE [ boolean ] |
     VERBOSE [ boolean ] |
     COSTS [ boolean ] |
     CPU [ boolean ] |
     DETAIL [ boolean ] |
     NODES [ boolean ] |
     NUM_NODES [ boolean ] |
     BUFFERS [ boolean ] |
     TIMING [ boolean ] |
     PLAN [ boolean ] |
     FORMAT { TEXT | XML | JSON | YAML } |
     BLOCKNAME [boolean] |
     OUTLINE [boolean]  |
     WARMUP |
     WARMUP HOT |
     SUMMARY [boolean] |
     SETTINGS [boolean]
  ```
- 显示SQL语句的执行计划，且要按顺序给出选项：
  ```
  EXPLAIN  { [  { ANALYZE  | ANALYSE  }  ] [ VERBOSE  ]  | PERFORMANCE  } statement;
  ```
- 显示复现SQL语句的执行计划所需的信息，通常用于定位问题。STATS选项必须单独使用：
  ```
  EXPLAIN ( STATS [ boolean ] ) statement;
  ```

 版权所有 © 华为云计算技术有限公司

- 显示DDL语句执行步骤的详细耗时信息。（该语法仅9.1.0及以上集群版本支持）
  ```
  EXPLAIN PERFORMANCE statement;
  ```
  支持的DDL语句有CREATE、CREATE INDEX、DROP、VACUUM FULL、ANALYZE及COPY。
- 执行预查询，并将预查询的数据缓存到本地磁盘，提升实际查询时的查询速度。（该语法仅9.1.0.200及以上集群版本支持）
  ```
  EXPLAIN WARMUP statement;
  EXPLAIN WARMUP HOT statement;
  ```

参数说明

表 14-7 EXPLAIN 参数说明

参数	描述	取值范围
statement	需要分析的SQL语句。	-

ANALYZE boolean | ANALYSE boolean	指定ANALYZE选项时，会实际执行SQL，根据实际的运行结果显示统计数据，包括每个计划节点内时间总开销（毫秒为单位）和实际返回的总行数。
	如果使用EXPLAIN分析INSERT，UPDATE，DELETE，CREATE TABLE AS或EXECUTE语句，但不改动数据（执行这些语句会影响数据），可以放在一个事务中，分析完成后可以直接回滚：
	`START TRANSACTION;`
	`EXPLAIN ANALYZE ...;`
	`ROLLBACK;`
		● TRUE：显示实际运行时间和其他统计数据。
		● FALSE：不显示。
		缺省值：TRUE

VERBOSE boolean	是否显示查询计划的附加信息。 附加信息包括查询计划中每个节点输出的列（Output），表的SCHEMA信息，函数的模式信息，表达式中列所属表的别名，被触发的触发器名称等。	● TRUE：显示附加信息。 ● FALSE：不显示。 缺省值：TRUE
COSTS boolean	是否显示每个计划节点的预估总成本，以及预估行数和每行宽度。	● TRUE：显示估计总成本和宽度。 ● FALSE：不显示。 缺省值：TRUE
CPU boolean	是否显示CPU的使用情况。	● TRUE：显示CPU的使用情况。 ● FALSE：不显示。 缺省值：TRUE

 版权所有 © 华为云计算技术有限公司

参数	描述	取值范围
DETAIL boolean	是否显示DN上的信息。 **说明** 8.2.1及以上集群版本支持explain打开Detail开关时，执行计划中会显示倾斜值比对耗时。	● TRUE：打印DN的信息。 ● FALSE：不打印。 缺省值：TRUE
NODES boolean	是否打印query执行的节点信息。	● TRUE：打印执行的节点的信息。 ● FALSE：不打印。 缺省值：TRUE
NUM_NODES boolean	是否打印执行中的节点的个数信息。	● TRUE：打印DN个数的信息。 ● FALSE：不打印。 缺省值：TRUE
BUFFERS boolean	是否显示缓冲区的使用信息。 缓冲区信息包括共享块（常规表或者索引块）、本地块（临时表或者索引块）和临时块（排序或者哈希等涉及到的短期存在的数据块）的命中块数，读取块数，更新块数等。	● TRUE：显示缓冲区的使用情况。 ● FALSE：不显示。 缺省值：FALSE
TIMING boolean	是否显示每个计划节点的实际启动时间和总的执行时间。	● TRUE：显示启动时间和花费在输出节点上的时间信息。 ● FALSE：不显示。 缺省值：TRUE
PLAN boolean	是否将执行计划存储在plan_table中。当该选项开启时，会将执行计划存储在PLAN_TABLE中，不打印到当前屏幕，因此该选项为on时，不能与其他选项同时使用。	● ON：将执行计划存储在plan_table中，不打印到当前屏幕。执行成功返回EXPLAIN SUCCESS。 ● OFF：不存储执行计划，将执行计划打印到当前屏幕。 缺省值：ON
FORMAT	指定输出格式。 各个格式输出的内容都是相同的，其中XML | JSON | YAML更有利于您通过程序解析SQL语句的查询计划。	支持四种格式：TEXT，XML，JSON和YAML。 缺省值：TEXT

版权所有 © 华为云计算技术有限公司

参数	描述	取值范围
BLOCKNAME boolean	是否显示算子的blockname信息。 **说明** 仅在explain_perf_mode取值为pretty时，打印blockname信息。	● TRUE：显示blockname信息。 ● FALSE：不显示blockname信息。 缺省值：TRUE
OUTLINE boolean	是否显示从计划中提取的outline信息。 **说明** 仅在explain_perf_mode取值为pretty时，打印outline信息。	● TRUE：显示outline信息。 ● FALSE：不显示outline信息。 缺省值：TRUE
PERFORMANCE	使用此选项时，即打印执行中的所有相关信息。	-
STATS boolean	是否打印复现SQL语句的执行计划所需的信息，包括对象定义、统计信息、配置参数等，通常用于定位问题。	● TRUE：显示复现SQL语句的执行计划所需的信息。 ● FALSE：不显示。 缺省值：TRUE
WARMUP	warmup将查询数据按照A1in > A1out > Am的顺序进行。该选项仅9.1.0.200及以上集群版本支持。	-
WARMUP HOT	warmup hot将查询的数据直接加入am列。该选项仅9.1.0.200及以上集群版本支持。	-
SUMMARY boolean	是否显示计划中explain的SUMMARY信息。 该选项仅9.1.1.100及以上集群版本支持。	● TRUE：显示summary信息。 ● FALSE：不显示summary信息。 缺省值：TRUE
SETTINGS boolean	是否打印有会话级修改的参数。目前只支持部分会话级参数。 该选项仅9.1.1.100及以上集群版本支持。	● TRUE：显示会话级修改的参数信息。 ● FALSE：不显示会话级修改的参数信息。 缺省值：FALSE

示例

创建一个表tpcds.customer_address_p1：

```
CREATE TABLE tpcds.customer_address_p1 AS TABLE tpcds.customer_address;
```

修改explain_perf_mode为normal：

 版权所有 © 华为云计算技术有限公司

```
SET explain_perf_mode=normal;
```

显示表简单查询的执行计划：

```
EXPLAIN SELECT * FROM tpcds.customer_address_p1;
                  QUERY PLAN
-------------------------------------------------------------------------------
 Data Node Scan on "__REMOTE_FQS_QUERY__"  (cost=0.00..0.00 rows=0 width=0)
   Node/s: All datanodes
(2 rows)
```

以JSON格式输出的执行计划（explain_perf_mode为normal时）：

```
EXPLAIN(FORMAT JSON) SELECT * FROM tpcds.customer_address_p1;
                   QUERY PLAN
-------------------------------------------------
 [                                      +
   {                                    +
     "Plan": {                            +
       "Node Type": "Data Node Scan",           +
       "RemoteQuery name": "__REMOTE_FQS_QUERY__",+
       "Alias": "__REMOTE_FQS_QUERY__",         +
       "Startup Cost": 0.00,                    +
       "Total Cost": 0.00,                      +
       "Plan Rows": 0,                          +
       "Plan Width": 0,                         +
       "Nodes": "All datanodes"                 +
     }                                  +
   }                                    +
 ]
(1 row)
```

如果有一个索引，当使用一个带索引WHERE条件的查询，可能会显示一个不同的计划：

```
EXPLAIN SELECT * FROM tpcds.customer_address_p1 WHERE ca_address_sk=10000;
                            QUERY PLAN
-------------------------------------------------------------------------------
 Data Node Scan on "__REMOTE_LIGHT_QUERY__"  (cost=0.00..0.00 rows=0 width=0)
   Node/s: datanode2
(2 rows)
```

以YAML格式输出的执行计划（explain_perf_mode为normal时）：

```
EXPLAIN(FORMAT YAML) SELECT * FROM tpcds.customer_address_p1 WHERE ca_address_sk=10000;
                   QUERY PLAN
-------------------------------------------------
 - Plan:                                +
     Node Type: "Data Node Scan"              +
     RemoteQuery name: "__REMOTE_LIGHT_QUERY__"+
     Alias: "__REMOTE_LIGHT_QUERY__"          +
     Startup Cost: 0.00                       +
     Total Cost: 0.00                         +
     Plan Rows: 0                             +
     Plan Width: 0                            +
     Nodes: "datanode2"
(1 row)
```

禁止开销估计的执行计划：

```
EXPLAIN(COSTS FALSE)SELECT * FROM tpcds.customer_address_p1 WHERE ca_address_sk=10000;
                 QUERY PLAN
---------------------------------------------
 Data Node Scan on "__REMOTE_LIGHT_QUERY__"
   Node/s: datanode2
(2 rows)
```

带有聚集函数查询的执行计划：

```
EXPLAIN SELECT SUM(ca_address_sk) FROM tpcds.customer_address_p1 WHERE ca_address_sk<10000;
                               QUERY PLAN
```

 版权所有 © 华为云计算技术有限公司

```
------------------------------------------------------------------------------------------
 Aggregate  (cost=18.19..14.32 rows=1 width=4)
   ->  Streaming (type: GATHER)  (cost=18.19..14.32 rows=3 width=4)
         Node/s: All datanodes
         ->  Aggregate  (cost=14.19..14.20 rows=3 width=4)
               ->  Seq Scan on customer_address_p1  (cost=0.00..14.18 rows=10 width=4)
                     Filter: (ca_address_sk < 10000)
(6 rows)
```

删除表tpcds.customer_address_p1：

```
DROP TABLE tpcds.customer_address_p1;
```

对ANALYZE语句执行EXPLAIN PERFORMANCE：

```
EXPLAIN PERFORMANCE ANALYZE t2_dist_row;
                           QUERY EXEC INFO
-------------------------------------------------------------------------
 lock FirstCN:
        coordinator1: actual time=0.240 loops=1
 estimate rows: actual time=[datanode3 0.000, datanode1 0.001]
        coordinator1: actual time=0.000 loops=1
        datanode1: actual time=0.001 loops=1
        datanode2: actual time=0.000 loops=1
        datanode3: actual time=0.000 loops=1
 sample rows: actual time=[datanode1 5.109, coordinator1 119.838]
        coordinator1: actual time=119.838 loops=1
        datanode1: actual time=5.109 loops=1
        datanode2: actual time=5.621 loops=1
        datanode3: actual time=5.342 loops=1
 fetch global stats:
        coordinator1: actual time=8.501 loops=1
 calc stats: actual time=[datanode3 80.794, datanode2 109.155]
        coordinator1: actual time=97.452 loops=1
        datanode1: actual time=94.375 loops=1
        datanode2: actual time=109.155 loops=1
        datanode3: actual time=80.794 loops=1
 calc column stats: actual time=[datanode2 0.938, datanode3 9.811]
        coordinator1: actual time=5.162 loops=2
        datanode1: actual time=1.453 loops=2
        datanode2: actual time=0.938 loops=2
        datanode3: actual time=9.811 loops=2
 calc index stats: actual time=[datanode3 12.392, coordinator1 36.113]
        coordinator1: actual time=36.113 loops=1
        datanode1: actual time=15.933 loops=1
        datanode2: actual time=13.419 loops=1
        datanode3: actual time=12.392 loops=1
 calc expr stats: actual time=[datanode3 41.665, datanode2 78.442]
        coordinator1: actual time=55.608 loops=1
        datanode1: actual time=63.179 loops=1
        datanode2: actual time=78.442 loops=1
        datanode3: actual time=41.665 loops=1
 sync stats:
        coordinator1: actual time=7.906 loops=1

 General Tracks
 CN build CN connection:
        coordinator1: actual time=0.002 loops=1
 CN build DN connection:
        coordinator1: actual time=0.070 loops=1
 -> execute ddl on other CN:
        coordinator1: actual time=0.001 loops=1
 -> execute ddl on other DN:
        coordinator1: actual time=0.000 loops=1
 Query Id: 72902018968225366
 Total runtime: 242.211 ms
(48 rows)
```

显示计划的blockname信息：

 版权所有 © 华为云计算技术有限公司

```
EXPLAIN (BLOCKNAME ON) SELECT SUM(ca_address_sk) FROM tpcds.customer_address_p1 WHERE
ca_address_sk<10000;
                                  QUERY PLAN
-------------------------------------------------------------------------------------------
 id |                 operation                 | E-rows | E-memory | E-width | E-costs
----+-------------------------------------------+--------+----------+---------+---------
  1 | ->  Aggregate                             |      1 |          |      12 | 16.14
  2 |    ->  Streaming (type: GATHER)           |      2 |          |      12 | 16.14
  3 |       ->  Aggregate                       |      2 | 1MB      |      12 | 10.14
  4 |          ->  Seq Scan on customer_address_p1 |      7 | 1MB      |       4 | 10.12

 Predicate Information (identified by plan id)
-----------------------------------------------
  4 --Seq Scan on customer_address_p1
        Filter: (ca_address_sk < 10000)

 Query Block Name / Object Alias (identified by plan id)
---------------------------------------------------------
  1 - sel$1
  4 - sel$1 / customer_address_p1@"sel$1"

  ====== Query Summary =====
-------------------------------
System available mem: 4710400KB
Query Max mem: 4710400KB
Query estimated mem: 2048KB
(22 rows)
```

显示计划的outline信息：

```
EXPLAIN (OUTLINE ON) SELECT SUM(ca_address_sk) FROM tpcds.customer_address_p1 WHERE
ca_address_sk<10000;
                                  QUERY PLAN
-------------------------------------------------------------------------------------------
 id |                 operation                 | E-rows | E-memory | E-width | E-costs
----+-------------------------------------------+--------+----------+---------+---------
  1 | ->  Aggregate                             |      1 |          |      12 | 16.14
  2 |    ->  Streaming (type: GATHER)           |      2 |          |      12 | 16.14
  3 |       ->  Aggregate                       |      2 | 1MB      |      12 | 10.14
  4 |          ->  Seq Scan on customer_address_p1 |      7 | 1MB      |       4 | 10.12

 Predicate Information (identified by plan id)
-----------------------------------------------
  4 --Seq Scan on customer_address_p1
        Filter: (ca_address_sk < 10000)

                     Outline Data
-------------------------------------------------------------
  /*+
     begin_outline_data
      TableScan(@"sel$1" tpcds.customer_address_p1@"sel$1")
     end_outline_data
  */

  ====== Query Summary =====
-------------------------------
System available mem: 4710400KB
Query Max mem: 4710400KB
Query estimated mem: 2048KB
(25 rows)
```

相关链接

ANALYZE | ANALYSE

 版权所有 © 华为云计算技术有限公司

14.6 EXPLAIN PLAN

功能描述

通过EXPLAIN PLAN命令可以将查询执行的计划信息存储于PLAN_TABLE表中。与EXPLAIN命令不同的是，EXPLAIN PLAN仅将计划信息进行存储，而不会打印到屏幕。

语法格式

```
EXPLAIN PLAN
[ SET STATEMENT_ID = string ]
FOR statement ;
```

参数说明

表 14-8 EXPLAIN PLAN 参数说明

参数	描述
PLAN	表示需要将计划信息存储于PLAN_TABLE中，存储成功将返回“EXPLAIN SUCCESS”。
STATEMENT_ID	用户可以对查询设置标签，输入的标签信息也将存储于PLAN_TABLE中。 **说明** 用户在执行EXPLAIN PLAN时，如果没有进行SET STATEMENT_ID，则默认为空值。同时，用户可输入的STATEMENT_ID最大长度为30个字节，超过长度将会产生报错。

注意事项

- EXPLAIN PLAN不支持在DN上执行。
- 对于执行错误的SQL无法进行计划信息的收集。
- PLAN_TABLE中的数据是session级生命周期并且session隔离和用户隔离，用户只能看到当前session、当前用户的数据。
- PLAN_TABLE无法与GDS外表进行关联查询。
- 对于不能下推的查询，无法收集到具体的object信息，object只能收集到REMOTE_QUERY或CTE等信息。详见**示例 2**。

示例 1

使用EXPLAIN PLAN收集SQL语句的执行计划，通常包括以下步骤：

步骤1 导入TPC-H样例数据。具体步骤参见**导入样例数据**。

步骤2 执行EXPLAIN PLAN。

 版权所有 © 华为云计算技术有限公司

说明

执行EXPLAIN PLAN后会将计划信息自动存储于PLAN_TABLE中，不支持对PLAN_TABLE进行INSERT、UPDATE、ANALYZE等操作。

PLAN_TABLE详细介绍见视图PLAN_TABLE。

```
explain plan set statement_id='TPCH-Q4' for
select
o_orderpriority,
count(*) as order_count
from
orders
where
o_orderdate >= '1993-07-01'::date
and o_orderdate < '1993-07-01'::date + interval '3 month'
and exists (
select
*
from
lineitem
where
l_orderkey = o_orderkey
and l_commitdate < l_receiptdate
)
group by
o_orderpriority
order by
o_orderpriority;
```

步骤3 查询PLAN_TABLE。

```
SELECT * FROM PLAN_TABLE;
```

```
 statement_id | plan_id  | id |      operation      |   options    | object_name | object_type | object_owner |                projection
--------------+----------+----+---------------------+--------------+-------------+-------------+--------------+-------------------------------------------
 TPCH-Q4      | 16781167 |  1 | ROW ADAPTER         |              |             |             |              | ORDERS.O_ORDERPRIORITY, (PG_CATALOG.COUNT(*))
 TPCH-Q4      | 16781167 |  2 | VECTOR SORT         |              |             |             |              | ORDERS.O_ORDERPRIORITY, (PG_CATALOG.COUNT(*))
 TPCH-Q4      | 16781167 |  3 | VECTOR AGGREGATE    | HASHED       |             |             |              | ORDERS.O_ORDERPRIORITY, PG_CATALOG.COUNT(*)
 TPCH-Q4      | 16781167 |  4 | VECTOR STREAMING    | GATHER       |             |             |              | ORDERS.O_ORDERPRIORITY, (COUNT(*))
 TPCH-Q4      | 16781167 |  5 | VECTOR AGGREGATE    | HASHED       |             |             |              | ORDERS.O_ORDERPRIORITY, COUNT(*)
 TPCH-Q4      | 16781167 |  6 | VECTOR NESTED LOOPS | SEMI         |             |             |              | ORDERS.O_ORDERPRIORITY
 TPCH-Q4      | 16781167 |  7 | TABLE ACCESS        | CSTORE SCAN  | ORDERS      | TABLE       | TPCH         | ORDERS.O_ORDERPRIORITY, ORDERS.O_ORDERKEY
 TPCH-Q4      | 16781167 |  8 | VECTOR MATERIALIZE  |              |             |             |              | LINEITEM.L_ORDERKEY
 TPCH-Q4      | 16781167 |  9 | TABLE ACCESS        | CSTORE SCAN  | LINEITEM    | TABLE       | TPCH         | LINEITEM.L_ORDERKEY
```

步骤4 清理PLAN_TABLE表中的数据。

```
DELETE FROM PLAN_TABLE WHERE xxx;
```

----结束

示例 2

对于不能下推的查询，执行explain plan后plan_table中object仅收集到REMOTE_QUERY或CTE等信息。

优化器生成下发语句的计划，此时仅能收集到REMOTE_QUERY。

```
 explain plan set statement_id = 'test remote query' for
 select
 current_user
 from
 customer;
```

查询PLAN_TABLE。

```
SELECT * FROM PLAN_TABLE;
```

```
   statement_id    | plan_id  | id |   operation    |  options  |   object_name    | object_type  | object_owner |  projection
-------------------+----------+----+----------------+-----------+------------------+--------------+--------------+--------------
 test remote query | 29360133 |  1 | NESTED LOOPS   | CARTESIAN |                  |              |              | 'apple'::name
 test remote query | 29360133 |  2 | DATA NODE SCAN |           | customer         | REMOTE_QUERY |              |
 test remote query | 29360133 |  3 | DATA NODE SCAN |           | customer_address | REMOTE_QUERY |              |
(3 rows)
```


版权所有 © 华为云计算技术有限公司

14.7 LOCK

功能描述

LOCK TABLE获取表级锁。

当自动获取引用表的命令的锁时，DWS会始终使用限制最小的锁模式。如果用户需要一种更为严格的锁模式，可以使用LOCK命令。例如，某个应用是在Read Committed隔离级别上运行事务，并且需要保证表中的数据在事务运行期间保持稳定。为实现这个目的，则可以在查询之前对表使用SHARE锁模式进行锁定。这样将防止并发数据更改，并确保后续的查询可以读到已提交的持久化的数据。因为SHARE锁模式与任何写操作需要的ROW EXCLUSIVE模式冲突，并且LOCK TABLE name IN SHARE MODE语句将等到所有当前持有ROW EXCLUSIVE模式锁的事务提交或回滚后才能执行。因此，一旦获得该锁，就不会存在未提交的写操作，此外其他操作也只能等到该锁释放之后才能开始。

注意事项

- LOCK TABLE只能在一个事务块的内部有用，因为锁在事务结束时就会被释放。出现在任意事务块外面的LOCK TABLE都会报错。
- 如果没有声明锁模式，缺省为最严格的模式ACCESS EXCLUSIVE。
- LOCK TABLE ... IN ACCESS SHARE MODE需要在目标表上有SELECT权限。所有其他形式的LOCK需要UPDATE和/或DELETE权限。
- 没有UNLOCK TABLE命令，锁总是在事务结束时释放。
- LOCK TABLE只处理表级的锁，因此那些带“ROW”字样的锁模式都是有歧义的。这些模式名字通常可理解为用户试图在一个被锁定的表中获取行级的锁。同样，ROW EXCLUSIVE模式也是一个可共享的表级锁。注意，只要是涉及到LOCK TABLE ，所有锁模式都有相同的语义，区别仅在于规则中锁与锁之间是否冲突，规则请参见**锁冲突关系**。

语法格式

```
LOCK [ TABLE ] {[ ONLY ] name [, ...]| {name [ * ]} [, ...]}
    [ IN {ACCESS SHARE | ROW SHARE | ROW EXCLUSIVE | SHARE UPDATE EXCLUSIVE | SHARE | SHARE ROW EXCLUSIVE | EXCLUSIVE | ACCESS EXCLUSIVE | UPDATE EXCLUSIVE} MODE ]
    [ NOWAIT ] [LOCAL COORDINATOR ONLY];
```

锁冲突关系

在DWS中，锁模式用于管理并发事务对资源的访问，以防止数据不一致和丢失。

表14-9列出了DWS支持的8种常规锁模式及其锁冲突关系：

表 14-9 锁模式间的冲突关系

锁模式名称	锁等级	锁用途	冲突关系
访问共享锁（AccessShareLock）	1	SELECT语句，允许其他事务读取数据。	8

版权所有 © 华为云计算技术有限公司

锁模式名称	锁等级	锁用途	冲突关系
行共享锁（RowShareLock）	2	SELECT FOR UPDATE或FOR SHARE，允许其他事务读取但阻止写入。	7|8
行排他锁（RowExclusiveLock）	3	INSERT、UPDATE、DELETE语句。	5|6|7|8
共享更新排他锁（ShareUpdateExclusiveLock）	4	VACUUM（非FULL）、ANALYZE、CREATE INDEX CONCURRENTLY、COMMENT ON，允许其他事务读取但阻止写入。	4|5|6|7|8
共享锁（ShareLock）	5	CREATE INDEX（非CONCURRENTLY），允许其他事务读取但阻止写入。	3|4|6|7|8
共享行排他锁（ShareRowExclusiveLock）	6	ROW SELECT...FOR UPDATE，类似RowExclusiveLock，但允许RowShareLock。	3|4|5|6|7|8
排他锁（ExclusiveLock）	7	阻止RowShareLock或SELECT ... FOR UPDATE操作。	2|3|4|5|6|7|8
访问排他锁（AccessExclusiveLock）	8	ALTER TABLE、DROP TABLE、VACUUM FULL，完全阻止其他事务读取和写入。	1|2|3|4|5|6|7|8

参数说明

表 14-10 LOCK 参数说明

参数	描述	取值范围
name	要锁定的表的名字，可以有模式修饰。 LOCK TABLE命令中声明的表的顺序就是上锁的顺序。	已存在的表名。
ONLY	如果指定ONLY，则只有指定表被锁定；如果没有指定ONLY，则指定表及其所有继承表都会被锁定。 该参数仅为了兼容PG保留，DWS不支持继承表。	-
ACCESS SHARE	ACCESS锁只允许对表进行读取，而禁止对表进行修改。所有对表进行读取而不修改的SQL语句都会自动请求这种锁。例如，SELECT命令会自动在被引用的表上请求一个这种锁。	-

版权所有 © 华为云计算技术有限公司

参数	描述	取值范围
ROW SHARE	ROW SHARE锁允许对表进行并发读取，禁止对表进行其他操作。 SELECT FOR UPDATE和SELECT FOR SHARE命令会自动在目标表上请求ROW SHARE锁（且所有被引用但不是FOR SHARE/FOR UPDATE的其他表上，还会自动加上ACCESS SHARE锁）。	-
ROW EXCLUSIVE	与ROW SHARE锁不同，ROW EXCLUSIVE允许并发读取表，也允许修改表中的数据。UPDATE，DELETE，INSERT命令会自动在目标表上请求这个锁（且所有被引用的其他表上还会自动加上的ACCESS SHARE锁）。通常情况下，所有会修改表数据的命令都会请求表的ROW EXCLUSIVE锁。	-
SHARE UPDATE EXCLUSIVE	这个模式保护一个表的模式不被并发修改，以及禁止在目标表上执行垃圾回收命令（VACUUM ）。 VACUUM（不带FULL选项），ANALYZE，CREATE INDEX CONCURRENTLY命令会自动请求这样的锁。	-
SHARE	SHARE锁允许并发的查询，但是禁止对表进行修改。 CREATE INDEX（不带CONCURRENTLY选项）语句会自动请求这种锁。	-
SHARE ROW EXCLUSIVE	SHARE ROW EXCLUSIVE锁禁止对表进行任何的并发修改，而且是独占锁，因此一个会话中只能获取一次。 任何SQL语句都不会自动请求这个锁模式。	-
EXCLUSIVE	EXCLUSIVE锁允许对目标表进行并发查询，但是禁止任何其他操作。 这个模式只允许并发加ACCESS SHARE锁，也就是说，只有对表的读动作可以和持有这个锁模式的事务并发执行。 任何SQL语句都不会在用户表上自动请求这个锁模式。然而在某些操作的时候，会在某些系统表上请求它。	-
ACCESS EXCLUSIVE	这个模式保证其所有者（事务）是可以访问该表的唯一事务。 ALTER TABLE，DROP TABLE，TRUNCATE，REINDEX，CLUSTER，VACUUM FULL命令会自动请求这种锁。	在LOCK TABLE命令没有明确声明需要的锁模式时，为缺省锁模式。

版权所有 © 华为云计算技术有限公司

参数	描述	取值范围
UPDATE EXCLUSIVE	UPDATE EXCLUSIVE锁允许并发（AUTO）VACUUM和（AUTO）ANALYZE，但不允许多个（AUTO）VACUUM并发。 **说明** ● 该参数仅8.2.1.200及以上集群版本支持。 ● UPDATE EXCLUSIVE锁模式仅在VACUUM语法中使用。	-
NOWAIT	声明LOCK TABLE不去等待任何冲突的锁释放，如果无法立即获取该锁，该命令退出并且发出一个错误信息。 在不指定NOWAIT的情况下获取表级锁时，如果有其他互斥锁存在，则等待其他锁的释放。	-
LOCAL COORDINATOR ONLY	声明LOCK TABLE只在接收当前会话请求的CN上执行，而不会下发给其他CN和所有DN。该选项仅针对元数据的操作，以提高效率。 **须知** ● 该参数仅8.2.0.100以上集群版本支持。 ● 当前仅支持ACCESS SHARE锁模式，其他锁模式会报错。	-

示例

准备数据：

```
DROP SCHEMA IF EXISTS tpcds CASCADE;
CREATE SCHEMA tpcds;
CREATE TABLE tpcds.reason
(
    r_reason_sk             bigint              not null,
    r_reason_id             char(16)             not null,
    r_reason_desc           char(100)
)
 with (orientation = column)
distribute by replication;
```

向一个外键表上插入数据时，在有主键的表上使用SHARE锁：

```
START TRANSACTION;

LOCK TABLE tpcds.reason IN SHARE MODE;

SELECT r_reason_desc FROM tpcds.reason WHERE r_reason_sk=5;

COMMIT;
```

在执行删除操作时对一个有主键的表进行SHARE ROW EXCLUSIVE锁：

```
DROP TABLE IF EXISTS tpcds.reason_t1;
CREATE TABLE tpcds.reason_t1 AS TABLE tpcds.reason;

START TRANSACTION;
```

 版权所有 © 华为云计算技术有限公司

```
LOCK TABLE tpcds.reason_t1 IN SHARE ROW EXCLUSIVE MODE;

DELETE FROM tpcds.reason_t1 WHERE r_reason_desc IN(SELECT r_reason_desc FROM tpcds.reason_t1
WHERE r_reason_sk < 6 );

DELETE FROM tpcds.reason_t1 WHERE r_reason_sk = 7;

COMMIT;
```

删除表tpcds.reason_t1：

```
DROP TABLE tpcds.reason_t1;
```

查看表定义时，对表加ACCESS SHARE锁模式、限制范围是当前CN：

```
DROP TABLE IF EXISTS lock_test;
CREATE TABLE lock_test (a int, b int);
BEGIN;
LOCK TABLE lock_test IN ACCESS SHARE MODE LOCAL COORDINATOR ONLY;
SELECT pg_get_tabledef('lock_test');
END;
```

14.8 MERGE INTO

功能描述

MERGE INTO语句将目标表和源表中数据针对关联条件进行匹配，若关联条件匹配时对目标表进行UPDATE，不匹配时对目标表执行INSERT。MERGE INTO合并执行UPDATE和INSERT，避免多次执行以提高执行效率。

注意事项

- 执行MERGE INTO操作的用户需要同时拥有目标表的UPDATE和INSERT权限，以及源表的SELECT权限。
- 不支持PREPARE。
- 不支持重分布过程中执行MERGE INTO。
- 不支持对包含触发器的目标表执行MERGE INTO。
- 对roundrobin表执行MERGE INTO时，建议关闭GUC参数allow_concurrent_tuple_update，否则会不支持部分MERGE INTO语句。
- 对于并发MERGE INTO场景，触发UPDATE时，传统列存储格式由于同时操作相同CU时会产生锁冲突所以并发性能较差，对于HStore表由于不需要阻塞等待，并发MERGE INTO性能可达到列存的百倍以上。

语法格式

```
MERGE [/*+ plan_hint */] INTO table_name [ partition_clause ] [ [ AS ] alias ]
USING { { table_name | view_name } | subquery } [ partition_clause ] [ [ AS ] alias ]
ON ( condition )
[
  WHEN MATCHED THEN
  UPDATE SET { column_name = { expression | DEFAULT } |
         ( column_name [, ...] ) = ( { expression | DEFAULT } [, ...] ) } [, ...]
  [ WHERE condition ]
]
[
  WHEN NOT MATCHED THEN
  INSERT { DEFAULT VALUES |
  [ ( column_name [, ...] ) ] VALUES ( { expression | DEFAULT } [, ...] ) [, ...] [ WHERE condition ] }
];
```

 版权所有 © 华为云计算技术有限公司

```
where partition_clause can be:
PARTITION ( partition_name ) | PARTITION FOR ( partition_key_value [, ...] )
```

参数说明

表 14-11 MERGE INTO 参数说明

参数	描述	取值范围
plan_hint子句	以/*+ */的形式在关键字后，用于对指定语句块生成的计划进行hint调优，详细用法请参见**使用Plan Hint进行调优**。	-
INTO子句	指定正在更新或插入的目标表。 • **table_name** 目标表的表名。 • **partition_clause** 用来指定目标表分区的子句。该参数9.1.0及以上集群版本支持。 只有当目标表是分区表时可以指定分区。 指定的分区名需要是目标表已存在的分区或者指定的分区键值所处的分区已存在。 • **alias** 目标表的别名。	有效的表名。
USING子句	指定源表，源表可以是表、视图或子查询。 • **partition_clause** 用来指定源表分区的子句。该参数9.1.0及以上集群版本支持。 只有当源表是分区表时可以指定分区，视图或者子查询不支持指定分区。 指定的分区名需要是源表已存在的分区或者指定的分区键值所处的分区已存在。 • **alias** 源表的别名。	有效的源表名。
ON子句	关联条件，用于指定目标表和源表的关联条件。不支持更新关联条件中的字段。ON关联条件可以是ctid，xc_node_id，tableoid这三个系统列。	-

版权所有 © 华为云计算技术有限公司

参数	描述	取值范围
WHEN MATCHED子句	当源表和目标表中数据根据关联条件可以匹配上时，选择WHEN MATCHED子句进行UPDATE操作。 ● 不支持更新分布列。不支持更新系统表、系统列。 ● 当指定目标表的分区时，若UPDATE目标表的分区键导致分区变化： – 目标表是列存表时，会报错“The inserted partition key is not mapped to the specified 'partition_name' partition”。 – 目标表是行存表时： 若目标表的表属性为ENABLE ROW MOVEMENT，则不会报错，更新的数据会属于新分区。 若目标表的表属性为DISABLE ROW MOVEMENT，会报错“fail to update partitioned table 'table_name' ”。	-
WHEN NOT MATCHED子句	当源表和目标表中数据根据关联条件无法匹配时，选择WHEN NOT MATCHED子句进行INSERT操作。 不支持INSERT子句中包含多个VALUES。 WHEN MATCHED和WHEN NOT MATCHED子句顺序可以交换，可以缺省其中一个，但不能同时缺省，不支持同时指定两个WHEN MATCHED或WHEN NOT MATCHED子句。 当指定目标表的分区时，若INSERT了超出该分区范围的数据，会报错“The inserted partition key is not mapped to the specified 'partition_name' partition”。	-
DEFAULT	用对应字段的缺省值填充该字段。	如果没有缺省值，则为NULL。
WHERE condition	UPDATE子句和INSERT子句的条件，只有在条件满足时才进行更新操作，可缺省。不支持WHERE条件中引用系统列。	-

 版权所有 © 华为云计算技术有限公司

示例

创建目标表products和源表newproducts，并插入数据：

```
DROP TABLE IF EXISTS products;
DROP TABLE IF EXISTS newproducts;


CREATE TABLE products
(
product_id INTEGER,
product_name VARCHAR2(60),
category VARCHAR2(60)
);

INSERT INTO products VALUES (1501, 'vivitar 35mm', 'electronics');
INSERT INTO products VALUES (1502, 'olympus is50', 'electronics');
INSERT INTO products VALUES (1600, 'play gym', 'toys');
INSERT INTO products VALUES (1601, 'lamaze', 'toys');
INSERT INTO products VALUES (1666, 'harry potter', 'dvd');

CREATE TABLE newproducts
(
product_id INTEGER,
product_name VARCHAR2(60),
category VARCHAR2(60)
);

INSERT INTO newproducts VALUES (1502, 'olympus camera', 'electronics');
INSERT INTO newproducts VALUES (1601, 'lamaze', 'toys');
INSERT INTO newproducts VALUES (1666, 'harry potter', 'toys');
INSERT INTO newproducts VALUES (1700, 'wait interface', 'books');
```

执行MERGE INTO操作，仅当product_id字段匹配且目标表product_name字段非 'play gym' 时UPDATE，product_id字段不匹配且源表category字段为books时INSERT。

```
MERGE INTO products p
USING newproducts np
ON (p.product_id = np.product_id)
WHEN MATCHED THEN
  UPDATE SET p.product_name = np.product_name, p.category = np.category WHERE p.product_name != 'play gym'
WHEN NOT MATCHED THEN
  INSERT VALUES (np.product_id, np.product_name, np.category) WHERE np.category = 'books';
```

查询更新后的结果显示：目标表products的对应product_id字段1502、1666行相关数据已更新，1700行为新插入的数据。

```
SELECT * FROM products ORDER BY product_id;
```

```
gaussdb=> SELECT * FROM products ORDER BY product_id;
 product_id |  product_name  |  category
------------+----------------+-------------
       1501 | vivitar 35mm   | electronics
       1502 | olympus camera | electronics
       1600 | play gym       | toys
       1601 | lamaze         | toys
       1666 | harry potter   | toys
       1700 | wait interface | books
(6 rows)
```

删除表：

```
DROP TABLE products;
DROP TABLE newproducts;
```

 版权所有 © 华为云计算技术有限公司

常见问题

执行MERGE INTO将源表内容根据匹配条件对目标表做更新报错unable to get a stable set of rows in the source table

14.9 INSERT 和 UPSERT

14.9.1 INSERT

功能描述

向表中添加一行或多行数据。

注意事项

- 使用INSERT请遵守INSERT开发规范。
- 只有拥有表INSERT权限的用户，才可以向表中插入数据。
- 如果使用RETURNING子句，用户必须要有该表的SELECT权限。
- 如果使用QUERY子句插入来自查询里的数据行，用户还需要拥有在查询里使用的表的SELECT权限。
- 如果使用OVERWRITE子句覆盖式插入数据，用户还需要拥有该表的SELECT和TRUNCATE权限（因为OVERWRITE操作是先清空目标表原有数据）。
- 当连接到TD兼容的数据库时，td_compatible_truncation参数设置为on时，将启用超长字符串自动截断功能，在后续的INSERT语句中（不包含外表的场景下），对目标表中char和varchar类型的列上插入超长字符串时，系统会自动按照目标表中相应列定义的最大长度对超长字符串进行截断。

  说明

  如果向字符集为字节类型编码（SQL_ASCII，LATIN1等）的数据库中插入多字节字符数据（如汉字等），且字符数据跨越截断位置，这种情况下，按照字节长度自动截断，自动截断后会在尾部产生非预期结果。如果用户有对于截断结果正确性的要求，建议用户采用UTF8等能够按照字符截断的输入字符集作为数据库的编码集。
- 当单次插入的数据量大于等于表级参数DELTAROW_THRESHOLD时，数据会直接插入主表生成CU（Compress Unit）。
- 当插入的数据量小于表级参数DELTAROW_THRESHOLD时，会在辅助Delta表上插入一条类型为I的记录，同时将数据序列化存储到这条记录的values字段。
- 插入到Delta表的数据跟主表使用全局统一分配的cuid。
- 插入到Delta表的数据依赖autovacuum来merge到主表CU。

语法格式

```
[ WITH [ RECURSIVE ] with_query [, ...] ]
INSERT [/*+ plan_hint */] [ IGNORE | OVERWRITE ] INTO table_name [ partition_clause ] [ AS alias ]
[ ( column_name [, ...] ) ]
    { DEFAULT VALUES
    | VALUES {( { expression | DEFAULT } [, ...] ) }[, ...]
    | query }
    [ ON DUPLICATE KEY duplicate_action | ON CONFLICT [ conflict_target ] conflict_action ]
    [ RETURNING {* | {output_expression [ [ AS ] output_name ] }[, ...]} ];
```

where partition_clause can be:

 版权所有 © 华为云计算技术有限公司

```
   PARTITION ( partition_name )
   | PARTITION FOR ( partition_key_value [, ...] )

and duplicate_action can be:

   UPDATE { column_name = { expression | DEFAULT } |
            ( column_name [, ...] ) = ( { expression | DEFAULT } [, ...] )
          } [, ...]

and conflict_target can be one of:

   ( { index_column_name | ( index_expression ) } [ COLLATE collation ] [ opclass ] [, ...] ) [ WHERE
index_predicate ]
   ON CONSTRAINT constraint_name

and conflict_action is one of:

   DO NOTHING
   DO UPDATE SET { column_name = { expression | DEFAULT } |
                   ( column_name [, ...] ) = ( { expression | DEFAULT } [, ...] )
                 } [, ...]
             [ WHERE condition ]
```

参数说明

表 14-12 INSERT 参数说明

参数	描述	取值范围或示例
WITH [RECURSIVE] with_query [, ...]	用于声明一个或多个可以在主查询中通过名字引用的子查询，相当于临时表。 如果声明了RECURSIVE，则允许SELECT子查询通过名字引用它自己。 其中with_query的详细格式为： `with_query_name [ ( column_name [, ...] ) ] AS ( {select | values | insert | update | delete} )` • with_query_name指定子查询生成的结果集名字，在查询中可使用该名称访问子查询的结果集。 • column_name指定子查询结果集中显示的列名。 • 每个子查询可以是SELECT、VALUES、INSERT、UPDATE或DELETE语句。	**示例：使用WITH语句生成临时表再插入另一张表**
plan_hint子句	以/*+ */的形式在关键字后，用于对指定语句块生成的计划进行hint调优，详细用法请参见**使用Plan Hint进行调优**。	-
IGNORE	用于主键或者唯一约束冲突时忽略冲突的数据。 详细介绍参见**UPSERT**。	**示例：使用IGNORE忽略主键冲突**

版权所有 © 华为云计算技术有限公司

参数	描述	取值范围或示例
OVERWRITE	**仅8.1.1及以上版本支持**，用于标识覆盖式插入方式，使用此种插入方式执行结束后，目标源数据被清空，只存在新插入的数据。 OVERWRITE支持指定列插入的功能，其他列为默认值，若无默认值则为NULL。 INSERT OVERWRITE约束： ● OVERWRITE不要和INSERT INTO这类实时写入的操作并发，否则实时写入数据有被意外清理的风险。 ● 复制表（Replication）不支持INSERT OVERWRITE。OVERWRITE适用于大批量数据导入场景，不建议用于少量数据的插入场景。 ● 避免对同一张表执行并发insert overwrite操作，否则会出现类似报错“tuple concurrently updated.”。 ● 如果集群正在扩缩容，且INSERT OVERWRITE的写入表需要执行数据重分布，则INSERT OVERWRITE会清除当前数据，并自动将插入的数据按扩缩容后的节点来进行数据分布。如果INSERT OVERWRITE和该表的数据重分布过程同时执行，INSERT OVERWRITE会中断该表的数据重分布过程。 ● 细粒度容灾表不支持INSERT OVERWRITE语法。 ● OVERWRITE支持覆盖式插入OBS外表，此方式会删除原目录中数据文件。 ● 外表支持OVERWRITE单分区，支持Value多级分区，该功能仅9.1.0.100及以上版本支持。 示例：INSERT OVERWRITE into table_name partition for (p1_value, p2_value...) ... ● 对同一张表同时执行INSERT OVERWRITE和ALTER TABLE...ADD PARTITION操作时，偶发可能会出现不同节点的分区数量不一致的情况，从而导致INSERT OVERWRITE报错。遇到这种情况时，建议重新执行INSERT OVERWRITE语句，即可解决。	-
table_name	要插入数据的目标表名。	已存在的表名。
AS	用于给目标表table_name指定别名。alias即为别名的名字。	-

版权所有 © 华为云计算技术有限公司

参数	描述	取值范围或示例
column_name	目标表中的字段名： • 字段名可以用子字段名或者数组下标修饰。 • 没有在字段列表中出现的每个字段，将由系统默认值，或者声明时的默认值填充，若都没有则用NULL填充。例如，向一个复合类型中的某些字段插入数据，其他字段将是NULL。 • 目标字段（column_name）可以按顺序排列。如果没有列出任何字段，则默认全部字段，且顺序为表声明时的顺序。 • 如果value子句和query中只提供了N个字段，则目标字段为前N个字段。 • value子句和query提供的值在表中从左到右关联到对应列。	已存在的字段名。
partition_name	分区名。	已存在的分区名。
partition_key_value	分区键值。 通过PARTITION FOR（partition_key_value [, ...])子句指定的这一组值，可以唯一确定一个分区。支持指定多个分区键值，多个分区键值以英文逗号分隔。	-
expression	赋予对应column的一个有效表达式或值： • 向表中字段插入单引号时需要使用单引号自身进行转义。 • 如果插入行的表达式不是正确的数据类型，系统试图进行类型转换，若转换不成功，则插入数据失败，系统返回错误信息。	**示例：向某列中插入一个表达式或值**
DEFAULT	对应字段名的缺省值。	如果没有缺省值，则为NULL，参见**示例：写入数据时对某个字段使用默认值**。
query	一个查询语句（SELECT语句），将查询结果作为插入的数据。	**示例：将查询结果插入表**
ON DUPLICATE KEY	用于主键或者唯一约束冲突时更新冲突的数据。 duplicate_action指定更新列和更新的数据。 详细介绍参见**UPSERT**。	**示例：使用ON DUPLICATE KEY更新主键或唯一约束冲突时插入数据**

版权所有 © 华为云计算技术有限公司

参数	描述	取值范围或示例
ON CONFLICT	用于主键或者唯一约束冲突时忽略或者更新冲突的数据。 conflict_target用于指定列名index_column_name 、包含多个列名的表达式index_expression或约束名字constraint_name。用于从列名、包含多个列名的表达式或者约束名推断是否有唯一索引。其中index_column_name和index_expression遵循CREATE INDEX的索引列格式。 conflict_action指定主键或者唯一约束冲突时执行的策略。有两种： • DO NOTHING冲突忽略。 • DO UPDATE SET冲突更新。 后面指定更新列和更新的数据。 详细介绍参见**UPSERT**。	**示例**
RETURNING	返回实际插入的行，RETURNING列表的语法与SELECT的输出列表一致。	**示例：使用RETURNING返回插入结果**
output_expression	INSERT命令在每一行都被插入之后用于计算输出结果的表达式。	该表达式可以使用table的任意字段。可以使用*返回被插入行的所有字段。
output_name	字段的输出名称。	字符串，符合标识符命名规范。

示例：插入单条或多条数据

创建表reason_t1：

```
DROP TABLE IF EXISTS reason_t1;
CREATE TABLE reason_t1
(
    TABLE_SK         INTEGER              ,
    TABLE_ID         VARCHAR(20)          ,
    TABLE_NA          VARCHAR(20)
);
```

向表中插入一条数据：

```
INSERT INTO reason_t1(TABLE_SK, TABLE_ID, TABLE_NA) VALUES (1, 'S01', 'StudentA');
```

向表中插入一条数据，和上一条语法等效：

```
INSERT INTO reason_t1 VALUES (1, 'S01', 'StudentA');
```

向表中插入TABLE_SK小于1的数据：

```
INSERT INTO reason_t1 SELECT * FROM reason_t1 WHERE TABLE_SK < 1;
```


版权所有 © 华为云计算技术有限公司

向表中插入多条数据：

```
INSERT INTO reason_t1 VALUES (1, 'S01', 'StudentA'),(2, 'T01', 'TeacherA'),(3, 'T02', 'TeacherB');
SELECT * FROM reason_t1 ORDER BY 1;
```

table_sk	table_id	table_na
1	S01	StudentA
1	S01	StudentA
1	S01	StudentA
2	T01	TeacherA
3	T02	TeacherB

示例：将查询结果插入表

将表reason_t1的数据插入到表reason_t1中：

```
INSERT INTO reason_t1 SELECT * FROM reason_t1;
```

示例：使用 RETURNING 返回插入结果

插入表数据，并返回插入的数据对应的字段table_id和table_na的取值。

```
INSERT INTO reason_t1 SELECT * FROM reason_t1 RETURNING table_id, table_na;
```

table_id	table_na
T01	TeacherA
T02	TeacherB
S01	StudentA
T02	TeacherB
T01	TeacherA

插入数据，并返回插入数据的所有字段取值。

```
INSERT INTO reason_t1 SELECT * FROM reason_t1 RETURNING *;
```

 版权所有 © 华为云计算技术有限公司

table_sk	table_id	table_na
2	T01	TeacherA
3	T02	TeacherB
2	T01	TeacherA
3	T02	TeacherB
1	S01	StudentA

插入数据，并返回重命名后的字段my_id取值。

```
INSERT INTO reason_t1 VALUES (1, 'S01', 'StudentA') RETURNING table_id AS my_id;
```

my_id
S01

示例：使用 INSERT OVERWRITE 覆盖式插入

使用INSERT OVERWRITE更新表中的数据，即覆盖式插入数据（先清空目标表数据）：

```
INSERT OVERWRITE INTO reason_t1 values (5, 'S02', 'StudentB');
SELECT * FROM reason_t1 ORDER BY 1;
```

table_sk	table_id	table_na
5	S02	StudentB

示例：写入数据时对某个字段使用默认值

```
DROP TABLE IF EXISTS products;
CREATE TABLE products (
    id INT PRIMARY KEY,
    name VARCHAR(50),
    price DECIMAL(10, 2) DEFAULT 0.00
);
```

插入一条数据，未定义price字段的取值，则默认使用0.00。

```
INSERT INTO products (id, name)VALUES (201, 'phone'); -- price字段将使用默认值0.00
```

插入一条数据，显式使用默认值。

```
INSERT INTO products (id, name, price)VALUES (202, 'pad', DEFAULT); -- 显式使用默认值
SELECT * FROM products;
```

id	name	price
201	phone	0.00
202	pad	0.00

版权所有 © 华为云计算技术有限公司

示例：使用 WITH 语句生成临时表再插入另一张表

将一个表中的部分数据插入到另一个表中：先通过WITH子查询得到一张临时表temp_t，然后将临时表temp_t中的所有数据插入另一张表reason_t1中：

```
WITH temp_t AS (SELECT * FROM reason_t1) INSERT INTO reason_t1 SELECT * FROM temp_t ORDER BY 1;
```

示例：使用 IGNORE 忽略主键冲突

```
DROP TABLE IF EXISTS users;
CREATE TABLE users (
    id INT PRIMARY KEY,
    username VARCHAR(50)
)
DISTRIBUTE BY HASH (id);
```

插入两条主键都是1的记录，提示冲突，无法执行。

```
INSERT  INTO users (id, username)
VALUES (1, 'john_doe'),
       (2, 'jane_smith'),
       (1, 'duplicate_user'); --主键冲突，无法插入
```

```
error_code: DWS.S0010010
error_msg: Constraint Violation. STATE: 23505, message: ERROR: dn_6005_6006: duplicate key value violates unique constraint "users_pkey"
Detail: Key (id)=(1) already exists.
```

使用IGNORE忽略冲突，冲突数据被忽略跳过。

```
INSERT IGNORE INTO users (id, username)
VALUES (1, 'john_doe'),
       (2, 'jane_smith'),
       (1, 'duplicate_user'); -- 冲突数据会被忽略
SELECT * FROM users;
```

id	username
1	john_doe
2	jane_smith

示例：使用 ON DUPLICATE KEY 更新主键或唯一约束冲突时插入数据

```
DROP TABLE IF EXISTS user_logins;
CREATE TABLE user_logins
 (user_id  int  primary key,
login_count  int)
DISTRIBUTE BY HASH (user_id);
```

插入2条数据，第二条主键冲突，插入失败。

```
INSERT INTO user_logins VALUES (1, 1);
INSERT INTO user_logins VALUES (1, 1);  --主键冲突，插入失败
SELECT * FROM user_logins;
```

user_id	login_count
1	1

 版权所有 © 华为云计算技术有限公司

使用ON DUPLICATE KEY声明，如果主键冲突了，则更新login_count字段，使其加1。

```
INSERT INTO user_logins VALUES (1, 1) ON DUPLICATE KEY UPDATE login_count = login_count + 1;
SELECT * FROM user_logins;
```

user_id	login_count
1	2

示例：向某列中插入一个表达式或值

```
CREATE TABLE tt01 (id int,content varchar(50));

INSERT INTO tt01 values (1,'Jack say ''hello''');  --插入Jack say 'hello'，加单引号进行转义
INSERT INTO tt01 values (2,'Rose do 50%');
INSERT INTO tt01 values (3,'Lilei say ''world''');  --插入Lilei say 'world'，加单引号进行转义
INSERT INTO tt01 values (4,'Hanmei do 100%');

SELECT * FROM tt01;
```

id	content
3	Lilei say 'world'
2	Rose do 50%
1	Jack say 'hello'
4	Hanmei do 100%

示例：向分区表的指定分区插入数据

向分区表的指定分区插入数据：

```
CREATE TABLE test_range_row(a int, d int)
DISTRIBUTE BY hash(a) PARTITION BY RANGE(d)
(
    PARTITION p1 values LESS THAN (60),
    PARTITION p2 values LESS THAN (75),
    PARTITION p3 values LESS THAN (90),
    PARTITION p4 VALUES LESS THAN (maxvalue)
);
INSERT OVERWRITE INTO test_range_row PARTITION(p1) VALUES(55,51);
INSERT OVERWRITE INTO test_range_row PARTITION(p3) VALUES(85,80);

DELETE FROM test_range_row PARTITION(p1);
```

向分区外表中的指定分区插入数据：

```
CREATE FOREIGN TABLE test_ft(c1 text, c2 text, c3 text, c4 text)
PARTITION BY (c1,c2,c3) AUTOMAPPED;
```

向一级(p1_1)分区中插入数据：

```
INSERT OVERWRITE INTO test_ft PARTITION FOR(p1_1) VALUES(p1_1,2,3,4);
```

向二级(p1_2,p2_2)分区中插入数据：

```
INSERT OVERWRITE INTO test_ft PARTITION FOR(p1_2,p2_2) VALUES(p1_2,p2_2,3,4);
```

向三级(p1_3,p2_3,p3_3)分区中插入数据：

 版权所有 © 华为云计算技术有限公司

```
INSERT OVERWRITE INTO test_ft PARTITION FOR(p1_3,p2_3,p3_3) VALUES(p1_3,p2_3,p3_3,4);
```

示例：MySQL 兼容模式下，通过设置 hive_compat_options 实现插入空串不被当作 0 处理

DWS在MySQL兼容模式下进行数据操作时，如果插入空串，系统会默认**将空串当作0处理**。

在金融行业，例如从Hive导入数据到DWS场景中，如果将空串当作0处理会导致数据被篡改（金融场景下空串和0具有不同的意义），影响数据的准确性和可靠性。

如何确保数据在导入过程中的完整性和一致性？通过设置GUC参数**hive_compat_options**值为**empty_str_to_null**实现转换逻辑：**当插入空串时，系统会将空串转为NULL**，从而保持数据的一致性和准确性，下面模拟这个过程。

创建一个兼容模式为MySQL的数据库。

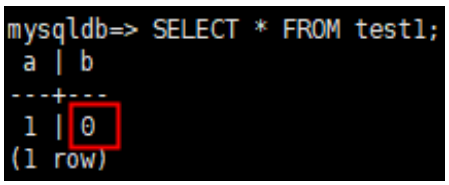
```
DROP DATABASE IF EXISTS mysqldb;
CREATE DATABASE mysqldb DBCOMPATIBILITY 'MYSQL';
```

重新连接新的数据库mysqldb，执行以下语句，并插入空串。

```
CREATE TABLE test1 (a int, b int);
INSERT INTO test1 VALUES (1, '');
```

查询数据，发现空串被设置为0。

```
SELECT * FROM test1;
```



将GUC参数hive_compat_options设置为**empty_str_to_null**。

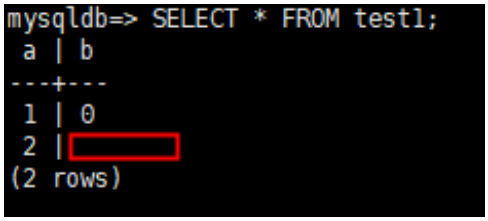
```
ALTER DATABASE mysqldb SET hive_compat_options TO empty_str_to_null;
```

再次插入数据。

```
INSERT INTO test1 VALUES (2, '');
```

查询数据，发现空串自动转换为NULL。

```
SELECT * FROM test1;
```



常见问题

- **分区表插入数据报错：inserted partition key does not map to any table partition**
- **插入数据时，VARCHAR(n)类型的字段中存在中文字符，提示value too long for type character varying**

 版权所有 © 华为云计算技术有限公司

14.9.2 UPSERT

功能描述

向表中添加一行或多行数据。当出现主键或者唯一约束冲突时，它可以选择更新冲突的数据，或者忽略这些冲突。

须知

UPSERT语法仅8.1.1及以上版本支持。

注意事项

⚠ 警告

- 避免UPDATE/UPSERT并发更新同一张列存表。
- 更多开发设计规范参见**总体开发设计规范**。

- **权限要求**：只有拥有表INSERT、UPDATE权限的用户，才可以通过UPSERT语句向表中插入或更新数据。
- **索引要求**：目标表上必须包含主键或者唯一索引才可以执行UPSERT的冲突更新语句。所有的唯一索引都不可用时不能执行UPSERT的冲突更新语句，重建索引后可以正常执行。
- **其他约束：**
  - 不支持与WITH子句同时使用。
  - 不支持与INSERT OVERWRITE同时使用。
  - UPDATE子句和UPDATE的WHERE子句不能有子查询。
  - UPDATE子句中VALUES(colname)用法不支持外层嵌套函数，即不支持类似sqrt(VALUES(colname))用法。如需支持，使用EXCLUDED.colname语法。
  - INSERT INTO ON CONFLICT(...) DO UPDATE必须有conflict_target。即必须指定列或者约束名。
  - 不能更新分布列。（**例外：当分布键与更新值相同时**）。

    ```
    CREATE TABLE t1(dist_key int PRIMARY KEY, a int, b int);
    INSERT INTO t1 VALUES(1,2,3) ON CONFLICT(dist_key) DO UPDATE SET dist_key = EXCLUDED.dist_key, a = EXCLUDED.a + 1;
    INSERT INTO t1 VALUES(1,2,3) ON CONFLICT(dist_key) DO UPDATE SET dist_key = dist_key, a = EXCLUDED.a + 1;
    ```

  - 不支持对包含触发器（触发事件为INSERT或UPDATE）的目标表执行UPSERT语句。
  - 不支持对可更新视图执行UPSERT语句。
  - UPDATE子句、UPDATE中WHERE子句或者索引条件表达式不能包含**不下推函数**。
  - 不支持延迟唯一索引。
  - 通过INSERT INTO SELECT语句执行UPSERT的更新操作时，需要注意SELECT语句的查询结果顺序。在分布式环境中未使用ORDER BY语句时每次执行相同

版权所有 © 华为云计算技术有限公司

的SELECT语句返回结果顺序可能不一样，这会导致UPSERT语句的执行结果不符合预期。

- 不支持多次更新。插入多组数据间如果有冲突，则会出现类似“INSERT ON CONFLICT DO UPDATE command cannot affect row a second time”报错（**例外：当查询计划是PGXC计划时**），详细案例参见**常见问题**。
- 与列存一样，当UPSERT触发更新操作时，当前版本HStore上的更新操作始终先DELETE再INSERT。

- **性能影响**：
  - 当在列存表上执行UPSERT时，建议开启DELTA表，能够有效防止执行UPSERT时产生小CU（大量的小CU会导致空间膨胀和查询性能差）。
  - 对于列存表上的UPSERT、UPDATE、DELETE并发场景，由于并发更新到同一个CU时需要等待CU锁，无法支持这几个操作的并发执行，开启DELTA也无法解决该问题。如果需要支持列存表的UPSERT、UPDATE、DELETE并发场景，请使用HStore表。
  - 对于并发UPSERT场景，传统列存储格式由于同时操作相同CU时会产生锁冲突所以并发性能较差，对于HStore表由于不需要阻塞等待，并发UPSERT性能可达到列存的百倍以上。
- **死锁问题**：可能存在分布式死锁导致查询hang问题。例如在一个事务中或者通过JDBC(setAutoCommit(false))批量执行多条UPSERT语句，多个类似任务同时执行时，由于不同线程在不同节点更新顺序可能不同，在存在并发更新同一行的场景里可能会有死锁问题。**解决方法**如下：

  a. 调小GUC参数**lockwait_timeout**值（默认20分钟）。分布式死锁会等待lockwait_timeout时间然后报错。通过减小此参数的数值，降低死锁造成的业务等待时间。

  b. 保证主键相同的数据只从一个数据库连接导入数据库，可以并发执行UPSERT语句。

  c. 每个事务中只执行一条UPSERT语句，这样也可以并发执行UPSERT语句。

  d. 单线程执行多条UPSERT语句，不能并发执行UPSERT语句。

  如上解决方法中，方法1只能降低等待时间，无法解决死锁问题，在业务中有UPSERT语句时，仍建议减小此参数值。方法2、3、4均可以解决死锁问题，但建议采用方法2，其性能优于其他两个方法。

语法格式

详细介绍请参见INSERT的**语法格式**。有两种UPSERT语法格式：

第一种**不指定索引**。会在所有主键或唯一索引上检查冲突，有冲突就会忽略或者更新。

第二种**指定索引**。会从ON CONFLICT子句中指定列名、包含列名的表达式或者约束名上推断主键或者唯一索引。

- **场景一：冲突更新语法**：当插入的数据发生冲突时，使用新数据更新已有数据。
  - **不指定索引的冲突更新**：
    ```
    INSERT INTO ON DUPLICATE KEY UPDATE
    ```
  - **指定索引的冲突更新**：
    ```
    INSERT INTO ON CONFLICT(...) DO UPDATE SET
    INSERT INTO ON CONFLICT ON CONSTRAINT con_name DO UPDATE SET
    ```
- **场景二：冲突忽略语法**：当插入的数据发生冲突时，直接忽略这次插入操作。

版权所有 © 华为云计算技术有限公司

- **不指定索引的冲突忽略**：
  ```
  INSERT IGNORE
  INSERT INTO ON CONFLICT DO NOTHING
  ```
- **指定索引的冲突忽略**：
  ```
  INSERT INTO ON CONFLICT(...) DO NOTHING
  INSERT INTO ON CONFLICT ON CONSTRAINT con_name DO NOTHING
  ```

表 14-13 UPSERT 参数说明

参数	描述	取值范围或示例
ON CONFLICT (...)	对于指定索引的语法形式，通过指定列名或者约束名推断主键或者唯一索引。列名可以是单一列名，或者由多个列名组成的表达式，比如（column1，column2，column3）。 由于创建索引时可以指定collation和opclass，所以此处列名后也可以指定。 COLLATE collation指定列的排序规则。opclass指定操作符类的名字。具体参考**CREATE INDEX**。 从指定列名的表达式中推断出唯一索引，整体原则是判断某唯一索引是否能够恰好完全包含conflict_target指定的列名。 ● 如果没有指定collation和opclass，那么只要列或者列名的表达式相同（不管索引列上指定的collation和opclass是什么），都认为匹配。 ● 如果指定collation和opclass，那么需要与索引的collation和opclass匹配才可以。	-
UPDATE子句	UPDATE子句可以通过VALUES(colname)或者EXCLUDED.colname引用插入的数据。EXCLUDED表示因冲突原本该排除的数据行，语法示例参见**示例**。	-
WHERE子句	● 用于在数据冲突时，判断是否满足指定条件。如果满足，则更新冲突数据；否则忽略。 ● 只有第二种语法形式的冲突更新语法可以指定WHERE子句。即INSERT INTO ON CONFLICT(...) DO UPDATE SET WHERE	-

示例

创建表reason_t2，并向表中插入数据：

```
CREATE TABLE reason_t2
(
  a    int primary key,
  b    int,
```

 版权所有 © 华为云计算技术有限公司

```
 c   int
);

INSERT INTO reason_t2 VALUES (1, 2, 3);
SELECT * FROM reason_t2 ORDER BY 1;
```

```
gaussdb=> SELECT * FROM reason_t2 ORDER BY 1;
 a | b | c
---+---+---
 1 | 2 | 3
(1 row)
```

向表reason_t2中插入两条数据，一条有冲突，一条无冲突。有冲突的数据进行忽略，无冲突的数据进行插入：

```
INSERT INTO reason_t2 VALUES (1, 4, 5),(2, 6, 7) ON CONFLICT(a) DO NOTHING;
SELECT * FROM reason_t2 ORDER BY 1;
```

```
gaussdb=> SELECT * FROM reason_t2 ORDER BY 1;
 a | b | c
---+---+---
 1 | 2 | 3
 2 | 6 | 7
(2 rows)
```

向表reason_t2中插入数据，一条有冲突，一条无冲突。有冲突的数据进行更新，无冲突的数据进行插入，EXCLUDED.b和EXCLUDED.c分别表示引用插入值的b列和c列，本例中为4，5。

```
INSERT INTO reason_t2 VALUES (1, 4, 5),(3, 8, 9) ON CONFLICT(a) DO UPDATE SET b = EXCLUDED.b, c =
EXCLUDED.c;
SELECT * FROM reason_t2 ORDER BY 1;
```

```
gaussdb=> SELECT * FROM reason_t2 ORDER BY 1;
 a | b | c
---+---+---
 1 | 4 | 5
 2 | 6 | 7
 3 | 8 | 9
(3 rows)
```

根据过滤条件筛选被更新的行：

```
INSERT INTO reason_t2 VALUES (2, 7, 8) ON CONFLICT (a) DO UPDATE SET b = excluded.b, c = excluded.c
WHERE reason_t2.c = 7;
SELECT * FROM reason_t2 ORDER BY 1;
```

```
gaussdb=> SELECT * FROM reason_t2 ORDER BY 1;
 a | b | c
---+---+---
 1 | 4 | 5
 2 | 7 | 8
 3 | 8 | 9
(3 rows)
```

向表reason_t中插入数据，有冲突的数据进行更新并调整更新映射关系，即c列更新到b，b列更新到c：

```
INSERT INTO reason_t2 VALUES (1, 2, 3) ON CONFLICT (a) DO UPDATE SET b = excluded.c, c = excluded.b;
SELECT * FROM reason_t2 ORDER BY 1;
```

 版权所有 © 华为云计算技术有限公司

```
gaussdb=> SELECT * FROM reason_t2 ORDER BY 1;
 a | b | c
---+---+---
 1 | 3 | 2
 2 | 7 | 8
 3 | 8 | 9
(3 rows)
```

常见问题

不支持多次更新。插入多组数据间如果有冲突，则会出现类似“INSERT ON CONFLICT DO UPDATE command cannot affect row a second time”报错（**例外：当查询计划是PGXC计划时**），以下是报错示例。

```
CREATE TABLE t1(id int PRIMARY KEY, a int, b int);
SET explain_perf_mode=pretty;
-- STREAM计划
EXPLAIN (COSTS OFF) INSERT INTO t1 VALUES(1,2,3),(1,5,6) ON CONFLICT(id) DO UPDATE SET a =
EXCLUDED.a + 1;
                 QUERY PLAN
-------------------------------------------------
 id |              operation
----+----------------------------------------
  1 | ->  Streaming (type: GATHER)
  2 |    ->  Insert on t1
  3 |       ->  Streaming(type: REDISTRIBUTE)
  4 |          ->  Values Scan on "*VALUES*"
Predicate Information (identified by plan id)
-----------------------------------------------
  2 --Insert on t1
       Conflict Resolution: UPDATE
       Conflict Arbiter Indexes: t1_pkey
  ====== Query Summary =====
-----------------------------
System available mem: 819200KB
Query Max mem: 819200KB
Query estimated mem: 3104KB
(18 rows)
INSERT INTO t1 VALUES(1,2,3),(1,5,6) ON CONFLICT(id) DO UPDATE SET a = EXCLUDED.a + 1;
ERROR:  dn_xxxx: INSERT ON CONFLICT DO UPDATE command cannot affect row a second time
HINT:  Ensure that no rows proposed for insertion within the same command have duplicate constrained
values.
-- 关闭stream，生成PGXC计划
set enable_stream_operator = off;
EXPLAIN (COSTS OFF) INSERT INTO t1 VALUES(1,2,3),(1,5,6) ON CONFLICT(id) DO UPDATE SET a =
EXCLUDED.a + 1;
                QUERY PLAN
-----------------------------------------------
 id |           operation
----+-------------------------------------
  1 | ->  Insert on t1
  2 |    ->  Values Scan on "*VALUES*"
Predicate Information (identified by plan id)
-----------------------------------------------
  1 --Insert on t1
       Conflict Resolution: UPDATE
       Conflict Arbiter Indexes: t1_pkey
       Node expr: id
(11 rows)
INSERT INTO t1 VALUES(1,2,3),(1,5,6) ON CONFLICT(id) DO UPDATE SET a = EXCLUDED.a + 1;
INSERT 0 2
```

 版权所有 © 华为云计算技术有限公司

14.10 UPDATE

功能描述

更新表中的数据。UPDATE修改满足条件的所有行中指定的字段值，WHERE子句声明条件，SET子句指定的字段会被修改，没有出现的字段则保持原值。

注意事项

- 使用UPDATE请遵循**UPDATE操作规范**。
- **权限要求**：要修改表，用户必须对该表有UPDATE权限，对expression或condition条件里涉及到的任何表要有SELECT权限。
- **不允许对表的分布列（distribute column）进行修改**。
- UPDATE操作频繁的表不建议创建为复制表。
- **列存表约束**：
  - 对于列存表，暂时不支持RETURNING子句。
  - 列存表不支持结果不确定的更新（non-deterministic update）。试图对列存表用多行数据更新一行时会报错。
  - 列存表的更新操作，旧记录空间不会回收，需要执行VACUUM FULL table_name进行清理。
  - 对于列存表，支持轻量化UPDATE操作。轻量化UPDATE只重写更新列，减少空间使用量。列存轻量化UPDATE通过GUC参数enable_light_colupdate控制是否开启，默认关闭。
  - 列存轻量化UPDATE在以下场景不能使用：更新索引列，更新主键列，更新分区列，更新PCK列和在线扩容，会自动转化为普通UPDATE方式。
  - 列存轻量化UPDATE与后台列存AUTOVACUUM并发会小概率报错，可以通过ALTER TABLE设置表级参数enable_column_autovacuum_garbage为off来避免。需要注意的是设置表级参数enable_column_autovacuum_garbage为off会关闭该表的后台列存AUTOVACUUM。

  HStore表约束：
  - 与列存一样，当前版本HStore上的UPDATE操作始终先DELETE再INSERT。全局GUC参数可控制打开HStore的轻量化UPDATE，当前版本默认关闭。
  - 对于并发UPDATE场景， 传统列存储格式由于同时操作相同CU时会产生锁冲突所以并发性能较差， 对于HStore表由于不需要阻塞等待，并发UPDATE性能可达到列存的百倍以上。

语法格式

```
   UPDATE [/*+ plan_hint */] [ ONLY ] table_name [ * ] [ [ AS ] alias ]
   [ PARTITION ( partition_name ) | PARTITION FOR ( partition_key_value [, ...] ) ]
SET {column_name = { expression | DEFAULT }
   |( column_name [, ...] ) = {( { expression | DEFAULT } [, ...] ) |sub_query }}[, ...]
   [ FROM from_list] [ WHERE condition ]
   [ RETURNING {*
            | {output_expression [ [ AS ] output_name ]} [, ...] }];

where sub_query can be:
SELECT [ ALL | DISTINCT [ ON ( expression [, ...] ) ] ]
{ * | {expression [ [ AS ] output_name ]} [, ...] }
[ FROM from_item [, ...] ]
```

 版权所有 © 华为云计算技术有限公司

```
[ WHERE condition ]
[ GROUP BY grouping_element [, ...] ]
[ HAVING condition [, ...] ]
```

参数说明

表 14-14 UPDATE 参数说明

参数	描述	示例
plan_hint子句	以/*+ */的形式在关键字后，用于对指定语句块生成的计划进行hint调优，详细用法请参见**使用Plan Hint进行调优**。	-
table_name	要更新的表名，可以带上schema名。	test.table1，其中test为schema名。
alias	为表指定一个简短的别名，方便引用，例如alias。字符串，符合标识符命名规范，详情请参见**标识符命名规范**。	更新别名为e的表employees： `UPDATE employees AS e SET e.salary = e.salary * 1.1 WHERE e.department = 'technology';`
partition_name	分区名称，该子句指定要更新的分区名称对应的数据，适用于范围分区、列表分区等已命名的分区表。仅8.2.1及以上集群版本支持该选项。	更新分区名为p1的数据： `UPDATE test_range_row PARTITION(p1) SET d = 41;`
partition_key_value	分区键值，该子句指定要更新的分区键值对应的数据。 通过PARTITION FOR (partition_key_value [, ...])子句指定的这一组值，可以唯一确定一个分区。	更新分区键值为41的数据行，将相应的b值改为20。 `UPDATE test_range_row PARTITION FOR (41) SET b = 20;`

column_name	要修改的字段名。
	支持使用目标表的表名加字段名来引用这个字段。例如：
	`UPDATE foo SET foo.col_name = 'DWS';`
	支持使用目标表的别名加字段名来引用这个字段。例如：
	`UPDATE foo AS f SET f.col_name = 'DWS';`
		-

expression	赋予字段的值或表达式。	给商品价格打8折： `UPDATE products SET price = price * 0.8 WHERE product_id = 101;`

版权所有 © 华为云计算技术有限公司

参数	描述	示例
DEFAULT	用对应字段的缺省值填充该字段。如果没有缺省值，则为NULL。	重置非活跃用户的状态和登录次数： UPDATE users SET status = DEFAULT, login_count = DEFAULT WHERE last_login < CURRENT_DATE - INTERVAL '30 days';
sub_query	子查询。使用同一个数据库里的其他表的查询结果，来更新目标表。 其中SELECT子句具体介绍请参考**SELECT**。	使用子查询更新部门平均工资： UPDATE departments d SET avg_salary = (SELECT AVG(salary) FROM employees e WHERE e.dept_id = d.dept_id);
from_list	另一个表的表达式列表，允许在WHERE条件里关联其他表字段来更新目标表。与在一个SELECT语句的FROM子句里声明表列表类似。 **须知** 目标表绝对不能出现在from_list里，除非在使用一个自连接（此时它必须以from_list的别名出现）。	**示例：使用FROM from_list子句，优先处理VIP等级高的客户的订单**
condition	跟在WHERE后面的一个返回boolean类型结果的表达式。只有这个表达式返回true的行才会被更新。	只更新数学成绩低于80分的学生： UPDATE student_scores SET math_score = math_score + 5, update_time = NOW() WHERE math_score < 80;
output_expression	在所有需要更新的行都被更新之后，用于计算返回值的表达式，便于确认更新结果。 取值范围：使用任何table以及FROM中列出的表的字段。*表示返回所有字段。	更新并返回修改前后的库存信息： UPDATE inventory SET stock = stock - 10 WHERE category = '电子产品' RETURNING item_id AS "商品ID", name AS "商品名称", stock AS "新库存量", stock + 10 AS "原库存量";
output_name	给RETURNING返回的列起的别名，将输出结果优化。字符串，符合标识符命名规范，详情请参见**标识符命名规范**。	-

版权所有 © 华为云计算技术有限公司

示例：常用 UPDATE 场景

创建样例表并插入数据：

```
DROP TABLE IF EXISTS products;
CREATE TABLE products (
    product_id INT PRIMARY KEY,         --商品ID
    name VARCHAR(100),                  --商品名称
    price DECIMAL(10,2),                --商品价格
    stock INT,                          --商品库存
    last_updated TIMESTAMP              --数据更新时间
);

INSERT INTO products VALUES
(101, 'Phone', 2999.00, 50, NULL),
(102, 'PC', 5999.00, 30, NULL),
(103, 'PAD', 1999.00, 20, NULL);
```

不含WHERE子句表示直接更新所有记录的值：所有产品价格上浮20%。

```
UPDATE products SET price = price * 1.2 , last_updated = NOW();
SELECT * FROM products ORDER BY 1;
```

product_id	name	price	stock	last_updated
101	Phone	3598.80	50	2025-05-27 21:44:13.759399
102	PC	7198.80	30	2025-05-27 21:44:13.759399
103	PAD	2398.80	20	2025-05-27 21:44:13.759399

使用WHERE子句更新指定的数据：将平板电脑的价格上涨500 USD。

```
UPDATE products SET price = price + 500 , last_updated = NOW()
WHERE name = 'PAD';
SELECT * FROM products ORDER BY 1;
```

product_id	name	price	stock	last_updated
101	Phone	3598.80	50	2025-05-27 21:44:13.759399
102	PC	7198.80	30	2025-05-27 21:44:13.759399
103	PAD	2898.80	20	2025-05-27 21:47:20.690807

使用WHERE子句更新特定范围的数据：商品清库存，需要将库存大于40的商品打8折处理。

```
UPDATE products SET price = price * 0.8 , last_updated = NOW()
WHERE stock > 40;
SELECT * FROM products ORDER BY 1;
```

product_id	name	price	stock	last_updated
101	Phone	2879.04	50	2025-05-27 21:53:31.996656
102	PC	7198.80	30	2025-05-27 21:44:13.759399
103	PAD	2898.80	20	2025-05-27 21:47:20.690807

版权所有 © 华为云计算技术有限公司

在一个UPDATE命令中更新多个字段，方法是在SET子句中列出更多赋值：调整PC的价格和库存。

```
UPDATE products SET price = price - 1000 , stock = stock - 10,  last_updated = NOW()
WHERE name = 'PC';
SELECT * FROM products ORDER BY 1;
```

product_id	name	price	stock	last_updated
101	Phone	2879.04	50	2025-05-27 21:53:31.996656
102	PC	6198.80	20	2025-05-27 22:00:07.851013
103	PAD	2898.80	20	2025-05-27 21:47:20.690807

示例：更新分区表的指定分区

创建样例表并导入数据：

```
DROP TABLE IF EXISTS test_range_row;
CREATE TABLE test_range_row(a int, b int,d int)
DISTRIBUTE BY hash(a) PARTITION BY RANGE(d)
(
    PARTITION p1 values LESS THAN (60),
    PARTITION p2 values LESS THAN (75),
    PARTITION p3 values LESS THAN (90),
    PARTITION p4 VALUES LESS THAN (maxvalue)
);

INSERT INTO test_range_row PARTITION(p1) VALUES(55,20,51);
INSERT INTO test_range_row PARTITION(p1) VALUES(54,18,59);
INSERT INTO test_range_row PARTITION(p2) VALUES(15,18,63);
INSERT INTO test_range_row PARTITION(p3) VALUES(12,21,80);
INSERT INTO test_range_row PARTITION(p3) VALUES(83,21,85);
INSERT INTO test_range_row PARTITION(p4) VALUES(36,21,97);

SELECT * FROM test_range_row ORDER BY d;
```

a	b	d
55	20	51
54	18	59
15	18	63
12	21	80
83	21	85
36	21	97

总条数：6

版权所有 © 华为云计算技术有限公司

更新分区名称为p1的行数据，将d值修改为41。

```
UPDATE test_range_row PARTITION(p1) SET d = 41;
SELECT * FROM test_range_row ORDER BY d;
```

a	b	d
54	18	41
55	20	41
15	18	63
12	21	80
83	21	85
36	21	97

总条数：6

从以上结果看，分区为p1，即d<60的行数据，相应的d值都被修改为41。

继续执行以下SQL，更新分区键值为41的数据行，将相应的b值改为20。

```
UPDATE test_range_row PARTITION FOR (41) SET b = 20;
SELECT * FROM test_range_row ORDER BY d;
```

a	b	d
54	20	41
55	20	41
15	18	63
12	21	80
83	21	85
36	21	97

总条数：6

 版权所有 © 华为云计算技术有限公司

示例：使用 FROM from_list 子句，优先处理 VIP 等级高的客户的订单

需要更新目标表：订单表orders，但是需要使用FROM子句关联另一张表customers的客户ID字段，来获取客户等级。

创建客户信息表和订单表，包含的字段如下：

```
DROP TABLE IF EXISTS customers;
CREATE TABLE customers (                --创建客户信息表
    customer_id INT PRIMARY KEY,         --客户ID
    name VARCHAR(100),                  --客户姓名
    vip_level INT                      --VIP等级
);
DROP TABLE IF EXISTS orders;
CREATE TABLE orders (                   --创建订单表
    order_id INT PRIMARY KEY,            --订单号
    customer_id INT ,                  --客户ID
    amount DECIMAL(10,2),                --消费金额
    status VARCHAR(20)                 --订单状态
);
```

插入测试数据，订单状态均为pending，即三个客户的订单都未处理。

```
INSERT INTO customers VALUES
(1, '张三', 1),
(2, '李四', 3),
(3, '王五', 2);
INSERT INTO orders VALUES
(1001, 1, 500.00, 'pending'),
(1002, 2, 1200.00, 'pending'),
(1003, 3, 800.00, 'pending');
SELECT * FROM orders;
```

order_id	customer_id	amount	status
1002	2	1200.00	pending
1001	1	500.00	pending
1003	3	800.00	pending

总条数：3

由于订单较多，现要求优先处理VIP等级高的客户的订单，此时需要通过FROM from_list子句，关联客户表的客户ID，获取客户VIP等级，本例为，优先处理VIP等级为3级以上的客户订单。

```
UPDATE orders o
SET status = 'priority'      --将订单状态修改为“高优先级”
FROM customers c
WHERE o.customer_id = c.customer_id AND c.vip_level >= 3;  --从客户表中获取等级VIP为3级以上的客户
SELECT * FROM orders;
```

 版权所有 © 华为云计算技术有限公司

order_id	customer_id	amount	status
1002	2	1200.00	priority
1001	1	500.00	pending
1003	3	800.00	pending

总条数：3

示例：使用 DEFAULT 重置非活跃用户的状态

创建用户表（带默认值），并插入数据：

```
DROP TABLE IF EXISTS users;
CREATE TABLE users (
user_id INT PRIMARY KEY,                --用户ID
username VARCHAR(50) ,                   --用户名
login_count INT DEFAULT 0,               --登录次数，默认为0
last_login TIMESTAMP,                   --最后一次登录的时间
status VARCHAR(20) DEFAULT 'active');     --用户状态，默认为active

INSERT INTO users (user_id, username, login_count, last_login) VALUES
(1, 'user1', 5, '2023-06-01 10:00:00'),
(2, 'user2', 3, '2023-06-15 14:30:00');
SELECT * FROM users;
```

user_id	username	login_count	last_login	status
1	user1	5	2023-06-01 10:00:00	active
2	user2	3	2023-06-15 14:30:00	active

总条数：2

重置非活跃用户的状态和登录次数：

```
UPDATE users SET
status = DEFAULT,      -- 重置为'active'
login_count = DEFAULT   -- 重置为0
WHERE last_login < CURRENT_DATE - INTERVAL '30 days';
SELECT * FROM users;
```

user_id	username	login_count	last_login	status
1	user1	0	2023-06-01 10:00:00	active
2	user2	0	2023-06-15 14:30:00	active

常见问题

- **执行SQL时报错：abort transaction due to concurrent update**

版权所有 © 华为云计算技术有限公司

- 如何判断表是否执行过UPDATE或DELETE
- 更新报错ERROR：Non-deterministic UPDATE

14.11 VALUES

功能描述

根据给定的值表达式计算一个或一组行的值。它通常用于在一个较大的命令内生成一个“常数表”。

注意事项

- 应当避免使用VALUES返回数量非常大的结果行，否则可能会遭遇内存耗尽或者性能低下。出现在INSERT中的VALUES是一个特殊情况，因为目标字段类型可以从INSERT的目标表获知，并不需要通过扫描VALUES列表来推测，所以此情况下可以处理非常大的结果行。
- 如果指定了多行，那么每一行都必须拥有相同的元素个数。

语法格式

```
VALUES {( expression [, ...] )} [, ...]
    [ ORDER BY { sort_expression [ ASC | DESC | USING operator ] } [, ...] ]
    [ { [ LIMIT { count | ALL } ] [ OFFSET start [ ROW | ROWS ] ] } | { LIMIT start, { count | ALL } } ]
    [ FETCH { FIRST | NEXT } [ count ] { ROW | ROWS } ONLY ];
```

参数说明

表 14-15 VALUES 参数说明

参数	描述	取值范围
expression	用于计算或插入结果表指定地点的常量或者表达式。 在一个出现在INSERT顶层的VALUES列表中，expression可以被DEFAULT替换以表示插入目的字段的缺省值。除此以外，VALUES在其他场景时是不能使用DEFAULT的。	-
sort_expression	一个表示如何排序结果行的表达式或者整数常量。	-
ASC	指定按照升序排列。	-
DESC	指定按照降序排列。	-
operator	一个排序操作符。	-
count	返回的最大行数。	-
start	开始返回行之前忽略的行数。	-

版权所有 © 华为云计算技术有限公司

参数	描述	取值范围
FETCH { FIRST | NEXT } [count] { ROW | ROWS } ONLY	FETCH子句限定返回查询结果从第一行开始的总行数。	count的缺省值为1。

示例：插入单条或多条数据

创建表reason_t1：

```
DROP TABLE IF EXISTS reason_t1;
CREATE TABLE reason_t1
(
    TABLE_SK          INTEGER               ,
    TABLE_ID          VARCHAR(20)           ,
    TABLE_NA          VARCHAR(20)
);
```

向表中插入一条数据：

```
INSERT INTO reason_t1(TABLE_SK, TABLE_ID, TABLE_NA) VALUES (1, 'S01', 'StudentA');
```

向表中插入一条数据，和上一条语法等效：

```
INSERT INTO reason_t1 VALUES (1, 'S01', 'StudentA');
```

向表中插入TABLE_SK小于1的数据：

```
INSERT INTO reason_t1 SELECT * FROM reason_t1 WHERE TABLE_SK < 1;
```

向表中插入多条数据：

```
INSERT INTO reason_t1 VALUES (1, 'S01', 'StudentA'),(2, 'T01', 'TeacherA'),(3, 'T02', 'TeacherB');
SELECT * FROM reason_t1 ORDER BY 1;
```

table_sk	table_id	table_na
1	S01	StudentA
1	S01	StudentA
1	S01	StudentA
2	T01	TeacherA
3	T02	TeacherB

版权所有 © 华为云计算技术有限公司

15 DCL 语法

15.1 DCL 语法一览表

DCL（Data Control Language数据控制语言），是用来设置或更改数据库用户或角色权限的语句。

授权

DWS提供了针对数据对象和角色授权的语句，请参考**GRANT**。

收回权限

DWS提供了收回权限的语句，请参考**REVOKE**。

设置默认权限

DWS允许设置应用于将来创建的对象的权限，请参考**ALTER DEFAULT PRIVILEGES**。

15.2 ALTER DEFAULT PRIVILEGES

功能描述

设置应用于将来要创建的对象的权限（不会影响现有对象的权限）。

用户只可以修改由用户本身或者用户本身所属的角色所创建的对象的默认权限，这些权限可以对全局范围设置（即数据库中创建的所有对象），也可以为指定模式下的对象设置。

查看有关数据库用户的默认权限的信息，可以查询PG_DEFAULT_ACL系统表。

注意事项

目前只支持表（包括视图）、序列、函数和类型（包括域）的权限更改。

版权所有 © 华为云计算技术有限公司

语法格式

```
ALTER DEFAULT PRIVILEGES
   [ FOR { ROLE | USER } target_role [, ...] ]
   [ IN SCHEMA schema_name [, ...] ]
   abbreviated_grant_or_revoke;
```

abbreviated_grant_or_revoke子句用于指定对哪些对象进行授权或回收权限，可以是以下选项之一：

- grant_on_tables_clause
- grant_on_functions_clause
- grant_on_types_clause
- grant_on_sequences_clause
- revoke_on_tables_clause
- revoke_on_functions_clause
- revoke_on_types_clause
- revoke_on_sequences_clause

grant_on_tables_clause子句用于对表授权。

```
GRANT { { SELECT | INSERT | UPDATE | DELETE | TRUNCATE | REFERENCES | TRIGGER | ANALYZE | ANALYSE
| VACUUM | ALTER | DROP }
   [, ...] | ALL [ PRIVILEGES ] }
   ON TABLES
   TO { [ GROUP ] role_name | PUBLIC } [, ...]
   [ WITH GRANT OPTION ]
```

grant_on_functions_clause子句用于对函数授权。

```
GRANT { { EXECUTE | ALTER | DROP } [, ...] | ALL [ PRIVILEGES ] }
   ON FUNCTIONS
   TO { [ GROUP ] role_name | PUBLIC } [, ...]
   [ WITH GRANT OPTION ]
```

grant_on_types_clause子句用于对类型授权。

```
GRANT { USAGE | ALL [ PRIVILEGES ] }
   ON TYPES
   TO { [ GROUP ] role_name | PUBLIC } [, ...]
   [ WITH GRANT OPTION ]
```

grant_on_sequences_clause子句用于对序列授权。

```
GRANT { { USAGE | SELECT | UPDATE }
   [, ...] | ALL [ PRIVILEGES ] }
   ON SEQUENCES
   TO { [ GROUP ] role_name | PUBLIC } [, ...]
   [ WITH GRANT OPTION ]
```

revoke_on_tables_clause子句用于回收表对象的权限。

```
REVOKE [ GRANT OPTION FOR ]
   { { SELECT | INSERT | UPDATE | DELETE | TRUNCATE | REFERENCES | TRIGGER | ANALYZE | ANALYSE |
VACUUM | ALTER | DROP }
   [, ...] | ALL [ PRIVILEGES ] }
   ON TABLES
   FROM { [ GROUP ] role_name | PUBLIC } [, ...]
   [ CASCADE | RESTRICT | CASCADE CONSTRAINTS ]
```

revoke_on_functions_clause子句用于回收函数的权限。

```
REVOKE [ GRANT OPTION FOR ]
   { { EXECUTE | ALTER | DROP } [, ...] | ALL [ PRIVILEGES ] }
   ON FUNCTIONS
   FROM { [ GROUP ] role_name | PUBLIC } [, ...]
   [ CASCADE | RESTRICT | CASCADE CONSTRAINTS ]
```

revoke_on_types_clause子句用于回收类型的权限。

 版权所有 © 华为云计算技术有限公司

```
REVOKE [ GRANT OPTION FOR ]
    { USAGE | ALL [ PRIVILEGES ] }
    ON TYPES
    FROM { [ GROUP ] role_name | PUBLIC } [, ...]
    [ CASCADE | RESTRICT | CASCADE CONSTRAINTS ]
```

revoke_on_sequences_clause子句用于回收序列的权限。

```
REVOKE [ GRANT OPTION FOR ]
    { { USAGE | SELECT | UPDATE }
    [, ...] | ALL [ PRIVILEGES ] }
    ON SEQUENCES
    FROM { [ GROUP ] role_name | PUBLIC } [, ...]
    [ CASCADE | RESTRICT | CASCADE CONSTRAINTS ]
```

参数说明

表 15-1 ALTER DEFAULT PRIVILEGES 参数说明

参数	描述	取值范围或示例

FOR { ROLE | USER } target_role [, ...]	FOR ROLE/USER 后面跟着的target_role是指将来要创建对象的角色或用户名称，即要访问谁未来创建的对象，就写这个owner对应的角色或用户名称。如果省略FOR ROLE/USER，则缺省值为当前角色/用户。
	target_role必须有schema_name的CREATE权限。查看角色/用户是否具有schema的CREATE权限可使用has_schema_privilege函数。
	`SELECT a.rolname, n.nspname FROM pg_authid as a, pg_namespace as n WHERE has_schema_privilege(a.oid, n.oid, 'CREATE');`
		已有角色或用户的名称。
		举例，授权用户jim可访问用户lily未来创建对象的权限。
		`ALTER DEFAULT PRIVILEGES FOR ROLE lily IN SCHEMA lily GRANT SELECT ON tables TO jim;`
		或者
		`ALTER DEFAULT PRIVILEGES FOR USER lily IN SCHEMA lily GRANT SELECT ON tables TO jim;`

schema_name	现有模式的名称。 如果指定了模式名，那么之后在这个模式下面创建的所有对象默认的权限都会被修改。如果IN SCHEMA被省略，那么全局权限会被修改。 如果模式名为public，需要指定for user target_role。	现有模式的名称。
role_name	被授予或者取消权限角色的名称。 **说明** 如果想删除一个被赋予了默认权限的角色，必须撤销其默认权限的更改或者使用DROP OWNED BY删除该角色的默认权限记录。	已有的角色名称。

示例

- **前提条件**：创建测试模式tpcds、表和用户，为以下示例准备样例数据。

```
CREATE SCHEMA tpcds;
CREATE TABLE tpcds.teama (a int,b int);
CREATE TABLE tpcds.teamb (a int,b int);
DROP USER IF EXISTS test1;
DROP USER IF EXISTS test2;
```

 版权所有 © 华为云计算技术有限公司

```
CREATE USER test1 PASSWORD '{Password}';
CREATE USER test2 PASSWORD '{Password}';
CREATE USER jack PASSWORD '{Password}';
```

- 将创建在模式tpcds里的所有表（和视图）的SELECT权限授予每一个用户：
  ```
  ALTER DEFAULT PRIVILEGES IN SCHEMA tpcds GRANT SELECT ON TABLES TO PUBLIC;
  ```
- 将tpcds下的所有表的插入权限授予用户jack：
  ```
  ALTER DEFAULT PRIVILEGES IN SCHEMA tpcds GRANT INSERT ON TABLES TO jack;
  ```
- 撤销上述权限：
  ```
  ALTER DEFAULT PRIVILEGES IN SCHEMA tpcds REVOKE SELECT ON TABLES FROM PUBLIC;
  ALTER DEFAULT PRIVILEGES IN SCHEMA tpcds REVOKE INSERT ON TABLES FROM jack;
  ```
- 假设有两个用户test1、test2，如果想要test2用户对test1用户未来创建的表都有查询权限可以用如下操作：
  - 首先，把test1的schema的权限赋权给test2用户。
    ```
    GRANT usage, create ON SCHEMA test1 TO test2;
    ```
  - 其次，把test1用户下的表的查询权限赋值给test2用户。**注**，以下的FOR USER也可以改成FOR ROLE，作用效果一致。
    ```
    ALTER DEFAULT PRIVILEGES FOR USER test1 IN SCHEMA test1 GRANT SELECT ON tables TO test2;
    ```
  - 然后，test1用户创建表。
    ```
    SET ROLE test1 password '{password}';
    CREATE TABLE test3( a int, b int);
    ```
  - 最后，用test2用户去查询。
    ```
    SET ROLE test2 password '{password}';
    SELECT * FROM test1.test3;
     a | b
    ---+---
    (0 rows)
    ```

相关链接

GRANT，**REVOKE**

15.3 ANALYZE | ANALYSE

功能描述

用于收集有关数据库中表内容的统计信息，统计信息分为表级和列级。其中，表级统计信息存储在PG_CLASS中，列级统计结果存储在系统表PG_STATISTIC下。执行计划生成器会使用这些统计数据，以确定最有效的执行计划。

它的主要作用是：

- **优化查询性能**：为查询优化器提供数据分布信息，帮助生成更高效的执行计划。
- **提高查询计划准确性**：使优化器能更准确地估算不同查询路径的成本。
- **支持索引选择**：帮助优化器决定是否使用索引以及使用哪个索引。

需要执行 ANALYZE 的场景

- **大规模数据变化**：

  大规模数据导入/UPDATE/DELETE等操作，会导致表数据行数变化，新增的大量数据也会导致数据特征发生大的变化，此时需要对表重新收集统计信息。
- **查询新增数据**：

 版权所有 © 华为云计算技术有限公司

常见于业务表新增数据查询场景，这个也是客户现场中最常出现的统计信息没有及时更新而导致的数据查询不准确的问题，这种场景最主要的特征如下：

- 存在一个按照时间增长的业务表。
- 业务表每天入库新一天的数据。
- 数据入库之后查询新增数据进行数据加工分析。

在最后步骤的数据加工分析时，最常用的方法就是使用Filter条件从分区表中筛选数据，如

```
passtime > '2020-01-19 00:00:00' AND passtime < '2020-01-20 00:00:00'
```

假如新增数据入库之后没有执行ANALYZE，优化器发现Filter条件中的passtime取值范围超过了统计信息中记录的passtime值的上边界，会把估算满足passtime > '2020-01-19 00:00:00' AND passtime < '2020-01-20 00:00:00' 的tuple个数为1条，导致估算行数验证失真。

注意事项

- 如果没有指定参数，ANALYZE会分析当前数据库中的每个表和分区表。同时也可以通过指定table_name、column和partition_name参数把分析限定在特定的表、列或分区表中。
- 能够执行ANALYZE特定表的用户，包括表的所有者、表所在数据库的所有者、通过GRANT被授予该表上ANALYZE权限的用户或者被授予了gs_role_analyze_any角色的用户以及有SYSADMIN属性的用户。
- 在百分比采样收集统计信息时，用户需要被授予ANALYZE和SELECT权限。
- 开启分区统计信息参数enable_analyze_partition，对分区表设置表级incremental_analyze参数，会先对无统计信息或有数据变化的分区执行ANALYZE，然后通过合并分区统计信息方式生成分区主表的统计信息。
- 开启analyze_options参数（9.1.1.100及以上集群版本支持）中的analyze_on_vw选项，支持对列存3.0表在弹性VW上执行ANALYZE，在弹性VW的数据节点上完成数据采样，在CN上使用采样数据进行统计信息的计算。
- 仅8.1.1及以上集群版本支持在匿名块、事务块、函数或存储过程内对单表进行ANALYZE操作。
- 对于ANALYZE全库，库中各表的ANALYZE处于不同的事务中，所以不支持在匿名块、事务块、函数或存储过程内对全库执行ANALYZE。
- 统计信息的回滚操作不支持PG_CLASS中相关字段的回滚。
- 全局临时表在每个会话的ANALYZE的统计信息独立。每个会话都有自己的统计信息。
- ANALYZE|ANALYSE VERIFY用于检测数据库中普通表（行存表、列存表）的数据文件是否损坏，目前此命令暂不支持HDFS表。
- ANALYZE VERIFY操作处理的大多为异常场景检测需要使用RELEASE版本。ANALYZE VERIFY场景不触发远程读，因此远程读参数不生效。对于关键系统表出现错误被系统检测出页面损坏时，将直接报错不再继续检测。

⚠ 警告

- 单次新增、修改量占表总量10%以上场景，需在业务中增加显式Analyze操作。
- 更多开发设计规范参见**总体开发设计规范**。

版权所有 © 华为云计算技术有限公司

语法格式

- 收集表的统计信息。
```
{ ANALYZE | ANALYSE } [ { VERBOSE | LIGHT | FORCE | PREDICATE } ]
    [ table_name [ ( column_name [, ...] ) ] ];
```
- 收集分区的统计信息。
```
{ ANALYZE | ANALYSE } [ { VERBOSE | LIGHT | FORCE } ]
    [ table_name [ ( column_name [, ...] ) ] ]
    PARTITION ( partition_name ) ;
```

注意

- 8.2.1.200及以上集群版本支持通过设置GUC参数enable_analyze_partition控制是否允许对表的指定分区收集统计信息。**该参数开启后，分区表的统计信息获取方式将发生变化。如需开启请联系技术支持。**
- 8.2.1.200以下的集群版本，普通分区表语法上支持对表的指定分区收集统计信息，但功能上暂未实现。对指定分区执行ANALYZE，会返回WARNING提示。
- 不支持使用临时采样表来收集分区的统计信息。
- 不支持分区上的多列统计信息和表达式统计信息。

- 收集外表的统计信息。
```
{ ANALYZE | ANALYSE } [ VERBOSE ]
    { foreign_table_name | FOREIGN TABLES };
```
- 收集多列统计信息
```
{ANALYZE | ANALYSE} [ VERBOSE ]
    table_name (( column_1_name, column_2_name [, ...] ));
```

说明

- 收集多列统计信息时，请设置GUC参数default_statistics_target为负数，以使用百分比采样方式。
- 每组多列统计信息最多支持32列。
- 不支持收集多列统计信息的表：系统表、HDFS外表复制表、全局临时表。

- 检测当前库的数据文件
```
{ANALYZE | ANALYSE} VERIFY {FAST|COMPLETE|DISTRIBUTION};
```

说明

- 支持对全库进行操作，由于涉及的表较多，建议以重定向保存结果**gsql -d database -p port -f "verify.sql"> verify_warning.txt 2>&1**。
- 不支持HDFS表（内表和外表），不支持临时表和UNLOGGED表。
- 只有外部表提示NOTICE，内部表的检测会包含在它所依赖的外部表，不对外显示。
- 该语句支持容错处理。
- 对于全库操作时，当关键系统表出现损坏则直接报错，不再继续执行。
- DISTRIBUTION选项仅9.1.1.200及以上集群版本支持。

- 检测表和索引的数据文件
```
{ANALYZE | ANALYSE} VERIFY {FAST|COMPLETE|DISTRIBUTION} table_name|index_name [CASCADE];
```


版权所有 © 华为云计算技术有限公司

说明

- 支持对普通表的操作和索引表的操作，但不支持对索引表index使用CASCADE操作。原因是CASCADE模式用于处理主表的所有索引表，当单独对索引表进行检测时，无需使用CASCADE模式。
- 不支持HDFS表（内表和外表），不支持临时表和UNLOGGED表。
- 对于主表的检测会同步检测主表的内部表，例如toast表、cudesc表等。
- 当提示索引表损坏时，建议使用reindex命令进行重建索引操作。
- DISTRIBUTION选项仅9.1.1.200及以上集群版本支持，DISTRIBUTION不支持CASCADE模式。

- 检测表分区的数据文件

```
{ANALYZE | ANALYSE} VERIFY {FAST|COMPLETE|DISTRIBUTION} table_name PARTITION {(partition_name)} [CASCADE];
```

说明

- 支持对表的单独分区进行检测操作，但不支持对索引表index使用CASCADE操作。
- 不支持HDFS表（内表和外表），不支持临时表和UNLOGGED表。
- DISTRIBUTION选项仅9.1.1.200及以上集群版本支持，DISTRIBUTION不支持CASCADE模式。

参数说明

表 15-2 ANALYZE | ANALYSE 参数说明

参数	描述	取值范围
VERBOSE	启用VERBOSE后，ANALYZE会发出进度信息，表明目前正在处理的表。各种有关表的统计信息也会打印出来。	-
LIGHT	轻量化模式下对表收集的统计信息保存到内存中，不写入系统表，执行时对表加一级锁，该选项请加括号使用，例如ANALYZE (LIGHT) t1。	-
FORCE	FORCE模式支持表的统计信息被锁定的情况下进行强制执行ANALYZE，该选项请加括号使用，例如ANALYZE (FORCE) t1。	-
PREDICATE	PREDICATE模式表示只计算当前识别到的**谓词列**的统计信息，该参数仅9.1.0.100及以上集群版本支持。是否支持谓词列的ANALYZE操作由GUC参数**analyze_predicate_column_threshold**控制，取值为0表示关闭，大于0表示开启谓词列收集功能，且仅对列数大于等于此取值的表进行谓词列ANALYZE。 该选项请加括号使用，例如ANALYZE (PREDICATE) xxx。 谓词信息是在查询解析阶段收集，动态采样也支持谓词列采样。	-

版权所有 © 华为云计算技术有限公司

参数	描述	取值范围
table_name	需要分析的特定表的表名（可能会带模式名），如果省略，将对数据库中的所有表（非外部表）进行分析。 对于ANALYZE收集统计信息，目前仅支持行存表、列存表、HDFS表、ORC格式的OBS外表、CARBONDATA格式的OBS外表、协同分析的外表。	表名长度不超过63个字节，以字母或下划线开头，可包含字母、数字、下划线、$、#。
column_name	需要分析特定列的列名，默认为所有列。	已有的列名，格式为：column_1_name，column_2_name，column_3_name....
partition_name	表示分析该分区表的统计信息。目前语法上支持分区表执行ANALYZE，但功能实现上暂不支持对指定分区统计信息的分析。	已有的分区表名。
foreign_table_name	需要分析的特定外表的表名（可能会带模式名），该表的数据存放于HDFS分布式文件系统中。	已有的外表名。
FOREIGN TABLES	分析所有当前用户权限下，数据位于HDFS分布式文件系统中的HDFS外表。	已有的HDFS外表。
index_name	需要分析的特定索引表的表名（可能会带模式名）。	已有的索引名。
FAST|COMPLETE|DISTRIBUTION	不同模式下，对行存表、列存表、哈希分布表的数据分布进行校验。 行存表： ● FAST模式下，主要对于行存表的CRC和page header进行校验，如果校验失败则会告警。 ● COMPLETE模式下，主要对行存表的指针、tuple进行解析校验。 列存表： ● FAST模式下，主要对于列存表的CRC和magic进行校验，如果校验失败则会告警。 ● COMPLETE模式下，主要对列存表的CU进行解析校验。 哈希分布表： DISTRIBUTION模式下，主要对于哈希分布表在DN上的数据分布正确性进行校验，如果校验失败则会告警。	-
CASCADE	CASCADE模式下会对当前表的所有索引进行检测处理。	-

版权所有 © 华为云计算技术有限公司

示例

- 使用ANALYZE语句更新表customer_info统计信息。
```
DROP TABLE IF EXISTS customer_info;
CREATE TABLE customer_info (name varchar(50), email varchar(100));
ANALYZE customer_info;
```
- 使用ANALYZE VERBOSE语句更新表customer_info统计信息，并输出表的相关信息。
```
ANALYZE VERBOSE customer_info;
```
```
gaussdb=> ANALYZE VERBOSE customer_info;
NOTICE:  dn_6005_6006: (dn_6005_6006 pid=48199)analyzing "public.customer_info"
NOTICE:  dn_6001_6002: (dn_6001_6002 pid=51641)analyzing "public.customer_info"
NOTICE:  dn_6003_6004: (dn_6003_6004 pid=105076)analyzing "public.customer_info"
NOTICE:  dn_6005_6006: (dn_6005_6006 pid=48199)analyzing "public.customer_info"
NOTICE:  dn_6001_6002: (dn_6001_6002 pid=51641)analyzing "public.customer_info"
NOTICE:  dn_6003_6004: (dn_6003_6004 pid=105076)analyzing "public.customer_info"
INFO:  (cn_5003 pid=50567)analyzing "public.customer_info"
ANALYZE
```
- 更新表customer_info中特定列name，email的统计信息。
```
ANALYZE customer_info (name, email);
```
- 使用轻量化模式更新表customer_info的统计信息。
```
ANALYZE (LIGHT) customer_info;
```
- 使用强制模式更新表customer_info的统计信息，即使表已被锁定。
```
ANALYZE (FORCE) customer_info;
```
- 只更新谓词列的统计信息，只会对当前为止收集到的谓词列进行采样，最后在查询业务稳定后使用。
```
ANALYZE (PREDICATE) customer_info;
```
通过函数**pg_stat_get_predicate_columns**可以查询当前已收集了表的哪些谓词列，注意以下示例的表名为“customer_info”，实际执行时请替换表名。
```
SELECT pr.attnum,pa.attname FROM
pg_catalog.pg_stat_get_predicate_columns('customer_info'::regclass) pr LEFT JOIN pg_attribute pa ON
pa.attrelid='customer_info'::regclass AND pa.attnum = pr.attnum;
```

15.4 DEALLOCATE

功能描述

DEALLOCATE用于删除先前编写的预备语句。如果未显式删除一个预备语句，则在会话结束时将其删除。

有关预备语句可参考**PREPARE**。

注意事项

无。

语法格式

```
DEALLOCATE [ PREPARE ] { name | ALL };
```


版权所有 © 华为云计算技术有限公司

参数说明

表 15-3 DEALLOCATE 参数说明

参数	描述	取值范围
PREPARE	创建一个预备语句，可选关键字，对操作没有影响。	-
name	将要删除的预备语句的名称。	字符串，需符合预备语句的命名规范。
ALL	删除所有预备语句。	字符串，需符合预备语句的格式。

示例

无。

15.5 DO

功能描述

执行匿名代码块。

代码块被视为没有参数的函数主体，返回值类型是void。它的解析和执行是同一时刻发生的。

注意事项

- 程序语言在使用之前，必须通过命令CREATE LANGUAGE安装到当前的数据库中。 plpgsql是默认的安装语言，其他语言安装时必须指定。
- 如果语言是不受信任的，用户必须有使用程序语言的USAGE权限，或者是系统管理员。

语法格式

```
DO [ LANGUAGE lang_name ] code;
```

参数说明

表 15-4 DO 参数说明

参数	描述	取值范围
lang_name	用来解析代码的程序语言的名字。	符合规范的程序语言，默认的语言是plpgsql。
code	可以被执行的程序语言代码。	字符串格式。

版权所有 © 华为云计算技术有限公司

示例

授予用户webuser对模式tpcds下视图的所有操作权限：

```
DO $$DECLARE r record;
BEGIN
   FOR r IN SELECT c.relname AS table_name ,n.nspname table_schema FROM pg_class c,pg_namespace n
         WHERE c.relnamespace = n.oid AND n.nspname = 'tpcds' AND relkind IN ('r','v')
   LOOP
      EXECUTE 'GRANT ALL ON ' || quote_ident(r.table_schema) || '.' || quote_ident(r.table_name) || ' TO
webuser';
   END LOOP;
END$$;
```

15.6 EXECUTE

功能描述

执行一个前面准备好的预备语句。由于预备语句只在会话的生命期里存在，那么该预备语句必须在当前会话中由一个更早执行的PREPARE语句创建。

注意事项

如果创建预备语句的PREPARE语句声明了一些参数，那么传递给EXECUTE语句的必须是一个兼容的参数集，否则就会生成一个错误。

语法格式

```
EXECUTE name [ ( parameter [, ...] ) ];
```

参数说明

表 15-5 EXECUTE 参数说明

参数	描述	取值范围
name	要执行的预备语句的名字。	字符串，需符合预备语句的命名规范。
parameter	给预备语句的一个参数的具体数值。必须是一个生成与创建预备语句时指定参数的数据类型相兼容的值的表达式。	字符串，与指定参数的数据类型相兼容的值的表达式。

示例

准备数据：

```
CREATE SCHEMA IF NOT EXISTS tpcds;
DROP TABLE IF EXISTS  tpcds.reason_t1;
CREATE TABLE tpcds.reason_t1 (
   r_reason_sk     integer,
   r_reason_id     character(16),
```

 版权所有 © 华为云计算技术有限公司

```
    r_reason_desc  character(100)
);
```

为一个INSERT语句创建一个预备语句，然后执行它：

```
PREPARE insert_reason(integer,character(16),character(100)) AS INSERT INTO tpcds.reason_t1
VALUES($1,$2,$3);
EXECUTE insert_reason(52, 'AAAAAAAADDAAAAAA', 'reason 52');
```

15.7 EXECUTE DIRECT

功能描述

在指定的节点上执行SQL语句。一般情况下，SQL语句的执行是由集群负载自动分配到合适的节点上，execute direct主要用于数据库维护和测试。

注意事项

- 只有系统管理员才能执行EXECUTE DIRECT。
- 为了各个节点上数据的一致性，SQL语句仅支持SELECT，不允许执行事务语句、DDL、DML。
- 使用此类型语句在指定的DN执行AVG聚集计算时，返回结果集是以数组形式返回，如{4,2}，表示sum结果为4，count结果为2。
- 由于CN节点不存储用户表数据，不允许指定CN节点执行用户表上的SELECT查询。
- 不允许执行嵌套的EXECUTE DIRECT语句，即执行的SQL语句不能同样是EXECUTE DIRECT语句，此时可直接执行最内层EXECUTE DIRECT语句代替。

语法格式

```
EXECUTE DIRECT ON ( nodename [, ... ] ) query ;
```

参数说明

表 15-6 EXECUTE DIRECT 参数说明

参数	描述	取值范围
nodename	节点名称。	字符串，已存在的节点。
query	要执行查询语句。	字符串，可以执行的SQL语句。

- **nodename**
  节点名称。
  取值范围：已存在的节点。
- **query**
  要执行查询的语句。

示例

查询节点dn_6001_6002上tpcds.customer_address记录：

 版权所有 © 华为云计算技术有限公司

```
EXECUTE DIRECT ON(dn_6001_6002) 'select count(*) from tpcds.customer_address';
```

```
gaussdb=> EXECUTE DIRECT ON(dn_6001_6002) 'select count(*) from tpcds.customer_address';
 count
-------
     0
(1 row)
```

15.8 GRANT

功能描述及注意事项

对角色和用户进行授权操作。

使用GRANT命令进行用户授权包括以下三种场景：

将系统权限(SYSADMIN、CREATEDB、CREATEROLE、AUDITADMIN 和 LOGIN)授权给角色或用户

系统权限又称为用户属性，包括SYSADMIN、CREATEDB、CREATEROLE、AUDITADMIN和LOGIN。

表 15-7 系统权限

系统权限	权限描述	是否支持通过GRANT/REVOKE授权或撤销
SYSADMIN	支持查看特殊系统表。	是
CREATEDB	支持创建数据库。	否
CREATEROLE	支持创建用户与角色。	
AUDITADMIN	支持查看审计日志。	
LOGIN	支持连接数据库。	

系统权限一般在创建或修改角色阶段，通过CREATE/ALTER ROLE语法来指定。

其中，SYSADMIN权限可以通过GRANT/REVOKE ALL PRIVILEGE授予或撤销。

但系统权限无法通过ROLE和USER的权限被继承，也无法授予PUBLIC。

将数据库对象(表和视图、指定字段、数据库、函数、模式、表空间等)授权给角色或用户

将数据库对象（表和视图、指定字段、数据库、函数、模式等）的相关权限授予特定角色或用户。

- **不允许对表分区进行GRANT操作**，否则会引起告警。
- 如果要重命名一个对象，用户除了必须是该对象的所有者外，还必须拥有该对象所在模式的CREATE权限。

 版权所有 © 华为云计算技术有限公司

- **将模式中的表或者视图对象授权给其他用户时，需要将表或视图所属的模式的USAGE权限同时授予该用户，**若没有该权限，则只能看到这些对象的名字，并不能实际进行对象访问。
- GRANT命令将数据库对象的特定权限授予一个或多个角色。这些权限会追加到已有的权限上。
- **关键字PUBLIC**表示该权限要赋予所有角色，包括以后创建的用户。PUBLIC可以看做是一个隐含定义好的组，它总是包括所有角色。任何角色或用户都将拥有通过GRANT直接赋予的权限和所属的权限，再加上PUBLIC的权限。
- 如果声明了WITH GRANT OPTION，则被授权的用户也可以将此权限赋予他人，否则就不能授权给他人。这个选项不能赋予PUBLIC，这是DWS特有的属性。
- DWS会将某些类型的对象上的权限授予PUBLIC。默认情况下，对表、表字段、序列、外部数据源、外部服务器、模式或表空间对象的权限不会授予PUBLIC，而以下这些对象的权限会授予PUBLIC：数据库的CONNECT权限和CREATE TEMP TABLE权限、函数的EXECUTE特权、语言和数据类型（包括域）的USAGE特权。对象拥有者可以撤销默认授予PUBLIC的权限并专门授予权限给其他用户。为了更安全，建议在同一个事务中创建对象并设置权限，这样其他用户就没有时间窗口使用该对象。另外，这些初始的默认权限可以使用ALTER DEFAULT PRIVILEGES命令修改。

将角色或用户的权限授权给其他角色或用户

将一个角色或用户的权限授予一个或多个其他角色或用户。在这种情况下，每个角色或用户都可视为拥有一个或多个数据库权限的集合。

- 当声明了WITH ADMIN OPTION，被授权的用户可以将该权限再次授予其他角色或用户，以及撤销所有由该角色或用户继承到的权限。
- 当授权的角色或用户发生变更或被撤销时，所有继承该角色或用户权限的用户拥有的权限都会随之发生变更。
- 数据库系统管理员可以给任何角色或用户授予/撤销任何权限。
- 拥有CREATEROLE权限的角色可以赋予或者撤销任何非系统管理员角色的权限。
- **将系统权限授权给角色或用户**

  系统权限又称为用户属性，包括SYSADMIN、CREATEDB、CREATEROLE、AUDITADMIN和LOGIN。

  系统权限一般通过CREATE/ALTER ROLE语法来指定。其中，SYSADMIN权限可以通过GRANT/REVOKE ALL PRIVILEGE授予或撤销。但系统权限无法通过ROLE和USER的权限被继承，也无法授予PUBLIC。

- **将数据库对象授权给角色或用户**

  将数据库对象（表和视图、指定字段、数据库、函数、模式等）的相关权限授予特定角色或用户。

  GRANT命令将数据库对象的特定权限授予一个或多个角色。这些权限会追加到已有的权限上。

  关键字PUBLIC表示该权限要赋予所有角色，包括以后创建的用户。PUBLIC可以看做是一个隐含定义好的组，它总是包括所有角色。任何角色或用户都将拥有通过GRANT直接赋予的权限和所属的权限，再加上PUBLIC的权限。

  如果声明了WITH GRANT OPTION，则被授权的用户也可以将此权限赋予他人，否则就不能授权给他人。这个选项不能赋予PUBLIC，这是DWS特有的属性。

  DWS会将某些类型的对象上的权限授予PUBLIC。默认情况下，对表、表字段、序列、外部数据源、外部服务器、模式或表空间对象的权限不会授予PUBLIC，而以

 版权所有 © 华为云计算技术有限公司

下这些对象的权限会授予PUBLIC：数据库的CONNECT权限和CREATE TEMP TABLE权限、函数的EXECUTE特权、语言和数据类型（包括域）的USAGE特权。对象拥有者可以撤销默认授予PUBLIC的权限并专门授予权限给其他用户。为了更安全，建议在同一个事务中创建对象并设置权限，这样其他用户就没有时间窗口使用该对象。另外，这些初始的默认权限可以使用ALTER DEFAULT PRIVILEGES命令修改。

- **将角色或用户的权限授权给其他角色或用户**

  将一个角色或用户的权限授予一个或多个其他角色或用户。在这种情况下，每个角色或用户都可视为拥有一个或多个数据库权限的集合。

  当声明了WITH ADMIN OPTION，被授权的用户可以将该权限再次授予其他角色或用户，以及撤销所有由该角色或用户继承到的权限。当授权的角色或用户发生变更或被撤销时，所有继承该角色或用户权限的用户拥有的权限都会随之发生变更。

  数据库系统管理员可以给任何角色或用户授予/撤销任何权限。拥有CREATEROLE权限的角色可以赋予或者撤销任何非系统管理员角色的权限。

权限分类

表 15-8 GRANT 的权限分类

参数	描述
SELECT	允许对指定的表、视图、序列执行SELECT语句。
INSERT	允许对指定的表执行INSERT语句。
UPDATE	允许对声明的表中任意字段执行UPDATE语句。SELECT… FOR UPDATE和SELECT… FOR SHARE除了需要SELECT权限外，还需要UPDATE权限。
DELETE	允许执行DELETE语句删除指定表中的数据。
TRUNCATE	允许执行TRUNCATE语句删除指定表中的所有记录。
REFERENCES	创建一个外键约束，必须拥有参考表和被参考表的REFERENCES权限。
TRIGGER	创建一个触发器，必须拥有表或视图的TRIGGER权限。
ANALYZE | ANALYSE	对表执行ANALYZE | ANALYSE操作来收集表的统计信息，必须拥有表的ANALYZE | ANALYSE权限。

 版权所有 © 华为云计算技术有限公司

参数	描述
CREATE	• 对于数据库，允许在数据库里创建新的模式。 • 对于模式，允许在模式中创建新的对象。如果要重命名一个对象，用户除了必须是该对象的所有者外，还必须拥有该对象所在模式的CREATE权限。 • 对于子集群，允许在子集群中创建表对象。
CONNECT	允许用户连接到指定的数据库。
TEMPORARY | TEMP	允许在使用指定数据库时创建临时表。
EXECUTE	允许使用指定的函数，以及利用这些函数实现的操作符。
USAGE	• 对于过程语言，允许用户在创建函数的时候指定过程语言。 • 对于模式，USAGE允许访问包含在指定模式中的对象，若没有该权限，则只能看到这些对象的名字。 • 对于序列，USAGE允许使用nextval函数。 • 对于子集群， 对包含在指定模式中的对象有访问权限时，USAGE允许访问指定子集群下的表对象。
COMPUTE	针对计算子集群，允许用户在具有compute权限的计算子集群上进行弹性计算。
ALTER	允许修改表、模式或函数。
DROP	允许删除表、模式或函数。
VACUUM	允许对表执行VACUUM。
REFRESH	允许对物化视图执行刷新。
ALL PRIVILEGES	一次赋予指定用户/角色所有可赋予的权限。只有系统管理员有权执行GRANT ALL PRIVILEGES。
WITH GRANT OPTION	如果声明了WITH GRANT OPTION，则被授权的用户也可以将此权限赋予他人，否则就不能授权给他人。这个选项不能赋予PUBLIC。 • NODE GROUP不支持WITH GRANT OPTION功能。 • 使用WITH GRANT OPTION时需确保**security_enable_options**参数值中设置了grant_with_grant_option。

版权所有 © 华为云计算技术有限公司

参数	描述
WITH ADMIN OPTION	指定权限是否允许转授。如果声明了WITH ADMIN OPTION，角色的成员又可以将角色的成员身份授予其他人。

语法格式

- 将**表或视图的访问权限**赋予指定的用户或角色。不允许对表分区进行GRANT操作，对表分区进行GRANT操作会引起告警。

```
GRANT { { SELECT | INSERT | UPDATE | DELETE | TRUNCATE | REFERENCES | TRIGGER | ANALYZE |
ANALYSE | VACUUM | ALTER | DROP } [, ...]
    | ALL [ PRIVILEGES ] }
  ON { [ TABLE ] table_name [, ...]
    | ALL TABLES IN SCHEMA schema_name [, ...] }
  TO { [ GROUP ] role_name | PUBLIC } [, ...]
  [ WITH GRANT OPTION ];
```

⚠ 注意

避免对同一张表或视图（包括ON TABLE和ON ALL TABLES IN SCHEMA）执行并发GRANT/REVOKE操作，或者对同一张表或视图同时执行GRANT/REVOKE操作和DDL操作，否则会出现类似报错“tuple concurrently updated.”。

- 将**表中字段的访问权限**赋予指定的用户或角色。

```
GRANT { {{ SELECT | INSERT | UPDATE | REFERENCES } ( column_name [, ...] )} [, ...]
    | ALL [ PRIVILEGES ] ( column_name [, ...] ) }
  ON [ TABLE ] table_name [, ...]
  TO { [ GROUP ] role_name | PUBLIC } [, ...]
  [ WITH GRANT OPTION ];
```

- 将**模式的访问权限**赋予指定的用户或角色。

```
GRANT { { CREATE | USAGE | ALTER | DROP } [, ...] | ALL [ PRIVILEGES ] }
  ON SCHEMA schema_name [, ...]
  TO { [ GROUP ] role_name | PUBLIC } [, ...]
  [ WITH GRANT OPTION ];
```

⚠ 注意

将模式中的表或者视图对象授权给其他用户时，需要将表或视图所属的模式的USAGE权限同时授予该用户，若没有该权限，则只能看到这些对象的名字，并不能实际进行对象访问。

- 将**数据库的访问权限**赋予指定的用户或角色。

```
GRANT { { CREATE | CONNECT | TEMPORARY | TEMP } [, ...]
    | ALL [ PRIVILEGES ] }
  ON DATABASE database_name [, ...]
  TO { [ GROUP ] role_name | PUBLIC } [, ...]
  [ WITH GRANT OPTION ];
```

- 将**角色的权限**赋予其他用户或角色的语法。

```
GRANT role_name [, ...]
  TO role_name [, ...]
  [ WITH ADMIN OPTION ];
```

- 将**所有权限**赋予指定的角色。

版权所有 © 华为云计算技术有限公司

```
GRANT ALL { PRIVILEGES | PRIVILEGE }
  TO role_name;
```

- 将域的访问权限赋予指定的用户或角色。

```
GRANT { USAGE | ALL [ PRIVILEGES ] }
   ON DOMAIN domain_name [, ...]
   TO { [ GROUP ] role_name | PUBLIC } [, ...]
   [ WITH GRANT OPTION ];
```

说明

当前版本暂时不支持赋予域的访问权限。

- 将外部数据源的访问权限赋予给指定的用户或角色。

```
GRANT { USAGE | ALL [ PRIVILEGES ] }
   ON FOREIGN DATA WRAPPER fdw_name [, ...]
   TO { [ GROUP ] role_name | PUBLIC } [, ...]
   [ WITH GRANT OPTION ];
```

- 将外部服务器的访问权限赋予给指定的用户或角色。

```
GRANT { USAGE | ALL [ PRIVILEGES ] }
   ON FOREIGN SERVER server_name [, ...]
   TO { [ GROUP ] role_name | PUBLIC } [, ...]
   [ WITH GRANT OPTION ];
```

- 将函数的访问权限赋予给指定的用户或角色。

```
GRANT { { EXECUTE | ALTER | DROP } [, ...] | ALL [ PRIVILEGES ] }
   ON { FUNCTION {function_name ( [ {[ argmode ] [ arg_name ] arg_type} [, ...] ] )} [, ...]
     | ALL FUNCTIONS IN SCHEMA schema_name [, ...] }
   TO { [ GROUP ] role_name | PUBLIC } [, ...]
   [ WITH GRANT OPTION ];
```

- 将过程语言的访问权限赋予给指定的用户或角色。

```
GRANT { USAGE | ALL [ PRIVILEGES ] }
   ON LANGUAGE lang_name [, ...]
   TO { [ GROUP ] role_name | PUBLIC } [, ...]
   [ WITH GRANT OPTION ];
```

说明

当前版本暂时不支持过程语言。

- 将大对象的访问权限赋予指定的用户或角色。

```
GRANT { { SELECT | UPDATE } [, ...] | ALL [ PRIVILEGES ] }
   ON LARGE OBJECT loid [, ...]
   TO { [ GROUP ] role_name | PUBLIC } [, ...]
   [ WITH GRANT OPTION ];
```

说明

当前版本暂时不支持大对象。

- 将序列的访问权限赋予指定的用户或角色。

```
GRANT { { SELECT | UPDATE | USAGE } [, ...] | ALL [ PRIVILEGES ] }
   ON { SEQUENCE sequence_name [, ...]
      | ALL SEQUENCES IN SCHEMA schema_name [, ...] }
   TO { [ GROUP ] role_name | PUBLIC } [, ...]
   [ WITH GRANT OPTION ];
```

- 将子集群的访问权限赋予指定的用户或角色。普通用户不能执行针对Node Group的GRANT/REVOKE操作。

```
GRANT { CREATE | USAGE | COMPUTE | ALL [ PRIVILEGES ] }
   ON NODE GROUP group_name [, ...]
   TO { [ GROUP ] role_name | PUBLIC } [, ...]
   [ WITH GRANT OPTION ];
```

- 将类型的访问权限赋予指定的用户或角色。

```
GRANT { USAGE | ALL [ PRIVILEGES ] }
   ON TYPE type_name [, ...]
   TO { [ GROUP ] role_name | PUBLIC } [, ...]
   [ WITH GRANT OPTION ];
```

 版权所有 © 华为云计算技术有限公司

> **说明**
>
> 当前版本暂时不支持赋予类型的访问权限。

- 将物化视图的权限赋予指定的角色。该语法仅9.1.2及以上集群版本支持。

```
GRANT { { SELECT | ALTER | DROP | REFRESH } [, ...]
    | ALL [ PRIVILEGES ] }
  ON MATERIALIZED VIEW materialized_view_name [, ...]
  TO {[GROUP] role_name | PUBLIC} [, ...] [WITH GRANT OPTION];
```

参数说明

表 15-9 GRANT 的参数说明

参数	描述	取值范围
role_name	已存在用户名称，可通过PG_USER视图查询用户列表。	字符串，要符合标识符命名规范。最多为63个字节。
table_name	已存在表名称，可通过PG_TABLES查询表名称。	字符串，长度不超过63个字节，以字母或下划线开头，可包含字母、数字、下划线、$、#。
column_name	已存在字段名称，可通过以下语句查询表字段名称。 `SELECT * FROM PG_GET_TABLEDEF('table name');`	字符串，要符合标识符命名规范。详情请参见标识符命名规范。
schema_name	已存在模式名称，可通过PG_NAMESPACE系统表查询模式名称。	字符串，要符合标识符命名规范。详情请参见标识符命名规范。
database_name	已存在数据库名称，可通过PG_DATABASE系统表查询数据库名称。	字符串，要符合标识符的命名规范：数据库名称是由大小写英文字母、数字、_和$组合而成，并且不能以数字和$开头。
function_name	已存在函数名称，可通过PG_PROC系统表查询函数名称。	字符串，要符合标识符命名规范。详情请参见标识符命名规范。
sequence_name	已存在序列名称，可通过PG_SEQUENCES系统视图（仅9.1.0及以上集群支持）查询序列名称。	字符串，仅可以使用小写字母（a~z）、大写字母（A~Z），数字和特殊字符"#"，"_"，"$"的组合。
domain_name	已存在域类型名称。	字符串，要符合标识符命名规范。详情请参见标识符命名规范。
fdw_name	已存在外部数据包名称。	字符串，要符合标识符命名规范。详情请参见标识符命名规范。

版权所有 © 华为云计算技术有限公司

参数	描述	取值范围
lang_name	已存在语言名称。	字符串，要符合标识符命名规范。详情请参见**标识符命名规范**。
type_name	已存在类型名称。	字符串，要符合标识符命名规范。详情请参见**标识符命名规范**。
group_name	已存在的子集群名称。	字符串，要符合标识符命名规范。详情请参见**标识符命名规范**。
argmode	函数的参数模式。	字符串，要符合标识符命名规范。详情请参见**标识符命名规范**。
arg_name	函数的参数名称。	字符串，要符合标识符命名规范。详情请参见**标识符命名规范**。
arg_type	函数的参数类型。	字符串，要符合标识符命名规范。详情请参见**标识符命名规范**。
loid	包含本页的大对象的标识符。	字符串，要符合标识符命名规范。详情请参见**标识符命名规范**。

示例

- **将系统权限授权给用户或者角色。**
  - 将sysadmin所有可用权限授权给joe用户：
    ```
    GRANT ALL PRIVILEGES TO joe;
    ```
    授权成功后，用户joe会拥有sysadmin的所有权限。
- **将对象权限授权给用户或者角色。**
  - 将表tpcds.reason的SELECT权限授权给用户joe：
    ```
    GRANT SELECT ON TABLE tpcds.reason TO joe;
    ```
  - 将表tpcds.reason的所有权限授权给用户kim：
    ```
    GRANT ALL PRIVILEGES ON tpcds.reason TO kim;
    ```
    授权成功后，kim用户就拥有了tpcds.reason表的所有权限，包括增删改查等权限。
  - 将模式tpcds的使用权限授权给用户joe：
    ```
    GRANT USAGE ON SCHEMA tpcds TO joe;
    ```
    授权成功后，joe用户就拥有了模式schema的USAGE权限，允许访问包含在指定模式schema中的对象。
  - 将tpcds.reason表中r_reason_sk、r_reason_id、r_reason_desc列的查询权限，r_reason_desc的更新权限授权给joe：
    ```
    GRANT select (r_reason_sk,r_reason_id,r_reason_desc),update (r_reason_desc) ON tpcds.reason TO joe;
    ```


版权所有 © 华为云计算技术有限公司

授权成功后，用户joe对tpcds.reason表中r_reason_sk，r_reason_id的查询权限会立即生效。

```
GRANT select (r_reason_sk, r_reason_id) ON tpcds.reason TO joe ;
```

  - 将函数func_add_sql的EXECUTE权限授权给用户joe。

```
CREATE FUNCTION func_add_sql(f1 integer,f2 integer) RETURNS integer
    AS 'select $1 + $2;'
    LANGUAGE SQL
    IMMUTABLE
    RETURNS NULL ON NULL INPUT;
GRANT EXECUTE ON FUNCTION func_add_sql(integer, integer) TO joe;
```

  - 将序列serial的UPDATE权限授权给joe用户。

```
GRANT UPDATE ON SEQUENCE serial TO joe;
```

  - 将数据库gaussdb的连接权限授权给用户joe，并给予其在gaussdb中创建schema的权限：

```
GRANT create,connect on database gaussdb TO joe ;
```

  - 将模式tpcds的访问权限授权给角色tpcds_manager，并授予该角色在tpcds下创建对象的权限，不允许该角色中的用户将权限授权给其他人：

```
GRANT USAGE,CREATE ON SCHEMA tpcds TO tpcds_manager;
```

- **将用户或者角色的权限授权给其他用户或角色。**
  - 将用户joe的权限授权给用户manager，并允许该角色将权限授权给其他人：

```
GRANT joe TO manager WITH ADMIN OPTION;
```

  - 将用户manager的权限授权给senior_manager用户：

```
GRANT manager TO senior_manager;
```

- **将物化视图的权限授权给其他用户或角色。**该语法仅9.1.2及以上集群版本支持。

  将物化视图mv1的刷新权限授权给用户manager，并允许该角色将权限授权给其他人：

```
GRANT REFRESH ON mv1 TO manager WITH ADMIN OPTION;
```

相关链接

REVOKE，ALTER DEFAULT PRIVILEGES

常见问题

- **如何给DWS指定用户赋予某张表的权限**
- **如何给DWS指定用户赋予某个SCHEMA的权限**
- **如何创建DWS数据库只读用户**
- **如何查看DWS某个用户有哪些表的权限**
- **将Schema中的表的查询权限赋给其他用户，赋权后仍无法查询Schema中的表**
- **某张表执行过grant select on table t1 to public，如何针对某用户回收权限**
- **普通用户创建或删除GDS/OBS外表语句时报错，提示没有权限或权限不足**
- **赋予用户schema的all权限后建表仍然报错ERROR: current user does not have privilege to role tom**
- **执行语句过程中报错：无权限操作**

版权所有 © 华为云计算技术有限公司

15.9 PREPARE

功能描述

创建一个预备语句。

预备语句是服务端的对象，可以用于优化性能。在执行PREPARE语句的时候，指定的查询被解析、分析、重写。当随后发出EXECUTE语句的时候，预备语句被规划和执行。这种设计避免了重复解析、分析工作。PREPARE语句创建后在整个数据库会话期间一直存在，一旦创建成功，即便是在事务块中创建，事务回滚，PREPARE也不会删除。只能通过显式调用**DEALLOCATE**进行删除，会话结束时，PREPARE也会自动删除。

注意事项

无。

语法格式

```
PREPARE name [ ( data_type [, ...] ) ] AS statement;
```

参数说明

表 15-10 PREPARE 参数说明

参数	描述	取值范围
name	指定预备语句的名称。该名称必须在该会话中是唯一的。	字符串。
data_type	参数的数据类型。	-
statement	可以执行的SQL语句。	任何SELECT INSERT、UPDATE、DELETE或VALUES语句。

- **name**
  指定预备语句的名称。该名称必须在该会话中是唯一的。
- **data_type**
  参数的数据类型。
- **statement**
  任何SELECT INSERT、UPDATE、DELETE或VALUES语句。

示例

准备数据。

```
DROP TABLE IF EXISTS reason_t1;
CREATE TABLE reason_t1 (
```

 版权所有 © 华为云计算技术有限公司

```
    r_reason_sk     integer,
    r_reason_id     character(16),
    r_reason_desc   character(100)
);
```

为一个INSERT语句创建一个预备语句，然后执行它：

```
PREPARE insert_reason(integer,character(16),character(100)) AS INSERT INTO reason_t1 VALUES($1,$2,$3);
EXECUTE insert_reason(52, 'AAAAAAAADDAAAAAA', 'reason 52');
```

相关链接

DEALLOCATE

15.10 REASSIGN OWNED

功能描述

修改数据库对象的所有者。

REASSIGN OWNED要求系统将所有old_role拥有的数据库对象的所有者更改为new_role。

注意事项

- REASSIGN OWNED常用于在删除角色之前的准备工作。由于不会影响其他数据库中的对象， 因此通常需要在每个数据库中执行此命令，该数据库包含要删除的角色所拥有的对象。
- 执行REASSIGN OWNED需要有原角色和目标角色上的权限。
- 资源管理不对该语法的数据切换进行监控，需调用SELECT GS_WLM_READJUST_USER_SPACE(0);手动校准监控数据。

语法格式

```
REASSIGN OWNED BY old_role [, ...] TO new_role;
```

参数说明

表 15-11 REASSIGN OWNED 参数说明

参数	描述	取值范围
old_role	原所有者角色名。	有效的角色名。
new_role	新所有者角色名。	有效的角色名。

示例

将由joe和jack角色拥有的所有数据库对象重新分配给admin。

```
REASSIGN OWNED BY joe, jack TO admin;
```

 版权所有 © 华为云计算技术有限公司

15.11 REVOKE

功能描述

REVOKE用于撤销一个或多个角色的权限。

注意事项

非对象所有者试图在对象上REVOKE权限，命令按照以下规则执行：

- 如果授权用户没有该对象上的权限，则命令立即失败。
- 如果授权用户有部分权限，则只撤销那些有授权选项的权限。
- 如果授权用户没有授权选项，REVOKE ALL PRIVILEGES形式将发出一个错误信息，而对于其他形式的命令而言，如果是命令中指定名字的权限没有相应的授权选项，该命令将发出一个警告。
- 不允许对表分区进行REVOKE操作，对分区表进行REVOKE操作会引起告警。

语法格式

- 撤销指定表和视图上权限。

```
REVOKE [ GRANT OPTION FOR ]
    { { SELECT | INSERT | UPDATE | DELETE | TRUNCATE | REFERENCES | TRIGGER | ANALYZE |
ANALYSE | VACUUM | ALTER | DROP }[, ...]
    | ALL [ PRIVILEGES ] }
    ON { [ TABLE ] table_name [, ...]
        | ALL TABLES IN SCHEMA schema_name [, ...] }
    FROM { [ GROUP ] role_name | PUBLIC } [, ...]
    [ CASCADE | RESTRICT ];
```

> **⚠ 注意**
>
> 避免对同一张表或视图（包括ON TABLE和ON ALL TABLES IN SCHEMA）执行并发GRANT/REVOKE操作，或者对同一张表或视图同时执行GRANT/REVOKE操作和DDL操作，否则会出现类似报错“tuple concurrently updated.”。

- 撤销表上指定字段权限。

```
REVOKE [ GRANT OPTION FOR ]
    { {{ SELECT | INSERT | UPDATE | REFERENCES } ( column_name [, ...] )}[, ...]
    | ALL [ PRIVILEGES ] ( column_name [, ...] ) }
    ON [ TABLE ] table_name [, ...]
    FROM { [ GROUP ] role_name | PUBLIC } [, ...]
    [ CASCADE | RESTRICT ];
```

- 撤销指定数据库上权限。

```
REVOKE [ GRANT OPTION FOR ]
    { { CREATE | CONNECT | TEMPORARY | TEMP } [, ...]
    | ALL [ PRIVILEGES ] }
    ON DATABASE database_name [, ...]
    FROM { [ GROUP ] role_name | PUBLIC } [, ...]
    [ CASCADE | RESTRICT ];
```

- 撤销指定函数上权限。

```
REVOKE [ GRANT OPTION FOR ]
    { { EXECUTE | ALTER | DROP } [, ...] | ALL [ PRIVILEGES ] }
    ON { FUNCTION {function_name ( [ {[ argmode ] [ arg_name ] arg_type} [, ...] ] )} [, ...]
        | ALL FUNCTIONS IN SCHEMA schema_name [, ...] }
```


版权所有 © 华为云计算技术有限公司

```
  FROM { [ GROUP ] role_name | PUBLIC } [, ...]
  [ CASCADE | RESTRICT ];
```

- 撤销指定大对象上权限。

```
REVOKE [ GRANT OPTION FOR ]
  { { SELECT | UPDATE } [, ...] | ALL [ PRIVILEGES ] }
  ON LARGE OBJECT loid [, ...]
  FROM { [ GROUP ] role_name | PUBLIC } [, ...]
  [ CASCADE | RESTRICT ];
```

- 撤销指定序列上权限。

```
REVOKE [ GRANT OPTION FOR ]
  { { SELECT | UPDATE | USAGE } [, ...] | ALL [ PRIVILEGES ] }
  ON SEQUENCE sequence_name [, ...]
  FROM { [ GROUP ] role_name | PUBLIC } [, ...]
  [ CASCADE | RESTRICT ];
```

- 撤销指定模式上权限。

```
REVOKE [ GRANT OPTION FOR ]
  { { CREATE | USAGE | ALTER | DROP } [, ...] | ALL [ PRIVILEGES ] }
  ON SCHEMA schema_name [, ...]
  FROM { [ GROUP ] role_name | PUBLIC } [, ...]
  [ CASCADE | RESTRICT ];
```

- 撤销指定子集群上权限

```
REVOKE [ GRANT OPTION FOR ]
  { CREATE | USAGE | COMPUTE | ALL [ PRIVILEGES ] }
  ON NODE GROUP group_name [, ...]
  FROM { [ GROUP ] role_name | PUBLIC } [, ...]
  [ CASCADE | RESTRICT ];
```

- 按角色撤销角色上的权限。

```
REVOKE [ ADMIN OPTION FOR ]
  role_name [, ...] FROM role_name [, ...]
  [ CASCADE | RESTRICT ];
```

- 撤销指定角色的sysadmin权限。

```
REVOKE ALL { PRIVILEGES | PRIVILEGE } FROM role_name;
```

- 撤销物化视图对象的权限。该语法仅9.1.2及以上集群版本支持。

```
REVOKE { { SELECT | ALTER | DROP | REFRESH } [, ...]
    | ALL [ PRIVILEGES ] }
  ON MATERIALIZED VIEW materialized_view_name [, ...]
  FROM {[GROUP] role_name | PUBLIC} [, ...] [WITH GRANT OPTION];
```

参数说明

关键字PUBLIC表示一个隐式定义的拥有所有角色的组。REVOKE权限类别和参数说明，请参见GRANT的**参数说明**。

 版权所有 © 华为云计算技术有限公司

说明

- 任何特定角色拥有的特权包括直接授予该角色的特权、从该角色作为其成员的角色中得到的权限以及授予给PUBLIC的权限。因此，从PUBLIC收回SELECT特权并不一定会意味着所有角色都会失去在该对象上的SELECT特权，那些直接被授予的或者通过另一个角色被授予的角色仍然会拥有它。类似地，从一个用户收回SELECT后，如果PUBLIC仍有SELECT权限，该用户还是可以使用SELECT。
- 依赖性权限：如用户A拥有某个表的UPDATE权限，及WITH GRANT OPTION选项，同时A把这个权限赋予了用户B，则用户B持有的权限称为依赖性权限。当用户A持有的权限或者授权选项被撤销时，依赖性权限仍然存在，但如果声明了CASCADE，则所有依赖性权限都被撤销。
- 一个用户只能撤销由它自己直接赋予的权限。例如，如果用户A被指定授权（WITH ADMIN OPTION）选项，且把一个权限赋予了用户B，然后用户B又赋予了用户C，则用户A不能直接将C的权限撤销。但是，用户A可以撤销用户B的授权选项，并且使用CASCADE。这样，用户C的权限就会自动被撤销。同样，如果A和B都赋予了C同样的权限，则A可以撤销他自己的授权选项，但是不能撤销B的，因此C仍然拥有该权限。
- 如果执行REVOKE的角色持有的权限是通过多层成员关系获得的，则具体是哪个包含的角色执行的该命令是不确定的。在这种场合下，最好的方法是使用SET ROLE成为特定角色，然后执行REVOKE，否则可能导致删除了不想删除的权限，或者是任何权限都没有删除。

示例

撤销joe用户的sysadmin权限：

```
REVOKE ALL PRIVILEGES FROM joe;
```

撤销指定模式上授予的权限：

```
REVOKE USAGE,CREATE ON SCHEMA tpcds FROM tpcds_manager;
```

撤销joe用户的CONNECT特权：

```
REVOKE CONNECT ON DATABASE mydatabase FROM joe;
```

从用户joe撤销角色admins中的成员资格：

```
REVOKE admins FROM joe;
```

撤销用户joe对视图myView具有的所有特权：

```
REVOKE ALL PRIVILEGES ON myView FROM joe;
```

撤销针对表customer_t1的public插入特权：

```
REVOKE INSERT ON customer_t1 FROM PUBLIC;
```

撤销用户joe对模式tpcds的使用权限：

```
REVOKE USAGE ON SCHEMA tpcds FROM joe;
```

撤销用户joe对tpcds.reason表中r_reason_sk，r_reason_id的查询权限：

```
REVOKE select (r_reason_sk, r_reason_id) ON tpcds.reason FROM joe;
```

撤销用户joe对函数func_add_sql的执行权限：

```
REVOKE execute ON FUNCTION func_add_sql(integer, integer) FROM joe CASCADE;
```

撤销用户joe对物化视图mv1的刷新权限：

```
REVOKE REFRESH ON mv1 FROM joe;
```

 版权所有 © 华为云计算技术有限公司

相关链接

GRANT

 版权所有 © 华为云计算技术有限公司

16 DQL 语法

16.1 DQL 语法一览表

DQL（Data Query Language数据查询语言），用于从表或视图中获取数据。

查询

DWS提供了用于从表或视图中获取数据的语句，请参考**SELECT**。

根据查询结果定义一个新表。

DWS提供了根据查询结果创建一个新表，并且将查询到的数据插入到新表中的语句，请参考**SELECT INTO**。

16.2 SELECT

功能描述

SELECT用于从表或视图中读取数据。

SELECT语句就像叠加在数据库表上的过滤器，利用SQL关键字从数据表中过滤出用户需要的数据。

注意事项

- 使用SELECT请遵循**SELECT操作规范**。
- SELECT支持普通表与HDFS之间的JOIN，不支持普通表与GDS外表之间的JOIN。即SELECT语句中不能同时出现普通表和GDS外表。
- 必须对每个在SELECT命令中使用的字段有SELECT权限，使用FOR UPDATE或FOR SHARE还要求UPDATE权限。
- 列存表与HStore表都暂不支持SELECT FOR UPDATE语法。
- 对HStore表执行SELECT查询时，会扫描列存主表CU上的数据、delta表上的I记录中的数据、内存中每行数据更新信息，并将三种信息合并后返回。

版权所有 © 华为云计算技术有限公司

- 在通过主键索引或唯一索引查询数据的场景中：

  对于传统列存表，唯一索引会同时存储行存Delta表上的数据位置信息（blocknum，offset）与列存主表的数据位置信息（cuid，offset），数据MERGE到主表后又会插入新的索引元组，索引会持续膨胀。

  对于HStore表，由于实现了全局CUID的统一分配，索引元组中始终只存储（cuid，offset）， 数据MERGE后不会产生新的索引元组。

语法格式

主句语法格式如下，其中包含的FROM子句、GROUP BY子句中的语法格式单独定义，如需了解各子句的含义，可以单击以下代码中的链接进入。

说明

另外，SELECT支持简化版查询语法，功能相当于select * from table_name。

```
TABLE { ONLY {(table_name)| table_name} | table_name [ * ]};
```

```
[ WITH [ RECURSIVE ] with_query [, ...] ]
SELECT [/*+ plan_hint */] [ ALL | DISTINCT [ ON ( expression [, ...] ) ] ]
{ * | {expression [ [ AS ] output_name ]} [, ...] }
[ FROM from_item [, ...] ]          --跳转到FROM子句
[ WHERE condition ]
[ GROUP BY grouping_element [, ...] ]       --跳转到GROUP BY子句
[ HAVING condition [, ...] ]
[ WINDOW {window_name AS ( window_definition )} [, ...] ]
[ { UNION | INTERSECT | EXCEPT | MINUS } [ ALL | DISTINCT ] select ]
[ ORDER BY {expression [ [ ASC | DESC | USING operator ] | nlssort_expression_clause ] [ NULLS { FIRST |
LAST } ]} [, ...] ]
[ { [ LIMIT { count | ALL } ] [ OFFSET start [ ROW | ROWS ] ] } | { LIMIT start, { count | ALL } } ]
[ FETCH { FIRST | NEXT } [ count ] { ROW | ROWS } ONLY ]
[ {FOR { UPDATE | SHARE } [ OF table_name [, ...] ] [ NOWAIT ]} [...] ];
```

须知

condition和expression中可以使用输出列中表达式的别名。

- 只能同一层引用。
- 只能引用输出列中的别名。
- 只能是后面的表达式引用前面的表达式。
- 不能包含volatile函数。
- 不能包含Window function函数。
- 不支持在join on条件中引用别名。
- 输出列中有多个要应用的别名则报错。

 版权所有 © 华为云计算技术有限公司

SELECT 主句参数说明

表 16-1 SELECT 参数说明

参数	描述	取值范围或示例
WITH [RECURSIVE] with_query [, ...]	用于声明一个或多个在主查询中通过名字引用的子查询（**公共表达式CTE**），相当于临时表，当需要多次引用相同子查询结果时，使用WITH可简化代码逻辑，提高编码效率。 如果声明了RECURSIVE，则允许递归查询，即允许SELECT子查询通过名字引用它自己。 其中with_query的详细格式为： with_query_name [(column_name [, ...])] AS [[NOT] MATERIALIZED] ({select | values | insert | update | delete})	• with_query_name指定子查询生成的结果集名称，可使用该名称访问子查询的结果集。 • [NOT] MATERIALIZED，指定该参数可改变默认行为： – 指定MATERIALIZED时，将子查询执行一次，并将其结果集进行物化，供主查询多次查询其结果集。 – 指定NOT MATERIALIZED时，with_query只被主查询引用一次，则不再单独执行，而是将其子查询替换到主查询中的引用处，随主查询一起执行。 – 如下几种情况会忽略NOT MATERIALIZED： – 子查询中含有volatile函数。 – 子查询为含有FOR UPDATE/FOR SHARE的SELECT/

版权所有 © 华为云计算技术有限公司

参数	描述	取值范围或示例
		VALUES语句。 – 子查询为INSERT/UPDATE/DELETE等语句。 – with_query为RECURSIVE。 – 被引用次数大于1的with_query2引用了外层自引用的with_query1，则with_query2不能被替换到引用处。例如下面示例中，tmp2被引用了两次，tmp2因为引用了外层自引用的tmp1，所以即使tmp2指定了NOT MATERIALIZED也会被物化。 with recursive tmp1(b) as (values(1) union all (with tmp2 as not materialized (select * from tmp1) select tt1.b + tt2.b from tmp2 tt1,

 版权所有 © 华为云计算技术有限公司

参数	描述	取值范围或示例
		tmp2 tt2)) select * from tmp1; • column_name指定子查询结果集中显示的列名。 • 每个子查询可以是SELECT，VALUES，INSERT，UPDATE或DELETE语句。
plan_hint子句	以/*+ */的形式在SELECT关键字后，用于对SELECT对应的语句块生成的计划进行hint调优 ，详细用法请参见**使用Plan Hint进行调优**。	-
ALL	声明返回所有符合条件的行，是默认行为，可以省略该关键字。	-
DISTINCT [ON (expression [, ...])]	从SELECT的结果集中删除所有重复的行，确保结果集中每行都是唯一的。 ON (expression [, ...]) 只保留那些在给出的表达式上运算出相同结果的行当中的第一行。 **须知** DISTINCT ON表达式是使用与ORDER BY相同的规则进行解释的。除非使用了ORDER BY来保证需要的行首先出现，否则，"第一行" 是不可预测的。	-
SELECT列表	指定查询表中列名，可以是部分列或者是全部（使用通配符*表示）。 通过使用子句AS output_name可以为输出字段取个别名，这个别名通常用于输出字段的显示。	列名可以用下面几种形式表达： • 手动输入列名，多个列之间用英文逗号（,）分隔。 • 可以是FROM子句里面计算出来的字段。 • 通常关键字AS可以省略，如select 2a，则会隐式转换为 select 2 as a。
FROM子句	为SELECT声明一个或者多个源表。具体请参见**表16-2**。	-

版权所有 © 华为云计算技术有限公司

参数	描述	取值范围或示例
WHERE子句	WHERE子句构成一个行选择表达式，用来缩小SELECT的查询范围。	• LIKE操作符 • "(+)"操作符，不建议使用。 具体使用约束请参见**WHERE子句使用约束**。
GROUP BY子句	将查询结果按某一列或多列的值分组，值相等的为一组。	详情请参见**表16-4**。
HAVING子句	用于过滤分组（GROUP BY）后的结果，它与WHERE类似，但是通常与GROUP BY子句配合使用来过滤出特殊的组。 HAVING子句将组的一些属性与一个常数值进行比较，只有满足HAVING子句中的逻辑表达式的组才会被提取出来。	-
WINDOW子句	用于定义窗口函数，一般格式为WINDOW window_name AS（window_definition）[， ...]，window_name是可以被随后的窗口定义所引用的窗口名称，window_definition可以是以下格式： [existing_window_name] [PARTITION BY expression [, ...]] [ORDER BY expression [ASC | DESC | USING operator] [NULLS { FIRST | LAST }] [, ...]] [frame_clause] frame_clause为窗口函数定义一个窗口框架window frame，即当前查询的一组相关行。frame_clause可以是以下的格式： [RANGE | ROWS] frame_start [RANGE | ROWS] BETWEEN frame_start AND frame_end frame_start和frame_end可以是： UNBOUNDED PRECEDING value PRECEDING (RANGE不支持) CURRENT ROW value FOLLOWING (RANGE不支持) UNBOUNDED FOLLOWING **须知** 对列存表的查询目前只支持row_number窗口函数，不支持frame_clause。	-

版权所有 © 华为云计算技术有限公司

参数	描述	取值范围或示例
UNION子句	UNION计算多个SELECT语句返回行集合的并集。 UNION子句有如下约束条件： ● 除非声明了ALL子句，否则缺省的UNION结果不包含重复的行。 ● 同一个SELECT语句中的多个UNION操作符是从左向右计算的，除非用圆括号进行了标识。 ● FOR UPDATE不能在UNION的结果或输入中声明。 一般表达式： `select_statement UNION [ALL] select_statement` ● select_statement可以是任何没有ORDER BY、LIMIT、FOR UPDATE子句的SELECT语句。 ● 如果用圆括号包围，ORDER BY和LIMIT可以附着在子表达式里。	-
INTERSECT子句	INTERSECT计算多个SELECT语句返回行集合的交集，不含重复的记录。 INTERSECT子句有如下约束条件： ● 同一个SELECT语句中的多个INTERSECT操作符是从左向右计算的，除非用圆括号进行了标识。 ● 当对多个SELECT语句的执行结果进行UNION和INTERSECT操作的时候，会优先处理INTERSECT。 一般形式： `select_statement INTERSECT select_statement` select_statement可以是任何没有FOR UPDATE子句的SELECT语句。	-

 版权所有 © 华为云计算技术有限公司

参数	描述	取值范围或示例
EXCEPT子句	EXCEPT操作符计算存在于左边SELECT语句的输出而不存在于右边SELECT语句输出的行。 EXCEPT子句有如下的通用形式： select_statement EXCEPT [ALL] select_statement ● select_statement是任何没有FOR UPDATE子句的SELECT表达式。 ● EXCEPT操作符计算存在于左边SELECT语句的输出而不存在于右边SELECT语句输出的行。 ● EXCEPT的结果不包含任何重复的行，除非声明了ALL选项。使用ALL时，一个在左边表中有m个重复而在右边表中有n个重复的行将在结果中出现 max(m-n,0) 次。 ● 除非用圆括号指明顺序，否则同一个SELECT语句中的多个EXCEPT操作符是从左向右计算的。EXCEPT和UNION的绑定级别相同。 ● 目前，不能给EXCEPT的结果或者任何EXCEPT的输入声明FOR UPDATE子句。	-
MINUS子句	与EXCEPT子句具有相同的功能和用法。	-
ORDER BY子句	对SELECT语句检索得到的数据进行升序或降序排序。对于ORDER BY表达式中包含多列的情况： ● 首先根据最左边的列进行排序，如果这一列的值相同，则根据下一个表达式进行比较，以此类推。 ● 如果对于所有声明的表达式都相同，则按随机顺序返回。 ● ORDER BY中排序的列必须包括在SELECT语句所检索的结果集的列中。 **须知** ● 如果未指定ORDER BY，则按数据库系统最快生成的顺序返回。 ● 可以选择在ORDER BY子句中的任何表达式之后添加关键字ASC（升序）或DESC（降序）。如果未指定，则默认使用ASC。 ● 如果要支持中文拼音排序和不区分大小写排序，需要在创建数据库时指定编码格式为UTF-8或GBK。	-

版权所有 © 华为云计算技术有限公司

参数	描述	取值范围或示例
nlssort_expression_clause	设置排序方式，其格式为： NLSSORT (column_name, ' NLS_SORT = { SCHINESE_PINYIN_M | generic_m_ci } ') NLS_SORT：指定某字段按照特殊方式排序。目前仅支持中文拼音格式排序和不区分大小写排序。取值如下： • SCHINESE_PINYIN_M，按照中文拼音排序（目前只支持GBK字符集内的一级汉字排序）。如果要支持此排序方式，在创建数据库时需要指定编码格式为“GBK”，否则排序无效。 • generic_m_ci，不区分大小写排序。	• **示例：通过NLSSORT子句实现查询结果按中文拼音首字母进行排序** • **示例：通过NLSSORT子句实现查询结果不区分大小写排序**
[{ [LIMIT { count | ALL }] [OFFSET start [ROW | ROWS]] } | { LIMIT start, { count | ALL } }]	LIMIT子句由两个独立的Limit子句、Offset子句和一个多参Limit子句构成： LIMIT { count | ALL } OFFSET start [ROW | ROWS] LIMIT start, { count | ALL } 其中，count声明返回的最大行数，而start声明跳过前面的start行数。如果这两个参数都指定了，会计算跳过start行之后返回的count行数。多参Limit子句不可和单参的Limit子句或Offset子句共同出现。 例如： LIMIT 10 OFFSET 20; --跳过前面20行，返回后面10行，即返回第21行到第30行 LIMIT 5,10; --跳过前面5行，返回后面10行，即返回第6行到第16行	-
FETCH { FIRST | NEXT } [count] { ROW | ROWS } ONLY	FETCH子句限定返回查询结果从第一行开始的总行数。	如果不指定count，默认值为1。

版权所有 © 华为云计算技术有限公司

参数	描述	取值范围或示例
FOR UPDATE子句	FOR UPDATE子句将对SELECT检索出来的行进行加锁，这样避免在当前事务结束前被其他事务修改或者删除。即执行UPDATE、 DELETE、 SELECT FOR UPDATE这些行的事务将被阻塞，直到当前事务结束。 NOWAIT选项：避免操作等待其他事务提交，如果被选择的行不能立即被锁，执行SELECT FOR UPDATE NOWAIT将会立即报错。 ● 如果在FOR UPDATE/SHARE中明确指定了表名，则只有这些指定的表被锁定，其他在SELECT中使用的表将不会被锁定。否则，将锁定该命令中所有使用的表。 ● 如果FOR UPDATE/SHARE应用于一个视图或者子查询，它同样将锁定所有该视图或子查询中使用到的表。 ● 多个FOR UPDATE/SHARE子句可以用于为不同的表指定不同的锁定模式。 ● 如果一个表中同时出现（或隐含同时出现）在FOR UPDATE/SHARE子句中，则按照FOR UPDATE处理。类似的，如果影响一个表的任意子句中出现了NOWAIT，该表将按照NOWAIT处理。 ● FOR SHARE/UPDATE：是对每个检索出来的行要求一个共享锁，而不是排他锁。一个共享锁阻塞其他事务执行UPDATE、DELETE、SELECT，而不阻塞SELECT FOR SHARE。 **须知** ● 对于FOR UPDATE/SHARE，执行计划不能下推的SQL，直接返回报错信息；对于执行计划可以下推的，下推到DN执行。 ● 对列存表的查询不支持FOR UPDATE/SHARE。	-

FROM 子句参数说明

FROM from_item 子句格式如下：

```
{[ ONLY ] table_name [ * ] [ partition_clause ] [ [ AS ] alias [ ( column_alias [, ...] ) ] ]
|( select ) [ AS ] alias [ ( column_alias [, ...] ) ]
|with_query_name [ [ AS ] alias [ ( column_alias [, ...] ) ] ]
|function_name ( [ argument [, ...] ] ) [ AS ] alias [ ( column_alias [, ...] | column_definition [, ...] ) ]
|function_name ( [ argument [, ...] ] ) AS ( column_definition [, ...] )
|from_item [ NATURAL ] join_type from_item [ ON join_condition | USING ( join_column [, ...] ) ]}
```

 版权所有 © 华为云计算技术有限公司

表 16-2 FROM 子句相关参数说明

参数	描述	取值范围或示例
table_name	表名或视图名，名称前可加上模式名，如：schema_name.table_name。	字符串，要符合标识符命名规范。 具体请参见**标识符命名规范**。
alias	给表或复杂的表引用起一个临时的表别名，以便被其余的查询引用。 别名用于缩写或者在自连接中消除歧义。如果提供了别名，它就会完全隐藏表的实际名字。	字符串，要符合标识符命名规范。 具体请参见**标识符命名规范**。
column_alias	列别名。	字符串，要符合标识符命名规范。 具体请参见**标识符命名规范**。
partition_clause	PARTITION子句，查询某个分区表中相应分区的数据。格式如下： `PARTITION { ( partition_name ) | FOR ( partition_value [, ...] ) }` • partition_name：分区名。 • partition_value：指定的分区键值。在创建分区表时，如果指定了多个分区键，可以通过PARTITION FOR子句指定的这一组分区键的值，唯一确定一个分区。 **须知** 指定分区只适合普通表。	示例： `SELECT * FROM tpcds.reason_p PARTITION (P_05_BEFORE);`
subquery	FROM子句中可以出现子查询，创建一个临时表保存子查询的输出。	例如，将子查询结果取别名为e，并作为FROM数据源： `SELECT d.department_name, e.avg_salary FROM ( SELECT department_id, AVG(salary) AS avg_salary FROM employees GROUP BY department_id) AS e JOIN departments d ON e.department_id = d.id;`

版权所有 © 华为云计算技术有限公司

参数	描述	取值范围或示例
with_query_name	WITH子句同样可以作为FROM子句的源，可以通过WITH查询的名字对其进行引用。	字符串，要符合标识符命名规范。 具体请参见**标识符命名规范**。 举例，WITH子句取名为cte，并作为FROM的引用源： `WITH cte AS (SELECT id, name FROM users) SELECT * FROM cte AS user_info;`
function_name	函数名称。函数调用也可以出现在FROM子句中。	字符串，要符合标识符命名规范。 具体请参见**标识符命名规范**。
join_type	JOIN的类型。	支持五种JOIN类型： • [INNER] JOIN • LEFT [OUTER] JOIN • RIGHT [OUTER] JOIN • FULL [OUTER] JOIN • CROSS JOIN 详情请参见**表16-3**。
ON join_condition	连接条件，用于限定连接中的哪些行是匹配的。如：ON left_table.a = right_table.a。	-
USING(join_column[, ...])	ON left_table.a = right_table.a AND left_table.b = right_table.b ... 的简写。要求对应的列必须同名。	-
NATURAL	NATURAL是具有相同名称的两个表的所有列的USING列表的简写。	-
from item	用于连接的查询源对象的名称。	-

版权所有 © 华为云计算技术有限公司

表 16-3 join_type 的五种类型

参数	描述
[INNER] JOIN	一个JOIN子句由两个FROM项组合而成。可使用圆括号以决定嵌套的顺序。如果没有圆括号，JOIN从左向右嵌套。在任何情况下，JOIN都比逗号分隔的FROM项绑定得更紧。
LEFT [OUTER] JOIN	返回笛卡尔积中所有符合连接条件的行，再加上左表中通过连接条件没有匹配到右表行中的那些行。这样，左边的行将扩展为生成表的全长，方法是在那些右表对应的字段位置填充上NULL。 需要注意的是，只在计算匹配的时候，才使用JOIN子句的条件，外层的条件是在计算完成之后施加的。
RIGHT [OUTER] JOIN	返回所有内连接的结果行，加上右表中未匹配的行（左表对应字段填充NULL）。右外连接等价于将左表和右表互换位置后的左外连接。
FULL [OUTER] JOIN	返回所有内连接的结果行，加上每个不匹配的左边行（右边用NULL扩展），再加上每个不匹配的右边行（左边用NULL扩展）。
CROSS JOIN	CROSS JOIN等效于INNER JOIN ON（TRUE），即没有被条件删除的行。与简单的FROM和WHERE的效果相同。 **说明** ● 必须为INNER和OUTER连接类型声明一个连接条件，即NATURAL ON，join_condition，USING (join_column [， ...])之一。但是它们不能出现在CROSS JOIN中。 ● CROSS JOIN和INNER JOIN生成一个简单的笛卡尔积，和在FROM的顶层列出两个项的结果相同。

GROUP BY 子句参数说明

其中group子句为：

```
( )
| expression
| ( expression [, ...] )
| ROLLUP ( { expression | ( expression [, ...] ) } [, ...] )
| CUBE ( { expression | ( expression [, ...] ) } [, ...] )
| GROUPING SETS ( grouping_element [, ...] )
```

 版权所有 © 华为云计算技术有限公司

表 16-4 GROUP BY 子句参数说明

参数	描述	示例

ROLLUP
({ expression |
(expression [, ...]) }
[, ...])
	ROLLUP是计算一个有序的分组列在GROUP BY中指定的标准聚集值，然后从右到左进一步创建高层次的部分和，最后创建了累积和。一个分组能够看做一系列的分组集。例如：
	`GROUP BY ROLLUP (a,b,c)`
	等价于：
	`GROUP BY GROUPING SETS((a,b,c), (a,b), (a), ( ))`
	ROLLUP子句中的元素可以是单独的字段或表达式，也可以是使用括号包含的列表。如果是括号中的列表，产生分组集时它们必须作为一个整体。例如：
	`GROUP BY ROLLUP ((a,b), (c,d))`
	等价于：
	`GROUPING SETS ((a,b,c,d), (a,b), (c,d ), ( ))`
		示例：使用GROUP BY进行分组查询（ROLLUP、CUBE、GROUPING SETS）

版权所有 © 华为云计算技术有限公司

参数	描述	示例

CUBE ({ expression | (expression [, ...]) } [, ...])	CUBE是自动对group by子句中列出的字段进行分组汇总，结果集将包含维度列中各值的所有可能组合，以及与这些维度值组合相匹配的基础行中的聚合值。它会为每个分组返回一行汇总信息， 用户可以使用CUBE来产生交叉表值。比如，在CUBE子句中给出三个表达式（n = 3），运算结果为$2^n = 2^3 = 8$组。 以n个表达式的值分组的行称为常规行，其余的行称为超级聚集行。例如：
	`GROUP BY CUBE (a,b,c)`
	等价于：
	`GROUP BY GROUPING SETS((a,b,c), (a,b), (a,c), (b,c), (a), (b), (c), ( ))`
	CUBE子句中的元素可以是单独的字段或表达式，也可以是使用括号包含的列表。如果是括号中的列表，产生分组集时它们必须作为一个整体。例如：
	`GROUP BY CUBE (a, (b, c), d)`
	等价于：
	`GROUP BY GROUPING SETS ((a,b,c,d), (a,b,c), (a), ( ))`
		示例：使用GROUP BY进行分组查询（ROLLUP、CUBE、GROUPING SETS）

版权所有 © 华为云计算技术有限公司

参数	描述	示例
GROUPING SETS (grouping_element [, ...])	GROUPING SETS子句是GROUP BY子句的进一步扩展，它可以使用户指定多个GROUP BY选项。选项用于定义分组集，每个分组集都需要包含在单独的括号中，空白的括号（()）表示将所有数据当作一个组处理。 这样做可以通过裁剪用户不需要的数据组来提高效率。 用户可以根据需要指定所需的数据组进行查询。 **须知** 如果SELECT列表的表达式中引用了那些没有分组的字段，则会报错，除非使用了聚集函数，因为对于未分组的字段，可能返回多个数值。	**示例：使用GROUP BY进行分组查询（ROLLUP、CUBE、GROUPING SETS）**

WHERE 子句使用约束

1. 对于WHERE子句的LIKE操作符，当LIKE中要查询特殊字符“%”、“_”、“\”的时候需要使用反斜杠“\”来进行转义。

```
CREATE TABLE tt01 (id int,content varchar(50));

INSERT INTO tt01 values (1,'Jack say ''hello''');
INSERT INTO tt01 values (2,'Rose do 50%');
INSERT INTO tt01 values (3,'Lilei say ''world''');
INSERT INTO tt01 values (4,'Hanmei do 100%');

SELECT * FROM tt01;
 id |      content
----+-------------------
  3 | Lilei say 'world'
  4 | Hanmei do 100%
  1 | Jack say 'hello'
  2 | Rose do 50%
(4 rows)

SELECT * FROM tt01 WHERE content like '%''he%';
 id |     content
----+------------------
  1 | Jack say 'hello'
```


版权所有 © 华为云计算技术有限公司

```
(1 row)

SELECT * FROM tt01 WHERE content like '%50\%%';
 id |   content
----+-------------
  2 | Rose do 50%
(1 row)
```

2. WHERE子句中可以通过指定"(+)"操作符的方法将表的连接关系转换为外连接。但是不建议用户使用这种用法，因为这并不是SQL的标准语法，在做平台迁移的时候可能面临语法兼容性的问题。使用"(+)"有很多限制如下：
   a. "(+)"只能出现在where子句中。
   b. 如果from子句中已经有指定表连接关系，那么"(+)"不能在where子句中使用。
   c. "(+)"只能作用在表或者视图的列上，不能作用在表达式上。
   d. 如果表A和表B有多个连接条件，那么必须在所有的连接条件中指定"(+)"，否则"(+)"将不会生效，表连接会转化成内连接，并且不给出任何提示信息。
   e. "(+)"作用的连接条件中的表不能跨查询或者子查询。如果"(+)"作用的表，不在当前查询或者子查询的from子句中，则会报错。如果"(+)"作用的对端的表不存在，则不报错，同时连接关系会转化为内连接。
   f. "(+)"作用的表达式不能直接通过"OR"连接。
   g. 如果"(+)"作用的列是和一个常量的比较关系， 那么这个表达式会成为join条件的一部分。
   h. 同一个表不能对应多个外表。
   i. "(+)"只能出现"比较表达式"，"NOT表达式"，“ANY表达式”，“ALL表达式”，“IN表达式”，“NULLIF表达式”，“IS DISTINCT FROM表达式”，“IS OF”表达式。"(+)"不能出现在其他类型表达式中，并且这些表达式中不允许出现通过“AND”和“OR”连接。
   j. "(+)"只能转化为左外连接或者右外连接，不能转化为全连接，即不能在一个表达式的两个表上同时指定"(+)"。

示例：使用 WITH 子句实现子查询

先通过子查询得到一张临时表temp_t（临时表不进行强制物化MATERIALIZED，是否物化由系统内部优化结果来决定），再查询表temp_t中的所有数据。

```
WITH temp_t(name,isdba) AS (SELECT usename,usesuper FROM pg_user) SELECT * FROM temp_t;
```

```
gaussdb=> WITH temp_t(name,isdba) AS (SELECT usename,usesuper FROM pg_user) SELECT * FROM temp_t;
  name   | isdba
---------+-------
 dbadmin | f
 Ruby    | t
(2 rows)
```

通过子查询得到一张临时表temp_t，并显式指定MATERIALIZED，即进行强制物化（强制把CTE结果存成临时表，执行一次，可重复使用）再查询表temp_t中的所有数据。

```
WITH temp_t(name,isdba) AS MATERIALIZED (SELECT usename,usesuper FROM pg_user) SELECT * FROM temp_t;
```

通过子查询得到一张临时表temp_t，并显式指定NOT MATERIALIZED，即不进行物化（不生成临时表，直接将子查询逻辑“内联”到主查询），再查询表temp_t中的所有数据。

 版权所有 © 华为云计算技术有限公司

```
WITH temp_t(name,isdba) AS NOT MATERIALIZED (SELECT usename,usesuper FROM pg_user)
 SELECT * FROM temp_t t1 WHERE name LIKE 'A%'
 UNION ALL
 SELECT * FROM temp_t t2 WHERE name LIKE 'B%';
```

示例：其他 SELECT 常用场景

准备数据。

```
DROP SCHEMA IF EXISTS tpcds CASCADE;
CREATE SCHEMA tpcds;
CREATE TABLE tpcds.reason ( r_reason_sk integer not null, r_reason_id char(16) not null, r_reason_desc
char(100) ) ;
INSERT INTO tpcds.reason (r_reason_sk, r_reason_id, r_reason_desc)  VALUES
  (1, 'RTRN001', 'Product quality problem or defective item'),
  (2, 'RTRN002', 'Size does not fit or incorrect size'),
  (3, 'RTRN003', 'Customer changed mind or did not like the product');
```

查询tpcds.reason表的所有r_reason_sk记录，且去除重复。

```
SELECT DISTINCT(r_reason_sk) FROM tpcds.reason;
```

LIMIT子句示例：获取表中一条记录。

```
SELECT * FROM tpcds.reason LIMIT 1;
```

LIMIT子句示例：获取表中第三条记录。

```
SELECT * FROM tpcds.reason LIMIT 1 OFFSET 2;
```

LIMIT子句示例：获取表中前两条记录。

```
SELECT * FROM tpcds.reason LIMIT 2;
```

查询所有记录，且按字母升序排列。

```
SELECT r_reason_desc FROM tpcds.reason ORDER BY r_reason_desc;
```

通过表别名，从pg_user和pg_user_status这两张表中获取数据。

```
SELECT a.usename,b.locktime FROM pg_user a,pg_user_status b WHERE a.usesysid=b.roloid;
```

FULL JOIN子句示例：将pg_user和pg_user_status这两张表的数据进行全连接显示，即数据的合集。

```
SELECT a.usename,b.locktime,a.usesuper FROM pg_user a FULL JOIN pg_user_status b on
a.usesysid=b.roloid;
```

GROUP BY子句示例：根据查询条件过滤，并对结果进行分组。

```
SELECT r_reason_id, AVG(r_reason_sk) FROM tpcds.reason GROUP BY r_reason_id HAVING
AVG(r_reason_sk) > 25;
```

GROUP BY子句示例：通过group by别名来对结果进行分组。

```
SELECT r_reason_id AS id FROM tpcds.reason GROUP BY id;
```

UNION子句示例：将表tpcds.reason里r_reason_desc字段中的内容以W开头和以N开头的进行合并。

```
SELECT r_reason_sk, tpcds.reason.r_reason_desc
   FROM tpcds.reason
   WHERE tpcds.reason.r_reason_desc LIKE 'W%'
UNION
SELECT r_reason_sk, tpcds.reason.r_reason_desc
   FROM tpcds.reason
   WHERE tpcds.reason.r_reason_desc LIKE 'N%';
```

 版权所有 © 华为云计算技术有限公司

示例：使用递归查询（WITH RECURSIVE）查询某公司三层部门的归属关系

以某公司的三层部门的递归为例，在一个包含一级、二级、三级部门的排序不规则的表中，通过递归查询（使用关键字RECURSIVE），查询出某个部门（例如技术部）下的所有二、三级部门。

准备数据：

```
DROP TABLE IF EXISTS departments;
CREATE TABLE departments
(id varchar(10),
pid varchar(10),
dept_name varchar(50));
INSERT INTO  departments VALUES('0',null,'Head office');
INSERT INTO  departments VALUES('1','0','Technology');     -- 一级部门
INSERT INTO  departments VALUES('2','0','Marketing');
INSERT INTO  departments VALUES('3','0','HR');
INSERT INTO  departments VALUES('4','0','Business');
INSERT INTO  departments VALUES('5','0','Maintenance');
INSERT INTO  departments VALUES('1-1','1','R&D Center');   -- 二级部门
INSERT INTO  departments VALUES('1-2','1','Pre-Research');
INSERT INTO  departments VALUES('1-3','1','Network');
INSERT INTO  departments VALUES('1-4','1','Test');
INSERT INTO  departments VALUES('1-5','1','Architecture');
INSERT INTO  departments VALUES('1-1-1','1-1','Frontend');   -- 三级部门
INSERT INTO  departments VALUES('1-1-2','1-1','Backend');
INSERT INTO  departments VALUES('1-1-3','1-1','Design');
INSERT INTO  departments VALUES('2-1','2','Branding');
INSERT INTO  departments VALUES('2-2','2','Pricing');
INSERT INTO  departments VALUES('2-3','2','Training');
```

查询表的原始数据，结果显示排序不规则，很难看出三层部门之间的归属关系。

```
SELECT * FROM departments;
```

id	pid	dept_name
1	0	Technology
4	0	Business
1-2	1	Pre-Research
1-5	1	Architecture
1-1-3	1-1	Design
2-3	2	Training
0	*(Null)*	Head office
3	0	HR

此时使用WITH RECURSIVE关键字，查询出技术部下面所有的二、三级部门，可以使用如下SQL：

版权所有 © 华为云计算技术有限公司

```
WITH RECURSIVE  t1 AS (         --定义一个公共表达式名称为t1
SELECT id,pid,dept_name FROM departments WHERE id='1'   --初始化递归的根节点：查询id为1的记录，即技术部
UNION ALL         --递归合并结果（保留重复项）
SELECT            --递归部分：逐层向下查询
t2.id,            --子节点id，各一级部门名称
t2.pid,           --父节点id，即总公司
t2.dept_name          --子节点名称
FROM departments t2
INNER JOIN t1 ON t2.pid = t1.id          --通过父节点ID连接递归结果
)
SELECT id,dept_name FROM t1 ORDER BY id;       --最终查询，按ID排序显示所有二、三级部门
```

id	dept_name
1	Technology
1-1	R&D Center
1-1-1	Frontend
1-1-2	Backend
1-1-3	Design
1-2	Pre-Research
1-3	Network
1-4	Test
1-5	Architecture

以上查询结果看起来还不太直观，可以将结果再稍微改进，将各个层级展示出来。

```
WITH RECURSIVE  t1 AS (

SELECT    --初始化递归的根节点，查询id为1的记录，即技术部，显示转换为text类型
 id,
 pid,
 dept_name::text   --显式转换为text类型（为后续路径拼接做准备）
FROM departments
WHERE id='1'

UNION ALL

SELECT    --递归部分：拼接层级路径
t2.id,
t2.pid,
(t1.dept_name||'->'||t2.dept_name)::text  AS dept_name   --拼接父级路径与当前节点名称，如Technology -> R&D Center
FROM departments t2
INNER JOIN t1 ON t2.pid = t1.id )

SELECT id,dept_name FROM t1 ORDER BY id;   --最终查询，显示带层级路径的结果
```


版权所有 © 华为云计算技术有限公司

id	dept_name
1	Technology
1-1	Technology->R&D Center
1-1-1	Technology->R&D Center->Frontend
1-1-2	Technology->R&D Center->Backend
1-1-3	Technology->R&D Center->Design
1-2	Technology->Pre-Research
1-3	Technology->Network
1-4	Technology->Test
1-5	Technology->Architecture

示例：通过 NLSSORT 子句实现查询结果按中文拼音首字母进行排序

NLS_SORT子句示例：中文拼音排序。该示例请在GBK编码的数据库中执行，可以先创建新的数据库。

```
CREATE DATABASE music ENCODING 'GBK' template = template0;
```

连接新的数据库music后再执行以下语句：

```
DROP TABLE IF EXISTS stu_pinyin_info;
CREATE TABLE stu_pinyin_info (id bigint, name text) DISTRIBUTE BY REPLICATION;
INSERT INTO stu_pinyin_info VALUES (1, '李雷'),(2, '石祥传'), (3,'安南鹏');
SELECT * FROM stu_pinyin_info ORDER BY NLSSORT (name, 'NLS_SORT = SCHINESE_PINYIN_M' );
```

```
music=> SELECT * FROM stu_pinyin_info ORDER BY NLSSORT (name, 'NLS_SORT = SCHINESE_PINYIN_M' );
 id |  name
----+--------
  3 | 安南鹏
  1 | 李雷
  2 | 石祥传
(3 rows)
```

示例：通过 NLSSORT 子句实现查询结果不区分大小写排序

不区分大小写排序:

```
DROP TABLE IF EXISTS stu_icase_info;
CREATE TABLE stu_icase_info (id bigint, name text) DISTRIBUTE BY REPLICATION;
INSERT INTO stu_icase_info VALUES (1, 'aaaa'),(2, 'AAAA');
SELECT * FROM stu_icase_info ORDER BY NLSSORT (name, 'NLS_SORT = generic_m_ci');
```

 版权所有 © 华为云计算技术有限公司

```
gaussdb=> SELECT * FROM stu_icase_info ORDER BY NLSSORT (name, 'NLS_SORT = generic_m_ci');
 id | name
----+------
  1 | aaaa
  2 | AAAA
(2 rows)
```

示例：查询分区表的数据

创建分区表tpcds.reason_p，并插入数据，再从tpcds.reason_p的表分区P_05_BEFORE中获取数据。

```
DROP SCHEMA IF EXISTS tpcds CASCADE;
CREATE SCHEMA tpcds;
CREATE TABLE tpcds.reason_p
(
  r_reason_sk integer,
  r_reason_id character(16),
  r_reason_desc character(100)
)
PARTITION BY RANGE (r_reason_sk)
(
  partition P_05_BEFORE values less than (05),
  partition P_15 values less than (15),
  partition P_25 values less than (25),
  partition P_35 values less than (35),
  partition P_45_AFTER values less than (MAXVALUE)
);

INSERT INTO tpcds.reason_p values(3,'AAAAAAAABAAAAAAA','reason 1'),
(10,'AAAAAAAABAAAAAAA','reason 2'),(4,'AAAAAAAABAAAAAAA','reason 3'),
(10,'AAAAAAAABAAAAAAA','reason 4'),(10,'AAAAAAAABAAAAAAA','reason 5'),
(20,'AAAAAAAACAAAAAAA','reason 6'),(30,'AAAAAAAACAAAAAAA','reason 7');

SELECT * FROM tpcds.reason_p PARTITION (P_05_BEFORE);
```

```
gaussdb=> SELECT * FROM tpcds.reason_p PARTITION (P_05_BEFORE);
 r_reason_sk |   r_reason_id    |                                                                  r_reason_desc
-------------+------------------+------------------------------------------------------------------------------------------------
           3 | AAAAAAAABAAAAAAA | reason 1
           4 | AAAAAAAABAAAAAAA | reason 3
(2 rows)
```

```
——查询分区P_15的行数：
SELECT count(*) FROM tpcds.reason_p PARTITION (P_15);
```

```
gaussdb=> SELECT count(*) FROM tpcds.reason_p PARTITION (P_15);
 count
-------
     3
(1 row)
```

GROUP BY子句示例：按r_reason_id分组统计tpcds.reason_p表中的记录数。

```
SELECT COUNT(*),r_reason_id FROM tpcds.reason_p GROUP BY r_reason_id;
```

 版权所有 © 华为云计算技术有限公司

```
gaussdb=> SELECT COUNT(*),r_reason_id FROM tpcds.reason_p GROUP BY r_reason_id;
 count |    r_reason_id
-------+------------------
     5 | AAAAAAAABAAAAAAA
     2 | AAAAAAAACAAAAAAA
(2 rows)
```

HAVING子句示例：按r_reason_id分组统计tpcds.reason_p表中的记录，并只显示r_reason_id个数大于2的信息。

```
SELECT COUNT(*) c,r_reason_id FROM tpcds.reason_p GROUP BY r_reason_id HAVING c>2;
```

```
gaussdb=> SELECT COUNT(*) c,r_reason_id FROM tpcds.reason_p GROUP BY r_reason_id HAVING c>2;
 c |    r_reason_id
---+------------------
 5 | AAAAAAAABAAAAAAA
(1 row)
```

IN子句示例：按r_reason_id分组统计tpcds.reason_p表中的r_reason_id个数，并只显示r_reason_id值为AAAAAAAABAAAAAAA或AAAAAAAADAAAAAAA的个数。

```
SELECT COUNT(*),r_reason_id FROM tpcds.reason_p GROUP BY r_reason_id HAVING r_reason_id
IN('AAAAAAAABAAAAAAA','AAAAAAAADAAAAAAA');
```

```
gaussdb=> SELECT COUNT(*),r_reason_id FROM tpcds.reason_p GROUP BY r_reason_id HAVING r_reason_id IN('AAAAAAAABAAAAAAA','AAAAAAAADAAAAAAA');
 count |   r_reason_id
-------+-----------------
     5 | AAAAAAAABAAAAAAA
(1 row)
```

INTERSECT子句示例：查询r_reason_id等于AAAAAAAABAAAAAAA，并且r_reason_sk小于5的信息。

```
SELECT * FROM tpcds.reason_p WHERE r_reason_id='AAAAAAAABAAAAAAA' INTERSECT SELECT * FROM
tpcds.reason_p WHERE r_reason_sk<5;
```

```
gaussdb=> SELECT * FROM tpcds.reason_p WHERE r_reason_id='AAAAAAAABAAAAAAA' INTERSECT SELECT * FROM tpcds.reason_p WHERE r_reason_sk<5;
 r_reason_sk |   r_reason_id    |                                                  r_reason_desc
-------------+------------------+----------------------------------------------------------------------------------------------------
           3 | AAAAAAAABAAAAAAA | reason 1
           4 | AAAAAAAABAAAAAAA | reason 3
(2 rows)
```

EXCEPT子句示例：查询r_reason_id等于AAAAAAAABAAAAAAA，并且去除r_reason_sk小于4的信息。

```
SELECT * FROM tpcds.reason_p WHERE r_reason_id='AAAAAAAABAAAAAAA' EXCEPT SELECT * FROM
tpcds.reason_p WHERE r_reason_sk<4;
```

```
gaussdb=> SELECT * FROM tpcds.reason_p WHERE r_reason_id='AAAAAAAABAAAAAAA' EXCEPT SELECT * FROM tpcds.reason_p WHERE r_reason_sk<4;
 r_reason_sk |   r_reason_id    |                                                  r_reason_desc
-------------+------------------+----------------------------------------------------------------------------------------------------
          10 | AAAAAAAABAAAAAAA | reason 2
           4 | AAAAAAAABAAAAAAA | reason 3
          10 | AAAAAAAABAAAAAAA | reason 4
          10 | AAAAAAAABAAAAAAA | reason 5
(4 rows)
```

在 WHERE 子句中指定(+)实现左、右连接

本例演示在WHERE子句中指定老式（+）来实现左右连接，注意，**该老式写法不再推荐，推荐使用标准LEFT JOIN/RIGHT JOIN标准 ANSI 语法，本示例仅供参考**。

 版权所有 © 华为云计算技术有限公司

准备数据。

```
DROP TABLE IF EXISTS customer;
DROP TABLE IF EXISTS store_returns;
CREATE TABLE customer (
    c_customer_sk integer PRIMARY KEY,
    c_customer_id char(16)
);
CREATE TABLE store_returns (
    sr_item_sk integer,
    sr_customer_sk integer
);

INSERT INTO customer (c_customer_sk, c_customer_id) VALUES (1, 'AAAAAAAAJNGEBAAA');
INSERT INTO store_returns (sr_item_sk, sr_customer_sk)  VALUES  (1, 0), (18000, 99);
```

通过在where子句中指定"(+)"来实现左连接。

```
select t1.sr_item_sk ,t2.c_customer_id from store_returns t1, customer t2 where t1.sr_customer_sk  =
t2.c_customer_sk(+)
order by 1 desc limit 1;
```

```
gaussdb=> SELECT t1.sr_item_sk ,t2.c_customer_id FROM store_returns t1, customer t2 WHERE t1.sr_customer_sk  = t2.c_customer_sk(+)
gaussdb-> ORDER BY 1 DESC LIMIT 1;
 sr_item_sk | c_customer_id
------------+---------------
      18000 |
(1 row)
```

通过在where子句中指定"(+)"来实现右连接。

```
select t1.sr_item_sk ,t2.c_customer_id from store_returns t1, customer t2 where t1.sr_customer_sk(+)  =
t2.c_customer_sk
order by 1 desc limit 1;
```

```
gaussdb=> SELECT t1.sr_item_sk ,t2.c_customer_id FROM store_returns t1, customer t2 WHERE t1.sr_customer_sk(+)  = t2.c_customer_sk
gaussdb-> ORDER BY 1 DESC LIMIT 1;
 sr_item_sk |  c_customer_id
------------+------------------
            | AAAAAAAAJNGEBAAA
(1 row)
```

通过在where子句中指定"(+)"来实现左连接，并且增加连接条件。

```
select t1.sr_item_sk ,t2.c_customer_id from store_returns t1, customer t2 where t1.sr_customer_sk  =
t2.c_customer_sk(+) and t2.c_customer_sk(+) < 1 order by 1  limit 1;
```

```
gaussdb=> SELECT t1.sr_item_sk ,t2.c_customer_id FROM store_returns t1, customer t2 WHERE t1.sr_customer_sk  = t2.c_customer_sk(+) and t2.c_customer_sk(+) < 1 ORDER BY 1  LIMIT 1;
 sr_item_sk | c_customer_id
------------+---------------
          1 |
(1 row)
```

不支持在where子句中指定"(+)"的同时使用内层嵌套AND/OR的表达式。

```
select t1.sr_item_sk ,t2.c_customer_id from store_returns t1, customer t2 where not(t1.sr_customer_sk  =
t2.c_customer_sk(+) and t2.c_customer_sk(+) < 1);
```

```
gaussdb=> SELECT t1.sr_item_sk ,t2.c_customer_id FROM store_returns t1, customer t2 WHERE NOT(t1.sr_customer_sk  = t2.c_customer_sk(+) AND t2.c_customer_sk(+) < 1);
ERROR:  Operator "(+)" cannot be used in nesting expression.
LINE 1: ...tomer_id FROM store_returns t1, customer t2 WHERE NOT(t1.sr_...
                                                             ^
```

where子句在不支持表达式宏指定"(+)"会报错。

```
select t1.sr_item_sk ,t2.c_customer_id from store_returns t1, customer t2 where (t1.sr_customer_sk  =
t2.c_customer_sk(+))::bool;
```

```
gaussdb=> SELECT t1.sr_item_sk ,t2.c_customer_id FROM store_returns t1, customer t2 WHERE (t1.sr_customer_sk  = t2.c_customer_sk(+))::bool;
ERROR:  Operator "(+)" can only be used in common expression.
```


版权所有 © 华为云计算技术有限公司

where子句在表达式的两边都指定"(+)"会报错。

```
select t1.sr_item_sk ,t2.c_customer_id from store_returns t1, customer t2 where t1.sr_customer_sk(+)  = t2.c_customer_sk(+);
```

```
gaussdb=> SELECT t1.sr_item_sk ,t2.c_customer_id FROM store_returns t1, customer t2 WHERE t1.sr_customer_sk(+)  = t2.c_customer_sk(+);
ERROR:  Operator "(+)" can't be specified on more than one relation in one join condition
HINT:  "t1", "t2"...are specified Operator "(+)" in one condition.
```

示例：使用 GROUP BY 进行分组查询（ROLLUP、CUBE、GROUPING SETS）

在一些报表场景中，经常会对数据做分组统计（group by），例如对一级部门下辖的二级部门员工数进行统计。

```
DROP TABLE IF EXISTS emp;
CREATE TABLE emp(
id int primary key,    --工号
name text,             --员工名
dep_1 text,            --一级部门
dep_2 text             --二级部门
);

INSERT INTO emp VALUES
(01, 'liwei', 'Product', 'SRE'),
(02, 'liming', 'Product', 'U1'),
(03, 'hanlei', 'Product', 'U1'),
(04, 'hanmeimei', 'Product', 'SRE'),
(05, 'dingjun', 'Product', 'SRE'),
(06, 'huxue', 'Product', 'E1'),
(07, 'liuyi', 'Product', 'E1'),
(08, 'luhua', 'Product', 'E1'),
(09, 'yangjing', 'Product', 'U1'),
(10, 'binbin', 'Product', 'U1'),
(11, 'jim', 'Product', 'SRE'),
(12, 'leo', 'Product', 'SER');


SELECT COUNT(*), dep_2 FROM emp GROUP BY dep_2;   --统计二级部门的员工人数
```

count	dep_2
1	SER
3	E1
4	SRE
4	U1

常见的统计报表业务中，通常需要进一步计算一级部门的“合计”人数，也就是二级部门各分组的累加，就可以借助于rollup，如下所示，比前面的分组计算结果多了一行合计的数据。如下图的“12 Null”，表示一级部门总人数12。

```
SELECT COUNT(*), dep_2 FROM emp GROUP BY ROLLUP(dep_2) ORDER BY 1;
```


版权所有 © 华为云计算技术有限公司

count	dep_2
1	SER
3	E1
4	U1
4	SRE
12	(Null)

ROLLUP，CUBE，GROUPING SETS的用法如下：

- **使用ROLLUP**：

  首先创建一张用户信息表customer：

```
DROP TABLE IF EXISTS customer;
CREATE TABLE customer
(
c_id char(16) not null,         --用户ID
c_name char(20) ,               --用户名
c_age integer ,                --年龄
c_country varchar(20) ,          --地区
c_class char(10),               --用户级别
c_gender text,                 --性别
c_balance numeric );            --余额

INSERT INTO customer VALUES(1, 'tom', '20', 'Region A', '1', 'male', 300);
INSERT INTO customer VALUES(2, 'jack', '30', 'Region B', '1', 'male', 100);
INSERT INTO customer VALUES(3, 'rose', '40', 'Region C', '1', 'female', 200);
INSERT INTO customer VALUES(4, 'Frank', '60', 'Region D', '1', 'male', 100);
INSERT INTO customer VALUES(5, 'Leon', '20', 'Region A', '2', 'male', 200);
INSERT INTO customer VALUES(6, 'Lucy', '20', 'Region A', '1', 'female', 500);
```

  ROLLUP是在分组计算基础上增加了合计，从字面意思理解，就是从最小聚合级开始，聚合单位逐渐扩大，例如如下语句。

```
--首先对(c_country, c_class)进行GROUP BY，接着对(c_country)进行GROUP BY,最后对全表进行GROUP BY操作。
SELECT c_country, c_class, sum(c_balance) FROM customer GROUP BY ROLLUP(c_country, c_class) ORDER BY 1,2,3;
```

 版权所有 © 华为云计算技术有限公司

c_country	c_class	sum
Region A	1	800
Region A	2	200
Region A	*(Null)*	1000
Region B	1	100
Region B	*(Null)*	100
Region C	1	200
Region C	*(Null)*	200
Region D	1	100
Region D	*(Null)*	100
(Null)	*(Null)*	1400

Total Records: 10

以上语句等价于：

```
SELECT c_country, c_class, SUM(c_balance) FROM customer GROUP BY c_country, c_class UNION ALL
SELECT c_country, null, SUM(c_balance) FROM customer GROUP BY c_country
UNION ALL
SELECT null, null, SUM(c_balance) FROM customer ORDER BY 1,2,3;
```

- **使用CUBE**：

  CUBE从字面意思理解就是各个维度的意思，也就是说全部组合，即聚合键中所有字段的组合的分组统计结果，例如如下语句：

  ```
  --首先对(c_country,c_class)进行GROUP BY，然后依次对(c_country)、(c_class)进行GROUP BY，最后对全表进行GROUP BY操作。
  SELECT c_country, c_class, SUM(c_balance) FROM customer GROUP BY CUBE(c_country, c_class)
  ORDER BY 1,2,3;
  ```

 版权所有 © 华为云计算技术有限公司

c_country	c_class	sum
Region A	1	800
Region A	2	200
Region A	*(Null)*	1000
Region B	1	100
Region B	*(Null)*	100
Region C	1	200
Region C	*(Null)*	200
Region D	1	100
Region D	*(Null)*	100
(Null)	1	1200
(Null)	2	200
(Null)	*(Null)*	1400

Total Records: 12

以上语句等价于：

```
SELECT c_country, c_class, SUM(c_balance) FROM customer GROUP BY c_country, c_class
union all
SELECT c_country, null, SUM(c_balance) FROM customer GROUP BY c_country
union all
SELECT null, null, SUM(c_balance) FROM customer
union all
SELECT NULL, c_class, SUM(c_balance) FROM customer GROUP BY c_class order by 1,2,3;
```

- **使用GROUPING SETS**:

  GROUPING SETS区别于ROLLUP和CUBE，并没有总体的合计功能，相当于从ROLLUP和CUBE的结果中提取出部分记录，例如如下语句：

  ```
  --分别对(c_class)、(c_country)进行GROUP BY计算
  SELECT c_country, c_class, SUM(c_balance) FROM customer GROUP BY GROUPING SETS(c_country, c_class) ORDER BY 1,2,3;
  ```

 版权所有 © 华为云计算技术有限公司

c_country	c_class	sum
Region A	(Null)	1000
Region B	(Null)	100
Region C	(Null)	200
Region D	(Null)	100
(Null)	1	1200
(Null)	2	200

Total Records: 6

以上语句等价于：

```
SELECT c_country, null, SUM(c_balance) FROM customer GROUP BY c_country
UNION ALL
SELECT null, c_class, sum(c_balance) FROM customer GROUP BY c_class ORDER BY 1,2,3;
```

16.3 SELECT INTO

功能描述

SELECT INTO用于根据查询结果创建一个新表，并且将查询到的数据插入到新表中。

数据并不返回给客户端，这一点和普通的SELECT不同。新表的字段具有和SELECT的输出字段相同的名字和数据类型。

注意事项

CREATE TABLE AS的作用和SELECT INTO类似，且提供了SELECT INTO所提供功能的超集。建议使用CREATE TABLE AS语法替代SELECT INTO，因为SELECT INTO不能在存储过程中使用并且SELECT INTO字段名，不支持接收空行。

语法格式

```
[ WITH [ RECURSIVE ] with_query [, ...] ]
SELECT [ ALL | DISTINCT [ ON ( expression [, ...] ) ] ]
    { * | {expression [ [ AS ] output_name ]} [, ...] }
    INTO [ [ GLOBAL | LOCAL | VOLATILE ] { TEMPORARY | TEMP } | UNLOGGED ] [ TABLE ] new_table
    [ FROM from_item [, ...] ]
    [ WHERE condition ]
    [ GROUP BY expression [, ...] ]
    [ HAVING condition [, ...] ]
    [ WINDOW {window_name AS ( window_definition )} [, ...] ]
    [ { UNION | INTERSECT | EXCEPT | MINUS } [ ALL | DISTINCT ] select ]
    [ ORDER BY {expression [ [ ASC | DESC | USING operator ] | nlssort_expression_clause ] [ NULLS { FIRST |
LAST } ]} [, ...] ]
    [ { [ LIMIT { count | ALL } ] [ OFFSET start [ ROW | ROWS ] ] } | { LIMIT start, { count | ALL } } ]
    [ FETCH { FIRST | NEXT } [ count ] { ROW | ROWS } ONLY ]
    [ {FOR { UPDATE | SHARE } [ OF table_name [, ...] ] [ NOWAIT ]} [...] ];
```

 版权所有 © 华为云计算技术有限公司

参数说明

INTO [[GLOBAL | LOCAL | VOLATILE] { TEMPORARY | TEMP } | UNLOGGED] [TABLE] new_table

[GLOBAL | LOCAL | VOLATILE] { TEMPORARY | TEMP }指定临时表类型，具体可参考**参数说明**章节中GLOBAL | LOCAL | VOLATILE相关说明。

UNLOGGED指定表为非日志表。写入非日志表中的数据不会写入到预写日志中，这样就会比普通表快很多。但是，非日志表也是不安全的，在冲突或异常关机后会被自动删截。非日志表中的内容也不会被复制到备用服务器中。在该类表中创建的索引也不会被自动记录。

new_table指定新建表的名字。

说明

SELECT INTO的其他参数可参考SELECT的**SELECT主句参数说明**。

示例

将reason_t表中TABLE_SK小于3的值加入到新建表中。

```
DROP TABLE IF EXISTS reason_t;
CREATE TABLE reason_t
(
   TABLE_SK         INTEGER             ,
   TABLE_ID         VARCHAR(20)          ,
   TABLE_NA          VARCHAR(20)
);
INSERT INTO reason_t VALUES (1, 'S01', 'StudentA'),(2, 'T01', 'TeacherA'),(3, 'T02', 'TeacherB'),(3, 'S02',
'StudentB');

SELECT * INTO reason_t_bck FROM reason_t WHERE TABLE_SK < 3;
SELECT * FROM reason_t_bck;
```

```
gaussdb=> SELECT * FROM reason_t_bck;
 table_sk | table_id | table_na
----------+----------+----------
        2 | T01      | TeacherA
        1 | S01      | StudentA
(2 rows)
```

相关链接

SELECT

版权所有 © 华为云计算技术有限公司

17 TCL 语法

17.1 TCL 语法一览表

TCL（Transaction Control Language事务控制语言 ），用来控制数据库操纵事务发生的时间及效果，对数据库实行监视等。

提交

DWS通过COMMIT或者END可完成提交事务的功能。请参考**COMMIT | END**。

设置保存点

DWS用于在当前事务里建立一个新的保存点。请参考**SAVEPOINT**。

回滚

DWS回滚当前事务状态回到上次最后提交的状态。请参考**ROLLBACK**。

17.2 ABORT

功能描述

回滚当前事务并且撤销所有当前事务中所做的更改。

作用等同于**ROLLBACK**，早期SQL有用ABORT，现在推荐使用ROLLBACK。

关于更多**事务管理**，请参见**事务管理**。

注意事项

在事务外部执行ABORT语句不会影响事务的执行，但是会产生一个警告信息。

语法格式

```
ABORT [ WORK | TRANSACTION ];
```

 版权所有 © 华为云计算技术有限公司

参数说明

WORK | TRANSACTION

可选关键字，对ABORT操作没有影响。

示例

以下演示，使用ABORT语句回滚撤销修改账号余额的过程。

步骤1 创建测试表并导入数据。

```
-- 创建测试表
DROP TABLE IF EXISTS user_account;
CREATE TABLE  user_account (
    id INT PRIMARY KEY,
    username VARCHAR(20),
    balance DECIMAL(10,2)
);
-- 插入初始数据
INSERT INTO user_account VALUES (1, 'lily', 100.00);
INSERT INTO user_account VALUES (2, 'lilei', 200.00);
-- 查看初始数据
SELECT * FROM user_account;
```

```
gaussdb=> SELECT * FROM user_account;
 id | username | balance
----+----------+---------
  1 | lily     |  100.00
  2 | lilei    |  200.00
(2 rows)
```

步骤2 事务内使用ABORT进行回滚。

1. 开启事务。
   ```
   BEGIN TRANSACTION;
   ```
2. 执行修改操作（扣减lily余额）。
   ```
   UPDATE user_account SET balance = balance - 50 WHERE id = 1;
   ```
3. 查看修改后的数据（事务内临时生效）。
   ```
   SELECT * FROM user_account;
   ```
   ```
   gaussdb=> SELECT * FROM user_account;
    id | username | balance
   ----+----------+---------
     2 | lilei    |  200.00
     1 | lily     |   50.00
   (2 rows)
   ```
4. 执行 ABORT 回滚（撤销所有修改）。
   ```
   ABORT;
   ```
   以上还有**等价写法**，例如ABORT WORK; 或者ABORT TRANSACTION; 现代写法用ROLLBACK;
5. 验证回滚结果：数据恢复为初始值。
   ```
   SELECT * FROM user_account;
   ```

 版权所有 © 华为云计算技术有限公司

```
gaussdb=> ABORT;
ROLLBACK
gaussdb=> SELECT * FROM user_account;
 id | username | balance
----+----------+---------
  1 | lily     |  100.00
  2 | lilei    |  200.00
(2 rows)
```

----结束

相关链接

事务管理，**SET TRANSACTION**，**COMMIT | END**，**ROLLBACK**

17.3 BEGIN

功能描述

BEGIN可以用于开始一个匿名块，也可以用于开始一个事务。本节描述用BEGIN开始匿名块的语法，以BEGIN开始事务的语法见**START TRANSACTION**。

匿名块是能够动态地创建和执行过程代码的结构，而不需要以持久化的方式将代码作为数据库对象储存在数据库中。

注意事项

无。

语法格式

- 开启匿名块

```
[DECLARE [declare_statements]]
BEGIN
execution_statements
END;
/
```

- 开启事务

```
BEGIN [ WORK | TRANSACTION ]
 [
  {
    ISOLATION LEVEL { READ COMMITTED | READ UNCOMMITTED | SERIALIZABLE | REPEATABLE
READ }
    | { READ WRITE | READ ONLY }
  } [, ...]
 ];
```


版权所有 © 华为云计算技术有限公司

参数说明

表 17-1 BEGIN 参数说明

参数	描述	取值范围
declare_statements	声明变量，包括变量名和变量类型，如“sales_cnt int”。	-
execution_statements	匿名块中要执行的语句。	已存在的函数名称。

示例

- 开始事务块：
  ```
  BEGIN;
  ```
- 要以可重复读隔离级别开始事务块：
  ```
  BEGIN TRANSACTION ISOLATION LEVEL REPEATABLE READ;
  ```
- 使用匿名块输出字符串：
  ```
  BEGIN
  dbms_output.put_line('Hello');
  END;
  /
  ```

相关链接

START TRANSACTION

17.4 CHECKPOINT

功能描述

检查点（CHECKPOINT）是一个事务日志中的点，所有数据文件都在该点被更新以反映日志中的信息，所有数据文件都将被刷新到磁盘。

设置事务日志检查点。预写式日志（WAL）缺省时在事务日志中每隔一段时间放置一个检查点。可以设置相关运行时参数（checkpoint_segments和checkpoint_timeout）来调整这个原子化检查点的间隔。

注意事项

- 只有系统管理员可以调用CHECKPOINT。
- CHECKPOINT强制立即进行检查，而不是等到下一次调度时的检查点。

语法格式

```
CHECKPOINT;
```

参数说明

无。

 版权所有 © 华为云计算技术有限公司

示例

设置检查点：

```
CHECKPOINT;
```

17.5 COMMIT | END

功能描述

通过COMMIT或者END可完成提交事务的功能，即提交事务的所有操作。

注意事项

执行COMMIT这个命令的时候，命令执行者必须是该事务的创建者或系统管理员，且创建和提交操作可以不在同一个会话中。

语法格式

```
{ COMMIT | END } [ WORK | TRANSACTION ] ;
```

参数说明

表 17-2 COMMIT | END 参数说明

参数	描述
COMMIT | END	通过COMMIT或END提交当前事务，让所有当前事务更改为其他事务可见。
WORK | TRANSACTION	可选关键字，仅提供可读性，对提交事务操作没有影响。

示例

提交事务，让所有更改永久化：

```
COMMIT;
```

相关链接

ROLLBACK

17.6 COMMIT PREPARED

功能描述

在分布式数据库的两阶段提交（Two-Phase Commit, 2PC）过程中，事务的提交分为两个阶段：第一阶段执行PREPARE TRANSACTION将事务置为“预备”状态，第二阶段执行COMMIT PREPARED或ROLLBACK PREPARED完成提交或回滚。当预备事务的

 版权所有 © 华为云计算技术有限公司

第一阶段已成功完成，且所有参与方均已确认可以提交时，可使用COMMIT PREPARED命令完成事务的最终提交，使事务中的数据变更持久化并对所有会话可见。

典型业务场景包括：

- **分布式事务正常提交**：两阶段提交的第一阶段（PREPARE）已成功完成，所有参与节点均回复“可以提交”，协调者发出COMMIT PREPARED完成最终提交。
- **故障恢复后提交**：协调者在第一阶段后发生故障，恢复后确认各参与者的预备事务仍在，发出COMMIT PREPARED完成延迟提交。
- **跨会话提交**：预备事务由一个会话创建（**PREPARE TRANSACTION**），由另一个会话提交（COMMIT PREPARED），适用于协调者与参与者不在同一会话的场景。

注意事项

- 该功能仅在维护模式（GUC参数xc_maintenance_mode为on时）下可用。该模式谨慎打开，一般供维护人员排查问题使用，普通用户不应使用该模式。
- 命令执行者必须是该事务的创建者或系统管理员，且创建和提交操作可以不在同一个会话中。
- 事务功能由数据库自动维护，不应显式使用事务功能。

语法格式

```
COMMIT PREPARED transaction_id ;
COMMIT PREPARED transaction_id WITH CSN;
```

参数说明

表 17-3 COMMIT PREPARED 参数说明

参数	描述	取值范围
transaction_id	待提交的预备事务的标识符。该标识符为执行PREPARE TRANSACTION时指定的全局事务ID（GID），必须为当前已存在的预备事务的标识符。	字符串类型，长度必须小于200字节。示例值：'txn_transfer_001'。
CSN(commit sequence number)	待提交事务的序列号（Commit Sequence Number），是一个64位递增无符号数字。使用WITH CSN语法时，系统按指定的CSN进行提交。	64位无符号整数。可通过查询pg_prepared_xacts视图或事务相关系统视图获取。

示例

以下示例演示DWS运维中的实际场景：查询数据库内部创建的残留预备事务，并通过COMMIT PREPARED完成最终提交。

步骤1 开启维护模式。（**开启后可能会造成数据不一致的风险**，需联系技术支持人员后开启）。

```
SET xc_maintenance_mode = on;
```

 版权所有 © 华为云计算技术有限公司

步骤2 查询当前存在的预备事务，获取需要提交的事务标识符（gid字段值）。

```
SELECT transaction, gid, prepared, owner, database FROM pg_prepared_xacts;
```

预期输出（假设存在一个残留的预备事务）：

```
transaction |       gid        |        prepared         |  owner   | database
------------+------------------+-------------------------+----------+----------
    1234567 | txn_distribute_01| 2024-01-15 14:30:00+08 | app_user | db_test
(1 row)
```

其中gid字段的值txn_distribute_01即为COMMIT PREPARED命令需要指定的transaction_id。

步骤3 执行COMMIT PREPARED，提交GID为’txn_distribute_01’的预备事务。

```
COMMIT PREPARED 'txn_distribute_01';
```

步骤4 验证预备事务是否已成功提交。再次查询pg_prepared_xacts视图，确认目标事务已不存在。

```
SELECT transaction, gid FROM pg_prepared_xacts WHERE gid = 'txn_distribute_01';
```

预期输出（查询返回空结果，表示预备事务已被提交终结）：

```
transaction | gid
------------+-----
(0 rows)
```

步骤5 关闭维护模式。

```
SET xc_maintenance_mode = off;
```

----结束

相关链接

PREPARE TRANSACTION，**ROLLBACK PREPARED**

常见问题

Q：如何查询当前存在的预备事务?

A：可以通过查询系统视图pg_prepared_xacts查看当前所有处于预备状态的事务：

```
SELECT transaction, gid, prepared, owner, database FROM pg_prepared_xacts;
```

其中gid字段的值即为各预备事务的标识符，用于后续COMMIT PREPARED或ROLLBACK PREPARED操作。

Q：COMMIT PREPARED 和 COMMIT PREPARED WITH CSN 有什么区别?

A：COMMIT PREPARED按默认方式提交预备事务；COMMIT PREPARED WITH CSN允许指定提交序列号（CSN），系统将按指定的CSN顺序进行提交。WITH CSN语法通常用于需要精确控制事务提交顺序的场景，一般使用默认的COMMIT PREPARED即可。

17.7 PREPARE TRANSACTION

功能描述

在分布式数据库中，当一个事务涉及多个节点或多个数据源时，需要通过两阶段提交（Two-Phase Commit, 2PC）协议来保证所有参与者要么全部提交、要么全部回滚。

 版权所有 © 华为云计算技术有限公司

PREPARE TRANSACTION命令用于完成两阶段提交的第一阶段——将当前事务的状态持久化到磁盘上，使其进入“预备”状态。

执行该命令后，事务不再与当前会话关联，其状态完全保存在磁盘上。即使在此后、正式提交请求之前数据库发生崩溃，该预备事务在数据库恢复后仍具有极高的提交成功率。

一旦事务进入预备状态，可以在稍后通过**COMMIT PREPARED**或**ROLLBACK PREPARED**命令分别进行提交或回滚。这两条命令可以从任意会话中发出，不限于最初执行事务的会话。

典型使用场景：

- **跨节点分布式事务**：事务涉及多个数据库节点，需要协调者先通知各参与者进入预备状态，再统一决定提交或回滚。
- **跨系统数据一致性**：事务涉及数据库与外部系统（如消息队列、缓存）的协同操作，外部系统确认准备就绪后，再通过COMMIT PREPARED完成最终提交。
- **故障恢复处理**：两阶段提交过程中协调者发生故障，参与者已进入预备状态并持有资源，等待协调者恢复后发出提交或回滚指令。

从发出PREPARE TRANSACTION命令的会话角度来看，该命令执行后当前事务即结束，不再有活跃的当前事务，且预备事务的数据变更效果在该会话中不可见（效果将在事务最终提交时重新可见）。

如果PREPARE TRANSACTION因某些原因执行失败，则当前事务会被自动回滚取消。

注意事项

- PREPARE TRANSACTION通常由分布式事务协调器自动调用，**一般用户不应在应用中显式使用此命令，此语法仅为参考**。
- 该功能仅在维护模式（GUC参数xc_maintenance_mode为on时）下可用。该模式谨慎打开，一般供维护人员排查问题使用，一般用户不应使用该模式。
- 在运行PREPARE TRANSACTION命令时，必须在postgresql.conf配置文件中增大max_prepared_transactions的数值。建议至少将其设置为等于max_connections，这样每个会话都可以有一个等待中的预备事务。

语法格式

```
PREPARE TRANSACTION transaction_id;
```

参数说明

表 17-4 PREPARE TRANSACTION 参数说明

参数	描述	取值范围
transaction_id	待提交事务的标识符，用于后面在COMMIT PREPARED或ROLLBACK PREPARED时标识这个事务。该标识符即为全局事务ID（GID）。	字符串类型，长度必须小于200字节。不能与当前其他预备事务已使用的标识符同名。示例值：'txn_transfer_001'。

版权所有 © 华为云计算技术有限公司

示例

以下示例以银行转账业务为背景，演示PREPARE TRANSACTION的语法用法和预备事务的验证方法。

⚠ 注意

在DWS分布式环境下，PREPARE TRANSACTION由内部分布式事务协调器自动调用，不支持通过客户端手动执行。以下示例中的PREPARE TRANSACTION步骤为语法说明，实际环境中预备事务由数据库内部自动创建。如需查看和终结内部创建的预备事务，请参见**COMMIT PREPARED**和**ROLLBACK PREPARED**。

步骤1 （可选）清理残留的同名预备事务。如果之前执行过相同GID的PREPARE TRANSACTION但未COMMIT或ROLLBACK，需先清理。

```
SET xc_maintenance_mode = on;
SELECT transaction, gid FROM pg_prepared_xacts WHERE gid = 'txn_transfer_001';
-- 如果存在残留记录，执行以下命令清理：
ROLLBACK PREPARED 'txn_transfer_001';
```

步骤2 创建测试表并插入初始数据。

```
DROP TABLE IF EXISTS accounts;
CREATE TABLE accounts (
    id      INT PRIMARY KEY,
    name    VARCHAR(50),
    balance NUMERIC(10,2)
);
INSERT INTO accounts VALUES (1, 'Alice', 1000.00);
INSERT INTO accounts VALUES (2, 'Bob', 500.00);
```

步骤3 开启事务，执行转账操作：Alice向Bob转账200元。

```
BEGIN;
UPDATE accounts SET balance = balance - 200.00 WHERE id = 1;
UPDATE accounts SET balance = balance + 200.00 WHERE id = 2;
```

步骤4 执行PREPARE TRANSACTION，将当前事务置为预备状态，并指定全局事务标识符为’txn_transfer_001’。(仅为参考，客户端无法执行)

```
PREPARE TRANSACTION 'txn_transfer_001';
```

执行成功后，当前会话不再有活跃事务。

步骤5 验证事务已进入预备状态。通过查询pg_prepared_xacts系统视图，确认预备事务已创建，并获取其事务标识符（gid字段值）。

```
SELECT transaction, gid, prepared, owner, database FROM pg_prepared_xacts;
```

预期输出：

```
transaction |       gid        |          prepared          |  owner   | database
------------+------------------+----------------------------+----------+----------
    1234567 | txn_transfer_001 | 2024-01-15 14:30:00+08 | app_user | db_test
(1 row)
```

其中gid字段的值txn_transfer_001即为后续执行COMMIT PREPARED或ROLLBACK PREPARED时需要指定的transaction_id。

步骤6 确认预备事务的数据变更在当前会话中不可见（因为事务尚未提交）。

```
SELECT * FROM accounts ORDER BY id;
```


版权所有 © 华为云计算技术有限公司

预期输出：

```
id | name  | balance
----+-------+---------
  1 | Alice | 1000.00
  2 | Bob   |  500.00
(2 rows)
```

后续操作请参见**COMMIT PREPARED**（提交预备事务）或**ROLLBACK PREPARED**（回滚预备事务）

----结束

相关链接

COMMIT PREPARED，**ROLLBACK PREPARED**

常见问题

Q：如何查询当前存在的预备事务?

A：可以通过查询系统视图pg_prepared_xacts查看当前所有处于预备状态的事务：

```
SELECT transaction, gid, prepared, owner, database FROM pg_prepared_xacts;
```

其中gid字段的值即为各预备事务的标识符，用于后续COMMIT PREPARED或ROLLBACK PREPARED操作。

17.8 SAVEPOINT

功能描述

SAVEPOINT用于在当前事务里建立一个新的保存点。

保存点是事务中的一个特殊记号，它允许将那些在它建立后执行的命令全部回滚，把事务的状态恢复到保存点所在的时刻。

注意事项

- 使用ROLLBACK TO SAVEPOINT回滚到一个保存点。使用RELEASE SAVEPOINT删除一个保存点，但是保留该保存点建立后执行的命令的效果。
- 保存点只能在一个事务块里面建立。在一个事务里面可以定义多个保存点。
- 函数、匿名块和存储过程中不支持使用SAVEPOINT语法。
- 由于节点故障或者通信故障引起的分布式节点线程或进程退出导致的报错，以及由于COPY FROM操作中源数据与目标表的表结构不一致导致的报错，均不能正常回滚到保存点之前，而是整个事务回滚。
- SQL标准要求，使用savepoint建立一个同名保存点时，需要自动删除前面那个同名保存点。在DWS数据库里，将保留旧的保存点，但是在回滚或者释放的时候，只使用最近的那个。释放了新的保存点将导致旧的再次成为ROLLBACK TO SAVEPOINT和RELEASE SAVEPOINT可以访问的保存点。除此之外，SAVEPOINT是完全符合SQL标准的。

语法格式

```
SAVEPOINT savepoint_name;
```


版权所有 © 华为云计算技术有限公司

参数说明

savepoint_name

新建保存点的名字。

示例

- 建立一个保存点，然后撤销建立保存点后执行的所有命令的效果：

```
DROP TABLE IF EXISTS table1;
CREATE TABLE table1 (a int);

START TRANSACTION;
INSERT INTO table1 VALUES (1);
SAVEPOINT my_savepoint;
INSERT INTO table1 VALUES (2);
ROLLBACK TO SAVEPOINT my_savepoint;
INSERT INTO table1 VALUES (3);
COMMIT;
```

查询表的内容，会同时看到1和3，不能看到2，因为2被回滚。

```
SELECT * FROM table1;
```

```
gaussdb=> SELECT * FROM table1;
 a
---
 3
 1
(2 rows)
```

- 建立并随后销毁一个保存点：

```
START TRANSACTION;
INSERT INTO table1 VALUES (3);
SAVEPOINT my_savepoint;
INSERT INTO table1 VALUES (4);
RELEASE SAVEPOINT my_savepoint;
COMMIT;
```

查询表的内容，会同时看到3和4。

```
SELECT * FROM table1;
```

```
gaussdb=> SELECT * FROM table1;
 a
---
 1
 3
 3
 4
(4 rows)
```

相关链接

RELEASE SAVEPOINT，**ROLLBACK TO SAVEPOINT**

 版权所有 © 华为云计算技术有限公司

17.9 SET TRANSACTION

功能描述

为当前事务设置特性。它对后面的事务没有影响。事务特性包括事务隔离级别、事务访问模式(读/写或者只读)。

注意事项

无。

语法格式

设置事务的隔离级别、读写模式。

```
{ SET [ LOCAL ] TRANSACTION|SET SESSION CHARACTERISTICS AS TRANSACTION }
  { ISOLATION LEVEL { READ COMMITTED | READ UNCOMMITTED | SERIALIZABLE | REPEATABLE READ }
  | { READ WRITE | READ ONLY } } [, ...]
```

参数说明

表 17-5 SET TRANSACTION 参数说明

参数	描述	取值范围
LOCAL	声明该命令只在当前事务中有效。	-
SESSION	声明该命令只对当前会话起作用。	字符串，要符合标识符的命名规范。
ISOLATION LEVEL	指定事务隔离级别，该参数决定当一个事务中存在其他并发运行事务时能够看到什么数据。 **说明** 在事务中第一个数据修改语句（INSERT，DELETE，UPDATE，FETCH，COPY）执行之后，事务隔离级别就不能再次设置。	• READ COMMITTED：读已提交隔离级别，只能读到已经提交的数据，而不会读到未提交的数据。缺省值为READ COMMITTED。 • READ UNCOMMITTED：读未提交隔离级别，DWS不支持READ UNCOMMITTED，如果设置了READ UNCOMMITTED，实际上使用的是READ COMMITTED。 • REPEATABLE READ：可重复读隔离级别，仅能看到事务开始之前提交的数据，不能看到未提交的数据以及在事务执行期间由其他并发事务提交的修改。 • SERIALIZABLE：事务可序列化，DWS不支持SERIALIZABLE，如果设置了SERIALIZABLE，实际上使用的是REPEATABLE READ。

版权所有 © 华为云计算技术有限公司

参数	描述	取值范围
READ WRITE | READ ONLY	指定事务访问模式。	读写模式或者只读模式。

示例

设置当前事务的隔离级别为READ COMMITTED，访问模式为READ ONLY：

```
START TRANSACTION;
SET LOCAL TRANSACTION ISOLATION LEVEL READ COMMITTED READ ONLY;
COMMIT;
```

17.10 START TRANSACTION

功能描述

通过START TRANSACTION启动事务。如果声明了隔离级别、读写模式，那么新事务就使用这些特性，类似执行了**SET TRANSACTION**。

注意事项

无。

语法格式

格式一：START TRANSACTION格式

```
START TRANSACTION
 [
  {
    ISOLATION LEVEL { READ COMMITTED | READ UNCOMMITTED | SERIALIZABLE | REPEATABLE READ }
    | { READ WRITE | READ ONLY }
  } [, ...]
 ];
```

格式二：BEGIN格式

```
BEGIN [ WORK | TRANSACTION ]
 [
  {
    ISOLATION LEVEL { READ COMMITTED | READ UNCOMMITTED | SERIALIZABLE | REPEATABLE READ }
    | { READ WRITE | READ ONLY }
  } [, ...]
 ];
```


版权所有 © 华为云计算技术有限公司

参数说明

表 17-6 START TRANSACTION 参数说明

参数	描述	取值范围
WORK | TRANSACTION	BEGIN格式中的可选关键字，对操作没有影响。	-
ISOLATION LEVEL	指定事务隔离级别，该参数决定当一个事务中存在其他并发运行事务时能够看到什么数据。 **说明** 在事务中第一个数据修改语句（INSERT，DELETE，UPDATE，FETCH，COPY）执行之后，事务隔离级别就不能再次设置。	• READ COMMITTED：读已提交隔离级别，只能读到已经提交的数据，而不会读到未提交的数据。缺省值为READ COMMITTED。 • READ UNCOMMITTED：读未提交隔离级别，DWS不支持READ UNCOMMITTED，如果设置了READ UNCOMMITTED，实际上使用的是READ COMMITTED。 • REPEATABLE READ：可重复读隔离级别，仅能看到事务开始之前提交的数据，不能看到未提交的数据以及在事务执行期间由其他并发事务提交的修改。 • SERIALIZABLE：事务可序列化，DWS不支持SERIALIZABLE，如果设置了SERIALIZABLE，实际上使用的是REPEATABLE READ。
READ WRITE | READ ONLY	指定事务访问模式。	读写模式或者只读模式。

示例

- 准备数据：

```
DROP SCHEMA IF EXISTS tpcds CASCADE;
CREATE SCHEMA tpcds;
CREATE TABLE tpcds.reason (a int, b int);
```

- 以默认方式启动事务：

```
START TRANSACTION;
SELECT * FROM tpcds.reason;
END;
```

- 以隔离级别为READ COMMITTED，读/写方式启动事务：

```
START TRANSACTION ISOLATION LEVEL READ COMMITTED READ WRITE;
SELECT * FROM tpcds.reason;
COMMIT;
```

相关链接

COMMIT | END，**ROLLBACK**，**SET TRANSACTION**

版权所有 © 华为云计算技术有限公司

17.11 ROLLBACK

功能描述

回滚当前事务并取消当前事务中的所有更新。

在事务运行的过程中发生了某种故障，事务不能继续执行，系统将事务中对数据库的所有已完成的操作全部撤销，数据库状态回到事务开始时。

注意事项

如果不在一个事务内部发出ROLLBACK不会有问题，但是将抛出一个警告信息。

语法格式

```
ROLLBACK [ WORK | TRANSACTION ];
```

参数说明

WORK | TRANSACTION

可选关键字。除了增加可读性，对ROLLBACK操作没有影响。

示例

取消当前事务中的所有更改：

```
ROLLBACK;
```

相关链接

COMMIT | END

17.12 RELEASE SAVEPOINT

功能描述

RELEASE SAVEPOINT删除当前事务先前定义的保存点。

保存点删除后便无法再作为回滚点使用，除此之外没有其他用户可见的行为。删除保存点并不能撤销在保存点建立起来之后执行的命令的影响。要撤销那些命令可以使用ROLLBACK TO SAVEPOINT 。在不再需要的时候删除保存点可以令系统在事务结束之前提前回收一些资源。

RELEASE SAVEPOINT也删除所有在指定的保存点建立之后的所有保存点。

注意事项

- 不能RELEASE一个没有定义的保存点，语法上会报错。
- 如果事务在回滚状态，则不能释放保存点。
- 如果多个保存点拥有同样的名字，只释放最近定义的保存点。

 版权所有 © 华为云计算技术有限公司

语法格式

```
RELEASE [ SAVEPOINT ] savepoint_name;
```

参数说明

savepoint_name

要删除的保存点的名字。

示例

准备数据：

```
DROP SCHEMA IF EXISTS tpcds CASCADE;
CREATE SCHEMA tpcds;
CREATE TABLE tpcds.table1 (a int);
```

建立并随后销毁一个保存点：

```
BEGIN;
    INSERT INTO tpcds.table1 VALUES (3);
    SAVEPOINT my_savepoint;
    INSERT INTO tpcds.table1 VALUES (4);
    RELEASE SAVEPOINT my_savepoint;
COMMIT;
```

相关链接

SAVEPOINT，**ROLLBACK TO SAVEPOINT**

17.13 ROLLBACK PREPARED

功能描述

在分布式数据库的两阶段提交（Two-Phase Commit, 2PC）过程中，事务的提交分为两个阶段：第一阶段执行**PREPARE TRANSACTION**将事务置为“预备”状态，第二阶段执行**COMMIT PREPARED**或ROLLBACK PREPARED完成提交或回滚。当预备事务因节点故障、网络异常或应用错误等原因未能正常进入第二阶段提交，而长时间停留在预备状态时，会持续持有锁和系统资源，影响其他事务的正常执行。此时，可使用ROLLBACK PREPARED命令取消该预备事务，释放其占用的资源。

典型业务场景包括：

- 分布式事务参与节点宕机后，预备事务无人终结，需要手动回滚清理。
- 应用层异常退出导致两阶段提交中断，预备事务残留需人工介入处理。
- 维护人员排查两阶段提交问题时，需要回滚特定预备事务以恢复系统正常状态。
- 两阶段提交第一阶段后，某个参与节点回复”无法提交”，协调者决定回滚整个事务。

注意事项

- 该功能仅在维护模式(GUC参数xc_maintenance_mode为on时)下可用。**该模式谨慎打开，一般供维护人员排查问题使用，一般用户不应使用该模式**。
- 要想回滚一个预备事务，必须是最初发起事务的用户，或者是系统管理员。

版权所有 © 华为云计算技术有限公司

- **事务功能由数据库自动维护，用户不应显式使用事务功能**。ROLLBACK PREPARED用于清理数据库内部创建的残留预备事务。在DWS分布式架构下，预备事务由内部分布式事务协调器（GTM）自动创建，不支持通过客户端手动执行PREPARE TRANSACTION。运维人员可通过查询pg_prepared_xacts视图查看残留的预备事务，并使用ROLLBACK PREPARED进行清理。

语法格式

```
ROLLBACK PREPARED transaction_id ;
```

参数说明

表 17-7 ROLLBACK PREPARED 参数说明

参数	描述	取值范围
transaction_id	待回滚的预备事务的标识符。该标识符为执行PREPARE TRANSACTION时指定的全局事务ID（GID），必须为当前已存在的预备事务的标识符。	字符串类型，最大长度为256个字符。不能与当前其他预备事务已使用的标识符冲突。示例值：'txn_transfer_002'。

示例

以下示例演示DWS运维中最常见的实际场景：查询数据库内部创建的残留预备事务，并通过ROLLBACK PREPARED命令回滚清理。

步骤1 开启维护模式（**开启后可能会造成数据不一致的风险**，需联系技术支持人员后开启）。

```
SET xc_maintenance_mode = on;
```

步骤2 查询当前所有残留的预备事务，获取需要回滚的事务标识符（gid字段值）。

```
SELECT transaction, gid, prepared, owner, database FROM pg_prepared_xacts;
```

预期输出（假设存在两个残留的预备事务）：

```
transaction |       gid        |        prepared         |  owner   | database
------------+------------------+-------------------------+----------+----------
    1234567 | txn_distribute_01| 2024-01-15 14:30:00+08 | app_user | db_test
    1234568 | txn_distribute_02| 2024-01-15 15:00:00+08 | app_user | db_test
(2 rows)
```

其中gid字段的值即为各预备事务的标识符，用于ROLLBACK PREPARED操作。

步骤3 回滚第一个残留的预备事务。

```
ROLLBACK PREPARED 'txn_distribute_01';
```

步骤4 验证该预备事务是否已成功回滚。

```
SELECT transaction, gid FROM pg_prepared_xacts WHERE gid = 'txn_distribute_01';
```

预期输出（查询返回空结果，表示已成功回滚）：

```
transaction | gid
------------+-----
(0 rows)
```

步骤5 关闭维护模式。

 版权所有 © 华为云计算技术有限公司

```
SET xc_maintenance_mode = off;
```

----结束

相关链接

COMMIT PREPARED，**PREPARE TRANSACTION**

常见问题

ROLLBACK PREPARED 执行报错“ERROR: prepared transaction with identifier xxx does not exist”如何处理?

A：此报错表示指定的transaction_id不存在于当前预备事务列表中。可能原因包括：

- 标识符拼写错误或大小写不匹配。
- 该预备事务已被其他会话提交或回滚。
- 该预备事务从未创建成功。

请通过查询pg_prepared_xacts视图确认当前存在的预备事务及其正确的gid值，注意标识符区分大小写，需使用单引号括起来。

ROLLBACK PREPARED 和 ROLLBACK 有什么区别?

A：ROLLBACK用于回滚当前会话中活跃的普通事务；ROLLBACK PREPARED用于回滚已经通过PREPARE TRANSACTION进入预备状态的事务。预备事务已脱离创建它的会话，必须通过ROLLBACK PREPARED（需维护模式）来回滚，普通的ROLLBACK命令无法影响预备事务。

ROLLBACK PREPARED 执行报错“Explicit prepare transaction is not supported”

A：未开启维护模式，请执行以下命令开启。

```
SET xc_maintenance_mode = on;
```

17.14 ROLLBACK TO SAVEPOINT

功能描述

ROLLBACK TO SAVEPOINT用于回滚到保存点，隐含地删除所有在该保存点之后建立的保存点。

回滚所有指定保存点建立之后执行的命令。保存点仍然有效，并且需要时可以再次回滚到该点。

注意事项

- 不能回滚到未定义的保存点，语法上会报错。
- 在保存点方面，游标有一些非事务性的行为。任何在保存点里打开的游标都会在回滚掉这个保存点之后关闭。如果一个前面打开了的游标在保存点里面，并且游标被一个FETCH命令影响，而这个保存点稍后回滚了，那么这个游标的位置仍然在FETCH让它指向的位置(也就是FETCH不会被回滚)。关闭一个游标的行为也不会被回滚给撤销掉。如果一个游标的操作导致事务回滚，那么这个游标就会置于

 版权所有 © 华为云计算技术有限公司

不可执行状态，所以，尽管一个事务可以用ROLLBACK TO SAVEPOINT重新恢复，但是游标不能再使用了。

- 使用ROLLBACK TO SAVEPOINT回滚到保存点。使用RELEASE SAVEPOINT删除保存点，但是保留该保存点建立后执行的命令的效果。

语法格式

```
ROLLBACK [ WORK | TRANSACTION ] TO [ SAVEPOINT ] savepoint_name;
```

参数说明

savepoint_name

回滚指定保存点的名称。

示例

撤销my_savepoint建立之后执行的命令（如下插入数据2的命令）的影响：

```
DROP SCHEMA IF EXISTS tpcds CASCADE;
CREATE SCHEMA tpcds;
CREATE TABLE tpcds.table1 (
    id integer
);

BEGIN;  -- 开启事务
    INSERT INTO tpcds.table1 VALUES (1);  -- 插入数据1
    SAVEPOINT my_savepoint;               -- 创建保存点
    INSERT INTO tpcds.table1 VALUES (2);  -- 插入数据2
    ROLLBACK TO SAVEPOINT my_savepoint;   -- 回滚到保存点（删除数据2）
    INSERT INTO tpcds.table1 VALUES (3);  -- 继续插入数据3
COMMIT;  -- 提交事务

SELECT * FROM tpcds.table1;  --验证数据发现，插入数据2的命令被撤销了。
```

```
gaussdb=> SELECT * FROM tpcds.table1;
 id
----
  3
  1
(2 rows)
```

游标位置不受保存点回滚的影响：

```
BEGIN;
DECLARE foo CURSOR FOR SELECT 1 UNION SELECT 2;
SAVEPOINT foo;
FETCH 1 FROM foo;
```

```
gaussdb=> FETCH 1 FROM foo;
 ?column?
----------
        1
(1 row)
```

```
ROLLBACK TO SAVEPOINT foo;
FETCH 1 FROM foo;
```

 版权所有 © 华为云计算技术有限公司

```
gaussdb=> FETCH 1 FROM foo;
 ?column?
----------
        2
(1 row)
```

```
COMMIT;
```

相关链接

SAVEPOINT，**RELEASE SAVEPOINT**

 版权所有 © 华为云计算技术有限公司