

数据仓库服务

# 开发指南

文档版本 04  
发布日期 2023-04-18



版权所有 © 华为云计算技术有限公司 2023。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



HUAWEI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

---

<b>1 欢迎</b>	<b>1</b>
1.1 文档面向的读者对象	1
1.2 阅读指引	2
1.3 文档表达约定	3
1.4 前置条件	3
<b>2 系统概述</b>	<b>5</b>
2.1 高可靠事务处理	5
2.2 查询高性能	5
2.3 相关概念	6
<b>3 定义数据库对象</b>	<b>8</b>
3.1 创建和管理数据库	8
3.2 创建和管理 Schema	9
3.3 创建和管理表	12
3.4 选择表存储模型	16
3.5 表分区定义	18
3.6 创建和管理索引	21
3.7 创建和使用序列	23
3.8 创建和管理视图	25
3.9 创建和管理定时任务	26
3.10 查看系统表	28
<b>4 数据迁移</b>	<b>31</b>
4.1 迁移数据到 GaussDB(DWS)	31
4.2 导入数据	34
4.2.1 使用外表从 OBS 并行导入数据	34
4.2.1.1 关于 OBS 并行导入	35
4.2.1.2 从 OBS 导入 CSV、TEXT 数据（方式一）	39
4.2.1.2.1 创建访问密钥（AK 和 SK）	39
4.2.1.2.2 上传数据到 OBS	41
4.2.1.2.3 创建 OBS 外表	42
4.2.1.2.4 执行导入数据	45
4.2.1.2.5 处理错误表	46
4.2.1.3 从 OBS 导入 ORC，PARQUET，CARBONDATA 等格式数据（方式二）	48

4.2.1.3.1 OBS 上的数据准备.....	48
4.2.1.3.2 创建外部服务器.....	49
4.2.1.3.3 创建外表.....	52
4.2.1.3.4 通过外表查询 OBS 上的数据.....	54
4.2.1.3.5 清除资源.....	55
4.2.1.3.6 支持的数据类型.....	57
4.2.2 使用 GDS 从远端服务器导入数据.....	60
4.2.2.1 关于 GDS 并行导入.....	60
4.2.2.2 准备源数据.....	64
4.2.2.3 安装配置和启动 GDS.....	64
4.2.2.4 创建 GDS 外表.....	69
4.2.2.5 执行导入数据.....	71
4.2.2.6 处理错误表.....	72
4.2.2.7 停止 GDS.....	75
4.2.2.8 GDS 导入示例.....	75
4.2.3 从 MRS 导入数据到集群.....	81
4.2.3.1 从 MRS 导入数据概述.....	81
4.2.3.2 MRS 集群上的数据准备.....	82
4.2.3.3 手动创建外部服务器.....	85
4.2.3.4 创建外表.....	89
4.2.3.5 执行数据导入.....	93
4.2.3.6 清除资源.....	94
4.2.3.7 错误处理.....	96
4.2.4 从 GaussDB(DWS)集群导入数据到新集群.....	96
4.2.5 基于 GDS 的跨集群互联互通.....	98
4.2.6 使用 gsql 元命令导入数据.....	101
4.2.7 使用 COPY FROM STDIN 导入数据.....	104
4.2.7.1 关于 COPY FROM STDIN 导入数据.....	104
4.2.7.2 CopyManager 类简介.....	104
4.2.7.3 示例：通过本地文件导入导出数据.....	105
4.2.7.4 示例：从 MySQL 向 GaussDB(DWS)进行数据迁移.....	107
4.3 整库迁移.....	108
4.3.1 使用 DRS 将数据导入 GaussDB(DWS).....	108
4.3.2 使用 CDM 迁移数据到 GaussDB(DWS).....	109
4.3.3 使用 DSC 工具迁移 SQL 脚本.....	109
4.4 元数据迁移.....	110
4.4.1 使用 gs_dump 和 gs_dumpall 命令导出元数据.....	110
4.4.1.1 概述.....	110
4.4.1.2 导出单个数据库.....	112
4.4.1.2.1 导出数据库.....	112
4.4.1.2.2 导出模式.....	114
4.4.1.2.3 导出表.....	117

4.4.1.3 导出所有数据库.....	120
4.4.1.3.1 导出所有数据库.....	120
4.4.1.3.2 导出全局对象.....	122
4.4.1.4 无权限角色导出数据.....	124
4.4.2 使用 gs_restore 命令导入数据.....	126
4.5 导出数据.....	131
4.5.1 使用外表导出数据到 OBS.....	131
4.5.1.1 关于 OBS 并行导出.....	131
4.5.1.2 导出 CSV、TEXT 数据到 OBS（方式一）.....	135
4.5.1.2.1 规划导出数据.....	135
4.5.1.2.2 创建 OBS 外表.....	136
4.5.1.2.3 执行导出.....	138
4.5.1.2.4 示例.....	139
4.5.1.3 导出 ORC 数据到 OBS（方式二）.....	142
4.5.1.3.1 规划导出数据.....	142
4.5.1.3.2 创建外部服务器.....	142
4.5.1.3.3 创建外表.....	142
4.5.1.3.4 执行导出.....	144
4.5.2 导出 ORC 数据到 MRS.....	145
4.5.2.1 导出 ORC 数据概述.....	145
4.5.2.2 规划导出数据.....	145
4.5.2.3 创建外部服务器.....	145
4.5.2.4 创建外表.....	145
4.5.2.5 执行导出.....	147
4.5.3 使用 GDS 导出数据到远端服务器.....	148
4.5.3.1 关于 GDS 并行导出.....	148
4.5.3.2 规划导出数据.....	151
4.5.3.3 安装配置和启动 GDS.....	151
4.5.3.4 创建 GDS 外表.....	152
4.5.3.5 执行导出数据.....	153
4.5.3.6 停止 GDS.....	153
4.5.3.7 GDS 导出示例.....	154
4.6 其他操作.....	158
4.6.1 GDS 管道文件常见问题.....	158
4.6.2 查看数据倾斜状态.....	159
4.6.3 分析表.....	162
<b>5 冷热数据管理.....</b>	<b>165</b>
<b>6 Oracle、Teradata 和 MySQL 语法兼容性差异.....</b>	<b>168</b>
<b>7 管理数据库安全.....</b>	<b>172</b>
7.1 管理用户及权限.....	172
7.1.1 数据库用户.....	172

7.1.2 用户管理.....	173
7.1.3 自定义密码策略.....	174
7.1.4 权限管理.....	179
7.1.5 三权分立.....	182
7.2 敏感数据管理.....	184
7.2.1 行级访问控制.....	184
7.2.2 数据脱敏.....	186
7.2.3 使用函数加解密.....	189
7.2.4 使用 pgcrypto 加密数据.....	192
<b>8 开发设计建议.....</b>	<b>201</b>
8.1 开发设计建议概述.....	201
8.2 数据库对象命名.....	201
8.3 数据库对象设计.....	202
8.3.1 Database 和 Schema 设计.....	202
8.3.2 表设计.....	203
8.3.3 字段设计.....	205
8.3.4 约束设计.....	206
8.3.5 视图和关联表设计.....	207
8.4 JDBC 配置.....	207
8.5 SQL 编写.....	208
8.6 自定义外部函数(pgSQL/Java)使用.....	211
8.7 PL/pgSQL 使用.....	212
<b>9 教程：使用 JDBC 或 ODBC 开发.....</b>	<b>215</b>
9.1 开发规范.....	215
9.2 驱动下载.....	215
9.3 基于 JDBC 开发.....	215
9.3.1 JDBC 包与驱动类.....	215
9.3.2 开发流程.....	216
9.3.3 加载驱动.....	216
9.3.4 连接数据库.....	217
9.3.5 执行 SQL 语句.....	220
9.3.6 处理结果集.....	222
9.3.7 关闭连接.....	224
9.3.8 示例：常用操作.....	224
9.3.9 示例：重新执行应用 SQL.....	228
9.3.10 示例：通过本地文件导入导出数据.....	231
9.3.11 示例：从 MySQL 向 GaussDB(DWS)进行数据迁移.....	232
9.3.12 JDBC 接口参考.....	234
9.3.12.1 java.sql.Connection.....	234
9.3.12.2 java.sql.CallableStatement.....	235
9.3.12.3 java.sql.DatabaseMetaData.....	236
9.3.12.4 java.sql.Driver.....	238

9.3.12.5 java.sql.PreparedStatement.....	239
9.3.12.6 java.sql.ResultSet.....	240
9.3.12.7 java.sql.ResultSetMetaData.....	242
9.3.12.8 java.sql.Statement.....	242
9.3.12.9 javax.sql.ConnectionPoolDataSource.....	243
9.3.12.10 javax.sql.DataSource.....	244
9.3.12.11 javax.sql.PooledConnection.....	244
9.3.12.12 javax.naming.Context.....	245
9.3.12.13 javax.naming.spi.InitialContextFactory.....	245
9.3.12.14 CopyManager.....	245
9.4 基于 ODBC 开发.....	247
9.4.1 ODBC 包及依赖的库和头文件.....	248
9.4.2 Linux 下配置数据源.....	248
9.4.3 Windows 下配置数据源.....	254
9.4.4 ODBC 开发示例.....	258
9.4.5 ODBC 接口参考.....	263
9.4.5.1 SQLAllocEnv.....	264
9.4.5.2 SQLAllocConnect.....	264
9.4.5.3 SQLAllocHandle.....	264
9.4.5.4 SQLAllocStmt.....	265
9.4.5.5 SQLBindCol.....	265
9.4.5.6 SQLBindParameter.....	266
9.4.5.7 SQLColAttribute.....	267
9.4.5.8 SQLConnect.....	269
9.4.5.9 SQLDisconnect.....	270
9.4.5.10 SQLExecDirect.....	270
9.4.5.11 SQLExecute.....	271
9.4.5.12 SQLFetch.....	272
9.4.5.13 SQLFreeStmt.....	273
9.4.5.14 SQLFreeConnect.....	273
9.4.5.15 SQLFreeHandle.....	273
9.4.5.16 SQLFreeEnv.....	274
9.4.5.17 SQLPrepare.....	274
9.4.5.18 SQLGetData.....	275
9.4.5.19 SQLGetDiagRec.....	276
9.4.5.20 SQLSetConnectAttr.....	278
9.4.5.21 SQLSetEnvAttr.....	279
9.4.5.22 SQLSetStmtAttr.....	280
<b>10 PostGIS Extension.....</b>	<b>282</b>
10.1 PostGIS 概述.....	282
10.2 PostGIS 使用.....	282
10.3 PostGIS 支持和限制.....	283

10.4 OPEN SOURCE SOFTWARE NOTICE (For PostGIS).....	287
<b>11 实时数仓.....</b>	<b>335</b>
11.1 实时数仓简介.....	335
11.2 支持与限制.....	338
11.3 实时数仓语法.....	339
11.3.1 CREATE TABLE.....	339
11.3.2 INSERT.....	342
11.3.3 DELETE.....	343
11.3.4 UPDATE.....	345
11.3.5 UPSERT.....	346
11.3.6 MERGE INTO.....	347
11.3.7 SELECT.....	349
11.3.8 ALTER TABLE.....	350
11.4 实时数仓函数.....	351
11.5 实时数仓 GUC 参数.....	352
<b>12 资源监控.....</b>	<b>354</b>
12.1 用户资源监控.....	354
12.2 资源池资源监控.....	356
12.3 内存资源监控.....	358
12.4 实例资源监控.....	359
12.5 实时 TopSQL.....	361
12.6 历史 TopSQL.....	363
12.7 TopSQL 查询示例.....	366
<b>13 用户自定义函数.....</b>	<b>369</b>
13.1 PL/Java 语言函数.....	369
13.2 PL/pgSQL 语言函数.....	378
<b>14 存储过程.....</b>	<b>380</b>
14.1 存储过程.....	380
14.2 数据类型.....	380
14.3 数据类型转换.....	380
14.4 数组和 record.....	381
14.4.1 数组.....	381
14.4.2 record.....	387
14.5 声明语法.....	390
14.5.1 基本结构.....	390
14.5.2 匿名块.....	390
14.5.3 子程序.....	391
14.6 基本语句.....	391
14.6.1 定义变量.....	392
14.6.2 赋值语句.....	393
14.6.3 调用语句.....	394



14.7 动态语句.....	395
14.7.1 执行动态查询语句.....	395
14.7.2 执行动态非查询语句.....	397
14.7.3 动态调用存储过程.....	398
14.7.4 动态调用匿名块.....	400
14.8 控制语句.....	401
14.8.1 返回语句.....	401
14.8.1.1 RETURN.....	401
14.8.1.2 RETURN NEXT 及 RETURN QUERY.....	402
14.8.2 条件语句.....	403
14.8.3 循环语句.....	404
14.8.4 分支语句.....	408
14.8.5 空语句.....	409
14.8.6 错误捕获语句.....	409
14.8.7 GOTO 语句.....	411
14.9 其他语句.....	413
14.9.1 锁操作.....	413
14.9.2 游标操作.....	413
14.10 游标.....	413
14.10.1 游标概述.....	413
14.10.2 显式游标.....	414
14.10.3 隐式游标.....	417
14.10.4 游标循环.....	418
14.11 高级包.....	419
14.11.1 DBMS_LOB.....	420
14.11.2 DBMS_RANDOM.....	428
14.11.3 DBMS_OUTPUT.....	429
14.11.4 UTL_RAW.....	430
14.11.5 DBMS_JOB.....	433
14.11.6 DBMS_SQL.....	439
14.12 调试.....	448
<b>15 系统表和系统视图.....</b>	<b>452</b>
15.1 系统表和系统视图概述.....	452
15.2 系统表.....	452
15.2.1 GS_BLOCKLIST_QUERY.....	452
15.2.2 GS_OBSCANINFO.....	453
15.2.3 GS_RESPOOL_RESOURCE_HISTORY.....	454
15.2.4 GS_WLM_INSTANCE_HISTORY.....	455
15.2.5 GS_WLM_OPERATOR_INFO.....	456
15.2.6 GS_WLM_SESSION_INFO.....	458
15.2.7 GS_WLM_USER_RESOURCE_HISTORY.....	458
15.2.8 PG_AGGREGATE.....	460

15.2.9 PG_AM.....	460
15.2.10 PG_AMOP.....	462
15.2.11 PG_AMPROC.....	463
15.2.12 PG_ATTRDEF.....	463
15.2.13 PG_ATTRIBUTE.....	464
15.2.14 PG_AUTHID.....	465
15.2.15 PG_AUTH_HISTORY.....	467
15.2.16 PG_AUTH_MEMBERS.....	467
15.2.17 PG_CAST.....	468
15.2.18 PG_CLASS.....	468
15.2.19 PG_COLLATION.....	472
15.2.20 PG_CONSTRAINT.....	473
15.2.21 PG_CONVERSION.....	475
15.2.22 PG_DATABASE.....	475
15.2.23 PG_DB_ROLE_SETTING.....	476
15.2.24 PG_DEFAULT_ACL.....	476
15.2.25 PG_DEPEND.....	477
15.2.26 PG_DESCRIPTION.....	479
15.2.27 PG_ENUM.....	479
15.2.28 PG_EXCEPT_RULE.....	480
15.2.29 PG_EXTENSION.....	480
15.2.30 PG_EXTENSION_DATA_SOURCE.....	481
15.2.31 PG_FINE_DR_INFO.....	481
15.2.32 PG_FOREIGN_DATA_WRAPPER.....	482
15.2.33 PG_FOREIGN_SERVER.....	482
15.2.34 PG_FOREIGN_TABLE.....	483
15.2.35 PG_INDEX.....	483
15.2.36 PG_INHERITS.....	485
15.2.37 PG_JOBS.....	485
15.2.38 PG_LANGUAGE.....	486
15.2.39 PG_LARGEOBJECT.....	487
15.2.40 PG_LARGEOBJECT_METADATA.....	488
15.2.41 PG_STAT_LAST_OPERATION.....	488
15.2.42 PG_NAMESPACE.....	489
15.2.43 PG_OBJECT.....	489
15.2.44 PG_OBSSCANINFO.....	490
15.2.45 PG_OPCLASS.....	491
15.2.46 PG_OPERATOR.....	492
15.2.47 PG_OPFAMILY.....	492
15.2.48 PG_PARTITION.....	493
15.2.49 PG_PLTEMPLATE.....	495
15.2.50 PG_PROC.....	495

15.2.51 PG_PUBLICATION.....	498
15.2.52 PG_PUBLICATION_NAMESPACE.....	499
15.2.53 PG_PUBLICATION_REL.....	499
15.2.54 PG_RANGE.....	500
15.2.55 PG_REDACTION_COLUMN.....	500
15.2.56 PG_REDACTION_POLICY.....	501
15.2.57 PG_RELFILENODE_SIZE.....	502
15.2.58 PG_RLSPOLICY.....	502
15.2.59 PG_RESOURCE_POOL.....	503
15.2.60 PG_REWRITE.....	504
15.2.61 PG_SECLABEL.....	504
15.2.62 PG_SHDEPEND.....	505
15.2.63 PG_SHDESCRIPTION.....	506
15.2.64 PG_SHSECLABEL.....	506
15.2.65 PG_STATISTIC.....	506
15.2.66 PG_STATISTIC_EXT.....	508
15.2.67 PG_SUBSCRIPTION.....	509
15.2.68 PG_SYNONYM.....	510
15.2.69 PG_TABLESPACE.....	510
15.2.70 PG_TRIGGER.....	510
15.2.71 PG_TS_CONFIG.....	511
15.2.72 PG_TS_CONFIG_MAP.....	512
15.2.73 PG_TS_DICT.....	512
15.2.74 PG_TS_PARSER.....	512
15.2.75 PG_TS_TEMPLATE.....	513
15.2.76 PG_TYPE.....	514
15.2.77 PG_USER_MAPPING.....	516
15.2.78 PG_USER_STATUS.....	517
15.2.79 PG_WORKLOAD_ACTION.....	517
15.2.80 PGXC_CLASS.....	518
15.2.81 PGXC_GROUP.....	518
15.2.82 PGXC_NODE.....	519
15.2.83 SNAPSHOT.....	520
15.2.84 TABLES_SNAP_TIMESTAMP.....	521
15.2.85 性能视图快照系统表.....	521
15.3 系统视图.....	522
15.3.1 ALL_ALL_TABLES.....	522
15.3.2 ALL_CONSTRAINTS.....	523
15.3.3 ALL_CONS_COLUMNS.....	523
15.3.4 ALL_COL_COMMENTS.....	524
15.3.5 ALL_DEPENDENCIES.....	524
15.3.6 ALL_IND_COLUMNS.....	525

15.3.7 ALL_IND_EXPRESSIONS.....	525
15.3.8 ALL_INDEXES.....	525
15.3.9 ALL_OBJECTS.....	526
15.3.10 ALL_PROCEDURES.....	526
15.3.11 ALL_SEQUENCES.....	527
15.3.12 ALL_SOURCE.....	527
15.3.13 ALL_SYNONYMS.....	528
15.3.14 ALL_TAB_COLUMNS.....	528
15.3.15 ALL_TAB_COMMENTS.....	529
15.3.16 ALL_TABLES.....	529
15.3.17 ALL_USERS.....	530
15.3.18 ALL_VIEWS.....	530
15.3.19 DBA_DATA_FILES.....	530
15.3.20 DBA_USERS.....	531
15.3.21 DBA_COL_COMMENTS.....	531
15.3.22 DBA_CONSTRAINTS.....	531
15.3.23 DBA_CONS_COLUMNS.....	532
15.3.24 DBA_IND_COLUMNS.....	532
15.3.25 DBA_IND_EXPRESSIONS.....	533
15.3.26 DBA_IND_PARTITIONS.....	533
15.3.27 DBA_INDEXES.....	534
15.3.28 DBA_OBJECTS.....	534
15.3.29 DBA_PART_INDEXES.....	535
15.3.30 DBA_PART_TABLES.....	536
15.3.31 DBA_PROCEDURES.....	536
15.3.32 DBA_SEQUENCES.....	536
15.3.33 DBA_SOURCE.....	537
15.3.34 DBA_SYNONYMS.....	537
15.3.35 DBA_TAB_COLUMNS.....	537
15.3.36 DBA_TAB_COMMENTS.....	538
15.3.37 DBA_TAB_PARTITIONS.....	539
15.3.38 DBA_TABLES.....	539
15.3.39 DBA_TABLESPACES.....	540
15.3.40 DBA_TRIGGERS.....	540
15.3.41 DBA_VIEWS.....	540
15.3.42 DUAL.....	540
15.3.43 GLOBAL_COLUMN_TABLE_IO_STAT.....	541
15.3.44 GLOBAL_REDO_STAT.....	541
15.3.45 GLOBAL_REL_IOSTAT.....	541
15.3.46 GLOBAL_ROW_TABLE_IO_STAT.....	541
15.3.47 GLOBAL_STAT_DATABASE.....	541
15.3.48 GLOBAL_TABLE_CHANGE_STAT.....	543

15.3.49 GLOBAL_TABLE_STAT.....	544
15.3.50 GLOBAL_WORKLOAD_SQL_COUNT.....	544
15.3.51 GLOBAL_WORKLOAD_SQL_ELAPSE_TIME.....	545
15.3.52 GLOBAL_WORKLOAD_TRANSACTION.....	546
15.3.53 GS_ALL_CONTROL_GROUP_INFO.....	546
15.3.54 GS_BLOCKLIST_QUERY.....	547
15.3.55 GS_CLUSTER_RESOURCE_INFO.....	547
15.3.56 GS_COLUMN_TABLE_IO_STAT.....	548
15.3.57 GS_GET_OBS_READ_TRAFFIC.....	548
15.3.58 GS_GET_OBS_WRITE_TRAFFIC.....	549
15.3.59 GS_INSTR_UNIQUE_SQL.....	550
15.3.60 GS_NODE_STAT_RESET_TIME.....	553
15.3.61 GS_OBS_LATENCY.....	553
15.3.62 GS_REL_IOSTAT.....	553
15.3.63 GS_RESPOOL_RUNTIME_INFO.....	553
15.3.64 GS_RESPOOL_RESOURCE_INFO.....	554
15.3.65 GS_ROW_TABLE_IO_STAT.....	556
15.3.66 GS_SESSION_CPU_STATISTICS.....	556
15.3.67 GS_SESSION_MEMORY_STATISTICS.....	557
15.3.68 GS_SQL_COUNT.....	558
15.3.69 GS_STAT_DB_CU.....	560
15.3.70 GS_STAT_SESSION_CU.....	560
15.3.71 GS_TABLE_CHANGE_STAT.....	560
15.3.72 GS_TABLE_STAT.....	561
15.3.73 GS_TOTAL_NODEGROUP_MEMORY_DETAIL.....	562
15.3.74 GS_USER_TRANSACTION.....	563
15.3.75 GS_VIEW_DEPENDENCY.....	563
15.3.76 GS_VIEW_DEPENDENCY_PATH.....	563
15.3.77 GS_VIEW_INVALID.....	564
15.3.78 GS_WAIT_EVENTS.....	564
15.3.79 GS_WLM_OPERATOR_INFO.....	566
15.3.80 GS_WLM_OPERATOR_HISTORY.....	566
15.3.81 GS_WLM_OPERATOR_STATISTICS.....	566
15.3.82 GS_WLM_SESSION_INFO.....	567
15.3.83 GS_WLM_SESSION_HISTORY.....	567
15.3.84 GS_WLM_SESSION_STATISTICS.....	571
15.3.85 GS_WLM_SQL_ALLOW.....	574
15.3.86 GS_WORKLOAD_SQL_COUNT.....	574
15.3.87 GS_WORKLOAD_SQL_ELAPSE_TIME.....	575
15.3.88 GS_WORKLOAD_TRANSACTION.....	575
15.3.89 MPP_TABLES.....	576
15.3.90 PG_AVAILABLE_EXTENSION_VERSIONS.....	576

15.3.91 PG_AVAILABLE_EXTENSIONS.....	577
15.3.92 PG_BULKLOAD_STATISTICS.....	577
15.3.93 PG_COMM_CLIENT_INFO.....	578
15.3.94 PG_COMM_DELAY.....	578
15.3.95 PG_COMM_STATUS.....	579
15.3.96 PG_COMM_RECV_STREAM.....	580
15.3.97 PG_COMM_SEND_STREAM.....	580
15.3.98 PG_COMM_QUERY_SPEED.....	581
15.3.99 PG_CONTROL_GROUP_CONFIG.....	582
15.3.100 PG_CURSORS.....	582
15.3.101 PG_EXT_STATS.....	583
15.3.102 PG_GET_INVALID_BACKENDS.....	584
15.3.103 PG_GET_SENDERS_CATCHUP_TIME.....	585
15.3.104 PG_GROUP.....	585
15.3.105 PG_INDEXES.....	586
15.3.106 PG_JOB.....	586
15.3.107 PG_JOB_PROC.....	588
15.3.108 PG_JOB_SINGLE.....	588
15.3.109 PG_LIFECYCLE_DATA_DISTRIBUTE.....	590
15.3.110 PG_LOCKS.....	590
15.3.111 PG_NODE_ENV.....	591
15.3.112 PG_OS_THREADS.....	592
15.3.113 PG_POOLER_STATUS.....	592
15.3.114 PG_PREPARED_STATEMENTS.....	593
15.3.115 PG_PREPARED_XACTS.....	593
15.3.116 PG_PUBLICATION_TABLES.....	594
15.3.117 PG_QUERYBAND_ACTION.....	594
15.3.118 PG_REPLICATION_SLOTS.....	595
15.3.119 PG_ROLES.....	595
15.3.120 PG_RULES.....	597
15.3.121 PG_RUNNING_XACTS.....	597
15.3.122 PG_SECLABELS.....	598
15.3.123 PG_SESSION_WLMSTAT.....	598
15.3.124 PG_SESSION_IOSTAT.....	600
15.3.125 PG_SETTINGS.....	600
15.3.126 PG_SHADOW.....	601
15.3.127 PG_SHARED_MEMORY_DETAIL.....	602
15.3.128 PG_STATS.....	603
15.3.129 PG_STAT_ACTIVITY.....	604
15.3.130 PG_STAT_ALL_INDEXES.....	607
15.3.131 PG_STAT_ALL_TABLES.....	607
15.3.132 PG_STAT_BAD_BLOCK.....	609

15.3.133 PG_STAT_BGWRITER.....	609
15.3.134 PG_STAT_DATABASE.....	610
15.3.135 PG_STAT_DATABASE_CONFLICTS.....	611
15.3.136 PG_STAT_GET_MEM_MBYTES_RESERVED.....	612
15.3.137 PG_STAT_USER_FUNCTIONS.....	612
15.3.138 PG_STAT_USER_INDEXES.....	613
15.3.139 PG_STAT_USER_TABLES.....	613
15.3.140 PG_STAT_REPLICATION.....	614
15.3.141 PG_STAT_SYS_INDEXES.....	615
15.3.142 PG_STAT_SYS_TABLES.....	616
15.3.143 PG_STAT_XACT_ALL_TABLES.....	617
15.3.144 PG_STAT_XACT_SYS_TABLES.....	618
15.3.145 PG_STAT_XACT_USER_FUNCTIONS.....	618
15.3.146 PG_STAT_XACT_USER_TABLES.....	619
15.3.147 PG_STATIO_ALL_INDEXES.....	619
15.3.148 PG_STATIO_ALL_SEQUENCES.....	620
15.3.149 PG_STATIO_ALL_TABLES.....	620
15.3.150 PG_STATIO_SYS_INDEXES.....	621
15.3.151 PG_STATIO_SYS_SEQUENCES.....	621
15.3.152 PG_STATIO_SYS_TABLES.....	622
15.3.153 PG_STATIO_USER_INDEXES.....	622
15.3.154 PG_STATIO_USER_SEQUENCES.....	623
15.3.155 PG_STATIO_USER_TABLES.....	623
15.3.156 PG_THREAD_WAIT_STATUS.....	624
15.3.157 PG_TABLES.....	633
15.3.158 PG_TDE_INFO.....	634
15.3.159 PG_TIMEZONE_ABBREVS.....	635
15.3.160 PG_TIMEZONE_NAMES.....	635
15.3.161 PG_TOTAL_MEMORY_DETAIL.....	635
15.3.162 PG_TOTAL_SCHEMA_INFO.....	638
15.3.163 PG_TOTAL_USER_RESOURCE_INFO.....	638
15.3.164 PG_USER.....	640
15.3.165 PG_USER_MAPPINGS.....	641
15.3.166 PG_VIEWS.....	641
15.3.167 PG_WLM_STATISTICS.....	642
15.3.168 PGXC_BULKLOAD_PROGRESS.....	643
15.3.169 PGXC_BULKLOAD_STATISTICS.....	643
15.3.170 PGXC_COLUMN_TABLE_IO_STAT.....	644
15.3.171 PGXC_COMM_CLIENT_INFO.....	644
15.3.172 PGXC_COMM_DELAY.....	645
15.3.173 PGXC_COMM_RECV_STREAM.....	645
15.3.174 PGXC_COMM_SEND_STREAM.....	646

15.3.175 PGXC_COMM_STATUS.....	647
15.3.176 PGXC_COMM_QUERY_SPEED.....	648
15.3.177 PGXC_DEADLOCK.....	648
15.3.178 PGXC_GET_STAT_ALL_TABLES.....	650
15.3.179 PGXC_GET_STAT_ALL_PARTITIONS.....	651
15.3.180 PGXC_GET_TABLE_SKEWNESS.....	652
15.3.181 PGXC_GTM_SNAPSHOT_STATUS.....	652
15.3.182 PGXC_INSTANCE_TIME.....	653
15.3.183 PGXC_LOCKWAIT_DETAIL.....	653
15.3.184 PGXC_INSTR_UNIQUE_SQL.....	655
15.3.185 PGXC_LOCK_CONFLICTS.....	655
15.3.186 PGXC_NODE_ENV.....	656
15.3.187 PGXC_NODE_STAT_RESET_TIME.....	656
15.3.188 PGXC_OS_RUN_INFO.....	656
15.3.189 PGXC_OS_THREADS.....	657
15.3.190 PGXC_PREPARED_XACTS.....	657
15.3.191 PGXC_REDO_STAT.....	657
15.3.192 PGXC_REL_IOSTAT.....	657
15.3.193 PGXC_REPLICATION_SLOTS.....	657
15.3.194 PGXC_RESPOOL_RUNTIME_INFO.....	658
15.3.195 PGXC_RESPOOL_RESOURCE_INFO.....	658
15.3.196 PGXC_RESPOOL_RESOURCE_HISTORY.....	660
15.3.197 PGXC_ROW_TABLE_IO_STAT.....	662
15.3.198 PGXC_RUNNING_XACTS.....	662
15.3.199 PGXC_SETTINGS.....	663
15.3.200 PGXC_SESSION_WLMSTAT.....	663
15.3.201 PGXC_STAT_ACTIVITY.....	664
15.3.202 PGXC_STAT_BAD_BLOCK.....	667
15.3.203 PGXC_STAT_BGWRITER.....	668
15.3.204 PGXC_STAT_DATABASE.....	668
15.3.205 PGXC_STAT_REPLICATION.....	668
15.3.206 PGXC_STAT_TABLE_DIRTY.....	668
15.3.207 PGXC_STAT_WAL.....	671
15.3.208 PGXC_SQL_COUNT.....	672
15.3.209 PGXC_TABLE_CHANGE_STAT.....	672
15.3.210 PGXC_TABLE_STAT.....	673
15.3.211 PGXC_THREAD_WAIT_STATUS.....	673
15.3.212 PGXC_TOTAL_MEMORY_DETAIL.....	674
15.3.213 PGXC_TOTAL_SCHEMA_INFO.....	676
15.3.214 PGXC_TOTAL_SCHEMA_INFO_ANALYZE.....	676
15.3.215 PGXC_TOTAL_USER_RESOURCE_INFO.....	677
15.3.216 PGXC_USER_TRANSACTION.....	679



15.3.217 PGXC_VARIABLE_INFO.....	680
15.3.218 PGXC_WAIT_DETAIL.....	680
15.3.219 PGXC_WAIT_EVENTS.....	682
15.3.220 PGXC_WLM_OPERATOR_HISTORY.....	683
15.3.221 PGXC_WLM_OPERATOR_INFO.....	683
15.3.222 PGXC_WLM_OPERATOR_STATISTICS.....	683
15.3.223 PGXC_WLM_SESSION_INFO.....	683
15.3.224 PGXC_WLM_SESSION_HISTORY.....	683
15.3.225 PGXC_WLM_SESSION_STATISTICS.....	683
15.3.226 PGXC_WLM_USER_RESOURCE_HISTORY.....	683
15.3.227 PGXC_WLM_WORKLOAD_RECORDS.....	686
15.3.228 PGXC_WORKLOAD_SQL_COUNT.....	687
15.3.229 PGXC_WORKLOAD_SQL_ELAPSE_TIME.....	687
15.3.230 PGXC_WORKLOAD_TRANSACTION.....	688
15.3.231 PLAN_TABLE.....	689
15.3.232 PLAN_TABLE_DATA.....	690
15.3.233 PV_FILE_STAT.....	690
15.3.234 PV_INSTANCE_TIME.....	691
15.3.235 PV_OS_RUN_INFO.....	692
15.3.236 PV_SESSION_MEMORY.....	692
15.3.237 PV_SESSION_MEMORY_DETAIL.....	692
15.3.238 PV_SESSION_STAT.....	694
15.3.239 PV_SESSION_TIME.....	694
15.3.240 PV_TOTAL_MEMORY_DETAIL.....	695
15.3.241 PV_REDO_STAT.....	697
15.3.242 PV_RUNTIME_ATTSTATS.....	697
15.3.243 PV_RUNTIME_RELSTATS.....	698
15.3.244 REDACTION_COLUMNS.....	698
15.3.245 REDACTION_POLICIES.....	699
15.3.246 REMOTE_TABLE_STAT.....	700
15.3.247 USER_COL_COMMENTS.....	700
15.3.248 USER_CONSTRAINTS.....	700
15.3.249 USER_CONS_COLUMNS.....	700
15.3.250 USER_INDEXES.....	701
15.3.251 USER_IND_COLUMNS.....	701
15.3.252 USER_IND_EXPRESSIONS.....	702
15.3.253 USER_IND_PARTITIONS.....	702
15.3.254 USER_JOBS.....	703
15.3.255 USER_OBJECTS.....	704
15.3.256 USER_PART_INDEXES.....	704
15.3.257 USER_PART_TABLES.....	705
15.3.258 USER_PROCEDURES.....	706

15.3.259 USER_SEQUENCES.....	706
15.3.260 USER_SOURCE.....	706
15.3.261 USER_SYNONYMS.....	706
15.3.262 USER_TAB_COLUMNS.....	707
15.3.263 USER_TAB_COMMENTS.....	708
15.3.264 USER_TAB_PARTITIONS.....	708
15.3.265 USER_TABLES.....	709
15.3.266 USER_TRIGGERS.....	709
15.3.267 USER_VIEWS.....	709
15.3.268 V\$SESSION.....	710
15.3.269 V\$SESSION_LONGOPS.....	710
<b>16 排序规则支持.....</b>	<b>711</b>
<b>17 GUC 参数.....</b>	<b>714</b>
17.1 查看 GUC 参数.....	714
17.2 设置 GUC 参数.....	715
17.3 GUC 使用说明.....	716
17.4 连接和认证.....	717
17.4.1 连接设置.....	717
17.4.2 安全和认证 ( postgresql.conf ) .....	718
17.4.3 通信库参数.....	724
17.5 资源消耗.....	730
17.5.1 内存.....	730
17.5.2 语句磁盘空间管控.....	738
17.5.3 内核资源使用.....	739
17.5.4 基于开销的清理延迟.....	739
17.5.5 异步 IO.....	741
17.6 并行导入.....	743
17.7 预写式日志.....	744
17.7.1 设置.....	744
17.7.2 检查点.....	748
17.7.3 归档.....	750
17.8 双机复制.....	751
17.8.1 发送端服务器.....	751
17.8.2 主服务器.....	752
17.8.3 备服务器.....	754
17.9 查询规划.....	754
17.9.1 优化器方法配置.....	754
17.9.2 优化器开销常量.....	764
17.9.3 基因查询优化器.....	766
17.9.4 其他优化器选项.....	768
17.10 错误报告和日志.....	781
17.10.1 记录日志的位置.....	781

17.10.2 记录日志的时间.....	782
17.10.3 记录日志的内容.....	785
17.11 告警检测.....	792
17.12 运行时统计.....	792
17.12.1 查询和索引统计收集器.....	792
17.12.2 性能统计.....	797
17.13 资源管理.....	797
17.14 自动清理.....	811
17.15 客户端连接缺省设置.....	816
17.15.1 语句行为.....	816
17.15.2 区域和格式化.....	822
17.15.3 其他缺省.....	826
17.16 锁管理.....	826
17.17 版本和平台兼容性.....	829
17.17.1 历史版本兼容性.....	829
17.17.2 平台和客户端兼容性.....	832
17.18 容错性.....	833
17.19 连接池参数.....	834
17.20 集群事务.....	836
17.21 开发人员选项.....	839
17.22 审计.....	856
17.22.1 审计开关.....	856
17.22.2 操作审计.....	858
17.23 事务监控.....	865
17.24 GTM 相关参数.....	866
17.25 其它选项.....	867
<b>18 术语表.....</b>	<b>893</b>

# 1 欢迎

## 1.1 文档面向的读者对象

数据库开发指南重点面向数据库的设计者、应用程序开发人员或DBA，提供设计、构建、查询和维护数据仓库所需的信息。

作为数据库管理员和应用程序开发人员，至少需要了解以下知识：

- 操作系统知识。这是一切的基础。
- SQL语法。这是操作数据库的必备能力。

### 声明

GaussDB(DWS)的作者们在进行文档写作时努力基于商用角度，从使用场景和任务完成角度给出内容指引。即使这样，文档中依然可能存在对Postgres内容的引用和参考。对于这类内容，遵从如下的Postgres Copyright：

Postgres-XC is Copyright © 1996-2013 by the PostgreSQL Global Development Group.

PostgreSQL is Copyright © 1996-2013 by the PostgreSQL Global Development Group.

Postgres95 is Copyright © 1994-5 by the Regents of the University of California.

IN NO EVENT SHALL THE UNIVERSITY OF CALIFORNIA BE LIABLE TO ANY PARTY FOR DIRECT, INDIRECT, SPECIAL, INCIDENTAL, OR CONSEQUENTIAL DAMAGES, INCLUDING LOST PROFITS, ARISING OUT OF THE USE OF THIS SOFTWARE AND ITS DOCUMENTATION, EVEN IF THE UNIVERSITY OF CALIFORNIA HAS BEEN ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

THE UNIVERSITY OF CALIFORNIA SPECIFICALLY DISCLAIMS ANY WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE. THE SOFTWARE PROVIDED HEREUNDER IS ON AN "AS-IS" BASIS, AND THE UNIVERSITY OF CALIFORNIA HAS NO OBLIGATIONS TO PROVIDE MAINTENANCE, SUPPORT, UPDATES, ENHANCEMENTS, OR MODIFICATIONS.

## 1.2 阅读指引

对于首次接触 GaussDB(DWS)的用户，建议先阅读以下部分：

- 介绍GaussDB(DWS)服务的特点、功能和适用场景。
- GaussDB(DWS)入门包含一个示例，引导您完成创建数据仓库集群、创建数据库表、上传数据和测试查询这一过程。

如果打算或正在将应用程序从其他数据仓库向GaussDB(DWS)迁移，您可能想知道GaussDB(DWS)在实施方式上有什么区别。

GaussDB(DWS)进行数据库应用程序开发过程中，下表将帮您找到对应的信息。

如果要..	查阅建议
快速开始使用 GaussDB(DWS)。	首先，按照《 <a href="#">数据仓库服务快速入门</a> 》中的步骤快速部署集群、连接到数据库并尝试进行一些查询。  准备好构建数据库后，将数据加载到表中并编写查询内容以操作数据仓库中的数据后，可以回到《 <a href="#">数据仓库服务数据库开发指南</a> 》。
了解 GaussDB(DWS) 数据仓库的内部架构。	如果您想要更全面地了解GaussDB(DWS)服务，请转到GaussDB(DWS)产品首页。
了解如何设计表以实现良好性能。	<a href="#">开发设计建议</a> 介绍数据库应用程序开发过程中，应当遵守的设计规范。依据这些规范进行建模，能够更好的契合 GaussDB(DWS)的分布式处理架构，输出更高效的业务SQL代码。  对业务的执行效率不满意，期望通过调优加快业务执行的情况下，可以参考 <a href="#">优化查询性能</a> 进行调优。性能调优是一项复杂的工程，有些时候无法系统性地说明和解释，而是依赖于DBA的经验判断。尽管如此， <a href="#">优化查询性能</a> 一节还是期望能尽量系统性的对性能调优方法加以说明，方便应用开发人员和刚接触 GaussDB(DWS)的DBA参考。
加载数据。	<a href="#">导入数据</a> 介绍数据入库GaussDB(DWS)的方法和途径。 <a href="#">导入最佳实践</a> 提供有关快速高效数据导入的经验提示。
管理用户、组和数据库安全。	<a href="#">管理数据库安全</a> 涵盖数据库安全主题。
监控和优化系统性能。	<a href="#">系统表和系统视图</a> 详细介绍您可以从中查询数据库状态并监控查询内容与流程的系统表和视图。  您还应该查阅 <a href="#">管理指南</a> 了解如何使用GaussDB(DWS)管理控制台检查系统运行状况、监控指标。

## 1.3 文档表达约定

### 举例约定

内容	说明
dbadmin	表示创建集群时指定的运行和维护GaussDB(DWS)的用户。
8000	表示GaussDB(DWS)监听客户端连接请求的端口号。

手册中的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（<http://www.tpc.org/tpcds/>），安装TPC-DS benchmark。

### SQL 语法文本格式约定

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。

格式	意义
大写	语法关键字（语句中保持不变、必须照输的部分）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。
[ ]	表示用“[ ]”括起来的部分是可选的。
...	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。

## 1.4 前置条件

使用本指南前，需要完成以下任务。

- 创建GaussDB(DWS)集群。

- 安装SQL客户端。
- 将SQL客户端连接到集群的默认数据库。

关于上述任务的详细指导，请参见《[数据仓库服务快速入门](#)》。

# 2 系统概述

## 2.1 高可靠事务处理

GaussDB(DWS)提供集群事务管理功能，此功能是集群HA、集群故障切换的基础，负责保证集群所有节点间事务的ACID特性，保证故障可恢复，以及恢复后满足数据的ACID（Atomicity, Consistency, Isolation, Durability）特性，并负责节点的并发控制。

### 故障恢复

为了在集群出现故障时尽可能地不中断服务，GaussDB(DWS)提供了高可靠机制。通过保护关键用户程序对外不间断提供服务，把因为硬件、软件和人为造成的故障对业务的影响程度降到最低，以保证业务的持续性。

- 硬件级高可靠：磁盘Raid、交换机堆叠及网卡bond、不间断电源UPS（Uninterruptible Power Supply）。
- 软件级高可靠：GaussDB(DWS)集群CN、GTM、DN等全方位HA。

### 事务管理

- 支持事务块，用户可以通过start transaction命令显式启动一个事务块。
- 支持单语句事务，用户不显式启动事务，则单条语句就是一个事务。
- 分布式事务管理。支持全局事务信息管理，包括gxid、snapshot、timestamp的管理，分布式事务状态管理，gxid溢出的处理。
- 分布式事务支持ACID特性。
- 支持分布式死锁预防，保证在出现死锁时自动解锁或者预防死锁。

## 2.2 查询高性能

GaussDB(DWS)通过如下功能来努力实现查询的高性能。

### 全并行的数据查询处理

GaussDB(DWS)是采用Shared-nothing架构的MPP系统，其由众多拥有独立且互不共享CPU、内存、存储等系统资源的逻辑节点组成。在这样的系统架构中，业务数据被



分散存储在多个节点上，数据分析任务被推送到数据所在位置就近执行，并行地完成大规模的数据处理工作，实现对数据处理的快速响应。

GaussDB(DWS)后台还通过算子并行执行、指令在寄存器并行执行、及LLVM动态编译剪枝冗余的条件逻辑判断，助力数据查询性能提升。

## 行列混合存储

GaussDB(DWS)支持行存储和列存储两种存储模型，用户可以根据应用场景，建表的时候选择行存储还是列存储表。

行列混合存储引擎可以同时为用户提供更优的数据压缩比（列存）、更好的索引性能（列存）、更好的点更新和点查询（行存）性能。

## 列存下的数据压缩

对于非活跃的早期数据可以通过压缩来减少空间占用，降低采购和运维成本。

GaussDB(DWS)列存储压缩支持Delta Value Encoding、Dictionary、RLE、LZ4、ZLIB等压缩算法，且能够根据数据特征自适应的选择压缩算法，平均压缩比7:1。压缩数据可直接访问，对业务透明，极大缩短历史数据访问的准备时间。

## 2.3 相关概念

### 数据库

数据库用于管理各类数据对象，与其他数据库隔离。创建数据库时可以指定对应的表空间，如果不指定相应的表空间，相关的对象会默认保存在PG\_DEFAULT空间中。数据库管理的对象可分布在多个表空间上。

### 实例

实例在GaussDB(DWS)中是运行在内存中的一组数据库进程，一个实例可以管理一个或多个数据库，这些数据库组成一个集簇。集簇是存储磁盘上的一个区域，这个区域在安装时初始化并由一个目录组成，所有数据都存储在这个目录中，这个目录被称为数据目录，使用initdb创建。理论上来说一个服务器上可以在不同的端口启动多个实例，但是GaussDB(DWS)一次只能管理一个实例，启动和停止都是依赖于具体的数据目录。以后由于兼容的需要不排除引入实例名这个概念的可能。

### 表空间

在GaussDB(DWS)中，表空间是一个目录，可以存在多个，里面存储的是它所包含的数据库的各种物理文件。由于表空间是一个目录，仅是起到了物理隔离的作用，其管理功能依赖于文件系统。

### 模式

GaussDB(DWS)的模式是对数据库做一个逻辑分割。所有的数据库对象都建立在模式下面。GaussDB(DWS)的模式和用户是弱绑定的，所谓的弱绑定是指虽然创建用户的同时会自动创建一个同名模式，但用户也可以单独创建模式，并且为用户指定其他的模式。

## 用户和角色

GaussDB(DWS)使用用户和角色来控制对数据库的访问。根据角色自身的设置不同，一个角色可以看做是一个数据库用户，或者一组数据库用户。在GaussDB(DWS)中角色和用户之间的区别只在于角色默认是没有LOGIN权限的。在GaussDB(DWS)中一个用户唯一对应一个角色，不过可以使用角色叠加来更灵活地进行管理。

## 事务管理

在事务管理上，GaussDB(DWS)采取了MVCC（多版本并发控制）结合两阶段锁的方式，其特点是读写之间不阻塞。GaussDB(DWS)的MVCC没有将历史版本数据统一存放，而是和当前元组的版本放在了一起。GaussDB(DWS)没有回滚段的概念，但是为了定期清除历史版本数据GaussDB(DWS)引入了一个VACUUM进程。一般情况下用户不用关注它，除非要做性能调优。此外，GaussDB(DWS)是自动提交事务。

# 3 定义数据库对象

## 3.1 创建和管理数据库

数据库 ( Database ) 是表、索引、视图、存储过程、操作符等对象的集合。GaussDB(DWS)支持创建多个数据库，但是客户端程序一次只能连接并访问一个数据库，无法跨数据库进行查询。

### 模板和默认数据

- GaussDB(DWS)提供了两个模板数据库template0、template1，以及一个默认的数据库gaussdb。
- 默认情况下，每个新创建的数据库都是基于一个模板数据库。GaussDB(DWS)数据库默认使用template1作为模板，编码格式为SQL\_ASCII，且不允许自定义字符编码。若创建数据库时需指定字符编码，请使用template0创建数据库。
- 请避免使用客户端或其他手段连接及操作两个模板数据库。

#### 说明

通过“show server\_encoding”命令可以查看当前数据库存储编码。

### 创建数据库

使用CREATE DATABASE命令创建一个新的数据库。

```
CREATE DATABASE mydatabase;
```

用户必须拥有数据库创建的权限或者是数据库的系统管理员权限才能创建数据库，赋予创建数据库的权限参见[管理用户及权限](#)。

## 📖 说明

- 创建数据库时，若数据库名称长度超过63字节，server端会对数据库名称进行截断，保留前63个字节，因此建议数据库名称长度不要超过63个字节，不要使用多字节字符做为对象名。（如果出现因为误操作导致在多字节字符的中间截断进而无法删除数据库对象的现象，请使用截断前的数据库对象名进行删除操作，或将该对象从各个数据库节点的相应系统表中依次删掉。）
- 数据库名称遵循SQL标识符的一般规则。当前用户自动成为此新数据库的所有者。
- 如果一个数据库系统用于承载相互独立的用户和项目，建议把它们放在不同的数据库里。
- 如果项目或者用户是相互关联的，并且可以相互使用对方的资源，则应该把它们放在同一个数据库里，但可以规划在不同的Schema中。
- GaussDB(DWS)允许创建的数据库总数目上限为128个。

## 查看数据库

查看数据库的方式：

- 使用\l元命令查看数据库系统的数据库列表。  

```
\l
```
- 使用如下命令通过系统表pg\_database查询数据库列表。  

```
SELECT datname FROM pg_database;
```

## 修改数据库

用户可以使用如下命令修改数据库属性（比如：owner、名称和默认的配置属性）。

- 使用以下命令为数据库设置默认的模式搜索路径。  

```
ALTER DATABASE mydatabase SET search_path TO pa_catalog,public;
```
- 使用如下命令为数据库重新命名。  

```
ALTER DATABASE mydatabase RENAME TO newdatabase;
```

## 删除数据库

用户可以使用DROP DATABASE命令删除数据库。该命令删除了数据库中的系统目录，并且删除了带有数据的磁盘上的数据库目录。用户必须是数据库的owner或者系统管理员才能删除数据库。当有用户连接数据库时，删除操作会失败。删除数据库时请先连接到其他的数据库。

使用如下命令删除数据库：  

```
DROP DATABASE newdatabase;
```

## 3.2 创建和管理 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。

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。

## 默认的“public” Schema

每个数据库都有一个名为public的Schema。如果用户没有创建任何Schema，对象会被创建在名为public的Schema中。所有的数据库角色（用户）都在public Schema中拥有CREATE和USAGE特权。在用户创建一个Schema时，用户应该为允许访问该Schema的用户授予特权。

## 创建 Schema

- 使用CREATE SCHEMA命令来创建一个新的Schema。

```
CREATE SCHEMA myschema;
```

如果需要在模式中创建或者访问对象，其完整的对象名称由模式名称和具体的对象名称组成。中间由符号“.”隔开。例如：myschema.table。

- 用户可以创建一个由他人拥有的schema。例如，创建名为myschema的Schema，并指定Schema的所有者为用户jack。

```
CREATE SCHEMA myschema AUTHORIZATION jack;
```

若不指定authorization username，则其所有者为执行该命令的用户。

## 修改 Schema

- 使用ALTER SCHEMA修改Schema名称，只有Schema所有者可以更改Schema。

```
ALTER SCHEMA schema_name RENAME TO new_name;
```

- 使用ALTER SCHEMA修改Schema所有者：

```
ALTER SCHEMA schema_name OWNER TO new_owner;
```

## 设置 Schema 搜索路径

GUC参数search\_path设置Schema的搜索顺序，参数取值形式为采用逗号分隔的Schema名称列表。如果创建对象时未指定目标Schema，则将该对象会被添加到搜索路径中列出的第一个Schema中。当不同Schema中存在同名的对象时，查询对象未指定Schema的情况下，将从搜索路径中包含该对象的第一个Schema中返回对象。

- 使用SHOW命令查看当前搜索路径。

```
SHOW SEARCH_PATH;  
search_path  
-----  
"$user",public  
(1 row)
```

search\_path参数的默认值为：“\$user”，public。\$user表示与当前会话用户名同名的Schema名，如果这样的模式不存在，\$user将被忽略。所以默认情况下，用户连接数据库后，如果数据库下存在同名Schema，则对象会添加到同名Schema下，否则对象被添加到Public Schema下。

- 使用SET命令修改当前会话的默认Schema。例如，将搜索路径设置为myschema、public，首先搜索myschema。

```
SET SEARCH_PATH TO myschema, public;
```

也可以使用ALTER ROLE命令为特定的角色（用户）设置search\_path。例如：

```
ALTER ROLE jack SET search_path TO myschema, public;
```

## 使用 Schema

在特定Schema下创建对象或者访问特定Schema下的对象，需要使用有Schema修饰的对象名。名称包含Schema名以及对象名，之间用“.”号分开。

- 在myschema下创建mytable表。以schema\_name.table\_name格式创建表。

```
CREATE TABLE myschema.mytable(id int, name varchar(20));
```

- 查询myschema下mytable表的所有数据。

```
SELECT * FROM myschema.mytable;  
id | name  
----+-----  
(0 rows)
```

## 查看 Schema

- 使用current\_schema()函数查看当前Schema:

```
SELECT current_schema();  
current_schema  
-----  
myschema  
(1 row)
```

- 要查看Schema所有者, 请对系统表PG\_NAMESPACE和PG\_USER执行如下关联查询。语句中的schema\_name请替换为实际要查找的Schema名称。

```
SELECT s.nspname,u.username AS nspowner FROM PG_NAMESPACE s, PG_USER u WHERE  
nspname='schema_name' AND s.nspowner = u.usesysid;
```

- 要查看所有Schema的列表, 请查询PG\_NAMESPACE系统表。

```
SELECT * FROM PG_NAMESPACE;
```

- 要查看属于某Schema下表的列表, 请查询系统视图PG\_TABLES。例如, 以下查询会返回Schema PG\_CATALOG中的表列表。

```
SELECT distinct(tablename),schemaname FROM PG_TABLES where schemaname = 'pg_catalog';
```

## Schema 的权限控制

默认情况下, 用户只能访问属于自己的Schema中的数据库对象。如需要访问其他Schema的对象, 则需赋予对应Schema的usage权限。

通过将模式的CREATE权限授予某用户, 被授权用户就可以在此模式中创建对象。

- 将myschema的usage权限赋给用户jack。

```
GRANT USAGE ON schema myschema TO jack;
```

- 将用户jack对于myschema的usage权限收回。

```
REVOKE USAGE ON schema myschema FROM jack;
```

## 删除 Schema

- 使用DROP SCHEMA命令删除一个空的Schema (即该Schema下没有数据库对象)。

```
DROP SCHEMA IF EXISTS myschema;
```

- 默认情况下, 删除一个Schema前, 它必须为空。要删除一个Schema及其包含的所有对象 (表、数据、函数等等), 需要使用CASCADE关键字。

```
DROP SCHEMA myschema CASCADE;
```

## 系统 Schema

- 每个数据库都包含一个pg\_catalog schema, 它包含系统表和所有内置数据类型、函数、操作符。pg\_catalog是搜索路径中的一部分, 始终在临时表所属的模式后面, 并在search\_path中所有模式的前面, 即具有第二搜索优先级。这样确保可以搜索到数据库内置对象。如果用户需要使用和系统内置对象重名的自定义对象, 可以在操作自定义对象时带上自己的模式。
- information\_schema由一个包含数据库中对象信息的视图集合组成。这些视图以一种标准化的方式从系统目录表中得到系统信息。

## 3.3 创建和管理表

### 创建表

CREATE TABLE命令创建一个表，创建表时可以定义以下内容：

- 表的列及[数据类型](#)。
- 表约束的定义，即任何用于限制列或者表中能包含的数据的表或者列约束。参见[表约束的定义](#)。
- 表分布的定义，即表的分布策略，它决定GaussDB(DWS)数据库如何在Segment之间划分数据。参见[表分布的定义](#)。
- 表存储格式。参见[选择表存储模型](#)。
- 分区表定义。参见[表分区定义](#)。

示例：CREATE TABLE创建了一个表web\_returns\_p1，使用wr\_item\_sk作为分布键，并基于wr\_returned\_date\_sk设置了range分布功能。

```
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 p2019 START(20191231) END(20221231) EVERY(10000),
  PARTITION p0 END(maxvalue)
);
```

### 表约束的定义

可以在列和表上定义约束来限制表中的数据，但是有以下一些限制：

1. 表中的主键约束和唯一约束必须包含分布列。
2. 列存表支持PARTIAL CLUSTER KEY、主键和唯一表级约束，不支持外键表级约束。
3. 列存表的字段约束只支持NULL、NOT NULL和DEFAULT常量值。

- 检查约束

检查约束允许用户指定一个特定列中的值必须满足一个布尔（真值）表达式。例如，要求产品价格为正：

```
CREATE TABLE products
(
  product_no integer,
  name text,
  price numeric CHECK (price > 0)
);
```

- 非空约束

非空约束指定一个列不能有空值。非空约束总是被写作为列约束。例如：

```
CREATE TABLE products
(
```

```
product_no integer NOT NULL,
name text NOT NULL,
price numeric
);
```

- 唯一约束

唯一约束确保一列或者一组列中包含的数据对于表中所有的行都是唯一的。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。

```
CREATE TABLE products
(
    product_no integer UNIQUE,
    name text,
    price numeric
)DISTRIBUTE BY HASH(product_no);
```

- 主键

主键约束是一个UNIQUE约束和一个NOT NULL约束的组合。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。如果一个表具有主键，这个列（或者这一组列）会被默认选中为该表的分布键。

例如：

```
CREATE TABLE products
(
    product_no integer PRIMARY KEY,
    name text,
    price numeric
)DISTRIBUTE BY HASH(product_no);
```

- 局部聚簇

局部聚簇通过min/max稀疏索引较快的实现基表扫描的filter过滤。Partial Cluster Key可以指定多列，但是一般不建议超过2列。例如：

```
CREATE TABLE products
(
    product_no integer,
    name text,
    price numeric,
    PARTIAL CLUSTER KEY(product_no)
) WITH (ORIENTATION = COLUMN);
```

## 表分布的定义

- GaussDB(DWS)支持的分布方式：复制表（Replication）、哈希表（Hash）和轮询表（Roundrobin）。

### 说明

轮询表（Roundrobin）分布方式仅8.1.2及以上集群版支持。



策略	描述	适用场景	优势与劣势
复制表 (Replication)	集群中每一个DN实例上都有一份全量表数据。	小表、维度表。	<ul style="list-style-type: none"> <li>Replication优点是每个DN上都有此表的全量数据，在join操作中可以避免数据重分布操作，从而减小网络开销，同时减少了plan segment(每个plan segment都会起对应的线程)</li> <li>Replication缺点是每个DN都保留了表的完整数据，造成数据的冗余。一般情况下只有较小的维度表才会定义为Replication表。</li> </ul>
哈希表 (Hash)	表数据通过hash方式散列到集群中的所有DN实例上。	数据量较大的事实表。	<ul style="list-style-type: none"> <li>在读/写数据时可以利用各个节点的IO资源，大大提升表的读/写速度。</li> <li>一般情况下大表(1000000条记录以上)定义为Hash表。</li> </ul>
轮询表 (Roundrobin)	表的每一行被轮番地发送给各个DN，数据会被均匀地分布在各个DN中。	数据量较大的事实表，且使用Hash分布时找不到合适的分布列。	<ul style="list-style-type: none"> <li>Roundrobin优点是保证了数据不会发生倾斜，从而提高了集群的空间利用率。</li> <li>Roundrobin缺点是无法像Hash表一样进行DN本地化优化，查询性能通常不如Hash表。</li> <li>一般在在大表无法找到合适的分布列时，定义为Roundrobin表，若大表能够找到合适的分布列，优先选择性能更好的Hash分布。</li> </ul>

• 选择分布列

采用Hash分布方式，需要为用户表指定一个分布列(distribute key)。当插入一条记录时，系统会根据分布列的值进行hash运算后，将数据存储在对应的DN中。所以Hash分布列选取至关重要，需要满足以下原则：

- 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。
- 在满足第一条原则的情况下尽量不要选取存在常量filter的列。**例如，表dwcjk相关的部分查询中出现dwcjk的列zqdh存在常量的约束(例如zqdh='000001')，那么就应当尽量不用zqdh做分布列。
- 在满足前两条原则的情况，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。

对于Hash分表策略，如果分布列选择不当，可能导致数据倾斜，查询时出现部分DN的I/O短板，从而影响整体查询性能。因此在采用Hash分表策略之后需对表的数据进行数据倾斜性检查，以确保数据在各个DN上是均匀分布的。可以使用以下SQL检查数据倾斜性

```
select
xc_node_id, count(1)
from tablename
group by xc_node_id
order by xc_node_id desc;
```

其中xc\_node\_id对应DN，一般来说，不同DN的数据量相差5%以上即可视为倾斜，如果相差10%以上就必须调整分布列。

- d. 一般不建议用户新增一列专门用作分布列，尤其不建议用户新增一列，然后用SEQUENCE的值来填充做为分布列。因为SEQUENCE可能会带来性能瓶颈和不必要的维护成本。

## 查看表数据

- 使用系统表pg\_tables查询数据库所有表的信息。  

```
SELECT * FROM pg_tables;
```
- 使用gsq的\d+命令查询表的属性。  

```
\d+ customer_t1;
```
- 执行如下命令查询表customer\_t1的数据量。  

```
SELECT count(*) FROM customer_t1;
```
- 执行如下命令查询表customer\_t1的所有数据。  

```
SELECT * FROM customer_t1;
```
- 执行如下命令只查询字段c\_customer\_sk的数据。  

```
SELECT c_customer_sk FROM customer_t1;
```
- 执行如下命令过滤字段c\_customer\_sk的重复数据。  

```
SELECT DISTINCT( c_customer_sk ) FROM customer_t1;
```
- 执行如下命令查询字段c\_customer\_sk为3869的所有数据。  

```
SELECT * FROM customer_t1 WHERE c_customer_sk = 3869;
```
- 执行如下命令按照字段c\_customer\_sk进行排序。  

```
SELECT * FROM customer_t1 ORDER BY c_customer_sk;
```

## 删除表数据

在使用表的过程中，可能会需要删除已过期的数据，删除数据必须从表中整行的删除。

SQL不能直接访问独立的行，只能通过声明被删除行匹配的条件进行。如果表中有一个主键，用户可以指定准确的行。用户可以删除匹配条件的一组行或者一次删除表中的所有行。

- 使用DELETE命令删除行，如果删除表customer\_t1中所有c\_customer\_sk为3869的记录：  

```
DELETE FROM customer_t1 WHERE c_customer_sk = 3869;
```
- 如果执行如下命令，会删除表中所有的行。  

```
DELETE FROM customer_t1;
TRUNCATE TABLE customer_t1;
```

### 说明

全表删除的场景下，建议使用truncate，不建议使用delete。

- 删除创建的表。

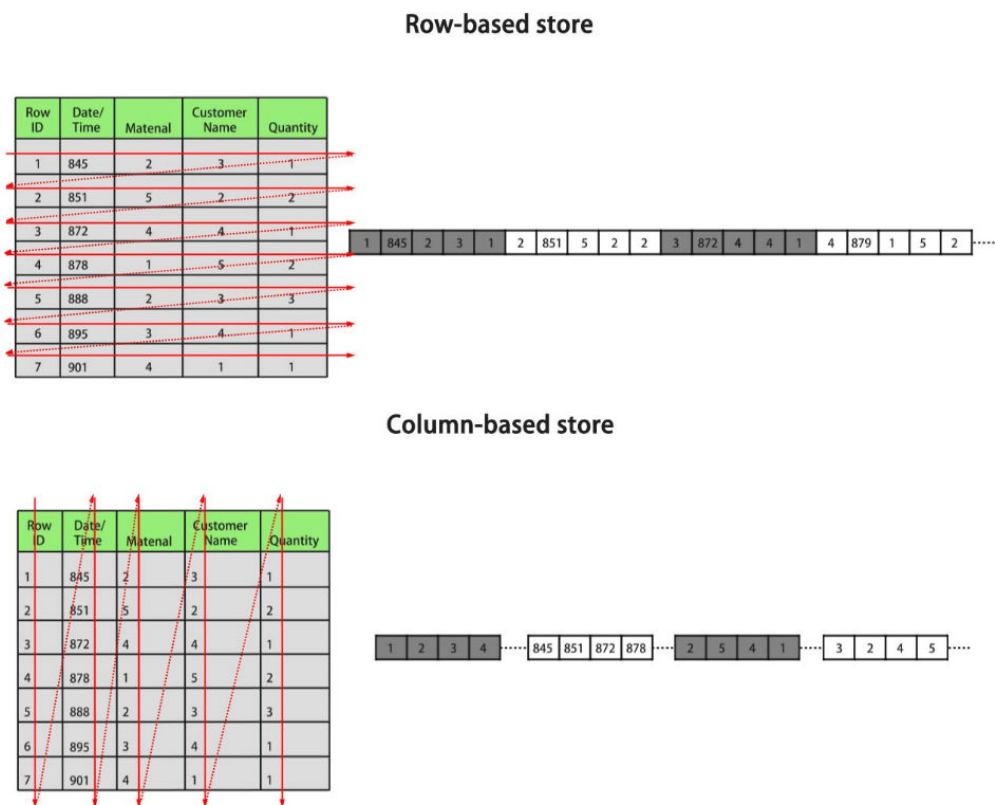
```
DROP TABLE customer_t1;
```

### 3.4 选择表存储模型

GaussDB(DWS)支持行列混合存储。当创建一个表时，可以选择表的存储格式为行存储或列存储。

行存储是指将表按行存储到硬盘分区上，列存储是指将表按列存储到硬盘分区上。默认情况下，创建的表为行存储。行存储和列存储的差异请参见图3-1。

图 3-1 行存储和列存储的差异



上图中，左上为行存表，右上为行存表在硬盘上的存储方式。左下为列存表，右下为列存表在硬盘上的存储方式。

GaussDB(DWS)表的行/列存储通过表定义的orientation属性定义。当指定orientation属性为row时，表为行存储；当指定orientation属性为column时，表为列存储；如果不指定，默认为行存储。行、列存储模型各有优劣，建议根据实际情况选择：

存储模型	优点	缺点	适用场景
行存	数据被保存在一起。INSERT/UPDATE容易。	选择(Selection)时即使只涉及某几列，所有数据也都会被读取。	<ul style="list-style-type: none"> <li>表的字段个数比较少，查询大部分字段。</li> <li>点查询(返回记录少，基于索引的简单查询)。</li> <li>增、删、改操作较多的场景。</li> </ul>
列存	<ul style="list-style-type: none"> <li>查询时只有涉及到的列会被读取。</li> <li>投影(Projection)很高效。</li> <li>任何列都能作为索引。</li> </ul>	<ul style="list-style-type: none"> <li>选择完成时，被选择的列要重新组装。</li> <li>INSERT/UPDATE比较麻烦。</li> </ul>	<ul style="list-style-type: none"> <li>表的字段比较多(大宽表)，查询中涉及到的列不多。</li> <li>统计分析类查询(关联、分组操作较多的场景)。</li> <li>即席查询(查询条件不确定，行存表扫描难以使用索引)。</li> </ul>

## 创建一个行存表

例如，创建一个名为customer\_t1的行存表：

```
CREATE TABLE customer_t1
(
  state_ID CHAR(2),
  state_NAME VARCHAR2(40),
  area_ID NUMBER
);
```

## 创建一个列存表

例如，创建一个名为customer\_t2的列存表：

```
CREATE TABLE customer_t2
(
  state_ID CHAR(2),
  state_NAME VARCHAR2(40),
  area_ID NUMBER
)
WITH (ORIENTATION = COLUMN);
```

## 使用压缩

表压缩可以在创建表时开启，压缩表能够使表中的数据以压缩格式存储，意味着占用相对少的内存。

对于I/O读写量大，CPU富足(计算相对小)的场景，选择高压缩比；反之选择低压缩比。建议依据此原则进行不同压缩下的测试和对比，以选择符合自身业务情况的最优压缩比。压缩比通过COMPRESSION参数指定，其支持的取值如下：

- 列存表为：YES/NO/LOW/MIDDLE/HIGH，默认值为LOW。
- 行存表为：YES/NO，默认值为NO。(行存表压缩功能暂未商用，如需使用请联系技术支持工程师)

各压缩级别所适用的业务场景说明如下：

压缩级别	所适用的业务场景
低级别压缩	系统CPU使用率高，存储磁盘空间充足。
中度压缩	系统CPU使用率适中，但存储磁盘空间不是特别充足。
高级别压缩	系统CPU使用率低，磁盘空间不充裕。

例如，创建一个名为customer\_t3的列存压缩表：

```
CREATE TABLE customer_t3
(
  state_ID CHAR(2),
  state_NAME VARCHAR2(40),
  area_ID NUMBER
)
WITH (ORIENTATION = COLUMN,COMPRESSION=middle);
```

## 3.5 表分区定义

分区表就是把逻辑上的一张表根据分区策略分成几张物理块库进行存储，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。当进行条件查询时，系统只会扫描满足条件的分区，避免全表扫描，从而提升查询性能。

分区表的优势：

- 改善查询性能。对分区对象的查询可以仅搜索自己关心的分区，提高检索效率。
- 增强可用性。如果分区表的某个分区出现故障，表在其他分区的数据仍然可用。
- 提升可维护性。对于需要周期性删除的过期历史数据，可以通过drop/truncate分区的方式快速高效处理。

### 支持的表分区类型

- 范围分区（Range Partitioning），基于一个数值型范围划分数据，例如按日期或价格区间定义。
- 列表分区（List Partitioning），基于一个值列表划分数据，例如按销售范围或产品属性定义。仅8.1.3及以上集群版本支持。

### 分区策略选择

当表有以下特征时，可以考虑使用表分区策略：

- 数据具有明显区间性的字段。  
分区表需要根据有明显区间性字段进行表分区。比如按照日期、区域、数值等字段进行分区，时间字段是最常见的分区字段。
- 业务查询有明显的区间范围特征。  
查询数据可落到区间范围指定的分区内，这样才能通过分区剪枝，只扫描查询需要的分区，从而提升数据扫描效率，降低数据扫描的IO开销。
- 表数据量比较大。

小表扫描本身耗时不大，分区表的性能收益不明显，因此只建议对大表采取分区策略。列存储模式下因为每个列是单独的文件出处，且最小的存储单元CU可存储6w行数据，因此对于列存分区表，建议每个分区的数据不小于DN个数\*6w。

## 创建范围(range)分区表

示例：创建一个按wr\_returned\_date\_sk范围分区的表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 VALUES LESS THAN(20161231),
  PARTITION p2017 VALUES LESS THAN(20171231),
  PARTITION p2018 VALUES LESS THAN(20181231),
  PARTITION p2019 VALUES LESS THAN(20191231),
  PARTITION pxxxx VALUES LESS THAN(maxvalue)
);
```

对于分区间隔固定、批量创建分区的场景。可使用如下示例：

```
CREATE TABLE web_returns_p2
(
  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(20161231) END(20191231) EVERY(10000),
  PARTITION p0 END(maxvalue)
);
```

## 创建列表(list)分区表

LIST分区表可以使用任意允许值比较的列作为分区键列。创建LIST分区表时，必须要为每一个分区声明每一个值分区。

示例：创建LIST分区表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 province1_202201 VALUES (('202201', 'city1'), ('202201', 'city2')),
  PARTITION province2_202201 VALUES (('202201', 'city3'), ('202201', 'city4'), ('202201', 'city5')),
  PARTITION rest VALUES (DEFAULT)
);
```

## 对已有的表进行分区

表只能在创建时被分区。如果用户有一个表想要分区，用户必须创建一个分过区的表，把原始表的数据载入到新表，再删除原始表并且把分过区的表重命名为原始表的名称。用户还必须重新授权表上的权限。例如：

```
CREATE TABLE web_returns_p2
(
  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(20161231) END(20191231) EVERY(10000),
  PARTITION p0 END(maxvalue)
);
INSERT INTO web_returns_p2 SELECT * FROM web_returns_p1;
DROP TABLE web_returns_p1;
ALTER TABLE web_returns_p2 RENAME TO web_returns_p1;
GRANT ALL PRIVILEGES ON web_returns_p1 TO dbadmin;
GRANT SELECT ON web_returns_p1 TO jack;
```

## 增加一个分区

使用ALTER TABLE语句为范围分区表增加一个分区。例如，为表web\_returns\_p1增加分区P2020。

```
ALTER TABLE web_returns_p1 ADD PARTITION P2020 VALUES LESS THAN (20201231);
```

## 分割一个分区

范围分区表和列表分区表分割分语法有所区别：

- 使用ALTER TABLE语句为范围分区表分割一个分区。例如，将表web\_returns\_p1分区pxxxx以20201231为分割点分割为p2020和p20xx两个分区。  

```
ALTER TABLE web_returns_p1 SPLIT PARTITION pxxxx AT(20201231) INTO (PARTITION p2020,PARTITION p20xx);
```
- 使用ALTER TABLE语句为列表分区表分割一个分区。例如，将表sales\_info分区province2\_202201分割为province3\_202201和province4\_202201两个分区。  

```
ALTER TABLE sales_info SPLIT PARTITION province2_202201 VALUES (('202201', 'city5')) INTO (PARTITION province3_202201,PARTITION province4_202201);
```

## 合并一个分区

使用ALTER TABLE语句为分区表合并一个分区。例如，将表web\_returns\_p1分区p2016和p2017合并为一个分区p20162017。

```
ALTER TABLE web_returns_p1 MERGE PARTITIONS p2016,p2017 INTO PARTITION p20162017;
```

## 删除一个分区

使用ALTER TABLE语句从分区表中删除一个分区。例如，删除表web\_returns\_p1的分区P2020。

```
ALTER TABLE web_returns_p1 DROP PARTITION P2020;
```

## 查询分区

- 查询分区p2019。  

```
SELECT * FROM web_returns_p1 PARTITION (p2019);  
SELECT * FROM web_returns_p1 PARTITION FOR (20201231);
```
- 查看分区表信息，可使用系统表dba\_tab\_partitions。  

```
SELECT * FROM dba_tab_partitions where table_name='web_returns_p1';
```

## 删除分区表

使用DROP TABLE语句删除一个分区表。

```
DROP TABLE web_returns_p1;
```

## 3.6 创建和管理索引

索引可以提高数据的访问速度，但同时也增加了插入、更新和删除操作的处理时间。所以是否要为表增加索引，索引建立在哪些字段上，是创建索引前必须要考虑的问题。需要分析应用程序的业务处理、数据使用、经常被用作查询的条件或者被要求排序的字段来确定是否建立索引。

### 索引类型

- 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。btree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。

### 索引的选择原则

索引建立在数据库表中的某些列上。因此，在创建索引时，应该仔细考虑在哪些列上创建索引。

- 在经常需要搜索查询的列上创建索引，可以加快搜索的速度。
- 在作为主键的列上创建索引，强制该列的唯一性和组织表中数据的排列结构。
- 在经常使用连接的列上创建索引，这些列主要是一些外键，可以加快连接的速度。
- 在经常需要根据范围进行搜索的列上创建索引，因为索引已经排序，其指定的范围是连续的。
- 在经常需要排序的列上创建索引，因为索引已经排序，这样查询可以利用索引的排序，加快排序查询时间。
- 在经常使用WHERE子句的列上创建索引，加快条件的判断速度。
- 为经常出现在关键字ORDER BY、GROUP BY、DISTINCT后面的字段建立索引。



 说明

- 索引创建成功后，系统会自动判断何时引用索引。当系统认为使用索引比顺序扫描更快时，就会使用索引。
- 索引创建成功后，必须和表保持同步以保证能够准确地找到新数据，这样就增加了数据操作的负荷。因此请定期删除无用的索引

## 创建索引

GaussDB(DWS)支持4种创建索引的方式请参见表3-1。

 说明

- 索引创建成功后，系统会自动判断何时引用索引。当系统认为使用索引比顺序扫描更快时，就会使用索引。
- 索引创建成功后，必须和表保持同步以保证能够准确地找到新数据，这样就增加了数据操作的负荷。因此请定期删除无用的索引。

表 3-1 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GaussDB(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，GaussDB(DWS)只有B-Tree可以创建唯一索引。
多字段索引	一个索引可以定义在表中的多个属性上。目前，GaussDB(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。

- 创建一个普通表。

```
CREATE TABLE tpcds.customer_address_bak AS TABLE tpcds.customer_address;
```

- 创建普通索引

如果对于tpcds.customer\_address\_bak表，需要经常进行以下查询。

```
SELECT ca_address_sk FROM tpcds.customer_address_bak WHERE ca_address_sk=14888;
```

通常，数据库系统需要逐行扫描整个tpcds.customer\_address\_bak表以寻找所有匹配的元组。如果表tpcds.customer\_address\_bak的规模很大，但满足WHERE条件的只有少数几个（可能是零个或一个），则这种顺序扫描的性能就比较差。如果让数据库系统在ca\_address\_sk属性上维护一个索引，用于快速定位匹配的元组，则数据库系统只需要在搜索树上查找少数的几层就可以找到匹配的元组，这将会大大提高数据查询的性能。同样，在数据库中进行更新和删除操作时，索引也可以提升这些操作的性能。

使用以下命令创建索引。

```
CREATE INDEX index_wr_returned_date_sk ON tpcds.customer_address_bak (ca_address_sk);
```

- 创建多字段索引

假如用户需要经常查询表tpcds.customer\_address\_bak中ca\_address\_sk是5050，且ca\_street\_number小于1000的记录，使用以下命令进行查询。

```
SELECT ca_address_sk,ca_address_id FROM tpcds.customer_address_bak WHERE ca_address_sk = 5050 AND ca_street_number < 1000;
```

使用以下命令在字段ca\_address\_sk和ca\_street\_number上定义一个多字段索引。

```
CREATE INDEX more_column_index ON tpcds.customer_address_bak(ca_address_sk,ca_street_number);
```

- 创建部分索引

如果只需要查询ca\_address\_sk为5050的记录，可以创建部分索引来提升查询效率。

```
CREATE INDEX part_index ON tpcds.customer_address_bak(ca_address_sk) WHERE ca_address_sk = 5050;
```

- 创建表达式索引

假如经常需要查询ca\_street\_number小于1000的信息，执行如下命令进行查询。

```
SELECT * FROM tpcds.customer_address_bak WHERE trunc(ca_street_number) < 1000;
```

可以为上面的查询创建表达式索引：

```
CREATE INDEX para_index ON tpcds.customer_address_bak (trunc(ca_street_number));
```

## 查看索引

- 执行如下命令查询系统和用户定义的所有索引。

```
SELECT RELNAME FROM PG_CLASS WHERE RELKIND='i';
```

- 执行如下命令查询指定索引的信息。

```
\di+ index_wr_returned_date_sk
```

## 重建索引

- 重建索引index\_wr\_returned\_date\_sk：

```
REINDEX INDEX index_wr_returned_date_sk;
```

- 重建表上的所有索引：

```
REINDEX TABLE tpcds.customer_address_bak;
```

## 删除索引

使用DROP INDEX命令删除索引：

```
DROP INDEX index_wr_returned_date_sk;
```

## 3.7 创建和使用序列

序列Sequence是用来产生唯一整数的数据库对象。序列的值是按照一定规则自增的整数。因为自增所以不重复，因此说Sequence具有唯一标识性。这也是Sequence常被用作主键的原因。

通过序列使某字段成为唯一标识符的方法有两种：

- 一种是声明字段的类型为序列整型，由数据库在后台自动创建一个对应的Sequence。

- 另一种是使用CREATE SEQUENCE自定义一个新的Sequence，然后将nextval('sequence\_name')函数读取的序列值，指定为某一字段的默认值，这样该字段就可以作为唯一标识符。

## 创建序列

方法一：声明字段类型为序列整型来定义标识符字段。例如：

```
CREATE TABLE T1
(
  id serial,
  name text
);
```

方法二：创建序列，并通过nextval('sequence\_name')函数指定为某一字段的默认值。这种方式更灵活，可以为序列定义cache，一次预申请多个序列值，减少与GTM的交互次数，来提高性能。

### 1. 创建序列

```
CREATE SEQUENCE seq1 cache 100;
```

### 2. 指定为某一字段的默认值，使该字段具有唯一标识属性。

```
CREATE TABLE T2
(
  id int not null default nextval('seq1'),
  name text
);
```

### 📖 说明

除了为序列指定了cache，方法二所实现的功能基本与方法一类似。但是一旦定义cache，序列将会产生空洞(序列值为不连贯的数值，如：1.4.5)，并且不能保序。另外为某序列指定从属列后，该列删除，对应的sequence也会被删除。虽然数据库并不限制序列只能为一列产生默认值，但最好不要多列共用同一个序列。

当前版本只支持在定义表的时候指定自增列，或者指定某列的默认值为nextval('seqname')，不支持在已有表中增加自增列或者增加默认值为nextval('seqname')的列。

## 修改一个序列

ALTER SEQUENCE命令更改现有序列的属性，包括修改修改拥有者、归属列和最大值。

- 指定序列与列的归属关系。

将序列和一个表的指定字段进行关联。在删除那个字段或其所在表的时候会自动删除已关联的序列。

```
ALTER SEQUENCE seq1 OWNED BY T2.id;
```

- 将序列serial的最大值修改为300：

```
ALTER SEQUENCE seq1 MAXVALUE 300;
```

## 删除一个序列

使用DROP SEQUENCE命令删除一个序列。例如，将删除名为seq1的序列：

```
DROP SEQUENCE seq1;
```

## 注意事项

新序列值的产生是靠GTM维护的，默认情况下，每申请一个序列值都要向GTM发送一次申请，GTM在当前值的基础上加上步长值作为产生的新值返回给调用者。GTM作为全局唯一的节点，势必成为性能的瓶颈，所以对于需要大量频繁产生序列号的操作，如使用Bulkload工具进行数据导入场景，是非常不推荐产生默认序列值的。比如，在下面所示的场景中，INSERT FROM SELECT语句的性能会非常慢。

```
CREATE SEQUENCE newSeq1;
CREATE TABLE newT1
```

```
(
  id int not null default nextval('newSeq1'),
  name text
);
INSERT INTO newT1(name) SELECT name from T1;
```

可以提高性能的写法是（假设T1表导入newT1表中的数据为10000行）：

```
INSERT INTO newT1(id, name) SELECT id,name from T1;
SELECT SETVAL('newSeq1',10000);
```

#### 📖 说明

序列操作函数nextval(), setval() 等均不支持回滚。另外setval设置的新值，会对当前会话的nextval立即生效，但对其他会话，如果定义了cache，不会立即生效，在用尽所有缓存的值后，其变动才被其他会话感知。所以为了避免产生重复值，要谨慎使用setval，设置的新值不能是已经产生的值或者在缓存中的值。

如果必须要在bulkload场景下产生默认序列值，则一定要为newSeq1定义足够大的cache，并且不要定义Maxvalue或者Minvalue。数据库会试图将nextval('sequence\_name')的调用下推到Data Node，以提高性能。目前GTM对并发的连接请求是有限制的，当Data Node很多时，将产生大量并发连接，这时一定要控制bulkload的并发数目，避免耗尽GTM的连接资源。如果目标表为复制表(DISTRIBUTE BY REPLICATION)时下推将不能进行。当数据量较大时，这对数据库将是个灾难。除了性能问题之外，空间也可能会剧烈膨胀，在导入结束后，需要用vacuum full来恢复。最好的方式还是如上建议的，不要在bulkload的场景中产生默认序列值。

另外，序列创建后，在每个节点上都维护了一张单行表，存储序列的定义及当前值，但此当前值并非GTM上的当前值，只是保存本节点与GTM交互后的状态。如果其他节点也向GTM申请了新值，或者调用了Setval修改了序列的状态，不会刷新本节点的单行表，但因每次申请序列值是向GTM申请，所以对序列正确性没有影响。

## 3.8 创建和管理视图

视图允许用户保存常用的或者复杂的查询。视图在磁盘上并没有被物理存储，当用户访问视图时查询会作为一个子查询运行。数据库中仅存放视图的定义，而不存放视图对应的数据，这些数据仍存放在原来的基本表中。若基本表中的数据发生变化，从视图中查询出的数据也随之改变。从这个意义上讲，视图就像一个窗口，透过它可以看到数据库中用户感兴趣的数据及变化。视图每次被引用的时候都会运行一次。

### 创建视图

使用CREATE VIEW命令创建新视图。

```
CREATE OR REPLACE VIEW MyView AS SELECT * FROM tpcds.customer WHERE c_customer_sk < 150;
```

#### 📖 说明

CREATE VIEW中的OR REPLACE可有可无，当存在OR REPLACE时，表示若以前存在该视图就进行替换。

### 查看视图

- 查看MyView视图，查询结果为当前实时数据。

```
SELECT * FROM myview;
```

- 查看当前用户下的视图。

```
SELECT * FROM user_views;
```

- 查看所有视图。  
`SELECT * FROM dba_views;`
- 查看某视图的具体信息。  
执行如下命令查询dba\_users视图的详细信息。

```
\d+ dba_users
View "PG_CATALOG.DBA_USERS"
Column |      Type      | Modifiers | Storage | Description
-----+-----+-----+-----+-----
USERNAME | CHARACTER VARYING(64) |          | extended |
View definition:
SELECT PG_AUTHID.ROLNAME::CHARACTER VARYING(64) AS USERNAME
FROM PG_AUTHID;
```

## 重建视图

使用ALTER VIEW命令在不键入查询语句的情况下重建视图。

```
ALTER VIEW myview REBUILD;
```

## 删除视图

使用DROP VIEW命令删除一个视图。

```
DROP VIEW myview;
```

DROP VIEW ... CASCADE命令也可以级联删除依赖此视图的对象。例如，如果A视图依赖于将要被删除的B视图，那么A视图也将被删除。如果没有CASCADE选项，这个DROP VIEW命令将会失败。

## 3.9 创建和管理定时任务

当客户在使用数据库过程中，如果白天执行一些耗时比较长的任务（例如：统计数据汇总之类或从其他数据库同步数据的任务），会对正常的业务有性能影响，所以客户经常选择在晚上执行，无形中增加了客户的工作量。因此数据库兼容Oracle数据库中定时任务的功能，可以由客户创建定时任务，当任务时间点到达后可以自动触发任务的执行，从而可以减少客户运维的工作量。

数据库兼容Oracle定时任务功能主要通过DBMS.JOB高级包提供的接口，可以实现定时任务的创建、任务到期自动执行、任务删除、修改任务属性（包括：任务id、任务的关闭开启、任务的触发时间、触发时间间隔、任务内容等）。

### 📖 说明

实时数仓（单机部署）暂不支持定时任务功能。

## 定时任务管理

### 步骤1 创建测试表：

```
CREATE TABLE test(id int, time date);
```

当结果显示为如下信息，则表示创建成功。

```
CREATE TABLE
```

### 步骤2 创建自定义存储过程：

```
CREATE OR REPLACE PROCEDURE PRC_JOB_1()
AS
N_NUM integer :=1;
```



- 修改JOB的Interval参数信息

--修改Job1的Interval为每隔1小时执行一次。

```
call dbms_job.interval(1,'sysdate + 1.0/24');
interval
-----
(1 row)
```

- 修改JOB的What参数信息

--修改Job1的What为执行SQL语句 “insert into public.test values(333, sysdate +5);”。

```
call dbms_job.what(1,'insert into public.test values(333, sysdate+5);');
what
-----
(1 row)
```

- 同时修改JOB的Next\_date、Interval、What等多个参数信息

```
call dbms_job.change(1, 'call public.prc_job_1();', sysdate, 'interval "1 minute"');
change
-----
(1 row)
```

#### 步骤8 删除JOB

```
call dbms_job.remove(1);
remove
-----
(1 row)
```

#### 步骤9 JOB的权限控制

- 当创建一个JOB时，该JOB会和创建该JOB的数据库和用户绑定（即：pg\_job系统视图新增的JOB记录中的dbname和log\_user）。
- 如果当前用户是DBA用户、系统管理员、该JOB的创建用户（即：pg\_job中的log\_user），那么该用户有权限通过高级包接口remove、change、next\_data、what、interval删除或修改JOB的参数信息。否则，会提示当前用户没有权限操作该JOB。
- 如果当前数据库是该JOB创建所属的数据库（即：为pg\_job系统视图中的dbname），那么连接到当前数据库上可以通过高级包接口remove、change、next\_data、what、interval删除或修改JOB的参数信息。
- 当删除JOB所属的数据库（即：为pg\_job系统视图中的dbname）时，系统会关联删除该数据库从属的JOB记录。
- 当删除JOB所属的用户（即：为pg\_job系统视图中的log\_user）时，系统会关联删除该用户从属的JOB记录。

----结束

## 3.10 查看系统表

除了创建的表以外，数据库还包含很多系统表。这些系统表包含集群安装信息以及 GaussDB(DWS)上运行的各种查询和进程的信息。可以通过查询系统表来收集有关数据库的信息。

“[系统表](#)和[系统视图](#)”中每个表的说明指出了表是对所有用户可见还是只对初始化用户可见。必须以初始化用户身份登录才能查询只对初始化用户可见的表。

## 查看数据库中包含的表

例如，在PG\_TABLES系统表中查看public schema中包含的所有表。

```
SELECT distinct(tablename) FROM pg_tables WHERE SCHEMANAME = 'public';
```

结果类似如下这样：

tablename
err_hr_staffs
test
err_hr_staffs_ft3
web_returns_p1
mig_seq_table
films4

(6 rows)

## 查看数据库用户

通过PG\_USER可以查看数据库中所有用户的列表，还可以查看用户ID（USESYSID）和用户权限。

```
SELECT * FROM pg_user;
```

username	usesysid	usecreatedb	usesuper	usecatupd	userepl	passwd	valbegin	valuntil	respool	parent	spacelimit	useconfig	nodegroup	tempspacelimit	spillspacelim	
Ruby	10	t	t	t	t	*****									default_pool	0
dbadmin	16393	f	f	f	f	*****									default_pool	0
lily	16691	f	f	f	f	*****									default_pool	0
jack	70694	f	f	f	f	*****									default_pool	0

(4 rows)

GaussDB(DWS)在内部使用Ruby执行日常管理和维护任务。可以向SELECT语句添加WHERE usesysid > 10来筛选查询，使其只显示用户定义的用户名称。

```
SELECT * FROM pg_user WHERE usesysid > 10;
```

username	usesysid	usecreatedb	usesuper	usecatupd	userepl	passwd	valbegin	valuntil	respool	parent	spacelimit	useconfig	nodegroup	tempspacelimit	spillspacelim	
dbadmin	16393	f	f	f	f	*****									default_pool	0
lily	16691	f	f	f	f	*****									default_pool	0
jack	70694	f	f	f	f	*****									default_pool	0

(3 rows)

## 查看和停止正在运行的查询语句

通过视图PG\_STAT\_ACTIVITY可以查看正在运行的查询语句。方法如下：

### 步骤1 设置参数track\_activities为on。

```
SET track_activities = on;
```

当此参数为on时，数据库系统才会收集当前活动查询的运行信息。



**步骤2** 查看正在运行的查询语句。以查看正在运行的查询语句所连接的数据库名、执行查询的用户、查询状态及查询对应的PID为例：

```
SELECT datname, username, state, pid FROM pg_stat_activity;
datname | username | state | pid
-----+-----+-----+-----
gaussdb | Ruby    | active | 140298793514752
gaussdb | Ruby    | active | 140298718004992
gaussdb | Ruby    | idle   | 140298650908416
gaussdb | Ruby    | idle   | 140298625742592
gaussdb | dbadmin | active | 140298575406848
(5 rows)
```

如果state字段显示为idle，则表明此连接处于空闲，等待用户输入命令。

如果仅需要查看非空闲的查询语句，则使用如下命令查看：

```
SELECT datname, username, state FROM pg_stat_activity WHERE state != 'idle';
```

**步骤3** 若需要取消运行时间过长的查询，通过PG\_TERMINATE\_BACKEND函数，根据线程ID结束会话。

```
SELECT PG_TERMINATE_BACKEND(139834759993104);
```

显示类似如下信息，表示结束会话成功。

```
PG_TERMINATE_BACKEND
-----
t
(1 row)
```

显示类似如下信息，表示用户执行了结束当前会话的操作。

```
FATAL: terminating connection due to administrator command
FATAL: terminating connection due to administrator command
```

### 📖 说明

gsq客户端使用PG\_TERMINATE\_BACKEND函数结束当前会话后台线程时，客户端不会退出而是自动重连。即还会返回“The connection to the server was lost. Attempting reset: Succeeded.”

```
FATAL: terminating connection due to administrator command
FATAL: terminating connection due to administrator command
The connection to the server was lost. Attempting reset: Succeeded.
```

----结束

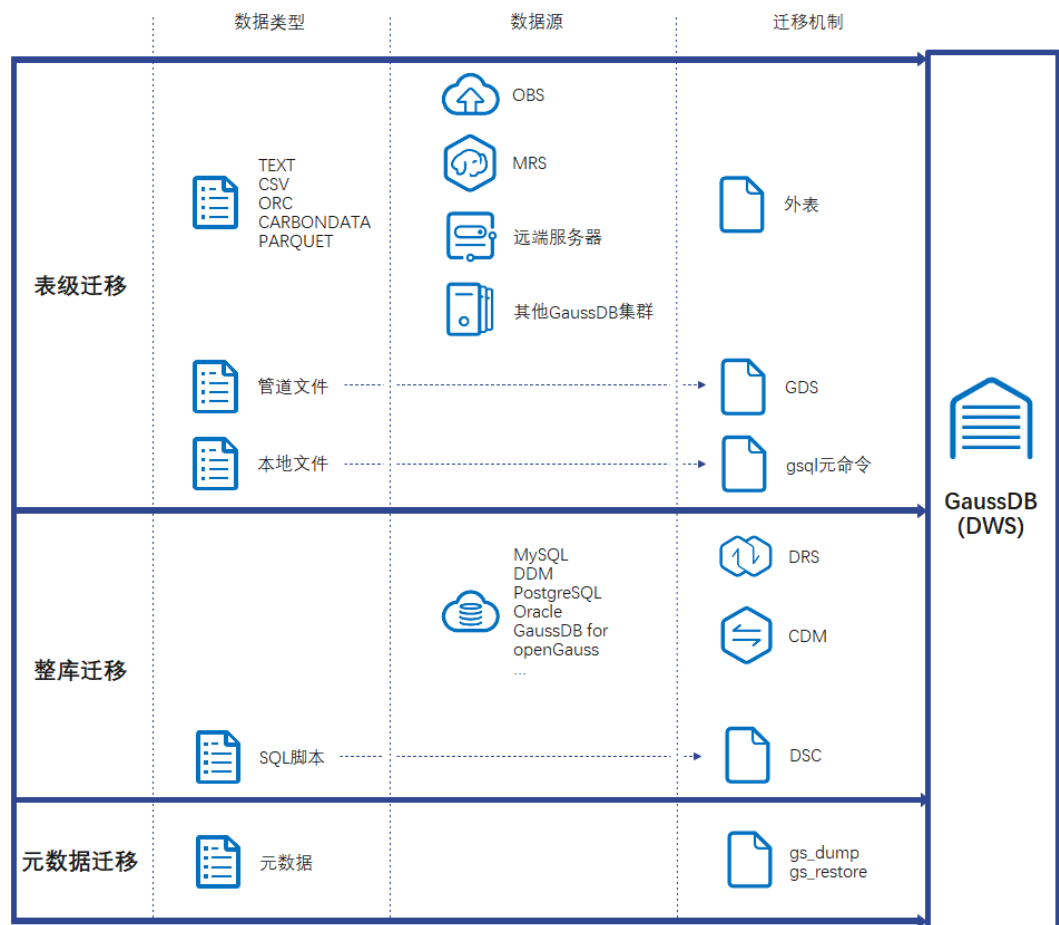
# 4 数据迁移

---

## 4.1 迁移数据到 GaussDB(DWS)

GaussDB(DWS)提供了灵活的数据入库方式，可以将多种数据源的数据导入到 GaussDB(DWS)中。各导入方式具有不同的特点，如[表4-1](#)所示，用户可以根据其特点自行选择。建议用户配合数据复制服务（Data Replication Service，简称DRS）、云数据迁移（Cloud Data Migration，简称CDM）和数据湖工厂（Data Lake Factory，简称DLF）一起使用，DRS用于数据实时同步，CDM用于批量数据迁移，DIS用于流数据接入，DLF可以对整个ETL过程进行编排调度，同时提供可视化的开发环境。

图 4-1 数据迁移示意图



**说明**

- DRS、CDM、OBS、MRS为云服务。
- GDS、DSC、gs\_restore、gs\_dump为内部工具。

表 4-1 数据导入方式说明

数据导入方式	数据源	说明	优势
使用外表从OBS并行导入数据	OBS	支持将存储在OBS上的TXT、CSV、ORC及CARBONDATA格式的数据并行导入到GaussDB(DWS)，支持导入后查询数据，也支持远程读OBS上的数据。 GaussDB(DWS)优先推荐的导入方式。	并行拉取方式，性能好，横向扩展。

数据导入方式	数据源	说明	优势
<a href="#">使用GDS从远端服务器导入数据</a>	Servers (即远端服务器)	使用GaussDB(DWS)提供的GDS工具, 利用多DN并行的方式, 将数据从远端服务器导入到GaussDB(DWS)。这种方式导入效率高, 适用于大批量数据入库。	
<a href="#">从MRS导入数据到集群</a>	MRS (HDFS)	配置一个GaussDB(DWS)集群连接到一个MRS集群, 然后将数据从MRS的HDFS中读取到GaussDB(DWS)。	并行拉取方式, 性能好, 横向扩展。
<a href="#">从GaussDB(DWS)集群导入数据到新集群</a>	-	支持两个GaussDB(DWS)集群之间的数据互访互通。通过Foreign Table方式实现跨DWS集群的数据访问和导入。	适用于多套DWS集群之间的数据同步。
<a href="#">基于GDS的跨集群互联互通</a>	-	通过GDS进行数据中转, 实现多个集群之间的数据同步。	适用于多套DWS集群之间的数据同步。
<a href="#">使用gsql元命令导入数据</a>	本地文件	与直接使用SQL语句COPY不同, 该命令读取/写入的文件只能是gsql客户端所在机器上的本地文件。	操作简单, 适用于小批量数据入库。
<a href="#">使用COPY FROM STDIN导入数据</a>	其他文件或数据库	使用Java语言开发应用程序时, 通过调用JDBC驱动的CopyManager接口, 从文件或其他数据库向GaussDB(DWS)写入数据。	从其他数据库直接写入GaussDB(DWS)的方式, 具有业务数据无需落地成文件的优势。
<a href="#">使用DRS将数据导入GaussDB(DWS)</a>	<ul style="list-style-type: none"> <li>• MySQL</li> <li>• DDM</li> <li>• PostgreSQL (公测)</li> <li>• Oracle (公测)</li> <li>• GaussDB分布式版 (公测)</li> </ul>	通过DRS实时同步功能, 将数据从一个数据源复制到GaussDB(DWS)数据仓库, 实现关键业务的数据实时流动。主要聚焦于表和数据的同步导入。	数据源丰富, 操作简单。

数据导入方式	数据源	说明	优势
<a href="#">使用CDM迁移数据到GaussDB(DWS)</a>	数据库、NoSQL、文件系统、大数据平台	CDM提供同构/异构数据源之间批量数据迁移的功能，帮助您实现从多种类型的数据源迁移数据到GaussDB(DWS)。CDM在迁移数据到GaussDB(DWS)时，采用的是Copy方式和GDS并行导入方式。	数据源丰富，操作简单。
<a href="#">使用DSC工具迁移SQL脚本</a>	数据库、NoSQL、文件系统、大数据平台	请参考第三方ETL工具的相关文档。 GaussDB(DWS)提供了DSC工具，可以将Teradata/Oracle脚本迁移到GaussDB(DWS)。	通过OBS中转，数据源丰富，数据转换能力强。
<a href="#">使用gs_dump和gs_dumpall命令导出元数据</a>	<ul style="list-style-type: none"> <li>• 纯文本格式</li> <li>• 自定义归档格式</li> <li>• 目录归档格式</li> <li>• tar归档格式</li> </ul>	gs_dump支持导出单个数据库或其内的对象，而gs_dumpall支持导出集群中所有数据库或各库的公共全局对象。 通过导入工具将导出的元数据信息导入至需要的数据库，可以完成数据库信息的迁移。	适用于元数据迁移。
<a href="#">使用gs_restore命令导入数据</a>	sql/tmp/tar文件格式	在数据库迁移场景下，支持使用gs_restore工具将事先使用gs_dump工具导出的文件格式，重新导入GaussDB(DWS)集群，实现表定义、数据库对象定义等元数据的导入。导入数据主要包括以下内容： <ul style="list-style-type: none"> <li>• 所有数据库对象定义。</li> <li>• 单个数据库对象定义。</li> <li>• 单个schema定义。</li> <li>• 单张表定义。</li> </ul>	

## 4.2 导入数据

### 4.2.1 使用外表从 OBS 并行导入数据

### 4.2.1.1 关于 OBS 并行导入

对象存储服务OBS ( Object Storage Service ) 是云上提供的一个基于对象的海量存储服务，为客户提供安全、高可靠、低成本的数据存储能力。OBS为用户提供了超大存储容量的能力，适合存放任意类型的文件。

数据仓库服务GaussDB(DWS)使用OBS作为集群数据与外部数据互相转化的平台，实现安全、高可靠和低成本存储需求。

GaussDB(DWS)支持通过外表将OBS上TXT、CSV、ORC、PARQUET、CARBONDATA以及JSON格式的数据导入到集群进行查询，也支持远程读OBS上的数据。因此对于经常查询的热数据建议直接导入GaussDB(DWS)后再做查询。偶尔查询的冷数据可以存储在OBS上直接远程读以节省成本。

目前，有两种外表可供OBS并行导入使用：

- 方式一：外表使用SERVER的FOREIGN DATA WRAPPER类型为dist\_fdw，具体可参考[CREATE FOREIGN TABLE \(OBS导入导出\)](#)章节。这种外表支持错误表，具有一定的容错能力，但只支持TXT、CSV格式的数据，参见[从OBS导入CSV、TEXT数据（方式一）](#)。
- 方式二：外表使用SERVER的FOREIGN DATA WRAPPER类型为dfs\_fdw，具体可参考[CREATE FOREIGN TABLE \(SQL on OBS or Hadoop\)](#)章节。这种外表查询性能较好，但不支持错误表，支持ORC、PARQUET、CARBONDATA、TXT、CSV以及JSON格式的数据，参见[从OBS导入ORC，PARQUET，CARBONDATA等格式数据（方式二）](#)。

#### 须知

- OBS导入导出数据时，不支持中文路径。
- OBS导入导出数据时，暂不支持跨Region进行OBS数据导入导出，必须确保OBS和DWS集群在同一个Region中。

## 概述

在数据迁移、ETL ( Extract-Transform-Load ) 过程中，需要向GaussDB(DWS)并行导入海量数据，使用普通方式会耗费大量的时间。GaussDB(DWS)提供了OBS ( Object Storage Service ) 及外表接口，通过OBS外表设置的导入URL路径、导入数据格式等信息来识别数据源文件，利用多DN ( Datanode ) 并行的方式，实现了数据的快速并行导入。

#### 优势：

- CN只负责任务的规划及下发，把数据导入的工作交给了DN，释放了CN的资源，使其有能力处理外部请求。
- 通过让各个DN都参与数据导入，充分利用各个设备的计算能力及网络带宽。
- 支持导入过程中对数据做预处理。
- 支持在导入过程中，针对数据格式错误设置导入容错性，并可在导入结束后根据错误信息定位错误数据。

#### 劣势：

需要创建OBS外表，并且要在OBS服务器上存放导入数据。

### 适用场景：

高并发、大数据量导入。

## 相关概念

- **数据源文件**：存储有数据的TEXT、CSV、ORC、PARQUET、CARBONDATA、JSON文件。文件中保存的是待并行导入数据库的数据。
- **OBS**：对象存储服务，是一种可存储文档、图片、音视频等非结构化数据的云存储服务。向GaussDB(DWS)并行导入数据时，数据对象放置在OBS服务器上。
- **桶 ( Bucket )**：对OBS中的一个存储空间的形象称呼，是存储对象的容器。
  - 对象存储是一种非常扁平化的存储方式，桶中存储的对象都在同一个逻辑层级，去除了文件系统中的多层级树形目录结构。
  - 在OBS中，桶名必须是全局唯一的且不能修改，即用户创建的桶不能与自己已创建的其他桶名称相同，也不能与其他用户创建的桶名称相同。每个桶在创建时都会生成默认的桶ACL ( Access Control List )，桶ACL列表的每项包含了对被授权用户授予什么样的权限，如读取权限、写入权限、完全控制权限等。用户只有对桶有相应的权限，才可以对桶进行操作，如创建、删除、显示、设置桶ACL等。
  - 一个用户最多可创建100个桶，但每个桶中存放的总数据容量和对象/文件数量没有限制。
- **对象**：是存储在OBS中的基本数据单位。用户上传的数据以对象的形式存储在OBS的桶中。对象的属性包括名称Key，Metadata，Data。

通常，将对象等同于文件来进行管理，但是由于OBS是一种对象存储服务，并没有文件系统中的文件和文件夹概念。为了使用户更方便进行管理数据，OBS提供了一种方式模拟文件夹。通过在对象的名称中增加“/”，如tpcds1000/stock.csv，tpcds1000可以等同于文件夹，stock.csv就可以等同于文件名，而对象名称 ( key ) 仍然是tpcds1000/stock.csv、对象的内容就是stock.csv数据文件的内容。
- **Key**：对象的名称 ( 键 )，为经过UTF-8编码的长度大于0且不超过1024的字符序列，一个桶里的每个对象必须拥有唯一的对象键值。用户可使用桶名+对象名来存储和获取对应的对象。
- **Metadata**：对象元数据，用来描述对象的信息。元数据又可分为系统元数据和用户元数据。这些元数据以键值对 ( Key-value ) 的形式随http头域一起上传到OBS系统。
  - 系统元数据由OBS系统产生，在处理对象数据时使用。系统元数据包括：Date, Content-length, last-modify, Content-MD5等。
  - 用户元数据由用户上传对象时指定，是用户自己对对象的一些描述信息。
- **Data**：对象的数据内容，OBS对于数据的内容是无感知的，即认为对象内的数据为无状态的二进制数据。
- **数据库普通表**：数据库中的普通表，数据源文件中的数据最终并行导入到这些表中存储，包括行存表、列存表。
- **外表**：用于识别数据源文件中的数据。外表中保存了数据源文件的位置、文件格式、编码格式、数据间的分隔符等信息。

## 导入数据原理

OBS导入原理如图4-2所示，CN负责任务的规划及下发，它是按文件给每个DN节点分配任务的。

分配算法如下：

例如，图4-2中，总共有4个节点DN0~DN3， OBS上有6个文件t1.data.0~t1.data.5，那么分配方式如下：

t1.data.0 -> DN0

t1.data.1 -> DN1

t1.data.2 -> DN2

t1.data.3 -> DN3

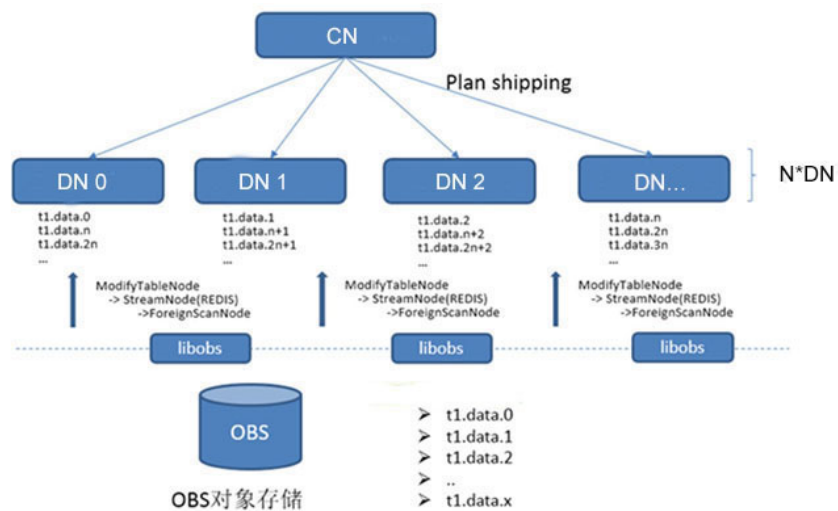
t1.data.4 -> DN0

t1.data.5 -> DN1

其中DN0 和DN1上分配了两个文件，其他DN分配了1个文件。

如果OBS上文件大小都相同时，OBS上的文件数与DN节点数的比例为1:1时导入性能为最好，因为每个DN分配的任务都相同。因此建议将数据文件存储到OBS前，尽可能均匀地将文件切分成多个，文件的数量以DN的整数倍更适合。

图 4-2 通过 OBS 外表并行导入数据





## 导入流程图

图 4-3 并行导入流程

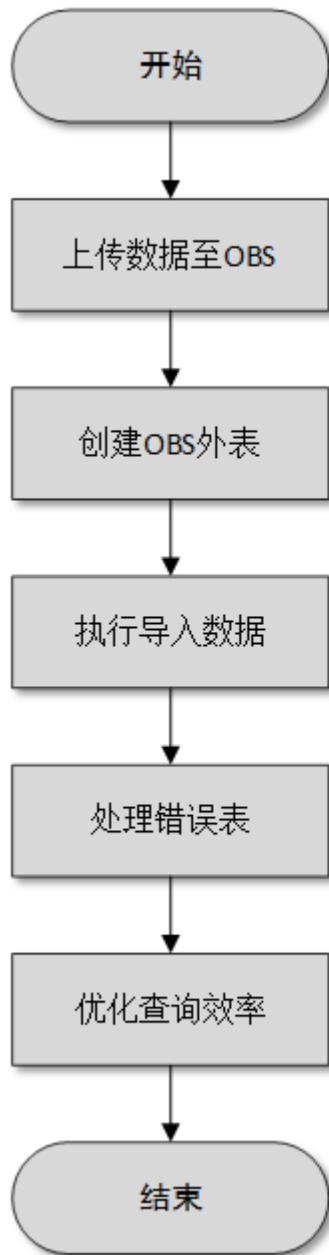


表 4-2 流程说明

流程	说明	子任务
上传数据至 OBS。	在 OBS 服务器上规划存储路径，并上传数据文件。 详细请参见 <a href="#">上传数据到 OBS</a> 。	-

流程	说明	子任务
创建OBS外表。	创建外表用于识别OBS服务器上的数据源文件。在OBS外表中保存了数据源在OBS服务器上的桶名、对象名，文件格式、存放位置、编码格式、数据间的分隔符等信息。 详细请参见 <a href="#">创建OBS外表</a> 。	-
执行导入数据。	在创建好外表后，通过INSERT语句，将数据快速、高效地导入到目标表中。 详细请参见 <a href="#">执行导入数据</a> 。	-
处理错误表。	在数据并行导入发生错误时，请根据错误信息， <a href="#">处理错误表</a> ，以保证导入数据的完整性。 <b>说明</b> <a href="#">使用外表从OBS并行导入数据</a> 中方式二的导入方式，不支持错误表。	-
优化查询效率。	导入数据后，通过ANALYZE语句生成表统计信息。ANALYZE语句会将统计结果自动存储在系统表PG_STATISTIC中。执行计划生成器会使用这些统计数据，以生成最有效的查询执行计划。	-

## 4.2.1.2 从 OBS 导入 CSV、TEXT 数据（方式一）

### 4.2.1.2.1 创建访问密钥（AK 和 SK）

在本示例中，将导入OBS数据到GaussDB(DWS)集群数据库中。云平台用户通过客户端或API、SDK等方式访问OBS时，需要通过AK/SK认证方式进行认证鉴权。因此，当您需要通过客户端或JDBC/ODBC应用程序等方式连接GaussDB(DWS)数据库访问OBS时，必须先获取访问密钥（AK和SK）。

- Access Key Id（AK）：访问密钥ID。与私有访问密钥关联的唯一标识符；访问密钥ID和私有访问密钥一起使用，对请求进行加密签名。
- Secret Access Key（SK）：与访问密钥ID结合使用的密钥，对请求进行加密签名，可标识发送方，并防止请求被修改。

### 创建访问密钥（AK 和 SK）

在创建访问密钥前，请确保登录控制台的帐号已通过实名认证。

通过管理控制台创建访问密钥（AK和SK）操作步骤如下：

**步骤1** 登录GaussDB(DWS) 管理控制台。

**步骤2** 单击右上角用户名，在下拉菜单中单击“我的凭证”。

**步骤3** 在左侧导航树单击“访问密钥”。

如果在访问密钥列表中已经有访问密钥了，可以直接使用现有的访问密钥。但是，在访问密钥列表中只能查看到访问密钥ID（即Access Key ID），只有在新增访问密钥时，用户才可以下载到含有Access Key ID和Secret Access Key的密钥文件。如果您没有该密钥文件，可以单击“新增访问密钥”重新创建。

#### 说明

- 每个用户最多可创建两个有效的访问密钥，如果当前已存在2个访问密钥，只能先删除现有的访问密钥，然后再重新创建。删除时，需要输入当前用户的登录密码、邮箱或手机短信的验证码，验证通过才能成功删除。
- 为了账号安全性，建议您定期更换并妥善保存访问密钥。

**步骤4** 单击“新增访问密钥”。

**步骤5** 在弹出的对话框中，输入登录密码和对应验证码，然后单击“确定”。

#### 说明

- 用户如果未绑定邮箱和手机，则只需输入登录密码。
- 用户如果同时绑定了邮箱和手机，可以选择其中一种方式进行验证。

**步骤6** 在弹出的“下载确认”提示框中，单击“确定”后，密钥会直接保存到浏览器默认的下载文件夹中。

#### 说明

- 为防止访问密钥泄露，建议您将其保存到安全的位置。
- 如果用户在此提示框中单击“取消”，则不会下载密钥，后续也将无法重新下载。如果需要使用访问密钥，可以重新创建新的访问密钥。

**步骤7** 打开下载下来的“credentials.csv”文件即可获取到访问密钥（AK和SK）。

----结束

## 注意事项

当用户发现访问密钥被异常使用（包括丢失，泄露等情况），或不再使用访问密钥时，建议在访问密钥列表中立即删除密钥或者通知管理员重置相关密钥。

删除访问密钥时，需要输入登录密码和邮箱或者手机验证码进行验证。

#### 说明

删除的访问密钥将永久删除且无法恢复。

## 4.2.1.2.2 上传数据到 OBS

### 操作场景

从OBS导入数据到集群之前，需要提前准备数据源文件，并将数据源文件上传到OBS。如果您的数据文件已经在OBS上了，则只需完成[上传数据到OBS](#)中的[步骤2~步骤3](#)。

### 准备数据文件

准备需要上传到OBS的CSV或TEXT数据源文件。

如果用户数据无法以CSV格式保存，可以选择以文本类型保存为其他任意格式后缀的文件。

#### 说明

根据[导入数据原理](#)，当数据源文件的数据量较大时，将数据文件存储到OBS前，尽可能均匀地将文件切分成多个，文件数量为DataNode的整数倍时，导入性能更好。

假设您已将3个CSV数据文件存储在OBS上，其原始数据分别如下：

- 数据文件 “product\_info.0”

示例数据如下所示：

```
100,XHDK-A-1293-#fJ3,2017-09-01,A,2017 Autumn New Shirt Women,red,M,328,2017-09-04,715,good!
205,KDKE-B-9947-#kL5,2017-09-01,A,2017 Autumn New Knitwear Women,pink,L,584,2017-09-05,406,very good!
300,JODL-X-1937-#pV7,2017-09-01,A,2017 autumn new T-shirt men,red,XL,1245,2017-09-03,502,Bad.
310,QQPX-R-3956-#aD8,2017-09-02,B,2017 autumn new jacket women,red,L,411,2017-09-05,436,It's really super nice.
150,ABEF-C-1820-#mC6,2017-09-03,B,2017 Autumn New Jeans Women,blue,M,1223,2017-09-06,1200,The seller's packaging is exquisite.
```

- 数据文件 “product\_info.1”

示例数据如下所示：

```
200,BCQP-E-2365-#qE4,2017-09-04,B,2017 autumn new casual pants men,black,L,997,2017-09-10,301,The clothes are of good quality.
250,EABE-D-1476-#oB1,2017-09-10,A,2017 autumn new dress women,black,S,841,2017-09-15,299,Follow the store for a long time.
108,CDXK-F-1527-#pL2,2017-09-11,A,2017 autumn new dress women,red,M,85,2017-09-14,22,It's really amazing to buy.
450,MMCE-H-4728-#nP9,2017-09-11,A,2017 autumn new jacket women,white,M,114,2017-09-14,22,Open the package and the clothes have no odor.
260,OCDA-G-2817-#bD3,2017-09-12,B,2017 autumn new woolen coat women,red,L,2004,2017-09-15,826,Very favorite clothes.
```

- 数据文件 “product\_info.2”

示例数据如下所示：

```
980,"ZKDS-J",2017-09-13,"B","2017 Women's Cotton Clothing","red","M",112,,
98,"FKQB-I",2017-09-15,"B","2017 new shoes men","red","M",4345,2017-09-18,5473
50,"DMQY-K",2017-09-21,"A","2017 pants men","red","37",28,2017-09-25,58,"good","good"
80,"GKLW-L",2017-09-22,"A","2017 Jeans Men","red","39",58,2017-09-25,72,"Very comfortable."
30,"HWEC-L",2017-09-23,"A","2017 shoes women","red","M",403,2017-09-26,607,"good!"
40,"IQPD-M",2017-09-24,"B","2017 new pants Women","red","M",35,2017-09-27,52,"very good."
50,"LPEC-N",2017-09-25,"B","2017 dress Women","red","M",29,2017-09-28,47,"not good at all."
60,"NQAB-O",2017-09-26,"B","2017 jacket women","red","S",69,2017-09-29,70,"It's beautiful."
70,"HWNB-P",2017-09-27,"B","2017 jacket women","red","L",30,2017-09-30,55,"I like it so much"
80,"JKHU-Q",2017-09-29,"C","2017 T-shirt","red","M",90,2017-10-02,82,"very good."
```

## 上传数据到 OBS

### 步骤1 上传数据到OBS。

将待导入的数据源文件存储在OBS桶中。

1. 登录OBS管理控制台。  
单击“服务列表”，选择“对象存储服务”，打开OBS管理控制台页面。
2. 创建桶。  
如何创建OBS桶，具体请参见《对象存储服务控制台指南》中的[创建桶](#)章节。  
例如，创建以下两个桶：“mybucket”和“mybucket02”。
3. 新建文件夹。  
具体请参见《对象存储服务控制台指南》中的[新建文件夹](#)章节。  
例如：
  - 在已创建的OBS桶“mybucket”中新建一个文件夹“input\_data”。
  - 在已创建的OBS桶“mybucket02”中新建一个文件夹“input\_data”。
4. 上传文件。  
具体请参见《对象存储服务控制台指南》的[上传文件](#)章节。  
例如：
  - 将以下数据文件上传到OBS桶“mybucket”的“input\_data”目录中。  
product\_info.0  
product\_info.1
  - 将以下数据文件上传到OBS桶“mybucket02”的“input\_data”目录中。  
product\_info.2

### 步骤2 获取数据源文件的OBS路径。

数据源文件在上传到OBS桶之后，会生成全局唯一的访问路径。数据源文件的OBS路径用于创建外表时location参数设置。

location参数中OBS文件的路径由“obs://”、桶名和文件路径组成，即为：obs://<bucket\_name>/<file\_path>/

例如，在本例中，location参数中数据文件的OBS路径分别为：

```
obs://mybucket/input_data/product_info.0
obs://mybucket/input_data/product_info.1
obs://mybucket02/input_data/product_info.2
```

### 步骤3 为导入用户设置OBS桶的读取权限。

在从OBS导入数据到集群时，执行导入操作的用户需要取得数据源文件所在OBS桶的读取权限。通过配置桶的ACL权限，可以将读取权限授予指定的用户帐号。

具体请参见《对象存储服务控制台指南》中的[配置桶ACL](#)章节。

----结束

## 4.2.1.2.3 创建 OBS 外表

### 操作步骤

- 步骤1** 根据[上传数据到OBS](#)中规划的路径，由此确定创建外表时使用的参数location的值。

**步骤2** 用户获取OBS访问协议对应的AK值和SK值。获取访问密钥，请登录管理控制台，将鼠标移至右上角的用户名，单击“我的凭证”，然后在左侧导航树单击“访问密钥”。在访问密钥页面，可以查看已有的访问密钥ID（即AK），如果要同时获取AK和SK，可以单击“新增访问密钥”创建并下载访问密钥。

**步骤3** 梳理待导入数据的格式信息，确定创建外表时使用的数据格式参数的值。需要收集的主要数据源格式信息如下：

- **format**: 外表中数据源文件的格式。OBS外表导入支持CSV、TEXT格式。缺省值为TEXT。
- **header**: 指定导出数据文件是否包含标题行，header只能用于CSV格式的文件中。
- **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。
- 外表可以识别的更多参数，详细使用请参见数据格式参数。

**步骤4** 规划并行导入容错性，以控制导入过程中处理错误的方式。

- **fill\_missing\_fields**: 数据入库时，数据源文件中某行的最后一个字段缺失时，请选择是直接将该字段设为Null，还是在错误表中报错提示。
- **ignore\_extra\_data**: 数据源文件中的字段比外表定义列数多时，请选择是忽略多出的列，还是在错误表中报错提示。
- **per node reject limit**: 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实例上录入错误表中的错误数量超过设定值时，本次导入失败，报错退出。可以选择不做限制，也可以根据所能容忍的错误数量选择一个上限值。
- **compatible\_illegal\_chars**: 导入时遇到非法字符，选择如何处理。是将非法字符按照转换规则转换后入库，还是报错中止导入。

非法字符容错转换规则如下：

- 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。
- 对非法字符进行容错转换时，如遇NULL、DELIMITER、QUOTE、ESCAPE也设置成了空格或问号，GaussDB(DWS)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。
- **error\_table\_name**: 用于记录数据格式错误信息的错误表表名。并行导入结束后查询此错误信息表，能够获取详细的错误信息。
- 更多参数，详细使用请参见容错性参数。

**步骤5** 根据前面步骤确定的参数，**创建OBS外表**。外表的创建语法以及详细使用，请参考CREATE FOREIGN TABLE (OBS导入导出)。

----结束

## 示例

在GaussDB(DWS)数据库中，创建一个外表。参数信息如下所示：

- **数据格式参数访问密钥（AK和SK）**
  - 用户获取OBS访问协议对应的AK值（access\_key）。
  - 用户获取OBS访问协议对应的SK值（secret\_access\_key）。

### 📖 说明

请根据用户实际获取的access\_key和secret\_access\_key的密钥替换示例中的对应内容。

- **设置数据格式参数**
  - **数据源文件格式 (format)** 为CSV。
  - **编码格式 (encoding)** 为UTF-8。
  - **是否使用加密 (encrypt)**，默认为'off'。
  - **字段分隔符 (delimiter)** 为','。
  - **引号字符 (quote)** 使用默认值双引号。
  - **null (数据文件中空值的表示)** 为“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)** 为false，当数据文件第一行不是标题行 (即表头)，不需要设置。

### 📖 说明

OBS导出数据时不支持该参数为true，使用缺省值false。

- **设置导入时的容错性参数**
  - **PER NODE REJECT LIMIT 'value'** 为unlimited，即接受导入过程中所有数据格式错误。
  - **LOG INTO error\_table\_name**指定为product\_info\_err，将数据导入过程中出现的数据格式错误信息写入表product\_info\_err。
  - **fill\_missing\_fields**为true，即当数据加载时，若数据源文件中一行数据的最后一个字段缺失，则把最后一个字段的值设置为NULL，不报错。
  - **ignore\_extra\_data**为true，当数据加载时，若数据源文件比外表定义列数多，则忽略行尾多出来的列，不报错。

根据以上信息，创建的外表如下所示：

```
DROP FOREIGN TABLE product_info_ext;

CREATE FOREIGN TABLE product_info_ext
(
  product_price      integer      not null,
  product_id         char(30)     not null,
  product_time       date         ,
  product_level      char(10)     ,
  product_name       varchar(200) ,
  product_type1      varchar(20)  ,
  product_type2      char(10)     ,
  product_monthly_sales_cnt integer ,
  product_comment_time date      ,
  product_comment_num integer     ,
  product_comment_content varchar(200)
)
SERVER gsmpp_server
OPTIONS(

LOCATION 'gsobs://mybucket/input_data/product_info | gsobs://mybucket02/input_data/product_info',
FORMAT 'CSV' ,
DELIMITER ';;',
encoding 'utf8',
header 'false',
ACCESS_KEY 'access_key_value_to_be_replaced',
SECRET_ACCESS_KEY 'secret_access_key_value_to_be_replaced',
fill_missing_fields 'true',
ignore_extra_data 'true'
)
READ ONLY
```

```
LOG INTO product_info_err  
PER NODE REJECT LIMIT 'unlimited';
```

返回如下信息表示创建成功：

```
CREATE FOREIGN TABLE
```

#### 4.2.1.2.4 执行导入数据

### 背景信息

在执行数据导入前，您可以参考以下优秀实践方法进行合理的设计部署，最大化的使用系统资源，以提高数据导入性能。

- OBS的数据导入性能，多数场景受限于网络的并发访问速率，因此在OBS服务器上最好部署多个桶，使用多桶并发导入，提高DN数据传输利用率。
- 并发导入场景，与单表导入相似，至少应保证I/O性能大于网络最大速率。
- 配置GUC参数“[raise\\_errors\\_if\\_no\\_files](#)”、“[partition\\_mem\\_batch](#)”和“[partition\\_max\\_cache\\_size](#)”，设置导入时是否区分“导入文件记录数为空”和“导入文件不存在”、导入时的缓存个数以及数据缓存区大小。
- 若导入表存在索引，在数据导入过程中，将增量更新索引信息，影响数据导入性能。建议在执行数据导入前，先删除相关表的索引。在数据导入完成后，再重新创建索引。

### 操作步骤

**步骤1** 在GaussDB(DWS)数据库中，创建目标表，用于存储从OBS导入的数据。建表语法请参考CREATE TABLE。

目标表的表结构和OBS上将要导入的数据源文件的字段要保持一一对应，即字段个数、字段类型要一致。并且，目标表和创建的外表的表结构也要保持一致，字段名称可以不一样。

**步骤2** （可选）若导入表存在索引，在数据导入过程中，将增量更新索引信息，影响数据导入性能。建议在执行数据导入前，先删除相关表的索引。在数据导入完成后，再重新创建索引。

**步骤3** 执行数据导入。

```
INSERT INTO [目标表名] SELECT * FROM [foreign table 表名];
```

- 若出现以下类似信息，说明数据导入成功。请查询错误信息表，查看是否存在数据格式错误，详细操作请参见[处理错误表](#)。

```
INSERT 0 20
```

- 若出现数据加载错误，请参见[处理错误表](#)，并重新执行数据导入。

----结束

### 示例

创建一个名为product\_info的表，示例如下：

```
DROP TABLE IF EXISTS product_info;  
CREATE TABLE product_info  
(  
    product_price    integer    not null,  
    product_id       char(30)   not null,  
    product_time     date       ,
```



```

product_level      char(10)      ,
product_name       varchar(200)  ,
product_type1      varchar(20)   ,
product_type2      char(10)      ,
product_monthly_sales_cnt integer      ,
product_comment_time date          ,
product_comment_num integer        ,
product_comment_content varchar(200)
)
with (
orientation = column,
compression=middle
)
DISTRIBUTE BY HASH (product_id);

```

执行以下命令将外表`product_info_ext`的数据导入到目标表`product_info`中：

```
INSERT INTO product_info SELECT * FROM product_info_ext;
```

#### 4.2.1.2.5 处理错误表

### 操作场景

当数据导入发生错误时，请根据本文指引信息进行处理。

### 查询错误信息

数据导入过程中发生的错误，一般分为数据格式错误和非数据格式错误。

- 数据格式错误

在创建外表时，通过设置参数“LOG INTO error\_table\_name”，将数据导入过程中出现的数据格式错误信息写入指定的错误信息表`error_table_name`中。您可以通过以下SQL，查询详细错误信息。

```
SELECT * FROM error_table_name;
```

错误信息表结构如表4-3所示。

表 4-3 错误信息表

列名称	类型	描述
nodeid	integer	报错节点编号。
begintime	timestamp with time zone	出现数据格式错误的时间。
filename	character varying	出现数据格式错误的数据库源文件名。
rownum	bigint	在数据库源文件中，出现数据格式错误的行号。
rawrecord	text	在数据库源文件中，出现数据格式错误的原始记录。
detail	text	详细错误信息。

- 非数据格式错误

对于非数据格式错误，一旦发生将导致整个数据导入失败。您可以根据执行数据导入过程中，界面提示的错误信息，帮助定位问题，处理错误表。

## 处理数据导入错误

根据获取的错误信息，请对照下表，处理数据导入错误。

表 4-4 处理数据导入错误

错误信息	原因	解决办法
missing data for column "r_reason_desc"	<ol style="list-style-type: none"> <li>1. 数据源文件中的列数比外表定义的列数少。</li> <li>2. 对于TEXT格式的数据源文件，由于转义字符（\）导致delimiter（分隔符）错位或者quote（引号字符）错位造成的错误。 示例：目标表存在3列字段，导入的数据如下所示。由于存在转义字符“\”，分隔符“ ”被转义为第二个字段的字段值，导致第三个字段值缺失。 BE Belgium 1</li> </ol>	<ol style="list-style-type: none"> <li>1. 由于列数少导致的报错，选择下列办法解决： <ul style="list-style-type: none"> <li>• 在数据源文件中，增加列“r_reason_desc”的字段值。</li> <li>• 在创建外表时，将参数“fill_missing_fields”设置为“on”。即当导入过程中，若数据源文件中一行数据的最后一个字段缺失，则把最后一个字段的值设置为NULL，不报错。</li> </ul> </li> <li>2. 对由于转义字符导致的错误，需检查报错的行中是否含有转义字符（\）。若存在，建议在创建外表时，将参数“noescaping”（是否不对\和后面的字符进行转义）设置为true。</li> </ol>
extra data after last expected column	数据源文件中的列数比外表定义的列数多。	<ul style="list-style-type: none"> <li>• 在数据源文件中，删除多余的字段值。</li> <li>• 在创建外表时，将参数“ignore_extra_data”设置为“on”。即在导入过程中，若数据源文件比外表定义的列数多，则忽略行尾多出来的列。</li> </ul>
invalid input syntax for type numeric: "a"	数据类型错误。	在数据源文件中，修改输入字段的数据类型。根据此错误信息，请将输入的数据类型修改为numeric。
null value in column "staff_id" violates not-null constraint	非空约束。	在数据源文件中，增加非空字段信息。根据此错误信息，请增加“staff_id”列的值。

错误信息	原因	解决办法
duplicate key value violates unique constraint "reg_id_pk"	唯一约束。	<ul style="list-style-type: none"> <li>删除数据源文件中重复的行。</li> <li>通过设置关键字“DISTINCT”，从SELECT结果集中删除重复的行，保证导入的每一行都是唯一的。</li> </ul> <pre>INSERT INTO reasons SELECT DISTINCT * FROM foreign_tpcds_reasons;</pre>
value too long for type character varying(16)	字段值长度超过限制。	在数据源文件中，修改字段值长度。根据此错误信息，字段值长度限制为VARCHAR2(16)。

### 4.2.1.3 从 OBS 导入 ORC, PARQUET, CARBONDATA 等格式数据（方式二）

#### 4.2.1.3.1 OBS 上的数据准备

##### 操作场景

使用SQL on OBS功能查询OBS数据之前：

- 假设您已将ORC数据存储到OBS上。  
例如，在使用Hive或Spark等组件时创建了ORC表，其表数据已经存储在OBS上的场景。  
假设有2个ORC数据文件“product\_info.0”和“product\_info.1”，其原始数据如[原始数据](#)所示，都已经存储在OBS桶“mybucket”的“demo.db/product\_info\_orc/”目录中。
- 如果您的数据文件已经在OBS上了，请执行[获取源数据的OBS路径并设置读取权限](#)中的步骤。

##### 说明

本小节以导入ORC格式为例，PARQUET、CARBONDATA、JSON数据的导入方法与ORC格式相似。

该方式同样也支持导入TEXT、CSV格式文件，但由于该方式不支持错误表，因此推荐使用[从OBS导入CSV、TEXT数据（方式一）](#)的导入方式。

##### 原始数据

假设您已将2个ORC数据文件存储在OBS上，其原始数据分别如下：

- 数据文件“product\_info.0”
- 数据文件“product\_info.1”

## 获取源数据的 OBS 路径并设置读取权限

### 步骤1 登录OBS管理控制台。

单击“服务列表”，选择“对象存储服务”，打开OBS管理控制台页面。

### 步骤2 获取数据源文件的OBS路径。

数据源文件在上传到OBS桶之后，会生成全局唯一的访问路径。在创建外表时需要指定数据源文件的OBS路径。

如何查看OBS路径，请参见《对象存储服务控制台指南》的[通过对象URL访问对象](#)章节。

例如，在本例中，查看到数据文件的OBS路径分别为：

```
https://obs.ap-southeast-1.myhuaweicloud.com/mybucket/demo.db/product_info_orc/product_info.0  
https://obs.ap-southeast-1.myhuaweicloud.com/mybucket/demo.db/product_info_orc/product_info.1
```

### 步骤3 为用户设置OBS桶的读取权限。

在使用SQL on OBS功能时，执行该功能的用户需要取得数据源文件所在OBS桶的读取权限。通过配置桶的ACL权限，可以将读取权限授予指定的用户帐号。

具体请参见《对象存储服务控制台指南》中的[配置桶ACL](#)章节。

----结束

### 4.2.1.3.2 创建外部服务器

创建外部服务器，用于定义OBS服务器的信息，供外表调用。创建外部服务器的详细语法，请参见CREATE SERVER。

#### (可选) 新建用户及数据库并授予外表权限

如果您将使用普通用户在自定义数据库中创建外部服务器和外表，由于普通用户没有外表权限无法创建，所以，您必须参照以下步骤新建用户和数据库，并授予该用户外表权限。

以下示例，是新建一个普通用户dbuser并创建一个数据库mydatabase，然后使用管理员用户授予dbuser外表权限。

#### 步骤1 使用数据库管理员通过GaussDB(DWS)提供的数据库客户端连接默认数据库gaussdb。

例如，使用gsq客户端的用户执行下面命令连接数据库：

```
gsq -d gaussdb -h 192.168.2.30 -U dbadmin -p 8000 -W password -r
```

#### 步骤2 新建一个普通用户，并用它创建一个数据库。

新建一个具有创建数据库权限的用户dbuser：

```
CREATE USER dbuser WITH CREATEDB PASSWORD 'password';
```

切换为新建的用户：

```
SET ROLE dbuser PASSWORD 'password';
```

执行以下命令创建数据库：

```
CREATE DATABASE mydatabase;
```

查询数据库：

```
SELECT * FROM pg_database;
```

返回结果中有mydatabase的信息表示创建成功：

datname	datdba	encoding	datcollate	datctype	datistemplate	dataallowconn	datconnlimit	datacl
template1	10	0	C	C	t	t	-1	14146   1351
1663	ORA							{=c/Ruby,Ruby=CTc/Ruby}
template0	10	0	C	C	t	f	-1	14146   1350
1663	ORA							{=c/Ruby,Ruby=CTc/Ruby}
gaussdb	10	0	C	C	f	t	-1	14146   1352
1663	ORA							{=Tc/Ruby,Ruby=CTc/Ruby,chaojun=C/Ruby,huobinru=C/Ruby}
mydatabase	17000	0	C	C	f	t	-1	14146   1351
1663	ORA							

(4 rows)

**步骤3** 使用管理员用户给普通用户赋予创建外部服务器的权限和使用外表的权限。

使用数据库管理员用户通过GaussDB(DWS)提供的数据库客户端连接新建的数据库。

例如，使用gsq客户端的用户可以直接使用如下语句切换为管理员用户去连接新建的数据库：

```
\c mydatabase dbadmin,
```

根据提示输入管理员用户密码。

#### 说明

注意，必须先使用管理员用户连接到**将要创建外部服务器和使用外表的数据库**，再对普通用户进行授权。

默认只有系统管理员才可以创建外部服务器，普通用户需要授权才可以创建，执行以下命令授权：

```
GRANT ALL ON SCHEMA public TO dbuser;
GRANT ALL ON FOREIGN DATA WRAPPER dfs_fdw TO dbuser;
```

其中fdw\_name的名字可以是hdfs\_fdw或者dfs\_fdw，dbuser为创建SERVER的用户名。

执行以下命令赋予用户使用外表的权限。

```
ALTER USER dbuser USEFT;
```

查看用户：

```
SELECT r.rolname, r.rolsuper, r.rolinherit,
       r.rolcreatorole, r.rolcreatedb, r.rolcanlogin,
       r.rolconnlimit, r.rolvalidbegin, r.rolvaliduntil,
       ARRAY(SELECT b.rolname
             FROM pg_catalog.pg_auth_members m
             JOIN pg_catalog.pg_roles b ON (m.roleid = b.oid)
             WHERE m.member = r.oid) as memberof
       , r.rolreplication
       , r.rolauditadmin
       , r.rolsystemadmin
       , r.roluseft
FROM pg_catalog.pg_roles r
ORDER BY 1;
```

返回结果中dbuser的信息中包含了UseFT权限，表示授权成功：

rolname	rolsuper	rolinherit	rolcreatorole	rolcreatedb	rolcanlogin	rolconnlimit	rolvalidbegin	rolvaliduntil	memberof	rolreplication	rolauditadmin	rolsystemadmin	roluseft
dbuser													UseFT





- **表字段定义**
    - **column\_name**: 外表中的字段名。
    - **type\_name**: 字段的数据类型。多个字段用“,” 隔开。  
外表的字段个数和字段类型，需要与OBS上保存的数据完全一致。
  - **SERVER dfs\_server**

外表的外部服务器名称，这个server必须存在。外表通过设置外部服务器连接OBS读取数据。  
此处应填写为参照[创建外部服务器](#)创建的外部服务器名称。
  - **OPTIONS参数**

用于指定外表数据的各类参数，关键参数如下所示。

    - “format”：表示对应的OBS服务上的文件格式支持"orc"、"parquet"、"carbodata"、"text"、"csv"、"json"。
    - “foldername”：必选参数。数据源文件的OBS路径，此处仅需要填写“/桶名/文件夹目录层级”。可以先通过[OBS上的数据准备](#)中的2获取数据源文件的完整的OBS路径，该路径为OBS服务的终端节点（Endpoint）。
  - “totalrows”：可选参数。该参数不是导入的总行数。由于OBS上文件可能很多，做analyze可能会很慢，通过“totalrows”参数，让用户来设置一个预估的值，使优化器能通过这个值做大小表的估计。一般预估值与实际值的数量级差不多时，查询效率较高。
  - “encoding”：外表中数据源文件的编码格式名称，缺省为utf8。对于OBS外表此参数为必选项。
- **DISTRIBUTE BY:**

这个子句是必须的，对于OBS外表，当前只支持**ROUNDROBIN**分布方式。  
表示外表在从数据源读取数据时，GaussDB(DWS)集群每一个节点随机读取一部分数据，并组成完整数据。
- **语法中的其他参数**

其他参数均为可选参数，用户可以根据自己的需求进行设置，在本例中不需要设置。

根据以上信息，创建外表命令如下所示：

建立不包含分区列的OBS外表，表关联的外部服务器为obs\_server，表对应的OBS服务上的文件格式为‘orc’，OBS上的数据存储路径为'/mybucket/data/'。

```
DROP FOREIGN TABLE IF EXISTS product_info_ext_obs;
CREATE FOREIGN TABLE product_info_ext_obs
(
  product_price      integer      not null,
  product_id         char(30)    not null,
  product_time       date        ,
  product_level      char(10)    ,
  product_name       varchar(200) ,
  product_type1      varchar(20) ,
  product_type2      char(10)    ,
  product_monthly_sales_cnt integer ,
  product_comment_time date      ,
  product_comment_num integer    ,
  product_comment_content varchar(200)
) SERVER obs_server
OPTIONS (
format 'orc',
```



```
foldername '/mybucket/demo.db/product_info_orc',  
encoding 'utf8',  
totalrows '10'  
)  
DISTRIBUTE BY ROUNDROBIN;
```

建立包含分区列的OBS外表，product\_info\_ext\_obs外表使用product\_manufacturer字段作为分区键，obs/mybucket/demo.db/product\_info\_orc/路径下有如下分区目录：

分区目录1：product\_manufacturer=10001

分区目录2：product\_manufacturer=10010

分区目录3：product\_manufacturer=10086

```
...  
DROP FOREIGN TABLE IF EXISTS product_info_ext_obs;  
CREATE FOREIGN TABLE product_info_ext_obs  
(  
    product_price          integer      not null,  
    product_id             char(30)     not null,  
    product_time           date         ,  
    product_level          char(10)    ,  
    product_name           varchar(200) ,  
    product_type1          varchar(20) ,  
    product_type2          char(10)    ,  
    product_monthly_sales_cnt integer    ,  
    product_comment_time   date        ,  
    product_comment_num    integer     ,  
    product_comment_content varchar(200) ,  
    product_manufacturer   integer  
) SERVER obs_server  
OPTIONS (  
    format 'orc',  
    foldername '/mybucket/demo.db/product_info_orc',  
    encoding 'utf8',  
    totalrows '10'  
)  
DISTRIBUTE BY ROUNDROBIN  
PARTITION BY (product_manufacturer) AUTOMAPPED;
```

#### 4.2.1.3.4 通过外表查询 OBS 上的数据

##### 直接查询外表查看 OBS 上的数据

如果数据量较少，可直接使用SELECT查询外表，即可查看到OBS上的数据。

**步骤1** 执行以下命令，则可以从外表查询数据。

```
SELECT * FROM product_info_ext_obs;
```

查询结果显示与[原始数据](#)显示相同，则表示导入成功。查询结果的结尾将显示以下信息：

```
(10 rows)
```

通过外表查询到数据后，用户可以将数据插入数据库的普通表。

----结束

##### 导入数据后查询数据

**步骤1** 在GaussDB(DWS)数据库中，创建导入数据的目标表，用于存储导入的数据。

该表的表结构必须与[创建外表](#)中创建的外表的表结构保持一致，即字段个数、字段类型要完全一致。

例如，创建一个名为product\_info的表，示例如下：

```
DROP TABLE IF EXISTS product_info;

CREATE TABLE product_info
(
  product_price      integer    not null,
  product_id        char(30)   not null,
  product_time       date      ,
  product_level      char(10)  ,
  product_name       varchar(200) ,
  product_type1      varchar(20) ,
  product_type2      char(10)   ,
  product_monthly_sales_cnt integer ,
  product_comment_time date    ,
  product_comment_num integer  ,
  product_comment_content varchar(200)
)
with (
  orientation = column,
  compression=middle
)
DISTRIBUTE BY HASH (product_id);
```

**步骤2** 执行“INSERT INTO .. SELECT ..”命令从外表导入数据到目标表。

示例：

```
INSERT INTO product_info SELECT * FROM product_info_ext_obs;
```

若出现以下类似信息，说明数据导入成功。

```
INSERT 0 10
```

**步骤3** 执行SELECT命令，查看从OBS导入到GaussDB(DWS)中的数据。

```
SELECT * FROM product_info;
```

查询结果显示如[原始数据](#)中所示的数据，表示导入成功。查询结果的结尾将显示以下信息：

```
(10 rows)
```

```
----结束
```

#### 4.2.1.3.5 清除资源

当完成本教程的示例后，如果您不再需要使用本示例中创建的资源，您可以删除这些资源，以免资源浪费或占用您的配额。步骤如下：

1. [删除外表和目标表](#)
2. [删除创建的外部服务器](#)
3. [删除数据库及其所属的用户](#)

如果您执行了（可选）[新建用户及数据库并授予外表权限](#)中的步骤，请删除数据库及其所属的用户。

### 删除外表和目标表

**步骤1** （可选）如果执行了[导入数据后查询数据](#)，请执行以下命令，删除目标表。

```
DROP TABLE product_info;
```

当结果显示为如下信息，则表示删除成功。

```
DROP TABLE
```

**步骤2** 执行以下命令，删除外表。

```
DROP FOREIGN TABLE product_info_ext_obs;
```

当结果显示为如下信息，则表示删除成功。

```
DROP FOREIGN TABLE
```

----结束

## 删除创建的外部服务器

**步骤1** 使用创建外部服务器的用户连接到外部服务器所在的数据库。

在本示例中，使用的是普通用户dbuser在数据库mydatabase中创建了一个外部服务器。用户需要通过GaussDB(DWS)提供的数据库客户端连接数据库。例如，使用gsq客户端的用户，可以通过以下两种方法中的一种进行连接：

- 如果已经登录了gsq客户端，可以执行以下命令进行切换：  

```
\c mydatabase dbuser;
```

根据提示输入密码。
- 如果已经登录了gsq客户端，您也可以执行`\q`退出gsq后，再执行以下命令重新进行连接：  

```
gsq -d mydatabase -h 192.168.2.30 -U dbuser -p 8000 -r
```

根据提示输入密码。

**步骤2** 删除创建的外部服务器。

执行以下命令进行删除，详细语法请参见DROP SERVER。

```
DROP SERVER obs_server;
```

返回以下信息表示删除成功：

```
DROP SERVER
```

查看外部服务器：

```
SELECT * FROM pg_foreign_server WHERE srvname='obs_server';
```

返回结果如下所示，表示已经删除成功：

```
srvname | srvowner | srfdw | srvtype | srvversion | srvacl | srvoptions  
-----+-----+-----+-----+-----+-----+-----  
(0 rows)
```

----结束

## 删除数据库及其所属的用户

如果您执行了（[可选](#)）[新建用户及数据库并授予外表权限](#)中的步骤，请参照以下步骤删除数据库及其所属的用户。

**步骤1** 删除自定义数据库。

通过GaussDB(DWS)提供的数据库客户端连接默认数据库gaussdb。

如果已经登录了gsq客户端，可以直接执行如下命令进行切换：

先切换到默认数据库：

```
\c gaussdb
```

根据界面提示输入密码。

执行以下命令，删除自定义数据库：

```
DROP DATABASE mydatabase;
```

返回以下信息表示删除成功：

```
DROP DATABASE
```

**步骤2** 使用管理员用户，删除本示例中创建的普通用户。

使用数据库管理员用户通过GaussDB(DWS)提供的数据库客户端连接数据库。

如果已经登录了gsql客户端，可以直接执行如下命令进行切换：

```
\c gaussdb dbadmin
```

执行以下命令回收创建外部服务器的权限：

```
REVOKE ALL ON FOREIGN DATA WRAPPER dfs_fdw FROM dbuser;
```

其中FOREIGN DATA WRAPPER的名字只能是dfs\_fdw，dbuser为创建SERVER的用户名。

执行以下命令删除用户：

```
DROP USER dbuser;
```

可使用\du命令查询用户，确认用户是否已经删除。

----结束

#### 4.2.1.3.6 支持的数据类型

目前大数据领域，主流文件格式为ORC。GaussDB(DWS)主要支持ORC文件格式。用户利用HIVE将数据导出存储为ORC文件格式，使用GaussDB(DWS)通过只读外表对ORC文件内的数据进行查询分析，因此，需要在ORC文件格式支持的数据类型与GaussDB(DWS)自身支持数据类型间进行匹配，匹配状况如表4-5所示。同理，GaussDB(DWS)可通过只写外表将数据导出存储为ORC文件格式，使用HIVE读取ORC文件内容，相互之间也需要类型匹配，匹配状况如表4-6所示：

表 4-5 ORC 格式的只读外表与 HIVE 数据类型匹配关系

类型名称	GaussDB(DWS)外表支持类型	HIVE建表类型
1字节整数	TINYINT (不推荐)	TINYINT
	SMALLINT (推荐)	TINYINT
2字节整数	SMALLINT	SMALLINT
4字节整数	INTEGER	INT
8字节整数	BIGINT	BIGINT
单精度浮点数	FLOAT4 (REAL)	FLOAT
双精度浮点型	FLOAT8(DOUBLE PRECISION)	DOUBLE

类型名称	GaussDB(DWS)外表支持类型	HIVE建表类型
科学数据类型	DECIMAL[p (,s)] 最大支持38位精度	DECIMAL最大支持38位 ( HIVE 0.11 )
日期类型	DATE	DATE
时间类型	TIMESTAMP	TIMESTAMP
BOOLEAN类型	BOOLEAN	BOOLEAN
Char类型	CHAR(n)	CHAR (n)
VarChar类型	VARCHAR(n)	VARCHAR (n)
字符串(文本大对象)	TEXT(CLOB)	STRING

表 4-6 ORC 格式的只写外表与 HIVE 数据类型匹配关系

类型名称	GaussDB(DWS)内表支持类型 (数据源表)	GaussDB(DWS)只写外表对应的类型	HIVE建表类型
1字节整数	TINYINT	TINYINT (不推荐)	SMALLINT
		SMALLINT (推荐)	SMALLINT
2字节整数	SMALLINT	SMALLINT	SMALLINT
4字节整数	INTEGER、 BINARY_INTEGER	INTEGER	INT
8字节整数	BIGINT	BIGINT	BIGINT
单精度浮点数	FLOAT4、REAL	FLOAT4、REAL	FLOAT
双精度浮点型	DOUBLE PRECISION、 FLOAT8、 BINARY_DOUBLE	DOUBLE PRECISION、 FLOAT8、 BINARY_DOUBLE	DOUBLE
科学数据类型	DECIMAL、 NUMERIC	DECIMAL[p (,s)] 最大支持38位精度	precision <=38时, DECIMAL, precision > 38时, STRING
日期类型	DATE	TIMESTAMP[(p)] [WITHOUT TIME ZONE]	TIMESTAMP

类型名称	GaussDB(DWS)内表支持类型 (数据源表)	GaussDB(DWS)只写外表对应的类型	HIVE建表类型
时间类型	TIME [(p)] [WITHOUT TIME ZONE]、 TIME [(p)] [WITH TIME ZONE]	TEXT	STRING
	TIMESTAMP[(p)] [WITHOUT TIME ZONE]、 TIMESTAMP[(p)] [WITH TIME ZONE]、 SMALLDATETIME	TIMESTAMP[(p)] [WITHOUT TIME ZONE]	TIMESTAMP
	INTERVAL DAY (l) TO SECOND (p)、 INTERVAL [FIELDS] [(p)]	VARCHAR(n)	VARCHAR(n)
BOOLEAN类型	BOOLEAN	BOOLEAN	BOOLEAN
Char类型	CHAR(n)、 CHARACTER(n)、 NCHAR(n)	CHAR(n)、 CHARACTER(n)、 NCHAR(n)	n<=255时, CHAR(n), n>255 时, STRING
VarChar类型	VARCHAR(n)、 CHARACTER VARYING(n)、 VARCHAR2(n)、	VARCHAR(n)	n<=65535时, VARCHAR(n), n>65535时, STRING
	NVARCHAR2(n)	TEXT	STRING
字符串(文本大对象)	TEXT、CLOB	TEXT、CLOB	STRING
货币类型	MONEY	NUMERIC	BIGINT

**须知**

1. GaussDB(DWS)外表支持NULL定义，HIVE数据表支持并采用相对应的NULL定义。
2. HIVE数据表中的TINYINT的取值范围为[-128,127]，而 GaussDB(DWS) 的TINYINT的取值范围为[0,255]，因此，HIVE表中的TINYINT类型在建GaussDB(DWS)只读外表时最好采用SMALLINT类型，如果使用TINYINT有可能存在读取值与实际值不一致的情况。同样，GaussDB(DWS)的TINYINT类型在导出时，只写外表和HIVE的建表类型也最好采用SMALLINT类型。
3. GaussDB(DWS)外表的日期和时间类型，不支持时区定义，HIVE不支持时区定义。
4. HIVE中date类型只有日期，没有时间，GaussDB(DWS)的date类型包含日期和时间。
5. GaussDB(DWS)支持ORC的压缩格式，包括ZLIB，SNAPPY，LZ4及NONE压缩方式。
6. 其中FLOAT4格式本身存在不精准问题，求和等操作在不同环境下可能产生不同的结果，在高精度要求场景下建议使用DECIMAL类型代替。
7. 兼容Teradata数据库模式下，外表不支持DATE类型。

## 4.2.2 使用 GDS 从远端服务器导入数据

### 4.2.2.1 关于 GDS 并行导入

INSERT和COPY方式执行数据导入时，是一个串行执行的过程，导入性能低，因此适用于小数据量的导入。对于大数据量的导入，GaussDB(DWS)支持使用GDS工具通过外表并行导入数据到集群。

当前版本的GDS已经支持从管道文件导入数据库，该功能使GDS的导入更加灵活多变。

- 当GDS用户的本地磁盘空间不足时：
  - 可直接将hdfs上的数据写入到管道文件而不需要占用额外的磁盘空间。
- 当用户导入前需要清洗数据时：
  - 用户可以根据自己的需求编写程序，将需要处理的数据流式实时的写入管道文件，完成导入的数据清洗工作。

**说明**

- 当前版本暂不支持SSL模式下GDS导入，请勿以SSL方式使用GDS。
- 本章涉及的所有管道文件都是指linux上的命名管道。

## 概述

并行导入将存储在服务器普通文件系统中的数据导入到GaussDB(DWS)数据库中。暂时不支持将存储在HDFS文件系统上的数据导入GaussDB(DWS)。

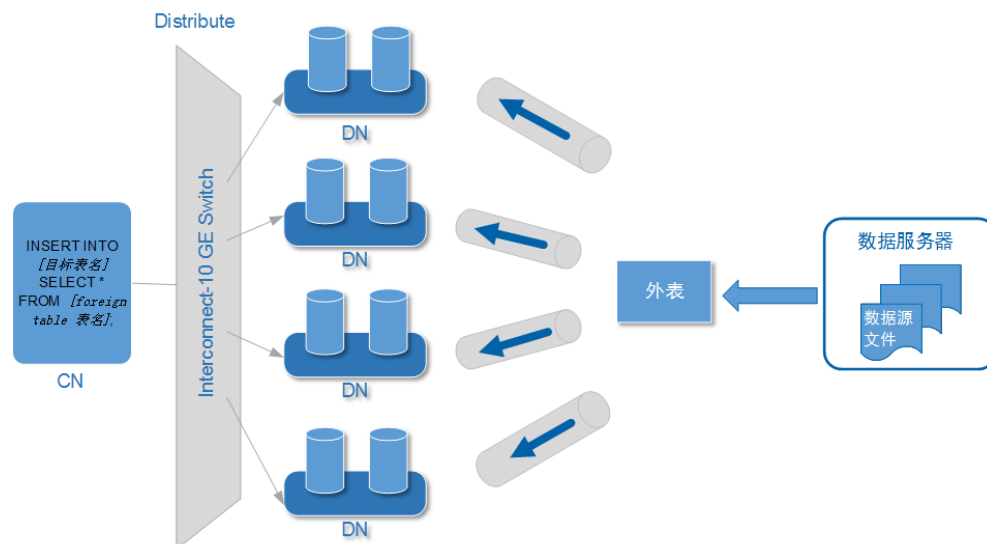
并行导入功能通过外表设置的导入策略、导入数据格式等信息来识别数据源文件，利用多DN并行的方式，将数据从数据源文件导入到数据库中，从而提高整体导入性能。如图4-4所示：

- CN只负责任务的规划及下发，把数据导入的工作交给了DN，释放了CN的资源，使其有能力处理其他外部请求。

- 所有DN都参与数据导入，这样可以充分利用各设备的计算能力及网络带宽，提升导入效率。

外表灵活的OPTION设置，有利于在数据入库前对数据做预处理，例如非法字符替换、容错处理等。

图 4-4 数据并行导入示意图



上图中所涉及的相关概念说明如下：

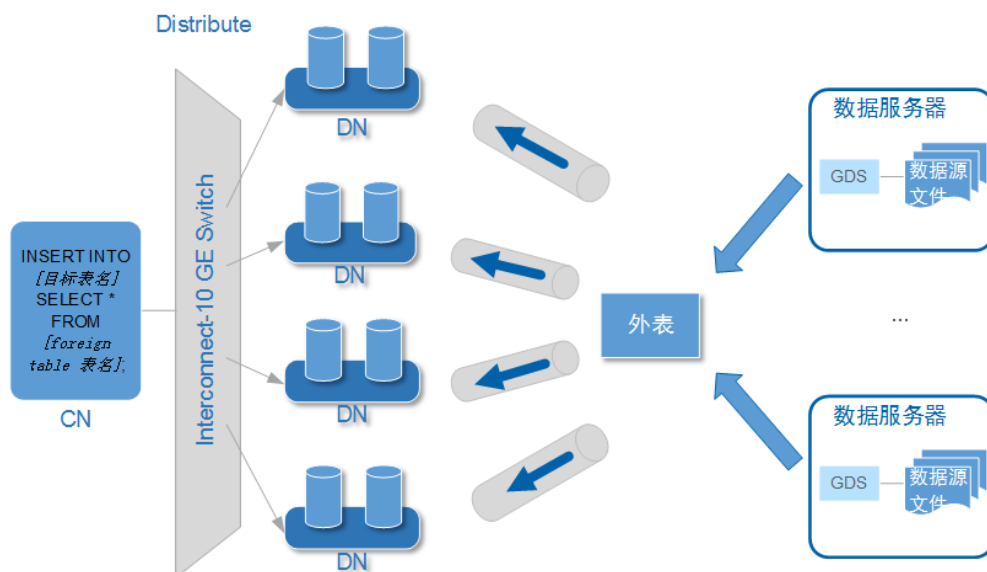
- **CN (Coordinator)**：GaussDB(DWS)协调节点。在导入场景下，接收到应用或客户端的导入SQL指令后，负责任务的规划及下发到DN。
- **DN (Datanode)**：GaussDB(DWS)数据节点。接收CN下发的导入任务，将数据源文件中的数据通过外表写入数据库目标表中。
- **数据源文件**：存有数据的文件。文件中保存的是待导入数据库的数据。
- **数据服务器**：数据源文件所在的服务器称为数据服务器。基于安全考虑，建议数据服务器和GaussDB(DWS)集群处于同一内网。
- **外表Foreign Table**：用于识别数据源文件的位置、文件格式、存放位置、编码格式、数据间的分隔符等信息。是关联数据文件与数据库实表（目标表）的对象。
- **目标表**：数据库中的实表。数据源文件中的数据最终导入到这些表中存储，包括行存表和列存表。

## GDS 并发导入

- 数据量大，数据存储多个服务器上时，在每个数据服务器上安装配置、启动GDS后，各服务器上的数据可以并行入库。如图4-5所示。



图 4-5 多数据服务器并行导入



### 须知

GDS进程数目不能超过DN数目。如果超过，会出现一个DN连接多个GDS进程的情形，可能会导致部分GDS异常运行。

- 数据存储在一台数据服务器上时，如果GaussDB(DWS)及数据服务器上的I/O资源均还有可利用空间时，可以采用GDS多线程来支持并发导入。

GDS是根据导入事务并发数来决定服务运行线程数的。也就是说即使启动GDS时设置了多线程，也并不会加速单个导入事务。未做过人为事务处理时，一条INSERT语句就是一个导入事务。

综上，多线程的使用场景如下：

- 多表并发导入时，采用多线程充分利用资源及提升并发导入效率。
- 对数据量大的某一事实表的导入进行提速。

将该事实表对应的数据拆分为多个数据文件，通过多外表同时入库的方式实现多线程并发导入。注意需确保每个外表所能读取的数据文件不重复。

## 导入流程

图 4-6 GDS 并行导入流程

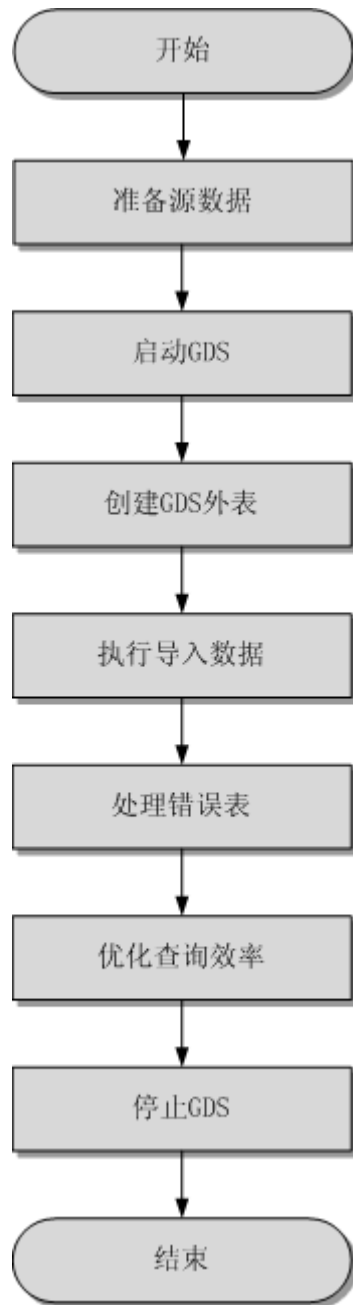


表 4-7 流程说明

流程	说明
准备源数据。	准备需要导入数据库的源数据文件，并上传至数据服务器。 详细内容请参见 <a href="#">准备源数据</a> 。
启动GDS。	在数据服务器上安装配置并启动GDS。 详细内容请参见 <a href="#">安装配置和启动GDS</a> 。

流程	说明
创建外表。	创建外表用于识别数据源文件中的数据。外表中保存了数据源文件的位置、文件格式、存放位置、编码格式、数据间的分隔符等信息。 详细内容请参见 <a href="#">创建GDS外表</a> 。
执行导入数据。	在创建好外表后，通过INSERT语句，将数据快速、高效地导入到目标表中。详细内容请参见 <a href="#">执行导入数据</a> 。
处理错误表。	在数据并行导入发生错误时，请根据具体的错误信息进行处理，以保证导入数据的完整性。 详细内容请参见 <a href="#">处理错误表</a> 。
优化查询效率。	导入数据后，通过ANALYZE语句生成表统计信息。ANALYZE语句会将统计结果自动存储在系统表PG_STATISTIC中。执行计划生成器会使用这些统计数据，以生成最有效的查询执行计划。
停止GDS	待数据导入完成后，登录每台数据服务器，分别停止GDS。 GDS的停止请参见 <a href="#">停止GDS</a> 。

### 4.2.2.2 准备源数据

#### 操作场景

通常在将数据导入数据库前，即将入库的数据已经在相关主机上了。这种保存着待入库数据的服务器为数据服务器。此时，只需检测以确认数据服务器和GaussDB(DWS)集群能够正常通信，并查看和记录数据在数据服务器上的存放目录备用。

如果待入库数据还没有就绪，则请先参考如下步骤，将数据上传到数据服务器上。

#### 操作步骤

**步骤1** 以root用户登录数据服务器。

**步骤2** 创建数据文件存放目录“/input\_data”。

```
mkdir -p /input_data
```

**步骤3** 将数据源文件上传至上一步所创建的目录中。

GDS并行导入支持CSV、TEXT格式的数据导入。请确保数据源文件符合格式要求。

----结束

### 4.2.2.3 安装配置和启动 GDS

#### 操作场景

GaussDB(DWS)提供了数据服务工具GDS来帮助分发待导入的用户数据及实现数据的高速导入。GDS需部署到数据服务器上。

数据量大，数据存储多个服务器上时，在每个数据服务器上安装配置、启动GDS后，各服务器上的数据可以并行入库。GDS在各台数据服务器上的安装配置和启动方法相同，本节以一台服务器为例进行说明。

## 背景信息

1. GDS支持在如下的操作系统中安装：

鲲鹏平台：

- Community Enterprise Operating System 7.6。
- EulerOS 2.0 SP8。
- Red Hat Enterprise Linux Server release 7.5。
- 中标麒麟7.5/7.6。

x86平台：

- SUSE Linux Enterprise Server 10 SP4 x86\_64。
- SUSE Linux Enterprise Server 11 SP1/SP2/SP3/SP4 x86\_64。
- SUSE Linux Enterprise Server 12 SP0/SP1/SP2/SP3/SP5 x86\_64。
- Red Hat Enterprise Linux Server release 6.4/6.5/6.6/6.7/6.8/6.9/7.0/7.1/7.2/7.3/7.4/7.5 x86\_64。
- Community Enterprise Operating System 6.4/6.5/6.6/6.7/6.8/6.9/7.0/7.1/7.2/7.3/7.4 x86\_64。
- EulerOS 2.5 x86\_64。

2. GDS的版本需与集群版本保持一致（如：GDS V100R008C00版本与DWS 1.3.X版本配套），否则可能会出现导入导出失败或导入导出进程停止响应等情况。

因此请勿使用历史版本的GDS进行导入。数据库版本升级后，请按照[操作步骤](#)中的办法下载GaussDB(DWS)软件包解压缩自带的GDS进行安装配置和启动。在导入导出开始时，GaussDB(DWS)也会进行两端的版本一致性检测，不一致时会打屏显示报错信息并终止对应操作。

GDS的版本号的查看办法为：在GDS工具的解压目录下执行如下命令。

```
gds -V
```

数据库版本的查看办法为：连接数据库后，执行如下SQL命令查看。

```
SELECT version();
```

## 操作步骤

- 步骤1** 在使用GDS导入/导出数据前，请先参考[步骤1：准备ECS作为GDS服务器](#)中的步骤：“准备弹性云服务器作为GDS服务器”、“下载GDS工具包和SSL证书”。

- 步骤2** 以root用户登录待安装GDS的数据服务器，创建存放GDS工具包的目录。

```
mkdir -p /opt/bin/dws
```

- 步骤3** 将GDS工具包上传至上一步所创建的目录中。

以上传SUSE Linux版本的工具包为例，将GDS工具包“dws\_client\_8.1.x\_suse\_x64.zip”上传至上一步所创建的目录中。

- 步骤4** （可选）如果使用SSL加密传输，请一并上传SSL证书至[步骤2](#)所创建的目录下。

- 步骤5** 在工具包所在目录下，解压工具包。

```
cd /opt/bin/dws  
unzip dws_client_8.1.x_suse_x64.zip
```

**步骤6** 创建GDS专有用户及其所属的用户组。此用户用于启动GDS及读取源数据。

```
groupadd gdsgrp
useradd -g gdsgrp gds_user
```

**步骤7** 分别修改工具包和数据源文件目录属主为GDS专有用户。

```
chown -R gds_user:gdsgrp /opt/bin/dws/gds
chown -R gds_user:gdsgrp /input_data
```

**步骤8** 切换到gds\_user用户。

```
su - gds_user
```

若当前集群版本为8.0.x及以前版本，请跳过**步骤9**，直接执行**步骤10**。

若当前集群版本为8.1.x及以上版本，则正常执行以下步骤。

**步骤9** 执行环境依赖脚本。（仅8.1.x及以上版本适用）

```
cd /opt/bin/dws/gds/bin
source gds_env
```

**步骤10** 启动GDS服务。

GDS是绿色软件，解压后启动即可。GDS启动方式有两种。

方式一：直接使用“gds”命令，在命令项中设置启动参数。

方式二：将启动参数写进配置文件“gds.conf”后，使用“gds\_ctl.py”命令启动。

对于集中一次性导入的场景推荐使用第一种方式。对于需要隔段时间再次导入的场景，推荐使用第二种方式以配置文件的形式提升启动效率。

- 方式一：直接使用“gds”命令，启动GDS。
  - 非SSL模式传输数据的情况下，启动GDS。

```
gds -d dir -p ip:port -H address_string -l log_file -D -t worker_num
```

示例：

```
/opt/bin/dws/gds/bin/gds -d /input_data/ -p 192.168.0.90:5000 -H 10.10.0.1/24 -l /opt/bin/dws/gds/gds_log.txt -D -t 2
```

- 使用SSL加密方式传输数据的情况下，启动GDS。

```
gds -d dir -p ip:port -H address_string -l log_file -D -t worker_num --enable-ssl --ssl-dir Cert_file
```

示例：

以**步骤4**中SSL证书以上传至/opt/bin为例，命令如下。

```
/opt/bin/dws/gds/bin/gds -d /input_data/ -p 192.168.0.90:5000 -H 10.10.0.1/24 -l /opt/bin/dws/gds/gds_log.txt -D --enable-ssl --ssl-dir /opt/bin/
```

命令中的斜体部分请根据实际替换。

- **-d dir**: 保存有待导入数据的数据文件所在目录。本教程中为“/input\_data/”。
- **-p ip:port**: GDS监听IP和监听端口。默认值为：127.0.0.1，需要替换为能跟GaussDB(DWS)通信的万兆网IP。监听端口的取值范围：1024~65535。默认值为：8098。本教程配置为：192.168.0.90:5000。
- **-H address\_string**: 允许哪些主机连接和使用GDS服务。参数需为CIDR格式。此参数配置的目的是允许GaussDB(DWS)集群可以访问GDS服务进行数据导入。所以请保证所配置的网段包含GaussDB(DWS)集群各主机。
- **-l log\_file**: 存放GDS的日志文件路径及文件名。本教程为“/opt/bin/dws/gds/gds\_log.txt”。
- **-D**: 后台运行GDS。仅支持Linux操作系统下使用。

- **-t worker\_num**: 设置GDS并发线程数。GaussDB(DWS)及数据服务器上的I/O资源均充足时，可以加大并发线程数。  
GDS是根据导入事务并发数来决定服务运行线程数的。也就是说即使启动GDS时设置了多线程，也并不会加速单个导入事务。未做过人为事务处理时，一条INSERT语句就是一个导入事务。
- **--enable-ssl**: 启用SSL加密方式传输数据。
- **--ssl-dir Cert\_file**: SSL证书所在目录。需与[步骤4](#)中的证书保存目录保持一致。
- 关于更多参数的设置信息请参考[gds命令简介](#)。
- 方式二：将启动参数写进配置文件“gds.conf”后，使用“gds\_ctl.py”命令启动。

- a. 使用如下命令，进入GDS工具包的“config”目录下，配置“gds.conf”文件。“gds.conf”配置详细信息请参考[表4-8](#)。

```
vim /opt/bin/dws/gds/config/gds.conf
```

示例：

配置“gds.conf”文件如下：

```
<?xml version="1.0"?>
<config>
<gds name="gds1" ip="192.168.0.90" port="5000" data_dir="/input_data/" err_dir="/err"
data_seg="100MB" err_seg="100MB" log_file="/log/gds_log.txt" host="10.10.0.1/24"
daemon='true' recursive="true" parallel="32"></gds>
</config>
```

配置文件信息如下：

- 数据服务器所在IP为192.168.0.90，GDS监听端口为5000。
  - 数据文件存放在“/input\_data/”目录下。
  - 错误日志文件存放在“/err”目录下。该目录需要拥有GDS读写权限的用户自行创建。
  - 单个数据文件大小为100MB。
  - 每个错误日志大小为100MB。
  - 日志保存在“/log/gds\_log.txt”文件中。该目录需要拥有GDS读写权限的用户自行创建。
  - 只允许IP为10.10.0.\*的节点进行连接。
  - GDS进程以后台方式运行。
  - 递归数据文件目录。
  - 指定并发导入工作线程数目为2。
- b. 执行如下命令启动GDS并确认GDS是否启动成功。

```
python3 gds_ctl.py start
```

示例：

```
cd /opt/bin/dws/gds/bin
python3 gds_ctl.py start
Start GDS gds1 [OK]
gds [options]:
-d dir Set data directory.
-p port Set GDS listening port.
```

```

ip:port    Set GDS listening ip address and port.
-l log_file Set log file.
-H secure_ip_range
           Set secure IP checklist in CIDR notation. Required for GDS to start.
-e dir     Set error log directory.
-E size    Set size of per error log segment.(0 < size < 1TB)
-S size    Set size of data segment.(1MB < size < 100TB)
-t worker_num Set number of worker thread in multi-thread mode, the upper limit is 32. If
without setting, the default value is 1.
-s status_file Enable GDS status report.
-D         Run the GDS as a daemon process.
-r         Read the working directory recursively.
-h         Display usage.
    
```

----结束

## gds.conf 参数说明

表 4-8 gds.conf 配置说明

属性	说明	取值范围
name	标识名。	-
ip	监听ip地址。	IP需为合法IP地址。 IP的默认值：127.0.0.1
port	监听端口号。	取值范围：1024~65535，正整数。 默认值：8098。
data_dir	数据文件目录。	-
err_dir	错误日志文件目录。	默认值：数据文件目录
log_file	日志文件路径。	-
host	设置允许连接到GDS的主机IP地址（参数为CIDR格式，仅支持linux系统）。	-
recursive	是否递归数据文件目录。	取值范围： <ul style="list-style-type: none"> <li>• true：递归。</li> <li>• false：不递归。</li> </ul> 默认值：false。
daemon	是否以DAEMON（后台）模式运行。	取值范围： <ul style="list-style-type: none"> <li>• true：以DAEMON模式运行。</li> <li>• false：不以DAEMON模式运行。</li> </ul> 默认值：false。
parallel	导入工作线程并发数目。	取值范围：0~32，正整数。 默认值：1。

#### 4.2.2.4 创建 GDS 外表

外表中配置了数据源格式信息、GDS服务的访问信息，从而GaussDB(DWS)最终可以通过外表将数据服务器上的数据引流进数据库实表中。

### 操作步骤

**步骤1** 收集数据源格式信息、GDS服务的访问信息。

需要收集的主要数据源格式信息如下：

- **format**: GDS外表导入支持CSV、TEXT和FIXED格式。请确认存放在数据服务器上待入库数据的格式。例如，假设待入库的数据为CSV格式。
- **header (仅支持CSV, FIXED格式)**: 确认数据文件是否包含标题行。
- **delimiter**: 确认数据文件中，字段间的分隔符。例如，假设是以英文逗号分隔的。
- **encoding**: 数据源文件的数据编码格式。例如，假设为UTF-8。
- **eol**: 确认数据文件中，行间的换行符。例如，默认的换行符，如0x0D0A、0X0A，或者自定义的换行符，如字符串!@#。该参数仅支持TEXT格式导入。
- 外表可识别的其他更多格式信息请参见数据格式参数。

需要收集的GDS服务的访问信息如下：

**location**: GDS服务的访问地址。例如以[安装配置和启动GDS](#)中的GDS信息为例。非SSL模式下的location为：**gsfs://192.168.0.90:5000/input\_data/**。SSL模式下的location为：**gsfss://192.168.0.90:5000/input\_data/**。其中，“192.168.0.90:5000”为GDS服务的IP及端口号；“input\_data”为GDS服务管理的数据源文件所在的路径。请根据实际情况替换。

**步骤2** 依据数据源文件中的数据情况，设计导入容错机制。

GaussDB(DWS)支持如下的数据容错性处理，相当于数据入库前对数据做初步的简单清洗。

- **fill\_missing\_fields**: 数据入库时，数据源文件中某行的最后一个字段缺失时，请选择是直接将该字段设为Null，还是在错误表中报错提示。
- **ignore\_extra\_data**: 数据源文件中的字段比外表定义列数多时，请选择是忽略多出的列，还是在错误表中报错提示。
- **per\_node\_reject\_limit**: 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实例上录入错误表中的错误数量超过设定值时，本次导入失败，报错退出。可以选择不做限制，也可以根据所能容忍的错误数量选择一个上限值。
- **compatible\_illegal\_chars**: 导入时遇到非法字符，选择如何处理。是将非法字符按照转换规则转换后入库，还是报错中止导入。

非法字符容错转换规则如下：

- 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。
- 对非法字符进行容错转换时，如遇NULL、DELIMITER、QUOTE、ESCAPE也设置成了空格或问号，GaussDB(DWS)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。



- **error\_table\_name**: 用于记录数据格式错误信息的错误表表名。并行导入结束后查询此错误信息表，能够获取详细的错误信息。
- **remote log 'name'**: 数据导入过程中的数据格式错误信息是否同时在GDS服务器上以文件方式保存。name为错误数据文件的文件名前缀。
- 关于容错性参数的更多信息请参考容错性参数。

**步骤3** 使用gsq或Data Studio连接数据库后，根据前面步骤所收集和规划的信息参数，创建GDS外表。

示例如下：

```
CREATE FOREIGN TABLE foreign_tpcds_reasons
(
  r_reason_sk integer not null,
  r_reason_id char(16) not null,
  r_reason_desc char(100)
)
SERVER gsmpp_server
OPTIONS
(
  LOCATION 'gsfs://192.168.0.90:5000/* | gsfs://192.168.0.91:5000/*',
  FORMAT 'CSV',
  DELIMITER ',',
  ENCODING 'utf8',
  HEADER 'false',
  FILL_MISSING_FIELDS 'true',
  IGNORE_EXTRA_DATA 'true'
)
LOG INTO product_info_err
PER NODE REJECT LIMIT 'unlimited';
```

示例中的各项说明如下：

- 外表字段需与数据库中即将存储数据的事实表保持一致。
- 对于GDS导入，SERVER gsmpp\_server请保持不变。
- location参数请根据**步骤1**中收集的GDS服务访问信息修改。注意GDS使用SSL加密传输时，需要将“gsfs”替换为“gsfss”。
- FORMAT、DELIMITER、ENCODING、HEADER请根据**步骤1**中收集的数据源格式信息填写。
- FILL\_MISSING\_FIELDS、IGNORE\_EXTRA\_DATA、LOG INTO及PER NODE REJECT LIMIT请根据**步骤2**中设计的导入容错机制填写。注意LOG INTO是指将数据格式错误录入哪个错误表，即其取值为错误表表名。

CREATE FOREIGN TABLE语法的更多信息，请参考CREATE FOREIGN TABLE (GDS导入导出)。

----结束

## 任务示例

除了以下示例，更多外表创建的示例请参考[GDS导入示例](#)。

- **示例1**: 创建GDS外表foreign\_tpcds\_reasons，数据格式为CSV。

```
CREATE FOREIGN TABLE foreign_tpcds_reasons
(
  r_reason_sk integer not null,
  r_reason_id char(16) not null,
  r_reason_desc char(100)
)
```

```
SERVER gsmpp_server OPTIONS (location 'gsfs://192.168.0.90:5000/* | gsfs://192.168.0.91:5000/*',  
FORMAT 'CSV',MODE 'Normal', ENCODING 'utf8', DELIMITER E'\x08', QUOTE E'\x1b', NULL '');
```

- **示例2: 创建GDS导入外表foreign\_tpcds\_reasons\_SSL, 使用SSL加密传输的模式传输, 数据格式为CSV。**

```
CREATE FOREIGN TABLE foreign_tpcds_reasons_SSL
```

```
(  
  r_reason_sk integer not null,  
  r_reason_id char(16) not null,  
  r_reason_desc char(100)  
)
```

```
SERVER gsmpp_server OPTIONS (location 'gsfs://192.168.0.90:5000/* | gsfs://192.168.0.91:5000/*',  
FORMAT 'CSV',MODE 'Normal', ENCODING 'utf8', DELIMITER E'\x08', QUOTE E'\x1b', NULL '');
```

- **示例3: 创建GDS外表foreign\_tpcds\_reasons, 数据格式为TEXT。**

```
CREATE FOREIGN TABLE foreign_tpcds_reasons
```

```
(  
  r_reason_sk integer not null,  
  r_reason_id char(16) not null,  
  r_reason_desc char(100)  
)
```

```
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',EOL '0x0D') WITH err_foreign_tpcds_reasons;
```

### 4.2.2.5 执行导入数据

完成GDS的安装部署及外表创建后, 本节介绍如何在GaussDB(DWS)数据库中创建事实表并将数据导入事实表中。

对于记录数超过千万条的表, 建议在执行全量数据导入前, 先导入部分数据, 以[进行数据倾斜检查和调整分布列](#), 避免导入大量数据后发现数据倾斜, 调整成本高。

### 前提条件

GDS服务器和GaussDB(DWS)集群之间网络可以互通。

- 需要创建一个弹性云服务器作为GDS服务器。
- 创建的弹性云服务器与GaussDB(DWS)集群应处于同一区域、同一虚拟私有云和子网。

### 操作步骤

**步骤1** 在GaussDB(DWS)中创建目标表, 用于存储导入的数据。建表语句请参见CREATE TABLE。

**步骤2** (可选) 若导入表存在索引, 在数据导入过程中, 将增量更新索引信息, 影响数据导入性能。建议在执行数据导入前, 先删除相关表的索引, 但是如果不能保证数据唯一性不建议删除唯一索引。在数据导入完成后, 再重新创建索引。

1. 假定在导入表“product\_info”上的“product\_id”字段上存在普通索引“product\_idx”。在执行数据导入前, 请先删除相关索引。

```
DROP INDEX product_idx;
```

2. 在数据导入完成后, 重建索引。

```
CREATE INDEX product_idx ON product_info(product_id);
```

**步骤3** 执行数据导入。

```
INSERT INTO [目标表名] SELECT * FROM [foreign table 表名];
```

- 若出现以下类似信息, 说明数据导入成功。请查询错误信息表, 查看是否存在数据格式错误, 详细操作请参见[处理错误表](#)。

```
INSERT 0 9
```

- 若出现数据加载错误，请参见[处理错误表](#)，并重新执行数据导入。

#### 📖 说明

- 若执行过程中出现数据加载错误，则数据全部导入失败，没有数据导入至目标表中。
- 编写批处理任务脚本，实现并发批量导入数据。并发量视机器资源使用情况而定。可通过几个表测试，监控资源利用率，根据结果提高或减少并发量。常用资源监控命令有：内存和CPU监控top命令，IO监控命令iostat，网络监控命令sar等。相关案例请参见。
- 在资源许可的情况下，多台GDS服务器并发导入会很大程度上提高数据导入效率。相关案例请参见[多数据服务器并行导入](#)。
- 对于高并发的GDS导入场景，为了保持GDS和DN间的数据连接稳定，可以将GDS服务器环境和DN所在环境的TCP Keepalive检测时间增长（推荐增长至5分钟）。调整集群环境的TCP Keepalive参数会影响故障检测的响应时间。

#### ----结束

## 任务示例

1. 创建一个名为reasons的目标表。

```
CREATE TABLE reasons
(
  r_reason_sk integer not null,
  r_reason_id char(16) not null,
  r_reason_desc char(100)
)
DISTRIBUTE BY HASH (r_reason_sk);
```

2. 在执行数据导入前，先删除相关表的索引。

假定在导入表“reasons”上的“r\_reason\_id”字段上存在普通索引“reasons\_idx”。在执行数据导入前，请先删除相关索引。

```
DROP INDEX reasons_idx;
```

3. 将数据源文件中的数据通过外表“foreign\_tpcds\_reasons”导入到表“reasons”中。

```
INSERT INTO reasons SELECT * FROM foreign_tpcds_reasons ;
```

4. 在数据导入完成后，再重新创建索引。

```
CREATE INDEX reasons_idx ON reasons(r_reasons_id);
```

### 4.2.2.6 处理错误表

## 操作场景

当数据导入发生错误时，请根据本文指引信息进行处理。

## 查询错误信息

数据导入过程中发生的错误，一般分为数据格式错误和非数据格式错误。

- 数据格式错误

在创建外表时，通过设置参数“LOG INTO error\_table\_name”，将数据导入过程中出现的数据格式错误信息写入指定的错误信息表error\_table\_name中。您可以通过以下SQL，查询详细错误信息。

```
SELECT * FROM error_table_name;
```

错误信息表结构如[表4-9](#)所示。

表 4-9 错误信息表

列名称	类型	描述
nodeid	integer	报错节点编号。
begintime	timestamp with time zone	出现数据格式错误的时间。
filename	character varying	出现数据格式错误的数据源文件名。 当GDS导入时，同时会包括对应GDS服务端的IP地址端口信息。
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。
detail	text	详细错误信息。

- 非数据格式错误

对于非数据格式错误，一旦发生将导致整个数据导入失败。您可以根据执行数据导入过程中，界面提示的错误信息，帮助定位问题，处理错误表。

## 处理数据导入错误

根据获取的错误信息，请对照下表，处理数据导入错误。

表 4-10 处理数据导入错误

错误信息	原因	解决办法
missing data for column "r_reason_desc"	<p>1. 数据源文件中的列数比外表定义的列数少。</p> <p>2. 对于TEXT格式的数据源文件，由于转义字符（\）导致delimiter（分隔符）错位或者quote（引号字符）错位造成的错误。  <b>示例：</b>目标表存在3列字段，导入的数据如下所示。由于存在转义字符“\”，分隔符“ ”被转义为第二个字段的字段值，导致第三个字段值缺失。                      BE Belgium\ 1</p>	<p>1. 由于列数少导致的报错，选择下列办法解决：</p> <ul style="list-style-type: none"> <li>在数据源文件中，增加列“r_reason_desc”的字段值。</li> <li>在创建外表时，将参数“fill_missing_fields”设置为“on”。即当导入过程中，若数据源文件中一行数据的最后一个字段缺失，则把最后一个字段的值设置为NULL，不报错。</li> </ul> <p>2. 对由于转义字符导致的错误，需检查报错的行中是否含有转义字符（\）。若存在，建议在创建外表时，将参数“noescaping”（是否不对\和后面的字符进行转义）设置为true。</p>
extra data after last expected column	数据源文件中的列数比外表定义的列数多。	<ul style="list-style-type: none"> <li>在数据源文件中，删除多余的字段值。</li> <li>在创建外表时，将参数“ignore_extra_data”设置为“on”。即在导入过程中，若数据源文件比外表定义的列数多，则忽略行尾多出来的列。</li> </ul>
invalid input syntax for type numeric: "a"	数据类型错误。	在数据源文件中，修改输入字段的数据类型。根据此错误信息，请将输入的数据类型修改为numeric。
null value in column "staff_id" violates not-null constraint	非空约束。	在数据源文件中，增加非空字段信息。根据此错误信息，请增加“staff_id”列的值。

错误信息	原因	解决办法
duplicate key value violates unique constraint "reg_id_pk"	唯一约束。	<ul style="list-style-type: none"> <li>删除数据源文件中重复的行。</li> <li>通过设置关键字“DISTINCT”，从SELECT结果集中删除重复的行，保证导入的每一行都是唯一的。</li> </ul> <pre>INSERT INTO reasons SELECT DISTINCT * FROM foreign_tpcds_reasons;</pre>
value too long for type character varying(16)	字段值长度超过限制。	在数据源文件中，修改字段值长度。根据此错误信息，字段值长度限制为VARCHAR2(16)。

#### 4.2.2.7 停止 GDS

##### 操作场景

待导入数据成功后，停止GDS。

##### 操作步骤

**步骤1** 以gds\_user用户登录安装GDS的数据服务器。

**步骤2** 请根据启动GDS的方式，选择停止GDS的方式。

- 若用户使用“gds”命令启动GDS，请使用以下方式停止GDS。

a. 执行如下命令，查询GDS进程号。

```
ps -ef|grep gds
```

示例：其中GDS进程号为128954。

```
ps -ef|grep gds
```

```
gds_user 128954 1 0 15:03 ? 00:00:00 gds -d /input_data/ -p 192.168.0.90:5000 -l /log/gds_log.txt -D
```

```
gds_user 129003 118723 0 15:04 pts/0 00:00:00 grep gds
```

b. 使用“kill”命令，停止GDS。其中128954为上一步骤中查询出的GDS进程号。

```
kill -9 128954
```

----结束

#### 4.2.2.8 GDS 导入示例

##### 多数据服务器并行导入

规划数据服务器与集群处于同一内网，数据服务器IP为192.168.0.90和192.168.0.91。数据源文件格式为CSV。

- 创建导入的目标表tpcds.reasons。

```
CREATE TABLE tpcds.reasons
(
  r_reason_sk integer not null,
  r_reason_id char(16) not null,
  r_reason_desc char(100)
);
```

2. 以root用户登录每台GDS数据服务器，在两台数据服务器上，分别创建数据文件存放目录“/input\_data”。以下以IP为192.168.0.90的数据服务器为例进行操作，剩余服务器上的操作与它一致。

```
mkdir -p /input_data
```

3. （可选）创建用户及其所属的用户组。此用户用于启动GDS。若该类用户及所属用户组已存在，可跳过此步骤。

```
groupadd gdsgrp
useradd -g gdsgrp gds_user
```

4. 将数据源文件均匀分发至相应数据服务器的“/input\_data”目录中。
5. 修改每台数据服务器上数据文件及数据文件目录“/input\_data”的属主为gds\_user。以下以IP为192.168.0.90的数据服务器为例，进行操作。

```
chown -R gds_user:gdsgrp /input_data
```

6. 以gds\_user用户登录每台数据服务器上分别启动GDS。

其中GDS安装路径为“/opt/bin/dws/gds”，数据文件存放在“/input\_data/”目录下，数据服务器所在IP为192.168.0.90和192.168.0.91，GDS监听端口为5000，以后台方式运行。

在IP为192.168.0.90的数据服务器上启动GDS。

```
/opt/bin/dws/gds/bin/gds -d /input_data -p 192.168.0.90:5000 -H 10.10.0.1/24 -D
```

在IP为192.168.0.91的数据服务器上启动GDS。

```
/opt/bin/dws/gds/bin/gds -d /input_data -p 192.168.0.91:5000 -H 10.10.0.1/24 -D
```

7. 创建外表tpcds.foreign\_tpcds\_reasons用于接收数据服务器上的数据。

其中设置导入模式信息如下所示：

- 导入模式为Normal模式。
- 由于启动GDS时，设置的数据源文件存放目录为“/input\_data”，GDS监听端口为5000，所以设置参数“location”为“gsfs://192.168.0.90:5000/\* | gsfs://192.168.0.91:5000/\*”。

设置数据格式信息是根据导出时设置的详细数据格式参数信息指定的，参数设置如下所示：

- 数据源文件格式（format）为CSV。
- 编码格式（encoding）为UTF-8。
- 字段分隔符（delimiter）为E'\x08'。
- 引号字符（quote）为E'\x1b'。
- 数据文件中空值（null）为没有引号的空字符串。
- 逃逸字符（escape）默认和quote相同。
- 数据文件是否包含标题行（header）为默认值false，即导入时数据文件第一行被识别为数据。

设置导入容错性如下所示：

- 允许出现的数据格式错误个数（PER NODE REJECT LIMIT 'value'）为unlimited，即接受导入过程中所有数据格式错误。
- 将数据导入过程中出现的数据格式错误信息（LOG INTO error\_table\_name）写入表err\_tpcds\_reasons。

根据以上信息，创建的外表如下所示：

```
CREATE FOREIGN TABLE tpcds.foreign_tpcds_reasons
(
  r_reason_sk integer not null,
  r_reason_id char(16) not null,
  r_reason_desc char(100)
)
SERVER gsmpp_server OPTIONS (location 'gsfs://192.168.0.90:5000/* | gsfs://192.168.0.91:5000/*',
format 'CSV',mode 'Normal', encoding 'utf8', delimiter E'\x08', quote E'\x1b', null '', fill_missing_fields
'false') LOG INTO err_tpcds_reasons PER NODE REJECT LIMIT 'unlimited';
```

8. 通过外表tpcds.foreign\_tpcds\_reasons，将数据导入目标表tpcds.reasons。

```
INSERT INTO tpcds.reasons SELECT * FROM tpcds.foreign_tpcds_reasons;
```

9. 查询错误信息表err\_tpcds\_reasons，处理数据导入错误。详细请参见[处理错误表](#)。

```
SELECT * FROM err_tpcds_reasons;
```

10. 待数据导入完成后，以gds\_user用户登录每台数据服务器，分别停止GDS。

以下以IP为192.168.0.90的数据服务器为例，停止GDS。其中GDS进程号为128954。

```
ps -ef|grep gds
gds_user 128954 1 0 15:03 ? 00:00:00 gds -d /input_data -p 192.168.0.90:5000 -D
gds_user 129003 118723 0 15:04 pts/0 00:00:00 grep gds
kill -9 128954
```

## 多线程导入

规划数据服务器与集群处于同一内网，数据服务器IP为192.168.0.90，导入的数据源文件格式为CSV，同时导入2个目标表。

1. 在数据库中创建导入的目标表tpcds.reasons1和tpcds.reasons2。

```
CREATE TABLE tpcds.reasons1
(
  r_reason_sk integer not null,
  r_reason_id char(16) not null,
  r_reason_desc char(100)
);
CREATE TABLE tpcds.reasons2
(
  r_reason_sk integer not null,
  r_reason_id char(16) not null,
  r_reason_desc char(100)
);
```

2. 以root用户登录GDS数据服务器，创建数据文件存放目录“/input\_data”，以及子目录“/input\_data/import1/”和“/input\_data/import2/”。

```
mkdir -p /input_data
```

3. 将目标表tpcds.reasons1的数据源文件存放在数据服务器“/input\_data/import1/”目录下，将目标表tpcds.reasons2的数据源文件存放在目录“/input\_data/import2/”下。

4. （可选）创建用户及其所属的用户组。此用户用于启动GDS。若该用户及所属用户组已存在，可跳过此步骤。

```
groupadd gdsgrp
useradd -g gdsgrp gds_user
```

5. 修改数据服务器上数据文件及数据文件目录“/input\_data”的属主为gds\_user。

```
chown -R gds_user:gdsgrp /input_data
```

6. 以gds\_user用户登录数据服务器上启动GDS。

其中GDS安装路径为“/gds”，数据文件存放在“/input\_data/”目录下，数据服务器所在IP为192.168.0.90，GDS监听端口为5000，以后台方式运行，设定并发度为2，并设定递归文件目录。

```
/gds/gds -d /input_data -p 192.168.0.90:5000 -H 10.10.0.1/24 -D -t 2 -r
```



7. 在数据库中创建外表tpcds.foreign\_tpcds\_reasons1和tpcds.foreign\_tpcds\_reasons2用于接收数据服务器上的数据。

以下以外表tpcds.foreign\_tpcds\_reasons1为例，讲解设置的导入外表参数信息。

其中设置的**导入模式信息**如下所示：

- 导入模式为Normal模式。
- 由于启动GDS时，设置的数据源文件存放目录为“/input\_data/”，GDS监听端口为5000，实际存放数据源文件目录为“/input\_data/import1/”，所以设置参数“location”为“gsfs://192.168.0.90:5000/import1/\*”。

设置的**数据格式信息**是根据导出时设置的详细数据格式参数信息指定的，参数设置如下所示：

- 数据源文件格式（format）为CSV。
- 编码格式（encoding）为UTF-8。
- 字段分隔符（delimiter）为E'\x08'。
- 引号字符（quote）为E'\x1b'。
- 数据文件中空值（null）为没有引号的空字符串。
- 逃逸字符（escape）默认和quote相同。
- 数据文件是否包含标题行（header）为默认值false，即导入时数据文件第一行被识别为数据。

设置的**导入容错性**如下所示：

- 允许出现的数据格式错误个数（PER NODE REJECT LIMIT 'value'）为unlimited，即接受导入过程中所有数据格式错误。
- 将数据导入过程中出现的数据格式错误信息（LOG INTO error\_table\_name）写入表err\_tpcds\_reasons1。
- 当数据源文件中一行的最后一个字段缺失（fill\_missing\_fields）时，自动设置为NULL。

根据以上信息，创建的外表tpcds.foreign\_tpcds\_reasons1如下所示：

```
CREATE FOREIGN TABLE tpcds.foreign_tpcds_reasons1
(
  r_reason_sk integer not null,
  r_reason_id char(16) not null,
  r_reason_desc char(100)
) SERVER gsmpp_server OPTIONS (location 'gsfs://192.168.0.90:5000/import1/*', format 'CSV', mode
'Normal', encoding 'utf8', delimiter E'\x08', quote E'\x1b', null '', fill_missing_fields 'on') LOG INTO
err_tpcds_reasons1 PER NODE REJECT LIMIT 'unlimited';
```

参考以上设置，创建的外表tpcds.foreign\_tpcds\_reasons2如下所示：

```
CREATE FOREIGN TABLE tpcds.foreign_tpcds_reasons2
(
  r_reason_sk integer not null,
  r_reason_id char(16) not null,
  r_reason_desc char(100)
) SERVER gsmpp_server OPTIONS (location 'gsfs://192.168.0.90:5000/import2/*', format 'CSV', mode
'Normal', encoding 'utf8', delimiter E'\x08', quote E'\x1b', null '', fill_missing_fields 'on') LOG INTO
err_tpcds_reasons2 PER NODE REJECT LIMIT 'unlimited';
```

8. 通过外表tpcds.foreign\_tpcds\_reasons1和tpcds.foreign\_tpcds\_reasons2将数据分别导入tpcds.reasons1和tpcds.reasons2。

```
INSERT INTO tpcds.reasons1 SELECT * FROM tpcds.foreign_tpcds_reasons1;
INSERT INTO tpcds.reasons2 SELECT * FROM tpcds.foreign_tpcds_reasons2;
```

9. 查询错误信息表err\_tpcds\_reasons1和err\_tpcds\_reasons2，处理数据导入错误。详细请参见[处理错误表](#)。

```
SELECT * FROM err_tpcds_reasons1;
SELECT * FROM err_tpcds_reasons2;
```

- 待数据导入完成后，以gds\_user用户登录数据服务器，停止GDS。  
其中GDS进程号为128954。

```
ps -ef|grep gds
gds_user 128954      1  0 15:03 ?        00:00:00 gds -d /input_data -p 192.168.0.90:5000 -D -t 2 -r
gds_user 129003 118723  0 15:04 pts/0    00:00:00 grep gds
kill -9 128954
```

## 单个管道文件导入

### 步骤1 启动GDS。

```
gds -d /gds_data/ -D -p 192.168.0.1:7789 -l /gds_log/aa.log -H 0/0 -t 10 -D
```

如果需要设置管道文件的超时时间，则使用--pipe-timeout参数设置。

### 步骤2 执行数据导入。

- 登录数据库创建内表。

```
CREATE TABLE test_pipe_1( id integer not null, sex text not null, name text );
```

- 创建只读外表。

```
CREATE FOREIGN TABLE foreign_test_pipe_tr( like test_pipe ) SERVER gsmpp_server OPTIONS
(LOCATION 'gsfs://192.168.0.1:7789/foreign_test_pipe.pipe', FORMAT 'text', DELIMITER ',', NULL '',
EOL '\0x0a',file_type 'pipe',auto_create_pipe 'false');
```

- 执行导入语句，此时语句会阻塞。

```
INSERT INTO test_pipe_1 select * from foreign_test_pipe_tr;
```

### 步骤3 通过GDS管道文件导入数据。

- 登录GDS服务器进入GDS数据目录。

```
cd /gds_data/
```

- 创建管道文件，如果auto\_create\_pipe设置为true跳过此步骤。

```
mkfifo foreign_test_pipe.pipe;
```

#### 说明

管道文件创建完成后，每执行完一次操作，业务会被自动清理。如果还需要执行其他业务，请参考该步骤重新创建管道文件。

- 向管道文件中写入数据。

```
cat postgres_public_foreign_test_pipe_tw.txt > foreign_test_pipe.pipe
```

- 若需要读取压缩文件到管道文件，执行：

```
gzip -d < out.gz > foreign_test_pipe.pipe
```

- 若需要读取hdfs文件到管道文件，执行：

```
hdfs dfs -cat - /user/hive/gds/test_pipe.txt > foreign_test_pipe.pipe
```

### 步骤4 查看导入语句返回的结果：

```
INSERT INTO test_pipe_1 select * from foreign_test_pipe_tr;
INSERT 0 4
SELECT * FROM test_pipe_1;
 id | sex | name
----+----+-----
 3 | 2   | 1111111111111111
 1 | 2   | 1111111111111111
 2 | 2   | 1111111111111111
 4 | 2   | 1111111111111111
(4 rows)
```

----结束

## 多进程管道文件导入

GDS支持多进程管道文件导入，即启动一个外表对应多个GDS。

以本地文件的导入为例：

**步骤1** 启动多个GDS，如果已经启动跳过此步骤。

```
gds -d /***gds_data/ -D -p 192.168.0.1:7789 -l /***gds_log/aa.log -H 0/0 -t 10 -D
gds -d /***gds_data_1/ -D -p 192.168.0.1:7790 -l /***gds_log_1/aa.log -H 0/0 -t 10 -D
```

如果需要设置管道文件的超时时间，则使用--pipe-timeout参数设置。

**步骤2** 执行数据导入。

1. 登录数据库创建内表。

```
CREATE TABLE test_pipe( id integer not null, sex text not null, name text );
```

2. 创建只读外表。

```
CREATE FOREIGN TABLE foreign_test_pipe_tr( like test_pipe ) SERVER gsmpp_server OPTIONS
(LOCATION 'gsfs://192.168.0.1:7789/foreign_test_pipe.pipe|gsfs://192.168.0.1:7790/
foreign_test_pipe.pipe', FORMAT 'text', DELIMITER ',', NULL '', EOL '0x0a', file_type 'pipe',
auto_create_pipe 'false');
```

3. 导入语句，此时语句会阻塞。

```
INSERT INTO test_pipe_1 select * from foreign_test_pipe_tr;
```

**步骤3** 通过GDS管道文件导入数据。

1. 登录GDS，分别进入GDS数据目录。

```
cd /***gds_data/
cd /***gds_data_1/
```

2. 创建管道文件，如果auto\_create\_pipe设置为true跳过此步骤。

```
mkfifo foreign_test_pipe.pipe;
```

3. 分别读取管道文件并写入新文件：

```
cat postgres_public_foreign_test_pipe_tw.txt > foreign_test_pipe.pipe
```

**步骤4** 查看导入语句返回的结果：

```
INSERT INTO test_pipe_1 select * from foreign_test_pipe_tr;
INSERT 0 4
SELECT * FROM test_pipe_1;
id | sex | name
-----+-----+-----
3 | 2 | 1111111111111111
1 | 2 | 1111111111111111
2 | 2 | 1111111111111111
4 | 2 | 1111111111111111
(4 rows)
```

----结束

## 集群间不落地数据导入

**步骤1** 启动GDS。（如果已经启动跳过此步骤）

```
gds -d /***gds_data/ -D -p GDS_IP:GDS_PORT -l /***gds_log/aa.log -H 0/0 -t 10 -D
```

如果需要设置管道文件的超时时间，则使用--pipe-timeout参数设置。

**步骤2** 源数据库数据导出。

1. 登录目标数据库创建内表，并写入数据。

```
CREATE TABLE test_pipe( id integer not null, sex text not null, name text );
INSERT INTO test_pipe values(1,2,'1111111111111111');
INSERT INTO test_pipe values(2,2,'1111111111111111');
INSERT INTO test_pipe values(3,2,'1111111111111111');
INSERT INTO test_pipe values(4,2,'1111111111111111');
INSERT INTO test_pipe values(5,2,'1111111111111111');
```

2. 创建只写外表。

```
CREATE FOREIGN TABLE foreign_test_pipe( id integer not null, age text not null, name text ) SERVER
gsmpp_server OPTIONS (LOCATION 'gsfs://GDS_IP:GDS_PORT/', FORMAT 'text', DELIMITER ',', NULL ",
EOL '0x0a',file_type 'pipe') WRITE ONLY;
```

3. 导入语句，此时语句会阻塞。

```
INSERT INTO foreign_test_pipe SELECT * FROM test_pipe;
```

### 步骤3 目标集群导入数据。

1. 创建内表。

```
CREATE TABLE test_pipe (id integer not null, sex text not null, name text);
```

2. 创建只读外表。

```
CREATE FOREIGN TABLE foreign_test_pipe(like test_pipe) SERVER gsmpp_server OPTIONS (LOCATION
'gsfs://GDS_IP:GDS_PORT/', FORMAT 'text', DELIMITER ',', NULL ", EOL '0x0a' , file_type 'pipe',
auto_create_pipe 'false');
```

3. 执行导入语句：

```
INSERT INTO test_pipe SELECT * FROM foreign_test_pipe;
```

### 步骤4 查看目标集群导入语句返回的结果：

```
SELECT * FROM test_pipe;
id | sex | name
-----+-----+-----
3 | 2 | 1111111111111111
6 | 2 | 1111111111111111
7 | 2 | 1111111111111111
1 | 2 | 1111111111111111
2 | 2 | 1111111111111111
4 | 2 | 1111111111111111
5 | 2 | 1111111111111111
8 | 2 | 1111111111111111
9 | 2 | 1111111111111111
(9 rows)
```

----结束

#### 📖 说明

GDS默认导出或者导入的管道文件命名规则为：“数据库名\_模式名\_外表名.pipe”，因此默认需要目标集群与源集群的数据库名及模式名保持一致。如果数据库或模式不一致，则可以在location的url中指定相同的管道文件。

示例：

- 只写外表指定管道名。

```
CREATE FOREIGN TABLE foreign_test_pipe(id integer not null, age text not null, name text)
SERVER gsmpp_server OPTIONS (LOCATION 'gsfs://GDS_IP:GDS_PORT/foreign_test_pipe.pipe',
FORMAT 'text', DELIMITER ',', NULL ", EOL '0x0a' ,file_type 'pipe') WRITE ONLY;
```

- 只读外表指定管道名。

```
CREATE FOREIGN TABLE foreign_test_pipe(like test_pipe) SERVER gsmpp_server OPTIONS
(LOCATION 'gsfs://GDS_IP:GDS_PORT/foreign_test_pipe.pipe', FORMAT 'text', DELIMITER ',', NULL
", EOL '0x0a' ,file_type 'pipe',auto_create_pipe 'false');
```

## 4.2.3 从 MRS 导入数据到集群

### 4.2.3.1 从 MRS 导入数据概述

MapReduce服务（MapReduce Service，简称MRS）是一个基于开源Hadoop生态环境而运行的大数据集群，对外提供大容量数据的存储和分析能力，可解决用户的数据存储和处理需求。具体信息可参考《[MapReduce服务用户指南](#)》。

用户可以将海量业务数据，存储在MRS的分析集群，即使用Hive/Spark组件保存。Hive/Spark的数据文件则保存在HDFS中。GaussDB(DWS)支持在相同网络中，配置一

个GaussDB(DWS)集群连接到一个MRS集群，然后将数据从HDFS中的文件读取到GaussDB(DWS)。

#### 须知

确保MRS跟DWS网络互联互通，主要分以下几种场景：

场景一：MRS与DWS在同一个区域、同一个VPC下，默认网络互通。

场景二：MRS与DWS在同一个区域，不同VPC下，需要建立VPC对等连接，参见[对接连接简介](#)。

场景三：MRS与DWS不在一个区域，需要通过“[云连接\(CC\)](#)”打通网络，请参见对应服务的用户指南。

场景四：MRS属于云下场景，需要通过“[云专线\(DC\)](#)”或“[虚拟专用网络\(VPN\)](#)”打通网络，请参见对应服务的用户指南。

## 从 MRS 导入数据到集群的流程

1. [MRS集群上的数据准备](#)
2. （可选）[手动创建外部服务器](#)
3. [创建外表](#)
4. [执行数据导入](#)
5. [清除资源](#)

### 4.2.3.2 MRS 集群上的数据准备

从MRS导入数据到GaussDB(DWS)集群之前，假设您已经完成了以下准备工作：

1. 已创建MRS集群。
2. 在MRS集群上创建了Hive/Spark ORC表，且表数据已经存储到该表对应的HDFS路径上。

如果您已经完成上述准备，则可以跳过本章节。

为方便起见，以在MRS集群上创建Hive ORC表作为示例，完成上述准备工作。在MRS集群上创建Spark ORC表的大致流程和SQL语法，同Hive类似，在本文中不再展开描述。

## 数据文件

假设有数据文件product\_info.txt，示例数据如下所示：

```
100,XHDK-A-1293-#fJ3,2017-09-01,A,2017 Autumn New Shirt Women,red,M,328,2017-09-04,715,good
205,KDKE-B-9947-#kL5,2017-09-01,A,2017 Autumn New Knitwear Women,pink,L,584,2017-09-05,406,very
good!
300,JODL-X-1937-#pV7,2017-09-01,A,2017 autumn new T-shirt men,red,XL,1245,2017-09-03,502,Bad.
310,QQPX-R-3956-#aD8,2017-09-02,B,2017 autumn new jacket women,red,L,411,2017-09-05,436,It's really
super nice
150,ABEF-C-1820-#mC6,2017-09-03,B,2017 Autumn New Jeans Women,blue,M,1223,2017-09-06,1200,The
seller's packaging is exquisite
200,BCQP-E-2365-#qE4,2017-09-04,B,2017 autumn new casual pants men,black,L,997,2017-09-10,301,The
clothes are of good quality.
250,EABE-D-1476-#oB1,2017-09-10,A,2017 autumn new dress women,black,S,841,2017-09-15,299,Follow
the store for a long time.
108,CDXK-F-1527-#pL2,2017-09-11,A,2017 autumn new dress women,red,M,85,2017-09-14,22,It's really
```

```
amazing to buy
450,MMCE-H-4728-#nP9,2017-09-11,A,2017 autumn new jacket women,white,M,114,2017-09-14,22,Open
the package and the clothes have no odor
260,OCDA-G-2817-#bD3,2017-09-12,B,2017 autumn new woolen coat women,red,L,
2004,2017-09-15,826,Very favorite clothes
980,ZKDS-J-5490-#cW4,2017-09-13,B,2017 Autumn New Women's Cotton Clothing,red,M,
112,2017-09-16,219,The clothes are small
98,FKQB-I-2564-#dA5,2017-09-15,B,2017 autumn new shoes men,green,M,4345,2017-09-18,5473,The
clothes are thick and it's better this winter.
150,DMQY-K-6579-#eS6,2017-09-21,A,2017 autumn new underwear men,yellow,
37,2840,2017-09-25,5831,This price is very cost effective
200,GKLW-I-2897-#wQ7,2017-09-22,A,2017 Autumn New Jeans Men,blue,39,5879,2017-09-25,7200,The
clothes are very comfortable to wear
300,HWEC-L-2531-#xP8,2017-09-23,A,2017 autumn new shoes women,brown,M,403,2017-09-26,607,good
100,IQPD-M-3214-#yQ1,2017-09-24,B,2017 Autumn New Wide Leg Pants Women,black,M,
3045,2017-09-27,5021,very good.
350,LPEC-N-4572-#zX2,2017-09-25,B,2017 Autumn New Underwear Women,red,M,239,2017-09-28,407,The
seller's service is very good
110,NQAB-O-3768-#sM3,2017-09-26,B,2017 autumn new underwear women,red,S,
6089,2017-09-29,7021,The color is very good
210,HWNB-P-7879-#tN4,2017-09-27,B,2017 autumn new underwear women,red,L,3201,2017-09-30,4059,I
like it very much and the quality is good.
230,JKHU-Q-8865-#uO5,2017-09-29,C,2017 Autumn New Clothes with Chiffon Shirt,black,M,
2056,2017-10-02,3842,very good
```

## 在 MRS 集群上创建 Hive ORC 表

1. 创建了MRS集群。

具体操作请参见《MapReduce服务管理指南》的[购买自定义集群](#)。

2. 下载客户端。
  - a. 在MRS集群页面，单击集群名称进入“概览”，单击“前往Manager”，如果提示绑定公网IP，请先绑定公网IP。
  - b. 输入MRS Manager的用户名admin和密码，密码为创建MRS集群时输入的admin密码。
  - c. 登录成功后，选择“服务管理 > 下载客户端”，“客户端类型”选择“仅配置文件”，“下载路径”选择“服务器端”。单击“确定”。

## 下载客户端

警告：生成客户端会占用大量的磁盘IO，不建议在集群处于安装中、启动中、打补丁中等非稳态场景进行“下载客户端”操作。

\* 客户端类型  完整客户端  仅配置文件

\* 下载路径  服务器端  远端主机

仅保存到服务器如下路径，如果存在客户端文件，会覆盖路径下已有的客户端文件。

客户端路径

确定

取消

3. 登录MRS集群的Hive客户端。
  - a. 登录Master节点。

具体操作，请参见《MapReduce服务用户指南》中的[登录集群节点](#)。
  - b. 执行以下命令切换用户。

```
sudo su - omm
```
  - c. 执行以下命令切换到客户端目录：

```
cd /opt/client
```
  - d. 执行以下命令配置环境变量：

```
source bigdata_env
```
  - e. 如果当前集群已启用Kerberos认证，执行以下命令认证当前用户，当前用户需要具有创建Hive表的权限，配置拥有对应权限的角色，具体操作请参见《MapReduce服务用户指南》中的[创建用户](#)。为用户绑定对应角色。如果当前集群未启用Kerberos认证，则无需执行此命令。

```
kinit MRS集群用户
```

例如，`kinit hiveuser`。
  - f. 执行以下命令启动Hive客户端：

```
beeline
```
4. 在Hive中创建数据库demo。

执行以下命令创建数据库：

```
CREATE DATABASE demo;
```
5. 在数据库demo中新建了一个Hive TEXTFILE类型的表product\_info，并将[数据文件](#)（product\_info.txt）导入到该表对应的HDFS路径中。

执行以下命令切换到demo数据库：

```
USE demo;
```

执行以下命令，创建表product\_info，表字段按照[数据文件](#)中的数据进行定义：

```
DROP TABLE product_info;
```

```
CREATE TABLE product_info
```

```
product_price      int      ,
product_id         char(30) ,
product_time       date      ,
product_level      char(10)  ,
product_name       varchar(200),
product_type1      varchar(20) ,
product_type2      char(10)  ,
product_monthly_sales_cnt int  ,
product_comment_time date    ,
product_comment_num int      ,
product_comment_content varchar(200)
)
row format delimited fields terminated by ';'
stored as TEXTFILE;
```

有关导入数据到MRS集群的操作，请参见《MapReduce服务用户指南》中的[管理数据文件](#)章节。

6. 在数据库demo中创建了一个Hive ORC表product\_info\_orc。

执行以下命令，创建Hive ORC表product\_info\_orc，表字段与上一步创建的表product\_info完全一致：

```
DROP TABLE product_info_orc;

CREATE TABLE product_info_orc
(
  product_price      int      ,
  product_id         char(30) ,
  product_time       date      ,
  product_level      char(10)  ,
  product_name       varchar(200),
  product_type1      varchar(20) ,
  product_type2      char(10)  ,
  product_monthly_sales_cnt int  ,
  product_comment_time date    ,
  product_comment_num int      ,
  product_comment_content varchar(200)
)
row format delimited fields terminated by ';'
stored as orc;
```

7. 将product\_info表的数据插入到Hive ORC表product\_info\_orc中。

```
insert into product_info_orc select * from product_info;
```

查询表product\_info\_orc：

```
select * from product_info_orc;
```

如果查询到如[数据文件](#)所示的数据，表示已经成功将数据插入到ORC表。

### 4.2.3.3 手动创建外部服务器

创建外表语法（CREATE FOREIGN TABLE (SQL on Hadoop or OBS)）中，需指定一个与MRS数据源连接相关联的外部服务器。

当您通过GaussDB(DWS)管理控制台创建MRS数据源连接时，数据库管理员dbadmin会在默认数据库postgres中自动创建一个外部服务器。因此，如果您希望在默认数据库postgres中创建外表读取MRS数据，可以跳过本章节。

如果您希望使用普通用户在自定义数据库中创建外表读取MRS数据，必须先在自定义数据库中手动创建一个外部服务器。本章节将为您介绍，如何使用普通用户在自定义数据库中创建外部服务器。步骤如下：

1. 请确保GaussDB(DWS)集群已创建MRS数据源连接。  
具体操作请参见《数据仓库服务管理指南》的[创建MRS数据源连接](#)。
2. [创建用户和数据库并授予外表权限](#)



### 3. 手动创建外部服务器

#### 📖 说明

需要注意的是，当您不再需要从该MRS数据源读取数据时，通过GaussDB(DWS)管理控制台删除MRS数据源，仅会删除在默认数据库postgres中自动创建的外部服务器，手动创建的外部服务器需要您手动删除，具体操作请参见[删除手动创建的外部服务器](#)中的描述。

## 创建用户和数据库并授予外表权限

以下示例，是新建一个普通用户dbuser并创建一个数据库mydatabase，然后使用管理员用户授予dbuser外表权限。

**步骤1** 使用数据库管理员通过GaussDB(DWS)提供的数据库客户端连接默认数据库gaussdb

例如，使用gsq客户端的用户通过如下语句连接数据库：

```
gsq -d gaussdb -h 192.168.2.30 -U dbadmin -p 8000 -W password -r
```

**步骤2** 新建一个普通用户，并用它创建一个数据库。

新建一个具有创建数据库权限的用户dbuser：

```
CREATE USER dbuser WITH CREATEDB PASSWORD 'password';
```

切换为新建的用户：

```
SET ROLE dbuser PASSWORD 'password';
```

执行以下命令创建数据库：

```
CREATE DATABASE mydatabase;
```

查询数据库：

```
SELECT * FROM pg_database;
```

返回结果中有mydatabase的信息表示创建成功：

datname	datdba	encoding	datcollate	datctype	datistemplate	datallowconn	datconnlimit	datlastsysoid	datfrozensid	dattablespace	datcompatibility	datacl
template1	10	0	C	t	t	-1	14146	1351				
1663	ORA		{=c/Ruby,omm=CTc/Ruby}									
template0	10	0	C	t	f	-1	14146	1350				
1663	ORA		{=c/Ruby,Ruby=CTc/Ruby}									
gaussdb	10	0	C	f	t	-1	14146	1352				
1663	ORA		{=Tc/Ruby,Ruby=CTc/Ruby,chaojun=C/Ruby,huobinru=C/Ruby}									
mydatabase	17000	0	C	f	t	-1	14146	1351				
1663	ORA											

(4 rows)

**步骤3** 使用管理员用户给普通用户赋予创建外部服务器的权限和使用外表的权限。

使用数据库管理员用户通过数据库客户端连接新建的数据库。

例如，使用gsq客户端的用户可以直接使用如下语句切换为管理员用户去连接新建的数据库：

```
\c mydatabase dbadmin,
```

根据提示输入用户密码。

## 说明

注意，必须先使用管理员用户连接到**将要创建外部服务器和使用外表的数据库**，再对普通用户进行授权。

默认只有系统管理员才可以创建外部服务器，普通用户需要授权才可以创建，执行以下命令授权：

```
GRANT ALL ON FOREIGN DATA WRAPPER hdf5_fdw TO dbuser;
```

其中FOREIGN DATA WRAPPER的名字只能是hdf5\_fdw，dbuser为创建SERVER的用户名。

执行以下命令赋予用户使用外表的权限。

```
ALTER USER dbuser USEFT;
```

查看用户：

```
SELECT r.rolname, r.rolsuper, r.rolinherit,
       r.rolcreatorole, r.rolcreatedb, r.rolcanlogin,
       r.rolconnlimit, r.rolvalidbegin, r.rolvaliduntil,
       ARRAY(SELECT b.rolname
             FROM pg_catalog.pg_auth_members m
             JOIN pg_catalog.pg_roles b ON (m.roleid = b.oid)
             WHERE m.member = r.oid) as memberof
FROM pg_catalog.pg_roles r
ORDER BY 1;
```

返回结果中，dbuser的信息中包含了UseFT权限，表示授权成功：

rolname	rolsuper	rolinherit	rolcreatorole	rolcreatedb	rolcanlogin	rolconnlimit	rolvalidbegin	rolvaliduntil	memberof	rolreplication	rolauditadmin	rolsystemadmin	roluseft
dbuser	f	t	f	t	t	-1			{}	f			f
lily	f	t	f	f	t	-1			{}	f			
Ruby	t	t	t	t	t	-1			{}	t			

----结束

## 手动创建外部服务器

**步骤1** 使用数据库管理员通过GaussDB(DWS)提供的数据库客户端连接默认数据库postgres。

例如：通过gsql客户端登录数据库的用户可以使用以下两种方法中的一种进行连接：

可以通过以下两种方法中的一种进行连接：

- 如果已经登录了gsql客户端，可以执行以下命令切换数据库和用户：  

```
\c postgres dbadmin;
```

 根据提示输入密码。
- 如果尚未登录gsql客户端，或者已经登录了gsql客户端执行\q退出gsql后，执行以下命令重新进行连接：  

```
gsql -d postgres -h 192.168.2.30 -U dbadmin -p 8000 -W password -r
```

**步骤2** 执行以下命令查询自动创建的外部服务器的信息。

```
SELECT * FROM pg_foreign_server;
```

返回结果如：

srvname	srvowner	srvfdw	srvtype	srvversion	srvacl
gsmpp_server		10   13673			
gsmpp_errorinfo_server		10   13678			
hdfs_server_8f79ada0_d998_4026_9020_80d6de2692ca		16476   13685			

```
{"address=192.168.1.245:25000,192.168.1.218:25000",hdfscfgpath=/MRS/8f79ada0-d998-4026-9020-80d6de2692ca,type=hdfs}
```

(3 rows)

查询结果中，每一行代表一个外部服务器的信息。与MRS数据源连接相关联的外部服务器包含以下信息：

- `srvname`值包含“hdfs\_server”字样以及MRS集群的ID，此ID与MRS管理控制台的集群列表MRS ID相同。
- `srvoptions`字段中的`address`参数为MRS集群的主备节点的IP地址及端口。

您可以根据上述信息找到您所要的外部服务器，并记录下它的`srvname`和`srvoptions`的值。

**步骤3** 切换为即将创建外部服务器的用户去连接其对应的数据库。

在本示例中，执行以下命令，使用[创建用户和数据库并授予外表权限](#)中创建的普通用户`dbuser`连接其创建的数据库`mydatabase`。

```
\c mydatabase dbuser;
```

**步骤4** 创建外部服务器。

创建外部服务器的详细语法，请参见CREATE SERVER。示例如下：

```
CREATE SERVER hdfs_server_8f79ada0_d998_4026_9020_80d6de2692ca FOREIGN DATA WRAPPER
HDFS_FDW
OPTIONS
(
address '192.168.1.245:25000,192.168.1.218:25000';
hdfscfgpath '/MRS/8f79ada0-d998-4026-9020-80d6de2692ca';
type 'hdfs'
);
```

以下为必选参数的说明：

- 外部服务器名称  
允许用户自定义名字。  
在本例中，指定为前面的步骤[步骤2](#)中记录下来的`srvname`字段的值，如'`hdfs_server_8f79ada0_d998_4026_9020_80d6de2692ca`'。  
不同的数据库之间资源是隔离的，因此在不同的数据库中外部服务器名称可以相同。
- FOREIGN DATA WRAPPER  
只能指定为HDFS\_FDW，它在数据库中已经存在。
- OPTIONS参数  
以下参数请分别指定为步骤[步骤2](#)中记录下来的`srvoptions`中的参数值。
  - `address`  
指定HDFS集群的主备节点所在的IP地址以及端口。

- hdfsconfigpath  
指定HDFS集群配置文件路径。该参数仅支持type为HDFS时设置。只能设置一个路径。
- type  
取值为'hdfs'，表示HDFS\_FDW连接的是HDFS。

#### 步骤5 查看外部服务器。

```
SELECT * FROM pg_foreign_server WHERE
srvname='hdfs_server_8f79ada0_d998_4026_9020_80d6de2692ca';
```

返回结果如下所示，表示已经创建成功：

srvname	srvowner	srvfdw	srvtype	srvversion	srvacl
hdfs_server_8f79ada0_d998_4026_9020_80d6de2692ca	16476	13685			

```

{"address=192.168.1.245:25000,192.168.1.218:25000",hdfsconfigpath=/MRS/8f79ada0-d998-4026-9020-80d6de2692ca,type=hdfs}
(1 row)

```

----结束

### 4.2.3.4 创建外表

在GaussDB(DWS)数据库中创建一个Hadoop外表，用来访问存储在MRS HDFS文件系统上的Hadoop结构化数据。Hadoop外表是只读的，只能用于查询操作，可直接使用SELECT查询其数据。

#### 前提条件

- 已创建MRS集群，并将数据导入Hive/Spark数据库中的ORC表。  
请参见[MRS集群上的数据准备](#)。
- GaussDB(DWS)集群已创建MRS数据源连接。  
具体操作请参见《数据仓库服务管理指南》的[创建MRS数据源连接](#)。

### 获取 MRS 数据源的 HDFS 路径

有两种方法可以查看：

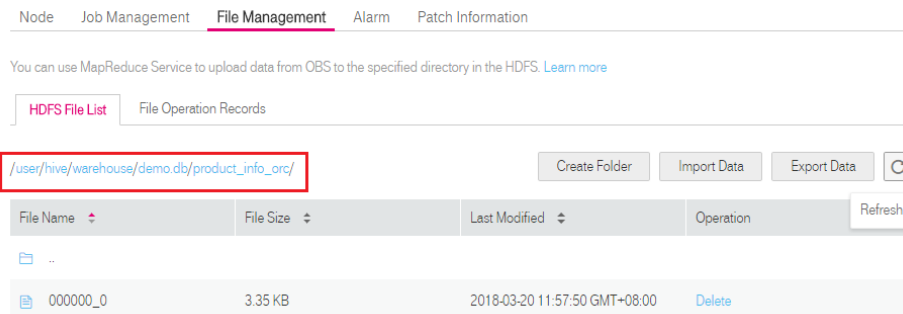
- **方法一：**  
对于Hive数据，可以登录MRS的Hive客户端（参见2），执行以下命令查看表的详细信息，并记录下location参数中的数据存储路径。  

```
use <database_name>;
desc formatted <table_name>;
```

 例如，返回结果中location参数值为“hdfs://hacluster/user/hive/warehouse/demo.db/product\_info\_orc/”，则记录HDFS路径为“/user/hive/warehouse/demo.db/product\_info\_orc/”。
- **方法二：**  
按以下步骤获取HDFS路径。
  - a. 登录MRS管理控制台。
  - b. 选择“集群列表 > 现有集群”，单击要查看的集群名称，进入集群基本信息页面。

- c. 单击“文件管理”，选择“HDFS文件列表”。
- d. 进入您要导入到GaussDB(DWS)集群的数据的存储目录，并记录其路径。

图 4-7 在 MRS 上查看数据存储路径



## 获取 MRS 数据源连接的外部服务器信息

**步骤1** 使用创建外部服务器的用户去连接其对应的数据库。

是否使用普通用户在自定义数据库中创建外表，请根据需求进行选择：

- 是
  - a. 请先确保，您已按照[手动创建外部服务器](#)章节中的步骤，创建了普通用户 dbuser和它的数据库mydatabase，并在mydatabase中手动创建了一个外部服务器。
  - b. 使用用户dbuser通过GaussDB(DWS)提供的数据库客户端连接数据库 mydatabase。

如果已经使用gsq客户端连接至数据库，可以直接执行如下命令进行用户和数据库切换：

```
\c mydatabase dbuser;
```

根据界面提示输入密码。

- 否
 

当您通过GaussDB(DWS)管理控制台创建MRS数据源连接时，数据库管理员 dbadmin会在默认数据库postgres中自动创建一个外部服务器。因此，如果使用数据库管理员dbadmin在默认数据库postgres中创建外表，需要通过 GaussDB(DWS)提供的数据库客户端工具连接数据库。例如，使用gsq客户端的用户通过如下命令连接数据库：

```
gsq -d postgres -h 192.168.2.30 -U dbadmin -p 8000 -W password -r
```

**步骤2** 执行以下命令，查看已创建的MRS数据源连接的外部服务器信息。

```
SELECT * FROM pg_foreign_server;
```

### 说明

也可以执行\desc+命令查看外部服务器信息。

返回结果如：

srvname	srvowner	srvfdw	srvtype	srvversion	srvacl
gsmpp_server	10	13673			
gsmpp_errorinfo_server	10	13678			

```
hdfs_server_8f79ada0_d998_4026_9020_80d6de2692ca | 16476 | 13685 | | | |  
{"address=192.168.1.245:25000,192.168.1.218:25000",hdfscfgpath=/MRS/8f79ada0-  
d998-4026-9020-80d6de2692ca,type=hdfs}  
(3 rows)
```

查询结果中，每一行代表一个外部服务器的信息。与MRS数据源连接相关联的外部服务器包含以下信息：

- `srvname`值包含“`hdfs_server`”字样以及MRS集群的ID，此ID与MRS管理控制台的集群列表MRS ID相同。
- `srvoptions`字段中的`address`参数为MRS集群的主备节点的IP地址及端口。

您可以根据上述信息找到您所要的外部服务器，并记录下它的`srvname`和`srvoptions`的值。

----结束

## 创建外表

当完成[获取MRS数据源连接的外部服务器信息](#)和[获取MRS数据源的HDFS路径](#)后，就可以创建一个外表，用于读取MRS数据源数据。

创建外表的语法格式如下，详细的描述请参见（`CREATE FOREIGN TABLE (SQL on Hadoop or OBS)`）。

```
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 dfs_server  
OPTIONS ( { option_name 'value' } [, ...] )  
DISTRIBUTE BY { ROUNDROBIN | REPLICATION }  
[ PARTITION BY ( column_name ) [ AUTOMAPPED ] ] ;
```

例如，创建一个名为“`foreign_product_info`”的外表，对语法中的参数按如下描述进行设置：

- **table\_name**  
必选。外表的表名。
- **表字段定义**
  - **column\_name**：外表中的字段名。
  - **type\_name**：字段的数据类型。多个字段用“,”隔开。  
外表的字段个数和字段类型，需要与MRS上保存的数据完全一致。定义字段的数据类型之前，您必须先了解[数据类型转换说明](#)。
- **SERVER dfs\_server**  
外表的外部服务器名称，这个server必须存在。外表通过设置外部服务器，从而关联MRS数据源连接并从MRS集群读取数据。  
此处应填写为通过[获取MRS数据源连接的外部服务器信息](#)查询到的“`srvname`”字段的值。
- **OPTIONS参数**  
用于指定外表数据的各类参数，关键参数如下所示。
  - **format**：必选参数。取值只支持“`orc`”。表示数据源文件的格式，只支持Hive的ORC数据文件。

- **foldername**: 必选参数。表示数据在HDFS的存储目录或数据文件路径。  
如果是启用了Kerberos认证的MRS分析集群，请确保MRS数据源连接的MRS用户，拥有此目录的读取权限。  
请按照[获取MRS数据源的HDFS路径](#)中的步骤获取HDFS路径，该路径作为**foldername**的参数值。
- **encoding**: 可选参数。外表中数据源文件的编码格式名称，缺省为utf8。
- **DISTRIBUTE BY**  
表示外表的数据读取方式。有以下两种方式供选择，在本例中选择ROUNDROBIN。
  - **ROUNDROBIN**: 表示外表在从数据源读取数据时，GaussDB(DWS)集群每一个节点读取随机一部分数据，并组成完整数据。
  - **REPLICATION**: 表示外表在从数据源读取数据时，GaussDB(DWS)集群每一个节点都读取一份完整数据。
- **语法中的其他参数**  
其他参数均为可选参数，用户可以根据自己的需求进行设置，在本例中不需要设置。

根据以上信息，创建外表命令如下所示：

```
DROP FOREIGN TABLE IF EXISTS foreign_product_info;  
  
CREATE FOREIGN TABLE foreign_product_info  
(  
    product_price          integer      ,  
    product_id            char(30)     ,  
    product_time          date         ,  
    product_level         char(10)    ,  
    product_name          varchar(200) ,  
    product_type1         varchar(20)  ,  
    product_type2         char(10)    ,  
    product_monthly_sales_cnt integer   ,  
    product_comment_time  date        ,  
    product_comment_num   integer     ,  
    product_comment_content varchar(200)  
) SERVER hdfs_server_8f79ada0_d998_4026_9020_80d6de2692ca  
OPTIONS (  
    format 'orc',  
    encoding 'utf8',  
    foldername '/user/hive/warehouse/demo.db/product_info_orc/'  
)  
DISTRIBUTE BY ROUNDROBIN;
```

## 数据类型转换说明

当前用户导入到Hive/Spark的数据在HDFS存储为ORC文件格式，GaussDB(DWS)实际读取HDFS中的ORC文件，并对文件内的数据进行查询分析。

由于Hive/Spark支持的数据类型与GaussDB(DWS)自身支持的数据类型存在差异，在创建外表定义表字段时，您需要了解这两者之间数据类型的对应关系，具体如[表4-11](#)所示：



表 4-11 数据类型匹配表

类型名称	GaussDB(DWS)的HDFS/OBS外表支持的字段类型	Hive表字段类型	Spark表字段类型
2字节整数	SMALLINT	SMALLINT	SMALLINT
4字节整数	INTEGER	INT	INT
8字节整数	BIGINT	BIGINT	BIGINT
单精度浮点数	FLOAT4 (REAL)	FLOAT	FLOAT
双精度浮点型	FLOAT8(DOUBLE PRECISION)	DOUBLE	FLOAT
科学数据类型	DECIMAL[p (,s)] 最大支持38位精度	DECIMAL 最大支持38位 ( Hive 0.11 )	DECIMAL
日期类型	DATE	DATE	DATE
时间类型	TIMESTAMP	TIMESTAMP	TIMESTAMP
Boolean类型	BOOLEAN	BOOLEAN	BOOLEAN
Char类型	CHAR(n)	CHAR (n)	STRING
VarChar类型	VARCHAR(n)	VARCHAR (n)	VARCHAR (n)
字符串	TEXT(CLOB)	STRING	STRING

#### 4.2.3.5 执行数据导入

##### 直接查询外表查看 MRS 数据源的数据

如果数据量较少，可直接使用SELECT查询外表，即可查看到MRS数据源的数据。

**步骤1** 执行以下命令，则可以从外表查询数据。

```
SELECT * FROM foreign_product_info;
```

查询结果显示如[数据文件](#)中所示的数据，表示导入成功。查询结果的结尾将显示以下信息：

```
(20 rows)
```

通过外表查询到数据后，用户可以将数据插入数据库的普通表。

----结束

##### 导入数据后查询数据

也可以将MRS数据导入GaussDB(DWS)后，再查询数据。

**步骤1** 在GaussDB(DWS)数据库中，创建导入数据的目标表，用于存储导入的数据。



该表的表结构必须与[创建外表](#)中创建的外表的表结构保持一致，即字段个数、字段类型要完全一致。

例如，创建一个名为product\_info的表，示例如下：

```
DROP TABLE IF EXISTS product_info;
CREATE TABLE product_info
(
  product_price      integer      ,
  product_id         char(30)     ,
  product_time       date         ,
  product_level      char(10)     ,
  product_name       varchar(200) ,
  product_type1      varchar(20)  ,
  product_type2      char(10)     ,
  product_monthly_sales_cnt integer ,
  product_comment_time date       ,
  product_comment_num integer     ,
  product_comment_content varchar(200)
)
with (
  orientation = column,
  compression=middle
)
DISTRIBUTE BY HASH (product_id);
```

**步骤2** 执行“INSERT INTO .. SELECT ..”命令从外表导入数据到目标表。

示例：

```
INSERT INTO product_info SELECT * FROM foreign_product_info;
```

若出现以下类似信息，说明数据导入成功。

```
INSERT 0 20
```

**步骤3** 执行SELECT命令，查看从MRS导入到GaussDB(DWS)中的数据。

```
SELECT * FROM product_info;
```

查询结果显示如[数据文件](#)中所示的数据，表示导入成功。查询结果的结尾将显示以下信息：

```
(20 rows)
```

----结束

### 4.2.3.6 清除资源

当完成本教程的示例后，如果您不再需要使用本示例中创建的资源，您可以删除这些资源，以免资源浪费或占用您的配额。

#### 删除外表和目标表

**步骤1** （可选）如果执行了[导入数据后查询数据](#)，请执行以下命令，删除目标表。

```
DROP TABLE product_info;
```

**步骤2** 执行以下命令，删除外表。

```
DROP FOREIGN TABLE foreign_product_info;
```

----结束

#### 删除手动创建的外部服务器

如果执行了[手动创建外部服务器](#)，请按照以下步骤删除外部服务器、数据库和用户。

**步骤1** 使用创建外部服务器的用户通过GaussDB(DWS)提供的数据库客户端连接到外部服务器所在的数据库。

例如，使用gsql客户端的用户可以通过以下两种方法中的一种进行连接：

- 如果已经登录了gsql客户端，可以执行以下命令进行切换：  

```
\c mydatabase dbuser;
```

根据提示输入密码。
- 如果已经登录了gsql客户端，您也可以执行\q退出gsql后，再执行以下命令重新进行连接：  

```
gsql -d mydatabase -h 192.168.2.30 -U dbuser -p 8000 -r
```

根据提示输入密码。

**步骤2** 删除手动创建的外部服务器。

执行以下命令进行删除，详细语法请参见DROP SERVER：

```
DROP SERVER hdfs_server_8f79ada0_d998_4026_9020_80d6de2692ca;
```

返回以下信息表示删除成功：

```
DROP SERVER
```

查看外部服务器：

```
SELECT * FROM pg_foreign_server WHERE  
srvname='hdfs_server_8f79ada0_d998_4026_9020_80d6de2692ca';
```

返回结果如下所示，表示已经删除成功：

```
srvname | srvowner | srvfdw | srvtype | srversion | srvacl | srvoptions  
-----+-----+-----+-----+-----+-----+-----  
(0 rows)
```

**步骤3** 删除自定义数据库。

通过GaussDB(DWS)提供的数据库客户端连接默认数据库postgres。

如果已经登录了gsql客户端，可以直接执行如下命令进行切换：

```
\c postgres
```

根据界面提示输入密码。

执行以下命令，删除自定义数据库：

```
DROP DATABASE mydatabase;
```

返回以下信息表示删除成功：

```
DROP DATABASE
```

**步骤4** 使用管理员用户，删除本示例中创建的普通用户。

使用数据库管理员用户通过GaussDB(DWS)提供的数据库客户端连接数据库。

如果已经登录了gsql客户端，可以直接执行如下命令进行切换：

```
\c postgres dbadmin
```

执行以下命令回收创建外部服务器的权限：

```
REVOKE ALL ON FOREIGN DATA WRAPPER hdfs_fdw FROM dbuser;
```

其中FOREIGN DATA WRAPPER的名字只能是hdfs\_fdw，dbuser为创建SERVER的用户名。

执行以下命令删除用户：

```
DROP USER dbuser;
```

可使用`\du`命令查询用户，确认用户是否已经删除。

---结束

### 4.2.3.7 错误处理

如下错误信息，表示GaussDB(DWS)期望读取ORC数据文件，但实际却是\*.txt类型的数据文件。请先创建Hive ORC类型的表，并将数据存储到该Hive ORC表中。

```
ERROR: dn_6009_6010: Error occurs while creating an orc reader for file /user/hive/warehouse/products_info.txt, detail can be found in dn log of dn_6009_6010.
```

## 4.2.4 从 GaussDB(DWS)集群导入数据到新集群

### 功能描述

通过在集群中创建Foreign Table的方式，实现在多个集群之间的关联查询和用来导入数据。

### 使用场景

- 将数据从一个GaussDB(DWS)集群导入到另外一个GaussDB(DWS)集群中。
- 多个集群之间的关联查询。

### 注意事项

- 两个集群必须在同一个Region、一个AZ内且VPC网络互通。
- 创建的外表与其对应的远端表的列名和类型名要完全一致，且远端表的类型为行存表、列存表、哈希表或者复制表。
- 如果关联的表在另外一个集群是复制表或者存在数据倾斜，性能可能会很差。
- 使用期间，两个集群的状态应为“Normal”。
- 使用期间，禁止对远端集群的源数据表做ddl修改和增、删、改操作，否则可能导致查询结果不一致。
- 两个集群都需要具备基于Foreign Table的SQL on other GaussDB数据处理功能。
- 建议配置LVS，如未配置，推荐使用多个CN作为server的地址，禁止将多个集群的CN地址写在一起。
- 请尽可能保证两端数据库的编码相同，否则可能出现报错或者收到的数据为乱码。
- 如果远端表已经做过统计信息收集，可以对外表执行analyze以获得更优的执行计划。
- 仅支持8.0.0及以上版本。

### 操作步骤

#### 步骤1 创建server。

```
CREATE SERVER server_remote FOREIGN DATA WRAPPER GC_FDW OPTIONS  
(address '10.180.157.231:8000,10.180.157.130:8000',
```

```
dbname 'gaussdb',
username 'xyz',
password 'xxxxxx'
);
```

### 说明

- server\_remote为server名字，供外表使用。
- address为远端集群CN的地址和端口号，如配置LVS，推荐只填写一个LVS地址，如未配置，推荐使用多个CN作为server的地址。
- dbname为远端集群的数据库名。
- username为连接远端集群使用的用户名，注意该用户不能为系统管理员。
- password为连接远端集群使用的用户名的密码。

### 步骤2 创建外表。

```
CREATE FOREIGN TABLE region
(
  R_REGIONKEY INT4,
  R_NAME TEXT,
  R_COMMENT TEXT
)
SERVER
  server_remote
OPTIONS
(
  schema_name 'test',
  table_name 'region',
  encoding 'gbk'
);
```

### 说明

- 外表的列不允许带任何约束。
- 外表的列名和列的类型要与远端集群对应的表的列名和列的类型完全一致。
- schema\_name为远端集群对应的表所在的schema，如果该option省略，则schema\_name预设该外表所在的schema。
- table\_name为远端集群对应的表所在的表名，如果该option省略，则table\_name预设该外表的表名。
- encoding为远端集群的编码，如果该option省略，则编码使用远端集群数据库的默认编码。

### 步骤3 查看建立的外表。

```
\d+ region

               Foreign table "public.region"
  Column   | Type   | Modifiers | FDW Options | Storage | Stats target | Description
-----+-----+-----+-----+-----+-----+-----
 r_regionkey | integer |           |             | plain   |              |
 r_name     | text   |           |             | extended |              |
 r_comment  | text   |           |             | extended |              |
Server: server_remote
FDW Options: (schema_name 'test', table_name 'region', encoding 'gbk')
FDW permission: read only
Has OIDs: no
Distribute By: ROUND ROBIN
Location Nodes: ALL DATANODES
```

### 步骤4 查看建立的server。

```
\des+ server_remote

List of foreign servers
  Name   | Owner | Foreign-data wrapper | Access privileges | Type | Version |
-----+-----+-----+-----+-----+-----+-----
                FDW Options | Description
```

```
+-----+
+-----+
server_remote | dbadmin | gc_fdw | | | | (address
'10.180.157.231:8000,10.180.157.130:8000', dbname 'gaussdb'
, username 'xyz', password 'xxxxxx') |
(1 row)
```

#### 步骤5 使用外表进行导入数据或者关联查询。

- 导入数据。

```
CREATE TABLE local_region
(
  R_REGIONKEY INT4,
  R_NAME TEXT,
  R_COMMENT TEXT
);
INSERT INTO local_region SELECT * FROM region;
```

##### 📖 说明

- 如遇到报错连接失败，请检查server的信息确认两个集群是否已经相互连通。
- 如遇到报错表不存在，请检查外表的option信息是否正确。
- 如遇到报错列信息不匹配，请检查外表的列信息是否与远端集群对应表的列信息是否一致。
- 如遇到报错版本不一致，请升级低版本的集群在继续使用。
- 如遇到乱码，请检查数据源的实际编码方式，并重新创建外表指定正确的编码。

- 关联查询。

```
SELECT * FROM region, local_region WHERE local_region.R_NAME = region.R_NAME;
```

##### 📖 说明

- 外表可以当做一个本地表来使用，执行复杂的作业。
- 如果远端集群已经有统计信息，请对该外表执行analyze以获得更优的执行计划。
- 如果本地集群的DN数量比远端集群的DN数量少，本地集群需要使用SMP来获得更佳的性能。

#### 步骤6 删除外表。

```
DROP FOREIGN TABLE region;
```

----结束

## 4.2.5 基于 GDS 的跨集群互联互通

### 功能描述

在“基于Foreign Table的数据处理”的基础上，通过GDS进行数据中转，实现多个集群之间的数据同步。

### 使用场景

将数据从一个集群同步到另外一个集群，支持全量数据同步、过滤条件数据同步。

### 注意事项

- 创建的互联互通外表与其对应的远端表的列名和类型名要完全一致，且远端表的类型为行存表或列存表。
- 执行同步语句时，要确保本地集群、远端集群的待同步表已存在。

- 使用期间，两个集群的状态应为Normal。
- 两个集群都需要具备基于GDS的跨集群互联互通功能。
- 建议两端集群的数据库编码保持一致，否则可能出现报错或者收到的数据为乱码。
- 两端集群所指定的数据库兼容类型要保持一致，否则可能报错或乱码。
- 确保执行数据同步的相关用户对待同步表有相应的访问权限。
- 互联互通外表只能用于跨集群数据同步场景，其他场景可能出错或无效。
- 互联互通外表不支持复杂的列上表达式，不支持复杂语法，包括join、排序、游标、with、集合等。
- 不下推的SQL语句无法使用本特性进行数据同步，否则会报错。
- 不支持EXPLAIN计划、逻辑集群。
- 当本地集群同步数据到远端集群时，只支持内表查询。
- Foreign Server的syncsrv选项指定的GDS不支持SSL模式。
- 数据同步结束时只校验数据行数，不校验数据内容。
- 业务最大并发数不能大于GDS启动参数-t的一半，同时也不能大于max\_active\_statements，否则可能会导致业务超时失败。

## 使用前准备

- 配置两个集群互连。
- 规划部署GDS服务器，确保所有的GDS服务器可以和上面配置的两个集群所有节点网络连通。部署GDS请参考[安装配置和启动GDS](#)。

## 操作步骤

假设远端集群的待同步表名称是tbl\_remote，用于数据同步的用户是user\_remote，该用户须对表tbl\_remote有访问权限；假设本地集群的待同步表名称是tbl\_local。

### 步骤1 创建server。

```
CREATE SERVER server_remote FOREIGN DATA WRAPPER GC_FDW OPTIONS(  
  address '192.168.178.207:8109',  
  dbname 'db_remote',  
  username 'user_remote',  
  password 'xxxxxxxx',  
  syncsrv 'gsfs://192.168.178.129:8789|gsfs://192.168.178.129:8790'  
);
```

- server\_remote为server名称，供互联互通外表使用。
- address为远端集群CN的IP地址和端口，仅允许填写一个地址。
- dbname为远端集群的数据库名。
- username为连接远端集群使用的用户名，注意该用户不能为系统管理员。
- password为连接远端集群使用的用户名的密码。
- syncsrv为GDS Server的IP地址和端口，如果有多个地址使用|分割，与GDS外表的location类似。

### 📖 说明

GaussDB(DWS)会对syncsrv所设置的GDS地址进行网络连接测试：

- 只能判断本地执行集群与GDS的网络情况，无法判断远端集群与GDS的网络情况，需要注意报错提示。
- 在移除不可用GDS后，从中选择不会导致业务hang的、数目适当的GDS进行数据同步。

### 步骤2 创建互联互通外表。

```
CREATE FOREIGN TABLE ft_tbl(  
  col_1 type_name,  
  col_2 type_name,  
  ...  
) SERVER server_remote OPTIONS (  
  schema_name 'schema_remote',  
  table_name 'tbl_remote',  
  encoding 'utf8'  
);
```

- schema\_name为远端集群表所属schema，如果该option缺省，则schema\_name预设为该外表所在的schema。
- table\_name为远端集群表名，如果该option缺省，则table\_name预设为该外表的表名。
- encoding为远端集群的编码，如果该option缺省，则编码使用本地集群数据库的默认编码。

### 📖 说明

- 选项schema\_name、table\_name大小写敏感，必须与远端schema、table的名字大小写保持一致。
- 互联互通外表的列不允许带任何约束。
- 互联互通外表的列名、列类型必须与远端集群的表tbl\_remote的列名和列类型完全一致。
- SERVER须设置为**步骤1**中新建的server，必须包含syncsrv属性。

### 步骤3 使用互联互通外表进行数据同步。

- 本地集群是目标集群时，发起数据同步业务：

全列全量数据同步：

```
INSERT INTO tbl_local SELECT * FROM ft_tbl;
```

全列过滤条件数据同步：

```
INSERT INTO tbl_local SELECT * FROM ft_tbl WHERE col_2 = XX;
```

部分列全量数据同步：

```
INSERT INTO tbl_local (col_1) SELECT col_1 FROM ft_tbl;
```

部分列过滤条件数据同步：

```
INSERT INTO tbl_local (col_1) SELECT col_1 FROM ft_tbl WHERE col_2 = XX;
```

- 本地集群是源集群时，发起数据同步业务：

单表数据同步：

```
INSERT INTO ft_tbl SELECT * FROM tbl_local;
```

join结果集数据同步：

```
INSERT INTO ft_tbl SELECT * FROM tbl_local1 join tbl_local2 ON XXX;
```

### 📖 说明

- 如遇到报错连接失败，请检查server的信息确认两个集群是否已经相互连通。
- 如遇到报错GDS连接失败，请检查syncsrv指定的GDS Server是否都已经启动，且与两个集群所有节点可以网络连通。
- 如遇到报错表不存在，请检查外表的option信息是否正确。
- 如遇到报错列不存在，请检查外表的列名是否与源表一致。
- 如遇到报错列重复定义，请检查是否相应列名超长，若超长建议使用AS别名精简。
- 如遇到报错无法解析列类型，请检查语句中是否有列上表达式。
- 如遇到报错列信息不匹配，请检查外表的列信息是否与远端集群对应表的列信息是否一致。
- 如遇到报错语法不支持，请检查是否使用了Join、distinct、排序等复杂用法。
- 如遇到乱码，请检查两端数据库的实际编码是否一致。
- 当本地集群是源集群时，存在极小的概率出现数据成功同步到远端集群，但是本地集群返回执行失败的情况，针对这种情况建议校验同步数据记录数。
- 当本地集群是源集群时，通过事务块、子事务等控制的数据同步，需要总事务提交后才能查询到数据同步结果。

#### 步骤4 删除互联互通外表。

```
DROP FOREIGN TABLE ft_tbl;
```

----结束

## 4.2.6 使用 gsql 元命令导入数据

GaussDB(DWS)的gsql工具提供了元命令\copy进行数据导入。

### \copy 命令

\copy命令格式以及说明参见[表 1 \copy元命令说明](#)。

表 4-12 \copy 元命令说明

语法	说明
<pre>\copy { table [ ( column_list ) ]   ( query ) } { from   to } { filename   stdin   stdout   pstdin   pstdout } [ with ] [ binary ] [ oids ] [ delimiter [ as ] 'character' ] [ null [ as ] 'string' ] [ csv [ header ] [ quote [ as ] 'character' ] [ escape [ as ] 'character' ] ] [ force quote column_list   * ] [ force not null column_list ] ]</pre>	<p>在任何gsql客户端登录数据库成功后，可以使用该命令进行数据的导入/导出。但是与SQL的COPY命令不同，该命令读取/写入的文件是本地文件，而非数据库服务器端文件；所以，要操作的文件的可访问性、权限等，都是受限于本地用户的权限。</p> <p><b>说明</b> \COPY只适合小批量，格式良好的数据导入，容错能力较差。导入数据应优先选择GDS或COPY。</p>

### 参数说明

- table



表的名字（可以有模式修饰）。

取值范围：已存在的表名。

- `column_list`  
可选的待复制字段列表。  
取值范围：任意字段。如果没有声明字段列表，将使用所有字段。
- `query`  
其结果将被复制。  
取值范围：一个必须用圆括弧包围的SELECT或VALUES命令。
- `filename`  
文件的绝对路径。执行copy命令的用户必须有此路径的写权限。
- `stdin`  
声明输入是来自标准输入。
- `stdout`  
声明输出打印到标准输出。
- `pstdin`  
声明输入是来自gsq的标准输入。
- `pstdout`  
声明输出打印到gsq的标准输出。
- `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。
- `oid`  
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。

取值范围：true/on，false/off。

默认值：false

- `delimiter [ as ] 'character'`  
指定数据文件行数据的字段分隔符。

#### 📖 说明

- 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：\abcdefghijklmnopqrstuvwxy0123456789。
- 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E'\x07'，E'\x08'，E'\x1b'等。

取值范围：支持多字符分隔符，但分隔符不能超过10个字节。

默认值：

- 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 [ as ] 'string'  
用来指定数据文件中空值的表示。  
取值范围：
  - 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。默认值：
  - 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。
- 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，则数据文本第一行会被识别为标题行，会忽略此行。如果header为off，而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是指定导出数据包含标题行的定义文件。如果header为off，则导出数据文件不包含标题行。  
取值范围：true/on，false/off。  
默认值：false
- quote [ as ] 'character'  
CSV格式文件下的引号字符。  
默认值：双引号。  
**说明**
  - 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E'\x07'，E'\x08'，E'\x1b'等。
- escape [ as ] '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。  
默认值：双引号。当与quote值相同时，会被替换为'\0'。
- force quote column\_list | \*  
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。  
取值范围：已存在的字段。
- force not null column\_list  
在CSV COPY FROM模式下，指定的字段输入不能为空。  
取值范围：已存在的字段。

## 任务示例

1. 创建目标表a。  

```
CREATE TABLE a(a int);
```

## 2. 导入数据。

### a. 从stdin复制数据到目标表a。

```
\copy a from stdin;
```

出现>>符号提示时，输入数据，输入\.时结束。

```
Enter data to be copied followed by a newline.  
End with a backslash and a period on a line by itself.  
>> 1  
>> 2  
>> \.
```

查询导入目标表a的数据。

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

### b. 从本地文件复制数据到目标表a。假设存在本地文件/home/omm/2.csv。

- 分隔符为 ‘，’ 。
- 在导入过程中，若数据源文件比外表定义的列数多，则忽略行尾多出来的列。

```
\copy a FROM '/home/omm/2.csv' WITH (delimiter',';IGNORE_EXTRA_DATA 'on');
```

## 4.2.7 使用 COPY FROM STDIN 导入数据

### 4.2.7.1 关于 COPY FROM STDIN 导入数据

这种方式适合数据写入量不太大，并发度不太高的场景。

用户可以使用以下方式通过COPY FROM STDIN语句直接向GaussDB(DWS)写入数据。

- 通过键盘输入向GaussDB(DWS)写入数据。
- 通过JDBC驱动的CopyManager接口从文件或者数据库向GaussDB(DWS)写入数据。此方法支持COPY语法中copy option的所有参数。

### 4.2.7.2 CopyManager 类简介

CopyManager是GaussDB(DWS) JDBC驱动中提供的一个API接口类，用于批量向GaussDB(DWS)集群中导入数据。

### CopyManager 的继承关系

CopyManager类位于org.postgresql.copy Package中，继承自java.lang.Object类，该类的声明如下：

```
public class CopyManager  
extends Object
```

### 构造方法

```
public CopyManager(BaseConnection connection)  
throws SQLException
```

## 常用方法

表 4-13 CopyManager 常用方法

返回值	方法	描述	throws
CopyIn	copyIn(String sql)	-	SQLException
long	copyIn(String sql, InputStream from)	使用COPY FROM STDIN从InputStream中快速向数据库中的表导入数据。	SQLException,IOException
long	copyIn(String sql, InputStream from, int bufferSize)	使用COPY FROM STDIN从InputStream中快速向数据库中的表导入数据。	SQLException,IOException
long	copyIn(String sql, Reader from)	使用COPY FROM STDIN从Reader中快速向数据库中的表导入数据。	SQLException,IOException
long	copyIn(String sql, Reader from, int bufferSize)	使用COPY FROM STDIN从Reader中快速向数据库中的表导入数据。	SQLException,IOException
CopyOut	copyOut(String sql)	-	SQLException
long	copyOut(String sql, OutputStream to)	将一个COPY TO STDOUT的结果集从数据库发送到OutputStream类中。	SQLException,IOException
long	copyOut(String sql, Writer to)	将一个COPY TO STDOUT的结果集从数据库发送到Writer类中。	SQLException,IOException

### 4.2.7.3 示例：通过本地文件导入导出数据

在使用JAVA语言基于GaussDB(DWS)进行二次开发时，可以使用CopyManager接口，通过流方式，将数据库中的数据导出到本地文件或者将本地文件导入数据库中，文件格式支持CSV、TEXT等格式。

样例程序如下，执行时需要加载GaussDB(DWS) jdbc驱动。

```
//以下用例以gsjdbc4.jar为例，如果要使用gsjdbc200.jar，请替换驱动类名（将代码中的“org.postgresql”替换成“com.huawei.gauss200.jdbc”）与连接URL串前缀（将“jdbc:postgresql”替换为“jdbc:gaussdb”）。
import java.sql.Connection;
```

```
import java.sql.DriverManager;
import java.io.IOException;
import java.io.FileInputStream;
import java.io.FileOutputStream;
import java.sql.SQLException;
import org.postgresql.copy.CopyManager;
import org.postgresql.core.BaseConnection;

public class Copy{

    public static void main(String[] args)
    {
        String urls = new String("jdbc:postgresql://10.180.155.74:8000/gaussdb"); //数据库URL
        String username = new String("jack"); //用户名
        String password = new String("*****"); //密码
        String tablename = new String("migration_table"); //定义表信息
        String tablename1 = new String("migration_table_1"); //定义表信息
        String driver = "org.postgresql.Driver";
        Connection conn = null;

        try {
            Class.forName(driver);
            conn = DriverManager.getConnection(urls, username, password);
        } catch (ClassNotFoundException e) {
            e.printStackTrace(System.out);
        } catch (SQLException e) {
            e.printStackTrace(System.out);
        }

        // 将SELECT * FROM migration_table查询结果导出到本地文件d:/data.txt
        try {
            copyToFile(conn, "d:/data.txt", "(SELECT * FROM migration_table)");
        } catch (SQLException e) {
            // TODO Auto-generated catch block
            e.printStackTrace();
        } catch (IOException e) {
            // TODO Auto-generated catch block
            e.printStackTrace();
        }

        //将d:/data.txt中的数据导入到migration_table_1中。
        try {
            copyFromFile(conn, "d:/data.txt", tablename1);
        } catch (SQLException e) {
            // TODO Auto-generated catch block
            e.printStackTrace();
        } catch (IOException e) {
            // TODO Auto-generated catch block
            e.printStackTrace();
        }

        // 将migration_table_1中的数据导出到本地文件d:/data1.txt
        try {
            copyToFile(conn, "d:/data1.txt", tablename1);
        } catch (SQLException e) {
            // TODO Auto-generated catch block
            e.printStackTrace();
        } catch (IOException e) {
            // TODO Auto-generated catch block
            e.printStackTrace();
        }
    }

    public static void copyFromFile(Connection connection, String filePath, String tableName)
        throws SQLException, IOException {

        FileInputStream fileInputStream = null;

        try {
            CopyManager copyManager = new CopyManager((BaseConnection)connection);
```

```
        fileInputStream = new FileInputStream(filePath);
        copyManager.copyIn("COPY " + tableName + " FROM STDIN", fileInputStream);
    } finally {
        if (fileInputStream != null) {
            try {
                fileInputStream.close();
            } catch (IOException e) {
                e.printStackTrace();
            }
        }
    }
}

public static void copyToFile(Connection connection, String filePath, String tableOrQuery)
    throws SQLException, IOException {

    FileOutputStream fileOutputStream = null;

    try {
        CopyManager copyManager = new CopyManager((BaseConnection)connection);
        fileOutputStream = new FileOutputStream(filePath);
        copyManager.copyOut("COPY " + tableOrQuery + " TO STDOUT", fileOutputStream);
    } finally {
        if (fileOutputStream != null) {
            try {
                fileOutputStream.close();
            } catch (IOException e) {
                e.printStackTrace();
            }
        }
    }
}
```

#### 4.2.7.4 示例：从 MySQL 向 GaussDB(DWS)进行数据迁移

下面示例演示如何通过CopyManager从mysql向GaussDB(DWS)进行数据迁移的过程。

```
//以下用例以gsjdbc4.jar为例，如果要使用gsjdbc200.jar，请替换驱动类名（将代码中的“org.postgresql”替换成“com.huawei.gauss200.jdbc”）与连接URL串前缀（将“jdbc:postgresql”替换为“jdbc:gaussdb”）。
import java.io.StringReader;
import java.sql.Connection;
import java.sql.DriverManager;
import java.sql.ResultSet;
import java.sql.SQLException;
import java.sql.Statement;

import org.postgresql.copy.CopyManager;
import org.postgresql.core.BaseConnection;

public class Migration{

    public static void main(String[] args) {
        String url = new String("jdbc:postgresql://10.180.155.74:8000/gaussdb"); //数据库URL
        String user = new String("jack"); //mppdb用户名
        String pass = new String("*****"); //mppdb密码
        String tablename = new String("migration_table"); //定义表信息
        String delimiter = new String("|"); //定义分隔符
        String encoding = new String("UTF8"); //定义字符集
        String driver = "org.postgresql.Driver";
        StringBuffer buffer = new StringBuffer(); //定义存放格式化数据的缓存

        try {
            //获取源数据库查询结果集
            ResultSet rs = getDataSet();

            //遍历结果集，逐行获取记录
            //将每条记录中各字段值，按指定分隔符分割，由换行符结束，拼成一个字符串
        }
    }
}
```

```
//把拼成的字符串，添加到缓存buffer
while (rs.next()) {
    buffer.append(rs.getString(1) + delimiter
        + rs.getString(2) + delimiter
        + rs.getString(3) + delimiter
        + rs.getString(4)
        + "\n");
}
rs.close();

try {
    //建立目标数据库连接
    Class.forName(driver);
    Connection conn = DriverManager.getConnection(url, user, pass);
    BaseConnection baseConn = (BaseConnection) conn;
    baseConn.setAutoCommit(false);

    //初始化表信息
    String sql = "Copy " + tablename + " from STDIN DELIMITER " + "" + delimiter + "" + "
ENCODING " + "" + encoding + "";

    //提交缓存buffer中的数据
    CopyManager cp = new CopyManager(baseConn);
    StringReader reader = new StringReader(buffer.toString());
    cp.copyIn(sql, reader);
    baseConn.commit();
    reader.close();
    baseConn.close();
} catch (ClassNotFoundException e) {
    e.printStackTrace(System.out);
} catch (SQLException e) {
    e.printStackTrace(System.out);
}

} catch (Exception e) {
    e.printStackTrace();
}
}

//*****
// 从源数据库返回查询结果集
//*****
private static ResultSet getDataSet() {
    ResultSet rs = null;
    try {
        Class.forName("com.mysql.jdbc.Driver").newInstance();
        Connection conn = DriverManager.getConnection("jdbc:mysql://10.119.179.227:3306/jack?
useSSL=false&allowPublicKeyRetrieval=true", "jack", "*****");
        Statement stmt = conn.createStatement();
        rs = stmt.executeQuery("select * from migration_table");
    } catch (SQLException e) {
        e.printStackTrace();
    } catch (Exception e) {
        e.printStackTrace();
    }
    return rs;
}
}
```

## 4.3 整库迁移

### 4.3.1 使用 DRS 将数据导入 GaussDB(DWS)

使用数据复制服务（Data Replication Service，简称DRS），可以将其他数据源的数据导入到GaussDB(DWS)集群的数据库中。当前支持导入的数据源主要包括以下：

- MySQL
- DDM
- PostgreSQL（公测）
- Oracle（公测）
- GaussDB分布式版（公测）

请参见[DRS实时同步](#)章节。

#### 📖 说明

其中PostgreSQL、Oracle、GaussDB分布式版的数据源属于公测阶段，请移步到DRS管理控制台，通过新建工单方式申请公测。

### 4.3.2 使用 CDM 迁移数据到 GaussDB(DWS)

使用云数据迁移服务（Cloud Data Migration，简称CDM），可以将其他数据源（例如MySQL）的数据迁移到GaussDB(DWS) 集群的数据库中。

使用CDM迁移数据到GaussDB(DWS) 的典型场景，请参见云数据迁移服务（简称CDM）的如下章节：

- [入门](#)：该入门场景为使用CDM迁移本地MySQL数据库到GaussDB(DWS)

### 4.3.3 使用 DSC 工具迁移 SQL 脚本

DSC（Database Schema Converter）是一款运行在Linux或Windows操作系统上的命令行工具，致力于向客户提供简单、快速、可靠的应用程序SQL脚本迁移服务，通过内置的语法迁移逻辑解析源数据库应用程序SQL脚本，并迁移为适用于GaussDB(DWS) 数据库的应用程序SQL脚本。DSC不需要连接数据库，可在离线模式下实现零停机迁移。在GaussDB(DWS) 中通过执行迁移后的SQL脚本即可恢复数据库，从而实现线下数据库轻松上云。

DSC支持迁移Teradata、Oracle、Netezza、MySQL和DB2数据库的SQL脚本。

### 下载 DSC SQL 语法迁移工具

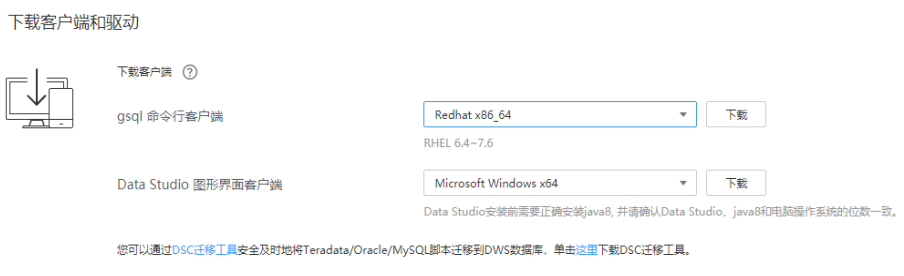
**步骤1** 登录GaussDB(DWS) 管理控制台。

**步骤2** 在左侧导航栏中，单击“连接管理”。

**步骤3** 在“下载客户端和驱动”区域，单击“这里”即可下载DSC迁移工具。

如果同时拥有不同版本的集群，系统会弹出对话框，提示您选择“集群版本”然后下载与集群版本相对应的客户端。在“集群管理”页面的集群列表中，单击指定集群的名称，再选择“基本信息”页签，可查看集群版本。

图 4-8 下载工具





**步骤4** 下载到本机后，使用WinSCP工具，将DSC工具上传到一个需安装工具的Linux主机上。

执行上传操作的用户需要对Linux主机的目标存放目录有完全控制权限。

----结束

## DSC SQL 语法迁移工具操作指导

详细指导请参见[DSC SQL语法迁移工具](#)。

## 4.4 元数据迁移

### 4.4.1 使用 gs\_dump 和 gs\_dumpall 命令导出元数据

#### 4.4.1.1 概述

GaussDB(DWS)提供的gs\_dump和gs\_dumpall工具，能够帮助用户导出需要的数据库对象或其相关信息。通过导入工具将导出的元数据信息导入至需要的数据库，可以完成数据库信息的迁移。gs\_dump支持导出单个数据库或其内的对象，而gs\_dumpall支持导出集群中所有数据库或各库的公共全局对象。详细的使用场景见[表4-14](#)。

表 4-14 适用场景

适用场景	支持的导出粒度	支持的导出格式	配套的导入方法
导出单个数据库	<b>数据库级导出。</b> <ul style="list-style-type: none"> <li>导出全量信息。 使用导出的全量信息可以创建一个与当前库相同的数据库，且库中数据也与当前库相同。</li> <li>仅导出库中所有对象的定义，包含库定义、函数定义、模式定义、表定义、索引定义和存储过程定义等。 使用导出的对象定义，可以快速创建一个相同的数据库，但是库中并无原数据库的数据。</li> <li>仅导出数据。</li> </ul>	<ul style="list-style-type: none"> <li>纯文本格式</li> <li>自定义归档格式</li> <li>目录归档格式</li> <li>tar归档格式</li> </ul>	<ul style="list-style-type: none"> <li>纯文本格式数据文件导入请参见<a href="#">使用gsq元命令导入数据</a>。</li> <li>自定义归档格式、目录归档格式和tar归档格式数据文件导入请参见<a href="#">使用gs_restore命令导入数据</a>。</li> </ul>
	<b>模式级导出。</b> <ul style="list-style-type: none"> <li>导出模式的全量信息。</li> <li>仅导出模式中数据。</li> <li>仅导出对象的定义，包含表定义、存储过程定义和索引定义等。</li> </ul>		

适用场景	支持的导出粒度	支持的导出格式	配套的导入方法
	<p><b>表级导出。</b></p> <ul style="list-style-type: none"> <li>导出表的全量信息。</li> <li>仅导出表中数据。</li> <li>仅导出表的定义。</li> </ul>		
导出所有数据库	<p><b>数据库级导出。</b></p> <ul style="list-style-type: none"> <li>导出全量信息。 使用导出的全量信息可以创建与当前集群相同的一个集群，拥有相同数据库和公共全局对象，且库中数据也与当前各库相同。</li> <li>仅导出各数据库中的对象定义，包含表空间、库定义、函数定义、模式定义、表定义、索引定义和存储过程定义等。 使用导出的对象定义，可以快速创建与当前集群相同的一个集群，拥有相同的数据库和表空间，但是库中并无原数据库的数据。</li> <li>仅导出数据。</li> </ul>	纯文本格式	数据文件导入请参见 <a href="#">使用gsq命令导入数据</a> 。
	<p><b>各库公共全局对象导出</b></p> <ul style="list-style-type: none"> <li>仅导出表空间信息。</li> <li>仅导出角色信息。</li> <li>导出角色与表空间。</li> </ul>		

gs\_dump和gs\_dumpall通过-U指定执行导出的用户帐户。如果当前使用的帐户不具备导出所要求的权限时，会无法导出数据。此时，可在导出命令中设置--role参数来指定具备权限的角色。在执行命令后，gs\_dump和gs\_dumpall会使用--role参数指定的角色，完成导出动作。可使用该功能的场景请参见[表4-14](#)，详细操作请参见[无权限角色导出数据](#)。

gs\_dump和gs\_dumpall通过对导出的数据文件加密，导入时对加密的数据文件进行解密，可以防止数据信息泄露，为数据库的安全提供保证。

gs\_dump和gs\_dumpall工具在进行数据导出时，其他用户可以访问集群数据库（读或写）。

gs\_dump和gs\_dumpall工具支持导出完整一致的数据。例如，T1时刻启动gs\_dump导出A数据库，或者启动gs\_dumpall导出整个集群数据库，那么导出数据结果将会是T1时刻A数据库或者该集群数据库的数据状态，T1时刻之后对A数据库或集群数据库的修改不会被导出。

gs\_dump和gs\_dumpall工具是通过“gsq命令行客户端”软件包解压缩获取。

## 注意事项

- 禁止修改导出的文件和内容，否则可能无法恢复成功。
- 为了保证数据一致性和完整性，导出工具会对需要转储的表设置共享锁。如果表在别的事务中设置了共享锁，gs\_dump和gs\_dumpall会等待锁释放后锁定表。如果无法在指定时间内锁定某个表，转储会失败。用户可以通过指定--lock-wait-timeout选项，自定义等待锁超时时间。
- 由于gs\_dumpall读取所有数据库中的表，因此必须以数据库集群管理员身份进行连接，才能导出完整文件。在使用gsq执行脚本文件导入时，同样需要管理员权限，以便添加用户和组，以及创建数据库。
- 由于GaussDB(DWS)数据库所有视图的定义都默认带有表名或别名的前缀（即tab.col的形式），因此可能与原始定义不符，导致在极少的场景下会发生重建视图字段对应基表不准确而报错的情况。为避免此情况建议在导出视图定义时，设置guc参数behavior\_compat\_options='compat\_display\_ref\_table'，使导出定义与原始语句一致。

### 4.4.1.2 导出单个数据库

#### 4.4.1.2.1 导出数据库

GaussDB(DWS)支持使用gs\_dump工具导出某个数据库级的内容，包含数据库的数据和所有对象定义。可根据需要自定义导出如下信息：

- 导出数据库全量信息，包含数据和所有对象定义。  
使用导出的全量信息可以创建一个与当前库相同的数据库，且库中数据也与当前库相同。
- 仅导出所有对象定义，包括：库定义、函数定义、模式定义、表定义、索引定义和存储过程定义等。  
使用导出的对象定义，可以快速创建一个相同的数据库，但是库中并无原数据库的数据。
- 仅导出数据，不包含所有对象定义。

## 操作步骤

**步骤1 准备ECS作为gsq客户端主机。**

**步骤2** 请参见[下载客户端](#)下载gsq客户端，并使用SSH文件传输工具（例如WinSCP工具），将客户端工具上传到一个待安装gsq的Linux主机上。

执行上传gsq操作的用户需要对客户端主机的目标存放目录有完全控制权限。

或者，您也可以先SSH远程登录到需要安装gsq的Linux主机，然后在Linux命令窗口，执行以下命令下载gsq客户端：

```
wget https://obs.myhuaweicloud.com/dws/download/dws_client_8.x.x_redhat_x64.zip --no-check-certificate
```

**步骤3** 执行以下命令解压客户端工具。

```
cd <客户端存放路径>  
unzip dws_client_8.x.x_redhat_x64.zip
```

其中：

- <客户端存放路径>：请替换为实际的客户端存放路径。

- `dws_client_8.1.x_redhat_x86.zip`: 这是“RedHat x86”对应的客户端工具包名称，请替换为实际下载的包名。

**步骤4** 执行以下命令配置客户端。

```
source gsql_env.sh
```

提示以下信息表示客户端已配置成功

```
All things done.
```

**步骤5** 使用`gs_dump`导出gaussdb数据库。

```
gs_dump -W password -U jack -f /home//backup/postgres_backup.tar -p 8000 gaussdb -h 10.10.10.100 -F t
```

**表 4-15** 常用参数说明

参数	参数说明	举例
-U	连接数据库的用户名，如果未填写则表示当前已连接的数据库用户。	-U jack
-W	指定用户连接的密码。 <ul style="list-style-type: none"> <li>• 如果主机的认证策略是trust，则不会对数据库管理员进行密码验证，即无需输入-W选项；</li> <li>• 如果没有-W选项，并且不是数据库管理员，会提示用户输入密码。</li> </ul>	-W password, 此处密码需要用户自定义。
-f	将导出文件发送至指定目录文件夹。如果这里省略，则使用标准输出。	-f /home//backup/postgres_backup.tar
-p	指定服务器所监听的TCP端口或本地Unix域套接字后缀，以确保连接。	-p 8000
-h	“集群地址”如果通过公网地址连接，请指定为集群“公网访问地址”或“公网访问域名”，如果通过内网地址连接，请指定为集群“内网访问地址”或“内网访问域名”。	-h 10.10.10.100
dbname	需要导出的数据库名称	gaussdb
-F	选择导出文件格式。-F参数值如下： <ul style="list-style-type: none"> <li>• p: 纯文本格式</li> <li>• c: 自定义归档</li> <li>• d: 目录归档格式</li> <li>• t: tar归档格式</li> </ul>	-F t

其他参数说明请参见《工具指南》中“gs\_dump”章节。

----结束

## 示例

示例一：执行gs\_dump，导出gaussdb数据库全量信息，并对导出文件进行压缩，导出文件格式为sql文本格式。

```
gs_dump -W password -U jack -f /home//backup/postgres_backup.sql -p 8000 -h 10.10.10.100 gaussdb -Z 8 -F p
gs_dump[port=""][gaussdb][2017-07-21 15:36:13]: dump database gaussdb successfully
gs_dump[port=""][gaussdb][2017-07-21 15:36:13]: total time: 3793 ms
```

示例二：执行gs\_dump，仅导出gaussdb数据库中的数据，不包含数据库对象定义，导出文件格式为自定义归档格式。

```
gs_dump -W Password -U jack -f /home//backup/postgres_data_backup.dmp -p 8000 -h 10.10.10.100 gaussdb -a -F c
gs_dump[port=""][gaussdb][2017-07-21 15:36:13]: dump database gaussdb successfully
gs_dump[port=""][gaussdb][2017-07-21 15:36:13]: total time: 3793 ms
```

示例三：执行gs\_dump，仅导出gaussdb数据库所有对象的定义，导出文件格式为sql文本格式。

--导出前，表nation有数据

```
select n_nationkey,n_name,n_regionkey from nation limit 3;
```

n_nationkey	n_name	n_regionkey
0	ALGERIA	0
3	CANADA	1
11	IRAQ	4

(3 rows)

```
gs_dump -W password -U jack -f /home//backup/postgres_def_backup.sql -p 8000 -h 10.10.10.100 gaussdb -s -F p
gs_dump[port=""][gaussdb][2017-07-20 15:04:14]: dump database gaussdb successfully
gs_dump[port=""][gaussdb][2017-07-20 15:04:14]: total time: 472 ms
```

示例四：执行gs\_dump，仅导出gaussdb数据库的所有对象的定义，导出文件格式为文本格式，并对导出文件进行加密。

```
gs_dump -W password -U jack -f /home//backup/postgres_def_backup.sql -p 8000 -h 10.10.10.100 gaussdb --with-encryption AES128 --with-key 1234567812345678 -s -F p
gs_dump[port=""][gaussdb][2018-11-14 11:25:18]: dump database gaussdb successfully
gs_dump[port=""][gaussdb][2018-11-14 11:25:18]: total time: 1161 ms
```

### 4.4.1.2.2 导出模式

GaussDB(DWS)目前支持使用gs\_dump工具导出模式级的内容，包含模式的数据和定义。用户可通过灵活的自定义方式导出模式内容，不仅支持选定一个模式或多个模式的导出，还支持排除一个模式或者多个模式的导出。可根据需要自定义导出如下信息：

- 导出模式全量信息，包含数据和对象定义。
- 仅导出数据，即模式包含表中的数据，不包含对象定义。
- 仅导出模式对象定义，包括：表定义、存储过程定义和索引定义等。

## 操作步骤

**步骤1** 准备ECS作为gsq客户端主机。

**步骤2** 请参见[下载客户端](#)下载gsq客户端，并使用SSH文件传输工具（例如WinSCP工具），将客户端工具上传到一个待安装gsq的Linux主机上。

执行上传gsq操作的用户需要对客户端主机的目标存放目录有完全控制权限。

或者，您也可以先SSH远程登录到需要安装gsq的Linux主机，然后在Linux命令窗口，执行以下命令下载gsq客户端：

```
wget https://obs.myhuaweicloud.com/dws/download/dws_client_8.x.x_redhat_x64.zip --no-check-certificate
```

**步骤3** 执行以下命令解压客户端工具。

```
cd <客户端存放路径>
unzip dws_client_8.x.x_redhat_x64.zip
```

其中：

- <客户端存放路径>：请替换为实际的客户端存放路径。
- dws\_client\_8.1.x\_redhat\_x86.zip：这是“RedHat x86”对应的客户端工具包名称，请替换为实际下载的包名。

**步骤4** 执行以下命令配置客户端。

```
source gsql_env.sh
```

提示以下信息表示客户端已配置成功

```
All things done.
```

**步骤5** 使用gs\_dump同时导出hr和public模式。

```
gs_dump -W Password -U jack -f /home//backup/MPPDB_schema_backup -p 8000 -h 10.10.10.100
human_resource -n hr -F d
```

**表 4-16** 常用参数说明

参数	参数说明	举例
-U	连接数据库的用户名，如果未填写则表示当前已连接的数据库用户。	-U jack
-W	指定用户连接的密码。 <ul style="list-style-type: none"> <li>• 如果主机的认证策略是trust，则不会对数据库管理员进行密码验证，即无需输入-W选项。</li> <li>• 如果没有-W选项，并且不是数据库管理员，会提示用户输入密码。</li> </ul>	-W password，此处密码需要用户自定义。
-f	将导出文件发送至指定目录文件夹。如果这里省略，则使用标准输出。	-f /home//backup/MPPDB_schema_backup
-p	指定服务器所监听的TCP端口或本地Unix域套接字后缀，以确保连接。	-p 8000

参数	参数说明	举例
-h	“集群地址”如果通过公网地址连接，请指定为集群“公网访问地址”或“公网访问域名”，如果通过内网地址连接，请指定为集群“内网访问地址”或“内网访问域名”。	-h 10.10.10.100
dbname	需要导出的数据库名称	human_resource
-n	只导出与模式名称匹配的模式，此选项包括模式本身和所有它包含的对象。 <ul style="list-style-type: none"> <li>单个模式: -n <i>schemaname</i></li> <li>多个模式: 多次输入 -n <i>schemaname</i></li> </ul>	<ul style="list-style-type: none"> <li>单个模式: -n hr</li> <li>多个模式: -n hr -n public</li> </ul>
-F	选择导出文件格式。-F参数值如下: <ul style="list-style-type: none"> <li>p: 纯文本格式</li> <li>c: 自定义归档</li> <li>d: 目录归档格式</li> <li>t: tar归档格式</li> </ul>	-F d

其他参数说明请参见《工具指南》中“gs\_dump”章节。

---结束

## 示例

示例一：执行gs\_dump，导出hr模式全量信息，并对导出文件进行压缩，导出文件格式为文本格式。

```
gs_dump -W password -U jack -f /home//backup/MPPDB_schema_backup.sql -p 8000 -h 10.10.10.100
human_resource -n hr -Z G -F p
gs_dump[port=""][human_resource][2017-07-21 16:05:55]: dump database human_resource successfully
gs_dump[port=""][human_resource][2017-07-21 16:05:55]: total time: 2425 ms
```

示例二：执行gs\_dump，仅导出hr模式的数据，导出文件格式为tar归档格式。

```
gs_dump -W password -U jack -f /home//backup/MPPDB_schema_data_backup.tar -p 8000 -h 10.10.10.100
human_resource -n hr -a -F t
gs_dump[port=""][human_resource][2018-11-14 15:07:16]: dump database human_resource successfully
gs_dump[port=""][human_resource][2018-11-14 15:07:16]: total time: 1865 ms
```

示例三：执行gs\_dump，仅导出hr模式的定义，导出文件格式为目录归档格式。

```
gs_dump -W password -U jack -f /home//backup/MPPDB_schema_def_backup -p 8000 -h 10.10.10.100
human_resource -n hr -s -F d
gs_dump[port=""][human_resource][2018-11-14 15:11:34]: dump database human_resource successfully
gs_dump[port=""][human_resource][2018-11-14 15:11:34]: total time: 1652 ms
```

示例四：执行gs\_dump，导出human\_resource数据库时，排除hr模式，导出文件格式为自定义归档格式。

```
gs_dump -W password -U jack -f /home//backup/MPPDB_schema_backup.dmp -p 8000 -h 10.10.10.100
human_resource -N hr -F c
```

```
gs_dump[port=""][human_resource][2017-07-21 16:06:31]: dump database human_resource successfully
gs_dump[port=""][human_resource][2017-07-21 16:06:31]: total time: 2522 ms
```

示例五：执行gs\_dump，同时导出hr和public模式，且仅导出模式定义，并对导出文件进行加密，导出文件格式为tar归档格式。

```
gs_dump -W password -U jack -f /home//backup/MPPDB_schema_backup1.tar -p 8000 -h 10.10.10.100
human_resource -n hr -n public -s --with-encryption AES128 --with-key 1234567812345678 -F t
gs_dump[port=""][human_resource][2017-07-21 16:07:16]: dump database human_resource successfully
gs_dump[port=""][human_resource][2017-07-21 16:07:16]: total time: 2132 ms
```

示例六：执行gs\_dump，导出human\_resource数据库时，排除hr和public模式，导出文件格式为自定义归档格式。

```
gs_dump -W password -U jack -f /home//backup/MPPDB_schema_backup2.dmp -p 8000 -h 10.10.10.100
human_resource -N hr -N public -F c
gs_dump[port=""][human_resource][2017-07-21 16:07:55]: dump database human_resource successfully
gs_dump[port=""][human_resource][2017-07-21 16:07:55]: total time: 2296 ms
```

示例七：执行gs\_dump，导出public模式下所有表（视图、序列和外表）和hr模式中staffs表，包含数据和表定义，导出文件格式为自定义归档格式。

```
gs_dump -W password -U jack -f /home//backup/MPPDB_backup3.dmp -p 8000 -h 10.10.10.100
human_resource -t public.* -t hr.staffs -F c
gs_dump[port=""][human_resource][2018-12-13 09:40:24]: dump database human_resource successfully
gs_dump[port=""][human_resource][2018-12-13 09:40:24]: total time: 896 ms
```

#### 4.4.1.2.3 导出表

GaussDB(DWS)支持使用gs\_dump工具导出表级的内容，包含表定义和表数据。视图、序列和外表属于特殊的表。用户可通过灵活的自定义方式导出表内容，不仅支持选定一个表或多个表的导出，还支持排除一个表或者多个表的导出。可根据需要自定义导出如下信息：

- 导出表全量信息，包含表数据和表定义。
- 仅导出数据，不包含表定义。
- 仅导出表定义。

### 操作步骤

**步骤1 准备ECS作为gsq客户端主机。**

**步骤2** 请参见[下载客户端](#)下载gsq客户端，并使用SSH文件传输工具（例如WinSCP工具），将客户端工具上传到一个待安装gsq的Linux主机上。

执行上传gsq操作的用户需要对客户端主机的目标存放目录有完全控制权限。

或者，您也可以先SSH远程登录到需要安装gsq的Linux主机，然后在Linux命令窗口，执行以下命令下载gsq客户端：

```
wget https://obs.myhuaweicloud.com/dws/download/dws_client_8.x.x_redhat_x64.zip --no-check-certificate
```

**步骤3** 执行以下命令解压客户端工具。

```
cd <客户端存放路径>
unzip dws_client_8.x.x_redhat_x64.zip
```

其中：

- <客户端存放路径>：请替换为实际的客户端存放路径。
- dws\_client\_8.1.x\_redhat\_x86.zip：这是“RedHat x86”对应的客户端工具包名称，请替换为实际下载的包名。



**步骤4** 执行以下命令配置客户端。

```
source gsql_env.sh
```

提示以下信息表示客户端已配置成功

```
All things done.
```

**步骤5** 使用gs\_dump同时导出指定表hr.staffs和hr employments。

```
gs_dump -W password -U jack -f /home//backup/MPPDB_table_backup -p 8000 -h 10.10.10.100  
human_resource -t hr.staffs -F d
```

**表 4-17** 常用参数说明

参数	参数说明	举例
-U	连接数据库的用户名，如果未填写则表示当前已连接的数据库用户。	-U jack
-W	指定用户连接的密码。 <ul style="list-style-type: none"> <li>如果主机的认证策略是trust，则不会对数据库管理员进行密码验证，即无需输入-W选项。</li> <li>如果没有-W选项，并且不是数据库管理员，会提示用户输入密码。</li> </ul>	-W password
-f	将导出文件发送至指定目录文件夹。如果这里省略，则使用标准输出。	-f /home//backup/ MPPDB_table_backup
-p	指定服务器所监听的TCP端口或本地Unix域套接字后缀，以确保连接。	-p 8000
-h	“集群地址”如果通过公网地址连接，请指定为集群“公网访问地址”或“公网访问域名”，如果通过内网地址连接，请指定为集群“内网访问地址”或“内网访问域名”。	-h 10.10.10.100
dbname	需要导出的数据库名称	human_resource
-t	指定导出的表（或视图、序列、外表），可以使用多个-t选项来选择多个表，也可以使用通配符指定多个表对象。当使用通配符指定多个表对象时，注意给pattern打引号，防止shell扩展通配符。 <ul style="list-style-type: none"> <li>单个表：-t schema.table</li> <li>多个表：多次输入-t schema.table</li> </ul>	<ul style="list-style-type: none"> <li>单个表：-t hr.staffs</li> <li>多个表：-t hr.staffs -t hr.employments</li> </ul>

参数	参数说明	举例
-F	选择导出文件格式。-F参数值如下： <ul style="list-style-type: none"> <li>• p: 纯文本格式</li> <li>• c: 自定义归档</li> <li>• d: 目录归档格式</li> <li>• t: tar归档格式</li> </ul>	-F d

其他参数说明请参见《工具指南》中“gs\_dump”章节。

---结束

## 示例

示例一：执行gs\_dump，导出表hr.staffs的定义和数据，并对导出文件进行压缩，导出文件格式为文本格式。

```
gs_dump -W password -U jack -f /home//backup/MPPDB_table_backup.sql -p 8000 -h 10.10.10.100
human_resource -t hr.staffs -Z 6 -F p
gs_dump[port=""][human_resource][2017-07-21 17:05:10]: dump database human_resource successfully
gs_dump[port=""][human_resource][2017-07-21 17:05:10]: total time: 3116 ms
```

示例二：执行gs\_dump，只导出表hr.staffs的数据，导出文件格式为tar归档格式。

```
gs_dump -W password -U jack -f /home//backup/MPPDB_table_data_backup.tar -p 8000 -h 10.10.10.100
human_resource -t hr.staffs -a -F t
gs_dump[port=""][human_resource][2017-07-21 17:04:26]: dump database human_resource successfully
gs_dump[port=""][human_resource][2017-07-21 17:04:26]: total time: 2570 ms
```

示例三：执行gs\_dump，导出表hr.staffs的定义，导出文件格式为目录归档格式。

```
gs_dump -W password -U jack -f /home//backup/MPPDB_table_def_backup -p 8000 -h 10.10.10.100
human_resource -t hr.staffs -s -F d
gs_dump[port=""][human_resource][2017-07-21 17:03:09]: dump database human_resource successfully
gs_dump[port=""][human_resource][2017-07-21 17:03:09]: total time: 2297 ms
```

示例四：执行gs\_dump，不导出表hr.staffs，导出文件格式为自定义归档格式。

```
gs_dump -W password -U jack -f /home//backup/MPPDB_table_backup4.dmp -p 8000 -h 10.10.10.100
human_resource -T hr.staffs -F c
gs_dump[port=""][human_resource][2017-07-21 17:14:11]: dump database human_resource successfully
gs_dump[port=""][human_resource][2017-07-21 17:14:11]: total time: 2450 ms
```

示例五：执行gs\_dump，同时导出两个表hr.staffs和hr employments，导出文件格式为文本格式。

```
gs_dump -W password -U jack -f /home//backup/MPPDB_table_backup1.sql -p 8000 -h 10.10.10.100
human_resource -t hr.staffs -t hr.employments -F p
gs_dump[port=""][human_resource][2017-07-21 17:19:42]: dump database human_resource successfully
gs_dump[port=""][human_resource][2017-07-21 17:19:42]: total time: 2414 ms
```

示例六：执行gs\_dump，导出时，排除两个表hr.staffs和hr employments，导出文件格式为文本格式。

```
gs_dump -W password -U jack -f /home//backup/MPPDB_table_backup2.sql -p 8000 -h 10.10.10.100
human_resource -T hr.staffs -T hr.employments -F p
gs_dump[port=""][human_resource][2017-07-21 17:21:02]: dump database human_resource successfully
gs_dump[port=""][human_resource][2017-07-21 17:21:02]: total time: 3165 ms
```

示例七：执行gs\_dump，导出表hr.staffs的定义和数据，只导出表hr employments的定义，导出文件格式为tar归档格式。

```
gs_dump -W password -U jack -f /home//backup/MPPDB_table_backup3.tar -p 8000 -h 10.10.10.100
human_resource -t hr.staffs -t hr.employments --exclude-table-data hr.employments -F t
```

```
gs_dump[port=""][human_resource][2018-11-14 11:32:02]: dump database human_resource successfully
gs_dump[port=""][human_resource][2018-11-14 11:32:02]: total time: 1645 ms
```

示例八：执行gs\_dump，导出表hr.staffs的定义和数据，并对导出文件进行加密，导出文件格式为文本格式。

```
gs_dump -W password -U jack -f /home//backup/MPPDB_table_backup4.sql -p 8000 -h 10.10.10.100
human_resource -t hr.staffs --with-encryption AES128 --with-key 1212121212121212 -F p
gs_dump[port=""][human_resource][2018-11-14 11:35:30]: dump database human_resource successfully
gs_dump[port=""][human_resource][2018-11-14 11:35:30]: total time: 6708 ms
```

示例九：执行gs\_dump，导出public模式下所有表（包括视图、序列和外表）和hr模式中staffs表，包含数据和表定义，导出文件格式为自定义归档格式。

```
gs_dump -W password -U jack -f /home//backup/MPPDB_table_backup5.dmp -p 8000 -h 10.10.10.100
human_resource -t public.* -t hr.staffs -F c
gs_dump[port=""][human_resource][2018-12-13 09:40:24]: dump database human_resource successfully
gs_dump[port=""][human_resource][2018-12-13 09:40:24]: total time: 896 ms
```

示例十：执行gs\_dump，仅导出依赖于t1模式下的test1表对象的视图信息，导出文件格式为目录归档格式。

```
gs_dump -W password -U jack -f /home//backup/MPPDB_view_backup6 -p 8000 -h 10.10.10.100
human_resource -t t1.test1 --include-depend-objs --exclude-self -F d
gs_dump[port=""][jack][2018-11-14 17:21:18]: dump database human_resource successfully
gs_dump[port=""][jack][2018-11-14 17:21:23]: total time: 4239 ms
```

### 4.4.1.3 导出所有数据库

#### 4.4.1.3.1 导出所有数据库

GaussDB(DWS)支持使用gs\_dumpall工具导出所有数据库的全量信息，包含集群中每个数据库信息和公共的全局对象信息。可根据需要自定义导出如下信息：

- 导出所有数据库全量信息，包含集群中每个数据库信息和公共的全局对象信息（包含角色和表空间信息）。  
使用导出的全量信息可以创建与当前集群相同的一个集群，拥有相同数据库和公共全局对象，且库中数据也与当前各库相同。
- 仅导出数据，即导出每个数据库中的数据，且不包含所有对象定义和公共的全局对象信息。
- 仅导出所有对象定义，包括：表空间、库定义、函数定义、模式定义、表定义、索引定义和存储过程定义等。

使用导出的对象定义，可以快速创建与当前集群相同的一个集群，拥有相同的数据库和表空间，但是库中并无原数据库的数据。

## 操作步骤

**步骤1 准备ECS作为gsq客户端主机。**

**步骤2** 请参见[下载客户端](#)下载gsq客户端，并使用SSH文件传输工具（例如WinSCP工具），将客户端工具上传到一个待安装gsq的Linux主机上。

执行上传gsq操作的用户需要对客户端主机的目标存放目录有完全控制权限。

或者，您也可以先SSH远程登录到需要安装gsq的Linux主机，然后在Linux命令窗口，执行以下命令下载gsq客户端：

```
wget https://obs.myhuaweicloud.com/dws/download/dws_client_8.x.x_redhat_x64.zip --no-check-certificate
```

**步骤3** 执行以下命令解压客户端工具。

```
cd <客户端存放路径>
unzip dws_client_8.x.x_redhat_x64.zip
```

其中：

- <客户端存放路径>：请替换为实际的客户端存放路径。
- dws\_client\_8.1.x\_redhat\_x86.zip：这是“RedHat x86”对应的客户端工具包名称，请替换为实际下载的包名。

**步骤4** 执行以下命令配置客户端。

```
source gsql_env.sh
```

提示以下信息表示客户端已配置成功

```
All things done.
```

**步骤5** 使用gs\_dumpall一次导出所有数据库信息。

```
gs_dumpall -W password -U dbadmin -f /home/dbadmin/backup/MPPDB_backup.sql -p 8000 -h 10.10.10.100
```

**表 4-18** 常用参数说明

参数	参数说明	举例
-U	连接数据库的用户名，需要是集群管理员用户。	-U dbadmin
-W	指定用户连接的密码。 <ul style="list-style-type: none"> <li>• 如果主机的认证策略是trust，则不会对数据库管理员进行密码验证，即无需输入-W选项；</li> <li>• 如果没有-W选项，并且不是数据库管理员，会提示用户输入密码。</li> </ul>	-W password，此处密码需要用户自定义。
-f	将导出文件发送至指定目录文件夹。如果这里省略，则使用标准输出。	-f /home/dbadmin/backup/MPPDB_backup.sql
-p	指定服务器所监听的TCP端口或本地Unix域套接字后缀，以确保连接。	-p 8000
-h	“集群地址”如果通过公网地址连接，请指定为集群“公网访问地址”或“公网访问域名”，如果通过内网地址连接，请指定为集群“内网访问地址”或“内网访问域名”。	-h 10.10.10.100

其他参数说明请参见《工具指南》中“gs\_dumpall”章节。

----结束

## 示例

示例一：执行gs\_dumpall，导出所有数据库全量信息（dbadmin用户为管理员用户），导出文件为文本格式。执行命令后，会有很长的打印信息，最终出现total time即代表执行成功。示例中将不体现中间的打印信息。

```
gs_dumpall -W password -U dbadmin -f /home/dbadmin/backup/MPPDB_backup.sql -p 8000 -h  
10.10.10.100  
gs_dumpall[port=""][2017-07-21 15:57:31]: dumpall operation successful  
gs_dumpall[port=""][2017-07-21 15:57:31]: total time: 9627 ms
```

示例二：执行gs\_dumpall，仅导出所有数据库定义（dbadmin用户为管理员用户），导出文件为文本格式。执行命令后，会有很长的打印信息，最终出现total time即代表执行成功。示例中将不体现中间的打印信息。

```
gs_dumpall -W password -U dbadmin -f /home/dbadmin/backup/MPPDB_backup.sql -p 8000 -h  
10.10.10.100 -s  
gs_dumpall[port=""][2018-11-14 11:28:14]: dumpall operation successful  
gs_dumpall[port=""][2018-11-14 11:28:14]: total time: 4147 ms
```

示例三：执行gs\_dumpall，仅导出所有数据库中数据，并对导出文件进行加密，导出文件为文本格式。执行命令后，会有很长的打印信息，最终出现total time即代表执行成功。示例中将不体现中间的打印信息。

```
gs_dumpall -W password -U dbadmin -f /home/dbadmin/backup/MPPDB_backup.sql -p 8000 -h  
10.10.10.100 -a --with-encryption AES128 --with-key 1234567812345678  
gs_dumpall[port=""][2018-11-14 11:32:26]: dumpall operation successful  
gs_dumpall[port=""][2018-11-14 11:23:26]: total time: 4147 ms
```

### 4.4.1.3.2 导出全局对象

GaussDB(DWS)支持使用gs\_dumpall工具导出所有数据库公共的全局对象，包含数据库用户和组，表空间及属性（例如：适用于数据库整体的访问权限）信息。

## 操作步骤

**步骤1 准备ECS作为gsq客户端主机。**

**步骤2** 请参见[下载客户端](#)下载gsq客户端，并使用SSH文件传输工具（例如WinSCP工具），将客户端工具上传到一个待安装gsq的Linux主机上。

执行上传gsq操作的用户需要对客户端主机的目标存放目录有完全控制权限。

或者，您也可以先SSH远程登录到需要安装gsq的Linux主机，然后在Linux命令窗口，执行以下命令下载gsq客户端：

```
wget https://obs.myhuaweicloud.com/dws/download/dws_client_8.x.x_redhat_x64.zip --no-check-certificate
```

**步骤3** 执行以下命令解压客户端工具。

```
cd <客户端存放路径>  
unzip dws_client_8.x.x_redhat_x64.zip
```

其中：

- <客户端存放路径>：请替换为实际的客户端存放路径。
- dws\_client\_8.1.x\_redhat\_x86.zip：这是“RedHat x86”对应的客户端工具包名称，请替换为实际下载的包名。

**步骤4** 执行以下命令配置客户端。

```
source gsql_env.sh
```

提示以下信息表示客户端已配置成功

All things done.

### 步骤5 使用gs\_dumpall导出表空间对象信息。

```
gs_dumpall -W password -U dbadmin -f /home/dbadmin/backup/MPPDB_tablespace.sql -p 8000 -h 10.10.10.100 -t
```

表 4-19 常用参数说明

参数	参数说明	举例
-U	连接数据库的用户名，需要是集群管理员用户。	-U dbadmin
-W	指定用户连接的密码。 <ul style="list-style-type: none"> <li>如果主机的认证策略是trust，则不会对数据库管理员进行密码验证，即无需输入-W选项；</li> <li>如果没有-W选项，并且不是数据库管理员，会提示用户输入密码。</li> </ul>	--W password, 此处密码需要用户自定义。
-f	将导出文件发送至指定目录文件夹。如果这里省略，则使用标准输出。	-f /home//backup/MPPDB_tablespace.sql
-p	指定服务器所监听的TCP端口或本地Unix域套接字后缀，以确保连接。	-p 8000
-h	“集群地址” 如果通过公网地址连接，请指定为集群“公网访问地址”或“公网访问域名”，如果通过内网地址连接，请指定为集群“内网访问地址”或“内网访问域名”。	-h 10.10.10.100
-t	或者--tablespaces-only，只转储表空间，不转储数据库或角色。	-

其他参数说明请参见《工具指南》中“gs\_dumpall”章节。

---结束

## 示例

示例一：执行gs\_dumpall，导出所有数据库的公共全局表空间信息和用户信息（dbadmin用户为管理员用户），导出文件为文本格式。

```
gs_dumpall -W password -U dbadmin -f /home/dbadmin/backup/MPPDB_globals.sql -p 8000 -h 10.10.10.100 -g
gs_dumpall[port=""][2018-11-14 19:06:24]: dumpall operation successful
gs_dumpall[port=""][2018-11-14 19:06:24]: total time: 1150 ms
```

示例二：执行gs\_dumpall，导出所有数据库的公共全局表空间信息（dbadmin用户为管理员用户），并对导出文件进行加密，导出文件为文本格式。

```
gs_dumpall -W password -U dbadmin -f /home/dbadmin/backup/MPPDB_tablespace.sql -p 8000 -h 10.10.10.100 -t --with-encryption AES128 --with-key 1212121212121212
gs_dumpall[port=""][2018-11-14 19:00:58]: dumpall operation successful
gs_dumpall[port=""][2018-11-14 19:00:58]: total time: 186 ms
```

示例三：执行gs\_dumpall，导出所有数据库的公共全局用户信息（dbadmin用户为管理员用户），导出文件为文本格式。

```
gs_dumpall -W password -U dbadmin -f /home/dbadmin/backup/MPPDB_user.sql -p 8000 -h 10.10.10.100 -r
gs_dumpall[port=""][2018-11-14 19:03:18]: dumpall operation successful
gs_dumpall[port=""][2018-11-14 19:03:18]: total time: 162 ms
```

#### 4.4.1.4 无权限角色导出数据

gs\_dump和gs\_dumpall通过-U指定执行导出的用户帐户。如果当前使用的帐户不具备导出所要求的权限时，会无法导出数据。此时，可在导出命令中设置--role参数来指定具备权限的角色。在执行命令后，gs\_dump和gs\_dumpall会使用--role参数指定的角色，完成导出动作。

### 操作步骤

**步骤1** 准备ECS作为gsqL客户端主机。

**步骤2** 请参见[下载客户端](#)下载gsqL客户端，并使用SSH文件传输工具（例如WinSCP工具），将客户端工具上传到一个待安装gsqL的Linux主机上。

执行上传gsqL操作的用户需要对客户端主机的目标存放目录有完全控制权限。

或者，您也可以先SSH远程登录到需要安装gsqL的Linux主机，然后在Linux命令窗口，执行以下命令下载gsqL客户端：

```
wget https://obs.myhuaweicloud.com/dws/download/dws_client_8.x.x_redhat_x64.zip --no-check-certificate
```

**步骤3** 执行以下命令解压客户端工具。

```
cd <客户端存放路径>
unzip dws_client_8.x.x_redhat_x64.zip
```

其中：

- <客户端存放路径>：请替换为实际的客户端存放路径。
- dws\_client\_8.1.x\_redhat\_x86.zip：这是“RedHat x86”对应的客户端工具包名称，请替换为实际下载的包名。

**步骤4** 执行以下命令配置客户端。

```
source gsqL_env.sh
```

提示以下信息表示客户端已配置成功

```
All things done.
```

**步骤5** 使用gs\_dump导出human\_resource数据库数据。

用户jack不具备导出数据库human\_resource的权限，而角色role1具备该权限，要实现导出数据库human\_resource，可以在导出命令中设置--role角色为role1，使用role1的权限，完成导出目的。导出文件格式为tar归档格式。

```
gs_dump -U jack -W password -f /home//backup/MPPDB_backup.tar -p 8000 -h 10.10.10.100 human_resource --role role1 --rolepassword password -F t
```

表 4-20 常用参数说明

参数	参数说明	举例dbadmin
-U	连接数据库的用户名。	-U jack
-W	指定用户连接的密码。 <ul style="list-style-type: none"> <li>如果主机的认证策略是trust，则不会对数据库管理员进行密码验证，即无需输入-W选项。</li> <li>如果没有-W选项，并且不是数据库管理员，会提示用户输入密码。</li> </ul>	-W <i>password</i> ，此处密码需要用户自定义。
-f	将导出文件发送至指定目录文件夹。如果这里省略，则使用标准输出。	-f /home//backup/MPPDB_backup.tar
-p	指定服务器所监听的TCP端口或本地Unix域套接字后缀，以确保连接。	-p 8000
-h	“集群地址”如果通过公网地址连接，请指定为集群“公网访问地址”或“公网访问域名”，如果通过内网地址连接，请指定为集群“内网访问地址”或“内网访问域名”。	-h 10.10.10.100
dbname	需要导出的数据库名称	human_resource
--role	指定导出使用的角色名。选择该选项，会使导出工具连接数据库后，发起一个SET ROLE角色名命令。当所授权用户（由-U指定）没有导出工具要求的权限时，该选项会起到作用，即切换到具备相应权限的角色。	-r role1
--rolepassword	指定具体角色用户的角色密码。	--rolepassword password
-F	选择导出文件格式。-F参数值如下： <ul style="list-style-type: none"> <li>p: 纯文本格式</li> <li>c: 自定义归档</li> <li>d: 目录归档格式</li> <li>t: tar归档格式</li> </ul>	-F t

其他参数说明请参见《工具指南》中“gs\_dump”或“gs\_dumpall”章节。

----结束



## 示例

示例一：执行gs\_dump导出数据，用户jack不具备导出数据库human\_resource的权限，而角色role1具备该权限，要实现导出数据库human\_resource，可以在导出命令中设置--role角色为role1，使用role1的权限，完成导出目的。导出文件格式为tar归档格式。

```
human_resource=# CREATE USER jack IDENTIFIED BY "password";

gs_dump -U jack -W password -f /home//backup/MPPDB_backup11.tar -p 8000 -h 10.10.10.100
human_resource --role role1 --rolepassword password -F t
gs_dump[port='8000'][human_resource][2017-07-21 16:21:10]: dump database human_resource successfully
gs_dump[port='8000'][human_resource][2017-07-21 16:21:10]: total time: 4239 ms
```

示例二：执行gs\_dump导出数据，用户jack不具备导出模式public的权限，而角色role1具备该权限，要实现导出模式public，可以在导出命令中设置--role角色为role1，使用role1的权限，完成导出目的。导出文件格式为tar归档格式。

```
human_resource=# CREATE USER jack IDENTIFIED BY "1234@abc";

gs_dump -U jack -W password -f /home//backup/MPPDB_backup12.tar -p 8000 -h 10.10.10.100
human_resource -n public --role role1 --rolepassword password -F t
gs_dump[port='8000'][human_resource][2017-07-21 16:21:10]: dump database human_resource successfully
gs_dump[port='8000'][human_resource][2017-07-21 16:21:10]: total time: 3278 ms
```

示例三：执行gs\_dumpall导出数据，用户jack不具备导出所有数据库的权限，而角色role1具备该权限，要实现导出所有数据库，可以在导出命令中设置--role角色为role1，使用role1的权限，完成导出目的。导出文件格式为文本归档格式。

```
human_resource=# CREATE USER jack IDENTIFIED BY "password";

gs_dumpall -U jack -W password -f /home//backup/MPPDB_backup.sql -p 8000 -h 10.10.10.100 --role role1
--rolepassword password
gs_dumpall[port='8000'][human_resource][2018-11-14 17:26:18]: dumpall operation successful
gs_dumpall[port='8000'][human_resource][2018-11-14 17:26:18]: total time: 6437 ms
```

## 4.4.2 使用 gs\_restore 命令导入数据

### 操作场景

gs\_restore是GaussDB(DWS)提供的与gs\_dump配套的导入工具。通过该工具，可将gs\_dump导出的文件导入至数据库。gs\_restore支持导入的文件格式包含自定义归档格式、目录归档格式和tar归档格式。

gs\_restore具备如下两种功能。

- 导入至数据库  
如果指定了数据库，则数据将被导入到指定的数据库中。其中，并行导入必须指定连接数据库的密码。
- 导入至脚本文件  
如果未指定导入数据库，则创建包含重建数据库所需的SQL语句脚本，并将其写入至文件或者标准输出。该脚本文件等效于gs\_dump导出的纯文本格式文件。

gs\_restore工具在导入时，允许用户选择需要导入的内容，并支持在数据导入前对等待导入的内容进行排序。

## 操作步骤

### 📖 说明

gs\_restore默认是以追加的方式进行数据导入。为避免多次导入造成数据异常，在进行导入时，建议使用"-e"和"-c"参数，即导入前删除已存在于待导入数据库中的数据库对象，同时当出现导入错误时，忽略当前错误，继续执行导入任务，并在导入后会显示相应的错误信息。

**步骤1** 以root用户登录到服务器，执行如下命令进入数据存放路径。

```
cd /opt/bin
```

**步骤2** 使用gs\_restore命令，从postgres整个数据库内容的导出文件中，将数据库的所有对象的定义导入到backupdb。

```
gs_restore -W Passwd@123-U jack /home//backup/MPPDB_backup.tar -p 8000 -h 10.10.10.100 -d backupdb -s -e -c
```

表 4-21 常用参数说明

参数	参数说明	举例
-U	连接数据库的用户名。	-U jack
-W	指定用户连接的密码。 <ul style="list-style-type: none"><li>如果主机的认证策略是trust，则不会对数据库管理员进行密码验证，即无需输入-W选项；</li><li>如果没有-W选项，并且不是数据库管理员，会提示用户输入密码。</li></ul>	-W Passwd@123
-d	连接数据库dbname，并将数据导入到该数据库中。	-d backupdb
-p	指定服务器所监听的TCP端口或本地Unix域套接字后缀，以确保连接。	-p 8000
-h	“集群地址”如果通过公网地址连接，请指定为集群“公网访问地址”或“公网访问域名”，如果通过内网地址连接，请指定为集群“内网访问地址”或“内网访问域名”。	-h 10.10.10.100
-e	当发送SQL语句到数据库时如果出现错误，退出当前出现错误的任务，并执行其他导入任务。即默认状态下会忽略错误任务并继续执行导入，且在导入后会显示一系列错误信息。	-
-c	在重新创建数据库对象前，清理（删除）已存在于将要导入的数据库中的数据库对象。	-

参数	参数说明	举例
-s	只导入模式定义，不导入数据。当前的序列值也不会被导入。	-

其他参数说明请参见《工具参考》中“服务端工具>gs\_restore”章节。

----结束

## 示例

示例一：执行gs\_restore，导入指定MPPDB\_backup.dmp文件（自定义归档格式）中postgres数据库的数据和对象定义。

```
gs_restore -W Passwd@123 backup/MPPDB_backup.dmp -p 8000 -h 10.10.10.100 -d backupdb
gs_restore[2017-07-21 19:16:26]: restore operation successful
gs_restore: total time: 13053 ms
```

示例二：执行gs\_restore，导入指定MPPDB\_backup.tar文件（tar归档格式）中postgres数据库的数据和对象定义。

```
gs_restore backup/MPPDB_backup.tar -p 8000 -h 10.10.10.100 -d backupdb
gs_restore[2017-07-21 19:21:32]: restore operation successful
gs_restore[2017-07-21 19:21:32]: total time: 21203 ms
```

示例三：执行gs\_restore，导入指定MPPDB\_backup目录文件（目录归档格式）中postgres数据库的数据和对象定义。

```
gs_restore backup/MPPDB_backup -p 8000 -h 10.10.10.100 -d backupdb
gs_restore[2017-07-21 19:26:46]: restore operation successful
gs_restore[2017-07-21 19:26:46]: total time: 21003 ms
```

示例四：执行gs\_restore，将postgres数据库的所有对象的定义导入至backupdb数据库。导入前，postgres存在完整的定义和数据，导入后，backupdb数据库只存在所有对象定义，表没有数据。

```
gs_restore -W Passwd@123 /home//backup/MPPDB_backup.tar -p 8000 -h 10.10.10.100 -d backupdb -s -e -c
gs_restore[2017-07-21 19:46:27]: restore operation successful
gs_restore[2017-07-21 19:46:27]: total time: 32993 ms
```

示例五：执行gs\_restore，导入MPPDB\_backup.dmp文件中PUBLIC模式的所有定义和数据。在导入时会先删除已经存在的对象，如果原对象存在跨模式的依赖则需手工强制干预。

```
gs_restore backup/MPPDB_backup.dmp -p 8000 -h 10.10.10.100 -d backupdb -e -c -n PUBLIC
gs_restore: [archiver (db)] Error while PROCESSING TOC:
gs_restore: [archiver (db)] Error from TOC entry 313; 1259 337399 TABLE table1 gaussdba
gs_restore: [archiver (db)] could not execute query: ERROR: cannot drop table table1 because other objects
depend on it
DETAIL: view t1.v1 depends on table table1
HINT: Use DROP ... CASCADE to drop the dependent objects too.
Command was: DROP TABLE public.table1;
```

手工删除依赖，导入完成后重新创建。

```
gs_restore backup/MPPDB_backup.dmp -p 8000 -h 10.10.10.100 -d backupdb -e -c -n PUBLIC
gs_restore[2017-07-21 19:52:26]: restore operation successful
gs_restore[2017-07-21 19:52:26]: total time: 2203 ms
```

示例六：执行gs\_restore，导入MPPDB\_backup.dmp文件中PUBLIC模式下表hr.staffs的定义。在导入之前，hr.staffs表不存在。

```
gs_restore backup/MPPDB_backup.dmp -p 8000 -h 10.10.10.100 -d backupdb -e -c -s -n PUBLIC -t hr.staffs
gs_restore[2017-07-21 19:56:29]: restore operation successful
gs_restore[2017-07-21 19:56:29]: total time: 21000 ms
```

示例七：执行gs\_restore，导入MPPDB\_backup.dmp文件中PUBLIC模式下表hr.staffs的数据。在导入之前，hr.staffs表不存在数据。

```
gs_restore backup/MPPDB_backup.dmp -p 8000 -h 10.10.10.100 -d backupdb -e -a -n PUBLIC -t hr.staffs
gs_restore[2017-07-21 20:12:32]: restore operation successful
gs_restore[2017-07-21 20:12:32]: total time: 20203 ms
```

示例八：执行gs\_restore，导入指定表hr.staffs的定义。在导入之前，hr.staffs表的数据是存在的。

```
human_resource=# select * from hr.staffs;
 staff_id | first_name | last_name | email | phone_number | hire_date | employment_id | salary |
-----+-----+-----+-----+-----+-----+-----+-----+
 200 | Jennifer | Whalen | JWHALEN | 515.123.4444 | 1987-09-17 00:00:00 | AD_ASST | 4400.00 |
 201 | Michael | Hartstein | MHARTSTE | 515.123.5555 | 1996-02-17 00:00:00 | MK_MAN | 13000.00 |
-----+-----+-----+-----+-----+
(2 rows)
```

```
gs_restore -W Passwd@123 /home//backup/MPPDB_backup.tar -p 8000 -h 10.10.10.100 -d human_resource -n hr -t staffs -s -e
restore operation successful
total time: 904 ms
```

```
human_resource=# select * from hr.staffs;
 staff_id | first_name | last_name | email | phone_number | hire_date | employment_id | salary |
-----+-----+-----+-----+-----+-----+-----+-----+
(0 rows)
```

示例九：执行gs\_restore，导入staffs和areas两个指定表的定义和数据。在导入之前，staffs和areas表不存在。

```
human_resource=# \d
List of relations
Schema | Name | Type | Owner | Storage
-----+-----+-----+-----+-----+
hr | employment_history | table | | {orientation=row,compression=no}
hr | employments | table | | {orientation=row,compression=no}
hr | places | table | | {orientation=row,compression=no}
hr | sections | table | | {orientation=row,compression=no}
hr | states | table | | {orientation=row,compression=no}
(5 rows)
```

```
gs_restore -W Passwd@123 /home/mppdb/backup/MPPDB_backup.tar -p 8000 -h 10.10.10.100 -d human_resource -n hr -t staffs -n hr -t areas
restore operation successful
total time: 724 ms
```

```
human_resource=# \d
List of relations
Schema | Name | Type | Owner | Storage
-----+-----+-----+-----+-----+
hr | areas | table | | {orientation=row,compression=no}
hr | employment_history | table | | {orientation=row,compression=no}
hr | employments | table | | {orientation=row,compression=no}
hr | places | table | | {orientation=row,compression=no}
(4 rows)
```

```
hr | sections | table | | {orientation=row,compression=no}
hr | staffs | table | | {orientation=row,compression=no}
hr | states | table | | {orientation=row,compression=no}
(7 rows)
```

```
human_resource=# select * from hr.areas;
```

```
area_id | area_name
```

```
-----+
```

```
4 | Iron
1 | Wood
2 | Lake
3 | Desert
```

```
(4 rows)
```

示例十：执行gs\_restore，导入hr的模式，包含模式下的所有对象定义和数据。

```
gs_restore -W Passwd@123 /home//backup/MPPDB_backup1.sql -p 8000 -h 10.10.10.100 -d backupdb -n hr -e -c
```

```
restore operation successful
```

```
total time: 702 ms
```

示例十一：执行gs\_restore，同时导入hr和hr1两个模式，仅导入模式下的所有对象定义。

```
gs_restore -W Passwd@123 /home//backup/MPPDB_backup2.dmp -p 8000 -h 10.10.10.100 -d backupdb -n hr -n hr1 -s
```

```
restore operation successful
```

```
total time: 665 ms
```

示例十二：执行gs\_restore，将human\_resource数据库导出文件进行解密并导入至backupdb数据库中。

```
create database backupdb;
```

```
gs_restore /home//backup/MPPDB_backup.tar -p 8000 -h 10.10.10.100 -d backupdb --with-key=1234567812345678
```

```
restore operation successful
```

```
total time: 23472 ms
```

```
gsql -d backupdb -p 8000 -r
```

```
gsql ((GaussDB 8.2.0 build 39137c2d) compiled at 2022-09-23 15:43:11 commit 3629 last mr 5138 release)
```

```
Non-SSL connection (SSL connection is recommended when requiring high-security)
```

```
Type "help" for help.
```

```
backupdb=# select * from hr.areas;
```

```
area_id | area_name
```

```
-----+
```

```
4 | Iron
1 | Wood
2 | Lake
3 | Desert
```

```
(4 rows)
```

示例十三：用户user1不具备将导出文件中数据导入至数据库backupdb的权限，而角色role1具备该权限，要实现将文件数据导入数据库backupdb，可以在导出命令中设置--role角色为role1，使用role1的权限，完成导出目的。

```
human_resource=# CREATE USER user1 IDENTIFIED BY 'password';
```

```
gs_restore -U user1 -W 1234@abc /home//backup/MPPDB_backup.tar -p 8000 -h 10.10.10.100 -d backupdb --role role1 --rolepassword password
```

```
restore operation successful
```

```
total time: 554 ms
```

```
gsql -d backupdb -p 8000 -r
```

```
gsql ((GaussDB 8.2.0 build 39137c2d) compiled at 2022-09-23 15:43:11 commit 3629 last mr 5138 release)
```

```
Non-SSL connection (SSL connection is recommended when requiring high-security)
```

```
Type "help" for help.
```

```
backupdb=# select * from hr.areas;
 area_id |   area_name
-----+-----
      4 | Iron
      1 | Wood
      2 | Lake
      3 | Desert
(4 rows)
```

## 4.5 导出数据

### 4.5.1 使用外表导出数据到 OBS

#### 4.5.1.1 关于 OBS 并行导出

##### 概述

GaussDB(DWS)数据库支持通过OBS外表并行导出数据：通过OBS外表设置的导出模式、导出数据格式等信息来指定导出的数据文件，利用多DN并行的方式，将数据从GaussDB(DWS)数据库导出到外部，存放在OBS对象存储服务器上，从而提高整体导出性能。

- CN只负责任务的规划及下发，把数据导出的工作交给了DN，释放了CN的资源，使其有能力处理外部请求。
- 通过让各个DN都参与数据导出，充分利用各个设备的计算能力及网络带宽。
- 支持多个OBS服务并发导出，导出的桶和对象的路径必须不同并且不能为空。
- 选择OBS服务器与集群节点处于联网状态，导出速率会受网络带宽影响。
- 支持数据文件格式：TEXT、CSV、ORC。单行数据大小需<1GB。
- ORC格式的数据仅8.1.0版本以后支持。

目前，有两种外表可供OBS并行导出使用：

- 方式一：外表使用SERVER的FOREIGN DATA WRAPPER类型为dist\_fdw，具体可参考[CREATE FOREIGN TABLE \(OBS导入导出\)](#)章节。支持导出CSV、TEXT格式的数据，参见[导出CSV、TEXT数据到OBS（方式一）](#)。
- 方式二：外表使用SERVER的FOREIGN DATA WRAPPER类型为dfs\_fdw，具体可参考[CREATE FOREIGN TABLE \(SQL on OBS or Hadoop\)](#)章节。仅支持导出ORC格式的数据，参见[导出ORC数据到OBS（方式二）](#)。

##### 须知

- OBS导入导出数据时，不支持中文路径。
- OBS导入导出数据时，暂不支持跨Region进行OBS数据导入导出，必须确保OBS和DWS集群在同一个Region中。

##### 相关概念

- **数据源文件**：存储有数据的TEXT、CSV、ORC文件。

- **OBS**: 对象存储服务, 是一种可存储文档、图片、音视频等非结构化数据的云存储服务。从GaussDB(DWS)并行导出数据时, 数据对象放置在OBS服务器上。
- **桶 ( Bucket )**: 对OBS中的一个存储空间的形象称呼, 是存储对象的容器。
  - 对象存储是一种非常扁平化的存储方式, 桶中存储的对象都在同一个逻辑层级, 去除了文件系统中的多层级树形目录结构。
  - 在OBS中, 桶名必须是全局唯一的且不能修改, 即用户创建的桶不能与自己已创建的其他桶名称相同, 也不能与其他用户创建的桶名称相同。每个桶在创建时都会生成默认的桶ACL ( Access Control List ), 桶ACL列表的每项包含了对被授权用户授予什么样的权限, 如读取权限、写入权限、完全控制权限等。用户只有对桶有相应的权限, 才可以对桶进行操作, 如创建、删除、显示、设置桶ACL等。
  - 一个用户最多可创建100个桶, 但每个桶中存放的总数据容量和对象/文件数量没有限制。
- **对象**: 是存储在OBS中的基本数据单位。用户上传的数据以对象的形式存储在OBS的桶中。对象的属性包括名称Key, Metadata, Data。

通常, 将对象等同于文件来进行管理, 但是由于OBS是一种对象存储服务, 并没有文件系统中的文件和文件夹概念。为了使用户更方便进行管理数据, OBS提供了一种方式模拟文件夹。通过在对象的名称中增加“/”, 如tpcds1000/stock.csv, tpcds1000可以等同于文件夹, stock.csv就可以等同于文件名, 而对象名称 ( key ) 仍然是tpcds1000/stock.csv、对象的内容就是stock.csv数据文件的内容。
- **Key**: 对象的名称 ( 键 ), 为经过UTF-8编码的长度大于0且不超过1024的字符序列, 一个桶里的每个对象必须拥有唯一的对象键值。用户可使用桶名+对象名来存储和获取对应的对象。
- **Metadata**: 对象元数据, 用来描述对象的信息。元数据又可分为系统元数据和用户元数据。这些元数据以键值对 ( Key-value ) 的形式随http头域一起上传到OBS系统。
  - 系统元数据由OBS系统产生, 在处理对象数据时使用。系统元数据包括: Date, Content-length, last-modify, Content-MD5等。
  - 用户元数据由用户上传对象时指定, 是用户自己对对象的一些描述信息。
- **Data**: 对象的数据内容, OBS对于数据的内容是无感知的, 即认为对象内的数据为无状态的二进制数据。
- **外表**: 用于识别数据源文件中的数据。外表中保存了数据源文件的位置、文件格式、存放位置、编码格式、数据间的分隔符等信息。

## 相关原理

下面分别从以下两类表介绍从集群导出数据到OBS的原理。

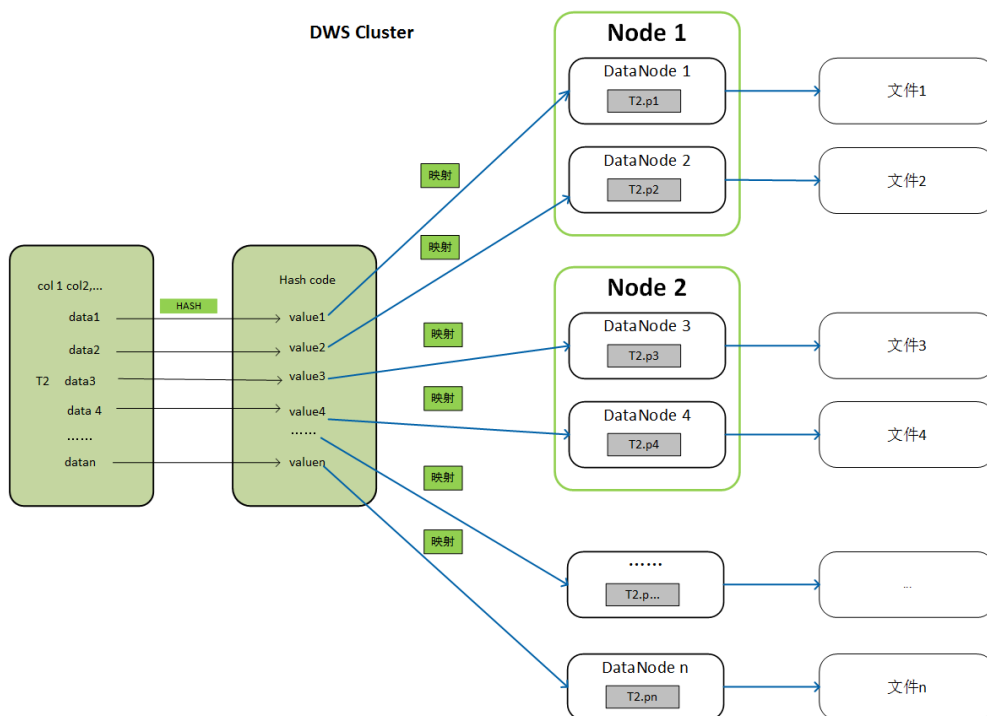
- **Hash分布表**: 在建表语句中指定了DISTRIBUTE BY HASH (Column\_Name)的表。

对于Hash分布表而言, 在存储表数据时, 采用的是散列 ( Hash ) 方式的存储原理, 如图4-9所示, 图中以将表 ( T2 ) 导出到OBS为例。

在存储表数据时, 将表 ( T2 ) 中指定的Hash字段 ( col2 ) 进行Hash运算后, 生成相应的Hash值 ( value ), 根据DN与Hash值的映射关系, 将该元组分发到相应的DN上进行存储。

在导出数据到OBS时, 每一个存储了导出表的 ( T2 ) 数据的DN会直接向OBS导出属于自己的数据文件。多个节点将并行导出原始数据。

图 4-9 Hash 分布原理



- Replication表：在建表语句中指定了DISTRIBUTE BY REPLICATION的表。  
Replication表在GaussDB(DWS)集群的每个节点上都会存储一份完整的表数据。因此，在导出数据到OBS时，GaussDB(DWS)只会随机选择一个DN节点向OBS导出数据。

## 导出文件的命名规则

GaussDB(DWS)向OBS导出数据的文件命名规则如下：

- 从DN节点导出数据时，以segment的格式存储在OBS服务中，文件命名规则为“表名称\_节点名称\_segment.n”。这里的“n”是从0开始按照自然数0、1、2、3递增。  
例如，表t1在datanode3里面的数据导出成文件“t1\_datanode3\_segment.0”、“t1\_datanode3\_segment.1”等等，以此类推。  
对于来自不同集群或不同数据库的数据，建议用户可以将数据导出到不同的OBS桶或者同一个OBS桶的不同路径下。
- 每个segment可以存储的最大数据为1GB，并且不能切断元组。如果segment超过1GB，超过1GB的数据会作为第二个segment进行存储。  
例如：  
datanode3节点将表（t1）导出到OBS时，一个segment里面已经存储了100条元组，大小是1023MB，如果再插入一条5MB的元组，大小就变成1028MB了，此时会以1023MB生成一个“t1\_datanode3\_segment.0”保存到OBS服务中，新插入的第101条元组作为下一个“t1\_datanode3\_segment.1”保存到OBS服务中。
- 导出Hash分布表时，每个DataNode节点生成的segment数量和集群的DataNode节点数无关，而是取决于每个DataNode节点上存储的数据量。按照Hash方式存储在各个DataNode节点上的数据分布不一定均匀。



例如，一个有6个DataNode节点的集群，DataNode1到DataNode6分别有1.5GB、0.7GB、0.6GB、0.8GB、0.4GB、0.5GB的数据，则导出时会生成7个OBS segment文件，其中DataNode1会生成1GB和0.5GB两个segment文件。

## 导出流程

图 4-10 并行导出流程

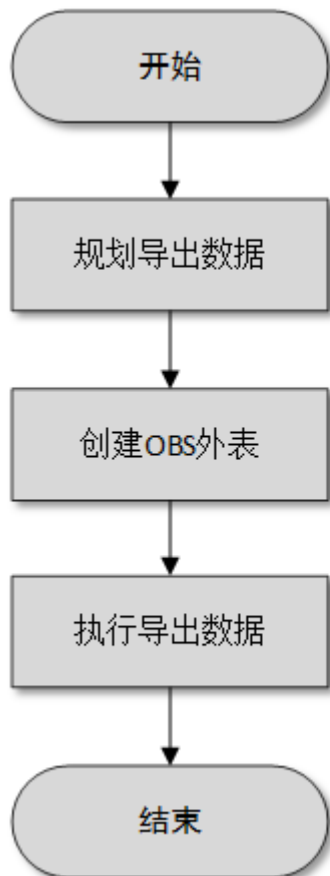


表 4-22 流程说明

流程	说明	子任务
规划导出数据	创建OBS桶，并在桶中创建导出后的数据文件的存放目录。 详细请参见 <a href="#">规划导出数据</a> 。	-
创建OBS外表。	创建外表用于帮助OBS指定的待导出数据文件。外表中保存了数据源文件导出后的位置、文件格式、编码格式、数据间的分隔符等信息。 详细内容请参见 <a href="#">创建OBS外表</a> 。	-

流程	说明	子任务
执行导出数据。	在创建好外表后，通过INSERT语句，将数据快速、高效地导出到数据文件中。 详细内容请参见 <a href="#">执行导出</a> 。	-

## 4.5.1.2 导出 CSV、TEXT 数据到 OBS（方式一）

### 4.5.1.2.1 规划导出数据

#### 操作场景

在OBS上规划导出数据存放的位置。

#### 规划 OBS 存储位置和文件

导出数据需要指定数据在OBS中的存储路径（需指定到目录），导出的数据可以按CSV解析格式保存到文件中。系统还支持TEXT类型的解析格式，将数据导出保存便于导入不同的应用程序。

导出路径的目标目录中不能存在任何文件。

#### 规划 OBS 桶权限

在导出数据时，执行导出操作的用户需要具备以下条件：

- 已开通OBS服务。
- 具备数据导出路径所在的OBS桶的写入权限。  
通过配置桶的ACL权限，可以将写入权限授予指定的用户帐号。  
具体操作请参见[根据规划准备OBS存储位置和OBS桶的写权限](#)。

#### 规划导出数据和外表

提前在数据库的表中准备好待导出的数据，且单行数据大小需要小于1GB。根据导出数据，规划匹配用户数据的外表，外表的字段、字段类型以及长度等属性需要能够对应用户数据。

### 根据规划准备 OBS 存储位置和 OBS 桶的写权限

**步骤1** 创建OBS桶，并在OBS桶中新建文件夹作为导出数据的存放目录。

1. 登录OBS管理控制台。  
单击“服务列表”，选择“对象存储服务”，打开OBS管理控制台页面。
2. 创建桶。  
如何创建OBS桶，具体请参见《对象存储服务控制台指南》中的[创建桶](#)章节。  
例如，创建桶：“mybucket”。

### 3. 新建文件夹。

在OBS桶中，创建导出数据的存放目录。

具体请参见《对象存储服务控制台指南》中的[新建文件夹](#)章节。

例如，在已创建的OBS桶“mybucket”中新建一个文件夹“output\_data”。

#### 步骤2 获取新建文件夹的OBS路径。

在创建外表时需要指定导出数据文件的OBS存放目录的路径，用于创建外表时location参数设置。

location参数中OBS文件夹的路径由“obs://”、桶名和文件路径组成，即为：obs://<bucket\_name>/<file\_path>/

例如，在本例中，location参数中OBS文件夹路径为：

```
obs://mybucket/output_data/
```

#### 说明

执行数据导出的时候，导出数据文件的OBS存放目录的路径必须为空。

#### 步骤3 为导出用户设置OBS桶的写权限。

在导出数据时，执行导出操作的用户需要具备数据导出路径所在的OBS桶的写入权限。通过配置桶的ACL权限，可以将写入权限授予指定的用户帐号。

具体请参见《对象存储服务控制台指南》中的[配置桶ACL](#)章节。

---结束

## 4.5.1.2.2 创建 OBS 外表

### 操作步骤

**步骤1** 根据[规划导出数据](#)中规划的路径，由此确定创建外表时使用的参数location的值。

**步骤2** 用户获取OBS访问协议对应的AK值和SK值。

获取访问密钥，请登录管理控制台，单击右上角的用户名并选择菜单“我的凭证”，然后在左侧导航树单击“管理访问密钥”。在访问密钥页面，可以查看已有的访问密钥ID（即AK），如果要同时获取AK和SK，可以单击“新增访问密钥”创建并下载访问密钥。

**步骤3** 梳理待导出数据的数据格式信息，确定创建外表时使用的数据格式参数的值。详细使用请参见数据格式参数。

**步骤4** 根据前面步骤确定的参数，**创建OBS外表**。外表的创建语法以及详细使用，请参考CREATE FOREIGN TABLE (OBS导入导出)。

---结束

### 示例一

例如，在GaussDB(DWS)数据库中，创建一个format参数为text的只写外表，用于导出text文件。设置的参数信息如下所示：

- **location**

在[规划导出数据](#)中，通过[获取数据源文件的OBS路径](#)已经获取到数据源文件的OBS路径。

因此，设置参数“location”为：

```
location 'obs://mybucket/output_data/';
```

- **访问密钥（AK和SK）**

- 用户获取OBS访问协议对应的AK值（access\_key）。
- 用户获取OBS访问协议对应的SK值（secret\_access\_key）。

**说明**

用户在创建用户时已经获取了access\_key和secret\_access\_key的密钥，请根据实际密钥替换示例中的内容。

- **设置数据格式参数**

- **数据源文件格式（format）**为TEXT。
- **编码格式（encoding）**为UTF-8。
- **是否使用加密（encrypt）**，默认为'off'。
- **字段分隔符（delimiter）**为'|'。

根据以上信息，创建的外表如下所示：

```
DROP FOREIGN TABLE IF EXISTS product_info_output_ext1;
CREATE FOREIGN TABLE product_info_output_ext1
(
  c_bigint bigint,
  c_char char(30),
  c_varchar varchar(30),
  c_nvarchar2 nvarchar2(30) ,
  c_data date,
  c_time time ,
  c_test varchar(30))
server gsmpp_server
options (
  LOCATION 'obs://mybucket/output_data/',
  ACCESS_KEY 'access_key_value_to_be_replaced',
  SECRET_ACCESS_KEY 'secret_access_key_value_to_be_replaced'
  format 'text',
  delimiter '|',
  encoding 'utf-8',
  encrypt 'on'
)
WRITE ONLY;
```

返回如下信息表示创建成功：

```
CREATE FOREIGN TABLE
```

## 示例二

例如，在GaussDB(DWS)数据库中，创建一个format参数为CSV的只写外表，用于导出CSV文件。设置的参数信息如下所示：

- **location**

在[规划导出数据](#)中，通过[获取数据源文件的OBS路径](#)，已经获取到数据源文件的OBS路径。

因此，设置参数“location”为：

```
location 'obs://mybucket/output_data/';
```

- **访问密钥（AK和SK）**

- 用户获取OBS访问协议对应的AK值（access\_key）。

- 用户获取OBS访问协议对应的SK值（secret\_access\_key）。

#### 📖 说明

用户在创建用户时已经获取了access\_key和secret\_access\_key的密钥，请根据实际密钥替换示例中的对应内容。

- **设置数据格式参数**

- **数据源文件格式（format）**为CSV。
- **编码格式（encoding）**为UTF-8。
- **是否使用加密（encrypt）**，默认为'off'。
- **字段分隔符（delimiter）**为','。
- **header（指定导出数据文件是否包含标题行）**

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，使用缺省值false，不需要设置，表示导出的数据文件第一行不是标题行（即表头）。

根据以上信息，创建的外表如下所示：

```
DROP FOREIGN TABLE IF EXISTS product_info_output_ext2;
CREATE FOREIGN TABLE product_info_output_ext2
(
  product_price      integer      not null,
  product_id         char(30)     not null,
  product_time       date         ,
  product_level      char(10)     ,
  product_name       varchar(200) ,
  product_type1      varchar(20)  ,
  product_type2      char(10)     ,
  product_monthly_sales_cnt integer  ,
  product_comment_time date       ,
  product_comment_num integer     ,
  product_comment_content varchar(200)
)
SERVER gsmpp_server
OPTIONS(
  location 'obs://mybucket/output_data/',
  FORMAT 'CSV' ,
  DELIMITER ';;',
  encoding 'utf8',
  header 'false',
  ACCESS_KEY 'access_key_value_to_be_replaced',
  SECRET_ACCESS_KEY 'secret_access_key_value_to_be_replaced'
)
WRITE ONLY ;
```

返回如下信息表示创建成功：

```
CREATE FOREIGN TABLE
```

### 4.5.1.2.3 执行导出

#### 操作步骤

##### 步骤1 执行数据导出。

```
INSERT INTO [foreign table 表名] SELECT * FROM [源表名];
```

----结束

## 执行导出数据示例

- **示例1**：将表product\_info\_output的数据通过外表product\_info\_output\_ext导出到数据文件中。

```
INSERT INTO product_info_output_ext SELECT * FROM product_info_output;
```

若出现以下类似信息，说明数据导出成功。

```
INSERT 0 10
```

- **示例2**：通过条件过滤（WHERE product\_price>500），向数据文件中导出部分数据。

```
INSERT INTO product_info_output_ext SELECT * FROM product_info_output WHERE product_price>500;
```

### 📖 说明

- 在需要执行多次数据导出时，导出数据的存放路径必须为空，否则将导出失败。
- 对于特殊的数据类型如RAW类型，在导出之后是一个二进制文本，导入工具无法识别。需使用RAWTOHEX()函数将其转换为16进制文本导出。

### 4.5.1.2.4 示例

## 单表导出操作步骤

通过创建外表，将数据库中的单表导出至OBS的两个桶中。

**步骤1** 用户通过管理控制台登录到OBS数据服务器。在OBS数据服务器上，分别创建数据文件存放的两个桶“/input-data1”“/input-data2”，并创建每个桶下面的data目录“/input-data1/data”“/input-data2/data”。

**步骤2** 在GaussDB(DWS)数据库上，创建外表tpcds.customer\_address\_ext1和tpcds.customer\_address\_ext2用于OBS数据服务器接收数据库导出数据。

OBS与集群处于同一区域，需要导出的表为GaussDB(DWS)示例表tpcds.customer\_address。

其中设置的**导出信息**如下所示：

- 由于OBS数据服务器上的数据源文件存放目录为“/input-data1/data/”和“/input-data2/data/”，所以设置tpcds.customer\_address\_ext1参数“location”为“obs://input-data1/data/”，设置tpcds.customer\_address\_ext2参数“location”为“obs://input-data2/data/”。

设置的**数据格式信息**是根据表从数据库导出时需要的详细数据格式参数信息指定的，参数设置如下所示：

- 数据源文件格式（format）为CSV。
- 编码格式（encoding）为UTF-8。
- 字段分隔符（delimiter）为'E'\x08'。
- 是否使用加密（encrypt），默认为'off'。
- 用户获取OBS访问协议对应的AK值（access\_key）。（必选）
- 用户获取OBS访问协议对应的SK值（secret\_access\_key）。（必选）

### 📖 说明

用户在创建用户时已经获取了access\_key和secret\_access\_key的密钥，请根据实际密钥替换示例中的内容。

根据以上信息，创建的外表如下所示：

```
CREATE FOREIGN TABLE tpcds.customer_address_ext1
(
ca_address_sk      integer           ,
ca_address_id      char(16)          ,
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)          )
SERVER gsmpp_server
OPTIONS(LOCATION 'obs://input-data1/data/',
FORMAT 'CSV',
ENCODING 'utf8',
DELIMITER E'\x08',
ENCRYPT 'off',
ACCESS_KEY 'access_key_value_to_be_replaced',
SECRET_ACCESS_KEY 'secret_access_key_value_to_be_replaced'
)Write Only;
CREATE FOREIGN TABLE tpcds.customer_address_ext2
(
ca_address_sk      integer           ,
ca_address_id      char(16)          ,
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)          )
SERVER gsmpp_server
OPTIONS(LOCATION 'obs://input-data2/data/',
FORMAT 'CSV',
ENCODING 'utf8',
DELIMITER E'\x08',
ENCRYPT 'off',
ACCESS_KEY 'access_key_value_to_be_replaced',
SECRET_ACCESS_KEY 'secret_access_key_value_to_be_replaced'
)Write Only;
```

**步骤3** 在GaussDB(DWS)数据库上，将数据表tpcds.customer\_address并发导出到外表tpcds.customer\_address\_ext1和tpcds.customer\_address\_ext2中。

```
INSERT INTO tpcds.customer_address_ext1 SELECT * FROM tpcds.customer_address;
INSERT INTO tpcds.customer_address_ext2 SELECT * FROM tpcds.customer_address;
```

### 说明

OBS外表在设计上禁止往非空的路径下导出文件，但是在并发场景下会出现同一路径导出文件的情况，此时会发生异常。

异常场景：假如用户使用同一张表的数据并发导出到同一个OBS的外表，在一条SQL语句执行在OBS服务器上没有生成文件时，另一条SQL语句也执行导出，最终执行结果为两条SQL语句均执行成功，产生数据覆盖现象，建议用户在执行OBS外表导出任务时，不要往同一OBS外表并发导出。

----**结束**

## 多表并发导出操作步骤

通过创建的两个外表，将数据库中的两个表分别导出至OBS的桶中。

**步骤1** 用户通过管理控制台登录到OBS数据服务器。在OBS数据服务器上，分别创建数据文件存放的两个桶“/input-data1”“/input-data2”，并创建每个桶下面的data目录“/input-data1/data”“/input-data2/data”。

**步骤2** 在GaussDB(DWS)数据库上，创建外表tpcds.customer\_address\_ext1和tpcds.customer\_address\_ext2分别用于OBS服务器接收导出的数据。

规划OBS与集群处于同一区域，需要导出的表为已存在的表tpcds.customer\_address和tpcds.customer\_demographics。

其中设置的导出信息如下所示：

- 由于OBS服务器上的数据源文件存放目录为“/input-data1/data/”和“/input-data2/data/”，所以设置tpcds.customer\_address\_ext1参数“location”为“obs://input-data1/data/”，设置tpcds.customer\_address\_ext2参数“location”为“obs://input-data2/data/”。

设置的**数据格式信息**是根据表从GaussDB(DWS)中导出时需要的详细数据格式参数信息指定的，参数设置如下所示：

- 数据源文件格式（format）为CSV。
- 编码格式（encoding）为UTF-8。
- 字段分隔符（delimiter）为'E'\x08'。
- 是否使用加密（encrypt），默认为'off'。
- 用户获取OBS访问协议对应的AK值（access\_key）。（必选）
- 用户获取OBS访问协议对应的SK值（secret\_access\_key）。（必选）

### 📖 说明

用户在创建用户时已经获取了access\_key和secret\_access\_key的密钥，请根据实际密钥替换示例中的内容。

根据以上信息，创建的外表如下所示：

```
CREATE FOREIGN TABLE tpcds.customer_address_ext1
(
  ca_address_sk          integer          ,
  ca_address_id         char(16)         ,
  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)
)
SERVER gsmpp_server
OPTIONS(LOCATION 'obs://input-data1/data/',
FORMAT 'CSV',
ENCODING 'utf8',
DELIMITER E'\x08',
ENCRYPT 'off',
ACCESS_KEY 'access_key_value_to_be_replaced');
```



```
SECRET_ACCESS_KEY 'secret_access_key_value_to_be_replaced'
)Write Only;
CREATE FOREIGN TABLE tpcds.customer_address_ext2
(
ca_address_sk          integer          ,
ca_address_id          char(16)         ,
ca_address_name        varchar(20)      ,
ca_address_code        integer         ,
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)    )
SERVER gsmpp_server
OPTIONS(LOCATION 'obs://input_data2/data/',
FORMAT 'CSV',
ENCODING 'utf8',
DELIMITER E'\x08',
QUOTE E'\x1b',
ENCRYPT 'off',
ACCESS_KEY 'access_key_value_to_be_replaced',
SECRET_ACCESS_KEY 'secret_access_key_value_to_be_replaced'
)Write Only;
```

**步骤3** 在GaussDB(DWS)数据库上，将数据表 tpcds.customer\_address和 tpcds.warehouse 并发导出到外表tpcds.customer\_address\_ext1和tpcds.customer\_address\_ext2中。

```
INSERT INTO tpcds.customer_address_ext1 SELECT * FROM tpcds.customer_address;
INSERT INTO tpcds.customer_address_ext2 SELECT * FROM tpcds.warehouse;
```

----结束

### 4.5.1.3 导出 ORC 数据到 OBS（方式二）

#### 4.5.1.3.1 规划导出数据

OBS导出数据准备：请参见[规划导出数据](#)完成OBS导出数据准备。

OBS导出支持的数据类型请参见[表4-6](#)。

HDFS导出数据准备：HDFS导出准备即配置MRS，具体信息可参考《[MapReduce服务用户指南](#)》。

#### 4.5.1.3.2 创建外部服务器

OBS创建外部服务器请参见[创建外部服务器](#)。

HDFS创建外部服务器请参见[手动创建外部服务器](#)。

#### 4.5.1.3.3 创建外表

当完成[创建外部服务器](#)后，在GaussDB(DWS)数据库中创建一个OBS/HDFS只写外表，用来访问存储在OBS/HDFS上的数据。此外表是只写的，只能用于导出操作。

创建外表的语法格式如下，详细的描述请参见[CREATE FOREIGN TABLE \(SQL on OBS or Hadoop\)](#)章节。

```
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 dfs_server
OPTIONS ( { option_name ' value ' } [ , ... ] )
[ {WRITE ONLY } ]
DISTRIBUTE BY {ROUNDROBIN | REPLICATION}
[ PARTITION BY ( column_name ) [ AUTOMAPPED ] ] ;
```

例如，创建一个名为"*product\_info\_ext\_obs*"的外表，对语法中的参数按如下描述进行设置：

- **table\_name**

外表的表名。

- **表字段定义**

- **column\_name**：外表中的字段名。

- **type\_name**：字段的数据类型。

多个字段用“,” 隔开。

- **SERVER dfs\_server**

外表的外部服务器名称，这个server必须存在。外表通过设置外部服务器连接OBS/HDFS读取数据。

此处应填写为参照[创建外部服务器](#)创建的外部服务器名称。

- **OPTIONS参数**

用于指定外表数据的各类参数，关键参数如下所示。

- “format”：表示导出的数据文件格式，支持“orc”格式。

- “foldername”：必选参数。外表中数据源文件目录。OBS：数据源文件的OBS路径，此处仅需要填写“/桶名/文件夹目录层级/”。HDFS：HDFS文件系统上的路径。此选项对WRITE ONLY外表为必选项。

- “encoding”：外表中数据源文件的编码格式名称，缺省为utf8。

- filesize

指定WRITE ONLY外表的文件大小，单位为MB。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的外表有效。

取值范围：[1, 1024]的整数。

#### 说明

filesize参数只对ORC格式的WRITE ONLY的HDFS外表有效。

- compression

指定ORC格式文件的压缩方式，此选项为可选项。

取值范围：zlib, snappy, lz4。缺省值为snappy。

- version

指定ORC格式的版本号，此选项为可选项。此语法仅对WRITE ONLY的外表有效。

取值范围：目前仅支持0.12。缺省值为0.12。

- 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外表有效。

取值范围：该数据库编码支持转换的数据编码。

### 说明

dataencoding参数只对ORC格式的WRITE ONLY的HDFS外表有效。

- **语法中的其他参数**

其他参数均为可选参数，用户可以根据自己的需求进行设置，在本例中不需要设置。

根据以上信息，创建外表命令如下所示：

```
DROP FOREIGN TABLE IF EXISTS product_info_ext_obs;
```

---建立不包含分区列的OBS外表，表关联的外部服务器为obs\_server，表对应的OBS服务上的文件格式为'orc'，OBS上的数据存储路径为'/mybucket/data/'。

```
CREATE FOREIGN TABLE product_info_ext_obs
(
  product_price      integer      ,
  product_id        char(30)     ,
  product_time      date         ,
  product_level     char(10)     ,
  product_name      varchar(200) ,
  product_type1     varchar(20)  ,
  product_type2     char(10)     ,
  product_monthly_sales_cnt integer ,
  product_comment_time date      ,
  product_comment_num integer    ,
  product_comment_content varchar(200)
) SERVER obs_server
OPTIONS (
  format 'orc',
  foldername '/mybucket/demo.db/product_info_orc',
  compression 'snappy',
  version '0.12'
) Write Only;
```

#### 4.5.1.3.4 执行导出

### 操作步骤

**步骤1** 执行数据导出。

```
INSERT INTO [foreign table 表名] SELECT * FROM [源表名];
```

----结束

### 执行导出数据示例

- **示例1：**将表product\_info\_output的数据通过外表product\_info\_output\_ext导出到数据文件中。

```
INSERT INTO product_info_output_ext SELECT * FROM product_info_output;
```

若出现以下类似信息，说明数据导出成功。

```
INSERT 0 10
```

- **示例2：**通过条件过滤（WHERE product\_price>500），向数据文件中导出部分数据。

```
INSERT INTO product_info_output_ext SELECT * FROM product_info_output WHERE product_price>500;
```

### 说明

对于特殊的数据类型如RAW类型，在导出之后是一个二进制文本，导入工具无法识别。需使用RAWTOHEX()函数将其转换为16进制文本导出。

## 4.5.2 导出 ORC 数据到 MRS

### 4.5.2.1 导出 ORC 数据概述

GaussDB(DWS)数据库支持通过HDFS外表导出ORC格式数据至MRS，通过外表设置的导出模式、导出数据格式等信息来指定导出的数据文件，利用多DN并行的方式，将数据从GaussDB(DWS)数据库导出到外部，存放在HDFS文件系统上，从而提高整体导出性能。

- CN只负责任务的规划及下发，把数据导出的工作交给了DN，释放了CN的资源，使其有能力处理外部请求。
- 通过让各个DN都参与数据导出，充分利用各个设备的计算能力及网络带宽。
- 支持多个hdfs server并发导出，导出的路径可以为空，命名规则需与导出文件一致。
- 选择MRS服务与集群节点处于联网状态，导出速率会受网络带宽影响。
- 支持数据文件格式：ORC。

### 导出文件命名规则

GaussDB(DWS)导出ORC数据的文件命名规则如下：

1. 导出至MRS (HDFS)：从DN节点导出数据时，以segment的格式存储在HDFS中，文件命名规则为“**mpp\_数据库名\_模式名\_表名称\_节点名称\_n.orc**”。这里的“n”是从0开始按照自然数0、1、2、3递增。
2. 对于来自不同集群或不同数据库的数据，建议用户可以将数据导出到不同路径下。ORC格式文件大小最大为128M，Stripe大小最大为64M。
3. 导出完成后会生成\_SUCCESS标记文件。

### 4.5.2.2 规划导出数据

MRS导出支持的数据类型请参见[表4-6](#)。

HDFS导出数据准备：HDFS导出准备即配置MRS，具体信息可参考《[MapReduce服务用户指南](#)》。

### 4.5.2.3 创建外部服务器

HDFS创建外部服务器请参见[手动创建外部服务器](#)。

### 4.5.2.4 创建外表

当完成[创建外部服务器](#)后，在GaussDB(DWS)数据库中创建一个HDFS只写外表，用来访问存储在HDFS上的数据。此外表是只写的，只能用于导出操作。

创建外表的语法格式如下，详细的描述请参见[CREATE FOREIGN TABLE \(SQL on OBS or Hadoop\)](#)章节。

```
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 dfs_server  
OPTIONS ( { option_name ' value ' } [ , ... ] )  
[ {WRITE ONLY } ]  
DISTRIBUTE BY {ROUNDROBIN | REPLICATION}  
[ PARTITION BY ( column_name ) [ AUTOMAPPED ] ] ;
```

例如，创建一个名为"*product\_info\_ext\_obs*"的外表，对语法中的参数按如下描述进行设置：

- **table\_name**  
外表的表名。
- **表字段定义**
  - **column\_name**：外表中的字段名。
  - **type\_name**：字段的数据类型。多个字段用“,” 隔开。
- **SERVER dfs\_server**  
外表的外部服务器名称，这个server必须存在。外表通过设置外部服务器连接OBS/HDFS读取数据。  
此处应参考[创建外部服务器](#)中创建的外部服务器名称填写。
- **OPTIONS参数**  
用于指定外表数据的各类参数，关键参数如下所示。
  - “format”：表示导出的数据文件格式，支持“orc”格式。
  - “foldername”：外表中数据源文件目录，即表数据目录在HDFS文件系统上对应的文件目录。此选项对WRITE ONLY外表为必选项，对READ ONLY外表为可选项。
  - “encoding”：外表中数据源文件的编码格式名称，缺省为utf8。
  - 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的外表有效。  
取值范围：[1, 1024]的整数。

#### 📖 说明

filesize参数只对ORC格式的WRITE ONLY的HDFS外表有效。

- compression  
指定ORC格式文件的压缩方式，此选项为可选项。此语法仅对WRITE ONLY的外表有效。  
取值范围：zlib, snappy, lz4。缺省值为snappy。
- version  
指定ORC格式的版本号，此选项为可选项。此语法仅对WRITE ONLY的外表有效。  
取值范围：目前仅支持0.12。缺省值为0.12。
- 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外表有效。

取值范围：该数据库编码支持转换的数据编码。

### 说明

dataencoding参数只对ORC格式的WRITE ONLY的HDFS外表有效。

- **语法中的其他参数**

其他参数均为可选参数，用户可以根据自己的需求进行设置，在本例中不需要设置。详细的描述请参见CREATE FOREIGN TABLE (SQL on Hadoop or OBS)。

根据以上信息，创建外表命令如下所示：

```
DROP FOREIGN TABLE IF EXISTS product_info_ext_obs;
```

---建立不包含分区列的HDFS外表，表关联的外部服务器为hdfs\_server，表对应的HDFS服务上的文件格式为‘orc’，HDFS上的数据存储路径为‘/user/hive/warehouse/product\_info\_orc/’。

```
CREATE FOREIGN TABLE product_info_ext_obs
(
  product_price      integer      ,
  product_id         char(30)     ,
  product_time       date         ,
  product_level      char(10)     ,
  product_name       varchar(200) ,
  product_type1      varchar(20)  ,
  product_type2      char(10)     ,
  product_monthly_sales_cnt integer ,
  product_comment_time date       ,
  product_comment_num integer     ,
  product_comment_content varchar(200)
) SERVER obs_server
OPTIONS (
  format 'orc',
  foldername '/user/hive/warehouse/product_info_orc/',
  compression 'snappy',
  version '0.12'
) Write Only;
```

## 4.5.2.5 执行导出

### 操作步骤

**步骤1** 执行数据导出。

```
INSERT INTO [foreign table 表名] SELECT * FROM [源表名];
```

----结束

### 执行导出数据示例

- **示例1**：将表product\_info\_output的数据通过外表product\_info\_output\_ext导出到数据文件中。

```
INSERT INTO product_info_output_ext SELECT * FROM product_info_output;
```

若出现以下类似信息，说明数据导出成功。

```
INSERT 0 10
```

- **示例2**：通过条件过滤（WHERE product\_price>500），向数据文件中导出部分数据。

```
INSERT INTO product_info_output_ext SELECT * FROM product_info_output WHERE product_price>500;
```

### 说明

对于特殊的数据类型如RAW类型，在导出之后是一个二进制文本，导入工具无法识别。需使用 RAWTOHEX()函数将其转换为16进制文本导出。

## 4.5.3 使用 GDS 导出数据到远端服务器

### 4.5.3.1 关于 GDS 并行导出

使用GDS工具将数据从数据库导出到普通文件系统中，适用于高并发、大量数据导出的场景。

当前版本的GDS支持从数据库导出到管道文件，该功能使GDS的导出更加灵活多变。

- 当GDS用户的本地磁盘空间不足时：
  - 通过管道文件将从GDS导出的数据进行压缩减少磁盘空间。
  - 通过管道直接将导出出来的数据放到hdfs服务器上。
- 当用户导出前需要清洗数据时：
  - 用户可以根据自己的需求编写程序，将需要处理的流式数据实时从管道中读取内容，完成导出的数据清洗工作。

### 说明

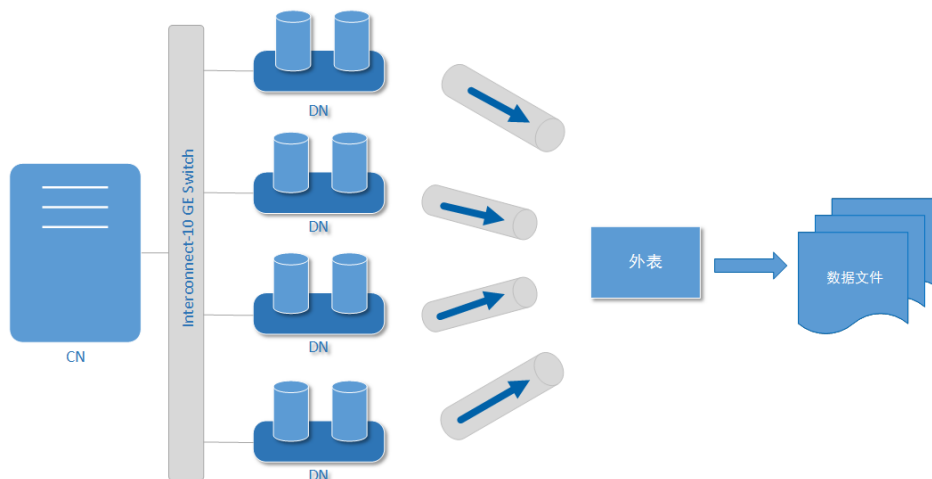
- 当前版本暂不支持SSL模式下GDS导出，请勿以SSL方式使用GDS。
- 本章涉及的所有管道文件都是指linux上的命名管道。

## 概述

**通过外表导出数据：**通过GDS外表设置的导出模式、导出数据格式等信息来指定待导出的数据文件，利用多DN并行的方式，将数据从数据库导出到数据文件中，从而提高整体导出性能。不支持直接导出文件到HDFS文件系统。

- CN只负责任务的规划及下发，把数据导出的工作交给了DN，释放了CN的资源，使其有能力处理外部请求。
- 通过让各个DN都参与数据导出，充分利用各个设备的计算能力及网络带宽。

图 4-11 通过外表导出数据



## 相关概念

- **数据文件**：存储有数据的TEXT、CSV或FIXED文件。文件中保存的是从GaussDB(DWS)数据库导出的数据。
- **外表**：用于规划导出数据文件的数据文件格式、存放位置、编码格式等信息。
- **GDS**：数据服务工具。在导出数据时，需要将此工具部署到数据文件所在的服务器上，使DN可以通过该工具导出数据。
- **表**：数据库中的表，包括行存表和列存表。数据文件中的数据从这些表中导出。
- **Remote导出模式**：将集群中的业务数据导出到集群之外的主机上。

## 导出模式

GaussDB(DWS)支持的导出模式有Remote模式。

- **Remote模式**：将集群中的业务数据导出到集群之外的主机上。
  - 支持多个GDS服务并发导出，但1个GDS在同一时刻，只能为1个集群提供导出服务。
  - 配置与集群节点处于统一内网的GDS服务，导出速率受网络带宽影响，推荐的网络配置为10GE。
  - 支持数据文件格式：TEXT、CSV。单行数据大小需<1GB。



## 导出流程

图 4-12 并行导出流程

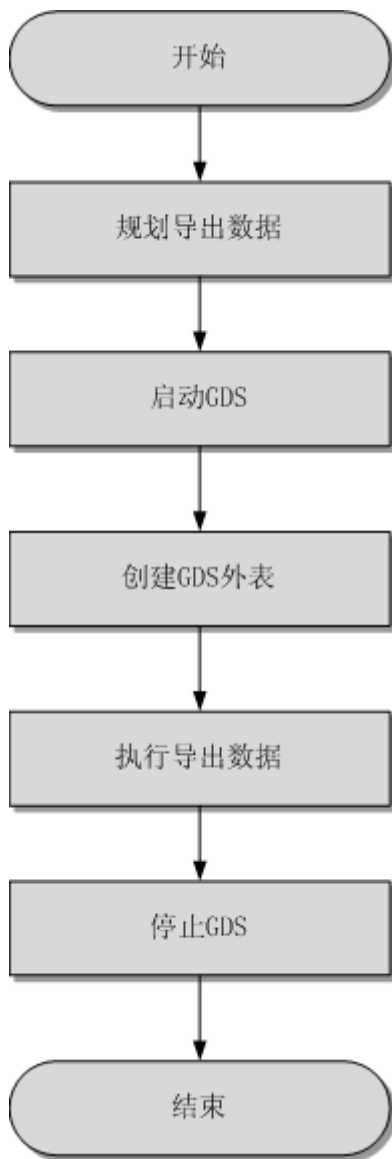


表 4-23 流程说明

流程	说明	子任务
规划导出数据。	根据所选模式，准备需要导出的数据并规划导出路径。 详细内容请参见 <a href="#">规划导出数据</a>	-
启动GDS。	若规划的导出模式为Remote模式，需在数据服务器上安装配置并启动GDS。 详细内容请参见 <a href="#">安装配置和启动GDS</a> 。	-

流程	说明	子任务
创建外表。	创建外表用于帮助GDS指定导出的数据文件。外表中保存了导出数据文件的位置、文件格式、编码格式、数据间的分隔符等信息。 详细内容请参见 <a href="#">创建GDS外表</a> 。	-
执行导出数据。	在创建好外表后，通过INSERT语句，将数据快速、高效地导出到数据文件中。 详细内容请参见 <a href="#">执行导出数据</a> 。	-
停止GDS。	数据导出完成后，停止GDS。 详细请参见 <a href="#">停止GDS</a> 。	-

### 4.5.3.2 规划导出数据

#### 操作场景

使用GDS从集群导出到数据之前，要提前准备需要导出的数据，并规划导出的路径。

#### 规划导出路径

- **Remote模式**

**步骤1** 以root用户登录GDS数据服务器，创建导出的数据文件存放目录“/output\_data”。

```
mkdir -p /output_data
```

**步骤2** （可选）创建用户及所属的用户组。此用户为启动GDS的用户，该用户需要拥有导出数据文件存放目录的写权限。

```
groupadd gdsgrp  
useradd -g gdsgrp gdsuser
```

若出现以下提示，说明数据库用户及所属用户组已存在，可跳过本步骤。

```
useradd: Account 'gdsuser' already exists.  
groupadd: Group 'gdsgrp' already exists.
```

**步骤3** 修改数据文件目录属主为gdsuser。

```
chown -R gdsuser:gdsgrp /output_data
```

----结束

### 4.5.3.3 安装配置和启动 GDS

GDS是GaussDB(DWS)提供的数据服务工具，通过和外表机制的配合，实现数据的高速导出。

详细内容请参见[安装配置和启动GDS](#)。

### 4.5.3.4 创建 GDS 外表

#### 操作步骤

**步骤1** 根据[规划导出数据](#)中规划的路径确定外表参数`location`的值。

- **Remote模式**

请通过URL方式设置参数“`location`”，用于指定导出的数据文件存放路径。

- 不需要指定文件名。
- 当有多个路径时，只有第一个路径有效。

**示例：**

GDS数据服务器IP为192.168.0.90，假定启动GDS时设置的监听端口为5000，设置的导出后文件存放目录为“`/output_data/`”。

根据以上情况，在创建外表时，指定参数“`location`”为“`gsfs://192.168.0.90:5000/`”。

**说明**

- `location`可以指定子目录如“`gsfs://192.168.0.90:5000/2019/11/`”实现同一张表根据日期导出到不同目录下。
- 现有版本在执行导出任务的时候会判断“`/output_data/2019/11/`”目录是否存在，不存在则创建。导出时会将文件写入此目录下，这样用户在创建或修改外表后就不需要再去手动执行“`mkdir -p /output_data/2019/11/`”。

**步骤2** 梳理待导出数据格式信息，确定创建外表时使用的数据格式参数的值。格式参数详细介绍，请参见[数据格式参数](#)。

**步骤3** 根据前面步骤确定的参数，**创建GDS外表**。外表的创建语法以及详细使用，请参考[CREATE FOREIGN TABLE \(GDS导入导出\)](#)。

----**结束**

#### 示例

- **示例：**创建GDS导出外表`foreign_tpcds_reasons`，待导出数据格式为CSV，用于接收数据服务器上的数据。

其中设置的**导出模式**信息如下所示：

规划数据服务器与集群处于同一内网，数据服务器IP为192.168.0.90，待导出的数据文件格式为CSV，选择并行导出模式为Remote模式。

假定启动GDS时，规划导出的数据文件存放目录为“`/output_data/`”，GDS监听端口为5000，所以设置参数“`location`”为“`gsfs://192.168.0.90:5000/`”。

设置导出的**数据格式信息**，参数设置如下所示：

- 导出数据文件格式（`format`）为CSV。
- 编码格式（`encoding`）为UTF-8。
- 字段分隔符（`delimiter`）为E'\x08'。
- 引号字符（`quote`）为E'\x1b'。
- 数据文件中空值（`null`）为没有引号的空字符串。
- 逃逸字符（`escape`）默认和`quote`相同。
- 数据文件是否包含标题行（`header`）为默认值false，即导出时数据文件第一行被识别为数据。

- 导出数据文件换行符样式（EOL）为0x0A。

创建的外表如下所示：

```
CREATE FOREIGN TABLE foreign_tpcds_reasons
(
  r_reason_sk integer not null,
  r_reason_id char(16) not null,
  r_reason_desc char(100)
)
SERVER gsmpp_server
OPTIONS (LOCATION 'gsfs://192.168.0.90:5000/',
        FORMAT 'CSV',
        DELIMITER E'\x08',
        QUOTE E'\x1b',
        NULL '',
        EOL '0x0a'
)
WRITE ONLY;
```

### 4.5.3.5 执行导出数据

#### 前提条件

需要确保每一个CN和DN所在服务器到GDS服务器的IP和端口是互通的。

#### 操作步骤

##### 步骤1 执行数据导出。

```
INSERT INTO [foreign table 表名] SELECT * FROM [源表名];
```

#### 📖 说明

编写批处理任务脚本，实现并发批量导出数据。并发量视机器资源使用情况而定。可通过几个表测试，监控资源利用率，根据结果提高或减少并发量。常用资源监控命令有：内存和CPU监控top命令，IO监控命令iostat，网络监控命令sar等。相关案例请参见[多线程导出](#)。

----结束

#### 任务示例

- **示例1：**将表reason的数据通过外表foreign\_tpcds\_reasons导出到数据文件中。  

```
INSERT INTO foreign_tpcds_reasons SELECT * FROM tpcds.reason;
```
- **示例2：**通过条件过滤（r\_reason\_sk = 1），向数据文件中导出部分数据。  

```
INSERT INTO foreign_tpcds_reasons SELECT * FROM tpcds.reason WHERE r_reason_sk=1;
```
- **示例3：**对于特殊的数据类型如RAW类型，在导出之后是一个二进制文本，导入工具无法识别。需使用RAWTOHEX()函数将其转换为16进制文本导出。  

```
INSERT INTO foreign_tpcds_reasons SELECT RAWTOHEX(c) FROM tpcds.reason;
```

### 4.5.3.6 停止 GDS

GDS是GaussDB(DWS)提供的数据服务工具，通过和外表机制的配合，实现数据的高速导出。

详细内容请参见[停止GDS](#)。

### 4.5.3.7 GDS 导出示例

#### Remote 模式导出

规划数据服务器与集群处于同一内网，数据服务器IP为192.168.0.90，导出数据文件格式为CSV，所以规划的并行导出模式为Remote模式。

Remote模式并行导出数据操作示例如下所示：

1. 以root用户登录GDS数据服务器，创建数据文件存放目录“/output\_data”，启动gds\_user用户及所属的用户组。

```
mkdir -p /output_data
```

2. （可选）创建用户及其所属的用户组。此用户用于启动GDS。若该类用户及所属用户组已存在，可跳过此步骤。

```
groupadd gdsgrp  
useradd -g gdsgrp gds_user
```

3. 修改数据服务器上数据文件目录“/output\_data”的属主为gds\_user。

```
chown -R gds_user:gdsgrp /output_data
```

4. 以gds\_user用户登录数据服务器上分别启动GDS。

其中GDS安装路径为“/opt/bin/dws/gds”，导出数据文件存放在“/output\_data/”目录下，数据服务器所在IP为192.168.0.90，GDS监听端口为5000，以后台方式运行。

```
/opt/bin/dws/gds/bin/gds -d /output_data -p 192.168.0.90:5000 -H 10.10.0.1/24 -D
```

5. 在数据库中创建外表foreign\_tpcds\_reasons用于接收数据服务器上的数据。

其中设置的**导出模式信息**如下所示：

- 由于启动GDS时，设置的导出数据文件存放目录为“/output\_data/”，GDS监听端口为5000。创建的导出数据文件存放目录为“/output\_data/”。所以设置参数“location”为“gsfs://192.168.0.90:5000/”。

设置导出的**数据文件格式**信息如下所示：

- 数据文件格式（format）为CSV。
- 编码格式（encoding）为UTF-8。
- 字段分隔符（delimiter）为E'\x08'。
- 引号字符（quote）为E'\x1b'。
- 数据文件中空值（null）为没有引号的空字符串。
- 逃逸字符（escape）默认和quote相同。
- 数据文件是否包含标题行（header）为默认值false，即导出时数据文件第一行被识别为数据。

根据以上信息，创建的外表如下所示：

```
CREATE FOREIGN TABLE foreign_tpcds_reasons  
(  
  r_reason_sk integer not null,  
  r_reason_id char(16) not null,  
  r_reason_desc char(100)  
) SERVER gsmpp_server OPTIONS (LOCATION 'gsfs://192.168.0.90:5000/', FORMAT 'CSV',ENCODING  
'utf8',DELIMITER E'\x08', QUOTE E'\x1b', NULL '') WRITE ONLY;
```

6. 在数据库上，通过外表foreign\_tpcds\_reasons，将数据导出到数据文件中。

```
INSERT INTO foreign_tpcds_reasons SELECT * FROM tpcds.reason;
```

7. 待数据导出完成后，以gds\_user用户登录数据服务器，停止GDS。

其中GDS进程号为128954。

```
ps -ef|grep gds  
gds_user 128954 1 0 15:03 ? 00:00:00 gds -d /output_data -p 192.168.0.90:5000 -D
```

```
gds_user 129003 118723 0 15:04 pts/0 00:00:00 grep gds  
kill -9 128954
```

## 多线程导出

规划数据服务器与集群处于同一内网，数据服务器IP为192.168.0.90，导出的数据文件格式为CSV，同时导出2个目标表，所以规划使用Remote模式进行多线程导出。

Remote模式多线程导出数据操作示例如下所示：

1. 以root用户登录GDS数据服务器，创建导出数据文件存放目录“/output\_data”，数据库用户及所属的用户组。  

```
mkdir -p /output_data  
groupadd gdsgrp  
useradd -g gdsgrp gds_user
```
2. 修改数据服务器上数据文件目录“/output\_data”的属主为gds\_user。  

```
chown -R gds_user:gdsgrp /output_data
```
3. 以gds\_user用户登录数据服务器上启动GDS。  
其中GDS安装路径为“/opt/bin/dws/gds”，导出数据文件存放在“/output\_data/”目录下，数据服务器所在IP为192.168.0.90，GDS监听端口为5000，以后台方式运行，设定并发度为2。  

```
/opt/bin/dws/gds/bin/gds -d /output_data -p 192.168.0.90:5000 -H 10.10.0.1/24 -D -t 2
```
4. 在GaussDB(DWS)上，创建外表foreign\_tpcds\_reasons1和foreign\_tpcds\_reasons2用于接收数据服务器上的数据。
  - 其中设置的**导出模式信息**如下所示：
    - 由于启动GDS时，设置的导出数据文件存放目录为“/output\_data/”，GDS监听端口为5000。创建的导出数据文件存放目录为“/output\_data/”。所以设置参数“location”为“gsfs://192.168.0.90:5000/”。
  - 设置导出的**数据文件格式信息**如下所示：
    - 数据文件格式（format）为CSV。
    - 编码格式（encoding）为UTF-8。
    - 字段分隔符（delimiter）为E'\x08'。
    - 引号字符（quote）为E'\x1b'。
    - 数据文件中空值（null）为没有引号的空字符串。
    - 逃逸字符（escape）默认和quote相同。
    - 数据文件是否包含标题行（header）为默认值false，即导出时数据文件第一行被识别为数据。

根据以上信息，创建的外表foreign\_tpcds\_reasons1如下所示：

```
CREATE FOREIGN TABLE foreign_tpcds_reasons1  
(  
  r_reason_sk integer not null,  
  r_reason_id char(16) not null,  
  r_reason_desc char(100)  
) SERVER gsmpp_server OPTIONS (LOCATION 'gsfs://192.168.0.90:5000/', FORMAT 'CSV', ENCODING 'utf8', DELIMITER E'\x08', QUOTE E'\x1b', NULL '') WRITE ONLY;
```

参考以上设置，创建的外表foreign\_tpcds\_reasons2如下所示：

```
CREATE FOREIGN TABLE foreign_tpcds_reasons2
(
  r_reason_sk integer not null,
  r_reason_id char(16) not null,
  r_reason_desc char(100)
) SERVER gsmpp_server OPTIONS (LOCATION 'gsfs://192.168.0.90:5000/', FORMAT 'CSV', DELIMITER
E'\x08', QUOTE E'\x1b', NULL '') WRITE ONLY;
```

5. 在数据库中通过外表foreign\_tpcds\_reasons1和foreign\_tpcds\_reasons2，将表reasons1和reasons2中的数据导出到目录“/output\_data”中。

```
INSERT INTO foreign_tpcds_reasons1 SELECT * FROM tpcds.reason;
INSERT INTO foreign_tpcds_reasons2 SELECT * FROM tpcds.reason;
```

6. 待数据导出完成后，以gds\_user用户登录数据服务器，停止GDS。

其中GDS进程号为128954。

```
ps -efl | grep gds
gds_user 128954 1 0 15:03 ? 00:00:00 gds -d /output_data -p 192.168.0.90:5000 -D -t 2
gds_user 129003 118723 0 15:04 pts/0 00:00:00 grep gds
kill -9 128954
```

## 单个管道文件导出

### 步骤1 启动GDS。

```
gds -d /gds_data/ -D -p 192.168.0.1:7789 -l /gds_log/aa.log -H 0/0 -t 10 -D
```

如果需要设置管道文件的超时时间，则使用--pipe-timeout参数设置。

### 步骤2 执行数据导出。

1. 登录数据库创建内表，并写入数据。

```
CREATE TABLE test_pipe( id integer not null, sex text not null, name text ) ;
```

```
INSERT INTO test_pipe values(1,2,'1111111111111111');
INSERT INTO test_pipe values(2,2,'1111111111111111');
INSERT INTO test_pipe values(3,2,'1111111111111111');
INSERT INTO test_pipe values(4,2,'1111111111111111');
INSERT 0 1
```

2. 创建只写外表。

```
CREATE FOREIGN TABLE foreign_test_pipe_tw( id integer not null, age text not null, name text )
SERVER gsmpp_server OPTIONS (LOCATION 'gsfs://192.168.0.1:7789/', FORMAT 'text', DELIMITER ',',
NULL '', EOL '\x0a', file_type 'pipe', auto_create_pipe 'false') WRITE ONLY;
```

3. 导出语句，此时语句会阻塞。

```
INSERT INTO foreign_test_pipe_tw select * from test_pipe;
```

### 步骤3 通过管道文件将数据从GDS导出。

1. 登录GDS，进入GDS数据目录。

```
cd /gds_data/
```

2. 创建管道文件，如果auto\_create\_pipe设置为true跳过此步骤。

```
mkfifo postgres_public_foreign_test_pipe_tw.pipe
```

#### 说明

管道文件创建完成后，每执行完一次操作，业务会被自动清理。如果还需要执行其他业务，请参考该步骤重新创建管道文件。

3. 读取管道文件并写入新文件。

```
cat postgres_public_foreign_test_pipe_tw.pipe > postgres_public_foreign_test_pipe_tw.txt
```

4. 若需要对导出的文件进行压缩执行：

```
gzip -9 -c < postgres_public_foreign_test_pipe_tw.pipe > out.gz
```

5. 若需要将管道文件的内容导出到hdfs服务器：

```
cat postgres_public_foreign_test_pipe_tw.pipe | hdfs dfs -put - /user/hive/gds/test_pipe.txt
```

**步骤4 验证导出的数据。**

1. 查看文件是否导出正确。  

```
cat postgres_public_foreign_test_pipe_tw.txt
3,2,1111111111111111
1,2,1111111111111111
2,2,1111111111111111
4,2,1111111111111111
```
2. 查看压缩后的文件。  

```
vim out.gz
3,2,1111111111111111
1,2,1111111111111111
2,2,1111111111111111
4,2,1111111111111111
```
3. 查看导出到hdfs服务器上的数据。  

```
hdfs dfs -cat /user/hive/***/test_pipe.txt
3,2,1111111111111111
1,2,1111111111111111
2,2,1111111111111111
4,2,1111111111111111
```

----结束

## 多进程管道文件导出

GDS也支持多进程管道文件导入导出，即启动一个外表对应多个GDS。

以本地文件的导出为例：

**步骤1 启动多个GDS。**

```
gds -d /***/gds_data/ -D -p 192.168.0.1:7789 -l /***/gds_log/aa.log -H 0/0 -t 10 -D
gds -d /***/gds_data_1/ -D -p 192.168.0.1:7790 -l /***/gds_log/aa.log -H 0/0 -t 10 -D
```

如果需要设置管道文件的超时时间，则使用--pipe-timeout参数设置。

**步骤2 执行数据导出。**

1. 登录数据库创建内表。  

```
CREATE TABLE test_pipe (id integer not null, sex text not null, name text);
```
2. 写入数据。  

```
INSERT INTO test_pipe values(1,2,'1111111111111111');
INSERT INTO test_pipe values(2,2,'1111111111111111');
INSERT INTO test_pipe values(3,2,'1111111111111111');
INSERT INTO test_pipe values(4,2,'1111111111111111');
```
3. 创建只写外表。  

```
CREATE FOREIGN TABLE foreign_test_pipe_tw( id integer not null, age text not null, name text )
SERVER gsmpp_server OPTIONS (LOCATION 'gsfs://192.168.0.1:7789/gsfss://192.168.0.1:7790/',
FORMAT 'text', DELIMITER ',', NULL '', EOL '0x0a',file_type 'pipe', auto_create_pipe 'false') WRITE
ONLY;
```
4. 导出语句，此时语句会阻塞。  

```
INSERT INTO foreign_test_pipe_tw select * from test_pipe;
```

**步骤3 通过管道文件将数据从GDS导出。**

1. 登录GDS，分别进入GDS数据目录。  

```
cd /***/gds_data/
cd /***/gds_data_1/
```
2. 创建管道文件，如果auto\_create\_pipe设置为true跳过此步骤。  

```
mkfifo postgres_public_foreign_test_pipe_tw.pipe
```
3. 分别读取管道文件并写入新文件。  

```
cat postgres_public_foreign_test_pipe_tw.pipe > postgres_public_foreign_test_pipe_tw.txt
```



**步骤4 验证导出的数据。**

```
cat /***/gds_data/postgres_public_foreign_test_pipe_tw.txt
3,2,1111111111111111
cat /***/gds_data_1/postgres_public_foreign_test_pipe_tw.txt
1,2,1111111111111111
2,2,1111111111111111
4,2,1111111111111111
```

----结束

## 4.6 其他操作

### 4.6.1 GDS 管道文件常见问题

#### 注意事项

- GDS支持并发导入导出，gds -t参数用于设置gds的工作线程池大小，控制并发场景下同时工作的工作线程数且不会加速单个sql任务。gds -t缺省值为8，上限值为200。在使用管道功能进行导入导出时，-t参数应不低于业务并发数。如果是双集群互联互通场景，-t参数应不低于业务并发数的两倍。
- 由于管道“读取即删除”的特点，需确保导入或导出过程中除GDS程序外无其他程序读取管道文件，避免导入过程中数据丢失或者任务报错及导出的文件内容混乱。
- 不支持对具有相同location的外表并发导入导出，即GDS的多个线程同时读取管道文件或者同时写入管道文件。
- GDS的单个导入导出任务只识别一个管道文件，因此不要对GDS外表设置带有通配符({}[]?)的location地址。如：

```
CREATE FOREIGN TABLE foreign_test_pipe_tr( like test_pipe ) SERVER gsmpp_server OPTIONS
(LLOCATION 'gsfs://192.168.0.1:7789/foreign_test_*', FORMAT 'text', DELIMITER ',', NULL '', EOL
'0x0a',file_type 'pipe',auto_create_pipe 'false');
```
- GDS启用-r递归参数时只识别一个管道文件，即GDS只会识别当前数据目录下的一个管道文件而不会递归寻找，因此-r参数在管道文件导入导出场景下不生效。
- 管道文件的导入导出不支持CN Retry，因为GDS无法控制对端用户和程序操作管道的行为。
- GDS导入时默认对端程序超过1小时未向管道中写入数据导入任务将会超时报错。
- GDS导出时默认对端程序超过1小时未从管道中读数据导出任务将会超时报错。
- 确保GDS版本和内核版本都已经支持管道文件导入导出功能。
- 当外表参数auto\_create\_pipe设置为true时，GDS自动创建管道文件可能存在延迟，因此操作管道文件时建议先判断自动创建的管道文件是否存在，且是否为管道文件类型。
- GDS管道文件的导入导出任务结束后会自动删除管道文件，但是手动终止任务时，管道文件的删除会有延迟，直到到达超时时间后才会被删除。

#### 常见问题和定位方法:

- 问题1: "/\*\*\*/postgres\_public\_foreign\_test\_pipe\_tr.pipe" must be named pipe.  
定位方法: GDS的外表file\_type类型为pipe但是操作的文件却是一个普通文件类型。应该排查postgres\_public\_foreign\_test\_pipe\_tr.pipe是否是为管道文件。

- 问题2: could not open pipe "/\*\*\*/postgres\_public\_foreign\_test\_pipe\_tw.pipe" cause by Permission denied.  
定位方法: GDS没有权限打开管道文件。
- 问题3: could not open source file /\*\*\*\*\*/postgres\_public\_foreign\_test\_pipe\_tw.pipe because timeout 300s for WRITING.  
定位方法: GDS导出时打开管道文件超时, 一般由于auto\_create\_pipe为false时候, 管道文件在300秒内未被创建, 或者创建了但是300秒内没有程序读取该管道文件。
- 问题4: could not open source file /\*\*\*\*\*/postgres\_public\_foreign\_test\_pipe\_tw.pipe because timeout 300s for READING.  
定位方法: GDS导出时打开管道文件超时, 一般由于auto\_create\_pipe为false时候, 管道文件在300秒内未被创建或者创建了但是300秒之内没有程序写入该管道文件。
- 问题5: could not poll writing source pipe file "/\*\*\*\*\*/postgres\_public\_foreign\_test\_pipe\_tw.pipe" timeout 300s.  
定位方法: GDS导出时超过300秒未等到管道上的写事件, 一般由于该管道文件超过300秒没有被读取。
- 问题6: could not poll reading source pipe file "/\*\*\*\*\*/postgres\_public\_foreign\_test\_pipe\_tw.pipe" timeout 300s.  
定位方法: GDS导入时超过300秒未等到管道上的读事件, 一般由于该管道文件超过300秒没有被写入。
- 问题7: could not open pipe file "/\*\*\*/postgres\_public\_foreign\_test\_pipe\_tw.pipe" for "WRITING" with error No such device or address.  
定位方法: 表示当前"/\*\*\*/postgres\_public\_foreign\_test\_pipe\_tw.pipe"管道文件没有程序正在读取导致GDS无法以写的方式打开管道文件。

## 4.6.2 查看数据倾斜状态

### 操作场景

数据倾斜会造成查询表性能下降。对于记录数超过千万条的表, 建议在执行全量数据导入前, 先导入部分数据, 以进行数据倾斜检查和调整分布列, 避免导入大量数据后发现数据倾斜, 调整成本高。

### 背景信息

GaussDB(DWS)是采用Shared-nothing架构的MPP ( Massive Parallel Processor, 大规模并发处理) 系统, 采用水平分布的方式, 将业务数据表的元组按合适的分布策略分散存储在所有的DN。

当前产品支持复制 ( Replication )、散列 ( Hash ) 和轮询 ( Roundrobin ) 三种用户表分布策略。

- Replication方式: 在每一个DN上存储一份全量表数据。对于数据量比较小的表建议采取Replication分布策略。
- Hash方式: 采用这种分布方式, 需要为用户表指定一个分布列 ( distribute key )。当插入一条记录时, 系统会根据分布列的值进行hash运算后, 将数据存储在对应的DN中。对于数据量比较大的表建议采取Hash分布策略。

- Roundrobin方式：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。对于数据量比较大的表，如果Hash分布找不到一个合适的分布列，建议采用Roundrobin分布策略。

对于Hash分布策略，如果分布列选择不当，可能导致数据倾斜。因此在采用Hash分布策略之后会对用户表的数据进行数据倾斜性检查，以确保数据在各个DN上是均匀分布的。一般情况下分布列都是选择键值重复度小，数据分布比较均匀的列。

## 操作步骤

**步骤1** 分析数据源特征，选择若干个键值重复度小，数据分布比较均匀的备选分布列。

**步骤2** 从**步骤1**中选择一个备选分布列创建目标表。

```
CREATE [ [ GLOBAL | LOCAL ] { TEMPORARY | TEMP } | UNLOGGED ] TABLE [ IF NOT EXISTS ] table_name
({ column_name data_type [ compress_mode ] [ COLLATE collation ] [ column_constraint [ ... ] ]
| table_constraint | LIKE source_table [ like_option [ ... ] ] }
[, ... ] ) [ WITH ( {storage_parameter = value} [, ... ] ) ]
[ ON COMMIT { PRESERVE ROWS | DELETE ROWS | DROP } ]
[ COMPRESS | NOCOMPRESS ] [ TABLESPACE tablespace_name ]
[ DISTRIBUTE BY { REPLICATION
| ROUNDROBIN
| { HASH ( column_name [,...] ) } } ];
```

**步骤3** 参照前面章节中的办法向目标表中导入小批量数据。

对于单个数据源文件，在导入时，可通过均匀切割，导入部分切割后的数据源文件来验证数据倾斜性。

**步骤4** 检验数据倾斜性。命令中的table\_name，请填入实际的目标表名。

```
SELECT a.count,b.node_name FROM (SELECT count(*) AS count,xc_node_id FROM table_name GROUP
BY xc_node_id) a, pgxc_node b WHERE a.xc_node_id=b.node_id ORDER BY a.count desc;
```

**步骤5** 若各DN上数据分布差小于10%，表明数据分布均衡，选择的分布列合适。请清理已导入小批量数据，导入全量数据，以完成数据迁移。

若各DN上数据分布差大于等于10%，表明数据分布倾斜，请从**步骤1**的备选分布列中删除该列，删除目标表，并重复**步骤2**、**步骤3**、**步骤4**和**步骤5**。

### 说明

此处的数据分布差表示实际查询到DN上的数据量与DN平均数据量的差异。

**步骤6**（可选）如果上述步骤不能选出适合的分布列，需要从备选分布列选择多个列的组合作为分布列来完成数据迁移。

----结束

## 示例

对目标表staffs选择合适的分布列。

1. 分析表staffs的数据源特征，选择数据重复度低且分布均匀的备选分布列staff\_ID、FIRST\_NAME和LAST\_NAME。
2. 先选择staff\_ID作为分布列，创建目标表staffs。

```
CREATE TABLE staffs
(
  staff_ID    NUMBER(6) not null,
  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)
)
DISTRIBUTE BY hash(staff_ID);
```

3. 向目标表staffs中导入部分数据。

根据以下查询所得，集群环境中主DN数为8个，则建议导入的记录数为80000条。

```
SELECT count(*) FROM pgxc_node where node_type='D';
count
-----
      8
(1 row)
```

4. 校验以staff\_ID为分布列的目标表staffs的数据倾斜性。

```
SELECT a.count,b.node_name FROM (select count(*) as count,xc_node_id FROM staffs GROUP BY
xc_node_id) a, pgxc_node b WHERE a.xc_node_id=b.node_id ORDER BY a.count desc;
count | node_name
-----+-----
11010 | datanode4
10000 | datanode3
12001 | datanode2
 8995 | datanode1
10000 | datanode5
 7999 | datanode6
 9995 | datanode7
10000 | datanode8
(8 rows)
```

5. 根据上一步骤查询所得，各DN上数据分布差大于10%，数据分布倾斜。所以从步骤1的备选分布列中删除该列，并删除目标表staffs。

```
DROP TABLE staffs;
```

6. 尝试选择staff\_ID、FIRST\_NAME和LAST\_NAME的组合为分布列，创建目标表staffs。

```
CREATE TABLE staffs
(
  staff_ID NUMBER(6) not null,
  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)
)
DISTRIBUTE BY hash(staff_ID,FIRST_NAME,LAST_NAME);
```

7. 校验以staff\_ID、FIRST\_NAME和LAST\_NAME的组合为分布列的目标表staffs的数据倾斜性。

```
SELECT a.count,b.node_name FROM (select count(*) as count,xc_node_id FROM staffs GROUP BY
xc_node_id) a, pgxc_node b WHERE a.xc_node_id=b.node_id ORDER BY a.count desc;
count | node_name
-----+-----
10010 | datanode4
10000 | datanode3
10001 | datanode2
 9995 | datanode1
10000 | datanode5
 9999 | datanode6
 9995 | datanode7
```

10000 | datanode8  
(8 rows)

8. 根据上一步骤查询所得，各DN上数据分布差小于10%，数据分布均衡，选择的分布列合适。
9. 清理已导入小批量数据。  
`TRUNCATE TABLE staffs;`
10. 导入全量数据，以完成数据迁移。

### 4.6.3 分析表

执行计划生成器需要使用表的统计信息，以生成最有效的查询执行计划，提高查询性能。因此数据导入完成后，建议执行ANALYZE语句生成最新的表统计信息。统计结果存储在系统表PG\_STATISTIC中。

#### 分析表

ANALYZE支持的表类型有行/列存表、HDFS表、ORC/CARBONDATA格式的OBS外表。ANALYZE同时也支持对本地表的指定列进行信息统计。下面以表的ANALYZE为例，更多关于ANALYZE的信息，请参见ANALYZE | ANALYSE。

##### 步骤1 更新表统计信息。

以表product\_info为例，ANALYZE命令如下：

```
ANALYZE product_info;
```

----结束

#### 表自动分析

GaussDB(DWS)提供了三种场景下表的自动分析。

- 当查询中存在“统计信息完全缺失”或“修改量达到analyze阈值”的表，且执行计划不采取FQS (Fast Query Shipping)执行时，则通过GUC参数**autoanalyze**控制此场景下表统计信息的自动收集。此时，查询语句会等待统计信息收集成功后，生成更优的执行计划，再执行原查询语句。
- 当autovacuum设置为on时，系统会定时启动autovacuum线程，对“修改量达到analyze阈值”的表在后台自动进行统计信息收集。

表 4-24 表自动分析

触发方式	触发条件	触发频率	控制参数	备注
同步	统计信息完全缺失	查询时	autoanalyze, autoanalyze_mode	truncate主表时会清空统计信息。
同步	数据修改量达到analyze阈值	查询时	autoanalyze, autoanalyze_mode	先触发analyze, 后选择最优计划。
异步	数据修改量达到analyze阈值	autovacuum线程轮询检查	autovacuum_mode, autovacuum_naptime	2s等锁超时, 5min执行超时。

**须知**

- autoanalyze只支持内存方式计算统计信息，不支持临时采样表方式计算统计信息。
- 多列统计信息仅支持临时采样表方式计算统计信息，因此autoanalyze不收集多列统计信息。
- 查询过程因表的“统计信息完全缺失”和“修改量达到analyze阈值”而自动触发autoanalyze的场景，当前不支持对外表触发autoanalyze，不支持对带有ON COMMIT [DELETE ROWS | DROP]选项的临时表触发autoanalyze。
- 修改量达到analyze阈值是指：表的修改量超过autovacuum\_analyze\_threshold + autovacuum\_analyze\_scale\_factor \* reltuples，其中reltuples是pg\_class中记录的表的估算行数。
- 基于定时启动的autovacuum线程触发的autoanalyze，仅支持行存表和列存表，不支持外表、HDFS表、OBS外表、临时表、unlogged表和toast表。
- 查询时触发analyze会对分区表的所有分区加四级锁，直到查询所在事务提交后才会放锁。四级锁不堵塞增删改查，但会堵塞分区的修改操作，比如分区的truncate，可以通过将object\_mtime\_record\_mode设置为disable\_partition，实现提前释放分区锁。
- autovacuum自动清理功能的生效还依赖于下面两个GUC参数：
  - **track\_counts**参数需要设置为on，开启收集数据库统计数据功能。
  - autovacuum\_max\_workers参数需要大于0，该参数表示能同时运行的自动清理线程的最大数量。

GaussDB(DWS)支持轻量化的autoanalyze，可通过[autoanalyze\\_mode](#)参数来设置。

表 4-25 两种 autoanalyze 的对比

类型	统计信息存储位置	统计信息是否持久化	加锁级别	是否向其它CN同步	是否影响autovacuum触发的analyze
普通 autoanalyze	系统表	是	四级锁	是	完成后，autovacuum不再触发analyze
轻量化 autoanalyze	内存	否	一级锁	否	完成后，autovacuum依然可以触发analyze

普通autoanalyze统计信息要存系统表，因此需要加四级锁，防止系统表的并发更新。触发普通autoanalyze的查询会从一级锁升四级锁，堵塞其它四级锁以上的操作。

轻量化autoanalyze对普通autoanalyze做了极简的设计：

- 统计信息仅写入内存，不写入系统表。
- 执行时只加一级锁。
- 统计信息不向其它CN同步。

查询触发的轻量化autoanalyze可以快速生成统计信息，并在一个CN内供多个查询共享使用。需要开启autovacuum后台触发的analyze，由其重新生成统计信息，实现统计信息的持久化和同步，并清理内存中的统计信息。

# 5 冷热数据管理

## 冷热数据简介

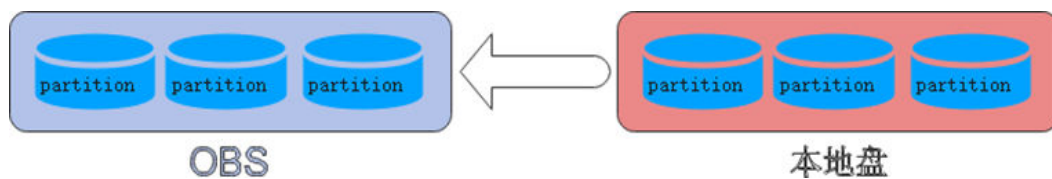
海量大数据场景下，随着业务和数据量的不断增长，数据存储与消耗的资源也日益增长。根据业务系统中用户对不同时期数据的不同使用需求，对膨胀的数据本身进行“冷热”分级管理，不仅可以提高数据分析性能还能降低业务成本。

例如，在网络流量分析系统中，用户可能对最近一个月内安全事件和网络访问情况感兴趣，而很少关注几个月前的数据。针对这样的一些场景，可以将数据按照时间分为：热数据、冷数据。

冷热数据主要从数据访问频率、更新频率进行划分。

- Hot（热数据）：访问、更新频率较高，未来被调用的概率较高的数据，对访问的响应时间要求很高的数据。
- Cold（冷数据）：不允许更新或更新频率比较低，访问频率比较低，对访问的响应时间要求不高的数据。

用户可以定义冷热管理表，将符合规则的冷数据切换至OBS上进行存储，可以按照分区自动进行冷热数据的判断和迁移。



## 冷热数据迁移

GaussDB(DWS)列存数据写入时，数据首先进入热分区进行存储，分区数据较多后，可通过手动或自动的方式，将符合冷数据规则的数据切换至OBS上进行存储。在数据切换至OBS上后，其元数据、Desc表信息以及索引信息仍在本地进行存储，保证了读取的性能。

## 冷热切换策略

目前冷热切换的策略名称支持LMT（last modify time）和HPN（hot partition number），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：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。

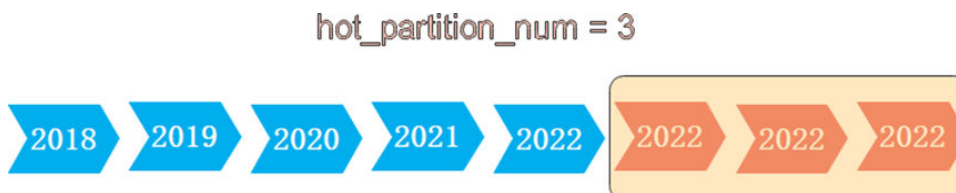


如下图中，设置day为2，即在冷热切换时，根据分区数据的最晚修改时间，保留2日内所修改的分区为热分区，其余数据为冷分区数据。假设当前时间为4月30日，4月30日对[4-26]分区进行了delete操作，4月29日对[4-27]分区进行了insert操作，故在冷热切换时，保留[4-26][4-27][4-29][4-30]四个分区为热分区。



- HPN：表示保留HPN个有数据的分区为热分区。分区顺序按照分区的Sequence ID来确定，分区的Sequence ID是根据分区边界值的大小，内置生成的序号，此序号不对外呈现。对于RANGE分区，分区的边界值越大，分区对应的Sequence ID越大；对于LIST分区，分区边界枚举值中的最大值越大，分区对应的Sequence ID越大。在冷热切换时，需要将数据迁移至OBS表空间中。其中HPN为整型，范围为[0,1600]。其中HPN为0时，表示不保留热分区，在进行冷热切换时，将所有有数据的分区都转为冷分区并存储在OBS上。

如下图中，设置HPN为3，即在冷热切换时，保留最新的3个有数据的分区为热分区数据，其余分区均切为冷分区。



## 注意事项

1. 冷热数据管理支持功能：
  - 支持对冷热表的insert、copy、delete、update、select等表相关的DML操作。
  - 支持对冷热表的权限管理等DCL操作。
  - 支持对冷热表进行analyze、vacuum、merge into等操作和一些分区的操作。
  - 支持从普通列存分区表升级为冷热数据表。
  - 支持带有冷热数据管理表的升级、扩容、缩容和重分布。
2. 冷热数据管理的使用约束限制：
  - 目前冷热表只支持列存2.0版本的分区表，外表不支持冷热分区。
  - 仅支持从热数据切换为冷数据，不支持从冷数据切换为热数据。对于已经切冷分区再次插入数据，数据直接会进入OBS，不会改变分区的冷热属性。
  - 对于同一分区在同一DN只会存在冷或热的一种情况，对于同一分区在不同DN可能存在部分DN为热数据，部分DN为冷数据。
  - 对于同时存在冷热分区的表，查询时会变慢，因为冷数据存储在OBS上，读写速度和时延都比在本地查询要慢。
  - 只支持修改冷热表的冷热切换策略，不支持修改冷热表的冷数据的表空间。
  - 冷热表的分区操作约束：
    - 不支持对冷分区的数据进行exchange操作。

- Merge partition分区只支持热分区和热分区合并、冷分区和冷分区合并，不支持冷热分区合并。
- ADD/Merge/Split Partition等分区操作不支持指定表空间为OBS表空间。
- 不支持创建时指定和修改冷热表分区的表空间。
- 冷热切换不是只要满足条件就立刻进行冷热数据切换，依赖用户手动调用切换命令，或者通过调度器调用切换命令后才真正进行数据切换。目前自动调度时间为每日0点，可进行修改。
- 目前冷热切换规则只支持LMT和HPN两种。
- 冷热数据表不支持物理细粒度备份和恢复，由于物理备份时只备份热数据，在备份恢复前后OBS上冷数据为同一份，不支持truncate和drop table等涉及删除文件操作语句的备份恢复操作。

## 使用示例

1. 创建列存冷热数据管理表，指定热数据有效期LMT为100天。

```
CREATE TABLE lifecycle_table(i int, val text) WITH (ORIENTATION = COLUMN, storage_policy = '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;
```
2. 切换冷数据至OBS表空间。
  - 自动切换：每日0点调度框架自动触发，无需关注切换情况；  
可自定义自动切换时间：根据业务情况调整自动触发时间，修改为每天早晨6点30分：；

```
select * from pg_obs_cold_refresh_time('lifecycle_table', '06:30:00');
```
  - 手动切换  
执行如下操作手动切换单表：

```
alter table lifecycle_table refresh storage;
```

  
执行如下操作批量切换所有冷热表：

```
select pg_catalog.pg_refresh_storage();
```
3. 查看冷热表数据分布情况。  
查看单表数据分布情况：

```
select * from pg_catalog.pg_lifecycle_table_data_distribute('lifecycle_table');
```

  
查看所有冷热表数据分布情况：

```
select * from pg_catalog.pg_lifecycle_node_data_distribute();
```

# 6 Oracle、Teradata 和 MySQL 语法兼容性差异

GaussDB(DWS)支持Oracle、Teradata和MySQL三种兼容模式，分别兼容Oracle、Teradata和MySQL语法，不同兼容模式下的语法行为有一些差异。

表 6-1 兼容项差异

兼容项	Oracle兼容	Teradata兼容	MySQL兼容
空串	只有null	区分空串和null	区分空串和null
空串转数字	null	转换为0	转换为0
超长字符自动截断	不支持	支持(GUC参数td_compatible_truncation设置为ON)	不支持
null拼接	非null对象与null拼接后返回非null对象。 例如, 'abc'    null 返回'abc'。	GUC参数behavior_compat_options增加strict_text_concat_td选项后, 兼容TD行为, null类型拼接后返回null。 例如, 'abc'    null返回null。	兼容MySQL行为, null类型拼接后返回null。 例如, 'abc'    null返回null。

兼容项	Oracle兼容	Teradata兼容	MySQL兼容
char(n)类型拼接	char(n)类型做拼接时移除右侧空格和占位。 例如, cast('a' as char(3))  'b'返回'ab'。	GUC参数 behavior_compat_options增加 bpchar_text_without_rtrim选项后, char(n)类型做拼接时, 保留空格, 并补足空格至指定的n长度。 当前不支持“比较字符串时忽略尾部空格”, 拼接后结果如果存在尾部空格, 进行比较时会对空格敏感。 例如, cast('a' as char(3))  'b'返回'a b'。	移除右侧空格和占位
concat(str1,str2)	返回所有非null字符串的连接	返回所有非null字符串的连接	入参中存在null时, 返回结果为null。
left和right负数处理	返回除最后/前 n 个字符以外的所有字符	返回除最后/前 n 个字符以外的所有字符	返回空串
lpad(string text, length int [, fill text]) rpad(string text, length int [, fill text])	通过填充字符 fill(缺省为空格), 把string填充到 length长度, 如果 string已经比 length长则将其尾部截断。如果fill为空串或者length为负数则返回null。	如果fill为空串且 string长度小于length时, 返回原字符串, 如果length为负数则返回空串。	如果fill为空串且 string长度小于length时, 返回空串, 如果length为负数则返回null。
substr(str, s[, n])	s = 0时, 返回前n个字符	s = 0时, 返回前n个字符。	s = 0时, 返回空串。
substring(str, s[, n]) substring(str [from s] [for n])	s = 0时, 返回前n - 1个字符 s < 0时, 返回前s + n - 1个字符 n < 0时, 报错。	s = 0时, 返回前n - 1个字符。 s < 0时, 返回前s + n - 1个字符。 n < 0时, 报错。	s = 0时, 返回空串。 s < 0时, 倒数第 s 个字符位置开始截取n个字符。 n < 0时, 返回空串。
trim、ltrim、rtrim、btrim(string[, characters])	从字符串string的指定位置删除只包含characters中字符(缺省为空格)的最长的字符串。	从字符串string的指定位置删除只包含characters中字符(缺省为空格)的最长的字符串。	从字符串string的指定位置删除等于characters的字符串(缺省为空格)。

兼容项	Oracle兼容	Teradata兼容	MySQL兼容
log(x)	以10为底的对数	以10为底的对数	自然对数
mod(x, 0)	除数为0时返回x	除数为0时返回x	除数为0时报错
数据类型date	date会转为timestamp, 包含年月日时分秒	只有年月	只有年月
to_char(date)	入参最大值仅支持timestamp类型的最大值, 不支持date类型的最大值; 返回值类型为timestamp	入参最大值仅支持timestamp类型的最大值, 不支持date类型的最大值; 返回值类型为date, 且格式为'YYYY/MM/DD'(GUC参数convert_empty_str_to_null_td打开)。	入参最大值仅支持timestamp类型的最大值, 不支持date类型的最大值; 返回值类型为date。
to_date, to_timestamp和to_number空串处理	返回null	返回null(GUC参数convert_empty_str_to_null_td打开)	to_date和to_timestamp返回null, to_number中参数为空串时, 返回0。
last_day和next_day返回类型	timestamp类型	timestamp类型	date类型
add_months返回类型	timestamp类型	timestamp类型	入参为date类型, 返回date类型。 入参为timestamp类型, 返回timestamp类型。 入参为timestampz类型, 返回timestampz类型。
CURRENT_TIME CURRENT_TIME(p)	获取当前事务的时间, 返回值类型为timetz	获取当前事务的时间, 返回值类型为timetz。	获取当前语句执行时的时间, 返回值类型为time。
CURRENT_TIMESTAMP CURRENT_TIMESTAMP(p)	获取当前语句执行时的时间, 返回值类型为timestampz	获取当前语句执行时的时间, 返回值类型为timestampz。	获取当前语句执行时的时间, 返回值类型为timestamp。
CURDATE	不支持	不支持	获取当前语句执行时的日期, 返回值类型为date。

兼容项	Oracle兼容	Teradata兼容	MySQL兼容
CURTIME(p)	不支持	不支持	获取当前语句执行时的时间，返回值类型为time。
LOCALTIME LOCALTIME(p)	获取当前事务的时间，返回值类型为time	获取当前事务的时间，返回值类型为time。	获取当前语句执行时的时间，返回值类型为time。
LOCALTIMEST AMP LOCALTIMEST AMP(p)	获取当前事务的时间，返回值类型为timestamp	获取当前事务的时间，返回值类型为timestamp。	获取当前语句执行时的时间，返回值类型为timestamp。
SYSDATE SYSDATE(p)	获取当前语句执行时的时间，返回值类型为timestamp(0)	获取当前语句执行时的时间，返回值类型为timestamp(0)。	获取当前系统的时间，返回值类型为timestamp(0)。此函数不可下推，建议用current_date代替。
now()	获取当前事务时间，返回值类型为timestamptz	获取当前事务时间，返回值类型为timestamptz。	获取语句执行的时间，返回值类型为timestamptz。
操作符'^'	幂运算	幂运算	异或
表达式 greatest、 least	返回所有非null入参的比较结果	返回所有非null入参的比较结果	入参中存在null时，返回结果为null。
表达式case、 coalesce、if、 ifnull入参类型 不同	报错	兼容TD行为，支持数字和字符串之间的类型转换，比如coalesce参数输入int和varchar类型，解析成varchar类型。	兼容MySQL行为，支持其他类型和字符串之间的类型转换，比如coalesce参数输入date、int和varchar类型，解析成varchar类型。
反引号	不支持	不支持	区分MySQL的保留字与普通字符。

# 7 管理数据库安全

## 7.1 管理用户及权限

### 7.1.1 数据库用户

在非三权分立模式下，GaussDB(DWS)支持管理员用户和普通用户两种类型的数据库帐号。三权分立下，用户类型和权限请参考[三权分立](#)。

- 管理员用户可以管理所有普通用户和数据库。
- 普通用户可连接和访问数据库，并在授权后进行特定的数据库操作以及执行SQL语句。

在用户登录GaussDB(DWS)数据库时会对其进行身份验证。用户可以拥有数据库和数据库对象（例如表），并且可以向用户和角色授予对这些对象的权限以控制谁可以访问哪个对象。除管理员外，具有CREATEDB属性的用户可以创建数据库并授予对这些数据库的权限。

#### 数据库用户类型

表 7-1 数据库用户类型

用户类型	描述	可进行的操作	如何创建
管理员 dbadmin	管理员也称作系统管理员，是指具有SYSADMIN属性的帐户。	非三权分立模式下，拥有系统的最高权限，能够执行所有的操作。系统管理员具有与对象所有者相同的权限。	<ul style="list-style-type: none"><li>• 在GaussDB(DWS) 管理控制台创建集群时创建的用户dbadmin是系统管理员。</li><li>• 使用CREATE USER或ALTER USER语法创建和设置管理员用户。 CREATE USER <i>sysadmin</i> WITH SYSADMIN password '<i>Password</i>'; ALTER USER <i>u1</i> SYSADMIN;</li></ul>

用户类型	描述	可进行的操作	如何创建
普通用户	普通用户	<ul style="list-style-type: none"> <li>使用工具连接数据库。</li> <li>拥有数据库系统特定操作的属性，如CREATEDB、CREATEROLE、SYSADMIN。</li> <li>访问数据库对象。</li> <li>执行SQL语句。</li> </ul>	使用CREATE USER语法创建普通用户。 <pre>CREATE USER u1 PASSWORD '{Password};</pre>
	私有用户	<p>在非三权分立模式下，创建的具有INDEPENDENT属性的私有用户。</p> <p>数据库管理员在未经其授权前，只能进行控制操作（DROP、ALTER、TRUNCATE），无权进行INSERT、DELETE、SELECT、UPDATE、COPY、GRANT、REVOKE、ALTER OWNER操作。</p>	使用CREATE USER语法创建私有用户。 <pre>CREATE USER user_independent WITH INDEPENDENT IDENTIFIED BY '{Password};</pre>

## 7.1.2 用户管理

使用CREATE USER和ALTER USER可以创建和管理数据库用户。

- 非三权分立下，GaussDB(DWS)用户帐户只能由系统管理员或拥有CREATEROLE属性的安全管理员创建和删除。
- 三权分立时，用户帐户只能由安全管理员创建。

### 创建用户

CREATE USER语句用于创建新的GaussDB(DWS)用户。创建新用户后，可以使用该用户连接数据库。

- 创建普通用户u1，并设置用户拥有CREATEDB属性。  

```
CREATE USER u1 WITH CREATEDB PASSWORD '{Password};
```
- 创建系统管理员mydbadmin，需指定参数SYSADMIN。  

```
CREATE USER mydbadmin sysadmin PASSWORD '{Password};
```
- 通过视图PG\_USER查看已创建的用户。  

```
SELECT * FROM pg_user;
```
- 要查看用户属性，请查询系统表PG\_AUTHID。  

```
SELECT * FROM pg_authid;
```

### 修改用户属性

ALTER USER语句用于更改用户属性（例如，更改用户密码或权限等内容）。



示例:

- 修改用户u1的密码:  

```
ALTER USER u1 IDENTIFIED BY 'newpassword' REPLACE 'oldpassword';
```
- 为用户u1赋予CREATEROLE权限:  

```
ALTER USER u1 CREATEROLE;
```

## 锁定用户

ALTER USER语句中ACCOUNT LOCK | ACCOUNT UNLOCK参数用于锁定或者解锁用户，被锁定的用户不允许登录。若管理员发现某帐户被盗、非法访问等异常情况，可手动锁定该帐户；当管理员认为帐户恢复正常后，可手动解锁该帐户。

示例:

- 锁定用户u1:  

```
ALTER USER u1 ACCOUNT LOCK;
```
- 解锁用户u1:  

```
ALTER USER u1 ACCOUNT UNLOCK;
```

## 删除用户

DROP USER语句用于删除一个或多个GaussDB(DWS)用户。当确认帐户不再使用，管理员可以删除用户帐户。用户删除后不可恢复。

- 同时删除多个用户时，用","隔开。
- 成功删除用户后，该用户的所有权限也会被一同删除。
- 当删除的用户正处于活动状态时，此会话状态不会立马断开，用户在会话状态断开后才会被完全删除。
- DROP USER语句指定CASCADE时，可级联删除依赖用户的表等对象。即删除owner是该用户的对象，并清理掉其他对象对该用户的授权信息。

示例:

- 删除用户u1:  

```
DROP USER u1;
```
- 级联删除帐户u2:  

```
DROP USER u2 CASCADE;
```

### 7.1.3 自定义密码策略

创建用户或修改用户时需要指定密码，GaussDB(DWS)有默认密码复杂度要求，也支持用户自定义数据库账户密码策略。

#### GaussDB(DWS)默认密码策略

GaussDB(DWS)默认进行密码复杂度校验（即GUC参数password\_policy默认为1），默认密码策略要求如下：

- 长度为8~32个字符。
- 至少包含大写字母、小写字母、数字或特殊字符中三种的组合。
- 不能是用户名和用户名反序，此条要求为非大小写敏感。
- 不能是当前密码、当前密码的反序。

## 自定义密码策略

密码策略包含：密码复杂度要求、密码有效期、密码重用设置以及密码的加密方式及密码重试与锁定，不同的策略项由对应的GUC参数控制，参见下表（详细内容也可参考[安全和认证（postgresql.conf）](#)）：

表 7-2 自定义密码策略及对应 GUC 参数

密码策略	对应参数名称	参数描述	参数取值范围	GaussDB(DWS)默认值
是否进行密码复杂度检查	password_policy	创建或者修改 GaussDB(DWS)帐户时，该参数决定是否进行密码复杂度检查。	整型，0、1 <ul style="list-style-type: none"> <li>0表示不采用任何密码复杂度策略。设置为0会存在安全风险，不建议设置为0</li> <li>1表示采用默认密码复杂度校验策略。</li> </ul>	1
密码复杂度要求	password_min_length	密码的最小长度。	整型，6~999	8
	password_max_length	密码的最大长度。	整型，6~999	32
	password_min_uppercase	包含大写字母（A-Z）的最少个数。	整型，0~999 <ul style="list-style-type: none"> <li>0表示没有限制。</li> <li>1~999表示创建账户所指定的密码中至少需要包含大写字母个数。</li> </ul>	0
	password_min_lowercase	包含小写字母（a-z）的最少个数。	整型，0~999 <ul style="list-style-type: none"> <li>0表示没有限制。</li> <li>1~999表示创建帐户所指定的密码中至少需要包含小写字母个数。</li> </ul>	0
	password_min_digital	包含数字（0-9）的最少个数。	整型，0~999 <ul style="list-style-type: none"> <li>0表示没有限制。</li> <li>1~999表示创建帐户所指定的密码中至少需要包含数字个数。</li> </ul>	0
password_min_special	包含特殊字符的最少个数（特殊字符的列表请参见表 7-3）。	整型，0~999 <ul style="list-style-type: none"> <li>0表示没有限制。</li> <li>1~999表示创建帐户所指定的密码中至少需要包含特殊字符个数。</li> </ul>	0	

密码策略	对应参数名称	参数描述	参数取值范围	GaussDB(DWS)默认值
密码有效期	password_effect_time	用户的密码有效期。当达到所设置的密码到期提醒天数 password_notify_time 时，系统会在用户登录数据库时提示用户修改密码。	浮点型，0~999，单位为天。 <ul style="list-style-type: none"> <li>0表示不开启有效期限限制功能。</li> <li>1~999表示创建帐户所指定的密码有效期，临近或超过有效期系统会提示用户修改密码。</li> </ul>	90
	password_notify_time	密码到期前提醒的天数。	整型，0~999，单位为天。 <ul style="list-style-type: none"> <li>0表示不开启提醒功能。</li> <li>1~999表示帐户密码到期前提醒的天数。</li> </ul>	7
密码重用设置	password_reuse_time	不可重用天数。	浮点型，0~3650，单位为天。 <ul style="list-style-type: none"> <li>0表示不检查密码可重用天数。</li> <li>正数表示新密码不能为该值指定的天数内使用过的密码。</li> </ul>	60
	password_reuse_max	不可重用次数。	整型，0~1000 <ul style="list-style-type: none"> <li>0表示不检查密码可重用次数。</li> <li>正整数表示新密码不能为该值指定的次数内使用过的密码。</li> </ul>	0
加密方式	password_encryption_type	采用何种加密方式对用户密码进行加密存储。	0、1、2 <ul style="list-style-type: none"> <li>0表示采用md5方式对密码加密。不建议用户使用。</li> <li>1表示采用sha256方式对密码加密，兼容postgres客户端的MD5用户认证方式。</li> <li>2表示采用sha256方式对密码加密。</li> </ul>	2

密码策略	对应参数名称	参数描述	参数取值范围	GaussDB(DWS)默认值
重试与锁定	password_lock_time	帐户被锁定后自动解锁的时间	浮点型，0~365，单位为天。 <ul style="list-style-type: none"> <li>0表示密码验证失败时，自动锁定功能不生效。</li> <li>正数表示帐户被锁定后，当锁定时间超过password_lock_time设定的值时，帐户将会被自行解锁。</li> </ul>	1
	failed_login_attempts	如果输入密码错误的次数达到failed_login_attempts则当前帐户被锁定，password_lock_time秒后被自动解锁。	整型，0~1000 <ul style="list-style-type: none"> <li>0表示自动锁定功能不生效。</li> <li>正整数表示当错误密码次数达到failed_login_attempts设定的值时，当前帐户将被锁定。</li> </ul>	10

表 7-3 特殊字符

编号	字符	编号	字符	编号	字符	编号	字符
1	~	9	*	17		25	<
2	!	10	(	18	[	26	.
3	@	11	)	19	{	27	>
4	#	12	-	20	}	28	/
5	\$	13	_	21	]	29	?
6	%	14	=	22	;	-	-
7	^	15	+	23	:	-	-
8	&	16	\	24	,	-	-

## 自定义密码策略示例

示例一：配置密码复杂度参数password\_policy。

1. 登录GaussDB(DWS) 管理控制台。

2. 在左侧导航栏中，单击“集群管理”。
3. 在集群列表中找到所需要的集群，单击集群名称，进入“集群详情”页面。
4. 单击“参数修改”页签，并在“参数列表”模块修改password\_policy参数值，然后单击“保存”。password\_policy参数无需进行重启集群操作，参数修改后立即生效。

图 7-1 password\_policy



### 示例二：配置密码有效期password\_effect\_time参数。

1. 登录GaussDB(DWS) 管理控制台。
2. 在左侧导航栏中，单击“集群管理”。
3. 在集群列表中找到所需要的集群，单击集群名称，进入“集群详情”页面。
4. 单击“参数修改”页签，并在“参数列表”模块修改password\_effect\_time参数值，然后单击“保存”。password\_effect\_time参数无需进行重启集群操作，参数修改后立即生效。

图 7-2 password\_effect\_time



## 密码设置和修改

- 建议系统管理员和普通用户都要定期修改自己的帐户密码，避免帐户密码被非法窃取。

以修改用户user1密码为例，使用管理员用户连接数据库并执行如下命令：

```
ALTER USER user1 IDENTIFIED BY 'newpassword' REPLACE 'oldpassword';
```

### 说明

密码要符合规则，否则会执行失败。

- 管理员可以修改自己的或者其他帐户的密码。通过修改其他帐户的密码，解决用户密码遗失所造成无法登录的问题。

以修改用户joe帐户密码为例，命令格式如下：

```
ALTER USER joe IDENTIFIED BY 'password';
```

## 📖 说明

- 系统管理员之间不允许互相修改对方密码。
- 系统管理员可以修改普通用户密码且不需要用户原密码。
- 系统管理员可以修改自己的密码但需要管理员原密码。
- 密码验证  
设置当前会话的用户和角色时，需要验证密码。如果输入密码与用户的存储密码不一致，则会报错。

以设置用户joe为例，命令格式如下：

```
SET ROLE joe PASSWORD 'password';
```

显示如下命令表示设置成功：

```
SET ROLE
```

## 7.1.4 权限管理

### 权限概述

**权限**表示用户访问某个数据库对象（包括模式、表、函数、序列等）的操作（包括增、删、改、查、创建等）是否被允许。

GaussDB(DWS)中的权限管理分为三种场景：

- 系统权限  
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。
- 数据对象权限  
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。
- 用户权限  
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。

### 层级权限管理

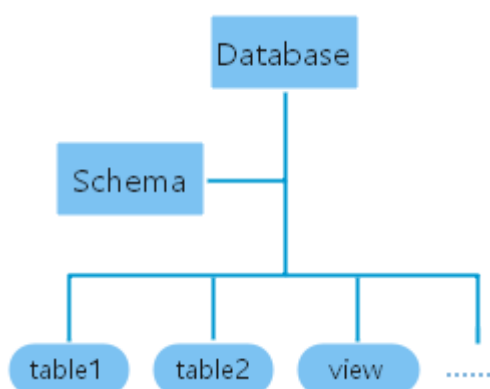
GaussDB(DWS)通过Database、Schema和数据对象权限实现层级权限管理。

- Database之间无法直接互访，通过连接隔离实现彻底的权限隔离。各个Database之间共享资源极少，可实现连接隔离、权限隔离等。数据库集群包含一个或多个已命名数据库。用户和角色在整个集群范围内是共享的，但是其数据并不共享。

即用户可以连接任何数据库，但当连接成功后，任何用户都只能访问连接请求里所声明的数据库。

- Schema隔离的方式共用资源较多，可以通过GRANT与REVOKE语法便捷地控制不同用户对各Schema及其下属对象的权限，从而赋给业务更多的灵活性。每个数据库包括一个或多个Schema。每个Schema包含表、函数等其他类型的对象。用户要访问包含在指定Schema中的对象，需要被授予Schema的USAGE权限。
- 对象创建后，默认只有对象所有者或者系统管理员可以查询、修改和删除对象。其他用户要访问包含具体的数据库对象，例如table1，需要首先被授予database的CONNECT权限，再被授予Schema的USAGE权限，最后授予table1的SELECT权限。用户要访问底层的对象，必须先赋予上层对象的权限。比如用户要创建或者删除Schema，需要首先被授予database的CREATE权限；

图 7-3 层级权限管理



## 角色

GaussDB(DWS)的权限管理模型，是一种典型的RBAC（基于角色的权限控制）的实现。其将用户、角色、权限通过此模型管理起来。

角色是一组权限的集合。

- “用户”概念和“角色”概念实际是等同的，唯一的区别在于“用户”拥有login权限，而“角色”拥有nologin权限。
- 按照数据库系统中承担的责任划分具有不同权限的角色。角色是数据库权限的集合，代表了一个数据库用户、或一组数据用户的行为约束。
- 角色和用户可以转换，通过ALTER将角色拥有登录权限。
- 通过GRANT把角色授予用户后，用户即具有了角色的所有权限。推荐使用角色进行高效权限分配。例如，可以为设计、开发和维护人员创建不同的角色，将角色GRANT给用户后，再向每个角色中的用户授予其所需数据的差异权限。在角色级别授予或撤销权限时，这些权限更改会对角色下的所有成员生效。
- 非三权分立时，只有系统管理员和具有CREATEROLE属性的用户才能创建、修改或删除角色。三权分立下，只有具有CREATEROLE属性的用户才能创建、修改或删除角色。

要查看所有角色，请查询系统表PG\_ROLES：

```
SELECT * FROM PG_ROLES;
```

具体的创建，修改和删除角色操作，请参考SQL语法参考中CREARE ROLE/ALTER ROLE/DROP ROLE。

## 预置角色

GaussDB(DWS)提供了一组预置角色，以“gs\_role\_”开头命名，提供对特定的、通常需要高权限的操作的访问，可以将这些角色授权予数据库中的其他用户或角色，使这些用户能够访问或使用特定的信息和功能。请谨慎使用预置角色，以确保预置角色权限的安全使用。

预置角色允许的权限范围可参考下表：

**表 7-4** 预置角色允许的权限范围

角色	权限描述
gs_role_signal_backend	具有调用函数pg_cancel_backend、pg_terminate_backend、pg_terminate_query、pg_cancel_query、pgxc_terminate_query、pgxc_cancel_query来取消或终止其他会话的权限，但不能操作属于初始用户的会话。
gs_role_read_all_stats	读取系统状态视图并且使用与扩展相关的各种统计信息，包括有些通常只对系统管理员可见的信息。包括： 资源管理类： <ul style="list-style-type: none"> <li>pgxc_wlm_operator_history</li> <li>pgxc_wlm_operator_info</li> <li>pgxc_wlm_operator_statistics</li> <li>pgxc_wlm_session_info</li> <li>pgxc_wlm_session_statistics</li> <li>pgxc_wlm_workload_records</li> <li>pgxc_workload_sql_count</li> <li>pgxc_workload_sql_elapse_time</li> <li>pgxc_workload_transaction</li> </ul> 状态信息类： <ul style="list-style-type: none"> <li>pgxc_stat_activity</li> <li>pgxc_get_table_skewness</li> <li>table_distribution</li> <li>pgxc_total_memory_detail</li> <li>pgxc_os_run_info</li> <li>pg_nodes_memory</li> <li>pgxc_instance_time</li> <li>pgxc_redo_stat</li> </ul>
gs_role_analyze_any	具有系统级ANALYZE权限类似系统管理员用户，跳过schema权限检查，对所有的表可以执行ANALYZE。
gs_role_vacuum_any	具有系统级VACUUM权限类似系统管理员用户，跳过schema权限检查，对所有的表可以执行VACUUM。



### 预置角色的使用约束：

- 以gs\_role\_开头的角色名作为数据库的预置角色保留字，禁止新建以“gs\_role\_”开头的用户/角色，也禁止将已有的用户/角色重命名为以“gs\_role\_”开头。
- 禁止对预置角色执行ALTER和DROP操作。
- 预置角色默认没有LOGIN权限，不设置预置登录密码。
- gsq命令\du和\dg不显示预置角色的相关信息，但若指定了PATTERN（用来指定要被显示的对象名称）则预置角色信息会显示。
- 三权分立关闭时，系统管理员和具有预置角色ADMIN OPTION权限的用户有权对预置角色执行GRANT/REVOKE管理；三权分立打开时，安全管理员（具有CREATEROLE属性）和具有预置角色ADMIN OPTION权限的用户有权对预置角色执行GRANT/REVOKE管理。例如：

```
GRANT gs_role_signal_backend TO user1;  
REVOKE gs_role_signal_backend FROM user1;
```

## 权限授予或撤销

数据库对象创建后，进行对象创建的用户就是该对象的所有者。集群安装后的默认情况下，未开启[三权分立](#)，数据库系统管理员具有与对象所有者相同的权限。

也就是说对象创建后，默认只有对象所有者或者系统管理员可以查询、修改和删除对象，以及通过GRANT将对象的权限授予其他用户。为使其他用户能够使用对象，可以由对象所有者或管理员通过GRANT/REVOKE对其他用户或角色授予与撤销。

- 使用GRANT语句授予权限。  
例如，将模式myschema的权限赋给角色u1后，将表myschema.t1的SELECT权限授予角色u1。  

```
GRANT USAGE ON SCHEMA myschema TO u1;  
GRANT SELECT ON TABLE myschema.t1 to u1;
```
- 使用REVOKE撤销已经授予的权限。  
例如：撤销用户u1在指定表myschema.t1上的所有权限。  

```
REVOKE ALL PRIVILEGES ON myschema.t1 FROM u1;
```

## 7.1.5 三权分立

默认情况下拥有SYSADMIN属性的系统管理员，具备系统最高权限。在实际业务管理中，为了避免系统管理员拥有过度集中的权利带来高风险，可以设置三权分立，将系统管理员的权限分立给安全管理员和审计管理员。

- 三权分立后，系统管理员将不再具有CREATEROLE属性（安全管理员）和AUDITADMIN属性（审计管理员）能力。即不再拥有创建角色和用户的权限，并不再拥有查看和维护数据库审计日志的权限。关于CREATEROLE属性和AUDITADMIN属性的更多信息请参考CREATE ROLE。
- 三权分立后，系统管理员只会对自己作为所有者的对象有权限。

三权分立的设置办法请参考[设置三权分立](#)章节。

三权分立前的权限详情及三权分立后的权限变化，请分别参见[表7-5](#)和[表7-6](#)。

表 7-5 默认的用户权限

对象名称	系统管理员	安全管理员	审计管理员	普通用户
表空间	对表空间有创建、修改、删除、访问、分配操作的权限。	不具有对表空间进行创建、修改、删除、分配的权限，访问需要被赋权。		
表	对所有表有所有的权限。	仅对自己的表有所有的权限，对其他用户的表无权限。		
索引	可以在所有的表上建立索引。	仅可以在自己的表上建立索引。		
模式	对所有模式有所有的权限。	仅对自己的模式有所有的权限，对其他用户的模式无权限。		
函数	对所有的函数有所有的权限。	仅对自己的函数有所有的权限，对其他用户放在public这个公共模式下的函数有调用的权限，对其他用户放在其他模式下的函数无权限。		
自定义视图	对所有的视图有所有的权限。	仅对自己的视图有所有的权限，对其他用户的视图无权限。		
系统表和系统视图	可以查看所有系统表和视图。	只可以查看部分系统表和视图。详细请参见 <a href="#">系统表和系统视图</a> 。		

表 7-6 三权分立较非三权分立权限变化说明

对象名称	系统管理员	安全管理员	审计管理员	普通用户
表空间	无变化	无变化。		
表	权限缩小。 只对自己的表有所有权限，对其他用户放在属于各自模式下的表无权限。	无变化。		
索引	权限缩小。 只可以在自己的表上建立索引。	无变化。		
模式	权限缩小。 只对自己的模式有所有的权限，对其他用户的模式无权限。	无变化。		
函数	权限缩小。 只对自己的函数有所有的权限，对其他用户放在属于各自模式下的函数无权限。	无变化。		

对象名称	系统管理员	安全管理员	审计管理员	普通用户
自定义视图	权限缩小。 只对自己的视图及其他用户放在public模式下的视图有所有的权限，对其他用户放在属于各自模式下的视图无权限。	无变化。		
系统表和系统视图	无变化。	无变化。	无变化。	无权查看任何系统表和视图。

## 7.2 敏感数据管理

### 7.2.1 行级访问控制

行级访问控制特性可以将数据库访问控制精确到数据表行级别，控制用户只能访问数据表的特定数据行，保证读写数据的安全。

#### 行级访问控制设置

行级访问控制的目的是控制表中行级数据可见性，通过在数据表上预定义Filter，在查询优化阶段将满足条件的表达式应用到执行计划上，影响最终的执行结果。当前受影响的SQL语句包括SELECT，UPDATE，DELETE。

- 用户可以使用**CREATE ROW LEVEL SECURITY POLICY**语句在表上创建行访问控制策略(Row Level Security)。  
该策略针对特定数据库用户、特定SQL操作生效的表达式。当数据库用户对数据表访问时，若SQL满足数据表特定的Row Level Security策略，在查询优化阶段将满足条件的表达式，按照属性(PERMISSIVE | RESTRICTIVE)类型，通过AND或OR方式拼接，应用到执行计划上。
- 对表创建了行访问控制策略后，只有打开该表的行访问控制开关(ALTER TABLE ... ENABLE ROW LEVEL SECURITY)，策略才能生效。

#### 行级访问控制示例

某业务场景中，表all\_data中汇总了不同用户的数据，使用行级访问控制功能实现：不同用户只能查看自身相关的数据信息，不能查看其他用户的数据信息。

##### 步骤1 创建用户alice, bob和peter。

```
CREATE ROLE alice PASSWORD '*****';
CREATE ROLE bob PASSWORD '*****';
CREATE ROLE peter PASSWORD '*****';
```

创建表all\_data并插入数据，表all\_data包含不同用户的数据信息。

```
CREATE TABLE public.all_data(id int, role varchar(100), data varchar(100));
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');
```

**步骤2** 将表all\_data的读取权限赋予用户alice, bob和peter。

```
GRANT SELECT ON all_data TO alice, bob, peter;
```

**步骤3** 使用ALTER TABLE *tablename* ENABLE ROW LEVEL SECURITY语句打开行访问控制策略开关。

```
ALTER TABLE all_data ENABLE ROW LEVEL SECURITY;
```

**步骤4** 使用CREATE ROW LEVEL SECURITY POLICY语句创建行访问控制策略, 使当前用户只能查看用户自己的数据。

```
CREATE ROW LEVEL SECURITY POLICY all_data_rls ON all_data USING(role = CURRENT_USER);
```

**步骤5** 查看表all\_data信息。

```
\d+ all_data
      Table "public.all_data"
  Column |          Type          | Modifiers | Storage | Stats target | Description
-----+-----+-----+-----+-----+-----
  id     | integer                |           |         |              |
  role   | character varying(100) |           | extended |              |
  data   | character varying(100) |           | extended |              |
Row Level Security Policies:
  POLICY "all_data_rls"
  USING (((role)::name = "current_user"()))
Has OIDs: no
Distribute By: ROUND ROBIN
Location Nodes: ALL DATANODES
Options: orientation=row, compression=no, enable_rowsecurity=true
```

**步骤6** 切换至用户alice, 查询表all\_data。查询结果显示行级访问控制策略生效: 用户alice只能查看自己的数据。

```
SET ROLE alice PASSWORD '*****';
```

```
SELECT * FROM all_data;
```

```
id | role | data
---+-----+-----
 1 | alice | alice data
```

显示表查询的执行计划, 提示访问all\_data表时受行级访问控制特性影响。

```
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: 4833280KB
Query Max mem: 4833280KB
Query estimated mem: 1024KB
(16 rows)
```

**步骤7** 切换至用户peter, 查询表all\_data。查询结果显示行级访问控制策略生效: 用户peter只能查看自己的数据。

```
SET ROLE peter PASSWORD '*****';
```

```
SELECT * FROM all_data;
id | role | data
-----+-----
 3 | peter | peter data
(1 row)
```

显示表查询的执行计划，提示查询all\_data表受行级访问控制特性影响。

```
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 = 'peter'::name)
Notice: This query is influenced by row level security feature

===== Query Summary =====
-----+-----
System available mem: 4833280KB
Query Max mem: 4833280KB
Query estimated mem: 1024KB
(16 rows)
```

----结束

## 7.2.2 数据脱敏

GaussDB(DWS)提供列级别的数据脱敏(Dynamic Data Masking)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，从而增强产品在数据安全和隐私保护方面的能力。

- 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，且支持按角色匹配脱敏规则。详细内容请参见《SQL语法参考》中“CREATE REDACTION POLICY”章节。
- 提供脱敏策略的修改语法，包括修改脱敏策略生效表达式、重命名脱敏策略，以及新增、修改、删除脱敏列信息。详细内容请参见《SQL语法参考》中“ALTER REDACTION POLICY”章节。
- 删除脱敏策略，可以一键式删除脱敏策略在表的所有列字段上的脱敏函数信息。详细内容请参见《SQL语法参考》中“DROP REDACTION POLICY”章节。
- GaussDB(DWS)提供MASK\_NONE、MASK\_FULL、MASK\_PARTIAL三种内置脱敏函数，也支持使用C语言、PL/PGSQL语言创建的用户自定义脱敏函数。详细规格请参见《SQL语法参考》中“数据脱敏函数”章节。
- 脱敏策略信息存储在系统表PG\_REDACTION\_POLICY，脱敏列信息存储在系统表PG\_REDACTION\_COLUMN。
- 用户可以通过系统视图REDACTION\_POLICIES和REDACTION\_COLUMNS，更加方便地查看脱敏策略及脱敏列信息。

## 📖 说明

- 通常，用户可以执行SELECT语句查看敏感信息的脱敏效果。如果语句有如下特征，则可能存在故意套取敏感数据的可能性，造成语句执行报错。
  - GROUP BY子句引用与目标列一样含有脱敏列的Target Entry作为分组。
  - DISTINCT作用在输出的脱敏列上。
  - 带有CTE的语句。
  - 涉及集合操作。
  - 子查询的目标列不是基表的脱敏列，而是基表脱敏列的表达式或者函数调用。
- 支持脱敏数据的COPY TO或者GDS导出功能。由于脱敏数据的不可逆性，针对脱敏数据的二次运算无任何实际意义。
- 禁止UPDATE、MERGE INTO、DELETE语句的目标列涉及脱敏列。
- UPSERT语句允许通过EXCLUDED更新插入数据。如果引用脱敏列更新基表数据，存在误改数据的可能，执行会报错。
- 8.2.0集群版本开始新增数据脱敏可算功能，通过guc参数`enable_redactcol_computable`控制。开启该功能后运算时使用原始数据，仅在导出数据库时对敏感数据进行脱敏处理。
- 数据脱敏可算功能支持DML语句的目标列涉及脱敏列。
  - 目标表如果是普通表则将源表脱敏策略同步到目标表，实际数据为未经过脱敏处理的原始数据，查询时会根据脱敏策略进行脱敏。
  - 目标表如果是外表，存在直接查看外表文件中数据的风险，所以实际数据为已经过脱敏处理的原始数据。

## 示例

以员工表emp，管理员用户alice以及普通用户matu、july为例，简要介绍数据脱敏过程。其中，用户alice是表emp的属主，表emp包含员工的姓名、手机号、邮箱、银行卡号、薪资等隐私数据。

1. 创建用户alice、matu和july。

```
CREATE ROLE alice PASSWORD 'password';
CREATE ROLE matu PASSWORD 'password';
CREATE ROLE july PASSWORD 'password';
```
2. 用户alice创建表emp并插入三条员工信息。

```
CREATE TABLE emp(id int, name varchar(20), phone_no varchar(11), card_no number, card_string
varchar(19), email text, salary numeric(100, 4), birthday date);
INSERT INTO emp VALUES(1, 'anny', '13420002340', 1234123412341234, '1234-1234-1234-1234',
'smithWu@163.com', 10000.00, '1999-10-02');
INSERT INTO emp VALUES(2, 'bob', '18299023211', 3456345634563456, '3456-3456-3456-3456',
'66allen_mm@qq.com', 9999.99, '1989-12-12');
INSERT INTO emp VALUES(3, 'cici', '15512231233', NULL, NULL, 'jonesishere@sina.com', NULL,
'1992-11-06');
```
3. 用户alice将表emp的读取权限授予用户matu、july。

```
GRANT SELECT ON emp TO matu, july;
```
4. 仅用户alice可查看所有员工信息，matu和july对员工所有银行卡号和薪资数据不可见，于是，对表emp创建脱敏策略，分布为字段card\_no、card\_string和salary绑定脱敏函数。

```
CREATE REDACTION POLICY mask_emp ON emp WHEN (current_user IN ('matu', 'july'))
ADD COLUMN card_no WITH mask_full(card_no),
ADD COLUMN card_string WITH mask_partial(card_string, 'VVVVFVVVVFVVVVFVVV', 'VVVV-VVVV-
VVVV-VVVV', '#', 1, 12),
ADD COLUMN salary WITH mask_partial(salary, '9', 1, length(salary) - 2);
```
5. 切换到用户matu和july，查看员工表emp。

```
SET ROLE matu PASSWORD 'password';
SELECT * FROM emp;
id | name | phone_no | card_no | card_string | email | salary | birthday
```

```

-----+-----+-----+-----+-----+-----+-----+-----+-----+-----
 1 | anny | 13420002340 |    0 | #####-#####-#####-1234 | smithWu@163.com | 99999.9990 |
1999-10-02 00:00:00
 2 | bob  | 18299023211 |    0 | #####-#####-#####-3456 | 66allen_mm@qq.com | 9999.9990 |
1989-12-12 00:00:00
 3 | cici | 15512231233 |      |                               | jonesishere@sina.com |           | 1992-11-06 00:00:00
(3 rows)
SET ROLE july PASSWORD 'password';
SELECT * FROM emp;
id | name | phone_no | card_no | card_string | email | salary | birthday
-----+-----+-----+-----+-----+-----+-----+-----+-----+-----
 1 | anny | 13420002340 |    0 | #####-#####-#####-1234 | smithWu@163.com | 99999.9990 |
1999-10-02 00:00:00
 2 | bob  | 18299023211 |    0 | #####-#####-#####-3456 | 66allen_mm@qq.com | 9999.9990 |
1989-12-12 00:00:00
 3 | cici | 15512231233 |      |                               | jonesishere@sina.com |           | 1992-11-06 00:00:00
(3 rows)

```

6. 用户matu也享有了员工所有信息的查看权限，只有july不可见，修改策略生效范围。

```
ALTER REDACTION POLICY mask_emp ON emp WHEN(current_user = 'july');
```

7. 切换到用户matu和july，重新查看员工表emp。

```

SET ROLE matu PASSWORD 'password';
SELECT * FROM emp;
id | name | phone_no | card_no | card_string | email | salary | birthday
-----+-----+-----+-----+-----+-----+-----+-----+-----+-----
 1 | anny | 13420002340 | 1234123412341234 | 1234-1234-1234-1234 | smithWu@163.com | 10000.0000 | 1999-10-02 00:00:00
 2 | bob  | 18299023211 | 3456345634563456 | 3456-3456-3456-3456 | 66allen_mm@qq.com | 9999.9900 | 1989-12-12 00:00:00
 3 | cici | 15512231233 |      |                               | jonesishere@sina.com |           | 1992-11-06 00:00:00
(3 rows)
SET ROLE july PASSWORD 'password';
SELECT * FROM emp;
id | name | phone_no | card_no | card_string | email | salary | birthday
-----+-----+-----+-----+-----+-----+-----+-----+-----+-----
 1 | anny | 13420002340 |    0 | #####-#####-#####-1234 | smithWu@163.com | 99999.9990 |
1999-10-02 00:00:00
 2 | bob  | 18299023211 |    0 | #####-#####-#####-3456 | 66allen_mm@qq.com | 9999.9990 |
1989-12-12 00:00:00
 3 | cici | 15512231233 |      |                               | jonesishere@sina.com |           | 1992-11-06 00:00:00
(3 rows)

```

8. 员工信息phone\_no、email和birthday也是隐私数据，更新脱敏策略mask\_emp，新增三个脱敏列。

```

ALTER REDACTION POLICY mask_emp ON emp ADD COLUMN phone_no WITH
mask_partial(phone_no, '*', 4);
ALTER REDACTION POLICY mask_emp ON emp ADD COLUMN email WITH mask_partial(email, '*', 1,
position('@' in email));
ALTER REDACTION POLICY mask_emp ON emp ADD COLUMN birthday WITH mask_full(birthday);

```

9. 切换到用户july，查看表emp数据。

```

SET ROLE july PASSWORD 'password';
SELECT * FROM emp;
id | name | phone_no | card_no | card_string | email | salary | birthday
-----+-----+-----+-----+-----+-----+-----+-----+-----+-----
 1 | anny | 134***** |    0 | #####-#####-#####-1234 | *****163.com | 99999.9990 | 1970-01-01 00:00:00
 2 | bob  | 182***** |    0 | #####-#####-#####-3456 | *****qq.com | 9999.9990 | 1970-01-01 00:00:00
 3 | cici | 155***** |      |                               | *****sina.com |           | 1970-01-01 00:00:00
(3 rows)

```

10. 查询视图redaction\_policies和redaction\_columns，查看当前脱敏策略mask\_emp的详细信息。

```

SELECT * FROM redaction_policies;
 object_schema | object_owner | object_name | policy_name | expression | enable |
 policy_description | inherited
-----+-----+-----+-----+-----+-----+-----
 public      | alice      | emp      | mask_emp   | ("current_user"() = 'july'::name) | t
 |          | f
(1 row)
SELECT object_name, column_name, function_info FROM redaction_columns;
 object_name | column_name |
 function_info
-----+-----
 emp      | card_no    | mask_full(card_no)
 emp      | card_string | mask_partial(card_string, 'VVVVFVVVVVVVVVVVVVVV'::text, 'VVVV-VVVV-
VVVV-VVVV'::text, '#'::text, 1, 12)
 emp      | email      | mask_partial(email, '*'::text, 1, "position"(email, '@'::text))
 emp      | salary     | mask_partial(salary, '9'::text, 1, (length((salary)::text) - 2))
 emp      | birthday   | mask_full(birthday)
 emp      | phone_no   | mask_partial(phone_no, '*'::text, 4)
(6 rows)

```

11. 新增一列salary\_info，若需要将文本类型的薪资信息统一脱敏成“\*.\*”，可以创建自定义脱敏函数实现。此处采用PL/PGSQL语言定义脱敏函数mask\_regexp\_salary，创建脱敏列时，只需自定义脱敏的函数名和参数列表，详细内容可参考[用户自定义函数](#)。

```

ALTER TABLE emp ADD COLUMN salary_info TEXT;
UPDATE emp SET salary_info = salary::text;
CREATE FUNCTION mask_regexp_salary(salary_info text) RETURNS text AS
$$
SELECT regexp_replace($1, '[0-9]+'::text, '*');
$$
LANGUAGE SQL
STRICT SHIPPABLE;
ALTER REDACTION POLICY mask_emp ON emp ADD COLUMN salary_info WITH
mask_regexp_salary(salary_info);
SET ROLE july PASSWORD 'password';
SELECT id, name, salary_info FROM emp;
 id | name | salary_info
-----+-----
 1 | anny | *.*
 2 | bob  | *.*
 3 | cici |
(3 rows)

```

12. 使用数据脱敏可算功能，确保GUC参数enable\_redactcol\_computable已提前开启。

```

SET ROLE july PASSWORD 'password';
SELECT id, name, salary_info FROM emp GROUP BY id, name, salary_info;
 id | name | salary_info
-----+-----
 1 | anny | *.*
 2 | bob  | *.*
 3 | cici |
(3 rows)

```

13. 无需为表emp设置敏感策略，删除脱敏策略mask\_emp。

```
DROP REDACTION POLICY mask_emp ON emp;
```

## 7.2.3 使用函数加解密

GaussDB(DWS)支持使用以下函数对字符串进行加解密。

- gs\_encrypt(encryptstr, keystr, cryptotype, cryptomode, hashmethod)  
描述：采用cryptotype和cryptomode组成的加密算法以及hashmethod指定的HMAC算法，以keystr为密钥对encryptstr字符串进行加密，返回加密后的字符串。支持的cryptotype：aes128, aes192, aes256, sm4。支持的cryptomode：



cbc。支持的hashmethod: sha256, sha384, sha512, sm3。支持的加密数据类型: 目前数据库支持的数值类型, 字符类型, 二进制类型中的RAW, 日期/时间类型中的DATE、TIMESTAMP、SMALLDATETIME。keystr的长度范围与加密算法相关, 为1~KeyLen字节。当cryptotype为aes128和sm4时, KeyLen为16, aes192时KeyLen为24, aes256时KeyLen为32。

返回值类型: text

返回值长度: 至少为  $4 * [(maclen + 56) / 3]$  字节, 不超过  $4 * [(Len + maclen + 56) / 3]$  字节, 其中Len为加密前数据长度(单位为字节), maclen为HMAC值的长度, 当hashmethod为sha256和sm3时maclen为32, sha384时maclen为48, sha512时maclen为64。即当hashmethod为sha256和sm3时, 返回值长度至少为120字节, 不超过  $4 * [(Len + 88) / 3]$  字节; 当hashmethod为sha384时, 返回值长度至少为140字节, 不超过  $4 * [(Len + 104) / 3]$  字节; 当hashmethod为sha512时, 返回值长度至少为160字节, 不超过  $4 * [(Len + 120) / 3]$  字节;

示例:

```
SELECT gs_encrypt('GaussDB(DWS)', '1234', 'aes128', 'cbc', 'sha256');
          gs_encrypt
-----
AAAAAAAAAAACcFjDcCSbop7D87sOa2nxTfrkE9RJQGK34ypgrOPsFJIqggI8tl
+eMDcQYT3po98wPCC7VBfhv7mdBy7IVnzdrp0rdMrD6/zTl8w0v9/s2OA==
(1 row)
```

#### 📖 说明

- 由于该函数的执行过程需要传入解密口令, 为了安全起见, gsql工具不会将该函数记录入执行历史。即无法在gsql里通过上下翻页功能找到该函数的执行历史。
- 在同一张数据表中, 加密函数ge\_encrypt与gs\_encrypt\_aes128不要混合使用。
- gs\_decrypt(decryptstr, keystr, cryptotype, cryptomode, hashmethod)

描述: 采用cryptotype和cryptomode组成的加密算法以及hashmethod指定的HMAC算法, 以keystr为密钥对decryptstr字符串进行解密, 返回解密后的字符串。解密使用的keystr必须保证与加密时使用的keystr一致才能正常解密。keystr不得为空。

返回值类型: text

示例:

```
SELECT gs_decrypt('AAAAAAAAAAACcFjDcCSbop7D87sOa2nxTfrkE9RJQGK34ypgrOPsFJIqggI8tl
+eMDcQYT3po98wPCC7VBfhv7mdBy7IVnzdrp0rdMrD6/zTl8w0v9/s2OA==', '1234', 'aes128', 'cbc',
'sha256');
          gs_decrypt
-----
GaussDB(DWS)
(1 row)
```

#### 📖 说明

- 由于该函数的执行过程需要传入解密口令, 为了安全起见, gsql工具不会将该函数记录入执行历史。即无法在gsql里通过上下翻页功能找到该函数的执行历史。
- 此函数需要结合gs\_encrypt加密函数共同使用, 且加密算法和HMAC算法要保证一致。
- gs\_encrypt\_aes128(encryptstr, keystr)

描述: 以keystr为密钥对encryptstr字符串进行加密, 返回加密后的字符串。keystr的长度范围为1~16字节。支持的加密数据类型: 目前数据库支持的数值类型, 字符类型, 二进制类型中的RAW, 日期/时间类型中的DATE、TIMESTAMP、SMALLDATETIME。

返回值类型: text

返回值长度：至少为92字节，不超过  $4 * [(Len+68)/3]$  字节，其中Len为加密前数据长度（单位为字节）。

示例：

```
SELECT gs_encrypt_aes128('MPPDB','1234');
-----
gs_encrypt_aes128
-----
gwditQLQG8NhFw4OuoKhhQJoXojhFLYkjeG0aYdSCtLCnIUgkNwwYI04KbuhmcGZp8jWizBdR1vU9Cspjuzl
0lbz12A=
(1 row)
```

#### 📖 说明

- 由于该函数的执行过程需要传入解密口令，为了安全起见，gsqL工具不会将该函数记录入执行历史。即无法在gsqL里通过上下翻页功能找到该函数的执行历史。
  - 在同一张数据表中，加密函数gs\_encrypt\_aes128与ge\_encrypt不要混合使用。
- gs\_decrypt\_aes128(decryptstr,keyststr)

描述：以keyststr为密钥对decryptstr字符串进行解密，返回解密后的字符串。解密使用的keyststr必须保证与加密时使用的keyststr一致才能正常解密。keyststr不得为空。

返回值类型：text

示例：

```
SELECT
gs_decrypt_aes128('gwditQLQG8NhFw4OuoKhhQJoXojhFLYkjeG0aYdSCtLCnIUgkNwwYI04KbuhmcGZp8j
WizBdR1vU9Cspjuzl0lbz12A=', '1234');
-----
gs_decrypt_aes128
-----
MPPDB
(1 row)
```

#### 📖 说明

- 由于该函数的执行过程需要传入解密口令，为了安全起见，gsqL工具不会将该函数记录入执行历史。即无法在gsqL里通过上下翻页功能找到该函数的执行历史。
  - 此函数需要结合gs\_encrypt\_aes128加密函数共同使用。
- gs\_hash(hashstr, hashmethod)

描述：以hashmethod算法对hashstr字符串进行信息摘要，返回信息摘要字符串。支持的hashmethod：sha256, sha384, sha512, sm3。

返回值类型：text

返回值长度：sha256和sm3返回64字节，sha384返回96字节，sha512返回128字节。

示例：

```
SELECT gs_hash('GaussDB(DWS)', 'sha256');
-----
gs_hash
-----
e59069daa6541ae20af7c747662702c731b26b8abd7a788f4d15611aa0db608efdbb5587ba90789a983f8
5dd51766609
(1 row)
```

- md5(string)

描述：将string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('ABC');
       md5
-----
902fbbd2b1df0c4f70b4a5d23525e932
(1 row)
```

## 7.2.4 使用 pgcrypto 加密数据

GaussDB(DWS)数据库自8.2.0集群版本开始内置加密解密模块pgcrypto。pgcrypto模块允许数据库用户以加密形式存储数据的某些列，为敏感数据增加了一层额外的保护。因此在没有加密密钥的情况下，任何人都无法读取以加密形式存储在GaussDB(DWS)数据库中的数据。

pgcrypto函数在数据库服务器内部运行，这意味着所有数据和密码都以明文形式在pgcrypto和客户端应用程序之间传输。为了获得最佳安全性，建议在客户端和GaussDB(DWS)服务器之间使用SSL连接。

有关pgcrypto模块中各个函数的详细信息如下:

### 通用哈希函数

- digest()

digest()函数可以根据不同的算法生成数据的二进制哈希值，语法如下:

```
digest(data text, type text) returns bytea
digest(data bytea, type text) returns bytea
```

其中，data是原始数据，type是加密算法包括md5、sha1、sha224、sha256、sha384、sha512和sm3；函数的返回结果为二进制字符串。

示例:

使用digest() 函数对字符串GaussDB(DWS)进行sha256加密存储:

```
select digest('GaussDB(DWS)', 'sha256');
       digest
-----
\xcc2d1b97c6adfb44bbce7386516f63f16fc6e6a10bd938861d3aba501ac8aab
(1 row)
```

- hmac()

hmac()函数可以根据不同的算法为带有密钥的数据计算出MAC值，语法如下:

```
hmac(data text, key text, type text) returns bytea
hmac(data bytea, key bytea, type text) returns bytea
```

其中，data是原始数据，key是加密密钥，type是加密算法，包括md5、sha1、sha224、sha256、sha384、sha512以及sm3；函数的返回结果为二进制字符串。

示例:

使用key123密钥和sha256加密算法对字符串GaussDB(DWS)计算MAC值:

```
select hmac('GaussDB(DWS)', 'key123', 'sha256');
       hmac
-----
\x14e1d9e110e9b11ab8379dc02b49533d50a6f4deafe6d6cd451d06c106c97d83
(1 row)
```

对于digest() 函数，如果同时被修改了原始数据和加密结果，无法进行识别；hmac() 函数只要密钥没有泄露的话，可以发现被篡改的数据。

如果该密钥大于哈希块的长度，它将先被哈希然后把结果用作密钥。

## 密码哈希函数

crypt()和gen\_salt()函数专用于哈希密码。crypt()执行哈希用于加密数据，gen\_salt()用于生成加盐哈希。

crypt()中的算法和普通的MD5或者SHA1哈希算法存在以下不同之处：

- crypt()中算法很慢。由于密码包含的数据量很小，这是增加暴力破解难度的唯一方法。
- 它们使用一个随机值（称为salt，即盐值），因此这样具有相同口令的用户将得到不同的密文口令。这也是针对破解算法提供一种额外的安全保护。
- 它们的结果中包括了算法类型，因此可以针对不同用户使用不同的算法对密码进行加密。
- 其中一些算法具有自适应性，意味着当计算机性能变得更快时，可以调整该算法使其变得更慢，而不会产生与已有密码的不兼容性。

crypt()函数所支持的算法如下表：

表 7-7 crypt()支持的算法

算法	密码最大长度	自适应性	Salt位数	输出结果长度	描述
bf	72	√	128	60	基于Blowfish的2a变种算法
md5	unlimited	×	48	34	基于MD5的加密算法
xdes	8	√	24	20	扩展DES
des	8	×	12	13	原生UNIX加密算法

- crypt()

crypt()语法格式如下：

```
crypt(password text, salt text) returns text
```

该函数返回password字符串crypt(3)格式的哈希值，salt参数由gen\_salt()函数生成。

对于相同的密码，crypt()函数每次也会返回不同的结果，因为gen\_salt()函数每次都会生成不同的salt。校验密码时可以将之前生成的哈希结果作为salt。

例如：设置一个新密码：

```
UPDATE ... SET pswhash = crypt('new password', gen_salt('bf',10));
```

通过比较存储的密码哈希值验证输入的密码的正确性。

```
SELECT (pswhash = crypt('entered password', pswhash)) AS pswmatch FROM ...;
```

如果输入的口令正确，这会返回true。

示例：

```
create table userpwd(userid int8, pwd text);
CREATE TABLE
```

```
insert into userpwd values (1, crypt('this is a pwd', gen_salt('bf',10)));
INSERT 0 1
```

```
select crypt('this is a pwd', pwd)=pwd as result from userpwd where userid =1;
result
-----
t
(1 row)

select crypt('this is a wrong pwd', pwd)=pwd as result from userpwd where userid =1;
result
-----
f
(1 row)
```

- **gen\_salt()**

gen\_salt()函数用来产生随机的参数给crypt，语法如下：

```
gen_salt(type text [, iter_count integer ]) returns text
```

该函数每次都会生成一个随机的盐值(salt)字符串，该字符串同时决定了crypt()函数使用的算法；type参数用于指定一个生成字符串的哈希算法，取值包括 des、xdes、md5 以及 bf。对于xdes和bf算法，iter\_count指迭代次数，数字越大加密时间越长，被破解需要的时间也越长。

```
SELECT gen_salt('des'), gen_salt('xdes'), gen_salt('md5'), gen_salt('bf');
gen_salt | gen_salt | gen_salt | gen_salt
-----+-----+-----+-----
qh      | _J9..uEUi | $1$SNgqyKAi | $2a$06$B/Etc3J8zYBV49LrDU97MO
(1 row)
```

每种算法生成的salt拥有固定的格式，例如bf算法结果中的\$2a\$06\$，2a表示Blowfish的2a变种算法，06表示迭代的次数。

如果忽略 iter\_count，将会使用默认的迭代次数。允许的iter\_count值与算法相关，如下表所示。对于xdes算法，迭代次数必须是一个奇数。

表 7-8 crypt()的迭代计数

算法	默认值	最小值	最大值
xdes	725	1	16777215
bf	6	4	31

## PGP 加密函数

GaussDB(DWS)的PGP加密函数遵循OpenPGP(RFC 4880)标准，包括对称密钥加密（私钥加密）和非对称密钥加密（公钥加密）。

加密后的PGP消息由两部分组成：

- 这个消息的会话密钥（加密过的对称密钥或者公钥）。
- 使用该会话密钥加密的数据。

对于对称密钥（也就是密码）加密：

1. 使用String2Key（S2K）算法对密钥进行加密，类似于执行一个特意减慢并且包含随机salt的crypt()算法，生成一个完整长度的二进制密钥。
2. 如果要求一个单独的会话密钥，生成一个随机的密钥。否则使用上面的S2K密钥直接作为会话密钥。
3. 如果直接使用S2K密钥，只将S2K设置加入会话密钥包中。否则，使用S2K密钥对会话密钥进行加密，然后放入会话密钥包中。

对于公钥加密：

1. 生成一个随机的会话密钥。
2. 使用公钥对其进行加密后放入会话密钥包中。

无论哪种情况，数据的加密过程如下：

1. 执行可选的数据操作：压缩、转换成UTF-8或者换行符的转换。
2. 在数据前面增加一个随机字节组成的块，相当于使用了一个随机的初始值（IV）。
3. 追加随机前缀和数据的SHA1哈希值到数据后面。
4. 将所有内容使用会话密钥进行加密后放入数据包中。

支持的PGP加密函数：

- `pgp_sym_encrypt()`

描述：用于对称密钥加密。

语法格式：

```
pgp_sym_encrypt(data text, psw text [, options text ]) returns bytea  
pgp_sym_encrypt_bytea(data bytea, psw text [, options text ]) returns bytea
```

其中，`data`是要加密的数据；`psw`是PGP对称密钥；`options`参数用于设置选项，参考表7-9。

- `pgp_sym_decrypt()`

描述：用于解密PGP对称密钥加密后的消息。

语法格式：

```
pgp_sym_decrypt(msg bytea, psw text [, options text ]) returns text  
pgp_sym_decrypt_bytea(msg bytea, psw text [, options text ]) returns bytea
```

其中，`msg`是要解密的消息；`psw`是PGP对称密钥；`options`参数用于设置选项，参考表7-9。为了避免输出无效的字符，不允许使用`pgp_sym_decrypt`函数对`bytea`数据进行解密；可以使用`pgp_sym_decrypt_bytea`对原始文本数据进行解密。

- `pgp_pub_encrypt()`

描述：用于公共密钥加密。

语法格式：

```
pgp_pub_encrypt(data text, key bytea [, options text ]) returns bytea  
pgp_pub_encrypt_bytea(data bytea, key bytea [, options text ]) returns bytea
```

其中，`data`是要加密的数据；`key`是PGP公钥，如果传入一个私钥将会返回错误；`options`参数用于设置选项，参考表7-9。

- `pgp_pub_decrypt()`

描述：用于解密PGP公共密钥加密后的消息。

语法格式：

```
pgp_pub_decrypt(msg bytea, key bytea [, psw text [, options text ]]) returns text  
pgp_pub_decrypt_bytea(msg bytea, key bytea [, psw text [, options text ]]) returns bytea
```

解密一个公共密钥加密的消息。`key`必须是对应于用来加密的公钥的私钥。如果私钥是用口令保护的，必须在`psw`中给出该口令。如果没有口令，但想要指定选项，需要给出一个空口令。

为了避免输出无效的字符，不允许使用`pgp_pub_decrypt`函数对`bytea`数据进行解密；可以使用`pgp_pub_decrypt_bytea`对原始文本数据进行解密。

其中，`key`是公共密钥对应的私钥；如果私钥使用了密码保护功能，必须在`psw`参数中指定密码；如果没有使用密码保护，想要指定`options`参数时必须指定一个空的`psw`。`options`参数用于设置选项，参考表7-9。

- `pgp_key_id()`  
描述：用于提取PGP公钥或者私钥的密钥ID；如果传入一个加密后的消息，将会返回加密该消息使用的密钥ID。  
语法格式：  
`pgp_key_id(bytea) returns text`  
该函数可能返回两个特殊密钥ID：
  - `SYMKEY`，表示该消息使用对称密钥进行加密。
  - `ANYKEY`，表示该消息使用公钥进行加密，但是密钥ID已经被删除。这意味着需要尝试所有的密钥，查找可以解密该消息的私钥。`pgcrypto`不会产生这种加密消息。

#### 📖 说明

不同的密钥可能拥有相同的ID，这种情况很少见但可能存在。客户端应用程序需要自己尝试使用不同的密钥进行解密，就像处理`ANYKEY`一样。

- `armor()`  
描述：用于将二进制数据转换为PGP ASCII-armor格式，相当于Base64加上CRC以及额外的格式化。  
语法格式：  
`armor(data bytea [ , keys text[], values text[] ]) returns text`
- `dearmor()`  
描述：用于执行相反的操作。  
语法格式：  
`dearmor(data text) returns bytea`  
将加密后的数据`bytea`，转换成PGP ASCII-armor格式或者反向转换。  
其中，`data`是需要转换的数据；如果指定了`keys`和`values`数值，每个`key/value`对都会生成一个armor header并添加到编码格式中；两个数组都是一维数组，长度相同，并且不能包含非ASCII字符。
- `pgp_armor_headers()`  
描述：函数用于返回数据中的armor header。  
`pgp_armor_headers(data text, key out text, value out text) returns setof record`  
返回结果是一个包含`key`和`value`两个字段的记录行集，如果其中包含任何非ASCII字符，都会被视作UTF-8字符。

#### 用GnuPG生成PGP 密钥

要生成一个新密钥：

```
gpg --gen-key
```

更好的密钥类型是“DSA和Elgamal”。

对于RSA密钥，你必须创建仅用于签名的DSA或RSA密钥作为主控密钥，然后用`gpg --edit-key`增加一个RSA加密子密钥。

要列举密钥：

```
gpg --list-secret-keys
```

要以ASCII-保护格式导出一个公钥：

```
gpg -a --export KEYID > public.key
```

要以ASCII-保护格式导出一个私钥：

```
gpg -a --export-secret-keys KEYID > secret.key
```

在把这些密钥交给PGP函数之前，需要对它们使用`dearmor()`。或者如果你能处理二进制数据，可以从命令中去掉`-a`。

**须知**

PGP加密函数的存在以下限制：

- 不支持签名。这也意味着它不会检查加密子密钥是否属于主控密钥。
- 不支持加密密钥作为主控密钥。由于通常并不鼓励这种用法，因此这应该不是问题。
- 不支持多个子密钥。由于实际应用中经常需要多个子密钥，这可能是个问题。另一方面，不要使用常规GPG/PGP密钥作为pgcrypto加密密钥，而应该创建新的密钥，因为这是非常不同的使用场景。

**PGP函数的选项**

pgcrypto函数中的选项名称和GnuPG类似，选项的值使用等号设置，每个选项使用逗号进行分隔。例如：

```
pgp_sym_encrypt(data, psw, 'compress-algo=1, cipher-algo=aes256')
```

除了convert-crlf之外，其他选项仅适用于加密函数。解密函数会从PGP数据中获取参数。

最常设置的选项包括compress-algo和unicode-mode。其他选项通常使用默认值。

表 7-9 pgcrypto 加密选项

选项	描述	默认值	取值	适用函数
cipher-algo	使用的密码算法。	aes128	bf, aes128, aes192, aes256, 3des, cast5	pgp_sym_encrypt, pgp_pub_encrypt
compress-algo	使用的压缩算法。	0	<ul style="list-style-type: none"> <li>● 0，表示不压缩。</li> <li>● 1，表示ZIP压缩。</li> <li>● 2，表示ZLIB压缩（= ZIP加上元数据和CRC）</li> </ul>	pgp_sym_encrypt, pgp_pub_encrypt
compress-level	压缩级别。级别越高压缩得越小但速度也越慢。0表示不压缩。	6	0, 1-9	pgp_sym_encrypt, pgp_pub_encrypt
convert-crlf	加密时是否将\n转换成\r\n并且解密时执行相反的转换是否把\r\n转换成\n。RFC4880指定文本数据存储时需要使用\r\n作为换行符。	0	0, 1	pgp_sym_encrypt, pgp_pub_encrypt, pgp_sym_decrypt, pgp_pub_decrypt



选项	描述	默认值	取值	适用函数
disable-mdc	不用 SHA-1 保护数据。仅用于兼容老旧的PGP产品。	0	0, 1	pgp_sym_encrypt, pgp_pub_encrypt
sess-key	使用单独的会话密钥。公钥加密总是使用一个单独的会话密钥。该选项用于对称密钥加密，因为对称密钥加密默认直接使用 S2K 密钥。	0	0, 1	pgp_sym_encrypt
s2k-mode	使用的S2K 算法。	3	<ul style="list-style-type: none"> <li>0，表示不使用 salt。不推荐！</li> <li>1，表示使用 salt，但是迭代固定次数。</li> <li>3，可变的迭代计数。</li> </ul>	pgp_sym_encrypt
s2k-count	S2K 算法的迭代次数。	65536 和 253952 之间的一个随机数值	大于等于1024并且小于等于65011712	pgp_sym_encrypt, 并且 s2k-mode=3
s2k-digest-algo	S2K计算时的摘要算法。	sha1	md5, sha1	pgp_sym_encrypt
s2k-cipher-algo	加密单独会话密钥时使用的密码。	默认使用 cipher-algo 的算法	bf, aes, aes128, aes192, aes256	pgp_sym_encrypt

选项	描述	默认值	取值	适用函数
unicode-mode	是否将文本数据在数据库内部编码和UTF-8之间来回转换。如果当前数据库已经是UTF-8，不会执行转换，但是消息将被标记为UTF-8。没有指定该选项就不会被标记	0	0, 1	pgp_sym_encrypt, pgp_pub_encrypt

## 原始加密函数

原始加密函数仅仅会对数据运行一次加密，不支持PGP加密的任何高级功能，因此存在以下问题：

- 直接将用户密钥作为加密密钥。
- 不提供任何完整性检查来校验加密后的数据是否被修改。
- 需要用户自己关联所有加密参数，包括初始值（IV）。
- 不支持处理文本数据。

因此，在引入了PGP加密后，不建议使用这些原始加密函数。

```
encrypt(data bytea, key bytea, type text) returns bytea
decrypt(data bytea, key bytea, type text) returns bytea
encrypt_iv(data bytea, key bytea, iv bytea, type text) returns bytea
decrypt_iv(data bytea, key bytea, iv bytea, type text) returns bytea
```

其中，data是需要加密的数据；type用于指定加密/解密方法。type参数的语法如下：

```
algorithm [ - mode ] [ /pad: padding ]
```

其中algorithm的取值范围如下：

- bf, Blowfish算法。包括近义词：BF, BF-CBC; BLOWFISH, BF-CBC; BLOWFISH-CBC, BF-CBC; BLOWFISH-ECB, BF-ECB; BLOWFISH-CFB, BF-CFB。
- aes, AES算法(Rijndael-128, -192或-256)。包括近义词：AES, AES-CBC; RIJNDAEL, AES-CBC; RIJNDAEL, AES-CBC; RIJNDAEL-CBC, AES-CBC; RIJNDAEL-ECB, AES-ECB。
- DES算法。包括近义词：DES, DES-CBC; 3DES, DES3-CBC; 3DES-ECB, DES3-ECB; 3DES-CBC, DES3-CBC
- sm4, SM4算法。包括近义词：SM4-CBC
- CAST5算法。包括近义词：CAST5-CBC

mode的可能取值范围如下：

- cbc, 下一个块依赖前一个块（默认值）
- ecb, 每个块独立加密（不推荐，仅用于测试）

padding的取值范围如下：

- pkcs, 数据可以是任意长度（默认值）
- none, 数据长度必须是密码块大小的倍数

例如，以下函数的加密结果相同：

```
encrypt(data, 'fooz', 'bf')  
encrypt(data, 'fooz', 'bf-cbc/pad:pkcs')
```

对于函数encrypt\_iv和decrypt\_iv，参数iv表示CBC模式的初始值，ECB模式会忽略该参数。如果它的长度不是准确的块大小，可能会被截断或者使用0进行填充。对于没有该参数的两个函数，默认全部使用0填充。

## 随机数据函数

- gen\_random\_bytes()函数用于生成具有强加密性的随机字节。  
gen\_random\_bytes(count integer) returns bytea

其中，count表示返回的字节数，取值为1到1024。

示例：

```
SELECT gen_random_bytes(16);  
       gen_random_bytes  
-----  
       \x1f1eddc11153afdde0f9e1229f8f4caf  
(1 row)
```

- gen\_random\_uuid()函数用于返回一个version 4的随机UUID。

```
SELECT gen_random_uuid();  
       gen_random_uuid  
-----  
       2bd664a2-b760-4859-8af6-8d09ccc5b830
```

# 8 开发设计建议

## 8.1 开发设计建议概述

本开发设计建议约定数据库建模和数据库应用程序开发过程中，应当遵守的设计规范。依据这些规范进行建模，能够更好的契合GaussDB(DWS)的分布式处理架构，输出更高效的业务SQL代码。

本开发设计建议中所陈述的“建议”和“关注”含义如下：

- **建议：**用户应当遵守的设计规则。遵守这些规则，能够保证业务的高效运行；违反这些规则，将导致业务性能的大幅下降或某些业务逻辑错误。
- **关注：**在业务开发过程中客户需要注意的细则。用于标识容易导致客户理解错误的知识点（实际上遵守SQL标准的SQL行为），或者程序中潜在的客户不易感知的默认行为。

## 8.2 数据库对象命名

数据库对象命名需要满足约束：长度不超过63个字符，以字母或下划线开头，中间字符可以是字母、数字、下划线、\$、#。

- 【建议】避免使用保留或者非保留关键字命名数据库对象。

### 📖 说明

可以使用select \* from pg\_get\_keywords()查询GaussDB(DWS)的关键字，或者在关键字章节中查看。

- 【建议】避免使用双引号括起来的字符串来定义数据库对象名称，除非需要限制数据库对象名称的大小写。数据库对象名称大小写敏感会使定位问题难度增加。
- 【建议】数据库对象命名风格务必保持一致。
  - 增量开发的业务系统或进行业务迁移的系统，建议遵守历史的命名风格。
  - 数据库对象名称由字母、数字和下划线组成，并且不能由数字开头。建议使用多个单词组成，以下划线分割。
  - 数据库对象名称最好能够望文知意，尽量避免使用自定义缩写（可以使用通用的术语缩写进行命名）。例如，在命名中可以使用具有实际业务含义的英文词汇或汉语拼音，但规则应该在集群范围内保持一致。

- 变量名的关键是要具有描述性，即变量名称要有一定的意义，变量名要有前缀标明该变量的类型。
- 【建议】表对象的命名应该可以表征该表的重要特征。例如，在表对象命名时区分该表是普通表、临时表还是非日志表：
  - 普通表名按照数据集的业务含义命名。
  - 临时表以“tmp\_+后缀”命名。
  - 非日志表以“ul\_+后缀”命名。
  - 外表以“f\_+后缀”命名。

## 8.3 数据库对象设计

### 8.3.1 Database 和 Schema 设计

GaussDB(DWS)中可以使用Database和Schema实现业务的隔离，区别在于Database的隔离更加彻底，各个Database之间共享资源极少，可实现连接隔离、权限隔离等，Database之间无法直接互访。Schema隔离的方式共用资源较多，可以通过grant与revoke语法便捷地控制不同用户对各Schema及其下属对象的权限。

- 从便捷性和资源共享效率上考虑，推荐使用Schema进行业务隔离。
- 建议系统管理员创建Schema和Database，再赋予相关用户对应的权限。

#### Database 设计建议

- 【建议】在实际业务中，根据需要创建新的Database，不建议直接使用集群默认的gaussdb数据库。
- 【建议】一个集群内，用户自定义的Database数量建议不超过3个。
- 【建议】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Database的时候使用UTF-8编码。
- 【关注】创建Database时，需要重点关注字符集编码(ENCODING)和兼容性(DBCOMPATIBILITY)两个配置项。GaussDB(DWS)支持Oracle、Teradata和MySQL三种兼容模式，分别兼容Oracle、Teradata和MySQL语法，不同兼容模式下的语法行为可能有一些差异。详细内容可参考[Oracle、Teradata和MySQL语法兼容性差异](#)。
- 【关注】Database的owner默认拥有该Database下所有对象的所有权限，包括删除权限。删除权限影响较大，请谨慎使用。

#### Schema 设计建议

- 【关注】如果该用户不具有sysadmin权限或者不是该Schema的owner，要访问Schema下的对象，需要同时给用户赋予Schema的usage权限和对象的相应权限。
- 【关注】如果要在Schema下创建对象，需要授予操作用户该Schema的create权限。
- 【关注】Schema的owner默认拥有该Schema下对象的所有权限，包括删除权限。删除权限影响较大，请谨慎使用。

## 8.3.2 表设计

GaussDB(DWS)是分布式架构。数据分布在各个DN上。总体上讲，良好的表设计需要遵循以下原则：

- 【关注】将表数据均匀分布在各个DN上。数据均匀分布，可以防止数据在部分DN上集中分布，从而导致因存储倾斜造成集群有效容量下降。通过选择合适的分布列，可以避免数据倾斜。
- 【关注】将表的扫描压力均匀分散在各个DN上。避免扫描压力集中在部分DN上，而导致性能瓶颈。例如，在事实表上使用等值过滤条件时，将会导致扫描压力不均匀。
- 【关注】减少需要扫描的数据量。通过分区表的剪枝机制可以大幅减少数据的扫描量。
- 【关注】尽量减少随机I/O。通过聚簇/局部聚簇可以实现热数据的连续存储，将随机I/O转换为连续I/O，从而减少扫描的I/O代价。
- 【关注】尽量避免数据shuffle。shuffle，是指在物理上，数据从一个节点，传输到另一个节点。shuffle占用了大量宝贵的网络资源，减小不必要的数据shuffle，可以减少网络压力，使数据的处理本地化，提高集群的性能和可支持的并发度。通过对关联条件和分组条件的仔细设计，能够尽可能的减少不必要的数据shuffle。

### 选择存储方案

【建议】表的存储类型是表定义设计的第一步，客户业务类型是决定表的存储类型的主要因素，表存储类型的选择依据请参考[表8-1](#)。

表 8-1 表的存储类型及场景

存储类型	适用场景
行存	<ul style="list-style-type: none"> <li>• 点查询(返回记录少，基于索引的简单查询)。</li> <li>• 增、删、改操作较多的场景。</li> </ul>
列存	<ul style="list-style-type: none"> <li>• 统计分析类查询(关联、分组操作较多的场景)。</li> <li>• 即席查询(查询条件不确定，行存表扫描难以使用索引)。</li> </ul>

### 选择分布方案

【建议】表的分布方式的选择一般遵循以下原则：

表 8-2 表的分布方式及使用场景

分布方式	描述	适用场景
Hash	表数据通过Hash方式散列到集群中的所有DN上。	数据量较大的事实表。
Replication	集群中每一个DN都有一份全量表数据。	维度表、数据量较小的事实表。

分布方式	描述	适用场景
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。	数据量较大的事实表，且使用Hash分布时找不到合适的分布列。

## 选择分区方案

当表中的数据量很大时，应当对表进行分区，一般需要遵循以下原则：

- 【建议】使用具有明显区间性的字段进行分区，比如日期、区域等字段上建立分区。
- 【建议】分区名称应当体现分区的数据特征。例如，关键字+区间特征。
- 【建议】将分区上边界的分区值定义为MAXVALUE，以防止可能出现的数据溢出。

典型的分区表定义如下：

```
CREATE TABLE staffs_p1
(
  staff_ID      NUMBER(6) not null,
  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(4,2),
  MANAGER_ID   NUMBER(6),
  section_ID   NUMBER(4)
)
PARTITION BY RANGE (HIRE_DATE)
(
  PARTITION HIRE_19950501 VALUES LESS THAN ('1995-05-01 00:00:00'),
  PARTITION HIRE_19950502 VALUES LESS THAN ('1995-05-02 00:00:00'),
  PARTITION HIRE_maxvalue VALUES LESS THAN (MAXVALUE)
);
```

## 选择分布键

Hash表的分布键选取至关重要，如果分布键选择不当，可能会导致数据倾斜，从而导致查询时，I/O负载集中在部分DN上，影响整体查询性能。因此，在确定Hash表的分布策略之后，需要对表数据进行倾斜性检查，以确保数据的均匀分布。分布键的选择一般需要遵循以下原则：

- 【建议】选作分布键的字段取值应该比较离散，以便数据能在各个DN上均匀分布。当单个字段无法满足离散条件时，可以考虑使用多个字段一起作为分布键。一般情况下，可以考虑选择表的主键作为分布键。例如，在人员信息表中选择证件号码作为分布键。
- 【建议】在满足第一条原则的情况下，尽量不要选取在查询中存在常量过滤条件的字段作为分布键。例如，在表dwckj相关的查询中，字段zqdh存在常量过滤条件“zqdh='000001'”，那么就应当尽量不选择zqdh字段做为分布键。
- 【建议】在满足前两条原则的情况，尽量选择查询中的关联条件为分布键。当关联条件作为分布键时，Join任务的相关数据都分布在DN本地，将极大减少DN之间的数据流动代价。

### 8.3.3 字段设计

#### 选择数据类型

在字段设计时，基于查询效率的考虑，一般遵循以下原则：

- 【建议】尽量使用高效数据类型。  
选择数值类型时，在满足业务精度的情况下，选择数据类型的优先级从高到低依次为整数、浮点数、NUMERIC。
- 【建议】当多个表存在逻辑关系时，表示同一含义的字段应该使用相同的数据类型。
- 【建议】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)、BPCHAR(n)、NCHAR(n)、CHARACTER(n)。  
关于字符串类型的详细说明，请参见[常用字符串类型介绍](#)。

#### 常用字符串类型介绍

在进行字段设计时，需要根据数据特征选择相应的数据类型。字符串类型在使用时比较容易混淆，下表列出了GaussDB(DWS)中常见的字符串类型：

表 8-3 常用字符串类型

名称	描述	最大存储空间
CHAR(n)	定长字符串，n描述了存储的字节长度，如果输入的字符串字节格式小于n，那么后面会自动用空字符补齐至n个字节。	10MB
CHARACTER(n)	定长字符串，n描述了存储的字节长度，如果输入的字符串字节格式小于n，那么后面会自动用空字符补齐至n个字节。	10MB
NCHAR(n)	定长字符串，n描述了存储的字节长度，如果输入的字符串字节格式小于n，那么后面会自动用空字符补齐至n个字节。	10MB
BPCHAR(n)	定长字符串，n描述了存储的字节长度，如果输入的字符串字节格式小于n，那么后面会自动用空字符补齐至n个字节。	10MB
VARCHAR(n)	变长字符串，n描述了可以存储的最大字节长度。	10MB
CHARACTER VARYING(n)	变长字符串，n描述了可以存储的最大字节长度；此数据类型和VARCHAR(n)是同一数据类型的不同表达形式。	10MB



名称	描述	最大存储空间
VARCHAR2(n)	变长字符串，n描述了可以存储的最大字节长度，此数据类型是为兼容Oracle类型新增的，行为和VARCHAR(n)一致。	10MB
NVARCHAR2(n)	变长字符串，n描述了可以存储的最大字符长度。	10MB
TEXT	不限长度(不超过1GB-8203字节)变长字符串。	1GB-8203字节

## 8.3.4 约束设计

### DEFAULT 和 NULL 约束

- 【建议】如果能够从业务层面补全字段值，那么，就不建议使用DEFAULT约束，避免数据加载时产生不符合预期的结果。
- 【建议】给明确不存在NULL值的字段加上NOT NULL约束，优化器会在特定场景下对其进行自动优化。
- 【建议】给可以显式命名的约束显式命名。除了NOT NULL和DEFAULT约束外，其他约束都可以显式命名。

### 局部聚簇

Partial Cluster Key（局部聚簇，简称PCK）是列存表的一种局部聚簇技术，在GaussDB(DWS)中，使用PCK可以通过min/max稀疏索引实现事实表快速过滤扫描。PCK的选取遵循以下原则：

- 【关注】一张表上只能建立一个PCK，一个PCK可以包含多列，但是一般不建议超过2列。
- 【建议】在查询中的简单表达式过滤条件上创建PCK。这种过滤条件一般形如col op const，其中col为列名，op为操作符 =、>、>=、<=、<，const为常量值。
- 【建议】在满足上面条件的前提下，选择distinct值比较少的列上建PCK。

### 唯一约束

- 【关注】行存表与列存表都支持唯一约束。
- 【建议】从命名上明确标识唯一约束，例如，命名为“UNI+构成字段”。

### 主键约束

- 【关注】行存表与列存表都支持主键约束。
- 【建议】从命名上明确标识主键约束，例如，将主键约束命名为“PK+字段名”。

## 检查约束

- 【关注】行存表支持检查约束，而列存表不支持。
- 【建议】从命名上明确标识检查约束，例如，将检查约束命名为“CK+字段名”。

## 8.3.5 视图和关联表设计

### 视图设计

- 【建议】除非视图之间存在强依赖关系，否则不建议视图嵌套。
- 【建议】视图定义中尽量避免排序操作。

### 关联表设计

- 【建议】表之间的关联字段应该尽量少。
- 【建议】关联字段的数据类型应该保持一致。
- 【建议】关联字段在命名上，应该可以明显体现出关联关系。例如，采用同样名称来命名。

## 8.4 JDBC 配置

目前，GaussDB(DWS)相关的第三方工具都是通过JDBC进行连接的，此部分将介绍工具配置时的注意事项。

### 连接参数

- 【关注】第三方工具通过JDBC连接GaussDB(DWS)时，JDBC向GaussDB(DWS)发起连接请求，会默认添加以下配置参数，详见JDBC代码ConnectionFactoryImpl类的实现。

```
params = {  
  { "user", user },  
  { "database", database },  
  { "client_encoding", "UTF8" },  
  { "DateStyle", "ISO" },  
  { "extra_float_digits", "2" },  
  { "TimeZone", createPostgresTimeZone() },  
};
```

这些参数可能会导致JDBC客户端的行为与gsq客户端的行为不一致，例如，Date数据显示方式、浮点数精度表示、timezone显示。

如果实际期望和这些配置不符，建议在java连接设置代码中显式设定这些参数。

- 【建议】通过JDBC连接数据库时，应该保证下面两个时区设置一致：
  - JDBC客户端所在主机的时区。
  - GaussDB(DWS)集群所在主机的时区。

### fetchsize

【关注】在应用程序中，如果需要使用fetchsize，必须关闭autocommit。开启autocommit，会令fetchsize配置失效。

## autocommit

【建议】在JDBC向GaussDB(DWS)申请连接的代码中，建议显式打开autocommit开关。如果基于性能或者其他方面考虑，需要关闭autocommit时，需要应用程序自己来保证事务的提交。例如，在指定的业务SQL执行完之后做显式提交，特别是客户端退出之前务必保证所有的事务已经提交。

## 释放连接

【建议】推荐使用连接池限制应用程序的连接数。每执行一条SQL就连接一次数据库，是一种不好SQL的编写习惯。

【建议】在应用程序完成作业任务之后，应当及时断开和GaussDB(DWS)的连接，释放资源。建议在任务中设置session超时时间参数。

【建议】使用JDBC连接池，在将连接释放给连接池前，需要执行以下操作，重置会话环境。否则，可能会因为历史会话信息导致的对象冲突。

- 如果在连接中设置了GUC参数，那么在将连接归还连接池之前，必须使用“SET SESSION AUTHORIZATION DEFAULT;RESET ALL;”将连接的状态清空。
- 如果使用了临时表，那么在将连接归还连接池之前，必须将临时表删除。

## CopyManager

【建议】在不使用ETL工具，数据入库实时性要求又比较高的情况下，建议在开发应用程序时，使用GaussDB(DWS) JDBC驱动的CopyManager接口进行微批导入。

CopyManager的使用方法请参见[CopyManager](#)。

# 8.5 SQL 编写

## DDL

- 【建议】在GaussDB(DWS)中，建议DDL（建表、comments等）操作统一执行，在批处理作业中尽量避免DDL操作。避免大量并发事务对性能的影响。
- 【建议】在非日志表（unlogged table）使用完后，立即执行数据清理（truncate）操作。因为在异常场景下，GaussDB(DWS)不保证非日志表（unlogged table）数据的安全性。
- 【建议】临时表和非日志表的存储方式建议和基表相同。当基表为行存（列存）表时，临时表和非日志表也推荐创建为行存（列存）表，可以避免行列混合关联带来的高计算代价。
- 【建议】索引字段的总长度不超过50字节。否则，索引大小会膨胀比较严重，带来较大的存储开销，同时索引性能也会下降。
- 【建议】不要使用DROP...CASCADE方式删除对象，除非已经明确对象间的依赖关系，以免误删。

## 数据加载和卸载

- 【建议】在insert语句中显式给出插入的字段列表。例如：

```
INSERT INTO task(name,id,comment) VALUES ('task1','100','第100个任务');
```
- 【建议】在批量数据入库之后，或者数据增量达到一定阈值后，建议对表进行analyze操作，防止统计信息不准确而导致的执行计划劣化。

- 【建议】如果要清理表中的所有数据，建议使用truncate table方式，不要使用delete table方式。delete table方式删除性能差，且不会释放那些已经删除了的数据占用的磁盘空间。

## 类型转换

- 【建议】在需要数据类型转换（不同数据类型进行比较或转换）时，使用强制类型转换，以防隐式类型转换结果与预期不符。
- 【建议】在查询中，对常量要显式指定数据类型，不要试图依赖任何隐式的数据类型转换。
- 【关注】在ORACLE兼容模式下，在导入数据时，空字符串会自动转化为NULL。如果需要保留空字符串需要新建兼容性为TD的数据库。

## 查询操作

- 【建议】除ETL程序外，应该尽量避免向客户端返回大量结果集的操作。如果结果集过大，应考虑业务设计是否合理。
- 【建议】使用事务方式执行DDL和DML操作。例如，truncate table、update table、delete table、drop table等操作，一旦执行提交就无法恢复。对于这类操作，建议使用事务进行封装，必要时可以进行回滚。
- 【建议】在查询编写时，建议明确列出查询涉及的所有字段，不建议使用“SELECT \*”这种写法。一方面基于性能考虑，尽量减少查询输出列；另一方面避免增删字段对前端业务兼容性的影响。
- 【建议】在访问表对象时带上schema前缀，可以避免因schema切换导致访问到非预期的表。
- 【建议】超过3张表或视图进行关联（特别是full join）时，执行代价难以估算。建议使用WITH TABLE AS语句创建中间临时表的方式增加SQL语句的可读性。
- 【建议】尽量避免使用笛卡尔积和Full join。这些操作会造成结果集的急剧膨胀，同时其执行性能也很低。
- 【关注】NULL值的比较只能使用IS NULL或者IS NOT NULL的方式判断，其他任何形式的逻辑判断都返回NULL。例如：NULL<>NULL、NULL=NULL和NULL<>1返回结果都是NULL，而不是期望的布尔值。
- 【关注】需要统计表中所有记录数时，不要使用count(col)来替代count(\*)。count(\*)会统计NULL值（真实行数），而count(col)不会统计。
- 【关注】在执行count(col)时，将“值为NULL”的记录行计数为0。在执行sum(col)时，当所有记录都为NULL时，最终将返回NULL；当不全为NULL时，“值为NULL”的记录行将被计数为0。
- 【关注】count(多个字段)时，多个字段名必须用圆括号括起来。例如，count( col1,col2,col3 )。注意：通过多字段统计行数时，即使所选字段都为NULL，该行也被计数，效果与count(\*)一致。
- 【关注】count(distinct col)用来计算该列不重复的非NULL的数量，NULL将不被计数。
- 【关注】count(distinct (col1,col2,...))用来统计多列的唯一值数量，当所有统计字段都为NULL时，也会被计数，同时这些记录被认为是相同的。
- 【关注】通过常量来过滤数据时，会根据常量的数据类型和匹配列的数据类型来查找用于这两种数据类型计算的函数，如果找不到对应的函数，则会相应的进行隐式数据类型转化，然后再根据转化后的数据类型查找用于转化后的数据类型计算的函数。

```
SELECT * FROM test WHERE timestamp_col = 20000101;
```

上述例子中，假设timestamp\_col是timestamp类型，则会先查找支持timestamp类型和int类型（常量数字认为是int类型）“等于”运算的函数，如果找不到，则把timestamp\_col和常量数字隐式类型转化成text类型来计算。

- 【建议】尽量避免标量子查询语句的出现。标量子查询是出现在select语句输出列表中的子查询，在下面例子中，括号内部分即为一个标量子查询语句：

```
SELECT id, (SELECT COUNT(*) FROM films f WHERE f.did = s.id) FROM staffs_p1 s;
```

标量子查询往往会导致查询性能急剧劣化，在应用开发过程中，应当根据业务逻辑，对标量子查询进行等价转换，将其写为表关联。

- 【建议】在where子句中，应当对过滤条件进行排序，把选择读较小（筛选出的记录数较少）的条件排在前面。
- 【建议】where子句中的过滤条件，尽量符合单边规则。即把字段名放在比较条件的一边，优化器在某些场景下会自动进行剪枝优化。形如col op expression，其中col为表的一个列，op为‘=’、‘>’的等比较操作符，expression为不含列名的表达式。例如，

```
SELECT id, from_image_id, from_person_id, from_video_id FROM face_data WHERE current_timestamp(6) - time < '1 days'::interval;
```

改写为：

```
SELECT id, from_image_id, from_person_id, from_video_id FROM face_data where time > current_timestamp(6) - '1 days'::interval;
```

- 【建议】尽量避免不必要的排序操作。排序需要耗费大量的内存及CPU，如果业务逻辑许可，可以组合使用order by和limit，减小资源开销。GaussDB(DWS)默认按照ASC & NULL LAST进行排序。
- 【建议】使用ORDER BY子句进行排序时，显式指定排序方式（ASC/DESC），NULL的排序方式（NULL FIRST/NULL LAST）。
- 【建议】不要单独依赖limit子句返回特定顺序的结果集。如果部分特定结果集，可以将ORDER BY子句与Limit子句组合使用，必要时也可以使用offset跳过特定结果。
- 【建议】在保障业务逻辑准确的情况下，建议尽量使用UNION ALL来代替UNION。
- 【建议】如果过滤条件只有OR表达式，可以将OR表达式转化为UNION ALL以提升性能。使用OR的SQL语句经常无法优化，导致执行速度慢。例如，将下面语句

```
SELECT * FROM scdc.pub_menu WHERE (cdp= 300 AND inline=301) OR (cdp= 301 AND inline=302) OR (cdp= 302 AND inline=301);
```

转换为：

```
SELECT * FROM scdc.pub_menu WHERE (cdp= 300 AND inline=301) union all SELECT * FROM scdc.pub_menu WHERE (cdp= 301 AND inline=302) union all SELECT * FROM tablename WHERE (cdp= 302 AND inline=301);
```

- 【建议】当in(val1, val2, val3...)表达式中字段较多时，建议使用in (values (va1), (val2), (val3)...)语句进行替换。优化器会自动把in约束转换为非关联子查询，从而提升查询性能。
- 【建议】在关联字段不存在NULL值的情况下，使用(not) exist代替(not) in。例如，在下面查询语句中，当T1.C1列不存在NULL值时，可以先为T1.C1字段添加NOT NULL约束，再进行如下改写。

```
SELECT * FROM T1 WHERE T1.C1 NOT IN (SELECT T2.C2 FROM T2);
```

可以改写为：

```
SELECT * FROM T1 WHERE NOT EXISTS (SELECT * FROM T1,T2 WHERE T1.C1=T2.C2);
```

 说明

- 如果不能保证T1.C1列的值为NOT NULL的情况下，就不能进行上述改写。
- 如果T1.C1为子查询的输出，要根据业务逻辑确认其输出是否为NOT NULL。
- 【建议】通过游标进行翻页查询，而不是使用LIMIT OFFSET语法，避免多次执行带来的资源开销。游标必须在事务中使用，执行完后务必关闭游标并提交事务。

## 8.6 自定义外部函数(pgSQL/Java)使用

- 【关注】Java UDF可以实现一些java逻辑计算，禁止在Java UDF中封装业务。
- 【关注】禁止在Java函数中使用任何方式连接数据库，包括但不限于JDBC。
- 【关注】只能选择下表中的数据类型，不支持自定义类型、复杂数据类型（Java Array类及派生类）等：
- 【关注】不支持UDAF（用户定义聚合函数），UDTF（用户自定义表生成函数）。

表 8-4 PL/Java 默认数据类型映射关系

GaussDB(DWS)	Java
BOOLEAN	boolean
"char"	byte
bytea	byte[]
SMALLINT	short
INTEGER	int
BIGINT	long
FLOAT4	float
FLOAT8	double
CHAR	java.lang.String
VARCHAR	java.lang.String
TEXT	java.lang.String
name	java.lang.String
DATE	java.sql.Timestamp
TIME	java.sql.Time (stored value treated as local time)
TIMETZ	java.sql.Time
TIMESTAMP	java.sql.Timestamp
TIMESTAMPTZ	java.sql.Timestamp

## 8.7 PL/pgSQL 使用

### 总体开发原则

1. 应完全按照设计文档进行开发。
2. 程序模块应做到高内聚低耦合。
3. 应有正确、全面的故障对策。
4. 程序编写应做到结构合理，条理清晰。
5. 程序名称命名应按照统一的命名规则进行命名。
6. 应充分考虑程序的运行效率，包括程序的执行效率和数据库的查询、存储效率，在保证应用的同时应使用效率高的处理方法。
7. 程序注释应详细、正确、规范。
8. 除非应用特别需要控制commit和rollback的提交时机，否则应在存储过程结束时执行显式的commit或者rollback操作。
9. 程序处理应支持7\*24小时；对于中断，应用程序应提供安全、简单的断点再续处理。
10. 应提供标准、简单的应用输出，为应用维护人员提供明确的进度显示、错误描述和运行结果；为业务人员提供明确、直观的报表、凭证输出。

### 程序编写原则

1. 在PL/PGSQL中的SQL语句宜使用绑定变量。
2. 在PL/PGSQL中的SQL语句宜使用RETURNING字句。
3. 存储过程使用原则：
  - a. 对于单个存储过程中Varchar或者Varchar2类型输出参数个数不应超过50个。
  - b. 不应使用long类型作为输入或输出参数。
  - c. 对于大小超过10MB的字符串类型输出，应使用CLOB类型。
4. 变量声明原则：
  - a. 变量声明时，如果含义和应用表某字段含义或某变量相同时，应使用%TYPE声明。
  - b. 记录声明时，如果含义和某应用表行数据相同时，应使用%ROWTYPE声明。
  - c. 变量声明每行应只包含一条语句。
  - d. 不应声明LONG类型的变量。
5. 游标使用类型：
  - a. 显式游标使用后应关闭。
  - b. 游标变量使用后应关闭，若游标变量需要传递数据给调用的应用程序，应在应用程序中进行游标关闭处理；若游标变量仅在存储过程中使用，应显式关闭游标。
  - c. 在使用DBMS\_SQL.CLOSE\_CURSOR关闭游标前，应使用DBMS\_SQL.IS\_OPEN判断游标是否已打开。
6. 集合使用原则：
  - a. 引用集合中的元素时宜使用FORALL语句，不宜使用FOR循环语句。

7. 动态语句使用原则：
  - a. 联机系统的交易程序不宜使用动态SQL。
  - b. PL/PGSQL中要实现DDL语句和系统控制命令，可使用动态SQL。
  - c. 宜尽量使用变量绑定。
8. 拼装SQL的使用原则：
  - a. 拼装SQL宜使用绑定变量。
  - b. 拼装SQL语句的条件如果有外部输入源，应对输入条件进行字符检查，防止攻击。
  - c. 在PL/PGSQL脚本中，单行代码的长度，不宜超过2499字符。
9. Trigger使用原则：
  - a. Trigger可用于实现增量数据日志等于业务处理无关的可用性设计场景。
  - b. 不应使用Trigger实现业务处理相关功能。

## 异常处理原则

任何在PL/pgSQL函数中发生的错误会中止该函数的执行，而且实际上会中止其周围的事务。你可以使用一个带有EXCEPTION子句的BEGIN块俘获错误并且从中恢复。

1. 在使用PL/PGSQL块中，如果使用了不能返回确定结果的SQL语句，宜在EXCEPTION中对程序可能出现的异常进行处理，避免出现未处理的出错被传递到外层块，导致程序逻辑错误。
2. 对于系统已经定义了了的异常，可以直接使用。DWS暂不支持自定义异常。
3. 进入和退出一个包含EXCEPTION子句的块要比不包含的块开销大的多。因此，非必要场景不应使用EXCEPTION。

## 书写规范

1. 变量命名规则：
  - a. 过程、函数的输入参数格式宜为：IN\_参数名，参数名宜使用大写。
  - b. 过程、函数的输出参数格式宜为：OUT\_参数名，参数名宜使用大写。
  - c. 过程、函数得输入输出参数格式宜为：IO\_参数名，参数名宜使用大写。
  - d. 过程、函数得程序中用到的变量宜由v\_变量名组成，变量名宜使用小写。
  - e. 将查询语句做成字符串拼接时，where语句的拼接变量名宜统一为v\_where，select语句的拼接变量名宜为v\_select。
  - f. 记录（RECORD）的类型（TYPE）命名宜由T+变量名组成，名称宜使用大写。
  - g. 游标命名宜由CUR+变量名组成，名称宜使用大写。
  - h. 引用游标（REF CURSOR）的命名宜由REF+变量名组成，名称宜使用大写。
2. 变量类型定义：
  - a. 变量类型声明时，如果其含义和应用表某字段含义相同时，应使用%TYPE声明。
  - b. 记录类型声明时，如果其含义和某应用表行数据相同时，应使用%ROWTYPE声明。
3. 注释规范：



- a. 注释应该是有意义的，而不应是重述代码。
  - b. 注释应简洁、易懂，以中文为主。为了表达准确，名词或操作上也可以使用英文。
  - c. 应在每个存储过程、函数得开始加入注释，内容应包括：本程序的简要功能描述、编写者、编写日期、程序版本号信息和程序变更信息，而且各存储过程开头注释应保持统一格式。
  - d. 应在输入输出参数的旁边添加注释，注明变量的意义。
  - e. 每个块或大分支的开始宜添加注释，描述块的简要功能，若使用算法，宜添加注释简单描述算法的目的和结果。
4. 变量声明格式：  
每行应只包含一条语句，如同时需要赋初始值，应在同一行书写。
  5. 大小写规范：  
除了变量名，应一律使用大写。
  6. 缩进规范：  
创建存储过程语句中，同一层的CREATE、AS/IS、BEGIN、END这几个关键字应位于同一列，其他部分依次缩进。
  7. 语句详述：
    - a. 变量定义语句。每行应只包含一条语句。
    - b. 同一层的IF、ELSEIF、ELSE和END关键字应开始于同一列，执行语句缩进。
    - c. CASE和END关键字应位于同一列，WHEN和ELSE关键字应缩进。
    - d. 同一层的LOOP和END LOOP关键字应位于同一列，层内语句或嵌套应依次缩进。

# 9 教程：使用 JDBC 或 ODBC 开发

## 9.1 开发规范

如果用户在APP的开发中，使用了连接池机制，那么需要遵循如下规范：

- 如果在连接中设置了GUC参数，那么在将连接归还连接池之前，必须使用“SET SESSION AUTHORIZATION DEFAULT;RESET ALL;”将连接的状态清空。
- 如果使用了临时表，那么在将连接归还连接池之前，必须将临时表删除。

否则，连接池里面的连接就是有状态的，会对用户后续使用连接池进行操作的正确性带来影响。

## 9.2 驱动下载

请参见[下载JDBC或ODBC驱动](#)。

## 9.3 基于 JDBC 开发

JDBC (Java Database Connectivity, java数据库连接) 是一种用于执行SQL语句的Java API, 可以为多种关系数据库提供统一访问接口, 应用程序可基于它操作数据。GaussDB(DWS)库提供了对JDBC 4.0特性的支持, 需要使用JDK1.6及以上版本编译程序代码, 不支持JDBC桥接ODBC方式。

### 9.3.1 JDBC 包与驱动类

#### JDBC 包

从管理控制台下载包名为dws\_8.1.x\_jdbc\_driver.zip, 具体下载方法请参见[驱动下载](#)。

解压后有两个JDBC的驱动jar包:

- gsjdbc4.jar: 与PostgreSQL保持兼容的驱动包, 其中类名、类结构与PostgreSQL驱动完全一致, 曾经运行于PostgreSQL的应用程序可以直接移植到当前系统使用。
- gsjdbc200.jar: 如果同一JVM进程内需要同时访问PostgreSQL及GaussDB(DWS)请使用此驱动包, 它的主类名为“com.huawei.gauss200.jdbc.Driver”(即将

“org.postgresql”替换为“com.huawei.gauss200.jdbc”），数据库连接的URL前缀为“jdbc:gaussdb”，其余与gsjdbc4.jar相同。

## 驱动类

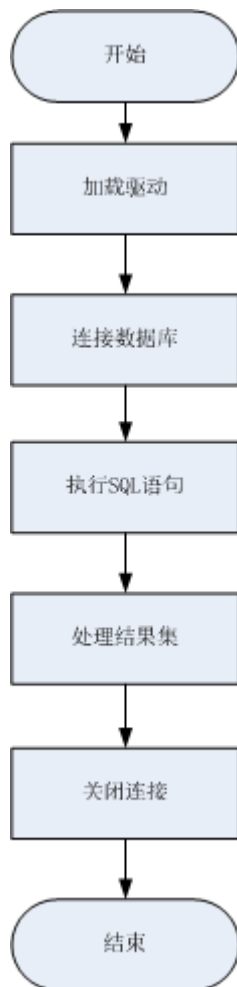
在创建数据库连接之前，需要加载数据库驱动类“org.postgresql.Driver”（对应包gsjdbc4.jar）或者“com.huawei.gauss200.jdbc.Driver”（对应gsjdbc200.jar）。

### 📖 说明

由于GaussDB(DWS)在JDBC的使用上与PG的使用方法保持兼容，所以同时在同一进程内使用两个JDBC的驱动的时候，可能会类名冲突。

## 9.3.2 开发流程

图 9-1 采用 JDBC 开发应用程序的流程



## 9.3.3 加载驱动

在创建数据库连接之前，需要先加载数据库驱动程序。

加载驱动有两种方法：

- 在代码中创建连接之前任意位置隐含装载：  
`Class.forName("org.postgresql.Driver");`
- 在JVM启动时参数传递：`java -Djdbc.drivers=org.postgresql.Driver jdbctest`

#### 📖 说明

- 上述jdbctest为测试用例程序的名称。
- 当使用gsjdbc200.jar时，上面的Driver类名相应修改为  
`"com.huawei.gauss200.jdbc.Driver"`

## 9.3.4 连接数据库

在创建数据库连接之后，才能使用它来执行SQL语句操作数据。

#### 📖 说明

如果您使用的是开源的JDBC驱动程序，应确保数据库参数password\_encryption\_type取值设置为1，如果参数值不为1，可能会出现连接失败，典型的报错信息比如：“none of the server's SASL authentication mechanisms are supported”，参见以下操作：

1. 将数据库参数password\_encryption\_type修改为1，修改方法参见[修改数据库参数](#)。
2. 新建一个数据库用户用于连接，或者重置准备使用的数据库用户的密码。
  - 如果您使用的是管理员账号，参见[重置密码](#)。
  - 如果是普通用户，可以先通过其他客户端工具（例如Data Studio）连接数据库后，使用ALTER USER语句来修改密码。
3. 再尝试连接数据库。

需要执行以上操作的原因：

- 调整参数的原因：当前MD5算法已被证实可以人工碰撞，已严禁将之用于密码校验算法。GaussDB(DWS) 采用默认安全设计，默认禁止MD5算法的密码校验，而PostgreSQL的开源libpq通信协议恰恰使用的是MD5算法。所以需要调整一下密码算法参数password\_encryption\_type，打开MD5算法。
- 修改密码的原因：GaussDB(DWS) 中是不会存储您的密码原文的，而是存储的密码HASH摘要（默认是SHA256摘要），在密码校验时该摘要会与客户端发来的密码摘要进行比对（中间会有加盐操作）。故当您只是单纯调整了密码算法策略时，数据库是无法还原您的密码进而再生成MD5的摘要值的，必须要求您手动修改一次密码或者创建一个新用户，这时新的密码将会采用您设置的HASH算法进行摘要存储，用于下次连接认证。

## 函数原型

JDBC提供了三个方法，用于创建数据库连接。

- `DriverManager.getConnection(String url);`
- `DriverManager.getConnection(String url, Properties info);`
- `DriverManager.getConnection(String url, String user, String password);`

## 参数

表 9-1 数据库连接参数

参数	描述
url	<p>gsjdbc4.jar数据库连接描述符。格式如下：</p> <ul style="list-style-type: none"> <li>• jdbc:postgresql:database</li> <li>• jdbc:postgresql://host/database</li> <li>• jdbc:postgresql://host:port/database</li> <li>• jdbc:postgresql://host:port[,host:port][...]/database</li> </ul> <p><b>说明</b> 使用gsjdbc200.jar时，将“jdbc:postgresql”修改为“jdbc:gaussdb”</p> <ul style="list-style-type: none"> <li>• database为要连接的数据库名称。</li> <li>• host为数据库服务器名称或IP地址。 连接GaussDB(DWS)的机器与GaussDB(DWS)不在同一网段时，host指定的IP地址应为Manager界面上所设的mppdb.coo.cooListenIp2（应用访问IP）的取值。 由于安全原因，数据库CN禁止集群内部其他节点无认证接入。如果要在集群内部访问CN，请将JDBC程序部署在CN所在机器，host使用"127.0.0.1"。否则可能会出现“FATAL: Forbid remote connection with trust method!”错误。 建议业务系统单独部署在集群外部，否则可能会影响数据库运行性能。</li> <li>• port为数据库服务器端口。缺省情况下，会尝试连接到localhost的8000端口的database。</li> <li>• 支持多ip端口配置形式，jdbc自动实现了负载均衡，多ip端口配置形式是采取随机访问+failover的方式，这个过程系统会自动忽略不可达IP。 以","隔开，例如jdbc:postgresql://10.10.0.13:8000,10.10.0.14:8000/database</li> </ul>

参数	描述
info	<p>数据库连接属性。常用的属性如下：</p> <ul style="list-style-type: none"> <li>• user: String类型。表示创建连接的数据库用户。</li> <li>• password: String类型。表示数据库用户的密码。</li> <li>• ssl: Boolean类型。表示是否使用SSL连接。</li> <li>• loggerLevel: string类型。为LogStream或LogWriter设置记录进DriverManager当前值的日志信息量。目前支持"OFF"、"DEBUG"和"TRACE"。值为"DEBUG"时，表示只打印DEBUG级别以上的日志，将记录非常少的信息。值等于TRACE时，表示打印DEBUG和TRACE级别的日志，将产生详细的日志信息。默认值为OFF，表示不打印日志。</li> <li>• prepareThreshold: integer类型。用于确定在转换为服务器端的预备语句之前，要求执行方法PreparedStatement的次数。缺省值是5。</li> <li>• batchSize: boolean类型，用于确定是否使用batch模式连接。</li> <li>• fetchsize: integer类型，用于设置数据库链接所创建statement的默认fetchsize。</li> <li>• ApplicationName: string类型。应用名称，在不做设置时，缺省值为PostgreSQL JDBC Driver。</li> <li>• allowReadOnly:boolean类型，用于设置connection是否允许设置readonly模式，默认为false，若该参数不被设置为true，则执行connection.setReadOnly不生效。</li> <li>• blobMode:string类型，用于设置setBinaryStream方法为不同的数据类型赋值，设置为on时表示为blob数据类型赋值，设置为off时表示为bytea数据类型赋值，默认为on。</li> <li>• connectionExtraInfo: Boolean类型。表示驱动是否上报当前驱动的部署路径、进程属主用户到数据库。</li> </ul> <p><b>说明</b> 取值范围: true或false，默认值为true。设置connectionExtraInfo为true，JDBC驱动会将当前驱动的部署路径、进程属主用户上报到数据库中，记录在connection_info参数（参见<a href="#">connection_info</a>）里；同时可以在<a href="#">PG_STAT_ACTIVITY</a>和<a href="#">PGXC_STAT_ACTIVITY</a>中查询到。</p>
user	数据库用户。
password	数据库用户的密码。

## 示例

//以下用例以gsjdbc4.jar为例，如果要使用gsjdbc200.jar，请替换驱动类名（将代码中的“org.postgresql”替换成“com.huawei.gauss200.jdbc”）与连接URL串前缀（将“jdbc:postgresql”替换为“jdbc:gaussdb”）。  
//以下代码将获取数据库连接操作封装为一个接口，可通过给定用户名和密码来连接数据库。

```
public static Connection GetConnection(String username, String passwd) {
    //驱动类。
    String driver = "org.postgresql.Driver";
    //数据库连接描述符。
    String sourceURL = "jdbc:postgresql://10.10.0.13:8000/postgres?currentSchema=test";
    Connection conn = null;
```

```
try {
    //加载驱动。
    Class.forName(driver);
} catch (ClassNotFoundException e ){
    e.printStackTrace();
    return null;
}

try {
    //创建连接。
    conn = DriverManager.getConnection(sourceURL, username, passwd);
    System.out.println("Connection succeed!");
} catch (SQLException e) {
    e.printStackTrace();
    return null;
}

return conn;
}
```

## 9.3.5 执行 SQL 语句

### 执行普通 SQL 语句

应用程序通过执行SQL语句来操作数据库的数据（不用传递参数的语句），需要按以下步骤执行：

**步骤1** 调用Connection的createStatement方法创建语句对象。

```
Statement stmt = con.createStatement();
```

**步骤2** 调用Statement的executeUpdate方法执行SQL语句。

```
int rc = stmt.executeUpdate("CREATE TABLE customer_t1(c_customer_sk INTEGER, c_customer_name VARCHAR(32));");
```

#### 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，事务块中不支持vacuum操作。如果其中有一个语句失败，那么整个请求都将会被回滚。

**步骤3** 关闭语句对象。

```
stmt.close();
```

----结束

### 执行预编译 SQL 语句

预编译语句是只编译和优化一次，然后通过设置不同的参数值多次使用。由于已经预先编译好，后续使用会减少执行时间。因此，如果多次执行一条语句，请选择使用预编译语句。可以按以下步骤执行：

**步骤1** 调用Connection的prepareStatement方法创建预编译语句对象。

```
PreparedStatement pstmt = con.prepareStatement("UPDATE customer_t1 SET c_customer_name = ? WHERE c_customer_sk = 1");
```

**步骤2** 调用PreparedStatement的setShort设置参数。

```
pstmt.setShort(1, (short)2);
```

**步骤3** 调用PreparedStatement的executeUpdate方法执行预编译SQL语句。

```
int rowcount = pstmt.executeUpdate();
```

**步骤4** 调用PreparedStatement的close方法关闭预编译语句对象。

```
pstmt.close();
```

----结束

## 调用存储过程

GaussDB(DWS)支持通过JDBC直接调用事先创建的存储过程，步骤如下：

**步骤1** 调用Connection的prepareCall方法创建调用语句对象。

```
CallableStatement cstmt = myConn.prepareCall("{? = CALL TESTPROC(?,?,?)}");
```

**步骤2** 调用CallableStatement的setInt方法设置参数。

```
cstmt.setInt(2, 50);  
cstmt.setInt(1, 20);  
cstmt.setInt(3, 90);
```

**步骤3** 调用CallableStatement的registerOutParameter方法注册输出参数。

```
cstmt.registerOutParameter(4, Types.INTEGER); //注册out类型的参数，类型为整型。
```

**步骤4** 调用CallableStatement的execute执行方法调用。

```
cstmt.execute();
```

**步骤5** 调用CallableStatement的getInt方法获取输出参数。

```
int out = cstmt.getInt(4); //获取out参数
```

示例：

```
//在数据库中已创建了如下存储过程，它带有out参数。  
create or replace procedure testproc  
(  
    psv_in1 in integer,  
    psv_in2 in integer,  
    psv_inout in out integer  
)  
as  
begin  
    psv_inout := psv_in1 + psv_in2 + psv_inout;  
end;  
/
```

**步骤6** 调用CallableStatement的close方法关闭调用语句。

```
cstmt.close();
```

### 说明

- 很多的数据库类如Connection、Statement和ResultSet都有close()方法，在使用完对象后应把它们关闭。要注意的是，Connection的关闭将间接关闭所有与它关联的Statement，Statement的关闭间接关闭了ResultSet。
- 一些JDBC驱动程序还提供命名参数的方法来设置参数。命名参数的方法允许根据名称而不是顺序来设置参数，若参数有默认值，则可以不用指定参数值就可以使用此参数的默认值。即使存储过程中参数的顺序发生了变更，也不必修改应用程序。目前GaussDB(DWS)数据库的JDBC驱动程序不支持此方法。
- GaussDB(DWS)数据库不支持带有输出参数的函数，也不支持存储过程和函数参数默认值。

----结束



**须知**

- 当游标作为存储过程的返回值时，如果使用JDBC调用该存储过程，返回的游标将不可用。
- 存储过程不能和普通SQL在同一条语句中执行。

## 执行批处理

用一条预处理语句处理多条相似的数据，数据库只创建一次执行计划，节省了语句的编译和优化时间。可以按如下步骤执行：

**步骤1** 调用Connection的prepareStatement方法创建预编译语句对象。

```
PreparedStatement pstmt = con.prepareStatement("INSERT INTO customer_t1 VALUES (?");
```

**步骤2** 针对每条数据都要调用setShort设置参数，以及调用addBatch确认该条设置完毕。

```
pstmt.setShort(1, (short)2);  
pstmt.addBatch();
```

**步骤3** 调用PreparedStatement的executeBatch方法执行批处理。

```
int[] rowcount = pstmt.executeBatch();
```

**步骤4** 调用PreparedStatement的close方法关闭预编译语句对象。

```
pstmt.close();
```

**说明**

在实际的批处理过程中，通常不终止批处理程序的执行，否则会降低数据库的性能。因此在批处理程序时，应该关闭自动提交功能，每几行提交一次。关闭自动提交功能的语句为：  
conn.setAutoCommit(false);

----结束

## 9.3.6 处理结果集

### 设置结果集类型

不同类型的结果集有各自的应用场景，应用程序需要根据实际情况选择相应的结果集类型。在执行SQL语句过程中，都需要先创建相应的语句对象，而部分创建语句对象的方法提供了设置结果集类型的功能。具体的参数设置如表9-2所示。涉及的Connection的方法如下：

```
//创建一个Statement对象，该对象将生成具有给定类型和并发性的ResultSet对象。  
createStatement(int resultSetType, int resultSetConcurrency);
```

```
//创建一个PreparedStatement对象，该对象将生成具有给定类型和并发性的ResultSet对象。  
prepareStatement(String sql, int resultSetType, int resultSetConcurrency);
```

```
//创建一个CallableStatement对象，该对象将生成具有给定类型和并发性的ResultSet对象。  
prepareCall(String sql, int resultSetType, int resultSetConcurrency);
```

表 9-2 结果集类型

参数	描述
resultSetType	<p>表示结果集的类型，具体有三种类型：</p> <ul style="list-style-type: none"> <li>• ResultSet.TYPE_FORWARD_ONLY: ResultSet只能向前移动。是缺省值。</li> <li>• ResultSet.TYPE_SCROLL_SENSITIVE: 在修改后重新滚动到修改所在行，可以看到修改后的结果。</li> <li>• ResultSet.TYPE_SCROLL_INSENSITIVE: 对可修改例程所做的编辑不进行显示。</li> </ul> <p><b>说明</b> 结果集从数据库中读取了数据之后，即使类型是 ResultSet.TYPE_SCROLL_SENSITIVE，也不会看到由其他事务在这之后引起的改变。调用ResultSet的refreshRow()方法，可进入数据库并从其中取得当前游标所指记录的最新数据。</p>
resultSetConcurrency	<p>表示结果集的并发，具体有两种类型：</p> <ul style="list-style-type: none"> <li>• ResultSet.CONCUR_READ_ONLY: 如果不从结果集中的数据建立一个新的更新语句，不能对结果集中的数据进行更新。</li> <li>• ResultSet.CONCUR_UPDATEABLE: 可改变的结果集。对于可滚动的结果集，可对结果集进行适当的改变。</li> </ul>

## 在结果集中定位

ResultSet对象具有指向其当前数据行的光标。最初，光标被置于第一行之前。next方法将光标移动到下一行；因为该方法在ResultSet对象没有下一行时返回false，所以可以在while循环中使用它来迭代结果集。但对于可滚动的结果集，JDBC驱动程序提供更多的定位方法，使ResultSet指向特定的行。定位方法如表9-3所示。

表 9-3 在结果集中定位的方法

方法	描述
next()	把ResultSet向下移动一行。
previous()	把ResultSet向上移动一行。
beforeFirst()	把ResultSet定位到第一行之前。
afterLast()	把ResultSet定位到最后一行之后。
first()	把ResultSet定位到第一行。
last()	把ResultSet定位到最后一行。
absolute(int)	把ResultSet移动到参数指定的行数。
relative(int)	向前或者向后移动参数指定的行。

## 获取结果集中光标的位置

对于可滚动的结果集，可能会调用定位方法来改变光标的位置。JDBC驱动程序提供了获取结果集中光标所在位置的方法。获取光标位置的方法如表9-4所示。

表 9-4 获取结果集光标的位置

方法	描述
isFirst()	是否在一行。
isLast()	是否在最后一行。
isBeforeFirst()	是否在第一行之前。
isAfterLast()	是否在最后一行之后。
getRow()	获取当前在第几行。

## 获取结果集中的数据

ResultSet对象提供了丰富的方法，以获取结果集中的数据。获取数据常用的方法如表9-5所示，其他方法请参考JDK官方文档。

表 9-5 ResultSet 对象的常用方法

方法	描述
int getInt(int columnIndex)	按列标获取int型数据。
int getInt(String columnLabel)	按列名获取int型数据。
String getString(int columnIndex)	按列标获取String型数据。
String getString(String columnLabel)	按列名获取String型数据。
Date getDate(int columnIndex)	按列标获取Date型数据
Date getDate(String columnLabel)	按列名获取Date型数据。

### 9.3.7 关闭连接

在使用数据库连接完成相应的数据操作后，需要关闭数据库连接。

关闭数据库连接可以直接调用其close方法即可。如：`conn.close()`

### 9.3.8 示例：常用操作

#### 示例 1

在完成以下示例前，需要先创建存储过程。

```
create or replace procedure testproc  
(
```

```
    psv_in1 in integer,
    psv_in2 in integer,
    psv_inout in out integer
)
as
begin
    psv_inout := psv_in1 + psv_in2 + psv_inout;
end;
/
```

此示例将演示如何基于 GaussDB(DWS)提供的 JDBC 接口开发应用程序。

```
//DBtest.java
//以下用例以gsjdbc4.jar为例，如果要使用gsjdbc200.jar，请替换驱动类名（将代码中的“org.postgresql”替换成“com.huawei.gauss200.jdbc”）与连接URL串前缀（将“jdbc:postgresql”替换为“jdbc:gaussdb”）。
//演示基于JDBC开发的主要步骤，会涉及创建数据库、创建表、插入数据等。
```

```
import java.sql.Connection;
import java.sql.DriverManager;
import java.sql.PreparedStatement;
import java.sql.SQLException;
import java.sql.Statement;
import java.sql.CallableStatement;

public class DBTest {

    //创建数据库连接。
    public static Connection GetConnection(String username, String passwd) {
        String driver = "org.postgresql.Driver";
        String sourceURL = "jdbc:postgresql://localhost:gaussdb";
        Connection conn = null;
        try {
            //加载数据库驱动。
            Class.forName(driver).newInstance();
        } catch (Exception e) {
            e.printStackTrace();
            return null;
        }

        try {
            //创建数据库连接。
            conn = DriverManager.getConnection(sourceURL, username, passwd);
            System.out.println("Connection succeed!");
        } catch (Exception e) {
            e.printStackTrace();
            return null;
        }

        return conn;
    };

    //执行普通SQL语句，创建customer_t1表。
    public static void CreateTable(Connection conn) {
        Statement stmt = null;
        try {
            stmt = conn.createStatement();

            //执行普通SQL语句。
            int rc = stmt
                .executeUpdate("CREATE TABLE customer_t1(c_customer_sk INTEGER, c_customer_name
                VARCHAR(32));");

            stmt.close();
        } catch (SQLException e) {
            if (stmt != null) {
                try {
                    stmt.close();
                } catch (SQLException e1) {
                    e1.printStackTrace();
                }
            }
        }
    }
}
```

```
    }
    e.printStackTrace();
  }
}

//执行预处理语句，批量插入数据。
public static void BatchInsertData(Connection conn) {
    PreparedStatement pst = null;

    try {
        //生成预处理语句。
        pst = conn.prepareStatement("INSERT INTO customer_t1 VALUES (?,?)");
        for (int i = 0; i < 3; i++) {
            //添加参数。
            pst.setInt(1, i);
            pst.setString(2, "data " + i);
            pst.addBatch();
        }
        //执行批处理。
        pst.executeBatch();
        pst.close();
    } catch (SQLException e) {
        if (pst != null) {
            try {
                pst.close();
            } catch (SQLException e1) {
                e1.printStackTrace();
            }
        }
        e.printStackTrace();
    }
}

//执行预编译语句，更新数据。
public static void ExecPreparedSQL(Connection conn) {
    PreparedStatement pstmt = null;
    try {
        pstmt = conn
            .prepareStatement("UPDATE customer_t1 SET c_customer_name = ? WHERE c_customer_sk = 1");
        pstmt.setString(1, "new Data");
        int rowcount = pstmt.executeUpdate();
        pstmt.close();
    } catch (SQLException e) {
        if (pstmt != null) {
            try {
                pstmt.close();
            } catch (SQLException e1) {
                e1.printStackTrace();
            }
        }
        e.printStackTrace();
    }
}

//执行存储过程。
public static void ExecCallableSQL(Connection conn) {
    CallableStatement cstmt = null;
    try {

        cstmt=conn.prepareCall("{? = CALL TESTPROC(?,?,?)");
        cstmt.setInt(2, 50);
        cstmt.setInt(1, 20);
        cstmt.setInt(3, 90);
        cstmt.registerOutParameter(4, Types.INTEGER); //注册out类型的参数，类型为整型。
        cstmt.execute();
        int out = cstmt.getInt(4); //获取out参数
        System.out.println("The CallableStatment TESTPROC returns:"+out);
        cstmt.close();
    }
}
```

```
} catch (SQLException e) {
    if (cstmt != null) {
        try {
            cstmt.close();
        } catch (SQLException e1) {
            e1.printStackTrace();
        }
    }
    e.printStackTrace();
}
}

/**
 * 主程序，逐步调用各静态方法。
 * @param args
 */
public static void main(String[] args) {
    //创建数据库连接。
    Connection conn = GetConnection("tester", "password");

    //创建表。
    CreateTable(conn);

    //批插数据。
    BatchInsertData(conn);

    //执行预编译语句，更新数据。
    ExecPreparedSQL(conn);

    //执行存储过程。
    ExecCallableSQL(conn);

    //关闭数据库连接。
    try {
        conn.close();
    } catch (SQLException e) {
        e.printStackTrace();
    }
}
}
```

## 示例 2 客户端内存占用过多解决

此示例主要使用setFetchSize来调整客户端内存使用，它的原理是通过数据库游标来分批获取服务器端数据，但它会加大网络交互，可能会损失部分性能。

由于游标事务内有效，故需要先关闭自动提交。

```
// 关闭掉自动提交
conn.setAutoCommit(false);
Statement st = conn.createStatement();

// 打开游标，每次获取50行数据
st.setFetchSize(50);
ResultSet rs = st.executeQuery("SELECT * FROM mytable");
while (rs.next()) {
    System.out.print("a row was returned.");
}
rs.close();

// 关闭服务器游标。
st.setFetchSize(0);
rs = st.executeQuery("SELECT * FROM mytable");
while (rs.next()) {
    System.out.print("many rows were returned.");
}
```

```
}  
rs.close();  
  
// Close the statement.  
st.close();
```

### 9.3.9 示例：重新执行应用 SQL

当主DN故障且40s未恢复时，GaussDB(DWS)会自动将对应的备DN升主，使集群正常运行。备升主期间正在运行的作业会失败；备升主后启动的作业不会再受影响。如果要做到DN主备切换过程中，上层业务不感知，可参考此示例构建业务层SQL重试机制。

//以下用例以gsjdbc4.jar为例，如果要使用gsjdbc200.jar，请替换驱动类名（将代码中的“org.postgresql”替换成“com.huawei.gauss200.jdbc”）与连接URL串前缀（将“jdbc:postgresql”替换为“jdbc:gaussdb”）。

```
import java.sql.Connection;  
import java.sql.DriverManager;  
import java.sql.PreparedStatement;  
import java.sql.ResultSet;  
import java.sql.SQLException;  
import java.sql.Statement;  
  
/**  
 *  
 *  
 */  
class ExitHandler extends Thread {  
    private Statement cancel_stmt = null;  
  
    public ExitHandler(Statement stmt) {  
        super("Exit Handler");  
        this.cancel_stmt = stmt;  
    }  
    public void run() {  
        System.out.println("exit handle");  
        try {  
            this.cancel_stmt.cancel();  
        } catch (SQLException e) {  
            System.out.println("cancel query failed.");  
            e.printStackTrace();  
        }  
    }  
}  
  
public class SQLRetry {  
    //创建数据库连接。  
    public static Connection GetConnection(String username, String passwd) {  
        String driver = "org.postgresql.Driver";  
        String sourceURL = "jdbc:postgresql://10.131.72.136:8000/gaussdb";  
        Connection conn = null;  
        try {  
            //加载数据库驱动。  
            Class.forName(driver).newInstance();  
        } catch (Exception e) {  
            e.printStackTrace();  
            return null;  
        }  
  
        try {  
            //创建数据库连接。  
            conn = DriverManager.getConnection(sourceURL, username, passwd);  
            System.out.println("Connection succeed!");  
        } catch (Exception e) {  
            e.printStackTrace();  
            return null;  
        }  
    }  
}
```

```
        return conn;
    }

    //执行普通SQL语句，创建jdbc_test1表。
    public static void CreateTable(Connection conn) {
        Statement stmt = null;
        try {
            stmt = conn.createStatement();

            // add ctrl+c handler
            Runtime.getRuntime().addShutdownHook(new ExitHandler(stmt));

            //执行普通SQL语句。?
            int rc2 = stmt
                .executeUpdate("DROP TABLE if exists jdbc_test1;");

            int rc1 = stmt
                .executeUpdate("CREATE TABLE jdbc_test1(col1 INTEGER, col2 VARCHAR(10));");

            stmt.close();
        } catch (SQLException e) {
            if (stmt != null) {
                try {
                    stmt.close();
                } catch (SQLException e1) {
                    e1.printStackTrace();
                }
            }
            e.printStackTrace();
        }
    }

    //执行预处理语句，批量插入数据。
    public static void BatchInsertData(Connection conn) {
        PreparedStatement pst = null;

        try {
            //生成预处理语句。
            pst = conn.prepareStatement("INSERT INTO jdbc_test1 VALUES (?,?)");
            for (int i = 0; i < 100; i++) {
                //添加参数。
                pst.setInt(1, i);
                pst.setString(2, "data " + i);
                pst.addBatch();
            }
            //执行批处理。
            pst.executeBatch();
            pst.close();
        } catch (SQLException e) {
            if (pst != null) {
                try {
                    pst.close();
                } catch (SQLException e1) {
                    e1.printStackTrace();
                }
            }
            e.printStackTrace();
        }
    }

    //执行预编译语句，更新数据。
    private static boolean QueryRedo(Connection conn){
        PreparedStatement pstmt = null;
        boolean retValue = false;
        try {
            pstmt = conn
                .prepareStatement("SELECT col1 FROM jdbc_test1 WHERE col2 = ?");

            pstmt.setString(1, "data 10");
        }
    }
}
```



```
        ResultSet rs = pstmt.executeQuery();

        while (rs.next()) {
            System.out.println("col1 = " + rs.getString("col1"));
        }
        rs.close();

        pstmt.close();
        retValue = true;
    } catch (SQLException e) {
        System.out.println("catch..... retValue " + retValue);
        if (pstmt != null) {
            try {
                pstmt.close();
            } catch (SQLException e1) {
                e1.printStackTrace();
            }
        }
        e.printStackTrace();
    }

    System.out.println("finesh.....");
    return retValue;
}

//查询语句，执行失败重试，重试次数可配置。
public static void ExecPreparedSQL(Connection conn) throws InterruptedException {
    int maxRetryTime = 50;
    int time = 0;
    String result = null;
    do {
        time++;
        try {
            System.out.println("time:" + time);
            boolean ret = QueryRedo(conn);
            if (ret == false) {
                System.out.println("retry, time:" + time);
                Thread.sleep(10000);
                QueryRedo(conn);
            }
        } catch (Exception e) {
            e.printStackTrace();
        }
    } while (null == result && time < maxRetryTime);
}

/**
 * 主程序，逐步调用各静态方法。
 * @param args
 * @throws InterruptedException
 */
public static void main(String[] args) throws InterruptedException {
    //创建数据库连接。
    Connection conn = GetConnection("testuser", "test@123");

    //创建表。
    CreateTable(conn);

    //批插数据。
    BatchInsertData(conn);

    //执行预编译语句，更新数据。
    ExecPreparedSQL(conn);

    //关闭数据库连接。
    try {
        conn.close();
    } catch (SQLException e) {
```

```
e.printStackTrace();
}

}

}
```

### 9.3.10 示例：通过本地文件导入导出数据

在使用JAVA语言基于GaussDB(DWS)进行二次开发时，可以使用CopyManager接口，通过流方式，将数据库中的数据导出到本地文件或者将本地文件导入数据库中，文件格式支持CSV、TEXT等格式。

样例程序如下，执行时需要加载GaussDB(DWS) jdbc驱动。

```
//以下用例以gsjdbc4.jar为例，如果要使用gsjdbc200.jar，请替换驱动类名（将代码中的“org.postgresql”替换成“com.huawei.gauss200.jdbc”）与连接URL串前缀（将“jdbc.postgresql”替换为“jdbc.gaussdb”）。
import java.sql.Connection;
import java.sql.DriverManager;
import java.io.IOException;
import java.io.FileInputStream;
import java.io.FileOutputStream;
import java.sql.SQLException;
import org.postgresql.copy.CopyManager;
import org.postgresql.core.BaseConnection;

public class Copy{

    public static void main(String[] args)
    {
        String urls = new String("jdbc:postgresql://10.180.155.74:8000/gaussdb"); //数据库URL
        String username = new String("jack"); //用户名
        String password = new String("*****"); //密码
        String tablename = new String("migration_table"); //定义表信息
        String tablename1 = new String("migration_table_1"); //定义表信息
        String driver = "org.postgresql.Driver";
        Connection conn = null;

        try {
            Class.forName(driver);
            conn = DriverManager.getConnection(urls, username, password);
        } catch (ClassNotFoundException e) {
            e.printStackTrace(System.out);
        } catch (SQLException e) {
            e.printStackTrace(System.out);
        }

        // 将SELECT * FROM migration_table查询结果导出到本地文件d:/data.txt
        try {
            copyToFile(conn, "d:/data.txt", "(SELECT * FROM migration_table)");
        } catch (SQLException e) {
            // TODO Auto-generated catch block
            e.printStackTrace();
        } catch (IOException e) {
            // TODO Auto-generated catch block
            e.printStackTrace();
        }

        //将d:/data.txt中的数据导入到migration_table_1中。
        try {
            copyFromFile(conn, "d:/data.txt", tablename1);
        } catch (SQLException e) {
            // TODO Auto-generated catch block
            e.printStackTrace();
        } catch (IOException e) {
            // TODO Auto-generated catch block
            e.printStackTrace();
        }
    }
}
```

```
// 将migration_table_1中的数据导出到本地文件d:/data1.txt
try {
    copyToFile(conn, "d:/data1.txt", tablename1);
} catch (SQLException e) {
    // TODO Auto-generated catch block
    e.printStackTrace();
} catch (IOException e) {
    // TODO Auto-generated catch block
    e.printStackTrace();
}
}

public static void copyFromFile(Connection connection, String filePath, String tableName)
    throws SQLException, IOException {

    FileInputStream fileInputStream = null;

    try {
        CopyManager copyManager = new CopyManager((BaseConnection)connection);
        fileInputStream = new FileInputStream(filePath);
        copyManager.copyIn("COPY " + tableName + " FROM STDIN", fileInputStream);
    } finally {
        if (fileInputStream != null) {
            try {
                fileInputStream.close();
            } catch (IOException e) {
                e.printStackTrace();
            }
        }
    }
}

public static void copyToFile(Connection connection, String filePath, String tableOrQuery)
    throws SQLException, IOException {

    FileOutputStream fileOutputStream = null;

    try {
        CopyManager copyManager = new CopyManager((BaseConnection)connection);
        fileOutputStream = new FileOutputStream(filePath);
        copyManager.copyOut("COPY " + tableOrQuery + " TO STDOUT", fileOutputStream);
    } finally {
        if (fileOutputStream != null) {
            try {
                fileOutputStream.close();
            } catch (IOException e) {
                e.printStackTrace();
            }
        }
    }
}
}
```

### 9.3.11 示例：从 MySQL 向 GaussDB(DWS)进行数据迁移

下面示例演示如何通过CopyManager从mysql向GaussDB(DWS)进行数据迁移的过程。

```
//以下用例以gsjdbc4.jar为例，如果要使用gsjdbc200.jar，请替换驱动类名（将代码中的“org.postgresql”替换成“com.huawei.gauss200.jdbc”）与连接URL串前缀（将“jdbc:postgresql”替换为“jdbc:gaussdb”）。
import java.io.StringReader;
import java.sql.Connection;
import java.sql.DriverManager;
import java.sql.ResultSet;
import java.sql.SQLException;
import java.sql.Statement;

import org.postgresql.copy.CopyManager;
import org.postgresql.core.BaseConnection;
```

```
public class Migration{

    public static void main(String[] args) {
        String url = new String("jdbc:postgresql://10.180.155.74:8000/gaussdb"); //数据库URL
        String user = new String("jack"); //mppdb用户名
        String pass = new String("*****"); //mppdb密码
        String tablename = new String("migration_table"); //定义表信息
        String delimiter = new String("|"); //定义分隔符
        String encoding = new String("UTF8"); //定义字符集
        String driver = "org.postgresql.Driver";
        StringBuffer buffer = new StringBuffer(); //定义存放格式化数据的缓存

        try {
            //获取源数据库查询结果集
            ResultSet rs = getDataSet();

            //遍历结果集，逐行获取记录
            //将每条记录中各字段值，按指定分隔符分割，由换行符结束，拼成一个字符串
            //把拼成的字符串，添加到缓存buffer
            while (rs.next()) {
                buffer.append(rs.getString(1) + delimiter
                    + rs.getString(2) + delimiter
                    + rs.getString(3) + delimiter
                    + rs.getString(4)
                    + "\n");
            }
            rs.close();

            try {
                //建立目标数据库连接
                Class.forName(driver);
                Connection conn = DriverManager.getConnection(url, user, pass);
                BaseConnection baseConn = (BaseConnection) conn;
                baseConn.setAutoCommit(false);

                //初始化表信息
                String sql = "Copy " + tablename + " from STDIN DELIMITER " + "'" + delimiter + "'" + "
ENCODING " + "'" + encoding + "'";

                //提交缓存buffer中的数据
                CopyManager cp = new CopyManager(baseConn);
                StringReader reader = new StringReader(buffer.toString());
                cp.copyIn(sql, reader);
                baseConn.commit();
                reader.close();
                baseConn.close();
            } catch (ClassNotFoundException e) {
                e.printStackTrace(System.out);
            } catch (SQLException e) {
                e.printStackTrace(System.out);
            }
        } catch (Exception e) {
            e.printStackTrace();
        }
    }

    //*****
    // 从源数据库返回查询结果集
    //*****
    private static ResultSet getDataSet() {
        ResultSet rs = null;
        try {
            Class.forName("com.mysql.jdbc.Driver").newInstance();
            Connection conn = DriverManager.getConnection("jdbc:mysql://10.119.179.227:3306/jack?
useSSL=false&allowPublicKeyRetrieval=true", "jack", "*****");
            Statement stmt = conn.createStatement();
            rs = stmt.executeQuery("select * from migration_table");
        }
    }
}
```

```

    } catch (SQLException e) {
        e.printStackTrace();
    } catch (Exception e) {
        e.printStackTrace();
    }
    return rs;
}
}

```

## 9.3.12 JDBC 接口参考

JDBC接口是一套提供给用户的API方法，本节将对部分常用接口做具体描述，若涉及其他接口可参考JDK1.6（软件包）/JDBC4.0中相关内容。

### 9.3.12.1 java.sql.Connection

java.sql.Connection是数据库连接接口。

表 9-6 对 java.sql.Connection 接口的支持情况

方法名	返回值类型	支持JDBC 4
close()	void	Yes
commit()	void	Yes
createStatement()	Statement	Yes
getAutoCommit()	boolean	Yes
getClientInfo()	Properties	Yes
getClientInfo(String name)	String	Yes
getTransactionIsolation()	int	Yes
isClosed()	boolean	Yes
isReadOnly()	boolean	Yes
prepareStatement(String sql)	PreparedStatement	Yes
rollback()	void	Yes
setAutoCommit(boolean autoCommit)	void	Yes
setClientInfo(Properties properties)	void	Yes
setClientInfo(String name,String value)	void	Yes

**须知**

接口内部默认使用自动提交模式，若通过setAutoCommit(false)关闭自动提交，将会导致后面执行的语句都受到显式事务包裹，数据库中不支持事务中执行的语句不能在此模式下执行。

### 9.3.12.2 java.sql.CallableStatement

java.sql.CallableStatement是存储过程执行接口。

表 9-7 对 java.sql.CallableStatement 的支持情况

方法名	返回值类型	支持JDBC 4
registerOutParameter(int parameterIndex, int type)	void	Yes
wasNull()	boolean	Yes
getString(int parameterIndex)	String	Yes
getBoolean(int parameterIndex)	boolean	Yes
getByte(int parameterIndex)	byte	Yes
getShort(int parameterIndex)	short	Yes
getInt(int parameterIndex)	int	Yes
getLong(int parameterIndex)	long	Yes
getFloat(int parameterIndex)	float	Yes
getDouble(int parameterIndex)	double	Yes
getBigDecimal(int parameterIndex)	BigDecimal	Yes
getBytes(int parameterIndex)	byte[]	Yes
getDate(int parameterIndex)	Date	Yes
getTime(int parameterIndex)	Time	Yes
getTimestamp(int parameterIndex)	Timestamp	Yes
getObject(int parameterIndex)	Object	Yes

**说明**

- 不允许含有OUT参数的语句执行批量操作。
- 以下方法是从java.sql.Statement继承而来：close, execute, executeQuery, executeUpdate, getConnection, getResultSet, getUpdateCount, isClosed, setMaxRows, setFetchSize。
- 以下方法是从java.sql.PreparedStatement继承而来：addBatch, clearParameters, execute, executeQuery, executeUpdate, getMetaData, setBigDecimal, setBoolean, setByte, setBytes, setDate, setDouble, setFloat, setInt, setLong, setNull, setObject, setString, setTime, setTimestamp。

### 9.3.12.3 java.sql.DatabaseMetaData

java.sql.DatabaseMetaData是数据库对象定义接口。

表 9-8 对 java.sql.DatabaseMetaData 的支持情况

方法名	返回值类型	支持JDBC 4
getTables(String catalog, String schemaPattern, String tableNamePattern, String[] types)	ResultSet	Yes
getColumns(String catalog, String schemaPattern, String tableNamePattern, String columnNamePattern)	ResultSet	Yes
getTableTypes()	ResultSet	Yes
getUserName()	String	Yes
isReadOnly()	boolean	Yes
nullsAreSortedHigh()	boolean	Yes
nullsAreSortedLow()	boolean	Yes
nullsAreSortedAtStart()	boolean	Yes
nullsAreSortedAtEnd()	boolean	Yes
getDatabaseProductName()	String	Yes
getDatabaseProductVersion()	String	Yes
getDriverName()	String	Yes
getDriverVersion()	String	Yes
getDriverMajorVersion()	int	Yes
getDriverMinorVersion()	int	Yes
usesLocalFiles()	boolean	Yes

方法名	返回值类型	支持JDBC 4
usesLocalFilePerTable()	boolean	Yes
supportsMixedCaseIdentifiers()	boolean	Yes
storesUpperCaseIdentifiers()	boolean	Yes
storesLowerCaseIdentifiers()	boolean	Yes
supportsMixedCaseQuotedIdentifiers()	boolean	Yes
storesUpperCaseQuotedIdentifiers()	boolean	Yes
storesLowerCaseQuotedIdentifiers()	boolean	Yes
storesMixedCaseQuotedIdentifiers()	boolean	Yes
supportsAlterTableWithAddColumn()	boolean	Yes
supportsAlterTableWithDropColumn()	boolean	Yes
supportsColumnAliasing()	boolean	Yes
nullPlusNonNullsNull()	boolean	Yes
supportsConvert()	boolean	Yes
supportsConvert(int fromType, int toType)	boolean	Yes
supportsTableCorrelationNames()	boolean	Yes
supportsDifferentTableCorrelationNames()	boolean	Yes
supportsExpressionsInOrderBy()	boolean	Yes
supportsOrderByUnrelated()	boolean	Yes
supportsGroupBy()	boolean	Yes
supportsGroupByUnrelated()	boolean	Yes
supportsGroupByBeyondSelect()	boolean	Yes
supportsLikeEscapeClause()	boolean	Yes
supportsMultipleResultSets()	boolean	Yes



方法名	返回值类型	支持JDBC 4
supportsMultipleTransactions()	boolean	Yes
supportsNonNullableColumns()	boolean	Yes
supportsMinimumSQLGrammar()	boolean	Yes
supportsCoreSQLGrammar()	boolean	Yes
supportsExtendedSQLGrammar()	boolean	Yes
supportsANSI92EntryLevelSQL()	boolean	Yes
supportsANSI92IntermediateSQL()	boolean	Yes
supportsANSI92FullSQL()	boolean	Yes
supportsIntegrityEnhancementFacility()	boolean	Yes
supportsOuterJoins()	boolean	Yes
supportsFullOuterJoins()	boolean	Yes
supportsLimitedOuterJoins()	boolean	Yes
isCatalogAtStart()	boolean	Yes
supportsSchemasInDataManipulation()	boolean	Yes
supportsSavepoints()	boolean	Yes
supportsResultSetHoldability(int holdability)	boolean	Yes
getResultSetHoldability()	int	Yes
getDatabaseMajorVersion()	int	Yes
getDatabaseMinorVersion()	int	Yes
getJDBCMajorVersion()	int	Yes
getJDBCMinorVersion()	int	Yes

### 9.3.12.4 java.sql.Driver

java.sql.Driver是数据库驱动接口。

表 9-9 对 java.sql.Driver 的支持情况

方法名	返回值类型	支持JDBC 4
acceptsURL(String url)	boolean	Yes
connect(String url, Properties info)	Connection	Yes
jdbcCompliant()	boolean	Yes
getMajorVersion()	int	Yes
getMinorVersion()	int	Yes

### 9.3.12.5 java.sql.PreparedStatement

java.sql.PreparedStatement是预处理语句接口。

表 9-10 对 java.sql.PreparedStatement 的支持情况

方法名	返回值类型	支持JDBC 4
clearParameters()	void	Yes
execute()	boolean	Yes
executeQuery()	ResultSet	Yes
executeUpdate()	int	Yes
getMetaData()	ResultSetMetaData	Yes
setBoolean(int parameterIndex, boolean x)	void	Yes
setBigDecimal(int parameterIndex, BigDecimal x)	void	Yes
setByte(int parameterIndex, byte x)	void	Yes
setBytes(int parameterIndex, byte[] x)	void	Yes
setDate(int parameterIndex, Date x)	void	Yes
setDouble(int parameterIndex, double x)	void	Yes
setFloat(int parameterIndex, float x)	void	Yes

方法名	返回值类型	支持JDBC 4
setInt(int parameterIndex, int x)	void	Yes
setLong(int parameterIndex, long x)	void	Yes
setNString(int parameterIndex, String value)	void	Yes
setShort(int parameterIndex, short x)	void	Yes
setString(int parameterIndex, String x)	void	Yes
addBatch()	void	Yes
executeBatch()	int[]	Yes
clearBatch()	void	Yes

#### 📖 说明

- addBatch()、execute()必须在clearBatch()之后才能执行。
- 调用executeBatch()方法并不会清除batch。用户必须显式使用clearBatch()清除。
- 在添加了一个batch的绑定变量后，用户若想重用这些值(再次添加一个batch)，无需再次使用set\*()方法。
- 以下方法是从java.sql.Statement继承而来：close，execute，executeQuery，executeUpdate，getConnection，getResultSet，getUpdateCount，isClosed，setMaxRows，setFetchSize。

### 9.3.12.6 java.sql.ResultSet

java.sql.ResultSet是执行结果集接口。

表 9-11 对 java.sql.ResultSet 的支持情况

方法名	返回值类型	支持JDBC 4
findColumn(String columnLabel)	int	Yes
getBigDecimal(int columnIndex)	BigDecimal	Yes
getBigDecimal(String columnLabel)	BigDecimal	Yes

方法名	返回值类型	支持JDBC 4
getBoolean(int columnIndex)	boolean	Yes
getBoolean(String columnLabel)	boolean	Yes
getBytes(int columnIndex)	byte[]	Yes
getBytes(String columnLabel)	byte[]	Yes
getDate(int columnIndex)	Date	Yes
getDate(String columnLabel)	Date	Yes
getDouble(int columnIndex)	double	Yes
getDouble(String columnLabel)	double	Yes
getFloat(int columnIndex)	float	Yes
getFloat(String columnLabel)	float	Yes
getInt(int columnIndex)	int	Yes
getInt(String columnLabel)	int	Yes
getLong(int columnIndex)	long	Yes
getLong(String columnLabel)	long	Yes
getShort(int columnIndex)	short	Yes
getShort(String columnLabel)	short	Yes
getString(int columnIndex)	String	Yes
getString(String columnLabel)	String	Yes
getTime(int columnIndex)	Time	Yes

方法名	返回值类型	支持JDBC 4
getTime(String columnLabel)	Time	Yes
getTimestamp(int columnIndex)	Timestamp	Yes
getTimestamp(String columnLabel)	Timestamp	Yes
isAfterLast()	boolean	Yes
isBeforeFirst()	boolean	Yes
isFirst()	boolean	Yes
next()	boolean	Yes

#### 说明

- 一个Statement不能有多个处于“open”状态的ResultSet。
- 用于遍历结果集(ResultSet)的游标(Cursor)在被提交后不能保持“open”的状态。

### 9.3.12.7 java.sql.ResultSetMetaData

java.sql.ResultSetMetaData是对ResultSet对象相关信息的具体描述。

表 9-12 对 java.sql.ResultSetMetaData 的支持情况

方法名	返回值类型	支持JDBC 4
getColumnCount()	int	Yes
getColumnName(int column)	String	Yes
getColumnType(int column)	int	Yes
getColumnTypeName(int column)	String	Yes

### 9.3.12.8 java.sql.Statement

java.sql.Statement是SQL语句接口。

表 9-13 对 java.sql.Statement 的支持情况

方法名	返回值类型	支持JDBC 4
close()	void	Yes

方法名	返回值类型	支持JDBC 4
execute(String sql)	boolean	Yes
executeQuery(String sql)	ResultSet	Yes
executeUpdate(String sql)	int	Yes
getConnection()	Connection	Yes
getResultSet()	ResultSet	Yes
getQueryTimeout()	int	Yes
getUpdateCount()	int	Yes
isClosed()	boolean	Yes
setQueryTimeout(int seconds)	void	Yes
setFetchSize(int rows)	void	Yes
cancel()	void	Yes

#### 📖 说明

通过setFetchSize可以减少结果集在客户端的内存占用情况。它的原理是通过将结果集打包成游标，然后分段处理，所以会加大数据库与客户端的通信量，会有性能损耗。

由于数据库游标是事务内有效，所以，在设置setFetchSize的同时，需要将连接设置为非自动提交模式，setAutoCommit(false)。同时在业务数据需要持久化到数据库中时，在连接上执行提交操作。

### 9.3.12.9 javax.sql.ConnectionPoolDataSource

javax.sql.ConnectionPoolDataSource是数据源连接池接口。

表 9-14 对 javax.sql.ConnectionPoolDataSource 的支持情况

方法名	返回值类型	支持JDBC 4
getLoginTimeout()	int	Yes
getLogWriter()	PrintWriter	Yes
getPooledConnection()	PooledConnection	Yes
getPooledConnection(String user,String password)	PooledConnection	Yes
setLoginTimeout(int seconds)	void	Yes

方法名	返回值类型	支持JDBC 4
setLogWriter(PrintWriter out)	void	Yes

### 9.3.12.10 javax.sql.DataSource

javax.sql.DataSource是数据源接口。

表 9-15 对 javax.sql.DataSource 接口的支持情况

方法名	返回值类型	支持JDBC 4
getConnection()	Connection	Yes
getConnection(String username,String password)	Connection	Yes
getLoginTimeout()	int	Yes
getLogWriter()	PrintWriter	Yes
setLoginTimeout(int seconds)	void	Yes
setLogWriter(PrintWriter out)	void	Yes

### 9.3.12.11 javax.sql.PooledConnection

javax.sql.PooledConnection是由连接池创建的连接接口。

表 9-16 对 javax.sql.PooledConnection 的支持情况

方法名	返回值类型	支持JDBC 4
addConnectionEventListener (ConnectionEventListener listener)	void	Yes
close()	void	Yes
getConnection()	Connection	Yes
removeConnectionEventListener (ConnectionEventListener listener)	void	Yes
addStatementEventListener (StatementEventListener listener)	void	Yes
removeStatementEventListener (StatementEventListener listener)	void	Yes

### 9.3.12.12 javax.naming.Context

javax.naming.Context是连接配置的上下文接口。

表 9-17 对 javax.naming.Context 的支持情况

方法名	返回值类型	支持JDBC 4
bind(Name name, Object obj)	void	Yes
bind(String name, Object obj)	void	Yes
lookup(Name name)	Object	Yes
lookup(String name)	Object	Yes
rebind(Name name, Object obj)	void	Yes
rebind(String name, Object obj)	void	Yes
rename(Name oldName, Name newName)	void	Yes
rename(String oldName, String newName)	void	Yes
unbind(Name name)	void	Yes
unbind(String name)	void	Yes

### 9.3.12.13 javax.naming.spi.InitialContextFactory

javax.naming.spi.InitialContextFactory是初始连接上下文工厂接口。

表 9-18 对 javax.naming.spi.InitialContextFactory 的支持情况

方法名	返回值类型	支持JDBC 4
getInitialContext(Hashtable<?,?> environment)	Context	Yes

### 9.3.12.14 CopyManager

CopyManager是GaussDB(DWS) JDBC驱动中提供的一个API接口类，用于批量向GaussDB(DWS)集群中导入数据。



## CopyManager 的继承关系

CopyManager类位于org.postgresql.copy Package中，继承自java.lang.Object类，该类的声明如下：

```
public class CopyManager
extends Object
```

## 构造方法

```
public CopyManager(BaseConnection connection)
```

```
throws SQLException
```

## 常用方法

表 9-19 CopyManager 常用方法

返回值	方法	描述	throws
CopyIn	copyIn(String sql)	-	SQLException
long	copyIn(String sql, InputStream from)	使用COPY FROM STDIN从InputStream中快速向数据库中的表加载数据。	SQLException,IOException
long	copyIn(String sql, InputStream from, int bufferSize)	使用COPY FROM STDIN从InputStream中快速向数据库中的表加载数据。	SQLException,IOException
long	copyIn(String sql, Reader from)	使用COPY FROM STDIN从Reader中快速向数据库中的表加载数据。	SQLException,IOException
long	copyIn(String sql, Reader from, int bufferSize)	使用COPY FROM STDIN从Reader中快速向数据库中的表加载数据。	SQLException,IOException
CopyOut	copyOut(String sql)	-	SQLException
long	copyOut(String sql, OutputStream to)	将一个COPY TO STDOUT的结果集从数据库发送到OutputStream类中。	SQLException,IOException

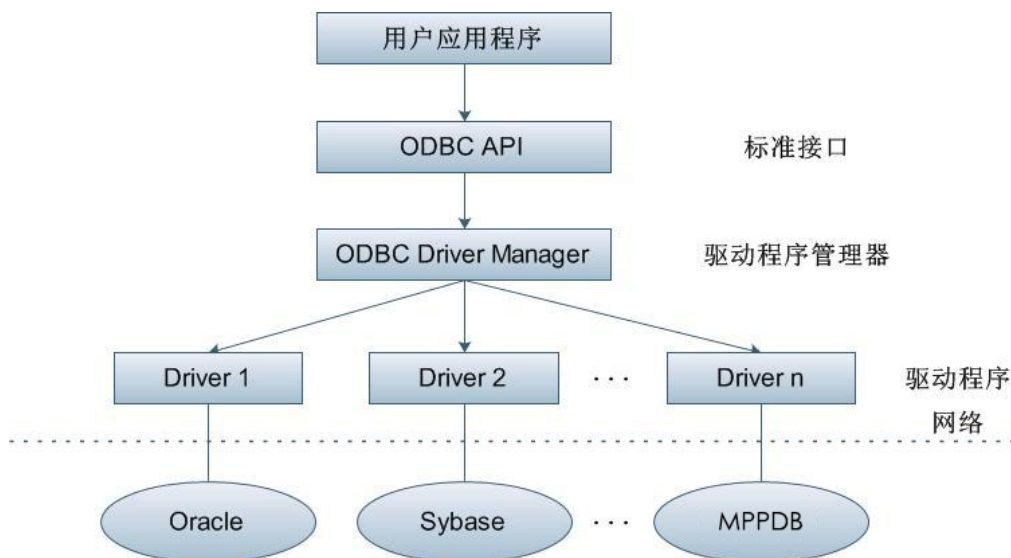
返回值	方法	描述	throws
long	copyOut(String sql, Writer to)	将一个COPY TO STDOUT的结果集从数据库发送到Writer类中。	SQLException,IOException

## 9.4 基于 ODBC 开发

ODBC ( Open Database Connectivity, 开放数据库互连 ) 是由Microsoft公司基于X/OPEN CLI提出的用于访问数据库的应用程序编程接口。应用程序通过ODBC提供的API与数据库进行交互, 在避免了应用程序直接操作数据库系统的同时, 增强了应用程序的可移植性、扩展性和可维护性。

ODBC的系统结构参见图9-2。

图 9-2 ODBC 系统机构



GaussDB(DWS)目前在以下环境中提供对ODBC3.5的支持。

表 9-20 ODBC 支持平台

操作系统	平台
SUSE Linux Enterprise Server 11 SP1/SP2/SP3/SP4 SUSE Linux Enterprise Server 12 及SP1/SP2/SP3/SP5	x86_64位
Red Hat Enterprise Linux 6.4/6.5/6.6/6.7/6.8/6.9/7.0/7.1/7.2/7.3/7.4/7.5	x86_64位
Red Hat Enterprise Linux 7.5	ARM64位
CentOS 6.4/6.5/6.6/6.7/6.8/6.9/7.0/7.1/7.2/7.3/7.4	x86_64位

操作系统	平台
CentOS 7.6	ARM64位
EulerOS 2.0 SP2/SP3	x86_64位
EulerOS 2.0 SP8	ARM64位
中标麒麟 7.5/7.6	ARM64位
Oracle Linux R7U4	x86_64位
Windows 7	32位
Windows 7	64位
Windows Server 2008	32位
Windows Server 2008	64位

以上操作系统平台是指ODBC程序所在的操作系统平台，可以与数据库部署的操作系统平台不同。

UNIX/Linux系统下的驱动程序管理器主要有unixODBC和iODBC，在这选择驱动管理器unixODBC-2.3.0作为连接数据库的组件。

Windows系统自带ODBC驱动程序管理器，在控制面板->管理工具中可以找到数据源（ODBC）选项。

#### 说明

当前数据库ODBC驱动基于开源版本，对于GaussDB(DWS)的数据类型，tinyint、smalldatetime、nvarchar2在获取数据类型的时候，可能会出现不兼容。

## 9.4.1 ODBC 包及依赖的库和头文件

### Linux 下的 ODBC 包

从发布包中获取，包名为dws\_8.1.x\_odbc\_driver\_for\_XXX\_XXX.zip。Linux环境下，开发应用程序要用到unixODBC提供的头文件（sql.h、sql.h等）和库libodbc.so。这些头文件和库可从unixODBC-2.3.0的安装包中获得。

### Windows 下的 ODBC 包

从发布包中获取，包名为dws\_8.1.x\_odbc\_driver\_for\_windows.zip。Windows环境下，开发应用程序用到的相关头文件和库文件都由系统自带。

## 9.4.2 Linux 下配置数据源

将GaussDB(DWS)提供的ODBC DRIVER（psqlodbcw.so）配置到数据源中便可使用。配置数据源需要配置“odbc.ini”和“odbcinst.ini”两个文件（在编译安装unixODBC过程中生成且默认放在“/usr/local/etc”目录下），并在服务器端进行配置。

## 操作步骤

### 步骤1 获取unixODBC源码包。

获取参考地址：<http://sourceforge.net/projects/unixodbc/files/unixODBC/2.3.0/unixODBC-2.3.0.tar.gz/download>

### 步骤2 目前不支持unixODBC-2.2.1版本。以unixODBC-2.3.0版本为例，在客户端执行如下命令安装unixODBC。默认安装到“/usr/local”目录下，生成数据源文件到“/usr/local/etc”目录下，库文件生成在“/usr/local/lib”目录。

```
tar zxvf unixODBC-2.3.0.tar.gz
cd unixODBC-2.3.0
#修改configure文件(如果不存在, 那么请修改configure.ac), 找到LIB_VERSION
#将它的值修改为"1:0:0", 这样将编译出*.so.1的动态库, 与psqlodbcw.so的依赖关系相同。
vim configure

./configure --enable-gui=no #如果要在TaiShan服务器上编译, 请追加一个configure参数: --build=aarch64-unknown-linux-gnu
make
#安装可能需要root权限
make install
```

安装unixODBC。如果机器上已经安装了其他版本的unixODBC，可以直接覆盖安装。

### 步骤3 替换客户端GaussDB(DWS)驱动程序。

将dws\_8.1.x\_odbc\_driver\_for\_XXX\_XXX.zip解压，在“/dws\_8.1.x\_odbc\_driver\_for\_XXX\_XXX/odbc/lib”目录下得到“psqlodbcw.la”和“psqlodbcw.so”两个文件。

### 步骤4 配置数据源。

#### 1. 配置ODBC驱动文件。

在“/usr/local/etc/odbcinst.ini”文件中追加以下内容。

```
[GaussMPP]
Driver64=/usr/local/lib/psqlodbcw.so
setup=/usr/local/lib/psqlodbcw.so
```

odbcinst.ini文件中的配置参数说明如表9-21所示。

表 9-21 odbcinst.ini 文件配置参数

参数	描述	示例
[DriverName]	驱动器名称，对应数据源DSN中的驱动名。	[DRIVER_N]
Driver64	驱动动态库的路径。	Driver64=/xxx/odbc/lib/psqlodbcw.so
setup	驱动安装路径，与Driver64中动态库的路径一致。	setup=/xxx/odbc/lib/psqlodbcw.so

#### 2. 配置数据源文件。

在“/usr/local/etc/odbc.ini”文件中追加以下内容。

```
[MPP_ODBC]
Driver=GaussMPP
Servername=10.10.0.13 (数据库Server IP)
Database=gaussdb (数据库名)
Username=dbadmin (数据库用户名)
```

```

Password= (数据库用户密码)
Port=8000 (数据库监听端口)
Sslmode=allow
    
```

odbc.ini文件配置参数说明如表9-22所示。

表 9-22 odbc.ini 文件配置参数

参数	描述	示例
[DSN]	数据源的名称。	[MPPODBC]
Driver	驱动名，对应odbcinst.ini中的DriverName。	Driver=DRIVER_N
Servename	服务器的IP地址。	Servename=10.145.130.26
Database	要连接的数据库的名称。	Database=gaussdb
Username	数据库用户名称。	Username=dbadmin
Password	数据库用户密码。	Password= <b>说明</b> ODBC驱动本身已经对内存密码进行过清理，以保证用户密码在连接后不会再在内存中保留。 但是如果配置了此参数，由于UnixODBC对数据源文件等进行缓存，可能导致密码长期保留在内存中。 推荐在应用程序连接时，将密码传递给相应API，而非写在数据源配置文件中。同时连接成功后，应当及时清理保存密码的内存段。
Port	服务器的端口号。	Port=8000
Sslmode	开启SSL模式	Sslmode=allow
UseServerSidePrepare	是否开启数据库端扩展查询协议。 可选值0或1，默认为1，表示打开扩展查询协议。	UseServerSidePrepare=1
UseBatchProtocol	是否开启批量查询协议（打开可提高DML性能）；可选值0或者1，默认为1。 当此值为0时，不使用批量查询协议（主要用于与早期数据库版本通信兼容）。 当此值为1，并且数据库support_batch_bind参数存在且为on时，将打开批量查询协议。	UseBatchProtocol=1

参数	描述	示例
ConnectionExtraInfo	GUC参数connection_info（参见 <a href="#">connection_info</a> ）中显示驱动部署路径和进程属主用户的开关。	ConnectionExtraInfo=1 <b>说明</b> 默认值为1。当设置为0时，ODBC驱动会将当前驱动的名称、驱动版本上报到数据库中；当设置为1时，ODBC驱动会将当前驱动的名称、部署路径、进程属主用户上报到数据库中，记录在connection_info参数（参见 <a href="#">connection_info</a> ）里；同时可以在 <a href="#">PG_STAT_ACTIVITY</a> 和 <a href="#">PGXC_STAT_ACTIVITY</a> 中查询到。
ForExtensionConnector	ETL工具性能优化参数，可进行内存优化，降低对端的CN内存占用，避免因CN内存使用过多导致系统不稳定。 可选值0或者1，默认为0，表示不开启优化项。 请勿在数据库系统之外的其他业务中配置此参数，以免影响业务的正确性。	ForExtensionConnector=1
KeepDisallowPremature	当UseDeclareFetch=1时，应用程序调用SQLPrepare后调用SQLNumResultCols、SQLDescribeCol或SQLColAttribute获取结果集列信息时，SQL语句中的游标是否具有with hold属性。 可选值0或者1，0表示具有with hold属性，1表示不具有with hold属性，默认为0。	KeepDisallowPremature=1 <b>说明</b> - UseServerSidePrepare=1时，KeepDisallowPremature参数不生效，使用时需要指定UseServerSidePrepare为0，例如： UseDeclareFetch=1 KeepDisallowPremature=1 UseServerSidePrepare=0

其中关于sslmode的选项的允许值，具体信息见下表：

表 9-23 sslmode 的可选项及其描述

sslmode	是否会启用SSL加密	描述
disable	否	不使用SSL安全连接。

sslmode	是否会启用SSL加密	描述
allow	可能	如果数据库服务器要求使用，则可以使用SSL安全加密连接，但不验证数据库服务器的真实性。
prefer	可能	如果数据库支持，那么首选使用SSL安全加密连接，但不验证数据库服务器的真实性。
require	是	必须使用SSL安全连接，但是只做了数据加密，而并不验证数据库服务器的真实性。
verify-ca	是	必须使用SSL安全连接，并且验证数据库是否具有可信证书机器签发的证书。
verify-full	是	必须使用SSL安全连接，在verify-ca的验证范围之外，同时验证数据库所在主机的主机名是否与证书内容一致。GaussDB(DWS)不支持此模式。

### 步骤5 SSL模式

如果需要使用SSL证书连接，那么请将GaussDB(DWS)安装包中的SSLCERT的证书包解压，在shell环境下，执行“source sslcert\_env.sh”，即在当前会话完成证书的默认位置的部署。

或者手动声明如下环境变量，同时保证client.key\*系列文件为600权限：

```
export PGSSLCERT="/YOUR/PATH/OF/client.crt" #请修改该路径到client.crt的绝对路径
export PGSSLKEY="/YOUR/PATH/OF/client.key" #请修改该路径到client.key的绝对路径
```

同时将数据源中的Sslmode选项调整至“verify-ca”。

**步骤6** 将客户端所在主机的IP网段加入GaussDB(DWS)的安全组规则，确保客户端主机与GaussDB(DWS)网络互通。

**步骤7** 配置环境变量。

```
vim ~/.bashrc
```

在配置文件中追加以下内容。

```
export LD_LIBRARY_PATH=/usr/local/lib/:$LD_LIBRARY_PATH
export ODBC_SYSINI=/usr/local/etc
export ODBCINI=/usr/local/etc/odbc.ini
```

**步骤8** 执行如下命令使设置生效。

```
source ~/.bashrc
```

----结束

## 测试数据源配置

执行isql -v GaussODBC(数据源名称)命令。

- 如果显示如下信息，表明配置正确，连接成功。

```
+-----+
| Connected!          |
|                    |
| sql-statement      |
| help [tablename]   |
| quit               |
+-----+
```

```
|  
+-----+  
SQL>
```

- 若显示ERROR信息，则表明配置错误。请检查上述配置是否正确。

## 常见问题处理

- [UnixODBC][Driver Manager]Can't open lib 'xxx/xxx/psqlodbcw.so' : file not found.

此问题的可能原因：

- odbcinst.ini文件中配置的路径不正确

确认的方法：'ls'一下错误信息中的路径，以确保该psqlodbcw.so文件存在，同时具有执行权限。

- psqlodbcw.so的依赖库不存在，或者不在系统环境变量中

确认的办法：ldd一下错误信息中的路径，如果是缺少libodbc.so.1等UnixODBC的库，那么按照“操作步骤”中的方法重新配置UnixODBC，并确保它的安装路径下的lib目录添加到了LD\_LIBRARY\_PATH中；如果是缺少其他库，请将ODBC驱动包中的lib目录添加到LD\_LIBRARY\_PATH中。

- [UnixODBC]connect to server failed: no such file or directory

此问题可能的原因：

- 配置了错误的/不可达的数据库地址，或者端口

请检查数据源配置中的Servername及Port配置项。

- 服务器监听不正确

如果确认Servername及Port配置正确，请根据“操作步骤”中数据库服务器的相关配置，确保数据库监听了合适的网卡及端口。

- 防火墙及网闸设备

请确认防火墙设置，将数据库的通信端口添加到可信端口中。

如果有网闸设备，请确认一下相关的设置。

- [unixODBC]The password-stored method is not supported.

此问题可能原因：

数据源中未配置sslmode配置项。

解决办法：

请配置该选项至allow或以上选项。此配置的更多信息，见[表9-23](#)。

- Server common name "xxxx" does not match host name "xxxxx"

此问题的原因：

使用了SSL加密的“verify-full”选项，驱动程序会验证证书中的主机名与实际部署数据库的主机名是否一致。

解决办法：

碰到此问题可以使用“verify-ca”选项，不再校验主机名；或者重新生成一套与数据库所在主机名相同的CA证书。

- Driver's SQLAllocHandle on SQL\_HANDLE\_DBC failed

此问题的可能原因：

可执行文件（比如UnixODBC的isql，以下都以isql为例）与数据库驱动（psqlodbcw.so）依赖于不同的odbc的库版本：libodbc.so.1或者libodbc.so.2。此问题可以通过如下方式确认：



```
ldd `which isql` | grep odbc  
ldd psqlodbcw.so | grep odbc
```

这时，如果输出的libodbc.so最后的后缀数字不同或者指向不同的磁盘物理文件，那么基本就可以断定是此问题。isql与psqlodbcw.so都会要求加载libodbc.so，这时如果它们加载的是不同的物理文件，便会导致两套完全同名的函数列表，同时出现在同一个可见域里（UnixODBC的libodbc.so.\*的函数导出列表完全一致），产生冲突，无法加载数据库驱动。

解决办法：

确定一个要使用的UnixODBC，然后卸载另外一个（比如卸载库版本号为.so.2的UnixODBC），然后将剩下的.so.1的库，新建一个同名但是后缀为.so.2的软链接，便可解决此问题。

- FATAL: Forbid remote connection with trust method!

由于安全原因，数据库CN禁止集群内部其他节点无认证接入。

如果要在集群内部访问CN，请将ODBC程序部署在CN所在机器，服务器地址使用"127.0.0.1"。建议业务系统单独部署在集群外部，否则可能会影响数据库运行性能。

- [unixODBC][Driver Manager]Invalid attribute value

在使用SQL on other GaussDB功能时碰到此问题，有可能是unixODBC的版本并非推荐版本，建议通过“odbcinst --version”命令排查环境中的unixODBC版本。

- authentication method 10 not supported.

使用开源客户端碰到此问题，可能原因：

数据库中存储的口令校验只存储了SHA256格式哈希，而开源客户端只识别MD5校验，双方校验方法不匹配报错。

#### 📖 说明

- 数据库并不存储用户口令，只存储用户口令的哈希码。
- 早期版本（V100R002C80SPC300之前的版本）的数据库只存储了SHA256格式的哈希，并未存储MD5的哈希，所以无法使用MD5做用户口令校验。
- 新版本（V100R002C80SPC300及之后版本）的数据库当用户更新用户口令或者新建用户时，会同时存储两种格式的哈希码，这时将兼容开源的认证协议。
- 但是当老版本升级到新版本时，由于哈希的不可逆性，所以数据库无法还原用户口令，进而生成新格式的哈希，所以仍然只保留了SHA256格式的哈希，导致仍然无法使用MD5做口令认证。

要解决该问题，可以更新用户口令；或者新建一个用户，赋予同等权限，使用新用户连接数据库。

- unsupported frontend protocol 3.51: server supports 1.0 to 3.0

目标数据库版本过低，或者目标数据库为开源数据库。请使用对应版本的数据库驱动连接目标数据库。

## 9.4.3 Windows 下配置数据源

Windows操作系统自带ODBC数据源管理器，无需用户手动安装管理器便可直接进行配置。

### 操作步骤

#### 步骤1 替换客户端GaussDB(DWS)驱动程序

将GaussDB-8.2.0-Windows-Odbc.tar.gz解压后，根据需要，双击psqlodbc.msi（32位）或者psqlodbc\_x64.msi（64位）进行驱动安装。

## 步骤2 打开驱动管理器

在配置数据源时，请使用对应的驱动管理器（假设操作系统安装盘符为C:盘，如果是其他盘符，请对路径做相应修改）：

- **64位操作系统上进行32位程序开发，安装32位驱动程序后，使用32位的驱动管理器：** C:\Windows\SysWOW64\odbcad32.exe  
请勿直接使用控制面板->管理工具->数据源(ODBC)。

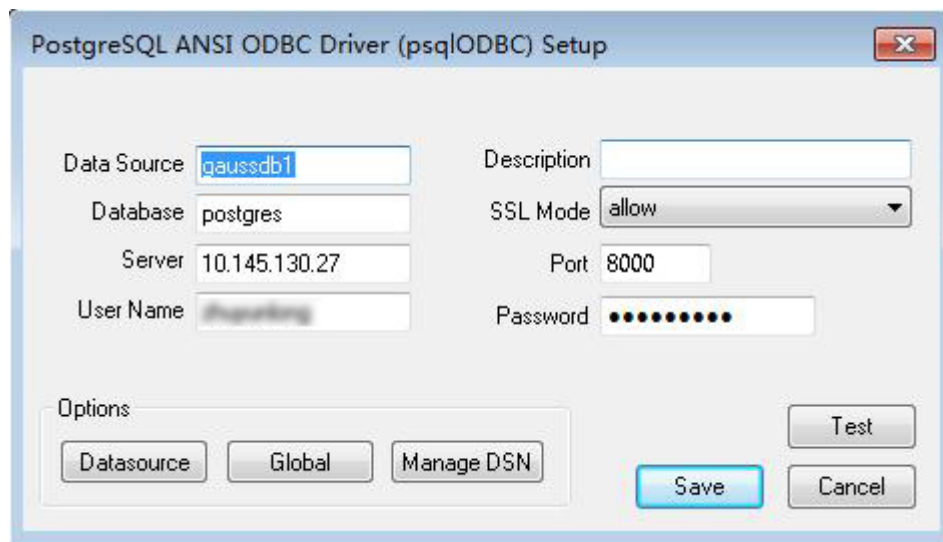
### 📖 说明

WoW64的全称是"Windows 32-bit on Windows 64-bit"，C:\Windows\SysWOW64存放的是64位系统上的32位运行环境。而C:\Windows\System32存放的是与操作系统一致的运行环境，具体的技术信息请查阅Windows的相关技术文档。

- **64位操作系统上进行64位程序开发，安装64位驱动程序后，使用64位的驱动管理器：** C:\Windows\System32\odbcad32.exe  
请勿直接使用控制面板->管理工具->数据源(ODBC)。
- **32位操作系统请使用：** C:\Windows\System32\odbcad32.exe  
或者单击计算机->控制面板->管理工具->数据源(ODBC)打开驱动管理器。

## 步骤3 配置数据源

在打开的驱动管理器上，选择用户DSN->添加->PostgreSQL Unicode（如果是64位驱动，将会有64位标识），然后进行配置：



### 须知

此界面上配置的用户名及密码信息，将会被记录在Windows注册表中，再次连接数据库时就不再需要输入认证信息。但是出于安全考虑，建议在单击"Save"按钮保存配置信息前，清空相关敏感信息；在使用ODBC的连接API时，再传入所需的用户名、密码信息。

#### 步骤4 SSL模式

如果需要使用SSL证书连接，那么请将GaussDB(DWS)安装包中的SSLCERT的证书包解压，双击"sslcert\_env.bat"文件，即可完成证书的默认位置的部署。

#### 须知

该sslcert\_env.bat为了保证证书环境的纯净，在%APPDATA%\postgresql目录存在时，会提示是否需要移除相关目录。如果有需要，请备份该目录中的文件。

或者手动将client.crt、client.key、client.key.cipher、client.key.rand文件放至%APPDATA%\postgresql(该目录需手动建立)目录下，并且将文件名中的client改为postgres，例如client.key修改为postgres.key；将cacert.pem文件放至%APPDATA%\postgresql目录，并更名为root.crt。

同时将步骤2中的设置窗口的“SSL Mode”选项调整至“verify-ca”。

表 9-24 sslmode 的可选项及其描述

sslmode	是否会启用SSL加密	描述
disable	否	不使用SSL安全连接。
allow	可能	如果数据库服务器要求使用，则可以使用SSL安全加密连接，但不验证数据库服务器的真实性。
prefer	可能	如果数据库支持，那么首选使用SSL安全加密连接，但不验证数据库服务器的真实性。
require	是	必须使用SSL安全连接，但是只做了数据加密，而并不验证数据库服务器的真实性。
verify-ca	是	必须使用SSL安全连接，并且验证数据库是否具有可信证书机器签发的证书。
verify-full	是	必须使用SSL安全连接，在verify-ca的验证范围之外，同时验证数据库所在主机的主机名是否与证书内容一致。 <b>说明</b> 不支持此模式。

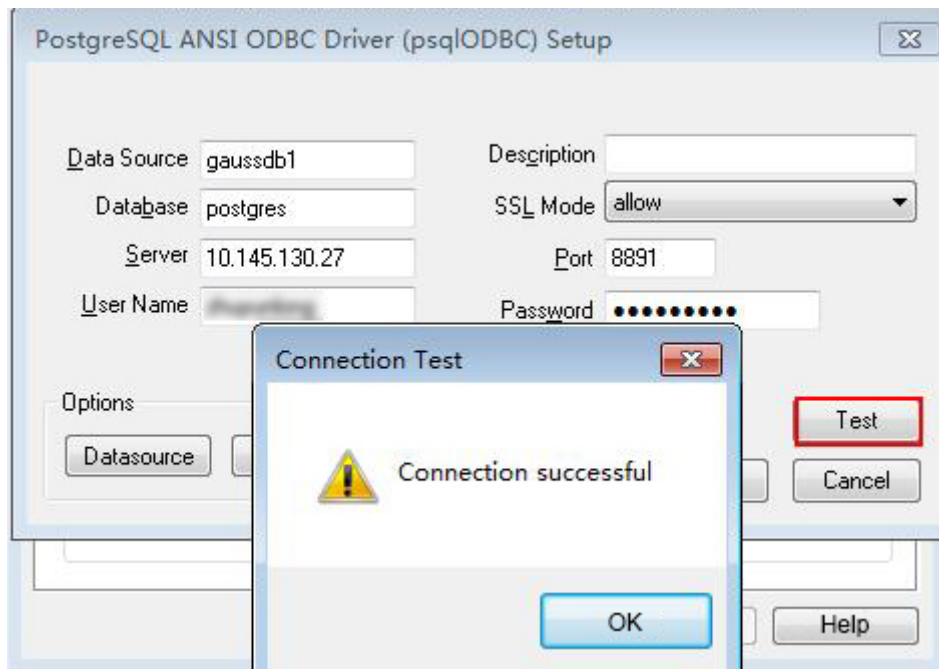
**步骤5** 将客户端所在主机的IP网段加入GaussDB(DWS)的安全组规则，确保客户端主机与GaussDB(DWS)网络互通。

----结束

### 测试数据源配置

单击Test进行测试。

- 如果显示如下，则表明配置正确，连接成功。



- 若显示ERROR信息，则表明配置错误。请检查上述配置是否正确。

## 常见问题处理

- Server common name "xxxx" does not match host name "xxxxx"  
此问题的原因是使用了SSL加密的“verify-full”选项，这时驱动程序会验证证书中的主机名与实际部署数据库的主机名是否一致。碰到此问题可以使用“verify-ca”选项，不再校验主机名；或者重新生成一套与数据库所在主机名相同的CA证书。
- connect to server failed: no such file or directory  
此问题可能的原因：
  - 配置了错误的/不可达的数据库地址，或者端口  
请检查数据源配置中的Servername及Port配置项。
  - 服务器监听不正确  
如果确认Servername及Port配置正确，请根据“操作步骤”中数据库服务器的相关配置，确保数据库监听了合适的网卡及端口。
  - 防火墙及网闸设备  
请确认防火墙设置，将数据库的通信端口添加到可信端口中。  
如果有网闸设备，请确认一下相关的设置。
- 在指定的DSN中，驱动程序和应用程序之间的体系结构不匹配  
此问题可能的原因：在64位程序中使用了32位驱动，或者相反。  
C:\Windows\SysWOW64\odbcad32.exe：这是32位ODBC驱动管理器。  
C:\Windows\System32\odbcad32.exe：这是64位ODBC驱动管理器。
- The password-stored method is not supported.  
此问题可能原因：  
数据源中未配置sslmode配置项，请调整此项至allow或以上级别，允许SSL连接，此选项的更多说明，请见表9-24。

- authentication method 10 not supported.

使用开源客户端碰到此问题，可能原因：

数据库中存储的口令校验只存储了SHA256格式哈希，而开源客户端只识别MD5校验，双方校验方法不匹配报错。

#### 📖 说明

- 数据库并不存储用户口令，只存储用户口令的哈希码。
- 早期版本（V100R002C80SPC300之前的版本）的数据库只存储了SHA256格式的哈希，并未存储MD5的哈希，所以无法使用MD5做用户口令校验。
- 新版本（V100R002C80SPC300及之后版本）的数据库当用户更新用户口令或者新建用户时，会同时存储两种格式的哈希码，这时将兼容开源的认证协议。
- 但是当老版本升级到新版本时，由于哈希的不可逆性，所以数据库无法还原用户口令，进而生成新格式的哈希，所以仍然只保留了SHA256格式的哈希，导致仍然无法使用MD5做口令认证。

要解决该问题，参见以下操作：

- a. 将数据库参数password\_encryption\_type修改为1，修改方法参见[修改数据库参数](#)。
- b. 新建一个数据库用户用于连接，或者重置准备使用的数据库用户的密码。
  - 如果您使用的是管理员账号，参见[重置密码](#)。
  - 如果是普通用户，可以先通过其他客户端工具（例如Data Studio）连接数据库后，使用ALTER USER语句来修改密码。
- c. 再尝试连接数据库。

- unsupported frontend protocol 3.51: server supports 1.0 to 3.0

目标数据库版本过低，或者目标数据库为开源数据库。请使用对应版本的数据库驱动连接目标数据库。

- FATAL: GSS authentication method is not allowed because XXXX user password is not disabled.

或：GSSAPI authentication not supported.

目标CN的pg\_hba.conf里配置了当前客户端IP使用“gss”方式来做认证，该认证算法不支持用作客户端的身份认证，请修改到“sha256”后再试。

同时请注意，数据库当前不支持在集群内跨节点连接数据库，如果是在集群内跨节点连接CN出现此问题，请将业务程序调整到集群外后重试。

## 9.4.4 ODBC 开发示例

### 常用功能示例代码

```
// 此示例演示如何通过ODBC方式获取GaussDB(DWS)中的数据。
// DBtest.c (compile with: libodbc.so)
#include <stdlib.h>
#include <stdio.h>
#include <sqlext.h>
#ifdef WIN32
#include <windows.h>
#endif
SQLHENV     V_OD_Env;      // Handle ODBC environment
SQLHSTMT    V_OD_hstmt;   // Handle statement
SQLHDBC     V_OD_hdbc;    // Handle connection
char         typename[100];
SQLINTEGER  value = 100;
```

```
SQLINTEGER V_OD_erg,V_OD_buffer,V_OD_err,V_OD_id;
SQLLEN V_StrLen_or_IndPtr;
int main(int argc,char *argv[])
{
    // 1. 申请环境句柄
    V_OD_erg = SQLAllocHandle(SQL_HANDLE_ENV,SQL_NULL_HANDLE,&V_OD_Env);
    if ((V_OD_erg != SQL_SUCCESS) && (V_OD_erg != SQL_SUCCESS_WITH_INFO))
    {
        printf("Error AllocHandle\n");
        exit(0);
    }
    // 2. 设置环境属性 ( 版本信息 )
    SQLSetEnvAttr(V_OD_Env, SQL_ATTR_ODBC_VERSION, (void*)SQL_OV_ODBC3, 0);
    // 3. 申请连接句柄
    V_OD_erg = SQLAllocHandle(SQL_HANDLE_DBC, V_OD_Env, &V_OD_hdbc);
    if ((V_OD_erg != SQL_SUCCESS) && (V_OD_erg != SQL_SUCCESS_WITH_INFO))
    {
        SQLFreeHandle(SQL_HANDLE_ENV, V_OD_Env);
        exit(0);
    }
    // 4. 设置连接属性
    SQLSetConnectAttr(V_OD_hdbc, SQL_ATTR_AUTOCOMMIT, SQL_AUTOCOMMIT_ON, 0);
    // 5. 连接数据库, 这里的 "userName" 与 "password" 分别表示连接数据库的用户名和用户密码, 请根据
    实际情况修改。
    // 如果odbc.ini文件中已经配置了用户名密码, 那么这里可以留空 ( "" ); 但是不建议这么做, 因为一旦
    odbc.ini权限管理不善, 将导致数据库用户密码泄露。
    V_OD_erg = SQLConnect(V_OD_hdbc, (SQLCHAR*) "gaussdb", SQL_NTS,
        (SQLCHAR*) "userName", SQL_NTS, (SQLCHAR*) "password", SQL_NTS);
    if ((V_OD_erg != SQL_SUCCESS) && (V_OD_erg != SQL_SUCCESS_WITH_INFO))
    {
        printf("Error SQLConnect %d\n",V_OD_erg);
        SQLFreeHandle(SQL_HANDLE_ENV, V_OD_Env);
        exit(0);
    }
    printf("Connected !\n");
    // 6. 设置语句属性
    SQLSetStmtAttr(V_OD_hstmt,SQL_ATTR_QUERY_TIMEOUT,(SQLPOINTER *)3,0);
    // 7. 申请语句句柄
    SQLAllocHandle(SQL_HANDLE_STMT, V_OD_hdbc, &V_OD_hstmt);
    // 8. 直接执行SQL语句。
    SQLExecDirect(V_OD_hstmt,"drop table IF EXISTS customer_t1",SQL_NTS);
    SQLExecDirect(V_OD_hstmt,"CREATE TABLE customer_t1(c_customer_sk INTEGER, c_customer_name
    VARCHAR(32));",SQL_NTS);
    SQLExecDirect(V_OD_hstmt,"insert into customer_t1 values(25,'li')",SQL_NTS);
    // 9. 准备执行
    SQLPrepare(V_OD_hstmt,"insert into customer_t1 values(?)",SQL_NTS);
    // 10. 绑定参数
    SQLBindParameter(V_OD_hstmt,1,SQL_PARAM_INPUT,SQL_C_SLONG,SQL_INTEGER,0,0,
        &value,0,NULL);
    // 11. 执行准备好的语句
    SQLExecute(V_OD_hstmt);
    SQLExecDirect(V_OD_hstmt,"select id from testtable",SQL_NTS);
    // 12. 获取结果集某一列的属性
    SQLColAttribute(V_OD_hstmt,
    1,SQL_DESC_TYPE_NAME,typename,sizeof(typename),NULL,NULL);
    printf("SQLColAttribute %s\n",typename);
    // 13. 绑定结果集
    SQLBindCol(V_OD_hstmt,1,SQL_C_SLONG, (SQLPOINTER)&V_OD_buffer,150,
        (SQLLEN *)&V_StrLen_or_IndPtr);
    // 14. 通过SQLFetch取结果集中数据
    V_OD_erg=SQLFetch(V_OD_hstmt);
    // 15. 通过SQLGetData获取并返回数据。
    while(V_OD_erg != SQL_NO_DATA)
    {
        SQLGetData(V_OD_hstmt,1,SQL_C_SLONG,(SQLPOINTER)&V_OD_id,0,NULL);
        printf("SQLGetData ----ID = %d\n",V_OD_id);
        V_OD_erg=SQLFetch(V_OD_hstmt);
    };
    printf("Done !\n");
}
```

```
// 16. 断开数据源连接并释放句柄资源
SQLFreeHandle(SQL_HANDLE_STMT,V_OD_hstmt);
SQLDisconnect(V_OD_hdbc);
SQLFreeHandle(SQL_HANDLE_DBC,V_OD_hdbc);
SQLFreeHandle(SQL_HANDLE_ENV, V_OD_Env);
return(0);
}
```

## 批量绑定示例代码

```
/******
* 请在数据源中打开UseBatchProtocol，同时指定数据库中参数support_batch_bind
* 为on
* CHECK_ERROR的作用是检查并打印错误信息。
* 此示例将与用户交互式获取DSN、模拟的数据量，忽略的数据量，并将最终数据入库到test_odbc_batch_insert
中
*****/
#include <stdio.h>
#include <stdlib.h>
#include <sql.h>
#include <sqlext.h>
#include <string.h>

#include "util.c"

void Exec(SQLHDBC hdbc, SQLCHAR* sql)
{
    SQLRETURN retcode;           // Return status
    SQLHSTMT hstmt = SQL_NULL_HSTMT; // Statement handle
    SQLCHAR loginfo[2048];

    // Allocate Statement Handle
    retcode = SQLAllocHandle(SQL_HANDLE_STMT, hdbc, &hstmt);
    CHECK_ERROR(retcode, "SQLAllocHandle(SQL_HANDLE_STMT)",
                hstmt, SQL_HANDLE_STMT);

    // Prepare Statement
    retcode = SQLPrepare(hstmt, (SQLCHAR*) sql, SQL_NTS);
    sprintf((char*)loginfo, "SQLPrepare log: %s", (char*)sql);
    CHECK_ERROR(retcode, loginfo, hstmt, SQL_HANDLE_STMT);

    retcode = SQLExecute(hstmt);
    sprintf((char*)loginfo, "SQLExecute stmt log: %s", (char*)sql);
    CHECK_ERROR(retcode, loginfo, hstmt, SQL_HANDLE_STMT);

    retcode = SQLFreeHandle(SQL_HANDLE_STMT, hstmt);
    sprintf((char*)loginfo, "SQLFreeHandle stmt log: %s", (char*)sql);
    CHECK_ERROR(retcode, loginfo, hstmt, SQL_HANDLE_STMT);
}

int main ()
{
    SQLHENV henv = SQL_NULL_HENV;
    SQLHDBC hdbc = SQL_NULL_HDBC;
    int batchCount = 1000;
    SQLLEN rowsCount = 0;
    int ignoreCount = 0;

    SQLRETURN retcode;
    SQLCHAR dsn[1024] = {'\0'};
    SQLCHAR loginfo[2048];

    // 交互获取数据源名称
    getStr("Please input your DSN", (char*)dsn, sizeof(dsn), 'N');
    // 交互获取批量绑定的数据量
    getInt("batchCount", &batchCount, 'N', 1);
    do
    {
        // 交互获取批量绑定的数据中，不要入库的数据量
    }
}
```



```
getInt("ignoreCount", &ignoreCount, 'N', 1);
if (ignoreCount > batchCount)
{
    printf("ignoreCount(%d) should be less than batchCount(%d)\n", ignoreCount, batchCount);
}
}while(ignoreCount > batchCount);

retcode = SQLAllocHandle(SQL_HANDLE_ENV, SQL_NULL_HANDLE, &henv);
CHECK_ERROR(retcode, "SQLAllocHandle(SQL_HANDLE_ENV)",
            henv, SQL_HANDLE_ENV);

// Set ODBC Verion
retcode = SQLSetEnvAttr(henv, SQL_ATTR_ODBC_VERSION,
                        (SQLPOINTER*)SQL_OV_ODBC3, 0);
CHECK_ERROR(retcode, "SQLSetEnvAttr(SQL_ATTR_ODBC_VERSION)",
            henv, SQL_HANDLE_ENV);

// Allocate Connection
retcode = SQLAllocHandle(SQL_HANDLE_DBC, henv, &hdbc);
CHECK_ERROR(retcode, "SQLAllocHandle(SQL_HANDLE_DBC)",
            henv, SQL_HANDLE_DBC);

// Set Login Timeout
retcode = SQLSetConnectAttr(hdbc, SQL_LOGIN_TIMEOUT, (SQLPOINTER)5, 0);
CHECK_ERROR(retcode, "SQLSetConnectAttr(SQL_LOGIN_TIMEOUT)",
            hdbc, SQL_HANDLE_DBC);

// Set Auto Commit
retcode = SQLSetConnectAttr(hdbc, SQL_ATTR_AUTOCOMMIT,
                            (SQLPOINTER)(1), 0);
CHECK_ERROR(retcode, "SQLSetConnectAttr(SQL_ATTR_AUTOCOMMIT)",
            hdbc, SQL_HANDLE_DBC);

// Connect to DSN
sprintf(loginInfo, "SQLConnect(DSN:%s)", dsn);
retcode = SQLConnect(hdbc, (SQLCHAR*) dsn, SQL_NTS,
                    (SQLCHAR*) NULL, 0, NULL, 0);
CHECK_ERROR(retcode, loginInfo, hdbc, SQL_HANDLE_DBC);

// init table info.
Exec(hdbc, "drop table if exists test_odbc_batch_insert");
Exec(hdbc, "create table test_odbc_batch_insert(id int primary key, col varchar2(50))");

// 下面的代码根据用户输入的数据量，构造出将要入库的数据：
{
    SQLRETURN retcode;
    SQLHSTMT hstmt;
    int i;
    SQLCHAR *sql = NULL;
    SQLINTEGER *ids = NULL;
    SQLCHAR *cols = NULL;
    SQLLEN *bufLenIds = NULL;
    SQLLEN *bufLenCols = NULL;
    SQLUSMALLINT *operptr = NULL;
    SQLUSMALLINT *statusptr = NULL;
    SQLULEN process = 0;

    // 这里是按列构造，每个字段的内存连续存放在一起。
    ids = (SQLINTEGER*)malloc(sizeof(ids[0]) * batchCount);
    cols = (SQLCHAR*)malloc(sizeof(cols[0]) * batchCount * 50);
    // 这里是每个字段中，每一行数据的内存长度。
    bufLenIds = (SQLLEN*)malloc(sizeof(bufLenIds[0]) * batchCount);
    bufLenCols = (SQLLEN*)malloc(sizeof(bufLenCols[0]) * batchCount);
    // 该行是否需要被处理，SQL_PARAM_IGNORE 或 SQL_PARAM_PROCEED
    operptr = (SQLUSMALLINT*)malloc(sizeof(operptr[0]) * batchCount);
    memset(operptr, 0, sizeof(operptr[0]) * batchCount);
    // 该行的处理结果。
    // 注：由于数据库中处理方式是同一语句隶属同一事务中，所以如果出错，那么待处理数据都将是出错的，
    并不会部分入库。
}
```



```
statusptr = (SQLUSMALLINT*)malloc(sizeof(statusptr[0]) * batchCount);
memset(statusptr, 88, sizeof(statusptr[0]) * batchCount);

if (NULL == ids || NULL == cols || NULL == bufLenCols || NULL == bufLenIds)
{
    fprintf(stderr, "FAILED:\tmalloc data memory failed\n");
    goto exit;
}

for (int i = 0; i < batchCount; i++)
{
    ids[i] = i;
    sprintf(cols + 50 * i, "column test value %d", i);
    bufLenIds[i] = sizeof(ids[i]);
    bufLenCols[i] = strlen(cols + 50 * i);
    operptr[i] = (i < ignoreCount) ? SQL_PARAM_IGNORE : SQL_PARAM_PROCEED;
}

// Allocate Statement Handle
retcode = SQLAllocHandle(SQL_HANDLE_STMT, hdbc, &hstmtinesrt);
CHECK_ERROR(retcode, "SQLAllocHandle(SQL_HANDLE_STMT)",
            hstmtinesrt, SQL_HANDLE_STMT);

// Prepare Statement
sql = (SQLCHAR*)"insert into test_odbc_batch_insert values(?, ?)";
retcode = SQLPrepare(hstmtinesrt, (SQLCHAR*) sql, SQL_NTS);
sprintf((char*)loginfo, "SQLPrepare log: %s", (char*)sql);
CHECK_ERROR(retcode, loginfo, hstmtinesrt, SQL_HANDLE_STMT);

retcode = SQLSetStmtAttr(hstmtinesrt, SQL_ATTR_PARAMSET_SIZE, (SQLPOINTER)batchCount,
sizeof(batchCount));
CHECK_ERROR(retcode, "SQLSetStmtAttr", hstmtinesrt, SQL_HANDLE_STMT);

retcode = SQLBindParameter(hstmtinesrt, 1, SQL_PARAM_INPUT, SQL_C_SLONG, SQL_INTEGER,
sizeof(ids[0]), 0,&(ids[0]), 0, bufLenIds);
CHECK_ERROR(retcode, "SQLBindParameter for id", hstmtinesrt, SQL_HANDLE_STMT);

retcode = SQLBindParameter(hstmtinesrt, 2, SQL_PARAM_INPUT, SQL_C_CHAR, SQL_CHAR, 50, 50,
cols, 50, bufLenCols);
CHECK_ERROR(retcode, "SQLBindParameter for cols", hstmtinesrt, SQL_HANDLE_STMT);

retcode = SQLSetStmtAttr(hstmtinesrt, SQL_ATTR_PARAMS_PROCESSED_PTR, (SQLPOINTER)&process,
sizeof(process));
CHECK_ERROR(retcode, "SQLSetStmtAttr for SQL_ATTR_PARAMS_PROCESSED_PTR", hstmtinesrt,
SQL_HANDLE_STMT);

retcode = SQLSetStmtAttr(hstmtinesrt, SQL_ATTR_PARAM_STATUS_PTR, (SQLPOINTER)statusptr,
sizeof(statusptr[0]) * batchCount);
CHECK_ERROR(retcode, "SQLSetStmtAttr for SQL_ATTR_PARAM_STATUS_PTR", hstmtinesrt,
SQL_HANDLE_STMT);

retcode = SQLSetStmtAttr(hstmtinesrt, SQL_ATTR_PARAM_OPERATION_PTR, (SQLPOINTER)operptr,
sizeof(operptr[0]) * batchCount);
CHECK_ERROR(retcode, "SQLSetStmtAttr for SQL_ATTR_PARAM_OPERATION_PTR", hstmtinesrt,
SQL_HANDLE_STMT);

retcode = SQLExecute(hstmtinesrt);
sprintf((char*)loginfo, "SQLExecute stmt log: %s", (char*)sql);
CHECK_ERROR(retcode, loginfo, hstmtinesrt, SQL_HANDLE_STMT);

retcode = SQLRowCount(hstmtinesrt, &rowsCount);
CHECK_ERROR(retcode, "SQLRowCount execution", hstmtinesrt, SQL_HANDLE_STMT);

if (rowsCount != (batchCount - ignoreCount))
{
    sprintf(loginfo, "(batchCount - ignoreCount)(%d) != rowsCount(%d)", (batchCount - ignoreCount),
rowsCount);
    CHECK_ERROR(SQL_ERROR, loginfo, NULL, SQL_HANDLE_STMT);
}
```

```
else
{
    sprintf(loginfo, "(batchCount - ignoreCount)(%d) == rowsCount(%d)", (batchCount - ignoreCount),
rowsCount);
    CHECK_ERROR(SQL_SUCCESS, loginfo, NULL, SQL_HANDLE_STMT);
}

if (rowsCount != process)
{
    sprintf(loginfo, "process(%d) != rowsCount(%d)", process, rowsCount);
    CHECK_ERROR(SQL_ERROR, loginfo, NULL, SQL_HANDLE_STMT);
}
else
{
    sprintf(loginfo, "process(%d) == rowsCount(%d)", process, rowsCount);
    CHECK_ERROR(SQL_SUCCESS, loginfo, NULL, SQL_HANDLE_STMT);
}

for (int i = 0; i < batchCount; i++)
{
    if (i < ignoreCount)
    {
        if (statusptr[i] != SQL_PARAM_UNUSED)
        {
            sprintf(loginfo, "statusptr[%d](%d) != SQL_PARAM_UNUSED", i, statusptr[i]);
            CHECK_ERROR(SQL_ERROR, loginfo, NULL, SQL_HANDLE_STMT);
        }
    }
    else if (statusptr[i] != SQL_PARAM_SUCCESS)
    {
        sprintf(loginfo, "statusptr[%d](%d) != SQL_PARAM_SUCCESS", i, statusptr[i]);
        CHECK_ERROR(SQL_ERROR, loginfo, NULL, SQL_HANDLE_STMT);
    }
}

retcode = SQLFreeHandle(SQL_HANDLE_STMT, hstmtinesrt);
sprintf((char*)loginfo, "SQLFreeHandle hstmtinesrt");
CHECK_ERROR(retcode, loginfo, hstmtinesrt, SQL_HANDLE_STMT);
}

exit:
printf ("\nComplete.\n");

// Connection
if (hdbc != SQL_NULL_HDBC) {
    SQLDisconnect(hdbc);
    SQLFreeHandle(SQL_HANDLE_DBC, hdbc);
}

// Environment
if (henv != SQL_NULL_HENV)
    SQLFreeHandle(SQL_HANDLE_ENV, henv);

return 0;
}
```

## 9.4.5 ODBC 接口参考

ODBC接口是一套提供给用户的API函数，本节将对部分常用接口做具体描述，若涉及其他接口可参考msdn（网址：[https://msdn.microsoft.com/en-us/library/windows/desktop/ms714177\(v=vs.85\).aspx](https://msdn.microsoft.com/en-us/library/windows/desktop/ms714177(v=vs.85).aspx)）中ODBC Programmer's Reference项的相关内容。

### 9.4.5.1 SQLAllocEnv

在ODBC 3.x版本中，ODBC 2.x的函数SQLAllocEnv已被SQLAllocHandle代替。有关详细信息请参阅[SQLAllocHandle](#)。

### 9.4.5.2 SQLAllocConnect

在ODBC 3.x版本中，ODBC 2.x的函数SQLAllocConnect已被SQLAllocHandle代替。有关详细信息请参阅[SQLAllocHandle](#)。

### 9.4.5.3 SQLAllocHandle

#### 功能描述

分配环境、连接、语句或描述符的句柄，它替代了ODBC 2.x函数SQLAllocEnv、SQLAllocConnect及SQLAllocStmt。

#### 原型

```
SQLRETURN SQLAllocHandle(SQLSMALLINT HandleType,
                          SQLHANDLE InputHandle,
                          SQLHANDLE *OutputHandlePtr);
```

#### 参数

表 9-25 SQLAllocHandle 参数

关键字	参数说明
HandleType	<p>由SQLAllocHandle分配的句柄类型。必须为下列值之一：</p> <ul style="list-style-type: none"> <li>SQL_HANDLE_ENV（环境句柄）</li> <li>SQL_HANDLE_DBC（连接句柄）</li> <li>SQL_HANDLE_STMT（语句句柄）</li> <li>SQL_HANDLE_DESC（描述句柄）</li> </ul> <p>申请句柄顺序为，先申请环境句柄，再申请连接句柄，最后申请语句句柄，后申请的句柄都要依赖它前面申请的句柄。</p>
InputHandle	<p>将要分配的新句柄的类型。</p> <ul style="list-style-type: none"> <li>如果HandleType为SQL_HANDLE_ENV，则这个值为SQL_NULL_HANDLE。</li> <li>如果HandleType为SQL_HANDLE_DBC，则这一定是一个环境句柄。</li> <li>如果HandleType为SQL_HANDLE_STMT或SQL_HANDLE_DESC，则它一定是一个连接句柄。</li> </ul>
OutputHandlePtr	<p><b>输出参数：</b>一个缓冲区的指针，此缓冲区以新分配的数据结构存放返回的句柄。</p>

## 返回值

- SQL\_SUCCESS：表示调用正确，
- SQL\_SUCCESS\_WITH\_INFO：表示会有一些警告信息，
- SQL\_ERROR：表示比较严重的错误，如：内存分配失败、建立连接失败等、
- SQL\_INVALID\_HANDLE：表示调用无效句柄。其他API的返回值同理。

## 注意事项

当分配的句柄并非环境句柄时，如果SQLAllocHandle返回的值为SQL\_ERROR，则它会将OutputHandlePtr的值设置为SQL\_NULL\_HDBC、SQL\_NULL\_HSTMT或SQL\_NULL\_HDESC。之后，通过调用带有适当参数的SQLGetDiagRec，其中HandleType和Handle被设置为InputHandle的值，可得到相关的SQLSTATE值，通过SQLSTATE值可以查出调用此函数的具体信息。

## 示例

参见：[示例](#)

### 9.4.5.4 SQLAllocStmt

在ODBC 3.x版本中，ODBC 2.x的函数SQLAllocStmt已被SQLAllocHandle代替。有关详细信息请参阅[SQLAllocHandle](#)。

### 9.4.5.5 SQLBindCol

## 功能描述

将应用程序数据缓冲区绑定到结果集的列中。

## 原型

```
SQLRETURN SQLBindCol(SQLHSTMT StatementHandle,
    SQLUSMALLINT ColumnNumber,
    SQLSMALLINT TargetType,
    SQLPOINTER TargetValuePtr,
    SQLLEN BufferLength,
    SQLLEN *StrLen_or_IndPtr);
```

## 参数

表 9-26 SQLBindCol 参数

关键字	参数说明
StatementHandle	语句句柄。
ColumnNumber	要绑定结果集的列号。起始列号为0，以递增的顺序计算列号，第0列是书签列。若未设置书签页，则起始列号为1。
TargetType	缓冲区中C数据类型的标识符。

关键字	参数说明
TargetValuePtr	<b>输出参数：</b> 指向与列绑定的数据缓冲区的指针。SQLFetch函数返回这个缓冲区中的数据。如果此参数为一个空指针，则 StrLen_or_IndPtr是一个有效值。
BufferLength	TargetValuePtr指向缓冲区的长度，以字节为单位。
StrLen_or_IndPtr	<b>输出参数：</b> 缓冲区的长度或指示器指针。若为空值，则未使用任何长度或指示器值。

## 返回值

- SQL\_SUCCESS：表示调用正确。
- SQL\_SUCCESS\_WITH\_INFO：表示会有一些警告信息。
- SQL\_ERROR：表示比较严重的错误，如：内存分配失败、建立连接失败等。
- SQL\_INVALID\_HANDLE：表示调用无效句柄。其他API的返回值同理。

## 注意事项

当SQLBindCol返回SQL\_ERROR或SQL\_SUCCESS\_WITH\_INFO时，通过调用[SQLGetDiagRec](#)函数，并将HandleType和Handle参数设置为SQL\_HANDLE\_STMT和StatementHandle，可得到一个相关的SQLSTATE值，通过SQLSTATE值可以查出调用此函数的具体信息。

## 示例

参见：[示例](#)

### 9.4.5.6 SQLBindParameter

#### 功能描述

将一条SQL语句中的一个参数标志和一个缓冲区绑定起来。

#### 原型

```
SQLRETURN SQLBindParameter(SQLHSTMT StatementHandle,
    SQLUSMALLINT ParameterNumber,
    SQLSMALLINT InputOutputType,
    SQLSMALLINT ValuetType,
    SQLSMALLINT ParameterType,
    SQLULEN ColumnSize,
    SQLSMALLINT DecimalDigits,
    SQLPOINTER ParameterValuePtr,
    SQLLEN BufferLength,
    SQLLEN *StrLen_or_IndPtr);
```

## 参数

表 9-27 SQLBindParameter

关键词	参数说明
StatementHandle	语句句柄。
ParameterNumber	参数序号，起始为1，依次递增。
InputOutputType	输入输出参数类型。
ValueType	参数的C数据类型。
ParameterType	参数的SQL数据类型。
ColumnSize	列的大小或相应参数标记的表达式。
DecimalDigits	列的十进制数字或相应参数标记的表达式。
ParameterValuePtr	指向存储参数数据缓冲区的指针。
BufferLength	ParameterValuePtr指向缓冲区的长度，以字节为单位。
StrLen_or_IndPtr	缓冲区的长度或指示器指针。若为空值，则未使用任何长度或指示器值。

## 返回值

- SQL\_SUCCESS：表示调用正确。
- SQL\_SUCCESS\_WITH\_INFO：表示会有一些警告信息。
- SQL\_ERROR：表示比较严重的错误，如：内存分配失败、建立连接失败等。
- SQL\_INVALID\_HANDLE：表示调用无效句柄。其他API的返回值同理。

## 注意事项

当SQLBindCol返回SQL\_ERROR或SQL\_SUCCESS\_WITH\_INFO时，通过调用[SQLGetDiagRec](#)函数，并将HandleType和Handle参数设置为SQL\_HANDLE\_STMT和StatementHandle，可得到一个相关的SQLSTATE值，通过SQLSTATE值可以查出调用此函数的具体信息。

## 示例

参见：[示例](#)

### 9.4.5.7 SQLColAttribute

## 功能描述

返回结果集中一列的描述符信息。

## 原型

```
SQLRETURN SQLColAttribute(SQLHSTMT StatementHandle,
    SQLUSMALLINT ColumnNumber,
    SQLUSMALLINT FieldIdentifier,
    SQLPOINTER CharacterAttributePtr,
    SQLSMALLINT BufferLength,
    SQLSMALLINT *StringLengthPtr,
    SQLPOINTER NumericAttributePtr);
```

## 参数

表 9-28 SQLColAttribute 参数

关键字	参数说明
StatementHandle	语句句柄。
ColumnNumber	要检索字段的列号，起始为1，依次递增。
FieldIdentifier	IRD中ColumnNumber行的字段。
CharacterAttributePtr	<b>输出参数：</b> 一个缓冲区指针，返回FieldIdentifier字段值。
BufferLength	<ul style="list-style-type: none"> <li>如果FieldIdentifier是一个ODBC定义的字段，而且CharacterAttributePtr指向一个字符串或二进制缓冲区，则此参数为该缓冲区的长度。</li> <li>如果FieldIdentifier是一个ODBC定义的字段，而且CharacterAttributePtr指向一个整数，则会忽略该字段。</li> </ul>
StringLengthPtr	<b>输出参数：</b> 缓冲区指针，存放*CharacterAttributePtr中字符类型数据的字节总数，对于非字符类型，忽略BufferLength的值。
NumericAttributePtr	<b>输出参数：</b> 指向一个整型缓冲区的指针，返回IRD中ColumnNumber行FieldIdentifier字段的值。

## 返回值

- SQL\_SUCCESS：表示调用正确。
- SQL\_SUCCESS\_WITH\_INFO：表示会有一些警告信息。
- SQL\_ERROR：表示比较严重的错误，如：内存分配失败、建立连接失败等。
- SQL\_INVALID\_HANDLE：表示调用无效句柄。其他API的返回值同理。

## 注意事项

当SQLColAttribute返回SQL\_ERROR或SQL\_SUCCESS\_WITH\_INFO时，通过调用[SQLGetDiagRec](#)函数，并将HandleType和Handle参数设置为SQL\_HANDLE\_STMT和StatementHandle，可得到一个相关的SQLSTATE值，通过SQLSTATE值可以查出调用此函数的具体信息。

## 示例

参见：[示例](#)

### 9.4.5.8 SQLConnect

#### 功能描述

在驱动程序和数据源之间建立连接。连接上数据源之后，可以通过连接句柄访问到所有有关连接数据源的信息，包括程序运行状态、事务处理状态和错误信息。

#### 原型

```
SQLRETURN SQLConnect(SQLHDBC ConnectionHandle,
    SQLCHAR *ServerName,
    SQLSMALLINT NameLength1,
    SQLCHAR *UserName,
    SQLSMALLINT NameLength2,
    SQLCHAR *Authentication,
    SQLSMALLINT NameLength3);
```

#### 参数

表 9-29 SQLConnect 参数

关键字	参数说明
ConnectionHandle	连接句柄，通过SQLAllocHandle获得。
ServerName	要连接数据源的名称。
NameLength1	ServerName的长度。
UserName	数据源中数据库用户名。
NameLength2	UserName的长度。
Authentication	数据源中数据库用户密码。
NameLength3	Authentication的长度。

#### 返回值

- SQL\_SUCCESS：表示调用正确。
- SQL\_SUCCESS\_WITH\_INFO：表示会有一些警告信息。
- SQL\_ERROR：表示比较严重的错误，如：内存分配失败、建立连接失败等。
- SQL\_INVALID\_HANDLE：表示调用无效句柄。其他API的返回值同理。
- SQL\_STILL\_EXECUTING：表示语句正在执行。

#### 注意事项

当调用SQLConnect函数返回SQL\_ERROR或SQL\_SUCCESS\_WITH\_INFO时，通过调用[SQLGetDiagRec](#)函数，并将HandleType和Handle参数设置为SQL\_HANDLE\_DBC和



ConnectionHandle，可得到一个相关的SQLSTATE值，通过SQLSTATE值可以查出调用此函数的具体信息。

## 示例

参见：[示例](#)

### 9.4.5.9 SQLDisconnect

#### 功能描述

关闭一个与特定连接句柄相关的连接。

#### 原型

```
SQLRETURN SQLDisconnect(SQLHDBC ConnectionHandle);
```

#### 参数

表 9-30 SQLDisconnect 参数

关键字	参数说明
ConnectionHandle	连接句柄，通过SQLAllocHandle获得。

#### 返回值

- SQL\_SUCCESS：表示调用正确。
- SQL\_SUCCESS\_WITH\_INFO：表示会有一些警告信息。
- SQL\_ERROR：表示比较严重的错误，如：内存分配失败、建立连接失败等。
- SQL\_INVALID\_HANDLE：表示调用无效句柄。其他API的返回值同理。

#### 注意事项

当调用SQLDisconnect函数返回SQL\_ERROR或SQL\_SUCCESS\_WITH\_INFO时，通过调用[SQLGetDiagRec](#)函数，并将HandleType和Handle参数设置为SQL\_HANDLE\_DBC和ConnectionHandle，可得到一个相关的SQLSTATE值，通过SQLSTATE值可以查出调用此函数的具体信息。

## 示例

参见：[示例](#)

### 9.4.5.10 SQLExecDirect

#### 功能描述

使用参数的当前值，执行一条准备好的语句。对于一次只执行一条SQL语句，SQLExecDirect是最快的执行方式。

## 原型

```
SQLRETURN SQLExecDirect(SQLHSTMT StatementHandle,  
                        SQLCHAR *StatementText,  
                        SQLINTEGER TextLength);
```

## 参数

表 9-31 SQLExecDirect 参数

关键字	参数说明
StatementHandle	语句句柄，通过SQLAllocHandle获得。
StatementText	要执行的SQL语句。不支持一次执行多条语句。
TextLength	StatementText的长度。

## 返回值

- SQL\_SUCCESS：表示调用正确。
- SQL\_SUCCESS\_WITH\_INFO：表示会有一些警告信息。
- SQL\_NEED\_DATA：在执行SQL语句前没有提供足够的参数。
- SQL\_ERROR：表示比较严重的错误，如：内存分配失败、建立连接失败等。
- SQL\_INVALID\_HANDLE：表示调用无效句柄。其他API的返回值同理。
- SQL\_STILL\_EXECUTING：表示语句正在执行。
- SQL\_NO\_DATA：表示SQL语句不返回结果集。

## 注意事项

当调用SQLExecDirect函数返回SQL\_ERROR或SQL\_SUCCESS\_WITH\_INFO时，通过调用[SQLGetDiagRec](#)函数，并将HandleType和Handle参数设置为SQL\_HANDLE\_STMT和StatementHandle，可得到一个相关的SQLSTATE值，通过SQLSTATE值可以查出调用此函数的具体信息。

## 示例

参见：[示例](#)

### 9.4.5.11 SQLExecute

#### 功能描述

如果语句中存在参数标记的话，SQLExecute函数使用参数标记参数的当前值，执行一条准备好的SQL语句。

## 原型

```
SQLRETURN SQLExecute(SQLHSTMT StatementHandle);
```

## 参数

表 9-32 SQLExecute 参数

关键字	参数说明
StatementHandle	要执行语句的语句句柄。

## 返回值

- SQL\_SUCCESS: 表示调用正确。
- SQL\_SUCCESS\_WITH\_INFO: 表示会有一些警告信息。
- SQL\_NEED\_DATA: 表示在执行SQL语句前没有提供足够的参数。
- SQL\_ERROR: 表示比较严重的错误，如：内存分配失败、建立连接失败等。
- SQL\_NO\_DATA: 表示SQL语句不返回结果集。
- SQL\_INVALID\_HANDLE: 表示调用无效句柄。其他API的返回值同理。
- SQL\_STILL\_EXECUTING: 表示语句正在执行。

## 注意事项

当SQLExecute函数返回SQL\_ERROR或SQL\_SUCCESS\_WITH\_INFO时，可通过调用 [SQLGetDiagRec](#) 函数，并将HandleType和Handle参数设置为SQL\_HANDLE\_STMT和StatementHandle，可得到一个相关的SQLSTATE值，通过SQLSTATE值可以查出调用此函数的具体信息。

## 示例

参见：[示例](#)

### 9.4.5.12 SQLFetch

## 功能描述

从结果集中取下一个行集的数据，并返回所有被绑定列的数据。

## 原型

```
SQLRETURN SQLFetch(SQLHSTMT StatementHandle);
```

## 参数

表 9-33 SQLFetch 参数

关键字	参数说明
StatementHandle	语句句柄，通过SQLAllocHandle获得。

## 返回值

- SQL\_SUCCESS：表示调用正确。
- SQL\_SUCCESS\_WITH\_INFO：表示会有一些警告信息。
- SQL\_ERROR：表示比较严重的错误，如：内存分配失败、建立连接失败等。
- SQL\_NO\_DATA：表示SQL语句不返回结果集。
- SQL\_INVALID\_HANDLE：表示调用无效句柄。其他API的返回值同理。
- SQL\_STILL\_EXECUTING：表示语句正在执行。

## 注意事项

当调用SQLFetch函数返回SQL\_ERROR或SQL\_SUCCESS\_WITH\_INFO时，通过调用[SQLGetDiagRec](#)函数，并将HandleType和Handle参数设置为SQL\_HANDLE\_STMT和StatementHandle，可得到一个相关的SQLSTATE值，通过SQLSTATE值可以查出调用此函数的具体信息。

## 示例

参见：[示例](#)

### 9.4.5.13 SQLFreeStmt

在ODBC 3.x版本中，ODBC 2.x的函数SQLFreeStmt已被SQLFreeHandle代替。有关详细信息请参阅[SQLFreeHandle](#)。

### 9.4.5.14 SQLFreeConnect

在ODBC 3.x版本中，ODBC 2.x的函数SQLFreeConnect已被SQLFreeHandle代替。有关详细信息请参阅[SQLFreeHandle](#)。

### 9.4.5.15 SQLFreeHandle

## 功能描述

释放与指定环境、连接、语句或描述符相关联的资源，它替代了ODBC 2.x函数SQLFreeEnv、SQLFreeConnect及SQLFreeStmt。

## 原型

```
SQLRETURN SQLFreeHandle(SQLSMALLINT HandleType,  
                        SQLHANDLE Handle);
```

## 参数

表 9-34 SQLFreeHandle 参数

关键字	参数说明
HandleType	SQLFreeHandle要释放的句柄类型。必须为下列值之一： <ul style="list-style-type: none"><li>• SQL_HANDLE_ENV</li><li>• SQL_HANDLE_DBC</li><li>• SQL_HANDLE_STMT</li><li>• SQL_HANDLE_DESC</li></ul> 如果HandleType不是这些值之一，SQLFreeHandle返回SQL_INVALID_HANDLE。
Handle	要释放的句柄。

## 返回值

- SQL\_SUCCESS：表示调用正确。
- SQL\_SUCCESS\_WITH\_INFO：表示会有一些警告信息。
- SQL\_ERROR：表示比较严重的错误，如：内存分配失败、建立连接失败等。
- SQL\_INVALID\_HANDLE：表示调用无效句柄。其他API的返回值同理。

## 注意事项

如果SQLFreeHandle返回SQL\_ERROR，句柄仍然有效。

## 示例

参见：[示例](#)

### 9.4.5.16 SQLFreeEnv

在ODBC 3.x版本中，ODBC 2.x的函数SQLFreeEnv已被SQLFreeHandle代替。有关详细信息请参阅[SQLFreeHandle](#)。

### 9.4.5.17 SQLPrepare

## 功能描述

准备一个将要进行的SQL语句。

## 原型

```
SQLRETURN SQLPrepare(SQLHSTMT StatementHandle,  
SQLCHAR *StatementText,  
SQLINTEGER TextLength);
```

## 参数

表 9-35 SQLPrepare 参数

关键字	参数说明
StatementHandle	语句句柄。
StatementText	SQL文本串。
TextLength	StatementText的长度。

## 返回值

- SQL\_SUCCESS：表示调用正确。
- SQL\_SUCCESS\_WITH\_INFO：表示会有一些警告信息。
- SQL\_ERROR：表示比较严重的错误，如：内存分配失败、建立连接失败等。
- SQL\_INVALID\_HANDLE：表示调用无效句柄。其他API的返回值同理。
- SQL\_STILL\_EXECUTING：表示语句正在执行。

## 注意事项

当SQLPrepare返回的值为SQL\_ERROR或SQL\_SUCCESS\_WITH\_INFO时，通过调用[SQLGetDiagRec](#)函数，并将HandleType和Handle参数分别设置为SQL\_HANDLE\_STMT和StatementHandle，可得到一个相关的SQLSTATE值，通过SQLSTATE值可以查出调用此函数的具体信息。

## 示例

参见：[示例](#)

### 9.4.5.18 SQLGetData

#### 功能描述

SQLGetData返回结果集中某一列的数据。可以多次调用它来部分地检索不定长度的数据。

#### 原型

```
SQLRETURN SQLGetData(SQLHSTMT StatementHandle,
SQLUSMALLINT Col_or_Param_Num,
SQLSMALLINT TargetType,
SQLPOINTER TargetValuePtr,
SQLLEN BufferLength,
SQLLEN *StrLen_or_IndPtr);
```

## 参数

表 9-36 SQLGetData 参数

关键字	参数说明
StatementHandle	语句句柄，通过SQLAllocHandle获得。
Col_or_Param_Num	要返回数据的列号。结果集的列按增序从1开始编号。书签列的列号为0。
TargetType	TargetValuePtr缓冲中的C数据类型的类型标识符。若TargetType为SQL_ARD_TYPE，驱动使用ARD中SQL_DESC_CONCISE_TYPE字段的类型标识符。若为SQL_C_DEFAULT，驱动根据源的SQL数据类型选择缺省的数据类型。
TargetValuePtr	<b>输出参数：</b> 指向返回数据所在缓冲区的指针。
BufferLength	TargetValuePtr所指向缓冲区的长度。
StrLen_or_IndPtr	<b>输出参数：</b> 指向缓冲区的指针，在此缓冲区中返回长度或标识符的值。

## 返回值

- SQL\_SUCCESS：表示调用正确。
- SQL\_SUCCESS\_WITH\_INFO：表示会有一些警告信息。
- SQL\_ERROR：表示比较严重的错误，如：内存分配失败、建立连接失败等。
- SQL\_NO\_DATA：表示SQL语句不返回结果集。
- SQL\_INVALID\_HANDLE：表示调用无效句柄。其他API的返回值同理。
- SQL\_STILL\_EXECUTING：表示语句正在执行。

## 注意事项

当调用SQLFetch函数返回SQL\_ERROR或SQL\_SUCCESS\_WITH\_INFO时，通过调用SQLGetDiagRec函数，并将HandleType和Handle参数分别设置为SQL\_HANDLE\_STMT和StatementHandle，可得到一个相关的SQLSTATE值，通过SQLSTATE值可以查出调用此函数的具体信息。

## 示例

参见：[示例](#)

### 9.4.5.19 SQLGetDiagRec

#### 功能描述

返回诊断记录的多个字段的当前值，其中诊断记录包含错误、警告及状态信息。

## 原型

```
SQLRETURN SQLGetDiagRec(SQLSMALLINT HandleType
                        SQLHANDLE Handle,
                        SQLSMALLINT RecNumber,
                        SQLCHAR *SQLState,
                        SQLINTEGER *NativeErrorPtr,
                        SQLCHAR *MessageText,
                        SQLSMALLINT BufferLength
                        SQLSMALLINT *TextLengthPtr);
```

## 参数

表 9-37 SQLGetDiagRec 参数

关键字	参数说明
HandleType	句柄类型标识符，它说明诊断所要求的句柄类型。必须为下列值之一： <ul style="list-style-type: none"> <li>• SQL_HANDLE_ENV</li> <li>• SQL_HANDLE_DBC</li> <li>• SQL_HANDLE_STMT</li> <li>• SQL_HANDLE_DESC</li> </ul>
Handle	诊断数据结构的句柄，其类型由HandleType来指出。如果HandleType是SQL_HANDLE_ENV，Handle可以是共享的或非共享的环境句柄。
RecNumber	指出应用从查找信息的状态记录。状态记录从1开始编号。
SQLState	<b>输出参数：</b> 指向缓冲区的指针，该缓冲区存储着有关RecNumber的五字符的SQLSTATE码。
NativeErrorPtr	<b>输出参数：</b> 指向缓冲区的指针，该缓冲区存储着本地的错误码。
MessageText	指向缓冲区的指针，该缓冲区存储着诊断信息文本串。
BufferLength	MessageText的长度。
TextLengthPtr	<b>输出参数：</b> 指向缓冲区的指针，返回MessageText中的字节总数。如果返回字节数大于BufferLength，则MessageText中的诊断信息文本被截断成BufferLength减去NULL结尾字符的长度。

## 返回值

- SQL\_SUCCESS：表示调用正确。
- SQL\_SUCCESS\_WITH\_INFO：表示会有一些警告信息。
- SQL\_ERROR：表示比较严重的错误，如：内存分配失败、建立连接失败等。
- SQL\_INVALID\_HANDLE：表示调用无效句柄。其他API的返回值同理。

## 注意事项

SQLGetDiagRec不发布自己的诊断记录。它用下列返回值来报告它自己的执行结果：



- SQL\_SUCCESS：函数成功返回诊断信息。
- SQL\_SUCCESS\_WITH\_INFO：\*MessageText太小以致不能容纳所请求的诊断信息。没有诊断记录生成。
- SQL\_INVALID\_HANDLE：由HandType和Handle所指出的句柄是不合法句柄。
- SQL\_ERROR：RecNumber小于等于0或BufferLength小于0。

如果调用ODBC函数返回SQL\_ERROR或SQL\_SUCCESS\_WITH\_INFO，可调用SQLGetDiagRec返回诊断信息值SQLSTATE，SQLSTATE值的如下表。

表 9-38 SQLSTATE 值

SQLSTATE	错误	描述
HY000	一般错误	未定义特定的SQLSTATE所产生的一个错误。
HY001	内存分配错误	驱动程序不能分配所需要的内存来支持函数的执行或完成。
HY008	取消操作	调用SQLCancel取消执行语句后，依然在StatementHandle上调用函数。
HY010	函数系列错误	在为执行中的所有数据参数或列发送数据前就调用了执行函数。
HY013	内存管理错误	不能处理函数调用，可能由当前内存条件差引起。
HYT01	连接超时	数据源响应请求之前，连接超时。
IM001	驱动程序不支持此函数	调用了StatementHandle相关的驱动程序不支持的函数

## 示例

参见：[示例](#)

### 9.4.5.20 SQLSetConnectAttr

#### 功能描述

设置控制连接各方面的属性。

#### 原型

```
SQLRETURN SQLSetConnectAttr(SQLHDBC ConnectionHandle,
                             SQLINTEGER Attribute,
                             SQLPOINTER ValuePtr,
                             SQLINTEGER StringLength);
```

## 参数

表 9-39 SQLSetConnectAttr 参数

关键字	参数说明
StatementHandle	连接句柄。
Attribute	设置属性。
ValuePtr	指向对应Attribute的值。依赖于Attribute的值，ValuePtr是32位无符号整型值或指向以空结束的字符串。注意，如果ValuePtr参数是驱动程序指定值。ValuePtr可能是有符号的整数。
StringLength	如果ValuePtr指向字符串或二进制缓冲区，这个参数是*ValuePtr长度，如果ValuePtr指向整型，忽略StringLength。

## 返回值

- SQL\_SUCCESS: 表示调用正确。
- SQL\_SUCCESS\_WITH\_INFO: 表示会有一些警告信息。
- SQL\_ERROR: 表示比较严重的错误，如：内存分配失败、建立连接失败等。
- SQL\_INVALID\_HANDLE: 表示调用无效句柄。其他API的返回值同理。

## 注意事项

当SQLSetConnectAttr的返回值为SQL\_ERROR或SQL\_SUCCESS\_WITH\_INFO时，通过借助SQL\_HANDLE\_DBC的HandleType和ConnectionHandle的Handle，调用[SQLGetDiagRec](#)可得到相关的SQLSTATE值，通过SQLSTATE值可以查出调用此函数的具体信息。

## 示例

参见：[示例](#)

### 9.4.5.21 SQLSetEnvAttr

## 功能描述

设置控制环境各方面的属性。

## 原型

```
SQLRETURN SQLSetEnvAttr(SQLHENV EnvironmentHandle,
                        SQLINTEGER Attribute,
                        SQLPOINTER ValuePtr,
                        SQLINTEGER StringLength);
```

## 参数

表 9-40 SQLSetEnvAttr 参数

关键字	参数说明
EnvironmentHandle	环境句柄。
Attribute	需设置的环境属性，可为如下值： <ul style="list-style-type: none"> <li>SQL_ATTR_ODBC_VERSION：指定ODBC版本。</li> <li>SQL_CONNECTION_POOLING：连接池属性。</li> <li>SQL_OUTPUT_NTS：指明驱动器返回字符串的形式。</li> </ul>
ValuePtr	指向对应Attribute的值。依赖于Attribute的值，ValuePtr可能是32位整型值，或为以空结束的字符串。
StringLength	如果ValuePtr指向字符串或二进制缓冲区，这个参数是*ValuePtr长度，如果ValuePtr指向整型，忽略StringLength。

## 返回值

- SQL\_SUCCESS：表示调用正确。
- SQL\_SUCCESS\_WITH\_INFO：表示会有一些警告信息。
- SQL\_ERROR：表示比较严重的错误，如：内存分配失败、建立连接失败等。
- SQL\_INVALID\_HANDLE：表示调用无效句柄。其他API的返回值同理。

## 注意事项

当SQLSetEnvAttr的返回值为SQL\_ERROR或SQL\_SUCCESS\_WITH\_INFO时，通过借助SQL\_HANDLE\_ENV的HandleType和EnvironmentHandle的Handle，调用[SQLGetDiagRec](#)可得到相关的SQLSTATE值，通过SQLSTATE值可以查出调用此函数的具体信息。

## 示例

参见：[示例](#)

### 9.4.5.22 SQLSetStmtAttr

## 功能描述

设置相关语句的属性。

## 原型

```
SQLRETURN SQLSetStmtAttr(SQLHSTMT StatementHandle,
                          SQLINTEGER Attribute,
                          SQLPOINTER ValuePtr,
                          SQLINTEGER StringLength);
```

## 参数

表 9-41 SQLSetStmtAttr 参数

关键字	参数说明
StatementHandle	语从句柄。
Attribute	需设置的属性。
ValuePtr	指向对应Attribute的值。依赖于Attribute的值，ValuePtr可能是32位无符号整型值，或指向以空结束的字符串，二进制缓冲区，或者驱动定义值。注意，如果ValuePtr参数是驱动程序指定值。ValuePtr可能是有符号的整数。
StringLength	如果ValuePtr指向字符串或二进制缓冲区，这个参数是*ValuePtr长度，如果ValuePtr指向整型，忽略StringLength。

## 返回值

- SQL\_SUCCESS：表示调用正确。
- SQL\_SUCCESS\_WITH\_INFO：表示会有一些警告信息。
- SQL\_ERROR：表示比较严重的错误，如：内存分配失败、建立连接失败等。
- SQL\_INVALID\_HANDLE：表示调用无效句柄。其他API的返回值同理。

## 注意事项

当SQLSetStmtAttr的返回值为SQL\_ERROR或SQL\_SUCCESS\_WITH\_INFO时，通过借助SQL\_HANDLE\_STMT的HandleType和StatementHandle的Handle，调用[SQLGetDiagRec](#)可得到相关的SQLSTATE值，通过SQLSTATE值可以查出调用此函数的具体信息。

## 示例

参见：[示例](#)

# 10 PostGIS Extension

## 10.1 PostGIS 概述

### 📖 说明

- PostGIS Extension依赖的第三方软件需要用户进行单独安装，用户如需使用PostGIS功能，请提交工单或联系技术支持人员提交申请。
- 如果用户在使用中出现“ERROR: EXTENSION is not yet supported.”这种报错，则表示没有安装PostGIS软件包，请联系技术支持进行获取。

GaussDB(DWS)提供PostGIS Extension（版本为PostGIS-2.4.2）。PostGIS Extension是PostgreSQL的空间数据库扩展，提供如下空间信息服务功能：空间对象、空间索引、空间操作函数和空间操作符。PostGIS Extension完全遵循OpenGIS规范。

GaussDB(DWS)中PostGIS Extension依赖第三方开源软件如下。

- Geos 3.6.2
- Proj 4.9.2
- Json 0.12.1
- Libxml2 2.7.1
- Gdal 1.11.0

## 10.2 PostGIS 使用

### 📖 说明

- PostGIS Extension依赖的第三方软件需要用户进行单独安装，用户如需使用PostGIS功能，请提交工单或联系技术支持人员提交申请。
- 如果用户在使用中出现“ERROR: EXTENSION is not yet supported.”这种报错，则表示没有安装PostGIS软件包，请联系技术支持进行获取。

### 创建 Extension

创建PostGIS Extension可直接使用CREATE EXTENSION命令进行创建：

```
CREATE EXTENSION postgis;
```

## 使用 Extension

PostGIS Extension函数调用格式为：

```
SELECT GisFunction (Param1, Param2,.....);
```

其中GisFunction为函数名，Param1、Param2等为函数参数名。下列SQL语句展示PostGIS的简单使用，对于各函数的具体使用，请参考《[PostGIS-2.4.2用户手册](#)》。

示例1：几何表的创建。

```
CREATE TABLE cities ( id integer, city_name varchar(50) );  
SELECT AddGeometryColumn('cities', 'position', 4326, 'POINT', 2);
```

示例2：几何数据的插入。

```
INSERT INTO cities (id, position, city_name) VALUES (1,ST_GeomFromText('POINT(-9.5 23)',4326),'CityA');  
INSERT INTO cities (id, position, city_name) VALUES (2,ST_GeomFromText('POINT(-10.6 40.3)',4326),'CityB');  
INSERT INTO cities (id, position, city_name) VALUES (3,ST_GeomFromText('POINT(20.8 30.3)',4326),'CityC');
```

示例3：计算三个城市间任意两个城市距离。

```
SELECT p1.city_name,p2.city_name,ST_Distance(p1.position,p2.position) FROM cities AS p1, cities AS p2  
WHERE p1.id > p2.id;
```

## 删除 Extension

在GaussDB(DWS)中删除PostGIS Extension的方法如下所示：

```
DROP EXTENSION postgis [CASCADE];
```

### 说明

如果Extension被其它对象依赖（如创建的几何表），需要加入CASCADE（级联）关键字，删除所有依赖对象。

## 10.3 PostGIS 支持和限制

### 支持数据类型

GaussDB(DWS)的PostGIS Extension支持如下数据类型：

- box2d
- box3d
- geometry\_dump
- geometry
- geography
- raster

### 说明

创建Postgis和使用Postgis非同一用户时，请设置如下GUC参数：  
SET behavior\_compat\_options = 'bind\_procedure\_searchpath';

## 支持的操作符和函数列表

表 10-1 PostGIS 支持的操作符和函数列表

函数分类	包含函数
Management Functions	AddGeometryColumn、DropGeometryColumn、DropGeometryTable、PostGIS_Full_Version、PostGIS_GEOS_Version、PostGIS_Liblwgeom_Version、PostGIS_Lib_Build_Date、PostGIS_Lib_Version、PostGIS_PROJ_Version、PostGIS_Scripts_Build_Date、PostGIS_Scripts_Installed、PostGIS_Version、PostGIS_LibXML_Version、PostGIS_Scripts_Released、Populate_Geometry_Columns、UpdateGeometrySRID
Geometry Constructors	ST_BdPolyFromText、ST_BdMPolyFromText、ST_Box2dFromGeoHash、ST_GeogFromText、ST_GeographyFromText、ST_GeogFromWKB、ST_GeomCollFromText、ST_GeomFromEWKB、ST_GeomFromEWKT、ST_GeometryFromText、ST_GeomFromGeoHash、ST_GeomFromGML、ST_GeomFromGeoJSON、ST_GeomFromKML、ST_GMLToSQL、ST_GeomFromText、ST_GeomFromWKB、ST_LineFromMultiPoint、ST_LineFromText、ST_LineFromWKB、ST_LinestringFromWKB、ST_MakeBox2D、ST_3DMakeBox、ST_MakeEnvelope、ST_MakePolygon、ST_MakePoint、ST_MakePointM、ST_MLineFromText、ST_MPointFromText、ST_MPolyFromText、ST_Point、ST_PointFromGeoHash、ST_PointFromText、ST_PointFromWKB、ST_Polygon、ST_PolygonFromText、ST_WKBToSQL、ST_WKTToSQL
Geometry Accessors	GeometryType、ST_Boundary、ST_CoordDim、ST_Dimension、ST_EndPoint、ST_Envelope、ST_ExteriorRing、ST_GeometryN、ST_GeometryType、ST_InteriorRingN、ST_IsClosed、ST_IsCollection、ST_IsEmpty、ST_IsRing、ST_IsSimple、ST_IsValid、ST_IsValidReason、ST_IsValidDetail、ST_M、ST_NDims、ST_NPoints、ST_NRings、ST_NumGeometries、ST_NumInteriorRings、ST_NumInteriorRing、ST_NumPatches、ST_NumPoints、ST_PatchN、ST_PointN、ST_SRID、ST_StartPoint、ST_Summary、ST_X、ST_XMax、ST_XMin、ST_Y、ST_YMax、ST_YMin、ST_Z、ST_ZMax、ST_Zmflag、ST_ZMin
Geometry Editors	ST_AddPoint、ST_Affine、ST_Force2D、ST_Force3D、ST_Force3DZ、ST_Force3DM、ST_Force4D、ST_ForceCollection、ST_ForceSFS、ST_ForceRHR、ST_LineMerge、ST_CollectionExtract、ST_CollectionHomogenize、ST_Multi、ST_RemovePoint、ST_Reverse、ST_Rotate、ST_RotateX、ST_RotateY、ST_RotateZ、ST_Scale、ST_Segmentize、ST_SetPoint、ST_SetSRID、ST_SnapToGrid、ST_Snap、ST_Transform、ST_Translate、ST_TransScale

函数分类	包含函数
Geometry Outputs	ST_AsBinary、ST_AsEWKB、ST_AsEWKT、ST_AsGeoJSON、ST_AsGML、ST_AsHEXEWKB、ST_AsKML、ST_AsLatLonText、ST_AsSVG、ST_AsText、ST_AsX3D、ST_GeoHash
Operators	&&、&&&、&<、&< 、&>、<<、<< 、=、>>、@、 &>、 >>、~、~=、<->、<#>
Spatial Relationships and Measurements	ST_3DClosestPoint、ST_3DDistance、ST_3DDWithin、ST_3DDFullyWithin、ST_3DIntersects、ST_3DLongestLine、ST_3DMaxDistance、ST_3DShortestLine、ST_Area、ST_Azimuth、ST_Centroid、ST_ClosestPoint、ST_Contains、ST_ContainsProperly、ST_Covers、ST_CoveredBy、ST_Crosses、ST_LineCrossingDirection、ST_Disjoint、ST_Distance、ST_HausdorffDistance、ST_MaxDistance、ST_DistanceSphere、ST_DistanceSpheroid、ST_DFullyWithin、ST_DWithin、ST_Equals、ST_HasArc、ST_Intersects、ST_Length、ST_Length2D、ST_3DLength、ST_Length_Spheroid、ST_Length2D_Spheroid、ST_3DLength_Spheroid、ST_LongestLine、ST_OrderingEquals、ST_Overlaps、ST_Perimeter、ST_Perimeter2D、ST_3DPerimeter、ST_PointOnSurface、ST_Project、ST_Relate、ST_RelateMatch、ST_ShortestLine、ST_Touches、ST_Within
Geometry Processing	ST_Buffer、ST_BuildArea、ST_Collect、ST_ConcaveHull、ST_ConvexHull、ST_CurveToLine、ST_DelaunayTriangles、ST_Difference、ST_Dump、ST_DumpPoints、ST_DumpRings、ST_FlipCoordinates、ST_Intersection、ST_LineToCurve、ST_MakeValid、ST_MemUnion、ST_MinimumBoundingCircle、ST_Polygonize、ST_Node、ST_OffsetCurve、ST_RemoveRepeatedPoints、ST_SharedPaths、ST_Shift_Longitude、ST_Simplify、ST_SimplifyPreserveTopology、ST_Split、ST_SymDifference、ST_Union、ST_UnaryUnion
Linear Referencing	ST_LineInterpolatePoint、ST_LineLocatePoint、ST_LineSubstring、ST_LocateAlong、ST_LocateBetween、ST_LocateBetweenElevations、ST_InterpolatePoint、ST_AddMeasure
Miscellaneous Functions	ST_Accum、Box2D、Box3D、ST_Expand、ST_Extent、ST_3DExtent、Find_SRID、ST_MemSize
Exceptional Functions	PostGIS_AddBBox、PostGIS_DropBBox、PostGIS_HasBBox
Raster Management Functions	AddRasterConstraints、DropRasterConstraints、AddOverviewConstraints、DropOverviewConstraints、PostGIS_GDAL_Version、PostGIS_Raster_Lib_Build_Date、PostGIS_Raster_Lib_Version、ST_GDALDrivers、UpdateRasterSRID



函数分类	包含函数
Raster Constructors	ST_AddBand、ST_AsRaster、ST_Band、ST_MakeEmptyRaster、ST_Tile、ST_FromGDALRaster
Raster Accessors	ST_GeoReference、ST_Height、ST_IsEmpty、ST_MetaData、ST_NumBands、ST_PixelHeight、ST_PixelWidth、ST_ScaleX、ST_ScaleY、ST_RasterToWorldCoord、ST_RasterToWorldCoordX、ST_RasterToWorldCoordY、ST_Rotation、ST_SkewX、ST_SkewY、ST_SRID、ST_Summary、ST_UpperLeftX、ST_UpperLeftY、ST_Width、ST_WorldToRasterCoord、ST_WorldToRasterCoordX、ST_WorldToRasterCoordY
Raster Band Accessors	ST_BandMetaData、ST_BandNoDataValue、ST_BandIsNoData、ST_BandPath、ST_BandPixelType、ST_HasNoBand
Raster Pixel Accessors and Setters	ST_PixelAsPolygon、ST_PixelAsPolygons、ST_PixelAsPoint、ST_PixelAsPoints、ST_PixelAsCentroid、ST_PixelAsCentroids、ST_Value、ST_NearestValue、ST_Neighborhood、ST_SetValue、ST_SetValues、ST_DumpValues、ST_PixelOfValue
Raster Editors	ST_SetGeoReference、ST_SetRotation、ST_SetScale、ST_SetSkew、ST_SetSRID、ST_SetUpperLeft、ST_Resample、ST_Rescale、ST_Reskew、ST_SnapToGrid、ST_Resize、ST_Transform
Raster Band Editors	ST_SetBandNoDataValue、ST_SetBandIsNoData
Raster Band Statistics and Analytics	ST_Count、ST_CountAgg、ST_Histogram、ST_Quantile、ST_SummaryStats、ST_SummaryStatsAgg、ST_ValueCount
Raster Outputs	ST_AsBinary、ST_AsGDALRaster、ST_AsJPEG、ST_AsPNG、ST_AsTIFF
Raster Processing	ST_Clip、ST_ColorMap、ST_Intersection、ST_MapAlgebra、ST_Reclass、ST_Union、ST_Distinct4ma、ST_InvDistWeight4ma、ST_Max4ma、ST_Mean4ma、ST_Min4ma、ST_MinDist4ma、ST_Range4ma、ST_StdDev4ma、ST_Sum4ma、ST_Aspect、ST_HillShade、ST_Roughness、ST_Slope、ST_TPI、ST_TRI、Box3D、ST_ConvexHull、ST_DumpAsPolygons、ST_Envelope、ST_MinConvexHull、ST_Polygon、ST_Contains、ST_ContainsProperly、ST_Covers、ST_CoveredBy、ST_Disjoint、ST_Intersects、ST_Overlaps、ST_Touches、ST_SameAlignment、ST_NotSameAlignmentReason、ST_Within、ST_DWithin、ST_DFullyWithin
Raster Operators	&&、&<、&>、=、@、~=、~

## 空间索引

GaussDB(DWS)数据库的PostGIS Extension支持GIST (Generalized Search Tree) 空间索引（分区表除外）。相比于B-tree索引，GIST索引适应于任意类型的非常规数据结构，可有效提高几何和地理数据信息的检索效率。

使用如下命令创建GIST索引：

```
CREATE INDEX indexname ON tablename USING GIST ( geometryfield );
```

## 扩展限制

- 只支持行存表。
- 只支持Oracle兼容格式数据库。
- 不支持拓扑对象管理模块Topology。
- 不支持BRIN索引。
- `spatial_ref_sys`表在扩容期间只支持查询操作。

## 10.4 OPEN SOURCE SOFTWARE NOTICE (For PostGIS)

This document contains open source software notice for the product. And this document is confidential information of copyright holder. Recipient shall protect it in due care and shall not disseminate it without permission.

### Warranty Disclaimer

This document is provided “as is” without any warranty whatsoever, including the accuracy or comprehensiveness. Copyright holder of this document may change the contents of this document at any time without prior notice, and copyright holder disclaims any liability in relation to recipient’s use of this document.

Open source software is provided by the author “as is” and any express or implied warranties, including, but not limited to, the implied warranties of merchantability and fitness for a particular purpose are disclaimed. In no event shall the author be liable for any direct, indirect, incidental, special, exemplary, or consequential damages (including, but not limited to, procurement of substitute goods or services; loss of data or profits; or business interruption) however caused and on any theory of liability, whether in contract, strict liability, or tort (including negligence or otherwise) arising in any way out of the use of open source software, even if advised of the possibility of such damage.

### Copyright Notice And License Texts

Software: postgis-2.4.2

Copyright notice:

"Copyright (C) 1996-2015 Free Software Foundation, Inc.

Copyright (C) 1989, 1991 Free Software Foundation, Inc.,

51 Franklin Street, Fifth Floor, Boston, MA 02110-1301

Copyright 2008 Kevin Neufeld

Copyright (c) 2009 Walter Bruce Sinclair  
Copyright 2006-2013 Stephen Woodbridge.  
Copyright (c) 2008 Walter Bruce Sinclair  
Copyright (c) 2012 TJ Holowaychuk <tj@vision-media.ca>  
Copyright (c) 2008, by Attractive Chaos <attractivechaos@aol.co.uk>  
Copyright (c) 2001-2012 Walter Bruce Sinclair  
Copyright (c) 2010 Walter Bruce Sinclair  
Copyright 2006 Stephen Woodbridge  
Copyright 2006-2010 Stephen Woodbridge.  
Copyright (c) 2006-2014 Stephen Woodbridge.  
Copyright (c) 2017, Even Rouault <even.rouault at spatialys.com>  
Copyright (C) 2004-2015 Sandro Santilli <strk@kbt.io>  
Copyright (C) 2008-2011 Paul Ramsey <pramsey@cleverelephant.ca>  
Copyright (C) 2008 Mark Cave-Ayland <mark.cave-ayland@siriusit.co.uk>  
Copyright 2015 Nicklas Avén <nicklas.aven@jordogskog.no>  
Copyright 2008 Paul Ramsey  
Copyright (C) 2012 Sandro Santilli <strk@kbt.io>  
Copyright 2012 Sandro Santilli <strk@kbt.io>  
Copyright (C) 2014 Sandro Santilli <strk@kbt.io>  
Copyright 2013 Olivier Courtin <olivier.courtin@oslandia.com>  
Copyright 2009 Paul Ramsey <pramsey@cleverelephant.ca>  
Copyright 2008 Paul Ramsey <pramsey@cleverelephant.ca>  
Copyright 2011 Sandro Santilli <strk@kbt.io>  
Copyright 2015 Daniel Baston  
Copyright 2009 Olivier Courtin <olivier.courtin@oslandia.com>  
Copyright 2014 Kashif Rasul <kashif.rasul@gmail.com> and  
Shoaib Burq <saburq@gmail.com>  
Copyright 2013 Sandro Santilli <strk@kbt.io>  
Copyright 2010 Paul Ramsey <pramsey@cleverelephant.ca>  
Copyright (C) 2017 Sandro Santilli <strk@kbt.io>  
Copyright (C) 2015 Sandro Santilli <strk@kbt.io>  
Copyright (C) 2009 Paul Ramsey <pramsey@cleverelephant.ca>  
Copyright (C) 2011 Sandro Santilli <strk@kbt.io>

Copyright 2010 Olivier Courtin <olivier.courtin@oslandia.com>  
Copyright 2014 Nicklas Avén  
Copyright 2011-2016 Regina Obe  
Copyright (C) 2008 Paul Ramsey  
Copyright (C) 2011-2015 Sandro Santilli <strk@kbt.io>  
Copyright 2010-2012 Olivier Courtin <olivier.courtin@oslandia.com>  
Copyright (C) 2015 Daniel Baston <dbaston@gmail.com>  
Copyright (C) 2013 Nicklas Avén  
Copyright (C) 2016 Sandro Santilli <strk@kbt.io>  
Copyright 2017 Darafei Praliaskouski <me@komzpa.net>  
Copyright (c) 2016, Paul Ramsey <pramsey@cleverelephant.ca>  
Copyright (C) 2011-2012 Sandro Santilli <strk@kbt.io>  
Copyright (C) 2011 Paul Ramsey <pramsey@cleverelephant.ca>  
Copyright (C) 2007-2008 Mark Cave-Ayland  
Copyright (C) 2001-2006 Refrations Research Inc.  
Copyright 2015 Daniel Baston <dbaston@gmail.com>  
Copyright 2009 David Skea <David.Skea@gov.bc.ca>  
Copyright (C) 2012-2015 Paul Ramsey <pramsey@cleverelephant.ca>  
Copyright (C) 2012-2015 Sandro Santilli <strk@kbt.io>  
Copyright 2001-2006 Refrations Research Inc.  
Copyright (C) 2004 Refrations Research Inc.  
Copyright 2011-2014 Sandro Santilli <strk@kbt.io>  
Copyright 2009-2010 Sandro Santilli <strk@kbt.io>  
Copyright 2015-2016 Daniel Baston <dbaston@gmail.com>  
Copyright 2011-2015 Sandro Santilli <strk@kbt.io>  
Copyright 2007-2008 Mark Cave-Ayland  
Copyright 2012-2013 Oslandia <infos@oslandia.com>  
Copyright (C) 2015-2017 Sandro Santilli <strk@kbt.io>  
Copyright (C) 2001-2003 Refrations Research Inc.  
Copyright 2016 Sandro Santilli <strk@kbt.io>  
Copyright 2011 Kashif Rasul <kashif.rasul@gmail.com>  
Copyright (C) 2014 Nicklas Avén  
Copyright (C) 2010 Paul Ramsey <pramsey@cleverelephant.ca>

Copyright (C) 2010-2015 Paul Ramsey <pramsey@cleverelephant.ca>  
Copyright (C) 2011 Sandro Santilli <strk@kbt.io>  
Copyright (C) 2011-2014 Sandro Santilli <strk@kbt.io>  
Copyright (C) 1984, 1989-1990, 2000-2015 Free Software Foundation, Inc.  
Copyright (C) 2011 Paul Ramsey  
Copyright 2001-2003 Refrations Research Inc.  
Copyright 2009-2010 Olivier Courtin <olivier.courtin@oslandia.com>  
Copyright 2010-2012 Oslandia  
Copyright 2006 Corporacion Autonoma Regional de Santander  
Copyright 2013 Nicklas Avén  
Copyright 2011-2016 Arrival 3D, Regina Obe  
Copyright (C) 2009 David Skea <David.Skea@gov.bc.ca>  
Copyright (C) 2017 Sandro Santilli <strk@kbt.io>  
Copyright (C) 2009-2012 Paul Ramsey <pramsey@cleverelephant.ca>  
Copyright (C) 2010 - Oslandia  
Copyright (C) 2006 Mark Leslie <mark.leslie@lisssoft.com>  
Copyright (C) 2008-2009 Mark Cave-Ayland <mark.cave-ayland@siriusit.co.uk>  
Copyright (C) 2009-2015 Paul Ramsey <pramsey@cleverelephant.ca>  
Copyright (C) 2010 Olivier Courtin <olivier.courtin@camptocamp.com>  
Copyright 2010 Nicklas Avén  
Copyright 2012 Paul Ramsey  
Copyright 2011 Nicklas Avén  
Copyright 2002 Thamer Alharbash  
Copyright 2011 OSGeo  
Copyright (C) 2009-2011 Paul Ramsey <pramsey@cleverelephant.ca>  
Copyright (C) 2008 Mark Cave-Ayland <mark.cave-ayland@siriusit.co.uk>  
Copyright (C) 2004-2007 Refrations Research Inc.  
Copyright 2010 LISAssoft Pty Ltd  
Copyright 2010 Mark Leslie  
Copyright (c) 1999, Frank Warmerdam  
Copyright 2009 Mark Cave-Ayland <mark.cave-ayland@siriusit.co.uk>  
Copyright (c) 2007, Frank Warmerdam  
Copyright 2008 OpenGeo.org

Copyright (C) 2008 OpenGeo.org  
Copyright (C) 2009 Mark Cave-Ayland <mark.cave-ayland@siriusit.co.uk>  
Copyright 2010 LISAsoft  
Copyright (C) 2010 Mark Cave-Ayland <mark.cave-ayland@siriusit.co.uk>  
Copyright (c) 1999, 2001, Frank Warmerdam  
Copyright (C) 2016-2017 Bjørn Harrtell <bjorn@wololo.org>  
Copyright (C) 2017 Danny Goette <danny.goette@fem.tu-ilmenau.de>  
Copyright 2009-2011 Paul Ramsey <pramsey@cleverelephant.ca>  
^copyright^  
Copyright 2012 (C) Paul Ramsey <pramsey@cleverelephant.ca>  
Copyright (C) 2006 Refrations Research Inc.  
Copyright 2009 Paul Ramsey <pramsey@opengeo.org>  
Copyright 2001-2009 Refrations Research Inc.  
Copyright (C) 2010 Olivier Courtin <olivier.courtin@oslandia.com>  
By Nathan Wagner, copyright disclaimed,  
this entire file is in the public domain  
Copyright 2009-2011 Olivier Courtin <olivier.courtin@oslandia.com>  
Copyright (C) 2001-2005 Refrations Research Inc.  
Copyright 2001-2011 Refrations Research Inc.  
Copyright 2009-2014 Sandro Santilli <strk@kbt.io>  
Copyright (C) 2008 Paul Ramsey <pramsey@cleverelephant.ca>  
Copyright (C) 2007 Refrations Research Inc.  
Copyright (C) 2010 Sandro Santilli <strk@kbt.io>  
Copyright 2012 J Smith <dark.panda@gmail.com>  
Copyright 2009 - 2010 Oslandia  
Copyright 2009 Oslandia  
Copyright 2001-2005 Refrations Research Inc.  
Copyright 2016 Paul Ramsey <pramsey@cleverelephant.ca>  
Copyright 2016 Daniel Baston <dbaston@gmail.com>  
Copyright (C) 2011 OpenGeo.org  
Copyright (c) 2003-2017, Troy D. Hanson <http://troydhanson.github.com/uthash/>  
Copyright (C) 2011 Regents of the University of California  
Copyright (C) 2011-2013 Regents of the University of California

Copyright (C) 2010-2011 Jorge Arevalo <jorge.arevalo@deimos-space.com>  
Copyright (C) 2010-2011 David Zwarg <dzwarg@azavea.com>  
Copyright (C) 2009-2011 Pierre Racine <pierre.racine@sbf.ulaval.ca>  
Copyright (C) 2009-2011 Mateusz Loskot <mateusz@loskot.net>  
Copyright (C) 2008-2009 Sandro Santilli <strk@kbt.io>  
Copyright (C) 2013 Nathaniel Hunter Clay <clay.nathaniel@gmail.com>  
Copyright (C) 2013 Nathaniel Hunter Clay <clay.nathaniel@gmail.com>  
Copyright (C) 2013 Bborie Park <dustymugs@gmail.com>  
Copyright (C) 2013 Nathaniel Hunter Clay <clay.nathaniel@gmail.com>  
(C) 2009 Mateusz Loskot <mateusz@loskot.net>  
Copyright (C) 2009 Mateusz Loskot <mateusz@loskot.net>  
Copyright (C) 2009-2010 Mateusz Loskot <mateusz@loskot.net>  
Copyright (C) 2009-2010 Jorge Arevalo <jorge.arevalo@deimos-space.com>  
Copyright (C) 2012 Regents of the University of California  
Copyright (C) 2013 Regents of the University of California  
Copyright (C) 2012-2013 Regents of the University of California  
Copyright (C) 2009 Sandro Santilli <strk@kbt.io>  
"

License: The GPL v2 License.

GNU GENERAL PUBLIC LICENSE

Version 2, June 1991

Copyright (C) 1989, 1991 Free Software Foundation, Inc.

51 Franklin St, Fifth Floor, Boston, MA 02110-1301

Everyone is permitted to copy and distribute verbatim copies of this license document, but changing it is not allowed.

Preamble

The licenses for most software are designed to take away your freedom to share and change it. By contrast, the GNU General Public License is intended to guarantee your freedom to share and change free software--to make sure the software is free for all its users. This General Public License applies to most of the Free Software Foundation's software and to any other program whose authors commit to using it. (Some other Free Software Foundation software is covered by the GNU Library General Public License instead.) You can apply it to your programs, too.

When we speak of free software, we are referring to freedom, not price. Our General Public Licenses are designed to make sure that you have the freedom to distribute copies of free software (and charge for this service if you wish), that you receive source code or can get it if you want it, that you can change the software or use pieces of it in new free programs; and that you know you can do these things.

To protect your rights, we need to make restrictions that forbid anyone to deny you these rights or to ask you to surrender the rights. These restrictions translate to certain responsibilities for you if you distribute copies of the software, or if you modify it.

For example, if you distribute copies of such a program, whether gratis or for a fee, you must give the recipients all the rights that you have. You must make sure that they, too, receive or can get the source code. And you must show them these terms so they know their rights.

We protect your rights with two steps: (1) copyright the software, and (2) offer you this license which gives you legal permission to copy, distribute and/or modify the software.

Also, for each author's protection and ours, we want to make certain that everyone understands that there is no warranty for this free software. If the software is modified by someone else and passed on, we want its recipients to know that what they have is not the original, so that any problems introduced by others will not reflect on the original authors' reputations.

Finally, any free program is threatened constantly by software patents. We wish to avoid the danger that redistributors of a free program will individually obtain patent licenses, in effect making the program proprietary. To prevent this, we have made it clear that any patent must be licensed for everyone's free use or not licensed at all.

The precise terms and conditions for copying, distribution and modification follow.?

## GNU GENERAL PUBLIC LICENSE

### TERMS AND CONDITIONS FOR COPYING, DISTRIBUTION AND MODIFICATION

0. This License applies to any program or other work which contains a notice placed by the copyright holder saying it may be distributed under the terms of this General Public License. The "Program", below, refers to any such program or work, and a "work based on the Program" means either the Program or any derivative work under copyright law: that is to say, a work containing the Program or a portion of it, either verbatim or with modifications and/or translated into another language. (Hereinafter, translation is included without limitation in the term "modification".) Each licensee is addressed as "you".

Activities other than copying, distribution and modification are not covered by this License; they are outside its scope. The act of running the Program is not restricted, and the output from the Program is covered only if its contents



constitute a work based on the Program (independent of having been made by running the Program). Whether that is true depends on what the Program does.

1. You may copy and distribute verbatim copies of the Program's source code as you receive it, in any medium, provided that you conspicuously and appropriately publish on each copy an appropriate copyright notice and disclaimer of warranty; keep intact all the notices that refer to this License and to the absence of any warranty; and give any other recipients of the Program a copy of this License along with the Program.

You may charge a fee for the physical act of transferring a copy, and you may at your option offer warranty protection in exchange for a fee.

2. You may modify your copy or copies of the Program or any portion of it, thus forming a work based on the Program, and copy and distribute such modifications or work under the terms of Section 1 above, provided that you also meet all of these conditions:

a) You must cause the modified files to carry prominent notices stating that you changed the files and the date of any change.

b) You must cause any work that you distribute or publish, that in whole or in part contains or is derived from the Program or any part thereof, to be licensed as a whole at no charge to all third parties under the terms of this License.

c) If the modified program normally reads commands interactively when run, you must cause it, when started running for such interactive use in the most ordinary way, to print or display an announcement including an appropriate copyright notice and a notice that there is no warranty (or else, saying that you provide a warranty) and that users may redistribute the program under these conditions, and telling the user how to view a copy of this License. (Exception: if the Program itself is interactive but does not normally print such an announcement, your work based on the Program is not required to print an announcement.)

These requirements apply to the modified work as a whole. If identifiable sections of that work are not derived from the Program, and can be reasonably considered independent and separate works in themselves, then this License, and its terms, do not apply to those sections when you distribute them as separate works. But when you distribute the same sections as part of a whole which is a work based on the Program, the distribution of the whole must be on the terms of this License, whose permissions for other licensees extend to the entire whole, and thus to each and every part regardless of who wrote it.

Thus, it is not the intent of this section to claim rights or contest your rights to work written entirely by you; rather, the intent is to exercise the right to control the distribution of derivative or collective works based on the Program.

In addition, mere aggregation of another work not based on the Program with the Program (or with a work based on the Program) on a volume of a storage or distribution medium does not bring the other work under the scope of this License.

3. You may copy and distribute the Program (or a work based on it, under Section 2) in object code or executable form under the terms of Sections 1 and 2 above provided that you also do one of the following:

a) Accompany it with the complete corresponding machine-readable source code, which must be distributed under the terms of Sections 1 and 2 above on a medium customarily used for software interchange; or,

b) Accompany it with a written offer, valid for at least three years, to give any third party, for a charge no more than your cost of physically performing source distribution, a complete machine-readable copy of the corresponding source code, to be distributed under the terms of Sections 1 and 2 above on a medium customarily used for software interchange; or,

c) Accompany it with the information you received as to the offer to distribute corresponding source code. (This alternative is allowed only for noncommercial distribution and only if you received the program in object code or executable form with such an offer, in accord with Subsection b above.)

The source code for a work means the preferred form of the work for making modifications to it. For an executable work, complete source code means all the source code for all modules it contains, plus any associated interface definition files, plus the scripts used to control compilation and installation of the executable. However, as a special exception, the source code distributed need not include anything that is normally distributed (in either source or binary form) with the major components (compiler, kernel, and so on) of the operating system on which the executable runs, unless that component itself accompanies the executable.

If distribution of executable or object code is made by offering access to copy from a designated place, then offering equivalent access to copy the source code from the same place counts as distribution of the source code, even though third parties are not compelled to copy the source along with the object code.

4. You may not copy, modify, sublicense, or distribute the Program except as expressly provided under this License. Any attempt otherwise to copy, modify, sublicense or distribute the Program is void, and will automatically terminate your rights under this License. However, parties who have received copies, or rights, from you under this License will not have their licenses terminated so long as such parties remain in full compliance.

5. You are not required to accept this License, since you have not signed it. However, nothing else grants you permission to modify or distribute the Program or its derivative works. These actions are prohibited by law if you do not accept this License. Therefore, by modifying or distributing the Program (or any work based on the Program), you indicate your acceptance of this License to do so, and all its terms and conditions for copying, distributing or modifying the Program or works based on it.

6. Each time you redistribute the Program (or any work based on the Program), the recipient automatically receives a license from the original licensor to copy,

distribute or modify the Program subject to these terms and conditions. You may not impose any further restrictions on the recipients' exercise of the rights granted herein. You are not responsible for enforcing compliance by third parties to this License.

7. If, as a consequence of a court judgment or allegation of patent infringement or for any other reason (not limited to patent issues), conditions are imposed on you (whether by court order, agreement or otherwise) that contradict the conditions of this License, they do not excuse you from the conditions of this License. If you cannot distribute so as to satisfy simultaneously your obligations under this License and any other pertinent obligations, then as a consequence you may not distribute the Program at all. For example, if a patent license would not permit royalty-free redistribution of the Program by all those who receive copies directly or indirectly through you, then the only way you could satisfy both it and this License would be to refrain entirely from distribution of the Program.

If any portion of this section is held invalid or unenforceable under any particular circumstance, the balance of the section is intended to apply and the section as a whole is intended to apply in other circumstances.

It is not the purpose of this section to induce you to infringe any patents or other property right claims or to contest validity of any such claims; this section has the sole purpose of protecting the integrity of the free software distribution system, which is implemented by public license practices. Many people have made generous contributions to the wide range of software distributed through that system in reliance on consistent application of that system; it is up to the author/donor to decide if he or she is willing to distribute software through any other system and a licensee cannot impose that choice.

This section is intended to make thoroughly clear what is believed to be a consequence of the rest of this License.

8. If the distribution and/or use of the Program is restricted in certain countries either by patents or by copyrighted interfaces, the original copyright holder who places the Program under this License may add an explicit geographical distribution limitation excluding those countries, so that distribution is permitted only in or among countries not thus excluded. In such case, this License incorporates the limitation as if written in the body of this License.

9. The Free Software Foundation may publish revised and/or new versions of the General Public License from time to time. Such new versions will be similar in spirit to the present version, but may differ in detail to address new problems or concerns.

Each version is given a distinguishing version number. If the Program specifies a version number of this License which applies to it and "any later version", you have the option of following the terms and conditions either of that version or of any later version published by the Free Software Foundation. If the Program does not specify a version number of this License, you may choose any version ever published by the Free Software Foundation.

10. If you wish to incorporate parts of the Program into other free programs whose distribution conditions are different, write to the author to ask for permission. For software which is copyrighted by the Free Software Foundation, write to the Free Software Foundation; we sometimes make exceptions for this. Our decision will be guided by the two goals of preserving the free status of all derivatives of our free software and of promoting the sharing and reuse of software generally.

#### NO WARRANTY

11. BECAUSE THE PROGRAM IS LICENSED FREE OF CHARGE, THERE IS NO WARRANTY FOR THE PROGRAM, TO THE EXTENT PERMITTED BY APPLICABLE LAW. EXCEPT WHEN OTHERWISE STATED IN WRITING THE COPYRIGHT HOLDERS AND/OR OTHER PARTIES PROVIDE THE PROGRAM "AS IS" WITHOUT WARRANTY OF ANY KIND, EITHER EXPRESSED OR IMPLIED, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE. THE ENTIRE RISK AS TO THE QUALITY AND PERFORMANCE OF THE PROGRAM IS WITH YOU. SHOULD THE PROGRAM PROVE DEFECTIVE, YOU ASSUME THE COST OF ALL NECESSARY SERVICING, REPAIR OR CORRECTION.

12. IN NO EVENT UNLESS REQUIRED BY APPLICABLE LAW OR AGREED TO IN WRITING WILL ANY COPYRIGHT HOLDER, OR ANY OTHER PARTY WHO MAY MODIFY AND/OR REDISTRIBUTE THE PROGRAM AS PERMITTED ABOVE, BE LIABLE TO YOU FOR DAMAGES, INCLUDING ANY GENERAL, SPECIAL, INCIDENTAL OR CONSEQUENTIAL DAMAGES ARISING OUT OF THE USE OR INABILITY TO USE THE PROGRAM (INCLUDING BUT NOT LIMITED TO LOSS OF DATA OR DATA BEING RENDERED INACCURATE OR LOSSES SUSTAINED BY YOU OR THIRD PARTIES OR A FAILURE OF THE PROGRAM TO OPERATE WITH ANY OTHER PROGRAMS), EVEN IF SUCH HOLDER OR OTHER PARTY HAS BEEN ADVISED OF THE POSSIBILITY OF SUCH DAMAGES.

#### END OF TERMS AND CONDITIONS

##### How to Apply These Terms to Your New Programs

If you develop a new program, and you want it to be of the greatest possible use to the public, the best way to achieve this is to make it free software which everyone can redistribute and change under these terms.

To do so, attach the following notices to the program. It is safest to attach them to the start of each source file to most effectively convey the exclusion of warranty; and each file should have at least the "copyright" line and a pointer to where the full notice is found.

<one line to give the program's name and a brief idea of what it does.>

Copyright (C) <year> <name of author>

This program is free software; you can redistribute it and/or modify it under the terms of the GNU General Public License as published by the Free Software Foundation; either version 2 of the License, or (at your option) any later version.

This program is distributed in the hope that it will be useful, but WITHOUT ANY WARRANTY; without even the implied warranty of MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE. See the GNU General Public License for more details.

You should have received a copy of the GNU General Public License along with this program; if not, write to the Free Software Foundation, Inc., 51 Franklin St, Fifth Floor, Boston, MA 02110-1301

Also add information on how to contact you by electronic and paper mail.

If the program is interactive, make it output a short notice like this when it starts in an interactive mode:

```
Gnomovision version 69, Copyright (C) year name of author
```

```
Gnomovision comes with ABSOLUTELY NO WARRANTY; for details type `show w'.
```

```
This is free software, and you are welcome to redistribute it under certain conditions; type `show c' for details.
```

The hypothetical commands `show w' and `show c' should show the appropriate parts of the General Public License. Of course, the commands you use may be called something other than `show w' and `show c'; they could even be mouse-clicks or menu items--whatever suits your program.

You should also get your employer (if you work as a programmer) or your school, if any, to sign a "copyright disclaimer" for the program, if necessary. Here is a sample; alter the names:

```
Yoyodyne, Inc., hereby disclaims all copyright interest in the program  
`Gnomovision' (which makes passes at compilers) written by James Hacker.
```

```
<signature of Ty Coon>, 1 April 1989 Ty Coon, President of Vice
```

This General Public License does not permit incorporating your program into proprietary programs. If your program is a subroutine library, you may consider it more useful to permit linking proprietary applications with the library. If this is what you want to do, use the GNU Library General Public License instead of this License.

```
Software:Geos
```

```
Copyright notice:
```

```
Copyright (C) 2009 Sandro Santilli <strk@keybit.net>
```

```
Copyright (C) 2006 Refrations Research Inc.
```

```
Copyright (C) 2013 Sandro Santilli <strk@keybit.net>
```

```
Copyright (C) 2011 Sandro Santilli <strk@keybit.net>
```

```
Copyright (C) 2009 Sandro Santilli <strk@keybit.net>
```

Copyright (C) 2011 Sandro Santilli <strk@keybit.net>  
Copyright (C) 2005-2011 Refrations Research Inc.  
Copyright (C) 2009 Ragi Y. Burhum <ragi@burhum.com>  
Copyright (C) 2010 Sandro Santilli <strk@keybit.net>  
Copyright (C) 2009 2011 Sandro Santilli <strk@keybit.net>  
Copyright (C) 2005 2006 Refrations Research Inc.  
Copyright (C) 2011 Sandro Santilli <strk@keybit.net>  
Copyright (C) 2006-2011 Refrations Research Inc.  
Copyright (C) 2011 Sandro Santilli <strk@keybit.net>  
Copyright (C) 2009-2011 Sandro Santilli <strk@keybit.net>  
Copyright (C) 2016 Daniel Baston  
Copyright (C) 2008 Sean Gillies  
Copyright (C) 2009 Sandro Santilli <strk@keybit.net>  
Copyright (C) 2006 Refrations Research Inc.  
Copyright (C) 2012 Sandro Santilli <strk@keybit.net>  
Copyright (C) 2009 Sandro Santilli <strk@keybit.net>  
Copyright (C) 2008-2010 Safe Software Inc.  
Copyright (C) 2006-2007 Refrations Research Inc.  
Copyright (C) 2005-2007 Refrations Research Inc.  
Copyright (C) 2007 Refrations Research Inc.  
Copyright (C) 2014 Mika Heiskanen <mika.heiskanen@fmi.fi>  
Copyright (C) 2009-2010 Sandro Santilli <strk@keybit.net>  
Copyright (C) 2009 2011 Sandro Santilli <strk@keybit.net>  
Copyright (C) 2010 Sandro Santilli <strk@keybit.net>  
Copyright (C) 2009 Mateusz Loskot  
Copyright (C) 2005-2009 Refrations Research Inc.  
Copyright (C) 2001-2009 Vivid Solutions Inc.  
Copyright (C) 2012 Sandro Santilli <strk@keybit.net>  
Copyright (C) 2006 Wu Yongwei  
Copyright (C) 2012 Excensus LLC.  
Copyright (C) 1996-2015 Free Software Foundation, Inc.  
Copyright (c) 1995 Olivier Devillers <Olivier.Devillers@sophia.inria.fr>  
Copyright (C) 2007-2010 Safe Software Inc.

Copyright (C) 2010 Safe Software Inc.

Copyright (C) 2006 Refrations Research

Copyright 2004 Sean Gillies, sgillies@frii.com

Copyright (C) 2011 Mateusz Loskot <mateusz@loskot.net>

Copyright (C) 2015 Nyal Dawson <nyall dot dawson at gmail dot com>

Original code (2.0 and earlier )copyright (c) 2000-2006 Lee Thomason  
(www.grinninglizard.com)

Original code (2.0 and earlier )copyright (c) 2000-2002 Lee Thomason  
(www.grinninglizard.com)

License: LGPL V2.1

GNU LESSER GENERAL PUBLIC LICENSE

Version 2.1, February 1999

Copyright (C) 1991, 1999 Free Software Foundation, Inc. 51 Franklin Street, Fifth  
Floor, Boston, MA 02110-1301

Everyone is permitted to copy and distribute verbatim copies of this license  
document, but changing it is not allowed.

Copyright (C) 2005-2011 Refrations Research Inc.

Copyright (C) 2009 Ragi Y. Burhum <ragi@burhum.com>

Copyright (C) 2010 Sandro Santilli <strk@keybit.net>

Copyright (C) 2009 2011 Sandro Santilli <strk@keybit.net>

Copyright (C) 2005 2006 Refrations Research Inc.

Copyright (C) 2011 Sandro Santilli <strk@keybit.net>

Copyright (C) 2006-2011 Refrations Research Inc.

Copyright (C) 2011 Sandro Santilli <strk@keybit.net

Copyright (C) 2009-2011 Sandro Santilli <strk@keybit.net>

Copyright (C) 2016 Daniel Baston

Copyright (C) 2008 Sean Gillies

Copyright (C) 2009 Sandro Santilli <strk@keybit.net>

Copyright (C) 2006 Refrations Research Inc.

Copyright (C) 2012 Sandro Santilli <strk@keybit.net>

Copyright (C) 2009 Sandro Santilli <strk@keybit.net>

Copyright (C) 2008-2010 Safe Software Inc.

Copyright (C) 2006-2007 Refrations Research Inc.

Copyright (C) 2005-2007 Refrations Research Inc.  
Copyright (C) 2007 Refrations Research Inc.  
Copyright (C) 2014 Mika Heiskanen <mika.heiskanen@fmi.fi>  
Copyright (C) 2009-2010 Sandro Santilli <strk@keybit.net>  
Copyright (C) 2009 2011 Sandro Santilli <strk@keybit.net>  
Copyright (C) 2010 Sandro Santilli <strk@keybit.net>  
Copyright (C) 2009 Mateusz Loskot  
Copyright (C) 2005-2009 Refrations Research Inc.  
Copyright (C) 2001-2009 Vivid Solutions Inc.  
Copyright (C) 2012 Sandro Santilli <strk@keybit.net>  
Copyright (C) 2006 Wu Yongwei  
Copyright (C) 2012 Excensus LLC.  
Copyright (C) 1996-2015 Free Software Foundation, Inc.  
Copyright (c) 1995 Olivier Devillers <Olivier.Devillers@sophia.inria.fr>  
Copyright (C) 2007-2010 Safe Software Inc.  
Copyright (C) 2010 Safe Software Inc.  
Copyright (C) 2006 Refrations Research  
Copyright 2004 Sean Gillies, sgillies@frii.com  
Copyright (C) 2011 Mateusz Loskot <mateusz@loskot.net>  
Copyright (C) 2015 Nyal Dawson <nyall dot dawson at gmail dot com>  
Original code (2.0 and earlier )copyright (c) 2000-2006 Lee Thomason  
(www.grinninglizard.com)  
Original code (2.0 and earlier )copyright (c) 2000-2002 Lee Thomason  
(www.grinninglizard.com)  
  
License: LGPL V2.1  
  
GNU LESSER GENERAL PUBLIC LICENSE  
Version 2.1, February 1999  
  
Copyright (C) 1991, 1999 Free Software Foundation, Inc. 51 Franklin Street, Fifth  
Floor, Boston, MA 02110-1301  
  
Everyone is permitted to copy and distribute verbatim copies of this license  
document, but changing it is not allowed.  
  
Copyright (C) 2005-2011 Refrations Research Inc.  
Copyright (C) 2009 Ragi Y. Burhum <ragi@burhum.com>



Copyright (C) 2010 Sandro Santilli <strk@keybit.net>  
Copyright (C) 2009 2011 Sandro Santilli <strk@keybit.net>  
Copyright (C) 2005 2006 Refrations Research Inc.  
Copyright (C) 2011 Sandro Santilli <strk@keybit.net>  
Copyright (C) 2006-2011 Refrations Research Inc.  
Copyright (C) 2011 Sandro Santilli <strk@keybit.net>  
Copyright (C) 2009-2011 Sandro Santilli <strk@keybit.net>  
Copyright (C) 2016 Daniel Baston  
Copyright (C) 2008 Sean Gillies  
Copyright (C) 2009 Sandro Santilli <strk@keybit.net>  
Copyright (C) 2006 Refrations Research Inc.  
Copyright (C) 2012 Sandro Santilli <strk@keybit.net>  
Copyright (C) 2009 Sandro Santilli <strk@keybit.net>  
Copyright (C) 2008-2010 Safe Software Inc.  
Copyright (C) 2006-2007 Refrations Research Inc.  
Copyright (C) 2005-2007 Refrations Research Inc.  
Copyright (C) 2007 Refrations Research Inc.  
Copyright (C) 2014 Mika Heiskanen <mika.heiskanen@fmi.fi>  
Copyright (C) 2009-2010 Sandro Santilli <strk@keybit.net>  
Copyright (C) 2009 2011 Sandro Santilli <strk@keybit.net>  
Copyright (C) 2010 Sandro Santilli <strk@keybit.net>  
Copyright (C) 2009 Mateusz Loskot  
Copyright (C) 2005-2009 Refrations Research Inc.  
Copyright (C) 2001-2009 Vivid Solutions Inc.  
Copyright (C) 2012 Sandro Santilli <strk@keybit.net>  
Copyright (C) 2006 Wu Yongwei  
Copyright (C) 2012 Excensus LLC.  
Copyright (C) 1996-2015 Free Software Foundation, Inc.  
Copyright (c) 1995 Olivier Devillers <Olivier.Devillers@sophia.inria.fr>  
Copyright (C) 2007-2010 Safe Software Inc.  
Copyright (C) 2010 Safe Software Inc.  
Copyright (C) 2006 Refrations Research  
Copyright 2004 Sean Gillies, sgillies@frii.com

Copyright (C) 2011 Mateusz Loskot <mateusz@loskot.net>

Copyright (C) 2015 Nyal Dawson <nyall dot dawson at gmail dot com>

Original code (2.0 and earlier )copyright (c) 2000-2006 Lee Thomason  
(www.grinninglizard.com)

Original code (2.0 and earlier )copyright (c) 2000-2002 Lee Thomason  
(www.grinninglizard.com)

License: LGPL V2.1

GNU LESSER GENERAL PUBLIC LICENSE

Version 2.1, February 1999

Copyright (C) 1991, 1999 Free Software Foundation, Inc. 51 Franklin Street, Fifth  
Floor, Boston, MA 02110-1301

Everyone is permitted to copy and distribute verbatim copies of this license  
document, but changing it is not allowed.

Copyright (C) 2005-2011 Refrations Research Inc.

Copyright (C) 2009 Ragi Y. Burhum <ragi@burhum.com>

Copyright (C) 2010 Sandro Santilli <strk@keybit.net>

Copyright (C) 2009 2011 Sandro Santilli <strk@keybit.net>

Copyright (C) 2005 2006 Refrations Research Inc.

Copyright (C) 2011 Sandro Santilli <strk@keybit.net>

Copyright (C) 2006-2011 Refrations Research Inc.

Copyright (C) 2011 Sandro Santilli <strk@keybit.net>

Copyright (C) 2009-2011 Sandro Santilli <strk@keybit.net>

Copyright (C) 2016 Daniel Baston

Copyright (C) 2008 Sean Gillies

Copyright (C) 2009 Sandro Santilli <strk@keybit.net>

Copyright (C) 2006 Refrations Research Inc.

Copyright (C) 2012 Sandro Santilli <strk@keybit.net>

Copyright (C) 2009 Sandro Santilli <strk@keybit.net>

Copyright (C) 2008-2010 Safe Software Inc.

Copyright (C) 2006-2007 Refrations Research Inc.

Copyright (C) 2005-2007 Refrations Research Inc.

Copyright (C) 2007 Refrations Research Inc.

Copyright (C) 2014 Mika Heiskanen <mika.heiskanen@fmi.fi>

Copyright (C) 2009-2010 Sandro Santilli <strk@keybit.net>  
Copyright (C) 2009 2011 Sandro Santilli <strk@keybit.net>  
Copyright (C) 2010 Sandro Santilli <strk@keybit.net>  
Copyright (C) 2009 Mateusz Loskot  
Copyright (C) 2005-2009 Refrations Research Inc.  
Copyright (C) 2001-2009 Vivid Solutions Inc.  
Copyright (C) 2012 Sandro Santilli <strk@keybit.net>  
Copyright (C) 2006 Wu Yongwei  
Copyright (C) 2012 Excensus LLC.  
Copyright (C) 1996-2015 Free Software Foundation, Inc.  
Copyright (c) 1995 Olivier Devillers <Olivier.Devillers@sophia.inria.fr>  
Copyright (C) 2007-2010 Safe Software Inc.  
Copyright (C) 2010 Safe Software Inc.  
Copyright (C) 2006 Refrations Research  
Copyright 2004 Sean Gillies, sgillies@frii.com  
Copyright (C) 2011 Mateusz Loskot <mateusz@loskot.net>  
Copyright (C) 2015 Nyal Dawson <nyall dot dawson at gmail dot com>  
Original code (2.0 and earlier )copyright (c) 2000-2006 Lee Thomason  
(www.grinninglizard.com)  
Original code (2.0 and earlier )copyright (c) 2000-2002 Lee Thomason  
(www.grinninglizard.com)

License: LGPL V2.1

GNU LESSER GENERAL PUBLIC LICENSE  
Version 2.1, February 1999

Copyright (C) 1991, 1999 Free Software Foundation, Inc. 51 Franklin Street, Fifth  
Floor, Boston, MA 02110-1301

Everyone is permitted to copy and distribute verbatim copies of this license  
document, but changing it is not allowed.

Copyright (C) 2005-2011 Refrations Research Inc.  
Copyright (C) 2009 Ragi Y. Burhum <ragi@burhum.com>  
Copyright (C) 2010 Sandro Santilli <strk@keybit.net>  
Copyright (C) 2009 2011 Sandro Santilli <strk@keybit.net>  
Copyright (C) 2005 2006 Refrations Research Inc.

Copyright (C) 2011 Sandro Santilli <strk@keybit.net>  
Copyright (C) 2006-2011 Refrations Research Inc.  
Copyright (C) 2011 Sandro Santilli <strk@keybit.net>  
Copyright (C) 2009-2011 Sandro Santilli <strk@keybit.net>  
Copyright (C) 2016 Daniel Baston  
Copyright (C) 2008 Sean Gillies  
Copyright (C) 2009 Sandro Santilli <strk@keybit.net>  
Copyright (C) 2006 Refrations Research Inc.  
Copyright (C) 2012 Sandro Santilli <strk@keybit.net>  
Copyright (C) 2009 Sandro Santilli <strk@keybit.net>  
Copyright (C) 2008-2010 Safe Software Inc.  
Copyright (C) 2006-2007 Refrations Research Inc.  
Copyright (C) 2005-2007 Refrations Research Inc.  
Copyright (C) 2007 Refrations Research Inc.  
Copyright (C) 2014 Mika Heiskanen <mika.heiskanen@fmi.fi>  
Copyright (C) 2009-2010 Sandro Santilli <strk@keybit.net>  
Copyright (C) 2009 2011 Sandro Santilli <strk@keybit.net>  
Copyright (C) 2010 Sandro Santilli <strk@keybit.net>  
Copyright (C) 2009 Mateusz Loskot  
Copyright (C) 2005-2009 Refrations Research Inc.  
Copyright (C) 2001-2009 Vivid Solutions Inc.  
Copyright (C) 2012 Sandro Santilli <strk@keybit.net>  
Copyright (C) 2006 Wu Yongwei  
Copyright (C) 2012 Excensus LLC.  
Copyright (C) 1996-2015 Free Software Foundation, Inc.  
Copyright (c) 1995 Olivier Devillers <Olivier.Devillers@sophia.inria.fr>  
Copyright (C) 2007-2010 Safe Software Inc.  
Copyright (C) 2010 Safe Software Inc.  
Copyright (C) 2006 Refrations Research  
Copyright 2004 Sean Gillies, sgillies@frie.com  
Copyright (C) 2011 Mateusz Loskot <mateusz@loskot.net>  
Copyright (C) 2015 Nyal Dawson <nyall dot dawson at gmail dot com>  
Original code (2.0 and earlier )copyright (c) 2000-2006 Lee Thomason  
(www.grinninglizard.com)

Original code (2.0 and earlier )copyright (c) 2000-2002 Lee Thomason  
(www.grinninglizard.com)

License: LGPL V2.1

GNU LESSER GENERAL PUBLIC LICENSE

Version 2.1, February 1999

Copyright (C) 1991, 1999 Free Software Foundation, Inc. 51 Franklin Street, Fifth  
Floor, Boston, MA 02110-1301

Everyone is permitted to copy and distribute verbatim copies of this license  
document, but changing it is not allowed.

Copyright (C) 2005-2011 Refrations Research Inc.

Copyright (C) 2009 Ragi Y. Burhum <ragi@burhum.com>

Copyright (C) 2010 Sandro Santilli <strk@keybit.net>

Copyright (C) 2009 2011 Sandro Santilli <strk@keybit.net>

Copyright (C) 2005 2006 Refrations Research Inc.

Copyright (C) 2011 Sandro Santilli <strk@keybit.net>

Copyright (C) 2006-2011 Refrations Research Inc.

Copyright (C) 2011 Sandro Santilli <strk@keybit.net

Copyright (C) 2009-2011 Sandro Santilli <strk@keybit.net>

Copyright (C) 2016 Daniel Baston

Copyright (C) 2008 Sean Gillies

Copyright (C) 2009 Sandro Santilli <strk@keybit.net>

Copyright (C) 2006 Refrations Research Inc.

Copyright (C) 2012 Sandro Santilli <strk@keybit.net>

Copyright (C) 2009 Sandro Santilli <strk@keybit.net>

Copyright (C) 2008-2010 Safe Software Inc.

Copyright (C) 2006-2007 Refrations Research Inc.

Copyright (C) 2005-2007 Refrations Research Inc.

Copyright (C) 2007 Refrations Research Inc.

Copyright (C) 2014 Mika Heiskanen <mika.heiskanen@fmi.fi>

Copyright (C) 2009-2010 Sandro Santilli <strk@keybit.net>

Copyright (C) 2009 2011 Sandro Santilli <strk@keybit.net>

Copyright (C) 2010 Sandro Santilli <strk@keybit.net>

Copyright (C) 2009 Mateusz Loskot

Copyright (C) 2005-2009 Refrations Research Inc.  
Copyright (C) 2001-2009 Vivid Solutions Inc.  
Copyright (C) 2012 Sandro Santilli <strk@keybit.net>  
Copyright (C) 2006 Wu Yongwei  
Copyright (C) 2012 Excensus LLC.  
Copyright (C) 1996-2015 Free Software Foundation, Inc.  
Copyright (c) 1995 Olivier Devillers <Olivier.Devillers@sophia.inria.fr>  
Copyright (C) 2007-2010 Safe Software Inc.  
Copyright (C) 2010 Safe Software Inc.  
Copyright (C) 2006 Refrations Research  
Copyright 2004 Sean Gillies, sgillies@fii.com  
Copyright (C) 2011 Mateusz Loskot <mateusz@loskot.net>  
Copyright (C) 2015 Nyal Dawson <nyall dot dawson at gmail dot com>  
Original code (2.0 and earlier )copyright (c) 2000-2006 Lee Thomason  
(www.grinninglizard.com)  
Original code (2.0 and earlier )copyright (c) 2000-2002 Lee Thomason  
(www.grinninglizard.com)

License: LGPL V2.1

#### GNU LESSER GENERAL PUBLIC LICENSE

Version 2.1, February 1999

Copyright (C) 1991, 1999 Free Software Foundation, Inc. 51 Franklin Street, Fifth Floor, Boston, MA 02110-1301

Everyone is permitted to copy and distribute verbatim copies of this license document, but changing it is not allowed.

Copyright (C) 2005-2011 Refrations Research Inc.  
Copyright (C) 2009 Ragi Y. Burhum <ragi@burhum.com>  
Copyright (C) 2010 Sandro Santilli <strk@keybit.net>  
Copyright (C) 2009 2011 Sandro Santilli <strk@keybit.net>  
Copyright (C) 2005 2006 Refrations Research Inc.  
Copyright (C) 2011 Sandro Santilli <strk@keybit.net>  
Copyright (C) 2006-2011 Refrations Research Inc.  
Copyright (C) 2011 Sandro Santilli <strk@keybit.net>  
Copyright (C) 2009-2011 Sandro Santilli <strk@keybit.net>

Copyright (C) 2016 Daniel Baston  
Copyright (C) 2008 Sean Gillies  
Copyright (C) 2009 Sandro Santilli <strk@keybit.net>  
Copyright (C) 2006 Refrations Research Inc.  
Copyright (C) 2012 Sandro Santilli <strk@keybit.net>  
Copyright (C) 2009 Sandro Santilli <strk@keybit.net>  
Copyright (C) 2008-2010 Safe Software Inc.  
Copyright (C) 2006-2007 Refrations Research Inc.  
Copyright (C) 2005-2007 Refrations Research Inc.  
Copyright (C) 2007 Refrations Research Inc.  
Copyright (C) 2014 Mika Heiskanen <mika.heiskanen@fmi.fi>  
Copyright (C) 2009-2010 Sandro Santilli <strk@keybit.net>  
Copyright (C) 2009 2011 Sandro Santilli <strk@keybit.net>  
Copyright (C) 2010 Sandro Santilli <strk@keybit.net>  
Copyright (C) 2009 Mateusz Loskot  
Copyright (C) 2005-2009 Refrations Research Inc.  
Copyright (C) 2001-2009 Vivid Solutions Inc.  
Copyright (C) 2012 Sandro Santilli <strk@keybit.net>  
Copyright (C) 2006 Wu Yongwei  
Copyright (C) 2012 Excensus LLC.  
Copyright (C) 1996-2015 Free Software Foundation, Inc.  
Copyright (c) 1995 Olivier Devillers <Olivier.Devillers@sophia.inria.fr>  
Copyright (C) 2007-2010 Safe Software Inc.  
Copyright (C) 2010 Safe Software Inc.  
Copyright (C) 2006 Refrations Research  
Copyright 2004 Sean Gillies, sgillies@frie.com  
Copyright (C) 2011 Mateusz Loskot <mateusz@loskot.net>  
Copyright (C) 2015 Nyal Dawson <nyall dot dawson at gmail dot com>  
Original code (2.0 and earlier )copyright (c) 2000-2006 Lee Thomason  
(www.grinninglizard.com)  
Original code (2.0 and earlier )copyright (c) 2000-2002 Lee Thomason  
(www.grinninglizard.com)  
  
License: LGPL V2.1

GNU LESSER GENERAL PUBLIC LICENSE

Version 2.1, February 1999

Copyright (C) 1991, 1999 Free Software Foundation, Inc. 51 Franklin Street, Fifth Floor, Boston, MA 02110-1301

Everyone is permitted to copy and distribute verbatim copies of this license document, but changing it is not allowed.

Copyright (C) 2005-2011 Refrations Research Inc.

Copyright (C) 2009 Ragi Y. Burhum <ragi@burhum.com>

Copyright (C) 2010 Sandro Santilli <strk@keybit.net>

Copyright (C) 2009 2011 Sandro Santilli <strk@keybit.net>

Copyright (C) 2005 2006 Refrations Research Inc.

Copyright (C) 2011 Sandro Santilli <strk@keybit.net>

Copyright (C) 2006-2011 Refrations Research Inc.

Copyright (C) 2011 Sandro Santilli <strk@keybit.net>

Copyright (C) 2009-2011 Sandro Santilli <strk@keybit.net>

Copyright (C) 2016 Daniel Baston

Copyright (C) 2008 Sean Gillies

Copyright (C) 2009 Sandro Santilli <strk@keybit.net>

Copyright (C) 2006 Refrations Research Inc.

Copyright (C) 2012 Sandro Santilli <strk@keybit.net>

Copyright (C) 2009 Sandro Santilli <strk@keybit.net>

Copyright (C) 2008-2010 Safe Software Inc.

Copyright (C) 2006-2007 Refrations Research Inc.

Copyright (C) 2005-2007 Refrations Research Inc.

Copyright (C) 2007 Refrations Research Inc.

Copyright (C) 2014 Mika Heiskanen <mika.heiskanen@fmi.fi>

Copyright (C) 2009-2010 Sandro Santilli <strk@keybit.net>

Copyright (C) 2009 2011 Sandro Santilli <strk@keybit.net>

Copyright (C) 2010 Sandro Santilli <strk@keybit.net>

Copyright (C) 2009 Mateusz Loskot

Copyright (C) 2005-2009 Refrations Research Inc.

Copyright (C) 2001-2009 Vivid Solutions Inc.

Copyright (C) 2012 Sandro Santilli <strk@keybit.net>



Copyright (C) 2006 Wu Yongwei  
Copyright (C) 2012 Excensus LLC.  
Copyright (C) 1996-2015 Free Software Foundation, Inc.  
Copyright (c) 1995 Olivier Devillers <Olivier.Devillers@sophia.inria.fr>  
Copyright (C) 2007-2010 Safe Software Inc.  
Copyright (C) 2010 Safe Software Inc.  
Copyright (C) 2006 Refrations Research  
Copyright 2004 Sean Gillies, sgillies@frii.com  
Copyright (C) 2011 Mateusz Loskot <mateusz@loskot.net>  
Copyright (C) 2015 Nyal Dawson <nyall dot dawson at gmail dot com>  
Original code (2.0 and earlier )copyright (c) 2000-2006 Lee Thomason  
(www.grinninglizard.com)  
Original code (2.0 and earlier )copyright (c) 2000-2002 Lee Thomason  
(www.grinninglizard.com)

License: LGPL V2.1

GNU LESSER GENERAL PUBLIC LICENSE  
Version 2.1, February 1999

Copyright (C) 1991, 1999 Free Software Foundation, Inc. 51 Franklin Street, Fifth  
Floor, Boston, MA 02110-1301

Everyone is permitted to copy and distribute verbatim copies of this license  
document, but changing it is not allowed.

Copyright (C) 2005-2011 Refrations Research Inc.  
Copyright (C) 2009 Ragi Y. Burhum <ragi@burhum.com>  
Copyright (C) 2010 Sandro Santilli <strk@keybit.net>  
Copyright (C) 2009 2011 Sandro Santilli <strk@keybit.net>  
Copyright (C) 2005 2006 Refrations Research Inc.  
Copyright (C) 2011 Sandro Santilli <strk@keybit.net>  
Copyright (C) 2006-2011 Refrations Research Inc.  
Copyright (C) 2011 Sandro Santilli <strk@keybit.net>  
Copyright (C) 2009-2011 Sandro Santilli <strk@keybit.net>  
Copyright (C) 2016 Daniel Baston  
Copyright (C) 2008 Sean Gillies  
Copyright (C) 2009 Sandro Santilli <strk@keybit.net>

Copyright (C) 2006 Refrations Research Inc.  
Copyright (C) 2012 Sandro Santilli <strk@keybit.net>  
Copyright (C) 2009 Sandro Santilli <strk@keybit.net>  
Copyright (C) 2008-2010 Safe Software Inc.  
Copyright (C) 2006-2007 Refrations Research Inc.  
Copyright (C) 2005-2007 Refrations Research Inc.  
Copyright (C) 2007 Refrations Research Inc.  
Copyright (C) 2014 Mika Heiskanen <mika.heiskanen@fmi.fi>  
Copyright (C) 2009-2010 Sandro Santilli <strk@keybit.net>  
Copyright (C) 2009 2011 Sandro Santilli <strk@keybit.net>  
Copyright (C) 2010 Sandro Santilli <strk@keybit.net>  
Copyright (C) 2009 Mateusz Loskot  
Copyright (C) 2005-2009 Refrations Research Inc.  
Copyright (C) 2001-2009 Vivid Solutions Inc.  
Copyright (C) 2012 Sandro Santilli <strk@keybit.net>  
Copyright (C) 2006 Wu Yongwei  
Copyright (C) 2012 Excensus LLC.  
Copyright (C) 1996-2015 Free Software Foundation, Inc.  
Copyright (c) 1995 Olivier Devillers <Olivier.Devillers@sophia.inria.fr>  
Copyright (C) 2007-2010 Safe Software Inc.  
Copyright (C) 2010 Safe Software Inc.  
Copyright (C) 2006 Refrations Research  
Copyright 2004 Sean Gillies, sgillies@fii.com  
Copyright (C) 2011 Mateusz Loskot <mateusz@loskot.net>  
Copyright (C) 2015 Nyal Dawson <nyall dot dawson at gmail dot com>  
Original code (2.0 and earlier )copyright (c) 2000-2006 Lee Thomason  
(www.grinninglizard.com)  
Original code (2.0 and earlier )copyright (c) 2000-2002 Lee Thomason  
(www.grinninglizard.com)  
  
License: LGPL V2.1  
  
GNU LESSER GENERAL PUBLIC LICENSE  
Version 2.1, February 1999

Copyright (C) 1991, 1999 Free Software Foundation, Inc. 51 Franklin Street, Fifth Floor, Boston, MA 02110-1301

Everyone is permitted to copy and distribute verbatim copies of this license document, but changing it is not allowed.

Copyright (C) 2005-2011 Refrations Research Inc.

Copyright (C) 2009 Ragi Y. Burhum <ragi@burhum.com>

Copyright (C) 2010 Sandro Santilli <strk@keybit.net>

Copyright (C) 2009 2011 Sandro Santilli <strk@keybit.net>

Copyright (C) 2005 2006 Refrations Research Inc.

Copyright (C) 2011 Sandro Santilli <strk@keybit.net>

Copyright (C) 2006-2011 Refrations Research Inc.

Copyright (C) 2011 Sandro Santilli <strk@keybit.net>

Copyright (C) 2009-2011 Sandro Santilli <strk@keybit.net>

Copyright (C) 2016 Daniel Baston

Copyright (C) 2008 Sean Gillies

Copyright (C) 2009 Sandro Santilli <strk@keybit.net>

Copyright (C) 2006 Refrations Research Inc.

Copyright (C) 2012 Sandro Santilli <strk@keybit.net>

Copyright (C) 2009 Sandro Santilli <strk@keybit.net>

Copyright (C) 2008-2010 Safe Software Inc.

Copyright (C) 2006-2007 Refrations Research Inc.

Copyright (C) 2005-2007 Refrations Research Inc.

Copyright (C) 2007 Refrations Research Inc.

Copyright (C) 2014 Mika Heiskanen <mika.heiskanen@fmi.fi>

Copyright (C) 2009-2010 Sandro Santilli <strk@keybit.net>

Copyright (C) 2009 2011 Sandro Santilli <strk@keybit.net>

Copyright (C) 2010 Sandro Santilli <strk@keybit.net>

Copyright (C) 2009 Mateusz Loskot

Copyright (C) 2005-2009 Refrations Research Inc.

Copyright (C) 2001-2009 Vivid Solutions Inc.

Copyright (C) 2012 Sandro Santilli <strk@keybit.net>

Copyright (C) 2006 Wu Yongwei

Copyright (C) 2012 Excensus LLC.

Copyright (C) 1996-2015 Free Software Foundation, Inc.

Copyright (c) 1995 Olivier Devillers <Olivier.Devillers@sophia.inria.fr>  
Copyright (C) 2007-2010 Safe Software Inc.  
Copyright (C) 2010 Safe Software Inc.  
Copyright (C) 2006 Refrations Research  
Copyright 2004 Sean Gillies, sgillies@frii.com  
Copyright (C) 2011 Mateusz Loskot <mateusz@loskot.net>  
Copyright (C) 2015 Nyal Dawson <nyall dot dawson at gmail dot com>  
Original code (2.0 and earlier )copyright (c) 2000-2006 Lee Thomason  
(www.grinninglizard.com)  
Original code (2.0 and earlier )copyright (c) 2000-2002 Lee Thomason  
(www.grinninglizard.com)

License: LGPL V2.1

GNU LESSER GENERAL PUBLIC LICENSE

Version 2.1, February 1999

Copyright (C) 1991, 1999 Free Software Foundation, Inc. 51 Franklin Street, Fifth  
Floor, Boston, MA 02110-1301

Everyone is permitted to copy and distribute verbatim copies of this license  
document, but changing it is not allowed.

Copyright (C) 2005-2011 Refrations Research Inc.  
Copyright (C) 2009 Ragi Y. Burhum <ragi@burhum.com>  
Copyright (C) 2010 Sandro Santilli <strk@keybit.net>  
Copyright (C) 2009 2011 Sandro Santilli <strk@keybit.net>  
Copyright (C) 2005 2006 Refrations Research Inc.  
Copyright (C) 2011 Sandro Santilli <strk@keybit.net>  
Copyright (C) 2006-2011 Refrations Research Inc.  
Copyright (C) 2011 Sandro Santilli <strk@keybit.net>  
Copyright (C) 2009-2011 Sandro Santilli <strk@keybit.net>  
Copyright (C) 2016 Daniel Baston  
Copyright (C) 2008 Sean Gillies  
Copyright (C) 2009 Sandro Santilli <strk@keybit.net>  
Copyright (C) 2006 Refrations Research Inc.  
Copyright (C) 2012 Sandro Santilli <strk@keybit.net>  
Copyright (C) 2009 Sandro Santilli <strk@keybit.net>

Copyright (C) 2008-2010 Safe Software Inc.  
Copyright (C) 2006-2007 Refrations Research Inc.  
Copyright (C) 2005-2007 Refrations Research Inc.  
Copyright (C) 2007 Refrations Research Inc.  
Copyright (C) 2014 Mika Heiskanen <mika.heiskanen@fmi.fi>  
Copyright (C) 2009-2010 Sandro Santilli <strk@keybit.net>  
Copyright (C) 2009 2011 Sandro Santilli <strk@keybit.net>  
Copyright (C) 2010 Sandro Santilli <strk@keybit.net>  
Copyright (C) 2009 Mateusz Loskot  
Copyright (C) 2005-2009 Refrations Research Inc.  
Copyright (C) 2001-2009 Vivid Solutions Inc.  
Copyright (C) 2012 Sandro Santilli <strk@keybit.net>  
Copyright (C) 2006 Wu Yongwei  
Copyright (C) 2012 Excensus LLC.  
Copyright (C) 1996-2015 Free Software Foundation, Inc.  
Copyright (c) 1995 Olivier Devillers <Olivier.Devillers@sophia.inria.fr>  
Copyright (C) 2007-2010 Safe Software Inc.  
Copyright (C) 2010 Safe Software Inc.  
Copyright (C) 2006 Refrations Research  
Copyright 2004 Sean Gillies, sgillies@frie.com  
Copyright (C) 2011 Mateusz Loskot <mateusz@loskot.net>  
Copyright (C) 2015 Nyal Dawson <nyall dot dawson at gmail dot com>  
Original code (2.0 and earlier )copyright (c) 2000-2006 Lee Thomason  
(www.grinninglizard.com)  
Original code (2.0 and earlier )copyright (c) 2000-2002 Lee Thomason  
(www.grinninglizard.com)

License: LGPL V2.1

GNU LESSER GENERAL PUBLIC LICENSE

Version 2.1, February 1999

Copyright (C) 1991, 1999 Free Software Foundation, Inc. 51 Franklin Street, Fifth  
Floor, Boston, MA 02110-1301

Everyone is permitted to copy and distribute verbatim copies of this license  
document, but changing it is not allowed.

Copyright (C) 2005-2011 Refrations Research Inc.  
Copyright (C) 2009 Ragi Y. Burhum <ragi@burhum.com>  
Copyright (C) 2010 Sandro Santilli <strk@keybit.net>  
Copyright (C) 2009 2011 Sandro Santilli <strk@keybit.net>  
Copyright (C) 2005 2006 Refrations Research Inc.  
Copyright (C) 2011 Sandro Santilli <strk@keybit.net>  
Copyright (C) 2006-2011 Refrations Research Inc.  
Copyright (C) 2011 Sandro Santilli <strk@keybit.net>  
Copyright (C) 2009-2011 Sandro Santilli <strk@keybit.net>  
Copyright (C) 2016 Daniel Baston  
Copyright (C) 2008 Sean Gillies  
Copyright (C) 2009 Sandro Santilli <strk@keybit.net>  
Copyright (C) 2006 Refrations Research Inc.  
Copyright (C) 2012 Sandro Santilli <strk@keybit.net>  
Copyright (C) 2009 Sandro Santilli <strk@keybit.net>  
Copyright (C) 2008-2010 Safe Software Inc.  
Copyright (C) 2006-2007 Refrations Research Inc.  
Copyright (C) 2005-2007 Refrations Research Inc.  
Copyright (C) 2007 Refrations Research Inc.  
Copyright (C) 2014 Mika Heiskanen <mika.heiskanen@fmi.fi>  
Copyright (C) 2009-2010 Sandro Santilli <strk@keybit.net>  
Copyright (C) 2009 2011 Sandro Santilli <strk@keybit.net>  
Copyright (C) 2010 Sandro Santilli <strk@keybit.net>  
Copyright (C) 2009 Mateusz Loskot  
Copyright (C) 2005-2009 Refrations Research Inc.  
Copyright (C) 2001-2009 Vivid Solutions Inc.  
Copyright (C) 2012 Sandro Santilli <strk@keybit.net>  
Copyright (C) 2006 Wu Yongwei  
Copyright (C) 2012 Excensus LLC.  
Copyright (C) 1996-2015 Free Software Foundation, Inc.  
Copyright (c) 1995 Olivier Devillers <Olivier.Devillers@sophia.inria.fr>  
Copyright (C) 2007-2010 Safe Software Inc.  
Copyright (C) 2010 Safe Software Inc.

Copyright (C) 2006 Refrations Research

Copyright 2004 Sean Gillies, sgillies@frii.com

Copyright (C) 2011 Mateusz Loskot <mateusz@loskot.net>

Copyright (C) 2015 Nyal Dawson <nyall dot dawson at gmail dot com>

Original code (2.0 and earlier )copyright (c) 2000-2006 Lee Thomason  
(www.grinninglizard.com)

Original code (2.0 and earlier )copyright (c) 2000-2002 Lee Thomason  
(www.grinninglizard.com)

License: LGPL V2.1

GNU LESSER GENERAL PUBLIC LICENSE

Version 2.1, February 1999

Copyright (C) 1991, 1999 Free Software Foundation, Inc. 51 Franklin Street, Fifth  
Floor, Boston, MA 02110-1301

Everyone is permitted to copy and distribute verbatim copies of this license  
document, but changing it is not allowed.

Copyright (C) 2005-2011 Refrations Research Inc.

Copyright (C) 2009 Ragi Y. Burhum <ragi@burhum.com>

Copyright (C) 2010 Sandro Santilli <strk@keybit.net>

Copyright (C) 2009 2011 Sandro Santilli <strk@keybit.net>

Copyright (C) 2005 2006 Refrations Research Inc.

Copyright (C) 2011 Sandro Santilli <strk@keybit.net>

Copyright (C) 2006-2011 Refrations Research Inc.

Copyright (C) 2011 Sandro Santilli <strk@keybit.net>

Copyright (C) 2009-2011 Sandro Santilli <strk@keybit.net>

Copyright (C) 2016 Daniel Baston

Copyright (C) 2008 Sean Gillies

Copyright (C) 2009 Sandro Santilli <strk@keybit.net>

Copyright (C) 2006 Refrations Research Inc.

Copyright (C) 2012 Sandro Santilli <strk@keybit.net>

Copyright (C) 2009 Sandro Santilli <strk@keybit.net>

Copyright (C) 2008-2010 Safe Software Inc.

Copyright (C) 2006-2007 Refrations Research Inc.

Copyright (C) 2005-2007 Refrations Research Inc.

Copyright (C) 2007 Refrations Research Inc.  
Copyright (C) 2014 Mika Heiskanen <mika.heiskanen@fmi.fi>  
Copyright (C) 2009-2010 Sandro Santilli <strk@keybit.net>  
Copyright (C) 2009 2011 Sandro Santilli <strk@keybit.net>  
Copyright (C) 2010 Sandro Santilli <strk@keybit.net>  
Copyright (C) 2009 Mateusz Loskot  
Copyright (C) 2005-2009 Refrations Research Inc.  
Copyright (C) 2001-2009 Vivid Solutions Inc.  
Copyright (C) 2012 Sandro Santilli <strk@keybit.net>  
Copyright (C) 2006 Wu Yongwei  
Copyright (C) 2012 Excensus LLC.  
Copyright (C) 1996-2015 Free Software Foundation, Inc.  
Copyright (c) 1995 Olivier Devillers <Olivier.Devillers@sophia.inria.fr>  
Copyright (C) 2007-2010 Safe Software Inc.  
Copyright (C) 2010 Safe Software Inc.  
Copyright (C) 2006 Refrations Research  
Copyright 2004 Sean Gillies, sgillies@frie.com  
Copyright (C) 2011 Mateusz Loskot <mateusz@loskot.net>  
Copyright (C) 2015 Nyal Dawson <nyall dot dawson at gmail dot com>  
Original code (2.0 and earlier )copyright (c) 2000-2006 Lee Thomason  
(www.grinninglizard.com)  
Original code (2.0 and earlier )copyright (c) 2000-2002 Lee Thomason  
(www.grinninglizard.com)

License: LGPL V2.1

GNU LESSER GENERAL PUBLIC LICENSE

Version 2.1, February 1999

Copyright (C) 1991, 1999 Free Software Foundation, Inc. 51 Franklin Street, Fifth Floor, Boston, MA 02110-1301

Everyone is permitted to copy and distribute verbatim copies of this license document, but changing it is not allowed.

Copyright (C) 2005-2011 Refrations Research Inc.

Copyright (C) 2009 Ragi Y. Burhum <ragi@burhum.com>

Copyright (C) 2010 Sandro Santilli <strk@keybit.net>



Copyright (C) 2009 2011 Sandro Santilli <strk@keybit.net>  
Copyright (C) 2005 2006 Refrations Research Inc.  
Copyright (C) 2011 Sandro Santilli <strk@keybit.net>  
Copyright (C) 2006-2011 Refrations Research Inc.  
Copyright (C) 2011 Sandro Santilli <strk@keybit.net>  
Copyright (C) 2009-2011 Sandro Santilli <strk@keybit.net>  
Copyright (C) 2016 Daniel Baston  
Copyright (C) 2008 Sean Gillies  
Copyright (C) 2009 Sandro Santilli <strk@keybit.net>  
Copyright (C) 2006 Refrations Research Inc.  
Copyright (C) 2012 Sandro Santilli <strk@keybit.net>  
Copyright (C) 2009 Sandro Santilli <strk@keybit.net>  
Copyright (C) 2008-2010 Safe Software Inc.  
Copyright (C) 2006-2007 Refrations Research Inc.  
Copyright (C) 2005-2007 Refrations Research Inc.  
Copyright (C) 2007 Refrations Research Inc.  
Copyright (C) 2014 Mika Heiskanen <mika.heiskanen@fmi.fi>  
Copyright (C) 2009-2010 Sandro Santilli <strk@keybit.net>  
Copyright (C) 2009 2011 Sandro Santilli <strk@keybit.net>  
Copyright (C) 2010 Sandro Santilli <strk@keybit.net>  
Copyright (C) 2009 Mateusz Loskot  
Copyright (C) 2005-2009 Refrations Research Inc.  
Copyright (C) 2001-2009 Vivid Solutions Inc.  
Copyright (C) 2012 Sandro Santilli <strk@keybit.net>  
Copyright (C) 2006 Wu Yongwei  
Copyright (C) 2012 Excensus LLC.  
Copyright (C) 1996-2015 Free Software Foundation, Inc.  
Copyright (c) 1995 Olivier Devillers <Olivier.Devillers@sophia.inria.fr>  
Copyright (C) 2007-2010 Safe Software Inc.  
Copyright (C) 2010 Safe Software Inc.  
Copyright (C) 2006 Refrations Research  
Copyright 2004 Sean Gillies, sgillies@frie.com  
Copyright (C) 2011 Mateusz Loskot <mateusz@loskot.net>

Copyright (C) 2015 Nyall Dawson <nyall dot dawson at gmail dot com>

Original code (2.0 and earlier )copyright (c) 2000-2006 Lee Thomason  
(www.grinninglizard.com)

Original code (2.0 and earlier )copyright (c) 2000-2002 Lee Thomason  
(www.grinninglizard.com)

License: LGPL V2.1

GNU LESSER GENERAL PUBLIC LICENSE

Version 2.1, February 1999

Copyright (C) 1991, 1999 Free Software Foundation, Inc. 51 Franklin Street, Fifth  
Floor, Boston, MA 02110-1301

Everyone is permitted to copy and distribute verbatim copies of this license  
document, but changing it is not allowed.

Copyright (C) 2005-2011 Refrations Research Inc.

Copyright (C) 2009 Ragi Y. Burhum <ragi@burhum.com>

Copyright (C) 2010 Sandro Santilli <strk@keybit.net>

Copyright (C) 2009 2011 Sandro Santilli <strk@keybit.net>

Copyright (C) 2005 2006 Refrations Research Inc.

Copyright (C) 2011 Sandro Santilli <strk@keybit.net>

Copyright (C) 2006-2011 Refrations Research Inc.

Copyright (C) 2011 Sandro Santilli <strk@keybit.net

Copyright (C) 2009-2011 Sandro Santilli <strk@keybit.net>

Copyright (C) 2016 Daniel Baston

Copyright (C) 2008 Sean Gillies

Copyright (C) 2009 Sandro Santilli <strk@keybit.net>

Copyright (C) 2006 Refrations Research Inc.

Copyright (C) 2012 Sandro Santilli <strk@keybit.net>

Copyright (C) 2009 Sandro Santilli <strk@keybit.net>

Copyright (C) 2008-2010 Safe Software Inc.

Copyright (C) 2006-2007 Refrations Research Inc.

Copyright (C) 2005-2007 Refrations Research Inc.

Copyright (C) 2007 Refrations Research Inc.

Copyright (C) 2014 Mika Heiskanen <mika.heiskanen@fmi.fi>

Copyright (C) 2009-2010 Sandro Santilli <strk@keybit.net>

Copyright (C) 2009 2011 Sandro Santilli <strk@keybit.net>  
Copyright (C) 2010 Sandro Santilli <strk@keybit.net>  
Copyright (C) 2009 Mateusz Loskot  
Copyright (C) 2005-2009 Refrations Research Inc.  
Copyright (C) 2001-2009 Vivid Solutions Inc.  
Copyright (C) 2012 Sandro Santilli <strk@keybit.net>  
Copyright (C) 2006 Wu Yongwei  
Copyright (C) 2012 Excensus LLC.  
Copyright (C) 1996-2015 Free Software Foundation, Inc.  
Copyright (c) 1995 Olivier Devillers <Olivier.Devillers@sophia.inria.fr>  
Copyright (C) 2007-2010 Safe Software Inc.  
Copyright (C) 2010 Safe Software Inc.  
Copyright (C) 2006 Refrations Research  
Copyright 2004 Sean Gillies, sgillies@frii.com  
Copyright (C) 2011 Mateusz Loskot <mateusz@loskot.net>  
Copyright (C) 2015 Nyal Dawson <nyall dot dawson at gmail dot com>  
Original code (2.0 and earlier )copyright (c) 2000-2006 Lee Thomason  
(www.grinninglizard.com)  
Original code (2.0 and earlier )copyright (c) 2000-2002 Lee Thomason  
(www.grinninglizard.com)

License: LGPL V2.1

GNU LESSER GENERAL PUBLIC LICENSE

Version 2.1, February 1999

Copyright (C) 1991, 1999 Free Software Foundation, Inc. 51 Franklin Street, Fifth Floor, Boston, MA 02110-1301

Everyone is permitted to copy and distribute verbatim copies of this license document, but changing it is not allowed.

Copyright (C) 2005-2011 Refrations Research Inc.

Copyright (C) 2009 Ragi Y. Burhum <ragi@burhum.com>

Copyright (C) 2010 Sandro Santilli <strk@keybit.net>

Copyright (C) 2009 2011 Sandro Santilli <strk@keybit.net>

Copyright (C) 2005 2006 Refrations Research Inc.

Copyright (C) 2011 Sandro Santilli <strk@keybit.net>

Copyright (C) 2006-2011 Refrations Research Inc.  
Copyright (C) 2011 Sandro Santilli <strk@keybit.net>  
Copyright (C) 2009-2011 Sandro Santilli <strk@keybit.net>  
Copyright (C) 2016 Daniel Baston  
Copyright (C) 2008 Sean Gillies  
Copyright (C) 2009 Sandro Santilli <strk@keybit.net>  
Copyright (C) 2006 Refrations Research Inc.  
Copyright (C) 2012 Sandro Santilli <strk@keybit.net>  
Copyright (C) 2009 Sandro Santilli <strk@keybit.net>  
Copyright (C) 2008-2010 Safe Software Inc.  
Copyright (C) 2006-2007 Refrations Research Inc.  
Copyright (C) 2005-2007 Refrations Research Inc.  
Copyright (C) 2007 Refrations Research Inc.  
Copyright (C) 2014 Mika Heiskanen <mika.heiskanen@fmi.fi>  
Copyright (C) 2009-2010 Sandro Santilli <strk@keybit.net>  
Copyright (C) 2009 2011 Sandro Santilli <strk@keybit.net>  
Copyright (C) 2010 Sandro Santilli <strk@keybit.net>  
Copyright (C) 2009 Mateusz Loskot  
Copyright (C) 2005-2009 Refrations Research Inc.  
Copyright (C) 2001-2009 Vivid Solutions Inc.  
Copyright (C) 2012 Sandro Santilli <strk@keybit.net>  
Copyright (C) 2006 Wu Yongwei  
Copyright (C) 2012 Excensus LLC.  
Copyright (C) 1996-2015 Free Software Foundation, Inc.  
Copyright (c) 1995 Olivier Devillers <Olivier.Devillers@sophia.inria.fr>  
Copyright (C) 2007-2010 Safe Software Inc.  
Copyright (C) 2010 Safe Software Inc.  
Copyright (C) 2006 Refrations Research  
Copyright 2004 Sean Gillies, sgillies@frie.com  
Copyright (C) 2011 Mateusz Loskot <mateusz@loskot.net>  
Copyright (C) 2015 Nyal Dawson <nyall dot dawson at gmail dot com>  
Original code (2.0 and earlier )copyright (c) 2000-2006 Lee Thomason  
(www.grinninglizard.com)

Original code (2.0 and earlier )copyright (c) 2000-2002 Lee Thomason  
(www.grinninglizard.com)

License: LGPL V2.1

## GNU LESSER GENERAL PUBLIC LICENSE

Version 2.1, February 1999

Copyright (C) 1991, 1999 Free Software Foundation, Inc. 51 Franklin Street, Fifth Floor, Boston, MA 02110-1301

Everyone is permitted to copy and distribute verbatim copies of this license document, but changing it is not allowed.

[This is the first released version of the Lesser GPL. It also counts as the successor of the GNU Library Public License, version 2, hence the version number 2.1.]

### Preamble

The licenses for most software are designed to take away your freedom to share and change it. By contrast, the GNU General Public

Licenses are intended to guarantee your freedom to share and change free software--to make sure the software is free for all its users.

This license, the Lesser General Public License, applies to some specially designated software packages--typically libraries--of the Free Software Foundation and other authors who decide to use it. You can use it too, but we suggest you first think carefully about whether this license or the ordinary General Public License is the better strategy to use in any particular case, based on the explanations below.

When we speak of free software, we are referring to freedom of use, not price. Our General Public Licenses are designed to make sure that you have the freedom to distribute copies of free software (and charge for this service if you wish); that you receive source code or can get it if you want it; that you can change the software and use pieces of it in new free programs; and that you are informed that you can do these things.

To protect your rights, we need to make restrictions that forbid distributors to deny you these rights or to ask you to surrender these rights. These restrictions translate to certain responsibilities for you if you distribute copies of the library or if you modify it.

For example, if you distribute copies of the library, whether gratis or for a fee, you must give the recipients all the rights that we gave you. You must make sure that they, too, receive or can get the source code. If you link other code with the library, you must provide complete object files to the recipients, so that they can relink them with the library after making changes to the library and recompiling it. And you must show them these terms so they know their rights.

We protect your rights with a two-step method: (1) we copyright the library, and (2) we offer you this license, which gives you legal permission to copy, distribute and/or modify the library.

To protect each distributor, we want to make it very clear that there is no warranty for the free library. Also, if the library is modified by someone else and passed on, the recipients should know that what they have is not the original version, so that the original author's reputation will not be affected by problems that might be introduced by others.

Finally, software patents pose a constant threat to the existence of any free program. We wish to make sure that a company cannot effectively restrict the users of a free program by obtaining a restrictive license from a patent holder. Therefore, we insist that any patent license obtained for a version of the library must be consistent with the full freedom of use specified in this license.

Most GNU software, including some libraries, is covered by the ordinary GNU General Public License. This license, the GNU Lesser General Public License, applies to certain designated libraries, and

is quite different from the ordinary General Public License. We use this license for certain libraries in order to permit linking those libraries into non-free programs.

When a program is linked with a library, whether statically or using a shared library, the combination of the two is legally speaking a combined work, a derivative of the original library. The ordinary General Public License therefore permits such linking only if the entire combination fits its criteria of freedom. The Lesser General Public License permits more lax criteria for linking other code with the library.

We call this license the "Lesser" General Public License because it does Less to protect the user's freedom than the ordinary General Public License. It also provides other free software developers Less of an advantage over competing non-free programs. These disadvantages are the reason we use the ordinary General Public License for many libraries. However, the Lesser license provides advantages in certain special circumstances.

For example, on rare occasions, there may be a special need to encourage the widest possible use of a certain library, so that it becomes a de-facto standard. To achieve this, non-free programs must be allowed to use the library. A more frequent case is that a free library does the same job as widely used non-free libraries. In this case, there is little to gain by limiting the free library to free software only, so we use the Lesser General Public License.

In other cases, permission to use a particular library in non-free programs enables a greater number of people to use a large body of free software. For example, permission to use the GNU C Library in

non-free programs enables many more people to use the whole GNU operating system, as well as its variant, the GNU/Linux operating system.

Although the Lesser General Public License is Less protective of the users' freedom, it does ensure that the user of a program that is linked with the Library has the

freedom and the wherewithal to run that program using a modified version of the Library.

The precise terms and conditions for copying, distribution and modification follow. Pay close attention to the difference between a "work based on the library" and a "work that uses the library". The

former contains code derived from the library, whereas the latter must be combined with the library in order to run.

## GNU LESSER GENERAL PUBLIC LICENSE

### TERMS AND CONDITIONS FOR COPYING, DISTRIBUTION AND MODIFICATION

0. This License Agreement applies to any software library or other program which contains a notice placed by the copyright holder or other authorized party saying it may be distributed under the terms of this Lesser General Public License (also called "this License"). Each licensee is addressed as "you".

A "library" means a collection of software functions and/or data prepared so as to be conveniently linked with application programs (which use some of those functions and data) to form executables.

The "Library", below, refers to any such software library or work which has been distributed under these terms. A "work based on the Library" means either the Library or any derivative work under

copyright law: that is to say, a work containing the Library or a portion of it, either verbatim or with modifications and/or translated straightforwardly into another language. (Hereinafter, translation is included without limitation in the term "modification".)

"Source code" for a work means the preferred form of the work for making modifications to it. For a library, complete source code means all the source code for all modules it contains, plus any associated interface definition files, plus the scripts used to control compilation and installation of the library.

Activities other than copying, distribution and modification are not covered by this License; they are outside its scope. The act of running a program using the Library is not restricted, and output from such a program is covered only if its contents constitute a work based on the Library (independent of the use of the Library in a tool for writing it). Whether that is true depends on what the Library does and what the program that uses the Library does.

1. You may copy and distribute verbatim copies of the Library's complete source code as you receive it, in any medium, provided that you conspicuously and appropriately publish on each copy an

appropriate copyright notice and disclaimer of warranty; keep intact all the notices that refer to this License and to the absence of any warranty; and distribute a copy of this License along with the

Library.

You may charge a fee for the physical act of transferring a copy, and you may at your option offer warranty protection in exchange for a fee.

2. You may modify your copy or copies of the Library or any portion of it, thus forming a work based on the Library, and copy and distribute such modifications or work under the terms of Section 1

above, provided that you also meet all of these conditions:

a) The modified work must itself be a software library.

b) You must cause the files modified to carry prominent notices stating that you changed the files and the date of any change.

c) You must cause the whole of the work to be licensed at no charge to all third parties under the terms of this License.

d) If a facility in the modified Library refers to a function or a table of data to be supplied by an application program that uses the facility, other than as an argument passed when the facility is invoked, then you must make a good faith effort to ensure that, in the event an application does not supply such function or table, the facility still operates, and performs whatever part of

its purpose remains meaningful.

(For example, a function in a library to compute square roots has a purpose that is entirely well-defined independent of the application. Therefore, Subsection 2d requires that any application-supplied function or table used by this function must be optional: if the application does not supply it, the square root function must still compute square roots.)

These requirements apply to the modified work as a whole. If identifiable sections of that work are not derived from the Library, and can be reasonably considered independent and separate works in

themselves, then this License, and its terms, do not apply to those sections when you distribute them as separate works. But when you distribute the same sections as part of a whole which is a work based on the Library, the distribution of the whole must be on the terms of this License, whose permissions for other licensees extend to the entire whole, and thus to each and every part regardless of who wrote it.

Thus, it is not the intent of this section to claim rights or contest your rights to work written entirely by you; rather, the intent is to exercise the right to control the distribution of derivative or

collective works based on the Library.

In addition, mere aggregation of another work not based on the Library with the Library (or with a work based on the Library) on a volume of a storage or distribution medium does not bring the other work under the scope of this License.



3. You may opt to apply the terms of the ordinary GNU General Public License instead of this License to a given copy of the Library. To do this, you must alter all the notices that refer to this License, so that they refer to the ordinary GNU General Public License, version 2, instead of to this License. (If a newer version than version 2 of the ordinary GNU General Public License has appeared, then you can specify that version instead if you wish.) Do not make any other change in these notices.

Once this change is made in a given copy, it is irreversible for that copy, so the ordinary GNU General Public License applies to all subsequent copies and derivative works made from that copy.

This option is useful when you wish to copy part of the code of the Library into a program that is not a library.

4. You may copy and distribute the Library (or a portion or derivative of it, under Section 2) in object code or executable form under the terms of Sections 1 and 2 above provided that you accompany

it with the complete corresponding machine-readable source code, which must be distributed under the terms of Sections 1 and 2 above on a medium customarily used for software interchange.

If distribution of object code is made by offering access to copy from a designated place, then offering equivalent access to copy the source code from the same place satisfies the requirement to

distribute the source code, even though third parties are not compelled to copy the source along with the object code.

5. A program that contains no derivative of any portion of the Library, but is designed to work with the Library by being compiled or linked with it, is called a "work that uses the Library". Such a

work, in isolation, is not a derivative work of the Library, and therefore falls outside the scope of this License.

However, linking a "work that uses the Library" with the Library creates an executable that is a derivative of the Library (because it contains portions of the Library), rather than a "work that uses the library". The executable is therefore covered by this License.

Section 6 states terms for distribution of such executables.

When a "work that uses the Library" uses material from a header file that is part of the Library, the object code for the work may be a derivative work of the Library even though the source code is not. Whether this is true is especially significant if the work can be linked without the Library, or if the work is itself a library. The threshold for this to be true is not precisely defined by law.

If such an object file uses only numerical parameters, data structure layouts and accessors, and small macros and small inline functions (ten lines or less in length), then the use of the object

file is unrestricted, regardless of whether it is legally a derivative work. (Executables containing this object code plus portions of the Library will still fall under Section 6.)

Otherwise, if the work is a derivative of the Library, you may distribute the object code for the work under the terms of Section 6. Any executables containing that work also fall under Section 6,

whether or not they are linked directly with the Library itself.

6. As an exception to the Sections above, you may also combine or link a "work that uses the Library" with the Library to produce a work containing portions of the Library, and distribute that work

under terms of your choice, provided that the terms permit modification of the work for the customer's own use and reverse engineering for debugging such modifications.

You must give prominent notice with each copy of the work that the Library is used in it and that the Library and its use are covered by this License. You must supply a copy of this License. If the work during execution displays copyright notices, you must include the copyright notice for the Library among them, as well as a reference directing the user to the copy of this License. Also, you must do one of these things:

a) Accompany the work with the complete corresponding machine-readable source code for the Library including whatever changes were used in the work (which must be distributed under Sections 1 and 2 above); and, if the work is an executable linked with the Library, with the complete machine-readable "work that uses the Library", as object code and/or source code, so that the user can modify the Library and then relink to produce a modified executable containing the modified Library. (It is understood that the user who changes the contents of definitions files in the Library will not necessarily be able to recompile the application to use the modified definitions.)

b) Use a suitable shared library mechanism for linking with the Library. A suitable mechanism is one that (1) uses at run time a copy of the library already present on the user's computer system,

rather than copying library functions into the executable, and (2) will operate properly with a modified version of the library, if the user installs one, as long as the modified version is interface-compatible with the version that the work was made with.

c) Accompany the work with a written offer, valid for at least three years, to give the same user the materials specified in Subsection 6a, above, for a charge no more than the cost of performing this distribution.

d) If distribution of the work is made by offering access to copy from a designated place, offer equivalent access to copy the above specified materials from the same place.

e) Verify that the user has already received a copy of these materials or that you have already sent this user a copy.

For an executable, the required form of the "work that uses the Library" must include any data and utility programs needed for reproducing the executable from it. However, as a special exception,

the materials to be distributed need not include anything that is normally distributed (in either source or binary form) with the major components (compiler, kernel, and so on) of the operating system on

which the executable runs, unless that component itself accompanies the executable.

It may happen that this requirement contradicts the license restrictions of other proprietary libraries that do not normally accompany the operating system. Such a contradiction means you cannot

use both them and the Library together in an executable that you distribute.

7. You may place library facilities that are a work based on the Library side-by-side in a single library together with other library facilities not covered by this License, and distribute such a combined library, provided that the separate distribution of the work based on the Library and of the other library facilities is otherwise permitted, and provided that you do these two things:

a) Accompany the combined library with a copy of the same work based on the Library, uncombined with any other library facilities. This must be distributed under the terms of the Sections above.

b) Give prominent notice with the combined library of the fact that part of it is a work based on the Library, and explaining where to find the accompanying uncombined form of the same work.

8. You may not copy, modify, sublicense, link with, or distribute the Library except as expressly provided under this License. Any attempt otherwise to copy, modify, sublicense, link with, or distribute the Library is void, and will automatically terminate your rights under this License. However, parties who have received copies, or rights, from you under this License will not have their licenses terminated so long as such parties remain in full compliance.

9. You are not required to accept this License, since you have not signed it. However, nothing else grants you permission to modify or distribute the Library or its derivative works. These actions are prohibited by law if you do not accept this License. Therefore, by modifying or distributing the Library (or any work based on the Library), you indicate your acceptance of this License to do so, and all its terms and conditions for copying, distributing or modifying the Library or works based on it.

10. Each time you redistribute the Library (or any work based on the Library), the recipient automatically receives a license from the original licensor to copy, distribute, link with or modify the Library subject to these terms and conditions.

You may not impose any further restrictions on the recipients' exercise of the rights granted herein.

You are not responsible for enforcing compliance by third parties with this License.

11. If, as a consequence of a court judgment or allegation of patent infringement or for any other reason (not limited to patent issues), conditions are imposed on you (whether by court order, agreement or otherwise) that contradict the conditions of this License, they do not excuse you from the conditions of this License. If you cannot distribute so as to satisfy simultaneously your obligations under this License and any other pertinent obligations, then as a consequence you may not distribute the Library at all. For example, if a patent license would not permit royalty-free redistribution of the Library by all those who receive copies directly or indirectly through you, then the only way you could satisfy both it and this License would be to refrain entirely from distribution of the Library.

If any portion of this section is held invalid or unenforceable under any particular circumstance, the balance of the section is intended to apply, and the section as a whole is intended to apply in other circumstances.

It is not the purpose of this section to induce you to infringe any patents or other property right claims or to contest validity of any such claims; this section has the sole purpose of protecting the

integrity of the free software distribution system which is implemented by public license practices. Many people have made generous contributions to the wide range of software distributed through that system in reliance on consistent application of that system; it is up to the author/donor to decide if he or she is willing to distribute software through any other system and a licensee cannot impose that choice.

This section is intended to make thoroughly clear what is believed to be a consequence of the rest of this License.

12. If the distribution and/or use of the Library is restricted in certain countries either by patents or by copyrighted interfaces, the original copyright holder who places the Library under this License may add an explicit geographical distribution limitation excluding those countries, so that distribution is permitted only in or among countries not thus excluded. In such case, this License incorporates the limitation as if written in the body of this License.

13. The Free Software Foundation may publish revised and/or new versions of the Lesser General Public License from time to time.

Such new versions will be similar in spirit to the present version, but may differ in detail to address new problems or concerns.

Each version is given a distinguishing version number. If the Library specifies a version number of this License which applies to it and "any later version", you have the option of following the terms and conditions either of that version or of any later version published by the Free Software Foundation. If the Library does

not specify a license version number, you may choose any version ever published by the Free Software Foundation.

14. If you wish to incorporate parts of the Library into other free programs whose distribution conditions are incompatible with these, write to the author to ask for permission. For software which is

copyrighted by the Free Software Foundation, write to the Free Software Foundation; we sometimes make exceptions for this. Our decision will be guided by the two goals of preserving the free status

of all derivatives of our free software and of promoting the sharing and reuse of software generally.

#### NO WARRANTY

15. BECAUSE THE LIBRARY IS LICENSED FREE OF CHARGE, THERE IS NO WARRANTY FOR THE LIBRARY, TO THE EXTENT PERMITTED BY APPLICABLE LAW. EXCEPT WHEN OTHERWISE STATED IN WRITING THE COPYRIGHT HOLDERS AND/OR OTHER PARTIES PROVIDE THE LIBRARY "AS IS" WITHOUT WARRANTY OF ANY KIND, EITHER EXPRESSED OR IMPLIED, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE. THE ENTIRE RISK AS TO THE QUALITY AND PERFORMANCE OF THE LIBRARY IS WITH YOU. SHOULD THE LIBRARY PROVE DEFECTIVE, YOU ASSUME THE COST OF ALL NECESSARY SERVICING, REPAIR OR CORRECTION.

16. IN NO EVENT UNLESS REQUIRED BY APPLICABLE LAW OR AGREED TO IN WRITING WILL ANY COPYRIGHT HOLDER, OR ANY OTHER PARTY WHO MAY MODIFY AND/OR REDISTRIBUTE THE LIBRARY AS PERMITTED ABOVE, BE LIABLE TO YOU FOR DAMAGES, INCLUDING ANY GENERAL, SPECIAL, INCIDENTAL OR CONSEQUENTIAL DAMAGES ARISING OUT OF THE USE OR INABILITY TO USE THE LIBRARY (INCLUDING BUT NOT LIMITED TO LOSS OF DATA OR DATA BEING RENDERED INACCURATE OR LOSSES SUSTAINED BY YOU OR THIRD PARTIES OR A FAILURE OF THE LIBRARY TO OPERATE WITH ANY OTHER SOFTWARE), EVEN IF SUCH HOLDER OR OTHER PARTY HAS BEEN ADVISED OF THE POSSIBILITY OF SUCH DAMAGES.

#### END OF TERMS AND CONDITIONS

#### How to Apply These Terms to Your New Libraries

If you develop a new library, and you want it to be of the greatest possible use to the public, we recommend making it free software that everyone can redistribute and change. You can do so by permitting redistribution under these terms (or, alternatively, under the terms of the ordinary General Public License).

To apply these terms, attach the following notices to the library. It is safest to attach them to the start of each source file to most effectively convey the exclusion of warranty; and each file should have at least the "copyright" line and a pointer to where the full notice is found.

<one line to give the library's name and a brief idea of what it does.>

Copyright (C) <year> <name of author>

This library is free software; you can redistribute it and/or modify it under the terms of the GNU Lesser General Public License as published by the Free Software Foundation; either version 2.1 of the License, or (at your option) any later version.

This library is distributed in the hope that it will be useful, but WITHOUT ANY WARRANTY; without even the implied warranty of MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE. See the GNU

Lesser General Public License for more details.

You should have received a copy of the GNU Lesser General Public License along with this library; if not, write to the Free Software Foundation, Inc., 51 Franklin Street, Fifth Floor, Boston, MA 02110-1301

Also add information on how to contact you by electronic and paper mail.

You should also get your employer (if you work as a programmer) or your school, if any, to sign a "copyright disclaimer" for the library, if necessary. Here is a sample; alter the names:

Yoyodyne, Inc., hereby disclaims all copyright interest in the library `Frob' (a library for tweaking knobs) written by James Random Hacker.

<signature of Ty Coon>, 1 April 1990

Ty Coon, President of Vice

That's all there is to it!

Software: JSON-C

Copyright notice:

Copyright (c) 2004, 2005 Metaparadigm Pte. Ltd.

Copyright (c) 2009-2012 Eric Haszlakiewicz

Copyright (c) 2004, 2005 Metaparadigm Pte Ltd

Copyright (c) 2009 Hewlett-Packard Development Company, L.P.

Copyright 2011, John Resig

Copyright 2011, The Dojo Foundation

Copyright (c) 2012 Eric Haszlakiewicz

Copyright (c) 2009-2012 Hewlett-Packard Development Company, L.P.

Copyright (c) 2008-2009 Yahoo! Inc. All rights reserved.

Copyright (C) 1996, 1997, 1998, 1999, 2000, 2001, 2003, 2004, 2005, 2006,  
2007, 2008, 2009, 2010, 2011 Free Software Foundation, Inc.

Copyright (c) 2013 Metaparadigm Pte. Ltd.

License: MIT License

Copyright (c) 2009-2012 Eric Haszlakiewicz

Permission is hereby granted, free of charge, to any person obtaining a copy of this software and associated documentation files (the "Software"), to deal in the Software without restriction, including without limitation the rights to use, copy, modify, merge, publish, distribute, sublicense, and/or sell copies of the Software, and to permit persons to whom the Software is furnished to do so, subject to the following conditions:

The above copyright notice and this permission notice shall be included in all copies or substantial portions of the Software.

THE SOFTWARE IS PROVIDED "AS IS", WITHOUT WARRANTY OF ANY KIND, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO THE WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND NONINFRINGEMENT. IN NO EVENT SHALL THE AUTHORS OR COPYRIGHT HOLDERS BE LIABLE FOR ANY CLAIM, DAMAGES OR OTHER LIABILITY, WHETHER IN AN ACTION OF CONTRACT, TORT OR OTHERWISE, ARISING FROM, OUT OF OR IN CONNECTION WITH THE SOFTWARE OR THE USE OR OTHER DEALINGS IN THE SOFTWARE.

-----  
Copyright (c) 2004, 2005 Metaparadigm Pte Ltd

Permission is hereby granted, free of charge, to any person obtaining a copy of this software and associated documentation files (the "Software"), to deal in the Software without restriction, including without limitation the rights to use, copy, modify, merge, publish, distribute, sublicense, and/or sell copies of the Software, and to permit persons to whom the Software is furnished to do so, subject to the following conditions:

The above copyright notice and this permission notice shall be included in all copies or substantial portions of the Software.

THE SOFTWARE IS PROVIDED "AS IS", WITHOUT WARRANTY OF ANY KIND, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO THE WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND NONINFRINGEMENT. IN NO EVENT SHALL THE AUTHORS OR COPYRIGHT HOLDERS BE LIABLE FOR ANY CLAIM, DAMAGES OR OTHER LIABILITY, WHETHER IN AN ACTION OF CONTRACT, TORT OR OTHERWISE, ARISING FROM, OUT OF OR IN CONNECTION WITH THE SOFTWARE OR THE USE OR OTHER DEALINGS IN THE SOFTWARE.

Software: proj  
Copyright notice:  
"Copyright (C) 2010 Mateusz Loskot <mateusz@loskot.net>  
Copyright (C) 2007 Douglas Gregor <doug.gregor@gmail.com>  
Copyright (C) 2007 Troy Straszheim  
CMake, Copyright (C) 2009-2010 Mateusz Loskot <mateusz@loskot.net> )  
Copyright (C) 2011 Nicolas David <nicolas.david@ign.fr>  
Copyright (c) 2000, Frank Warmerdam  
Copyright (c) 2011, Open Geospatial Consortium, Inc.  
Copyright (C) 1996, 1997, 1998, 1999, 2000, 2001, 2003, 2004, 2005, 2006,  
2007, 2008, 2009, 2010, 2011 Free Software Foundation, Inc.  
Copyright (c) Charles Karney (2012-2015) <charles@karney.com> and licensed  
Copyright (c) 2005, Antonello Andrea  
Copyright (c) 2010, Frank Warmerdam  
Copyright (c) 1995, Gerald Evenden  
Copyright (c) 2000, Frank Warmerdam <warmerdam@pobox.com>  
Copyright (c) 2010, Frank Warmerdam <warmerdam@pobox.com>  
Copyright (c) 2013, Frank Warmerdam  
Copyright (c) 2003 Gerald I. Evenden  
Copyright (c) 2012, Frank Warmerdam <warmerdam@pobox.com>  
Copyright (c) 2002, Frank Warmerdam  
Copyright (c) 2004 Gerald I. Evenden  
Copyright (c) 2012 Martin Raspaud  
Copyright (c) 2001, Thomas Flemming, tf@ttqv.com  
Copyright (c) 2002, Frank Warmerdam <warmerdam@pobox.com>  
Copyright (c) 2009, Frank Warmerdam  
Copyright (c) 2003, 2006 Gerald I. Evenden  
Copyright (c) 2011, 2012 Martin Lambers <marlam@marlam.de>  
Copyright (c) 2006, Andrey Kiselev  
Copyright (c) 2008-2012, Even Rouault <even dot rouault at mines-paris dot org>  
Copyright (c) 2001, Frank Warmerdam  
Copyright (c) 2001, Frank Warmerdam <warmerdam@pobox.com>  
Copyright (c) 2008 Gerald I. Evenden



"

License: MIT License

Please see above

Software: libxml2

Copyright notice:

"See Copyright for the status of this software.

Copyright (C) 1998-2003 Daniel Veillard. All Rights Reserved.

Copyright (C) 2003 Daniel Veillard.

copy: see Copyright for the status of this software.

copy: see Copyright for the status of this software

copy: see Copyright for the status of this software.

Copyright (C) 2000 Bjorn Reese and Daniel Veillard.

Copy: See Copyright for the status of this software.

See COPYRIGHT for the status of this software

Copyright (C) 2000 Gary Pennington and Daniel Veillard.

Copyright (C) 1996, 1997, 1998, 1999, 2000, 2001, 2003, 2004, 2005, 2006,  
2007 Free Software Foundation, Inc.

Copyright (C) 1998 Bjorn Reese and Daniel Stenberg.

Copyright (C) 2001 Bjorn Reese <breese@users.sourceforge.net>

Copyright (C) 2000 Bjorn Reese and Daniel Stenberg.

Copyright (C) 2001 Bjorn Reese and Daniel Stenberg.

See Copyright for the status of this software

"

License: MIT License

Please see above

# 11 实时数仓

## 11.1 实时数仓简介

实时数仓需要支持将insert+upsert+update等操作实时快速入库，数据来源于上游的其他数据库或者应用，同时要求入库后的数据要能及时查询，对于查询的效率要求很高。

目前GaussDB(DWS)标准数仓已有的行存表或者列存表都无法同时满足实时入库和实时查询两个诉求。其中行存表实时入库能力强，支持高并发更新，但是磁盘占用高，查询效率低；列存表数据压缩率高，AP查询性能好，但是不能很好的支持并发更新，并发入库存在严重的锁冲突。

为了解决上面的问题，需要在使用列存储格式尽量降低磁盘占用的同时，支持高并发的更新操作入库以及高性能的查询效率。GaussDB(DWS)的实时数仓中的HStore表就是针对这种情况设计和实现的，面向对于实时入库和实时查询有较强诉求的场景，同时拥有处理传统TP场景的事务能力。

GaussDB(DWS)提供的实时数仓中实现了一种全新的HStore表，可以做到单条或者小批量IUD操作的高并发实时入库，也可以支持大批量的定期入库。数据入库提交后即可查询，无任何时延。支持主键等传统索引能力去重和加速点查，也支持分区、多维字典、局部排序等方式进一步加速AP查询，也可以在TPCC这种强事务压力场景下保证数据强一致性。

### 📖 说明

- 实时数仓的HStore表仅8.2.0.100及以上集群版本支持。
- HStore表当前的冷热数据与时序表暂未商用，后续版本会支持。
- HStore表是实时数仓中设计的一种表类型，与SQL参数hstore没有任何关系。

## 与标准数仓的区别

实时数仓与标准数仓是GaussDB(DWS)的两种不同类型产品，在使用上也存在一定差异，具体可参考[表11-1](#)进行对比分析。

表 11-1 实时数仓与标准数仓的差异

数仓类型	标准数仓	实时数仓
适用场景	融合分析业务，一体化OLAP分析场景。主要应用于金融、政企、电商、能源等领域。	实时入库+分析混合业务，上游数据实时入库+数据入库后实时高效查询场景。主要用于电商、金融等实时入库要求高的场景。
产品优势	性价比高，使用场景广泛。支持冷热数据分析，存储、计算弹性伸缩，无限算力、无限容量等。	混合负载，入库性能强。提供与列存相当的高性能查询效率与高压缩率的数据压缩能力。同时拥有处理传统TP场景的事务能力。
功能特点	支持海量数据离线处理和交互查询，数据规模大、复杂数据挖掘具有很好的性能优势。	支持海量数据高并发的更新操作入库以及高性能的查询效率。在数据规模大、入库并发高、查询要求高的场景下具有很好的性能优势。
SQL语法	SQL语法兼容性高，语法通用，易于使用。	兼容列存语法。
GUC参数	丰富的GUC参数，根据客户业务场景适配最适合客户的数仓环境。	兼容标准数仓GUC参数，同时支持实时数仓调优参数。

## 技术特点

- 完整的事务一致性**

体现在数据插入或者更新后提交即可见，不存在时延；并发更新后数据保证强一致，不会出现乱序导致的结果预期不一致。
- 查询性能好**

多表关联等复杂AP查询场景下，更完善的分布式查询计划与分布式执行器带来的性能优势，支持复杂的子查询和存储过程。
- 入库快**

彻底解决列存CU锁冲突问题，支持高并发的更新入库操作，典型场景下，并发更新性能是之前的百倍以上。
- 高压缩**

数据在MERGE进入列存主表后，按列存储具有天然的压缩优势，能极大地节省磁盘空间与IO资源。
- 查询加速**

支持主键等传统索引能力去重和加速点查，也支持分区、多维字典、局部排序等方式进一步加速AP查询。

## 行存、列存、HStore 表对比

表 11-2 行存、列存、HStore 表对比

表类型	行存表	列存表	HStore表
数据存储方式	以元组为单位，将每一条数据的所有属性值存储到临近的空间里。	以CU（Compress Unit）为单位，将单个属性的所有值存储到临近的空间里。	数据主要以CU形式存储在列存主表上，对于被更新的列、小批量插入的数据将被序列化后存储到新设计的Delta表上。
数据写入	行存压缩暂未商用，数据按原始状态存储，磁盘空间占用较大。	按列存储时，由于属性值类型相同具有天然的压缩优势。数据写入时能极大节省IO资源与磁盘空间占用。	批量插入的数据直接写入CU，具有与列存一致的压缩优势。 被更新的列、小批量插入的数据会序列化后压缩。同时定期merge到主表CU。
数据更新	数据按行更新，没有CU锁问题，并发更新（update/insert/delete等）性能好。	即使更新单条数据，也要获取整个CU的锁，基本无法支持并发更新（update/insert/delete等）。	彻底解决列存更新的CU锁问题，并发更新（update/insert/delete等）的性能达到行存的60%以上。
数据读取	按行读取，即使只需访问某一列的数据，也需要将一整行的数据取出。查询性能较差。	按列读取时只需访问该列的CU，再加上CU的压缩优势导致需要占用的IO资源更少，读取性能很好。	对于列存主表的数据按列读取，对于被更新的列、小批量插入的数据会反序列化后取出，数据merge到主表后具有与列存一致的数据读取优势。
优点	并发更新性能好。	查询性能好，磁盘占用空间少。	并发更新性能好，数据MERGE后具有与列存一致的查询性能优势与压缩优势。
缺点	占用磁盘空间多，查询性能差。	基本无法支持并发更新。	需要后台常驻线程对HStore表进行merge清理操作。先merge到CU主表再进行清理，与SQL语法中的Merge无关。
适用场景	<ol style="list-style-type: none"> <li>更新删除操作频繁的TP事务场景。</li> <li>点查询（基于索引的、返回数据量小的简单查询）。</li> </ol>	<ol style="list-style-type: none"> <li>查询分析为主的AP场景。</li> <li>数据量大，存入后的更新删除操作少。</li> </ol>	<ol style="list-style-type: none"> <li>实时并发入库场景。</li> <li>需要支持高并发的更新入库操作以及高性能的查询效率。</li> </ol>

## 11.2 支持与限制

实时数仓兼容所有列存语法，具体情况如下：

表 11-3 支持的语法

语法	是否支持
CREATE TABLE	支持
CREATE TABLE LIKE	支持
DROP TABLE	支持
INSERT	支持
COPY	支持
SELECT	支持
TRUNCATE	支持
EXPLAIN	支持
ANALYZE	支持
VACUUM	支持
ALTER TABLE DROP PARTITION	支持
ALTER TABLE ADD PARTITION	支持
ALTER TABLE SET WITH OPTION	支持
ALTER TABLE DROP COLUMN	支持
ALTER TABLE ADD COLUMN	支持
ALTER TABLE ADD NODELIST	支持
ALTER TABLE CHANGE OWNER	支持
ALTER TABLE RENAME COLUMN	支持
ALTER TABLE TRUNCATE PARTITION	支持
CREATE INDEX	支持
DROP INDEX	支持
DELETE	支持
ALTER TABLE 其他	支持
ALTER INDEX	支持
MERGE	支持

语法	是否支持
SELECT INTO	支持
UPDATE	支持
CREATE TABLE AS	支持

## 约束限制

1. 当需要使用HStore表时，需要同步修改以下几个参数的默认值，否则会导致HStore表性能严重劣化。  
推荐的参数修改配置是：autovacuum\_max\_workers\_hstore=3，autovacuum\_max\_workers=6，autovacuum=true。
2. 当前HStore与列存都不支持使用vacuum清理索引脏数据，在频繁update场景可能会导致索引膨胀，后续版本会支持。

## 11.3 实时数仓语法

### 11.3.1 CREATE TABLE

#### 功能描述

在当前数据库中创建一个新的空白HStore表，该表由命令执行者所有。

实时数仓提供创建HStore表DDL语句。创建HStore表DDL需要指定enable\_hstore为true，同时需要将orientation属性设置为column。

#### 注意事项

- 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）。

#### 语法格式

```
CREATE TABLE [ IF NOT EXISTS ] table_name
({ column_name data_type
  | LIKE source_table [like_option [...]] }
)
[, ... ]
[ WITH ( {storage_parameter = value} [, ... ] ) ]
[ TABLESPACE tablespace_name ]
[ DISTRIBUTE BY HASH ( column_name [...]) ]
[ TO { GROUP groupname | NODE ( nodename [, ... ] ) } ]
[ PARTITION BY {
  {RANGE (partition_key) ( partition_less_than_item [, ... ] ) }
} [ { ENABLE | DISABLE } ROW MOVEMENT ] ];
其中like选项like_option为:
{ INCLUDING | EXCLUDING } { DEFAULTS | CONSTRAINTS | INDEXES | STORAGE | COMMENTS | PARTITION
| RELOPTIONS | DISTRIBUTION | ALL }
```

## 列存表的 Delta 表差异

表 11-4 HStore 表与列存表的辅助 Delta 表差异

数仓类型	列存的delta表	HStore的delta表
表结构	与列存主表的表定义一致	与主表表定义不一样。
功能	用于暂存小批量insert的数据，满阈值后再merge到主表，避免直接insert到主表产生大量小CU。	用于持久化存储update/delete/insert信息。在故障拉起后用于恢复内存更新链等管理并发更新的内存结构。
缺陷	来不及merge导致delta表膨胀，影响查询性能，同时无法解决并发update的锁冲突问题	依赖后台常驻autovacuum来做merge操作。

## 参数说明

- 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。
- table\_name**  
 要创建的表名。  
 表名长度不超过63个字符，以字母或下划线开头，可包含字母、数字、下划线、\$、#。
- column\_name**  
 新表中要创建的字段名。  
 字段名长度不超过63个字符，以字母或下划线开头，可包含字母、数字、下划线、\$、#。
- data\_type**  
 字段的数据类型。
- 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字段名及其数据类型。  
 新表与原表之间在创建动作完毕之后是完全无关的。在原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原表的时候包含新表的数据。  
 被复制的列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。  
 HStore表只能从HStore表中进行继承。
- WITH ( { storage\_parameter = value } [, ... ] )**  
 这个子句为表指定一个可选的存储参数。
  - ORIENTATION**  
 指定表数据的存储方式，即时序方式、行存方式、列存方式，该参数设置成功后就不再支持修改。对于HStore表，应当使用列存方式，同时设置enable\_hstore为true。

取值范围:

- TIMESERIES, 表示表的数据将以时序方式存储。
- COLUMN, 表示表的数据将以列存方式存储。
- ROW, 表示表的数据将以行方式存储。

默认值: ROW。

- 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。

取值范围:

- HStore表和列存表的有效值为YES/NO和/LOW/MIDDLE/HIGH, 默认值为LOW。
- 行存表的有效值为YES/NO, 默认值为NO。

- 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。该参数只对时序表和列存表有效。

取值范围: 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。该参数只对时序表和列存表有效。

取值范围: 10000~60000

默认值60000

- PARTIAL\_CLUSTER\_ROWS

指定了在数据加载过程中进行将局部聚簇存储的记录数目。该参数只对时序表和列存表有效。

取值范围: 600000~2147483647

- enable\_delta

指定了在列存表是否开启delta表。对HStore表不能开启该参数。

默认值: off

- SUB\_PARTITION\_COUNT

指定二级分区的个数。该参数用于设置在导入阶段二级分区个数。在建表时进行设置, 建表后不支持修改。不建议用户随意设置该默认值, 可能会影响导入和查询的性能。

取值范围: 1 ~ 1024

默认值: 32

- DELTAROW\_THRESHOLD

指定HStore表导入时小于多少行( $SUB\_PARTITION\_COUNT * DELTAROW\_THRESHOLD$ )的数据进入delta表。



- 取值范围：0~60000  
默认值为60000
- COLVERSION  
指定存储格式的版本。HStore表只支持2.0版本。  
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。  
默认值：2.0
  - DISTRIBUTE BY  
指定表如何在节点之间分布或者复制。  
取值范围：  
HASH (column\_name )：对指定的列进行Hash，通过映射，把数据分布到指定DN。
  - 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，一般用户不应该使用。
  - PARTITION BY  
指定HStore表的初始分区。

## 示例

创建简单的HStore表:

```
CREATE TABLE warehouse_t1
(
  W_WAREHOUSE_SK      INTEGER          NOT NULL,
  W_WAREHOUSE_ID     CHAR(16)          NOT NULL,
  W_WAREHOUSE_NAME    VARCHAR(20)
  W_WAREHOUSE_SQ_FT   INTEGER
  W_STREET_NUMBER     CHAR(10)
  W_STREET_NAME       VARCHAR(60)
  W_STREET_TYPE       CHAR(15)
  W_SUITE_NUMBER      CHAR(10)
  W_CITY              VARCHAR(60)
  W_COUNTY            VARCHAR(30)
  W_STATE             CHAR(2)
  W_ZIP               CHAR(10)
  W_COUNTRY           VARCHAR(20)
  W_GMT_OFFSET        DECIMAL(5,2)
)WITH(ORIENTATION=COLUMN, ENABLE_HSTORE=ON);

CREATE TABLE warehouse_t2 (LIKE warehouse_t1 INCLUDING ALL);
```

## 11.3.2 INSERT

### 功能描述

往HStore表中插入一行或多行数据。

### 注意事项

- 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DELTA\_ROW\_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

```
INSERT [/*+ plan_hint */] [ IGNORE | OVERWRITE ] INTO table_name [ AS alias ] [ ( column_name [, ...] ) ]
{ DEFAULT VALUES
| VALUES {{ ( expression | DEFAULT ) [, ...] } [, ...] } | query }
```

## 参数说明

- **table\_name**  
要插入数据的目标表名。  
取值范围：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。
- **column\_name**  
目标表中的字段名。
- **query**  
一个查询语句（SELECT语句），将查询结果作为插入的数据。

## 示例

创建表reason\_t1:

```
-- 创建表reason_t1。
CREATE TABLE reason_t1
(
    TABLE_SK      INTEGER          ,
    TABLE_ID      VARCHAR(20)      ,
    TABLE_NAME    VARCHAR(20)
)WITH(ORIENTATION=COLUMN, ENABLE_HSTORE=ON);
```

向表中插入一条记录:

```
INSERT INTO reason_t1(TABLE_SK, TABLE_ID, TABLE_NAME) VALUES (1, 'S01', 'StudentA');
```

向表中插入多条记录:

```
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_NAME
```

```
-----+-----+-----
1 | S01 | StudentA
2 | T01 | TeacherA
3 | T02 | TeacherB
(3 rows)
```

## 11.3.3 DELETE

### 功能描述

删除HStore表中的数据。

## 注意事项

- 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。
- HStore表上的单条Delete操作，会往Delta中插入一条type是D的记录，同时在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。
- 语法完全兼容列存，更多信息可以参考UPDATE语法。

## 语法格式

```
DELETE FROM [ ONLY ] table_name [ * ] [ [ AS ] alias ]  
[ USING using_list ]  
[ WHERE condition ]
```

## 参数说明

- **ONLY**  
如果指定ONLY则只有该表被删除；如果没有声明，则该表和它的所有子表将都被删除。
- **table\_name**  
目标表的名字（可以有模式修饰）。  
取值范围：已存在的表名。
- **alias**  
目标表的别名。  
取值范围：字符串，符合标识符命名规范。
- **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。

## 示例

创建表reason\_t2:

```
CREATE TABLE reason_t2  
(  
  TABLE_SK    INTEGER      ,  
  TABLE_ID    VARCHAR(20)  ,  
  TABLE_NA    VARCHAR(20)  
)WITH(ORIENTATION=COLUMN, ENABLE_HSTORE=ON);  
INSERT INTO reason_t2 VALUES (1, 'S01', 'StudentA'),(2, 'T01', 'TeacherA'),(3, 'T02', 'TeacherB');
```

使用WHERE 条件删除:

```
DELETE FROM reason_t2 WHERE TABLE_SK = 2;  
DELETE FROM reason_t2 AS rt2 WHERE rt2.TABLE_SK = 2;
```

使用IN语法删除:

```
DELETE FROM reason_t2 WHERE TABLE_SK in (1,3);
```

## 11.3.4 UPDATE

### 功能描述

更新HStore表上指定的数据。

### 注意事项

- 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 ]
SET {column_name = { expression | DEFAULT }
    | ( column_name [, ...] ) = {( { expression | DEFAULT } [, ...] ) |sub_query }}[, ...]
[ FROM from_list ] [ WHERE condition ];
```

### 参数说明

- **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，可以使用模式修饰。  
取值范围：已存在的表名称。
- **alias**  
目标表的别名。  
取值范围：字符串，符合标识符命名规范。
- **expression**  
赋给字段的值或表达式。
- **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。
- **from\_list**  
一个表的表达式列表，允许在WHERE条件里使用其他表的字段。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）。

- **condition**  
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。

## 示例

创建表reason\_update:

```
CREATE TABLE reason_update
(
  TABLE_SK    INTEGER
  , TABLE_ID  VARCHAR(20)
  , TABLE_NA  VARCHAR(20)
)WITH(ORIENTATION=COLUMN, ENABLE_HSTORE=ON);
```

向表reason\_update中插入数据:

```
INSERT INTO reason_update VALUES (1, 'S01', 'StudentA'),(2, 'T01', 'TeacherA'),(3, 'T02', 'TeacherB');
```

对表reason\_update执行UPDATE操作:

```
UPDATE reason_update SET TABLE_NA = 'TeacherD' where TABLE_SK = 3;
```

## 11.3.5 UPSERT

### 功能描述

HStore兼容UPSERT语法，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更新或者忽略冲突的数据。

### 注意事项

- 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。
- 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。
- 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。

### 语法格式

表 11-5 UPSERT 语法格式

语法格式	冲突更新	冲突忽略
第一种：不指定索引	INSERT INTO ON DUPLICATE KEY UPDATE	INSERT IGNORE INSERT INTO ON CONFLICT DO NOTHING
第二种：从指定列名或者约束上可以推断唯一约束	INSERT INTO ON CONFLICT(...) DO UPDATE SET INSERT INTO ON CONFLICT ON CONSTRAINT con_name DO UPDATE SET	INSERT INTO ON CONFLICT(...) DO NOTHING INSERT INTO ON CONFLICT ON CONSTRAINT con_name DO NOTHING

### 参数说明

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。

- 唯一索引推断  
对于第二种语法形式，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。
- 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。
- 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\_upsert并向表中插入数据：

```
CREATE TABLE reason_upsert
(
  a int primary key,
  b int,
  c int
)WITH(ORIENTATION=COLUMN, ENABLE_HSTORE=ON);
INSERT INTO reason_upsert VALUES (1, 2, 3);
```

忽略冲突的数据：

```
INSERT INTO reason_upsert VALUES (1, 4, 5),(2, 6, 7) ON CONFLICT(a) DO NOTHING;
```

更新冲突的数据：

```
INSERT INTO reason_upsert VALUES (1, 4, 5),(3, 8, 9) ON CONFLICT(a) DO UPDATE SET b = EXCLUDED.b,
c = EXCLUDED.c;
```

## 11.3.6 MERGE INTO

### 功能描述

通过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，无法匹配时对目标表执行INSERT。此语法可以很方便地用来合并执行UPDATE和INSERT，避免多次执行。

### 注意事项

对于并发MERGEINTO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 INTO table_name [ [ AS ] alias ]
USING { { table_name | view_name } | subquery } [ [ 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 ] }
];
```

## 参数说明

- **INTO子句**

指定正在更新或插入的目标表。

- **table\_name**  
目标表的表名。
- **alias**  
目标表的别名。  
取值范围：字符串，符合标识符命名规范。

- **USING子句**

指定源表，源表可以为表、视图或子查询。

- **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。

- **WHEN MATCHED子句**

当源表和目标表中数据针对关联条件可以匹配上时，选择WHEN MATCHED子句进行UPDATE操作。

 **说明**

不支持更新分布列、系统表以及系统列。

- **WHEN NOT MATCHED子句**

当源表和目标表中数据针对关联条件无法匹配时，选择WHEN NOT MATCHED子句进行INSERT操作。

 **说明**

- 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。

## 示例

创建进行MERGE INTO的目标：

```
CREATE TABLE target(a int, b int)WITH(ORIENTATION = COLUMN, ENABLE_HSTORE = ON);
INSERT INTO target VALUES(1, 1),(2, 2);
```

创建数据源表：

```
CREATE TABLE source(a int, b int)WITH(ORIENTATION = COLUMN, ENABLE_HSTORE = ON);
INSERT INTO source VALUES(1, 1),(2, 2),(3, 3),(4, 4),(5, 5);
```

执行MERGE INTO操作：

```
MERGE INTO target t
USING source s
ON (t.a = s.a)
WHEN MATCHED THEN
  UPDATE SET t.b = t.b + 1
```

```
WHEN NOT MATCHED THEN  
INSERT VALUES (s.a, s.b) WHERE s.b % 2 = 0;
```

## 11.3.7 SELECT

### 功能描述

从HStore表读取数据。

### 注意事项

- 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。
- 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

```
[ WITH [ RECURSIVE ] with_query [, ... ]  
SELECT [ /*+ plan_hint */ ] [ ALL | DISTINCT [ ON ( expression [, ... ] ) ] ]  
{ * | { expression [ [ AS ] output_name ] } [, ... ] }  
[ FROM from_item [, ... ] ]  
[ WHERE condition ]  
[ GROUP BY grouping_element [, ... ] ]  
[ HAVING condition [, ... ] ]  
[ { 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 } } ]
```

### 参数说明

- **DISTINCT [ ON ( expression [, ... ] ) ]**  
从SELECT的结果集中删除所有重复的行，使结果集中的每行都是唯一的。  
ON ( expression [, ... ] ) 只保留那些在给出的表达式上运算出相同结果的行集合中的第一行。
- **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。
- **FROM子句**  
为SELECT声明一个或者多个源表。  
FROM子句涉及的元素如下所示。
- **WHERE子句**  
WHERE子句构成一个行选择表达式，用来缩小SELECT查询的范围。condition是返回值为布尔型的任意表达式，任何不满足该条件的行都不会被检索。



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。同时，使用"(")有很多限制：

- **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。
- **HAVING子句**  
与GROUP BY子句配合用来选择特殊的组。HAVING子句将组的一些属性与一个常数值比较，只有满足HAVING子句中的逻辑表达式的组才会被提取出来。
- **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：

## 示例

创建表reason\_select并向表中插入数据：

```
CREATE TABLE reason_select
(
  r_reason_sk integer,
  r_reason_id integer,
  r_reason_desc character(100)
)WITH(ORIENTATION = COLUMN, ENABLE_HSTORE=ON);
INSERT INTO reason_select values(3, 1,'reason 1'),(10, 2,'reason 2'),(4, 3,'reason 3'),(10, 4,'reason 4');
```

执行GROUP BY分组操作：

```
SELECT COUNT(*), r_reason_sk FROM reason_select GROUP BY r_reason_sk;
```

执行HAVING过滤操作：

```
SELECT COUNT(*) c,r_reason_sk FROM reason_select GROUP BY r_reason_sk HAVING c > 1;
```

执行ORDER BY操作：

```
SELECT * FROM reason_select ORDER BY r_reason_sk;
```

## 11.3.8 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、添加/更新多个列、将列存改为HStore表等。

### 注意事项

- 通过ALTER修改enable\_hstore值可以将列存表变成HStore表，或者将HStore修改成列存表。但需要注意enable\_delta为on时，无法设置enable\_hstore为on。
- 对于部分ALTER操作（修改列类型，分区合并，添加NOT NULL约束，添加主键约束），HStore表需要先将数据MERGE到主表然后再执行原有的ALTER逻辑，这可能会带来额外的性能开销，性能影响大小与delta表的数据量相关。
- 当需要增加列时，建议不要与其它类型的ALTER（比如修改列类型等）组合使用，在一条ALTER里只有ADD COLUMN情况下，由于不需要做FULL MERGE，性能会有很大提升。
- 不支持修改存储参数ORIENTATION。

### 修改表属性

使用语法：

```
ALTER TABLE [ IF EXISTS ] <table_name> SET ( {ENABLE_HSTORE = ON} [, ... ] );
```

将列存表修改成HStore表：

```
CREATE TABLE alter_test(a int, b int) WITH(ORIENTATION = COLUMN);  
ALTER TABLE alter_test SET (ENABLE_HSTORE = ON);
```

### 须知

当需要使用HStore表时，一定要同步修改如下参数的默认值，否则会导致HStore性能严重劣化，推荐的默认配置是

autovacuum\_max\_workers\_hstore=3, autovacuum\_max\_workers=6,  
autovacuum=true。

## 增加列

使用语法：

```
ALTER TABLE [ IF EXISTS ] <table_name> ADD COLUMN <new_column> <data_type> [ DEFAULT  
<default_value>];
```

示例：

创建表alter\_test2并对其增加列：

```
CREATE TABLE alter_test2(a int, b int) WITH(ORIENTATION = COLUMN,ENABLE_HSTORE = ON);  
ALTER TABLE alter_test ADD COLUMN c int;
```

### 说明

增加列时不建议在同一个SQL中与其它ALTER组合使用。

## 重命名

使用语法：

```
ALTER TABLE [ IF EXISTS ] <table_name> RENAME TO <new_table_name>;
```

示例：

创建表alter\_test3并对其重命名为alter\_new：

```
CREATE TABLE alter_test3(a int, b int) WITH(ORIENTATION = COLUMN,ENABLE_HSTORE = ON);  
ALTER TABLE alter_test3 RENAME TO alter_new;
```

## 11.4 实时数仓函数

### hstore\_light\_merge(rel\_name text)

描述：该函数用于手动对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)

描述：该函数用于手动对HStore表进行全量清理操作。

返回值类型：int

### 须知

- 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。

示例：

```
SELECT hstore_full_merge('reason_select');
```

## 11.5 实时数仓 GUC 参数

### autovacuum

**参数说明：**控制是否启动数据库自动清理进程（autovacuum）。

**参数类型：**SIGHUP

**取值范围：**布尔型

- on表示开启数据库自动清理进程。
- off表示关闭数据库自动清理进程。

**默认值：**on

### autovacuum\_max\_workers

**参数说明：**设置能同时运行的自动清理线程的最大数量，该参数的取值上限与max\_connections和job\_queue\_processes大小有关。

**参数类型：**SIGHUP

**取值范围：**整型

- 最小值为0，表示不会自动进行autovacuum。
- 理论最大值为262143，实际最大值为动态值。计算公式为“262143 - max\_inner\_tool\_connections - max\_connections - job\_queue\_processes - 辅助线程数 - autovacuum的launcher线程数 - 1”，其中辅助线程数和autovacuum的launcher线程数由两个宏来指定，当前版本的默认值分别为20和2。

**默认值：**3

### autovacuum\_max\_workers\_hstore

**参数说明：**设置Autovacuum\_max\_workers里面，能同时运行的专用于清理hstore的自动清理线程的最大数量。

**参数类型：** SIGHUP

**取值范围：** 整型

**默认值：** 0

#### 说明

当需要使用hstore表时，一定要同步修改一下几个参数的默认值，否则会导致hstore表性能严重劣化，推荐的修改配置是：

autovacuum\_max\_workers\_hstore=3, autovacuum\_max\_workers=6, autovacuum=true。

## enable\_hstore\_lightupdate

**参数说明：** 用于控制是否开启对hstore表上的轻量化UPDATE（对Hstore表执行UPDATE时会自动判断是否需要轻量化UPDATE）。

**参数类型：** SIGHUP

**取值范围：** 布尔型

- on表示开启对hstore表上的轻量化UPDATE。
- off表示关闭对hstore表上的轻量化UPDATE。

**默认值：** off

## enable\_hstore\_merge\_keepgtm

**参数说明：** 用于控制列存、hstore表上的autovacuum中的merge是否在GTM(全局事务管理)占槽位。

**参数类型：** SIGHUP

**取值范围：** 布尔型

- true表示在GTM（全局事务管理）占槽位。
- false表示在GTM（全局事务管理）不占槽位。

**默认值：** true

## hstore\_buffer\_size

**参数说明：** 用于控制HStore的CU槽位数量，该槽位用于存储每个CU的更新链，能显著提升更新与查询效率。

为避免占用内存过大，系统会根据内存大小计算出一个槽位值，再与该参数相比取最小的值。

**参数类型：** POSTMASTER

**取值范围：** 整型，100~100000

**默认值：** true

# 12 资源监控

GaussDB(DWS)为用户提供了多维度的资源监控视图。可以查看作业的实时资源记录和历史资源记录。

## 12.1 用户资源监控

### 背景信息

在多租户管理的框架下，用户可以实时查询所有用户资源（包括内存、CPU核数、存储空间、临时空间、算子落盘空间和IO）实时使用情况，也可以查询用户资源的历史使用情况。

#### 说明

- 用户实时资源相关视图/函数为：[PG\\_TOTAL\\_USER\\_RESOURCE\\_INFO](#)、[GS\\_WLM\\_USER\\_RESOURCE\\_INFO](#)和[PGXC\\_TOTAL\\_USER\\_RESOURCE\\_INFO](#)；用户历史资源相关表为：[GS\\_WLM\\_USER\\_RESOURCE\\_HISTORY](#)和[PGXC\\_WLM\\_USER\\_RESOURCE\\_HISTORY](#)。
- 用户监控可以同时监控快慢车道所有作业的CPU、IO和内存使用情况，不再受限于仅监控慢车道作业；
- 当前快车道作业内存和CPU不受控，在快车道运行作业占用资源较多情况下，可能出现已用资源大于资源限制的情况；
- DN监控视图中，IO、内存和CPU显示的是本DN上资源池资源使用和资源限制信息；
- CN监控视图中，IO、内存和CPU显示的是集群内所有DN资源池资源使用和资源限制的累积和；
- DN每隔5s更新一次监控信息，CN每隔5s从DN收集一次用户监控信息，因为各实例单独更新/收集用户监控信息，因此各实例监控信息更新时间可能不一致；
- 辅助线程中每隔30s自动调用持久化函数，持久化用户监控数据，正常情况下不需要用户单独调用持久化函数持久化用户监控数据；
- 当用户数量较多，集群规模较大时，查询此类实时视图，因CN/DN间实时通信开销，会有一些的网络延时；
- 初始管理用户不进行资源监控。

### 操作步骤

- 查询所有用户的资源限额和资源实时使用情况。

```
SELECT * FROM PG_TOTAL_USER_RESOURCE_INFO;
```

得到的结果视图如下:

username	used_memory	total_memory	used_cpu	total_cpu	used_space	total_space	used_temp_space	total_temp_space	used_spill_space	total_spill_space	read_kbytes	write_kbytes	read_counts	write_counts	read_speed	write_speed
perfdm	0	0	0	0	0	-1	0	0	-1	0	0	0	-1	0		
usern	0	17250	0	48	0	-1	0	0	-1	0	0	0	-1	0		
userg	814955731	15525	23.53	48	0	-1	0	0	-1	0	143233	42678				
userg1	814972419	13972	23.53	48	0	-1	0	0	-1	0	143233	42710				

(4 rows)

其中, IO资源监控字段(read\_kbytes、write\_kbytes、read\_counts、write\_counts、read\_speed和write\_speed)需要在GUC参数enable\_user\_metric\_persistent开启时才有监控数据。

所查各字段说明详见[PG\\_TOTAL\\_USER\\_RESOURCE\\_INFO](#)。

- 查询具体某个用户的资源限额和资源实时使用情况。

```
SELECT * FROM GS_WLM_USER_RESOURCE_INFO('username');
```

查询结果如下:

userid	used_memory	total_memory	used_cpu	total_cpu	used_space	total_space	used_temp_space	total_temp_space	used_spill_space	total_spill_space	read_kbytes	write_kbytes	read_counts	write_counts	read_speed	write_speed
16407	18	1655	6	19	13787176	-1	0	0	-1	0	0	0	-1	0		

(1 row)

- 查询所有用户的资源限额和资源历史使用情况。

```
SELECT * FROM GS_WLM_USER_RESOURCE_HISTORY;
```

查询结果如下:

username	timestamp	used_memory	total_memory	used_cpu	total_cpu	used_space	total_space	used_temp_space	total_temp_space	used_spill_space	total_spill_space	read_kbytes	write_kbytes	read_counts	write_counts	read_speed	write_speed
usern	2020-01-08 22:56:06.456855+08	0	17250	0	48	0	0	0	0	0	0	0	0	0	0	0	0
userg	2020-01-08 22:56:06.458659+08	814955731	15525	23.53	48	0	0	0	0	0	143233	42678					
userg1	2020-01-08 22:56:06.460252+08	814972419	13972	23.53	48	0	0	0	0	0	143233	42710					

对于系统表[GS\\_WLM\\_USER\\_RESOURCE\\_HISTORY](#), 仅当GUC参数enable\_user\_metric\_persistent开启时, 才会定期将视图[PG\\_TOTAL\\_USER\\_RESOURCE\\_INFO](#)中的数据保存到历史表中。

所查各字段说明详见[GS\\_WLM\\_USER\\_RESOURCE\\_HISTORY](#)。

## 12.2 资源池资源监控

### 简介

多租户管理框架下，用户关联资源池执行查询，用户执行查询占用的资源将汇总至关联资源池上，通过资源池监控视图用户可以直观的查询到所有资源池的实时资源使用情况，同时也可以通过资源池监控历史表查询资源池资源的历史使用情况。

资源池监控数据每5s更新一次，但是因为CN和DN时间差，实际监控数据更新时间可能会大于5s，正常不会超过10s。资源池监控数据每30s持久化一次，资源池监控和用户监控逻辑基本一致，因此共用GUC参数控制持久化和老化，使用GUC参数 `enable_user_metric_persistent` 控制是否进行资源池监控数据持久化，使用GUC参数 `user_metric_retention_time` 控制资源池监控数据老化。

资源池监控的资源包含：快慢车道作业运行和排队信息，CPU、内存以及逻辑IO资源监控信息。涉及的监控视图和历史表如下所示：

1. 资源池实时运行信息监控视图(单CN)：[GS\\_RESPPOOL\\_RUNTIME\\_INFO](#)；
2. 资源池实时运行信息监控视图(所有CN)：[PGXC\\_RESPPOOL\\_RUNTIME\\_INFO](#)；
3. 资源池实时资源监控视图(单CN)：[GS\\_RESPPOOL\\_RESOURCE\\_INFO](#)；
4. 资源池实时资源监控视图(所有CN)：[PGXC\\_RESPPOOL\\_RESOURCE\\_INFO](#)；
5. 资源池历史资源监控表(单CN)：[GS\\_RESPPOOL\\_RESOURCE\\_HISTORY](#)；
6. 资源池历史资源监控视图(所有CN)：[PGXC\\_RESPPOOL\\_RESOURCE\\_HISTORY](#)；

#### 说明

- 资源池监控可以同时监控快慢车道所有作业的CPU、IO和内存使用情况，不再受限于仅监控慢车道作业；
- 当前快车道作业内存和CPU不受控，在快车道运行作业占用资源较多情况下，可能出现已用资源大于资源限制的情况；
- DN资源池监控视图中，IO、内存和CPU显示的是本DN上资源池资源使用和资源限制信息；
- CN资源池监控视图中，IO、内存和CPU显示的是集群内所有DN资源池资源使用和资源限制的累积和；
- DN每隔5s更新一次资源池监控信息，CN每隔5s从DN收集一次资源池监控信息，因为各实例单独更新/收集资源池监控信息，因此各实例监控信息更新时间可能不一致；
- 辅助线程中每隔30s自动调用持久化函数，持久化资源池监控数据，正常情况下不需要用户单独调用持久化函数持久化资源池监控数据。

### 操作步骤

- 查询资源池的作业实时运行情况。  
`SELECT * FROM GS_RESPPOOL_RUNTIME_INFO;`

得到的结果视图如下：

nodegroup	rpname	ref_count	fast_run	fast_wait	slow_run	slow_wait
vc1	p2	10	0	0	0	0
vc2	p3	10	5	5	0	0
vc2	p4	0	0	0	0	0
vc1	default_pool	0	0	0	0	0
vc2	default_pool	0	0	0	0	0
vc1	p1	20	5	5	3	7

(6 rows)

其中:

- ref\_count为引用当前资源池信息的作业数，语句从进入管控到结束一直占用该计数；
  - fast\_run和slow\_run为负载管理记账信息，只有管控(fast\_limit/slow\_limit大于0)时该值才有效；
  - 该视图仅在CN上有效，持久化信息保存在GS\_RESPOOL\_RESOURCE\_HISTORY中；
  - 各字段说明详见[GS\\_RESPOOL\\_RUNTIME\\_INFO](#)。
- 查询资源池的资源限额和资源实时使用情况。  
SELECT \* FROM GS\_RESPOOL\_RESOURCE\_INFO;

得到的结果视图如下:

nodegroup	rpname	cgroup	ref_count	fast_run	fast_wait	fast_limit	slow_run	slow_wait	slow_limit	used_cpu	cpu_limit	used_mem	estimate_mem	mem_limit	read_kbytes	write_kbytes	read_counts	write_counts	read_speed	write_speed
vc1	p2	DefaultClass:Rush	10	0	0	-1	0	0	0	9.97	48	20	0	11555	8	2880	1	360	1	589
vc2	p3	DefaultClass:Rush	10	5	5	5	0	0	0	4.98	48	11	0	11555	0	848	0	106	0	173
vc2	p4	DefaultClass:Rush	0	0	0	-1	0	0	0	0	48	0	0	11555	0	0	0	0	0	0
vc1	default_pool	DefaultClass:Medium	0	0	0	-1	0	0	0	-1	0	48	0	0	11555	0	0	0	0	0
vc2	default_pool	DefaultClass:Medium	0	0	0	-1	0	0	0	-1	0	48	0	0	11555	0	0	0	0	0
vc1	p1	DefaultClass:Rush	20	5	5	5	3	7	3	7.98	48	16	768	11555	8	2656	1	332	1	543

(6 rows)

- 该视图在CN和DN上均有效，DN上CPU、内存和IO为本DN资源消耗情况，CN上CPU、内存和IO为集群内所有DN上资源消耗的累加和；
  - estimate\_mem仅在动态负载管理情况下CN上有效，显示资源池估算内存记账情况；
  - IO监控信息仅在enable\_logical\_io\_statistics开启时才会记录；
  - 各字段说明详见[GS\\_RESPOOL\\_RESOURCE\\_INFO](#)。
- 查询资源池的资源限额和资源历史使用情况。  
SELECT \* FROM GS\_RESPOOL\_RESOURCE\_HISTORY ORDER BY timestamp DESC;

得到的结果视图如下:

timestamp	nodegroup	rpname	cgroup	ref_count	fast_run	fast_wait	fast_limit	slow_run	slow_wait	slow_limit	used_cpu	cpu_limit	used_mem	estimate_mem	mem_limit	read_kbytes	write_kbytes	read_counts	write_counts	read_speed	write_speed
2022-03-04 09:41:57.53739+08	vc1	p2	DefaultClass:Rush	10	0	0	-1	0	0	9.97	48	20	0	11555	8	2880	1	360	1	589	
2022-03-04 09:41:57.53739+08	vc1	p1	DefaultClass:Rush	20	5	5	5	3	7	7.98	48	16	768	11555	8	2656	1	332	1	543	
2022-03-04 09:41:57.53739+08	vc2	default_pool	DefaultClass:Medium	0	0	0	-1	0	0	-1	0	48	0	0	11555	0	0	0	0	0	



0	-1	0	0	-1	0	48	0	0	11555	0
0	0	0	0	0	0					
2022-03-04 09:41:57.53739+08	vc1		default_pool	DefaultClass:Medium		0	0			
0	-1	0	0	-1	0	48	0	0	11555	0
0	0	0	0	0	0					
2022-03-04 09:41:57.53739+08	vc2		p4	DefaultClass:Rush		0	0	0		
-1	0	0	10	0	48	0	0	11555	0	0
0	0	0	0	0	0					
2022-03-04 09:41:57.53739+08	vc2		p3	DefaultClass:Rush		10	5	5		
5	0	0	10	4.99	48	11	0	11555	0	880
0	110	0	180							
2022-03-04 09:41:27.335234+08	vc2		p3	DefaultClass:Rush		10	5	5		
5	0	0	10	4.98	48	11	0	11555	0	856
0	107	0	175							

- 该监控信息来自资源池监控历史表，enable\_user\_metric\_persistent开启时每30秒记录一次；
- 该表数据保存时间由GUC参数user\_metric\_retention\_time控制；
- 各字段说明详见[GS\\_RESPOOL\\_RESOURCE\\_HISTORY](#)。

## 12.3 内存资源监控

### 内存监控

GaussDB(DWS)提供了监控整个集群内存使用状态的视图：

查询pgxc\_total\_memory\_detail视图，必须具有sysadmin权限。

```
SELECT * FROM pgxc_total_memory_detail;
```

如果查询该视图时出现以下错误，请开启内存管理功能。

```
SELECT * FROM pgxc_total_memory_detail;
```

```
ERROR: unsupported view for memory protection feature is disabled.
```

```
CONTEXT: PL/pgSQL function pgxc_total_memory_detail() line 12 at FOR over EXECUTE statement
```

开启内存管理功能用户可通过GaussDB(DWS) 控制台设置enable\_memory\_limit和max\_process\_memory参数，方法如下：

- 登录GaussDB(DWS) 管理控制台。
- 在左侧导航栏中，单击“集群管理”。
- 在集群列表中找到所需要的集群，单击集群名称，进入集群“基本信息”页面。
- 单击“参数修改”页签，修改参数“enable\_memory\_limit”的值为on，然后单击“保存”。
- 修改参数“max\_process\_memory”的值为合适的值，修改建议请参见[max\\_process\\_memory](#)，然后单击“保存”。
- 在“修改预览”窗口，确认修改无误后，单击“保存”。修改完成后需要重启集群，参数才会生效。

### 共享内存监控

用户可以通过视图pg\_shared\_memory\_detail查询共享内存上下文信息：

```
SELECT * FROM pg_shared_memory_detail;
```

contextname	level	parent	totalsize	freesize	usedsize
ProcessMemory	0		24576	9840	14736
Workload manager memory context	1	ProcessMemory		2105400	7304
wlm collector hash table	2	Workload manager memory context	8192	3736	4456
Resource pool hash table	2	Workload manager memory context	24576	15968	8608

```
wlm cgroup hash table | 2 | Workload manager memory context | 24576 | 15968 | 8608  
(5 rows)
```

该视图列举了内存使用的上下文名称、级别、上级内存上下文、共享内存总大小等。

另外，在数据库中，GUC参数“memory\_tracking\_mode”用来设置内存信息统计的模式，共支持四种模式：

- none，不启动内存统计功能。
- normal，仅做内存实时统计，不生成文件。
- executor，生成统计文件，包含执行层使用过的所有已分配内存的上下文信息。

当为executor模式时，将在DN进程的pg\_log目录下生成csv格式文件，命名方式为：memory\_track\_<DN名称>\_query\_<queryid>.csv。作业执行时，执行器postgres线程和所有stream线程执行的算子信息，都将输入该文件。

文件内容根据以下面内容组成实例如下：

```
0, 0, ExecutorState, 0, PortalHeapMemory, 0, 40K, 602K, 23  
1, 3, CStoreScan_29360131_25, 0, ExecutorState, 1, 265K, 554K, 23  
2, 128, cstore scan per scan memory context, 1, CStoreScan_29360131_25, 2, 24K, 24K, 23  
3, 127, cstore scan memory context, 1, CStoreScan_29360131_25, 2, 264K, 264K, 23  
4, 7, InitPartitionMapTmpMemoryContext, 1, CStoreScan_29360131_25, 2, 31K, 31K, 23  
5, 2, VecPartIterator_29360131_24, 0, ExecutorState, 1, 16K, 16K, 23  
0, 0, ExecutorState, 0, PortalHeapMemory, 0, 24K, 1163K, 20  
1, 3, CStoreScan_29360131_22, 0, ExecutorState, 1, 390K, 1122K, 20  
2, 20, cstore scan per scan memory context, 1, CStoreScan_29360131_22, 2, 476K, 476K, 20  
3, 19, cstore scan memory context, 1, CStoreScan_29360131_22, 2, 264K, 264K, 20  
4, 7, InitPartitionMapTmpMemoryContext, 1, CStoreScan_29360131_22, 2, 23K, 23K, 20  
5, 2, VecPartIterator_29360131_21, 0, ExecutorState, 1, 16K, 16K, 20
```

其中各字段分别为：输出序号、线程内分配内存上下文的序号、当前内存上下文的名称、父内存上下文的输出序号、父内存上下文的名称、内存上下文树形层次级别号、当前内存上下文使用的内存峰值、当前内存上下文及其所有子内存上下文使用的内存峰值、当前线程所在query的plannodeid。

在本例中，记录“1, 3, CStoreScan\_29360131\_22, 0, ExecutorState, 1, 390K, 1122K, 20”和Explain Analyze的对应关系如下：

- CstoreScan\_29360131\_22代表CstoreScan算子。
- 1122K代表CstoreScan算子的PeakMemory。

- fullexec，生成文件包含执行层申请过的所有内存上下文信息。

当设置为fullexec模式时，输出信息和executor模式相同，但可能增加部分内存上下文分配信息，因为所有申请内存（无论是否申请成功）相关的信息都会被打印出来。由于仅记录内存申请信息，故记录中内存上下文使用的峰值均为0。

## 12.4 实例资源监控

GaussDB(DWS)提供了监控CN、DN实例资源使用状态（包括内存，CPU，磁盘IO，进程物理IO和进程逻辑IO）的系统表及监控整个集群资源使用状态的系统表。

关于系统表GS\_WLM\_INSTANCE\_HISTORY的详细介绍，请参考[GS\\_WLM\\_INSTANCE\\_HISTORY](#)。

### 📖 说明

系统表GS\_WLM\_INSTANCE\_HISTORY中的数据分布在对应的实例中，CN实例监控数据保存在CN实例中，DN实例监控数据保存在DN实例中；DN实例由于有备机，当主DN实例异常时，该DN实例的监控数据能够从备机恢复；但CN实例无备机，当某CN实例异常再恢复时，该CN实例的监控数据会丢失。

## 操作步骤

- 查询当前实例最近的资源使用情况。

```
SELECT * FROM GS_WLM_INSTANCE_HISTORY ORDER BY TIMESTAMP DESC;
```

查询结果如下:

instancename	timestamp	used_cpu	free_mem	used_mem	io_await	io_util	disk_read	disk_write	process_read	process_write	logical_read	logical_write	read_counts	write_counts
dn_6015_6016	2022-01-10 17:29:17.329495+08	0	14570	8982	662.923	99.9601	697666	93655.5	183104	30082	285659	30079	357717	37667
dn_6015_6016	2022-01-10 17:29:07.312049+08	0	14578	8974	883.102	99.9801	756228	81417.4	189722	30786	285681	30780	358103	38584
dn_6015_6016	2022-01-10 17:28:57.284472+08	0	14583	8969	727.135	99.9801	648581	88799.6	177120	31176	252161	31175	316085	39079
dn_6015_6016	2022-01-10 17:28:47.256613+08	0	14591	8961	679.534	100.08	655360	169962	179404	30424	242002	30422	303351	38136

- 查询当前实例某一段时间内的资源使用情况。

```
SELECT * FROM GS_WLM_INSTANCE_HISTORY WHERE TIMESTAMP > '2022-01-10' AND TIMESTAMP < '2020-01-11' ORDER BY TIMESTAMP DESC;
```

查询结果如下:

instancename	timestamp	used_cpu	free_mem	used_mem	io_await	io_util	disk_read	disk_write	process_read	process_write	logical_read	logical_write	read_counts	write_counts
dn_6015_6016	2022-01-10 17:29:17.329495+08	0	14570	8982	662.923	99.9601	697666	93655.5	183104	30082	285659	30079	357717	37667
dn_6015_6016	2022-01-10 17:29:07.312049+08	0	14578	8974	883.102	99.9801	756228	81417.4	189722	30786	285681	30780	358103	38584
dn_6015_6016	2022-01-10 17:28:57.284472+08	0	14583	8969	727.135	99.9801	648581	88799.6	177120	31176	252161	31175	316085	39079
dn_6015_6016	2022-01-10 17:28:47.256613+08	0	14591	8961	679.534	100.08	655360	169962	179404	30424	242002	30422	303351	38136

- 查询集群最近的资源使用情况，可以在CN节点上调用 `pgxc_get_wlm_current_instance_info` 存储过程函数。

```
SELECT * FROM pgxc_get_wlm_current_instance_info('ALL');
```

查询结果如下:

instancename	timestamp	used_cpu	free_mem	used_mem	io_await	io_util	disk_read	disk_write	process_read	process_write	logical_read	logical_write	read_counts	write_counts
coordinator2	2020-01-14 21:58:29.290894+08	0	12010	278	16.0445	7.19561	184.431	27959.3	0	10	0	0	0	0
coordinator3	2020-01-14 21:58:27.567655+08	0	12000	288	.964557	3.40659	332.468	3375.02	26	13	0	0	0	0
datanode1	2020-01-14 21:58:23.900321+08	0	11899	389	1.17296	3.25	329.6	2870.4	28	8	13	3	18	6
datanode2	2020-01-14 21:58:32.832989+08	0	11904	384	17.948	8.52148	214.186	25894.1	28	10	13	3	18	6
datanode3	2020-01-14 21:58:24.826694+08	0	11894	394	1.16088	3.15	2868.8	25	10	13	3	18	6	328
coordinator1	2020-01-14 21:58:33.367649+08	0	11988	300	9.53286	10.05	43.2	55232	0	0	0	0	0	0
coordinator1	2020-01-14 21:58:23.216645+08	0	11988	300	1.17085	3.21182	324.729	2831.13	8	13	0	0	0	0

(7 rows)

- 查询集群历史的资源使用情况，可以在CN节点上调用 `pgxc_get_wlm_history_instance_info` 存储过程函数。

```
SELECT * FROM pgxc_get_wlm_history_instance_info('ALL', '2020-01-14 21:00:00', '2020-01-14 22:00:00', 3);
```

查询结果如下:

instancename	timestamp	used_cpu	free_mem	used_mem	io_await	io_util	disk_read	disk_write	process_read	process_write	logical_read	logical_write	read_counts	write_counts
coordinator2	2020-01-14 21:50:49.778902+08	0	12020	268	.127371	.789211	15.984	3994.41	0	0	0	0	0	0
coordinator2	2020-01-14 21:53:49.043646+08	0	12018	270	30.2902	8.65404	276.77	16741.8	3	1	0	0	0	0
coordinator2	2020-01-14 21:57:09.202654+08	0	12018	270	.16051	.979021	59.9401	5596	0	0	0	0	0	0
coordinator3	2020-01-14 21:38:48.948646+08	0	12012	276	.0769231	.00999001	0	35.1648	0	1	0	0	0	0
coordinator3	2020-01-14 21:40:29.061178+08	0	12012	276	.118421	.0199601	0	970.858	0	0	0	0	0	0
coordinator3	2020-01-14 21:50:19.612777+08	0	12010	278	24.411	11.7665	8.78244	44641.1	0	0	0	0	0	0
datanode1	2020-01-14 21:49:42.758649+08	0	11909	379	.798776	8.02	51.2	20924.8	0	0	0	0	0	0
datanode1	2020-01-14 21:49:52.760188+08	0	11909	379	23.8972	14.1	0	74760	0	0	0	0	0	0
datanode1	2020-01-14 21:50:22.769226+08	0	11909	379	39.5868	7.4	19760.8	0	0	0	0	0	0	0
datanode2	2020-01-14 21:58:02.826185+08	0	11905	383	.351648	.32	20.8	504.8	0	0	0	0	0	0
datanode2	2020-01-14 21:56:42.80793+08	0	11906	382	.559748	.04	326.4	0	0	0	0	0	0	0
datanode2	2020-01-14 21:45:21.632407+08	0	11901	387	12.1313	4.55544	3.1968	45177.2	0	0	0	0	0	0
datanode3	2020-01-14 21:58:14.823317+08	0	11898	390	.378205	.99	48	23353.6	0	0	0	0	0	0
datanode3	2020-01-14 21:47:50.665028+08	0	11901	387	1.07494	1.19	0	15506.4	0	0	0	0	0	0
datanode3	2020-01-14 21:51:21.720117+08	0	11903	385	10.2795	3.11	0	11031.2	0	0	0	0	0	0
coordinator1	2020-01-14 21:42:59.121945+08	0	12020	268	.0857143	.0699301	0	6579.02	0	0	0	0	0	0
coordinator1	2020-01-14 21:41:49.042646+08	0	12020	268	20.9039	11.3786	6042.76	57903.7	0	0	0	0	0	0
coordinator1	2020-01-14 21:41:09.007652+08	0	12020	268	.0446429	.03996	0	1109.29	0	0	0	0	0	0

(18 rows)

## 12.5 实时 TopSQL

系统提供了不同级别的资源监控实时视图用来查询实时TopSQL。资源监控实时视图记录了查询作业运行时的资源使用情况(包括内存、下盘、CPU时间等)以及性能告警信息。

实时视图具体的对外接口如下表所示:

表 12-1 资源监控实时视图

视图级别	节点范围	查询视图
query级别/perf级别	当前CN	GS_WLM_SESSION_STATISTICS
	所有CN	PGXC_WLM_SESSION_STATISTICS
operator级别	当前CN	GS_WLM_OPERATOR_STATISTICS
	所有CN	PGXC_WLM_OPERATOR_STATISTICS

## 📖 说明

- 视图级别取决于资源监控的等级，即参数`resource_track_level`的配置。
- `perf`和`operator`级别会影响`GS_WLM_SESSION_STATISTICS/PGXC_WLM_SESSION_INFO`中的`query_plan`和`warning`字段的取值，详细内容参见[SQL自诊断](#)。
- 对外接口通过不同的前缀(`gs`与`pgxc`)来区分单CN查询视图以及集群级别查询视图。普通用户仅支持登录到集群的某个CN查询以`gs`为前缀的视图。
- 查询此类实时视图时，因需要获取作业运行实时资源使用情况，会有一些的网络延时。
- 实例故障时，实时TopSQL视图有可能记录不全。
- 实时TopSQL中能够记录的SQL语句的规格是：
  - 不记录特殊数据定义语句，如：`SET`、`RESET`、`SHOW`、`ALTER SESSION SET`、`SET CONSTRAINTS`语句；
  - 记录数据定义语句，例如：执行`CREATE`、`ALTER`、`DROP`、`GRANT`、`REVOKE`和`VACUUM`语句；
  - 记录数据操作语句，例如：
    - 执行`SELECT`、`INSERT`、`UPDATE`和`DELETE`语句。
    - 执行`explain analyze`和`explain performance`场景。
    - 使用`query`级别/`perf`级别视图。
  - 记录函数与存储过程的调用入口语句，当GUC参数`enable_track_record_subsql`开启的情况下，可记录存储过程的部分内部语句(`declare`定义语句除外)，仅会记录其中下发到DN执行的内部语句，未下发到DN执行的内部语句会被过滤掉；
  - 记录匿名块语句，当GUC参数`enable_track_record_subsql`开启的情况下，可记录匿名块中的部分内部语句，仅会记录其中下发到DN执行的内部语句，未下发到DN执行的内部语句会被过滤掉；
  - 记录游标语句，当游标并非从缓存中读取数据，而确实触发语句下发到DN上执行的条件下，该游标语句会被记录，并且会进行语句、执行计划增强，但当游标从缓存中读取数据时，不进行记录；当游标语句在匿名块或者函数中使用，当游标从DN上读取较多数据但不完全使用时，因当前架构限制，无法记录该游标在DN上的监控信息。对于`With Hold`游标，该语法执行逻辑特殊，会在事务提交阶段执行实际查询动作，当语句在该阶段执行报错时，作业的`aborted`状态无法反馈到TopSQL历史表中。
  - 重分布过程中的作业不统计；
  - JDBC执行的带占位符语句，通常会补齐参数内容，但如果参数和原语句合起来长度超过64KB，则不记录参数，或者如果是轻量化语句，直接下发到DN上执行，不记录参数。

## 前提条件

- GUC参数`enable_resource_track`为`on`（默认为`on`）。
- GUC参数`resource_track_level`为`query`、`perf`或`operator`（默认为`query`）。
- 监控作业的类型为：
  - 优化器估算的执行代价大于或等于`resource_track_cost`取值的作业。
- Cgroups功能正常加载，可通过`gs_cgroup -P`查看控制组信息。
- GUC参数`enable_track_record_subsql`控制是否记录存储过程、匿名块内部语句。

在上述条件中，`enable_resource_track`为系统级参数，用于设置是否开启资源监控功能。`resource_track_level`为session级参数，可以对某个session的资源监控级别进行灵活设置。这两个参数的设置方法如下表：

表 12-2 设置资源监控信息统计级别

enable_resource_track	resource_track_level	query级别信息	算子级别信息
on(default)	none	不统计	不统计
	query(default)	统计	不统计
	perf	统计	不统计
	operator	统计	统计
off	none/query/operator	不统计	不统计

## 操作步骤

**步骤1** 通过视图gs\_session\_cpu\_statistics查询实时CPU信息。

```
SELECT * FROM gs_session_cpu_statistics;
```

**步骤2** 通过视图gs\_session\_memory\_statistics查询实时memory信息。

```
SELECT * FROM gs_session_memory_statistics;
```

**步骤3** 通过视图gs\_wlm\_session\_statistics查询当前CN的实时资源。

```
SELECT * FROM gs_wlm_session_statistics;
```

**步骤4** 通过视图pgxc\_wlm\_session\_statistics查询所有CN的实时资源。

```
SELECT * FROM pgxc_wlm_session_statistics;
```

**步骤5** 通过视图gs\_wlm\_operator\_statistics查询当前CN作业算子执行实时资源信息。

```
SELECT * FROM gs_wlm_operator_statistics;
```

**步骤6** 通过视图pgxc\_wlm\_operator\_statistics查询所有CN作业算子执行实时资源信息。

```
SELECT * FROM pgxc_wlm_operator_statistics;
```

**步骤7** 通过视图pg\_session\_wlmstat查询当前用户执行作业正在运行时的负载管理信息。

```
SELECT * FROM pg_session_wlmstat;
```

**步骤8** 通过视图pgxc\_wlm\_workload\_records（动态负载功能开启，即enable\_dynamic\_workload为on时该视图有效）查询当前用户在每个CN上作业执行时的状态信息。

```
SELECT * FROM pgxc_wlm_workload_records;
```

----结束

## 12.6 历史 TopSQL

系统提供了资源监控历史视图用例查询历史TopSQL。资源监控历史视图记录了查询作业运行结束时的资源使用情况(包括内存、下盘、CPU时间等)和运行状态信息(包括报错、终止、异常等)以及性能告警信息。但对于由于FATAL、PANIC错误导致查询异常结束时，状态信息列只显示aborted，无法记录详细异常信息。特别的，对于查询解析，优化阶段的状态信息则无法监控。

历史视图具体的对外接口如下表所示：

视图级别	节点范围	查询视图	
query级别/perf级别	当前CN	历史（ Database Manager 接口）	<a href="#">GS_WLM_SESSION_HISTORY</a>
		历史（内部转储接口）	<a href="#">GS_WLM_SESSION_INFO</a>
	所有CN	历史（ Database Manager 接口）	<a href="#">PGXC_WLM_SESSION_HISTORY</a>
		历史（内部转储接口）	<a href="#">PGXC_WLM_SESSION_INFO</a>
operator级别	当前CN	历史（ Database Manager 接口）	<a href="#">GS_WLM_OPERATOR_HISTORY</a>
		历史（内部转储接口）	<a href="#">GS_WLM_OPERATOR_INFO</a>
	所有CN	历史（ Database Manager 接口）	<a href="#">PGXC_WLM_OPERATOR_HISTORY</a>
		历史（内部转储接口）	<a href="#">PGXC_WLM_OPERATOR_INFO</a>

### 📖 说明

- 视图级别取决于资源监控的等级，即参数[resource\\_track\\_level](#)的配置。
- perf和operator级别会影响[GS\\_WLM\\_SESSION\\_STATISTICS/PGXC\\_WLM\\_SESSION\\_INFO](#)中的query\_plan和warning字段的取值，详细内容参见[SQL自诊断](#)。
- 对外接口通过不同的前缀(gs与pgxc)来区分单CN查询视图以及集群级别查询视图。普通用户仅支持登录到集群的某个CN查询以gs为前缀的视图。
- 实例故障时，历史TopSQL视图有可能记录不全。
- 在某些异常的情况下，历史TopSQL中的状态信息列可能会显示为unknown，其记录的监控信息会导致不准确。
- 历史TopSQL能够记录的SQL语句的规格与实时TopSQL能够记录的SQL语句的规格一致。请参考[实时TopSQL中能够记录的SQL语句的规格](#)。
- 历史TopSQL只有在GUC参数enable\_resource\_record开启时才会记录数据。
- 查询历史TopSQL Query以及算子级别数据时，仅能通过postgres数据库进行访问。
- 历史TopSQL侧重于查询性能的定位定界辅助分析，不作为审计功能使用，不记录语法分析报错类语句。

### 前提条件

- GUC参数enable\_resource\_track为on（默认为on）。
- GUC参数[resource\\_track\\_level](#)为query、perf或operator（默认为query）。设置方法详见[表12-2](#)。
- GUC参数enable\_resource\_record为on（默认为on）。
- GUC参数[resource\\_track\\_duration](#)小于作业执行时间（默认为60s）。
- GUC参数enable\_track\_record\_subsql控制是否记录存储过程、匿名块内部语句（默认为off）。



- 监控作业类型为：
  - 资源监控实时视图（参见表12-1）中记录的作业结束时的执行时间大于或等于 `resource_track_duration` 的作业。
- Cgroups功能正常加载，可通过 `gs_cgroup -P` 查看控制组信息。

## 操作步骤

**步骤1** 通过视图 `gs_wlm_session_history` 查询当前CN最近执行作业结束后的负载记录。

```
SELECT * FROM gs_wlm_session_history;
```

**步骤2** 通过视图 `pgxc_wlm_session_history` 查询所有CN最近执行作业结束后的负载记录。

```
SELECT * FROM pgxc_wlm_session_history;
```

**步骤3** 通过数据表 `gs_wlm_session_info` 查询当前CN作业执行结束后的负载记录。要查到历史记录，必须保证 `enable_resource_record` 为 on。

```
SELECT * FROM gs_wlm_session_info;
```

- 消耗内存最多的10个Query

```
SELECT * FROM gs_wlm_session_info order by max_peak_memory desc limit 10;
```

- 消耗CPU最多的10个Query

```
SELECT * FROM gs_wlm_session_info order by total_cpu_time desc limit 10;
```

**步骤4** 通过视图 `pgxc_wlm_session_info` 查询所有CN的作业执行结束后的负载记录。要查到历史记录，必须保证 `enable_resource_record` 为 on。

```
SELECT * FROM pgxc_wlm_session_info;
```

- 查询所有CN消耗时间最多的10个query

```
SELECT * FROM pgxc_wlm_session_info order by duration desc limit 10;
```

**步骤5** 通过函数 `pgxc_get_wlm_session_info_bytime` 对视图 `pgxc_wlm_session_info` 进行筛选查询，要查到历史记录，必须保证 `enable_resource_record` 为 on。在统计数据量很大的场景中，建议使用该函数进行查询。

### 说明

GaussDB(DWS)集群默认使用时区为UTC时间，与系统时间存在8h时差，请确保数据库时间与系统时间一致后进行以下查询。

- 查询所有CN上开始时间介于“2019-09-10 15:30:00”和“2019-09-10 15:35:00”之间的query，每个CN最多返回10条记录

```
SELECT * FROM pgxc_get_wlm_session_info_bytime('start_time', '2019-09-10 15:30:00', '2019-09-10 15:35:00', 10);
```

- 查询所有CN上结束时间介于“2019-09-10 15:30:00”和“2019-09-10 15:35:00”之间的query，每个CN最多返回10条记录

```
SELECT * FROM pgxc_get_wlm_session_info_bytime('finish_time', '2019-09-10 15:30:00', '2019-09-10 15:35:00', 10);
```

**步骤6** 通过视图 `gs_wlm_operator_history` 查询当前CN作业算子最近执行资源信息。要查到记录，必须保证 `resource_track_level` 为 operator。

```
SELECT * FROM gs_wlm_operator_history;
```

**步骤7** 通过视图 `pgxc_wlm_operator_history` 查询所有CN作业算子最近执行资源信息。要查到记录，必须保证 `resource_track_level` 为 operator。

```
SELECT * FROM pgxc_wlm_operator_history;
```

**步骤8** 通过数据表 `gs_wlm_operator_info` 查询当前CN作业算子历史执行资源信息。要查到记录，必须保证 `resource_track_level` 为 operator 和 `enable_resource_record` 为 on。



```
SELECT * FROM gs_wlm_operator_info;
```

**步骤9** 通过视图pgxc\_wlm\_operator\_info查询所有CN作业算子历史执行资源信息。要查到记录，必须保证resource\_track\_level为operator和enable\_resource\_record为on。

```
SELECT * FROM pgxc_wlm_operator_info;
```

---结束

### 📖 说明

- 对于上述的视图信息，由于预设内存的限制，内存中能够保留的数据记录数量有限，实时查询在结束后会导入到历史相关的视图中。关于记录上限，对于query级别视图，当新的需要记录的查询超过内存约束记录数上限时，则当前查询无法记录，下条查询重新进行规则判断；在每个CN上，记query级别历史视图的内存占用（默认100MB）可通过PG\_TOTAL\_MEMORY\_DETAIL视图进行查询。
- 对于算子级别视图，当需要记录的查询的plan\_node数量加上当前内存中已有的记录数量超过内存约束记录数上限时，则当前查询的所有算子节点不记录，下条查询重新按照算子规则判定。在每个CN上，记算子级别视图在内存中可记录的最大实时和历史记录数分别为max\_oper\_realt\_num(当前系统值为56987)，max\_oper\_hist\_num(113975)；记当前用户业务系统的平均每个查询的节点数为num\_plan\_node，则在每个CN上，实时视图允许客户执行的最大并发数： $num\_realt\_active = max\_oper\_realt\_num / num\_plan\_node$ ；历史视图允许客户执行的最大并发数： $num\_hist\_active = max\_oper\_hist\_num / (180 / run\_time) / num\_plan\_node$ 。
- 如果并发过高，避免需要记录的查询数量超过query级别视图和算子级别视图的值，可以通过参数session\_history\_memory修改历史查询视图的内存，内存增大和查询数量成正比。

## 12.7 TopSQL 查询示例

本章节以查询TPC-DS样例数据的作业为例，演示如何查看**实时TopSQL**和**历史TopSQL**。

### 配置集群参数

查询TopSQL资源监控信息之前，需要先配置相关的GUC参数，以便能查询到作业的资源监控历史信息或归档信息。步骤如下：

1. 登录GaussDB(DWS)管理控制台。
2. 在“集群管理”页面，找到所需要的集群，单击集群名称，进入集群详情页面。
3. 单击“参数修改”标签页，可以看到当前集群的参数值。
4. 修改参数resource\_track\_duration值为合适的值，单击“保存”按钮进行保存。

### 📖 说明

- enable\_resource\_record开关打开后，会引起存储空间膨胀及轻微性能影响，不用时请关闭。
5. 返回集群管理页面，单击右上角的刷新按钮，等待集群参数配置完成。

### TopSQL 查询示例

本示例以TPC-DS样例数据为例。

**步骤1** 打开SQL客户端工具，连接到您的数据库。

**步骤2** 使用explain语句查询所要执行的SQL语句的预估代价，从而可以确定该SQL语句是否会进行资源监控。

默认执行代价大于 **resource\_track\_cost** 的查询才会进行资源监控，用户才可以查询到相关的资源监控信息。

例如，执行如下语句查询该SQL语句的预估执行代价：

```
SET CURRENT_SCHEMA = tpcds;
EXPLAIN WITH customer_total_return AS
( SELECT sr_customer_sk as ctr_customer_sk,
sr_store_sk as ctr_store_sk,
sum(SR_FEE) as ctr_total_return
FROM store_returns, date_dim
WHERE sr_returned_date_sk = d_date_sk AND d_year =2000
GROUP BY sr_customer_sk, sr_store_sk )
SELECT c_customer_id
FROM customer_total_return ctr1, store, customer
WHERE ctr1.ctr_total_return > (select avg(ctr_total_return)*1.2
FROM customer_total_return ctr2
WHERE ctr1.ctr_store_sk = ctr2.ctr_store_sk)
AND s_store_sk = ctr1.ctr_store_sk
AND s_state = 'TN'
AND ctr1.ctr_customer_sk = c_customer_sk
ORDER BY c_customer_id
limit 100;
```

查询结果如下所示，第一行的E-costs列的值即为当前语句的预估代价。

图 12-1 explain 查询结果

id	operation	E-rows	E-width	E-costs
1	-> Row Adapter	6	20	153.06
2	-> Vector Limit	6	20	153.06
3	-> Vector Streaming (type: GATHER)	6	20	153.06
4	-> Vector Limit	6	20	152.84
5	-> Vector Sort	6	20	152.84
6	-> Vector Hash Join (7,26)	6	20	152.83
7	-> Vector Streaming(type: REDISTRIBUTE)	6	4	134.57
8	-> Vector Hash Join (9,18)	6	4	134.46
9	-> Vector Hash Join (10,11)	1	44	97.33
10	-> CStore Scan on store	1	4	60.23
11	-> Vector Subquery Scan on ctr1	6	40	37.07
12	-> Vector Hash Aggregate	6	54	37.06
13	-> Vector Streaming(type: REDISTRIBUTE)	6	22	37.04
14	-> Vector Hash Join (15,17)	6	22	37.00
15	-> Vector Streaming(type: BROADCAST)	6	4	18.74
16	-> CStore Scan on date_dim	1	4	18.06
17	-> CStore Scan on store_returns	60	26	18.02
18	-> Vector Hash Aggregate	6	68	37.09
19	-> Vector Subquery Scan on ctr2	6	36	37.07
20	-> Vector Hash Aggregate	6	54	37.06
21	-> Vector Streaming(type: REDISTRIBUTE)	6	22	37.04
22	-> Vector Hash Join (23,25)	6	22	37.00
23	-> Vector Streaming(type: BROADCAST)	6	4	18.74
24	-> CStore Scan on date_dim	1	4	18.06
25	-> CStore Scan on store_returns	60	26	18.02
26	-> CStore Scan on customer	60	24	18.02

本示例为了演示TopSQL的资源监控功能，需要将 **resource\_track\_cost** 参数设置为比 explain 查询结果中的预估代价小的一个值，例如100，设置方法请参见 **resource\_track\_cost**。

### 说明

在完成本示例后，仍然要将 **resource\_track\_cost** 设置为原始的默认值100000或者一个比较合理的值，否则参数值太小会影响数据库性能。

### 步骤3 执行SQL语句。

```
SET CURRENT_SCHEMA = tpcds;
WITH customer_total_return AS
(SELECT sr_customer_sk as ctr_customer_sk,
sr_store_sk as ctr_store_sk,
sum(SR_FEE) as ctr_total_return
FROM store_returns,date_dim
WHERE sr_returned_date_sk = d_date_sk
AND d_year =2000
```

```
GROUP BY sr_customer_sk ,sr_store_sk)
SELECT c_customer_id
FROM customer_total_return ctr1, store, customer
WHERE ctr1.ctr_total_return > (select avg(ctr_total_return)*1.2
FROM customer_total_return ctr2
WHERE ctr1.ctr_store_sk = ctr2.ctr_store_sk)
AND s_store_sk = ctr1.ctr_store_sk
AND s_state = 'TN'
AND ctr1.ctr_customer_sk = c_customer_sk
ORDER BY c_customer_id
limit 100;
```

**步骤4** 在SQL语句执行期间，查询该条SQL语句在当前CN上的Memory峰值的实时信息。

```
SELECT query,max_peak_memory,average_peak_memory,memory_skew_percent FROM
gs_wlm_session_statistics ORDER BY start_time DESC;
```

含义：查询**query级别**的SQL语句的Memory峰值**实时信息**（语句在所有DN上的每秒最大Memory峰值，在所有DN上的每秒平均Memory峰值，在DN间的Memory倾斜率）

实时TopSQL资源监控信息的更多查询示例，请参见[实时TopSQL](#)。

**步骤5** 等待**步骤3**中的SQL执行完成，然后查询该语句执行期间的资源监控历史信息。

```
select query,start_time,finish_time,duration,status from gs_wlm_session_history order by start_time desc;
```

含义：查询**query级别**的SQL语句执行期间的**历史信息**（语句执行的开始时间，结束时间，实际执行时间，执行状态），时间单位为ms。

历史TopSQL资源监控信息的更多查询示例，请参见[历史TopSQL](#)。

**步骤6** 等待**步骤3**中的SQL执行结束的三分钟后，在info视图中查询该语句的资源监控历史信息。

如果设置参数enable\_resource\_record为“on”，且**步骤3**中SQL的执行时间不小于resource\_track\_duration所设置的值，该条语句的历史信息将会在三分钟后被归档到gs\_wlm\_session\_info视图中。

对于info视图，只支持在连接postgres数据库时查询。因此，请切换为连接postgres数据库后，再执行以下语句进行查询：

```
select query,start_time,finish_time,duration,status from gs_wlm_session_info order by start_time desc;
```

----结束

# 13 用户自定义函数

## 说明

- 实时数仓（单机部署）暂不支持用户自定义函数。
- 实时数仓（单机部署）8.2.0.100及以上集群版本支持OBS导入导出。

## 13.1 PL/Java 语言函数

使用GaussDB(DWS)数据库的PL/Java函数，用户可以使用自己喜欢的Java IDE编写Java方法，并将包含这些方法的jar文件安装到GaussDB(DWS)数据库中，然后使用该方法。GaussDB(DWS) PL/Java基于开源PL/Java 1.5.5开发，所使用的JRE版本为1.8.0\_322。

### 使用限制

Java UDF可以实现一些java逻辑计算，强烈建议不要在Java UDF中封装业务

- 强烈建议不要在Java函数中使用任何方式连接数据库，包括但不限于JDBC。
- 暂不支持的数据类型：除表13-1内容之外的数据类型，包括自定义类型，复杂数据类型（Java Array类及派生类）。
- 暂不支持UDAF，UDTF。

### 示例

使用PL/Java函数时，需要首先将Java方法的实现打包为jar包并且部署到数据库中，然后使用数据库管理员账号创建函数，考虑兼容性问题，请使用1.8.0\_322版本的JRE进行编译。

#### 步骤1 编译jar包。

Java方法的实现和出包可以借助IDE来实现，以下是一个通过命令行来进行编译和出包的简单的示例，通过这个简单示例可以创建一个包含单个方法的jar包文件。

首先，编写一个Example.java文件，在此文件中实现子字符串大写转换的方法，本例中类名为Example，方法名为upperString，内容如下：

```
public class Example
{
    public static String upperString (String text, int beginIndex, int endIndex)
```

```
{
    return text.substring(beginIndex, endIndex).toUpperCase();
}
```

然后，创建manifest.txt清单文件，文件内容如下：

```
Manifest-Version: 1.0
Main-Class: Example
Specification-Title: "Example"
Specification-Version: "1.0"
Created-By: 1.6.0_35-b10-428-11M3811
Build-Date: 08/14/2018 10:09 AM
```

其中，Manifest-Version定义了manifest文件的版本，Main-Class定义了jar文件的入口类，Specification-Title和Specification-Version属于包的扩展属性，Specification-Title定义了扩展规范的标题，Specification-Version定义了扩展规范的版本，Created-By声明了该文件的生成者，Build-Date声明了该文件构建日期。

最后，编译java文件并打包得到javaudf-example.jar

```
javac Example.java
jar cfm javaudf-example.jar manifest.txt Example.class
```

### 须知

jar包的命名规则应符合JDK命名要求，如果含有非法字符，在部署或者使用函数时将出错。

## 步骤2 部署jar包。

Jar包首先需要放置到OBS服务器中，放置方法具体请参见《[对象存储服务控制台指南](#)》的上传文件章节。接着创建访问密钥AK/SK，获取访问密钥的具体步骤，请参见《[数据仓库服务用户指南](#)》中的“[创建访问密钥（AK和SK）](#)”章节。登录数据库运行gs\_extend\_library函数，将文件导入到GaussDB(DWS)中：

```
SELECT gs_extend_library('addjar', 'obs://bucket/path/javaudf-example.jar
accesskey=access_key_value_to_be_replaced secretkey=secret_access_key_value_to_be_replaced
region=region_name libraryname=example');
```

gs\_extend\_library函数如何使用请参见[管理jar包和文件](#)。函数中的AK/SK值，请用户根据实际获取值替换。region\_name请用户根据实际所在的区域名称替换。

## 步骤3 使用PL/Java函数。

首先，使用拥有sysadmin权限的数据库用户（例如：dbadmin）登录数据库并创建java\_upperstring函数如下：

```
CREATE FUNCTION java_upperstring(VARCHAR, INTEGER, INTEGER)
RETURNS VARCHAR
AS 'Example.upperString'
LANGUAGE JAVA;
```

### 说明

- 函数java\_upperstring中定义的数据类型为GaussDB(DWS)的数据类型。该数据类型需要和步骤1中java定义的方法upperString中数据类型一一对应。GaussDB(DWS)与Java数据类型的对应关系，请参见表13-1。
- AS子句用于指定该函数所调用的Java方法的类名和static方法名，格式为“类名.方法名”。该字段需要和步骤1中java定义的类名和方法名一致。
- 使用PL/Java函数时，LANGUAGE字段应指定为JAVA。
- CREATE FUNCTION更多说明，请参见创建函数。

然后，执行java\_upperstring函数：

```
SELECT java_upperstring('test', 0, 1);
```

得到预期结果为：

```
java_upperstring
-----
T
(1 row)
```

#### 步骤4 授权普通用户使用PL/Java函数。

创建普通用户，名称为udf\_user。

```
CREATE USER udf_user PASSWORD 'password';
```

授权普通用户udf\_user对java\_upperstring函数的使用权限。注意，此处需要把函数所在模式和函数的使用权限同时赋予给用户，用户才可以使用此函数。

```
GRANT ALL PRIVILEGES ON SCHEMA public TO udf_user;
GRANT ALL PRIVILEGES ON FUNCTION java_upperstring(VARCHAR, INTEGER, INTEGER) TO udf_user;
```

以普通用户udf\_user登录数据库。

```
SET SESSION AUTHORIZATION udf_user PASSWORD 'password';
```

执行java\_upperstring函数：

```
SELECT public.java_upperstring('test', 0, 1);
```

得到预期结果为：

```
java_upperstring
-----
T
(1 row)
```

#### 步骤5 删除函数。

如果不再使用该函数可以进行删除：

```
DROP FUNCTION java_upperstring;
```

#### 步骤6 卸载jar包。

使用gs\_extend\_library函数卸载jar包：

```
SELECT gs_extend_library('rmjar', 'libraryname=example');
```

----结束

## SQL 定义与使用

- 管理jar包和文件

拥有sysadmin权限的数据库用户可以使用gs\_extend\_library函数来部署、查看和删除数据库中的jar包，函数语法如下：

```
SELECT gs_extend_library('[action]', '[operation]');
```

### 📖 说明

- **action**表明操作动作，值可以为：
  - ls表示查看数据库中jar包，会对各节点的文件进行MD5值一致性检查。
  - addjar表示将OBS服务器中的jar包部署到数据库中。
  - rmjar表示将数据库中的jar包删除。
- **operation**表明操作字符串，格式为：  
obs://[bucket]/[source\_filepath] accesskey=[accesskey] secretkey=[secretkey]  
region=[region] libraryname=[libraryname]
  - bucket: OBS文件所属桶名，不可缺省。
  - source\_filepath: OBS服务器上的文件路径，仅支持jar文件。
  - accesskey: obs服务获得的accesskey，不可缺省。
  - secret\_key: obs服务获得的secretkey，不可缺省。
  - region: 自定义函数jar包所存放的OBS桶所属的区域，不可缺省。
  - libraryname: 自定义库名，此自定义命名用于GaussDB(DWS)内对jar文件的调用。当action为addjar和rmjar时，该参数不可缺省，当action为ls时，该参数可以缺省。注意，自定义库名不允许含有/;&\$<>'\{}"()[]~\*?!等字符。

- **创建函数**

PL/Java函数通过CREATE FUNCTION语法创建，并且定义为LANGUAGE JAVA，且包含RETURNS和AS子句。

- CREATE FUNCTION时指定所创建函数的名称，以及参数类型；
- RETURNS子句用于指定该函数的返回类型；
- AS子句与用于指定该函数所调用的Java方法的类名和static方法名，如果需要向Java方法传递NULL值作为入参，还需要指定该参数类型所对应的Java封装类名（详见[NULL值处理](#)）。
- 更多语法说明，请参见CREATE FUNCTION。

```
CREATE [ OR REPLACE ] FUNCTION function_name
( [ { argname [ argmode ] argtype [ { DEFAULT | := | = } expression ] } [, ...] ] )
[ RETURNS rettype [ DETERMINISTIC ] ]
LANGAUGE JAVA
[
  { IMMUTABLE | STATBLE | VOLATILE }
  | [ NOT ] LEAKPROOF
  | WINDOW
  | { CALLED ON NULL INPUT | RETURNS NULL ON NULL INPUT | STRICT }
  | [ { EXTERNAL } SECURITY INVOKER | [ EXTERNAL ] SECURITY DEFINER | AUTHID DEFINER |
AUTHID CURRENT_USER}
  | { FENCED }
  | COST execution_cost
  | ROWS result_rows
  | SET configuration_parameter { { TO |= } value | FROM CURRENT }
] [ ... ]
{
  AS 'class_name.method_name' ( { argtype } [, ...] )
}
```

- **使用函数**

执行时，PL/Java会根据jar包名的字母序列，在所有部署的jar包中寻找函数指定的Java类，并调用首次找到的类中函数所指定的Java方法，并返回调用结果。

- **删除函数**



PL/Java函数通过DROP FUNCTION语法删除函数。更多语法说明，请参见DROP FUNCTION。

```
DROP FUNCTION [ IF EXISTS ] function_name ( ( [ { [ argmode ] [ argname ] argtype } [, ...] ] )
[ CASCADE | RESTRICT ] );
```

需要注意的是，如果所删除的函数为重载函数（详见[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。

- **函数授权**

非sysadmin户无法创建PL/Java函数，sysadmin用户可以赋予其他类型用户使用函数的权限。更多语法说明，请参见GRANT。

```
GRANT { EXECUTE | 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 ];
```

## 基本数据类型映射关系

表 13-1 PL/Java 默认数据类型映射关系

GaussDB(DWS)	Java
BOOLEAN	boolean
"char"	byte
bytea	byte[]
SMALLINT	short
INTEGER	int
BIGINT	long
FLOAT4	float
FLOAT8	double
CHAR	java.lang.String
VARCHAR	java.lang.String
TEXT	java.lang.String
name	java.lang.String
DATE	java.sql.Timestamp
TIME	java.sql.Time (stored value treated as local time)
TIMETZ	java.sql.Time
TIMESTAMP	java.sql.Timestamp
TIMESTAMPTZ	java.sql.Timestamp



## 数组类型处理

GaussDB(DWS)支持基础数组类型的转换，只需要在创建函数时在数据类型后追加 [] 即可，例如：

```
CREATE FUNCTION java_arrayLength(INTEGER[])
  RETURNS INTEGER
  AS 'Example.getArrayLength'
LANGUAGE JAVA;
```

Java代码类似于：

```
public class Example
{
  public static int getArrayLength(Integer[] intArray)
  {
    return intArray.length;
  }
}
```

那么下面的调用的语句后：

```
SELECT java_arrayLength(ARRAY[1, 2, 3]);
```

得到预期结果应该如下所示：

```
java_arrayLength
-----
3
(1 row)
```

## NULL 值处理

对于默认与Java的简单类型进行映射转换的那些GaussDB(DWS)数据类型，是无法处理NULL值的，如果希望在Java方法里能够获得并处理从GaussDB(DWS)中传入的NULL值，可以使用Java的封装类，并通过以下方式在AS子句中指定该Java封装类：

```
CREATE FUNCTION java_countnulls(INTEGER[])
  RETURNS INTEGER
  AS 'Example.countNulls(java.lang.Integer[])'
LANGUAGE JAVA;
```

Java代码类似于：

```
public class Example
{
  public static int countNulls(Integer[] intArray)
  {
    int nullCount = 0;
    for (int idx = 0; idx < intArray.length; ++idx)
    {
      if (intArray[idx] == null)
        nullCount++;
    }
    return nullCount;
  }
}
```

那么下面的调用的语句后：

```
SELECT java_countNulls(ARRAY[null, 1, null, 2, null]);
```

得到的预期结果应该如下所示：

```
java_countNulls
-----
3
(1 row)
```

## 重载函数

PL/Java支持重载函数，因此可以创建同名函数，或者调用Java代码中的重载方法。步骤如下：

### 步骤1 创建重载函数

例如，在Java中可以实现两个方法名相同，输入参数类型不同的方法dummy(int) 和 dummy(String)

```
public class Example
{
    public static int dummy(int value)
    {
        return value*2;
    }
    public static String dummy(String value)
    {
        return value;
    }
}
```

并在GaussDB(DWS)中创建两个同名函数分别指定为上述两个方法：

```
CREATE FUNCTION java_dummy(INTEGER)
    RETURNS INTEGER
    AS 'Example.dummy'
LANGUAGE JAVA;

CREATE FUNCTION java_dummy(VARCHAR)
    RETURNS VARCHAR
    AS 'Example.dummy'
LANGUAGE JAVA;
```

### 步骤2 调用重载函数

在调用重载函数时，GaussDB(DWS)会根据输入的参数类型去调用匹配该类型的Java方法。因此上述两个函数的调用结果如下所示：

```
SELECT java_dummy(5);
java_dummy
-----
          10
(1 row)

SELECT java_dummy('5');
java_dummy
-----
          5
(1 row)
```

需要注意的是，由于GaussDB(DWS)对数据类型存在隐式转换的情况，因此建议在调用重载函数时，指定输入参数的类型，例如：

```
SELECT java_dummy(5::varchar);
java_dummy
-----
          5
(1 row)
```

此时会优先匹配所指定的参数类型，如果不存在指定参数类型的Java方法，则会对参数进行隐式转换匹配转换后的参数类型对应的Java方法。

```
SELECT java_dummy(5::INTEGER);
java_dummy
-----
          10
```

```
(1 row)

DROP FUNCTION java_dummy(INTEGER);

SELECT java_dummy(5::INTEGER);
 java_dummy
-----
5
(1 row)
```

### 须知

隐式转换的数据类型包括：

- 可以默认转换为INTEGER类型的包括：SMALLINT
- 可以默认转换为BIGINT类型的包括：SMALLINT, INTEGER
- 可以默认转换为BOOL类型的包括：TINYINT, SMALLINT, INTEGER, BIGINT
- 可以默认转换为TEXT类型的包括：CHAR, NAME, BIGINT, INTEGER, SMALLINT, TINYINT, RAW, FLOAT4, FLOAT8, BPCHAR, VARCHAR, NVARCHAR2, DATE, TIMESTAMP, TIMESTAMPTZ, NUMERIC, SMALLDATETIME
- 可以默认转换为VARCHAR类型的包括：TEXT, CHAR, BIGINT, INTEGER, SMALLINT, TINYINT, RAW, FLOAT4, FLOAT8, BPCHAR, DATE, NVARCHAR2, TIMESTAMP, NUMERIC, SMALLDATETIME

### 步骤3 删除重载函数

对于重载函数，删除时需要指定函数的参数类型，否则无法删除。

```
DROP FUNCTION java_dummy(INTEGER);
```

----结束

## 相关 GUC 参数

- **pljava\_vmoptions**

会话级别的GUC参数，该参数用于设置JVM的启动参数，例如：

```
SET pljava_vmoptions='-Xmx64m -Xms2m -XX:MaxMetaspaceSize=8m';
```

pljava\_vmoptions可接受的参数包括：

- JDK8 JVM启动参数
- JDK8 JVM系统属性参数（以-D开头，如：-Djava.ext.dirs）。

### 须知

不建议用户设置任何包含目录的参数，可能会导致不可预期的行为。

- 用户自定义参数（以-D开头，如：-Duser.defined.option）

**须知**

如果所设置的参数不在上述范围内，视为非法参数，会在调用函数时报错。

```
SET pljava_vmoptions=' illegal.option';
SET
SELECT java_dummy(5::int);
ERROR: UDF Error:cannot use PL/Java before successfully completing its setup.Please check if
your pljava_vmooption is set correctly,since we do not ignore illegal parameters.Or check the log
for more messages.
```

**• FencedUDFMemoryLimit**

会话级别的GUC参数，用户限制会话发起的单个Fenced UDF Worker进程的最大虚拟内存使用量，设置方法如下：

```
SET FencedUDFMemoryLimit='512MB';
```

该参数的取值范围为 (150MB, 1G]，当设置大于1G时会立即报错，当设置小于等于150MB时，则会在调用函数时报错。

**须知**

- FencedUDFMemoryLimit设置为0，表示不控制Fenced UDF Worker的虚拟内存使用量。
- 建议通过设置udf\_memory\_limit控制Fenced UDF Worker使用的物理内存量。不建议用户使用FencedUDFMemoryLimit，尤其在使用Java UDF时不建议用户设置此参数。但是如果用户非常清楚设置该参数带来的影响，可以参考下列信息进行设置：
  - C UDF worker启动之后，占用的虚拟内存约为200MB，占用的物理内存约为16MB。
  - Java UDF worker启动之后，占用的虚拟内存约为2.5GB，占用的物理内存约为50MB。

## 异常处理

如果在JVM中发生异常，PL/Java的异常处理机制会将异常时JVM的堆栈信息输出到客户端。

## 日志

PL/Java使用标准的Java Logger。因此，用户可以通过如下方式记录日志：

```
Logger.getAnonymousLogger().config( "Time is " + new
Date(System.currentTimeMillis()));
```

初始化的Java Logger类会默认设置为CONFIG级别，对应为GaussDB(DWS)的LOG级别。Java Logger类输出的日志消息都会重定向到GaussDB(DWS)后端，并写入到服务器日志或显示在用户界面上。MPPDB服务器日志将记录LOG、WARNING、ERROR级别的信息，而SQL用户界面将显示WARNING和ERROR级别的日志消息。Java Logger级别与GaussDB(DWS)的日志级别对应关系见下表。

表 13-2 PL/Java 日志级别

java.util.logging.Level	GaussDB(DWS) 日志级别
SERVER	ERROR
WARNING	WARNING
CONFIG	LOG
INFO	INFO
FINE	DEBUG1
FINER	DEBUG2
FINEST	DEBUG3

用户可以通过以下方式更改Java Logger的记录级别。例如通过下面的Java代码修改Java Logger级别为SEVERE，此时再记录WARNING级别的日志时，日志消息（msg）就不会再写入到GaussDB(DWS)日志中。

```
Logger log = Logger.getAnonymousLogger();  
Log.setLevel(Level.SEVERE);  
log.log(Level.WARNING, msg);
```

## 安全问题

在GaussDB(DWS)中，PL/Java是一种untrusted语言，PL/Java函数只能由数据库sysadmin用户进行创建，通过GRANT方式赋予其他用户使用权限（详见[函数授权](#)）。

同时PL/Java控制用户对文件系统的访问权限，不允许用户在Java方法中对大部分的系统文件进行读操作，不允许所有的写、删除和执行操作。

## 13.2 PL/pgSQL 语言函数

PL/pgSQL类似于Oracle的PL/SQL，是一种可载入的过程语言。

用PL/pgSQL创建的函数可以被用在任何可以使用内建函数的地方。例如，可以创建复杂条件的计算函数并且后面用它们来定义操作符或把它们用于索引表达式。

SQL被大多数数据库用作查询语言。它是可移植的并且容易学习。但是每一个SQL语句必须由数据库服务器单独执行。

这意味着客户端应用必须发送每一个查询到数据库服务器、等待它被处理、接收并处理结果、做一些计算，然后发送更多查询给服务器。如果客户端和数据库服务器不在同一台机器上，所有这些会引起进程间通信并且将带来网络负担。

通过PL/pgSQL，可以将一整块计算和一系列查询分组在数据库服务器内部，这样就有了一种过程语言的能力并且使SQL更易用，同时能节省的客户端/服务器通信开销。

- 客户端和服务端之间的额外往返通信被消除。
- 客户端不需要的中间结果不必被整理或者在服务器和客户端之间传送。
- 多轮的查询解析可以被避免。

PL/pgSQL可以使用SQL中所有的数据类型、操作符和函数。

应用PL/pgSQL创建函数的语法为CREATE FUNCTION。正如前面所说，PL/pgSQL类似于Oracle的PL/SQL，是一种可载入的过程语言。其应用方法与[存储过程](#)相似，只是存储过程无返回值，函数有返回值。

# 14 存储过程

## 14.1 存储过程

商业规则和业务逻辑可以通过程序存储在GaussDB(DWS)中，这个程序就是存储过程。

存储过程是SQL，PL/SQL，Java语句的组合。存储过程使执行商业规则的代码可以从应用程序中移动到数据库。从而，代码存储一次能够被多个程序使用。

存储过程的创建及调用办法请参考[CREATE PROCEDURE](#)。

[PL/pgSQL语言函数](#)节所提到的PL/pgSQL语言创建的函数与存储过程的应用方法相通。下面各节中，除非特别声明，否则内容通用于存储过程和PL/pgSQL语言函数。

## 14.2 数据类型

数据类型是一组值的集合以及定义在这个值集上的一组操作。GaussDB(DWS)数据库是由表的集合组成的，而各表中的列定义了该表，每一列都属于一种数据类型，GaussDB(DWS)根据数据类型有相应函数对其内容进行操作，例如GaussDB(DWS)可对数值型数据进行加、减、乘、除操作。

## 14.3 数据类型转换

数据库中允许有些数据类型进行隐式类型转换（赋值、函数调用的参数等），有些数据类型间不允许进行隐式数据类型转换，可尝试使用GaussDB(DWS)提供的类型转换函数，例如CAST进行数据类型强转。

GaussDB(DWS)数据库常见的隐式类型转换，请参见[表14-1](#)。

### 须知

GaussDB(DWS)支持的DATE的效限范围是：公元前4713年到公元294276年。

表 14-1 隐式类型转换表

原始数据类型	目标数据类型	备注
CHAR	VARCHAR2	-
CHAR	NUMBER	原数据必须由数字组成。
CHAR	DATE	原数据不能超出合法日期范围。
CHAR	RAW	-
CHAR	CLOB	-
VARCHAR2	CHAR	-
VARCHAR2	NUMBER	原数据必须由数字组成。
VARCHAR2	DATE	原数据不能超出合法日期范围。
VARCHAR2	CLOB	-
NUMBER	CHAR	-
NUMBER	VARCHAR2	-
DATE	CHAR	-
DATE	VARCHAR2	-
RAW	CHAR	-
RAW	VARCHAR2	-
CLOB	CHAR	-
CLOB	VARCHAR2	-
CLOB	NUMBER	原数据必须由数字组成。
INT4	CHAR	-

## 14.4 数组和 record

### 14.4.1 数组

#### 数组类型的使用

在使用数组之前，需要自定义一个数组类型。

在存储过程中紧跟AS关键字后面定义数组类型。定义方法为：  
`TYPE array_type IS VARRAY(size) OF data_type [NOT NULL];`

其中：



- array\_type: 要定义的数组类型名。
- VARRAY: 表示要定义的数组类型。
- size: 取值为正整数, 表示可以容纳的成員的最大数量。
- data\_type: 要创建的数组中成員的类型。
- NOT NULL: 可选约束, 可以约束该数组中的元素均不为NULL。

#### 📖 说明

- 在GaussDB(DWS)中, 数组会自动增长, 访问越界会返回一个NULL, 不会报错。越界写入数组会提示: Subscript outside of limit.
- 在存储过程中定义的数组类型, 其作用域仅在该存储过程中。
- 建议选择上述定义方法的一种来自定义数组类型, 当同时使用两种方法定义同名的数组类型时, GaussDB(DWS)会优先选择存储过程中定义的数组类型来声明数组变量。

GaussDB(DWS) 8.1.0之前版本, 由于数组可以自动增长, 系统不会校验数组越界以及数组元素的长度限制。当前版本为了兼容Oracle的用法增加了相关约束。如果已经存在越界写入等场景, 可通过在`behavior_compat_options`参数中配置`varray_verification`, 来兼容之前不校验的行为。

示例:

```
--演示在存储过程中对数组声明操作。
CREATE OR REPLACE PROCEDURE array_proc
AS
    TYPE ARRAY_INTEGER IS VARRAY(1024) OF INTEGER;--定义数组类型
    TYPE ARRAY_INTEGER_NOT_NULL IS VARRAY(1024) OF INTEGER NOT NULL;--定义非空数组类型
    ARRINT ARRAY_INTEGER := ARRAY_INTEGER(); --声明数组类型的变量
BEGIN
    ARRINT.extend(10);
    FOR I IN 1..10 LOOP
        ARRINT(I) := I;
    END LOOP;
    DBMS_OUTPUT.PUT_LINE(ARRINT.COUNT);
    DBMS_OUTPUT.PUT_LINE(ARRINT(1));
    DBMS_OUTPUT.PUT_LINE(ARRINT(10));
    DBMS_OUTPUT.PUT_LINE(ARRINT(ARRINT.FIRST));
    DBMS_OUTPUT.PUT_LINE(ARRINT(ARRINT.last));
END;
/

--调用该存储过程。
CALL array_proc();
10
1
10
1
10

--删除存储过程。
DROP PROCEDURE array_proc;
```

## ROWTYPE 类型数组的声明及使用

除了上述示例中普通数组以及not null数组的声明及使用, 数组中同时支持rowtype类型数组的声明及使用。

示例:

```
--演示在存储过程中对数组COUNT函数的用法。
CREATE TABLE tbl (a int, b int);
INSERT INTO tbl VALUES(1, 2),(2, 3),(3, 4);
CREATE OR REPLACE PROCEDURE array_proc
AS
```

```
CURSOR all_tbl IS SELECT * FROM tbl ORDER BY a;
TYPE tbl_array_type IS varray(50) OF tbl%rowtype; --定义rowtype类型的数组，tbl是任意表。
tbl_array tbl_array_type;
tbl_item tbl%rowtype;
inx1 int;
BEGIN
tbl_array := tbl_array_type();
inx1 := 0;
FOR tbl_item IN all_tbl LOOP
inx1 := inx1 + 1;
tbl_array(inx1) := tbl_item;
END LOOP;
WHILE inx1 IS NOT NULL LOOP
DBMS_OUTPUT.PUT_LINE('tbl_array(inx1).a=' || tbl_array(inx1).a || ' tbl_array(inx1).b=' ||
tbl_array(inx1).b);
inx1 := tbl_array.PRIOR(inx1);
END LOOP;
END;
/
```

执行结果：

```
call array_proc();
tbl_array(inx1).a=3 tbl_array(inx1).b=4
tbl_array(inx1).a=2 tbl_array(inx1).b=3
tbl_array(inx1).a=1 tbl_array(inx1).b=2
```

## 数组相关函数使用

在GaussDB(DWS)中，提供了类似Oracle相关的数组函数的支持，可以通过以下函数获取数组的一些属性或者对数组内容进行操作。

### COUNT

COUNT函数可以返回当前数组元素的数量，仅统计初始化过的元素或者经过EXTEND函数扩展后的元素。

用法如下：

*varray*.COUNT或*varray*.COUNT()

示例：

```
--演示在存储过程中对数组COUNT函数的用法。
CREATE OR REPLACE PROCEDURE test_varray
AS
TYPE varray_type IS VARRAY(20) OF INT;
v_varray varray_type;
BEGIN
v_varray := varray_type(1, 2, 3);
DBMS_OUTPUT.PUT_LINE('v_varray.count=' || v_varray.count);
v_varray.extend;
DBMS_OUTPUT.PUT_LINE('v_varray.count=' || v_varray.count);
END;
/
```

执行结果：

```
call test_varray();
v_varray.count=3
v_varray.count=4
```

### FIRST 和 LAST

FIRST函数可以返回第一个元素的下标。LAST函数可以返回最后一个元素的下标。

用法如下：

*varray*.FIRST或*varray*.FIRST()

*varray*.LAST或*varray*.LAST()

示例：

```
--演示在存储过程中对数组FIRST和LAST函数的用法。
CREATE OR REPLACE PROCEDURE test_varray
AS
    TYPE varray_type IS VARRAY(20) OF INT;
    v_varray varray_type;
BEGIN
    v_varray := varray_type(1, 2, 3);
    DBMS_OUTPUT.PUT_LINE('v_varray.first=' || v_varray.first);
    DBMS_OUTPUT.PUT_LINE('v_varray.last=' || v_varray.last);
END;
/
```

执行结果：

```
call test_varray();
v_varray.first=1
v_varray.last=3
```

## EXTEND

### 说明

EXTEND函数主要是为了兼容Oracle的两种用法。在GaussDB(DWS)中，数组会自动增长，EXTEND函数不是必须的。如果是新写的存储过程，完全没有必要使用EXTEND函数。

EXTEND函数可以对数组进行扩展，EXTEND有两种调用方式。

- 方式一：

EXTEND包含一个整形入参，表示数组向后扩展size大小的长度，EXTEND后COUNT和LAST函数的值也会有相应的变化。

用法如下：

*varray*.EXTEND(size)

其中*varray*.EXTEND这种无参的调用默认会向后扩展1位等价于*varray*.EXTEND(1)

- 方式二：

EXTEND包含两个整形入参，第一个参数代表向后扩展size大小的长度，第二个参数表示扩展后的数组元素值和之下标为index的元素相同。

用法如下：

*varray*.EXTEND(size, index)

示例：

```
--演示在存储过程中对数组EXTEND函数的用法。
CREATE OR REPLACE PROCEDURE test_varray
AS
    TYPE varray_type IS VARRAY(20) OF INT;
    v_varray varray_type;
BEGIN
    v_varray := varray_type(1, 2, 3);
    v_varray.extend(3);
    DBMS_OUTPUT.PUT_LINE('v_varray.count=' || v_varray.count);
    v_varray.extend(2,3);
    DBMS_OUTPUT.PUT_LINE('v_varray.count=' || v_varray.count);
    DBMS_OUTPUT.PUT_LINE('v_varray(7)=' || v_varray(7));
END;
```

```
DBMS_OUTPUT.PUT_LINE('v_varray(8)=' || v_varray(7));  
END;  
/
```

执行结果:

```
call test_varray();  
v_varray.count=6  
v_varray.count=8  
v_varray(7)=3  
v_varray(8)=3
```

## NEXT 和 PRIOR

NEXT函数和PRIOR函数主要用于数组的循环遍历中，NEXT函数会根据入参index值，返回下一个数组元素的下标，若已经到达数组下标最大值则返回NULL。PRIOR函数会根据入参index值，返回上一个数组元素的下标，若已经到达数组下标最小值则返回NULL。

用法如下:

*vvarray*.NEXT(index)

*vvarray*.PRIOR(index)

示例:

```
--演示在存储过程中对数组NEXT和PRIOR函数的用法。  
CREATE OR REPLACE PROCEDURE test_varray  
AS  
  TYPE vvarray_type IS VARRAY(20) OF INT;  
  v_varray vvarray_type;  
  i int;  
BEGIN  
  v_varray := vvarray_type(1, 2, 3);  
  
  i := v_varray.COUNT;  
  WHILE i IS NOT NULL LOOP  
    DBMS_OUTPUT.PUT_LINE('test prior v_varray('||i||')=' || v_varray(i));  
    i := v_varray.PRIOR(i);  
  END LOOP;  
  
  i := 1;  
  WHILE i IS NOT NULL LOOP  
    DBMS_OUTPUT.PUT_LINE('test next v_varray('||i||')=' || v_varray(i));  
    i := v_varray.NEXT(i);  
  END LOOP;  
END;  
/
```

执行结果:

```
call test_varray();  
test prior v_varray(3)=3  
test prior v_varray(2)=2  
test prior v_varray(1)=1  
test next v_varray(1)=1  
test next v_varray(2)=2  
test next v_varray(3)=3
```

## EXISTS

EXISTS函数可以判断数组下标是否存在。

用法如下:

### *varray*.EXISTS(index)

示例:

```
--演示在存储过程中对数组EXISTS函数的用法。
CREATE OR REPLACE PROCEDURE test_varray
AS
  TYPE varray_type IS VARRAY(20) OF INT;
  v_varray varray_type;
BEGIN
  v_varray := varray_type(1, 2, 3);
  IF v_varray.EXISTS(1) THEN
    DBMS_OUTPUT.PUT_LINE('v_varray.EXISTS(1)');
  END IF;
  IF NOT v_varray.EXISTS(10) THEN
    DBMS_OUTPUT.PUT_LINE('NOT v_varray.EXISTS(10)');
  END IF;
END;
/
```

执行结果:

```
call test_varray();
v_varray.EXISTS(1)
NOT v_varray.EXISTS(10)
```

## TRIM

TRIM函数可以从数组尾部删除指定数量的元素。

用法如下:

*varray*.TRIM(size)

其中*varray*.TRIM这种无参的调用会默认入参为1，等价于*varray*.TRIM(1)

示例:

```
--演示在存储过程中对数组TRIM函数的用法。
CREATE OR REPLACE PROCEDURE test_varray
AS
  TYPE varray_type IS VARRAY(20) OF INT;
  v_varray varray_type;
BEGIN
  v_varray := varray_type(1, 2, 3, 4, 5);
  v_varray.trim(3);
  DBMS_OUTPUT.PUT_LINE('v_varray.count' || v_varray.count);
  v_varray.trim;
  DBMS_OUTPUT.PUT_LINE('v_varray.count:' || v_varray.count);
END;
/
```

执行结果:

```
call test_varray();
v_varray.count:2
v_varray.count:1
```

## DELETE

DELETE函数可以从数组删除数组中的所有元素。

用法如下:

*varray*.DELETE或*varray*.DELETE()

示例:

```
--演示在存储过程中对数组DELETE函数的用法。
CREATE OR REPLACE PROCEDURE test_varray
AS
    TYPE varray_type IS VARRAY(20) OF INT;
    v_varray varray_type;
BEGIN
    v_varray := varray_type(1, 2, 3, 4, 5);
    v_varray.delete;
    DBMS_OUTPUT.PUT_LINE('v_varray.count:' || v_varray.count);
END;
/
```

执行结果：

```
call test_varray();
v_varray.count:0
```

## LIMIT

LIMIT函数可以返回数组的最大长度限制。

用法如下：

*varray*.LIMIT或*varray*.LIMIT()

示例：

```
--演示在存储过程中对数组LIMIT函数的用法。
CREATE OR REPLACE PROCEDURE test_varray
AS
    TYPE varray_type IS VARRAY(20) OF INT;
    v_varray varray_type;
BEGIN
    v_varray := varray_type(1, 2, 3, 4, 5);
    DBMS_OUTPUT.PUT_LINE('v_varray.limit:' || v_varray.limit);
END;
/
```

执行结果：

```
call test_varray();
v_varray.limit:20
```

## 14.4.2 record

### record 类型的变量

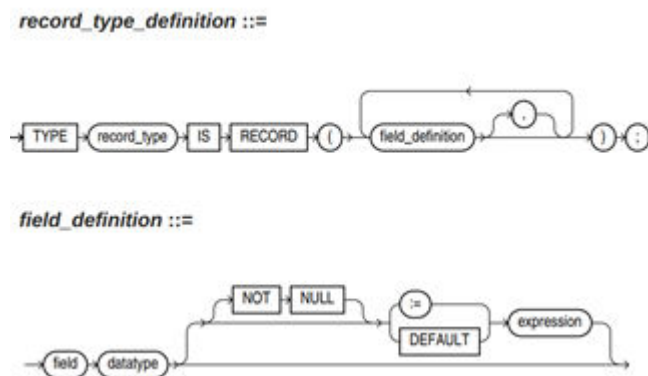
创建一个record变量的方式：

定义一个record类型，然后使用该类型来声明一个变量。

### 语法

record类型的语法参见图14-1。

图 14-1 record 类型的语法



对以上语法格式的解释如下：

- record\_type：声明的类型名称。
- field：record类型中的成员名称。
- datatype：record类型中成员的类型。
- expression：设置默认值的表达式。

### 说明

在GaussDB(DWS)中：

- record类型的变量的赋值支持，
  - 在函数或存储过程的声明阶段，声明一个record类型，并且可以在该类型中定义成员变量。
  - 一个record变量到另一个record变量的赋值。
  - SELECT INTO和FETCH向一个record类型的变量中赋值。
  - 将一个NULL值赋值给一个record变量。
- 不支持INSERT和UPDATE语句使用record变量进行插入数据和更新数据。
- 如果成员有复合类型，在声明阶段不支持指定默认值，该行为同声明阶段的变量一样。

## 示例

下面存储过程中用到的表定义如下：

```
CREATE TABLE emp_rec
(
  empno      numeric(4,0),
  ename      character varying(10),
  job        character varying(9),
  mgr        numeric(4,0),
  hiredate   timestamp(0) without time zone,
  sal        numeric(7,2),
  comm       numeric(7,2),
  deptno     numeric(2,0)
)
with (orientation = column,compression=middle)
distribute by hash (sal);
\d emp_rec
```

Column	Type	Modifiers
empno	numeric(4,0)	not null
ename	character varying(10)	
job	character varying(9)	

```
mgr      | numeric(4,0)          |
hiredate | timestamp(0) without time zone |
sal      | numeric(7,2)          |
comm     | numeric(7,2)          |
deptno   | numeric(2,0)          |

--演示在存储过程中对数组进行操作。
CREATE OR REPLACE FUNCTION regress_record(p_w VARCHAR2)
RETURNS
VARCHAR2 AS $$
DECLARE

    --声明一个record类型。
    type rec_type is record (name varchar2(100), epno int);
    employer rec_type;

    --使用%type声明record类型
    type rec_type1 is record (name emp_rec.ename%type, epno int not null :=10);
    employer1 rec_type1;

    --声明带有默认值的record类型
    type rec_type2 is record (
        name varchar2 not null := 'SCOTT',
        epno int not null :=10);
    employer2 rec_type2;
    CURSOR C1 IS select ename,empno from emp_rec order by 1 limit 1;

BEGIN
    --对一个record类型的变量的成员赋值。
    employer.name := 'WARD';
    employer.epno = 18;
    raise info 'employer name: % , epno:%', employer.name, employer.epno;

    --将一个record类型的变量赋值给另一个变量。
    employer1 := employer;
    raise info 'employer1 name: % , epno: %', employer1.name, employer1.epno;

    --将一个record类型变量赋值为NULL。
    employer1 := NULL;
    raise info 'employer1 name: % , epno: %', employer1.name, employer1.epno;

    --获取record变量的默认值。
    raise info 'employer2 name: % ,epno: %', employer2.name, employer2.epno;

    --在for循环中使用record变量
    for employer in select ename,empno from emp_rec order by 1 limit 1
    loop
        raise info 'employer name: % , epno: %', employer.name, employer.epno;
    end loop;

    --在select into 中使用record变量。
    select ename,empno into employer2 from emp_rec order by 1 limit 1;
    raise info 'employer name: % , epno: %', employer2.name, employer2.epno;

    --在cursor中使用record变量。
    OPEN C1;
    FETCH C1 INTO employer2;
    raise info 'employer name: % , epno: %', employer2.name, employer2.epno;
    CLOSE C1;
    RETURN employer.name;
END;
$$
LANGUAGE plpgsql;

--调用该存储过程。
CALL regress_record('abc');
INFO: employer name: WARD , epno:18
INFO: employer1 name: WARD , epno: 18
INFO: employer1 name: <NULL> , epno: <NULL>
```



```
INFO: employer2 name: SCOTT ,epno: 10
--删除存储过程。
DROP PROCEDURE regress_record;
```

## 14.5 声明语法

### 14.5.1 基本结构

#### 结构

PL/SQL块中可以包含子块，子块可以位于PL/SQL中任何部分。PL/SQL块的结构如下：

- 声明部分：声明PL/SQL用到的变量，类型及游标，以及局部的存储过程和函数。  
DECLARE

#### 说明

不涉及变量声明时声明部分可以没有。

- 对匿名块来说，没有变量声明部分时，可以省去DECLARE关键字。
- 对存储过程来说，没有DECLARE， AS相当于DECLARE。即便没有变量声明的部分，关键字AS也必须保留。
- 执行部分：过程及SQL语句，程序的主要部分。必选。  
BEGIN
- 执行异常部分：错误处理。可选。  
EXCEPTION
- 结束  
END;  
/

#### 须知

禁止在PL/SQL块中使用连续的Tab，连续的Tab可能会造成在使用gsqI工具带“-r”参数执行PL/SQL块时出现异常。

#### 分类

PL/SQL块可以分为以下几类：

- 匿名块：动态构造，只能执行一次。语法请参考图14-2。
- 子程序：存储在数据库中的存储过程、函数和操作符及高级包等。当在数据库上建立好后，可以在其他程序中调用它们。

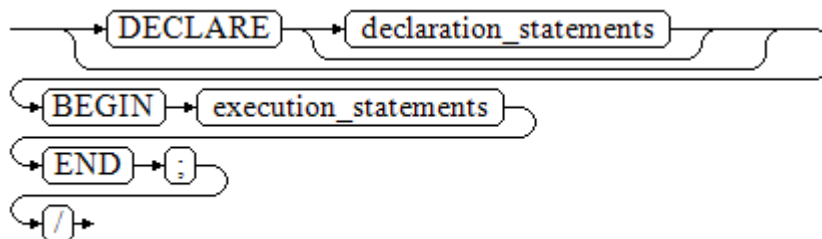
### 14.5.2 匿名块

匿名块（Anonymous Block）一般用于不频繁执行的脚本或不重复进行的活动。它们在一个会话中执行，并不被存储。

#### 语法

匿名块的语法参见图14-2。

图 14-2 anonymous\_block::=



对以上语法图的解释如下：

- 匿名块程序实施部分，以BEGIN语句开始，以END语句停顿，以一个分号结束。输入“/”按回车执行它。

#### 须知

最后的结束符“/”必须独占一行，不能直接跟在END后面。

- 声明部分包括变量定义、类型、游标定义等。
- 最简单的匿名块不执行任何命令。但一定要在任意实施块里至少有一个语句，甚至是一个NULL语句。

## 示例

下面列举了基本的匿名块程序：

```
--空语句块
BEGIN
  NULL;
END;
/

--将信息打印到控制台：
BEGIN
  dbms_output.put_line('hello world!');
END;
/

--将变量内容打印到控制台：
DECLARE
  my_var VARCHAR2(30);
BEGIN
  my_var := 'world';
  dbms_output.put_line('hello'||my_var);
END;
/
```

### 14.5.3 子程序

存储在数据库中的存储过程、函数和操作符及高级包等。当在数据库上建立好后，可以在其他程序中调用它们。

## 14.6 基本语句

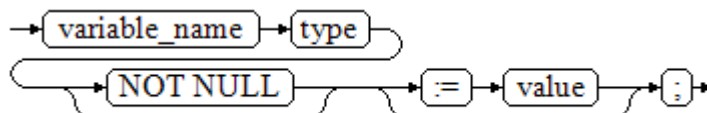
## 14.6.1 定义变量

介绍PL/SQL中变量的声明，以及该变量在代码中的作用域。

### 变量声明

变量声明语法请参见图14-3。

图 14-3 declare\_variable::=



对以上语法格式的解释如下：

- variable\_name，为变量名。
- type，为变量类型。
- value，是该变量的初始值（如果不给定初始值，则初始为NULL）。value也可以是表达式。

### 示例

```
DECLARE
  emp_id INTEGER := 7788; --定义变量并赋值
BEGIN
  emp_id := 5*7784; --变量赋值
END;
/
```

变量类型除了支持基本类型，还可使用%TYPE和%ROWTYPE去声明一些与其他表字段或表结构本身相关的变量。

### %TYPE 属性

%TYPE主要用于声明某个与其他变量类型（例如，表中某列的类型）相同的变量。假如想定义一个my\_name变量，它的变量类型与employee的firstname类型相同，可使用如下定义：

```
my_name employee.firstname%TYPE
```

这样定义可以带来两个好处，首先，不用预先知道employee表的firstname类型具体是什么。其次，即使之后firstname类型有了变化，也不需要再次修改my\_name的类型。

### %ROWTYPE 属性

%ROWTYPE属性主要用于对一组数据的类型声明，用于存储表中的一行数据，或从游标匹配的结果。假如需要一组数据，该组数据的字段名称与字段类型都与employee表相同。可以通过如下定义：

```
my_employee employee%ROWTYPE
```

**须知**

多个CN的环境下，存储过程中无法声明临时表的%ROWTYPE及%TYPE属性。因为临时表仅在当前session有效，在编译阶段其他CN无法看到当前CN的临时表。故多个CN的环境下，会提示该临时表不存在。

**变量作用域**

变量的作用域表示变量在代码块中的可访问性和可用性。只有在它的作用域内，变量才有效。

- 变量必须在declare部分声明，即必须建立BEGIN-END块。块结构也强制变量必须先声明后使用，即变量在过程内有不同作用域、不同的生存期。
- 同一变量可以在不同的作用域内定义多次，内层的定义会覆盖外层的定义。
- 在外部块定义的变量，可以在嵌套块中使用。但外部块不能访问嵌套块中的变量。

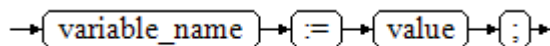
**示例**

```
DECLARE
  emp_id INTEGER :=7788; --定义变量并赋值
  outer_var INTEGER :=6688; --定义变量并赋值
BEGIN
  DECLARE
    emp_id INTEGER :=7799; --定义变量并赋值
    inner_var INTEGER :=6688; --定义变量并赋值
  BEGIN
    dbms_output.put_line('inner emp_id ='||emp_id); --显示值为7799
    dbms_output.put_line('outer_var ='||outer_var); --引用外部块的变量
  END;
  dbms_output.put_line('outer emp_id ='||emp_id); --显示值为7788
END;
/
```

**14.6.2 赋值语句****语法**

给变量赋值的语法请参见图14-4。

图 14-4 assignment\_value::=



对以上语法格式的解释如下：

- variable\_name，为变量名。
- value，可以是值或表达式。值value的类型需要和变量variable\_name的类型兼容才能正确赋值。

**示例**

```
DECLARE
  emp_id INTEGER := 7788;--赋值
BEGIN
```

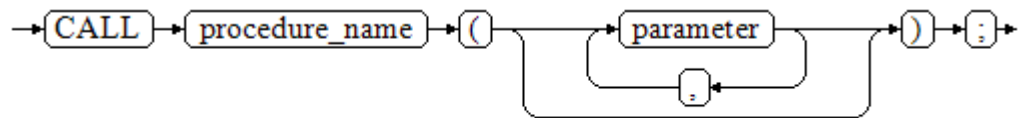
```
emp_id := 5;--赋值
emp_id := 5*7784;
END;
/
```

### 14.6.3 调用语句

#### 语法

调用一个语句的语法请参见图14-5。

图 14-5 call\_clause::=



对以上语法格式的解释如下：

- procedure\_name，为存储过程名。
- parameter，为存储过程的参数，可以没有或者有多个参数。

#### 示例

```
--创建存储过程proc_staffs
CREATE OR REPLACE PROCEDURE proc_staffs
(
section NUMBER(6),
salary_sum out NUMBER(8,2),
staffs_count out INTEGER
)
IS
BEGIN
SELECT sum(salary), count(*) INTO salary_sum, staffs_count FROM staffs where section_id = section;
END;
/

--创建存储过程proc_return.
CREATE OR REPLACE PROCEDURE proc_return
AS
v_num NUMBER(8,2);
v_sum INTEGER;
BEGIN
proc_staffs(30, v_sum, v_num); --调用语句
dbms_output.put_line(v_sum||'#'||v_num);
RETURN; --返回语句
END;
/

--调用存储过程proc_return.
CALL proc_return();

--清除存储过程
DROP PROCEDURE proc_staffs;
DROP PROCEDURE proc_return;

--创建函数func_return.
CREATE OR REPLACE FUNCTION func_return returns void
language plpgsql
AS $$
```

```

DECLARE
v_num INTEGER := 1;
BEGIN
dbms_output.put_line(v_num);
RETURN; --返回语句
END $$;

-- 调用函数func_return
CALL func_return();
1

-- 清除函数
DROP FUNCTION func_return;
    
```

## 14.7 动态语句

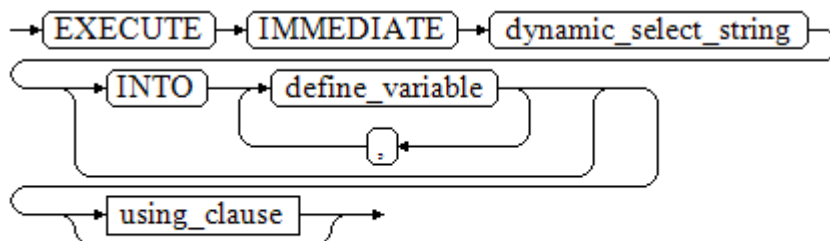
### 14.7.1 执行动态查询语句

介绍执行动态查询语句。GaussDB(DWS)提供两种方式：使用EXECUTE IMMEDIATE、OPEN FOR实现动态查询。前者通过动态执行SELECT语句，后者结合了游标的使用。当需要将查询的结果保存在一个数据集用于提取时，可使用OPEN FOR实现动态查询。

#### EXECUTE IMMEDIATE

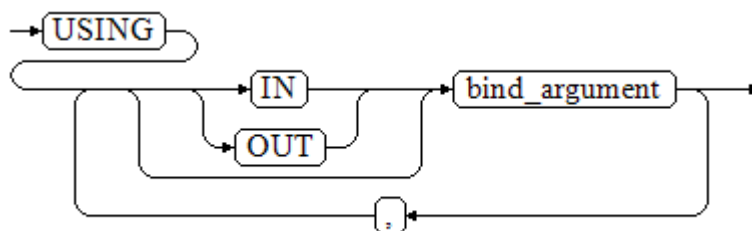
语法图请参见图14-6。

图 14-6 EXECUTE IMMEDIATE dynamic\_select\_clause::=



using\_clause子句的语法图参见图14-7。

图 14-7 using\_clause-1



对以上语法格式的解释如下：

- define\_variable, 用于指定存放单行查询结果的变量。
- USING IN bind\_argument, 用于指定存放传递给动态SQL值的变量, 即在dynamic\_select\_string中存在占位符时使用。
- USING OUT bind\_argument, 用于指定存放动态SQL返回值的变量。

#### 须知

- 查询语句中, into和out不能同时存在;
- 占位符命名以“:”开始, 后面可跟数字、字符或字符串, 与USING子句的bind\_argument一一对应;
- bind\_argument只能是值、变量或表达式, 不能是表名、列名、数据类型等数据库对象, 即不支持使用bind\_argument为动态SQL语句传递模式对象。如果存储过程需要通过声明参数传递数据库对象来构造动态SQL语句(常见于执行DDL语句时), 建议采用连接运算符“||”拼接dynamic\_select\_clause;
- 动态PL/SQL块允许出现重复的占位符, 即相同占位符只能与USING子句的一个bind\_argument按位置对应。

#### 示例

```
--从动态语句检索值 ( INTO 子句 ) :
DECLARE
  staff_count VARCHAR2(20);
BEGIN
  EXECUTE IMMEDIATE 'select count(*) from staffs'
    INTO staff_count;
  dbms_output.put_line(staff_count);
END;
/

--传递并检索值 ( INTO子句用在USING子句前 ) :
CREATE OR REPLACE PROCEDURE dynamic_proc
AS
  staff_id   NUMBER(6) := 200;
  first_name VARCHAR2(20);
  salary     NUMBER(8,2);
BEGIN
  EXECUTE IMMEDIATE 'select first_name, salary from staffs where staff_id = :1'
    INTO first_name, salary
    USING IN staff_id;
  dbms_output.put_line(first_name || ' ' || salary);
END;
/

--调用存储过程
CALL dynamic_proc();

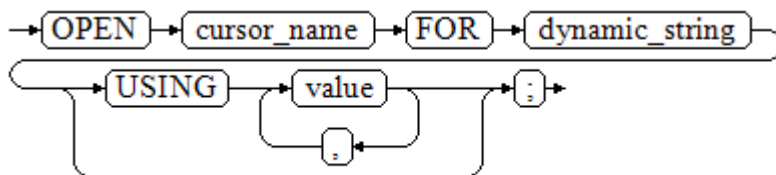
--删除存储过程
DROP PROCEDURE dynamic_proc;
```

## OPEN FOR

动态查询语句还可以使用OPEN FOR打开动态游标来执行。

语法参见图14-8。

图 14-8 open\_for::=



参数说明：

- cursor\_name：要打开的游标名。
- dynamic\_string：动态查询语句。
- USING value：在dynamic\_string中存在占位符时使用。

游标的使用请参考[游标](#)。

### 示例

```

DECLARE
  name      VARCHAR2(20);
  phone_number VARCHAR2(20);
  salary    NUMBER(8,2);
  sqlstr    VARCHAR2(1024);

  TYPE app_ref_cur_type IS REF CURSOR; --定义游标类型
  my_cur app_ref_cur_type; --定义游标变量

BEGIN
  sqlstr := 'select first_name,phone_number,salary from staffs
            where section_id = :1';
  OPEN my_cur FOR sqlstr USING '30'; --打开游标, using是可选的
  FETCH my_cur INTO name, phone_number, salary; --获取数据
  WHILE my_cur%FOUND LOOP
    dbms_output.put_line(name||'#'||phone_number||'#'||salary);
    FETCH my_cur INTO name, phone_number, salary;
  END LOOP;
  CLOSE my_cur; --关闭游标
END;
/

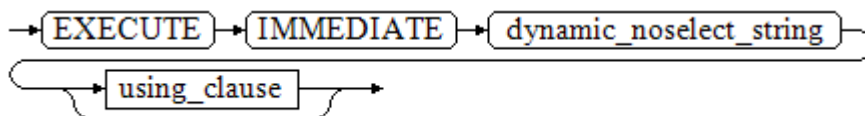
```

## 14.7.2 执行动态非查询语句

### 语法

语法请参见[图14-9](#)。

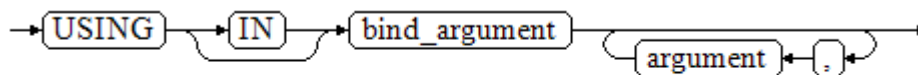
图 14-9 noselect::=



using\_clause子句的语法参见[图14-10](#)。



图 14-10 using\_clause-2



对以上语法格式的解释如下：

USING IN bind\_argument用于指定存放传递给动态SQL值的变量，在dynamic\_noselect\_string中存在占位符时使用，即动态SQL语句执行时，bind\_argument将替换相对应的占位符。要注意的是，bind\_argument只能是值、变量或表达式，不能是表名、列名、数据类型等数据库对象。如果存储过程需要通过声明参数传递数据库对象来构造动态SQL语句（常见于执行DDL语句时），建议采用连接运算符“||”拼接dynamic\_select\_clause。另外，动态语句允许出现重复的占位符，相同占位符只能与唯一一个bind\_argument按位置一一对应。

## 示例

```
--创建表
CREATE TABLE sections_t1
(
  section      NUMBER(4) ,
  section_name VARCHAR2(30),
  manager_id   NUMBER(6),
  place_id     NUMBER(4)
)
DISTRIBUTE BY hash(manager_id);

--声明变量
DECLARE
  section      NUMBER(4) := 280;
  section_name VARCHAR2(30) := 'Info support';
  manager_id   NUMBER(6) := 103;
  place_id     NUMBER(4) := 1400;
  new_colname  VARCHAR2(10) := 'sec_name';
BEGIN
--执行查询
  EXECUTE IMMEDIATE 'insert into sections_t1 values(:1, :2, :3, :4)'
    USING section, section_name, manager_id, place_id;
--执行查询（重复占位符）
  EXECUTE IMMEDIATE 'insert into sections_t1 values(:1, :2, :3, :1)'
    USING section, section_name, manager_id;
--执行ALTER语句（建议采用“||”拼接数据库对象构造DDL语句）
  EXECUTE IMMEDIATE 'alter table sections_t1 rename section_name to ' || new_colname;
END;
/

--查询数据
SELECT * FROM sections_t1;

--删除表
DROP TABLE sections_t1;
```

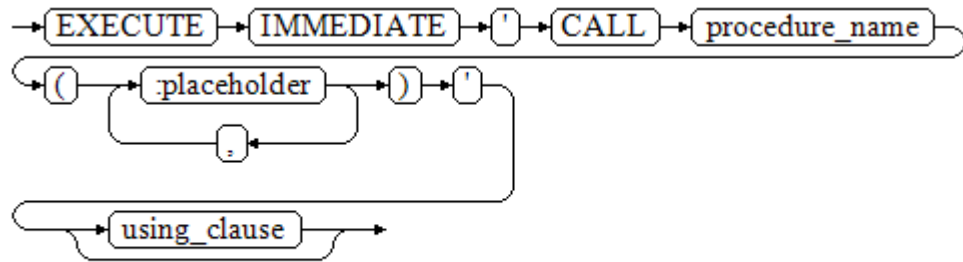
### 14.7.3 动态调用存储过程

动态调用存储过程必须使用匿名的语句块将存储过程或语句块包在里面，使用EXECUTE IMMEDIATE...USING语句后面带IN、OUT来输入、输出参数。

## 语法

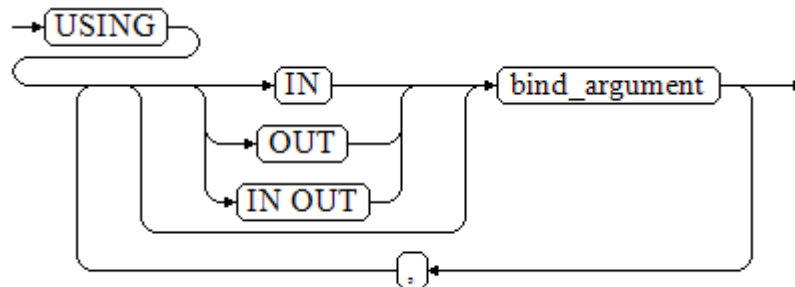
语法请参见图14-11。

图 14-11 call\_procedure::=



using\_clause子句的语法参见图14-12。

图 14-12 using\_clause-3



对以上语法格式的解释如下：

- CALL procedure\_name，调用存储过程。
- [:placeholder1, :placeholder2, …]，存储过程参数占位符列表。占位符个数与参数个数相同。
- USING [IN|OUT|IN OUT] bind\_argument，用于指定存放传递给存储过程参数值的变量。bind\_argument前的修饰符与对应参数的修饰符一致。

## 示例

```
--创建存储过程proc_add。
CREATE OR REPLACE PROCEDURE proc_add
(
    param1 in INTEGER,
    param2 out INTEGER,
    param3 in INTEGER
)
AS
BEGIN
    param2:= param1 + param3;
END;
/

DECLARE
    input1 INTEGER:=1;
    input2 INTEGER:=2;
    statement VARCHAR2(200);
    param2 INTEGER;
BEGIN
    --声明调用语句
```

```
statement := 'call proc_add(:col_1, :col_2, :col_3)';
--执行语句
EXECUTE IMMEDIATE statement
    USING IN input1, OUT param2, IN input2;
dbms_output.put_line('result is: '||to_char(param2));
END;
/

--删除存储过程
DROP PROCEDURE proc_add;
```

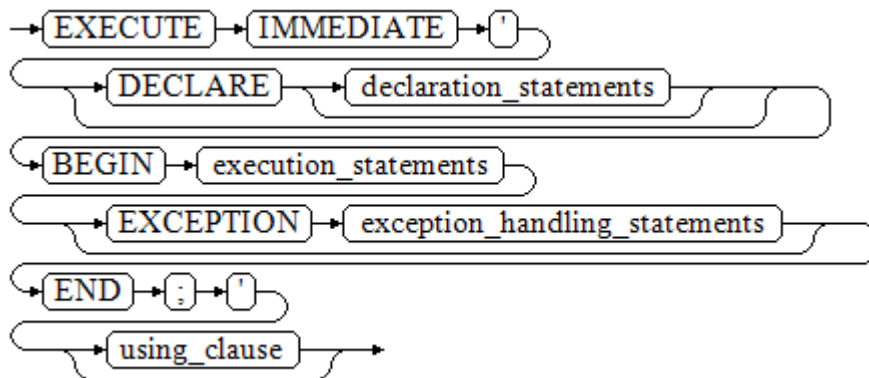
## 14.7.4 动态调用匿名块

动态调用匿名块是指在动态语句中执行匿名块，使用EXECUTE IMMEDIATE…USING语句后面带IN、OUT来输入、输出参数。

### 语法

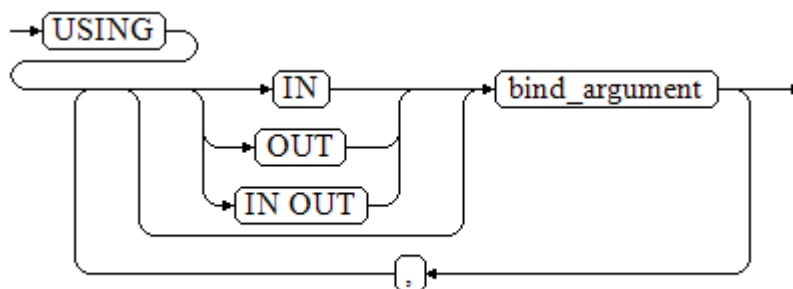
语法请参见图14-13。

图 14-13 call\_anonymous\_block::=



using\_clause子句的语法参见图14-14。

图 14-14 using\_clause-4



对以上语法格式的解释如下：

- 匿名块程序实施部分，以BEGIN语句开始，以END语句停顿，以一个分号结束。
- USING [IN|OUT|IN OUT] bind\_argument，用于指定存放传递给存储过程参数值的变量。bind\_argument前的修饰符与对应参数的修饰符一致。

- 匿名块中间的输入输出参数使用占位符来指明，要求占位符个数与参数个数相同，并且占位符所对应参数的顺序和USING中参数的顺序一致。
- 目前GaussDB(DWS)在动态语句调用匿名块时，EXCEPTION语句中暂不支持使用占位符进行输入输出参数的传递。

## 示例

```
--创建存储过程dynamic_proc
CREATE OR REPLACE PROCEDURE dynamic_proc
AS
  staff_id   NUMBER(6) := 200;
  first_name VARCHAR2(20);
  salary     NUMBER(8,2);
BEGIN
  --执行匿名块
  EXECUTE IMMEDIATE 'begin select first_name, salary into :first_name, :salary from staffs where
staff_id= :dno; end;'
  USING OUT first_name, OUT salary, IN staff_id;
  dbms_output.put_line(first_name|| ' ' || salary);
END;
/

--调用存储过程
CALL dynamic_proc();

--删除存储过程
DROP PROCEDURE dynamic_proc;
```

## 14.8 控制语句

### 14.8.1 返回语句

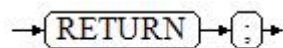
GaussDB(DWS)提供两种方式返回数据：RETURN或RETURN NEXT及RETURN QUERY。其中，RETURN NEXT和RETURN QUERY只适用于函数，不适用存储过程。

#### 14.8.1.1 RETURN

### 语法

返回语句的语法请参见图14-15。

图 14-15 return\_clause::=



对以上语法的解释如下：

用于将控制从存储过程或函数返回给调用者。

### 示例

请参见调用语句的示例。

## 14.8.1.2 RETURN NEXT 及 RETURN QUERY

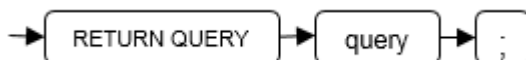
### 语法

创建函数时需要指定返回值SETOF datatype。

return\_next\_clause::=



return\_query\_clause::=



对以上语法的解释如下：

当需要函数返回一个集合时，使用RETURN NEXT或者RETURN QUERY向结果集追加结果，然后继续执行函数的下一条语句。随着后续的RETURN NEXT或RETURN QUERY命令的执行，结果集中会有多个结果。函数执行完成后会一起返回所有结果。

RETURN NEXT可用于标量和复合数据类型。

RETURN QUERY有一种变体RETURN QUERY EXECUTE，后面还可以增加动态查询，通过USING向查询插入参数。

### 示例

```
CREATE TABLE t1(a int);
INSERT INTO t1 VALUES(1),(10);

--RETURN NEXT
CREATE OR REPLACE FUNCTION fun_for_return_next() RETURNS SETOF t1 AS $$
DECLARE
  r t1%ROWTYPE;
BEGIN
  FOR r IN select * from t1
  LOOP
    RETURN NEXT r;
  END LOOP;
  RETURN;
END;
$$ LANGUAGE PLPGSQL;
call fun_for_return_next();
a
---
1
10
(2 rows)

-- RETURN QUERY
CREATE OR REPLACE FUNCTION fun_for_return_query() RETURNS SETOF t1 AS $$
DECLARE
  r t1%ROWTYPE;
BEGIN
  RETURN QUERY select * from t1;
END;
$$
language plpgsql;
call fun_for_return_next();
a
---
```

```
1
10
(2 rows)
```

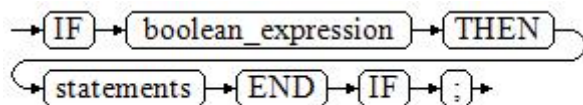
## 14.8.2 条件语句

条件语句的主要作用判断参数或者语句是否满足已给定的条件，根据判定结果执行相应的操作。

GaussDB(DWS)有五种形式的IF：

- IF\_THEN

图 14-16 IF\_THEN::=



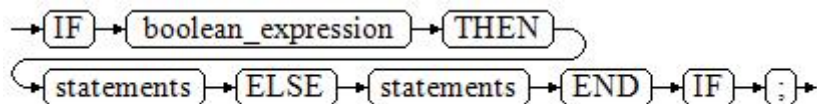
IF\_THEN语句是IF的最简单形式。如果条件为真，statements将被执行。否则，将忽略它们的结果使该IF\_THEN语句执行结束。

### 示例

```
IF v_user_id <> 0 THEN
  UPDATE users SET email = v_email WHERE user_id = v_user_id;
END IF;
```

- IF\_THEN\_ELSE

图 14-17 IF\_THEN\_ELSE::=



IF\_THEN\_ELSE语句增加了ELSE的分支，可以声明在条件为假的时候执行的语句。

### 示例

```
IF parentid IS NULL OR parentid = ''
THEN
  RETURN;
ELSE
  hp_true_filename(parentid);--表示调用存储过程
END IF;
```

- IF\_THEN\_ELSE IF

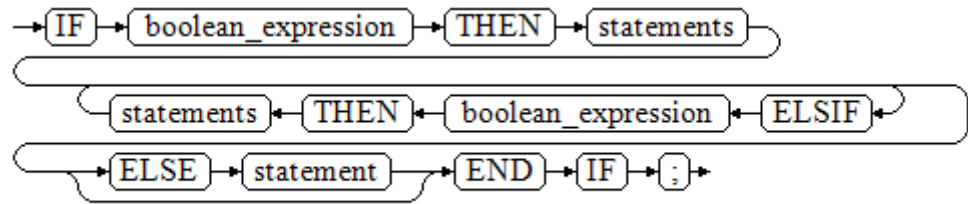
IF语句可以嵌套，嵌套方式如下：

```
IF sex = 'm' THEN
  pretty_sex := 'man';
ELSE
  IF sex = 'f' THEN
    pretty_sex := 'woman';
  END IF;
END IF;
```

这种形式实际上就是在一个IF语句的ELSE部分嵌套了另一个IF语句。因此需要一个END IF语句给每个嵌套的IF，另外还需要一个END IF语句结束父IF-ELSE。如果有多个选项，可使用下面的形式。

- IF\_THEN\_ELSIF\_ELSE

图 14-18 IF\_THEN\_ELSIF\_ELSE::=



### 示例

```
IF number_tmp = 0 THEN
    result := 'zero';
ELSIF number_tmp > 0 THEN
    result := 'positive';
ELSIF number_tmp < 0 THEN
    result := 'negative';
ELSE
    result := 'NULL';
END IF;
```

- IF\_THEN\_ELSEIF\_ELSE

ELSEIF是ELSIF的别名。

### 综合示例

```
CREATE OR REPLACE PROCEDURE proc_control_structure(i in integer)
AS
BEGIN
    IF i > 0 THEN
        raise info 'i:% is greater than 0. ',i;
    ELSIF i < 0 THEN
        raise info 'i:% is smaller than 0. ',i;
    ELSE
        raise info 'i:% is equal to 0. ',i;
    END IF;
    RETURN;
END;
/

CALL proc_control_structure(3);

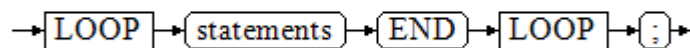
--删除存储过程
DROP PROCEDURE proc_control_structure;
```

## 14.8.3 循环语句

### 简单 LOOP 语句

#### 语法图

图 14-19 loop::=



## 示例

```
CREATE OR REPLACE PROCEDURE proc_loop(i in integer, count out integer)
AS
BEGIN
    count:=0;
    LOOP
        IF count > i THEN
            raise info 'count is %.', count;
            EXIT;
        ELSE
            count:=count+1;
        END IF;
    END LOOP;
END;
/

CALL proc_loop(10,5);
```

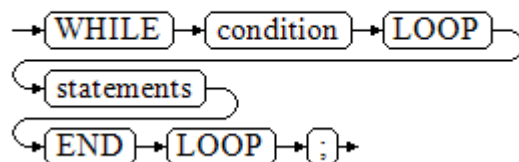
## 须知

该循环必须要结合EXIT使用，否则将陷入死循环。

## WHILE\_LOOP 语句

### 语法图

图 14-20 while\_loop::=



只要条件表达式为真，WHILE语句就会不停的在一系列语句上进行循环，在每次进入循环体的时候进行条件判断。

## 示例

```
CREATE TABLE integertable(c1 integer) DISTRIBUTE BY hash(c1);
CREATE OR REPLACE PROCEDURE proc_while_loop(maxval in integer)
AS
DECLARE
    i int :=1;
BEGIN
    WHILE i < maxval LOOP
        INSERT INTO integertable VALUES(i);
        i:=i+1;
    END LOOP;
END;
/

--调用函数
CALL proc_while_loop(10);

--删除存储过程和表
```

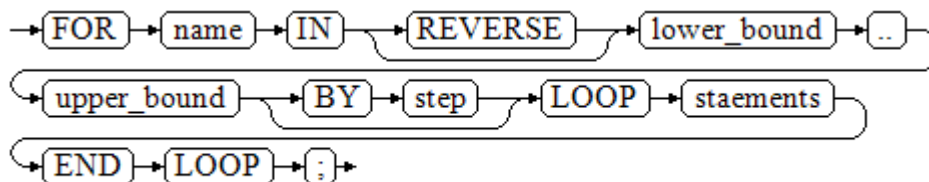


```
DROP PROCEDURE proc_while_loop;
DROP TABLE integertable;
```

## FOR\_LOOP ( integer 变量 ) 语句

### 语法图

图 14-21 for\_loop::=



### 说明

- 变量name会自动定义为integer类型并且只在此循环里存在。变量name介于lower\_bound和upper\_bound之间。
- 当使用REVERSE关键字时，lower\_bound必须大于等于upper\_bound，否则循环体不会被执行。

### 示例

```
--从0到5进行循环
CREATE OR REPLACE PROCEDURE proc_for_loop()
AS
BEGIN
  FOR I IN 0..5 LOOP
    DBMS_OUTPUT.PUT_LINE('It is '||to_char(I) || ' time;');
  END LOOP;
END;
/

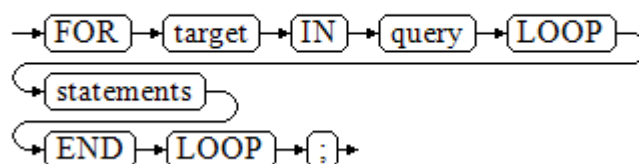
--调用函数
CALL proc_for_loop();

--删除存储过程
DROP PROCEDURE proc_for_loop;
```

## FOR\_LOOP 查询语句

### 语法图

图 14-22 for\_loop\_query::=



### 说明

变量target会自动定义，类型和query的查询结果的类型一致，并且只在此循环中有效。target的取值就是query的查询结果。

### 示例

```
--循环输出查询结果。
CREATE OR REPLACE PROCEDURE proc_for_loop_query()
AS
    record VARCHAR2(50);
BEGIN
    FOR record IN SELECT spcname FROM pg_tablespace LOOP
        dbms_output.put_line(record);
    END LOOP;
END;
/

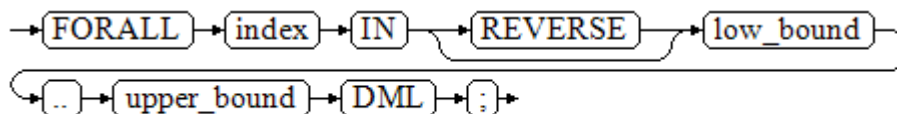
--调用函数
CALL proc_for_loop_query();

--删除存储过程
DROP PROCEDURE proc_for_loop_query;
```

## FORALL 批量查询语句

### 语法图

图 14-23 forall::=



### 说明

变量index会自动定义为integer类型并且只在此循环里存在。index的取值介于low\_bound和upper\_bound之间。

### 示例

```
CREATE TABLE hdfs_t1 (
    title NUMBER(6),
    did VARCHAR2(20),
    data_peroid VARCHAR2(25),
    kind VARCHAR2(25),
    interval VARCHAR2(20),
    time DATE,
    isModified VARCHAR2(10)
)
DISTRIBUTE BY hash(did);

INSERT INTO hdfs_t1 VALUES( 8, 'Donald', 'OConnell', 'DOCONNEL', '650.507.9833', to_date('21-06-1999',
'dd-mm-yyyy'), 'SH_CLERK' );

CREATE OR REPLACE PROCEDURE proc_forall()
AS
BEGIN
    FORALL i IN 100..120
        insert into hdfs_t1(title) values(i);
END;
/
```

```
--调用函数
CALL proc_forall();

--查询存储过程调用结果
SELECT * FROM hdfs_t1 WHERE title BETWEEN 100 AND 120;

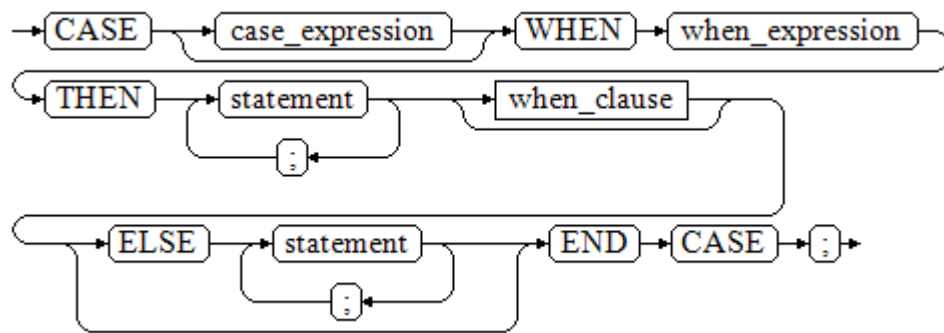
--删除存储过程和表
DROP PROCEDURE proc_forall;
DROP TABLE hdfs_t1;
```

## 14.8.4 分支语句

### 语法

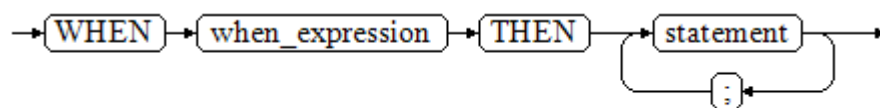
分支语句的语法请参见图14-24。

图 14-24 case\_when::=



when\_clause子句的语法图参见图14-25。

图 14-25 when\_clause::=



参数说明：

- case\_expression: 变量或表达式。
- when\_expression: 常量或者条件表达式。
- statement: 执行语句。

### 示例

```
CREATE OR REPLACE PROCEDURE proc_case_branch(pi_result in integer, pi_return out integer)
AS
BEGIN
CASE pi_result
WHEN 1 THEN
pi_return := 111;
WHEN 2 THEN
pi_return := 222;
```

```
WHEN 3 THEN
    pi_return := 333;
WHEN 6 THEN
    pi_return := 444;
WHEN 7 THEN
    pi_return := 555;
WHEN 8 THEN
    pi_return := 666;
WHEN 9 THEN
    pi_return := 777;
WHEN 10 THEN
    pi_return := 888;
ELSE
    pi_return := 999;
END CASE;
raise info 'pi_return : %',pi_return ;
END;
/

CALL proc_case_branch(3,0);

--删除存储过程
DROP PROCEDURE proc_case_branch;
```

## 14.8.5 空语句

在PL/SQL程序中，可以用NULL语句来说明“不用做任何事情”，相当于一个占位符，可以使某些语句变得有意义，提高程序的可读性。

### 语法

空语句的用法如下：

```
DECLARE
...
BEGIN
...
    IF v_num IS NULL THEN
        NULL; -- 不需要处理任何数据。
    END IF;
END;
/
```

## 14.8.6 错误捕获语句

缺省时，当PL/SQL函数执行过程中发生错误时退出函数执行，并且周围的事务也会回滚。可以用一个带有EXCEPTION子句的BEGIN块捕获错误并且从中恢复。其语法是正常的BEGIN块语法的一个扩展：

```
[<<label>>]
[DECLARE
    declarations]
BEGIN
    statements
EXCEPTION
    WHEN condition [OR condition ...] THEN
        handler_statements
    [WHEN condition [OR condition ...] THEN
        handler_statements
    ...]
END;
```

如果没有发生错误，这种形式的块只是简单地执行所有语句，然后转到END之后的下一个语句。但是如果在执行的语句内部发生了一个错误，则这个语句将会回滚，然后转到EXCEPTION列表。寻找匹配错误的第一个条件。若找到匹配，则执行对应的

handler\_statements，然后转到END之后的下一个语句。如果没有找到匹配，则会向事务的外层报告错误，和没有EXCEPTION子句一样。

也就是说该错误可以被一个包围块用EXCEPTION捕获，如果没有包围块，则进行退出函数处理。

condition的名字遵循SQL标准错误码编号说明的任意值。特殊的条件名OTHERS匹配除了QUERY\_CANCELED之外的所有错误类型。

如果在选中的handler\_statements里发生了新错误，则不能被这个EXCEPTION子句捕获，而是向事务的外层报告错误。一个外层的EXCEPTION子句可以捕获它。

如果一个错误被EXCEPTION捕获，PL/SQL函数的局部变量保持错误发生时的原值，但是所有该块中想写入数据库中的状态都回滚。

示例：

```
CREATE TABLE mytab(id INT,firstname VARCHAR(20),lastname VARCHAR(20)) DISTRIBUTE BY hash(id);

INSERT INTO mytab(firstname, lastname) VALUES('Tom', 'Jones');

CREATE FUNCTION fun_exp() RETURNS INT
AS $$
DECLARE
  x INT :=0;
  y INT;
BEGIN
  UPDATE mytab SET firstname = 'Joe' WHERE lastname = 'Jones';
  x := x + 1;
  y := x / 0;
EXCEPTION
  WHEN division_by_zero THEN
    RAISE NOTICE 'caught division_by_zero';
    RETURN x;
END;$$
LANGUAGE plpgsql;

call fun_exp();
NOTICE: caught division_by_zero
 fun_exp
-----
      1
(1 row)

select * from mytab;
 id | firstname | lastname
-----+-----+-----
   1 | Tom       | Jones
(1 row)

DROP FUNCTION fun_exp();
DROP TABLE mytab;
```

当控制到达给y赋值的的地方时，会有一个division\_by\_zero错误失败。这个错误将被EXCEPTION子句捕获。而在RETURN语句里返回的数值将是x的增量值。

### 📖 说明

进入和退出一个包含EXCEPTION子句的块要比不包含的块开销大的多。因此，不必要的时候不要使用EXCEPTION。

在下列场景中，无法捕获处理异常，整个存储过程回滚：节点故障、网络故障引起的存储过程参与节点线程退出以及COPY FROM操作中源数据与目标表的表结构不一致造成的异常。

示例：UPDATE/INSERT异常

这个例子根据使用异常处理器执行恰当的UPDATE或INSERT。

```
CREATE TABLE db (a INT, b TEXT);

CREATE FUNCTION merge_db(key INT, data TEXT) RETURNS VOID AS
$$
BEGIN
    LOOP
        --第一次尝试更新key
        UPDATE db SET b = data WHERE a = key;
        IF found THEN
            RETURN;
        END IF;
        --不存在，所以尝试插入key，如果其他人同时插入相同的key，可能得到唯一key失败。
        BEGIN
            INSERT INTO db(a,b) VALUES (key, data);
            RETURN;
        EXCEPTION WHEN unique_violation THEN
            --什么也不做，并且循环尝试再次更新。
            END;
        END LOOP;
    END;
$$
LANGUAGE plpgsql;

SELECT merge_db(1, 'david');
SELECT merge_db(1, 'dennis');

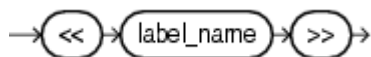
--删除FUNCTION和TABLE
DROP FUNCTION merge_db;
DROP TABLE db ;
```

## 14.8.7 GOTO 语句

GOTO语句可以实现从GOTO位置到目标语句的无条件跳转。GOTO语句会改变原本的执行逻辑，因此应该慎重使用，或者也可以使用EXCEPTION处理特殊场景。当执行GOTO语句时，目标Label必须是唯一的。

### 语法

label declaration ::=



goto statement ::=



### 示例

```
CREATE OR REPLACE PROCEDURE GOTO_test()
AS
DECLARE
    v1 int;
BEGIN
    v1 := 0;
    LOOP
        EXIT WHEN v1 > 100;
        v1 := v1 + 2;
        if v1 > 25 THEN
            GOTO pos1;
        END IF;
    END LOOP;
<<pos1>>
```

```
v1 := v1 + 10;
raise info 'v1 is %.', v1;
END;
/

call GOTO_test();
DROP PROCEDURE GOTO_test();
```

## 限制场景

GOTO使用有以下限制场景

- 不支持有多个相同的GOTO labels目标场景，无论是否在同一个block中。

```
BEGIN
  GOTO pos1;
  <<pos1>>
  SELECT * FROM ...
  <<pos1>>
  UPDATE t1 SET ...
END;
```

- 不支持GOTO跳转到IF语句，CASE语句，LOOP语句中。

```
BEGIN
  GOTO pos1;
  IF valid THEN
    <<pos1>>
    SELECT * FROM ...
  END IF;
END;
```

- 不支持GOTO语句从一个IF子句跳转到另一个IF子句，或从一个CASE语句的WHEN子句跳转到另一个WHEN子句。

```
BEGIN
  IF valid THEN
    GOTO pos1;
    SELECT * FROM ...
  ELSE
    <<pos1>>
    UPDATE t1 SET ...
  END IF;
END;
```

- 不支持从外部块跳转到内部的BEGIN-END块。

```
BEGIN
  GOTO pos1;
  BEGIN
    <<pos1>>
    UPDATE t1 SET ...
  END;
END;
```

- 不支持从异常处理部分跳转到当前的BEGIN-END块。但可以跳转到上层BEGIN-END块。

```
BEGIN
  <<pos1>>
  UPDATE t1 SET ...
  EXCEPTION
    WHEN condition THEN
      GOTO pos1;
END;
```

- 如果从GOTO到一个不包含执行语句的位置，需要添加NULL语句。

```
DECLARE
  done BOOLEAN;
BEGIN
  FOR i IN 1..50 LOOP
    IF done THEN
      GOTO end_loop;
    END IF;
```

```
<<end_loop>> -- not allowed unless an executable statement follows
NULL; -- add NULL statement to avoid error
END LOOP; -- raises an error without the previous NULL
END;
/
```

## 14.9 其他语句

### 14.9.1 锁操作

GaussDB(DWS)提供了多种锁模式用于控制对表中数据的并发访问。这些模式可以在MVCC（多版本并发控制）无法给出期望行为的场合。同样，大多数GaussDB(DWS)命令自动施加恰当的锁，以保证被引用的表在命令的执行过程中不会以一种不兼容的方式被删除或者修改。比如，在存在其他并发操作的时候，ALTER TABLE是不能在同一个表上执行的。

### 14.9.2 游标操作

GaussDB(DWS)中游标（cursor）是系统为用户开设的一个数据缓冲区，存放着SQL语句的执行结果。每个游标区都有一个名字。用户可以用SQL语句逐一从游标中获取记录，并赋给主变量，交由主语言进一步处理。

游标的操作主要有游标的定义、打开、获取和关闭。

完整的游标操作示例可参考[显式游标](#)。

## 14.10 游标

### 14.10.1 游标概述

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。

#### 须知

当游标作为存储过程的返回值时，如果使用JDBC调用该存储过程，返回的游标将不可用。

游标的使用分为显式游标和隐式游标。对于不同的SQL语句，游标的使用情况不同，详细信息请参见[表14-2](#)。

表 14-2 游标使用情况

SQL语句	游标
非查询语句	隐式的
结果是单行的查询语句	隐式的或显式的
结果是多行的查询语句	显式的



## 14.10.2 显式游标

显式游标主要用于对查询语句的处理，尤其是在查询结果为多条记录的情况下。

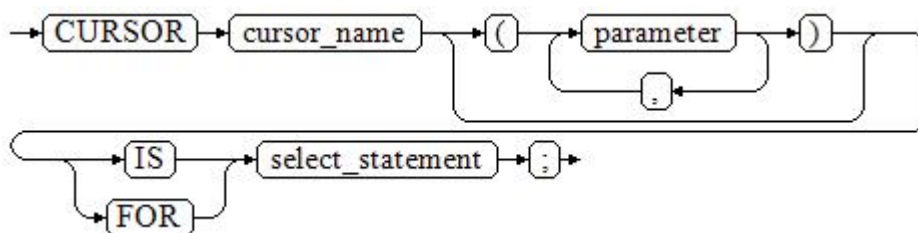
### 处理步骤

显式游标处理需六个PL/SQL步骤：

**步骤1 定义静态游标：**就是定义一个游标名，以及与其相对应的SELECT语句。

定义静态游标的语法图，请参见图14-26。

图 14-26 static\_cursor\_define::=



参数说明：

- cursor\_name：定义的游标名。
- parameter：游标参数，只能为输入参数，其格式为：  
parameter\_name datatype
- select\_statement：查询语句。

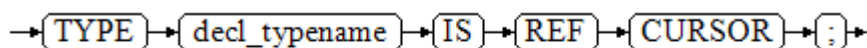
#### 说明

根据执行计划的不同，系统会自动判断该游标是否可以用于以倒序的方式检索数据行。

**定义动态游标：**指ref游标，可以通过一组静态的SQL语句动态的打开游标。首先定义ref游标类型，然后定义该游标类型的游标变量，在打开游标时通过OPEN FOR动态绑定SELECT语句。

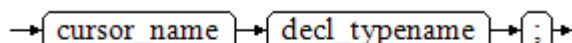
定义动态游标的语法图，请参见图14-27和图14-28。

图 14-27 cursor\_typename::=



GaussDB(DWS)支持sys\_refcursor动态游标类型，函数或存储过程可以通过sys\_refcursor参数传入或传出游标结果集合，函数也可以通过返回sys\_refcursor来返回游标结果集合。

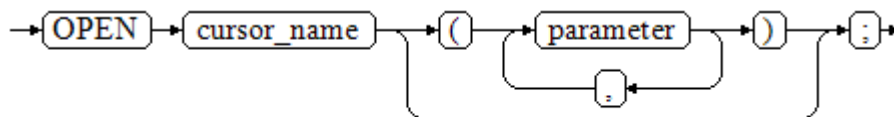
图 14-28 dynamic\_cursor\_define::=



**步骤2 打开静态游标：**就是执行游标所对应的SELECT语句，将其查询结果放入工作区，并且指针指向工作区的首部，标识游标结果集合。如果游标查询语句中带有FOR UPDATE选项，OPEN语句还将锁定数据库表中游标结果集合对应的数据行。

打开静态游标的语法图，请参见图14-29。

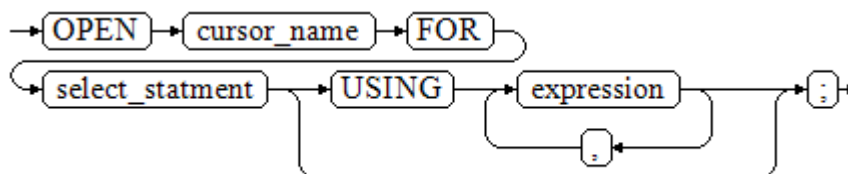
图 14-29 open\_static\_cursor::=



**打开动态游标：**可以通过OPEN FOR语句打开动态游标，动态绑定SQL语句。

打开动态游标的语法图，请参见图14-30。

图 14-30 open\_dynamic\_cursor::=

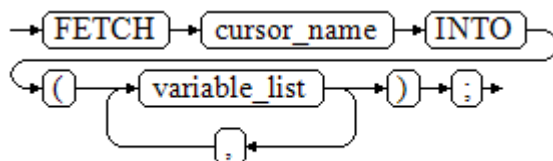


PL/SQL程序不能用OPEN语句重复打开一个游标。

**步骤3 提取游标数据：**检索结果集中的数据行，放入指定的输出变量中。

提取游标数据的语法图，请参见图14-31。

图 14-31 fetch\_cursor::=



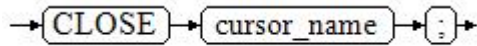
**步骤4** 对该记录进行处理。

**步骤5** 继续处理，直到活动集中没有记录。

**步骤6 关闭游标：**当提取和处理完游标结果集合数据后，应及时关闭游标，以释放该游标所占用的系统资源，并使该游标的工作区变成无效，不能再使用FETCH语句获取其中数据。关闭后的游标可以使用OPEN语句重新打开。

关闭游标的语法图，请参见图14-32。

图 14-32 close\_cursor::=



----结束

## 属性

游标的属性用于控制程序流程或者了解程序的状态。当运行DML语句时，PL/SQL打开一个内建游标并处理结果，游标是维护查询结果的内存中的一个区域，游标在运行DML语句时打开，完成后关闭。显式游标的属性为：

- %FOUND布尔型属性：当最近一次读记录时成功返回，则值为TRUE。
- %NOTFOUND布尔型属性：与%FOUND相反。
- %ISOPEN布尔型属性：当游标已打开时返回TRUE。
- %ROWCOUNT数值型属性：返回已从游标中读取的记录数。

## 示例

```

--游标参数的传递方法。
CREATE OR REPLACE PROCEDURE cursor_proc1()
AS
DECLARE
  DEPT_NAME VARCHAR(100);
  DEPT_LOC NUMBER(4);
  --定义游标
  CURSOR C1 IS
    SELECT section_name, place_id FROM sections WHERE section_id <= 50;
  CURSOR C2(sect_id INTEGER) IS
    SELECT section_name, place_id FROM sections WHERE section_id <= sect_id;
  TYPE CURSOR_TYPE IS REF CURSOR;
  C3 CURSOR_TYPE;
  SQL_STR VARCHAR(100);
BEGIN
  OPEN C1;--打开游标
  LOOP
    --通过游标取值
    FETCH C1 INTO DEPT_NAME, DEPT_LOC;
    EXIT WHEN C1%NOTFOUND;
    DBMS_OUTPUT.PUT_LINE(DEPT_NAME||'---'||DEPT_LOC);
  END LOOP;
  CLOSE C1;--关闭游标

  OPEN C2(10);
  LOOP
    FETCH C2 INTO DEPT_NAME, DEPT_LOC;
    EXIT WHEN C2%NOTFOUND;
    DBMS_OUTPUT.PUT_LINE(DEPT_NAME||'---'||DEPT_LOC);
  END LOOP;
  CLOSE C2;

  SQL_STR := 'SELECT section_name, place_id FROM sections WHERE section_id <= :DEPT_NO;';
  OPEN C3 FOR SQL_STR USING 50;
  LOOP
    FETCH C3 INTO DEPT_NAME, DEPT_LOC;
    EXIT WHEN C3%NOTFOUND;
    DBMS_OUTPUT.PUT_LINE(DEPT_NAME||'---'||DEPT_LOC);
  END LOOP;
  CLOSE C3;
END;
/
  
```

```
CALL cursor_proc1();

DROP PROCEDURE cursor_proc1;
--给工资低于3000的员工增加工资500。
CREATE TABLE staffs_t1 AS TABLE staffs;

CREATE OR REPLACE PROCEDURE cursor_proc2()
AS
DECLARE
    V_EMPNO NUMBER(6);
    V_SAL NUMBER(8,2);
    CURSOR C IS SELECT staff_id, salary FROM staffs_t1;
BEGIN
    OPEN C;
    LOOP
        FETCH C INTO V_EMPNO, V_SAL;
        EXIT WHEN C%NOTFOUND;
        IF V_SAL<=3000 THEN
            UPDATE staffs_t1 SET salary =salary + 500 WHERE staff_id = V_EMPNO;
        END IF;
    END LOOP;
    CLOSE C;
END;
/

CALL cursor_proc2();

--删除存储过程
DROP PROCEDURE cursor_proc2;
DROP TABLE staffs_t1;
--SYS_REFCURSOR类型做为函数参数
CREATE OR REPLACE PROCEDURE proc_sys_ref(O OUT SYS_REFCURSOR)
IS
C1 SYS_REFCURSOR;
BEGIN
OPEN C1 FOR SELECT section_ID FROM sections ORDER BY section_ID;
O := C1;
END;
/

DECLARE
C1 SYS_REFCURSOR;
TEMP NUMBER(4);
BEGIN
proc_sys_ref(C1);
LOOP
    FETCH C1 INTO TEMP;
    DBMS_OUTPUT.PUT_LINE(C1%ROWCOUNT);
    EXIT WHEN C1%NOTFOUND;
END LOOP;
END;
/

--删除存储过程
DROP PROCEDURE proc_sys_ref;
```

### 14.10.3 隐式游标

对于非查询语句，如修改、删除操作，则由系统自动地为这些操作设置游标并创建其工作区，这些由系统隐含创建的游标称为隐式游标，隐式游标的名字为SQL，这是由系统定义的。

#### 简介

对于隐式游标的操作，如定义、打开、取值及关闭操作，都由系统自动地完成，无需用户进行处理。用户只能通过隐式游标的相关属性，来完成相应的操作。在隐式游标

的工作区中，所存放的数据是最新处理的一条SQL语句所包含的数据，与用户自定义的显式游标无关。

格式调用为：SQL%

#### 说明

INSERT, UPDATE, DROP, SELECT语句中不必明确定义游标。

## 属性

隐式游标属性为：

- SQL%FOUND布尔型属性：当最近一次读记录时成功返回，则值为TRUE。
- SQL%NOTFOUND布尔型属性：与%FOUND相反。
- SQL%ROWCOUNT数值型属性：返回已从游标中读取得记录数。
- SQL%ISOPEN布尔型属性：取值总是FALSE。SQL语句执行完毕立即关闭隐式游标。

## 示例

```
--删除EMP表中某部门的所有员工，如果该部门中已没有员工，则在DEPT表中删除该部门。
CREATE TABLE staffs_t1 AS TABLE staffs;
CREATE TABLE sections_t1 AS TABLE sections;

CREATE OR REPLACE PROCEDURE proc_cursor3()
AS
  DECLARE
    V_DEPTNO NUMBER(4) := 100;
  BEGIN
    DELETE FROM staffs WHERE section_ID = V_DEPTNO;
    --根据游标状态做进一步处理
    IF SQL%NOTFOUND THEN
      DELETE FROM sections_t1 WHERE section_ID = V_DEPTNO;
    END IF;
  END;
/

CALL proc_cursor3();

--删除存储过程和临时表
DROP PROCEDURE proc_cursor3;
DROP TABLE staffs_t1;
DROP TABLE sections_t1;
```

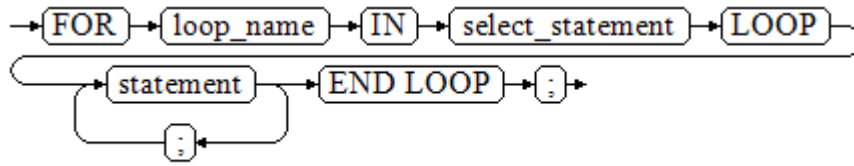
### 14.10.4 游标循环

游标在WHILE语句、LOOP语句中的使用称为游标循环，一般这种循环都需要使用OPEN、FETCH和CLOSE语句。下面要介绍的一种循环不需要这些操作，可以简化游标循环的操作，这种循环方式适用于静态游标的循环，不用执行静态游标的四个步骤。

## 语法

FOR AS循环的语法请参见图14-33。

图 14-33 FOR\_AS\_loop::=



## 注意事项

- 不能在该循环语句中对查询的表进行更新操作。
- 变量loop\_name会自动定义且只在此循环中有效，类型和select\_statement的查询结果类型一致。loop\_name的取值就是select\_statement的查询结果。
- 游标的属性中%FOUND、%NOTFOUND、%ROWCOUNT在GaussDB(DWS)数据库中都是访问同一个内部变量，事务和匿名块不支持多个游标同时访问。

## 示例

```

BEGIN
FOR ROW_TRANS IN
  SELECT first_name FROM staffs
  LOOP
    DBMS_OUTPUT.PUT_LINE (ROW_TRANS.first_name );
  END LOOP;
END;
/

--创建表
CREATE TABLE integerTable1( A INTEGER) DISTRIBUTE BY hash(A);
CREATE TABLE integerTable2( B INTEGER) DISTRIBUTE BY hash(B);
INSERT INTO integerTable2 VALUES(2);

--多游标共享游标属性的标量
DECLARE
  CURSOR C1 IS SELECT A FROM integerTable1;--声明游标
  CURSOR C2 IS SELECT B FROM integerTable2;
  PL_A INTEGER;
  PL_B INTEGER;
BEGIN
  OPEN C1;--打开游标
  OPEN C2;
  FETCH C1 INTO PL_A; ---- C1%FOUND 和 C2%FOUND 值为 FALSE
  FETCH C2 INTO PL_B; ---- C1%FOUND 和 C2%FOUND 的值都为 TRUE
  --判断游标状态
  IF C1%FOUND THEN
    IF C2%FOUND THEN
      DBMS_OUTPUT.PUT_LINE('Dual cursor share parameter. ');
    END IF;
  END IF;
  CLOSE C1;--关闭游标
  CLOSE C2;
END;
/

--删除临时表
DROP TABLE integerTable1;
DROP TABLE integerTable2;
  
```

## 14.11 高级包

## 14.11.1 DBMS\_LOB

### 接口介绍

高级功能包DBMS\_LOB支持的所有接口请参见表14-3。

表 14-3 DBMS\_LOB

接口名称	描述
<b>DBMS_LOB.GETLENGTH</b>	获取并返回指定的LOB类型对象的长度。
<b>DBMS_LOB.OPEN</b>	打开一个LOB返回一个LOB的描述符。
<b>DBMS_LOB.READ</b>	根据指定的长度及起始位置偏移读取LOB内容的一部分到BUFFER缓冲区。
<b>DBMS_LOB.WRITE</b>	根据指定长度及起始位置偏移将BUFFER中内容写入到LOB中。
<b>DBMS_LOB.WRITEAPPEND</b>	根据指定长度将BUFFER中内容写入到LOB的尾部。
<b>DBMS_LOB.COPY</b>	根据指定长度及起始位置偏移将BLOB内容写入到另一个BLOB中。
<b>DBMS_LOB.ERASE</b>	根据指定长度及起始位置偏移删除BLOB中的内容。
<b>DBMS_LOB.CLOSE</b>	关闭已经打开的LOB描述符。
<b>DBMS_LOB.INSTR</b>	返回一个字符串在LOB中第N次出现的位置。
<b>DBMS_LOB.COMPARE</b>	比较两个LOB或者两个LOB的某一部分。
<b>DBMS_LOB.SUBSTR</b>	用于读取一个LOB的子串，返回读取的字节个数或者字符个数。
<b>DBMS_LOB.TRIM</b>	用于截断指定长度的LOB，执行完会将LOB的长度设置为newlen参数指定的长度。
<b>DBMS_LOB.CREATETEMPORARY</b>	创建一个临时的BLOB或者CLOB。
<b>DBMS_LOB.APPEND</b>	将原LOB的内容拼接到目的LOB中。

- **DBMS\_LOB.GETLENGTH**  
存储过程GETLENGTH获取并返回指定的LOB类型对象的长度。

DBMS\_LOB.GETLENGTH函数原型为：

```
DBMS_LOB.GETLENGTH (
lob_loc IN BLOB)
RETURN INTEGER;
```

```
DBMS_LOB.GETLENGTH (
lob_loc IN CLOB)
RETURN INTEGER;
```

表 14-4 DBMS\_LOB.GETLENGTH 接口参数说明

参数	描述
lob_loc	待获取长度的指定的LOB类型的对象。

- DBMS\_LOB.OPEN

存储过程打开一个LOB，并返回一个LOB描述符，该过程无实际意义，仅用于兼容。

DBMS\_LOB.OPEN函数原型为：

```
DBMS_LOB.LOB (
lob_loc INOUT BLOB,
open_mode IN BINARY_INTEGER);
```

```
DBMS_LOB.LOB (
lob_loc INOUT CLOB,
open_mode IN BINARY_INTEGER);
```

表 14-5 DBMS\_LOB.OPEN 接口参数说明

参数	描述
lob_loc	被打开的一个BLOB或者CLOB描述符。
open_mode IN BINARY_INTEG ER	打开的模式，目前支持DBMS_LOB.LOB_READWRITE

- DBMS\_LOB.READ

存储过程READ根据指定长度及起始位置偏移读取LOB内容的一部分到BUFFER缓冲区。

DBMS\_LOB.READ函数原型为：

```
DBMS_LOB.READ (
lob_loc IN BLOB,
amount IN INTEGER,
offset IN INTEGER,
buffer OUT RAW);
```

```
DBMS_LOB.READ (
lob_loc IN CLOB,
amount IN OUT INTEGER,
offset IN INTEGER,
buffer OUT VARCHAR2);
```

表 14-6 DBMS\_LOB.READ 接口参数说明

参数	说明
lob_loc	待读入的指定的LOB类型的对象。
amount	读入长度。 <b>说明</b> 如果读入长度为负，会收到错误提示“ERROR: argument 2 is null, invalid, or out of range.”



参数	说明
offset	指定从LOB内容的哪个位置开始读取的偏移（即相对LOB内容起始位置的字节数）。
buffer	读取LOB内容后存放的目标缓冲区。

- DBMS\_LOB.WRITE

存储过程WRITE根据指定长度及起始位置偏移将BUFFER中内容写入到LOB变量中。

DBMS\_LOB.WRITE函数原型为：

```
DBMS_LOB.WRITE (
lob_loc IN OUT BLOB,
amount IN INTEGER,
offset IN INTEGER,
buffer IN RAW);
```

```
DBMS_LOB.WRITE (
lob_loc IN OUT CLOB,
amount IN INTEGER,
offset IN INTEGER,
buffer IN VARCHAR2);
```

表 14-7 DBMS\_LOB.WRITE 接口参数说明

参数	说明
lob_loc	待写入的指定的LOB类型的对象。
amount	写入长度。 <b>说明</b> 如果写入长度小于1或写入长度大于待写入的内容长度，则报错。
offset	指定从LOB内容的哪个位置开始写入的偏移（即相对LOB内容起始位置的字节数）。 <b>说明</b> 如果偏移量小于1或偏移量大于LOB类型最大长度，则报错。
buffer	待写入的内容。

- DBMS\_LOB.WRITEAPPEND

存储过程WRITEAPPEND根据指定长度将BUFFER中内容写入到LOB的尾部。

DBMS\_LOB.WRITEAPPEND函数原型为：

```
DBMS_LOB.WRITEAPPEND (
lob_loc IN OUT BLOB,
amount IN INTEGER,
buffer IN RAW);
```

```
DBMS_LOB.WRITEAPPEND (
lob_loc IN OUT CLOB,
amount IN INTEGER,
buffer IN VARCHAR2);
```

表 14-8 DBMS\_LOB.WRITEAPPEND 接口参数说明

参数	说明
lob_loc	待写入的指定的LOB类型的对象。
amount	写入长度。 <b>说明</b> 如果写入长度小于1或写入长度大于待写入的内容长度，则报错。
buffer	待写入的内容。

- DBMS\_LOB.COPY

存储过程COPY根据指定长度及起始位置偏移将BLOB内容复制到另一个BLOB中。

DBMS\_LOB.COPY函数原型为：

```
DBMS_LOB.COPY (
dest_lob  IN OUT  BLOB,
src_lob   IN     BLOB,
amount   IN     INTEGER,
dest_offset IN   INTEGER DEFAULT 1,
src_offset IN   INTEGER DEFAULT 1);
```

表 14-9 DBMS\_LOB.COPY 接口参数说明

参数	说明
dest_lob	待拷入的BLOB类型对象。
src_lob	待拷出的BLOB类型对象。
amount	复制长度。 <b>说明</b> 如果拷入长度小于1或拷入长度大于BLOB类型最大长度，则报错。
dest_offset	指定从BLOB内容的哪个位置开始拷入的偏移（即相对BLOB内容起始位置的字节数）。 <b>说明</b> 如果偏移量小于1或偏移量大于BLOB类型最大长度，则报错。
src_offset	指定从BLOB内容的哪个位置开始拷出的偏移（即相对BLOB内容起始位置的字节数）。 <b>说明</b> 如果偏移量小于1或偏移量大于复制来源BLOB的长度，则报错。

- DBMS\_LOB.ERASE

存储过程ERASE根据指定长度及起始位置偏移删除BLOB中的内容。

DBMS\_LOB.ERASE函数原型为：

```
DBMS_LOB.ERASE (
lob_loc   IN OUT  BLOB,
amount   IN OUT  INTEGER,
offset   IN     INTEGER DEFAULT 1);
```

表 14-10 DBMS\_LOB.ERASE 接口参数说明

参数	说明
lob_loc	待删除内容的BLOB类型对象。
amount	待删除的长度。 <b>说明</b> 如果删除长度小于1或删除长度大于BLOB类型最大长度，则报错。
offset	指定从BLOB内容的哪个位置开始删除的偏移（即相对BLOB内容起始位置的字节数）。 <b>说明</b> 如果偏移量小于1或偏移量大于BLOB类型最大长度，则报错。

- DBMS\_LOB.CLOSE

存储过程CLOSE根据指定长度及起始位置偏移关闭已经打开的LOB内容。

DBMS\_LOB.CLOSE函数原型为：

```
DBMS_LOB.CLOSE(
src_loc IN BLOB);

DBMS_LOB.CLOSE (
src_loc IN CLOB);
```

表 14-11 DBMS\_LOB.CLOSE 接口参数说明

参数	说明
src_loc	待关闭的LOB类型对象。

- DBMS\_LOB.INSTR

该函数返回在LOB中第N次出现的位置，如果输入的是一些无效值会返回NULL值。offset < 1 or offset > LOBMAXSIZE, nth < 1, nth > LOBMAXSIZE。

DBMS\_LOB.INSTR函数原型为：

```
DBMS_LOB.INSTR (
lob_loc IN BLOB,
pattern IN RAW,
offset IN INTEGER := 1,
nth IN INTEGER := 1)
RETURN INTEGER;

DBMS_LOB.INSTR (
lob_loc IN CLOB,
pattern IN VARCHAR2 ,
offset IN INTEGER := 1,
nth IN INTEGER := 1)
RETURN INTEGER;
```

表 14-12 DBMS\_LOB.INSTR 接口参数说明

参数	说明
lob_loc	要查找的LOB描述符。

参数	说明
pattern	要匹配的模式，对于BLOB是由一组RAW类型的数据组成，对于CLOB是由一组text类型的数据组成。
offset	对于BLOB是以字节为单位的绝对偏移量，对于CLOB是以字符为单位的偏移量，模式匹配的起始位置是1。
nth	模式匹配的次数，最小值为1。

- DBMS\_LOB.COMPARE

这个函数比较两个LOB或者两个LOB的一部分。

- 如果比较的结果相等返回0，否则返回非零的值。
- 如果第一个CLOB比第二个小，返回-1；如果第一个CLOB比第二个大，返回1。
- 如果amount, offset\_1, offset\_2这几个参数有无效参数返回NULL，有效的偏移量范围是1~LOBMAXSIZE。

DBMS\_LOB.READ函数原型为：

```
DBMS_LOB.COMPARE (
lob_1 IN BLOB,
lob_2 IN BLOB,
amount IN INTEGER := DBMS_LOB.LOBMAXSIZE,
offset_1 IN INTEGER := 1,
offset_2 IN INTEGER := 1)
RETURN INTEGER;

DBMS_LOB.COMPARE (
lob_1 IN CLOB,
lob_2 IN CLOB,
amount IN INTEGER := DBMS_LOB.LOBMAXSIZE,
offset_1 IN INTEGER := 1,
offset_2 IN INTEGER := 1)
RETURN INTEGER;
```

表 14-13 DBMS\_LOB.COMPARE 接口参数说明

参数	说明
lob_1	第一个要比较的LOB描述符。
lob_2	第二个要比较的LOB描述符。
amount	要比较的字符数或者字节数，最大值为DBMS_LOB.LOBMAXSIZE。
offset_1	第一个LOB描述符的偏移量，初始位置是1。
offset_2	第二个LOB描述符的偏移量，初始位置是1。

- DBMS\_LOB.SUBSTR

用于读取一个LOB的子串，返回读取的字节个数或者字符个数，当amount > 1或者amount < 32767, offset < 1或者offset > LOBMAXSIZE的时候返回值是NULL。

DBMS\_LOB.SUBSTR函数原型为：

```

DBMS_LOB.SUBSTR (
lob_loc IN BLOB,
amount IN INTEGER := 32767,
offset IN INTEGER := 1)
RETURN RAW;

DBMS_LOB.SUBSTR (
lob_loc IN CLOB,
amount IN INTEGER := 32767,
offset IN INTEGER := 1)
RETURN VARCHAR2;

```

**表 14-14** DBMS\_LOB.SUBSTR 接口参数说明

参数	说明
lob_loc	将要读取子串的LOB描述符，对于BLOB类型的返回值是读取的字节个数，对于CLOB类型的返回值是字符个数。
offset	要读取的字节数或者字符数量。
buffer	从开始位置偏移的字符数或者字节数量。

- DBMS\_LOB.TRIM

这个存储过程用于截断指定长度的LOB，执行完这个存储过程会将LOB的长度设置为newlen参数指定的长度。如果对一个空的LOB执行截断操作，不会有任何执行结果；如果指定的长度比LOB的长度长，会产生一个异常。

DBMS\_LOB.TRIM函数原型为：

```

DBMS_LOB.TRIM (
lob_loc IN OUT BLOB,
newlen IN INTEGER);

DBMS_LOB.TRIM (
lob_loc IN OUT CLOB,
newlen IN INTEGER);

```

**表 14-15** DBMS\_LOB.TRIM 接口参数说明

参数	说明
lob_loc	待读入的指定BLOB类型的对象。
newlen	截断后LOB的新长度对于BLOB是字节数，对于CLOB是字符数。

- DBMS\_LOB.CREATETEMPORARY

这个存储过程创建一个临时的BLOB或者CLOB，这个存储过程仅用于语法上的兼容，并无实际意义。

DBMS\_LOB.CREATETEMPORARY函数原型为：

```

DBMS_LOB.CREATETEMPORARY (
lob_loc IN OUT BLOB,
cache IN BOOLEAN,
dur IN INTEGER);

DBMS_LOB.CREATETEMPORARY (
lob_loc IN OUT CLOB,
cache IN BOOLEAN,
dur IN INTEGER);

```

**表 14-16** DBMS\_LOB.CREATETEMPORARY 接口参数说明

参数	说明
lob_loc	LOB描述符。
cache	仅用于语法上的兼容。
dur	仅用于语法上的兼容。

- DBMS\_LOB.APPEND

存储过程READ根据指定长度及起始位置偏移读取BLOB内容的一部分到BUFFER缓冲区。

DBMS\_LOB.APPEND函数原型为：

```
DBMS_LOB.APPEND (
dest_lob IN OUT BLOB,
src_lob IN BLOB);
```

```
DBMS_LOB.APPEND (
dest_lob IN OUT CLOB,
src_lob IN CLOB);
```

**表 14-17** DBMS\_LOB.APPEND 接口参数说明

参数	说明
dest_lob	要写入的LOB描述符。
src_lob	读取的LOB描述符。

## 示例

```
--获取字符串的长度
SELECT DBMS_LOB.GETLENGTH('12345678');

DECLARE
myraw RAW(100);
amount INTEGER :=2;
buffer INTEGER :=1;
begin
DBMS_LOB.READ('123456789012345',amount,buffer,myraw);
dbms_output.put_line(myraw);
end;
/

CREATE TABLE blob_Table (t1 blob) DISTRIBUTE BY REPLICATION;
CREATE TABLE blob_Table_bak (t2 blob) DISTRIBUTE BY REPLICATION;
INSERT INTO blob_Table VALUES('abcdef');
INSERT INTO blob_Table_bak VALUES('22222');

DECLARE
str varchar2(100) := 'abcdef';
source raw(100);
dest blob;
copyto blob;
amount int;
PSV_SQL varchar2(100);
PSV_SQL1 varchar2(100);
a int :=1;
len int;
```

```

BEGIN
source := utl_raw.cast_to_raw(str);
amount := utl_raw.length(source);

PSV_SQL := 'select * from blob_Table for update';
PSV_SQL1 := 'select * from blob_Table_bak for update';

EXECUTE IMMEDIATE PSV_SQL into dest;
EXECUTE IMMEDIATE PSV_SQL1 into copyto;

DBMS_LOB.WRITE(dest, amount, 1, source);
DBMS_LOB.WRITEAPPEND(dest, amount, source);

DBMS_LOB.ERASE(dest, a, 1);
DBMS_OUTPUT.PUT_LINE(a);
DBMS_LOB.COPY(copyto, dest, amount, 10, 1);
DBMS_LOB.CLOSE(dest);
RETURN;
END;
/

--删除表
DROP TABLE blob_Table;
DROP TABLE blob_Table_bak;
    
```

## 14.11.2 DBMS\_RANDOM

### 接口介绍

高级功能包DBMS\_RANDOM支持的所有接口请参见[表14-18](#)。

表 14-18 DBMS\_RANDOM 接口参数说明

接口名称	描述
<b>DBMS_RANDOM.SEED</b>	设置一个随机数的种子。
<b>DBMS_RANDOM.VALUE</b>	生成一个大小介于指定的low及high之间的随机数。

- **DBMS\_RANDOM.SEED**  
存储过程SEED用于设置一个随机数的种子。DBMS\_RANDOM.SEED函数原型为：  
DBMS\_RANDOM.SEED (seed IN INTEGER);

表 14-19 DBMS\_RANDOM.SEED 接口参数说明

参数	描述
seed	用于产生一个随机数的种子。

- **DBMS\_RANDOM.VALUE**  
存储过程VALUE生成一个大小介于指定的low及high之间的随机数。  
DBMS\_RANDOM.VALUE函数原型为：  
DBMS\_RANDOM.VALUE(  
low IN NUMBER,

```
high IN NUMBER)
RETURN NUMBER;
```

表 14-20 DBMS\_RANDOM.VALUE 接口参数说明

参数	描述
low	指定随机数大小的下边界，生成的随机数大于或等于low。
high	指定随机数大小的上边界，生成的随机数小于high。

### 说明

实际上，只要求这里的参数类型是NUMERIC即可，对于左右边界的大小并没有要求。

## 示例

```
--产生0到1之间的随机数:
SELECT DBMS_RANDOM.VALUE(0,1);

--对于指定范围内的整数，要加入参数low和high，并从结果中截取较小的数（最大值不能被作为可能的值）。所以
对于0到99之间的整数，使用下面的代码:
SELECT TRUNC(DBMS_RANDOM.VALUE(0,100));
```

## 14.11.3 DBMS\_OUTPUT

### 接口介绍

高级功能包DBMS\_OUTPUT支持的所有接口请参见表14-21。

表 14-21 DBMS\_OUTPUT

接口名称	描述
<b>DBMS_OUTPUT.PUT_LINE</b>	输出指定的文本，文本长度不能超过32767字节。
<b>DBMS_OUTPUT.PUT</b>	将指定的文本输出到指定文本的前面，不添加换行符，文本长度不能超过32767字节。
<b>DBMS_OUTPUT.ENABLE</b>	设置输出缓冲区的大小。若不指定，缓冲区最大只能容纳20000字节，缓冲区最小可设置为2000字节，若设置小于2000字节将按2000字节处理。

- DBMS\_OUTPUT.PUT\_LINE

存储过程PUT\_LINE向消息缓冲区写入一行带有行结束符的文本。  
DBMS\_OUTPUT.PUT\_LINE函数原型为：

```
DBMS_OUTPUT.PUT_LINE (
item IN VARCHAR2);
```



表 14-22 DBMS\_OUTPUT.PUT\_LINE 接口参数说明

参数	描述
item	写入消息缓冲区的文本。

- DBMS\_OUTPUT.PUT

存储过程PUT将指定的文本输出到指定文本的前面，不添加换行符。  
DBMS\_OUTPUT.PUT函数原型为：

```
DBMS_OUTPUT.PUT (
item IN VARCHAR2);
```

表 14-23 DBMS\_OUTPUT.PUT 接口参数说明

参数	描述
item	写入指定文本前的文本。

- DBMS\_OUTPUT.ENABLE

存储过程ENABLE设置输出缓冲区的大小，如果不指定的话缓冲区最大只能容纳20000字节。DBMS\_OUTPUT.ENABLE函数原型为：

```
DBMS_OUTPUT.ENABLE (
buf IN INTEGER);
```

表 14-24 DBMS\_OUTPUT.ENABLE 接口参数说明

参数	描述
buf	设置输出缓冲区的大小。

## 示例

```
BEGIN
  DBMS_OUTPUT.ENABLE(50);
  DBMS_OUTPUT.PUT ('hello, ');
  DBMS_OUTPUT.PUT_LINE('database!');--输出hello, database!
END;
/
```

## 14.11.4 UTL\_RAW

### 接口介绍

高级功能包UTL\_RAW支持的所有接口请参见[表14-25](#)。

表 14-25 UTL\_RAW

接口名称	描述
<code>UTL_RAW.CAST_FROM_BINARY_INTEGER</code>	将INTEGER类型值转换为二进制表示形式（即RAW类型）。
<code>UTL_RAW.CAST_TO_BINARY_INTEGER</code>	将二进制表示形式的整型值（即RAW类型）转换为INTEGER类型。
<code>UTL_RAW.LENGTH</code>	获取RAW类型对象的长度。
<code>UTL_RAW.CAST_TO_RAW</code>	将VARCHAR2类型值转化为二进制表示形式（即RAW类型）。

### 须知

RAW类型的外部表现形式是十六进制，内部存储形式是二进制。例如一个RAW类型的数据11001011的表现形式为‘CB’，即在实际的类型转换中输入的是‘CB’。

- `UTL_RAW.CAST_FROM_BINARY_INTEGER`  
存储过程`CAST_FROM_BINARY_INTEGER`将INTEGER类型值转换为二进制表示形式（即RAW类型）。

`UTL_RAW.CAST_FROM_BINARY_INTEGER`函数原型为：

```
UTL_RAW.CAST_FROM_BINARY_INTEGER (
n          IN INTEGER,
endianess IN INTEGER)
RETURN RAW;
```

表 14-26 UTL\_RAW.CAST\_FROM\_BINARY\_INTEGER 接口参数说明

参数	描述
n	待转成RAW类型的整型数值。
endianess	表示字节序的整型值1或2（1代表BIG_ENDIAN，2代表LITTLE-ENDIAN）。

- `UTL_RAW.CAST_TO_BINARY_INTEGER`  
存储过程`CAST_TO_BINARY_INTEGER`将二进制表示形式的整型值（即RAW类型）转换为INTEGER类型。

`UTL_RAW.CAST_TO_BINARY_INTEGER`函数原型为：

```
UTL_RAW.CAST_TO_BINARY_INTEGER (
r          IN RAW,
endianess IN INTEGER)
RETURN BINARY_INTEGER;
```

表 14-27 UTL\_RAW.CAST\_TO\_BINARY\_INTEGER 接口参数说明

参数	描述
r	二进制表示形式的整型值（即RAW类型）。
endianess	表示字节序的整型值1或2（1代表BIG_ENDIAN，2代表LITTLE-ENDIAN）。

- UTL\_RAW.LENGTH  
存储过程LENGTH返回RAW类型对象的长度。

UTL\_RAW.LENGTH函数原型为：

```
UTL_RAW.LENGTH(  
r IN RAW)  
RETURN INTEGER;
```

表 14-28 UTL\_RAW.LENGTH 接口参数说明

参数	描述
r	RAW类型对象

- UTL\_RAW.CAST\_TO\_RAW  
存储过程CAST\_TO\_RAW将VARCHAR2类型的对象转换成RAW类型。

UTL\_RAW.CAST\_TO\_RAW函数原型为：

```
UTL_RAW.CAST_TO_RAW(  
c IN VARCHAR2)  
RETURN RAW;
```

表 14-29 UTL\_RAW.CAST\_TO\_RAW 接口参数说明

参数	描述
c	待转换的VARCHAR2类型对象

## 示例

```
--在存储过程中操作RAW数据  
CREATE OR REPLACE PROCEDURE proc_raw  
AS  
str varchar2(100) := 'abcdef';  
source raw(100);  
amount integer;  
BEGIN  
source := utl_raw.cast_to_raw(str);--类型转换  
amount := utl_raw.length(source);--获取长度  
dbms_output.put_line(amount);  
END;  
/  
  
--调用存储过程  
CALL proc_raw();  
  
--删除存储过程  
DROP PROCEDURE proc_raw;
```

## 14.11.5 DBMS\_JOB

### 接口介绍

高级功能包DBMS\_JOB支持的所有接口请参见表14-30。

表 14-30 DBMS\_JOB

接口名称	描述
<b>DBMS_JOB.SUBMIT</b>	提交一个定时任务。作业号由系统自动生成。
<b>DBMS_JOB.ISUBMIT</b>	提交一个定时任务。作业号由用户指定。
<b>DBMS_JOB.REMOVE</b>	通过作业号来删除定时任务。
<b>DBMS_JOB.BROKEN</b>	禁用或者启用定时任务。
<b>DBMS_JOB.CHANGE</b>	修改定时任务的属性，包括任务内容、下次执行时间、执行间隔。
<b>DBMS_JOB.WHAT</b>	修改定时任务的任务内容属性。
<b>DBMS_JOB.NEXT_DATE</b>	修改定时任务的下次执行时间属性。
<b>DBMS_JOB.INTERVAL</b>	修改定时任务的执行间隔属性。
<b>DBMS_JOB.CHANGE_OWNER</b>	修改定时任务的属主。

- **DBMS\_JOB.SUBMIT**  
存储过程SUBMIT提交一个系统提供的定时任务。

DBMS\_JOB.SUBMIT函数原型为：

```
DBMS_JOB.SUBMIT(
  what      IN TEXT,
  next_date IN TIMESTAMP DEFAULT sysdate,
  job_interval IN TEXT DEFAULT 'null',
  job       OUT INTEGER);
```

#### 说明

当创建一个定时任务（DBMS\_JOB）时，系统默认将当前数据库和用户名与当前创建的定时任务（DBMS\_JOB）绑定起来。该接口函数可以通过call或select调用，如果通过select调用，可以不填写出参。如果在存储过程中则需要用通过perform调用该接口函数。

表 14-31 DBMS\_JOB.SUBMIT 接口参数说明

参数	类型	入参/出参	是否可以空	描述
what	text	IN	否	要执行的SQL语句。支持一个或多个‘DML’，‘匿名块’，‘调用存储过程的语句’或3种混合的场景。
next_date	timestamp	IN	否	下次作业运行时间。默认值为当前系统时间（sysdate）。如果是过去时间，在提交作业时表示立即执行。
interval	text	IN	是	用来计算下次作业运行时间的时间表达式，可以是interval表达式，也可以是sysdate加上一个numeric值（例如：sysdate+1.0/24）。如果为空值或字符串“null”表示只执行一次，执行后JOB状态STATUS变成'd'不再执行。
job	integer	OUT	否	作业号。范围为1~32767。当使用select调用dbms.submit时，该参数可以省略。

示例：

```
select DBMS_JOB.SUBMIT('call pro_xxx()',' to_date('20180101','yyyymmdd'),'sysdate+1');
select DBMS_JOB.SUBMIT('call pro_xxx()',' to_date('20180101','yyyymmdd'),'sysdate+1.0/24');
CALL DBMS_JOB.SUBMIT('INSERT INTO T_JOB VALUES(1); call pro_1(); call pro_2()';
add_months(to_date('201701','yyyymm'),1), 'date_trunc("day",SYSDATE) + 1 +(8*60+30.0)/
(24*60) ',:jobid);
```

- DBMS\_JOB.ISUBMIT

ISUBMIT与SUBMIT语法功能相同，但其第一个参数是入参，即指定的作业号，SUBMIT最后一个参数是出参，表示系统自动生成的作业号。

示例：

```
CALL dbms_job.isubmit(101, 'insert_msg_statistic1;', sysdate, 'sysdate+3.0/24');
```

- DBMS\_JOB.REMOVE

存储过程REMOVE删除指定的定时任务。

DBMS\_JOB.REMOVE函数原型为：

```
REMOVE(job IN INTEGER);
```

表 14-32 DBMS\_JOB.REMOVE 接口参数说明

参数	类型	入参/出参	是否可以空	描述
job	integer	IN	否	指定的作业号。

示例:

```
CALL dbms_job.remove(101);
```

- DBMS\_JOB.BROKEN

存储过程BROKEN禁用或者启用定时任务。

DBMS\_JOB.BROKEN函数原型为:

```
DBMS_JOB.BROKEN(
job      IN  INTEGER,
broken   IN  BOOLEAN,
next_date IN  TIMESTAMP DEFAULT sysdate);
```

表 14-33 DBMS\_JOB.BROKEN 接口参数说明

参数	类型	入参/ 出参	是否 可以为空	描述
job	integer	IN	否	指定的作业号。
broken	boolean	IN	否	状态标志位，true代表禁用，false代表启用。具体true或false值更新当前job；如果为空值，则不改变原有job的状态。
next_date	timestamp	IN	是	下次运行时间，默认为当前系统时间。如果参数broken状态为true，则更新该参数为'4000-1-1'；如果参数broken状态为false，且如果参数next_date不为空值，则更新指定job的next_date值，如果next_date为空值，则不更新next_date值。该参数可以省略，为默认值。

示例:

```
CALL dbms_job.broken(101, true);
CALL dbms_job.broken(101, false, sysdate);
```

- DBMS\_JOB.CHANGE

存储过程CHANGE修改定时任务的属性，包括任务内容、下次执行时间、执行间隔。

DBMS\_JOB.CHANGE函数原型为:

```
DBMS_JOB.CHANGE(
job      IN  INTEGER,
what     IN  TEXT,
next_date IN  TIMESTAMP,
interval IN  TEXT);
```

表 14-34 DBMS\_JOB.CHANGE 接口参数说明

参数	类型	入参/出参	是否可以 为空	描述
job	integer	IN	否	指定的作业号。
what	text	IN	是	执行的存储过程名或者sql语句块。如果该参数为空值，则不更新指定job的what值，否则更新指定job的what值。
next_date	timestamp	IN	是	下次运行时间。如果该参数为空值，则不更新指定job的next_date值，否则更新指定job的next_date值。
interval	text	IN	是	用来计算下次作业运行时间的时间表达式。如果该参数为空值，则不更新指定job的interval值；如果该参数不为空值，会校验interval是否为有效的时间类型或interval类型，则更新指定job的interval值。如果为字符串"null"表示只执行一次，执行后JOB状态STATUS变成'd'不再执行。

示例：

```
CALL dbms_job.change(101, 'call userproc();', sysdate, 'sysdate + 1.0/1440');
CALL dbms_job.change(101, 'insert into tbl_a values(sysdate);', sysdate, 'sysdate + 1.0/1440');
```

- DBMS\_JOB.WHAT

存储过程WHAT修改定时任务的任务内容属性。

DBMS\_JOB.WHAT函数原型为：

```
DBMS_JOB.WHAT(
job          IN  INTEGER,
what         IN  TEXT);
```

表 14-35 DBMS\_JOB.WHAT 接口参数说明

参数	类型	入参/出参	是否可以 为空	描述
job	integer	IN	否	指定的作业号。
what	text	IN	否	执行的存储过程调用或者sql语句块。

 说明

- 当what参数是一个或多个可以执行成功的sql语句/程序块/调用存储过程时，该接口函数才能被执行成功，否则会执行失败。
- 若what参数为一个简单的insert、update等语句，需要在表前加模式名。

示例：

```
CALL dbms_job.what(101, 'call userproc();');
CALL dbms_job.what(101, 'insert into tbl_a values(sysdate);');
```

- DBMS\_JOB.NEXT\_DATE

存储过程NEXT\_DATE修改定时任务的下次执行时间属性。

DBMS\_JOB.NEXT\_DATE函数原型为：

```
DBMS_JOB.NEXT_DATE(
job      IN  INTEGER,
next_date IN  TIMESTAMP);
```

表 14-36 DBMS\_JOB.NEXT\_DATE 接口参数说明

参数	类型	入参/出参	是否可以空	描述
job	integer	IN	否	指定的作业号。
next_date	timestamp	IN	否	下次运行时间。

 说明

如果输入的next\_date的值小于当前日期值，该job会立即执行一次。

示例：

```
CALL dbms_job.next_date(101, sysdate);
```

- DBMS\_JOB.INTERVAL

存储过程INTERVAL修改定时任务的执行间隔属性。

DBMS\_JOB.INTERVAL函数原型为：

```
DBMS_JOB.INTERVAL(
job      IN  INTEGER,
interval IN  TEXT);
```

表 14-37 DBMS\_JOB.INTERVAL 接口参数说明

参数	类型	入参/出参	是否可以空	描述
job	integer	IN	否	指定的作业号。
interval	text	IN	是	用来计算下次作业运行时间的时间表表达式。如果为空值或字符串"null"表示只执行一次，执行后JOB状态STATUS变成'd' 不再执行。interval是否为有效的时间类型或interval类型。



示例:

```
CALL dbms_job.interval(101, 'sysdate + 1.0/1440');
```

### 📖 说明

对于指定job正在运行状态（即job\_status为'r'）时，不允许通过remove、change、next\_date、what、interval等接口删除或修改job的参数信息。

- **DBMS\_JOB.CHANGE\_OWNER**  
存储过程CHANGE\_OWNER修改定时任务的属主。

DBMS\_JOB.CHANGE\_OWNER函数原型为:

```
DBMS_JOB.CHANGE_OWNER(  
job          IN  INTEGER,  
new_owner    IN  NAME);
```

**表 14-38** DBMS\_JOB.CHANGE\_OWNER 接口参数说明

参数	类型	入参/出参	是否可以 为空	描述
job	integer	IN	否	指定的作业号。
new_owner	name	IN	否	新的用户名。

示例:

```
CALL dbms_job.change_owner(101, 'alice');
```

## 约束说明

1. 创建一个新job后，该job从属于当前coordinator（即：该job仅在当前coordinator上调度和执行），其他coordinator不会调度和执行该job。所有coordinator都可以查看、修改、删除其他CN创建的job。
2. job只能通过dbms\_job高级包提供的接口进行创建、更新、删除操作，因为高级包的接口中会考虑所有CN间job信息的同步和pg\_jobs表主键的关联操作，如果通过DML语句对pg\_jobs表进行增删改，会导致job信息在CN间不一致和系统表无法关联变更的混乱问题，会严重影响job内部的管理。
3. 由于用户创建的每个任务和CN绑定，若不开启CN故障自动迁移功能，当任务运行过程中，该CN故障，则该任务的状态无法实时刷新。如果在任务未执行时CN故障，则该CN上的任务都得不到正常的调度和执行。建议开启CN故障自动迁移功能，故障CN上的作业会迁移至其他CN继续调度。
4. job在定时执行过程中，需要在当前job所属的CN上实时更新该job的运行状态、最近执行开始时间、最近执行结束时间、下次开始时间、失败次数（如果job执行失败）等相关参数信息到pg\_jobs系统表中，并同步到其他CN，保证job信息的一致性。如果其他CN存在节点故障，那么job所属CN会同步超时重发的处理，导致job执行时间变长，但CN间同步超时失败后，原CN上pg\_jobs表中job的相关信息仍然能正常更新，且job能正常执行成功。当故障CN恢复正常后，可能出现该CN上pg\_jobs表中当前job的执行时间、运行状态等参数与原CN上不一致的情况，需要原CN上再次执行该job后才能保证job信息的同步。

- 对于并发同时有多个job到达执行时间的场景，由于会为每个job创建一个线程来执行job，由于系统内部启动每个线程的时间会有延迟，因此会导致同时并发执行的job的开始时间有延迟，每个job的延迟时间在0.1ms左右。

## 14.11.6 DBMS\_SQL

### 接口介绍

高级功能包DBMS\_SQL支持的接口请参见表1 DBMS\_SQL。

表 14-39 DBMS\_SQL

接口名称	描述
DBMS_SQL.OPEN_CURSOR	打开一个游标。
DBMS_SQL.CLOSE_CURSOR	关闭一个已打开的游标。
DBMS_SQL.PARSE	向游标传递一组SQL语句，目前只支持SELECT。
DBMS_SQL.EXECUTE	在游标上执行一组动态定义操作。
DBMS_SQL.FETCH_ROWS	读取游标一行数据。
DBMS_SQL.DEFINE_COLUMN	动态定义一个列。
DBMS_SQL.DEFINE_COLUMN_CHAR	动态定义一个char类型的列。
DBMS_SQL.DEFINE_COLUMN_INT	动态定义一个int类型的列。
DBMS_SQL.DEFINE_COLUMN_LONG	动态定义一个long类型的列。
DBMS_SQL.DEFINE_COLUMN_RAW	动态定义一个raw类型的列。
DBMS_SQL.DEFINE_COLUMN_TEXT	动态定义一个text类型的列。
DBMS_SQL.DEFINE_COLUMN_UNKNOWN	动态定义一个未知列（类型不识别时入此接口）。
DBMS_SQL.COLUMN_VALUE	读取一个已动态定义的列值。
DBMS_SQL.COLUMN_VALUE_CHAR	读取一个已动态定义的列值（指定char类型）。
DBMS_SQL.COLUMN_VALUE_INT	读取一个已动态定义的列值（指定int类型）。
DBMS_SQL.COLUMN_VALUE_LONG	读取一个已动态定义的列值（指定long类型）。
DBMS_SQL.COLUMN_VALUE_RAW	读取一个已动态定义的列值（指定raw类型）。
DBMS_SQL.COLUMN_VALUE_TEXT	读取一个已动态定义的列值（指定text类型）。

接口名称	描述
<code>DBMS_SQL.COLUMN_VALUE_UNKNOWN</code>	读取一个已动态定义的列值（类型不识别时入此接口）。
<code>DBMS_SQL.IS_OPEN</code>	检查游标是否已打开。

### 📖 说明

- 建议使用`dbms_sql.define_column`及`dbms_sql.column_value`定义参数列。
- 当结果集大于`work_mem`设定值时会触发结果集临时下盘，但最大阈值不超过512MB。
- `DBMS_SQL.OPEN_CURSOR`  
该函数用来打开一个游标，是后续`dbms_sql`各项操作的前提。该函数不传入任何参数，内部自动递增生成游标ID，并作为返回值返回给`integer`定义的变量。

`DBMS_SQL.OPEN_CURSOR`函数原型为：

```
DBMS_SQL.OPEN_CURSOR (
)
RETURN INTEGER;
```

- `DBMS_SQL.CLOSE_CURSOR`  
该函数用来关闭一个游标，是`dbms_sql`各项操作的结束。如果在存储过程结束时没有调用该函数，则该游标占用的内存仍然会保存，因此关闭游标非常重要。由于异常情况的发生会中途退出存储过程，导致游标未能关闭，因此建议存储过程中有异常处理，将该接口包含在内。

`DBMS_SQL.CLOSE_CURSOR`函数原型为：

```
DBMS_SQL.CLOSE_CURSOR (
cursorid IN INTEGER
)
RETURN INTEGER;
```

**表 14-40** `DBMS_SQL.CLOSE_CURSOR` 接口说明

参数名称	描述
<code>cursorid</code>	打算关闭的游标ID号

- `DBMS_SQL.PARSE`  
该函数用来解析给定游标的查询语句，被传入的查询语句会立即执行。目前仅支持`SELECT`查询语句的解析，且语句参数仅可通过`text`类型传递，长度不大于1G。

`DBMS_SQL.PARSE`函数的原型为：

```
DBMS_SQL.PARSE (
cursorid IN INTEGER,
query_string IN TEXT,
label IN INTEGER
)
RETURN BOOLEAN;
```

表 14-41 DBMS\_SQL.PARSE 接口说明

参数名称	描述
cursorid	执行查询语句解析的游标ID
query_string	执行的查询语句
language_flag	版本语言号，目前只支持1

- DBMS\_SQL.EXECUTE

该函数用来执行一个给定的游标。该函数接收一个游标ID，运行后获得的数据用于后续操作。目前仅支持SELECT查询语句的执行。

DBMS\_SQL.EXECUTE函数的原型为：

```
DBMS_SQL.EXECUTE(
  cursorid IN INTEGER,
)
RETURN INTEGER;
```

表 14-42 DBMS\_SQL.EXECUTE 接口说明

参数名称	描述
cursorid	执行查询语句解析的游标ID

- DBMS\_SQL.FETCH\_ROWS

该函数返回符合查询条件的数据行数，每一次运行该接口都会获取到新的行数的集合，直到数据读取完毕获取不到新行为止。

DBMS\_SQL.FETCH\_ROWS函数的原型为：

```
DBMS_SQL.FETCH_ROWS(
  cursorid IN INTEGER,
)
RETURN INTEGER;
```

表 14-43 DBMS\_SQL.FETCH\_ROWS 接口说明

参数名称	描述
curosorid	执行的游标ID

- DBMS\_SQL.DEFINE\_COLUMN

该函数用来定义从给定游标返回的列，该接口只能应用于SELECT定义的游标中。定义的列通过查询列表的相对位置来标识，传入变量的数据类型决定了该列被定义的类型。

DBMS\_SQL.DEFINE\_COLUMN函数的原型为：

```
DBMS_SQL.DEFINE_COLUMN(
  cursorid IN INTEGER,
  position IN INTEGER,
  column_ref IN ANYELEMENT,
  column_size IN INTEGER default 1024
)
RETURN INTEGER;
```

表 14-44 DBMS\_SQL.DEFINE\_COLUMN 接口说明

参数名称	描述
cursorid	执行的游标ID
position	动态定义列在查询中的位置
column_ref	任意类型的变量，可根据变量类型选择适当的接口动态定义列
column_size	定义的列的长度

- DBMS\_SQL.DEFINE\_COLUMN\_CHAR

该函数用来定义从给定游标返回的CHAR类型的列，该接口只能应用于SELECT定义的游标中。定义的列通过查询列表的相对位置来标识，传入变量的数据类型决定了该列被定义的类型。

DBMS\_SQL.DEFINE\_COLUMN\_CHAR函数的原型为：

```
DBMS_SQL.DEFINE_COLUMN_CHAR(
cursorid IN INTEGER,
position IN INTEGER,
column IN TEXT,
column_size IN INTEGER
)
RETURN INTEGER;
```

表 14-45 DBMS\_SQL.DEFINE\_COLUMN\_CHAR 接口说明

参数名称	描述
cursorid	执行的游标ID
position	动态定义列在查询中的位置
column	需要定义的某类型的参数变量
column_size	动态定义列长度

- DBMS\_SQL.DEFINE\_COLUMN\_INT

该函数用来定义从给定游标返回的INT类型的列，该接口只能应用于SELECT定义的游标中。定义的列通过查询列表的相对位置来标识，传入变量的数据类型决定了该列被定义的类型。

DBMS\_SQL.DEFINE\_COLUMN\_INT函数的原型为：

```
DBMS_SQL.DEFINE_COLUMN_INT(
cursorid IN INTEGER,
position IN INTEGER
)
RETURN INTEGER;
```

表 14-46 DBMS\_SQL.DEFINE\_COLUMN\_INT 接口说明

参数名称	描述
cursorid	执行的游标ID
position	动态定义列在查询中的位置

- **DBMS\_SQL.DEFINE\_COLUMN\_LONG**

该函数用来定义从给定游标返回的长列类型（非数据类型long）的列，该接口只能应用于SELECT定义的游标中。定义的列通过查询列表的相对位置来标识，传入变量的数据类型决定了该列被定义的类型。长列的大小限制为1G。

DBMS\_SQL.DEFINE\_COLUMN\_LONG函数的原型为：

```
DBMS_SQL.DEFINE_COLUMN_LONG(
cursorid   IN INTEGER,
position   IN INTEGER
)
RETURN INTEGER;
```

**表 14-47** DBMS\_SQL.DEFINE\_COLUMN\_LONG 接口说明

参数名称	描述
cursorid	执行的游标ID
position	动态定义列在查询中的位置

- **DBMS\_SQL.DEFINE\_COLUMN\_RAW**

该函数用来定义从给定游标返回的RAW类型的列，该接口只能应用于SELECT定义的游标中。定义的列通过查询列表的相对位置来标识，传入变量的数据类型决定了该列被定义的类型。

DBMS\_SQL.DEFINE\_COLUMN\_RAW函数的原型为：

```
DBMS_SQL.DEFINE_COLUMN_RAW(
cursorid   IN INTEGER,
position   IN INTEGER,
column     IN BYTEA,
column_size IN INTEGER
)
RETURN INTEGER;
```

**表 14-48** DBMS\_SQL.DEFINE\_COLUMN\_RAW 接口说明

参数名称	描述
cursorid	执行的游标ID
position	动态定义列在查询中的位置
column	RAW类型的参数变量
column_size	列的长度

- **DBMS\_SQL.DEFINE\_COLUMN\_TEXT**

该函数用来定义从给定游标返回的TEXT类型的列，该接口只能应用于SELECT定义的游标中。定义的列通过查询列表的相对位置来标识，传入变量的数据类型决定了该列被定义的类型。

DBMS\_SQL.DEFINE\_COLUMN\_TEXT函数的原型为：

```
DBMS_SQL.DEFINE_COLUMN_CHAR(
cursorid   IN INTEGER,
position   IN INTEGER,
max_size   IN INTEGER
)
RETURN INTEGER;
```

表 14-49 DBMS\_SQL.DEFINE\_COLUMN\_TEXT 接口说明

参数名称	描述
cursorid	执行的游标ID
position	动态定义列在查询中的位置
max_size	定义的TEXT类型的最大长度

- DBMS\_SQL.DEFINE\_COLUMN\_UNKNOWN

该函数用来处理从给定游标返回的未知数据类型的列，该接口仅用于类型不识别时的报错退出。

DBMS\_SQL.DEFINE\_COLUMN\_UNKNOWN函数的原型为：

```
DBMS_SQL.DEFINE_COLUMN_CHAR(
cursorid IN INTEGER,
position IN INTEGER,
column IN TEXT
)
RETURN INTEGER;
```

表 14-50 DBMS\_SQL.DEFINE\_COLUMN\_UNKNOWN 接口说明

参数名称	描述
cursorid	执行的游标ID
position	动态定义列在查询中的位置
column	动态定义的参数

- DBMS\_SQL.COLUMN\_VALUE

该函数用来返回给定游标给定位置的游标元素值，该接口访问的是 DBMS\_SQL.FETCH\_ROWS 获取的数据。

DBMS\_SQL.COLUMN\_VALUE函数的原型为：

```
DBMS_SQL.COLUMN_VALUE(
cursorid IN INTEGER,
position IN INTEGER,
column_value INOUT ANYELEMENT
)
RETURN ANYELEMENT;
```

表 14-51 DBMS\_SQL.COLUMN\_VALUE 接口说明

参数名称	描述
cursorid	执行的游标ID
position	动态定义列在查询中的位置
column_value	定义的列的返回值

- DBMS\_SQL.COLUMN\_VALUE\_CHAR

该函数用来返回给定游标给定位置的游标CHAR类型的值，该接口访问的是 DBMS\_SQL.FETCH\_ROWS 获取的数据。

DBMS\_SQL.COLUMN\_VALUE\_CHAR函数的原型为：

```
DBMS_SQL.COLUMN_VALUE_CHAR(
cursorid      IN  INTEGER,
position      IN  INTEGER,
column_value  INOUT CHARACTER,
err_num       INOUT NUMERIC default 0,
actual_length INOUT INTEGER default 1024
)
RETURN RECORD;
```

表 14-52 DBMS\_SQL.COLUMN\_VALUE\_CHAR 接口说明

参数名称	描述
cursorid	执行的游标ID
position	动态定义列在查询中的位置
column_value	返回值
err_num	错误号。传出参数，须传入变量做参数。目前未实现，固定传出-1。
actual_length	返回值的实际长度

- DBMS\_SQL.COLUMN\_VALUE\_INT

该函数用来返回给定游标给定位置的游标INT类型的值，该接口访问的是DBMS\_SQL.FETCH\_ROWS获取的数据。DBMS\_SQL.COLUMN\_VALUE\_INT函数的原型为：

```
DBMS_SQL.COLUMN_VALUE_INT(
cursorid      IN  INTEGER,
position      IN  INTEGER
)
RETURN INTEGER;
```

表 14-53 DBMS\_SQL.COLUMN\_VALUE\_INT 接口说明

参数名称	描述
cursorid	执行的游标ID
position	动态定义列在查询中的位置

- DBMS\_SQL.COLUMN\_VALUE\_LONG

该函数用来返回给定游标给定位置的游标长列（非long/bigint整型）类型的值，该接口访问的是DBMS\_SQL.FETCH\_ROWS获取的数据。

DBMS\_SQL.COLUMN\_VALUE\_LONG函数的原型为：

```
DBMS_SQL.COLUMN_VALUE_LONG(
cursorid      IN  INTEGER,
position      IN  INTEGER,
length        IN  INTEGER,
off_set       IN  INTEGER,
column_value  INOUT TEXT,
actual_length INOUT INTEGER default 1024
)
RETURN RECORD;
```



表 14-54 DBMS\_SQL.COLUMN\_VALUE\_LONG 接口说明

参数名称	描述
cursorid	执行的游标ID
position	动态定义列在查询中的位置
length	返回值的长度
off_set	返回值的起始位置
column_value	返回值
actual_length	实际返回值的长度

- DBMS\_SQL.COLUMN\_VALUE\_RAW

该函数用来返回给定游标给定位置的游标RAW类型的值，该接口访问的是 DBMS\_SQL.FETCH\_ROWS 获取的数据。

DBMS\_SQL.COLUMN\_VALUE\_RAW函数的原型为：

```
DBMS_SQL.COLUMN_VALUE_RAW(
cursorid          IN  INTEGER,
position          IN  INTEGER,
column_value      INOUT BYTEA,
err_num           INOUT NUMERIC default 0,
actual_length     INOUT INTEGER default 1024
)
RETURN RECORD;
```

表 14-55 DBMS\_SQL.COLUMN\_VALUE\_RAW 接口说明

参数名称	描述
cursorid	执行的游标ID
position	动态定义列在查询中的位置
column_value	返回的列值
err_num	错误号。传出参数，须传入变量做参数。目前未实现，固定传出-1。
actual_length	返回值的实际长度，不能长于此值，否则截断

- DBMS\_SQL.COLUMN\_VALUE\_TEXT

该函数用来返回给定游标给定位置的游标TEXT类型的值，该接口访问的是 DBMS\_SQL.FETCH\_ROWS 获取的数据。

DBMS\_SQL.COLUMN\_VALUE\_TEXT函数的原型为：

```
DBMS_SQL.COLUMN_VALUE_TEXT(
cursorid          IN  INTEGER,
position          IN  INTEGER
)
RETURN TEXT;
```

表 14-56 DBMS\_SQL.COLUMN\_VALUE\_TEXT 接口说明

参数名称	描述
cursorid	执行的游标ID
position	动态定义列在查询中的位置

- DBMS\_SQL.COLUMN\_VALUE\_UNKNOWN

该函数用来返回给定游标给定位置的游标未知类型的值，该接口为类型不支持时的报错处理接口。

DBMS\_SQL.COLUMN\_VALUE\_UNKNOWN函数的原型为：

```
DBMS_SQL.COLUMN_VALUE_UNKNOWN(
cursorid          IN  INTEGER,
position          IN  INTEGER,
COLUMN_TYPE      IN  TEXT
)
RETURN TEXT;
```

表 14-57 DBMS\_SQL.COLUMN\_VALUE\_UNKNOWN 接口说明

参数名称	描述
cursorid	执行的游标ID
position	动态定义列在查询中的位置
column_type	返回的参数类型

- DBMS\_SQL.IS\_OPEN

该函数用来返回游标的当前状态：打开、解析、执行、定义。取值是为TRUE，关闭后为FALSE，未知时报错，其余默认为关闭。

DBMS\_SQL.IS\_OPEN函数的原型为：

```
DBMS_SQL.IS_OPEN(
cursorid          IN  INTEGER
)
RETURN BOOLEAN;
```

表 14-58 DBMS\_SQL.IS\_OPEN 接口说明

参数名称	描述
cursorid	被查询的游标ID

## 示例

```
--在存储过程中操作raw数据
create or replace procedure pro_dbms_sql_all_02(in_raw raw,v_in int,v_offset int)
as
cursorid int;
v_id int;
v_info bytea :=1;
query varchar(2000);
execute_ret int;
```

```

define_column_ret_row bytea := '1';
define_column_ret int;
begin
drop table if exists pro_dbms_sql_all_tb1_02 ;
create table pro_dbms_sql_all_tb1_02(a int ,b blob);
insert into pro_dbms_sql_all_tb1_02 values(1,HEXTORAW('DEADBEEF'));
insert into pro_dbms_sql_all_tb1_02 values(2,in_row);
query := 'select * from pro_dbms_sql_all_tb1_02 order by 1';
--打开游标
cursorid := dbms_sql.open_cursor();
--编译游标
dbms_sql.parse(cursorid, query, 1);
--定义列
define_column_ret:= dbms_sql.define_column(cursorid,1,v_id);
define_column_ret_row:= dbms_sql.define_column_raw(cursorid,2,v_info,10);
--执行
execute_ret := dbms_sql.execute(cursorid);
loop
exit when (dbms_sql.fetch_rows(cursorid) <= 0);
--获取值
dbms_sql.column_value(cursorid,1,v_id);
dbms_sql.column_value_raw(cursorid,2,v_info,v_in,v_offset);
--输出结果
dbms_output.put_line('id: || v_id || ' info: ' || v_info);
end loop;
--关闭游标
dbms_sql.close_cursor(cursorid);
end;
/
--调用存储过程
call pro_dbms_sql_all_02(HEXTORAW('DEADBEEF'),0,1);
--删除存储过程
DROP PROCEDURE pro_dbms_sql_all_02;

```

## 14.12 调试

### 语法

RAISE有以下五种语法格式:

图 14-34 raise\_format::=

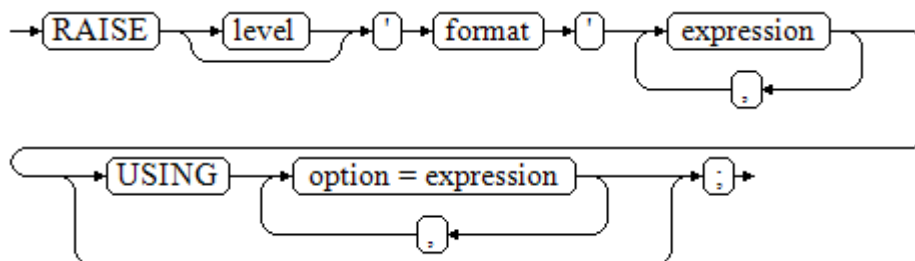


图 14-35 raise\_condition::=

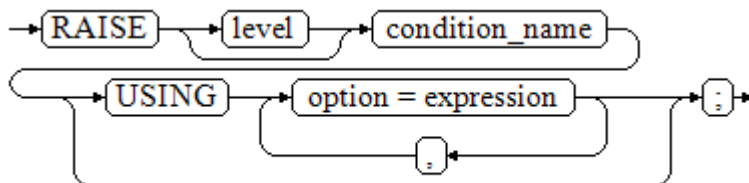


图 14-36 raise\_sqlstate::=

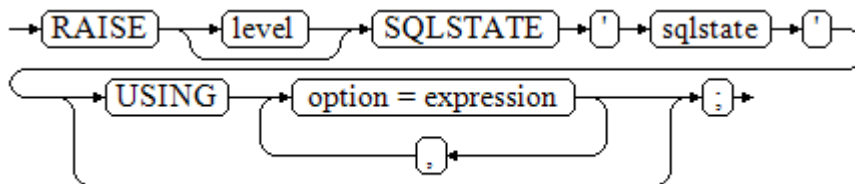


图 14-37 raise\_option::=

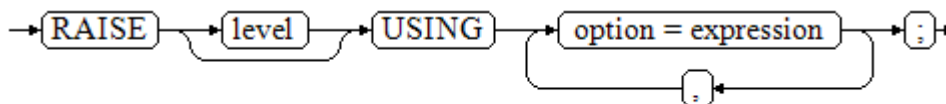
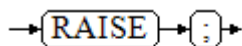


图 14-38 raise::=



### 参数说明:

- level选项用于指定错误级别，有DEBUG，LOG，INFO，NOTICE，WARNING以及EXCEPTION（默认值）。EXCEPTION抛出一个正常终止当前事务的异常，其他的仅产生不同异常级别的信息。特殊级别的错误信息是否报告到客户端、写到服务器日志由log\_min\_messages和client\_min\_messages这两个配置参数控制。
- format: 格式字符串，指定要报告的错误消息文本。格式字符串后可跟表达式，用于向消息文本中插入。在格式字符串中，%由format后面跟着的参数的值替换，%%用于打印出%。例如：  
--v\_job\_id 将替换字符串中的 %:  
RAISE NOTICE 'Calling cs\_create\_job(%)',v\_job\_id;
- option = expression: 向错误报告中添加另外的信息。关键字option可以是MESSAGE、DETAIL、HINT以及ERRCODE，并且每一个expression可以是任意的字符串。
  - MESSAGE，指定错误消息文本，这个选项不能用于在USING前包含一个格式字符串的RAISE语句中。
  - DETAIL，说明错误的详细信息。
  - HINT，用于打印出提示信息。

- ERRCODE，向报告中指定错误码（SQLSTATE）。可以使用条件名称或者直接用五位字符的SQLSTATE错误码。
- condition\_name：错误码对应的条件名。
- sqlstate：错误码。

如果在RAISE EXCEPTION命令中既没有指定条件名也没有指定SQLSTATE，默认用RAISE EXCEPTION (P0001)。如果没有指定消息文本，默认用条件名或者SQLSTATE作为消息文本。

### 须知

当由SQLSTATE指定了错误码，则不局限于已定义的错误码，可以选择任意包含五个数字或者大写的ASCII字母的错误码，而不是00000。建议避免使用以三个0结尾的错误码，因为这种错误码是类别码，会被整个种类捕获。

### 说明

图14-38所示的语法不接任何参数。这种形式仅用于一个BEGIN块中的EXCEPTION语句，它使得错误重新被处理。

## 示例

终止事务时，给出错误和提示信息：

```
CREATE OR REPLACE PROCEDURE proc_raise1(user_id in integer)
AS
BEGIN
RAISE EXCEPTION 'Noexistence ID --> %',user_id USING HINT = 'Please check your user ID';
END;
/

call proc_raise1(300011);

--执行结果
ERROR: Noexistence ID --> 300011
HINT: Please check your user ID
```

两种设置SQLSTATE的方式：

```
CREATE OR REPLACE PROCEDURE proc_raise2(user_id in integer)
AS
BEGIN
RAISE 'Duplicate user ID: %',user_id USING ERRCODE = 'unique_violation';
END;
/

\set VERBOSITY verbose
call proc_raise2(300011);

--执行结果
ERROR: Duplicate user ID: 300011
SQLSTATE: 23505
LOCATION: exec_stmt_raise, pl_exec.cpp:3482
```

如果主要的参数是条件名或者是SQLSTATE，可以使用：

```
RAISE division_by_zero;
```

```
RAISE SQLSTATE '22012';
```

例如：

```
CREATE OR REPLACE PROCEDURE division(div in integer, dividend in integer)
AS
DECLARE
res int;
BEGIN
IF dividend=0 THEN
RAISE division_by_zero;
RETURN;
ELSE
res := div/dividend;
RAISE INFO 'division result: %', res;
RETURN;
END IF;
END;
/
call division(3,0);

--执行结果
ERROR: division_by_zero
```

或者另一种方式:

```
RAISE unique_violation USING MESSAGE = 'Duplicate user ID: ' || user_id;
```

# 15 系统表和系统视图

## 15.1 系统表和系统视图概述

系统表是GaussDB(DWS)存放结构元数据，是GaussDB(DWS)数据库系统运行控制信息的来源，也是数据库系统的核心组成部分。系统表包含集群安装信息以及GaussDB(DWS)上运行的各种查询和进程的信息。可以通过查询系统表来收集有关数据库的信息。

系统视图提供了查询系统表和访问数据库内部状态的方法。当用户对数据库中的一张或者多张表的某些字段的组合感兴趣，而又不想每次键入这些查询时，用户就可以定义一个视图来解决这个问题。视图与基本表不同，不是物理上实际存在的，是一个虚表。数据库中仅存放视图的定义，而不存放视图对应的数据，这些数据仍存放在原来的基本表中。若基本表中的数据发生变化，从视图中查询出的数据也随之改变。从这个意义上讲，视图就像一个窗口，透过它可以看到数据库中用户感兴趣的数据及变化。视图每次被引用的时候都会运行一次。

三权分立下，非管理员无权查看系统表和视图。非三权分立下，系统表和系统视图要么只对管理员可见，要么对所有用户可见。下面的系统表和视图有些标识了需要管理员权限，这些系统表和视图只有管理员可以查询。

### 须知

禁止对系统表或系统视图进行增删改等操作，手动对系统表或系统视图的修改或破坏可能会导致系统信息不一致，造成系统控制异常甚至出现集群不可用的情况。

## 15.2 系统表

### 15.2.1 GS\_BLOCKLIST\_QUERY

GS\_BLOCKLIST\_QUERY系统表存储作业黑名单信息和异常信息，该表以unique\_sql\_id作为唯一索引，进行作业异常信息统计和黑名单记录，可通过与GS\_WLM\_SESSION\_INFO进行关联获取作业的query字段和执行信息。

GaussDB(DWS)同时提供了GS\_BLOCKLIST\_QUERY视图用于查询作业黑名单和异常信息，该视图可直接显示query字段信息，不过因为该视图与

GS\_WLM\_SESSION\_INFO为依赖关系，因此在GS\_WLM\_SESSION\_INFO表较大的情况下，查询可能需要消耗较长的时间。

表 15-1 GS\_BLOCKLIST\_QUERY 字段

名字	类型	引用	描述
unique_sql_id	bigint	-	基于查询解析树生成的查询唯一ID。
block_list	boolean	-	查询作业是否属于黑名单。
except_num	integer	-	查询作业异常次数。
except_time	timestamp	-	查询作业最近一次异常时间。

### 说明

- 此系统表的schema为dbms\_om。
- 此系统表包含唯一索引，使用哈希分布方式分布在DN上，分布列为unique\_sql\_id。
- 此系统表仅限在gaussdb数据库中查询，其它数据库中查询会直接报错。
- GS\_BLOCKLIST\_QUERY视图存储在pg\_catalog下。
- 通常对于DML语句，在计算Unique SQL ID的过程中会忽略常量值。但对于DDL、DCL以及设置参数等语句，常量值不可以忽略。因此一个unique\_sql\_id可能会对应一个或多个查询。

## 15.2.2 GS\_OBSSCANINFO

GS\_OBSSCANINFO系统表定义了云上加速场景中，使用加速集群时扫描OBS数据的运行时信息，每条记录对应一个query中单个OBS外表的运行时信息。

表 15-2 GS\_OBSSCANINFO 字段

名字	类型	引用	描述
query_id	bigint	-	查询标识。
user_id	text	-	执行该查询的数据库用户。
table_name	text	-	OBS外表的表名。
file_type	text	-	底层数据保存的文件格式。
time_stamp	time_stamp	-	扫描操作开始的时间。
actual_time	double	-	扫描操作执行时间，单位为秒。
file_scanned	bigint	-	扫描的文件数量。
data_size	double	-	扫描的数据量，单位为字节。
billing_info	text	-	保留字段。



### 15.2.3 GS\_RESPOOL\_RESOURCE\_HISTORY

GS\_RESPOOL\_RESOURCE\_HISTORY表记录资源池监控历史信息，CN和DN上均进行记录。

表 15-3 GS\_RESPOOL\_RESOURCE\_HISTORY 字段

名称	类型	描述
timestamp	timestamp	资源池监控信息持久化时间
nodegroup	name	资源池所属逻辑集群名称，默认集群显示"installation"
rpname	name	资源池名称
cgroup	name	资源池关联控制组名称
ref_count	int	资源池引用作业数，作业经过资源池不管是否管控都会计数，仅CN上有效
fast_run	int	资源池快车道运行作业数，只在CN上有效
fast_wait	int	资源池快车道排队作业数，只在CN上有效
fast_limit	int	资源池快车道作业并发限制，只在CN上有效
slow_run	int	资源池慢车道运行作业数，只在CN上有效
slow_wait	int	资源池慢车道排队作业数，只在CN上有效
slow_limit	int	资源池慢车道作业并发限制，只在CN上有效
used_cpu	double	资源池5s监控周期内使用CPU个数平均值，保留小数点后2位 <ul style="list-style-type: none"> <li>• DN：显示当前DN上资源池使用的CPU个数</li> <li>• CN：显示所有DN上资源池使用CPU的累积和</li> </ul>
cpu_limit	int	资源池可用CPU的上限，CPU配额管控情况下为GaussDB可用CPU，CPU限额管控情况下为关联控制组CPU可用CPU <ul style="list-style-type: none"> <li>• DN：显示当前DN上资源池可用CPU上限</li> <li>• CN：显示所有DN上资源池可用CPU上限的累积和</li> </ul>
used_mem	int	资源池使用的内存大小，单位：'MB' <ul style="list-style-type: none"> <li>• DN：显示当前DN上资源池使用的内存大小</li> <li>• CN：显示所有DN上资源池使用内存的累积和</li> </ul>
estimate_memory	int	当前CN上，资源池运行作业的估算内存之和，只在CN上有效

名称	类型	描述
mem_limit	int	资源池可用内存上限，单位：'MB' <ul style="list-style-type: none"> <li>DN：显示当前DN上资源池可用内存上限</li> <li>CN：显示所有DN上资源池可用内存上限的累积和</li> </ul>
read_kbytes	bigint	资源池5s监控周期内逻辑读字节数，单位：'KB' <ul style="list-style-type: none"> <li>DN：显示当前DN上资源池逻辑读字节数</li> <li>CN：显示所有DN上资源池逻辑读字节的累积和</li> </ul>
write_kbytes	bigint	资源池5s监控周期内逻辑写字节数，单位：'KB' <ul style="list-style-type: none"> <li>DN：显示当前DN上资源池逻辑写字节数</li> <li>CN：显示所有DN上资源池逻辑写字节的累积和</li> </ul>
read_counts	bigint	资源池5s监控周期内逻辑读次数 <ul style="list-style-type: none"> <li>DN：显示当前DN上资源池逻辑读次数</li> <li>CN：显示所有DN上资源池逻辑读次数的累积和</li> </ul>
write_counts	bigint	资源池5s监控周期内逻辑写次数 <ul style="list-style-type: none"> <li>DN：显示当前DN上资源池逻辑写次数</li> <li>CN：显示所有DN上资源池逻辑写次数的累积和</li> </ul>
read_speed	double	资源池5s监控周期内逻辑读速率的平均值 <ul style="list-style-type: none"> <li>DN：显示当前DN上资源池逻辑读速率</li> <li>CN：显示所有DN上资源池逻辑读速率的累积和</li> </ul>
write_speed	double	资源池5s监控周期内逻辑写速率平均值 <ul style="list-style-type: none"> <li>DN：显示当前DN上资源池逻辑写速率</li> <li>CN：显示所有DN上资源池逻辑写速率的累积和</li> </ul>

## 15.2.4 GS\_WLM\_INSTANCE\_HISTORY

GS\_WLM\_INSTANCE\_HISTORY系统表存储与实例(CN或DN)相关的资源使用相关信息。该系统表里每条记录都是对应时间点某实例资源使用情况，包括：内存、CPU核数、磁盘IO、进程物理IO和进程逻辑IO信息。

表 15-4 GS\_WLM\_INSTANCE\_HISTORY 字段

名称	类型	描述
instancename	text	实例名称。
timestamp	timestamp with time zone	时间戳。
used_cpu	int	实例使用CPU所占用的百分比。
free_mem	int	实例未使用的内存大小，单位MB。
used_mem	int	实例已使用的内存大小，单位MB。
io_await	real	实例所使用磁盘的io_await值（10秒均值）。
io_util	real	实例所使用磁盘的io_util值（10秒均值）。
disk_read	real	实例所使用磁盘的读速率（10秒均值），单位KB/s。
disk_write	real	实例所使用磁盘的写速率（10秒均值），单位KB/s。
process_read	bigint	实例对应进程从磁盘读数据的读速率(不包括从磁盘pagecache中读取的字节数，10秒均值)，单位KB/s。
process_write	bigint	实例对应进程向磁盘写数据的写速率(不包括向磁盘pagecache中写入的字节数，10秒均值)，单位KB/s。
logical_read	bigint	CN实例：不统计。 DN实例：该实例在本次统计间隙（10秒）内逻辑读字节速率，单位KB/s。
logical_write	bigint	CN实例：不统计。 DN实例：该实例在本次统计间隙（10秒）内逻辑写字节速率，单位KB/s。
read_counts	bigint	CN实例：不统计。 DN实例：该实例在本次统计间隙（10秒）内逻辑读操作次数之和，单位次。
write_counts	bigint	CN实例：不统计。 DN实例：该实例在本次统计间隙（10秒）内逻辑写操作次数之和，单位次。

## 15.2.5 GS\_WLM\_OPERATOR\_INFO

GS\_WLM\_OPERATOR\_INFO系统表显示执行作业结束后的算子相关的记录。此数据是从内核中转储到系统表中的数据。当设置GUC参数`enable_resource_record`为on时，

系统会定时将GS\_WLM\_OPERATOR\_HISTORY中的记录导入此系统表，开启此功能会占用系统存储空间并对性能有一定影响，不建议用户使用。

**说明**

- 此系统表的schema是dbms\_om。
- 此系统表在gaussdb数据库中有分布列，分布列是queryid，其它数据库中无分布列。
- pg\_catalog下存在GS\_WLM\_OPERATOR\_INFO视图。

**表 15-5 GS\_WLM\_OPERATOR\_INFO 的字段**

名称	类型	描述
nodename	text	执行语句的CN实例名称。
queryid	bigint	语句执行使用的内部query_id。
pid	bigint	后端线程id。
plan_node_id	integer	查询对应的执行计划的plan node id。
plan_node_name	text	对应于plan_node_id的算子的名称。
start_time	timestamp with time zone	该算子处理第一条数据的开始时间。
duration	bigint	该算子到结束时候总的执行时间(ms)。
query_dop	integer	当前算子执行时的并行度。
estimated_rows	bigint	优化器估算的行数信息。
tuple_processed	bigint	当前算子返回的元素个数。
min_peak_memory	integer	当前算子在所有DN上的最小内存峰值(MB)。
max_peak_memory	integer	当前算子在所有DN上的最大内存峰值(MB)。
average_peak_memory	integer	当前算子在所有DN上的平均内存峰值(MB)。
memory_skew_percent	integer	当前算子在各DN间的内存使用倾斜率。
min_spill_size	integer	若发生下盘，所有下盘DN的最小下盘数据量(MB)，默认为0。
max_spill_size	integer	若发生下盘，所有下盘DN的最大下盘数据量(MB)，默认为0。
average_spill_size	integer	若发生下盘，所有下盘DN的平均下盘数据量(MB)，默认为0。

名称	类型	描述
spill_skew_percent	integer	若发生下盘，DN间下盘倾斜率。
min_cpu_time	bigint	该算子在所有DN上的最小执行时间(ms)。
max_cpu_time	bigint	该算子在所有DN上的最大执行时间(ms)。
total_cpu_time	bigint	该算子在所有DN上的总执行时间(ms)。
cpu_skew_percent	integer	DN间执行时间的倾斜率。
warning	text	主要显示如下几类告警信息： 1. Sort/SetOp/HashAgg/HashJoin spill 2. Spill file size large than 256MB 3. Broadcast size large than 100MB 4. Early spill 5. Spill times is greater than 3 6. Spill on memory adaptive 7. Hash table conflict

## 15.2.6 GS\_WLM\_SESSION\_INFO

GS\_WLM\_SESSION\_INFO系统表显示所有CN执行作业结束后的负载管理记录。此数据是从内核中转储到系统表中的数据。当设置GUC参数`enable_resource_record`为on时，系统会定时将GS\_WLM\_SESSION\_HISTORY中的记录导入此系统表，开启此功能会占用系统存储空间并对性能有一定影响，不建议用户使用。具体的字段请参考表15-137。

### 📖 说明

- 此系统表的schema是dbms\_om。
- 此系统表在gaussdb数据库中有分布列，分布列是queryid，其它数据库中无分布列。
- pg\_catalog下存在GS\_WLM\_SESSION\_INFO视图。

## 15.2.7 GS\_WLM\_USER\_RESOURCE\_HISTORY

GS\_WLM\_USER\_RESOURCE\_HISTORY系统表存储与用户使用资源相关的信息，该表在CN和DN上均存有数据。该系统表的每条记录都是对应时间点某用户的资源使用情况，包括：内存、CPU核数、存储空间、临时空间、算子落盘空间、逻辑IO流量、逻辑IO次数和逻辑IO速率信息。其中，内存、CPU、IO相关监控项仅记录用户复杂作业的资源使用情况。

GS\_WLM\_USER\_RESOURCE\_HISTORY系统表的数据来源于PG\_TOTAL\_USER\_RESOURCE\_INFO视图。

表 15-6 GS\_WLM\_USER\_RESOURCE\_HISTORY 字段

名称	类型	描述
username	text	用户名。
timestamp	timestamp with time zone	时间戳。
used_memory	int	正在使用的内存大小，单位MB。
total_memory	int	可以使用的内存大小，单位MB。值为0表示未限制最大可用内存，其限制取决于数据库最大可用内存。
used_cpu	real	正在使用的CPU核数。
total_cpu	int	该机器节点上，用户关联控制组的CPU核数总和。
used_space	bigint	已使用的存储空间大小，单位KB。
total_space	bigint	可使用的存储空间大小，单位KB，值为-1表示未限制最大存储空间。
used_temp_space	bigint	已使用的临时存储空间大小，单位KB。
total_temp_space	bigint	可使用的临时存储空间大小，单位KB，值为-1表示未限制最大临时存储空间。
used_spill_space	bigint	已使用的算子落盘存储空间大小，单位KB。
total_spill_space	bigint	可使用的算子落盘存储空间大小，单位KB，值为-1表示未限制最大算子落盘存储空间。
read_kbytes	bigint	监控周期内，读操作的字节流量，单位KB。
write_kbytes	bigint	监控周期内，写操作的字节流量，单位KB。
read_counts	bigint	监控周期内，读操作的次数，单位次。
write_counts	bigint	监控周期内，写操作的次数，单位次。
read_speed	real	监控周期内，读操作的字节速率，单位KB/s。
write_speed	real	监控周期内，写操作的字节速率，单位KB/s。

## 15.2.8 PG\_AGGREGATE

PG\_AGGREGATE系统表存储与聚集函数有关的信息。PG\_AGGREGATE里的每条记录都是一条pg\_proc里面的记录的扩展。PG\_PROC记录承载该聚集的名字、输入和输出数据类型，以及其它一些和普通函数类似的信息。

表 15-7 PG\_AGGREGATE 字段

名字	类型	引用	描述
aggfnoid	regproc	PG_PROC.oid	此聚集函数的PG_PROC OID。
aggtransfn	regproc	PG_PROC.oid	转换函数。
aggcollectfn	regproc	PG_PROC.oid	收集函数。
aggfinalfn	regproc	PG_PROC.oid	最终处理函数（如果没有则为0）。
aggstoptop	oid	PG_OPERATOR.oid	关联排序操作符（如果没有则为0）。
aggtranstype	oid	PG_TYPE.oid	此聚集函数的内部转换（状态）数据的数据类型。
agginitval	text	-	转换状态的初始值。这是一个文本数据域，它包含初始值的外部字符串表现形式。如果数据域是null，则转换状态值从null开始。
agginitcollect	text	-	收集状态的初始值。这是一个文本数据域，它包含初始值的外部字符串表现形式。如果数据域是null，则收集状态值从null开始。

## 15.2.9 PG\_AM

PG\_AM系统表存储有关索引访问方法的信息。系统支持的每种索引访问方法都有一行。

表 15-8 PG\_AM 字段

名字	类型	引用	描述
oid	oid	-	行标识符（隐藏属性，必须明确选择才会显示）。
amname	name	-	访问方法的名称。

名字	类型	引用	描述
amstrategies	smallint	-	访问方法的操作符策略个数，或者如果访问方法没有一个固定的操作符策略集则为0。
amsupport	smallint	-	访问方法的支持过程个数。
amcanorder	boolean	-	这种访问方式是否支持通过索引字段值的命令扫描排序。
amcanorderbyop	boolean	-	这种访问方式是否支持通过索引字段上操作符的结果的命令扫描排序。
amcanbackward	boolean	-	访问方式是否支持向后扫描。
amcanunique	boolean	-	访问方式是否支持唯一索引。
amcanmulticol	boolean	-	访问方式是否支持多字段索引。
amoptionalkey	boolean	-	访问方式是否支持第一个索引字段上没有任何约束的扫描。
amsearcharray	boolean	-	访问方式是否支持 ScalarArrayOpExpr 搜索。
amsearchnulls	boolean	-	访问方式是否支持 IS NULL/NOT NULL 搜索。
amstorage	boolean	-	允许索引存储的数据类型与列的数据类型是否不同。
amclusterable	boolean	-	是否允许在一个这种类型的索引上集群。
ampredlocks	boolean	-	是否允许这种类型的一个索引管理细粒度的谓词锁定。
amkeytype	oid	PG_TYPE.oid	存储在索引里数据的类型，如果不是一个固定的类型则为0。
aminsert	regproc	PG_PROC.oid	“插入此行”函数。
ambeginscan	regproc	PG_PROC.oid	“准备索引扫描”函数。
amgettupple	regproc	PG_PROC.oid	“下一个有效行”函数，如果没有则为0。
amgetbitmap	regproc	PG_PROC.oid	“抓取所有的有效行”函数，如果没有则为0。
amrescan	regproc	PG_PROC.oid	“（重新）开始索引扫描”函数。
amendscan	regproc	PG_PROC.oid	“索引扫描后清理”函数。
ammarkpos	regproc	PG_PROC.oid	“标记当前扫描位置”函数。
amrestrpos	regproc	PG_PROC.oid	“恢复已标记的扫描位置”函数。



名字	类型	引用	描述
ammerge	regproc	PG_PROC.oid	“归并多个索引对象”函数。
ambuild	regproc	PG_PROC.oid	“建立新索引”函数。
ambuildempty	regproc	PG_PROC.oid	“建立空索引”函数。
ambulkdelete	regproc	PG_PROC.oid	批量删除函数。
amvacuumclean up	regproc	PG_PROC.oid	VACUUM后的清理函数。
amcanreturn	regproc	PG_PROC.oid	检查是否索引支持唯一索引扫描的函数，如果没有则为0。
amcostestimate	regproc	PG_PROC.oid	估计一个索引扫描开销的函数。
amoptions	regproc	PG_PROC.oid	用于分析和验证索引的reloptions函数。

## 15.2.10 PG\_AMOP

PG\_AMOP系统表存储有关和访问方法操作符族关联的信息。如果一个操作符是一个操作符族中的成员，则在这个表中会占据一行。一个族成员是一个search操作符或一个ordering操作符。一个操作符可以在多个族中出现，但是不能在一个族中的多个搜索位置或多个排序位置中出现。

表 15-9 PG\_AMOP 字段

名字	类型	引用	描述
oid	oid	-	行标识符（隐藏属性，必须明确选择才会显示）。
amopfamily	oid	PG_OPFAMILY.oid	该项的操作符族。
amoplefttype	oid	PG_TYPE.oid	操作符的左输入类型。
amoprightright	oid	PG_TYPE.oid	操作符的右输入类型。
amopstrategy	smallint	-	操作符策略数。
amoppurpose	"char"	-	操作符目的，s为搜索或o为排序。
amopopr	oid	PG_OPERATOR.oid	该操作符的OID。
amopmethod	oid	PG_AM.oid	使用此操作符族的索引访问方法。
amopsortfamily	oid	PG_OPFAMILY.oid	如果是一个排序操作符，则该项会按照btree操作符族排序；如果是一个搜索操作符，则为0。

search操作符表明这个操作符族的一个索引可以被搜索，找到所有满足WHERE indexed\_column operator constant的行。显然，这样的操作符必须返回布尔值，并且它的左输入类型必须匹配索引的字段数据类型。

ordering操作符表明这个操作符族的一个索引可以被扫描，返回以ORDER BY indexed\_column operator constant顺序表示的行。这样的操作符可以返回任意可排序的数据类型，它的左输入类型也必须匹配索引的字段数据类型。ORDER BY的确切语义是由amopsortfamily字段指定的，该字段必须为操作符的返回类型引用一个btree操作符族。

## 15.2.11 PG\_AMPROC

PG\_AMPROC系统表存储有关与访问方法操作符族相关联的支持过程的信息。每个属于某个操作符族的支持过程都占有一行。

表 15-10 PG\_AMPROC 字段

名字	类型	引用	描述
oid	oid	-	行标识符（隐藏属性，必须明确选择才会显示）
amprocfamily	oid	<a href="#">PG_OPFAMILY.oid</a>	该项的操作符族
amproclefttype	oid	<a href="#">PG_TYPE.oid</a>	相关操作符的左输入数据类型
amprocrighttype	oid	<a href="#">PG_TYPE.oid</a>	相关操作符的右输入数据类型
amprocnum	smallint	-	支持过程编号
amproc	regproc	<a href="#">PG_PROC.oid</a>	过程的OID

amproclefttype和amprocrighttype字段的习惯解释，标识一个特定支持过程所支持的操作符的左右输入类型。对于某些访问方式，匹配支持过程本身的输入数据类型，对其他的则不会匹配。有一个对索引的“缺省”支持过程的概念，amproclefttype和amprocrighttype都等于索引操作符类的opcintype。

## 15.2.12 PG\_ATTRDEF

PG\_ATTRDEF系统表存储字段的默认值。

表 15-11 PG\_ATTRDEF 字段

名称	类型	描述
adrelid	oid	该字段所属的表
adnum	smallint	字段编号
adbin	pg_node_tree	字段缺省值的内部表现形式
adsrc	text	人类可读的缺省值的内部表现形式

名称	类型	描述
adbin_on_update	pg_node_tree	字段on_update_expr值的内部表现形式
adsrc_on_update	text	人类可读的on_update_expr值的内部表现形式

## 15.2.13 PG\_ATTRIBUTE

PG\_ATTRIBUTE系统表存储关于表字段的信息。

表 15-12 PG\_ATTRIBUTE 字段

名称	类型	描述
attrelid	oid	该字段所属的表。
attname	name	字段名。
atttypid	oid	字段类型。
attstattarget	integer	控制ANALYZE为该字段设置的统计细节的级别。 <ul style="list-style-type: none"> <li>零值表示不收集统计信息。</li> <li>负数表示使用系统缺省的统计对象。</li> <li>正数值的确切信息是和数据类型相关的。</li> </ul> 对于标量数据类型，ATTSTATTARGET既是要收集的“最常用数值”的目标数目，也是要创建的柱状图的目标数量。
attlen	smallint	是本字段类型pg_type.typlen的复制。
attnum	smallint	字段编号。
attndims	integer	如果该字段是数组，该值表示数组的维数，否则是0。
attcacheoff	integer	在磁盘上总是-1，但是如果加载入内存中的行描述器中，它可能会被更新为缓冲在行中字段的偏移量。
atttypmod	integer	记录创建新表时支持的类型特定的数据（比如，varchar字段的最大长度）。它传递给类型相关的输入和长度转换函数当做第三个参数。其值对那些不需要ATTTYPMOD的类型通常为-1。
attbyval	boolean	pg_type.typbyval字段值的复制。
attstorage	"char"	pg_type.typstorage字段值的复制。
attalign	"char"	pg_type.typalign字段值的复制。

名称	类型	描述
attnotnull	boolean	代表一个非空约束。可以改变这个字段来打开或者关闭该约束。
atthasdef	boolean	该字段是否存在缺省值，此时它对应pg_attrdef表里实际定义此值的记录。
attisdropped	boolean	该字段是否已经被删除，不再有效。如果被删除，该字段物理上仍然存在表中，但会被分析器忽略，因此不能再通过SQL访问。
attislocal	boolean	该字段是否局部定义在对象中。一个字段可以同时是局部定义和继承的。
attcmprmode	tinyint	对某一列指定压缩方式。压缩方式包括： <ul style="list-style-type: none"> <li>• ATT_CMPR_NOCOMPRESS</li> <li>• ATT_CMPR_DELTA</li> <li>• ATT_CMPR_DICTIONARY</li> <li>• ATT_CMPR_PREFIX</li> <li>• ATT_CMPR_NUMSTR</li> </ul>
attinhcount	integer	该字段所拥有的直接父表的个数。如果一个字段的父表个数非零，则它就不能被删除或重命名。
atcollation	oid	对此列定义的校对列。
attacl	aclitem[]	列级访问权限控制。
attoptions	text[]	属性级可选项。
attfdwoptions	text[]	属性级外数据选项。
attinitdefval	bytea	存储了此列默认的值表达式。行存表的ADD COLUMN需要使用此字段。
attkvtype	tinyint	该字段的kv_type属性。取值如下： <ul style="list-style-type: none"> <li>• 0表示默认值，用于非时序表</li> <li>• 1表示维度属性(TSTAG)，仅用于时序表</li> <li>• 2表示指标属性(TSFIELD)，仅用于时序表</li> <li>• 3时间属性(TSTIME)，仅用于时序表</li> </ul>

## 15.2.14 PG\_AUTHID

PG\_AUTHID系统表存储有关数据库认证标识符（角色）的信息。角色把“用户”的概念包含在内。一个用户实际上就是一个rolcanlogin标志被设置的角色。任何角色（不管rolcanlogin设置与否）都能够把其他角色作为成员。

在一个集群中只有一份pg\_authid，不是每个数据库有一份。需要有系统管理员权限才可以访问此系统表。

表 15-13 PG\_AUTHID 字段

名称	类型	描述
oid	oid	行标识符（隐藏属性，必须明确选择才会显示）。
rolname	name	角色名称。
rolsuper	boolean	角色是否是拥有最高权限的初始系统管理员。
rolinherit	boolean	角色是否自动继承其所属角色的权限。
rolcreatorole	boolean	角色是否可以创建更多角色。
rolcreatedb	boolean	角色是否可以创建数据库。
rolcatupdate	boolean	角色是否可以更新系统表。只有usesysid=10的初始系统管理员拥有此权限。其他用户无法获得此权限。
rolcanlogin	boolean	角色是否可以登录，即该角色是否能够作为初始会话授权标识符。
rolreplication	boolean	角色是一个复制的角色（适配作用，没有实际的功能）。
rolauditadmin	boolean	审计用户。
rolsystemadmin	boolean	管理员用户。
rolconnlimit	integer	对于可以登录的角色，限制其最大并发连接数量。-1 表示没有限制。
rolpassword	text	口令(可能是加密的)，如果没有口令，则为 NULL。
rolvalidbegin	timestamp with time zone	帐户的有效开始时间，如果没有开始时间，则为 NULL。
rolvaliduntil	timestamp with time zone	帐户的有效结束时间，如果没有结束时间，则为 NULL。
rolrespool	name	用户所能够使用的resource pool。
roluseft	boolean	角色是否可以操作外表。
rolparentid	oid	用户所在组用户的OID。
roltabspace	Text	用户永久表存储空间限额。
rolkind	char	特殊用户种类，包括私有用户、逻辑集群管理员、普通用户。
rolnodegroup	oid	用户所关联的Node Group OID，该Node Group 必须是逻辑集群。

名称	类型	描述
roltemp space	Text	用户临时表存储空间限额。
rolspill space	Text	用户算子落盘空间限额。
rolexcp data	text	保留字段未使用。
rolauthinfo	text	用户采用LDAP认证时的额外信息。如果是其他认证模式，则为NULL。
rolpwdexpire	integer	用户口令过期时间，在口令未过期时，用户自己修改口令。过期后请管理员修改口令。-1表示没有过期时间限制。
rolpwdtime	timestamp with time zone	口令的创建时间

## 15.2.15 PG\_AUTH\_HISTORY

PG\_AUTH\_HISTORY系统表记录了角色的认证历史。需要有系统管理员权限才可以访问此系统表。

表 15-14 PG\_AUTH\_HISTORY 字段

名称	类型	描述
roloid	oid	角色标识
passwordtime	timestamp with time zone	创建和修改密码的时间
rolpassword	text	角色密码，使用MD5、SHA256加密或者不加密

## 15.2.16 PG\_AUTH\_MEMBERS

PG\_AUTH\_MEMBERS系统表存储显示角色之间的成员关系。

表 15-15 PG\_AUTH\_MEMBERS 字段

名称	类型	描述
roleid	oid	拥有成员的角色ID
member	oid	属于ROLEID角色的成员角色ID
grantor	oid	赋予此成员关系的角色ID

名称	类型	描述
admin_option	boolean	如果有权限可以把ROLEID角色的成员关系赋予其他角色，则为真

## 15.2.17 PG\_CAST

PG\_CAST系统表存储数据类型之间的转化关系。

表 15-16 PG\_CAST 字段

名称	类型	描述
castsource	oid	源数据类型的OID。
casttarget	oid	目标数据类型的OID。
castfunc	oid	转化函数的OID。0表示不需要转化函数。
castcontext	"char"	源数据类型和目标数据类型间的转化方式： <ul style="list-style-type: none"> <li>• e表示只能进行显式转化（使用CAST或::语法）。</li> <li>• i表示只能进行隐式转化。</li> <li>• a表示类型间同时支持隐式和显式转化。</li> </ul>
castmethod	"char"	转化方法： <ul style="list-style-type: none"> <li>• f表示使用castfunc字段中指定的函数进行转化。</li> <li>• b表示类型间是二进制强制转化，不使用castfunc。</li> </ul>

## 15.2.18 PG\_CLASS

PG\_CLASS系统表存储数据库对象信息及其之间的关系。

表 15-17 PG\_CLASS 字段

名称	类型	描述
oid	oid	行标识符（隐藏属性，必须明确选择才会显示）。
relname	name	表、索引、视图等对象的名称。
relnamespace	oid	包含该关系的命名空间的OID。
reltype	oid	对应该表的行类型的数据类型（索引为零，因为索引没有pg_type记录）。
reloftype	oid	复合类型的OID，0表示其他类型。
relowner	oid	关系所有者。

名称	类型	描述
relam	oid	如果行是索引，则就是所用的访问模式（B-tree，hash等）。
relfilenode	oid	该关系在磁盘上的文件的名称，如果没有则为0。
reltablespace	oid	该关系存储所在的表空间。如果为0，则使用该数据库的缺省表空间。如果关系无磁盘文件，该字段无意义。
relpages	double precision	以页(大小为BLCKSZ)为单位的此表在磁盘上的大小，只是优化器使用的一个近似值。
reltuples	double precision	表中行的数目，只是优化器使用的一个估计值。
relallvisible	integer	被标识为全可见的表中的页数。此字段是优化器用来做SQL执行优化使用的。VACUUM、ANALYZE和一些DDL语句（例如，CREATE INDEX）会引起此字段更新。
reltoastrelid	oid	与此表关联的TOAST表的OID，如果没有则为0。TOAST表在一个从属表里“离线”存储大字段。
reltoastidxid	oid	对于TOAST表是它的索引的OID，如果不是TOAST表则为0。
reldeltarelid	oid	Delta表的OID。 Delta表附属于列存表。用于存储数据导入过程中的甩尾数据。
reldeltaidx	oid	Delta表的索引表OID。
relcudescrelid	oid	CU描述表的OID。 CU描述表（Desc表）附属于列存表。用于控制表目录中存储数据的可见性。
relcudescidx	oid	CU描述表的索引表OID。
relhasindex	boolean	如果对象是一个表且至少有（或者最近建有）一个索引，则为真。 由CREATE INDEX设置，但DROP INDEX不会立即将它清除。如果VACUUM进程检测一个表没有索引，会清理relhasindex字段，将relhasindex值设置为假。
relisshared	boolean	如果该表在整个集群中由所有数据库共享则为真。只有某些系统表（比如pg_database）是共享的。
relpersistence	"char"	<ul style="list-style-type: none"> <li>• p表示永久表。</li> <li>• u表示非日志表。</li> <li>• t表示临时表。</li> </ul>



名称	类型	描述
relkind	"char"	<ul style="list-style-type: none"> <li>• r表示普通表。</li> <li>• i表示索引。</li> <li>• S表示序列。</li> <li>• v表示视图。</li> <li>• c表示复合类型。</li> <li>• t表示TOAST表。</li> <li>• f表示外表。</li> </ul>
relnatts	smallint	关系中用户字段数目（除了系统字段以外）。在pg_attribute里肯定有相同数目对应行。
relchecks	smallint	表上检查约束的数目，参见PG_CONSTRAINT。
relhasoids	boolean	如果为关系中每行都生成一个OID，则为真。
relhaspkey	boolean	如果该表有一个（或曾有）主键，则为真。
relhasrules	boolean	如果表有规则，则为真。是否有规则可参考系统表PG_REWRITE。
relhastriggers	boolean	如果表有（或曾有）触发器，则为真。参见PG_TRIGGER。
relhassubclass	boolean	如果表有（或曾有）任何继承的子表，则为真。
relcmprs	tinyint	<p>表示是否启用表的压缩特性。需要特别注意，当且仅当批量插入才会触发压缩，普通的CRUD并不能够触发压缩。</p> <ul style="list-style-type: none"> <li>• 0表示其他不支持压缩的表（主要是指系统表，不支持压缩属性的修改操作）。</li> <li>• 1表示表数据的压缩特性为NOCOMPRESS或者无指定关键字。</li> <li>• 2表示表数据的压缩特性为COMPRESS。</li> </ul>
relhasclusterkey	boolean	是否有局部聚簇存储。
relrowmovement	boolean	<p>针对分区表进行update操作时，是否允许行迁移。</p> <ul style="list-style-type: none"> <li>• true：表示允许行迁移。</li> <li>• false：表示不允许行迁移。</li> </ul>
parttype	"char"	<p>表或者索引是否具有分区表的性质。</p> <ul style="list-style-type: none"> <li>• p表示带有分区表性质。</li> <li>• n表示没有分区表特性。</li> <li>• v表示该表为HDFS的Value分区表。</li> </ul>

名称	类型	描述
relfrozenxid	xid32	该表中所有在此之间的事务ID已经被替换为一个固定的 ("frozen") 事务ID。该字段用于跟踪表是否需要为了防止事务ID重叠 ( 或者允许收缩pg_clog ) 而进行清理。如果该关系不是表则为0 ( InvalidTransactionId )。 为保持前向兼容, 保留此字段, 新增relfrozenxid64用于记录此信息。
relacl	aclitem[]	访问权限。 查询的回显结果为以下形式: rolename=xxxx/yyyy --赋予一个角色的权限 =xxxx/yyyy --赋予public的权限 xxxx表示赋予的权限, yyyy表示授予该权限的角色。 权限的参数说明请参见表15-18。
reloptions	text[]	特定的访问方法选项, 用"keyword=value"字符串形式表示。
relfrozenxid64	xid	该表中所有在此之前的事务ID已经被替换为一个固定的 ("frozen") 事务ID。该字段用于跟踪表是否需要为了防止事务ID重叠 ( 或者允许收缩pg_clog ) 而进行清理。如果该关系不是表则为0 ( InvalidTransactionId )。

表 15-18 权限的参数说明

参数	参数说明
r	SELECT ( 读 )
w	UPDATE ( 写 )
a	INSERT ( 插入 )
d	DELETE
D	TRUNCATE
x	REFERENCES
t	TRIGGER
X	EXECUTE
U	USAGE
C	CREATE
c	CONNECT
T	TEMPORARY
A	ANALYZE ANALYSE

参数	参数说明
L	ALTER
P	DROP
v	VACUUM
arwdDxtA, vLP	ALL PRIVILEGES (用于表)
*	给前面权限的授权选项

## 应用示例

查看某张表的oid及relfilenode：

```
select oid,relname,relfilenode from pg_class where relname = 'table_name';
```

统计行存表数量：

```
select 'row count:'||count(1) as point from pg_class where relkind = 'r' and oid > 16384 and reloptions::text not like '%column%' and reloptions::text not like '%internal_mask%';
```

统计列存表数量：

```
select 'column count:'||count(1) as point from pg_class where relkind = 'r' and oid > 16384 and reloptions::text like '%column%';
```

## 15.2.19 PG\_COLLATION

PG\_COLLATION系统表描述可用的排序规则，本质上从一个SQL名字映射到操作系统本地类别。

表 15-19 PG\_COLLATION 字段

名字	类型	引用	描述
oid	oid	-	行标识符（隐藏属性，必须明确选择才会显示）
collname	name	-	排序规则名（每个命名空间和编码唯一）
collnamespace	oid	<a href="#">PG_NAMESPACE</a> .oid	包含该排序规则的命名空间的OID
collowner	oid	<a href="#">PG_AUTHID</a> .oid	排序规则的所有者
collencoding	integer	-	排序规则可用的编码，如果适用于任意编码为-1。 <b>说明</b> 使用函数pg_encoding_to_char()可以将编号转换为对应的编码名称。
collcollate	name	-	排序规则对象的LC_COLLATE

名字	类型	引用	描述
collctype	name	-	排序规则对象的LC_CTYPE

## 15.2.20 PG\_CONSTRAINT

PG\_CONSTRAINT系统表存储表上的检查约束、主键、唯一约束和外键约束。

表 15-20 PG\_CONSTRAINT 字段

名称	类型	描述
conname	name	约束名称（不一定是唯一的）。
connamespace	oid	包含约束的命名空间的OID。
contype	"char"	<ul style="list-style-type: none"> <li>• c = 检查约束</li> <li>• f = 外键约束</li> <li>• p = 主键约束</li> <li>• u = 唯一约束</li> <li>• t = 触发器约束</li> </ul>
condeferrable	boolean	该约束是否可以推迟。
condeferred	boolean	缺省时该约束是否可以推迟。
convalidated	boolean	约束是否有效。目前，只有外键和CHECK约束可将其设置为FALSE。
conrelid	oid	该约束所在的表；如果不是表约束则为0。
contypid	oid	该约束所在的域；如果不是一个域约束则为0。
conindid	oid	与约束关联的索引ID。
confrelid	oid	如果是外键，则为参考的表；否则为0。
confupdtype	"char"	外键更新动作代码。 <ul style="list-style-type: none"> <li>• a = 没动作</li> <li>• r = 限制</li> <li>• c = 级联</li> <li>• n = 设置为null</li> <li>• d = 设置为缺省</li> </ul>

名称	类型	描述
confdeltype	"char"	外键删除动作代码。 <ul style="list-style-type: none"> <li>• a = 没动作</li> <li>• r = 限制</li> <li>• c = 级联</li> <li>• n = 设置为null</li> <li>• d = 设置为缺省</li> </ul>
confmatchtype	"char"	外键匹配类型。 <ul style="list-style-type: none"> <li>• f = 全部</li> <li>• p = 部分</li> <li>• u = 简单 (未指定)</li> </ul>
conislocal	boolean	是否是为关系创建的本地约束。
coninhcount	integer	约束直接继承父表的数目。继承父表数非零时，不能删除或重命名该约束。
connoinherit	boolean	是否可以被继承。
consoft	boolean	是否为信息约束 (Informational Constraint)。
conopt	boolean	是否使用信息约束优化执行计划。
conkey	smallint[]	如果是表约束，则是约束控制的字段列表。
confkey	smallint[]	如果是一个外键，则是参考的字段的列表。
conpfeqop	oid[]	如果是一个外键，是做PK=FK比较的相等操作符ID的列表。
conppeqop	oid[]	如果是一个外键，是做PK=PK比较的相等操作符ID的列表。
conffeqop	oid[]	如果是一个外键，是做FK=FK比较的相等操作符ID的列表。
conexclp	oid[]	如果是一个排他约束，是列的排他操作符ID列表。
conbin	pg_node_tree	如果是检查约束，则是其表达式的内部形式。
consrc	text	如果是检查约束，则是表达式的人类可读形式。

**须知**

- 当被引用的对象改变时，consrc不能被更新。例如，它不会跟踪字段的重命名。最好还是使用pg\_get\_constraintdef()来抽取一个检查约束的定义，而不是依赖这个字段。
- pg\_class.relchecks需要和每个关系在此目录中的检查约束数量保持一致。

## 15.2.21 PG\_CONVERSION

PG\_CONVERSION系统表描述编码转换信息。

表 15-21 PG\_CONVERSION 字段

名称	类型	引用	描述
oid	oid	-	行标识符（隐藏属性，必须明确选择才会显示）
conname	name	-	转换名（在一个命名空间里唯一）
connamespace	oid	<a href="#">PG_NAMESPACE</a> .oid	包含此转换的命名空间的OID
conowner	oid	<a href="#">PG_AUTHID</a> .oid	编码转换的属主
conforencoding	integer	-	源编码ID
contoencoding	integer	-	目的编码ID
conproc	regproc	<a href="#">PG_PROC</a> .oid	转换过程
condefault	boolean	-	如果为缺省转换则为真

## 15.2.22 PG\_DATABASE

PG\_DATABASE系统表存储关于可用数据库的信息。

表 15-22 PG\_DATABASE 字段

名称	类型	描述
datname	name	数据库名称。
datdba	oid	数据库所有者，通常为其创建者。
encoding	integer	数据库的字符编码方式。 pg_encoding_to_char()可以将此编号转换为编码名称。
datcollate	name	数据库使用的排序顺序。

名称	类型	描述
datctype	name	数据库使用的字符分类。
datistemplate	boolean	是否允许作为模板数据库。
dataallowconn	boolean	如果为假，则没有用户可以连接到这个数据库。这个字段用于保护template0数据库不被更改。
datconlimit	integer	该数据库上允许的最大并发连接数，-1表示无限制。
datlastsysoid	oid	数据库里最后一个系统OID。
datfrozensid	xid32	用于跟踪该数据库是否需要为了防止事务ID重叠而进行清理。 为保持前向兼容，保留此字段，新增datfrozensid64用于记录此信息。
dattablespace	oid	数据库的缺省表空间。
datcompatibility	name	数据库兼容模式。
datacl	aclitem[]	访问权限。
datfrozensid64	xid	用于跟踪该数据库是否需要为了防止事务ID重叠而进行清理。

### 15.2.23 PG\_DB\_ROLE\_SETTING

PG\_DB\_ROLE\_SETTING系统表存储数据库运行时每个角色与数据绑定的配置项的默认值。

表 15-23 PG\_DB\_ROLE\_SETTING 字段

名称	类型	描述
setdatabase	oid	配置项所对应的数据库，如果未指定数据库，则为0。
setrole	oid	配置项所对应的角色，如果未指定角色，则为0。
setconfig	text[]	运行时配置项的默认值。

### 15.2.24 PG\_DEFAULT\_ACL

PG\_DEFAULT\_ACL系统表存储为新建对象设置的初始权限。

表 15-24 PG\_DEFAULT\_ACL 字段

名称	类型	描述
defaclrole	oid	与此权限相关的角色ID。
defaclnamespace	oid	与此权限相关的命名空间，如果没有，则为0。
defaclobjtype	"char"	此权限的对象类型。 <ul style="list-style-type: none"> <li>• r表示表或视图。</li> <li>• S表示序列。</li> <li>• f表示函数。</li> <li>• T表示类型。</li> </ul>
defaclacl	aclitem[]	创建该类型时所拥有的访问权限。

## 应用示例

查看新建用户role1的初始权限：

```
select * from PG_DEFAULT_ACL;
defaclrole | defaclnamespace | defaclobjtype | defaclacl
-----+-----+-----+-----
16820 | 16822 | r | {role1=r/user1}
```

也可使用如下语句进行转换后更直观的查看：

```
SELECT pg_catalog.pg_get_userbyid(d.defaclrole) AS "Granter", n.nspname AS "Schema", CASE
d.defaclobjtype WHEN 'r' THEN 'table' WHEN 'S' THEN 'sequence' WHEN 'f' THEN 'function' WHEN 'T'
THEN 'type' END AS "Type", pg_catalog.array_to_string(d.defaclacl, E', ') AS "Access privileges" FROM
pg_catalog.pg_default_acl d LEFT JOIN pg_catalog.pg_namespace n ON n.oid = d.defaclnamespace ORDER
BY 1, 2, 3;
```

输出结果如下，表示通过用户user1授予用户role1对模式“user1”有读的权限。

```
Granter | Schema | Type | Access privileges
-----+-----+-----+-----
user1 | user1 | table | role1=r/user1
(1 row)
```

## 15.2.25 PG\_DEPEND

PG\_DEPEND系统表记录数据库对象之间的依赖关系。这些信息允许DROP命令找出哪些其它对象必须由DROP CASCADE删除，或者是在DROP RESTRICT的情况下避免删除。

另请参考PG\_SHDEPEND，对于记录那些在数据库集群之间共享的对象之间的依赖性关系提供了相似的功能。

表 15-25 PG\_DEPEND 字段

名称	类型	引用	描述
classid	oid	<a href="#">PG_CLASS.oid</a>	依赖对象所在系统表的OID。



名称	类型	引用	描述
objid	oid	任意OID属性	指定依赖对象的OID。
objsubid	integer	-	对于表字段，是该属性的字段数（objid和classid引用表本身）。对于所有其它对象类型，此字段是0。
refclassid	oid	<a href="#">PG_CLASS</a> .oid	被引用对象所在的系统表的OID。
refobjid	oid	任意OID属性	指定的被引用对象的OID。
refobjsubid	integer	-	对于表字段，是该字段的字段号（refobjid和refclassid引用表本身）。对于所有其它对象类型，此字段是零。
deptype	"char"	-	定义此依赖关系特定语义的代码。

在所有情况下，一个PG\_DEPEND记录表示被引用对象不能在没有删除依赖对象的情况下被删除。但是其中也有几种由deptype定义的情况：

- **DEPENDENCY\_NORMAL (n)**：独立创建的对象之间的一般关系。依赖对象可以在不影响被引用对象的情况下删除。被引用对象只能通过指定CASCADE被删除，这种情况下依赖对象也会被删除。例如：一个表字段对其数据类型有一般依赖关系。
- **DEPENDENCY\_AUTO (a)**：依赖对象可以和被引用对象分别删除，且在被引用对象被删除时应自动被删除（不管是RESTRICT或CASCADE模式）。例如：一个表上的命名约束是该表上的自动依赖关系，因此如果删除了表，它也会被删除。
- **DEPENDENCY\_INTERNAL (i)**：依赖对象作为被引用对象过程的一部分创建，且是其内部实现的一部分。DROP依赖对象是不会直接允许的（会给户发出一个针对被引用对象的DROP）。不管是否指定CASCADE，一个被引用对象的DROP将被传播来删除其依赖对象。例如：一个用于强制外键约束的触发器将被设置为内部依赖于其约束的[PG\\_CONSTRAINT](#)项。
- **DEPENDENCY\_EXTENSION (e)**：依赖对象作为被依赖对象extension的一个成员（请参见[PG\\_EXTENSION](#)）。依赖对象可以通过在被依赖对象上DROP EXTENSION删除。在功能上，这种依赖类型和内部依赖的作用相同，其存在只是为了清晰和简化gs\_dump。
- **DEPENDENCY\_PIN (p)**：没有依赖对象。这种类型的记录标志着系统本身依赖于被引用对象，因此这个对象决不能被删除。这种类型的记录只有在initdb的时候创建。有依赖对象的字段都为0。

## 应用示例

查询名为serial1的数据库对象sequence和哪个表有依赖关系。

1. 先通过系统表PG\_CLASS查询序列名为serial1的oid。

```
SELECT oid FROM pg_class WHERE relname = 'serial1';
oid
-----
17815
(1 row)
```

2. 使用系统表PG\_DEPEND根据所查询的序列serial1的oid获取依赖该序列的对象。

```
SELECT * FROM pg_depend WHERE objid = '17815';
classid | objid | objsubid | refclassid | refobjid | refobjsubid | deptype
```

```
-----+-----+-----+-----+-----+-----+-----
1259 | 17815 | 0 | 2615 | 2200 | 0 | n
1259 | 17815 | 0 | 1259 | 17812 | 1 | a
(2 rows)
```

3. 根据字段refobjid获取依赖该序列serial1的表的OID，并查询到表名。其结果表示，序列serial1依赖于表customer\_address。

```
SELECT relname FROM pg_class where oid='17812';
```

```
relname
-----
customer_address
(1 row)
```

## 15.2.26 PG\_DESCRIPTION

PG\_DESCRIPTION系统表可以给每个数据库对象存储一个可选的描述（注释）。许多内置的系统对象的描述提供了PG\_DESCRIPTION的初始内容。

这个表的功能类似PG\_SHDESCRIPTION，用于记录整个集群范围内共享对象的注释。

表 15-26 PG\_DESCRIPTION 字段

名称	类型	引用	描述
objoid	oid	任意OID属性	描述所属对象的OID。
classoid	oid	PG_CLASS.oid	对象显示的系统表的OID。
objsubid	integer	-	对于一个表字段的注释，为字段号（objoid和classoid指向表自身）。对于其它对象类型，为0。
description	text	-	对该对象描述的任意文本。

## 15.2.27 PG\_ENUM

PG\_ENUM系统表包含显示每个枚举类型值和标签的记录。给定枚举类型的内部表示实际上是PG\_ENUM里面相关行的OID。

表 15-27 PG\_ENUM 字段

名称	类型	引用	描述
oid	oid	-	行标识符（隐藏属性，必须明确选择才会显示）
enumtypid	oid	PG_TYPE.oid	包含此枚举值的pg_type项的OID
enumsortorder	real	-	此枚举值在其枚举类型中的排序位置
enumlabel	name	-	此枚举值的文本标签

PG\_ENUM行的OID值遵循一种特殊规则：OID的数值被保证按照其枚举类型一样的排序顺序排序。即如果两个偶数OID属于同一枚举类型，那么较小的OID必然具有较小

enumsortorder值。奇数OID不需要遵循排序顺序。这种规则使得枚举比较例程在很多常见情况下可以避免系统目录查找。创建和修改枚举类型的例程尝试尽可能地为枚举值分配偶数OID。

当一个枚举类型被创建后，其成员会被分配排序顺序位置1到n。但是后面增加的成员可能会分配负值或者分数值的enumsortorder。对于这些值的唯一要求是它们必须被正确的排序且在每个枚举类型中保持唯一。

## 15.2.28 PG\_EXCEPT\_RULE

PG\_EXCEPT\_RULE系统表存储关于异常规则的信息。一个异常规则集合由多个名称相同的异常规则组成。

表 15-28 PG\_EXCEPT\_RULE

名称	类型	描述
name	name	异常规则集合的名称。
rule	name	该异常规则集中某一个具体规则类型或者触发当前异常规则集时采取的操作。（如blocktime/elapsedtime/spillsize等类型或者触发异常规则后的操作）
value	name	该异常规则对应的规则阈值。如果是触发异常规则后的操作，那么该字段为abort。

## 15.2.29 PG\_EXTENSION

PG\_EXTENSION系统表存储关于所安装扩展的信息。GaussDB(DWS)默认有十二个扩展，即PLPGSQL、DIST\_FDW、FILE\_FDW、HDFS\_FDW、HSTORE、PLDBGAPI、DIMSEARCH、PACKAGES、GC\_FDW、UUID-OSSP、LOG\_FDW和ROACH\_API。

表 15-29 PG\_EXTENSION

名称	类型	描述
extname	name	扩展名
extowner	oid	扩展的所有者
extnamespace	oid	扩展导出对象的命名空间
extrelocatable	boolean	如果扩展能够重定位到其他schema，则为true
extversion	text	扩展的版本号
extconfig	oid[]	扩展的配置信息
extcondition	text[]	扩展配置信息的过滤条件

### 15.2.30 PG\_EXTENSION\_DATA\_SOURCE

PG\_EXTENSION\_DATA\_SOURCE系统表存储外部数据源对象的信息。一个外部数据源对象（Data Source）包含了外部数据库的一些口令编码等信息，主要配合Extension Connector使用。

表 15-30 PG\_EXTENSION\_DATA\_SOURCE 字段

名字	类型	引用	描述
oid	oid	-	行标识符（隐藏属性，必须明确选择才会显示）。
srcname	name	-	外部数据源对象的名称。
srcowner	oid	PG_AUTH ID.oid	外部数据源对象的所有者。
srctype	text	-	外部数据源对象的类型，缺省为空。
srcversion	text	-	外部数据源对象的版本，缺省为空。
srcacl	aclitem[]	-	访问权限。
srcoptions	text[]	-	外部数据源对象的指定选项，使用“keyword=value”格式的字符串。

### 15.2.31 PG\_FINE\_DR\_INFO

PG\_FINE\_DR\_INFO系统表用于记录细粒度容灾备表的回放状态。该系统表仅8.2.0.100及以上集群版本支持。

表 15-31 PG\_FINE\_DR\_INFO 字段

名字	类型	描述
oid	oid	行标识符（隐藏属性，必须明确选择才会显示）。
relid	oid	细粒度容灾备表的oid。
lastcsn	xid	上一次成功回放时的csn。
lastxmin	xid	上一次成功回放时的xmin。
lastxmax	xid	上一次成功回放时的xmax。
laststarttime	timestamp with time zone	上一次成功回放的开始时间。
lastendtime	timestamp with time zone	上一次成功回放的结束时间。

## 应用示例

在容灾集群上查看备表回放状态：

```
SELECT * FROM pg_fine_dr_info;
relid | lastcsn | lastxmin | lastxmax | laststarttime | lastendtime
-----+-----+-----+-----+-----+-----
21132 | 1251610 | 1251609 | 1251611 | 2023-01-04 20:51:58.375136+08 | 2023-01-04 20:51:58.393986+08
(1 row)
```

### 15.2.32 PG\_FOREIGN\_DATA\_WRAPPER

PG\_FOREIGN\_DATA\_WRAPPER系统表存储外部数据封装器定义。一个外部数据封装器是在外部服务器上驻留外部数据的机制，是可以访问的。

表 15-32 PG\_FOREIGN\_DATA\_WRAPPER 字段

名字	类型	引用	描述
oid	oid	-	行标识符(隐藏属性，必须明确选择才会显示)。
fdwname	name	-	外部数据封装器名。
fdwowner	oid	PG_AUTHID.oid	外部数据封装器的所有者。
fdwhandler	oid	PG_PROC.oid	引用一个负责为外部数据封装器提供扩展例程的处理函数。如果没有提供处理函数则为0。
fdwvalidator	oid	PG_PROC.oid	引用一个验证器函数，这个验证器函数负责验证给予外部数据封装器的选项、外部服务器选项和使用外部数据封装器的用户映射的有效性。如果没有提供验证器函数则为0。
fdwacl	aclitem[]	-	访问权限。
fdwoptions	text[]	-	外部数据封装器指定选项，使用“keyword=value”格式的字符串。

### 15.2.33 PG\_FOREIGN\_SERVER

PG\_FOREIGN\_SERVER系统表存储外部服务器定义。一个外部服务器描述了一个外部数据源，例如一个远程服务器。外部服务器通过外部数据封装器访问。

表 15-33 PG\_FOREIGN\_SERVER 字段

名字	类型	引用	描述
oid	oid	-	行标识符（隐藏属性，必须明确选择才会显示）
srvname	name	-	外部服务器名

名字	类型	引用	描述
srvowner	oid	<a href="#">PG_AUTHID.oid</a>	外部服务器的所有者
srvfdw	oid	<a href="#">PG_FOREIGN_DATA_WRAPPER.oid</a>	此外部服务器的外部数据封装器的OID
srvtype	text	-	服务器的类型（可选）
srvversion	text	-	服务器的版本（可选）
srvacl	aclitem[]	-	访问权限
srvoptions	text[]	-	外部服务器指定选项，使用“keyword=value”格式的字符串

## 15.2.34 PG\_FOREIGN\_TABLE

PG\_FOREIGN\_TABLE系统表存储外部表的辅助信息。

表 15-34 PG\_FOREIGN\_TABLE 字段

名称	类型	描述
ftrelid	oid	外部表的OID
ftserver	oid	外部表的所在服务器的OID
ftwriteonly	boolean	外部表是否可写
ftoptions	text[]	外部表的可选项

## 15.2.35 PG\_INDEX

PG\_INDEX系统表存储索引的一部分信息，其他的信息大多数在PG\_CLASS中。

表 15-35 PG\_INDEX 字段

名称	类型	描述
indexrelid	oid	此索引的pg_class项的OID。
indrelid	oid	使用该索引的表在pg_class项的OID。
indnatts	smallint	索引中的字段数目。
indisunique	boolean	如果为真，为唯一索引。
indisprimary	boolean	如果为真，该索引为该表的主键。该字段为真时，indisunique也总是为真。

名称	类型	描述
indisexclusion	boolean	如果为真，该索引支持排他约束。
indimmediate	boolean	如果为真，在插入数据时会立即执行唯一性检查。
indisclustered	boolean	如果为真，则该表最后以此索引进行了聚簇。
indisusable	boolean	如果为真，此索引对INSERT/SELECT可用。
indisvalid	boolean	如果为真，则此索引可以用于查询。如果为假，则该索引可能不完整，仍然必须在INSERT/UPDATE操作时进行更新，但不能安全的被用于查询。如果是唯一索引，则唯一属性也不为真。
indcheckxmin	boolean	如果为真，查询不能使用此索引，直到pg_index此行的xmin低于其快照的TransactionXmin，因为该表可能包含它们可见的不兼容行断开的热链。
indisready	boolean	如果为真，表示此索引对插入数据可用。如果为假，在插入或修改数据时忽略此索引。
indkey	int2vector	这是一个包含indnatts值的数组，这些数组值表示此索引所建立的表字段。比如一个值为1 3的意思是第一个字段和第三个字段组成这个索引键。数组中的0表示对应的索引属性是一个表字段上的表达式，而不是一个简单的字段引用。
indcollation	oidvector	索引用到的各列的ID。
indclass	oidvector	对于索引键中的每个字段，该字段都包含要使用的操作符类的OID，详见PG_OPCLASS。
indoption	int2vector	存储列前标识位，该标识位是由索引的访问方法定义。
indexprs	pg_node_tree	表达式树（以nodeToString()形式表现）用于那些非简单字段引用的索引属性。它是一个列表，个数与INDKEY中的零值个数相同。如果所有索引属性都是简单的引用，则为空。
indpred	pg_node_tree	部分索引谓词的表达式树（以nodeToString()的形式表现）。如果不是部分索引，则为空。

名称	类型	描述
indnullstreatment	tinyint	<p>表示唯一索引中NULL值的处理方式，该字段只有当indisunique为真时才起作用。</p> <p>取值范围：</p> <ul style="list-style-type: none"> <li>0 表示NULLS DISTINCT，NULL互不相等，NULL值可重复插入。</li> <li>1 表示NULLS NOT DISTINCT，NULL严格相等，NULL值不可重复插入。</li> <li>2 表示NULLS IGNORE，在等值比较时忽略NULL值列。若索引列全为NULL，则NULL值可重复插入；若部分索引列为NULL，只有非NULL列不相等，才可插入。</li> </ul> <p>默认值：0</p> <p>说明</p> <ul style="list-style-type: none"> <li>若当前集群为低版本升级到8.2.0.100版本，对于之前已存在索引，该字段值均为NULL；对于新创建的索引，则根据[ NULLS [NOT] DISTINCT   NULLS IGNORE ]字段的指定情况确定该字段值，默认值为0。</li> <li>若当前集群为新装的8.2.0.100版本，对于新创建的索引，则根据[ NULLS [NOT] DISTINCT   NULLS IGNORE ]字段的指定情况确定该字段值，默认值为0。</li> </ul>

## 15.2.36 PG\_INHERITS

PG\_INHERITS系统表记录关于表继承层次的信息。数据库里每个直接的子系表都有一条记录。间接的继承可以通过追溯记录链来判断。

表 15-36 PG\_INHERITS 字段

名字	类型	引用	描述
inhrelid	oid	<a href="#">PG_CLASS.oid</a>	子表的OID。
inhparent	oid	<a href="#">PG_CLASS.oid</a>	父表的OID。
inhseqno	integer	-	如果一个子表存在多个直系父表（多重继承），这个数字表示此继承字段的排列顺序。计数从1开始。

## 15.2.37 PG\_JOBS

PG\_JOBS系统表存储用户创建的定时任务的任务详细信息，定时任务线程定时轮询pg\_jobs系统表中的时间，当任务到期会触发任务的执行。该系统表属于Shared Relation，所有创建的job记录对所有数据库可见。



表 15-37 PG\_JOBS 字段

名字	类型	描述
job_id	integer	作业ID, 主键, 是唯一的 (有唯一索引)。
what	text	作业内容。
log_user	oid	创建者的UserID。
priv_user	oid	作业执行者的UserID。
job_db	oid	标识作业执行的数据库OID。
job_nsp	oid	标识作业运行时所在的命名空间OID。
job_node	oid	标识当前作业是在哪个CN上创建和执行。
is_broken	boolean	标识当前作业是否为失效状态, 当作业连续失败16次后, 会将is_broken自动设置为true, 后续不再执行该作业。
start_date	timestamp without time zone	作业第一次开始执行时间, 时间精确到毫秒。
next_run_date	timestamp without time zone	下次定时执行任务的时间, 时间精确到毫秒。
failure_count	smallint	失败计数, 作业连续执行失败16次, 不再继续执行。
interval	text	作业执行的重复时间间隔。
last_start_date	timestamp without time zone	上次运行开始时间, 时间精确到毫秒。
last_end_date	timestamp without time zone	上次运行的结束时间, 时间精确到毫秒。
last_suc_date	timestamp without time zone	上次成功运行的开始时间, 时间精确到毫秒。
this_run_date	timestamp without time zone	正在运行任务的开始时间, 时间精确到毫秒。

## 15.2.38 PG\_LANGUAGE

PG\_LANGUAGE系统表登记编程语言, 用户可以用这些语言或接口写函数或者存储过程。

表 15-38 PG\_LANGUAGE 字段

名字	类型	引用	描述
oid	oid	-	行标识符（隐藏属性，必须明确选择才会显示）。
lanname	name	-	语言的名称。
lanowner	oid	<a href="#">PG_AUTHID.oid</a>	语言的所有者。
lanispl	boolean	-	内部语言为假（比如SQL），用户定义语言为真。目前， <code>gs_dump</code> 仍然使用该字段判断哪些语言需要转储，但可能在将来会被其它机制所取代。
lanpltrusted	boolean	-	如果是可信语言则为真，即系统相信它不会被授予任何正常SQL执行环境之外的权限。只有初始用户可以在用非可信的语言创建函数。
lanplcallfoid	oid	<a href="#">PG_PROC.oid</a>	对于非内部语言，这是指该语言处理器的引用，语言处理器是一个特殊函数，负责执行以某种语言写的所有函数。
laninline	oid	<a href="#">PG_PROC.oid</a>	此字段引用一个负责执行“inline”匿名代码块的函数（DO块）。如果不支持内联块则为0。
lanvalidator	oid	<a href="#">PG_PROC.oid</a>	此字段引用一个语言校验器函数，它负责检查新创建函数的语法和有效性。如果没有提供校验器，则为0。
lanacl	aclitem[]	-	访问权限。

## 15.2.39 PG\_LARGEOBJECT

PG\_LARGEOBJECT系统表保存那些标记着“大对象”的数据。一个大对象是使用其创建时分配的OID标识的。每个大对象都分解成足够小的小段或者“页面”以便以行的形式存储在PG\_LARGEOBJECT里。每页的数据定义为LOBKSIZE。

需要有系统管理员权限才可以访问此系统表。

表 15-39 PG\_LARGEOBJECT 字段

名字	类型	引用	描述
loid	oid	<a href="#">PG_LARGEOBJECT_METADATA.oid</a>	包含本页的大对象的标识符。
pageno	integer	-	本页在其大对象数据中的页码，从0开始计算。

名字	类型	引用	描述
data	bytea	-	实际存储在大对象中的数据。这些数据不会超过LOBLKSIZE字节，且可能更小。

PG\_LARGEOBJECT的每一行保存一个大对象的一个页面，从该对象内部的字节偏移（pageno \* LOBLKSIZE）开始。这种实现允许稀疏存储：页面可能丢失，并且可以比LOBLKSIZE字节少（即使它们不是对象的最后一页）。大对象中丢失的区域会被读为0。

## 15.2.40 PG\_LARGEOBJECT\_METADATA

PG\_LARGEOBJECT\_METADATA系统表存储与大数据相关的元数据。实际的大对象数据存储在PG\_LARGEOBJECT里。

表 15-40 PG\_LARGEOBJECT\_METADATA 字段

名字	类型	引用	描述
oid	oid	-	行标识符（隐藏属性，必须明确选择才会显示）
lomowner	oid	PG_AUTHID.oid	大对象的所有者
lomacl	aclitem[]	-	访问权限

## 15.2.41 PG\_STAT\_LAST\_OPERATION

PG\_STAT\_LAST\_OPERATION系统表存储表、视图、函数等数据库对象上一次执行DDL时相关的信息。该系统表仅8.2.0.100及以上集群版本支持。

表 15-41 PG\_STAT\_LAST\_OPERATION 字段

名称	类型	引用	描述
oid	oid	-	行标识符（隐藏属性，必须明确选择才会显示）。
classid	oid	-	包含对象的系统目录的OID。
objid	oid	-	对象在其系统目录内的对象OID。
staactionname	name	-	在一个对象上采取的操作名。
stasysid	oid	PG_AUTHID.oid	DDL操作实施者的oid。

名称	类型	引用	描述
statime	timestamp with time zone	-	DDL操作的时间戳。
csn	xid	-	DDL操作执行时的当前的不推进的 csn。
stasubtype	text	-	被执行操作的对象类型或者被执行操作的子类。

## 应用示例

查看表上一次执行DDL时相关的信息：

```
SELECT * FROM PG_STAT_LAST_OPERATION;
classid | objid | staactionname | stasysid | statime | csn | stasubtype
-----+-----+-----+-----+-----+-----+-----
 1259 | 20163 | CREATE | 10 | 2023-01-04 14:37:31.477995+08 | 282450 | TABLE
 1259 | 20163 | VACUUM | 10 | 2023-01-05 18:03:19.579291+08 | 1282717 | VACUUM FULL
(2 rows)
```

## 15.2.42 PG\_NAMESPACE

PG\_NAMESPACE系统表存储命名空间，即存储schema相关的信息。

表 15-42 PG\_NAMESPACE 字段

名称	类型	描述
nspname	name	命名空间的名称。
nspowner	oid	命名空间的所有者。
nsptimeline	bigint	在DN上创建此命名空间时的时间线。此字段为内部使用，仅在DN上有效。
nspacl	aclitem[]	访问权限。具体请参见GRANT和REVOKE。
permspace	bigint	schema永久表空间限额。
usedspace	bigint	schema已用永久表空间大小。

## 15.2.43 PG\_OBJECT

PG\_OBJECT系统表存储限定类型对象(object\_type中存在的类型)的创建用户、创建时间、最后修改时间和最后analyze时间。

表 15-43 PG\_OBJECT 字段

名称	类型	描述
object_oid	oid	对象标识符。
object_type	"char"	对象类型： <ul style="list-style-type: none"> <li>• r表示表，包括普通表和临时表</li> <li>• i表示索引</li> <li>• s表示序列</li> <li>• v表示视图</li> <li>• p表示存储过程和函数</li> <li>• f表示外表</li> </ul>
creator	oid	创建用户的标识符。
ctime	timestamp with time zone	对象创建时间。
mtime	timestamp with time zone	对象最后修改时间，默认记录修改行为包括ALTER操作、COMMENT、GRANT/REVOKE和TRUNCATE操作。 <b>object_mtime_record_mode</b> 参数可以细粒度控制ALTER、COMMENT、GRANT/REVOKE和TRUNCATE操作是否被记录。
last_analyze_time	timestamp with time zone	对象进行最后一次analyze的时间。

#### 须知

- 仅针对用户正常操作行为进行记录，无法记录对象升级以前和initdb过程中的行为。
- ctime和mtime的时间记录为本次操作的事务起始时间。
- 由扩容引起的对象修改时间也会被记录。

## 15.2.44 PG\_OBSSCANINFO

PG\_OBSSCANINFO系统表定义了云上加速场景中，使用加速集群时扫描OBS数据的运行时信息，每条记录对应一个query中单个OBS外表的运行时信息。

表 15-44 PG\_OBSSCANINFO 字段

名字	类型	引用	描述
query_id	bigint	-	查询标识

名字	类型	引用	描述
user_id	text	-	执行该查询的数据库用户
table_name	text	-	OBS外表的表名
file_type	text	-	底层数据保存的文件格式
time_stamp	time_stam	-	扫描操作开始的时间
actual_time	double	-	扫描操作执行时间，单位为秒
file_scanned	bigint	-	扫描的文件数量
data_size	double	-	扫描的数据量，单位为字节
billing_info	text	-	保留字段

## 15.2.45 PG\_OPCLASS

PG\_OPCLASS系统表定义索引访问方法操作符类。

每个操作符类为一种特定数据类型和一种特定索引访问方法定义索引字段的语义。一个操作符类本质上指定一个特定的操作符族适用于一个特定的可索引的字段数据类型。索引的字段实际可用的族中的操作符集是接受字段的数据类型作为它们的左边的输入的那个。

表 15-45 PG\_OPCLASS 字段

名字	类型	引用	描述
oid	oid	-	行标识符（隐藏属性，必须明确选择才会显示）。
opcmethod	oid	<a href="#">PG_AM.oid</a>	操作符类所属的索引访问方法。
opcname	name	-	操作符类的名称。
opcnamespace	oid	<a href="#">PG_NAMESPACE.oid</a>	操作符类所属的命名空间。
opowner	oid	<a href="#">PG_AUTHID.oid</a>	操作符类所有者。
opcfamily	oid	<a href="#">PG_OPFAMILY.oid</a>	包含此操作符类的操作符族。
opcintype	oid	<a href="#">PG_TYPE.oid</a>	操作符类索引的数据类型。
opcdefault	boolean	-	如果操作符类是opcintype的缺省，则为真。
opckeytype	oid	<a href="#">PG_TYPE.oid</a>	索引数据的类型，如果和opcintype相同则为0。

一个操作符类的opcmethod必须匹配包含它的操作符族的opfmethod。同样，对于任意给定的opcmethod和opcintype的组合，不能有超过一个PG\_OPCLASS行有opcdefault为真。

## 15.2.46 PG\_OPERATOR

PG\_OPERATOR系统表存储有关操作符的信息。

表 15-46 PG\_OPERATOR 字段

名字	类型	引用	描述
oid	oid	-	行标识符（隐藏属性，必须明确选择才会显示）
oprname	name	-	操作符的名称
oprnamespace	oid	<a href="#">PG_NAMESPACE.oid</a>	包含此操作符的命名空间的OID
oprowner	oid	<a href="#">PG_AUTHID.oid</a>	操作符所有者
oprkind	"char"	-	<ul style="list-style-type: none"> <li>• b=infix =中缀（两边）</li> <li>• l=前缀（左边）</li> <li>• r=后缀（右边）</li> </ul>
oprcanmerge	boolean	-	该操作符是否支持合并连接
oprcanhash	boolean	-	该操作符是否支持Hash连接
oprleft	oid	<a href="#">PG_TYPE.oid</a>	左操作数的类型
oprright	oid	<a href="#">PG_TYPE.oid</a>	右操作数的类型
oprresult	oid	<a href="#">PG_TYPE.oid</a>	结果类型
oprcom	oid	<a href="#">PG_OPERATOR.oid</a>	此操作符的交换符（如果存在）
oprnegate	oid	<a href="#">PG_OPERATOR.oid</a>	此操作符的反转器（如果存在）
oprcode	regproc	<a href="#">PG_PROC.oid</a>	实现该操作符的函数
oprrest	regproc	<a href="#">PG_PROC.oid</a>	此操作符的约束选择性计算函数
oprjoin	regproc	<a href="#">PG_PROC.oid</a>	此操作符的连接选择性计算函数

## 15.2.47 PG\_OPFAMILY

PG\_OPFAMILY系统表定义操作符族。

每个操作符族是操作符和相关支持例程的集合，这些例程实现了为特定索引访问方法指定的语义。此外，按照访问方法指定的某种方式，一个族内的操作符都是“兼容

的”。操作符族允许跨数据类型操作符与索引一起使用，并可以推理使用访问方法语义相关内容。

表 15-47 PG\_OPFAMILY 字段

名字	类型	引用	描述
oid	oid	-	行标识符（隐藏属性，必须明确选择才会显示）
opfmethod	oid	<a href="#">PG_AM</a> .oid	操作符族使用的索引方法
opfname	name	-	操作符族的名称
opfnamespace	oid	<a href="#">PG_NAMESPACE</a> .oid	操作符族的命名空间
opfowner	oid	<a href="#">PG_AUTHID</a> .oid	操作符族的所有者

定义一个操作符族的大多数信息不在PG\_OPFAMILY，而是在相关的[PG\\_AMOP](#)，[PG\\_AMPROC](#)和[PG\\_OPCLASS](#)中。

## 15.2.48 PG\_PARTITION

PG\_PARTITION系统表存储数据库内所有分区表(partitioned table)、分区(table partition)、分区上toast表和分区索引(index partition)四类对象的信息。分区表索引(partitioned index)的信息不在PG\_PARTITION系统表中保存。

表 15-48 PG\_PARTITION 字段

名称	类型	描述
relname	name	分区表、分区、分区上toast表和分区索引的名称。
parttype	"char"	对象类型： <ul style="list-style-type: none"> <li>• 'r': partitioned table</li> <li>• 'p': table partition</li> <li>• 'x': index partition</li> <li>• 't': toast table</li> </ul>
parentid	oid	当对象为分区表或分区时，此字段表示分区表在PG_CLASS中的OID。 当对象为index partition时，此字段表示所属分区表索引(partitioned index)的OID。
rangenum	integer	保留字段。
intervalnum	integer	保留字段。



名称	类型	描述
partstrategy	"char"	分区表分区策略，现在仅支持： 'r'：范围分区。 'v'：数值分区。 'l'：列表分区。
relfilenode	oid	table partition、index partition、分区上toast表的物理存储位置。
reltablespace	oid	table partition、index partition、分区上toast表所属表空间的OID。
relpages	double precision	统计信息：table partition、index partition的数据页数。
reltuples	double precision	统计信息：table partition、index partition的元组数。
relallvisible	integer	统计信息：table partition、index partition的可见数据页数。
reltoastrelid	oid	table partition所对应toast表的OID。
reltoastidxid	oid	table partition所对应toast表的索引的OID。
indextblid	oid	index partition对应table partition的OID。
indisusable	boolean	分区索引是否可用。
reldeltarelid	oid	Delta表的OID。
reldeltaidx	oid	Delta表的索引表的OID。
relcudescrelid	oid	CU描述表的OID。
relcudescidx	oid	CU描述表的索引表的OID。
relfrozenxid	xid32	冻结事务ID号。 为保持前向兼容，保留此字段，新增relfrozenxid64用于记录此信息。
intspnum	integer	间隔分区所属表空间的个数。
partkey	int2vector	分区键的列号。
intervaltablespace	oidvector	间隔分区所属的表空间，间隔分区以round-robin方式落在这些表空间内。
interval	text[]	间隔分区的间隔值。
boundaries	text[]	范围分区和间隔分区的上边界。
transit	text[]	间隔分区的跳转点。

名称	类型	描述
reloptions	text[]	设置partition的存储属性，与pg_class.reloptions的形态一样，用"keyword=value"格式的字符串来表示，目前用于在线扩容的信息搜集。
relfrozenxid64	xid	冻结事务ID号。
boundexprs	pg_node_tree	分区边界表达式。 <ul style="list-style-type: none"> <li>对于范围分区来说是分区上边界表达式。</li> <li>对于列表分区来说是分区边界枚举值集合。</li> </ul> pg_node_tree数据类型是不可读的，可用如下表达式pg_get_expr把当前字段单翻译为可读信息。 <pre>SELECT pg_get_expr(boundexprs, 0) FROM pg_partition WHERE relname = 'country_202201'; pg_get_expr</pre> <hr/> <pre>ROW(202201, 'city1'::text), ROW(202201, 'city2'::text) (1 row)</pre>

## 15.2.49 PG\_PLTEMPLATE

PG\_PLTEMPLATE系统表存储过程语言的“模板”信息。

表 15-49 PG\_PLTEMPLATE 字段

名称	类型	描述
tmplname	name	该模板所应用的语言的名称。
tmpltrusted	boolean	如果语言被认为是可信的，则为真。
tmpldbacreate	boolean	如果语言是由数据库所有者创建的，则为真。
tmplhandler	text	调用处理器函数的名称。
tmplinline	text	匿名块处理器的名称，如果没有则为NULL。
tmplvalidator	text	校验函数的名称，如果没有则为NULL。
tmpllibrary	text	实现语言的共享库的路径。
tmplacl	aclitem[]	模板的访问权限（未使用）。

## 15.2.50 PG\_PROC

PG\_PROC系统表存储函数或过程的信息。

表 15-50 PG\_PROC 字段

名称	类型	描述
proname	name	函数名。
pronamespace	oid	此函数所在命名空间的OID。
proowner	oid	函数的所有者。
prolang	oid	实现语言或函数的调用接口。
procost	real	估计执行成本。
prorows	real	结果行估计数。
provariadic	oid	参数元素的数据类型。
protransform	regproc	此函数的简化调用方式。
proisagg	boolean	函数是否为聚集函数。
proiswindow	boolean	函数是否为窗口函数。
prosecdef	boolean	函数是否为一个安全定义器（例如，一个“setuid”函数）。
proleakproof	boolean	函数有无其他影响。如果函数没有对参数进行防泄露处理，则会抛出错误。
proisstrict	boolean	如果任意调用参数为空，函数是否返回空值。这种情况下函数实际上根本不会被调用。非“strict”的函数必须准备处理空值输入。
proretset	boolean	函数是否返回一个集合（即，指定数据类型的多个数值）。
provolatile	"char"	说明该函数的结果是只依赖于它的输入参数，或者还会被外接因素影响。 <ul style="list-style-type: none"> <li>• i表示“不可变的”（immutable）函数，对于相同的输入总是输出相同的结果。</li> <li>• s表示稳定的”（stable）函数，对于固定输入其结果在一次扫描里不变。</li> <li>• v表示“易变”（volatile）函数，其结果可能在任何时候都变化。</li> </ul>
pronargs	smallint	参数个数。
pronargdefaults	smallint	有默认值的参数个数。
prorettype	oid	返回参数类型的OID。
proargtypes	oidvector	函数参数的数据类型的数组。数组里只包括输入参数（包括INOUT参数），因此也表现了函数的调用特征。

名称	类型	描述
proallargtypes	oid[]	函数参数的数据类型的数组。数组里包括所有参数的类型（包括OUT和INOUT参数），如果所有参数都是IN参数，则这个字段就会为空。注意数组下标是以1为起点的，而因为历史原因，proargtypes的下标起点为0。
proargmodes	"char"[]	函数参数模式的数组。 <ul style="list-style-type: none"> <li>• i表示IN参数</li> <li>• o表示OUT参数</li> <li>• b表示INOUT参数</li> </ul> 如果所有参数都是IN参数，则这个字段为空。注意此数组下标对应的是proallargtypes的位置，而不是proargtypes。
proargnames	text[]	函数参数的名字的数组。没有名字的参数在数组里设置为空字符串。如果没有一个参数有名字，这个字段为空。注意此数组的下标对应proallargtypes而不是proargtypes。
proargdefaults	pg_node_tree	默认值的表达式树。是PRONARGDEFAULTS元素的列表。
prosrc	text	描述函数或存储过程的定义。例如，对于解释型语言来说就是函数的源程序，或者一个链接符号，一个文件名，或者函数和存储过程创建时指定的其他任何函数体内容，具体取决于语言/调用习惯的实现。
probin	text	关于如何调用该函数的附加信息。同样，其含义也是和语言相关的。
proconfig	text[]	函数针对运行时配置变量的本地设置。
proacl	aclitem[]	访问权限。具体请参见GRANT和REVOKE。
prodefaultargpos	int2vector	函数默认值的位置，不局限于能最后几个参数才有默认值。
fencedmode	boolean	函数的执行模式，表示函数是在fence还是not fence模式下执行。如果是fence执行模式，函数的执行会在重新fork的进程中执行。默认值是fence。
proshippable	boolean	函数是否可以下推到DN上执行，默认值是false。 <ul style="list-style-type: none"> <li>• 对于IMMUTABLE类型的函数，函数始终可以下推到DN上执行。</li> <li>• 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。</li> </ul>

名称	类型	描述
propackage	boolean	该函数是否支持重载，主要针对Oracle风格的函数，默认值是false。

## 应用示例

查询指定函数的OID。例如，获取函数justify\_days的OID为1295。

```
SELECT oid FROM pg_proc where proname = 'justify_days';
oid
-----
1295
(1 row)
```

查询指定函数是否为聚集函数。例如，查询justify\_days函数为非聚集函数。

```
SELECT proisagg FROM pg_proc where proname = 'justify_days';
proisagg
-----
f
(1 row)
```

### 15.2.51 PG\_PUBLICATION

PG\_PUBLICATION系统表存储当前数据库中创建的所有发布。该系统表仅8.2.0.100及以上集群版本支持。

表 15-51 PG\_PUBLICATION 字段

名称	类型	引用	描述
oid	oid	-	行标识符（隐藏属性，必须明确选择才会显示）。
pubname	name	-	发布的名称。
pubowner	oid	<a href="#">PG_AUTHID.oid</a>	发布的拥有者。
puballtables	boolean	-	如果为true，这个发布自动包括数据库中的所有表，包括未来将会创建的任何表。
pubinsert	boolean	-	如果为true，为发布中的表复制INSERT操作。
pubupdate	boolean	-	如果为true，为发布中的表复制UPDATE操作。
pubdelete	boolean	-	如果为true，为发布中的表复制DELETE操作。
pubtruncate	boolean	-	如果为true，为发布中的表复制TRUNCATE操作。

## 应用示例

查看所有发布：

```
SELECT * FROM pg_publication;
pubname | pubowner | puballtables | pubinsert | pubupdate | pubdelete | pubtruncate
-----+-----+-----+-----+-----+-----+-----
myspub  | 10      | t           | t         | t         | t         | t
(1 row)
```

### 15.2.52 PG\_PUBLICATION\_NAMESPACE

PG\_PUBLICATION\_NAMESPACE系统表存储当前数据库中的发布和模式之间的映射，这是一种多对多映射。该系统表仅8.2.0.100及以上集群版本支持。

表 15-52 PG\_PUBLICATION\_NAMESPACE 字段

名字	类型	引用	描述
oid	oid	-	行标识符（隐藏属性，必须明确选择才会显示）。
prpubid	oid	<a href="#">PG_PUBLICATION.oid</a>	映射的发布的OID。
pnnspid	oid	<a href="#">PG_NAMESPACE.oid</a>	映射的模式的OID。

## 应用示例

查看所有发布和模式的映射：

```
SELECT * FROM pg_publication_namespace;
pnpubid | pnnspid
-----+-----
16797   | 16796
(1 row)
```

### 15.2.53 PG\_PUBLICATION\_REL

PG\_PUBLICATION\_REL系统表存储当前数据库中的发布和表之间的映射，这是一种多对多映射。该系统表仅8.2.0.100及以上集群版本支持。

#### 📖 说明

查询时推荐使用视图[PG\\_PUBLICATION\\_TABLES](#)，可以展现更详细的信息。

表 15-53 PG\_PUBLICATION\_REL 字段

名字	类型	引用	描述
oid	oid	-	行标识符（隐藏属性，必须明确选择才会显示）。
prpubid	oid	<a href="#">PG_PUBLICATION.oid</a>	映射的发布的OID。
prrelid	oid	<a href="#">PG_CLASS.oid</a>	映射的表的OID。

## 应用示例

查看所有发布和表的映射：

```
postgres=# SELECT * FROM pg_publication_rel;
 prpubid | prrelid
-----+-----
 16797 | 16757
 16797 | 16776
(2 rows)
```

### 15.2.54 PG\_RANGE

PG\_RANGE系统表存储关于范围类型的信息。

除了PG\_TYPE里类型的记录。

表 15-54 PG\_RANGE 字段

名字	类型	引用	描述
rngtypid	oid	PG_TYPE.oid	范围类型的OID。
rngsubtype	oid	PG_TYPE.oid	该范围类型的元素类型（子类型）的OID。
rngcollation	oid	PG_COLLATION.oid	用于范围比较的排序规则的OID，如果没有则为0。
rngsubopc	oid	PG_OPCLASS.oid	用于范围比较的子类型的操作符类的OID。
rngcanonical	regproc	PG_PROC.oid	转换范围类型为规范格式的函数的OID，如果没有则为0。
rngsubdiff	regproc	PG_PROC.oid	返回两个double precision元素值的不同的函数的OID，如果没有则为0。

rngsubopc（如果元素类型是可排序的，则加上rngcollation）决定用于范围类型的排序顺序。当元素类型是时使用rngcanonical用于离散类型的元素类型。

### 15.2.55 PG\_REDACTION\_COLUMN

PG\_REDACTION\_COLUMN系统表存储脱敏列的信息。

表 15-55 PG\_REDACTION\_COLUMN 字段

名称	类型	描述
object_oid	oid	脱敏对象OID。
column_attrno	smallint	脱敏列attrno。

名称	类型	描述
function_type	integer	脱敏类型。 <b>说明</b> 保留字段，仅为向前兼容低版本的脱敏列信息，可取值为0（NONE）、1（FULL）。
function_parameters	text	脱敏类型为partial类型时的参数。（保留字段，无实际意义）
regexp_pattern	text	脱敏类型为regexp时，格式化字符串。（保留字段，无实际意义）
regexp_replace_string	text	脱敏类型为regexp时，替换串。（保留字段，无实际意义）
regexp_position	integer	脱敏类型为regexp时，起始替换位置。（保留字段，无实际意义）
regexp_occurrence	integer	脱敏类型为regexp时，替换次数。（保留字段，无实际意义）
regexp_match_parameter	text	脱敏类型为regexp时，正则控制参数。（保留字段，无实际意义）
column_description	text	脱敏列描述信息。
function_expr	pg_node_tree	脱敏函数的内部表现形式。
inherited	bool	说明脱敏列是否是“继承”自其他脱敏列。

## 15.2.56 PG\_REDACTION\_POLICY

PG\_REDACTION\_POLICY系统表提供了脱敏对象的信息。

表 15-56 PG\_REDACTION\_POLICY 字段

名称	类型	描述
object_oid	oid	脱敏对象OID。
policy_name	name	脱敏策略名称。
enable	boolean	策略状态（开启、关闭）。 <b>说明</b> enable的取值： <ul style="list-style-type: none"> <li>• true表示开启</li> <li>• false表示关闭</li> </ul>
expression	pg_node_tree	策略生效表达式（针对用户）。



名称	类型	描述
policy_description	text	策略描述信息。
inherited	bool	说明脱敏策略是否“继承”自其他脱敏策略。

## 15.2.57 PG\_RELFILENODE\_SIZE

PG\_RELFILENODE\_SIZE系统表提供了文件级空间统计能力，表中的每一条记录则对应着磁盘上相应的物理文件和该文件的文件大小。

表 15-57 PG\_RELFILENODE\_SIZE 字段

名称	类型	描述
databaseid	oid	该物理文件所属database对应的OID。如果是跨库共享系统表，该值为0。
tablespaceid	oid	该物理文件所属表空间对应的OID。
relfilenode	oid	该物理文件的物理文件编号。
backendid	integer	创建该物理文件的后台线程号，通常为-1。
type	integer	该物理文件的文件类型。 <ul style="list-style-type: none"> <li>0为数据类型。</li> <li>1为FSM文件类型。</li> <li>2为VM文件类型。</li> <li>3为BCM文件类型。</li> <li>大于4为列存表对应列的数据文件和BCM文件大小之和。</li> </ul>
filesize	bigint	该物理文件的文件大小，单位Byte。

## 15.2.58 PG\_RLSPOLICY

PG\_RLSPOLICY系统表提供了行级访问控制策略的信息。

表 15-58 PG\_RLSPOLICY 字段

名称	类型	描述
polname	name	行访问控制策略名称。

名称	类型	描述
polrelid	oid	行访问控制策略的表OID。
polcmd	char	行访问控制策略影响的SQL操作，包括：*(ALL)、r(SELECT)、w(UPDATE)、d(DELETE)。
polpermissive	boolean	行访问控制策略的类型。 <b>说明</b> polpermissive的取值： <ul style="list-style-type: none"> <li>• true表示PERMISSIVE，表示行访问控制策略是宽容性策略。</li> <li>• false表示RESTRICTIVE，表示行访问控制策略是限制性策略。</li> </ul>
polroles	oid[]	行访问控制策略影响的数据库用户OID。
polqual	pg_node_tree	行访问控制策略的SQL条件表达式。

## 15.2.59 PG\_RESOURCE\_POOL

PG\_RESOURCE\_POOL系统表提供了数据库资源池的信息。

表 15-59 PG\_RESOURCE\_POOL 字段

名称	类型	描述
respool_name	name	资源池名称。
mem_percent	integer	内存配置的百分比。
cpu_affinity	bigint	CPU绑定core的数值。
control_group	name	资源池所在的control group名字。
active_statements	integer	资源池上最大的并发数。
max_dop	integer	最大并发度，保留字段。
memory_limit	name	资源池最大的内存。
parentid	oid	父资源池OID。
io_limits	integer	保留字段，无实际意义。
io_priority	text	保留字段，无实际意义。
is_foreign	boolean	表示资源池是否用于逻辑集群之外的用户。如果为true，表示资源池用来控制不属于当前资源池的普通用户的资源。

## 15.2.60 PG\_REWRITE

PG\_REWRITE系统表存储为表和视图定义的重写规则。

表 15-60 PG\_REWRITE 字段

名称	类型	描述
rulename	name	规则名称
ev_class	oid	使用该规则的表名
ev_attr	smallint	该规则适用的字段（目前总是为0，表示整个表）
ev_type	"char"	规则适用的事件类型： <ul style="list-style-type: none"> <li>• 1 = SELECT</li> <li>• 2 = UPDATE</li> <li>• 3 = INSERT</li> <li>• 4 = DELETE</li> </ul>
ev_enabled	"char"	用于控制复制的触发 <ul style="list-style-type: none"> <li>• O = “origin” 和 “local” 模式时触发</li> <li>• D = 禁用触发</li> <li>• R = “replica” 时触发</li> <li>• A = 任何模式都会触发</li> </ul>
is_instead	boolean	如果是INSTEAD规则，则为真
ev_qual	pg_node_tree	规则条件的表达式树（以nodeToString() 形式存在）
ev_action	pg_node_tree	规则动作的查询树（以nodeToString() 形式存在）

## 15.2.61 PG\_SECLABEL

PG\_SECLABEL系统表存储数据对象上的安全标签。

**PG\_SHSECLABEL**的作用类似，只是用于在一个数据库集群内共享的数据库对象的安全标签上。

表 15-61 PG\_SECLABEL 字段

名字	类型	引用	描述
objoid	oid	任意OID属性	此安全标签所属的对象的OID
classoid	oid	<b>PG_CLASS</b> .oid	此对象的系统目录的OID
objsubid	integer	-	出现在此对象中的列的序号

名字	类型	引用	描述
provider	text	-	与此标签相关的标签提供者
label	text	-	应用于此对象的安全标签

## 15.2.62 PG\_SHDEPEND

PG\_SHDEPEND系统表记录数据库对象和共享对象（比如角色）之间的依赖关系。这些信息使得GaussDB(DWS)可以确保对象在被删除时没有被其他对象引用。

**PG\_DEPEND**的作用类似，只是它是针对单个数据库中对象之间的依赖。

和大多数其他系统表不同，PG\_SHDEPEND在集群的所有数据库之间共享：每个数据库集群只有一个PG\_SHDEPEND，并非每个数据库一个。

表 15-62 PG\_SHDEPEND 字段

名字	类型	引用	描述
dbid	oid	<b>PG_DATABASE.oid</b>	依赖对象所在的数据库的OID，如果是共享对象，则为0。
classid	oid	<b>PG_CLASS.oid</b>	依赖对象所在的系统表的OID。
objid	oid	任意OID属性	指定的依赖对象的OID。
objsubid	integer	-	对于一个表字段，为字段号（objid和classid参考表本身）。对于所有其他对象类型，该字段为0。
refclassid	oid	<b>PG_CLASS.oid</b>	被引用对象所在的系统表的OID(必须是一个共享表)。
refobjid	oid	任意OID属性	指定的被引用对象的OID。
deptype	"char"	-	定义该依赖关系的特定语义的代码见表后说明。
objfile	text	-	用户定义C函数库文件路径。

在任何情况下，一条PG\_SHDEPEND记录就表明被引用的对象不能在未删除依赖对象的前提下被删除。但是其中也有几种依赖类型由deptype定义的情况：

- SHARED\_DEPENDENCY\_OWNER (o)  
被引用的对象（必须是一个角色）是依赖对象的所有者。
- SHARED\_DEPENDENCY\_ACL (a)  
在依赖对象的ACL（访问控制列表，也就是权限列表）中提到被引用的对象（必须是一个角色）。不会为对象的所有者创建SHARED\_DEPENDENCY\_ACL，因为所有者将具有SHARED\_DEPENDENCY\_OWNER记录。
- SHARED\_DEPENDENCY\_PIN (p)

没有依赖对象。这类记录标识系统自身依赖于被依赖对象，因此这种对象绝对不能被删除。此类型的记录只能被initdb创建，依赖对象的字段都为0。

### 15.2.63 PG\_SHDESCRIPTION

PG\_SHDESCRIPTION系统表存储共享数据库对象的可选注释。可以使用COMMENT命令操作注释的内容，使用psql的\d命令查看注释内容。

PG\_DESCRIPTION提供了类似的功能，它记录了单个数据库中对象的注释。

不同于大多数系统表，PG\_SHDESCRIPTION在集群中所有数据库之间共享：每个数据库集群只有一个PG\_SHDESCRIPTION，而不是每个数据库一个。

表 15-63 PG\_SHDESCRIPTION 字段

名字	类型	引用	描述
objoid	oid	任意OID属性	此描述所属的对象的OID
classoid	oid	<a href="#">PG_CLASS.oid</a>	此对象所在的系统目录的OID
description	text	-	作为对该对象描述的任意文本

### 15.2.64 PG\_SHSECLABEL

PG\_SHSECLABEL系统表存储在共享数据库对象上的安全标签。安全标签可以用SECURITY LABEL命令操作。

查看安全标签的简单点的方法，请参阅[PG\\_SECLABELS](#)。

[PG\\_SECLABEL](#)的作用类似，只是它是用于在单个数据库内部的对象的安全标签的。

不同于大多数的系统表，PG\_SHSECLABEL在一个集群中的所有数据库中共享：每个数据库集群只有一个PG\_SHSECLABEL，而不是每个数据库一个。

表 15-64 PG\_SHSECLABEL 字段

名字	类型	引用	描述
objoid	oid	任意OID属性	此安全标签所属对象的OID
classoid	oid	<a href="#">PG_CLASS.oid</a>	对象所属系统目录的OID
provider	text	-	与此标签关联的标签提供者
label	text	-	应用于该对象的安全标签

### 15.2.65 PG\_STATISTIC

PG\_STATISTIC系统表存储有关该数据库中表和索引列的统计数据。需要有系统管理员权限才可以访问此系统表。

表 15-65 PG\_STATISTIC 字段

名称	类型	描述
starelid	oid	所描述的字段所属的表或者索引。
starekind	"char"	所属对象的类型。
staattnum	smallint	所描述的字段在表中的编号，从1开始。
stainherit	boolean	是否统计有继承关系的对象。
stanullfrac	real	该字段中为NULL的记录的比例。
stawidth	integer	非NULL记录的平均存储宽度，以字节计。
stadistinct	real	标识全局统计信息中所有DN上字段里唯一的非NULL数据值的数目。 <ul style="list-style-type: none"> <li>• 大于0的值是独立数值的实际数目。</li> <li>• 小于0的值是表中行数的分数的负数（比如，一个字段的数值平均出现概率为两次，则可以表示为 stadistinct=-0.5）。</li> <li>• 0表示独立数值的数目未知。</li> </ul>
stakindN	smallint	一个编码，表示这种类型的统计存储在pg_statistic行的第n个“槽位”。 n的取值范围：1~5
staopN	oid	用于生成这些存储在第n个“槽位”的统计信息的操作符。比如，一个柱面图槽位会显示<操作符，该操作符定义了该数据的排序顺序。 n的取值范围：1~5
stanumbersN	real[]	第n个“槽位”的相关类型的数值类型统计，如果该槽位和数值类型没有关系，则就是NULL。 n的取值范围：1~5
stavaluesN	anyarray	第n个“槽位”类型的字段数据值，如果该槽位类型不存储任何数据值，则就是NULL。每个数组的元素值实际上都是指定字段的数据类型，因此，除了把这些字段的类型定义成anyarray之外，没有更好的办法。 n的取值范围：1~5
stadndistinct	real	标识dn1上字段里唯一的非NULL数据值的数目。 <ul style="list-style-type: none"> <li>• 大于0的值是独立数值的实际数目。</li> <li>• 小于0的值是表中行数的分数的负数（比如，一个字段的数值平均出现概率为两次，则可以表示为 stadistinct=-0.5）。</li> <li>• 0表示独立数值的数目未知。</li> </ul>
staextinfo	text	统计信息的扩展信息。预留字段。

## 15.2.66 PG\_STATISTIC\_EXT

PG\_STATISTIC\_EXT系统表存储有关该数据库中表的扩展统计数据，包括多列统计数据  
和表达式统计数据（后续支持）。收集哪些扩展统计数据是由用户指定的。需要有系  
统管理员权限才可以访问此系统表。

表 15-66 PG\_STATISTIC\_EXT 字段

名称	类型	描述
starelid	oid	所描述的字段所属的表或者索引。
starelkind	"char"	所属对象的类型。
stainherit	boolean	是否统计有继承关系的对象。
stanullfrac	real	该字段中为NULL的记录的比例。
stawidth	integer	非NULL记录的平均存储宽度，以字节计。
stadistinct	real	标识全局统计信息中所有DN上字段里唯一的非NULL数据值的数目。 <ul style="list-style-type: none"> <li>• 一个大于零的数值是独立数值的实际数目。</li> <li>• 一个小于零的数值是表中行数的分数的负数（比如，一个字段的数值平均出现概率为两次，则可以表示为stadistinct=-0.5）。</li> <li>• 0表示独立数值的数目未知。</li> </ul>
stadndistinct	real	标识dn1上字段里唯一的非NULL数据值的数目。 <ul style="list-style-type: none"> <li>• 一个大于零的数值是独立数值的实际数目。</li> <li>• 一个小于零的数值是表中行数的分数的负数（比如，一个字段的数值平均出现概率为两次，则可以表示为stadistinct=-0.5）。</li> <li>• 0表示独立数值的数目未知。</li> </ul>
stakindN	smallint	一个编码，表示这种类型的统计存储在pg_statistic行的第n个“槽位”。 n的取值范围：1~5
staopN	oid	一个用于生成这些存储在第n个“槽位”的统计信息的操作符。比如，一个柱面图槽位会显示<操作符，该操作符定义了该数据的排序顺序。 n的取值范围：1~5
stakey	int2vector	所描述的字段编号的数组。
stanumbers N	real[]	第n个“槽位”的相关类型的数值类型统计，如果该槽位和数值类型没有关系，则就是NULL。 n的取值范围：1~5

名称	类型	描述
stavaluesN	anyarray	第n个“槽位”类型的字段数据值，如果该槽位类型不存储任何数据值，则为NULL。每个数组的元素值实际上都是指定字段的数据类型，因此，除了把这些字段的类型定义成anyarray之外，没有更好的办法。 n的取值范围：1~5
staexprs	pg_node_tree	扩展统计信息对应的表达式。

## 15.2.67 PG\_SUBSCRIPTION

PG\_SUBSCRIPTION系统表存储所有现有的订阅。

表 15-67 PG\_SUBSCRIPTION 字段

名称	类型	引用	描述
oid	oid	-	行标识符（隐藏属性，必须明确选择才会显示）。
subdbid	oid	<a href="#">PG_DATABASE.oid</a>	订阅所在的数据库的OID。
subname	name	-	订阅的名称。
subowner	oid	<a href="#">PG_AUTHID.oid</a>	订阅的拥有者。
subenabled	boolean	-	如果为真，订阅被启用并且应该被复制。
subconninfo	text	-	到发布端数据库的连接信息。
subslotname	text	-	发布端数据库中复制槽的名称。空表示为NONE。
subpublications	text[]	-	被订阅的发布名称的数组。这些引用的是发布者服务器上的发布。

### 应用示例

查看所有订阅：

```
SELECT * FROM pg_subscription;
 subdbid | subname | subowner | subenabled |
 subconninfo | subslotname | subpublications
-----+-----+-----+-----+-----+-----+-----
 15992 | mysub | 10 | t | host=1.1.1.1,2.2.2.2 port=10000,20000 dbname=postgres user=repusr1
 password=password_123 | mysub | {mypub}
(1 row)
```



## 15.2.68 PG\_SYNONYM

PG\_SYNONYM系统表存储同义词对象名与其他数据库对象名间的映射信息。

表 15-68 PG\_SYNONYM 字段

名称	类型	描述
synname	name	同义词名称。
synnamespace	oid	该同义词所在的命名空间的OID。
synowner	oid	同义词的所有者，通常是其创建者OID。
synobjschema	name	关联对象指定的模式名。
synobjname	name	关联对象名。

## 15.2.69 PG\_TABLESPACE

PG\_TABLESPACE系统表存储表空间信息。

表 15-69 PG\_TABLESPACE 字段

名称	类型	描述
spcname	name	表空间名。
spcowner	oid	表空间的所有者，通常是其创建者。
spcacl	aclitem[]	访问权限。具体请参见GRANT和REVOKE。
spcoptions	text[]	表空间的选项。
spcmaxsize	text	可使用的最大磁盘空间大小，单位Byte。

## 15.2.70 PG\_TRIGGER

PG\_TRIGGER系统表存储触发器信息。

名称	类型	描述
tgrelid	oid	触发器所在表的OID。
tgname	name	触发器名。
tgfoid	oid	触发器OID。
tgtype	smallint	触发器类型。

名称	类型	描述
tgenabled	"char"	O表示触发器在“origin”和“local”模式下触发。 D表示触发器被禁用。 R表示触发器在“replica”模式下触发。 A表示触发器始终触发。
tgisinternal	boolean	内部触发器标识，如果为true表示内部触发器。
tgconstrelid	oid	完整性约束引用的表。
tgconstrindid	oid	完整性约束的索引。
tgconstraint	oid	约束触发器在pg_constraint中的OID。
tgdeferrable	boolean	约束触发器是为DEFERRABLE类型。
tginitdeferred	boolean	约束触发器是否为INITIALLY DEFERRED类型。
tgnargs	smallint	触发器函数入参个数。
tgattr	int2vector	当触发器指定列时的列号，未指定则为空数组。
tgargs	bytea	传递给触发器的参数。
tgqual	pg_node_tree	表示触发器的WHEN条件，如果没有则为null。

## 15.2.71 PG\_TS\_CONFIG

PG\_TS\_CONFIG系统表包含表示文本搜索配置的选项。一个配置指定一个特定的文本搜索解析器和一个用于解析器输出类型的字典列表。

解析器在PG\_TS\_CONFIG记录中显示，但是字典映射的标记是由PG\_TS\_CONFIG\_MAP中的辅助记录定义的。

表 15-70 PG\_TS\_CONFIG 字段

名字	类型	引用	描述
oid	oid	-	行标识符（隐藏属性，必须明确选择才会显示）
cfgname	name	-	文本搜索配置名
cfgnamespace	oid	PG_NAMESPACE.oid	此配置所在的命名空间的OID
cfgowner	oid	PG_AUTHID.oid	配置的所有者
cfgparser	oid	PG_TS_PARSER.oid	此配置的文本搜索解析器的OID
cfgoptions	text[]	-	分词相关配置选项

## 15.2.72 PG\_TS\_CONFIG\_MAP

PG\_TS\_CONFIG\_MAP系统表包含为每个文本搜索配置的解析器的每种输出符号类型，显示有哪些文本搜索字典可供查询以及以哪种顺序搜索。

表 15-71 PG\_TS\_CONFIG\_MAP 字段

名字	类型	引用	描述
mapcfg	oid	<a href="#">PG_TS_CONFIG</a> .oid	拥有此映射记录的 <b>PG_TS_CONFIG</b> 的OID
maptokentype	integer	-	由配置的解析器发出的一个符号类型
mapseqno	integer	-	查询该项的顺序
mapdict	oid	<a href="#">PG_TS_DICT</a> .oid	查询的文本搜索字典的OID

## 15.2.73 PG\_TS\_DICT

PG\_TS\_DICT系统表包含定义文本搜索字典的项。字典取决于文本搜索模板，该模板显示所有需要实现的功能。字典本身提供了用户可设置参数的模板。

即允许字典通过非权限用户创建。参数由文本字符串dictinitoption指定，参数的格式和意义取决于模板。

表 15-72 PG\_TS\_DICT 字段

名字	类型	引用	描述
oid	oid	-	行标识符（隐藏属性，必须明确选择才会显示）
dictname	name	-	文本搜索字典名
dictnamespace	oid	<a href="#">PG_NAMESPACE</a> .oid	此字典所在的命名空间的OID
dictowner	oid	<a href="#">PG_AUTHID</a> .oid	字典的所有者
dicttemplate	oid	<a href="#">PG_TS_TEMPLATE</a> .oid	此字典的文本搜索模板的OID
dictinitoption	text	-	模板的初始化选项字符串

## 15.2.74 PG\_TS\_PARSER

PG\_TS\_PARSER系统表包含定义文本解析器的项。解析器负责分割输入文本为词位，并且为每个词位分配标记类型。因为解析器必须通过C语言级别的函数实现，所以新解析器必须由数据库系统管理员创建。

表 15-73 PG\_TS\_PARSER 字段

名字	类型	引用	描述
oid	oid	-	行标识符（隐藏属性，必须明确选择才会显示）
prsname	name	-	文本搜索解析器名称
prsnamespace	oid	<a href="#">PG_NAMESPACE.oid</a>	解析器所在的命名空间的OID
prsstart	regproc	<a href="#">PG_PROC.oid</a>	解析器的启动函数的OID
prstoken	regproc	<a href="#">PG_PROC.oid</a>	解析器的下一个标记函数的OID
prsend	regproc	<a href="#">PG_PROC.oid</a>	解析器的关闭函数的OID
prsheadline	regproc	<a href="#">PG_PROC.oid</a>	解析器的标题函数的OID
prsllexype	regproc	<a href="#">PG_PROC.oid</a>	解析器的lexype函数的OID

## 15.2.75 PG\_TS\_TEMPLATE

PG\_TS\_TEMPLATE系统表包含定义文本搜索模板的项。模板是文本搜索字典的类的实现框架。因为模板必须通过C语言级别的函数实现，索引新模板的创建必须由数据库系统管理员创建。

表 15-74 PG\_TS\_TEMPLATE 字段

名字	类型	引用	描述
oid	oid	-	行标识符（隐藏属性，必须明确选择才会显示）
tmplname	name	-	文本搜索模板名
tmplnamespace	oid	<a href="#">PG_NAMESPACE.oid</a>	模板所属的命名空间的OID
tmplinit	regproc	<a href="#">PG_PROC.oid</a>	模板的初始化函数的OID
tmpllexize	regproc	<a href="#">PG_PROC.oid</a>	模板的lexize函数的OID

## 15.2.76 PG\_TYPE

PG\_TYPE系统表存储数据类型的相关信息。

表 15-75 PG\_TYPE 字段

名称	类型	描述
typename	name	数据类型名称。
typnamespace	oid	此类型所在的命名空间的OID。
typowner	oid	此类型的所有者。
typlen	smallint	对于定长类型是该类型内部表现形式的字节数。对于变长类型typlen为负值。 <ul style="list-style-type: none"> <li>-1表示一种“变长”（有长度字属性的数据）。</li> <li>-2表示一个以NULL结尾的C字符串。</li> </ul>
typbyval	boolean	指定内部传递这个类型的数值时是传值还是传引用。如果该类型的TYPLEN不是1, 2, 4, 8, TYPBYVAL最好为假。变长类型通常是传引用。即使TYPLEN允许传值, TYPBYVAL也可以为假。
typtype	"char"	<ul style="list-style-type: none"> <li>b表示基础类型。</li> <li>c表示复合类型（比如，一个表的行类型）。</li> <li>e表示枚举类型。</li> <li>p表示伪类型。</li> </ul> 参见typrelid和typbasetype。
typcategory	"char"	数据类型的模糊分类，可用于解析器使用的数据转换依据。
typispreferred	boolean	如果为真，则数据符合TYPCATEGORY所指定的转换规则时进行转换。
typisdefined	boolean	如果定义了类型，则为真。如果是一种尚未定义的类型占位符，则为假。如果为假，则除了该类型名称，命名空间和OID之外没有可依赖的对象。
typdelim	"char"	分析数组输入时，分隔两个此类型数值的字符。请注意，分隔符是与数组元素数据类型相关联，而不是与数组数据类型相关联。
typrelid	oid	如果是复合类型（请参见typtype），则此字段指向pg_class中定义该表的行。对于独立的复合类型，pg_class记录并不表示一个表，但是总需要它来查找该类型连接的pg_attribute记录。非复合类型为0。

名称	类型	描述
typelem	oid	如果不为0，则它标识pg_type中的另一行。当前类型可以像一个产生类型为typelem的数组来描述。 “true”数组类型是变长的（typlen= -1），但是某些定长（typlen > 0）类型也有非零的typelem（比如name和point）。如果一个定长类型有typelem，则其内部形式必须是typelem数据类型的某个数目的个数值，不能有其他数据。变长数组类型有一个该数组子过程定义的头（文件）。
typarray	oid	如果不为0，则表示在pg_type中有对应的类型记录。
typinput	regproc	输入转换函数（文本格式）。
typoutput	regproc	输出转换函数（文本格式）。
typreceive	regproc	输入转换函数（二进制格式），如果没有则为0。
typsend	regproc	输出转换函数（二进制格式），如果没有则为0。
typmodin	regproc	输入类型修改符函数，如果为0，则不支持。
typmodout	regproc	输出类型修改符函数，如果为0，则不支持。
typanalyze	regproc	自定义的ANALYZE函数，如果使用标准函数，则为0。
typalign	"char"	<p>当存储此类型的数值时要求的对齐方式。适用于磁盘存储以及该值在数据库中的大多数形式。如果数值是连续存储的，比如在磁盘上以完全的裸数据的形式存放时，则先在此类型的数据前填充空白，这样它就可以按照要求的边界存储。对齐引用是该序列中第一个数据的开头。可能的值包含：</p> <ul style="list-style-type: none"> <li>• c = char对齐，即不需要对齐。</li> <li>• s = short对齐（在大多数机器上是2字节）</li> <li>• i = int对齐（在大多数机器上是4字节）</li> <li>• d = double对齐（在大多数机器上是8字节，但不一定是全部）</li> </ul> <p><b>须知</b> 对于系统表里使用的类型，在pg_type里定义的尺寸和对齐方式要和编译器在表示表行的结构中布局方式保持一致。</p>

名称	类型	描述
typstorage	"char"	指明一个变长类型（那些有typlen = -1）是否准备好应付非常规值，以及对这种属性的类型的缺省策略是什么。可能的值包含： <ul style="list-style-type: none"> <li>• p: 数值总是以简单方式存储。</li> <li>• e: 数值可以存储在一个"次要"关系中（如果有该关系，请参见pg_class.reltoastrelid）。</li> <li>• m: 数值可以以内联压缩方式存储。</li> <li>• x: 数值可以以内联压缩方式或者在"次要"表里存储。</li> </ul> <b>须知</b> m域也可以移到从属表里存储，但只是最后的解决方法（首先移动e和x域）。
typenotnull	boolean	表示在某类型上的一个NOTNULL约束。目前只用于域。
typbasetype	oid	如果这是一个衍生类型（请参见typtype），则该标识作为这个类型的基础的类型。如果不是衍生类型则为零。
typtypmod	integer	域使用typtypmod记录要应用于其基础类型上的typmod（如果基础类型不使用typmod，则为-1）。如果此类型不是域，则为-1。
typndims	integer	如果一个域是数组，则typndims是数组维数的数值（即typbasetype是一个数组类型；域的类型elem将匹配基本类型的typelem）。除了数组类型的域以外的类型为0。
typcollation	oid	指定类型的排序规则。如果为0，则表示不支持排序规则。
typdefaultbin	pg_node_tree	如果不为NULL，则为该类型缺省表达式的nodeToString()表现形式。目前这个字段只用于域。
typdefault	text	如果某类型没有相关缺省值，则为NULL。如果typdefaultbin不为NULL，则typdefault必须包含一个typdefaultbin代表的缺省表达式的人类可读版本。如果typdefaultbin为NULL但typdefault不为NULL，typdefault则是该类型缺省值的外部表现形式，可以将其输入到类型的转换器生成一个常量。
typacl	aclitem[]	访问权限。

## 15.2.77 PG\_USER\_MAPPING

PG\_USER\_MAPPING系统表存储从本地用户到远程的映射。

需要有系统管理员权限才可以访问此系统表。普通用户可以使用视图 [PG\\_USER\\_MAPPINGS](#) 进行查询。

表 15-76 PG\_USER\_MAPPING 字段

名字	类型	引用	描述
oid	oid	-	行标识符（隐藏属性，必须明确选择才会显示）。
umuser	oid	PG_AUTHID.oid	被映射的本地用户的OID，如果用户映射是公共的则为0。
umserver	oid	PG_FOREIGN_SERVER.oid	包含此映射的外部服务器的OID。
umoptions	text[]	-	用户映射指定选项，使用“keyword=value”格式的字符串。

## 15.2.78 PG\_USER\_STATUS

PG\_USER\_STATUS系统表提供了访问数据库用户的状态。需要有系统管理员权限才可以访问此系统表

表 15-77 PG\_USER\_STATUS 字段

名称	类型	描述
rolid	oid	角色的标识。
failcount	integer	尝试失败次数。
locktime	timestamp with time zone	角色被锁定的时间点。
rolstatus	smallint	角色的状态。 <ul style="list-style-type: none"> <li>0：正常状态。</li> <li>1：由于登录失败次数超过阈值被锁定了一定的时间。</li> <li>2：被管理员锁定。</li> </ul>
permpspace	bigint	角色在当前实例上已经使用的永久表存储空间大小。
tempSPACE	bigint	角色在当前实例上已经使用的临时表存储空间大小。

## 15.2.79 PG\_WORKLOAD\_ACTION

PG\_WORKLOAD\_ACTION系统表存储query\_band的信息。



表 15-78 PG\_WORKLOAD\_ACTION 字段

名称	类型	描述
qband	name	query band键值对
class	name	query band关联行为类别
object	name	query band关联行为
action	name	query band关联行为表现

## 15.2.80 PGXC\_CLASS

PGXC\_CLASS系统表存储每张表的复制或分布信息。

表 15-79 PGXC\_CLASS 字段

名称	类型	描述
pcrelid	oid	表的OID
pclocatortype	"char"	定位器类型 <ul style="list-style-type: none"> <li>• H: hash</li> <li>• M: Modulo</li> <li>• N: Round Robin</li> <li>• R: Replicate</li> </ul>
pchashalgorithm	smallint	使用哈希算法分布元组
pchashbuckets	smallint	哈希容器的值
pgroup	name	节点组名称
redistributed	"char"	表已经完成重分布
redis_order	integer	重分布的顺序
pcttnum	int2vector	用作分布键的列标号
nodeoids	oidvector_extend	表分布的节点OID列表
options	text	系统内部保留字段，存储扩展状态信息

## 15.2.81 PGXC\_GROUP

PGXC\_GROUP系统表存储节点组信息。

表 15-80 PGXC\_GROUP 字段

名称	类型	描述
group_name	name	节点组名称
in_redistribution	"char"	是否需要重分布 <ul style="list-style-type: none"> <li>• n表示NodeGroup没有再进行重分布</li> <li>• y表示NodeGroup是重分布过程中的源节点组</li> <li>• t表示NodeGroup是重分布过程中的目的节点组</li> </ul>
group_members	oidvector_extend	节点组的节点OID列表
group_buckets	text	分布数据桶的集合
is_installation	boolean	是否安装子集群
group_acl	aclitem[]	访问权限
group_kind	"char"	node group类型 <ul style="list-style-type: none"> <li>• i表示installation node group</li> <li>• n表示普通非逻辑集群node group</li> <li>• v表示逻辑集群node group</li> <li>• e表示弹性集群</li> </ul>

## 15.2.82 PGXC\_NODE

PGXC\_NODE系统表存储集群节点信息。

表 15-81 PGXC\_NODE 字段

名称	类型	描述
node_name	name	节点名称。
node_type	"char"	节点类型。 C: 协调节点。 D: 数据节点。
node_port	integer	节点的端口号。
node_host	name	节点的主机名称或者IP（如配置为虚拟IP，则为虚拟IP）。
node_port1	integer	复制节点的端口号。
node_host1	name	复制节点的主机名称或者IP（如配置为虚拟IP，则为虚拟IP）。

名称	类型	描述
hostis_primary	boolean	表明当前节点是否发生主备切换。
nodeis_primary	boolean	在replication表下，是否优选当前节点作为优先执行的节点进行非查询操作。
nodeis_preferre d	boolean	在replication表下，是否优选当前节点作为首选的节点进行查询。
node_id	integer	节点标识符。
sctp_port	integer	主节点使用TCP代理通信库或SCTP通信库的数据通道监听端口。
control_port	integer	主节点使用TCP代理通信库或SCTP通信库的控制通道监听端口。
sctp_port1	integer	备节点使用TCP代理通信库或SCTP通信库的数据通道监听端口。
control_port1	integer	备节点使用TCP代理通信库或SCTP通信库的控制通道监听端口。
nodeis_central	boolean	当前节点为中心控制节点。

## 应用示例

查询集群的CN和DN信息：

```
select * from pgxc_node;
 node_name | node_type | node_port | node_host | node_port1 | node_host1 | hostis_primary |
 nodeis_primary | nodeis_preferred
 | node_id | sctp_port | control_port | sctp_port1 | control_port1 | nodeis_central
-----+-----+-----+-----+-----+-----+-----+-----+-----+
+-----+
dn_6001_6002 | D      | 40000 | 192.**.**.**1 | 45000 | 192.**.**.**2 | t      | f      | f
 | 1644780306 | 40002 | 40003 | 45002 | 45003 | f
dn_6003_6004 | D      | 40000 | 192.**.**.**2 | 45000 | 192.**.**.**3 | t      | f      | f
 | -966646068 | 40002 | 40003 | 45002 | 45003 | f
dn_6005_6006 | D      | 40000 | 192.**.**.**3 | 45000 | 192.**.**.**1 | t      | f      | f
 | 868850011 | 40002 | 40003 | 45002 | 45003 | f
cn_5001 | C      | 8000 | 192.**.**.**1 | 8000 | 192.**.**.**1 | t      | f      | f
 | 1120683504 | 8002 | 8003 | 0 | 0 | f
cn_5002 | C      | 8000 | 192.**.**.**2 | 8000 | 192.**.**.**2 | t      | f      | f
 | -1736975100 | 8002 | 8003 | 0 | 0 | f
cn_5003 | C      | 8000 | localhost | 8000 | localhost | t      | f      | f
 | -125853378 | 8002 | 8003 | 0 | 0 | t
(6 rows)
```

### 15.2.83 SNAPSHOT

SNAPSHOT系统表记录每次创建性能视图快照的起止时间，设置enable\_wdr\_snapshot为on后，该表由后台快照线程创建并维护。需要有系统管理员权限才可以访问此系统表。

表 15-82 dbms\_om.snapshot 字段

名称	类型	描述
snapshot_id	name	快照ID（此字段为主键和分布键）
start_ts	timestamp with time zone	快照开始时间
end_ts	timestamp with time zone	快照结束时间

#### 须知

- 此系统表的schema是dbms\_om。
- 禁止从外部修改或删除此表，否则可能引起视图快照相关功能异常。

## 15.2.84 TABLES\_SNAP\_TIMESTAMP

TABLES\_SNAP\_TIMESTAMP系统表记录每次对每个性能视图创建快照的起止时间，设置enable\_wdr\_snapshot为on后，该表由后台快照线程创建并维护。需要有系统管理员权限才可以访问此系统表。

表 15-83 dbms\_om.tables\_snap\_timestamp 字段

名称	类型	描述
snapshot_id	name	快照ID（此字段为主键和分布键）
db_name	text	视图所属数据库名称
tablename	text	视图名称
start_ts	timestamp with time zone	快照开始时间
end_ts	timestamp with time zone	快照结束时间

#### 须知

- 此系统表的schema是dbms\_om。
- 禁止从外部修改或删除此表，否则可能引起视图快照相关功能异常。

## 15.2.85 性能视图快照系统表

设置enable\_wdr\_snapshot为on后，后台快照线程会创建并维护以“SNAP\_+视图名称”方式命名的系统表，用以记录各性能视图的快照结果。需要有系统管理员权限才可以访问下列系统表。

- SNAP\_PGXC\_OS\_RUN\_INFO
- SNAP\_PGXC\_WAIT\_EVENTS
- SNAP\_PGXC\_INSTR\_UNIQUE\_SQL
- SNAP\_PGXC\_STAT\_BAD\_BLOCK
- SNAP\_PGXC\_STAT\_BGWRITER
- SNAP\_PGXC\_STAT\_REPLICATION
- SNAP\_PGXC\_REPLICATION\_SLOTS
- SNAP\_PGXC\_SETTINGS
- SNAP\_PGXC\_INSTANCE\_TIME
- SNAP\_GLOBAL\_WORKLOAD\_TRANSACTION
- SNAP\_PGXC\_WORKLOAD\_SQL\_COUNT
- SNAP\_PGXC\_STAT\_DATABASE
- SNAP\_GLOBAL\_STAT\_DATABASE
- SNAP\_PGXC\_REDO\_STAT
- SNAP\_GLOBAL\_REDO\_STAT
- SNAP\_PGXC\_REL\_IOSTAT
- SNAP\_GLOBAL\_REL\_IOSTAT
- SNAP\_PGXC\_TOTAL\_MEMORY\_DETAIL
- SNAP\_PGXC\_NODE\_STAT\_RESET\_TIME
- SNAP\_PGXC\_SQL\_COUNT
- SNAP\_GLOBAL\_TABLE\_STAT
- SNAP\_GLOBAL\_TABLE\_CHANGE\_STAT
- SNAP\_GLOBAL\_COLUMN\_TABLE\_IO\_STAT
- SNAP\_GLOBAL\_ROW\_TABLE\_IO\_STAT

此类系统表除增加snapshot\_id字段（bigint类型）外，其余的字段定义与对应视图相同，且各表的分布键均为snapshot\_id。

例如，SNAP\_PGXC\_OS\_RUN\_INFO，用于存储PGXC\_OS\_RUN\_INFO视图的快照，其字段新增了snapshot\_id，其余字段含义均与PGXC\_OS\_RUN\_INFO视图相同。

#### 须知

- 以上系统表的schema都是dbms\_om。
- 禁止从外部修改或删除以上系统表，否则可能引起视图快照相关功能异常。

## 15.3 系统视图

### 15.3.1 ALL\_ALL\_TABLES

ALL\_ALL\_TABLES视图存储当前用户所能访问的表或视图。

表 15-84 ALL\_ALL\_TABLES 字段

名称	类型	描述
owner	name	表或视图的所有者
table_name	name	表或视图的名称
tablespace_name	name	表或视图所在的表空间

## 15.3.2 ALL\_CONSTRAINTS

ALL\_CONSTRAINTS视图存储当前用户可访问的约束的信息。

表 15-85 ALL\_CONSTRAINTS 字段

名称	类型	描述
constraint_name	vcharacter varying(64)	约束名
constraint_type	text	约束类型 <ul style="list-style-type: none"> <li>• c表示检查约束</li> <li>• f表示外键约束</li> <li>• p表示主键约束</li> <li>• u表示唯一约束</li> </ul>
table_name	character varying(64)	约束相关的表名
index_owner	character varying(64)	约束相关的索引的所有者（只针对唯一约束和主键约束）
index_name	character varying(64)	约束相关的索引名（只针对唯一约束和主键约束）

## 15.3.3 ALL\_CONS\_COLUMNS

ALL\_CONS\_COLUMNS视图存储当前用户可访问的约束字段的信息。

表 15-86 ALL\_CONS\_COLUMNS 字段

名称	类型	描述
table_name	character varying(64)	约束相关的表名
column_name	character varying(64)	约束相关的列名

名称	类型	描述
constraint_name	character varying(64)	约束名
position	smallint	表中列的位置

### 15.3.4 ALL\_COL\_COMMENTS

ALL\_COL\_COMMENTS视图存储当前用户可访问的表和视图中字段的注释信息。

表 15-87 ALL\_COL\_COMMENTS 字段

名称	类型	描述
column_name	character varying(64)	列名
table_name	character varying(64)	表名或视图名
owner	character varying(64)	表或视图的所有者
comments	text	注释

### 15.3.5 ALL\_DEPENDENCIES

ALL\_DEPENDENCIES视图存储了当前用户可访问的函数、高级包之间的依赖关系。

#### 须知

因为相关信息的限制，目前GaussDB(DWS)中，此表为空表，表内没有任何记录。

表 15-88 ALL\_DEPENDENCIES 字段

名称	类型	描述
owner	character varying(30)	对象的所有者
name	character varying(30)	对象的名称
type	character varying(17)	对象的类型
referenced_owner	character varying(30)	引用对象的所有者
referenced_name	character varying(64)	引用对象的名称
referenced_type	character varying(17)	引用对象的类型
referenced_link_name	character varying(128)	引用对象的链接的名称
schemaid	numeric	当前schema的ID

名称	类型	描述
dependency_type	character varying(4)	依赖类型（REF或HARD）

### 15.3.6 ALL\_IND\_COLUMNS

ALL\_IND\_COLUMNS视图存储了当前用户可访问的所有索引的字段信息。

表 15-89 ALL\_IND\_COLUMNS 字段

名称	类型	描述
index_owner	character varying(64)	索引的所有者
index_name	character varying(64)	索引名
table_owner	character varying(64)	表的所有者
table_name	character varying(64)	表名
column_name	name	列名
column_position	smallint	索引中列的位置

### 15.3.7 ALL\_IND\_EXPRESSIONS

ALL\_IND\_EXPRESSIONS视图存储了当前用户可访问的表达式索引的信息。

表 15-90 ALL\_IND\_EXPRESSIONS 字段

名称	类型	描述
index_owner	character varying(64)	索引的所有者
index_name	character varying(64)	索引名
table_owner	character varying(64)	表的所有者
table_name	character varying(64)	表名
column_expression	text	定义列的基于函数的索引表达式
column_position	smallint	索引中列的位置

### 15.3.8 ALL\_INDEXES

ALL\_INDEXES视图存储了当前用户可访问的索引信息。



表 15-91 ALL\_INDEXES 字段

名称	类型	描述
owner	character varying(64)	索引的所有者
index_name	character varying(64)	索引名
table_name	character varying(64)	索引对应的表名
uniqueness	text	表示索引是否为唯一索引
generated	character varying(1)	表示索引的名称是否为系统生成
partitioned	character(3)	表示索引是否具有分区表的性质

### 15.3.9 ALL\_OBJECTS

ALL\_OBJECTS视图记录了当前用户可访问的数据库对象。

表 15-92 ALL\_OBJECTS 字段

名称	类型	描述
owner	name	对象的所有者
object_name	name	对象的名称
object_id	oid	对象的OID
object_type	name	对象的类型
namespace	oid	对象所在的命名空间
created	timestamp with time zone	对象的创建时间
last_ddl_time	timestamp with time zone	对象的最后修改时间

#### 须知

created和last\_ddl\_time支持的范围参见PG\_OBJECT中的记录范围。

### 15.3.10 ALL\_PROCEDURES

ALL\_PROCEDURES视图存储了当前用户可访问的所有存储过程或函数信息。

表 15-93 ALL\_PROCEDURES 字段

名称	类型	描述
owner	name	对象的所有者
object_name	name	对象名称

### 15.3.11 ALL\_SEQUENCES

ALL\_SEQUENCES视图存储当前用户能够访问的所有序列。

表 15-94 ALL\_SEQUENCES 字段

名称	类型	描述
sequence_owner	name	序列所有者
sequence_name	name	序列的名称
min_value	bigint	序列最小值
max_value	bigint	序列最大值
increment_by	bigint	序列的增量
cycle_flag	character(1)	表示序列是否是循环序列，取值为Y或N <ul style="list-style-type: none"> <li>• Y表示是循环序列</li> <li>• N表示不是循环序列</li> </ul>

### 15.3.12 ALL\_SOURCE

ALL\_SOURCE视图存储当前用户可访问的存储过程或函数信息，且提供存储过程或函数定义的字段。

表 15-95 ALL\_SOURCE 字段

名称	类型	描述
owner	name	对象的所有者
name	name	对象名称
type	name	对象的类型
text	text	对象的定义

### 15.3.13 ALL\_SYNONYMS

ALL\_SYNONYMS视图存储了当前用户可访问的所有同义词信息。

表 15-96 ALL\_SYNONYMS 字段

名称	类型	描述
owner	text	同义词的所有者
schema_name	text	同义词所属模式名
synonym_name	text	同义词的名称
table_owner	text	关联对象的所有者
table_schema_name	text	关联对象所属模式名
table_name	text	关联对象名

### 15.3.14 ALL\_TAB\_COLUMNS

ALL\_TAB\_COLUMNS视图存储了当前用户可访问的表和视图的列的描述信息。

表 15-97 ALL\_TAB\_COLUMNS 字段

名称	类型	描述
owner	character varying(64)	表或视图的所有者。
table_name	character varying(64)	表名或视图名。
column_name	character varying(64)	列名。
data_type	character varying(128)	列的数据类型。
column_id	integer	对象创建或增加列时列的序号。
data_length	integer	列的字节长度。
avg_column_len	numeric	列的平均长度（单位字节）。

名称	类型	描述
nullable	bpchar	该列是否允许为空，对于主键约束和非空约束，该值为n。
data_precision	integer	数据类型的精度，对于numeric数据类型有效，其他类型为NULL。
data_scale	integer	小数点右边的位数，对于numeric数据类型有效，其他类型为0。
char_length	numeric	列的长度（单位字符），只对varchar, nvarchar2, bpchar, char类型有效。
schema	character varying(64)	包含该表或视图的命名空间。
kind	text	当前记录所属的种类，如果此列属于表，则此字段显示为table；如果此列属于视图，则此字段显示为view。

### 15.3.15 ALL\_TAB\_COMMENTS

ALL\_TAB\_COMMENTS视图存储当前用户可访问的所有表和视图的注释信息。

表 15-98 ALL\_TAB\_COMMENTS 字段

名称	类型	描述
owner	character varying(64)	表或视图的所有者
table_name	character varying(64)	表或视图的名称
comments	text	注释

### 15.3.16 ALL\_TABLES

ALL\_TABLES视图存储当前用户可访问的所有表。

表 15-99 ALL\_TABLES 字段

名称	类型	描述
owner	character varying(64)	表的所有者
table_name	character varying(64)	表名
tablespace_name	character varying(64)	表所在的表空间名称
status	character varying(8)	当前记录是否有效

名称	类型	描述
temporary	character(1)	表是否为临时表 <ul style="list-style-type: none"> <li>Y表示是临时表</li> <li>N表示不是临时表</li> </ul>
dropped	character varying	当前记录是否已删除 <ul style="list-style-type: none"> <li>YES表示已删除</li> <li>NO表示未删除</li> </ul>
num_rows	numeric	表的估计行数

### 15.3.17 ALL\_USERS

ALL\_USERS视图存储记录数据库中所有用户，但不对用户信息进行详细的描述。

表 15-100 ALL\_USERS 字段

名称	类型	描述
username	name	用户名
user_id	oid	用户的OID

### 15.3.18 ALL\_VIEWS

ALL\_VIEWS视图存储了当前用户可访问的所有视图描述信息。

表 15-101 ALL\_VIEWS 字段

名称	类型	描述
owner	name	视图的所有者
view_name	name	视图的名称
text_length	integer	视图文本长度
text	text	视图文本

### 15.3.19 DBA\_DATA\_FILES

DBA\_DATA\_FILES视图存储关于数据库文件的描述。需要有系统管理员权限才可以访问。

表 15-102 DBA\_DATA\_FILES 字段

名称	类型	描述
tablespace_name	name	文件所属的表空间的名称
bytes	double precision	文件的字节长度

### 15.3.20 DBA\_USERS

DBA\_USERS视图存储关于数据库所有用户名信息。需要有系统管理员权限才可以访问。

表 15-103 DBA\_USERS 字段

名称	类型	描述
username	character varying(64)	用户名

### 15.3.21 DBA\_COL\_COMMENTS

DBA\_COL\_COMMENTS视图存储关于数据库中表和视图中字段的注释信息。需要有系统管理员权限才可以访问。

名称	类型	描述
column_name	character varying(64)	列名
table_name	character varying(64)	表名或视图名
owner	character varying(64)	表或视图的所有者
comments	text	注释

### 15.3.22 DBA\_CONSTRAINTS

DBA\_CONSTRAINTS视图存储关于数据库表中约束的信息。需要有系统管理员权限才可以访问。

名称	类型	描述
constraint_name	vcharacter varying(64)	约束名

名称	类型	描述
constraint_type	text	约束类型 <ul style="list-style-type: none"> <li>• c表示检查约束</li> <li>• f表示外键约束</li> <li>• p表示主键约束</li> <li>• u表示唯一约束</li> </ul>
table_name	character varying(64)	约束相关的表名
index_owner	character varying(64)	约束相关的索引的所有者（只针对唯一约束和主键约束）
index_name	character varying(64)	约束相关的索引名（只针对唯一约束和主键约束）

### 15.3.23 DBA\_CONS\_COLUMNS

DBA\_CONS\_COLUMNS视图存储关于数据库表中约束字段的信息。需要有系统管理员权限才可以访问。

名称	类型	描述
table_name	character varying(64)	约束相关的表名
column_name	character varying(64)	约束相关的列名
constraint_name	character varying(64)	约束名
position	smallint	表中列的位置

### 15.3.24 DBA\_IND\_COLUMNS

DBA\_IND\_COLUMNS视图存储关于数据库中所有索引的字段信息。需要有系统管理员权限才可以访问。

名称	类型	描述
index_owner	character varying(64)	索引的所有者
index_name	character varying(64)	索引名
table_owner	character varying(64)	表的所有者
table_name	character varying(64)	表名

名称	类型	描述
column_name	name	列名
column_position	smallint	索引中列的位置

### 15.3.25 DBA\_IND\_EXPRESSIONS

DBA\_IND\_EXPRESSIONS视图存储了数据库中的表达式索引的信息。需要有系统管理员权限才可以访问。

名称	类型	描述
index_owner	character varying(64)	索引的所有者
index_name	character varying(64)	索引名
table_owner	character varying(64)	表的所有者
table_name	character varying(64)	表名
column_expression	text	定义列的基于函数的索引表达式
column_position	smallint	索引中列的位置

### 15.3.26 DBA\_IND\_PARTITIONS

DBA\_IND\_PARTITIONS视图存储数据库中所有索引分区的信息。数据库中每个分区表的每个索引分区（如果存在的话）在DBA\_IND\_PARTITIONS里都会有一行记录。需要有系统管理员权限才可以访问。

名称	类型	描述
index_owner	character varying(64)	索引分区所属分区表索引的所有者的名称
schema	character varying(64)	索引分区所属分区表索引的模式
index_name	character varying(64)	索引分区所属分区表索引的名称
partition_name	character varying(64)	索引分区的名称
index_partition_usable	boolean	索引分区是否可用



名称	类型	描述
high_value	text	索引分区所对应的表分区的边界(范围分区为上边界, 列表分区为边界值集合)。前向兼容的保留字段, 8.1.3集群版本新增pretty_high_value用于记录此信息。
pretty_high_value	text	索引分区所对应的表分区的边界(范围分区为上边界, 列表分区为边界值集合)。查询结果为表分区对应边界表达式的即时反编译输出。该字段的输出比high_value的信息更详细, 根据实际使用场景可输出collaton、字段数据类型等信息。
def_tablespace_name	name	索引分区的表空间名称

### 15.3.27 DBA\_INDEXES

DBA\_INDEXES视图存储关于数据库下的所有索引信息。需要有系统管理员权限才可以访问。

名称	类型	描述
owner	character varying(64)	索引的所有者
index_name	character varying(64)	索引名
table_name	character varying(64)	索引对应的表名
uniqueness	text	表示索引是否为唯一索引
generated	character varying(1)	表示索引名称是否为系统生成
partitioned	character(3)	表示索引是否具有分区表的性质

### 15.3.28 DBA\_OBJECTS

DBA\_OBJECTS视图存储了数据库中所有数据库对象。需要有系统管理员权限才可以访问。

名称	类型	描述
owner	name	对象的所有者
object_name	name	对象的名称
object_id	oid	对象的OID

名称	类型	描述
object_type	name	对象的类型
namespace	oid	对象所在的命名空间
created	timestamp with time zone	对象的创建时间
last_ddl_time	timestamp with time zone	对象的最后修改时间

#### 须知

created和last\_ddl\_time支持的范围参见PG\_OBJECT中的记录范围。

### 15.3.29 DBA\_PART\_INDEXES

DBA\_PART\_INDEXES视图存储数据库中所有分区表索引的信息。需要有系统管理员权限才可以访问。

名称	类型	描述
index_owner	character varying(64)	分区表索引的所有者名称
schema	character varying(64)	分区表索引的模式
index_name	character varying(64)	分区表索引的名称
table_name	character varying(64)	分区表索引所属的分区表名称
partitioning_type	text	分区表的分区策略 <b>说明</b> 当前分区表策略仅支持范围分区（Range Partitioning）和列表分区（List Partitioning）。
partition_count	bigint	分区表索引的索引分区的个数
def_tablespace_name	name	分区表索引的表空间名称
partitioning_key_count	integer	分区表的分区键个数

### 15.3.30 DBA\_PART\_TABLES

DBA\_PART\_TABLES视图存储数据中所有分区表的信息。需要有系统管理员权限才可以访问。

名称	类型	描述
table_owner	character varying(64)	分区表的所有者名称
schema	character varying(64)	分区表的模式
table_name	character varying(64)	分区表的名称
partitioning_type	text	分区表的分区策略 <b>说明</b> 当前分区表策略仅支持范围分区（Range Partitioning）和列表分区（List Partitioning）。
partition_count	bigint	分区表的分区个数
def_tablespace_name	name	分区表的表空间名称
partitioning_key_count	integer	分区表的分区键个数

### 15.3.31 DBA\_PROCEDURES

DBA\_PROCEDURES视图存储关于数据库下的所有存储过程或函数信息。需要有系统管理员权限才可以访问。

名称	类型	描述
owner	character varying(64)	存储过程或函数的所有者
object_name	character varying(64)	存储过程或函数名称
argument_number	smallint	存储过程入参个数

### 15.3.32 DBA\_SEQUENCES

DBA\_SEQUENCES视图存储关于数据库下的所有序列信息。需要有系统管理员权限才可以访问。

名称	类型	描述
sequence_owner	character varying(64)	序列的所有者
sequence_name	character varying(64)	序列的名称

### 15.3.33 DBA\_SOURCE

DBA\_SOURCE视图存储关于数据库下的所有存储过程或函数信息，且提供存储过程或函数定义的字段。需要有系统管理员权限才可以访问。

名称	类型	描述
owner	character varying(64)	存储过程或函数的所有者
name	character varying(64)	存储过程或函数的名称
text	text	存储过程或函数的定义

### 15.3.34 DBA\_SYNONYMS

DBA\_SYNONYMS视图存储关于数据库下的所有同义词信息。需要有系统管理员权限才可以访问。

表 15-104 DBA\_SYNONYMS 字段

名称	类型	描述
owner	text	同义词的所有者
schema_name	text	同义词所属模式名
synonym_name	text	同义词的名称
table_owner	text	关联对象的所有者
table_schema_name	text	关联对象所属模式名
table_name	text	关联对象名

### 15.3.35 DBA\_TAB\_COLUMNS

DBA\_TAB\_COLUMNS视图存储关于表和视图的字段的信息。数据库里每个表的每个字段都在DBA\_TAB\_COLUMNS里有一行。需要有系统管理员权限才可以访问。

名称	类型	描述
owner	character varying(64)	表或视图的所有者
table_name	character varying(64)	表名或视图名

名称	类型	描述
column_name	character varying(64)	列名
data_type	character varying(128)	列的数据类型
column_id	integer	创建表或视图时列的序号
data_length	integer	列的字节长度
comments	text	注释
avg_column_len	numeric	列的平均长度（单位字节）
nullable	bpchar	该列是否允许为空，对于主键约束和非空约束，该值为n
data_precision	integer	数据类型的精度，对于numeric数据类型有效，其他类型为NULL
data_scale	integer	小数点右边的位数，对于numeric数据类型有效，其他类型为0
char_length	numeric	列的长度（以字符计），只对varchar，nvarchar2，bpchar，char类型有效
schema	character varying(64)	包含该表或视图的命名空间。
kind	text	当前记录所属的种类，如果此列属于表，则此字段显示为table；如果此列属于视图，则此字段显示为view。

### 15.3.36 DBA\_TAB\_COMMENTS

DBA\_TAB\_COMMENTS视图存储关于数据库下的所有表和视图的注释信息。需要有系统管理员权限才可以访问。

名称	类型	描述
owner	character varying(64)	表或视图的所有者
table_name	character varying(64)	表或视图的名称
comments	text	注释

### 15.3.37 DBA\_TAB\_PARTITIONS

DBA\_TAB\_PARTITIONS视图提供数据库中所有分区的信息。

名称	类型	描述
table_owner	character varying(64)	分区所在表的所有者
schema	character varying(64)	分区表模式
table_name	character varying(64)	表名
partition_name	character varying(64)	分区的名称
high_value	text	范围分区的上边界，或列表分区的边界值集合。前向兼容的保留字段，8.1.3集群版本新增pretty_high_value用于记录此信息。
pretty_high_value	text	范围分区的上边界，或列表分区的边界值集合。查询结果为表分区对应边界表达式的即时反编译输出。该字段的输出比high_value的信息更详细，根据实际使用场景可输出collaton、字段数据类型等信息。
tablespace_name	name	分区所在表空间的名称

### 15.3.38 DBA\_TABLES

DBA\_TABLES视图存储关于数据库下的所有表信息。需要有系统管理员权限才可以访问。

名称	类型	描述
owner	character varying(64)	表的所有者
table_name	character varying(64)	表名
tablespace_name	character varying(64)	表所在的表空间的名称
status	character varying(8)	当前记录是否有效
temporary	character(1)	是否为临时表 <ul style="list-style-type: none"> <li>• Y表示是临时表</li> <li>• N表示不是临时表</li> </ul>

名称	类型	描述
dropped	character varying	当前记录是否已删除 <ul style="list-style-type: none"> <li>• YES表示已删除</li> <li>• NO表示未删除</li> </ul>
num_rows	numeric	表的估计行数

### 15.3.39 DBA\_TABLESPACES

DBA\_TABLESPACES视图存储有关可用的表空间的信息。需要有系统管理员权限才可以访问。

表 15-105 DBA\_TABLESPACES 字段

名称	类型	描述
tablespace_name	character varying(64)	表空间的名称

### 15.3.40 DBA\_TRIGGERS

DBA\_TRIGGERS视图存储关于数据库内的触发器信息。需要有系统管理员权限才可以访问。

名称	类型	描述
trigger_name	character varying(64)	触发器名称
table_name	character varying(64)	定义触发器的表的名称
table_owner	character varying(64)	定义触发器的表的所有者

### 15.3.41 DBA\_VIEWS

DBA\_VIEWS视图存储关于数据库内的视图信息。需要有系统管理员权限才可以访问。

名称	类型	描述
owner	character varying(64)	视图的所有者
view_name	character varying(64)	视图的名称

### 15.3.42 DUAL

DUAL视图是数据库根据数据字典自动创建的，它只有一个文本字段，且只有一行，用于保存表达式计算结果。任何用户都可以访问它。

表 15-106 DUAL 字段

名称	类型	描述
dummy	text	表达式计算结果

### 15.3.43 GLOBAL\_COLUMN\_TABLE\_IO\_STAT

GLOBAL\_COLUMN\_TABLE\_IO\_STAT视图提供当前数据库所有列存表的IO统计数据。其字段的名称、类型和顺序与GS\_COLUMN\_TABLE\_IO\_STAT视图相同，具体的字段请参考[GS\\_COLUMN\\_TABLE\\_IO\\_STAT](#)。各统计字段为所有节点对应字段之和。

### 15.3.44 GLOBAL\_REDO\_STAT

GLOBAL\_REDO\_STAT视图显示集群中所有节点上XLOG重做过程中的统计信息总和。除avgiotim（表示所有节点平均的重做写入时间）外，其余字段名称和[PV\\_REDO\\_STAT](#)视图相同，但其余字段含义为各节点上PV\_REDO\_STAT视图同名字段的数值之和。

需要有系统管理员权限才可以访问此视图。

### 15.3.45 GLOBAL\_REL\_IOSTAT

GLOBAL\_REL\_IOSTAT视图显示集群中所有节点上磁盘读写统计信息的总和。其各字段的名称与[GS\\_REL\\_IOSTAT](#)视图相同，但含义为各节点上GS\_REL\_IOSTAT视图同名字段的数值之和。需要有系统管理员权限才可以访问此视图。

### 15.3.46 GLOBAL\_ROW\_TABLE\_IO\_STAT

GLOBAL\_ROW\_TABLE\_IO\_STAT视图提供当前数据库所有行存表的IO统计数据。其字段的名称、类型和顺序与GS\_ROW\_TABLE\_IO\_STAT视图相同，具体的字段请参考[GS\\_ROW\\_TABLE\\_IO\\_STAT](#)。各统计字段为所有节点对应字段之和。

### 15.3.47 GLOBAL\_STAT\_DATABASE

GLOBAL\_STAT\_DATABASE视图显示集群中所有节点上数据库的状态和统计信息之和。

- CN上查询GLOBAL\_STAT\_DATABASE视图，返回的结果除stats\_reset字段（当前CN上的状态重置时间）之外，其余字段表示在集群内相关节点上的数值之和。需注意，因GLOBAL\_STAT\_DATABASE视图中各字段的逻辑含义不同，求和的范围也有所不同。
- 在DN上查询视图GLOBAL\_STAT\_DATABASE，所得结果与[表15-107](#)相同。

表 15-107 GLOBAL\_STAT\_DATABASE 字段

名称	类型	描述	求和范围
datid	oid	数据库OID。	-
datname	name	数据库名。	-



名称	类型	描述	求和范围
numbackends	integer	当前节点上连接到该数据库的后端数。这是该视图中唯一一个反映目前状态值的列；所有列均返回自上次重置以来的累积值。	CN
xact_commit	bigint	当前节点上该数据库中已经提交的事务数。	CN
xact_rollback	bigint	当前节点上该数据库中已经回滚的事务数。	CN
blks_read	bigint	当前节点上该数据库中读取的磁盘块的数量。	DN
blks_hit	bigint	当前节点上高速缓存中发现的磁盘块的个数，即缓存中命中的块数（只包括GaussDB(DWS)缓冲区高速缓存，不包括文件系统的缓存）。	DN
tup_returned	bigint	当前节点上该数据库查询返回的行数。	DN
tup_fetched	bigint	当前节点上该数据库查询抓取的行数。	DN
tup_inserted	bigint	当前节点上该数据库插入的行数。	DN
tup_updated	bigint	当前节点上该数据库更新的行数。	DN
tup_deleted	bigint	当前节点上该数据库删除的行数。	DN
conflicts	bigint	当前节点上由于数据库恢复冲突取消的查询数量（只在备用服务器上发生）。请参见 <a href="#">PG_STAT_DATABASE_CONFLICTS</a> 获取更多信息。	CN和DN
temp_files	bigint	当前节点上该数据库创建的临时文件个数。计算所有临时文件，不论为什么创建临时文件（比如排序或者哈希），而且不考虑 log_temp_files 设置。	DN
temp_bytes	bigint	当前节点上该数据库写入临时文件的大小。计算所有临时文件，不论为什么创建临时文件，而且不考虑 log_temp_files 设置。	DN
deadlocks	bigint	当前节点上该数据库中发生的死锁数量。	CN和DN
blk_read_time	double precision	当前节点上该数据库后端读取数据文件块花费的时间，以毫秒计算。	DN

名称	类型	描述	求和范围
blk_write_time	double precision	当前节点上该数据库后端写入数据文件块花费的时间，以毫秒计算。	DN
stats_reset	timestamp with time zone	当前节点上该数据库统计重置的时间。	-

### 15.3.48 GLOBAL\_TABLE\_CHANGE\_STAT

GLOBAL\_TABLE\_CHANGE\_STAT视图显示当前数据库中所有表格（不包括外表）变更情况。表示次数的各字段为实例启动以来的累计值。

表 15-108 GLOBAL\_TABLE\_CHANGE\_STAT 字段

名称	类型	描述
schemaname	name	表的命名空间
relname	name	表的名称
last_vacuum	timestamp with time zone	最后一次手动Vacuum的时间。
vacuum_count	bigint	手动Vacuum的次数。为各CN节点上次数之和。
last_autovacuum	timestamp with time zone	最后一次自动Vacuum的时间。
autovacuum_count	bigint	自动Vacuum的次数。为各CN节点上次数之和。
last_analyze	timestamp with time zone	最后一次分析（包括手动和自动）的时间。
analyze_count	bigint	分析（包括手动和自动）的次数。由于analyze会同时所有节点上进行，该字段为所有CN节点上的最大值。
last_autoanalyze	timestamp with time zone	最后一次自动分析的时间。
autoanalyze_count	bigint	自动分析的次数。为各CN节点上次数之和。
last_change	bigint	最后一次修改（INSERT，UPDATE或DELETE）的时间。

### 15.3.49 GLOBAL\_TABLE\_STAT

GLOBAL\_TABLE\_STAT视图显示当前数据库中所有表格（不包括外表）的统计信息。除 live\_tuples和dead\_tuples为当前实时值外，其余各统计字段为实例启动以来的累计值。

表 15-109 GLOBAL\_TABLE\_STAT 字段

名称	类型	描述
schemaname	name	表的命名空间
relname	name	表的名称
distribute_mode	“char”	表的分布方式，与系统表pgxc_class中的plocator_type字段含义相同。
seq_scan	bigint	顺序扫描的次数。只统计行存表。如果是分区表，显示各个分区扫描次数的和。
seq_tuple_read	bigint	顺序扫描的行数。只统计行存表。
index_scan	bigint	索引扫描的次数。只统计行存表。
index_tuple_read	bigint	索引扫描的行数。只统计行存表。
tuple_inserted	bigint	插入的行数。如果是复制表，显示各节点最大值；如果是分布表，显示各节点之和。
tuple_updated	bigint	更新的行数。如果是复制表，显示各节点最大值；如果是分布表，显示各节点之和。
tuple_deleted	bigint	删除的行数。如果是复制表，显示各节点最大值；如果是分布表，显示各节点之和。
tuple_hot_updated	bigint	热更新的行数。如果是复制表，显示各节点最大值；如果是分布表，显示各节点之和。
live_tuples	bigint	活元组数量。显示各节点最大值；如果是分布表，显示各节点之和。 只适用行存表。
dead_tuples	bigint	死元组数量。显示各节点最大值；如果是分布表，显示各节点之和。 只适用行存表。

### 15.3.50 GLOBAL\_WORKLOAD\_SQL\_COUNT

GLOBAL\_WORKLOAD\_SQL\_COUNT视图显示集群中所有Workload控制组内SQL语句执行次数的统计信息，包括SELECT、UPDATE、INSERT、DELETE语句的执行次数统计，以及DDL、DML、DCL类型语句的执行次数统计。

表 15-110 GLOBAL\_WORKLOAD\_SQL\_COUNT 字段

名称	类型	描述
workload	name	Workload控制组名称
select_count	bigint	SELECT数量
update_count	bigint	UPDATE数量
insert_count	bigint	INSERT数量
delete_count	bigint	DELETE数量
ddl_count	bigint	DDL数量
dml_count	bigint	DML数量
dcl_count	bigint	DCL数量

### 15.3.51 GLOBAL\_WORKLOAD\_SQL\_ELAPSE\_TIME

GLOBAL\_WORKLOAD\_SQL\_ELAPSE\_TIME视图显示集群中所有Workload控制组内SQL语句执行的响应时间的统计信息，包括SELECT、UPDATE、INSERT、DELETE语句的最大、最小、平均、以及总响应时间，单位为微秒。

表 15-111 GLOBAL\_WORKLOAD\_SQL\_ELAPSE\_TIME 字段

名称	类型	描述
workload	name	Workload控制组名称
total_select_elapse	bigint	SELECT总响应时间
max_select_elapse	bigint	SELECT最大响应时间
min_select_elapse	bigint	SELECT最小响应时间
avg_select_elapse	bigint	SELECT平均响应时间
total_update_elapse	bigint	UPDATE总响应时间
max_update_elapse	bigint	UPDATE最大响应时间
min_update_elapse	bigint	UPDATE最小响应时间
avg_update_elapse	bigint	UPDATE平均响应时间
total_insert_elapse	bigint	INSERT总响应时间
max_insert_elapse	bigint	INSERT最大响应时间
min_insert_elapse	bigint	INSERT最小响应时间
avg_insert_elapse	bigint	INSERT平均响应时间
total_delete_elapse	bigint	DELETE总响应时间

名称	类型	描述
max_delete_elapse	bigint	DELETE最大响应时间
min_delete_elapse	bigint	DELETE最小响应时间
avg_delete_elapse	bigint	DELETE平均响应时间

### 15.3.52 GLOBAL\_WORKLOAD\_TRANSACTION

GLOBAL\_WORKLOAD\_TRANSACTION视图提供集群所有CN上WORKLOAD控制组相关的事务信息的总和。需要有系统管理员权限才可以访问。该视图仅在资源实时监控功能开启，即enable\_resource\_track为on时有效。

表 15-112 GLOBAL\_WORKLOAD\_TRANSACTION 字段

名称	类型	描述
workload	name	WORKLOAD控制组名称
commit_counter	bigint	各CN上提交次数总和
rollback_counter	bigint	各CN上回滚次数总和
resp_min	bigint	集群总体最小响应时间
resp_max	bigint	集群总体最大响应时间
resp_avg	bigint	各CN上平均响应时间
resp_total	bigint	各CN上响应时间总和

### 15.3.53 GS\_ALL\_CONTROL\_GROUP\_INFO

GS\_ALL\_CONTROL\_GROUP\_INFO视图显示数据库内所有的控制组信息。

表 15-113 GS\_ALL\_CONTROL\_GROUP\_INFO 字段

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

名称	类型	描述
limits	bigint	控制组分配的CPU资源限额
wdlevel	bigint	Workload控制组层级
cpucoros	text	控制组使用的CPU核的信息

### 15.3.54 GS\_BLOCKLIST\_QUERY

GS\_BLOCKLIST\_QUERY视图用于查询作业黑名单信息和异常信息，此视图是由系统表GS\_BLOCKLIST\_QUERY和GS\_WLM\_SESSION\_INFO关联所得，同时对查询结果进行了去重筛选，因此在GS\_WLM\_SESSION\_INFO表较大的情况下，查询可能需要消耗较长时间。

表 15-114 GS\_BLOCKLIST\_QUERY 视图字段

名字	类型	引用	描述
unique_sql_id	bigint	-	基于查询解析树生成的查询唯一ID。
block_list	boolean	-	查询作业是否属于黑名单。
except_num	integer	-	查询作业异常次数。
except_time	timestamp	-	查询作业最近一次异常时间。
query	text	-	执行的查询语句。

#### 说明

- 此视图仅限在gaussdb数据库中查询，其它数据库中查询会直接报错。
- 通常对于DML语句，在计算Unique SQL ID的过程中会忽略常量值。但对于DDL、DCL以及设置参数等语句，常量值不可以忽略。因此一个unique\_sql\_id可能会对应一个或多个查询。

### 15.3.55 GS\_CLUSTER\_RESOURCE\_INFO

GS\_CLUSTER\_RESOURCE\_INFO视图显示的是所有DN资源的汇总信息。

表 15-115 GS\_CLUSTER\_RESOURCE\_INFO 字段

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

名称	类型	描述
min_io_util	integer	DN最小IO使用率
max_io_util	integer	DN最大IO使用率
used_mem_rate	integer	物理节点最大内存使用率

### 15.3.56 GS\_COLUMN\_TABLE\_IO\_STAT

GS\_COLUMN\_TABLE\_IO\_STAT视图显示当前数据库中所有列存表在当前节点上的IO情况。各统计字段为实例启动以来的累计值。

表 15-116 GS\_COLUMN\_TABLE\_IO\_STAT 字段

名称	类型	描述
schemaname	name	表的命名空间
relname	name	表的名称
heap_read	bigint	堆逻辑读块数
heap_hit	bigint	堆命中块数
idx_read	bigint	索引逻辑读块数
idx_hit	bigint	索引命中块数
cu_read	bigint	Compression Unit逻辑读个数
cu_hit	bigint	Compression Unit命中个数
cidx_read	bigint	Compression Unit Index逻辑读个数
cidx_hit	bigint	Compression Unit Index命中个数

### 15.3.57 GS\_GET\_OBS\_READ\_TRAFFIC

GS\_GET\_OBS\_READ\_TRAFFIC视图，统计OBS读流量和平均读带宽，统计结果按10分钟聚集。该视图仅8.2.0及以上集群版本支持。

名称	类型	描述
nodename	TEXT	集群节点
hostname	TEXT	主机节点
traffic_mb	float8	logtime之前10分钟OBS读流量统计
bandwidth_m b_per_s	float8	平均带宽，单位MB/s

名称	类型	描述
reqcount	bigint	logtime之前10分钟OBS读次数
logtime	timestamp with time zone	记录统计信息时刻

## 应用示例

查询当前视图，统计OBS读流量和平均读带宽，统计结果按10分钟聚集。

```
select * from gs_obs_read_traffic;
nodename | hostname | traffic_mb | bandwidth_mb_per_s | reqcount | logtime
-----+-----+-----+-----+-----+-----
dn_1 | rhel_10_90_45_56 | 101.959338188171 | 5.14830159670447 | 23 | 2022-11-26 09:50:00+08
(1 row)
```

## 15.3.58 GS\_GET\_OBS\_WRITE\_TRAFFIC

GS\_GET\_OBS\_WRITE\_TRAFFIC视图，统计OBS写流量和平均写带宽，统计结果按10分钟聚集。该视图仅8.2.0及以上集群版本支持。

名称	类型	描述
nodename	TEXT	集群节点
hostname	TEXT	主机节点
traffic_mb	float8	logtime之前10分钟OBS写流量统计
bandwidth_m b_per_s	float8	平均带宽，单位MB/s
reqcount	bigint	logtime之前10分钟OBS写次数
logtime	timestamp with time zone	记录统计信息时刻

## 应用示例

查询当前视图，统计OBS写流量和平均写带宽，统计结果按10分钟聚集。

```
select * from gs_obs_write_traffic;
nodename | hostname | traffic_mb | bandwidth_mb_per_s | reqcount | logtime
-----+-----+-----+-----+-----+-----
dn_1 | rhel_10_90_45_56 | .000738143920898438 | .000289970820362525 | 12 | 2022-10-24 16:10:00+08
dn_1 | rhel_10_90_45_56 | .000354766845703125 | .000386063466694153 | 7 | 2022-10-24 18:50:00+08
dn_1 | rhel_10_90_45_56 | 9.34600830078125e-05 | .000143659648687162 | 2 | 2022-11-07 09:20:00+08
dn_1 | rhel_10_90_45_56 | 4.10079956054688e-05 | .000186667253592502 | 1 | 2022-11-07 09:30:00+08
dn_1 | rhel_10_90_45_56 | 2048.17834663391 | 27.2766632219637 | 2 | 2022-11-22 16:10:00+08
dn_1 | rhel_10_90_45_56 | 3747.23722648621 | 28.0842938534546 | 4 | 2022-11-22
```



```
16:20:00+08  
(6 row)
```

## 15.3.59 GS\_INSTR\_UNIQUE\_SQL

### Unique SQL 定义

数据库将接收到的每个SQL的文本字符串，都进行解析并生成内部解析树，遍历解析树并忽略其中的常数值，以一定的算法计算出来一个整数值作为Unique SQL ID，用来唯一标识这一类SQL，Unique SQL ID相同的一类SQL就叫做Unique SQL。

### 示例

假如，用户先后输入SQL：

```
select * from t1 where id = 1;  
select * from t1 where id = 2;
```

那么，这两条SQL的统计信息会汇聚到同一个Unique SQL上：

```
select * from t1 where id = ?;
```

### GS\_INSTR\_UNIQUE\_SQL 视图

GS\_INSTR\_UNIQUE\_SQL视图显示当前节点收集的Unique SQL的执行信息，主要包括以下内容：

- Unique SQL ID以及归一化后的SQL文本字符串，归一化后的SQL文本如**示例**中所示。通常对于DML语句，在计算Unique SQL ID的过程中会忽略常量值。但对于DDL、DCL以及设置参数等语句，常量值不可以忽略。
- 执行次数（成功执行的次数），响应时间（数据库内部的SQL执行时间，包括最大、最小和总时间）。
- Cache/IO信息，包含block的物理读、逻辑读次数，仅统计执行成功的SQL在各DN节点上的相关信息。该统计值与查询执行当时所处理的数据量、所使用的内存、是否多次执行、内存管理策略、是否有其他并发查询等因素相关，反映整个查询执行过程中的buffer块物理读和逻辑读次数，不同时间执行可能统计值不同。
- 行活动，包含SELECT语句的结果集返回行数、更新行、插入行、删除行、顺序扫描行、随机扫描行等信息。除结果集返回行数与该SELECT语句的结果集行数一致、且仅在CN上记录外，其他行活动信息均在DN上记录，且统计数值反应的是整个查询执行过程中的行活动，包括对相关系统表、元数据表、数据表等做必要的扫描和修改，与对应数据量以及相关参数设置相关，即统计数值将会大于等于对实际数据表的扫描和修改。
- 时间分布，包含：DB\_TIME/CPU\_TIME/EXECUTION\_TIME/PARSE\_TIME/PLAN\_TIME/REWRITE\_TIME/PL\_EXECUTION\_TIME/PL\_COMPILATION\_TIME/NET\_SEND\_TIME/DATA\_IO\_TIME，相关定义见**表1**。该信息在CN和DN节点均有统计，视图查询时将汇总展示。
- 软硬解析次数，包含软解析（缓存计划）、硬解析（生成计划）的次数，即如果本次执行的是之前缓存的计划，软解析次数+1，如果本次执行的计划是重新生成的，则硬解析次数+1。该次数在CN和DN节点上都会统计，视图查询时将汇总展示。

Unique SQL收集功能存在以下约束：

- 只有执行成功的SQL才会显示其详细的统计信息，否则可能只记录query、node、user等信息。

- 如果开启Unique SQL收集功能，CN节点将对所有接收到的查询进行统计收集，包括工具和用户的查询等。
- 若一条SQL语句内部包含执行多条SQL语句、类似存储过程执行等场景，仅会对最外层SQL生成一条Unique SQL，所有子SQL的统计信息都会汇总到该Unique SQL记录上。
- Unique SQL的响应时间统计中不完全包含NET\_SEND\_TIME阶段的时间，所以EXECUTION\_TIME和elapsed\_time等时间之间不存在大小比较关系。
- 对于类似begin;...;commit;等形式的事务块，当前不支持统计子句的解析时间（parse\_time）。

普通用户访问GS\_INSTR\_UNIQUE\_SQL视图，只能看到该用户相关的Unique SQL信息，管理员用户可以看到当前节点所有的Unique SQL信息。CN和DN上均可查询GS\_INSTR\_UNIQUE\_SQL视图，DN上显示的是本节点内的Unique SQL统计信息，CN上显示的是本节点Unique SQL完整统计信息，即该CN节点会收集其他CN和DN上对应该CN的Unique SQL的执行信息，进行汇总展示。通过查询GS\_INSTR\_UNIQUE\_SQL视图，能够定位由于消耗不同资源导致的Top SQL，为集群性能调优和维护提供依据。

GUC参数instr\_unique\_sql\_timeout设置了Unique SQL的超时时间，单位是小时。后台线程每隔1小时检查一次所有的Unique SQL，将last\_time在instr\_unique\_sql\_timeout小时之前的Unique SQL删除。

表 15-117 GS\_INSTR\_UNIQUE\_SQL 字段

名称	类型	描述
node_name	name	接收SQL的CN节点名称
node_id	integer	节点ID，等同于pgxc_node表中node_id
user_name	name	用户名称
user_id	oid	用户ID
unique_sql_id	bigint	归一化的UNIQUE SQL ID
query	text	归一化的SQL文本
n_calls	bigint	成功执行次数
min_elapse_time	bigint	SQL在数据库内的最小运行时间（单位：微秒）
max_elapse_time	bigint	SQL在数据库内的最大运行时间（单位：微秒）
total_elapse_time	bigint	SQL在数据库内的总运行时间（单位：微秒）
n_returned_rows	bigint	行活动-SELECT语句返回的结果集行数
n_tuples_fetched	bigint	行活动-随机扫描行（列存表/外表不统计）
n_tuples_returned	bigint	行活动-顺序扫描行（列存表/外表不统计）

名称	类型	描述
n_tuples_inserted	bigint	行活动-插入行数
n_tuples_updated	bigint	行活动-更新行数
n_tuples_deleted	bigint	行活动-删除行数
n_blocks_fetched	bigint	buffer的块访问次数，即物理读/IO
n_blocks_hit	bigint	buffer的块命中次数，即逻辑读/Cache
n_soft_parse	bigint	软解析次数（缓存计划）
n_hard_parse	bigint	硬解析次数（生成计划）
db_time	bigint	有效的DB执行时间，包含等待时间、网络发送时间等，若查询执行涉及到多线程，DB_TIME是多个线程的DB_TIME之和（单位：微秒）
cpu_time	bigint	CPU的执行时间，不包含sleep时间（单位：微秒）
execution_time	bigint	查询执行器内的SQL执行时间，DDL语句、以及某些不经过执行器执行的语句（例如Copy语句）不计数（单位：微秒）
parse_time	bigint	SQL解析时间（单位：微秒）
plan_time	bigint	SQL生成计划时间（单位：微秒）
rewrite_time	bigint	SQL重写时间（单位：微秒）
pl_execution_time	bigint	plpgsql过程化语言函数上的执行时间（单位：微秒）
pl_compilation_time	bigint	plpgsql过程化语言函数上的编译时间（单位：微秒）
net_send_time	bigint	网络时间，包含CN向客户端发送数据、DN向CN发送数据等时间（单位：微秒）
data_io_time	bigint	IO时间，文件IO耗时（单位：微秒）
first_time	timestamp with time zone	该SQL第一次执行的时间
last_time	timestamp with time zone	该SQL上一次执行的时间

### 15.3.60 GS\_NODE\_STAT\_RESET\_TIME

GS\_NODE\_STAT\_RESET\_TIME视图提供当前节点的统计信息重置时间，返回带时区的时间戳。

详细含义参考[get\\_node\\_stat\\_reset\\_time\(\)](#)函数。

### 15.3.61 GS\_OBS\_LATENCY

GS\_OBS\_LATENCY记录logtime之前10分钟内OBS的平均延迟信息，延迟数据是根据相关OBS的操作进行估算的结果。该视图仅8.2.0及以上集群版本支持。

表 15-118 GS\_OBS\_LATENCY 字段

名称	类型	描述
nodename	text	集群节点。
hostname	text	主机节点。
latency_ms	double precision	logtime之前10分钟内OBS的平均延迟，单位ms。
reqcount	bigint	logtime之前10分钟内OBS的请求次数。
logtime	timestamp with time zone	记录延迟信息的时刻。

### 15.3.62 GS\_REL\_IOSTAT

GS\_REL\_IOSTAT视图提供当前节点上磁盘读写的统计信息。当前版本中，每次读/写磁盘只读/写一页，所以读写次数与页数相等。

表 15-119 GS\_REL\_IOSTAT 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

### 15.3.63 GS\_RESPOOL\_RUNTIME\_INFO

GS\_RESPOOL\_RUNTIME\_INFO视图显示当前CN所有资源池作业运行信息。

表 15-120 GS\_RESPOOL\_RUNTIME\_INFO 字段

名称	类型	描述
nodegroup	name	资源池所属逻辑集群名称，默认集群显示 "installation"
rpname	name	资源池名称
ref_count	int	资源池引用作业数，作业经过资源池不管是否管控都会计数
fast_run	int	资源池快车道运行作业数
fast_wait	int	资源池快车道排队作业数
slow_run	int	资源池慢车道运行作业数
slow_wait	int	资源池慢车道排队作业数

### 15.3.64 GS\_RESPOOL\_RESOURCE\_INFO

GS\_RESPOOL\_RESOURCE\_INFO视图显示CN上所有资源池作业运行信息以及当前实例(CN/DN)所有资源池资源使用信息。

#### 说明

DN上仅显示当前DN所属逻辑集群的资源池监控信息。

表 15-121 GS\_RESPOOL\_RESOURCE\_INFO 字段

名称	类型	描述
nodegroup	name	资源池所属逻辑集群名称，默认集群显示 "installation"
rpname	name	资源池名称
cgroup	name	资源池关联控制组名称
ref_count	int	资源池引用作业数，作业经过资源池不管是否管控都会计数，只在CN上有效
fast_run	int	资源池快车道运行作业数，只在CN上有效
fast_wait	int	资源池快车道排队作业数，只在CN上有效
fast_limit	int	资源池快车道作业并发限制，只在CN上有效
slow_run	int	资源池慢车道运行作业数，只在CN上有效
slow_wait	int	资源池慢车道排队作业数，只在CN上有效
slow_limit	int	资源池慢车道作业并发限制，只在CN上有效

名称	类型	描述
used_cpu	double	资源池5s监控周期内使用CPU个数平均值，保留小数点后2位 <ul style="list-style-type: none"> <li>• DN：显示当前DN上资源池使用的CPU个数</li> <li>• CN：显示所有DN上资源池使用CPU的累积和</li> </ul>
cpu_limit	int	资源池可用CPU的上限，CPU配额管控情况下为GaussDB可用CPU，CPU限额管控情况下为关联控制组CPU可用CPU <ul style="list-style-type: none"> <li>• DN：显示当前DN上资源池可用CPU上限</li> <li>• CN：显示所有DN上资源池可用CPU上限的累积和</li> </ul>
used_mem	int	资源池当前使用的内存大小，单位：'MB' <ul style="list-style-type: none"> <li>• DN：显示当前DN上资源池使用的内存大小</li> <li>• CN：显示所有DN上资源池使用内存的累积和</li> </ul>
estimate_mem	int	当前CN上，资源池运行作业的估算内存之和，只在CN上有效
mem_limit	int	资源池可用内存上限，单位：'MB' <ul style="list-style-type: none"> <li>• DN：显示当前DN上资源池可用内存上限</li> <li>• CN：显示所有DN上资源池可用内存上限的累积和</li> </ul>
read_kbytes	bigint	资源池5s监控周期内逻辑读字节数，单位：'KB' <ul style="list-style-type: none"> <li>• DN：显示当前DN上资源池逻辑读字节数</li> <li>• CN：显示所有DN上资源池逻辑读字节的累积和</li> </ul>
write_kbytes	bigint	资源池5s监控周期内逻辑写字节数，单位：'KB' <ul style="list-style-type: none"> <li>• DN：显示当前DN上资源池逻辑写字节数</li> <li>• CN：显示所有DN上资源池逻辑写字节的累积和</li> </ul>
read_counts	bigint	资源池5s监控周期内逻辑读次数 <ul style="list-style-type: none"> <li>• DN：显示当前DN上资源池逻辑读次数</li> <li>• CN：显示所有DN上资源池逻辑读次数的累积和</li> </ul>
write_counts	bigint	资源池5s监控周期内逻辑写次数 <ul style="list-style-type: none"> <li>• DN：显示当前DN上资源池逻辑写次数</li> <li>• CN：显示所有DN上资源池逻辑写次数的累积和</li> </ul>

名称	类型	描述
read_speed	double	资源池5s监控周期内逻辑读速率的平均值 <ul style="list-style-type: none"> <li>• DN: 显示当前DN上资源池逻辑读速率</li> <li>• CN: 显示所有DN上资源池逻辑读速率的累积和</li> </ul>
write_speed	double	资源池5s监控周期内逻辑写速率平均值 <ul style="list-style-type: none"> <li>• DN: 显示当前DN上资源池逻辑写速率</li> <li>• CN: 显示所有DN上资源池逻辑写速率的累积和</li> </ul>

### 15.3.65 GS\_ROW\_TABLE\_IO\_STAT

GS\_ROW\_TABLE\_IO\_STAT视图显示当前数据库中所有行存表在当前节点上的IO情况。各统计字段为实例启动以来的累计值。

表 15-122 GS\_ROW\_TABLE\_IO\_STAT 字段

名称	类型	描述
schemaname	name	表的命名空间
relname	name	表的名称
heap_read	bigint	堆逻辑读块数
heap_hit	bigint	堆命中块数
idx_read	bigint	索引逻辑读块数
idx_hit	bigint	索引命中块数
toast_read	bigint	TOAST表逻辑读块数
toast_hit	bigint	TOAST表命中块数
tidx_read	bigint	TOAST表Index逻辑读个数
tidx_hit	bigint	TOAST表Index命中个数

### 15.3.66 GS\_SESSION\_CPU\_STATISTICS

GS\_SESSION\_CPU\_STATISTICS视图显示和当前用户执行复杂作业正在运行时的负载管理CPU使用的信息。

表 15-123 GS\_SESSION\_CPU\_STATISTICS 字段

名称	类型	描述
datid	oid	连接后端的数据库OID
username	name	登录到该后端的用户名
pid	bigint	后端线程ID
start_time	timestamp with time zone	语句执行的开始时间
min_cpu_time	bigint	语句在所有DN上的最小CPU时间，单位为ms
max_cpu_time	bigint	语句在所有DN上的最大CPU时间，单位为ms
total_cpu_time	bigint	语句在所有DN上的CPU总时间，单位为ms
query	text	正在执行的语句
node_group	text	语句所属用户对应的逻辑集群

### 15.3.67 GS\_SESSION\_MEMORY\_STATISTICS

GS\_SESSION\_MEMORY\_STATISTICS视图显示和当前用户执行复杂作业正在运行时的负载管理内存使用的信息。

表 15-124 GS\_SESSION\_MEMORY\_STATISTICS 字段

名称	类型	描述
datid	oid	连接后端的数据库OID
username	name	登录到该后端的用户名
pid	bigint	后端线程ID
start_time	timestamp with time zone	语句执行的开始时间
min_peak_memory	integer	语句在所有DN上的最小内存峰值大小，单位MB。
max_peak_memory	integer	语句在所有DN上的最大内存峰值大小，单位MB。



名称	类型	描述
spill_info	text	语句在所有DN上的下盘信息 None: 所有DN均未下盘 All: 所有DN均下盘 [a:b]: 数量为b个DN中有a个DN下盘
query	text	正在执行的语句
node_group	text	语句所属用户对应的逻辑集群

### 15.3.68 GS\_SQL\_COUNT

GS\_SQL\_COUNT视图显示数据库当前节点当前时刻执行的五类语句（SELECT、INSERT、UPDATE、DELETE、MERGE INTO）统计信息，包括执行次数和响应时间（除MERGE INTO语句外，统计其他四类语句的最大、最小、平均和总响应时间，单位为微秒），以及DDL、DML、DCL类型语句的执行次数。

GS\_SQL\_COUNT视图对DDL、DML、DCL类型语句分类与SQL语法中略有不同，具体如下：

- CREATE/ALTER/DROP USER，CREATE/ALTER/DROP ROLE等用户相关语句属于DCL类型。
- BEGIN/COMMIT/SET CONSTRAINTS/ROLLBACK/SAVEPOINT/START等事务相关语句属于DCL类型。
- ALTER SYSTEM KILL SESSION等价于SELECT pg\_terminate\_backend()语句，属于DML类型。

其余语句的分类与SQL语法中定义类似。

普通用户查询GS\_SQL\_COUNT视图仅能看到该用户当前节点的统计信息。管理员权限用户查询GS\_SQL\_COUNT视图则能看到所有用户当前节点的统计信息；当集群或该节点重启时，计数会清零，并重新开始计数。计数以节点收到的查询数为准，包括集群内部进行的查询；GS\_SQL\_COUNT视图涉及的统计信息只在CN上统计，且不统计从其他CN发送过来的SQL。在DN上查询该视图返回结果为空。

表 15-125 GS\_SQL\_COUNT 字段

名称	类型	描述
node_name	name	节点名称
user_name	name	用户名
select_count	bigint	SELECT数量
update_count	bigint	UPDATE数量
insert_count	bigint	INSERT数量
delete_count	bigint	DELETE数量
mergeinto_count	bigint	MERGE INTO数量

名称	类型	描述
ddl_count	bigint	DDL数量
dml_count	bigint	DML数量
dcl_count	bigint	DCL数量
total_select_elapse	bigint	SELECT总响应时间
avg_select_elapse	bigint	SELECT平均响应时间
max_select_elapse	bigint	SELECT最大响应时间
min_select_elapse	bigint	SELECT最小响应时间
total_update_elapse	bigint	UPDATE总响应时间
avg_update_elapse	bigint	UPDATE平均响应时间
max_update_elapse	bigint	UPDATE最大响应时间
min_update_elapse	bigint	UPDATE最小响应时间
total_delete_elapse	bigint	DELETE总响应时间
avg_delete_elapse	bigint	DELETE平均响应时间
max_delete_elapse	bigint	DELETE最大响应时间
min_delete_elapse	bigint	DELETE最小响应时间
total_insert_elapse	bigint	INSERT总响应时间
avg_insert_elapse	bigint	INSERT平均响应时间
max_insert_elapse	bigint	INSERT最大响应时间
min_insert_elapse	bigint	INSERT最小响应时间

### 15.3.69 GS\_STAT\_DB\_CU

GS\_STAT\_DB\_CU视图查询集群各个节点，每个数据库的CU命中情况。可以通过gs\_stat\_reset()进行清零。

表 15-126 GS\_STAT\_DB\_CU 字段

名称	类型	描述
node_name1	text	节点名称
db_name	text	数据库名称
mem_hit	bigint	内存命中次数
hdd_sync_read	bigint	硬盘同步读次数
hdd_asyn_read	bigint	硬盘异步读次数

### 15.3.70 GS\_STAT\_SESSION\_CU

GS\_STAT\_SESSION\_CU视图查询当前集群各个节点，当前运行session的CU命中情况。session退出相应的统计数据会清零。集群重启后，统计数据也会清零。

表 15-127 GS\_STAT\_SESSION\_CU 字段

名称	类型	描述
node_name1	text	节点名称
mem_hit	integer	内存命中次数
hdd_sync_read	integer	硬盘同步读次数
hdd_asyn_read	integer	硬盘异步读次数

### 15.3.71 GS\_TABLE\_CHANGE\_STAT

GS\_TABLE\_CHANGE\_STAT视图显示当前数据库中所有表格（不包括外表）在当前节点上的变更情况。表示次数的各字段为实例启动以来的累计值。

表 15-128 GS\_TABLE\_CHANGE\_STAT 字段

名称	类型	描述
schemaname	name	表的命名空间

名称	类型	描述
relname	name	表的名称
last_vacuum	timestamp with time zone	最后一次手动vacuum的时间。
vacuum_count	bigint	手动Vacuum的次数。
last_autovacuum	timestamp with time zone	最后一次自动vacuum的时间。
autovacuum_count	bigint	自动vacuum的次数。
last_analyze	timestamp with time zone	最后一次分析（包括手动和自动）的时间。
analyze_count	bigint	分析（包括手动和自动）的次数。
last_autoanalyze	timestamp with time zone	最后一次自动分析的时间。
autoanalyze_count	bigint	自动分析的次数。
last_change	bigint	最后一次修改（INSERT，UPDATE或DELETE）的时间。

### 15.3.72 GS\_TABLE\_STAT

GS\_TABLE\_STAT视图显示当前数据库中所有表格（不包括外表）在当前节点上的统计信息。除live\_tuples和dead\_tuples为当前实时值外，其余各统计字段为实例启动以来的累计值。

表 15-129 GS\_TABLE\_STAT 字段

名称	类型	描述
schemaname	name	表的命名空间
relname	name	表的名称
seq_scan	bigint	顺序扫描的次数。只统计行存表。如果是分区表，显示各个分区扫描次数的和。
seq_tuple_read	bigint	顺序扫描的行数。只统计行存表。
index_scan	bigint	索引扫描的次数。只统计行存表。
index_tuple_read	bigint	索引扫描的行数。只统计行存表。

名称	类型	描述
tuple_inserted	bigint	插入的行数。
tuple_updated	bigint	更新的行数。
tuple_deleted	bigint	删除的行数。
tuple_hot_updated	bigint	热更新的行数。
live_tuples	bigint	活元组数量。CN上查询该视图，analyze后显示该表格总的活元组数量，未analyze的情况下显示0。只适用行存表。
dead_tuples	bigint	死元组数量。CN上查询该视图，analyze后显示该表格总的死元组数量，未analyze的情况下显示0。只适用行存表。

### 15.3.73 GS\_TOTAL\_NODEGROUP\_MEMORY\_DETAIL

GS\_TOTAL\_NODEGROUP\_MEMORY\_DETAIL视图统计当前数据库逻辑集群使用内存的信息，单位为MB。

表 15-130 GS\_TOTAL\_NODEGROUP\_MEMORY\_DETAIL 字段

名称	类型	描述
ngname	text	逻辑集群名称
memorytype	text	内存类型，包括以下几种： <ul style="list-style-type: none"> <li>ng_total_memory：该逻辑集群的总内存大小</li> <li>ng_used_memory：该逻辑集群的实际使用内存大小</li> <li>ng_estimate_memory：该逻辑集群的估算使用内存大小</li> <li>ng_foreignrp_memsize：该逻辑集群的外部资源池的总内存大小</li> <li>ng_foreignrp_usedsize：该逻辑集群的外部资源池实际使用内存大小</li> <li>ng_foreignrp_peaksize：该逻辑集群的外部资源池使用内存的峰值</li> <li>ng_foreignrp_mempct：该逻辑集群的外部资源池占该逻辑集群总内存大小的百分比</li> <li>ng_foreignrp_estmsize：该逻辑集群的外部资源池估算使用内存大小</li> </ul>
memorybytes	integer	内存类型分配内存的大小

### 15.3.74 GS\_USER\_TRANSACTION

GS\_USER\_TRANSACTION视图提供查询单CN上用户相关的事务信息。数据库记录每个用户事务提交和回滚的次数及事务提交和回滚的响应时间，单位是微秒。

表 15-131 GS\_USER\_TRANSACTION 字段

名称	类型	描述
username	name	用户名称
commit_counter	bigint	提交次数
rollback_counter	bigint	回滚次数
resp_min	bigint	最小响应时间
resp_max	bigint	最大响应时间
resp_avg	bigint	平均响应时间
resp_total	bigint	响应时间总和

### 15.3.75 GS\_VIEW\_DEPENDENCY

GS\_VIEW\_DEPENDENCY视图提供查询当前用户可见的所有视图的直接依赖关系。

表 15-132 GS\_VIEW\_DEPENDENCY 字段

名称	类型	描述
objschema	name	视图空间名称
objname	name	视图名称
refobjschema	name	依赖对象的空间名称
refobjname	name	依赖对象的名称
relobjkind	char	依赖对象的类型 <ul style="list-style-type: none"> <li>• r表示依赖对象为表</li> <li>• v表示依赖对象为视图</li> </ul>

### 15.3.76 GS\_VIEW\_DEPENDENCY\_PATH

GS\_VIEW\_DEPENDENCY\_PATH视图提供查询当前用户可见的所有视图的直接依赖关系。如果该视图依赖的基础表存在且各级视图依赖关系正常，通过该视图可以查询自基础表开始的各级视图的依赖关系。

表 15-133 GS\_VIEW\_DEPENDENCY\_PATH 字段

名称	类型	描述
objschema	name	视图空间名称
objname	name	视图名称
refobjschema	name	依赖对象的空间名称
refobjname	name	依赖对象的名称
path	text	依赖路径

### 15.3.77 GS\_VIEW\_INVALID

GS\_VIEW\_INVALID视图提供查询当前用户可见的所有不可用的视图。如果该视图依赖的基础表或函数或同义词存在异常，该视图validtype列显示为“invalid”。

表 15-134 GS\_VIEW\_INVALID 字段

名称	类型	描述
oid	oid	视图OID
schemaname	name	视图空间名称
viewname	name	视图名称
viewowner	name	视图的所有者
definition	text	视图定义
validtype	text	视图有效性标识

### 15.3.78 GS\_WAIT\_EVENTS

GS\_WAIT\_EVENTS视图显示当前节点上各类等待状态和事件的统计信息。

仅在GUC参数enable\_track\_wait\_event为on的情况下，视图中各统计字段的数值才会被累加。若在运行过程中将enable\_track\_wait\_event设置为off，则不再累加统计数值，但已有数值不受影响。enable\_track\_wait\_event为off，查询该视图返回0行。

表 15-135 GS\_WAIT\_EVENTS 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。

名称	类型	描述
event	text	事件名称，可参考 <a href="#">PG_THREAD_WAIT_STATUS</a> 视图。
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。

当前版本中，对于type='LOCK\_EVENT'，'LWLOCK\_EVENT'和'IO\_EVENT'的事件，GS\_WAIT\_EVENTS视图显示范围与[PG\\_THREAD\\_WAIT\\_STATUS](#)视图对应事件相同。

对于type='STATUS'的事件GS\_WAIT\_EVENTS包含的等待状态列如下，其详细含义参见[PG\\_THREAD\\_WAIT\\_STATUS](#)视图。

- acquire lwlock
- acquire lock
- wait io
- wait pooler get conn
- wait pooler abort conn
- wait pooler clean conn
- wait transaction sync
- wait wal sync
- wait data sync
- wait producer ready
- create index
- analyze
- vacuum
- vacuum full
- gtm connect
- gtm begin trans
- gtm commit trans
- gtm rollback trans
- gtm create sequence
- gtm alter sequence



- gtm get sequence val
- gtm set sequence val
- gtm drop sequence
- gtm rename sequence

### 15.3.79 GS\_WLM\_OPERATOR\_INFO

本视图显示当前CN上已经完成执行的query语句中的算子执行信息，此系统视图信息来源于系统表dbms\_om.gs\_wlm\_operator\_info。

### 15.3.80 GS\_WLM\_OPERATOR\_HISTORY

GS\_WLM\_OPERATOR\_HISTORY视图显示的是当前用户在当前CN上执行作业结束后的算子的相关记录。

此视图用于Database Manager从内核中查询数据，内核中的数据会定时被清理。当GUC参数enable\_resource\_record为on时，视图中的记录每隔3分钟被转储到系统表GS\_WLM\_OPERATOR\_INFO中一次，同时视图中的记录被删除；当GUC参数enable\_resource\_record为off时，记录在视图中的存留时间达到超期时间后会被删除。记录的数据同表15-5。

### 15.3.81 GS\_WLM\_OPERATOR\_STATISTICS

GS\_WLM\_OPERATOR\_STATISTICS视图显示当前用户正在执行的作业的算子相关信息。

表 15-136 GS\_WLM\_OPERATOR\_STATISTICS 的字段

名称	类型	描述
queryid	bigint	语句执行使用的内部query_id。
pid	bigint	后端线程ID。
plan_node_id	integer	查询对应的执行计划的plan node id。
plan_node_name	text	对应于plan_node_id的算子的名称。
start_time	timestamp with time zone	该算子处理第一条数据的开始时间。
duration	bigint	该算子直到结束时总的执行时间(ms)。
status	text	当前算子的执行状态，包括finished和running。
query_dop	integer	当前算子执行时的并行度。
estimated_rows	bigint	优化器估算的行数信息。
tuple_processed	bigint	当前算子返回的元素个数。

名称	类型	描述
min_peak_memory	integer	当前算子在所有DN上的最小内存峰值(MB)。
max_peak_memory	integer	当前算子在所有DN上的最大内存峰值(MB)。
average_peak_memory	integer	当前算子在所有DN上的平均内存峰值(MB)。
memory_skew_percent	integer	当前算子在各DN间的内存使用倾斜率。
min_spill_size	integer	若发生下盘, 所有下盘DN的最小下盘数据量(MB), 默认为0。
max_spill_size	integer	若发生下盘, 所有下盘DN的最大下盘数据量(MB), 默认为0。
average_spill_size	integer	若发生下盘, 所有下盘DN的平均下盘数据量(MB), 默认为0。
spill_skew_percent	integer	若发生下盘, DN间下盘倾斜率。
min_cpu_time	bigint	该算子在所有DN上的最小执行时间(ms)。
max_cpu_time	bigint	该算子在所有DN上的最大执行时间(ms)。
total_cpu_time	bigint	该算子在所有DN上的总执行时间(ms)。
cpu_skew_percent	integer	DN间执行时间的倾斜率。
warning	text	主要显示如下几类告警信息: 1. Sort/SetOp/HashAgg/HashJoin spill 2. Spill file size large than 256MB 3. Broadcast size large than 100MB 4. Early spill 5. Spill times is greater than 3 6. Spill on memory adaptive 7. Hash table conflict

### 15.3.82 GS\_WLM\_SESSION\_INFO

本视图显示当前CN上已经完成执行的query语句的执行信息, 此系统视图信息来源于系统表dbms\_om.[gs\\_wlm\\_session\\_info](#)。

### 15.3.83 GS\_WLM\_SESSION\_HISTORY

GS\_WLM\_SESSION\_HISTORY视图显示当前用户在当前CN上执行作业结束后的负载管理记录。此视图用于Database Manager从GaussDB(DWS)中查询数据, 仅当GUC参

数enable\_resource\_track为on时，视图会查询GS\_WLM\_SESSION\_INFO表中3分钟内的数据进行返回。

表 15-137 GS\_WLM\_SESSION\_HISTORY 的字段

名称	类型	描述
datid	oid	连接后端的数据库OID。
dbname	text	连接后端的数据库名称。
schemaname	text	模式名。
nodename	text	语句执行的CN名称。
username	text	连接到后端的用户名。
application_name	text	连接到后端的应用名。
client_addr	inet	连接到后端的客户端的IP地址。如果此字段是null，它表明通过服务器机器上UNIX套接字连接客户端或者这是内部进程，如autovacuum。
client_hostname	text	客户端的主机名，这个字段是通过client_addr的反向DNS查找得到。这个字段只有在启动log_hostname且使用IP连接时才非空。
client_port	integer	客户端用于与后端通讯的TCP端口号，如果使用Unix套接字，则为-1。
query_band	text	用于标示作业类型，可通过GUC参数query_band进行设置，默认为空字符串。
block_time	bigint	语句执行前的阻塞时间，包含语句解析和优化时间，单位ms。
start_time	timestamp with time zone	语句执行的开始时间。
finish_time	timestamp with time zone	语句执行的结束时间。
duration	bigint	语句实际执行的时间，单位ms。
estimate_total_time	bigint	语句预估执行时间，单位ms。
status	text	语句执行结束状态：正常为finished，异常为aborted。该处记录的语句状态应为数据库服务端执行状态，当服务器端执行成功，结果集返回时报错，该语句应为finished。
abort_info	text	语句执行结束状态为aborted时显示异常信息。
resource_pool	text	用户使用的资源池。

名称	类型	描述
control_group	text	语句所使用的Cgroup。
estimate_memory	integer	语句在单个实例上预估使用的内存，单位MB。该字段只有当GUC参数 <a href="#">enable_dynamic_workload</a> 为on时才有效。
min_peak_memory	integer	语句在所有DN上的最小内存峰值，单位MB。
max_peak_memory	integer	语句在所有DN上的最大内存峰值，单位MB。
average_peak_memory	integer	语句执行过程中的内存使用平均值，单位MB。
memory_skew_percent	integer	语句各DN间的内存使用倾斜率。
spill_info	text	语句在所有DN上的下盘信息： None: 所有DN均未下盘。 All: 所有DN均下盘。 [a:b]: 数量为b个DN中有a个DN下盘。
min_spill_size	integer	若发生下盘，所有下盘DN的最小下盘数据量(MB)，默认为0。
max_spill_size	integer	若发生下盘，所有下盘DN的最大下盘数据量(MB)，默认为0。
average_spill_size	integer	若发生下盘，所有下盘DN的平均下盘数据量(MB)，默认为0。
spill_skew_percent	integer	若发生下盘，DN间下盘倾斜率。
min_dn_time	bigint	语句在所有DN上的最小执行时间，单位ms。
max_dn_time	bigint	语句在所有DN上的最大执行时间，单位ms。
average_dn_time	bigint	语句在所有DN上的平均执行时间，单位ms。
dn_time_skew_percent	integer	语句在各DN间的执行时间倾斜率。
min_cpu_time	bigint	语句在所有DN上的最小CPU时间，单位ms。
max_cpu_time	bigint	语句在所有DN上的最大CPU时间，单位ms。
total_cpu_time	bigint	语句在所有DN上的CPU总时间，单位ms。
cpu_skew_percent	integer	语句在DN间的CPU时间倾斜率。

名称	类型	描述
min_peak_iops	integer	语句在所有DN上的每秒最小IO峰值（列存单位是次/s，行存单位是万次/s）。
max_peak_iops	integer	语句在所有DN上的每秒最大IO峰值（列存单位是次/s，行存单位是万次/s）。
average_peak_iops	integer	语句在所有DN上的每秒平均IO峰值（列存单位是次/s，行存单位是万次/s）。
iops_skew_percent	integer	语句在DN间的IO倾斜率。
warning	text	主要显示如下几类告警信息以及SQL自诊断调优相关告警： 1. Spill file size large than 256MB 2. Broadcast size large than 100MB 3. Early spill 4. Spill times is greater than 3 5. Spill on memory adaptive 6. Hash table conflict
queryid	bigint	语句执行使用的内部query id。
query	text	执行的语句。
query_plan	text	语句的执行计划。
node_group	text	语句所属用户对应的逻辑集群。
pid	bigint	语句的后端线程的pid。
lane	text	语句执行时所在的快慢车道。
unique_sql_id	bigint	归一化的Unique SQL ID。
session_id	text	在数据库系统中唯一标记一个session，格式：session_start_time.tid.node_name。
min_read_bytes	bigint	语句在所有DN上的最小IO读字节数，单位Bytes。
max_read_bytes	bigint	语句在所有DN上的最大IO读字节数，单位Bytes。
average_read_bytes	bigint	语句在所有DN上的平均IO读字节数，单位Bytes。
min_write_bytes	bigint	语句在所有DN上的最小IO写字节数，单位Bytes。
max_write_bytes	bigint	语句在所有DN上的最大IO写字节数，单位Bytes。

名称	类型	描述
average_write_bytes	bigint	语句在所有DN上的平均IO写字节数，单位Bytes。
recv_pkg	bigint	语句对应的通信包接收总量，单位packages。
send_pkg	bigint	语句对应的通信包发送总量，单位packages。
recv_bytes	bigint	语句对应的通信流接收数据总量，单位Byte。
send_bytes	bigint	语句对应的通信流发送数据总量，单位Byte。

### 15.3.84 GS\_WLM\_SESSION\_STATISTICS

GS\_WLM\_SESSION\_STATISTICS视图显示当前用户在当前CN上正在执行的作业的负载管理记录。

表 15-138 GS\_WLM\_SESSION\_STATISTICS 的字段

名称	类型	描述
datid	oid	连接后端的数据OID。
dbname	name	连接后端的数据库名称。
schemaname	text	模式名。
nodename	text	语句执行的CN节点名称。
username	name	连接到后端的用户名。
application_name	text	连接到后端的应用名。
client_addr	inet	连接到后端的客户端的IP地址。如果此字段是null，它表明通过服务器机器上UNIX套接字连接客户端或者这是内部进程，如autovacuum。
client_hostname	text	客户端的主机名，这个字段是通过client_addr的反向DNS查找得到。这个字段只有在启动log_hostname且使用IP连接时才非空。
client_port	integer	客户端用于与后端通讯的TCP端口号，如果使用Unix套接字，则为-1。
query_band	text	用于标示作业类型，可通过GUC参数query_band进行设置，默认为空字符串。
pid	bigint	后端线程ID。
block_time	bigint	语句执行前的阻塞时间，单位ms。

名称	类型	描述
start_time	timestamp with time zone	语句执行的开始时间。
duration	bigint	语句已经执行的时间，单位ms。
estimate_total_time	bigint	语句执行预估总时间，单位ms。
estimate_left_time	bigint	语句执行预估剩余时间，单位ms。
enqueue	text	工作负载管理资源状态。
resource_pool	name	用户使用的资源池。
control_group	text	语句所使用的Cgroup。
estimate_memory	integer	语句在单个实例上预估使用的内存，单位MB。该字段只有当GUC参数 <a href="#">enable_dynamic_workload</a> 为on时才有效。
min_peak_memory	integer	语句在所有DN上的最小内存峰值，单位MB。
max_peak_memory	integer	语句在所有DN上的最大内存峰值，单位MB。
average_peak_memory	integer	语句执行过程中的内存使用平均值，单位MB。
memory_skew_percent	integer	语句在各DN间的内存使用倾斜率。
spill_info	text	语句在所有DN上的下盘信息： None: 所有DN均未下盘。 All: 所有DN均下盘。 [a:b]: 数量为b个DN中有a个DN下盘。
min_spill_size	integer	若发生下盘，所有下盘DN的最小下盘数据量(MB)，默认为0。
max_spill_size	integer	若发生下盘，所有下盘DN的最大下盘数据量(MB)，默认为0。
average_spill_size	integer	若发生下盘，所有下盘DN的平均下盘数据量(MB)，默认为0。
spill_skew_percent	integer	若发生下盘，DN间下盘倾斜率。
min_dn_time	bigint	语句在所有DN上的最小执行时间，单位ms。
max_dn_time	bigint	语句在所有DN上的最大执行时间，单位ms。

名称	类型	描述
average_dn_time	bigint	语句在所有DN上的平均执行时间，单位ms。
dntime_skew_percent	integer	语句在各DN间的执行时间倾斜率。
min_cpu_time	bigint	语句在所有DN上的最小CPU时间，单位ms。
max_cpu_time	bigint	语句在所有DN上的最大CPU时间，单位ms。
total_cpu_time	bigint	语句在所有DN上的CPU总时间，单位ms。
cpu_skew_percent	integer	语句在各DN间的CPU时间倾斜率。
min_peak_iops	integer	语句在所有DN上的每秒最小IO峰值（列存单位是次/s，行存单位是万次/s）。
max_peak_iops	integer	语句在所有DN上的每秒最大IO峰值（列存单位是次/s，行存单位是万次/s）。
average_peak_iops	integer	语句在所有DN上的每秒平均IO峰值（列存单位是次/s，行存单位是万次/s）。
iops_skew_percent	integer	语句在DN间的IO倾斜率。
min_read_speed	integer	一个监控周期（5s）内，语句在所有DN上的最小IO读速率，单位KB/s。
max_read_speed	integer	一个监控周期（5s）内，语句在所有DN上的最大IO读速率，单位KB/s。
average_read_speed	integer	一个监控周期（5s）内，语句在所有DN上的平均IO读速率，单位KB/s。
min_write_speed	integer	一个监控周期（5s）内，语句在所有DN上的最小IO写速率，单位KB/s。
max_write_speed	integer	一个监控周期（5s）内，语句在所有DN上的最大IO写速率，单位KB/s。
average_write_speed	integer	一个监控周期（5s）内，语句在所有DN上的平均IO写速率，单位KB/s。
recv_pkg	bigint	语句对应的通信包接收总量，单位packages。
send_pkg	bigint	语句对应的通信包发送总量，单位packages。
recv_bytes	bigint	语句对应的通信流接收数据总量，单位Byte。
send_bytes	bigint	语句对应的通信流发送数据总量，单位Byte。



名称	类型	描述
warning	text	主要显示如下几类告警信息以及SQL自诊断调优相关告警： 1. Spill file size large than 256MB 2. Broadcast size large than 100MB 3. Early spill 4. Spill times is greater than 3 5. Spill on memory adaptive 6. Hash table conflict
queryid	bigint	语句执行使用的内部query id。
query	text	正在执行的语句。
query_plan	text	语句的执行计划。
node_group	text	语句所属用户对应的逻辑集群。

### 15.3.85 GS\_WLM\_SQL\_ALLOW

GS\_WLM\_SQL\_ALLOW视图显示已经设置的资源管理SQL白名单，包括两部分内容：系统默认的SQL白名单和通过GUC参数[wlm\\_sql\\_allow\\_list](#)设置的SQL白名单。

### 15.3.86 GS\_WORKLOAD\_SQL\_COUNT

GS\_WORKLOAD\_SQL\_COUNT视图显示当前节点上Workload控制组内的SQL语句执行次数的统计信息，包括SELECT、UPDATE、INSERT、DELETE语句的执行次数统计，以及DDL、DML、DCL类型语句的执行次数统计。

表 15-139 GS\_WORKLOAD\_SQL\_COUNT 字段

名称	类型	描述
workload	name	Workload控制组名称
select_count	bigint	SELECT数量
update_count	bigint	UPDATE数量
insert_count	bigint	INSERT数量
delete_count	bigint	DELETE数量
ddl_count	bigint	DDL数量
dml_count	bigint	DML数量
dcl_count	bigint	DCL数量

### 15.3.87 GS\_WORKLOAD\_SQL\_ELAPSE\_TIME

GS\_WORKLOAD\_SQL\_ELAPSE\_TIME视图显示当前节点上Workload控制组内SQL语句执行的响应时间的统计信息，包括SELECT、UPDATE、INSERT、DELETE语句的最大、最小、平均、以及总响应时间，单位为微秒。

表 15-140 GS\_WORKLOAD\_SQL\_ELAPSE\_TIME 字段

名称	类型	描述
workload	name	Workload控制组名称
total_select_elapse	bigint	SELECT总响应时间
max_select_elapse	bigint	SELECT最大响应时间
min_select_elapse	bigint	SELECT最小响应时间
avg_select_elapse	bigint	SELECT平均响应时间
total_update_elapse	bigint	UPDATE总响应时间
max_update_elapse	bigint	UPDATE最大响应时间
min_update_elapse	bigint	UPDATE最小响应时间
avg_update_elapse	bigint	UPDATE平均响应时间
total_insert_elapse	bigint	INSERT总响应时间
max_insert_elapse	bigint	INSERT最大响应时间
min_insert_elapse	bigint	INSERT最小响应时间
avg_insert_elapse	bigint	INSERT平均响应时间
total_delete_elapse	bigint	DELETE总响应时间
max_delete_elapse	bigint	DELETE最大响应时间
min_delete_elapse	bigint	DELETE最小响应时间
avg_delete_elapse	bigint	DELETE平均响应时间

### 15.3.88 GS\_WORKLOAD\_TRANSACTION

GS\_WORKLOAD\_TRANSACTION视图提供查询单CN上Workload控制组相关的事务信息。数据库记录每个Workload控制组事务提交和回滚的次数及事务提交和回滚的响应时间，单位是微秒。

表 15-141 GS\_WORKLOAD\_TRANSACTION 字段

名称	类型	描述
workload	name	Workload控制组名称

名称	类型	描述
commit_counter	bigint	提交次数
rollback_counter	bigint	回滚次数
resp_min	bigint	最小响应时间
resp_max	bigint	最大响应时间
resp_avg	bigint	平均响应时间
resp_total	bigint	响应时间总和

### 15.3.89 MPP\_TABLES

MPP\_TABLES视图显示PGXC\_CLASS中的表信息。

表 15-142 MPP\_TABLES 字段

名称	类型	描述
schemaname	name	包含表的模式名。
tablename	name	表名。
tableowner	name	表的所有者。
tablespace	name	表所在的表空间。
pgroup	name	节点群的名称。
nodeoids	oidvector_extend	表分布的节点OID列表。

### 15.3.90 PG\_AVAILABLE\_EXTENSION\_VERSIONS

PG\_AVAILABLE\_EXTENSION\_VERSIONS视图显示数据库中某些特性的扩展版本信息。

表 15-143 PG\_AVAILABLE\_EXTENSION\_VERSIONS 字段

名称	类型	描述
name	name	扩展名
version	text	版本名
installed	boolean	如果此扩展的版本当前已经安装，则为真
superuser	boolean	如果只允许系统管理员安装此扩展，则为真
relocatable	boolean	如果扩展可以重新加载到另一个模式，则为真

名称	类型	描述
schema	name	扩展必须安装到的模式名，如果部分或全部可重新定位，则为NULL
requires	name[]	必备扩展的名称，如果没有则为NULL
comment	text	扩展的控制文件的注释字符串

### 15.3.91 PG\_AVAILABLE\_EXTENSIONS

PG\_AVAILABLE\_EXTENSIONS视图显示数据库中某些特性的扩展信息。

表 15-144 PG\_AVAILABLE\_EXTENSIONS 字段

名称	类型	描述
name	name	扩展名
default_version	text	缺省版本名，如果没有指定则为NULL
installed_version	text	当前安装的扩展版本，如果未安装则为NULL
comment	text	扩展控制文件的注释字符串

### 15.3.92 PG\_BULKLOAD\_STATISTICS

在集群任一正常节点上，通过查询PG\_BULKLOAD\_STATISTICS视图可以获取当前登录节点正在进行的导入导出业务执行情况，其中每一个导入/导出业务对应一条记录。需要有系统管理员权限才可以访问此视图

表 15-145 PG\_BULKLOAD\_STATISTICS 字段

名称	类型	描述
node_name	text	节点名称。
db_name	text	数据库名称。
query_id	bigint	查询ID，对应debug_query_id。
tid	bigint	当前线程的线程号。
lwtid	integer	当前线程的轻量级线程号。
session_id	bigint	GDS的会话ID。
direction	text	业务类型，取值包括：gds to file、gds from file、gds to pipe、gds from pipe、copy from、copy to。
query	text	查询语句。

address	text	当前导入导出外表的location。
query_start	timestamp with time zone	导入/导出开始时间。
total_bytes	bigint	待处理数据的总大小。 仅GDS普通文件导入时，且该行记录来自CN节点才会显示，否则为空。
phase	text	当前业务导入导出执行阶段，包括：INITIALIZING、TRANSFER_DATA、RELEASE_RESOURCE。
done_lines	bigint	已传输行数。
done_bytes	bigint	已传输字节数。

### 15.3.93 PG\_COMM\_CLIENT\_INFO

PG\_COMM\_CLIENT\_INFO视图存储单个节点客户端连接信息（DN上查询该视图显示CN连接DN的信息）。

表 15-146 PG\_COMM\_CLIENT\_INFO 字段

名称	类型	描述
node_name	text	当前节点的名称。
app	text	客户端应用名。
tid	bigint	当前线程的线程号。
lwtid	integer	当前线程的轻量级线程号。
query_id	bigint	查询ID，对应debug_query_id。
socket	integer	如果是物理连接，显示socket。
remote_ip	text	对端节点IP。
remote_port	text	对端节点port。
logic_id	integer	如果是逻辑连接，显示sid，显示-1时表示当前连接是物理连接。

### 15.3.94 PG\_COMM\_DELAY

PG\_COMM\_DELAY视图展示单个DN的通信库时延状态。

表 15-147 PG\_COMM\_DELAY 字段

名称	类型	描述
node_name	text	节点名称。
remote_name	text	连接对端时延最大的节点名称。
remote_host	text	连接对端IP地址。
stream_num	integer	当前物理连接使用的stream逻辑连接数量。
min_delay	integer	当前物理连接探测到的最小时延，单位微秒。
average	integer	当前物理连接探测时延的平均值，单位微秒。
max_delay	integer	当前物理连接探测到的最大时延，单位微秒。 <b>说明</b> 取值为-1，表示时延探测超时失败，请重新建立节点间连接后再执行查询。

### 15.3.95 PG\_COMM\_STATUS

PG\_COMM\_STATUS视图展示单个DN的通信库状态。

表 15-148 PG\_COMM\_STATUS 字段

名称	类型	描述
node_name	text	节点名称。
rxpck/s	integer	节点通信库接收速率，单位Byte/s。
txpck/s	integer	节点通信库发送速率，单位Byte/s。
rxkB/s	bigint	节点通信库接收速率，单位KByte/s。
txkB/s	bigint	节点通信库发送速率，单位KByte/s。
buffer	bigint	cmailbox的buffer大小。
memKB(libcomm)	bigint	libcomm进程通信内存大小，单位KByte。
memKB(libpq)	bigint	libpq进程通信内存大小，单位KByte。
%USED(PM)	integer	postmaster线程实时使用率。
%USED (sflow)	integer	gs_sender_flow_controller线程实时使用率。
%USED (rflow)	integer	gs_receiver_flow_controller线程实时使用率。
%USED (rloop)	integer	多个gs_receivers_loop线程中最高的实时使用率。
stream	integer	当前使用的逻辑连接总数。

## 15.3.96 PG\_COMM\_RECV\_STREAM

PG\_COMM\_RECV\_STREAM视图展示单个DN上所有的通信库接收流状态。

表 15-149 PG\_COMM\_RECV\_STREAM 字段

名称	类型	描述
node_name	text	节点名称。
local_tid	bigint	使用此通信流的线程ID。
remote_name	text	连接对端节点名称。
remote_tid	bigint	连接对端线程ID。
idx	integer	通信对端DN在本DN内的标识编号。
sid	integer	通信流在物理连接中的标识编号。
tcp_sock	integer	通信流所使用的tcp通信socket。
state	text	通信流当前的状态： <ul style="list-style-type: none"> <li>• UNKNOWN：表示当前逻辑连接状态未知。</li> <li>• READY：表示逻辑连接已就绪。</li> <li>• RUN：表示逻辑连接接收报文正常。</li> <li>• HOLD：表示逻辑连接接收报文等待中。</li> <li>• CLOSED：表示关闭逻辑连接。</li> <li>• TO_CLOSED：表示将会关闭逻辑连接。</li> <li>• WRITING：表示正在写入数据。</li> </ul>
query_id	bigint	通信流对应的debug_query_id编号。
pn_id	integer	通信流所执行查询的plan_node_id编号。
send_smp	integer	通信流所执行查询send端的smpid编号。
recv_smp	integer	通信流所执行查询recv端的smpid编号。
recv_bytes	bigint	通信流接收的数据总量，单位Byte。
time	bigint	通信流当前生命周期使用时长，单位ms。
speed	bigint	通信流的平均接收速率，单位Byte/s。
quota	bigint	通信流当前的通信配额值，单位Byte。
buff_usize	bigint	通信流当前缓存的数据大小，单位Byte。

## 15.3.97 PG\_COMM\_SEND\_STREAM

PG\_COMM\_SEND\_STREAM视图展示单个DN上所有的通信库发送流状态。

表 15-150 PG\_COMM\_SEND\_STREAM 字段

名称	类型	描述
node_name	text	节点名称。
local_tid	bigint	使用此通信流的线程ID。
remote_name	text	连接对端节点名称。
remote_tid	bigint	连接对端线程ID。
idx	integer	通信对端DN在本DN内的标识编号。
sid	integer	通信流在物理连接中的标识编号。
tcp_sock	integer	通信流所使用的tcp通信socket。
state	text	通信流当前的状态。 <ul style="list-style-type: none"> <li>• UNKNOWN: 表示当前逻辑连接状态未知。</li> <li>• READY: 表示逻辑连接已就绪。</li> <li>• RUN: 表示逻辑连接发送报文正常。</li> <li>• HOLD: 表示逻辑连接发送报文等待中。</li> <li>• CLOSED: 表示关闭逻辑连接。</li> <li>• TO_CLOSED: 表示将会关闭逻辑连接。</li> <li>• WRITING: 表示正在写入数据。</li> </ul>
query_id	bigint	通信流对应的debug_query_id编号。
pn_id	integer	通信流所执行查询的plan_node_id编号。
send_smp	integer	通信流所执行查询send端的smpid编号。
recv_smp	integer	通信流所执行查询recv端的smpid编号。
send_bytes	bigint	通信流发送的数据总量, 单位Byte。
time	bigint	通信流当前生命周期使用时长, 单位ms。
speed	bigint	通信流的平均发送速率, 单位Byte/s。
quota	bigint	通信流当前的通信配额值, 单位Byte。
wait_quota	bigint	通信流等待quota值产生的额外时间开销, 单位ms。

### 15.3.98 PG\_COMM\_QUERY\_SPEED

PG\_COMM\_QUERY\_SPEED视图展示单个节点上所有query对应的流量信息。



表 15-151 PG\_COMM\_QUERY\_SPEED 字段

名称	类型	描述
node_name	text	节点名称。
query_id	bigint	通信流对应的debug_query_id编号。
rxkB/s	bigint	查询对应的通信流接收速率，单位Byte/s。
txkB/s	bigint	查询对应的通信流发送速率，单位Byte/s。
rxkB	bigint	查询对应的通信流接收数据总量，单位Byte。
txkB	bigint	查询对应的通信流发送数据总量，单位Byte。
rxpck/s	bigint	查询对应的通信包接收速率，单位 packages/s。
txpck/s	bigint	查询对应的通信包发送速率，单位 packages/s。
rxpck	bigint	查询对应的通信包接收总量，单位packages。
txpck	bigint	查询对应的通信包发送总量，单位packages。

### 15.3.99 PG\_CONTROL\_GROUP\_CONFIG

PG\_CONTROL\_GROUP\_CONFIG视图存储系统的控制组配置信息。

表 15-152 PG\_CONTROL\_GROUP\_CONFIG 字段

名称	类型	描述
pg_control_group_config	text	控制组的配置信息

### 15.3.100 PG\_CURSORS

PG\_CURSORS视图列出了当前可用的游标。

表 15-153 PG\_CURSORS 字段

名称	类型	描述
name	text	游标名。
statement	text	声明改游标时的查询语句。
is_holdable	boolean	如果该游标是持久的（就是在声明该游标的事务结束后仍然可以访问该游标）则为TRUE，否则为FALSE。

名称	类型	描述
is_binary	boolean	如果该游标被声明为BINARY则为TRUE，否则为FALSE。
is_scrollable	boolean	如果该游标可以滚动（就是允许以不连续的方式检索）则为TRUE，否则为FALSE。
creation_time	timestamp with time zone	声明该游标的时间戳。

### 15.3.101 PG\_EXT\_STATS

PG\_EXT\_STATS视图提供对存储在PG\_STATISTIC\_EXT表里面的扩展统计信息的访问。扩展统计信息目前包括多列统计信息。

表 15-154 PG\_EXT\_STATS 字段

名称	类型	引用	描述
schemaname	name	PG_NAMESP ACE.nspname	包含表的模式名。
tablename	name	PG_CLASS.rel name	表名。
attname	int2vector	PG_STATISTI C_EXT.stakey	统计信息扩展的多列信息。
inherited	boolean	-	如果为真，则包含继承的子列，否则只是指定表的字段。
null_frac	real	-	记录中字段组合为空的百分比。
avg_width	integer	-	字段组合记录以字节记的平均宽度。
n_distinct	real	-	<ul style="list-style-type: none"> <li>如果大于0，表示字段组合中独立数值的估计数目。</li> <li>如果小于0，表示独立数值的数目被行数除的负数。</li> </ul> 用负数形式是因为ANALYZE认为独立数值的数目是随着表增长而增长； 正数的形式用于在字段看上去好像有固定的可能值数目的情况下。比如，-1表示一个字段组合中独立数值的个数和行数相同。 <ul style="list-style-type: none"> <li>如果等于0，表示独立数值的数目未知。</li> </ul>

名称	类型	引用	描述
n_dndistinct	real	-	标识dn1上字段组合中非NULL数据的唯一值的数目。 <ul style="list-style-type: none"> <li>• 如果大于0，表示独立数值的实际数目。</li> <li>• 如果小于0，表示独立数值的数目被行数除的负数。（比如，一个字段组合的数值平均出现概率为两次，则可以表示为 n_dndistinct=-0.5）。</li> <li>• 如果等于0，表示独立数值的数目未知。</li> </ul>
most_common_vals	anyarray	-	一个字段组合里最常用数值的列表。如果该字段组合不存在最常用数值，则为NULL。本列保存的多列常用数值均不为NULL。
most_common_freqs	real[]	-	一个最常用数值组合的频率的列表，即每个出现的次数除以行数。如果most_common_vals是NULL，则为NULL。
most_common_vals_null	anyarray	-	一个字段组合里最常用数值的列表。如果该字段组合不存在最常用数值，则为NULL。本列保存的多列常用数值中至少有一个值为NULL。
most_common_freqs_null	real[]	-	一个最常用数值组合的频率的列表，即每个出现的次数除以行数。如果most_common_vals_null是NULL，则为NULL。

### 15.3.102 PG\_GET\_INVALID\_BACKENDS

PG\_GET\_INVALID\_BACKENDS视图提供显示CN上连接到当前DN备机的后端线程信息。

表 15-155 PG\_GET\_INVALID\_BACKENDS 字段

名称	类型	描述
pid	bigint	线程ID
node_name	text	后端线程中连接的节点信息
dbname	name	当前连接的数据库

名称	类型	描述
backend_start	timestamp with time zone	后端线程启动的时间
query	text	后端线程正在执行的查询语句

### 15.3.103 PG\_GET\_SENDERS\_CATCHUP\_TIME

PG\_GET\_SENDERS\_CATCHUP\_TIME视图显示单个DN上当前活跃的主备发送线程的追赶信息。

表 15-156 PG\_GET\_SENDERS\_CATCHUP\_TIME 字段

名称	类型	描述
pid	bigint	当前sender的线程ID。
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。
type	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当前执行的bcm文件。
catchup_bcm_finished	integer	catchup已经操作完成的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。

### 15.3.104 PG\_GROUP

PG\_GROUP视图查看数据库认证角色及角色之间的成员关系。

表 15-157 PG\_GROUP 字段

名称	类型	描述
groname	name	组的名称
grosysid	oid	组的ID
grolist	oid[]	一个数组，包含这个组里面所有角色的ID

### 15.3.105 PG\_INDEXES

PG\_INDEXES视图提供对数据库中每个索引的有用信息的访问。

表 15-158 PG\_INDEXES 字段

名称	类型	引用	描述
schemaname	name	<a href="#">PG_NAMESPACE.nspname</a>	包含表和索引的模式名
tablename	name	<a href="#">PG_CLASS.relname</a>	此索引所服务的表名
indexname	name	<a href="#">PG_CLASS.relname</a>	索引名
tablespace	name	<a href="#">PG_TABLESPACE.spcname</a>	包含索引的表空间名称
indexdef	text	-	索引定义（一个重建的CREATE INDEX命令）

### 15.3.106 PG\_JOB

PG\_JOB视图用于代替之前版本的PG\_JOB系统表，提供对之前版本的前向兼容。原PG\_JOB系统表已经变更为PG\_JOBS系统表，关于PG\_JOBS系统表的描述详见[PG\\_JOBS](#)。

表 15-159 PG\_JOB 字段

名字	类型	描述
job_id	bigint	作业ID。
current_postgres_pid	bigint	如果当前任务已被执行，那么此处记录运行此任务的postgres线程ID。默认为-1，表示此任务未被执行过。
log_user	name	创建者的UserName。
priv_user	name	作业执行者的UserName。

名字	类型	描述
dbname	name	标识作业执行的数据库名。
node_name	name	标识当前作业是在哪个CN上创建和执行。
job_status	text	<p>当前任务的执行状态，取值范围：('r', 's', 'f', 'd')，默认为's'，取值含义：</p> <ul style="list-style-type: none"> <li>• r=running</li> <li>• s=successfully finished</li> <li>• f=job failed</li> <li>• d=disable</li> </ul> <p>当job连续执行失败16次，会将job_status自动设置为失效状态'd'，后续不再执行该job。</p> <p><b>说明</b></p> <ul style="list-style-type: none"> <li>• 当用户将定时任务关闭（即job_queue_processes为0时），由于监控job执行的线程不会启动，所以job_status不会根据job的实时状态进行设置，用户不需要关注job_status。</li> <li>• 只有当开启定时任务功能（job_queue_processes为非0时），系统才会根据当前job的实时状态刷新job_status的值。</li> </ul>
start_date	timestamp without time zone	作业第一次开始执行时间，时间精确到毫秒。
next_run_date	timestamp without time zone	下次定时执行任务的时间，时间精确到毫秒。
failure_count	smallint	失败计数，作业连续执行失败16次，不再继续执行。
interval	text	作业执行的重复时间间隔。
last_start_date	timestamp without time zone	上次运行开始时间，时间精确到毫秒。
last_end_date	timestamp without time zone	上次运行的结束时间，时间精确到毫秒。
last_suc_date	timestamp without time zone	上次成功运行的开始时间，时间精确到毫秒。
this_run_date	timestamp without time zone	正在运行任务的开始时间，时间精确到毫秒。
nspname	name	作业运行时所在的命名空间的名称。

名字	类型	描述
what	text	作业内容。

### 15.3.107 PG\_JOB\_PROC

PG\_JOB\_PROC视图用于代替之前版本的PG\_JOB\_PROC系统表，提供对之前版本的前向兼容。原PG\_JOB\_PROC系统表已经和原PG\_JOB系统表一同并入当前版本的PG\_JOBS系统表，关于PG\_JOBS系统表的描述详见[PG\\_JOBS](#)。

表 15-160 PG\_JOB\_PROC 字段

名字	类型	描述
job_id	bigint	作业ID
what	text	作业内容

### 15.3.108 PG\_JOB\_SINGLE

PG\_JOB\_SINGLE视图用于显示当前节点的作业信息。

表 15-161 PG\_JOB\_SINGLE 字段

名字	类型	描述
job_id	bigint	作业ID。
current_postgres_pid	bigint	如果当前任务已被执行，那么此处记录运行此任务的postgres线程ID。默认为-1，表示此任务未被执行过。
log_user	name	创建者的UserName。
priv_user	name	作业执行者的UserName。
dbname	name	标识作业执行的数据库名。
node_name	name	标识当前作业是在哪个CN上创建和执行。

名字	类型	描述
job_status	text	<p>当前任务的执行状态，取值范围：('r', 's', 'f', 'd')，默认为's'，取值含义：</p> <ul style="list-style-type: none"> <li>• r=running</li> <li>• s=successfully finished</li> <li>• f=job failed</li> <li>• d=disable</li> </ul> <p>当job连续执行失败16次，会将job_status自动设置为失效状态'd'，后续不再执行该job。</p> <p><b>说明</b></p> <ul style="list-style-type: none"> <li>• 当用户将定时任务关闭（即job_queue_processes为0时），由于监控job执行的线程不会启动，所以job_status不会根据job的实时状态进行设置，用户不需要关注job_status。</li> <li>• 只有当开启定时任务功能（job_queue_processes为非0时），系统才会根据当前job的实时状态刷新job_status的值。</li> </ul>
start_date	timestamp without time zone	作业第一次开始执行时间，时间精确到毫秒。
next_run_date	timestamp without time zone	下次定时执行任务的时间，时间精确到毫秒。
failure_count	smallint	失败计数，作业连续执行失败16次，不再继续执行。
interval	text	作业执行的重复时间间隔。
last_start_date	timestamp without time zone	上次运行开始时间，时间精确到毫秒。
last_end_date	timestamp without time zone	上次运行的结束时间，时间精确到毫秒。
last_suc_date	timestamp without time zone	上次成功运行的开始时间，时间精确到毫秒。
this_run_date	timestamp without time zone	正在运行任务的开始时间，时间精确到毫秒。
nspname	name	作业运行时所在的命名空间的名称。
what	text	作业内容。



### 15.3.109 PG\_LIFECYCLE\_DATA\_DISTRIBUTE

PG\_LIFECYCLE\_DATA\_DISTRIBUTE视图查询OBS多温表中冷热数据分布情况。

表 15-162 PG\_LIFECYCLE\_DATA\_DISTRIBUTE 字段

名称	类型	描述
schemaname	name	模式名
tablename	name	当前表名
nodename	name	节点名
hotpartition	text	该DN节点上的热分区
coldpartition	text	该DN节点上的冷分区
switchablepartition	text	该DN节点上的可切分区
hotdatasize	text	该DN节点上的热分区数据大小
colddatasize	text	该DN节点上的冷分区数据大小
switchabledatasize	text	该DN节点上的可切分区数据大小

### 15.3.110 PG\_LOCKS

PG\_LOCKS视图存储各打开事务所持有的锁信息。

表 15-163 PG\_LOCKS 字段

名称	类型	引用	描述
locktype	text	-	被锁定对象的类型：relation, extend, page, tuple, transactionid, virtualxid, object, userlock, advisory。
database	oid	<a href="#">PG_DATABASE.oid</a>	被锁定对象所在数据库的OID。 <ul style="list-style-type: none"> <li>如果被锁定的对象是共享对象，则OID为0。</li> <li>如果是一个事务ID，则为NULL。</li> </ul>

名称	类型	引用	描述
relation	oid	<a href="#">PG_CLASS.oid</a>	被锁定对象关系的OID，如果锁定的对象不是关系，也不是关系的一部分，则为NULL。
page	integer	-	关系内部的页面编号，如果对象不是关系页或者不是行页，则为NULL。
tuple	smallint	-	页面里边的行编号，如果对象不是行，则为NULL。
virtualxid	text	-	事务的虚拟ID，如果对象不是一个虚拟事务ID，则为NULL。
transactionid	xid	-	事务的ID，如果对象不是一个事务ID，则为NULL。
classid	oid	<a href="#">PG_CLASS.oid</a>	包含该对象的系统表的OID，如果对象不是普通的数据库对象，则为NULL。
objid	oid	-	对象在其系统表内的OID，如果对象不是普通数据库对象，则为NULL。
objsubid	smallint	-	对于表的某个字段对应为字段编号；对于其他对象类型，该字段为0；如果该对象不是普通数据库对象，则为NULL。
virtualtransaction	text	-	持有此锁或者在等待此锁的事务的虚拟ID。
pid	bigint	-	持有此锁或者等待此锁的服务器线程的逻辑ID。如果锁被一个预备事务持有，则为NULL。
mode	text	-	此线程持有的或者是期望持有的锁模式。更多有关锁模式的内容请参见 <a href="#">LOCK</a> 。
granted	boolean	-	<ul style="list-style-type: none"> <li>● 如果锁是持有锁，则为TRUE。</li> <li>● 如果锁是等待锁，则为FALSE。</li> </ul>
fastpath	boolean	-	如果通过fast-path获得锁，则为TRUE；如果通过主锁表获得，则为FALSE。

### 15.3.111 PG\_NODE\_ENV

PG\_NODE\_ENV视图提供获取当前节点的环境变量信息。

表 15-164 PG\_NODE\_ENV 字段

名称	类型	描述
node_name	text	当前节点名称
host	text	当前节点的主机名称
process	integer	当前节点的进程号
port	integer	当前节点的端口号
installpath	text	当前节点的安装目录
datapath	text	当前节点的数据目录
log_directory	text	当前节点的日志目录

### 15.3.112 PG\_OS\_THREADS

PG\_OS\_THREADS视图提供当前节点下所有线程的状态信息。

表 15-165 PG\_OS\_THREADS 字段

名称	类型	描述
node_name	text	当前节点名称
pid	bigint	当前节点进程中正在运行的线程号
lwpid	integer	与pid对应的轻量级线程号
thread_name	text	与pid对应的线程名称
creation_time	timestamp with time zone	与pid对应的线程创建的时间

### 15.3.113 PG\_POOLER\_STATUS

PG\_POOLER\_STATUS视图查询pooler中的缓存连接状态。该视图只能在CN上执行查询，显示本地CN的pooler模块的连接缓存信息。

表 15-166 PG\_POOLER\_STATUS 字段

名称	类型	描述
database	text	数据库名称
user_name	text	用户名
tid	bigint	连接CN的线程ID

名称	类型	描述
node_oid	bigint	连接的实例节点OID
node_name	name	连接的实例节点名称
in_use	boolean	连接是否正被使用 <ul style="list-style-type: none"> <li>• t ( true ) : 表示连接正在使用</li> <li>• f ( false ) : 表示连接没有使用</li> </ul>
fdsock	bigint	对端socket
remote_pid	bigint	对端线程号
session_params	text	由此连接下发的GUC session参数

### 15.3.114 PG\_PREPARED\_STATEMENTS

PG\_PREPARED\_STATEMENTS视图显示当前会话所有可用的预备语句。

表 15-167 PG\_PREPARED\_STATEMENTS 字段

名称	类型	描述
name	text	预备语句的标识符。
statement	text	创建该预备语句的查询字符串。对于从SQL创建的预备语句而言是客户端提交的PREPARE语句；对于通过前/后端协议创建的预备语句而言是预备语句自身的文本。
prepare_time	timestamp with time zone	创建该预备语句的时间戳。
parameter_types	regtype[]	该预备语句期望的参数类型，以regtype类型的数组格式出现。与该数组元素相对应的OID可以通过把regtype转换为oid值得到。
from_sql	boolean	<ul style="list-style-type: none"> <li>• 如果该预备语句是通过PREPARE语句创建的则为true。</li> <li>• 如果是通过前/后端协议创建的则为false。</li> </ul>

### 15.3.115 PG\_PREPARED\_XACTS

PG\_PREPARED\_XACTS视图显示当前准备好进行两阶段提交的事务的信息。

表 15-168 PG\_PREPARED\_XACTS 字段

名称	类型	引用	描述
transaction	xid	-	预备事务的数字事务标识
gid	text	-	赋予该事务的全局事务标识
prepared	timestamp with time zone	-	事务准备好提交的时间
owner	name	PG_AUTHID.rolname	执行该事务的用户名
database	name	PG_DATABASE.datname	执行该事务所在的数据库名

### 15.3.116 PG\_PUBLICATION\_TABLES

PG\_PUBLICATION\_TABLES视图显示发布与其所发布的表之间的映射信息。和底层的系统表PG\_PUBLICATION\_REL不同，这个视图展开了定义为FOR ALL TABLES和FOR ALL TABLES IN SCHEMA的发布，对这类发布来说，每一个可发布的表都有一行。该系统视图仅8.2.0.100及以上集群版本支持。

表 15-169 PG\_PUBLICATION\_TABLES 字段

名称	类型	描述
pubname	name	发布的名称。
schemaname	name	表的模式的名称。
tablename	name	表的名称。

#### 应用示例

查询所有发布表。

```
SELECT * FROM PG_PUBLICATION_TABLES;
pubname | schemaname | tablename
-----+-----+-----
mypub   | public     | t1
mypub   | public     | t2
(2 rows)
```

### 15.3.117 PG\_QUERYBAND\_ACTION

PG\_QUERYBAND\_ACTION视图显示query\_band关联行为和次序。

表 15-170 PG\_QUERYBAND\_ACTION 字段

名称	类型	描述
qband	text	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	integer	query_band搜索次序

### 15.3.118 PG\_REPLICATION\_SLOTS

PG\_REPLICATION\_SLOTS视图查看复制节点的信息。

表 15-171 PG\_REPLICATION\_SLOTS 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

### 15.3.119 PG\_ROLES

PG\_ROLES视图提供访问数据库角色的相关信息。

表 15-172 PG\_ROLES 字段

名称	类型	引用	描述
rolname	name	-	角色名。

名称	类型	引用	描述
rolsuper	boolean	-	该角色是否是拥有最高权限的初始系统管理员。
rolinherit	boolean	-	该角色是否继承角色的权限。
rolcreaterole	boolean	-	该角色是否可以创建其他的角色。
rolcreatedb	boolean	-	该角色是否可以创建数据库。
rolcatupdate	boolean	-	该角色是否可以更新系统表。只有usesysid=10的初始系统管理员拥有此权限。其他用户无法获得此权限。
rolcanlogin	boolean	-	该角色是否可以登录数据库。
rolreplication	boolean	-	该角色是否可以复制。
rolauditadmin	boolean	-	该角色是否为审计管理员。
rolsystemadmin	boolean	-	该角色是否为系统管理员。
rolconnlimit	integer	-	对于可以登录的角色，rolconnlimit限制了该角色允许发起的最大并发连接数。-1表示无限制。
rolpassword	text	-	不是口令，总是*****。
rolvalidbegin	timestamp with time zone	-	帐户的有效开始时间；如果没有设置有效开始时间，则为NULL。
rolvaliduntil	timestamp with time zone	-	帐户的有效结束时间；如果没有设置有效结束时间，则为NULL。
rolrespool	name	-	用户所能够使用的resource pool。
rolparentid	oid	<a href="#">PG_AUTHID.rolparentid</a>	用户所在组用户的OID。
roltabspace	text	-	用户永久表存储空间限额。
roltemp space	text	-	用户临时表存储空间限额。
rolspill space	text	-	用户算子落盘空间限额。
rolconfig	text[]	-	运行时配置变量的会话缺省。
oid	oid	<a href="#">PG_AUTHID.oid</a>	角色的ID。
roluseft	boolean	<a href="#">PG_AUTHID.roluseft</a>	角色是否可以操作外表。

名称	类型	引用	描述
nodegroup	name	-	角色所关联的逻辑集群名字，如果没有关联逻辑集群，该值为空。

### 15.3.120 PG\_RULES

PG\_RULES视图提供对查询重写规则的有效信息访问的接口。

表 15-173 PG\_RULES 字段

名称	类型	描述
schemaname	name	包含表的模式名
tablename	name	规则作用的表名
rulename	name	规则的名称
definition	text	规则定义（一个重新构造的创建命令）

### 15.3.121 PG\_RUNNING\_XACTS

PG\_RUNNING\_XACTS视图主要功能是显示当前节点运行事务的信息。

表 15-174 PG\_RUNNING\_XACTS 字段

名称	类型	描述
handle	integer	事务在GTM对应的句柄
gxid	xid	事务ID号
state	tinyint	事务状态（3: prepared或者0: starting）
node	text	节点名称
xmin	xid	节点上当前数据涉及的最小事务号xmin
vacuum	boolean	标志当前事务是否是lazy vacuum事务
timeline	bigint	标志数据库重启次数
prepare_xid	xid	处于prepared状态的事务的ID号，若不在prepared状态，值为0
pid	bigint	事务对应的线程id
next_xid	xid	CN传给DN的事务id号



## 15.3.122 PG\_SECLABELS

PG\_SECLABELS视图提供关于安全标签的信息。

表 15-175 PG\_SECLABELS 字段

名字	类型	引用	描述
objoid	oid	任意OID属性	安全标签所属的对象的OID。
classoid	oid	<a href="#">PG_CLASS</a> .oid	此对象的系统表的OID。
objsubid	integer	-	对于某个在表字段上的安全标签，为字段编号（引用表本身的objoid和classoid）。对于所有其他对象类型，该字段为0。
objtype	text	-	标签出现的对象的类型。
objnamespace	oid	<a href="#">PG_NAMESPACE</a> .oid	对象的命名空间的OID，如果适用；否则为NULL。
objname	text	-	标签适用的对象名。
provider	text	<a href="#">PG_SECLABEL</a> .provider	与标签相关的标签提供者。
label	text	<a href="#">PG_SECLABEL</a> .label	应用于此对象的安全标签。

## 15.3.123 PG\_SESSION\_WLMSTAT

PG\_SESSION\_WLMSTAT视图显示和当前用户执行作业正在运行时的负载管理相关信息。

表 15-176 PG\_SESSION\_WLMSTAT 字段

名称	类型	描述
datid	oid	连接后端的数据库OID。
datname	name	连接后端的数据库名称。
threadid	bigint	后端线程ID。
processid	integer	后端线程的pid。
usesysid	oid	登录后端的用户OID。
appname	text	连接到后端的应用名。
username	name	登录到该后端的用户名。
priority	bigint	语句所在Cgroups的优先级。

名称	类型	描述
attribute	text	语句的属性： <ul style="list-style-type: none"> <li>• Ordinary: 语句发送到数据库后被解析前的默认属性。</li> <li>• Simple: 简单语句。</li> <li>• Complicated: 复杂语句。</li> <li>• Internal: 数据库内部语句。</li> </ul>
block_time	bigint	语句当前为止的pending的时间，单位s。
elapsed_time	bigint	语句当前为止的实际执行时间，单位s。
total_cpu_time	bigint	语句在上一时间周期内的DN上CPU使用的总时间，单位s。
cpu_skew_percent	integer	语句在上一时间周期内的DN上CPU使用的倾斜率。
statement_mem	integer	语句执行所需要的估算内存，预留字段。
active_points	integer	语句占用的资源池并发点数。
dop_value	integer	语句的从资源池中获取的dop值。
control_group	text	语句当前所使用的Cgroups。
status	text	语句当前的状态，包括： <ul style="list-style-type: none"> <li>• pending: 执行前状态。</li> <li>• running: 执行进行状态。</li> <li>• finished: 执行正常结束。（当enqueue字段为StoredProc或Transaction时，仅代表语句中的部分作业已经执行完毕，该状态会持续到该语句完全执行完毕。）</li> <li>• aborted: 执行异常终止。</li> <li>• active: 非以上四种状态外的正常状态。</li> <li>• unknown: 未知状态。</li> </ul>
enqueue	text	语句当前的排队情况，包括： <ul style="list-style-type: none"> <li>• Global: 全局排队。</li> <li>• Respool: 资源池排队。</li> <li>• CentralQueue: 在中心协调节点(CCN)中排队。</li> <li>• Transaction: 语句处于一个事务块中。</li> <li>• StoredProc: 语句处于一个存储过程中。</li> <li>• None: 未在排队。</li> <li>• Forced None: 事务块语句或存储过程语句由于超出设定的等待时间而强制执行。</li> </ul>
resource_pool	name	语句当前所在的资源池。

名称	类型	描述
query	text	该后端的最新查询。如果state状态是active，此字段显示当前正在执行的查询。所有其他情况表示上一个查询。
isplana	bool	逻辑集群模式下，语句当前是否占用其他逻辑集群的资源执行。该值默认为f，表示不占用其他逻辑集群的资源执行。
node_group	text	语句所属用户对应的逻辑集群。
lane	text	表示语句查询的快慢车道。 <ul style="list-style-type: none"> <li>fast: 快车道。</li> <li>slow: 慢车道。</li> <li>none: 未管控。</li> </ul>

### 15.3.124 PG\_SESSION\_IOSTAT

PG\_SESSION\_IOSTAT视图8.1.2版本中已废弃，为兼容历史版本功能保留该视图，当前版本查询无效。

表 15-177 PG\_SESSION\_IOSTAT 字段

名称	类型	描述
query_id	bigint	作业ID。
mincurriops	integer	该作业当前io在各DN中的最小值。
maxcurriops	integer	该作业当前io在各DN中的最大值。
minpeakioops	integer	在作业运行时，作业io峰值中，各DN的最小值。
maxpeakioops	integer	在作业运行时，作业io峰值中，各DN的最大值。
io_limits	integer	该作业所设GUC参数io_limits。
io_priority	text	该作业所设GUC参数io_priority。
query	text	作业。
node_group	text	作业所属用户对应的逻辑集群。

### 15.3.125 PG\_SETTINGS

PG\_SETTINGS视图显示数据库运行时参数的相关信息。

表 15-178 PG\_SETTINGS 字段

名称	类型	描述
name	text	参数名称。
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, backend, superuser, user。
vartype	text	参数类型，包括bool, enum, integer, real, 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。

### 15.3.126 PG\_SHADOW

PG\_SHADOW视图显示了所有在PG\_AUTHID中标记了rolcanlogin的角色的属性。

这个系统表的名字来自于该表是不可读的，因为它包含口令。**PG\_USER**是一个在PG\_SHADOW上公开可读的视图，只是把口令域填成了空白。

表 15-179 PG\_SHADOW 字段

名字	类型	引用	描述
username	name	<b>PG_AUTHID</b> .rolname	用户名。

名字	类型	引用	描述
usesysid	oid	<a href="#">PG_AUTHID.oid</a>	用户的ID。
usecreatedb	boolean	-	用户可以创建数据库。
usesuper	boolean	-	用户是系统管理员。
usecatupd	boolean	-	用户可以更新系统表。即使是系统管理员，如果此字段不为真，也不能更新系统表。
userepl	boolean	-	用户可以初始化流复制和使系统处于或不处于备份模式。
passwd	text	-	口令（可能是加密的）；如果没有则为null。参阅 <a href="#">PG_AUTHID</a> 获取加密的口令是如何存储的信息。
valbegin	timestamp with time zone	-	帐户的有效开始时间；如果没有设置有效开始时间，则为NULL。
valuntil	timestamp with time zone	-	帐户的有效结束时间；如果没有设置有效结束时间，则为NULL。
respool	name	-	用户使用的资源池。
parent	oid	-	父资源池。
spacelimit	text	-	永久表存储空间限额。
tempspacelimit	text	-	临时表存储空间限额。
spillspacelimit	text	-	算子落盘空间限额。
useconfig	text[ ]	-	运行时配置变量的会话缺省。

### 15.3.127 PG\_SHARED\_MEMORY\_DETAIL

PG\_SHARED\_MEMORY\_DETAIL视图查询所有已产生的共享内存上下文的使用信息。

表 15-180 PG\_SHARED\_MEMORY\_DETAIL 字段

名字	类型	描述
contextname	text	内存上下文的名字
level	smallint	当前上下文在整体内存上下文中的层级
parent	text	上级内存上下文
totalsize	bigint	共享内存总大小, 单位Byte
freesize	bigint	共享内存剩余大小, 单位Byte
usedsize	bigint	共享内存使用大小, 单位Byte

### 15.3.128 PG\_STATS

PG\_STATS视图提供对存储在pg\_statistic表里面的单列统计信息的访问。

表 15-181 PG\_STATS 字段

名称	类型	引用	描述
schemaname	name	<b>PG_NAMESPACE</b> .nspname	包含表的模式名。
tablename	name	<b>PG_CLASS</b> .relname	表名。
attname	name	<b>PG_ATTRIBUTE</b> .attname	字段名。
inherited	boolean	-	如果为真, 则包含继承的子列, 否则只是指定表的字段。
null_frac	real	-	记录中字段为空的百分比。
avg_width	integer	-	字段记录以字节记的平均宽度。
n_distinct	real	-	<ul style="list-style-type: none"> <li>如果大于0, 表示字段中独立数值的估计数目。</li> <li>如果小于0, 表示独立数值的数目被行数除的负数。</li> </ul> 用负数形式是因为ANALYZE认为独立数值的数目是随着表增长而增长; 正数的形式用于在字段看上去好像有固定的可能值数目的情况下。比如, -1表示一个唯一字段, 独立数值的个数和行数相同。

名称	类型	引用	描述
n_dndistinct	real	-	标识dn1上字段中非NULL数据的唯一值的数目。 <ul style="list-style-type: none"> <li>• 如果大于0，表示独立数值的实际数目。</li> <li>• 如果小于0，表示独立数值的数目被行数除的负数。（比如，一个字段的数值平均出现概率为两次，则可以表示为 n_dndistinct=-0.5）。</li> <li>• 如果等于0，表示独立数值的数目未知。</li> </ul>
most_common_vals	anyarray	-	一个字段组合里最常用数值的列表。如果该字段组合不存在最常用数值，则为NULL。
most_common_freqs	real[]	-	一个最常用数值的频率的列表，即每个出现的次数除以行数。如果 most_common_vals 是 NULL，则为 NULL。
histogram_buckets	anyarray	-	一个数值的列表，它把字段的数值分成几组大致相同的组。如果在 most_common_vals 里有数值，则在此饼图的计算中省略。如果字段数据类型没有 <操作符 或者 most_common_vals 列表代表了整个分布性，则此字段为 NULL。
correlation	real	-	统计与字段值的物理行序和逻辑行序有关。它的范围从-1到+1。在数值接近-1或者+1的时候，在字段上的索引扫描将被认为比它接近零的时候开销更少，因为减少了对磁盘的随机访问。如果字段数据类型没有 <操作符，则这个字段为 NULL。
most_common_elems	anyarray	-	一个最常用的非空元素的列表。
most_common_elem_freqs	real[]	-	一个最常用元素的频率的列表。
elem_count_histogram	real[]	-	对于独立的非空元素的统计直方图。

### 15.3.129 PG\_STAT\_ACTIVITY

PG\_STAT\_ACTIVITY视图显示和当前用户查询相关的信息。若有管理员权限或预置角色权限可以显示和所有用户查询相关的信息。

表 15-182 PG\_STAT\_ACTIVITY 字段

名称	类型	描述
datid	oid	用户会话在后端连接到的数据库OID。
datname	name	用户会话在后端连接到的数据库名称。
pid	bigint	后端线程ID。
lwtid	integer	轻量级线程ID。
usesysid	oid	登录该后端的用户OID。
username	name	登录该后端的用户名。
application_name	text	连接到该后端的应用名。
client_addr	inet	连接到该后端的客户端的IP地址。如果此字段是null，则表示通过服务器机器上UNIX套接字连接客户端或者这是内部进程，如autovacuum。
client_hostname	text	客户端的主机名，此字段是通过client_addr的反向DNS查找得到。此字段只有在启动log_hostname且使用IP连接时才非空。
client_port	integer	客户端用于与后端通讯的TCP端口号，如果使用Unix套接字，则为-1。
backend_start	timestamp with time zone	后端进程启动时间，即客户端连接服务器的时间。
xact_start	timestamp with time zone	当前事务的启动时间，如果没有事务是活跃的，则为null。如果当前查询是首个事务，则这列等同于query_start列。
query_start	timestamp with time zone	开始当前活跃查询的时间，如果state的值不是active，则这个值是上一个查询的开始时间。
state_change	timestamp with time zone	状态最后一次改变的时间。
waiting	boolean	如果后端当前正等待锁或者等待节点则为true。



名称	类型	描述
enqueue	text	<p>语句当前排队状态。可能值是：</p> <ul style="list-style-type: none"> <li>waiting in queue：表示语句在排队中。</li> <li>waiting in global queue：表示语句在全局排队中。</li> <li>waiting in respool queue：表示语句在资源池排队中。</li> <li>waiting in ccn queue：表示作业在CCN排队中。</li> <li>空或no waiting queue：表示语句正在运行。</li> </ul>
state	text	<p>后端当前总体状态。可能值是：</p> <ul style="list-style-type: none"> <li>active：后台正在执行查询。</li> <li>idle：后台正在等待新的客户端命令。</li> <li>idle in transaction：后端在事务中，但事务中没有语句在执行。</li> <li>idle in transaction (aborted)：后端在事务中，但事务中有语句执行失败。</li> <li>fastpath function call：后端正在执行一个fast-path函数。</li> <li>disabled：如果后端禁用track_activities，则报告此状态。</li> </ul> <p><b>说明</b> 普通用户只能查看到自己帐户所对应的会话状态。即其他帐户的state信息为空。例如以judy用户连接数据库后，在pg_stat_activity中查看到的普通用户joe及初始用户omm的state信息为空：</p> <pre>SELECT datname, username, usesysid, state,pid FROM pg_stat_activity;  datname   username   usesysid   state    pid -----+-----+-----+-----+-----  postgres   omm             10           139968752121616  postgres   omm             10           139968903116560  db_tpcds   judy        16398   active    139968391403280  postgres   omm             10           139968643069712  postgres   omm             10           139968680818448  postgres   joe         16390           139968563377936 (6 rows)</pre>
resource_pool	name	用户使用的资源池。

名称	类型	描述
stmt_type	text	语句类型。
query_id	bigint	查询语句的ID。
query	text	此后端的最新查询。如果state状态是active（活跃的），此字段显示当前正在执行的查询。其他情况表示上一个查询。
connection_info	text	json格式字符串，记录当前连接数据库的驱动类型、驱动版本号、当前驱动的部署路径、进程属主用户等信息（参见 <a href="#">connection_info</a> ）。

### 15.3.130 PG\_STAT\_ALL\_INDEXES

PG\_STAT\_ALL\_INDEXES视图将包含当前数据库中的每个索引行，显示访问特定索引的统计。

索引可以通过简单的索引扫描或"位图"索引扫描进行使用。位图扫描中几个索引的输出可以通过AND或者OR规则进行组合，因此当使用位图扫描的时候，很难将独立堆行抓取与特定索引进行组合，因此，一个位图扫描增加pg\_stat\_all\_indexes.idx\_tup\_read使用索引计数，并且增加pg\_stat\_all\_tables.idx\_tup\_fetch表计数，但不影响pg\_stat\_all\_indexes.idx\_tup\_fetch。

表 15-183 PG\_STAT\_ALL\_INDEXES 字段

名称	类型	描述
relid	oid	索引的表的OID
indexrelid	oid	索引的OID
schemaname	name	索引中模式名
relname	name	索引的表名
indexrelname	name	索引名
idx_scan	bigint	索引上开始的索引扫描数
idx_tup_read	bigint	通过索引上扫描返回的索引项数
idx_tup_fetch	bigint	通过使用索引的简单索引扫描抓取的活表行数

### 15.3.131 PG\_STAT\_ALL\_TABLES

PG\_STAT\_ALL\_TABLES视图包含当前数据库中每个表每行的信息（包括TOAST表），显示访问特定表的统计信息。

表 15-184 PG\_STAT\_ALL\_TABLES 字段

名称	类型	描述
relid	oid	表的OID。
schemaname	name	此表的模式名。
relname	name	表名。
seq_scan	bigint	在此表上启动的顺序扫描数。
seq_tup_read	bigint	顺序扫描抓取的有live数据行的数目。
idx_scan	bigint	索引扫描的次数。
idx_tup_fetch	bigint	索引扫描抓取的有live数据行的数目。
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（即不需要单独的索引更新）。
n_live_tup	bigint	live行估计数。
n_dead_tup	bigint	dead行估计数。
last_vacuum	timestamp with time zone	最后一次手动vacuum时间（不计算VACUUM FULL）。
last_autovacuum	timestamp with time zone	最后一次autovacuum时间。
last_analyze	timestamp with time zone	最后一次analyze时间。
last_autoanalyze	timestamp with time zone	最后一次autovacuum时间。
vacuum_count	bigint	vacuum次数（不计算VACUUM FULL）。
autovacuum_count	bigint	autovacuum次数。
analyze_count	bigint	analyze次数。
autoanalyze_count	bigint	autoanalyze次数。

名称	类型	描述
last_data_changed	timestamp with time zone	记录该表最后一次数据发生变化的时间（引起数据变化的操作包括INSERT/UPDATE/DELETE、EXCHANGE/TRUNCATE/DROP partition），该列数据仅在本地CN记录。

### 15.3.132 PG\_STAT\_BAD\_BLOCK

PG\_STAT\_BAD\_BLOCK视图显示自节点启动后，读取数据时出现Page/CU校验失败的统计信息。

表 15-185 PG\_STAT\_BAD\_BLOCK 字段

名字	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

### 15.3.133 PG\_STAT\_BGWRITER

PG\_STAT\_BGWRITER视图显示关于后端写进程活动的统计信息。

表 15-186 PG\_STAT\_BGWRITER 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。

名称	类型	描述
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。
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。

### 15.3.134 PG\_STAT\_DATABASE

PG\_STAT\_DATABASE视图显示当前节点上每个数据库的状态和统计信息。

表 15-187 PG\_STAT\_DATABASE 字段

名称	类型	描述
datid	oid	数据库OID。
datname	name	数据库名。
numbackends	integer	当前节点上连接到该数据库的后端数。这是该视图中唯一一个反映目前状态值的列；所有列均返回自上次重置以来的累积值。
xact_commit	bigint	当前节点上该数据库中已经提交的事务数。
xact_rollback	bigint	当前节点上该数据库中已经回滚的事务数。
blks_read	bigint	当前节点上该数据库中读取的磁盘块的数量。
blks_hit	bigint	当前节点上高速缓存中发现的磁盘块的个数，即缓存中命中的块数（只包括GaussDB(DWS)缓冲区高速缓存，不包括文件系统的缓存）。
tup_returned	bigint	当前节点上该数据库查询返回的行数。

名称	类型	描述
tup_fetched	bigint	当前节点上该数据库查询抓取的行数。
tup_inserted	bigint	当前节点上该数据库插入的行数。
tup_updated	bigint	当前节点上该数据库更新的行数。
tup_deleted	bigint	当前节点上该数据库删除的行数。
conflicts	bigint	当前节点上由于数据库恢复冲突取消的查询数量（只在备用服务器上发生）。可参见 <a href="#">PG_STAT_DATABASE_CONFLICTS</a> 。
temp_files	bigint	当前节点上该数据库创建的临时文件个数。计算所有临时文件，不论为什么创建临时文件（比如排序或者哈希），而且不考虑log_temp_files设置。
temp_bytes	bigint	当前节点上该数据库写入临时文件的大小。计算所有临时文件，不论为什么创建临时文件，而且不考虑log_temp_files设置。
deadlocks	bigint	当前节点上该数据库中发生的死锁数量。
blk_read_time	double precision	当前节点上该数据库后端读取数据文件块花费的时间，以毫秒计算。
blk_write_time	double precision	当前节点上该数据库后端写入数据文件块花费的时间，以毫秒计算。
stats_reset	timestamp with time zone	当前节点上该数据库统计重置的时间。

### 15.3.135 PG\_STAT\_DATABASE\_CONFLICTS

PG\_STAT\_DATABASE\_CONFLICTS视图显示数据库冲突状态的统计信息。

表 15-188 PG\_STAT\_DATABASE\_CONFLICTS 字段

名称	类型	描述
datid	oid	数据库OID
datname	name	数据库名
confl_tablespace	bigint	冲突的表空间的数目
confl_lock	bigint	冲突的锁数目
confl_snapshot	bigint	冲突的快照数目

名称	类型	描述
confl_bufferpi n	bigint	冲突的缓冲区数目
confl_deadloc k	bigint	冲突的死锁数目

### 15.3.136 PG\_STAT\_GET\_MEM\_MBYTES\_RESERVED

PG\_STAT\_GET\_MEM\_MBYTES\_RESERVED视图显示线程在内存中保存的当前活动信息。该函数在调用时需要指定线程ID，线程ID的选取请参考PG\_STAT\_ACTIVITY中的pid，线程ID为0时表示选取当前线程ID，例如：

```
SELECT pg_stat_get_mem_mbytes_reserved(0);
```

表 15-189 PG\_STAT\_GET\_MEM\_MBYTES\_RESERVED 信息

名称	描述
ConnectInfo	连接信息
ParctlManager	并发管理信息
GeneralParams	基本参数信息
GeneralParams RPDATA	基本资源池信息
ExceptionManager	异常管理信息
CollectInfo	收集信息
GeneralInfo	基本信息
ParctlState	并发状态信息
CPU INFO	CPU信息
ControlGroup	控制组信息
IOSTATE	IO状态信息

### 15.3.137 PG\_STAT\_USER\_FUNCTIONS

PG\_STAT\_USER\_FUNCTIONS视图显示命名空间中用户自定义函数（函数语言为非内部语言）的状态信息。

表 15-190 PG\_STAT\_USER\_FUNCTIONS 字段

名称	类型	描述
funcid	oid	函数OID

名称	类型	描述
schemaname	name	模式名
funcname	name	函数名
calls	bigint	函数被调用的次数
total_time	double precision	函数的总执行时长
self_time	double precision	当前线程调用函数的总的时长

### 15.3.138 PG\_STAT\_USER\_INDEXES

PG\_STAT\_USER\_INDEXES视图显示数据库中用户自定义普通表和toast表的索引状态信息。

表 15-191 PG\_STAT\_USER\_INDEXES 字段

名称	类型	描述
relid	oid	此索引表的OID
indexrelid	oid	索引的OID
schemaname	name	索引中模式名
relname	name	索引的表名
indexrelname	name	索引名
idx_scan	bigint	通过索引扫描的次数
idx_tup_read	bigint	通过索引上扫描返回的索引条目数量
idx_tup_fetch	bigint	索引扫描抓取的有live数据行的数目

### 15.3.139 PG\_STAT\_USER\_TABLES

PG\_STAT\_USER\_TABLES视图显示所有命名空间中用户自定义普通表和toast表的状态信息。

表 15-192 PG\_STAT\_USER\_TABLES 字段

名称	类型	描述
relid	oid	表的OID
schemaname	name	表的模式名
relname	name	表名
seq_scan	bigint	在此表上表启动的顺序扫描的次数



名称	类型	描述
seq_tup_read	bigint	顺序扫描抓取的有live数据行的数目
idx_scan	bigint	索引扫描的次数
idx_tup_fetch	bigint	索引扫描抓取的有live数据行的数目
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（即不需要单独的索引更新）
n_live_tup	bigint	live行估计数
n_dead_tup	bigint	dead行估计数
last_vacuum	timestamp with time zone	最后一次手动vacuum时间（不计算VACUUM FULL）
last_autovacuum	timestamp with time zone	最后一次autovacuum时间
last_analyze	timestamp with time zone	最后一次analyze时间
last_autoanalyze	timestamp with time zone	最后一次autoanalyze时间
vacuum_count	bigint	vacuum的次数（不计算VACUUM FULL）
autovacuum_count	bigint	autovacuum的次数
analyze_count	bigint	analyze的次数
autoanalyze_count	bigint	autoanalyze的次数

### 15.3.140 PG\_STAT\_REPLICATION

PG\_STAT\_REPLICATION视图用于描述日志同步状态信息，如发起端发送日志位置，收端接收日志位置等。

表 15-193 PG\_STAT\_REPLICATION 字段

名称	类型	描述
pid	bigint	线程的PID
usesysid	oid	用户系统ID
username	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，还是一致的流状态）
sender_sent_location	text	发送端发送日志位置
receiver_writelocation	text	接收端write日志位置
receiver_flush_location	text	接收端flush日志位置
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，还是潜在同步）

### 15.3.141 PG\_STAT\_SYS\_INDEXES

PG\_STAT\_SYS\_INDEXES视图显示pg\_catalog、information\_schema模式中所有系统表的索引状态信息。

表 15-194 PG\_STAT\_SYS\_INDEXES 字段

名称	类型	描述
relid	oid	此索引表的OID
indexrelid	oid	索引的OID

名称	类型	描述
schemaname	name	索引中模式名
relname	name	索引的表名
indexrelname	name	索引名
idx_scan	bigint	通过索引扫描的次数
idx_tup_read	bigint	通过索引上扫描返回的索引条目数
idx_tup_fetch	bigint	索引扫描抓取的有live数据行的数目

### 15.3.142 PG\_STAT\_SYS\_TABLES

PG\_STAT\_SYS\_TABLES视图显示pg\_catalog、information\_schema模式的所有命名空间中系统表的统计信息。

表 15-195 PG\_STAT\_SYS\_TABLES 字段

名称	类型	描述
relid	oid	表的OID
schemaname	name	表的模式名
relname	name	表名
seq_scan	bigint	在此表上表启动的顺序扫描的次数
seq_tup_read	bigint	顺序扫描抓取的有live数据行的数目
idx_scan	bigint	索引扫描的次数
idx_tup_fetch	bigint	索引扫描抓取的有live数据行的数目
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（比如没有更新所需的单独索引）
n_live_tup	bigint	live行估计数
n_dead_tup	bigint	dead行估计数
last_vacuum	timestamp with time zone	最后一次手动vacuum时间（不计算VACUUM FULL）

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analyze时间
last_autoanalyze	timestamp with time zone	最后一次autoanalyze时间
vacuum_count	bigint	vacuum的次数（不计算VACUUM FULL）
autovacuum_count	bigint	autovacuum的次数
analyze_count	bigint	analyze的次数
autoanalyze_count	bigint	autoanalyze的次数

### 15.3.143 PG\_STAT\_XACT\_ALL\_TABLES

PG\_STAT\_XACT\_ALL\_TABLES视图显示命名空间中所有普通表和toast表的事务状态信息。

表 15-196 PG\_STAT\_XACT\_ALL\_TABLES 字段

名称	类型	描述
relid	oid	表的OID
schemaname	name	此表的模式名
relname	name	表名
seq_scan	bigint	在此表上启动的顺序扫描数
seq_tup_read	bigint	顺序扫描抓取的活跃行数
idx_scan	bigint	在此表上启动的索引扫描数
idx_tup_fetch	bigint	索引扫描抓取的活跃行数
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

名称	类型	描述
n_tup_hot_upd	bigint	HOT更新行数（比如没有更新所需的单独索引）

### 15.3.144 PG\_STAT\_XACT\_SYS\_TABLES

PG\_STAT\_XACT\_SYS\_TABLES视图显示命名空间中系统表的事务状态信息。

表 15-197 PG\_STAT\_XACT\_SYS\_TABLES 字段

名称	类型	描述
relid	oid	表的OID
schemaname	name	此表的模式名
relname	name	表名
seq_scan	bigint	在此表上启动的顺序扫描数
seq_tup_read	bigint	顺序扫描抓取的活跃行数
idx_scan	bigint	在此表上启动的索引扫描数
idx_tup_fetch	bigint	索引扫描抓取的活跃行数
n_tup_ins	bigint	插入行数
n_tup_upd	bigint	更新行数
n_tup_del	bigint	删除行数
n_tup_hot_upd	bigint	HOT更新行数（比如没有更新所需的单独索引）

### 15.3.145 PG\_STAT\_XACT\_USER\_FUNCTIONS

PG\_STAT\_XACT\_USER\_FUNCTIONS视图包含每个跟踪函数的行，显示关于函数执行的统计。

表 15-198 PG\_STAT\_XACT\_USER\_FUNCTIONS 字段

名称	类型	描述
funcid	oid	函数OID
schemaname	name	模式名
funcname	name	函数名
calls	bigint	函数被调用的次数

名称	类型	描述
total_time	double precision	函数的总执行时长
self_time	double precision	当前线程调用函数的总的时长

### 15.3.146 PG\_STAT\_XACT\_USER\_TABLES

PG\_STAT\_XACT\_USER\_TABLES视图显示命名空间中用户表的事务状态信息。

表 15-199 PG\_STAT\_XACT\_USER\_TABLES 字段

名称	类型	描述
relid	oid	表的OID
schemaname	name	此表的模式名
relname	name	表名
seq_scan	bigint	在该表上启动的顺序扫描数
seq_tup_read	bigint	顺序扫描抓取的活跃行数
idx_scan	bigint	在该表上启动的索引扫描数
idx_tup_fetch	bigint	索引扫描抓取的活跃行数
n_tup_ins	bigint	插入行数
n_tup_upd	bigint	更新行数
n_tup_del	bigint	删除行数
n_tup_hot_upd	bigint	HOT更新行数（比如没有更新所需的单独索引）

### 15.3.147 PG\_STATIO\_ALL\_INDEXES

PG\_STATIO\_ALL\_INDEXES视图将包含当前数据库中的每个索引行，显示特定索引的I/O的统计。

表 15-200 PG\_STATIO\_ALL\_INDEXES 字段

名称	类型	描述
relid	oid	此索引表的OID
indexrelid	oid	索引的OID

名称	类型	描述
schemaname	name	索引中模式名
relname	name	索引的表名
indexrelname	name	索引名
idx_blks_read	bigint	从索引中读取的磁盘块数
idx_blks_hit	bigint	索引缓冲区命中数量

### 15.3.148 PG\_STATIO\_ALL\_SEQUENCES

PG\_STATIO\_ALL\_SEQUENCES视图包含当前数据库中每个序列的每一行，显示特定序列关于I/O的统计。

表 15-201 PG\_STATIO\_ALL\_SEQUENCES 字段

名称	类型	描述
relid	oid	序列OID
schemaname	name	序列中模式名
relname	name	序列名
blks_read	bigint	从序列中读取的磁盘块数
blks_hit	bigint	序列缓冲区命中数量

### 15.3.149 PG\_STATIO\_ALL\_TABLES

PG\_STATIO\_ALL\_TABLES视图将包含当前数据库中每个表的一行（包括TOAST表），显示出特定表I/O的统计。

表 15-202 PG\_STATIO\_ALL\_TABLES 字段

名称	类型	描述
relid	oid	表OID
schemaname	name	此表模式名
relname	name	表名
heap_blks_read	bigint	从该表中读取的磁盘块数
heap_blks_hit	bigint	此表缓冲区命中数
idx_blks_read	bigint	从表中所有索引读取的磁盘块数

名称	类型	描述
idx_blks_hit	bigint	表中所有索引命中缓冲区数
toast_blks_read	bigint	此表的TOAST表读取的磁盘块数（如果存在）
toast_blks_hit	bigint	此表的TOAST表命中缓冲区数（如果存在）
tidx_blks_read	bigint	此表的TOAST表索引读取的磁盘块数（如果存在）
tidx_blks_hit	bigint	此表的TOAST表索引命中缓冲区数（如果存在）

### 15.3.150 PG\_STATIO\_SYS\_INDEXES

PG\_STATIO\_SYS\_INDEXES视图显示命名空间中所有系统表索引的IO状态信息。

表 15-203 PG\_STATIO\_SYS\_INDEXES 字段

名称	类型	描述
relid	oid	此索引表的OID
indexrelid	oid	索引的OID
schemaname	name	索引的模式名
relname	name	索引的表名
indexrelname	name	索引名
idx_blks_read	bigint	从索引中读取的磁盘块数
idx_blks_hit	bigint	索引缓冲区命中数量

### 15.3.151 PG\_STATIO\_SYS\_SEQUENCES

PG\_STATIO\_SYS\_SEQUENCES视图显示命名空间中所有系统表为序列的IO状态信息。

表 15-204 PG\_STATIO\_SYS\_SEQUENCES 字段

名称	类型	描述
relid	oid	序列OID
schemaname	name	序列中模式名
relname	name	序列名
blks_read	bigint	从序列中读取的磁盘块数
blks_hit	bigint	序列缓冲区命中数量



### 15.3.152 PG\_STATIO\_SYS\_TABLES

PG\_STATIO\_SYS\_TABLES视图显示命名空间中所有系统表的IO状态信息。

表 15-205 PG\_STATIO\_SYS\_TABLES 字段

名称	类型	描述
relid	oid	表OID
schemaname	name	表模式名
relname	name	表名
heap_blks_read	bigint	从表中读取的磁盘块数
heap_blks_hit	bigint	此表缓冲区命中数
idx_blks_read	bigint	从表中所有索引读取的磁盘块数
idx_blks_hit	bigint	表中所有索引命中缓冲区数
toast_blks_read	bigint	此表的TOAST表读取的磁盘块数（如果存在）
toast_blks_hit	bigint	此表的TOAST表命中缓冲区数（如果存在）
tidx_blks_read	bigint	此表的TOAST表索引读取的磁盘块数（如果存在）
tidx_blks_hit	bigint	此表的TOAST表索引命中缓冲区数（如果存在）

### 15.3.153 PG\_STATIO\_USER\_INDEXES

PG\_STATIO\_USER\_INDEXES视图显示命名空间中所有用户关系表索引的IO状态信息。

表 15-206 PG\_STATIO\_USER\_INDEXES 字段

名称	类型	描述
relid	oid	索引的表的OID
indexrelid	oid	该索引的OID
schemaname	name	该索引的模式名
relname	name	该索引的表名
indexrelname	name	索引名称
idx_blks_read	bigint	从索引中读取的磁盘块数

名称	类型	描述
idx_blks_hit	bigint	索引命中缓冲区数

### 15.3.154 PG\_STATIO\_USER\_SEQUENCES

PG\_STATIO\_USER\_SEQUENCES视图显示命名空间中所有用户关系表类型为序列的IO状态信息。

表 15-207 PG\_STATIO\_USER\_SEQUENCES 字段

名称	类型	描述
relid	oid	序列OID
schemaname	name	序列中模式名
relname	name	序列名
blks_read	bigint	从序列中读取的磁盘块数
blks_hit	bigint	序列中缓存命中数

### 15.3.155 PG\_STATIO\_USER\_TABLES

PG\_STATIO\_USER\_TABLES视图显示命名空间中所有用户关系表的IO状态信息。

表 15-208 PG\_STATIO\_USER\_TABLES 字段

名称	类型	描述
relid	oid	表OID
schemaname	name	该表模式名
relname	name	表名
heap_blks_read	bigint	从该表中读取的磁盘块数
heap_blks_hit	bigint	此表缓冲区命中数
idx_blks_read	bigint	从表中所有索引读取的磁盘块数
idx_blks_hit	bigint	表中所有索引缓冲区命中数
toast_blks_read	bigint	此表的TOAST表读取的磁盘块数（如果存在）
toast_blks_hit	bigint	此表的TOAST表命中缓冲区数（如果存在）
tidx_blks_read	bigint	此表的TOAST表索引读取的磁盘块数（如果存在）

名称	类型	描述
tidx_blks_hit	bigint	此表的TOAST表索引命中缓冲区数（如果存在）

### 15.3.156 PG\_THREAD\_WAIT\_STATUS

通过PG\_THREAD\_WAIT\_STATUS视图可以检测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。

表 15-209 PG\_THREAD\_WAIT\_STATUS 字段

名称	类型	描述
node_name	text	当前节点的名称。
db_name	text	数据库名称。
thread_name	text	线程名称。
query_id	bigint	查询ID，对应debug_query_id。
tid	bigint	当前线程的线程号。
lwtid	integer	当前线程的轻量级线程号。
ptid	integer	streaming线程的父线程。
tlevel	integer	streaming线程的层级。
smpid	integer	并行线程的ID。
wait_status	text	当前线程的等待状态。等待状态的详细信息请参见表 15-210。
wait_event	text	如果wait_status是acquire lock、acquire lwlock、wait io三种类型，此列描述具体的锁、轻量级锁、IO的信息。否则为空。

wait\_status列的等待状态有以下状态。

表 15-210 等待状态列表

wait_status值	含义
none	未等任意事件。
acquire lock	等待加锁，要么加锁成功，要么加锁等待超时。
acquire lwlock	等待获取轻量级锁。
wait io	等待IO完成。

wait_status值	含义
wait cmd	等待完成读取网络通信包。
wait pooler get conn	等待pooler完成获取连接。
wait pooler abort conn	等待pooler完成终止连接。
wait pooler clean conn	等待pooler完成清理连接。
pooler create conn: [nodename], total N	等待pooler建立连接，当前正在与nodename指定节点建立连接，且仍有N个连接等待建立。
get conn	获取到其他节点的连接。
set cmd: [nodename]	在连接上执行SET/RESET/TRANSACTION BLOCK LEVEL PARA SET/SESSION LEVEL PARA SET，当前正在nodename指定节点上执行。
cancel query	取消某连接上正在执行的SQL语句。
stop query	停止某连接上正在执行的查询。
wait node: [nodename](plevel), total N, [phase]	等待接收与某节点的连接上的数据，当前正在等待nodename节点plevel线程的数据，且仍有N个连接的数据待返回。如果状态包含phase信息，则可能的阶段状态有： <ul style="list-style-type: none"> <li>• begin：表示处于事务开始阶段。</li> <li>• commit：表示处于事务提交阶段。</li> <li>• rollback：表示处于事务回滚阶段。</li> </ul>
wait transaction sync: xid	等待xid指定事务同步。
wait wal sync	等待特定LSN的wal log完成到备机的同步。
wait data sync	等待完成数据页到备机的同步。
wait data sync queue	等待把行存的数据页或列存的CU放入同步队列。
flush data: [nodename](plevel), [phase]	等待向网络中nodename指定节点的plevel对应线程发送数据。如果状态包含phase信息，则可能的阶段状态为wait quota，即当前通信流正在等待quota值。
stream get conn: [nodename], total N	初始化stream flow时，等待与nodename节点的consumer对象建立连接，且当前有N个待建连对象。
wait producer ready: [nodename](plevel), total N	初始化stream flow时，等待每个producer都准备好，当前正在等待nodename节点plevel对应线程的producer对象准备好，且仍有N个producer对象处于等待状态。

wait_status值	含义
synchronize quit	steam plan结束时，等待stream线程组内的线程统一退出。
nodegroup destroy	steam plan结束时，等待销毁stream node group。
wait active statement	等待作业执行，正在资源负载管控中。
wait global queue	等待作业执行，正在全局队列排队。
wait respool queue	等待作业执行，正在资源池上排队。
wait ccn queue	等待作业执行，正在中心协调节点(CCN)中排队。
gtm connect	等待与GTM建连。
gtm get gxid	等待从GTM获取事务xid。
gtm get snapshot	等待从GTM获取事务快照snapshot。
gtm begin trans	等待GTM开始事务。
gtm commit trans	等待GTM提交事务。
gtm rollback trans	等待GTM执行事务回滚。
gtm create sequence	等待GTM创建sequence。
gtm alter sequence	等待GTM修改sequence。
gtm get sequence val	等待从GTM获取sequence的下一个值。
gtm set sequence val	等待GTM设置sequence的值。
gtm drop sequence	等待GTM删除sequence。
gtm rename sequece	等待GTM重命名sequence。
analyze: [relname], [phase]	当前正在对表relname执行analyze。如果状态包含phase信息，则为autovacuum，表示是数据库自动开启AutoVacuum线程执行的analyze分析操作。
vacuum: [relname], [phase]	当前正在对表relname执行vacuum。如果状态包含phase信息，则为autovacuum，表示是数据库自动开启AutoVacuum线程执行的vacuum清理操作。
vacuum full: [relname]	当前正在对表relname执行vacuum full清理。
create index	当前正在创建索引。

wait_status值	含义
HashJoin - [ build hash   write file ]	当前是HashJoin算子，主要关注耗时的执行阶段。 <ul style="list-style-type: none"> <li>• build hash：表示当前HashJoin算子正在建立哈希表。</li> <li>• write file：表示当前HashJoin算子正在将数据写入磁盘。</li> </ul>
HashAgg - [ build hash   write file ]	当前是HashAgg算子，主要关注耗时的执行阶段。 <ul style="list-style-type: none"> <li>• build hash：表示当前HashAgg算子正在建立哈希表。</li> <li>• write file：表示当前HashAgg算子正在将数据写入磁盘。</li> </ul>
HashSetop - [build hash   write file ]	当前是HashSetop算子，主要关注耗时的执行阶段。 <ul style="list-style-type: none"> <li>• build hash：表示当前HashSetop算子正在建立哈希表。</li> <li>• write file：表示当前HashSetop算子正在将数据写入磁盘。</li> </ul>
Sort   Sort - write file	当前是Sort算子做排序，write file表示Sort算子正在将数据写入磁盘。
Material   Material - write file	当前是Material算子，write file表示Material算子正在将数据写入磁盘。
wait sync consumer next step	Consumer接收端同步等待下一轮迭代。
wait sync producer next step	Producer发送端同步等待下一轮迭代。
wait agent release	正在释放当前agent(8.1.2及以上版本支持)。
wait stream task	stream线程等待被复用(8.1.2及以上版本支持)。

当wait\_status为acquire lwlock、acquire lock或者wait io时，表示有等待事件。正在等待获取wait\_event列对应类型的轻量级锁、事务锁，或者正在进行IO。

其中，wait\_status值为acquire lwlock（轻量级锁）时对应的wait\_event等待事件类型与描述信息如下。（wait\_event为extension时，表示此时的轻量级锁是动态分配的锁，未被监控。）

表 15-211 轻量级锁等待事件列表

wait_event类型	类型描述
ShmemIndexLock	用于保护共享内存中的主索引哈希表。

wait_event类型	类型描述
OidGenLock	用于避免不同线程产生相同的OID。
XidGenLock	用于避免两个事务获得相同的xid。
ProcArrayLock	用于避免并发访问或修改ProcArray共享数组。
SInvalReadLock	用于避免与清理失效消息并发执行。
SInvalWriteLock	用于避免与其它写失效消息、清理失效消息并发执行。
WALInsertLock	用于避免与其它WAL插入操作并发执行。
WALWriteLock	用于避免并发WAL写盘。
ControlFileLock	用于避免pg_control文件的读写并发、写写并发。
CheckpointLock	用于避免多个checkpoint并发执行。
CLogControlLock	用于避免并发访问或者修改Clog控制数据结构。
MultiXactGenLock	用于串行分配唯一MultiXact id。
MultiXactOffsetControlLock	用于避免对pg_multixact/offset的写写并发和读写并发。
MultiXactMemberControlLock	用于避免对pg_multixact/members的写写并发和读写并发。
RelCacheInitLock	用于失效消息场景对init文件进行操作时加锁。
CheckpointInterCommLock	用于向checkpointer发起文件刷盘请求场景，需要串行的向请求队列插入请求结构。
TwoPhaseStateLock	用于避免并发访问或者修改两阶段信息共享数组。
TablespaceCreateLock	用于确定tablespace是否已经存在。
BtreeVacuumLock	用于防止vacuum清理B-tree中还在使用的页面。
AutovacuumLock	用于串行化访问autovacuum worker数组。
AutovacuumScheduleLock	用于串行化分配需要vacuum的table。
SyncScanLock	用于确定heap扫描时某个relfilenode的起始位置。
NodeTableLock	用于保护存放CN和DN节点信息的共享结构。
PoolerLock	用于保证两个线程不会同时从连接池里取到相同的连接。
RelationMappingLock	用于等待更新系统表到存储位置之间映射的文件。
AsyncCtlLock	用于避免并发访问或者修改共享通知状态。
AsyncQueueLock	用于避免并发访问或者修改共享通知信息队列。

wait_event类型	类型描述
SerializableXactHashLock	用于避免对于可串行事务共享结构的写写并发和读写并发。
SerializableFinishedListLock	用于避免对于已完成可串行事务共享链表的写写并发和读写并发。
SerializablePredicateLockListLock	用于保护对于可串行事务持有的锁链表。
OldSerXidLock	用于保护记录冲突可串行事务的结构。
FileStatLock	用于保护存储统计文件信息的数据结构。
SyncRepLock	用于在主备复制时保护xlog同步信息。
DataSyncRepLock	用于在主备复制时保护数据页同步信息。
CStoreColspaceCacheLock	用于保护列存表的CU空间分配。
CStoreCUCacheSweepLock	用于列存CU Cache循环淘汰。
MetaCacheSweepLock	用于元数据循环淘汰。
DfsConnectorCacheLock	用于保护缓存HDFS连接的句柄的全局哈希表。
dummyServerInfoCacheLock	用于保护缓存加速集群连接信息的全局哈希表。
ExtensionConnectorLibLock	用于初始化ODBC连接场景，在加载与卸载特定动态库时进行加锁。
SearchServerLibLock	用于GPU加速场景初始化加载特定动态库时，对读文件操作进行加锁。
DfsUserLoginLock	用于保护HDFS用户信息的全局链表。
DfsSpaceCacheLock	用于控制HDFS表导入时文件ID单调递增。
LsnXlogChkFileLock	用于串行更新特定结构中记录的主备机的xlog flush位置点。
GTMHostInfoLock	用于避免并发访问或者修改GTM主机信息。
ReplicationSlotAllocationLock	用于主备复制时保护主机端的流复制槽的分配。
ReplicationSlotControlLock	用于主备复制时避免并发更新流复制槽状态。
ResourcePoolHashLock	用于避免并发访问或者修改资源池哈希表。
WorkloadStatHashLock	用于避免并发访问或者修改包含CN侧的SQL请求构成的哈希表。
WorkloadIoStatHashLock	用于避免并发访问或者修改用于统计当前DN的IO信息的哈希表。



wait_event类型	类型描述
WorkloadCGroupHashLock	用于避免并发访问或者修改Cgroup信息构成的哈希表。
OBSGetPathLock	用于避免对obs路径的写写并发和读写并发。
WorkloadUserInfoLock	用于避免并发访问或修改负载管理的用户信息哈希表。
WorkloadRecordLock	用于避免并发访问或修改在内存自适应管理时对CN收到请求构成的哈希表。
WorkloadIOUtilLock	用于保护记录iostat, CPU等负载信息的结构。
WorkloadNodeGroupLock	用于避免并发访问或者修改内存中的nodegroup信息构成的哈希表。
JobShmemLock	用于MPP兼容ORACLE定时任务功能中保护定时读取的全局变量。
OBSRuntimeLock	用于获取环境变量, 如GAUSSHOME。
LLVMDumpIRLock	用于导出动态生成函数所对应的汇编语言。
LLVMParseIRLock	用于在查询开始处从IR文件中编译并解析已写好的IR函数。
RPNNumberLock	用于加速集群的DN对正在执行计划的任务线程的计数。
ClusterRPLock	用于加速集群的CCN中维护的集群负载数据的并发存取控制。
CriticalCacheBuildLock	用于从共享或者本地缓存初始化文件中加载cache的场景。
WaitCountHashLock	用于保护用户语句计数功能场景中的共享结构。
BufMappingLock	用于保护对共享缓冲映射表的操作。
LockMgrLock	用于保护常规锁结构信息。
PredicateLockMgrLock	用于保护可串行事务锁结构信息。
OperatorRealTLock	用于避免并发访问或者修改记录算子级实时数据的全局结构。
OperatorHistLock	用于避免并发访问或者修改记录算子级历史数据的全局结构。
SessionRealTLock	用于避免并发访问或者修改记录query级实时数据的全局结构。
SessionHistLock	用于避免并发访问或者修改记录query级历史数据的全局结构。
CacheSlotMappingLock	用于保护CU Cache全局信息。

wait_event类型	类型描述
BarrierLock	用于保证当前只有一个线程在创建Barrier。

当wait\_status值为wait io时对应的wait\_event等待事件类型与描述信息如下。

表 15-212 IO 等待事件列表

wait_event类型	类型描述
BufFileRead	从临时文件中读取数据到指定buffer。
BufFileWrite	向临时文件中写入指定buffer中的内容。
ControlFileRead	读取pg_control文件。主要在数据库启动、执行checkpoint和主备校验过程中发生。
ControlFileSync	将pg_control文件持久化到磁盘。数据库初始化时发生。
ControlFileSyncUpdate	将pg_control文件持久化到磁盘。主要在数据库启动、执行checkpoint和主备校验过程中发生。
ControlFileWrite	写入pg_control文件。数据库初始化时发生。
ControlFileWriteUpdate	更新pg_control文件。主要在数据库启动、执行checkpoint和主备校验过程中发生。
CopyFileRead	copy文件时读取文件内容。
CopyFileWrite	copy文件时写入文件内容。
DataFileExtend	扩展文件时向文件写入内容。
DataFileFlush	将表数据文件持久化到磁盘
DataFileImmediateSync	将表数据文件立即持久化到磁盘。
DataFilePrefetch	异步读取表数据文件。
DataFileRead	同步读取表数据文件。
DataFileSync	将表数据文件的修改持久化到磁盘。
DataFileTruncate	表数据文件truncate。
DataFileWrite	向表数据文件写入内容。
LockFileAddToDataDirRead	读取"postmaster.pid"文件。
LockFileAddToDataDirSync	将"postmaster.pid"内容持久化到磁盘。
LockFileAddToDataDirWrite	将pid信息写到"postmaster.pid"文件。

wait_event类型	类型描述
LockFileCreateRead	读取LockFile文件"%s.lock"。
LockFileCreateSync	将LockFile文件"%s.lock"内容持久化到磁盘。
LockFileCreateWRITE	将pid信息写到LockFile文件"%s.lock"。
RelationMapRead	读取系统表到存储位置之间的映射文件
RelationMapSync	将系统表到存储位置之间的映射文件持久化到磁盘。
RelationMapWrite	写入系统表到存储位置之间的映射文件。
ReplicationSlotRead	读取流复制槽文件。重新启动时发生。
ReplicationSlotRestoreSync	将流复制槽文件持久化到磁盘。重新启动时发生。
ReplicationSlotSync	checkpoint时将流复制槽临时文件持久化到磁盘。
ReplicationSlotWrite	checkpoint时写流复制槽临时文件。
SLRUFlushSync	将pg_clog、pg_subtrans和pg_multixact文件持久化到磁盘。主要在执行checkpoint和数据库停机时发生。
SLRURead	读取pg_clog、pg_subtrans和pg_multixact文件。
SLRUSync	将脏页写入文件pg_clog、pg_subtrans和pg_multixact并持久化到磁盘。主要在执行checkpoint和数据库停机时发生。
SLRUWrite	写入pg_clog、pg_subtrans和pg_multixact文件。
TimelineHistoryRead	读取timeline history文件。在数据库启动时发生。
TimelineHistorySync	将timeline history文件持久化到磁盘。在数据库启动时发生。
TimelineHistoryWrite	写入timeline history文件。在数据库启动时发生。
TwophaseFileRead	读取pg_twophase文件。在两阶段事务提交、两阶段事务恢复时发生。
TwophaseFileSync	将pg_twophase文件持久化到磁盘。在两阶段事务提交、两阶段事务恢复时发生。
TwophaseFileWrite	写入pg_twophase文件。在两阶段事务提交、两阶段事务恢复时发生。
WALBootstrapSync	将初始化的WAL文件持久化到磁盘。在数据库初始化发生。
WALBootstrapWrite	写入初始化的WAL文件。在数据库初始化发生。
WALCopyRead	读取已存在的WAL文件并进行复制时产生的读操作。在执行归档恢复完后发生。

wait_event类型	类型描述
WALCopySync	将复制的WAL文件持久化到磁盘。在执行归档恢复完后发生。
WALCopyWrite	读取已存在WAL文件并进行复制时产生的写操作。在执行归档恢复完后发生。
WALInitSync	将新初始化的WAL文件持久化磁盘。在日志回收或写日志时发生。
WALInitWrite	将新创建的WAL文件初始化为0。在日志回收或写日志时发生。
WALRead	从xlog日志读取数据。两阶段文件redo相关的操作产生。
WALSyncMethodAssign	将当前打开的所有WAL文件持久化到磁盘。
WALWrite	写入WAL文件。

当wait\_status值为acquire lock（事务锁）时对应的wait\_event等待事件类型与描述信息如下。

表 15-213 事务锁等待事件列表

wait_event类型	类型描述
relation	对表加锁
extend	对表扩展空间时加锁
partition	对分区表加锁
partition_seq	对分区表的分区加锁
page	对表页面加锁
tuple	对页面上的tuple加锁
transactionid	对事务ID加锁
virtualxid	对虚拟事务ID加锁
object	加对象锁
cstore_freespace	对列存空闲空间加锁
userlock	加用户锁
advisory	加advisory锁

## 15.3.157 PG\_TABLES

PG\_TABLES视图提供了对数据库中每个表访问的有用信息。

表 15-214 PG\_TABLES 字段

名称	类型	引用	描述
schemaname	name	<a href="#">PG_NAMESPACE</a> .nspname	包含表的模式名。
tablename	name	<a href="#">PG_CLASS</a> .relname	表名。
tableowner	name	pg_get_userbyid( <a href="#">PG_CLASS</a> .relowner)	表的所有者。
tablespace	name	<a href="#">PG_TABLESPACE</a> .spcname	包含表的表空间，默认为 NULL。
hasindexes	boolean	<a href="#">PG_CLASS</a> .relhasindex	如果表上有索引（或者最近拥有）则为 TRUE，否则为 FALSE。
hasrules	boolean	<a href="#">PG_CLASS</a> .relhasrules	如果表上有规则，则为 TRUE，否则为 FALSE。
hasindexes	boolean	<a href="#">PG_CLASS</a> .RELHASTRIGGERS	如果表上有触发器，则为 TRUE，否则为 FALSE。
tablecreator	name	pg_get_userbyid( <a href="#">PG_OBJECT</a> .creator)	表创建用户，如创建用户已删除，则返回空。
created	timestamp with time zone	<a href="#">PG_OBJECT</a> .ctime	表创建时间。
last_ddl_time	timestamp with time zone	<a href="#">PG_OBJECT</a> .mtime	表最后修改时间。

### 15.3.158 PG\_TDE\_INFO

PG\_TDE\_INFO 视图提供了当前集群加密信息。

表 15-215 PG\_TDE\_INFO 字段

名称	类型	描述
is_encrypt	text	是否加密集群 <ul style="list-style-type: none"> <li>● f: 非加密集群</li> <li>● t: 加密集群</li> </ul>
g_tde_algo	text	加密算法 <ul style="list-style-type: none"> <li>● SM4-CTR-128</li> <li>● AES-CTR-128</li> </ul>

名称	类型	描述
remain	text	保留字段

## 应用示例

查看当前集群是否加密/查看当前集群的加密算法：

```
SELECT * FROM PG_TDE_INFO;
is_encrypt | g_tde_algo | remain
-----+-----+-----
f          | AES-CTR-128 | remain
(1 row)
```

### 15.3.159 PG\_TIMEZONE\_ABBREVS

PG\_TIMEZONE\_ABBREVS视图提供了输入例程能够识别的所有时区缩写。

表 15-216 PG\_TIMEZONE\_ABBREVS 字段

名称	类型	描述
abbrev	text	时区缩写
utc_offset	interval	相对于UTC的偏移量
is_dst	boolean	如果这是一个夏时制时区缩写则为TRUE，否则为FALSE

### 15.3.160 PG\_TIMEZONE\_NAMES

PG\_TIMEZONE\_NAMES视图提供了显示了所有能够被SET TIMEZONE识别的时区名及其缩写、UTC偏移量、是否夏时制。

表 15-217 PG\_TIMEZONE\_NAMES 字段

名称	类型	描述
name	text	时区名
abbrev	text	时区名缩写
utc_offset	interval	相对于UTC的偏移量
is_dst	boolean	如果当前正处于夏令时范围则为TRUE，否则为FALSE

### 15.3.161 PG\_TOTAL\_MEMORY\_DETAIL

PG\_TOTAL\_MEMORY\_DETAIL视图显示某个数据库节点内存使用情况。

表 15-218 PG\_TOTAL\_MEMORY\_DETAIL 字段

名称	类型	描述
nodename	text	节点名称

名称	类型	描述
memorytype	text	<p>内存的名称，包括以下几种：</p> <ul style="list-style-type: none"> <li>• max_process_memory: GaussDB(DWS)集群实例所占用的内存大小。</li> <li>• process_used_memory: GaussDB(DWS)进程所使用的内存大小。</li> <li>• max_dynamic_memory: 最大动态内存。</li> <li>• dynamic_used_memory: 已使用的动态内存。</li> <li>• dynamic_peak_memory: 内存的动态峰值。</li> <li>• dynamic_used_shrctx: 最大动态共享内存上下文。</li> <li>• dynamic_peak_shrctx: 共享内存上下文的动态峰值。</li> <li>• max_shared_memory: 最大共享内存。</li> <li>• shared_used_memory: 已使用的共享内存。</li> <li>• max_cstore_memory: 列存所允许使用的最大内存。</li> <li>• cstore_used_memory: 列存已使用的内存大小。</li> <li>• max_sctpcomm_memory: 通信库所允许使用的最大内存。</li> <li>• sctpcomm_used_memory: 通信库已使用的内存大小。</li> <li>• sctpcomm_peak_memory: 通信库的内存峰值。</li> <li>• max_topsql_memory: TopSQL记录历史作业监控信息允许使用的最大内存。</li> <li>• topsql_used_memory: TopSQL记录历史作业监控信息已使用的内存大小。</li> <li>• topsql_peak_memory: TopSQL记录历史作业监控信息的内存峰值。</li> <li>• other_used_memory: 其他已使用的内存大小。</li> <li>• gpu_max_dynamic_memory: GPU内存最大值。</li> <li>• gpu_dynamic_used_memory: 当前GPU可用内存和当前临时GPU内存之和。</li> <li>• gpu_dynamic_peak_memory: GPU内存使用的最大内存。</li> <li>• pooler_conn_memory: pooler连接占用内存大小。</li> </ul>



名称	类型	描述
		<ul style="list-style-type: none"> <li>pooler_freeconn_memory: pooler空闲连接占用的内存大小。</li> <li>storage_compress_memory: 列存压缩和解压缩使用的内存大小。</li> <li>udf_reserved_memory: 为UDF Worker进程预留的内存大小。</li> <li>mmap_used_memory: mmap使用的内存大小。</li> </ul>
memorybytes	integer	内存使用的大小, 单位MB

### 15.3.162 PG\_TOTAL\_SCHEMA\_INFO

PG\_TOTAL\_SCHEMA\_INFO视图显示各个数据库下所有Schema的存储资源使用情况。此视图在参数use\_workload\_manager为on时才有效。

名称	类型	描述
schemaid	oid	Schema的OID
schemaname	text	Schema名称
databaseid	oid	数据库的OID
databasename	name	数据库名称
usedspace	bigint	该Schema已使用的永久表存储空间大小, 单位Byte
permspace	bigint	该Schema的永久表存储空间上限值, 单位Byte

### 15.3.163 PG\_TOTAL\_USER\_RESOURCE\_INFO

PG\_TOTAL\_USER\_RESOURCE\_INFO视图显示所有用户资源使用情况, 需要使用管理员用户进行查询。此视图在参数use\_workload\_manager为on时才有效。

表 15-219 PG\_TOTAL\_USER\_RESOURCE\_INFO 字段

名称	类型	描述
username	name	用户名。
used_memory	integer	正在使用的内存大小, 单位MB。

名称	类型	描述
total_memory	integer	可以使用的内存大小，单位MB。值为0表示未限制最大可用内存，其限制取决于数据库最大可用内存。
used_cpu	double precision	正在使用的CPU核数（仅统计非默认资源池上复杂作业的CPU使用情况，且该值为相关控制组的CPU使用统计值）。
total_cpu	integer	在该机器节点上，用户关联控制组的CPU核数总和。
used_space	bigint	已使用的永久表存储空间大小，单位KB。
total_space	bigint	可使用的永久表存储空间大小，单位KB，值为-1表示未限制永久表存储空间。
used_temp_space	bigint	已使用的临时表存储空间大小，单位KB。
total_temp_space	bigint	可使用的临时表存储空间大小，单位KB，值为-1表示未限制临时表存储空间。
used_spill_space	bigint	已使用的算子落盘空间大小，单位KB。
total_spill_space	bigint	可使用的算子落盘空间大小，单位KB，值为-1表示未限制算子落盘空间。
read_kbytes	bigint	CN：过去5秒内，该用户在所有DN上复杂作业read的字节总数。(单位KB) DN：实例启动至当前时间为止，该用户复杂作业read的字节总数。(单位KB)
write_kbytes	bigint	CN：过去5秒内，该用户在所有DN上复杂作业write的字节总数。(单位KB) DN：实例启动至当前时间为止，该用户复杂作业write的字节总数。(单位KB)
read_counts	bigint	CN：过去5秒内，该用户在所有DN上复杂作业read的次数之和。(单位次) DN：实例启动至当前时间为止，该用户复杂作业read的次数之和。(单位次)
write_counts	bigint	CN：过去5秒内，该用户在所有DN上复杂作业write的次数之和。(单位次) DN：实例启动至当前时间为止，该用户复杂作业write的次数之和。(单位次)
read_speed	double precision	CN：过去5秒内，该用户在单个DN上复杂作业read平均速率。(单位KB/s) DN：过去5秒内，该用户在该DN上复杂作业read平均速率。(单位KB/s)

名称	类型	描述
write_speed	double precision	CN: 过去5秒内, 该用户在单个DN上复杂作业write平均速率。(单位KB/s) DN: 过去5秒内, 该用户在该DN上复杂作业write平均速率。(单位KB/s)

## 15.3.164 PG\_USER

PG\_USER视图提供了访问数据库用户的信息。

表 15-220 PG\_USER 字段

名称	类型	描述
username	name	用户名。
usesysid	oid	此用户的ID。
usecreatedb	boolean	用户是否可以创建数据库。
usesuper	boolean	用户是否是拥有最高权限的初始系统管理员。
usecatupd	boolean	用户是否可以直接更新系统表。只有usesysid=10的初始系统管理员拥有此权限。其他用户无法获得此权限。
userepl	boolean	用户是否可以复制数据流。
passwd	text	密文存储后的用户口令, 始终为*****。
valbegin	timestamp with time zone	帐户的有效开始时间; 如果没有设置有效开始时间, 则为NULL。
valuntil	timestamp with time zone	帐户的有效结束时间; 如果没有设置有效结束时间, 则为NULL。
respool	name	用户所在的资源池。
parentid	oid	父用户OID
spacelimit	text	永久表存储空间限额。
tempspacelimit	text	临时表存储空间限额。
spillspacelimit	text	算子落盘空间限额。
useconfig	text[]	运行时配置参数的会话缺省。
nodegroup	name	用户关联的逻辑集群名字, 如果该用户没有管理逻辑集群, 则该字段为空。

## 应用示例

查看当前数据库用户列表：

```
SELECT username FROM pg_user;
username
-----
dbadmin
u1
u2
u3
(4 rows)
```

### 15.3.165 PG\_USER\_MAPPINGS

PG\_USER\_MAPPINGS视图提供访问关于用户映射的信息的接口。

这个视图只是一个**PG\_USER\_MAPPING**的可读部分的视图化表现，如果用户无权使用它则查询此表时，有些选项字段会显示为空。

表 15-221 PG\_USER\_MAPPINGS 字段

名字	类型	引用	描述
umid	oid	<a href="#">PG_USER_MAPPING</a> .oid	用户映射的OID。
srvid	oid	<a href="#">PG_FOREIGN_SERVER</a> .oid	包含这个映射的外部服务器的OID。
srvname	name	<a href="#">PG_FOREIGN_SERVER</a> .srvname	外部服务器的名字。
umuser	oid	<a href="#">PG_AUTHID</a> .oid	被映射的本地角色的OID，如果用户映射是公共的则为0。
username	name	-	被映射的本地用户的名字。
umoptions	text[]	-	如果当前用户是外部服务器的所有者，则为用户映射指定选项，使用“keyword=value”字符串，否则为null。

### 15.3.166 PG\_VIEWS

PG\_VIEWS视图提供访问数据库中每个视图的有用信息。

表 15-222 PG\_VIEWS 字段

名称	类型	引用	描述
schemaname	name	<a href="#">PG_NAMESPACE</a> .nspname	包含视图的模式名

名称	类型	引用	描述
viewname	name	<b>PG_CLASS</b> .relname	视图的名称
viewowner	name	<b>PG_AUTHID</b> .Erolname	视图的所有者
definition	text	-	视图的定义

## 应用示例

查询指定模式下的所有视图：

```
select * from pg_views where schemaname = 'myschema';
schemaname | viewname | viewowner | definition
-----+-----+-----+-----
myschema  | myview  | dbadmin  | SELECT * FROM pg_tablespace WHERE (pg_tablespace.spcname =
'pg_default'::name);
myschema  | v1      | dbadmin  | SELECT * FROM t1 WHERE (t1.c1 > 200);
(2 rows)
```

## 15.3.167 PG\_WLM\_STATISTICS

PG\_WLM\_STATISTICS视图显示作业结束后或已被处理异常后的负载管理相关信息。该视图8.1.2版本中已废弃。

表 15-223 PG\_WLM\_STATISTICS 字段

名称	类型	描述
statement	text	执行了异常处理的语句。
block_time	bigint	语句执行前的阻塞时间。
elapsed_time	bigint	语句的实际执行时间。
total_cpu_time	bigint	语句执行异常处理时DN上CPU使用的总时间。
qualification_time	bigint	语句检查倾斜率的时间周期。
cpu_skew_percent	integer	语句在执行异常处理时DN上CPU使用的倾斜率。
control_group	text	语句执行异常处理时所使用的Cgroups。
status	text	语句执行异常处理后的状态，包括： <ul style="list-style-type: none"> <li>pending：执行前预备状态。</li> <li>running：执行进行状态。</li> <li>finished：执行正常结束。</li> <li>abort：执行异常终止。</li> </ul>

名称	类型	描述
action	text	语句执行的异常处理动作，包括： <ul style="list-style-type: none"><li>• abort：执行终止操作。</li><li>• adjust：执行Cgroups调整操作，目前只有降级操作。</li><li>• finish：正常结束。</li></ul>
queryid	bigint	语句执行所使用的内部query_id。
threadid	bigint	后端线程ID。

### 15.3.168 PGXC\_BULKLOAD\_PROGRESS

PGXC\_BULKLOAD\_PROGRESS显示导入业务的执行进度，仅支持GDS普通文件导入业务。需要有系统管理员权限才可以访问此视图

表 15-224 PGXC\_BULKLOAD\_PROGRESS 字段

名称	类型	描述
session_id	bigint	GDS的会话ID。
query_id	bigint	查询ID，对应debug_query_id。
query	text	查询语句。
progress	text	进度百分比。

### 15.3.169 PGXC\_BULKLOAD\_STATISTICS

通过CN查看PGXC\_BULKLOAD\_STATISTICS视图，可获取GDS、COPY、\COPY等业务执行过程中的实时统计信息。该视图汇总当前集群上各个节点正在执行的导入/导出类业务的实时执行情况，从而可以监控导入导出类业务的实时进度，辅助进行性能问题排查。

PGXC\_BULKLOAD\_STATISTICS视图与PG\_BULKLOAD\_STATISTICS视图列定义完全相同。这是由于PGXC\_BULKLOAD\_STATISTICS视图本质是到集群中各个节点上查询PG\_BULKLOAD\_STATISTICS视图汇总的结果。

需要有系统管理员权限才可以访问此视图。

表 15-225 PGXC\_BULKLOAD\_STATISTICS 字段

名称	类型	描述
node_name	text	节点名称。
db_name	text	数据库名称。

query_id	bigint	查询ID，对应debug_query_id。
tid	bigint	当前线程的线程号。
lwtid	integer	当前线程的轻量级线程号。
session_id	bigint	GDS的会话ID。
direction	text	业务类型，取值包括：gds to file、gds from file、gds to pipe、gds from pipe、copy from、copy to。
query	text	查询语句。
address	text	当前导入导出外表的location。
query_start	timestamp with time zone	导入/导出开始时间。
total_bytes	bigint	待处理数据的总大小。 仅GDS普通文件导入时，且该行记录来自CN节点才会显示，否则为空。
phase	text	当前阶段，包括：INITIALIZING、TRANSFER_DATA、RELEASE_RESOURCE。
done_lines	bigint	已传输行数。
done_bytes	bigint	已传输字节数。

### 15.3.170 PGXC\_COLUMN\_TABLE\_IO\_STAT

PGXC\_COLUMN\_TABLE\_IO\_STAT视图提供集群所有CN和DN节点上当前数据库所有列存表的IO统计数据。除在每一行前面增加name类型的nodename字段外，其余字段的名称、类型和顺序与GS\_COLUMN\_TABLE\_IO\_STAT视图相同，具体的字段请参考[GS\\_COLUMN\\_TABLE\\_IO\\_STAT](#)。

### 15.3.171 PGXC\_COMM\_CLIENT\_INFO

PGXC\_COMM\_CLIENT\_INFO视图存储所有节点客户端连接信息（DN上查询该视图显示CN连接DN的信息）。

表 15-226 PGXC\_COMM\_CLIENT\_INFO 字段

名称	类型	描述
node_name	text	当前节点的名称。
app	text	客户端应用名。
tid	bigint	当前线程的线程号。
lwtid	integer	当前线程的轻量级线程号。

名称	类型	描述
query_id	bigint	查询ID，对应debug_query_id。
socket	integer	如果是物理连接，展示socket。
remote_ip	text	对端节点IP。
remote_port	text	对端节点port。
logic_id	integer	如果是逻辑连接，展示sid，显示-1时表示当前连接是物理连接。

### 15.3.172 PGXC\_COMM\_DELAY

PGXC\_COMM\_DELAY视图展示所有DN的通信库时延状态。

表 15-227 PGXC\_COMM\_DELAY 字段

名称	类型	描述
node_name	text	节点名称
remote_name	text	连接对端节点的对端时延最大的节点名称
remote_host	text	连接对端IP的对端地址
stream_num	integer	当前物理连接使用的stream逻辑连接数量
min_delay	integer	当前物理连接探测到的最小时延，单位微秒
average	integer	当前物理连接探测时延的平均值，单位微秒
max_delay	integer	当前物理连接探测到的最大时延，单位微秒 <b>说明</b> 取值为-1，表示时延探测超时失败，请重新建立节点间连接后再执行查询。

### 15.3.173 PGXC\_COMM\_RECV\_STREAM

PGXC\_COMM\_RECV\_STREAM视图展示所有DN上的通信库接收流状态。

表 15-228 PGXC\_COMM\_RECV\_STREAM 字段

名称	类型	描述
node_name	text	节点名称。
local_tid	bigint	使用此通信流的线程ID。
remote_name	text	连接对端节点名称。



名称	类型	描述
remote_tid	bigint	连接对端线程ID。
idx	integer	通信对端DN在本DN内的标识编号。
sid	integer	通信流在物理连接中的标识编号。
tcp_sock	integer	通信流所使用的tcp通信socket。
state	text	通信流当前的状态： <ul style="list-style-type: none"> <li>UNKNOWN：表示当前逻辑连接状态未知。</li> <li>READY：表示逻辑连接已就绪。</li> <li>RUN：表示逻辑连接接收报文正常。</li> <li>HOLD：表示逻辑连接接收报文等待中。</li> <li>CLOSED：表示关闭逻辑连接。</li> <li>TO_CLOSED：表示将会关闭逻辑连接。</li> <li>WRITING：表示正在写入数据。</li> </ul>
query_id	bigint	通信流对应的debug_query_id编号。
pn_id	integer	通信流所执行查询的plan_node_id编号。
send_smp	integer	通信流所执行查询send端的smpid编号。
recv_smp	integer	通信流所执行查询recv端的smpid编号。
recv_bytes	bigint	通信流接收的数据总量，单位Byte。
time	bigint	通信流当前生命周期使用时长，单位ms。
speed	bigint	通信流的平均接收速率，单位Byte/s。
quota	bigint	通信流当前的通信配额值，单位Byte。
buff_usize	bigint	通信流当前缓存的数据大小，单位Byte。

### 15.3.174 PGXC\_COMM\_SEND\_STREAM

PGXC\_COMM\_SEND\_STREAM视图展示所有DN上的通信库发送流状态。

表 15-229 PGXC\_COMM\_SEND\_STREAM 字段

名称	类型	描述
node_name	text	节点名称。
local_tid	bigint	使用此通信流的线程ID。
remote_name	text	连接对端节点名称。
remote_tid	bigint	连接对端线程ID。

名称	类型	描述
idx	integer	通信对端DN在本DN内的标识编号。
sid	integer	通信流在物理连接中的标识编号。
tcp_sock	integer	通信流所使用的tcp通信socket。
state	text	通信流当前的状态： <ul style="list-style-type: none"> <li>UNKNOWN：表示当前逻辑连接状态未知。</li> <li>READY：表示逻辑连接已就绪。</li> <li>RUN：表示逻辑连接发送报文正常。</li> <li>HOLD：表示逻辑连接发送报文等待中。</li> <li>CLOSED：表示关闭逻辑连接。</li> <li>TO_CLOSED：表示将会关闭逻辑连接。</li> <li>WRITING：表示正在写入数据。</li> </ul>
query_id	bigint	通信流对应的debug_query_id编号。
pn_id	integer	通信流所执行查询的plan_node_id编号。
send_smp	integer	通信流所执行查询send端的smpid编号。
recv_smp	integer	通信流所执行查询recv端的smpid编号。
send_bytes	bigint	通信流发送的数据总量，单位Byte。
time	bigint	通信流当前生命周期使用时长，单位ms。
speed	bigint	通信流的平均发送速率，单位Byte/s。
quota	bigint	通信流当前的通信配额值，单位Byte。
wait_quota	bigint	通信流等待quota值产生的额外时间开销，单位ms。

### 15.3.175 PGXC\_COMM\_STATUS

PGXC\_COMM\_STATUS视图展示所有DN的通信库状态。

表 15-230 PGXC\_COMM\_STATUS 字段

名称	类型	描述
node_name	text	节点名称。
rxpck/s	integer	节点通信库接收速率，单位Byte/s。
txpck/s	integer	节点通信库发送速率，单位Byte/s。
rxkB/s	bigint	节点通信库接收速率，单位KByte/s。
txkB/s	bigint	节点通信库发送速率，单位KByte/s。

名称	类型	描述
buffer	bigint	cmailbox的buffer大小。
memKB(libcomm)	bigint	libcomm进程通信内存大小，单位KByte。
memKB(libpq)	bigint	libpq进程通信内存大小，单位KByte。
%USED(PM)	integer	postmaster线程实时使用率。
%USED (sflow)	integer	gs_sender_flow_controller线程实时使用率。
%USED (rflow)	integer	gs_receiver_flow_controller线程实时使用率。
%USED (rloop)	integer	多个gs_receivers_loop线程中最高的实时使用率。
stream	integer	当前使用的逻辑连接总数。

### 15.3.176 PGXC\_COMM\_QUERY\_SPEED

PGXC\_COMM\_QUERY\_SPEED视图展示所有节点上所有query对应的流量信息。

表 15-231 PGXC\_COMM\_QUERY\_SPEED 字段

名称	类型	描述
node_name	text	节点名称。
query_id	bigint	通信流对应的debug_query_id编号。
rxkB/s	bigint	查询对应的通信流接收速率，单位Byte/s。
txkB/s	bigint	查询对应的通信流发送速率，单位Byte/s。
rxkB	bigint	查询对应的通信流接收数据总量，单位Byte。
txkB	bigint	查询对应的通信流发送数据总量，单位Byte。
rxpck/s	bigint	查询对应的通信包接收速率，单位 packages/s。
txpck/s	bigint	查询对应的通信包发送速率，单位 packages/s。
rxpck	bigint	查询对应的通信包接收总量，单位packages。
txpck	bigint	查询对应的通信包发送总量，单位packages。

### 15.3.177 PGXC\_DEADLOCK

PGXC\_DEADLOCK视图获取导致分布式死锁产生的锁等待信息。

目前，PGXC\_DEADLOCK视图只收集locktype为relation、partition、page、tuple和transactionid的锁等待信息。

表 15-232 PGXC\_DEADLOCK 字段

名称	类型	描述
locktype	text	被锁定对象的类型。
nodename	name	被锁定对象的节点的名称。
dbname	name	被锁定对象的数据库的名称。如果被锁定对象是事务，则为NULL。
nspname	name	被锁定对象的命名空间的名称。
relname	name	被锁定对象对应的关系的名称。如果被锁定对象既不是关系，也不是关系的一部分，则为NULL。
partname	name	被锁定对象对应的分区的名称。如果被锁定对象不是分区，则为NULL。
page	integer	被锁定对象对应的页面的编号。如果被锁定对象既不是页面，也不是元组，则为NULL。
tuple	smallint	被锁定对象对应的元组的编号。如果被锁定对象不是元组，则为NULL。
transactionid	xid	被锁定对象对应的事务的ID。如果被锁定对象不是事务，则为NULL。
waitusername	name	等待锁的用户的名称。
waitgxid	xid	等待锁的事务的ID。
waitxactstart	timestamp with time zone	等待锁的事务的开始时间。
waitqueryid	bigint	等待锁的线程的最新查询ID。
waitquery	text	等待锁的线程的最新查询语句。
waitpid	bigint	等待锁的线程的ID。
waitmode	text	等待的锁的级别。
holdusername	name	持有锁的用户的名称。
holdgxid	xid	持有锁的事务的ID。
holdxactstart	timestamp with time zone	持有锁的事务的开始时间。
holdqueryid	bigint	持有锁的线程的最新查询ID。
holdquery	text	持有锁的线程的最新查询语句。
holdpid	bigint	持有锁的线程的ID。
holdmode	text	持有的锁的级别。

## 15.3.178 PGXC\_GET\_STAT\_ALL\_TABLES

PGXC\_GET\_STAT\_ALL\_TABLES视图获取各表的插入、更新、删除以及脏页率信息。

对于高脏页率的系统表，建议在确认当前没有用户操作该系统表时，再执行VACUUM FULL。

建议对脏页率超过80%的非系统表执行VACUUM FULL，用户也可根据业务场景自行选择是否执行VACUUM FULL。

### 说明

8.2.0.100及以上集群版本，查询脏页率推荐使用PGXC\_STAT\_TABLE\_DIRTY。

表 15-233 PGXC\_GET\_STAT\_ALL\_TABLES 字段

名称	类型	描述
relid	oid	表的OID
relname	name	表名
schemaname	name	表的模式名
n_tup_ins	numeric	插入的元组条数
n_tup_upd	numeric	更新的元组条数
n_tup_del	numeric	删除的元组条数
n_live_tup	numeric	live元组的条数
n_dead_tup	numeric	dead元组的条数
dirty_page_rate	numeric(5,2)	表的脏页率信息(%)

同时GaussDB(DWS)提供了函数pgxc\_get\_stat\_dirty\_tables(int dirty\_percent, int n\_tuples)和pgxc\_get\_stat\_dirty\_tables(int dirty\_percent, int n\_tuples, text schema)可以快速筛选出脏页率大于dirty\_percent，dead元组数大于n\_tuples，模式名是schema的表。

详细内容可参考[其他函数](#)章节。

## 应用示例

使用PGXC\_GET\_STAT\_ALL\_TABLES视图查询数据库内脏页率大于30%的表：

```
select * from PGXC_GET_STAT_ALL_TABLES where dirty_page_rate>30;
relid | relname | schemaname | n_tup_ins | n_tup_upd | n_tup_del | n_live_tup | n_dead_tup |
dirty_page_rate
-----+-----+-----+-----+-----+-----+-----+-----+
2840 | pg_toast_2619 | pg_toast | 7415 | 0 | 7415 | 0 | 291 | 88.00
9001 | pgxc_class | pg_catalog | 56331 | 3 | 56285 | 54 | 143 | 72.59
53860 | reason | dbadmin | 9 | 19 | 0 | 9 | 19 | 67.86
9025 | pg_object | pg_catalog | 112858 | 1179707 | 112619 | 246 | 429 | 63.56
```

```

9015 | pgxc_node       | pg_catalog | 15 | 24 | 0 | 15 | 24 | 61.54
2606 | pg_constraint   | pg_catalog | 78 | 0 | 42 | 36 | 42 | 53.85
1260 | pg_authid       | pg_catalog | 6 | 6 | 0 | 6 | 6 | 50.00
(7 rows)

```

也可以使用pgxc\_get\_stat\_dirty\_tables函数查询数据库内脏页率大于10%，脏数据行数大于1000行的表：

```

select a.schemaname,a.relname,pg_size_pretty(pg_table_size(b.oid)),a.dirty_page_rate from
pgxc_get_stat_dirty_tables(10,1000) a,pg_catalog.pg_class b where a.relname = b.relname order by
pg_table_size(b.oid) desc;
schemaname | relname | pg_size_pretty | dirty_page_rate
-----+-----+-----+-----
pg_catalog | pg_attribute | 2792 KB | 12.09
pg_catalog | pg_class | 568 KB | 15.36
pg_catalog | pg_type | 368 KB | 12.17
(3 rows)

```

### 15.3.179 PGXC\_GET\_STAT\_ALL\_PARTITIONS

PGXC\_GET\_STAT\_ALL\_PARTITIONS视图获取各分区表分区的插入、更新、删除以及脏页率信息。

该视图的统计信息依赖于ANALYZE，为获取最准确的信息请先对分区表进行ANALYZE。

#### 📖 说明

8.2.0.100及以上集群版本，查询脏页率推荐使用[PGXC\\_STAT\\_TABLE\\_DIRTY](#)。

**表 15-234** PGXC\_GET\_STAT\_ALL\_PARTITIONS 字段

名称	类型	描述
relid	oid	表的oid。
partid	oid	分区的oid
schemaname	name	表schemaname。
relname	name	表名。
partname	name	分区名。
n_tup_ins	numeric	插入的元组条数。
n_tup_upd	numeric	更新的元组条数。
n_tup_del	numeric	删除的元组条数。
n_live_tup	numeric	存活元组的条数。
n_dead_tup	numeric	死亡元组的条数。
page_dirty_rate	numeric(5,2)	表的脏页率信息(%)。

## 15.3.180 PGXC\_GET\_TABLE\_SKEWNESS

PGXC\_GET\_TABLE\_SKEWNESS视图展示当前库中表的数据分布倾斜情况。需要有系统管理员权限或预置角色gs\_role\_read\_all\_stats权限才可以访问此视图。

表 15-235 PGXC\_GET\_TABLE\_SKEWNESS 字段

名称	类型	描述
schemaname	name	表所在的模式名。
tablename	name	表名。
totalsize	numeric	表的总大小，单位Byte。
avgsize	numeric(1000, 0)	表大小平均值(totalsize/DN个数，该值为平均分布的理想情况下，表在各DN占用空间大小)。
maxratio	numeric(4,3)	单DN表大小最大值占比（表在各DN占用空间的最大值/totalsize）。
minratio	numeric(4,3)	单DN表大小最小值占比（表在各DN占用空间的最小值/totalsize）。
skewsize	bigint	表分布倾斜值（单DN表大小最大值 - 单DN表大小最小值）。
skewratio	numeric(4,3)	表分布倾斜率（skewsize/totalsize）。
skewstddev	numeric(1000, 0)	表分布标准方差（在表大小一定的情况下，该值越大表明表的整体分布情况越倾斜）。

## 15.3.181 PGXC\_GTM\_SNAPSHOT\_STATUS

PGXC\_GTM\_SNAPSHOT\_STATUS视图用于查看当前GTM上事务信息。

表 15-236 PGXC\_GTM\_SNAPSHOT\_STATUS 字段

名称	类型	描述
xmin	xid	仍在运行的最小事务号。
xmax	xid	已完成的事务号最大的事务的下一个事务号。
csn	integer	待提交事务的序列号。
oldestxmin	xid	当前最早的活跃事务在其取快照时，所有运行事务号最小的事务。
xcnt	integer	当前活跃的事务个数。

名称	类型	描述
running_xids	text	当前活跃的事务号。

### 15.3.182 PGXC\_INSTANCE\_TIME

PGXC\_INSTANCE\_TIME视图显示集群中各节点上进程的运行时间信息及各执行阶段所消耗时间，除新增node\_name（节点名称）字段外，其余字段内容和PV\_INSTANCE\_TIME视图相同。需要有系统管理员权限或预置角色gs\_role\_read\_all\_stats权限才可以访问此视图。

### 15.3.183 PGXC\_LOCKWAIT\_DETAIL

PGXC\_LOCKWAIT\_DETAIL视图显示集群中每个节点中锁等待链详细信息。如果节点中有多级的锁等待关系，会依次将整个锁等待链按照等待顺序显示出来。

该视图仅8.1.3.200及以上集群版本支持。

表 15-237 PGXC\_LOCKWAIT\_DETAIL 字段

名称	类型	描述
level	integer	锁等待链中的层级，以1开始，每显示一层等待关系level会加1。
node_name	name	节点名称，对应pgxc_node表中的node_name列。
lock_wait_hierarchy	text	锁等待链，以节点名称：进程号->等待进程号->等待进程号->...
lock_type	text	被锁定对象的类型。
database	oid	被锁定对象所在数据库的oid。
relation	oid	被锁定对象关系的oid。
page	integer	关系内部的页面编号。
tuple	smallint	页面的行编号。
virtual_xid	text	事务的虚拟id。
transaction_id	xid	事务id。
class_id	oid	包含该对象的系统表的oid。
obj_id	oid	对象在其系统表内的oid。
obj_subid	smallint	对于表的列字段编号。
virtual_transaction	text	持有此锁或者在等待此锁的事务的虚拟id。
pid	bigint	持有此锁或者等待此锁的线程号。



名称	类型	描述
mode	text	锁级别。
granted	boolean	是否持有锁。
fastpath	boolean	是否通过fastpath机制获得锁。
wait_for_pid	bigint	锁冲突线程的线程号。
conflict_mode	text	锁冲突线程持有的冲突锁级别。
query_id	bigint	查询语句的id。
query	text	查询语句。
application_name	text	连接到该后端的应用名。
backend_start	timestamp with time zone	后端进程启动时间，即客户端连接服务器的时间。
xact_start	timestamp with time zone	当前事务的启动时间。
query_start	timestamp with time zone	开始当前活跃查询的时间。
state	text	后端当前总体状态。

## 应用实例

**步骤1** 连接dn节点，开启一个事务，执行查询：

```
begin;select * from t1;
```

**步骤2** 另一个窗口连接cn，对表t1执行truncate：

```
truncate t1;
```

此时truncate会被阻塞。

**步骤3** 另开一个窗口连接cn，执行select \* from pgxc\_lockwait\_detail;

```
SELECT * FROM PGXC_LOCKWAIT_DETAIL;
level | node_name | lock_wait_hierarchy | lock_type | database | relation | page | tuple |
virtual_xid | transaction_id | class_id | obj_id | obj_subid | virtual_transaction | p
id | mode | granted | fastpath | wait_for_pid | conflict_mode | query_id
| query | application_name | backend_start |
xact_start | query_start | state
-----+-----+-----+-----+-----+-----+-----+-----+
+-----+-----+-----+-----+-----+-----+-----+-----+
+-----+-----+-----+-----+-----+-----+-----+-----+
+-----+-----+-----+-----+-----+-----+-----+-----+
1 | datanode1 | datanode1:140378619314976 | | relation | 16049 | 2147484411 | |
| | 673638 | | | | 19/297 | | 1403786
19314976 | AccessExclusiveLock | f | f | | 140378619263840 | AccessShareLock | 73183493945504391
```

```

| TRUNCATE t1 | coordinator1 | 2023-03-13 12:13:52.530602+08 | 2
023-03-13 14:52:16.1456+08 | 2023-03-13 14:52:16.148693+08 | active
2 | datanode1 | datanode1:140378619314976 -> 140378619263840 | relation | 16049 | 2147484411 |
| | | | | | | 23/16067 | 1403786
19263840 | AccessShareLock | t | f | | | | 0 | begin;select * from t1;
| gsql | 2023-03-13 14:19:26.325602+08 | 2
023-03-13 14:52:12.042741+08 | 2023-03-13 14:52:12.042741+08 | idle in transaction
(2 rows)
    
```

----结束

## 15.3.184 PGXC\_INSTR\_UNIQUE\_SQL

PGXC\_INSTR\_UNIQUE\_SQL视图展示集群中所有CN节点的Unique SQL的完整统计信息。

需要有系统管理员权限才可以访问此视图，具体的字段请参考[GS\\_INSTR\\_UNIQUE\\_SQL](#)。

## 15.3.185 PGXC\_LOCK\_CONFLICTS

PGXC\_LOCK\_CONFLICTS视图提供集群中有冲突的锁的信息。

当某一个锁正在等待另一个锁，或正在被另一个锁等待，即该锁是有冲突的。

目前，PGXC\_LOCK\_CONFLICTS视图只收集locktype为relation、partition、page、tuple和transactionid的锁的信息。

表 15-238 PGXC\_LOCK\_CONFLICTS 字段

名称	类型	描述
locktype	text	被锁定对象的类型。
nodename	name	被锁定对象的节点的名称。
dbname	name	被锁定对象的数据库的名称。如果被锁定对象是事务，则为NULL。
nspname	name	被锁定对象的命名空间的名称。
relname	name	被锁定对象对应的关系的名称。如果被锁定对象既不是关系，也不是关系的一部分，则为NULL。
partname	name	被锁定对象对应的分区的名称。如果被锁定对象不是分区，则为NULL。
page	integer	被锁定对象对应的页面的编号。如果被锁定对象既不是页面，也不是元组，则为NULL。
tuple	smallint	被锁定对象对应的元组的编号。如果被锁定对象不是元组，则为NULL。
transactionid	xid	被锁定对象对应的事务的ID。如果被锁定对象不是事务，则为NULL。
username	name	申请锁的用户的名称。

名称	类型	描述
gxid	xid	申请锁的事务的ID。
xactstart	timestamp with time zone	申请锁的事务的开始时间。
queryid	bigint	申请锁的线程的最新查询ID。
query	text	申请锁的线程的最新查询语句。
pid	bigint	申请锁的线程的ID。
mode	text	锁的级别。
granted	boolean	<ul style="list-style-type: none"> <li>如果锁已被持有，则为TRUE。</li> <li>如果锁还在等待其它锁，则为FALSE。</li> </ul>

### 15.3.186 PGXC\_NODE\_ENV

PGXC\_NODE\_ENV视图提供获取集群中所有节点的环境变量信息。

表 15-239 PGXC\_NODE\_ENV 字段

名称	类型	描述
node_name	text	集群中所有节点的名称
host	text	集群中所有节点的主机名称
process	integer	集群中所有节点的进程号
port	integer	集群中所有节点的端口号
installpath	text	集群中所有节点的安装目录
datapath	text	集群中所有节点的数据目录
log_directory	text	集群中所有节点的日志目录

### 15.3.187 PGXC\_NODE\_STAT\_RESET\_TIME

PGXC\_NODE\_STAT\_RESET\_TIME视图显示集群中各节点的统计信息重置时间，除新增node\_name（节点名称）字段外，其余字段内容和[GS\\_NODE\\_STAT\\_RESET\\_TIME](#)视图相同。需要有系统管理员权限才可以访问此视图。

### 15.3.188 PGXC\_OS\_RUN\_INFO

PGXC\_OS\_RUN\_INFO视图显示集群中各节点上操作系统运行的状态信息，除新增node\_name（节点名称）字段外，其余字段内容和[PV\\_OS\\_RUN\\_INFO](#)视图相同。需要有系统管理员权限或预置角色gs\_role\_read\_all\_stats权限才可以访问此视图。

## 15.3.189 PGXC\_OS\_THREADS

PGXC\_OS\_THREADS视图提供当前集群中所有正常节点下的线程状态信息。

表 15-240 PGXC\_OS\_THREADS 字段

名称	类型	描述
node_name	text	当前集群中所有正常节点的名称
pid	bigint	当前集群中所有正常节点进程中正在运行的线程号
lwpid	integer	与pid对应的轻量级线程号
thread_name	text	与pid对应的线程名称
creation_time	timestamp with time zone	与pid对应的线程创建的时间

## 15.3.190 PGXC\_PREPARED\_XACTS

PGXC\_PREPARED\_XACTS视图显示当前处于prepared阶段的两阶段事务。

表 15-241 PGXC\_PREPARED\_XACTS 字段

名字	类型	描述
pgxc_prepared_xact	text	查看当前处于prepared阶段的两阶段事务。

## 15.3.191 PGXC\_REDO\_STAT

视图PGXC\_REDO\_STAT显示集群中各节点上XLOG重做过程中的统计信息，除新增node\_name（节点名称）字段外，其余字段内容和PV\_REDO\_STAT视图相同。需要有系统管理员权限或预置角色gs\_role\_read\_all\_stats权限才可以访问此视图。

## 15.3.192 PGXC\_REL\_IOSTAT

PGXC\_REL\_IOSTAT视图显示集群中各节点上磁盘读写的统计信息。除新增node\_name（节点名称）字段外，其余字段内容和GS\_REL\_IOSTAT视图相同。需要有系统管理员权限才可以访问此视图。

## 15.3.193 PGXC\_REPLICATION\_SLOTS

PGXC\_REPLICATION\_SLOTS视图显示集群中DN上的复制信息，除增加表示节点名称的node\_name字段外，其余字段内容和PG\_REPLICATION\_SLOTS视图相同。需要有系统管理员权限才可以访问此视图。

## 15.3.194 PGXC\_RESPOOL\_RUNTIME\_INFO

PGXC\_RESPOOL\_RUNTIME\_INFO视图显示所有CN上所有资源池作业运行信息。

表 15-242 PGXC\_RESPOOL\_RUNTIME\_INFO 字段

名称	类型	描述
nodename	name	CN名称
nodegroup	name	资源池所属逻辑集群名称，默认集群显示"installation"
rpname	name	资源池名称
ref_count	int	资源池引用作业数，作业经过资源池不管是否管控都会计数
fast_run	int	资源池快车道运行作业数
fast_wait	int	资源池快车道排队作业数
slow_run	int	资源池慢车道运行作业数
slow_wait	int	资源池慢车道排队作业数

## 15.3.195 PGXC\_RESPOOL\_RESOURCE\_INFO

PGXC\_RESPOOL\_RESOURCE\_INFO视图显示所有实例上资源池实时监控信息。

### 说明

- DN上仅显示当前DN所属逻辑集群的资源池监控信息。
- 从8.2.0集群版本开始，新增了内存负反馈机制功能，CCN节点会根据DN节点的实际内存使用情况辅助反向调节语句的估算内存统计值，缓解语句估算内存高估场景，当CCN节点触发通过负反馈机制来减少估算内存统计值来增发作业时，由于CCN节点更新了语句缩减后的估算内存统计值，而CN节点还保留了原先的估算内存统计值，会导致资源池监控视图中的估算内存超过资源池的上限。
- 语句的算子分为重内存算子和非重内存算子，对于语句的内存管控是重内存算子，非重内存算子的内存开销较小；由于非重内存算子的内存开销、线程初始化的开销、表达式的开销等这些内存不做管控，会导致资源池的used\_mem会在一定程度上超过mem\_limit值。

表 15-243 PGXC\_RESPOOL\_RESOURCE\_INFO 字段

名称	类型	描述
nodename	name	实例名称，包含CN和DN。
nodegroup	name	资源池所属逻辑集群名称，默认集群显示"installation"。
rpname	name	资源池名称。
cgroup	name	资源池关联控制组名称。

名称	类型	描述
ref_count	int	资源池引用作业数，作业经过资源池不管是否管控都会计数，仅CN上有效。
fast_run	int	资源池快车道运行作业数，只在CN上有效。
fast_wait	int	资源池快车道排队作业数，只在CN上有效。
fast_limit	int	资源池快车道作业并发限制，只在CN上有效。
slow_run	int	资源池慢车道运行作业数，只在CN上有效。
slow_wait	int	资源池慢车道排队作业数，只在CN上有效。
slow_limit	int	资源池慢车道作业并发限制，只在CN上有效。
used_cpu	double	资源池5s监控周期内使用CPU个数平均值，保留小数点后2位。 <ul style="list-style-type: none"> <li>• DN：显示当前DN上资源池使用的CPU个数。</li> <li>• CN：显示所有DN上资源池使用CPU的累积和。</li> </ul>
cpu_limit	int	资源池可用CPU的上限，CPU配额管控情况下为GaussDB(DWS)可用CPU，CPU限额管控情况下为关联控制组CPU可用CPU。 <ul style="list-style-type: none"> <li>• DN：显示当前DN上资源池可用CPU上限。</li> <li>• CN：显示所有DN上资源池可用CPU上限的累积和。</li> </ul>
used_mem	int	资源池当前使用的内存大小，单位MB。 <ul style="list-style-type: none"> <li>• DN：显示当前DN上资源池使用的内存大小。</li> <li>• CN：显示所有DN上资源池使用内存的累积和。</li> </ul>
estimate_mem	int	当前CN上，资源池运行作业的估算内存之和，只在CN上有效。
mem_limit	int	资源池可用内存上限，单位MB。 <ul style="list-style-type: none"> <li>• DN：显示当前DN上资源池可用内存上限。</li> <li>• CN：显示所有DN上资源池可用内存上限的累积和。</li> </ul>
read_kbytes	bigint	资源池5s监控周期内逻辑读字节数，单位KB。 <ul style="list-style-type: none"> <li>• DN：显示当前DN上资源池逻辑读字节数。</li> <li>• CN：显示所有DN上资源池逻辑读字节的累积和。</li> </ul>
write_kbytes	bigint	资源池5s监控周期内逻辑写字节数，单位KB。 <ul style="list-style-type: none"> <li>• DN：显示当前DN上资源池逻辑写字节数。</li> <li>• CN：显示所有DN上资源池逻辑写字节的累积和。</li> </ul>

名称	类型	描述
read_counts	bigint	资源池5s监控周期内逻辑读次数。 <ul style="list-style-type: none"> <li>DN: 显示当前DN上资源池逻辑读次数。</li> <li>CN: 显示所有DN上资源池逻辑读次数的累积和。</li> </ul>
write_counts	bigint	资源池5s监控周期内逻辑写次数。 <ul style="list-style-type: none"> <li>DN: 显示当前DN上资源池逻辑写次数。</li> <li>CN: 显示所有DN上资源池逻辑写次数的累积和。</li> </ul>
read_speed	double	资源池5s监控周期内逻辑读速率的平均值。 <ul style="list-style-type: none"> <li>DN: 显示当前DN上资源池逻辑读速率。</li> <li>CN: 显示所有DN上资源池逻辑读速率的累积和。</li> </ul>
write_speed	double	资源池5s监控周期内逻辑写速率平均值。 <ul style="list-style-type: none"> <li>DN: 显示当前DN上资源池逻辑写速率。</li> <li>CN: 显示所有DN上资源池逻辑写速率的累积和。</li> </ul>

### 15.3.196 PGXC\_RESPOOL\_RESOURCE\_HISTORY

PGXC\_RESPOOL\_RESOURCE\_HISTORY用于查询所有实例上资源池监控历史信息。

表 15-244 PGXC\_RESPOOL\_RESOURCE\_HISTORY 字段

名称	类型	描述
nodename	name	实例名称, 包含CN和DN
timestamp	timestamp	资源池监控信息持久化时间
nodegroup	name	资源池所属逻辑集群名称, 默认集群显示 "installation"
rpname	name	资源池名称
cgroup	name	资源池关联控制组名称
ref_count	int	资源池引用作业数, 作业经过资源池不管是否管控都会计数, 仅CN上有效
fast_run	int	资源池快车道运行作业数, 只在CN上有效
fast_wait	int	资源池快车道排队作业数, 只在CN上有效
fast_limit	int	资源池快车道作业并发限制, 只在CN上有效
slow_run	int	资源池慢车道运行作业数, 只在CN上有效

名称	类型	描述
slow_wait	int	资源池慢车道排队作业数，只在CN上有效
slow_limit	int	资源池慢车道作业并发限制，只在CN上有效
used_cpu	double	资源池5s监控周期内使用CPU个数平均值，保留小数点后2位 <ul style="list-style-type: none"> <li>• DN：显示当前DN上资源池使用的CPU个数</li> <li>• CN：显示所有DN上资源池使用CPU的累积和</li> </ul>
cpu_limit	int	资源池可用CPU的上限，CPU配额管控情况下为GaussDB可用CPU，CPU限额管控情况下为关联控制组CPU可用CPU <ul style="list-style-type: none"> <li>• DN：显示当前DN上资源池可用CPU上限</li> <li>• CN：显示所有DN上资源池可用CPU上限的累积和</li> </ul>
used_mem	int	资源池使用的内存大小，单位：'MB' <ul style="list-style-type: none"> <li>• DN：显示当前DN上资源池使用的内存大小</li> <li>• CN：显示所有DN上资源池使用内存的累积和</li> </ul>
estimate_mem	int	当前CN上，资源池运行作业的估算内存之和，只在CN上有效
mem_limit	int	资源池可用内存上限，单位：'MB' <ul style="list-style-type: none"> <li>• DN：显示当前DN上资源池可用内存上限</li> <li>• CN：显示所有DN上资源池可用内存上限的累积和</li> </ul>
read_kbytes	bigint	资源池5s监控周期内逻辑读字节数，单位：'KB' <ul style="list-style-type: none"> <li>• DN：显示当前DN上资源池逻辑读字节数</li> <li>• CN：显示所有DN上资源池逻辑读字节的累积和</li> </ul>
write_kbytes	bigint	资源池5s监控周期内逻辑写字节数，单位：'KB' <ul style="list-style-type: none"> <li>• DN：显示当前DN上资源池逻辑写字节数</li> <li>• CN：显示所有DN上资源池逻辑写字节的累积和</li> </ul>
read_counts	bigint	资源池5s监控周期内逻辑读次数 <ul style="list-style-type: none"> <li>• DN：显示当前DN上资源池逻辑读次数</li> <li>• CN：显示所有DN上资源池逻辑读次数的累积和</li> </ul>
write_counts	bigint	资源池5s监控周期内逻辑写次数 <ul style="list-style-type: none"> <li>• DN：显示当前DN上资源池逻辑写次数</li> <li>• CN：显示所有DN上资源池逻辑写次数的累积和</li> </ul>



名称	类型	描述
read_speed	double	资源池5s监控周期内逻辑读速率的平均值 <ul style="list-style-type: none"> <li>DN: 显示当前DN上资源池逻辑读速率</li> <li>CN: 显示所有DN上资源池逻辑读速率的累积和</li> </ul>
write_speed	double	资源池5s监控周期内逻辑写速率平均值 <ul style="list-style-type: none"> <li>DN: 显示当前DN上资源池逻辑写速率</li> <li>CN: 显示所有DN上资源池逻辑写速率的累积和</li> </ul>

### 15.3.197 PGXC\_ROW\_TABLE\_IO\_STAT

PGXC\_ROW\_TABLE\_IO\_STAT视图提供集群所有CN和DN节点上当前数据库所有行存表的IO统计数据。除在每一行前面增加name类型的nodename字段外，其余字段的名称、类型和顺序与GS\_ROW\_TABLE\_IO\_STAT视图相同，具体的字段请参考[GS\\_ROW\\_TABLE\\_IO\\_STAT](#)。

### 15.3.198 PGXC\_RUNNING\_XACTS

PGXC\_RUNNING\_XACTS视图主要功能是显示集群中各个节点运行事务的信息，字段内容和[PG\\_RUNNING\\_XACTS](#)相同。

表 15-245 PGXC\_RUNNING\_XACTS 字段

名称	类型	描述
handle	integer	事务在GTM对应的句柄。
gxid	xid	事务ID号。
state	tinyint	事务状态（3: prepared或者0: starting）。
node	text	节点名称。
xmin	xid	节点上当前数据涉及的最小事务号xmin。
vacuum	boolean	标志当前事务是否是lazy vacuum事务。
timeline	bigint	标识数据库重启次数。
prepare_xid	xid	处于prepared状态事务的id号，若不在prepared状态，值为0。
pid	bigint	事务对应的线程ID。
next_xid	xid	CN传给DN的事务ID号。

## 15.3.199 PGXC\_SETTINGS

PGXC\_SETTINGS视图显示集群中各节点数据库运行时参数的相关信息，除新增node\_name（节点名称）字段外，其余字段内容和PG\_SETTINGS视图相同。需要有系统管理员权限才可以访问此视图。

## 15.3.200 PGXC\_SESSION\_WLMSTAT

PGXC\_SESSION\_WLMSTAT视图显示当前集群中各CN节点用户执行作业正在运行时的负载管理相关信息。

表 15-246 PGXC\_SESSION\_WLMSTAT 字段

名称	类型	描述
nodename	name	节点名称
datid	oid	连接后端的数据库OID。
datname	name	连接后端的数据库名称。
threadid	bigint	后端线程ID。
processid	integer	后端线程的pid。
usesysid	oid	登录后端的用户OID。
appname	text	连接到后端的应用名。
username	name	登录到该后端的用户名。
priority	bigint	语句所在Cgroups的优先级。
attribute	text	语句的属性： <ul style="list-style-type: none"><li>• Ordinary: 语句发送到数据库后被解析前的默认属性。</li><li>• Simple: 简单语句。</li><li>• Complicated: 复杂语句。</li><li>• Internal: 数据库内部语句。</li></ul>
block_time	bigint	语句当前为止的pending的时间，单位s。
elapsed_time	bigint	语句当前为止的实际执行时间，单位s。
total_cpu_time	bigint	语句在上一时间周期内的DN上CPU使用的总时间，单位s。
cpu_skew_percent	integer	语句在上一时间周期内的DN上CPU使用的倾斜率。
statement_mem	integer	语句执行所需要的估算内存，预留字段。
active_points	integer	语句占用的资源池并发点数。
dop_value	integer	从资源池中获取语句的dop值。

名称	类型	描述
control_group	text	语句当前所使用的Cgroups。
status	text	语句当前的状态，包括： <ul style="list-style-type: none"> <li>• pending: 执行前状态。</li> <li>• running: 执行进行状态。</li> <li>• finished: 执行正常结束。（当enqueue字段为StoredProc或Transaction时，仅代表语句中的部分作业已经执行完毕，该状态会持续到该语句完全执行完毕。）</li> <li>• aborted: 执行异常终止。</li> <li>• active: 非以上四种状态外的正常状态。</li> <li>• unknown: 未知状态。</li> </ul>
enqueue	text	语句当前的排队情况，包括： <ul style="list-style-type: none"> <li>• Global: 全局排队。</li> <li>• Respool: 资源池排队。</li> <li>• CentralQueue: 在中心协调节点(CCN)中排队。</li> <li>• Transaction: 语句处于一个事务块中。</li> <li>• StoredProc : 语句处于一个存储过程中。</li> <li>• None: 未在排队。</li> <li>• Forced None : 事务块语句或存储过程语句由于超出设定的等待时间而强制执行。</li> </ul>
resource_pool	name	语句当前所在的资源池。
query	text	该后端的最新查询。如果state状态是active，此字段显示当前正在执行的查询。所有其他情况表示上一个查询。
isplana	bool	逻辑集群模式下，语句当前是否占用其他逻辑集群的资源执行。该值默认为f，表示不占用其他逻辑集群的资源执行。
node_group	text	语句所属用户对应的逻辑集群。
lane	text	表示语句查询的快慢车道。 <ul style="list-style-type: none"> <li>• fast: 快车道。</li> <li>• slow: 慢车道。</li> <li>• none: 未管控。</li> </ul>

### 15.3.201 PGXC\_STAT\_ACTIVITY

PGXC\_STAT\_ACTIVITY视图显示当前集群下所有CN的当前用户查询相关的信息。

表 15-247 PGXC\_STAT\_ACTIVITY 字段

名称	类型	描述
coorname	text	当前集群下的CN名称。
datid	oid	用户会话在后端连接到的数据库OID。
datname	name	用户会话在后端连接到的数据库名称。
pid	bigint	后端线程ID。
lwtid	integer	后端线程的轻量级线程号。
usesysid	oid	登录此后端的用户OID。
username	name	登录此后端的用户名。
application_name	text	连接到此后端的应用名。
client_addr	inet	连接到此后端的客户端的IP地址。如果此字段是null，则表示通过服务器机器上UNIX套接字连接客户端或者这是内部进程，如autovacuum。
client_hostname	text	客户端的主机名，此字段是通过client_addr的反向DNS查找得到。仅在启动log_hostname且使用IP连接时才非空。
client_port	integer	客户端用于与后端通讯的TCP端口号，如果使用Unix套接字，则为-1。
backend_start	timestamp with time zone	后端进程启动时间，即客户端连接服务器的时间。
xact_start	timestamp with time zone	当前事务的启动时间，如果没有事务是活跃的，则为null。如果当前查询是首个事务，则这列等同于query_start列。
query_start	timestamp with time zone	开始当前活跃查询的时间，如果state的值不是active，则这个值是上一个查询的开始时间。
state_change	timestamp with time zone	状态最后一次改变的时间。
waiting	boolean	如果后端当前正等待锁或者等待节点则为true。

名称	类型	描述
enqueue	text	<p>语句当前排队状态。可能值是：</p> <ul style="list-style-type: none"> <li>waiting in global queue: 表示语句在全局排队中。</li> <li>waiting in respool queue: 表示语句在资源池排队中。</li> <li>waiting in ccn queue: 表示作业在CCN排队中。</li> <li>空或no waiting queue: 表示语句正在运行。</li> </ul>
state	text	<p>后端当前总体状态。可能值是：</p> <ul style="list-style-type: none"> <li>active: 后端正在执行一个查询。</li> <li>idle: 后端正在等待一个新的客户端命令。</li> <li>idle in transaction: 后端在事务中，但事务中没有语句在执行。</li> <li>idle in transaction (aborted): 后端在事务中，但事务中有语句执行失败。</li> <li>fastpath function call: 后端正在执行一个fast-path函数。</li> <li>disabled: 如果后端禁用track_activities，则报告此状态。</li> </ul> <p><b>说明</b> 只有系统管理员能查看到自己帐户所对应的会话状态。其他帐户的state信息为空。例如以judy用户连接数据库后，在pgxc_stat_activity中查看到的普通用户joe及初始用户dbadmin的state信息为空：</p> <pre>SELECT datname, username, usesysid, state,pid FROM pgxc_stat_activity;  datname   username   usesysid   state   pid -----+-----+-----+-----+----- +-----+ gaussdb   dbadmin   10      139968752121616 gaussdb   dbadmin   10      139968903116560 db_tpcds   judy   16398   active    139968391403280 gaussdb   dbadmin   10      139968643069712 gaussdb   dbadmin   10      139968680818448 gaussdb   joe   16390      139968563377936 (6 rows)</pre>
resource_pool	name	用户使用的资源池。
stmt_type	text	用户语句的语句类型。
query_id	bigint	查询语句的ID。

名称	类型	描述
query	text	此后端的最新查询。如果state状态是 active（活跃的），此字段显示当前正在执行的查询。其他情况表示上一个查询。
connection_info	text	json格式字符串，记录当前连接数据库的驱动类型、驱动版本号、当前驱动的部署路径、进程属主用户等信息（参见 <a href="#">connection_info</a> ）。

## 应用实例

查看当前处于阻塞状态的查询语句。

```
SELECT datname,username,state,query FROM PGXC_STAT_ACTIVITY WHERE waiting = true;
```

查看快照线程的工作状态。

```
SELECT application_name,backend_start,state_change,state,query FROM PGXC_STAT_ACTIVITY WHERE application_name='WDRSnapshot';
```

查看正在运行的查询语句。

```
SELECT datname,username,state,pid FROM PGXC_STAT_ACTIVITY;
```

```
datname | username | state | pid
-----+-----+-----+-----
gaussdb | Ruby    | active | 140298793514752
gaussdb | Ruby    | active | 140298718004992
gaussdb | Ruby    | idle   | 140298650908416
gaussdb | Ruby    | idle   | 140298625742592
gaussdb | dbadmin | active | 140298575406848
(5 rows)
```

查看指定数据库postgres上已使用的会话连接数。其中1表示数据库postgres上已使用的会话连接数。

```
SELECT COUNT(*) FROM PGXC_STAT_ACTIVITY WHERE DATNAME='postgres';
```

```
count
-----
1
(1 row)
```

### 15.3.202 PGXC\_STAT\_BAD\_BLOCK

PGXC\_STAT\_BAD\_BLOCK视图显示集群所有节点从启动后，在读取数据时出现Page/CU校验失败的统计信息。

表 15-248 PGXC\_STAT\_BAD\_BLOCK 字段

名字	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

名字	类型	描述
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

### 15.3.203 PGXC\_STAT\_BGWRITER

PGXC\_STAT\_BGWRITER视图显示集群中各节点上后端写进程活动的统计信息，除新增node\_name（节点名称）字段外，其余字段内容和PG\_STAT\_BGWRITER视图相同。需要有系统管理员权限才可以访问此视图。

### 15.3.204 PGXC\_STAT\_DATABASE

视图PGXC\_STAT\_DATABASE显示集群中各节点上数据库的状态和统计信息，除新增node\_name（节点名称）字段外，其余字段内容和PG\_STAT\_DATABASE视图相同。需要有系统管理员权限才可以访问此视图。

### 15.3.205 PGXC\_STAT\_REPLICATION

PGXC\_STAT\_REPLICATION视图显示集群中各节点上日志同步的状态信息，除新增node\_name（节点名称）字段外，其余字段内容和PG\_STAT\_REPLICATION视图相同。需要有系统管理员权限才可以访问此视图。

### 15.3.206 PGXC\_STAT\_TABLE\_DIRTY

PGXC\_STAT\_TABLE\_DIRTY显示当前集群中所有节点（CN和DN）上全部表的统计信息，并展示表在单节点（单CN级或者单DN级）的脏页率。该视图仅8.1.3及以上集群版本支持。

#### 说明

该视图的统计信息依赖于ANALYZE，为获取最准确的信息请先对表进行ANALYZE。

表 15-249 PGXC\_STAT\_TABLE\_DIRTY 字段

名称	类型	描述
nodename	text	节点名。
schema	name	表的模式名。
tablename	name	表名。
partname	name	分区表的分区名。

名称	类型	描述
last_vacuum	timestampwith time zone	最后一次手动vacuum时间。
last_autovacuum	timestampwith time zone	最后一次autovacuum时间。
last_analyze	timestampwith time zone	最后一次手动analyze时间。
last_antoanalyze	timestampwith 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	autoanalyze_count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）。
last_data_changed	timestampwith time zone	记录表最后一次数据变化的时间。

## 使用建议

- 对于高脏页率的系统表，建议在确认当前没有用户操作该系统表时，再执行VACUUM FULL。
- 建议对脏页率超过80%的非系统表执行VACUUM FULL，用户也可根据业务场景自行选择是否执行VACUUM FULL。

## 使用场景

1. 查询全库所有用户表的整体脏页率：

```
select
  t1.schema,
  t1.tablename,
  t1.total_ins,
```



```

t1.total_upd,
t1.total_del,
t1.total_tup_hot_upd,
t1.total_change,
t1.total_live,
t1.total_dead,
t1.total_dirty_rate,
t1.max_dirty,
t2.max_node,
t1.min_dirty,
t2.min_node
from
(select
  a.schema,
  a.tablename,
  sum(a.n_tup_ins) as total_ins,
  sum(a.n_tup_upd) as total_upd,
  sum(a.n_tup_del) as total_del,
  sum(a.n_tup_hot_upd) as total_tup_hot_upd,
  sum(a.n_tup_change) as total_change,
  sum(a.n_live_tup) as total_live,
  sum(a.n_dead_tup) as total_dead,
  Round((total_dead / (total_dead + total_live + 0.0001) * 100),2) AS total_dirty_rate,
  max(a.dirty_rate) as max_dirty,
  min(a.dirty_rate) as min_dirty
  from pg_catalog.pgxc_stat_table_dirty a where a.partname is null and a.schema not in
('pg_toast','cstore','gs_logical_cluster','sys','dbms_om','information_schema','pg_catalog','dbms_output',
dbms_random','utl_raw','utl_raw dbms_sql','dbms_lob') group by a.tablename, a.schema
) t1,
(select distinct
  tablename, schema,
  first_value(nodename) over(partition by tablename, schema order by dirty_rate) as min_node,
  first_value(nodename) over(partition by tablename, schema order by dirty_rate desc) as max_node
  from (select * from pg_catalog.pgxc_stat_table_dirty)) t2
where t1.tablename = t2.tablename and t1.schema = t2.schema;

```

## 2. 查询全库所有表(用户表+系统表)的整体脏页率:

```

select
  t1.schema,
  t1.tablename,
  t1.total_ins,
  t1.total_upd,
  t1.total_del,
  t1.total_tup_hot_upd,
  t1.total_change,
  t1.total_live,
  t1.total_dead,
  t1.total_dirty_rate,
  t1.max_dirty,
  t2.max_node,
  t1.min_dirty,
  t2.min_node
from
(select
  a.schema,
  a.tablename,
  sum(a.n_tup_ins) as total_ins,
  sum(a.n_tup_upd) as total_upd,
  sum(a.n_tup_del) as total_del,
  sum(a.n_tup_hot_upd) as total_tup_hot_upd,
  sum(a.n_tup_change) as total_change,
  sum(a.n_live_tup) as total_live,
  sum(a.n_dead_tup) as total_dead,
  Round((total_dead / (total_dead + total_live + 0.0001) * 100),2) AS total_dirty_rate,
  max(a.dirty_rate) as max_dirty,
  min(a.dirty_rate) as min_dirty
  from pg_catalog.pgxc_stat_table_dirty a where a.partname is null group by a.tablename, a.schema
) t1,
(select distinct
  tablename, schema,

```

```
first_value(nodename) over(partition by tablename, schema order by dirty_rate) as min_node,
first_value(nodename) over(partition by tablename, schema order by dirty_rate desc) as max_node
from (select * from pg_catalog.pgxc_stat_table_dirty)) t2
where t1.tablename = t2.tablename and t1.schema = t2.schema;
```

3. 查询全库系统表信息:

```
select * from pgxc_stat_table_dirty where schema in
('pg_toast','cstore','gs_logical_cluster','sys','dbms_om','information_schema','pg_catalog','dbms_output','
dbms_random','utl_raw','utl_raw dbms_sql','dbms_lob');
```

### 15.3.207 PGXC\_STAT\_WAL

PGXC\_STAT\_WAL视图显示当前query的wal日志和数据页的流量信息，该视图仅8.2.0及以上集群版本支持。

表 15-250 PGXC\_STAT\_WAL 字段

名称	类型	描述
query_id	bigint	当前query的ID。
query_start	timestamp	query起始时间。
global_wal	bigint	当前query在集群产生的wal日志总量，单位是Byte。
global_avg_wal_speed	bigint	当前query在集群产生wal日志的平均速率，单位是Byte/s。
global_datapage	bigint	当前query在集群产生的数据页总量，单位是Byte。
global_avg_data_page_speed	bigint	当前query在集群产生数据页的平均速率，单位是Byte/s。
min_wal_node	Text	当前query产生wal日志量最小的实例组名。
min_wal	bigint	最小node产生的wal日志量，单位是Byte。
max_wal_node	Text	当前query产生wal日志量最大的实例组名。
max_wal	bigint	最大node产生的wal日志量，单位是Byte。
min_datapage_node	Text	当前query产生数据页量最小的实例组名。
min_data_page	bigint	最小node产生的数据页量，单位是Byte。
max_datapage_node	Text	当前query产生数据页量最大的实例组名。
max_data_page	bigint	最大node产生的数据页量，单位是Byte。
avg_wal_per_node	bigint	平均每个node产生的wal日志量，单位是Byte。
avg_datapage_per_node	bigint	平均每个node产生的数据页量，单位是Byte。
query	Text	当前执行的语句。

📖 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。

应用示例

查询集群当前正在执行的语句，以及这些语句分别产生的wal日志的总量和平均速率、数据页总量和平均速率，以及这些信息各自DN上的分布情况。

```
SELECT * FROM PGXC_STAT_WAL;
  query_id | query_start | global_wal | global_avg_wal_speed | global_datapage |
global_avg_datapage_speed | min_wal_node | min_wal | max_wal_node | max_wal | min_datapage_node |
min_data_page | max_datapage_node | max_data_page | avg_wal_per_node |
avg_datapage_per_node | query
-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
72620543991351767 | 2022-11-10 16:49:47.743291+08 | 7579052 | 419000 | 284057600 |
15740000 | datanode1 | 7579052 | datanode1 | 7579052 | datanode1 | 284057600 |
datanode1 | 284057600 | 7579052 |
284057600 | insert into mpptest3 select * from mpptest3;
72620543991351781 | 2022-11-10 16:50:00.616697+08 | 55022176 | 10638000 | 0 |
0 | datanode1 | 55022176 | datanode1 | 55022176 | datanode1 | 0 |
datanode1 | 0 | 55022176 |
0 | insert into mpptest1 select * from mpptest1;
(2 rows)
```

15.3.208 PGXC\_SQL\_COUNT

通过PGXC\_SQL\_COUNT视图，可以实时显示集群中各CN节点上SELECT、INSERT、UPDATE、DELETE、MERGE INTO五种SQL、以及DDL、DML、DCL语句的节点级和用户级统计结果，识别当前业务负载较重的query类型，衡量整个集群和单个节点执行某种类型查询的能力。通过对以上几类SQL查询进行计数和响应时间统计，获得指定时刻的统计结果，经计算可以得到指定QPS等统计信息。例如，T1时刻，USER1的SELECT计数结果为X1，T2时刻为X2，则可计算得到该用户SELECT查询的QPS值为(X2-X1)/(T2-T1)。由此，可获得集群用户级QPS曲线图和集群吞吐情况，监测每个用户的业务负载是否发生剧烈变化。如果有剧烈变化，可以定位具体的语句类型(SELECT/INSERT/UPDATE/DELETE/MERGE INTO)。同时观测QPS曲线可以获知问题发生时间点，结合其它工具，定位问题点。能够为集群性能调优、问题定位等提供依据。

PGXC\_SQL\_COUNT视图的字段与GS\_SQL\_COUNT一致，详见[表15-125](#)。

📖 说明

当执行用户的MERGE INTO语句时，若能下推，在DN上收到的是MERGE INTO语句，将在DN节点上进行MERGE INTO类型计数，相应mergeinto\_count列计数增加；若不能下推，在DN上收到的是UPDATE或INSERT语句，将在DN节点上进行UPDATE或INSERT类型计数，相应的update\_count列或insert\_count列计数增加。

15.3.209 PGXC\_TABLE\_CHANGE\_STAT

PGXC\_TABLE\_CHANGE\_STAT视图提供集群所有CN节点上当前数据库所有表的变更情况。除在每一行前面增加name类型的nodename字段外，其余字段的名称、类型和顺序与GS\_TABLE\_CHANGE\_STAT视图相同，具体的字段请参考[GS\\_TABLE\\_CHANGE\\_STAT](#)。

### 15.3.210 PGXC\_TABLE\_STAT

PGXC\_TABLE\_STAT视图提供集群所有CN和DN节点上当前数据库所有表的统计信息。除在每一行前面增加name类型的nodename字段外，其余字段的名称、类型和顺序与GS\_TABLE\_STAT视图相同，具体的字段请参考[GS\\_TABLE\\_STAT](#)。

### 15.3.211 PGXC\_THREAD\_WAIT\_STATUS

通过CN节点查看PGXC\_THREAD\_WAIT\_STATUS视图，可以查看集群全局各个节点上所有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，从而更容易定位进程停止响应问题以及类似现象的原因。

PGXC\_THREAD\_WAIT\_STATUS视图和PG\_THREAD\_WAIT\_STATUS视图列定义完全相同，这是由于PGXC\_THREAD\_WAIT\_STATUS视图本质是到集群中各个节点上查询PG\_THREAD\_WAIT\_STATUS视图汇总的结果。

表 15-251 PGXC\_THREAD\_WAIT\_STATUS 字段

名称	类型	描述
node_name	text	当前节点的名称。
db_name	text	数据库名称。
thread_name	text	线程名称。
query_id	bigint	查询ID，对应debug_query_id。
tid	bigint	当前线程的线程号。
lwtid	integer	当前线程的轻量级线程号。
ptid	integer	streaming线程的父线程。
tlevel	integer	streaming线程的层级。
smpid	integer	并行线程的ID。
wait_status	text	当前线程的等待状态。等待状态的详细信息请参见 <a href="#">表15-210</a> 。
wait_event	text	如果wait_status是acquire lock、acquire lwlock、wait io三种类型，此列描述具体的锁、轻量级锁、IO的信息。否则是空。

例如：

在coordinator1执行一条语句之后长时间没有响应。可以创建另外一个连接到coordinator1上，查询coordinator1上的线程状态。

```
select * from pg_thread_wait_status where query_id > 0;
 node_name | db_name | thread_name | query_id | tid | lwtid | ptid | tlevel | smpid |
wait_status | wait_event
-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
coordinator1 | gaussdb | gsql | 20971544 | 140274089064208 | 22579 | | 0 | 0 | wait node:
datanode4 |
(1 rows)
```

此外，可以查看该语句在全局范围内各个节点上的工作情况。如下所示，每个DN上都没有在等待的阻塞资源，因为读取的数据太多而执行较慢。

```
select * from pgxc_thread_wait_status where query_id=20971544;
 node_name | db_name | thread_name | query_id | tid | lwtid | ptid | tlevel | smpid |
 wait_status | wait_event
-----+-----+-----+-----+-----+-----+-----+-----+-----+-----
 datanode1 | gaussdb | coordinator1 | 20971544 | 139902867994384 | 22735 | | 0 | 0 | wait
 node: datanode3 |
 datanode1 | gaussdb | coordinator1 | 20971544 | 139902838634256 | 22970 | 22735 | 5 | 0 |
 synchronize quit |
 datanode1 | gaussdb | coordinator1 | 20971544 | 139902607947536 | 22972 | 22735 | 5 | 1 |
 synchronize quit |
 datanode2 | gaussdb | coordinator1 | 20971544 | 140632156796688 | 22736 | | 0 | 0 | wait
 node: datanode3 |
 datanode2 | gaussdb | coordinator1 | 20971544 | 140632030967568 | 22974 | 22736 | 5 | 0 |
 synchronize quit |
 datanode2 | gaussdb | coordinator1 | 20971544 | 140632081299216 | 22975 | 22736 | 5 | 1 |
 synchronize quit |
 datanode3 | gaussdb | coordinator1 | 20971544 | 140323627988752 | 22737 | | 0 | 0 | wait
 node: datanode3 |
 datanode3 | gaussdb | coordinator1 | 20971544 | 140323523131152 | 22976 | 22737 | 5 | 0 | net
 flush data |
 datanode3 | gaussdb | coordinator1 | 20971544 | 140323548296976 | 22978 | 22737 | 5 | 1 | net
 flush data |
 datanode4 | gaussdb | coordinator1 | 20971544 | 140103024375568 | 22738 | | 0 | 0 | wait
 node: datanode3 |
 datanode4 | gaussdb | coordinator1 | 20971544 | 140102919517968 | 22979 | 22738 | 5 | 0 |
 synchronize quit |
 datanode4 | gaussdb | coordinator1 | 20971544 | 140102969849616 | 22980 | 22738 | 5 | 1 |
 synchronize quit |
 coordinator1 | gaussdb | gsql | 20971544 | 140274089064208 | 22579 | | 0 | 0 | wait node:
 datanode4 |
 (13 rows)
```

### 15.3.212 PGXC\_TOTAL\_MEMORY\_DETAIL

PGXC\_TOTAL\_MEMORY\_DETAIL视图显示集群内存使用情况。需要有系统管理员权限或预置角色gs\_role\_read\_all\_stats权限才可以访问此视图。

表 15-252 PGXC\_TOTAL\_MEMORY\_DETAIL 字段

名称	类型	描述
nodename	text	节点名称。

名称	类型	描述
memorytype	text	<p>内存使用的名称。</p> <ul style="list-style-type: none"> <li>max_process_memory: GaussDB(DWS) 集群实例所占用的内存大小。</li> <li>process_used_memory: GaussDB(DWS) 进程所使用的内存大小。</li> <li>max_dynamic_memory: 最大动态内存。</li> <li>dynamic_used_memory: 已使用的动态内存。</li> <li>dynamic_peak_memory: 内存的动态峰值。</li> <li>dynamic_used_shrctx: 最大动态共享内存上下文。</li> <li>dynamic_peak_shrctx: 共享内存上下文的动态峰值。</li> <li>max_shared_memory: 最大共享内存。</li> <li>shared_used_memory: 已使用的共享内存。</li> <li>max_cstore_memory: 列存所允许使用的最大内存。</li> <li>cstore_used_memory: 列存已使用的内存大小。</li> <li>max_sctpcomm_memory: 通信库所允许使用的最大内存。</li> <li>sctpcomm_used_memory: 通信库已使用的内存大小。</li> <li>sctpcomm_peak_memory: 通信库的内存峰值。</li> <li>other_used_memory: 其他已使用的内存大小。</li> <li>gpu_max_dynamic_memory: GPU内存最大值。</li> <li>gpu_dynamic_used_memory: 当前GPU可用内存和当前临时GPU内存之和。</li> <li>gpu_dynamic_peak_memory: GPU内存使用的最大内存。</li> <li>pooler_conn_memory: pooler连接占用内存大小。</li> <li>pooler_freeconn_memory: pooler空闲连接占用的内存大小。</li> <li>storage_compress_memory: 列存压缩和解压缩使用的内存大小。</li> </ul>

名称	类型	描述
		<ul style="list-style-type: none"> <li>udf_reserved_memory: 为UDF Worker 进程预留的内存大小。</li> <li>mmap_used_memory: mmap使用的内存大小。</li> </ul>
memorybytes	integer	内存使用的大小, 单位为MB。

### 15.3.213 PGXC\_TOTAL\_SCHEMA\_INFO

PGXC\_TOTAL\_SCHEMA\_INFO视图提供了集群所有实例上的Schema空间信息, 便于用户获悉集群各个实例上的Schema空间使用情况, 仅支持在CN节点上查询。

表 15-253 PGXC\_TOTAL\_SCHEMA\_INFO 字段

名称	类型	描述
schemaname	text	模式名称
schemaid	oid	模式OID
databasename	text	数据库名称
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

### 15.3.214 PGXC\_TOTAL\_SCHEMA\_INFO\_ANALYZE

PGXC\_TOTAL\_SCHEMA\_INFO\_ANALYZE视图提供了集群整体的Schema空间信息, 包括: 集群空间总值、各实例空间平均值、倾斜率、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, 便于用户获悉集群整体的Schema空间使用情况, 仅支持在CN节点上查询。

表 15-254 PGXC\_TOTAL\_SCHEMA\_INFO\_ANALYZE 字段

名称	类型	描述
schemaname	text	模式名称
databasename	text	数据库名称
nodegroup	text	节点组名称

名称	类型	描述
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

### 15.3.215 PGXC\_TOTAL\_USER\_RESOURCE\_INFO

PGXC\_TOTAL\_USER\_RESOURCE\_INFO视图显示所有实例上用户实时资源消耗信息。该视图仅8.2.0及以上集群版本支持。

表 15-255 PGXC\_TOTAL\_USER\_RESOURCE\_INFO 字段

名称	类型	描述
nodename	name	实例名称，包含CN和DN。
username	name	用户名。
used_memory	integer	正在使用的内存大小，单位MB。 DN：显示当前DN上对应用户正在使用的内存大小。 CN：显示所有DN上对应用户正在使用的内存累加和。
total_memory	integer	可以使用的内存大小，单位MB。值为0表示未限制最大可用内存，其限制取决于数据库最大可用内存。 DN：显示当前DN上对应用户可以使用的内存大小。 CN：显示所有DN上对应用户可以使用的内存大小之和。
used_cpu	double precision	正在使用的CPU核数（仅统计非默认资源池上复杂作业的CPU使用情况，且该值为相关控制组的CPU使用统计值）。 DN：显示当前DN上对应用户正在使用的CPU核数。 CN：显示所有DN上对应用户正在使用的CPU核数之和。



名称	类型	描述
total_cpu	integer	用户关联控制组的CPU核数总和。 DN: 显示当前DN上对应用户能够使用的CPU核数。 CN: 显示所有DN上对应用户能够使用的CPU核数之和。
used_space	bigint	已使用的永久表存储空间大小, 单位KB。 DN: 显示当前DN上对应用户已使用的永久表存储空间大小。 CN: 显示所有DN上对应用户已使用的永久表存储空间大小之和。
total_space	bigint	可使用的永久表存储空间大小, 单位KB, 值为-1表示未限制永久表存储空间。 DN: 显示当前DN上对应用户可使用的永久表存储空间大小。 CN: 显示所有DN上对应用户可使用的永久表存储空间大小之和。
used_temp_space	bigint	已使用的临时表存储空间大小, 单位KB。 DN: 显示当前DN上对应用户已使用的临时表存储空间大小。 CN: 显示所有DN上对应用户已使用的临时表存储空间大小之和。
total_temp_space	bigint	可使用的临时表存储空间大小, 单位KB, 值为-1表示未限制临时表存储空间。 DN: 显示当前DN上对应用户可使用的临时表存储空间大小。 CN: 显示所有DN上对应用户可使用的临时表存储空间大小之和。
used_spill_space	bigint	已使用的算子落盘空间大小, 单位KB。 DN: 显示当前DN上对应用户已使用的算子落盘空间大小 CN: 显示所有DN上对应用户已使用的算子落盘空间大小之和
total_spill_space	bigint	可使用的算子落盘空间大小, 单位KB, 值为-1表示未限制算子落盘空间。 DN: 显示当前DN上对应用户可使用的算子落盘空间大小。 CN: 显示所有DN上对应用户可使用的算子落盘空间大小之和。

名称	类型	描述
read_kbytes	bigint	CN: 过去5秒内, 该用户在所有DN上复杂作业read的字节总数。(单位KB) DN: 实例启动至当前时间为止, 该用户复杂作业read的字节总数。(单位KB)
write_kbytes	bigint	CN: 过去5秒内, 该用户在所有DN上复杂作业write的字节总数。(单位KB) DN: 实例启动至当前时间为止, 该用户复杂作业write的字节总数。(单位KB)
read_counts	bigint	CN: 过去5秒内, 该用户在所有DN上复杂作业read的次数之和。 DN: 实例启动至当前时间为止, 该用户复杂作业read的次数之和。
write_counts	bigint	CN: 过去5秒内, 该用户在所有DN上复杂作业write的次数之和。 DN: 实例启动至当前时间为止, 该用户复杂作业write的次数之和。
read_speed	double precision	CN: 过去5秒内, 该用户在单个DN上复杂作业read平均速率。(单位KB/s) DN: 过去5秒内, 该用户在该DN上复杂作业read平均速率。(单位KB/s)
write_speed	double precision	CN: 过去5秒内, 该用户在单个DN上复杂作业write平均速率。(单位KB/s) DN: 过去5秒内, 该用户在该DN上复杂作业write平均速率。(单位KB/s)

### 15.3.216 PGXC\_USER\_TRANSACTION

PGXC\_USER\_TRANSACTION视图提供查询所有CN上用户相关的事务信息。需要有系统管理员权限才可以访问此视图。该视图仅在资源实时监控功能开启, 即 [enable\\_resource\\_track](#) 为on时有效。

表 15-256 PGXC\_USER\_TRANSACTION 字段

名称	类型	描述
node_name	name	节点名称
username	name	用户名称
commit_counter	bigint	提交次数
rollback_counter	bigint	回滚次数
resp_min	bigint	最小响应时间

名称	类型	描述
resp_max	bigint	最大响应时间
resp_avg	bigint	平均响应时间
resp_total	bigint	响应时间总和

### 15.3.217 PGXC\_VARIABLE\_INFO

PGXC\_VARIABLE\_INFO视图用于查询集群中所有节点的xid、oid的状态。

表 15-257 PGXC\_VARIABLE\_INFO 字段

名称	类型	描述
node_name	text	节点名称
nextOid	oid	该节点下一次生成的OID
nextXid	xid	该节点下一次生成的事务号
oldestXid	xid	该节点最早的事务号。
xidVacLimit	xid	强制autovacuum的临界点
oldestXidDB	oid	该节点datafrozenid最小的数据库OID
lastExtendCSNLogpage	integer	最后一次扩展csnlog的页面号
startExtendCSNLogpage	integer	csnlog扩展的起始页面号
nextCommitSeqNo	integer	该节点下次生成的csn号
latestCompletedXid	xid	该节点提交或者回滚后节点上的最新事务号
startupMaxXid	xid	该节点关机前的最后一个事务号

### 15.3.218 PGXC\_WAIT\_DETAIL

PGXC\_WAIT\_DETAIL视图显示集群中所有节点SQL的详细等待链信息。该视图仅8.1.3.200及以上集群版本支持。

表 15-258 PGXC\_WAIT\_DETAIL 字段

名称	类型	描述
level	integer	等待链中的层级，以1开始，每显示一层等待关系level会加1。
lock_wait_hierarchy	text	等待链，以节点名称：进程号->几点名称：等待进程号->节点名称：等待进程号->...
node_name	text	节点名称。
db_name	text	database名称。
thread_name	text	线程名称。
query_id	bigint	查询语句的id。
tid	bigint	当前线程的线程号。
lwtid	integer	当前线程的轻量级线程号。
ptid	integer	streaming线程的父线程。
tlevel	integer	streaming线程的层级。
smpid	integer	并行线程的ID。
wait_status	text	当前线程的等待状态。
wait_event	text	持有此锁或者在等待此锁的事务的虚拟id。
exec_cn	boolean	是否执行sql语句的cn节点。
wait_node	text	锁级别级别。
query	text	查询语句。
application_name	text	连接到该后端的应用名。
backend_start	timestamp with time zone	后端进程启动时间，即客户端连接服务器的时间。
xact_start	timestamp with time zone	当前事务的启动时间。
query_start	timestamp with time zone	开始当前活跃查询的时间。
waiting	boolean	是否正处于等待状态。
state	text	后端当前总体状态。



### 15.3.220 PGXC\_WLM\_OPERATOR\_HISTORY

PGXC\_WLM\_OPERATOR\_HISTORY视图显示在所有CN上执行作业结束时的算子信息。此视图用于Database Manager从数据库中查询数据，数据库中的数据会被定时清理，清理周期为3分钟。

需要有系统管理员权限或预置角色gs\_role\_read\_all\_stats权限才可以访问此视图。具体的字段请参考表15-5。

### 15.3.221 PGXC\_WLM\_OPERATOR\_INFO

PGXC\_WLM\_OPERATOR\_INFO视图显示在所有CN上执行作业结束时的算子信息。此视图的数据直接从系统表GS\_WLM\_OPERATOR\_INFO获取。

需要有系统管理员权限或预置角色gs\_role\_read\_all\_stats权限才可以访问此视图。具体的字段请参考表15-5。

### 15.3.222 PGXC\_WLM\_OPERATOR\_STATISTICS

PGXC\_WLM\_OPERATOR\_STATISTICS视图显示在所有CN上正在执行作业的算子信息。

需要有系统管理员权限或预置角色gs\_role\_read\_all\_stats权限才可以访问此视图。具体的字段请参考表15-136。

### 15.3.223 PGXC\_WLM\_SESSION\_INFO

PGXC\_WLM\_SESSION\_INFO视图显示在所有CN上执行作业结束后的负载管理记录。此视图的数据直接从系统表GS\_WLM\_SESSION\_INFO获取。

具体的字段请参考表15-137。

### 15.3.224 PGXC\_WLM\_SESSION\_HISTORY

PGXC\_WLM\_SESSION\_HISTORY视图显示在所有CN上执行作业结束后的负载管理记录。此视图用于Database Manager从数据库中查询数据，数据库中的数据会被定时清理，清理周期为3分钟，详见GS\_WLM\_SESSION\_HISTORY视图介绍。

具体的字段请参考表15-137。

### 15.3.225 PGXC\_WLM\_SESSION\_STATISTICS

PGXC\_WLM\_SESSION\_STATISTICS视图显示在所有CN上正在执行的作业的负载管理信息。

具体的字段请参考表15-138。

### 15.3.226 PGXC\_WLM\_USER\_RESOURCE\_HISTORY

PGXC\_WLM\_USER\_RESOURCE\_HISTORY视图显示所有用户在对实例上资源消耗的历史信息。该视图仅8.2.0及以上集群版本支持。

表 15-259 PGXC\_WLM\_USER\_RESOURCE\_HISTORY 字段

名称	类型	描述
nodename	name	实例名称，包含CN和DN。
username	text	用户名。
timestamp	timestamp with time zone	时间戳。
used_memory	integer	正在使用的内存大小，单位MB。 DN：显示当前DN上对应用户正在使用的内存大小。 CN：显示所有DN上对应用户正在使用的内存累加和。
total_memory	integer	可以使用的内存大小，单位MB。值为0表示未限制最大可用内存，其限制取决于数据库最大可用内存。 DN：显示当前DN上对应用户可以使用的内存大小。 CN：显示所有DN上对应用户可以使用的内存大小之和。
used_cpu	double precision	正在使用的CPU核数（仅统计非默认资源池上复杂作业的CPU使用情况，且该值为相关控制组的CPU使用统计值）。 DN：显示当前DN上对应用户正在使用的CPU核数。 CN：显示所有DN上对应用户正在使用的CPU核数之和。
total_cpu	integer	用户关联控制组的CPU核数总和。 DN：显示当前DN上对应用户能够使用的CPU核数。 CN：显示所有DN上对应用户能够使用的CPU核数之和。
used_space	bigint	已使用的永久表存储空间大小，单位KB。 DN：显示当前DN上对应用户已使用的永久表存储空间大小。 CN：显示所有DN上对应用户已使用的永久表存储空间大小之和。
total_space	bigint	可使用的永久表存储空间大小，单位KB，值为-1表示未限制永久表存储空间。 DN：显示当前DN上对应用户可使用的永久表存储空间大小。 CN：显示所有DN上对应用户可使用的永久表存储空间大小之和。

名称	类型	描述
used_temp_space	bigint	已使用的临时表存储空间大小，单位KB。 DN：显示当前DN上对应用户已使用的临时表存储空间大小。 CN：显示所有DN上对应用户已使用的临时表存储空间大小之和。
total_temp_space	bigint	可使用的临时表存储空间大小，单位KB，值为-1表示未限制临时表存储空间。 DN：显示当前DN上对应用户可使用的临时表存储空间大小 CN：显示所有DN上对应用户可使用的临时表存储空间大小之和
used_spill_space	bigint	已使用的算子落盘空间大小，单位KB。 DN：显示当前DN上对应用户已使用的算子落盘空间大小。 CN：显示所有DN上对应用户已使用的算子落盘空间大小之和。
total_spill_space	bigint	可使用的算子落盘空间大小，单位KB，值为-1表示未限制算子落盘空间。 DN：显示当前DN上对应用户可使用的算子落盘空间大小。 CN：显示所有DN上对应用户可使用的算子落盘空间大小之和。
read_kbytes	bigint	CN：过去5秒内，该用户在所有DN上复杂作业read的字节总数，单位KB。 DN：实例启动至当前时间为止，该用户复杂作业read的字节总数，单位KB。
write_kbytes	bigint	CN：过去5秒内，该用户在所有DN上复杂作业write的字节总数，单位KB。 DN：实例启动至当前时间为止，该用户复杂作业write的字节总数，单位KB。
read_counts	bigint	CN：过去5秒内，该用户在所有DN上复杂作业read的次数之和。 DN：实例启动至当前时间为止，该用户复杂作业read的次数之和。
write_counts	bigint	CN：过去5秒内，该用户在所有DN上复杂作业write的次数之和。 DN：实例启动至当前时间为止，该用户复杂作业write的次数之和。



名称	类型	描述
read_speed	double precision	CN: 过去5秒内, 该用户在单个DN上复杂作业read平均速率, 单位KB/s。 DN: 过去5秒内, 该用户在该DN上复杂作业read平均速率, 单位KB/s。
write_speed	double precision	CN: 过去5秒内, 该用户在单个DN上复杂作业write平均速率, 单位KB/s。 DN: 过去5秒内, 该用户在该DN上复杂作业write平均速率, 单位KB/s。

### 15.3.227 PGXC\_WLM\_WORKLOAD\_RECORDS

PGXC\_WLM\_WORKLOAD\_RECORDS视图显示当前用户在每个CN上执行作业时CN上的状态信息。需要有系统管理员权限或预置角色gs\_role\_read\_all\_stats权限才可以访问此视图。该视图仅在动态负载功能开启, 即enable\_dynamic\_workload为on时有效。

表 15-260 PGXC\_WLM\_WORKLOAD\_RECORDS 字段

名称	类型	描述
node_name	text	作业执行所在的CN的name
thread_id	bigint	后端线程ID
processid	integer	线程的lwpid
timestamp	bigint	语句执行的开始时间
username	name	登录到该后端的用户名
memory	integer	语句所需的内存大小
active_points	integer	语句在资源池上消耗的资源点数
max_points	integer	资源在资源池上的最大资源数
priority	integer	作业的优先级
resource_pool	text	作业所在资源池
status	text	作业执行的状态, 包括: pending: 阻塞状态 running: 执行状态 finished: 结束状态 aborted: 终止状态 unkown: 未知状态
control_group	text	作业所使用的Cgroups

名称	类型	描述
enqueue	text	作业的排队信息，包括： GLOBAL: 全局排队 RESPOOL: 资源池排队 ACTIVE: 不排队
query	text	正在执行的语句

### 15.3.228 PGXC\_WORKLOAD\_SQL\_COUNT

PGXC\_WORKLOAD\_SQL\_COUNT视图显示集群中所有CN节点上的Workload控制组内的SQL语句执行次数的统计信息，包括SELECT、UPDATE、INSERT、DELETE语句的执行次数统计，以及DDL、DML、DCL类型语句的执行次数统计。需要有系统管理员权限或预置角色gs\_role\_read\_all\_stats权限才可以访问此视图。

表 15-261 PGXC\_WORKLOAD\_SQL\_COUNT 字段

名称	类型	描述
node_name	name	节点名称
workload	name	Workload控制组名称
select_count	bigint	SELECT数量
update_count	bigint	UPDATE数量
insert_count	bigint	INSERT数量
delete_count	bigint	DELETE数量
ddl_count	bigint	DDL数量
dml_count	bigint	DML数量
dcl_count	bigint	DCL数量

### 15.3.229 PGXC\_WORKLOAD\_SQL\_ELAPSE\_TIME

PGXC\_WORKLOAD\_SQL\_ELAPSE\_TIME视图显示集群中所有CN节点上Workload控制组内SQL语句执行的响应时间的统计信息，包括SELECT、UPDATE、INSERT、DELETE语句的最大、最小、平均、以及总响应时间，单位为微秒。需要有系统管理员权限或预置角色gs\_role\_read\_all\_stats权限才可以访问此视图。

表 15-262 PGXC\_WORKLOAD\_SQL\_ELAPSE\_TIME 字段

名称	类型	描述
node_name	name	节点名称

名称	类型	描述
workload	name	Workload控制组名称
total_select_elapse	bigint	SELECT总响应时间
max_select_elapse	bigint	SELECT最大响应时间
min_select_elapse	bigint	SELECT最小响应时间
avg_select_elapse	bigint	SELECT平均响应时间
total_update_elapse	bigint	UPDATE总响应时间
max_update_elapse	bigint	UPDATE最大响应时间
min_update_elapse	bigint	UPDATE最小响应时间
avg_update_elapse	bigint	UPDATE平均响应时间
total_insert_elapse	bigint	INSERT总响应时间
max_insert_elapse	bigint	INSERT最大响应时间
min_insert_elapse	bigint	INSERT最小响应时间
avg_insert_elapse	bigint	INSERT平均响应时间
total_delete_elapse	bigint	DELETE总响应时间
max_delete_elapse	bigint	DELETE最大响应时间
min_delete_elapse	bigint	DELETE最小响应时间
avg_delete_elapse	bigint	DELETE平均响应时间

### 15.3.230 PGXC\_WORKLOAD\_TRANSACTION

PGXC\_WORKLOAD\_TRANSACTION视图提供查询所有CN上Workload控制组相关的事务信息。需要有系统管理员权限或预置角色gs\_role\_read\_all\_stats权限才可以访问此视图。该视图仅在资源实时监控功能开启，即[enable\\_resource\\_track](#)为on时有效。

表 15-263 PGXC\_WORKLOAD\_TRANSACTION 字段

名称	类型	描述
node_name	name	节点名称
workload	name	Workload控制组名称
commit_counter	bigint	提交次数
rollback_counter	bigint	回滚次数
resp_min	bigint	最小响应时间，单位微秒
resp_max	bigint	最大响应时间，单位微秒

名称	类型	描述
resp_avg	bigint	平均响应时间，单位微秒
resp_total	bigint	响应时间总和，单位微秒

### 15.3.231 PLAN\_TABLE

PLAN\_TABLE显示用户通过执行EXPLAIN PLAN收集到的计划信息。计划信息的生命周期是session级别，session退出后相应的数据将被清除。同时不同session和不同user间的数据是相互隔离的。

表 15-264 PLAN\_TABLE 字段

名称	类型	描述
statement_id	varchar2(30)	用户输入的查询标签。
plan_id	bigint	查询标识。
id	int	查询生成的计划中的每一个执行算子的编号。
operation	varchar2(30)	计划中算子的操作描述。
options	varchar2(255)	操作选项。
object_name	name	操作对应的对象名，非查询中使用到的对象别名。来自于用户定义。
object_type	varchar2(30)	对象类型。
object_owner	name	对象所属schema，来自于用户定义。
projection	varchar2(4000)	操作输出的列信息。

#### 📖 说明

- object\_type取值范围为PG\_CLASS中定义的relkind类型（TABLE普通表，INDEX索引，SEQUENCE序列，VIEW视图，FOREIGN TABLE外表，COMPOSITE TYPE复合类型，TOASTVALUE TOAST表）和计划使用到的rtekind(SUBQUERY, JOIN, FUNCTION, VALUES, CTE, REMOTE\_QUERY)。
- object\_owner对于RTE来说是计划中使用的对象描述，非用户定义的类型不存在object\_owner。
- statement\_id、object\_name、object\_owner、projection字段内容遵循用户定义的大小写存储，其它字段内容采用大写存储。
- 支持用户对PLAN\_TABLE进行SELECT和DELETE操作，不支持其它DML操作。

### 15.3.232 PLAN\_TABLE\_DATA

PLAN\_TABLE\_DATA存储了用户通过执行EXPLAIN PLAN收集到的计划信息。与PLAN\_TABLE视图不同的是PLAN\_TABLE\_DATA表存储了所有session和user执行EXPLAIN PLAN收集的计划信息。

表 15-265 PLAN\_TABLE 字段

名称	类型	描述
session_id	text	表示插入该条数据的会话，由服务线程启动时间戳和服务线程ID组成。受非空约束限制。
user_id	oid	用户ID，用于标识触发插入该条数据的用户。受非空约束限制。
statement_id	varchar2(30)	用户输入的查询标签。
plan_id	bigint	查询标识。
id	int	计划中的节点编号。
operation	varchar2(30)	操作描述。
options	varchar2(255)	操作选项。
object_name	name	操作对应的对象名，来自于用户定义。
object_type	varchar2(30)	对象类型。
object_owner	name	对象所属schema，来自于用户定义。
projection	varchar2(4000)	操作输出的列信息。

#### 说明

- PLAN\_TABLE\_DATA中包含了当前节点所有用户、所有会话的数据，仅管理员有访问权限。普通用户可以通过PLAN\_TABLE视图查看属于自己的数据。
- 对于不活跃（已退出）的会话，其在PLAN\_TABLE\_DATA中的数据会在一定时间（默认5min）后被gs\_clean清理。用户也可以手动执行gs\_clean -C选项对表中不活跃的会话数据进行清理。
- PLAN\_TABLE\_DATA中的数据是用户通过执行EXPLAIN PLAN命令后由系统自动插入表中，因此禁止用户手动对数据进行插入或更新，否则会引起表中的数据混乱。需要对表中数据删除时，建议通过PLAN\_TABLE视图。
- statement\_id、object\_name、object\_owner和projection字段内容遵循用户定义的大小写存储，其它字段内容采用大写存储。

### 15.3.233 PV\_FILE\_STAT

PV\_FILE\_STAT视图通过对数据文件IO的统计，反映数据的IO性能，用以发现IO操作异常等性能问题。

表 15-266 PV\_FILE\_STAT 字段

名称	类型	描述
filenum	oid	文件标识
dbid	oid	数据库标识
spcid	oid	表空间标识
phyrds	bigint	读物理文件的数目
phywrts	bigint	写物理文件的数目
phyblkrd	bigint	读物理文件块的数目
phyblkwrt	bigint	写物理文件块的数目
readtim	bigint	读文件的总时长，单位微秒
writetim	bigint	写文件的总时长，单位微秒
avgiotim	bigint	读写文件的平均时长，单位微秒
lstiotim	bigint	最后一次读文件时长，单位微秒
miniotim	bigint	读写文件的最小时长，单位微秒
maxiowtm	bigint	读写文件的最大时长，单位微秒

### 15.3.234 PV\_INSTANCE\_TIME

PV\_INSTANCE\_TIME视图用于统计进程的运行时间信息及各执行阶段所消耗时间，单位为微秒。

提供当前节点下的各种时间消耗信息，主要分为以下类型：

- DB\_TIME: 作业在多核下的有效时间花费。
- CPU\_TIME: CPU时间的消耗。
- EXECUTION\_TIME: 执行器内花费的时间。
- PARSE\_TIME: SQL解析的时间花费。
- PLAN\_TIME: 生成Plan的时间花费。
- REWRITE\_TIME: SQL重写的时间消耗。
- PL\_EXECUTION\_TIME : plpgsql（存储过程）的执行时间。
- PL\_COMPILATION\_TIME: plpgsql（存储过程）编译时间。
- NET\_SEND\_TIME: 网络上的时间花销。
- DATA\_IO\_TIME: IO时间上的花销。

表 15-267 PV\_INSTANCE\_TIME 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
value	bigint	运行时间值

### 15.3.235 PV\_OS\_RUN\_INFO

PV\_OS\_RUN\_INFO视图显示当前操作系统运行的状态信息。

表 15-268 PV\_OS\_RUN\_INFO 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

### 15.3.236 PV\_SESSION\_MEMORY

PV\_SESSION\_MEMORY视图统计Session级别的内存使用情况，包含执行作业在数据节点上Postgres线程和Stream线程分配的所有内存。

表 15-269 PV\_SESSION\_MEMORY 字段

名称	类型	描述
sessid	text	线程启动时间+线程标识。
init_mem	integer	当前正在执行作业进入执行器前已分配的内存，单位MB。
used_mem	integer	当前正在执行作业已分配的内存，单位MB。
peak_mem	integer	当前正在执行作业已分配的内存峰值，单位MB。

### 15.3.237 PV\_SESSION\_MEMORY\_DETAIL

PV\_SESSION\_MEMORY\_DETAIL统计线程的内存使用情况，以MemoryContext节点来统计。

其中内存上下文“TempSmallContextGroup”，记录当前线程中所有内存上下文字段“totalsize”小于8192字节的信息汇总，并且内存上下文统计计数记录到“usedsize”字段中。所以在视图中，“TempSmallContextGroup”内存上下文中的“totalsize”和“freesize”是该线程中所有内存上下文“totalsize”小于8192字节的汇总总和，usedsize字段表示统计的内存上下文个数。

可通过“select \* from pv\_session\_memctx\_detail(threadid, ');”将某个线程所有内存上下文信息记录到“/tmp/dumpmem”目录下的“threadid\_timestamp.log”文件中。其中threadid可通过下表sessid中获得。

**表 15-270 PV\_SESSION\_MEMORY\_DETAIL 字段**

名称	类型	描述
sessid	text	线程启动时间+线程标识（字符串信息为timestamp.threadid）。
sesstype	text	线程名称。
contextname	text	内存上下文名称。
level	smallint	当前上下文在整体内存上下文中的层级。
parent	text	父内存上下文名称。
totalsize	bigint	当前内存上下文的内存总数，单位Byte。
freesize	bigint	当前内存上下文中已释放的内存总数，单位Byte。
usedsize	bigint	当前内存上下文中已使用的内存总数，单位Byte；“TempSmallContextGroup”内存上下文中该字段含义为统计计数。

## 应用示例

查询当前节点上所有MemoryContext的使用情况。

根据sessid定位到该MemoryContext是在哪个线程中创建和使用的，依据totalsize, freesize及usedsize来确认内存的使用情况是否符合预期，预先判断是否可能存在内存泄露的风险。

```
SELECT * FROM PV_SESSION_MEMORY_DETAIL order by totalsize desc;
      sessid      |      sesstype      |      contextname      | level |      parent
-----+-----+-----+-----+-----
| totalsize | freesize | usedsize
-----+-----+-----+-----+-----
+-----+-----+-----+-----+-----
0.139975915622720 | postmaster      |      gs_signal      |      1 |
TopMemoryContext | 17209904 | 8081136 | 9128768
1667462258.139973631031040 | postgres      |      SRF multi-call context      |      5 |
FunctionScan_139973631031040 | 1725504 | 3168 | 1722336
1667461280.139973666686720 | postgres      |      CacheMemoryContext      |      1 |
TopMemoryContext | 1472544 | 284456 | 1188088
1667450443.139973877479168 | postgres      |      CacheMemoryContext      |      1 |
TopMemoryContext | 1472544 | 356088 | 1116456
1667462258.139973631031040 | postgres      |      CacheMemoryContext      |      1 |
TopMemoryContext | 1472544 | 128216 | 1344328
1667461250.139973915236096 | postgres      |      CacheMemoryContext      |      1 |
TopMemoryContext | 1472544 | 226352 | 1246192
```



1667450439.139974010144512	WLMarbitr	CacheMemoryContext	1
TopMemoryContext	1472544   386736   1085808		
1667450439.139974151726848	WDRSnapshot	CacheMemoryContext	1
TopMemoryContext	1472544   159720   1312824		
1667450439.139974026925824	WLMmonitor	CacheMemoryContext	1
TopMemoryContext	1472544   297976   1174568		
1667451036.139973746386688	postgres	CacheMemoryContext	1
TopMemoryContext	1472544   208064   1264480		
1667461250.139973950891776	postgres	CacheMemoryContext	1
TopMemoryContext	1472544   270016   1202528		
1667450439.139974076212992	WLMCalSpaceInfo	CacheMemoryContext	1
TopMemoryContext	1472544   393952   1078592		
1667450439.139974092994304	WLMCollectWorker	CacheMemoryContext	1
TopMemoryContext	1472544   94848   1377696		
1667461254.139973971343104	postgres	CacheMemoryContext	1
TopMemoryContext	1472544   338544   1134000		
1667461280.139973822945024	postgres	CacheMemoryContext	1
TopMemoryContext	1472544   284456   1188088		
1667450439.139974202070784	JobScheduler	CacheMemoryContext	1
TopMemoryContext	1472544   216728   1255816		
1667450454.139973860697856	postgres	CacheMemoryContext	1
TopMemoryContext	1472544   388384   1084160		
0.139975915622720	postmaster	Postmaster	1
TopMemoryContext	1004288   88792   915496		
1667450439.139974218852096	AutoVacLauncher	CacheMemoryContext	1
TopMemoryContext	948256   183488   764768		
1667461250.139973915236096	postgres	TempSmallContextGroup	0
584448   148032   119			
1667462258.139973631031040	postgres	TempSmallContextGroup	0
579712   162128   123			

### 15.3.238 PV\_SESSION\_STAT

PV\_SESSION\_STAT视图以会话线程或AutoVacuum线程为单位，统计会话状态信息。

表 15-271 PV\_SESSION\_STAT 字段

名称	类型	描述
sessid	text	线程标识+线程启动时间
statid	integer	统计编号
statname	text	统计会话名称
statunit	text	统计会话单位
value	bigint	统计会话值

### 15.3.239 PV\_SESSION\_TIME

PV\_SESSION\_TIME视图用于统计会话线程的运行时间信息及各执行阶段所消耗时间，单位为微秒。

表 15-272 PV\_SESSION\_TIME 字段

名称	类型	描述
sessid	text	线程标识+线程启动时间

名称	类型	描述
stat_id	integer	统计编号
stat_name	text	运行时间类型名称
value	bigint	运行时间值

### 15.3.240 PV\_TOTAL\_MEMORY\_DETAIL

PV\_TOTAL\_MEMORY\_DETAIL视图统计当前数据库节点使用内存的信息，单位为MB。

表 15-273 PV\_TOTAL\_MEMORY\_DETAIL 字段

名称	类型	描述
nodename	text	节点名称。

名称	类型	描述
memorytype	text	<p>内存类型，包括以下几种：</p> <ul style="list-style-type: none"> <li>• max_process_memory: GaussDB(DWS)集群实例所占用的内存大小。</li> <li>• process_used_memory: GaussDB(DWS)进程所使用的内存大小。</li> <li>• max_dynamic_memory: 最大动态内存。</li> <li>• dynamic_used_memory: 已使用的动态内存。</li> <li>• dynamic_peak_memory: 内存的动态峰值。</li> <li>• dynamic_used_shrctx: 最大动态共享内存上下文。</li> <li>• dynamic_peak_shrctx: 共享内存上下文的动态峰值。</li> <li>• max_shared_memory: 最大共享内存。</li> <li>• shared_used_memory: 已使用的共享内存。</li> <li>• max_cstore_memory: 列存所允许使用的最大内存。</li> <li>• cstore_used_memory: 列存已使用的内存大小。</li> <li>• max_sctpcomm_memory: 通信库所允许使用的最大内存。</li> <li>• sctpcomm_used_memory: 通信库已使用的内存大小。</li> <li>• sctpcomm_peak_memory: 通信库的内存峰值。</li> <li>• other_used_memory: 其他已使用的内存大小。</li> <li>• gpu_max_dynamic_memory: GPU内存最大值。</li> <li>• gpu_dynamic_used_memory: 当前GPU可用内存和当前临时GPU内存之和。</li> <li>• gpu_dynamic_peak_memory: GPU内存使用的最大内存。</li> <li>• pooler_conn_memory: pooler连接占用内存大小。</li> <li>• pooler_freeconn_memory: pooler空闲连接占用的内存大小。</li> <li>• storage_compress_memory: 列存压缩和解压缩使用的内存大小。</li> <li>• udf_reserved_memory: 为UDF Worker进程预留的内存大小。</li> <li>• mmap_used_memory: mmap使用的内存大小。</li> </ul>
memorybytes	integer	内存类型分配内存的大小。

## 15.3.241 PV\_REDO\_STAT

PV\_REDO\_STAT视图提供当前节点上XLOG重做过程中的统计信息。

表 15-274 PV\_REDO\_STAT 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

## 15.3.242 PV\_RUNTIME\_ATTSTATS

PV\_RUNTIME\_ATTSTATS视图显示autoanalyze产生的内存中表级统计信息，各字段含义与PG\_STATS视图一样。

表 15-275 PV\_RUNTIME\_ATTSTATS 字段

名称	类型	描述
nspname	name	模式名
relname	name	表、索引等对象的名称
relpages	double precision	以页(大小为BLCKSZ)为单位的此表在磁盘上的大小，只是优化器使用的一个近似值。
reltuples	double precision	表中行的数目，只是优化器使用的一个估计值。
relallvisible	integer	被标识为全可见的表中的页数。此字段是优化器用来做SQL执行优化使用的。
relhasindex	boolean	如果对象是一个表且至少有（或者最近建有）一个索引，则为真。 由CREATE INDEX设置，但DROP INDEX不会立即将它清除。如果VACUUM进程检测一个表没有索引，会清理relhasindex字段，将relhasindex值设置为假。
changes	bigint	触发轻量化autoanalyze时，表的历史累计修改条数。

名称	类型	描述
level	text	轻量化autoanalyze生成的内存统计信息当前所处的阶段。包含: local, sendlist, global三个阶段。

### 15.3.243 PV\_RUNTIME\_RELSTATS

PV\_RUNTIME\_RELSTATS视图显示autoanalyze产生的内存中表级统计信息，各字段含义与PG\_CLASS视图一样。

表 15-276 PV\_RUNTIME\_RELSTATS 字段

名称	类型	描述
nspname	name	模式名
relname	name	表、索引等对象的名称
relpages	double precision	以页(大小为BLCKSZ)为单位的此表在磁盘上的大小，只是优化器使用的一个近似值。
reltuples	double precision	表中行的数目，只是优化器使用的一个估计值。
relallvisible	integer	被标识为全可见的表中的页数。此字段是优化器用来做SQL执行优化使用的。
relhasindex	boolean	如果对象是一个表且至少有（或者最近建有）一个索引，则为真。 由CREATE INDEX设置，但DROP INDEX不会立即将它清除。如果VACUUM进程检测一个表没有索引，会清理relhasindex字段，将relhasindex值设置为假。
changes	bigint	触发轻量化autoanalyze时，表的历史累计修改条数。
level	text	轻量化autoanalyze生成的内存统计信息当前所处的阶段。包含: local, sendlist, global三个阶段。

### 15.3.244 REDACTION\_COLUMNS

REDACTION\_COLUMNS视图展示当前数据库内所有脱敏列信息。

表 15-277 REDACTION\_COLUMNS 字段

名称	类型	描述
object_owner	name	脱敏对象owner

名称	类型	描述
object_name	name	脱敏对象名称
column_name	name	脱敏列名称
function_type	integer	脱敏类型
function_parameters	text	脱敏类型为partial类型时的参数（保留字段，无实际意义）
regexp_pattern	text	脱敏类型为regexp时，pattern串（保留字段，无实际意义）
regexp_replace_string	text	脱敏类型为regexp时，替换串（保留字段，无实际意义）
regexp_position	integer	脱敏类型为regexp时，起始替换位置（保留字段，无实际意义）
regexp_occurrence	integer	脱敏类型为regexp时，替换次数（保留字段，无实际意义）
regexp_match_parameter	text	脱敏类型为regexp时，正则控制参数（保留字段，无实际意义）
function_info	text	脱敏函数信息
column_description	text	脱敏列描述信息
inherited	bool	说明脱敏列是否是“继承”自其他脱敏列。

### 15.3.245 REDACTION\_POLICIES

REDACTION\_POLICIES视图展示当前数据库内所有脱敏对象信息。

表 15-278 REDACTION\_POLICIES 字段

名称	类型	描述
object_owner	name	脱敏对象owner
object_name	name	脱敏对象名称
policy_name	name	脱敏策略名称
expression	text	策略生效表达式（针对用户）
enable	boolean	策略状态（开启、关闭）
policy_description	text	策略描述信息
inherited	bool	说明脱敏列是否是“继承”自其他脱敏列。

### 15.3.246 REMOTE\_TABLE\_STAT

REMOTE\_TABLE\_STAT视图提供集群所有DN节点上当前数据库所有表的统计信息。除在每一行前面增加name类型的nodename字段外，其余字段的名称、类型和顺序与GS\_TABLE\_STAT视图相同，具体的字段请参考[GS\\_TABLE\\_STAT](#)。

### 15.3.247 USER\_COL\_COMMENTS

USER\_COL\_COMMENTS视图存储当前用户下表和视图的列注释信息。

名称	类型	描述
column_name	character varying(64)	列名
table_name	character varying(64)	表名或视图名
owner	character varying(64)	表或视图的所有者
comments	text	注释

### 15.3.248 USER\_CONSTRAINTS

USER\_CONSTRAINTS视图存储当前用户下表中约束的信息。

名称	类型	描述
constraint_name	vcharacter varying(64)	约束名
constraint_type	text	约束类型 <ul style="list-style-type: none"><li>• c表示检查约束</li><li>• f表示外键约束</li><li>• p表示主键约束</li><li>• u表示唯一约束</li></ul>
table_name	character varying(64)	约束相关的表名
index_owner	character varying(64)	约束相关的索引的所有者（只针对唯一约束和主键约束）
index_name	character varying(64)	约束相关的索引名（只针对唯一约束和主键约束）

### 15.3.249 USER\_CONS\_COLUMNS

USER\_CONS\_COLUMNS视图存储当前用户下表中约束列的信息。

名称	类型	描述
table_name	character varying(64)	约束相关的表名
column_name	character varying(64)	约束相关的列名
constraint_name	character varying(64)	约束名
position	smallint	表中列的位置

### 15.3.250 USER\_INDEXES

USER\_INDEXES视图存储关于本模式下的索引信息。

名称	类型	描述
owner	character varying(64)	索引的所有者
index_name	character varying(64)	索引名
table_name	character varying(64)	索引对应的表名
uniqueness	text	表示索引是否为唯一索引
generated	character varying(1)	表示索引名称是否为系统生成
partitioned	character(3)	表示索引是否具有分区表的性质

### 15.3.251 USER\_IND\_COLUMNS

USER\_IND\_COLUMNS视图存储当前用户下所有索引的字段信息。

名称	类型	描述
index_owner	character varying(64)	索引的所有者
index_name	character varying(64)	索引名
table_owner	character varying(64)	表的所有者
table_name	character varying(64)	表名
column_name	name	列名
column_position	smallint	索引中列的位置



### 15.3.252 USER\_IND\_EXPRESSIONS

USER\_IND\_EXPRESSIONS视图存储当前用户下基于函数的表达式索引的信息。

名称	类型	描述
index_owner	character varying(64)	索引的所有者
index_name	character varying(64)	索引名
table_owner	character varying(64)	表的所有者
table_name	character varying(64)	表名
column_expression	text	定义列的基于函数的索引表达式
column_position	smallint	索引中列的位置

### 15.3.253 USER\_IND\_PARTITIONS

USER\_IND\_PARTITIONS视图存储当前用户下的索引分区信息。

名称	类型	描述
index_owner	character varying(64)	索引分区所属分区表索引的所有者的名称
schema	character varying(64)	索引分区所属分区表索引的模式
index_name	character varying(64)	索引分区所属分区表索引的名称
partition_name	character varying(64)	索引分区的名称
index_partition_usable	boolean	索引分区是否可用
high_value	text	索引分区所对应的表分区的边界(范围分区为上边界, 列表分区为边界值集合)。 前向兼容的保留字段, 8.1.3集群版本新增pretty_high_value用于记录此信息。
pretty_high_value	text	索引分区所对应的表分区的边界(范围分区为上边界, 列表分区为边界值集合)。 查询结果为表分区对应边界表达式的即时反编译输出。该字段的输出比high_value的信息更详细, 根据实际使用场景可输出collaton、字段数据类型等信息。
def_tablespace_name	name	索引分区的表空间名称

## 15.3.254 USER\_JOBS

USER\_JOBS视图为当前用户所属定时任务的详细信息。

表 15-279 USER\_JOBS 字段

名字	类型	描述
job	int4	作业ID。
log_user	name not null	创建者的UserName。
priv_user	name not null	作业执行者的UserName。
dbname	name not null	作业创建数据库名字。
start_date	timestamp without time zone	作业的开始时间。
start_suc	text	作业成功执行的开始时间。
last_date	timestamp without time zone	上次运行开始时间。
last_suc	text	上次成功运行的开始时间。
this_date	timestamp without time zone	正在运行任务的开始时间。
this suc	text	正在运行任务成功的开始时间。
next_date	timestamp without time zone	任务下次执行时间。
next suc	text	任务下次成功执行时间。
broken	text	任务状态 如果为Y，不尝试运行此任务。 如果为N，将尝试执行此任务。

名字	类型	描述
status	char	当前任务的执行状态，取值范围：('r', 's', 'f', 'd')，默认为'r'，取值含义： <ul style="list-style-type: none"> <li>● r=running</li> <li>● s=successfully finished</li> <li>● f= job failed</li> <li>● d=aborted</li> </ul>
interval	text	用来计算下次运行时间的时间表达式，如果为nul，则表示定时任务只执行一次。
failures	smallint	失败计数，作业连续执行失败16次，不再继续执行。
what	text	可执行的作业。

### 15.3.255 USER\_OBJECTS

USER\_OBJECTS视图描述了当前用户拥有的数据库对象。

名称	类型	描述
owner	name	对象的所有者
object_name	name	对象的名称
object_id	oid	对象的OID
object_type	name	对象的类型
namespace	oid	对象所在的命名空间
created	timestamp with time zone	对象的创建时间
last_ddl_time	timestamp with time zone	对象的最后修改时间

#### 须知

created和last\_ddl\_time支持的范围参见PG\_OBJECT中的记录范围。

### 15.3.256 USER\_PART\_INDEXES

USER\_PART\_INDEXES视图存储当前用户下分区表索引的信息。

名称	类型	描述
index_owner	character varying(64)	分区表索引的所有者名称
schema	character varying(64)	分区表索引的模式
index_name	character varying(64)	分区表索引的名称
table_name	character varying(64)	分区表索引所属的分区表名称
partitioning_type	text	分区表的分区策略 <b>说明</b> 当前分区表策略仅支持范围分区（Range Partitioning）和列表分区（List Partitioning）。
partition_count	bigint	分区表索引的索引分区的个数
def_tablespace_name	name	分区表索引的表空间名称
partitioning_key_count	integer	分区表的分区键个数

### 15.3.257 USER\_PART\_TABLES

USER\_PART\_TABLES视图存储当前用户下分区表的信息。

名称	类型	描述
table_owner	character varying(64)	分区表的所有者名称
schema	character varying(64)	分区表的模式
table_name	character varying(64)	分区表的名称
partitioning_type	text	分区表的分区策略 <b>说明</b> 当前分区表策略仅支持范围分区（Range Partitioning）和列表分区（List Partitioning）。
partition_count	bigint	分区表的分区个数
def_tablespace_name	name	分区表的表空间名称
partitioning_key_count	integer	分区表的分区键个数

### 15.3.258 USER\_PROCEDURES

USER\_PROCEDURES视图存储关于本模式下的存储过程或函数信息。

名称	类型	描述
owner	character varying(64)	存储过程或函数的所有者
object_name	character varying(64)	存储过程或函数名称
argument_number	smallint	存储过程入参个数

### 15.3.259 USER\_SEQUENCES

USER\_SEQUENCES视图存储关于本模式下的序列信息。

名称	类型	描述
sequence_owner	character varying(64)	序列的所有者
sequence_name	character varying(64)	序列的名称

### 15.3.260 USER\_SOURCE

USER\_SOURCE视图存储关于本模式下的存储过程或函数信息，且提供存储过程或函数定义的字段。

名称	类型	描述
owner	character varying(64)	存储过程或函数的所有者
name	character varying(64)	存储过程或函数的名称
text	text	存储过程或函数的定义

### 15.3.261 USER\_SYNONYMS

USER\_SYNONYMS视图存储当前用户可访问的同义词信息。

表 15-280 USER\_SYNONYMS 字段

名称	类型	描述
schema_name	text	同义词所属模式名
synonym_name	text	同义词的名称
table_owner	text	关联对象的所有者

名称	类型	描述
table_schema_name	text	关联对象所属模式名
table_name	text	关联对象名

### 15.3.262 USER\_TAB\_COLUMNS

USER\_TAB\_COLUMNS视图存储当前用户可访问的表和视图字段信息。

名称	类型	描述
owner	character varying(64)	表或视图的所有者
table_name	character varying(64)	表名或视图名
column_name	character varying(64)	列名
data_type	character varying(128)	列的数据类型
column_id	integer	创建表或视图时列的序号
data_length	integer	列的字节长度
comments	text	注释
avg_column_len	numeric	列的平均长度（单位字节）
nullable	bpchar	该列是否允许为空，对于主键约束和非空约束，该值为n
data_precision	integer	数据类型的精度，对于numeric数据类型有效，其他类型为NULL
data_scale	integer	小数点右边的位数，对于numeric数据类型有效，其他类型为0
char_length	numeric	列的长度（以字符计），只对varchar, nvarchar2, bpchar, char类型有效

名称	类型	描述
schema	character varying(64)	包含该表或视图的命名空间。
kind	text	当前记录所属的种类，如果此列属于表，则此字段显示为table；如果此列属于视图，则此字段显示为view。

### 15.3.263 USER\_TAB\_COMMENTS

USER\_TAB\_COMMENTS视图存储当前用户所有表和视图的注释信息。

名称	类型	描述
owner	character varying(64)	表或视图的所有者
table_name	character varying(64)	表或视图的名称
comments	text	注释

### 15.3.264 USER\_TAB\_PARTITIONS

USER\_TAB\_PARTITIONS视图存储当前用户下所有分区的信息。当前用户下每个分区表的每个分区在USER\_TAB\_PARTITIONS中都会有一条记录。

名称	类型	描述
table_owner	character varying(64)	分区所在表的所有者
schema	character varying(64)	分区表模式
table_name	character varying(64)	表名
partition_name	character varying(64)	分区的名称
high_value	text	范围分区的上边界，或列表分区的边界值集合。前向兼容的保留字段，8.1.3集群版本新增pretty_high_value用于记录此信息。
pretty_high_value	text	范围分区的上边界，或列表分区的边界值集合。查询结果为表分区对应边界表达式的即时反编译输出。该字段的输出比high_value的信息更详细，根据实际使用场景可输出collaton、字段数据类型等信息。

名称	类型	描述
tablespace_name	name	分区所在表空间的名称

### 15.3.265 USER\_TABLES

USER\_TABLES视图存储关于当前模式下的表信息。

名称	类型	描述
owner	character varying(64)	表的所有者
table_name	character varying(64)	表名
tablespace_name	character varying(64)	表所在的表空间的名称
status	character varying(8)	当前记录是否有效
temporary	character(1)	是否为临时表 <ul style="list-style-type: none"> <li>• Y表示是临时表</li> <li>• N表示不是临时表</li> </ul>
dropped	character varying	当前记录是否已删除 <ul style="list-style-type: none"> <li>• YES表示已删除</li> <li>• NO表示未删除</li> </ul>
num_rows	numeric	表的估计行数

### 15.3.266 USER\_TRIGGERS

USER\_TRIGGERS视图存储关于当前用户下的触发器信息。

名称	类型	描述
trigger_name	character varying(64)	触发器名称
table_name	character varying(64)	定义触发器的表的名称
table_owner	character varying(64)	定义触发器的表的所有者

### 15.3.267 USER\_VIEWS

USER\_VIEWS视图存储关于当前模式下的所有视图信息。



名称	类型	描述
owner	character varying(64)	视图的所有者
view_name	character varying(64)	视图的名称

### 15.3.268 V\$SESSION

V\$SESSION视图存储当前会话的所有会话信息。

表 15-281 V\$SESSION 字段

名称	类型	描述
sid	bigint	当前活动的后台进程的OID。
serial#	integer	当前活动的后台进程的序号，在GaussDB(DWS)中为0。
user#	oid	登录此后台进程的用户的OID。
username	name	登录此后台进程的用户名。

### 15.3.269 V\$SESSION\_LONGOPS

V\$SESSION\_LONGOPS视图存储当前正在执行的操作的进度。

表 15-282 V\$SESSION\_LONGOPS 字段

名称	类型	描述
sid	bigint	当前正在执行的后台进程的OID。
serial#	integer	当前正在执行的后台进程的序号，在GaussDB(DWS)中为0。
sofar	integer	目前完成的工作量，在GaussDB(DWS)中为空。
totalwork	integer	工作总量，在GaussDB(DWS)中为空。

# 16 排序规则支持

排序规则(collation)是在字符集中指定数据排序顺序及对数据进行分类的规则。排序规则支持不再受限于数据库的LC\_COLLATE和LC\_CTYPE设置创建后就不能更改的约束。

## 概述

一种可排序数据类型的每一种表达式都有一个排序规则（系统内部的可排序数据类型可以是text、varchar和char等字符类型。用户定义的基础类型也可以被标记为可排序的，并且在一种可排序数据类型上的域也是可排序的）。如果该表达式是一个列引用，该表达式的排序规则就是列所定义的排序规则。如果该表达式是一个常量，排序规则就是该常量数据类型的默认排序规则。更复杂表达式的排序规则根据其输入的排序规则得来。

## 排序规则组合原则

- 当表达式的collation未指定时，则认为是默认的排序规则default，它表示数据库的区域设置。表达式的collation也可能是不确定的，此时，排序操作和其他不确定的排序规则的操作就会失败。
- 对于函数或操作符调用，其排序规则将通过检查所有参数的collation来决定。如果该函数或操作符调用的结果是一种可排序的数据类型，若有外层表达式要用到排序规则，那么该外层的表达式将继承对应函数和操作符所调用结果集的排序规则。
- 表达式的排序规则派生可以是显式或隐式。该区别会影响多个不同的排序规则出现在同一个表达式中时如何对collation进行组合。当执行语句使用COLLATE子句时，将发生显式派生，否则为隐式派生。当多个排序规则组合时，规则如下：
  - 如果输入表达式中存在显式COLLATE派生，则在输入表达式之间的所有显式派生的COLLATE必须相同，否则将产生冲突错误。如果存在显式COLLATE，那它就是排序规则组合的结果。
  - 如果不存在显式COLLATE，那所有输入表达式必须具有相同的隐式COLLATE或默认COLLATE。如果存在非默认COLLATE，那它就是排序规则组合的结果。否则，结果是默认COLLATE。
  - 如果在输入表达式之间存在多个冲突的非默认COLLATE，则组合被认为是具有不确定排序规则，这并非一种错误。如果被调用的函数或表达式需要用到排序规则，运行时将产生排序规则未知的错误。
- CASE表达式中，比较行为使用的规则以WHEN子句中的COLLATE设置为准。
- 显示COLLATE的派生仅在当前查询（CTE或SUBQUERY）中生效，查询外则降为隐式派生。

## 排序规则使用建议

- 同一条查询语句中，避免使用多种排序规则，可能导致非预期的结果集。
- 使用collate子句指定排序规则时，避免连续使用多个collate子句变更排序规则。

## 大小写不敏感排序规则支持

从集群8.1.3版本开始，GaussDB(DWS)增加内置排序规则case\_insensitive，即对字符类型的大小写不敏感行为（如排序、比较、哈希）。

约束条件：

- 支持字符类型：char/character/nchar、varchar/character varying/varchar2/nvarchar2/clob/text。
- 不支持字符类型：“char”和name。
- 不支持的编码：PG\_EUC\_JIS\_2004、PG\_MULE\_INTERNAL、PG\_LATIN10、PG\_WIN874。
- 不支持CREATE DATABASE时指定到LC\_COLLATE。
- 不支持正则表达式。
- 不支持字符类型的record比较（如record\_eq）。
- 不支持时序表。
- 不支持倾斜优化。
- 不支持RoughCheck优化。

## 示例

```
--语句中显示指定COLLATE子句。
SELECT 'a' = 'A', 'a' = 'A' COLLATE case_insensitive;
?column? | ?column?
-----+-----
f      | t
(1 row)
--建表时指定列属性为case_insensitive。
CREATE TABLE t1 (a text collate case_insensitive);
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
\d t1
      Table "public.t1"
Column | Type | Modifiers
-----+-----
a      | text | collate case_insensitive

INSERT INTO t1 values('a'),('A'),('b'),('B');
INSERT 0 4
--建表时指定，查询时无需指定。
SELECT a, a='a' FROM t1;
a | ?column?
---+-----
A | t
B | f
a | t
b | f
(4 rows)
SELECT a, count(1) FROM t1 GROUP BY a;
a | count
---+-----
a | 2
B | 2
```

```
(2 rows)
--CASE表达式，以WHEN子句中的COLLATE设置为准。
SELECT a,case a when 'a' collate case_insensitive then 'case1' when 'b' collate "C" then 'case2' else 'case3'
end from t1;
a | case
---+-----
A | case1
B | case3
a | case1
b | case2
(4 rows)
--跨子查询隐式派生。
SELECT * from (SELECT a collate "C" from t1) where a in ('a','b');
a
---
a
b
(2 rows)
SELECT * from t1,(SELECT a collate "C" from t1) t2 where t1.a=t2.a;
ERROR: could not determine which collation to use for string hashing
HINT: Use the COLLATE clause to set the collation explicitly.
```

---

**⚠ 注意**

- 由于collate case\_insensitive为不敏感排序，结果集不确定，再使用敏感排序筛选，会有结果集不稳定的问题，因此语句中避免出现敏感排序和不敏感排序混用。
  - 使用collate case\_insensitive指定字符类型行为为大小写不敏感后，性能较使用前会有所下降，因此性能敏感场景需谨慎评估后使用。
-

# 17 GUC 参数

## 17.1 查看 GUC 参数

GaussDB(DWS)的GUC参数影响数据库的系统行为，用户可根据业务场景和数据量查看并调整GUC参数取值。

- 查看GUC参数方式一：集群创建成功后，用户可在GaussDB(DWS) 管理控制台上查看常用的数据库参数。

名称	CN参数值	DN参数值	取值范围	是否强制设置	备注
fencedUDFMemoryLimit	0	0	0 - 2,147,483,647	否	控制每个fenced udf worker在每使用的虚拟内存。建议值: 0。
UDFWorkerMemoryLimit	1048576	104857600	0 - 2,147,483,647	是	控制fencedUDFMemoryLimit的最大值。单位: KiB。建议值: 1048576。
agg_redistribute_enhancement	off	off	-	否	当进行Agg操作时，如果包含多个group by列表且不分布列，进行重分布时会选择某一group by列表。
alarm_report_interval	100	1000	0 - 2,147,483,647	否	指定报警上报的时间间隔。建议值: 10。
allocate_mem_cost	0	0	0 - 1,7986+308	否	设置优化器计算Hash Join创建Hash表开销内存空间所需的开销。但Hash Join在算不表的优化使用...
allow_concurrent_tuple_update	on	on	-	否	设置是否允许并发更新。建议值: on。
analyze_options	ALLonLLVM_COMPILE	ALLonLLVM_COMPILE	-	否	通过开启分析器中并行功能启用相应的分析功能。包括数据校验、性能统计等。默认关闭。
archive_command			-	否	由管理员设置的用于归档WAL日志的命令。建议归档路径为绝对路径。默认值: 不设置。
archive_mode	off	off	-	否	表示是否开启归档模式。建议值: off。
archive_timeout	0	0	0 - 1,073,741,823	否	表示归档超时。建议值: 0。

- 查看GUC参数方式二：成功连接集群后，通过SQL命令的方式查看数据库GUC参数。

- 使用SHOW命令。

使用如下命令查看单个参数：

```
SHOW server_version;
```

server\_version显示数据库版本信息的参数。

使用如下命令查看所有参数：

```
SHOW ALL;
```

- 使用pg\_settings视图。

使用如下命令查看单个参数：

```
SELECT * FROM pg_settings WHERE NAME='server_version';
```

使用如下命令查看所有参数：

```
SELECT * FROM pg_settings;
```

## 17.2 设置 GUC 参数

为确保 GaussDB(DWS) 的最优性能，用户可根据业务需求对数据库中的 GUC 参数进行调整。

### 参数类型和值

- GaussDB(DWS) 的 GUC 参数类型分为以下五类：
  - SUSET，数据库管理员参数。设置后立即生效，无需重启集群。若在当前会话中设置该类型参数仅当前会话生效。
  - USERSET，普通用户参数。设置后立即生效，无需重启集群。若在当前会话中设置该类型参数仅当前会话生效。
  - POSTMASTER，数据库服务端参数。设置后需要重启集群才能生效，确认修改后系统会提示集群状态为待重启，建议在非业务高峰期手动重启集群，使参数生效。
  - SIGHUP，数据库全局参数。设置后全局生效，无法会话级生效。
  - BACKEND，数据库全局参数。设置后全局生效，无法会话级生效。
- 所有的参数名称不区分大小写。参数取值有整型、浮点型、字符串、布尔型和枚举型五类。
  - 布尔值可以是 (on, off)、(true, false)、(yes, no) 或者 (1, 0)，且不区分大小写。
  - 枚举类型的取值由系统表 pg\_settings 的 enumvals 字段取值所定义。
- 对于有单位的参数，在设置时请指定单位，否则将使用默认的单位。
  - 参数的默认单位由系统表 pg\_settings 的 unit 字段所定义。
  - 内存单位有：KB（千字节）、MB（兆字节）和 GB（吉字节）。
  - 时间单位：ms（毫秒）、s（秒）、min（分钟）、h（小时）和 d（天）。

### GUC 参数设置

GUC 参数设置有两种方式：

- 方式一：集群创建成功后，用户可以登录 GaussDB(DWS) 管理控制台，根据实际需要修改集群的数据库参数。具体操作请参见中的“[修改数据库参数](#)”章节。
- 方式二：成功连接集群后，通过 SQL 命令的方式设置 SUSET 或 USERSET 类型的参数。

修改指定数据库，用户，会话级别的参数。

- 设置数据库级别的参数

```
ALTER DATABASE dbname SET paraname TO value;
```

在下次会话中生效。

- 设置用户级别的参数

```
ALTER USER username SET paraname TO value;
```

在下次会话中生效。

- 设置会话级别的参数

```
SET paraname TO value;
```

修改本次会话中的取值。退出会话后，设置将失效。

## 操作步骤

设置参数，以设置explain\_perf\_mode参数为例。

### 步骤1 查看explain\_perf\_mode参数。

```
SHOW explain_perf_mode;
explain_perf_mode
-----
normal
(1 row)
```

### 步骤2 设置explain\_perf\_mode参数。

使用以下任意方式进行设置：

- 设置数据库级别的参数  
`ALTER DATABASE gaussdb SET explain_perf_mode TO pretty;`

当结果显示为如下信息，则表示设置成功。

```
ALTER DATABASE
```

在下次会话中生效。

- 设置用户级别的参数  
`ALTER USER dbadmin SET explain_perf_mode TO pretty;`

当结果显示为如下信息，则表示设置成功。

```
ALTER USER
```

在下次会话中生效。

- 设置会话级别的参数  
`SET explain_perf_mode TO pretty;`

当结果显示为如下信息，则表示设置成功。

```
SET
```

### 步骤3 检查参数设置的正确性。

```
SHOW explain_perf_mode;
explain_perf_mode
-----
pretty
(1 row)
```

----结束

## 17.3 GUC 使用说明

数据库提供了许多运行参数，配置这些参数可以影响数据库系统的行为。在修改这些参数时请确保用户理解了这些参数对数据库的影响，否则可能会导致无法预料的结果。

### 注意事项

- 参数中如果取值范围为字符串，此字符串应遵循操作系统的路径和文件名命名规则。
- 取值范围最大值为INT\_MAX的参数，此选项最大值跟所在的操作系统有关。
- 取值范围最大值为DBL\_MAX的参数，此选项最大值跟所在的操作系统有关。

## 17.4 连接和认证

### 17.4.1 连接设置

介绍设置客户端和服务端连接方式相关的参数。

#### max\_connections

**参数说明：**允许和数据库连接的最大并发连接数。此参数会影响集群的并发能力。

**参数类型：**POSTMASTER

**取值范围：**整型。CN最小值为1，最大值为16384；DN最小值为1，最大值为262143，由于集群内部存在着各种连接，设置时通常达不到最大值，若日志中出现'invalid value for parameter "max\_connections"'，需要调小DN的max\_connections值。

**默认值：**CN节点为800，DN节点为5000，如果该默认值超过内核支持的最大值（在执行gs\_initdb的时候判断），系统会提示错误。

**设置建议：**

CN中此参数建议保持默认值。DN中此参数按照如下公式计算：

$dop\_limit * 20 * 6 + 24$ ，公式中的dop\_limit为集群中每个DN对应的CPU数，计算公式为： $dop\_limit = \text{单机器的CPU逻辑核数} / \text{单机器的DN数}$ 。

最小值5000。

增大这个参数可能导致GaussDB(DWS)要求更多的SystemV共享内存或者信号量，可能超过操作系统缺省配置的最大值。这种情况下，请酌情对数值加以调整。

#### 须知

max\_connections取值的设置受max\_prepared\_transactions的影响，在设置max\_connections之前，应确保max\_prepared\_transactions的值大于或等于max\_connections的值，这样可确保每个会话都有一个等待中的预备事务。

#### sysadmin\_reserved\_connections

**参数说明：**为管理员用户预留的最少连接数。

**参数类型：**POSTMASTER

**取值范围：**整型，0~262143

**默认值：**3

#### application\_name

**参数说明：**连接数据库的客户端程序名称。

**参数类型：**USERSET



**取值范围：**字符串。

**默认值：**gsql

## connection\_info

**参数说明：**连接数据库的驱动类型、驱动版本号、当前驱动的部署路径和进程属主用户。（运维类参数，不建议用户设置）

**参数类型：**USERSET

**取值范围：**字符串

**默认值：**空字符串

### 说明

- 空字符串，表示当前连接数据库的驱动不支持自动设置connection\_info参数或应用程序未设置。
- 驱动连接数据库的时候自行拼接的connection\_info参数格式如下：

```
{ "driver_name": "ODBC", "driver_version": "(GaussDB 8.2.0 build 39137c2d) compiled at 2022-09-23 15:43:11 commit 3629 last mr 5138 debug", "driver_path": "/usr/local/lib/psqlodbcw.so", "os_user": "omm" }
```

ODBC, JDBC, gsql连接默认显示driver\_name和driver\_version, driver\_path, os\_user, 其他接口连接默认显示driver\_name和driver\_version, driver\_path和os\_user的显示由用户控制。

## 17.4.2 安全和认证 ( postgresql.conf )

介绍设置客户端和服务器的安全认证方式的相关参数。

### authentication\_timeout

**参数说明：**完成客户端认证的最长时间。如果一个客户端没有在这段时间里完成与服务器端的认证，则服务器自动中断与客户端的连接，这样就避免了出问题的客户端无限地占用连接数。

**参数类型：**SIGHUP

**取值范围：**整型，1~600，最小单位为秒（s）。

**默认值：**1min

### auth\_iteration\_count

**参数说明：**认证加密信息生成过程中使用的迭代次数。

**参数类型：**SIGHUP

**取值范围：**整型，2048-134217728

**默认值：**10000

### 须知

迭代次数设置过大会导致认证、用户创建等涉及口令加密的场景性能劣化，请根据实际硬件条件合理设置迭代次数。

## session\_timeout

**参数说明：**表明与服务器建立链接后，不进行任何操作的最长时间。

**参数类型：**USERSET

**取值范围：**整型，0-86400，最小单位为秒（s），0表示关闭超时设置。

**默认值：**10min

---

### 须知

- GaussDB(DWS) gsql客户端中有自动重连机制，所以针对初始化用户本地连接，超时后gsql表现的现象为断开后重连。
  - pooler连接池到其它CN和DN的连接，不受session\_timeout参数控制。
- 

## ssl

**参数说明：**启用SSL连接。

**参数类型：**POSTMASTER

**取值范围：**布尔型

- on表示启用SSL连接。
- off表示不启用SSL连接。

---

### 须知

GaussDB(DWS)目前支持SSL的场景为客户端连接CN场景，该参数目前建议只在CN中开启。

---

**默认值：**on

## ssl\_ciphers

**参数说明：**指定SSL支持的加密算法列表。

**参数类型：**POSTMASTER

**取值范围：**字符串，如果指定多个加密算法，加密算法之间需要以分号分割。

**默认值：**ALL

### 📖 说明

- 配置参数`ssl_ciphers`的默认值为ALL，表示支持下列所有加密算法。如果对加密算法没有特殊要求，建议用户使用该默认值。
  - TLS1\_3\_RFC\_AES\_128\_GCM\_SHA256
  - TLS1\_3\_RFC\_AES\_256\_GCM\_SHA384
  - TLS1\_3\_RFC\_CHACHA20\_POLY1305\_SHA256
  - TLS1\_3\_RFC\_AES\_128\_CCM\_SHA256
  - TLS1\_3\_RFC\_AES\_128\_CCM\_8\_SHA256
- SSL连接认证目前只支持配置标准协议TLS1.3的加密算法，TLS1.3的性能更优、安全性更好。同时也兼容与标准协议TLS1.2的客户端之间的SSL连接认证。

## ssl\_renegotiation\_limit

**参数说明：**指定在会话密钥重新协商之前，通过SSL加密通道可以传输的流量。这个重新协商流量限制机制可以减少攻击者针对大量数据使用密码分析法破解密钥的几率，但是也带来较大的性能损失。流量是指发送和接受的流量总和。

**参数类型：**USERSET

### 📖 说明

参数建议保持默认设置，即禁用重协商机制。不建议通过`gs_guc`工具或其他方式直接在`postgresql.conf`文件中设置`ssl_renegotiation_limit`参数，即使设置也不会生效。

**取值范围：**整型，0~INT\_MAX，单位为KB。其中0表示禁用重新协商机制。

**默认值：**0

## password\_policy

**参数说明：**在使用CREATE ROLE/USER或者ALTER ROLE/USER命令创建或者修改GaussDB(DWS)帐户时，该参数决定是否进行密码复杂度检查。

**参数类型：**SIGHUP

### 须知

从安全性考虑，请勿关闭密码复杂度策略。

**取值范围：**整型，0、1

- 0表示不采用任何密码复杂度策略。
- 1表示采用默认密码复杂度校验策略。

**默认值：**1

## password\_reuse\_time

**参数说明：**在使用ALTER USER或者ALTER ROLE修改用户密码时，该参数指定是否对新密码进行可重用天数检查。

**参数类型：**SIGHUP

**须知**

修改密码时会检查配置参数 `password_reuse_time` 和 `password_reuse_max`。

- 当 `password_reuse_time` 和 `password_reuse_max` 都为正数时，只要满足其中一个，即可认为密码可以重用。
- 当 `password_reuse_time` 为 0 时，表示不限制密码重用天数，仅限制密码重用次数。
- 当 `password_reuse_max` 为 0 时，表示不限制密码重用次数，仅限制密码重用天数。
- 当 `password_reuse_time` 和 `password_reuse_max` 都为 0 时，表示不对密码重用进行限制。

**取值范围：**浮点型，0~3650，单位为天。

- 0 表示不检查密码可重用天数。
- 正数表示新密码不能为该值指定的天数内使用过的密码。

**默认值：**60

## `password_reuse_max`

**参数说明：**在使用 ALTER USER 或者 ALTER ROLE 修改用户密码时，该参数指定是否对新密码进行可重用次数检查。

**参数类型：**SIGHUP

**须知**

修改密码时会检查配置参数 `password_reuse_time` 和 `password_reuse_max`。

- 当 `password_reuse_time` 和 `password_reuse_max` 都为正数时，只要满足其中一个，即可认为密码可以重用。
- 当 `password_reuse_time` 为 0 时，表示不限制密码重用天数，仅限制密码重用次数。
- 当 `password_reuse_max` 为 0 时，表示不限制密码重用次数，仅限制密码重用天数。
- 当 `password_reuse_time` 和 `password_reuse_max` 都为 0 时，表示不对密码重用进行限制。

**取值范围：**整型，0~1000

- 0 表示不检查密码可重用次数。
- 正整数表示新密码不能为该值指定的次数内使用过的密码。

**默认值：**0

## `password_lock_time`

**参数说明：**该参数指定帐户被锁定后自动解锁的时间。

**参数类型：** SIGHUP

#### 须知

`password_lock_time`和`failed_login_attempts`必须都为正数时锁定和解锁功能才能生效。

**取值范围：** 浮点型，0~365，单位为天。

- 0表示密码验证失败时，自动锁定功能不生效。
- 正数表示帐户被锁定后，当锁定时间超过`password_lock_time`设定的值时，帐户将会被自行解锁。

**默认值：** 1

## failed\_login\_attempts

**参数说明：** 在任意时候，如果输入密码错误的次数达到`failed_login_attempts`则当前帐户被锁定，`password_lock_time`秒后被自动解锁。例如，登录时输入密码失败，ALTER USER时修改密码失败等。

**参数类型：** SIGHUP

**取值范围：** 整型，0~1000

- 0表示自动锁定功能不生效。
- 正整数表示当错误密码次数达到`failed_login_attempts`设定的值时，当前帐户将被锁定。

**默认值：** 10

#### 须知

- `failed_login_attempts`和`password_lock_time`必须都为正数时锁定和解锁功能才能生效。
- `failed_login_attempts`会与客户端SSL连接模式共同决定用户的密码错误次数。当PGSSLMODE取值是`allow`或`prefer`时，客户的一次密码连接请求会生成两次连接请求：一次是尝试SSL连接，另一次是尝试非SSL连接。此时，用户感知到的密码错误次数是`failed_login_attempts`除以2。

## password\_encryption\_type

**参数说明：** 该字段决定采用何种加密方式对用户密码进行加密存储。

**参数类型：** SIGHUP

**取值范围：** 整型，0、1、2

- 0表示采用md5方式对密码加密。
- 1表示采用sha256方式对密码加密，兼容postgres客户端的MD5用户认证方式。
- 2表示采用sha256方式对密码加密。

### 须知

- md5为不安全的加密算法，不建议用户使用。
- 如果当前集群为8.0.0及以下版本升级到当前版本，该参数的默认值为保持前向兼容和原低版本集群一致。例如，8.0.0版本的password\_encryption\_type默认值为1，在8.0.0集群升级到8.1.1版本后，password\_encryption\_type默认值保持向前兼容仍旧是1。

默认值：2

## password\_min\_length

**参数说明：**该字段决定帐户密码的最小长度。

**参数类型：**SIGHUP

**取值范围：**整型，6~999

**默认值：**8

## password\_max\_length

**参数说明：**该字段决定帐户密码的最大长度。

**参数类型：**SIGHUP

**取值范围：**整型，6~999

**默认值：**32

## password\_min\_uppercase

**参数说明：**该字段决定帐户密码中至少需要包含大写字母个数。

**参数类型：**SIGHUP

**取值范围：**整型，0~999

- 0表示没有限制。
- 1~999表示创建账户所指定的密码中至少需要包含大写字母个数。

**默认值：**0

## password\_min\_lowercase

**参数说明：**该字段决定帐户密码中至少需要包含小写字母的个数。

**参数类型：**SIGHUP

**取值范围：**整型，0~999

- 0表示没有限制。
- 1~999表示创建帐户所指定的密码中至少需要包含小写字母个数。

**默认值：**0

## password\_min\_digital

**参数说明：**该字段决定帐户密码中至少需要包含数字的个数。

**参数类型：**SIGHUP

**取值范围：**整型，0~999

- 0表示没有限制。
- 1~999表示创建帐户所指定的密码中至少需要包含数字个数。

**默认值：**0

## password\_min\_special

**参数说明：**该字段决定帐户密码中至少需要包含个数。

**参数类型：**SIGHUP

**取值范围：**整型，0~999

- 0表示没有限制。
- 1~999表示创建帐户所指定的密码中至少需要包含特殊字符个数。

**默认值：**0

## password\_effect\_time

**参数说明：**该字段决定帐户密码的有效时间。

**参数类型：**SIGHUP

**取值范围：**浮点型，0~999，单位为天。

- 0表示不开启有效期限限制功能。
- 1~999表示创建帐户所指定的密码有效期，临近或超过有效期系统会提示用户修改密码。

**默认值：**90

## password\_notify\_time

**参数说明：**该字段决定帐户密码到期前提醒的天数。

**参数类型：**SIGHUP

**取值范围：**整型，0~999，单位为天。

- 0表示不开启提醒功能。
- 1~999表示帐户密码到期前提醒的天数。

**默认值：**7

## 17.4.3 通信库参数

本节介绍通信库相关的参数设置及取值范围等内容。

## comm\_tcp\_mode

**参数说明：**通信库使用TCP或SCTP协议建立数据通道的切换开关，重启集群生效。

**参数类型：**POSTMASTER

**取值范围：**布尔型，CN设置为on表示使用TCP模式连接DN，DN设置为on表示DN间使用TCP代理通信。

**默认值：**on

## comm\_sctp\_port

**参数说明：**通信库使用的TCP或SCTP协议监听端口。

**参数类型：**POSTMASTER

---

### 须知

集群部署时会自动分配此端口号，请不要轻易修改此参数，如端口号配置不正确会导致数据库通信失败。

---

**取值范围：**整型，0~65535

**默认值：**port+本机主DN数\*2+本DN在本机DN序号

## comm\_control\_port

**参数说明：**通信库使用的TCP协议监听端口。

**参数类型：**POSTMASTER

**取值范围：**整型，0~65535

**默认值：**port+本机主DN数\*2+本DN在本机DN序号+1

---

### 须知

集群部署时会自动分配此端口号，请不要轻易修改此参数，如端口号配置不正确会导致数据库通信失败。

---

## comm\_max\_datanode

**参数说明：**通信库支持的最大DN数。

**参数类型：**USERSET

**取值范围：**整型，1~8192

**默认值：**实际DN数



### 须知

DN数由小改大时立即生效，由大改小时需重启集群生效。

## comm\_max\_stream

**参数说明：**通信库支持的最大并发数据流数。该参数值必须大于**并发数\*每并发平均stream算子数\*（smp的平方）**。

**参数类型：**POSTMASTER

**取值范围：**整型，1~60000

**默认值：**通过公式 $\min(\text{query\_dop\_limit} * \text{query\_dop\_limit} * 2 * 20, \text{max\_process\_memory}(\text{字节}) * 0.025 / (\text{最大CN数} + \text{当前DN数}) / 260)$ 计算，小于1024按照1024取值，其中， $\text{query\_dop\_limit} = \text{单个机器CPU核数} / \text{单个机器DN数}$ 。

### 说明

- 不建议该参数值设置过大，因为comm\_max\_stream会占用内存（占用内存=256byte\*comm\_max\_stream\*comm\_max\_datanode），若并发数据流数过大，查询较为复杂及smp过大都会导致内存不足。
- 如果**comm\_max\_datanode**参数值较小，进程内存充足，可以适当将comm\_max\_stream值调大。

## comm\_max\_receiver

**参数说明：**通信库内部接收线程数量。

**参数类型：**POSTMASTER

**取值范围：**整型，1~50

**默认值：**4

## comm\_quota\_size

**参数说明：**通信库最大可连续发送包总大小。使用1GE网卡时，建议取较小值，推荐设置为20KB~40KB。

**参数类型：**USERSET

**取值范围：**整型，0~102400，默认单位为KB。0表示不使用quota机制。

**默认值：**1MB

## comm\_memory\_pool\_percent

**参数说明：**单个DN内通信库可使用内存池资源的百分比，用于自适应负载预留通信库通信消耗的内存大小。

**参数类型：**POSTMASTER

**取值范围：**整型，0~100

**默认值：**0

**须知**

此参数需根据实际业务情况做调整，若通信库使用内存小，可设置该参数数值较小，反之设置数值较大。

## comm\_client\_bind

**参数说明：**通信库客户端发起连接时是否使用bind绑定指定IP。

**参数类型：**USERSET

**取值范围：**布尔型

- on表示绑定指定IP。
- off表示不绑定指定IP。

**须知**

如果集群某一节点存在多个IP处于同一通信网段时，需设置为on。此时将绑定本地listen\_addresses指定的IP发起通信，随机端口号不能重复使用，集群并发数量会受到可用随机端口号数量的限制。

**默认值：**off

## comm\_no\_delay

**参数说明：**是否使用通信库连接的NO\_DELAY属性，重启集群生效。

**参数类型：**USERSET

**取值范围：**布尔型

**默认值：**off

**须知**

如果集群出现因每秒接收数据包过多导致的丢包时，需设置为off，以便小包合并成大大包发送，减少数据包总数。

## comm\_debug\_mode

**参数说明：**通信库debug模式开关，该参数设置是否打印通信层详细日志，session级别生效。

**须知**

设置为on时，打印日志量较大，会增加额外的overhead并降低数据库性能，仅在调试时打开，打开后及时关闭。

**参数类型：** USERSET

**取值范围：** 布尔型

- on表示打印通信库详细debug日志。
- off表示不打印通信库详细debug日志。

**默认值：** off

## comm\_ackchk\_time

**参数说明：** 无数据包接收情况下，该参数设置通信库服务端主动ACK触发时长。

**参数类型：** USERSET

**取值范围：** 整型，0~20000，单位为毫秒（ms）。取值为0表示关闭此功能。

**默认值：** 2000

## comm\_timer\_mode

**参数说明：** 通信库timer模式开关，该参数设置是否打印通信层各阶段时间桩，session级别生效。

---

### 须知

设置为on时，打印日志量较大，会增加额外的overhead并降低数据库性能，仅在调试时打开，打开后及时关闭。

---

**参数类型：** USERSET

**取值范围：** 布尔型

- on表示打印通信库详细时间桩日志。
- off表示不打印通信库详细时间桩日志。

**默认值：** off

## comm\_stat\_mode

**参数说明：** 通信库stat模式开关，该参数设置是否打印通信层的统计信息，session级别生效。

---

### 须知

设置为on时，打印日志量较大，会增加额外的overhead并降低数据库性能，仅在调试时打开，打开后及时关闭。

---

**参数类型：** USERSET

**取值范围：** 布尔型

- on表示打印通信库统计信息日志。

- off表示不打印通信库统计信息日志。

**默认值:** off

## enable\_stateless\_pooler\_reuse

**参数说明:** pooler复用切换开关，重启集群生效。

**参数类型:** POSTMASTER

**取值范围:** 布尔型

- on/true表示使用pooler复用模式。
- off/false表示关闭pooler复用模式。

### 须知

CN和DN需要同步设置。如果CN设置enable\_stateless\_pooler\_reuse为off，DN设置enable\_stateless\_pooler\_reuse为on会导致集群不能正常通信，因此必须对该参数做CN和DN全局相同的配置，重启集群生效。

**默认值:** off

## comm\_cn\_dn\_logic\_conn

**参数说明:** CN和DN间逻辑连接特性开关，重启集群生效。

**参数类型:** POSTMASTER

**取值范围:** 布尔型

- on/true表示CN和DN之间连接为逻辑链接，使用libcomm组件。
- off/false表示CN和DN之间连接为物理连接，使用libpq组件。

### 须知

如果CN设置comm\_cn\_dn\_logic\_conn为off，DN设置comm\_cn\_dn\_logic\_conn为on会导致集群不能正常通信，因此必须对该参数做CN和DN全局相同的配置，重启集群生效。

**默认值:** off

## client\_connection\_check\_interval

**参数说明:** 客户端连接状态检测时间间隔。该参数仅8.2.0及以上集群版本支持。

**参数类型:** USERSET

**取值范围:** 整型，0 ~ INT\_MAX，单位为毫秒。0表示不检测客户端连接状态。

**默认值:** 10000

**须知**

通过gsqll/jdbc/odbc等客户端直连CN执行长查询，在长查询执行期间：

- CN每隔client\_connection\_check\_interval时间检测一次客户端连接状态，若检测到客户端与CN的连接已经断开，则服务端主动终止长查询的执行，释放相关资源，避免集群资源浪费。
- DN每隔client\_connection\_check\_interval时间检测一次CN与DN的连接状态，若检测到CN与DN的连接已经断开，则DN主动终止长查询的执行，释放相关资源，避免集群资源浪费。

## 17.5 资源消耗

### 17.5.1 内存

介绍与内存相关的参数设置。

**须知**

本节涉及的参数仅在数据库服务重新启动后生效。

#### enable\_memory\_limit

**参数说明：**启用逻辑内存管理模块。

**参数类型：**POSTMASTER

**取值范围：**布尔型

- on表示启用逻辑内存管理模块。
- off表示不启用逻辑内存管理模块。

**默认值：**on

**须知**

- 若max\_process\_memory-max\_shared\_memory-cstore buffers少于2G，GaussDB(DWS)强制把enable\_memory\_limit设置为off。
- max\_shared\_memory参数与shared\_buffer、max\_connections以及max\_prepared\_transactions参数强相关，如果max\_shared\_memory过大，可以通过调整这三个参数减小。

#### max\_process\_memory

**参数说明：**设置一个数据库节点可用的最大物理内存。

**参数类型：**SIGHUP

**取值范围：**整型， $2*1024*1024 \sim INT\_MAX/2$ ，单位为KB。

**默认值：**非从备DN节点自动适配，一个机器部署多个DN情况下，公式为（物理内存大小）\* 0.8 / (1+主DN个数)；一个机器部署单个DN情况下，公式为（物理内存大小）\* 0.6；当结果不足2GB时，默认取2GB。从备DN默认为12GB。

**设置建议：**

- DN上该数值需要根据系统物理内存及单节点部署主DN个数决定的。一个机器部署多个DN情况下，`max_process_memory`计算公式如下：（物理内存大小 - `vm.min_free_kbytes`）\* 0.8 / (n+主DN个数)；一个机器部署单个DN情况下，`max_process_memory`计算公式如下：（物理内存大小 - `vm.min_free_kbytes`）\* 0.6。该参数目的是尽可能保证系统的可靠性，不会因数据库内存膨胀导致节点OOM。这个公式中提到`vm.min_free_kbytes`，其含义是预留操作系统内存供内核使用，通常用作操作系统内核中通信收发内存分配，至少为5%内存。即，`max_process_memory`=物理内存\*0.8/(n+主DN个数)，其中，当集群规模小于256时，n=1；当集群规模大于256且小于512时，n=2；当集群规模超过512时，n=3。
- 不推荐设置为最小阈值。
- CN上该数值内存可设置与DN数值一样。
- RAM：集群规划时分配给集群的最大使用内存。
- GaussDB(DWS)从8.2.0版本开始，为了提升内存资源利用率，增大了单机器单DN部署形态的`max_process_memory`初始值，但当出现集群状态不均衡的情况下，两个主DN节点会同时在一个机器上，如果仍然使用`max_process_memory`初始值，机器可能出现OOM报错的情况。为此，8.2.0版本开始，`max_process_memory`参数更改为SIGHUP类型，可以通过手动设置的方式来动态调整；同时，新增`max_process_memory_auto_adjust`参数，当出现集群状态不均衡的情况下，CM会根据集群状态来动态调整`max_process_memory`，调整`max_process_memory`的计算公式为：（物理内存大小 - `vm.min_free_kbytes`）\* 0.8 / 主DN个数。

## max\_process\_memory\_auto\_adjust

**参数说明：**设置是否开启`max_process_memory`参数的自动调整功能（该参数仅8.2.0及以上集群版本支持）。对于单机器单DN部署形态的集群，开启情况下，CM会在主备切换情况下，动态调整对应DN节点上的`max_process_memory`参数值，调整`max_process_memory`的计算公式为：（物理内存大小 - `vm.min_free_kbytes`）\* 0.8 / 主DN个数。

**参数类型：**SIGHUP

**取值范围：**布尔型

**默认值：**on

**设置建议：**建议设置为on，对于单机器单DN部署形态的集群，为提高内存资源利用率，`max_process_memory`参数的初始值在8.2.0集群版本开始进行了提升，但在主备切换场景下，两个主DN节点会同时在一个机器上，如果仍然使用`max_process_memory`初始值，机器可能出现OOM报错的情况，因此，需要CM介入来动态调整`max_process_memory`。

### 📖 说明

- 若max\_process\_memory\_auto\_adjust参数为off, max\_dynamic\_memory的值不会随着max\_process\_memory的值调整而修改。需要重启集群生效。
- 若max\_process\_memory\_auto\_adjust参数为on, max\_dynamic\_memory的值会随着max\_process\_memory的值调整而在线修改。

## shared\_buffers

**参数说明:** 设置GaussDB(DWS)使用的共享内存大小。增加此参数的值会使GaussDB(DWS)比系统默认设置需要更多的System V共享内存。

**参数类型:** POSTMASTER

**取值范围:** 整型, 128~INT\_MAX, 单位为8KB。

改变BLCKSZ的值会改变最小值。

**默认值:** CN节点为DN节点值的1/2, DN节点取公式计算:  
 $POWER(2, ROUND(\log(\max\_process\_memory * 1024 / 18, 2), 0))$ 。如果操作系统支持的共享内存小于32MB, 则在初始化数据存储区时会自动调整为操作系统支持的最大值。

**设置建议:**

由于GaussDB(DWS)大部分查询下推, 建议DN中此参数设置比CN大。

建议设置shared\_buffers值为内存的40%以内。行存列存分开对待。行存设大, 列存设小。列存: (单服务器内存/单服务器DN个数)\*0.4\*0.25。

如果设置较大的shared\_buffers需要同时增加checkpoint\_segments的值, 因为写入大量新增、修改数据需要消耗更多的时间周期。

## bulk\_write\_ring\_size

**参数说明:** 数据并行导入使用的环形缓冲区大小。

**参数类型:** USERSET

**取值范围:** 整型, 16384~INT\_MAX, 单位为KB。

**默认值:** 2GB

**设置建议:** 建议导入压力大的场景中增加DN中此参数配置。

## buffer\_ring\_ratio

**参数说明:** 设置并行导出时使用环形缓冲区的阈值大小。

**参数类型:** USERSET

**取值范围:** 整型, 1~1000

**默认值:** 250

### 📖 说明

- 默认值表示阈值为shared\_buffers的250/1000即1/4。
- 最小为shared\_buffers的1/1000。
- 最大为shared\_buffers的大小。

**设置建议：**导出时出现缓存命中率不符合预期的场景建议在DN中设置此参数。

## enable\_cstore\_ring\_buffer

**参数说明：**设置列存RingBuffer的开关，该参数仅8.2.0及以上集群版本支持。

**参数类型：**USERSET

**取值范围：**布尔型

**默认值：**off

**设置建议：**建议在业务运行一段时间，当CStoreBuffer内已存入大量高频查询的数据后，进行低频大表查询前打开，查询完大表后关闭。

## temp\_buffers

**参数说明：**设置每个数据库会话使用的LOCAL临时缓冲区的大小。

**参数类型：**USERSET

**取值范围：**整型，800~INT\_MAX/2，单位为8KB。

**默认值：**8MB

### 📖 说明

- 在每个会话的第一次使用临时表之前可以改变temp\_buffers的值，之后的设置将是无效的。
- 一个会话将按照temp\_buffers给出的限制，根据需要分配临时缓冲区。如果在一个并不需要大量临时缓冲区的会话里设置一个大的数值，其开销只是一个缓冲区描述符的大小。当缓冲区被使用，就会额外消耗8192字节。

## max\_prepared\_transactions

**参数说明：**设置可以同时处于“预备”状态的事务的最大数目。增加此参数的值会使GaussDB(DWS)比系统默认设置需要更多的System V共享内存。

当GaussDB(DWS)部署为主备双机时，在备机上此参数的设置必须要高于或等于主机上的，否则无法在备机上进行查询操作。

**参数类型：**POSTMASTER

**取值范围：**整型，0~536870911，其中CN取值为0表示关闭预备事务的特性。

**默认值：**CN节点为800，DN节点为800

### 📖 说明

为避免在准备步骤失败，此参数的值不能小于[max\\_connections](#)。



## work\_mem

**参数说明：**设置内部排序操作和Hash表在开始写入临时磁盘文件之前使用的内存大小。ORDER BY，DISTINCT和merge joins都要用到排序操作。Hash表在散列连接、散列为基础的聚集、散列为基础的IN子查询处理中都要用到。

对于复杂的查询，可能会同时并发运行好几个排序或者散列操作，每个都可以使用此参数所声明的内存量，不足时会使用临时文件。同样，好几个正在运行的会话可能会同时进行排序操作。因此使用的总内存可能是work\_mem的好几倍。

**参数类型：**USERSET

**取值范围：**整型，64~INT\_MAX，单位为KB。

**默认值：**小规格内存为512MB，大规格内存为2GB（`max_process_memory`大于等于30GB为大规格内存，否则为小规格内存）。

**设置建议：**

依据查询特点和并发来确定，一旦work\_mem限定的物理内存不够，算子运算数据将写入临时表空间，带来5-10倍的性能下降，查询响应时间从秒级下降到分钟级。

- 对于串行无并发的复杂查询场景，平均每个查询有5-10关联操作，建议work\_mem=50%内存/10。
- 对于串行无并发的简单查询场景，平均每个查询有2-5个关联操作，建议work\_mem=50%内存/5。
- 对于并发场景，建议work\_mem=串行下的work\_mem/物理并发数。

## query\_mem

**参数说明：**设置执行作业所使用的内存。如果设置的query\_mem值大于0，在生成执行计划时，优化器会将作业的估算内存调整为该值。

**参数类型：**USERSET

**取值范围：**整型，0，或大于32MB的整型，默认单位为KB。如果设置值小于32MB，系统会自动将该参数设置为默认值0，此时优化器不会根据该值调整作业的估算内存。

**默认值：**0

## query\_max\_mem

**参数说明：**设置执行作业所能够使用的最大内存。如果设置的query\_max\_mem值大于0，在生成执行计划时，优化器会根据该值来设置算子的可用内存。当作业执行时所使用内存超过该值时，将报错退出。

**参数类型：**USERSET

**取值范围：**整型，0，或大于32MB的整型，单位为KB。如果设置值为小于32MB，系统会自动将该参数设置为默认值0，此时优化器不会根据该值限制作业的内存使用。

**默认值：**0

## agg\_max\_mem

**参数说明：**设置执行作业中的Agg算子的聚集列超过5列时，该Agg算子所能够使用的最大内存。当agg\_max\_mem大于0时生效。（该参数仅8.1.3.200及以上集群版本支持）

**参数类型：**USERSET

**取值范围：**整型，0，或大于32MB的整型，单位为KB。如果设置值小于32MB，系统会自动将该参数设置为默认值0，此时不会根据该值限制Agg算子的内存使用。

**默认值：**

- 若当前集群为低版本升级到8.1.3版本，继承升级前参数，默认值为INT\_MAX。
- 若当前集群为新装的8.1.3版本，默认值为2GB。

## enable\_rowagg\_memory\_control

**参数说明：**控制行存agg算子内存使用上限。

**参数类型：**USERSET

**取值范围：**布尔型

- on表示开启控制行存agg算子内存使用限制。设置为on可避免行存agg内存过度使用导致系统内存不可用，但可能导致agg性能劣化。
- off表示关闭控制行存agg算子内存使用限制。设置为off可能导致系统内存不可用。

**默认值：**on

## maintenance\_work\_mem

**参数说明：**设置在维护性操作（比如VACUUM、CREATE INDEX、ALTER TABLE ADD FOREIGN KEY等）中可使用的最大的内存。该参数的设置会影响VACUUM、VACUUM FULL、CLUSTER、CREATE INDEX的执行效率。

**参数类型：**USERSET

**取值范围：**整型，1024~INT\_MAX，单位为KB。

**默认值：**小规格内存为512MB，大规格内存为2GB（[max\\_process\\_memory](#)大于等于30GB为大规格内存，否则为小规格内存）。

**设置建议：**

- 建议设置此参数的值等于[work\\_mem](#)，可以改进清理和恢复数据库转储的速度。因为在一个数据库会话里，任意时刻只有一个维护性操作可以执行，并且在执行维护性操作时不会有太多的会话。
- 当[自动清理](#)进程运行时，[autovacuum\\_max\\_workers](#)倍数的内存将会被分配，所以此时设置maintenance\_work\_mem的值应该不小于[work\\_mem](#)。
- 如果进行大数据量的cluster等，可以在session中调大该值。

## psort\_work\_mem

**参数说明：**设置列存表在进行局部排序中在开始写入临时磁盘文件之前使用的内存大小。带partial cluster key的表、带索引的表插入，创建表索引，删除表和更新表都会用到。

**参数类型：**USERSET

**须知**

多个正在运行的会话可能会同时进行表的局部排序操作，因此使用的总内存可能是 `psort_work_mem` 的好几倍。

**取值范围：** 整型，64~INT\_MAX，单位为KB。

**默认值：** 512MB

## max\_loaded\_cudesc

**参数说明：** 设置列存表在做扫描时，每列缓存 `cudesc` 信息的个数。增大设置会提高查询性能，但也会增加内存占用，特别是当列存表的列非常多时。

**参数类型：** USERSET

**取值范围：** 整型，100~INT\_MAX/2

**默认值：** 1024

**须知**

`max_loaded_cudesc` 设置过高时，有可能引起内存分配不足。

## max\_stack\_depth

**参数说明：** 设置 GaussDB(DWS) 执行堆栈的最大安全深度。需要这个安全界限是因为在服务器里，并非所有程序都检查了堆栈深度，只是在可能递规的过程，比如表达式计算这样的过程里面才进行检查。

**参数类型：** SUSERSET

**设置原则：**

- 此参数的最佳设置是等于操作系统内核允许的最大值（就是 `ulimit -s` 的设置）。
- 如果设置此参数的值大于实际的内核限制，则一个正在运行的递归函数可能会导致一个独立的服务器进程崩溃。在 GaussDB(DWS) 能够检测内核限制的操作系统上（SLES 上），将自动限制设置为一个不安全的值。
- 因为并非所有的操作都能够检测，所以建议用户在此设置一个明确的值。

**取值范围：** 整型，100~INT\_MAX，单位为KB。

**默认值：** 2MB

**说明**

默认值 2MB，这个值相对比较小，不容易导致系统崩溃。但是可能会因为该值较小，导致无法执行复杂的函数。

## cstore\_buffers

**参数说明：** 设置列存和 OBS、HDFS 外表列存格式（`orc`、`parquet`、`carbodata`）所使用的共享缓冲区的大小。

**参数类型：** POSTMASTER

**取值范围：** 整型，16384~INT\_MAX，单位为KB。

**默认值：** CN为32MB，DN取公式计算：  
POWER(2,ROUND(LOG(max\_process\_memory\*1024/18,2),0))

**设置建议：**

列存表使用cstore\_buffers设置的共享缓冲区，几乎不用shared\_buffers。因此在列存表为主的场景中，应减少shared\_buffers，增加cstore\_buffers。

OBS、HDFS外表使用cstore\_buffers设置ORC、Parquet、Carbondata的元数据和数据的缓存，元数据缓存大小为cstore\_buffers的1/4，最大不超过2GB，其余缓存空间为列存数据和外表列存格式数据共享使用。

## enable\_orc\_cache

**参数说明：** 设置是否允许在初始化cstore\_buffers时，将1/4的cstore\_buffers空间预留，用于缓存orc元数据。

**参数类型：** POSTMASTER

**取值范围：** 布尔型

**默认值：**

- on表示开启缓存orc元数据，可提升hdfs表的查询性能，但是会占用列存buffer资源，导致列存性能下降。
- off表示关闭缓存orc元数据。

## schedule\_splits\_threshold

**参数说明：** 设置HDFS外表schedule阶段能够在内存中存储的最大文件个数，超过该限制时，将把文件列表下盘处理。

**参数类型：** USERSET

**取值范围：** 整型，1~INT\_MAX

**默认值：** 60000

## bulk\_read\_ring\_size

**参数说明：** 并行导出，使用的环形缓冲区大小。

**参数类型：** USERSET

**取值范围：** 整型，256~INT\_MAX，单位为KB。

**默认值：** 16MB

## check\_cu\_size\_threshold

**参数说明：** 列存表插入时，如果一个CU中已插入的数据量大于该参数时，开始进行行级大小校验，避免生成非压缩态大于1G的CU。

**参数类型：** USERSET

**取值范围：**整型，0~1048576，单位为KB。

**默认值：**1GB

## max\_volatile\_memory

**参数说明：**指定所有session中volatile临时表相关context占用的最大使用总内存。在每个创建临时表的query执行时检查内存大小，所占内存不得超过该参数设置的最大内存，若超过限制则报错。该参数仅8.2.0及以上集群版本支持。

**参数类型：**SIGHUP

**取值范围：**整型，1024~INT\_MAX，单位为KB。

**默认值：**1GB

## 17.5.2 语句磁盘空间管控

介绍与语句磁盘空间管控相关的参数，用于限制语句磁盘空间使用。

### sql\_use\_spacelimit

**参数说明：**限制单个SQL在单个DN上，触发写盘操作时，所有类型写盘文件的总空间大小，管控的空间包括普通表、临时表以及中间结果集落盘占用的空间。系统管理员用户也受该参数限制。

**参数类型：**USERSET

**取值范围：**整型，-1~INT\_MAX，单位为KB。其中-1表示没有限制。

**默认值：**配置sql\_use\_spacelimit为实例所在磁盘空间总容量的10%。

#### 📖 说明

例如，执行语句中配置参数sql\_use\_spacelimit=100，当出现单DN写盘超过100kB时，DWS会主动终止该query的运行，并提示用户单DN写盘量超阈值。

```
insert into user1.t1 select * from user2.t1;  
ERROR: The space used on DN (104 kB) has exceeded the sql use space limit (100 kB).
```

建议处理方式：

- 优化语句，减少语句写盘占用空间。
- 如果磁盘空间充足可以适当调大该参数。

### temp\_file\_limit

**参数说明：**语句执行过程中触发落盘操作时，限制语句中单个线程落盘文件的总空间大小。例如，排序和哈希表使用的临时文件或者游标占用的临时文件。

此设置为会话级别的落盘文件控制。

**参数类型：**SUSET

**取值范围：**整型，-1~INT\_MAX，单位为KB。其中-1表示没有限制。

**默认值：**配置temp\_file\_limit为实例所在磁盘空间总容量的10%。

**须知**

SQL查询执行时使用的临时表空间不在此限制。

## bi\_page\_reuse\_factor

**参数说明：**行存表批量插入场景下，主备DN使用页复制进行数据同步时，可以复用的旧页面空闲空间的百分比。

**参数类型：**USERSET

**取值范围：**整型，0~100，单位为%。其中0表示不对页面进行复用，全部申请新页面。

**默认值：**70

**须知**

- 不建议将此值设置为50以下（0除外），如果复用页面的空闲空间较小的话，会使主备DN间传输过多的旧页面数据，从而导致批量插入性能下降。
- 不建议将此值设置为90以上，如果此值设置过高，会导致频繁查询空闲页面，但又无法复用旧页面，得不偿失。

## 17.5.3 内核资源使用

介绍与操作系统内核相关的参数，这些参数是否生效依赖于操作系统的设置。

### max\_files\_per\_process

**参数说明：**设置每个服务器进程允许同时打开的最大文件数目。如果操作系统内核强制一个合理的数目，则不需要设置。

但是在一些平台上（特别是大多数BSD系统），内核允许独立进程打开比系统真正可以支持的数目大得多得文件数。如果用户发现有的“Too many open files”这样的失败现象，请尝试缩小这个设置。通常情况下需要满足，系统FD（file descriptor）数量  $\geq$  最大并发数 \* 当前物理机主DN个数 \* max\_files\_per\_process \* 3。

**参数类型：**POSTMASTER

**取值范围：**整型，25~INT\_MAX

**默认值：**1000

## 17.5.4 基于开销的清理延迟

基于开销的清理延迟的目的是允许管理员减少VACUUM和ANALYZE语句在并发活动的数据库上的I/O影响。例如，VACUUM和ANALYZE此类维护语句并不需要迅速完成，且运行时要求此类语句不能严重干扰系统执行其他的数据库操作，管理员就可以使用该功能实现此目的。

**须知**

有些清理操作会持有关键的锁，这些操作应该尽快结束并释放锁。所以 GaussDB(DWS)的机制是，在这类操作过程中，基于开销的清理延迟不会发生作用。为了避免在这种情况下的长延时，实际的开销限制取下面两者之间的较大值：

- $\text{vacuum\_cost\_delay} * \text{accumulated\_balance} / \text{vacuum\_cost\_limit}$
- $\text{vacuum\_cost\_delay} * 4$

**背景信息**

在ANALYZE | ANALYSE和VACUUM语句执行过程中，系统维护一个内部的计数器，跟踪所执行的各种I/O操作的近似开销。如果积累的开销达到了vacuum\_cost\_limit声明的限制，则执行这个操作的进程将睡眠vacuum\_cost\_delay指定的时间。然后它会重置计数器然后继续执行。

这个特性是缺省关闭的。要想打开它，把vacuum\_cost\_delay变量设置为一个非零值。

**vacuum\_cost\_delay**

**参数说明：**指定开销超过vacuum\_cost\_limit的值时，进程睡眠的时间。

**参数类型：**USERSET

**取值范围：**整型，0~100，单位为毫秒（ms）。正数值表示打开基于开销的清理延迟特性；0表示关闭基于开销的清理延迟特性。

**默认值：**0

**须知**

- 许多系统上，睡眠的有效分辨率是10毫秒。因此把vacuum\_cost\_delay设置为一个不是10的整数倍的数值与将它设置为下一个10的整数倍作用相同。
- 此参数一般设置较小，常见的设置是10或20毫秒。调整此特性资源占用率时，最好是调整其他参数，而不是该参数。

**vacuum\_cost\_page\_hit**

**参数说明：**清理一个在共享缓存里找到的缓冲区的预计开销。他代表锁住缓冲池、查找共享的Hash表、扫描页面内容的开销。

**参数类型：**USERSET

**取值范围：**整型，0~10000，单位为毫秒（ms）。

**默认值：**1

**vacuum\_cost\_page\_miss**

**参数说明：**清理一个要从磁盘上读取的缓冲区的预计开销。他代表锁住缓冲池、查找共享Hash表、从磁盘读取需要的数据块、扫描它的内容的开销。

**参数类型：**USERSET



**取值范围：**整型，0~10000，单位为毫秒（ms）。

**默认值：**2

### **vacuum\_cost\_page\_dirty**

**参数说明：**清理修改一个原先是干净的块的预计开销。他代表把一个脏的磁盘块再次刷新到磁盘上的额外开销。

**参数类型：**USERSET

**取值范围：**整型，0~10000，单位为毫秒（ms）。

**默认值：**20

### **vacuum\_cost\_limit**

**参数说明：**导致清理进程休眠的开销限制。

**参数类型：**USERSET

**取值范围：**整型，1~10000，单位为毫秒（ms）。

**默认值：**200

## **17.5.5 异步 IO**

### **enable\_adio\_debug**

**参数说明：**允许维护人员输出一些与ADIO相关的日志，便于定位ADIO相关问题。开发人员专用，不建议普通用户使用。

**参数类型：**SUSET

**取值范围：**布尔型

- on/true表示开启此日志开关。
- off/false表示关闭此日志开关。

**默认值：**off

### **enable\_fast\_allocate**

**参数说明：**磁盘空间快速分配开关。只有在XFS文件系统上才能开启该开关。

**参数类型：**SUSET

**取值范围：**布尔型

- on/true表示开启此功能。
- off/false表示关闭此功能。

**默认值：**off

### **prefetch\_quantity**

**参数说明：**描述行存储使用ADIO预读取IO量的大小。



**参数类型:** USERSET

**取值范围:** 整型, 1024 ~ 1048576, 单位为8KB。

**默认值:** 32MB

## **backwrite\_quantity**

**参数说明:** 描述行存储使用ADIO写入IO量的大小。

**参数类型:** USERSET

**取值范围:** 整型, 1024 ~ 1048576, 单位为8KB。

**默认值:** 8MB

## **cstore\_prefetch\_quantity**

**参数说明:** 描述列存储使用ADIO预取IO量的大小。

**参数类型:** USERSET

**取值范围:** 整型, 1024 ~ 1048576, 单位为KB。

**默认值:** 32MB

## **cstore\_backwrite\_quantity**

**参数说明:** 描述列存储使用ADIO写入IO量的大小。

**参数类型:** USERSET

**取值范围:** 整型, 1024 ~ 1048576, 单位为KB。

**默认值:** 8MB

## **cstore\_backwrite\_max\_threshold**

**参数说明:** 描述列存储使用ADIO写入数据库可缓存最大的IO量。

**参数类型:** USERSET

**取值范围:** 整型, 4096 ~ INT\_MAX/2, 单位为KB。

**默认值:** 2GB

## **fast\_extend\_file\_size**

**参数说明:** 描述列存储使用ADIO预扩展磁盘的大小。

**参数类型:** SUSERSET

**取值范围:** 整型, 1024 ~ 1048576, 单位为KB。

**默认值:** 8MB

## **effective\_io\_concurrency**

**参数说明:** 磁盘子系统可以同时有效处理的请求数。对于RAID阵列, 此参数应该是阵列中驱动器主轴的数量。

**参数类型：**USERSET  
**取值范围：**整型，0~1000  
**默认值：**1

## 17.6 并行导入

GaussDB(DWS)提供了并行导入功能，以快速、高效地完成大量数据导入。介绍 GaussDB(DWS)并行导入的相关参数。

### raise\_errors\_if\_no\_files

**参数说明：**导入时是否区分“导入文件记录数为空”和“导入文件不存在”。  
raise\_errors\_if\_no\_files=TRUE，则“导入文件不存在”的时候，GaussDB(DWS)将抛出“文件不存在的”错误。

**参数类型：**SUSET  
**取值范围：**布尔型

- on表示导入时区分“导入文件记录数为空”和“导入文件不存在”。
- off表示导入时不区分“导入文件记录数为空”和“导入文件不存在”。

**默认值：**off

### partition\_max\_cache\_size

**参数说明：**为了优化对列存分区表的批量插入，在批量插入过程中会对数据进行缓存后再批量写盘。通过partition\_max\_cache\_size可指定数据缓存区大小。该值设置过大，将消耗较多系统内存资源；设置过小，将降低列存分区表批量插入性能。

**参数类型：**USERSET  
**取值范围：**4096~ INT\_MAX / 2，最小单位为KB。  
**默认值：**2GB

### partition\_mem\_batch

**参数说明：**为了优化对列存分区表的批量插入，在批量插入过程中会对数据进行缓存后再批量写盘。在partition\_max\_cache\_size设置的情况下，通过partition\_mem\_batch可指定缓存个数。该参数值设置越大，每个分区可用的缓存越小，降低列存分区表批量插入性能；设置越小，每个分区可用的缓存会越大，会消耗较多系统内存资源。

**参数类型：**USERSET  
**取值范围：**1~ 65535  
**默认值：**256

### gds\_debug\_mod

**参数说明：**为了增强对Gauss Data Service（以下简称GDS）相关问题的分析定位能力，可以通过此参数选择是否开启GDS的debug功能。参数开启后，将在集群节点对

应的日志中输出GDS每次收发的包裹类型、命令交互的对端以及其他交互相关的细节信息，方便记录Gaussdb端状态机的状态跳转，以及目前所处的状态信息。此参数打开会输出额外日志，增加日志IO开销，进而影响性能和日志的信息有效性，因此请仅在定位GDS问题时开启。

**参数类型：** USERSET

**取值范围：** 布尔型

- on表示开启GDS debug功能。
- off表示不开启GDS debug功能。

**默认值：** off

## enable\_delta\_store

**参数说明：** 该参数现已废弃。为了保留前向兼容，可以设置成功，但是实际不起任何作用。

开启列存delta表功能的相关内容，可参考表级参数enable\_delta。

**参数类型：** POSTMASTER

**取值范围：** 布尔型

- on表示开启列存delta表功能。
- off表示不开启列存delta表功能。

**默认值：** off

## enable\_parallel\_batch\_insert

**参数说明：** 控制行存表和列存表是否开启并发入库功能。该参数仅8.2.0.102及以上集群版本支持。

**参数类型：** USERSET

**取值范围：** 布尔型

- on表示开启并发入库功能。开启并发入库功能可以有效提升行存表和列存表insert、upsert、update和delete语句入库执行速度。
- off表示不开启并发入库功能。

**默认值：** off

# 17.7 预写式日志

## 17.7.1 设置

### wal\_level

**参数说明：** 设置写入WAL信息量的级别。

**参数类型：** POSTMASTER

**取值范围：**枚举类型

- minimal

优点：一些重要操作（包括创建表、创建索引、簇操作和表的复制）都能安全的跳过，这样就可以使操作变得更快。

缺点：WAL仅提供从数据库服务器崩溃或者紧急关闭状态恢复时所需要的基本信息，无法用WAL归档日志恢复数据。

- archive

这个参数增加了WAL归档需要的日志信息，从而可以支持数据库的归档恢复。

- hot\_standby

- 这个参数进一步增加了在备机上运行的SQL查询的信息，这个参数只能在数据库服务重新启动后生效。

- 为了在备机上开启只读查询，wal\_level必须在主机上设置成hot\_standby，并且备机必须打开hot\_standby参数。hot\_standby和archive级别之间的性能只有微小的差异，如果它们的设置对产品的性能影响有明显差异，欢迎反馈。

**默认值：** hot\_standby

**须知**

- 如果需要启用WAL日志归档和主备机的数据流复制，必须将此参数设置为archive或者hot\_standby。
- 如果此参数设置为archive，hot\_standby必须设置为off，否则将导致数据库无法启动。

## synchronous\_commit

**参数说明：**设置当前事务的同步方式。

**参数类型：**USERSET

**取值范围：**枚举类型

- on表示将备机的同步日志刷新到磁盘。
- off表示异步提交。
- local表示为本地提交。
- remote\_write表示要备机的同步日志写到磁盘。
- remote\_receive表示要备机同步日志接收数据。

**默认值：** on

## wal\_buffers

**参数说明：**设置用于存放WAL数据的共享内存空间的XLOG\_BLCKSZ数，XLOG\_BLCKSZ的大小默认为8KB。

**参数类型：**POSTMASTER

**取值范围：**-1~2<sup>18</sup>，单位为8KB。

- 如果设置为-1，表示wal\_buffers的大小随着参数shared\_buffers自动调整，为shared\_buffers的1/32，最小值为8个XLOG\_BLCKSZ，最大值为2048个XLOG\_BLCKSZ。
- 如果设置为其他值，当小于8时，会被默认设置为8；当大于2048的时，会被强制设置为2048。

**默认值：** 256MB

**设置建议：** 每次事务提交时，WAL缓冲区的内容都写入到磁盘中，因此设置为很大的值不会带来明显的性能提升。如果将它设置成几百兆，就可以在有很多即时事务提交的服务器上提高写入磁盘的性能。根据经验来说，默认值可以满足大多数的情况。

## enable\_wal\_decelerate

**参数说明：** 设置wal日志限速的总开关，该参数仅8.2.0及以上集群版本支持。

**参数类型：** SIGHUP

**取值范围：** 布尔型

- on表示启用此特性。
- off表示关闭此特性。

**默认值：** on

## wal\_decelerate\_policy

**参数说明：** 触发限速后的行为策略，该参数仅8.2.0及以上集群版本支持。

**参数类型：** USERSET

**取值范围：** 枚举类型

- warning表示触发限速条件后仅告警，但不降速。
- decelerate触发限速条件后，根据配速对当前业务进行降速处理。

**默认值：** warning

### 说明

warning不影响性能，decelerate会根据配速，对当前业务进行降速处理。

## wal\_write\_speed

**参数说明：** 单DN的每个query每秒最大允许作业触发WAL写入速率（byte/s），该参数仅8.2.0及以上集群版本支持。

**参数类型：** USERSET

**取值范围：** 整型，1024~10240000，单位为KB

**默认值：** 30MB

### 说明

对于带索引Copy、Delete等大量的作业，会根据配速进行作业速率限制。

## wal\_decelerate\_trigger\_threshold

**参数说明：**单DN每个query触发wal写入限速的阈值，该参数仅8.2.0及以上集群版本支持。

**参数类型：**USERSET

**取值范围：**整型，1024~100000000000，单位为KB

**默认值：**128MB

### 📖 说明

对于单query产生的xlog量大于该参数设置值时，即触发wal写入限速，对于DDL、少量的DML操作无影响。

## commit\_delay

**参数说明：**表示一个已经提交的数据在WAL缓冲区中存放的时间。

**参数类型：**USERSET

**取值范围：**整型，0~100000（微秒），其中0表示无延迟。

**默认值：**0

### 须知

- 设置为非0值时事务执行commit后不会立即写入WAL中，而仍存放在WAL缓冲区中，等待WalWriter进程周期性写入磁盘。
- 如果系统负载很高，在延迟时间内，其他事务可能已经准备好提交。但如果没有事务准备提交，这个延迟就是在浪费时间。

## commit\_siblings

**参数说明：**当一个事务发出提交请求时，如果数据库中正在执行的事务数量大于此参数的值，则该事务将等待一段时间（[commit\\_delay](#)的值），否则该事务则直接写入WAL。

**参数类型：**USERSET

**取值范围：**整型，0~1000

**默认值：**5

## enable\_xlog\_group\_insert

**参数说明：**控制WAL日志是否启动group的插入方式。仅鲲鹏架构支持。

**参数类型：**SIGHUP

**取值范围：**布尔型

- on表示开启。
- off表示关闭。

默认值: on

## wal\_compression

**参数说明:** 控制是否对FPI页面进行压缩。

**参数类型:** USERSET

**取值范围:** 布尔型

- on表示开启FPI压缩。
- off表示关闭FPI压缩。

默认值: on

### 须知

- 当前压缩算法为zlib，暂不支持设置为其他压缩算法。
- 对于通过从低版本升级成为当前版本的集群，此参数默认关闭（off）。如果用户需要，可以通过gs\_guc命令打开FPI压缩功能。
- 当前版本若为全新安装版本，此参数默认打开（on）。
- 从低版本升级上来的集群，如果手动开启了此参数，不允许再进行集群回滚操作。

## wal\_compression\_level

**参数说明:** 当打开wal\_compression参数时，设置zlib压缩算法的压缩级别。

**参数类型:** USERSET

**取值范围:** 整型，0~9

- 0表示不压缩。
- 1表示最低的压缩率。
- 9表示最高的压缩率。

默认值: 9

## 17.7.2 检查点

### checkpoint\_segments

**参数说明:** 设置[checkpoint\\_timeout](#)周期内所保留的最少WAL日志段文件数量。每个日志文件大小为16MB。

**参数类型:** SIGHUP

**取值范围:** 整型，最小值1

默认值: 64

**须知**

提升此参数可加快大数据的导入速度，但需要结合[checkpoint\\_timeout](#)、[shared\\_buffers](#)这两个参数统一考虑。这个参数同时影响WAL日志段文件复用数量，通常情况下pg\_xlog文件夹下最大的复用文件个数为2倍的checkpoint\_segments个，复用的文件被改名为后续即将使用的WAL日志段文件，不会被真正删除。

## checkpoint\_timeout

**参数说明：**设置自动WAL检查点之间的最长时间。

**参数类型：**SIGHUP

**取值范围：**整型，30~3600（秒）

**默认值：**15min

**须知**

在提升[checkpoint\\_segments](#)以加快大数据导入的场景也需将此参数调大，同时这两个参数提升会加大[shared\\_buffers](#)的负担，需要综合考虑。

## checkpoint\_completion\_target

**参数说明：**指定检查点完成的目标。

**参数类型：**SIGHUP

**取值范围：**0.0~1.0，其中默认值0.5表示每个checkpoint需要在checkpoints间隔时间的50%内完成。

**默认值：**0.5

## checkpoint\_warning

**参数说明：**如果由于填充检查点段文件导致检查点发生的时间间隔接近这个参数表示的秒数，就向服务器日志发送一个建议增加[checkpoint\\_segments](#)值的消息。

**参数类型：**SIGHUP

**取值范围：**整型（秒），其中0表示关闭警告。

**默认值：**5min

**推荐值：**5min

## checkpoint\_wait\_timeout

**参数说明：**设置请求检查点等待checkpointer线程启动的最长时间。

**参数类型：**SIGHUP

**取值范围：**整型，2~3600（秒）

**默认值：**1min



## 17.7.3 归档

### archive\_mode

**参数说明：**该参数控制是否将已完成的WAL段文件进行归档存储。

**参数类型：**SIGHUP

**取值范围：**布尔值

- on表示进行归档。
- off表示不进行归档。

**默认值：**off

#### 须知

当wal\_level设置成minimal时，archive\_mode参数无法使用。

### archive\_command

**参数说明：**由管理员设置的用于归档WAL日志的命令，建议归档路径为绝对路径。

**参数类型：**SIGHUP

**取值范围：**字符串

**默认值：**(disabled)

#### 须知

- 字符串中任何%p都被要归档的文件的绝对路径代替，而任何%f都只被该文件名代替（相对路径都相对于数据目录的）。如果需要在命令里嵌入%字符就必须双写%。
- 这个命令当且仅当成功的时候才返回零。示例如下：

```
archive_command = 'cp --remove-destination %p /mnt/server/archivedir/%f'
archive_command = 'copy %p /mnt/server/archivedir/%f' # Windows
```
- --remove-destination选项作用为：复制前如果目标文件已存在，会先删除已存在的目标文件，然后执行复制操作。
- 当archive\_mode为on, archive\_mode未设置时，系统会在pg\_xlog目录下新建backup目录，并将wal日志复制压缩至pg\_xlog/backup目录。

### max\_xlog\_backup\_size

**参数说明：**设置pg\_xlog/backup目录下备份WAL日志的大小。

**参数类型：**SIGHUP

**取值范围：**整型，1048576 ~ 104857600，单位为KB。

**默认值：**2097152

**须知**

- 当且仅当archive\_mode开启，archive\_command为NULL时，max\_xlog\_backup\_size参数设置才会生效。
- 每分钟检查pg\_xlog/backup目录下备份WAL日志大小，当超过max\_xlog\_backup\_size设置值时，删除最早的备份wal日志，直至备份WAL日志小于max\_xlog\_backup\_size \* 0.9。

## archive\_timeout

**参数说明：**表示归档周期。

**参数类型：**SIGHUP

**取值范围：**整型，0 ~ INT\_MAX，单位为秒。其中0表示禁用该功能。

**默认值：**0

**须知**

- 超过该参数设定的时间时强制切换WAL段。
- 由于强制切换而提早关闭的归档文件仍然与完整的归档文件长度相同。因此，将archive\_timeout设为很小的值将导致占用巨大的归档存储空间，建议将archive\_timeout设置为60秒。

## 17.8 双机复制

### 17.8.1 发送端服务器

#### wal\_keep\_segments

**参数说明：**Xlog日志文件段数量。设置“pg\_xlog”目录下保留事务日志文件的最小数目，备机通过获取主机的日志进行流复制。

**参数类型：**SIGHUP

**取值范围：**整型，2 ~ INT\_MAX

**默认值：**128

**设置建议：**

- 当服务器开启日志归档或者从检查点恢复时，保留的日志文件数量可能大于wal\_keep\_segments设定的值。
- 如果此参数设置过小，则在备机请求事务日志时，此事务日志可能已经被产生的新事务日志覆盖，导致请求失败，主备关系断开。
- 当双机为异步传输时，以COPY方式连续导入4G以上数据需要增大wal\_keep\_segments配置。以T6000单板为例，如果导入数据量为50G，建议调整参数为1000。您可以在导入完成并且日志同步正常后，动态恢复此参数设置。

## max\_replication\_slots

**参数说明：**设置主机端的日志复制slot个数。

**参数类型：**POSTMASTER

**取值范围：**整型，0 ~ 262143

**默认值：**8

物理流复制槽提供了一种自动化的方法来确保主DN在所有备DN或从备DN收到xlog之前，xlog不会被移除。也就是说物理流复制槽用于支撑集群HA。集群所需要的物理流复制槽数为：一组DN中，备加从备的和与主DN之间的比例。例如，假设集群的DN高可用方案为1主、1备、1从备，则所需物理流复制槽数为2。

关于逻辑复制槽数，请按如下规则考虑。

- 一个逻辑复制槽只能解码一个Database的修改，如果需要解码多个Database，则需要创建多个逻辑复制槽。
- 如果需要多路逻辑复制同步给多个目标数据库，在源端数据库需要创建多个逻辑复制槽，每个逻辑复制槽对应一条逻辑复制链路。

## max\_build\_io\_limit

**参数说明：**用于限制主机在提供备机重建（build）会话时，一秒时间内所允许磁盘读取的数据流量。

**参数类型：**SIGHUP

**取值范围：**整型，0~1048576，单位为KB。

**默认值：**0，表示主机对备机build无IO流控限制。

**设置建议：**可参考磁盘带宽和作业模型。无限制时或无作业干扰时，全量build在性能良好的磁盘（如SSD盘）下占磁盘带宽比例较小，磁盘IO未达到瓶颈，对业务性能影响较小，不需要设置阈值限制。在普通10000RPM转速的SAS盘下，如果build过程中，发现业务性能明显下降，可对该参数进行设置，当前推荐设置为20MB。

此设置将直接对build的进行速度和完成时间产生影响，不建议设置过低（10MB以下不建议）。在业务低峰时，建议及时取消限制，恢复build的正常速度。

### 📖 说明

- 该参数可在业务高峰期或主机磁盘IO压力较大场景时，通过限制备机build的流速阈值以减少对主机业务的影响。待业务高峰期过后，可取消限制或重新设置流速阈值。
- 具体业务场景以及磁盘性能状况，建议选择合适的阈值。

## 17.8.2 主服务器

### vacuum\_defer\_cleanup\_age

**参数说明：**指定VACUUM使用的事务数，VACUUM会延迟清除无效的行存表记录，延迟的事务个数通过vacuum\_defer\_cleanup\_age进行设置。即VACUUM和VACUUM FULL操作不会立即清理刚刚被删除元组。

**参数类型：**SIGHUP

**取值范围：**整型，0~1000000，值为0表示不延迟。

默认值: 0

## data\_replicate\_buffer\_size

**参数说明:** 发送端与接收端传递数据页时, 队列占用内存的大小。此参数会影响主备之间复制的缓冲大小。

**参数类型:** POSTMASTER

**取值范围:** 整型, 4~1023, 单位为MB。

**默认值:** CN为16MB, DN为128MB

## enable\_data\_replicate

**参数说明:** 当数据库在数据导入行存表时, 主机与备机的数据同步方式可以进行选择。

**参数类型:** USERSET

**取值范围:** 布尔型

- on表示导入数据行存表时主备数据采用数据页的方式进行同步。当 replication\_type参数为1时, 不允许设置为on。
- off表示导入数据行存表时主备数据采用日志 (Xlog) 方式进行同步。

**默认值:** on

## enable\_incremental\_catchup

**参数说明:** 控制主备之间数据追赶 (catchup) 的方式。

**参数类型:** SIGHUP

**取值范围:** 布尔类型

- on表示备机catchup时用增量catchup方式, 即从从备本地数据文件扫描获得主备差异数据文件列表, 进行主备之间的catchup。
- off表示备机catchup时用全量catchup方式, 即从主机本地所有数据文件扫描获得主备差异数据文件列表, 进行主备之间的catchup。

**默认值:** on

## wait\_dummy\_time

**参数说明:** 同时控制增量数据追赶 (catchup) 时, 集群主备从按顺序启动时等待从备启动的最长时间以及等待从备发回扫描列表的最长时间。

**参数类型:** SIGHUP

**取值范围:** 整型, 范围1~INT\_MAX, 单位为秒。

**默认值:** 300s



**注意**

单位只能设置为秒。

---

## 17.8.3 备服务器

### build\_backup\_param

**参数说明：**控制增量build过程中用于磁盘备份的最小满足规格。

**参数类型：**SIGHUP

**取值范围：**字符串

**默认值：**( 1%, 1G, 1G )

#### 📖 说明

该参数只会影响增量build过程中是否生成pg\_rewind\_bak目录，字符串生效需要采用'x %yG,zG'格式配置，**gs\_guc set**设置有效值时该参数才有效，否则不生效。其中x代表最小剩余空间百分比，y代表最小剩余空间，z代表当前磁盘总容量。

当前build实例需同时满足以下两个条件才会进行build备份生产pg\_rewind\_bak，占用额外空间；不满足则不进行build备份，不占用额外空间。

- 条件一：当前磁盘总容量  $\geq zG$ 。如果不满足该条件则不进行备份；若满足该条件会继续对条件二进行判断。
- 条件二：磁盘剩余空间  $\geq yG$ 且当前磁盘剩余空间百分比  $\geq x\%$ 。

## 17.9 查询规划

### 17.9.1 优化器方法配置

这些配置参数提供了影响查询优化器选择查询规划的原始方法。如果优化器为特定的查询选择的缺省规划并不是最优的，可以通过使用这些配置参数强制优化器选择一个不同的规划来临时解决这个问题。更好的方法包括调节优化器开销常量、手动运行ANALYZE、增加配置参数[default\\_statistics\\_target](#)的值、增加使用ALTER TABLE SET STATISTICS为指定列增加收集的统计信息。

### enable\_bitmapscan

**参数说明：**控制优化器对位图扫描规划类型的使用。

**参数类型：**USERSET

**取值范围：**布尔型

- on表示使用。
- off表示不使用。

**默认值：**on

### enable\_hashagg

**参数说明：**控制优化器对Hash聚集规划类型的使用。

**参数类型：**USERSET

**取值范围：**布尔型

- on表示使用。
- off表示不使用。

**默认值：** on

## enable\_mixedagg

**参数说明：** 控制优化器对Mixed Agg聚集规划类型的使用。（该参数仅8.2.0及以上集群版本支持）

**参数类型：** USERSET

**取值范围：** 布尔型

- on表示使用，为符合条件的Grouping Sets语句（包括Rollup或Cube）生成Mixed Agg查询计划。
- off表示不使用。

**默认值：** off

### 须知

通常在大数量场景（单个DN表的数据量在100GB以上），采用Mixed Agg查询计划可以提升语句的执行性能。

不支持Mixed Agg的场景如下：

- GROUP BY子句里列的数据类型不支持哈希操作。
- 聚集函数中包含DISTINCT去重或ORDER BY排序。
- 使用GROUPING SETS子句时不包含空的分组。

## enable\_hashjoin

**参数说明：** 控制优化器对Hash连接规划类型的使用。

**参数类型：** USERSET

**取值范围：** 布尔型

- on表示使用。
- off表示不使用。

**默认值：** on

## enable\_indexscan

**参数说明：** 控制优化器对索引扫描规划类型的使用。

**参数类型：** USERSET

**取值范围：** 布尔型

- on表示使用。
- off表示不使用。

**默认值：** on

## enable\_indexonlyscan

**参数说明：** 控制优化器对仅索引扫描规划类型的使用。

**参数类型：** USERSET

**取值范围：** 布尔型

- on表示使用。
- off表示不使用。

**默认值：** on

## enable\_material

**参数说明：** 控制优化器对实体化的使用。消除整个实体化是不可能的，但是可以关闭这个变量以防止优化器插入实体节点。

**参数类型：** USERSET

**取值范围：** 布尔型

- on表示使用。
- off表示不使用。

**默认值：** on

## enable\_mergejoin

**参数说明：** 控制优化器对融合连接规划类型的使用。

**参数类型：** USERSET

**取值范围：** 布尔型

- on表示使用。
- off表示不使用。

**默认值：** off

## enable\_nestloop

**参数说明：** 控制优化器对内表全表扫描嵌套循环连接规划类型的使用。完全消除嵌套循环连接是不可能的，但是关闭这个变量就会让优化器在存在其他方法的时候优先选择其他方法。

**参数类型：** USERSET

**取值范围：** 布尔型

- on表示使用。
- off表示不使用。

**默认值：** off

## enable\_index\_nestloop

**参数说明：**控制优化器对内表参数化索引扫描嵌套循环连接规划类型的使用。

**参数类型：**USERSET

**取值范围：**布尔型

- on表示使用。
- off表示不使用。

**默认值：**新安装集群的场景下是on，如果集群是从R8C10版本升级上来的，则保持前向兼容。如果是从R7C10或者更早的版本升级上来，则默认值是off。

## enable\_seqscan

**参数说明：**控制优化器对顺序扫描规划类型的使用。完全消除顺序扫描是不可能的，但是关闭这个变量会让优化器在存在其他方法的时候优先选择其他方法。

**参数类型：**USERSET

**取值范围：**布尔型

- on表示使用。
- off表示不使用。

**默认值：**on

## enable\_sort

**参数说明：**控制优化器使用的排序步骤。完全消除明确的排序是不可能的，但是关闭这个变量可以让优化器在存在其他方法的时候优先选择其他方法。

**参数类型：**USERSET

**取值范围：**布尔型

- on表示使用。
- off表示不使用。

**默认值：**on

## enable\_tidscan

**参数说明：**控制优化器对TID扫描规划类型的使用。

**参数类型：**USERSET

**取值范围：**布尔型

- on表示使用。
- off表示不使用。

**默认值：**on



## enable\_kill\_query

**参数说明：**CASCADE模式删除用户时，会删除此用户拥有的所有对象。此参数标识是否允许在删除用户的时候，取消锁定此用户所属对象的query。

**参数类型：**SUSET

**取值范围：**布尔型

- on表示允许取消锁定。
- off表示不允许取消锁定。

**默认值：**off

## enforce\_oracle\_behavior

**参数说明：**控制正则表达式的规则匹配模式。

**参数类型：**USERSET

**取值范围：**布尔型

- on表示正则表达式采用ORACLE格式的匹配规则。
- off表示正则表达式采用POSIX格式的匹配规则。

**默认值：**on

## enable\_stream\_concurrent\_update

**参数说明：**控制优化器在并发更新场景下对stream的使用，该参数受限于[enable\\_stream\\_operator](#)参数。

**参数类型：**USERSET

**取值范围：**布尔型

- on表示允许优化器对update语句生成stream计划。
- off表示优化器对update语句仅能生成非stream计划。

**默认值：**on

## enable\_stream\_ctescan

**参数说明：**控制stream计划是否支持ctescan。

**参数类型：**USERSET

**取值范围：**布尔型

- on表示stream计划下支持ctescan。
- off表示stream计划下不支持ctescan。

**默认值：**off

## enable\_stream\_operator

**参数说明：**控制优化器对stream的使用。

**参数类型：** USERSET

**取值范围：** 布尔型

- on表示使用。
- off表示不使用。

**默认值：** on

## enable\_stream\_recursive

**参数说明：** 控制是否将with-recursive关联查询下推DN分布式执行。

**参数类型：** USERSET

**取值范围：** 布尔型

- on表示支持使用with-recursive关联查询下推DN分布式执行。
- off表示不支持使用with\_recursive下推。

**默认值：** on

## enable\_value\_redistribute

**参数说明：** 控制是否开启生成value redistribute优化计划，8.2.0及以上集群版本中，该参数针对不带Partition by子句的rank、dense\_rank、row\_number是否生成value redistribute优化计划生效。

**参数类型：** USERSET

**取值范围：** 布尔型

- on表示支持使用value redistribute生成优化计划。
- off表示不支持使用value redistribute生成优化计划。

**默认值：** on

## max\_recursive\_times

**参数说明：** 控制with recursive的最大迭代次数。

**参数类型：** USERSET

**取值范围：** 整型，0~INT\_MAX。

**默认值：** 200

## enable\_vector\_engine

**参数说明：** 控制优化器对向量化执行引擎的使用。

**参数类型：** USERSET

**取值范围：** 布尔型

- on表示使用。
- off表示不使用。

默认值: on

## enable\_broadcast

**参数说明:** 控制优化器对stream代价估算时对broadcast分布方式的使用。

**参数类型:** USERSET

**取值范围:** 布尔型

- on表示使用。
- off表示不使用。

默认值: on

## enable\_change\_hjcost

**参数说明:** 控制优化器在Hash Join代价估算路径选择时, 是否使用将内表运行时代价排除在Hash Join节点运行时代价外的估算方式。如果使用, 则有利于选择条数少, 但运行代价大的表做内表。

**参数类型:** USERSET

**取值范围:** 布尔型

- on表示使用。
- off表示不使用。

默认值: off

## enable\_fstream

**参数说明:** 控制优化器下发语句时对stream的使用, 该参数只限定于HDFS外表使用。该参数现在已经废弃。为了保留前向兼容, 可以设置成功, 但是实际不起任何作用。

**参数类型:** USERSET

**取值范围:** 布尔型

- on表示使用。
- off表示不使用。

默认值: off

## enable\_hashfilter

**参数说明:** 控制是否允许对包含复制表(包括dual表和常量表)的计划生成hashfilter。该参数仅8.2.0及以上集群版本支持。

**参数类型:** USERSET

**取值范围:** 布尔型

- on表示允许生成hashfilter。
- off表示任何情况下都不允许生成hashfilter。

默认值: on

## best\_agg\_plan

**参数说明：**对于stream下的Agg操作，优化器会生成三种计划：

1. hashagg+gather(redistribute)+hashagg。
2. redistribute+hashagg(+gather)。
3. hashagg+redistribute+hashagg(+gather)。

本参数用于控制优化器生成哪种hashagg的计划。

**参数类型：**USERSET

**取值范围：**0, 1, 2, 3

- 取值为1时，强制生成第一种计划。
- 取值为2时，如果group by列可以重分布，强制生成第二种计划，否则生成第一种计划。
- 取值为3时，如果group by列可以重分布，强制生成第三种计划，否则生成第一种计划。
- 取值为0时，优化器会根据以上三种计划的估算cost选择最优的一种计划生成。

**默认值：**0

## volatile\_shipping\_version

**参数说明：**控制volatile函数下推执行的范围。该参数仅8.2.0及以上集群版本支持。

**参数类型：**USERSET

**取值范围：**0, 1

- 取值为1时，扩展支持nextval、uuid\_generate\_v1、sys\_guid、uuid函数出现在语句目标列时的完全下推。
- 取值为0时，支持random类函数的完全下推，nextval、uuid\_generate\_v1函数仅在INSERT含有简单查询语句的部分下推。

**默认值：**1

## agg\_redistribute\_enhancement

**参数说明：**当进行Agg操作时，如果包含多个group by列且均不为分布列，进行重分布时会选择某一group by列进行重分布。本参数控制选择重分布列的策略。

**参数类型：**USERSET

**取值范围：**布尔型

- on表示会选择估算distinct值最多的一个可重分布列作为重分布列。
- off表示会选择第一个可重分布列为重分布列。

**默认值：**off

## enable\_valuepartition\_pruning

**参数说明：**是否对DFS分区表进行静态/动态优化。

**参数类型：**USERSET

**取值范围：**布尔型

- on表示对DFS分区表进行静态/动态优化。
- off表示不对DFS分区表进行静态/动态优化。

**默认值：**on

## expected\_computing\_nodegroup

**参数说明：**标识选定的计算Node Group模式或目标计算Node Group。Node Group目前为内部用机制，用户无需设置。

共4种计算Node Group模式，用于关联操作和聚集操作时选定计算Node Group。在每一种模式中，优化器有针对性地选定几个候选计算Node Group，然后根据代价，从中为当前算子挑选更合适的计算Node Group。

**参数类型：**USERSET

**取值范围：**字符串

- optimal：候选计算Node Group列表包含算子操作对象所在的Node Group和由当前用户具有COMPUTE权限的所有Node Group包含的所有DN构成的Node Group。
- query：候选计算Node Group列表包含算子操作对象所在的Node Group和由当前查询涉及的所有基表所在Node Group包含的所有DN构成的Node Group。
- bind：当前session用户是逻辑集群用户时，候选计算Node Group为当前用户关联的逻辑集群的Node Group；当session用户不是逻辑集群用户时，候选计算Node Group选取规则和参数设置为query时的规则一致。
- Node Group名：
  - **enable\_nodegroup\_debug**为off时：候选计算Node Group列表包含算子操作对象所在的Node Group和该指定的Node Group。
  - **enable\_nodegroup\_debug**为on时：候选计算Node Group为指定的Node Group。

**默认值：**bind

## enable\_nodegroup\_debug

**参数说明：**控制优化器在多Node Group环境下，是否使用强制弹性计算。Node Group目前为内部用机制，用户无需设置。

该参数只在**expected\_computing\_nodegroup**被设置为具体Node Group时生效。

**参数类型：**USERSET

**取值范围：**布尔型

- on表示强制将计算弹性到expected\_computing\_nodegroup所指定的Node Group进行计算。
- off表示不强制使用某个Node Group进行计算。

**默认值：**off

## stream\_multiple

**参数说明：**设置优化器计算Stream算子的开销时的加权。

在原代价模型的基础上，最终Stream代价将被乘以此加权参数。

**参数类型：**USERSET

**取值范围：**浮点型，0~DBL\_MAX。

**默认值：**1

### 须知

此参数仅对Redistribute和Broadcast类型的Stream有效。

## qrw\_inlist2join\_optmode

**参数说明：**控制是否使用inlist-to-join查询重写。

**参数类型：**USERSET

**取值范围：**字符串

- disable：关闭inlist2join查询重写。
- cost\_base：基于代价的inlist2join查询重写。
- rule\_base：基于规则的inlist2join查询重写，即强制使用inlist2join查询重写。
- 任意正整数：inlist2join查询重写阈值，即list内元素个数大于该阈值，进行inlist2join查询重写。

**默认值：**cost\_base

## setop\_optmode

**参数说明：**控制不带ALL选项的集合操作(UNION/EXCEPT/INTERSECT)语句的各个查询分支语句是否执行去重操作。该参数仅8.2.0及以上集群版本支持。

**参数类型：**USERSET

**取值范围：**枚举型

- disable：查询分支不执行去重操作。
- force：强制查询分支执行去重操作。
- cost：优化器在查询分支去重和不去重这两种执行方式中，选择代价比较小的执行方式。

**默认值：**disable

### 须知

此参数配置仅在SQL语句的执行计划满足以下条件时生效：

- SQL语句中的UNION/EXCEPT/INTERSECT操作不带ALL选项。
- 执行集合操作的各个查询分支在进行集合操作前执行过数据重分布动作。

## skew\_option

**参数说明：**控制是否使用优化策略。

**参数类型：**USERSET

**取值范围：**字符串

- off：关闭策略。
- normal：采用激进策略。对于不确定是否出现倾斜的场景，认为存在倾斜，并进行相应优化。
- lazy：采用保守策略。对于不确定是否出现倾斜场景，认为不存在倾斜，不进行优化。

**默认值：**normal

## 17.9.2 优化器开销常量

介绍优化器开销常量。这里描述的开销可以按照任意标准度量。只关心其相对值，因此以相同的系数缩放它们将不会对优化器的选择产生任何影响。缺省时，它们以抓取顺序页的开销为基本单位。也就是说将seq\_page\_cost设为1.0，同时其他开销参数以它为基准设置。也可以使用其他基准，比如以毫秒计的实际执行时间。

## seq\_page\_cost

**参数说明：**设置优化器计算一次顺序磁盘页面抓取的开销。

**参数类型：**USERSET

**取值范围：**浮点型，0 ~ DBL\_MAX。

**默认值：**1

## random\_page\_cost

**参数说明：**设置优化器计算一次非顺序抓取磁盘页面的开销。

**参数类型：**USERSET

**取值范围：**浮点型，0 ~ DBL\_MAX。

**默认值：**4

### 📖 说明

- 虽然服务器允许将random\_page\_cost设置的比seq\_page\_cost小，但是物理上实际不受影响。如果所有数据库都位于随机访问内存中时，两者设置为相等很合理。因为在此种情况下，非顺序抓取页并没有副作用。同样，在缓冲率很高的数据库上，应该相对于CPU参数同时降低这两个值，因为获取内存中的页要比通常情况下开销小很多。
- 对于特别表空间中的表和索引，可以通过设置同名的表空间的参数来覆盖这个值。
- 相对于seq\_page\_cost，减少这个值将导致系统更倾向于使用索引扫描，而增加这个值使得索引扫描开销比较高。可以通过同时增加或减少这两个值来调整磁盘I/O相对于CPU的开销。

## cpu\_tuple\_cost

**参数说明：**设置优化器计算在一次查询中处理每一行数据的开销。

**参数类型：**USERSET

**取值范围：**浮点型，0~DBL\_MAX。

**默认值：**0.01

## cpu\_index\_tuple\_cost

**参数说明：**设置优化器计算在一次索引扫描中处理每条索引的开销。

**参数类型：**USERSET

**取值范围：**浮点型，0~DBL\_MAX。

**默认值：**0.005

## cpu\_operator\_cost

**参数说明：**设置优化器计算一次查询中执行一个操作符或函数的开销。

**参数类型：**USERSET

**取值范围：**浮点型，0~DBL\_MAX。

**默认值：**0.0025

## effective\_cache\_size

**参数说明：**设置优化器在一次单一的查询中可用的磁盘缓冲区的有效大小。

设置这个参数，还要考虑GaussDB(DWS)的共享缓冲区以及内核的磁盘缓冲区。另外，还要考虑预计的在不同表之间的并发查询数目，因为它们将共享可用的空间。

这个参数对GaussDB(DWS)分配的共享内存大小没有影响，它也不会使用内核磁盘缓冲，它只用于估算。数值是用磁盘页来计算的，通常每个页面是8192字节。

**参数类型：**USERSET

**取值范围：**整型，1~INT\_MAX，单位为8KB。

比默认值高的数值可能会导致使用索引扫描，更低的数值可能会导致选择顺序扫描。

**默认值：**128MB



## allocate\_mem\_cost

**参数说明：**设置优化器计算Hash Join创建Hash表开辟内存空间所需的开销，供Hash join估算不准时调优使用。

**参数类型：**USERSET

**取值范围：**浮点型，0~DBL\_MAX。

**默认值：**0

## smp\_thread\_cost

**参数说明：**设置优化器计算一个算子并行线程所需的开销，用于在生成query\_dop不合适时调优使用。（该参数仅8.2.0及以上集群版本支持）

**参数类型：**USERSET

**取值范围：**浮点型，1~10000。

**默认值：**1000

## 17.9.3 基因查询优化器

介绍基因查询优化器相关的参数。基因查询优化器（GEQO）是一种启发式的查询规划算法。这个算法减少了对复杂查询规划的时间，而且生成规划的开销有时也小于正常的详尽的查询算法。

### geqo

**参数说明：**控制基因查询优化的使用。

**参数类型：**USERSET

**取值范围：**布尔型

- on表示使用。
- off表示不使用。

**默认值：**on

#### 须知

通常情况下在执行过程中不要关闭，geqo\_threshold变量提供了更精细的控制GEQO的方法。

### geqo\_threshold

**参数说明：**如果执行语句的数量超过设计的FROM的项数，则会使用基因查询优化来执行查询。

**参数类型：**USERSET

**取值范围：**整型，2~INT\_MAX。

**默认值：**12

---

#### 须知

- 对于简单的查询，通常用详尽搜索方法，当涉及多个表的查询的时候，用GEQO可以更好的管理查询。
  - 一个FULL OUTER JOIN构造仅作为一个FROM项。
- 

### geqo\_effort

**参数说明：**控制GEQO在规划时间和规划质量之间的平衡。

**参数类型：**USERSET

**取值范围：**整型，1 ~ 10

**默认值：**5

---

#### 须知

- 比默认值大的数值增加了查询规划的时间，但是也增加了选中有效查询的几率。
  - geqo\_effort实际上并没有直接作用，只是用于计算其他影响GEQO的变量的默认值。如有需求，可以手动设置其他相关参数。
- 

### geqo\_pool\_size

**参数说明：**控制GEQO使用池的大小，也就是基因全体中的个体数量。

**参数类型：**USERSET

**取值范围：**整型，0 ~ INT\_MAX

---

#### 须知

至少是2，且有用的值一般在100到1000之间。设置为0，表示使用系统自适应方式，GaussDB(DWS)会基于geqo\_effort和表的个数选取合适的值。

---

**默认值：**0

---

### geqo\_generations

**参数说明：**控制GEQO使用的算法的迭代次数。

**参数类型：**USERSET

**取值范围：**整型，0 ~ INT\_MAX

---

#### 须知

必须至少是1，且有用的值介于100和1000之间。如果设置为0，则基于geqo\_pool\_size选取合适的值。

---

默认值: 0

## geqo\_selection\_bias

**参数说明:** 控制GEQO的选择性偏好, 即就是一个种群中的选择性压力。

**参数类型:** USERSET

**取值范围:** 浮点型, 1.5 ~ 2.0

**默认值:** 2

## geqo\_seed

**参数说明:** 控制GEQO使用的随机数生产器的初始化值, 用来从顺序连接在一起的查询空间中查找随机路径。

**参数类型:** USERSET

**取值范围:** 浮点型, 0.0 ~ 1.0

---

### 须知

不同的值会改变搜索的连接路径, 从而影响了所找路径的优劣。

---

默认值: 0

## 17.9.4 其他优化器选项

### default\_statistics\_target

**参数说明:** 为没有用ALTER TABLE SET STATISTICS设置字段目标的表设置缺省统计目标。此参数设置为正数是代表统计信息的样本数量, 为负数时, 代表使用百分比的形式设置统计目标, 负数转换为对应的百分比, 即-5代表5%。采样时, 会将 default\_statistics\_target \* 300作为随机抽样的大小, 例如默认值为100时, 会读取 100\* 300 个页面来完成随机抽样。

**参数类型:** USERSET

**取值范围:** 整型, -100 ~ 10000。

**须知**

- 比默认值大的正数数值增加了ANALYZE所需的时间，但是可能会改善优化器的估计质量。
- 调整此参数可能存在性能劣化的风险，如果某个查询劣化，可以考虑
  1. 恢复默认统计信息。
  2. 使用plan hint来调整到之前的查询计划。
- 当此guc参数设置为负数时，如果计算的采样样本数大于等于总数据量的2%，且用户表的数据量小于1600000时，ANALYZE所需时间相比guc参数为默认值的时间会有所增加。
- autoanalyze不支持临时表采样方式设置采样大小，采样过程使用参数默认值。
- 当强制使用内存方式计算统计信息时，采样大小受maintenance\_work\_mem参数限制。

**默认值：** 100

## random\_function\_version

**参数说明：** 控制analyze在进行数据采样时选取的random函数版本。该参数仅8.1.2及以上版本支持。

**参数类型：** USERSET

**取值范围：** 枚举类型

- 0 表示采用C标准库提供的random函数。
- 1 表示采用优化增强的random函数。

**默认值：**

- 若当前集群为低版本升级到8.2.0.100及以上集群版本，为保持和前向兼容，默认值为0。
- 若当前集群为新装的8.2.0.100及以上集群版本，默认值为1。

## constraint\_exclusion

**参数说明：** 控制查询优化器使用表约束查询的优化。

**参数类型：** USERSET

**取值范围：** 枚举类型

- on表示检查所有表的约束。
- off表示不检查约束。
- partition表示只检查继承的子表和UNION ALL子查询。

**须知**

当constraint\_exclusion为on，优化器用查询条件和表的CHECK约束比较，并且在查询条件和约束冲突的时候忽略对表的扫描。

**默认值:** partition

#### 📖 说明

目前, constraint\_exclusion缺省被打开, 通常用来实现表分区。为所有的表打开它时, 对于简单的查询强加了额外的规划, 并且对简单查询没有什么好处。如果不用分区表, 可以关掉它。

## cursor\_tuple\_fraction

**参数说明:** 优化器估计游标获取行数在总行数中的占比。

**参数类型:** USERSET

**取值范围:** 浮点型, 0.0 ~ 1.0。

#### 须知

比默认值小的值与使用“fast start”为游标规划的值相偏离, 从而使得前几行恢复的很快而抓取全部的行需要很长的时间。比默认值大的值加大了总的估计的时间。在最大的值1.0处, 像正常的查询一样规划游标, 只考虑总的估计时间和传送第一行的时间。

**默认值:** 0.1

## from\_collapse\_limit

**参数说明:** 根据生成的FROM列表的项数来判断优化器是否将把子查询合并到上层查询, 如果FROM列表项个数小于等于该参数值, 优化器会将子查询合并到上层查询。

**参数类型:** USERSET

**取值范围:** 整型, 1 ~ INT\_MAX。

#### 须知

比默认值小的数值将降低规划时间, 但是可能生成差的执行计划。

**默认值:** 8

## join\_collapse\_limit

**参数说明:** 根据得出的列表项数来判断优化器是否执行把除FULL JOINS之外的JOIN构造重写到FROM列表中。

**参数类型:** USERSET

**取值范围:** 整型, 1 ~ INT\_MAX。

**须知**

- 设置为1会避免任何JOIN重排。这样就使得查询中指定的连接顺序就是实际的连接顺序。查询优化器并不是总能选取最优的连接顺序，高级用户可以选择暂时把这个变量设置为1，然后指定它们需要的连接顺序。
- 比默认值小的数值减少规划时间但也降低了执行计划的质量。

默认值：8

## enable\_from\_collapse\_hint

**参数说明：**标识是否优先以hint生效的方式重写FROM列表，其次再根据from\_collapse\_limit、join\_collapse\_limit参数进行重写。该参数仅8.2.0及以上集群版本支持。

**参数类型：**USERSET

**取值范围：**布尔型

- on表示优先以hint生效的方式重写FROM列表。
- off表示无差别重写FROM列表。

**须知**

- 该参数启用时，优化器会优先以hint生效的方式重写FROM列表，但hint是否生效在计划生成后才能知道，因此有可能出现该hint不生效的情况。
- 该参数关闭时，回退到8.2.0版本之前的计划，即生成计划时不区分表是否有hint。

默认值：on

## plan\_mode\_seed

**参数说明：**该参数为调测参数，目前仅支持OPTIMIZE\_PLAN和RANDOM\_PLAN两种。其中：OPTIMIZE\_PLAN表示通过动态规划算法进行代价估算的最优plan，参数值设置为0；RANDOM\_PLAN表示随机生成的plan；如果设置为-1，表示用户不指定随机数的种子标识符seed值，由优化器随机生成[1, 2147483647]范围整型值的随机数，并根据随机数生成随机的执行计划；如果用户指定guc参数值为[1, 2147483647]范围的整型值，表示指定的生成随机数的种子标识符seed，优化器需要根据seed值生成随机的执行计划。

**参数类型：**USERSET

**取值范围：**整型，-1~ 2147483647

默认值：0

**须知**

- 当该参数设置为随机执行计划模式时，优化器会生成不同的随机执行计划，该执行计划可能不是最优计划。因此在随机计划模式下，会对查询性能产生影响，所以建议在升级、扩容、缩容等正常业务操作或运维过程中将该参数保持为默认值0。
- 当该参数不为0时，查询指定的plan hint不会生效。

## enable\_hdfs\_predicate\_pushdown

**参数说明：**标示是否启用谓词下推至原生数据层的功能。

**参数类型：**SUSET

**取值范围：**布尔型

- on表示启用谓词下推至原生数据层的功能。
- off表示不启用谓词下推至原生数据层的功能。

**默认值：**on

## windowagg\_pushdown\_enhancement

**参数说明：**标识是否在聚集场景下启用窗口函数谓词下推增强功能。（该参数仅8.2.0及以上集群版本支持）

**参数类型：**SUSET

**取值范围：**布尔型

- on表示聚集场景启用窗口函数谓词下推增强功能。
- off表示聚集场景不启用窗口函数谓词下推增强功能。

**默认值：**on

## implied\_quality\_optmode

**参数说明：**标识语句中等值列的条件传递优化策略。（该参数仅8.2.0及以上集群版本支持）

**参数类型：**SUSET

**取值范围：**枚举类型

- normal表示向前兼容8.1.3及以前版本，即推导出的表达式行为优化。
- negative表示推导出的表达式行为不优化。
- positive在normal的基础上增加了类型转换表达式的优化。

**默认值：**normal

## enable\_random\_datanode

**参数说明：**标示是否允许开启复制表DN随机查找功能，复制表在每个DN存放一份完整数据，随机选取可以缓解节点压力。

**参数类型：**USERSET

**取值范围：**布尔型

- on表示允许开启复制表DN随机查找功能。
- off表示不允许开启复制表DN随机查找功能。

**默认值：**on

## hashagg\_table\_size

**参数说明：**用于设置执行HASH AGG操作时HASH表的大小。

**参数类型：**USERSET

**取值范围：**整型，0~INT\_MAX/2。

**默认值：**0

## enable\_codegen

**参数说明：**标识是否允许开启代码生成优化，目前代码生成使用的是LLVM优化。

**参数类型：**USERSET

**取值范围：**布尔型

- on表示允许开启代码生成优化。
- off表示不允许开启代码生成优化。

---

### 须知

目前LLVM优化仅支持向量化执行引擎特性和SQL on Hadoop特性，在其他场景下建议关闭此参数。

---

**默认值：**on

## codegen\_strategy

**参数说明：**标识在表达式codegen化过程中所使用的代码生成优化策略。

**参数类型：**USERSET

**取值范围：**枚举类型

- partial表示当所计算表达式中即使包含部分未被codegen化的函数时，仍可借助表达式全codegen框架调用LLVM动态编译优化策略。
- pure表示当所计算表达式整体可被codegen化时，才考虑调用LLVM动态编译优化策略。

---

### 须知

在开启代码生成优化会导致查询性能下降的场景下可以设置此参数为pure，其他场景下建议不改变此参数的默认值partial。

---



**默认值:** partial

## enable\_codegen\_print

**参数说明:** 标识是否允许在log日志中打印所生成的LLVM IR函数。

**参数类型:** USERSET

**取值范围:** 布尔型

- on表示允许在log日志中打印IR函数。
- off表示不允许在log日志中打印IR函数。

**默认值:** off

## codegen\_cost\_threshold

**参数说明:** 由于LLVM编译生成最终的可执行机器码需要一定时间，因此只有当实际执行的代价大于编译生成机器码所需要的代码和优化后的执行代价之和时，利用代码生成才有收益。codegen\_cost\_threshold标识代价的阈值，当执行估算代价大于该代价时，使用LLVM优化。

**参数类型:** USERSET

**取值范围:** 整型，0~INT\_MAX

**默认值:** 10000

## enable\_constraint\_optimization

**参数说明:** 标识是否允许HDFS外表使用信息约束（Informational Constraint）优化执行计划。

**参数类型:** SUSERSET

**取值范围:** 布尔型

- on表示允许使用信息约束优化执行计划。
- off表示不允许使用信息约束优化执行计划。

**默认值:** on

## enable\_bloom\_filter

**参数说明:** 标识是否允许使用BloomFilter优化。

**参数类型:** USERSET

**取值范围:** 布尔型

- on表示允许使用BloomFilter优化。
- off表示不允许使用BloomFilter优化。

**默认值:** on

**须知**

适用场景：外表侧同线程包含有HDFS内外表或列存表的HASH JOIN会触发Bloom Filter。

使用限制：

1. JOIN类型仅限于INNER JOIN、SEMI JOIN、RIGHT JOIN、RIGHT SEMI JOIN、RIGHT ANTI JOIN、RIGHT ANTI FULL JOIN。
2. JOIN内表侧的数据不能超过5万行。
3. JOIN内表侧关联条件：对于HDFS内外表不能为表达式；对于列存表可以为表达式，但仅限于非JOIN层计算的表达式。
4. JOIN外表侧关联条件必须为简单列关联。
5. JOIN内表侧与外表侧关联条件均为简单列关联时，计划层估算必须可以去掉1/3以上的数据（仅针对HDFS内外表）。
6. JOIN不能包含null值关联。
7. JOIN层未出现下盘。
8. 数据类型：
  - HDFS内外表字段类型支持SMALLINT、INTEGER、BIGINT、REAL/FLOAT4、DOUBLE PRECISION/FLOAT8、CHAR(n)/CHARACTER(n)/NCHAR(n)、VARCHAR(n)/CHARACTER VARYING(n)、CLOB、TEXT。
  - 列存表字段类型支持SMALLINT、INTEGER、BIGINT、OID、“char”、CHAR(n)/CHARACTER(n)/NCHAR(n)、VARCHAR(n)/CHARACTER VARYING(n)、NVARCHAR2(n)、CLOB、TEXT、DATE、TIME、TIMESTAMP、TIMESTAMPZ，其中字符类型其排序规则必须指定为“C”。

**enable\_extrapolation\_stats**

**参数说明：**标识是否允许基于历史统计信息使用推理估算的逻辑。使用该逻辑对于未及时收集统计信息的表可以增大估算准确的可能性，但也存在错误推理导致估算过大的可能性。

**参数类型：**USERSET

**取值范围：**布尔型

- on表示允许基于历史统计信息使用推理估算的逻辑。
- off表示不允许基于历史统计信息使用推理估算的逻辑。

**默认值：**

- 若当前集群为低版本升级到8.2.0.100及以上集群版本，为保持和前向兼容，默认值为off。
- 若当前集群为新装的8.2.0.100及以上集群版本，默认值为on。

**autoanalyze**

**参数说明：**标识是否允许在生成计划的时候，对于“统计信息完全缺失”或“修改量达到analyze阈值”的表进行统计信息自动收集，当前不支持对外表触发autoanalyze，不支持对带有ON COMMIT [DELETE ROWS|DROP]选项的临时表触发autoanalyze，如需收集，需用户手动执行analyze操作。如果在auto analyze某个表的过程中数据库

发生异常，当数据库正常运行之后再执行语句有可能仍提示需要收集此表的统计信息。此时需要用户对该表手动执行一次analyze操作，以同步统计信息数据。

#### 须知

表的修改量达到analyze阈值是指：表的修改量超过 $\text{autovacuum\_analyze\_threshold} + \text{autovacuum\_analyze\_scale\_factor} * \text{reltuples}$ ，其中reltuples是pg\_class中记录的表的估算行数。

**参数类型：** SUSERSET

**取值范围：** 布尔型

- on表示允许自动进行统计信息收集。
- off表示不允许自动进行统计信息收集。

**默认值：** on

## query\_dop

**参数说明：** 用户自定义的查询并行度。

**参数类型：** USERSET

**取值范围：** 整型，-64-64

[1,64]：打开固定SMP功能，系统会使用固定并行度。

0：打开SMP自适应功能，系统会根据资源情况和计划特征动态为每个查询选取[1,8]之间（x86平台），[1,64]之间（鲲鹏平台）的最优的并行度。

[-64,-1]：打开SMP自适应功能，并限制自适应选取的最大并行度。

#### 📖 说明

- 对于短查询为主的TP类业务中，如果不能通过CN轻量化或下发语句进行业务的调优，则生成SMP计划的时间较长，建议设置query\_dop=1。对于AP类复杂语句的场景，建议设置query\_dop=0。
- 在开启并行查询后，请保证系统CPU、内存、网络、I/O等资源充足，以达到良好效果。
- 为了避免用户设置不合理的过大值造成性能劣化，系统会计算出该DN可用最大CPU核数，并以此来作为query\_dop的上限。如果用户设置query\_dop超过4并且同时超过该上限，那么系统会重置query\_dop为该上限值。

**默认值：** 1

## query\_dop\_ratio

**参数说明：** 用于当query\_dop取值为0时，设置在系统中给定的最优dop基础上，调整dop的倍数。即最终 $\text{dop} = \text{系统设置dop} * \text{query\_dop\_ratio}$ （最小值为1，最大值为64）。当设置为1时，表示不调整。

**参数类型：** USERSET

**取值范围：** 浮点型，0-64

**默认值：** 1

## debug\_group\_dop

**参数说明:** 当query\_dop取值为0时，针对生成的执行计划划分的以Stream算子为顶点的group，均分配统一的dop并行度。此参数用于人为指定特定group的dop进行性能调优，格式为G1,D1,G2,D2,...，其中：G1,G2为group的ID，可以从日志中获得，D1,D2为指定的dop值，可以为任意正整数。

**参数类型:** USERSET

**取值范围:** 字符型

**默认值:** 空

### 须知

该参数仅供内部调优使用，不允许用户进行设置，建议保持默认值。

## enable\_analyze\_check

**参数说明:** 标识是否允许在生成计划的时候，对于在pg\_class中显示reltuples和relpages均为0的表，检查该表是否曾进行过统计信息收集。

**参数类型:** SUSERSET

**取值范围:** 布尔型

- on表示允许检查。
- off表示不允许检查。

**默认值:** on

## enable\_sonic\_hashagg

**参数说明:** 标识是否依据规则约束使用基于面向列的hash表设计的Hash Agg算子。

**参数类型:** USERSET

**取值范围:** 布尔型

- on表示在满足约束条件时使用基于面向列的hash表设计的Hash Agg算子。
- off表示不使用面向列的hash表设计的Hash Agg算子。

### 说明

- 在开启enable\_sonic\_hashagg，且查询达到约束条件使用基于面向列的hash表设计的Hash Agg算子时，查询对应的Hash Agg算子内存使用通常可获得精简。但对于代码生成技术可获得显著性能提升的场景(enable\_codegen打开后获得较大性能提升)，对应的算子查询性能可能会出现劣化。
- 开启enable\_sonic\_hashagg，且查询达到约束条件使用基于面向列的hash表设计的Hash Agg算子时，在Explain Analyze/Performance的执行计划和执行信息中，算子显示为“Sonic Hash Aggregation”，而未达到该约束条件时，算子名称将显示为“Hash Aggregation”。

**默认值:** on

## enable\_sonic\_hashjoin

**参数说明：**标识是否依据规则约束使用基于面向列的hash表设计的Hash Join算子。

**参数类型：**USERSET

**取值范围：**布尔型

- on表示在满足约束条件时使用基于面向列的hash表设计的Hash Join算子。
- off表示不使用面向列的hash表设计的Hash Join算子。

### 📖 说明

- 当前开关仅适用于Inner Join的场景。
- 在开启enable\_sonic\_hashjoin，查询对应的Hash Inner算子内存使用通常可获得精简。但对于代码生成技术可获得显著性能提升的场景，对应的算子查询性能可能会出现劣化。
- 开启enable\_sonic\_hashjoin，且查询达到约束条件使用基于面向列的hash表设计的Hash Join算子时，在Explain Analyze/Performance的执行计划和执行信息中，算子显示为“Sonic Hash Join”，而未达到该约束条件时，算子名称将显示为“Hash Join”。

**默认值：**on

## enable\_sonic\_optspill

**参数说明：**标识是否优化sonic场景下HashJoin或者HashAgg的下盘文件个数。仅在enable\_sonic\_hashjoin或者 enable\_sonic\_hashagg开启情况下生效。

**参数类型：**USERSET

**取值范围：**布尔型

- on表示开启下盘文件数优化。
- off表示关闭下盘文件数优化。

### 📖 说明

满足sonic条件下的HashJoin或者HashAgg算子，在关闭此参数（设置为off）时每列会产生1个下盘文件。开启此参数（设置为on）时如果不同列数据类型相似，只有1个下盘文件（最多5个文件）。

**默认值：**on

## expand\_hashtable\_ratio

**参数说明：**控制Hash Agg和Hash Join算子执行过程中hash表的大小扩大比例。

**参数类型：**USERSET

**取值范围：**浮点型，0, 0.5~10

### 📖 说明

- 默认值设置为0时表示hash表大小会根据当前内存进行自适应扩展。
- 默认值设置为0.5~10之间时，显式的是指定hash表扩大的倍数，通常hash表越大性能越好，但会占用更多内存空间，在内存不足场景可能造成数据提前下盘，带来性能劣化。

**默认值：**0

## plan\_cache\_mode

**参数说明：**标识在prepare语句中，选择生成执行计划的策略。

**参数类型：**USERSET

**取值范围：**枚举类型

- auto表示按照默认的方式选择custom plan或者generic plan。
- force\_generic\_plan表示强制走generic plan。
- force\_custom\_plan表示强制走custom plan。

### 📖 说明

- 此参数只对prepare语句生效，一般用在prepare语句中参数化字段存在比较严重的数据倾斜的场景下。
- custom plan是指对于prepare语句，在执行execute的时候，把execute语句中的参数嵌套到语句之后生成的计划。custom plan会根据execute语句中具体的参数生成计划，这种方案的优点是每次都按照具体的参数生成优选计划，执行性能比较好；缺点是每次执行前都需要重新生成计划，存在大量的重复的优化器开销。
- generic plan是指对于prepare语句生成计划，该计划策略会在执行execute语句的时候把参数bind到plan中，然后执行计划。这种方案的优点是每次执行可以省去重复的优化器开销；缺点是当bind参数字段上数据存在倾斜时该计划可能不是最优的，部分bind参数场景下执行性能较差。

**默认值：**auto

## wlm\_query\_accelerate

**参数说明：**标识在短查询加速打开时，查询是否需要加速。

**参数类型：**USERSET

**取值范围：**整型，-1~1

- -1：短查询由快车道管控，长查询由慢车道管控。
- 0：查询不加速，短查询和长查询均由慢车道管控。
- 1：查询加速，短查询和长查询均由快车道管控。

**默认值：**-1

## show\_unshippable\_warning

**参数说明：**标识是否将语句不下推的告警打印到客户端。

**参数类型：**USERSET

**取值范围：**布尔型

- on：将语句不下推的原因以WARNING打印到日志和客户端
- off：仅将语句不下推的原因以LOG打印到日志

**默认值：**off

## hashjoin\_spill\_strategy

**参数说明：**选择hashjoin下盘策略。（该参数8.1.2及以上版本支持）

**参数类型：**USERSET

**取值范围：**整型，0~6

- 0：当内表较大且无法在数据库可用内存放下所有数据时，会将数据划分成不同的子分区，直到多次划分后无法分开且仍无法在内存放下所有数据时，尝试外表是否可以放到可用内存中建立哈希表。若外表可以放到可用内存中建立哈希表，则执行HashJoin。反之，则执行NestLoop。
- 1：当内表较大且无法在数据库可用内存放下所有数据时，会将数据划分成不同的子分区，直到多次划分后无法分开且仍无法在内存放下所有数据时，尝试外表是否可以放到可用内存中建立哈希表。如果内外表均很大，强制执行HashJoin。
- 2：当内表较大，并且多次下盘无法分开时，强制执行HashJoin。
- 3：当内表较大，并且多次下盘无法分开时，尝试外表是否可以放到数据库可用内存建立哈希表。如果内外表均很大，则报错。
- 4：当内表较大，并且多次下盘无法分开时，则报错。
- 5：当内表较大且无法在数据库可用内存放下所有数据时，如果外表数据可以放到内存中，则使用外表建立哈希表执行HashJoin。如果外表数据无法存放到内存中，则将数据划分成不同的子分区，直到内外表多次划分均无法分开时，执行NestLoop。
- 6：当内表较大且无法在数据库可用内存放下所有数据时，如果外表数据可以放到内存中，则使用外表建立哈希表执行HashJoin。如果外表数据无法存放到内存中，则将数据划分成不同的子分区，直到内外表多次划分均无法分开时，强制执行HashJoin。

#### 说明

- 此参数只对向量化HashJoin生效。
- 对于数据distinct值很少且数据量很大场景，可能出现无法下盘导致使用内存过大产生内存不受控的问题。取值0时通过尝试内外表交换或者Nestloop可以避免出现此类内存不受控问题。执行Nestloop可能造成某些场景性能劣化。遇到此种场景，该参数可取值1、2、6强制执行HashJoin。
- 取值0对向量化Full Join不生效，行为与取值1相同。只尝试外表是否可建立哈希表，不执行NestLoop。
- 取值5和6相对于取值0和1的优势是如果内表数据量大到无法直接放于可用内存中，但是外表可以，则直接使用外表进行Hashjoin，减少后续多次下盘划分数据的时间消耗。当外表数据distinct较少时，使用外表建立哈希表可能导致性能劣化。此时可以将参数取值调整到0或者1。

**默认值：**0

## max\_streams\_per\_query

**参数说明：**控制查询计划中Stream节点的数目。（该参数仅8.1.1及以上集群版本支持）

**参数类型：**SUSET

**取值范围：**整型，-1~10000

- -1，表示查询计划中Stream节点数目无限制。
- 0~10000，表示查询计划中Stream节点数目超过设定值后报错，查询计划不会被执行。



### 📖 说明

- 此参数只控制DN上的Stream节点，不考虑CN上的Gather节点。
- 此参数不影响Explain查询计划，但是对Explain analyze和Explain performance有影响。

默认值：-1

## enable\_agg\_limit\_opt

**参数说明：**标识是否对select distinct col from table limit N场景优化，其中N小于16384时生效，table为列存表。该参数仅8.2.0及以上集群版本支持。

**参数类型：**USERSET

**取值范围：**布尔型

- on表示开启该优化。开启后可保证查询结果来自不同DN，且各个DN无需全部建立全量hash表，可显著提升查询性能。
- off表示关闭该优化。

默认值：off

## 17.10 错误报告和日志

### 17.10.1 记录日志的位置

#### log\_truncate\_on\_rotation

**参数说明：**logging\_collector设置为on时，log\_truncate\_on\_rotation设置日志消息的写入方式。

**参数类型：**SIGHUP

**取值范围：**布尔型

- on表示GaussDB(DWS)以覆盖写入的方式写服务器日志消息。
- off表示GaussDB(DWS)将日志消息附加到同名的现有日志文件上。

默认值：off

### 📖 说明

示例：

假设日志需要保留7天，每天生成一个日志文件，日志文件名设置为server\_log.Mon、server\_log.Tue等。第二周的周二生成的日志消息会覆盖写入到server\_log.Tue。设置方法：将log\_filename设置为server\_log.%a，log\_truncate\_on\_rotation设置为on，log\_rotation\_age设置为1440，即日志有效时间为1天。

#### log\_rotation\_age

**参数说明：**logging\_collector设置为on时，log\_rotation\_age决定创建一个新日志文件的时间间隔。当现在的时间减去上次创建一个服务器日志的时间超过了log\_rotation\_age的值时，将生成一个新的日志文件。



**参数类型:** SIGHUP

**取值范围:** 整型, 0 ~ 24d, 单位为min, h, d。其中0表示关闭基于时间的新日志文件的创建。

**默认值:** 1d

## log\_rotation\_size

**参数说明:** logging\_collector设置为on时, log\_rotation\_size决定服务器日志文件的最大容量。当日志消息的总量超过日志文件容量时, 服务器将生成一个新的日志文件。

**参数类型:** SIGHUP

**取值范围:** 整型, INT\_MAX / 1024, 单位为KB。

0表示关闭基于容量的新日志文件的创建。

**默认值:** 20MB

## event\_source

**参数说明:** log\_destination设置为eventlog时, event\_source设置在日志中GaussDB(DWS)日志消息的标识。

**参数类型:** POSTMASTER

**取值范围:** 字符串

**默认值:** PostgreSQL

## 17.10.2 记录日志的时间

### client\_min\_messages

**参数说明:** 控制发送到客户端的消息级别。每个级别都包含排在它后面的所有级别中的信息。级别越低, 发送给客户端的消息就越少。

**参数类型:** USERSET

---

#### 须知

当client\_min\_messages和log\_min\_messages取相同值时, 其值所代表的级别不同。

---

**取值范围:** 枚举类型, 有效值有debug5、debug4、debug3、debug2、debug1、info、log、notice、warning、error。参数的详细信息请参见表17-1。

**默认值:** notice

### log\_min\_messages

**参数说明:** 控制写到服务器日志文件中的消息级别。每个级别都包含排在它后面的所有级别中的信息。级别越低, 服务器运行日志中记录的消息就越少。

**参数类型:** SUSERSET

**须知**

当`client_min_messages`和`log_min_messages`取相同值log时所代表的消息级别不同。

**取值范围：**枚举类型，有效值有debug5、debug4、debug3、debug2、debug1、info、log、notice、warning、error、fatal、panic。参数的详细信息请参见表17-1。

**默认值：**warning

## log\_min\_error\_statement

**参数说明：**控制在服务器日志中记录错误的SQL语句。

**参数类型：**SUSET

**取值范围：**枚举类型，有效值有debug5、debug4、debug3、debug2、debug1、info、log、notice、warning、error、fatal、panic。参数的详细信息请参见表17-1。

**说明**

- 设置为error，表示导致错误、日志消息、致命错误、panic的语句都将被记录。
- 设置为panic，表示关闭此特性。

**默认值：**error

## log\_min\_duration\_statement

**参数说明：**当某条语句的持续时间大于或者等于特定的毫秒数时，`log_min_duration_statement`参数用于控制记录每条完成语句的持续时间。

设置`log_min_duration_statement`可以很方便地跟踪需要优化的查询语句。对于使用扩展查询协议的客户端，语法分析、绑定、执行每一步所花时间被独立记录。

**参数类型：**SUSET

**须知**

当此选项与`log_statement`同时使用时，已经被`log_statement`记录的语句文本不会被重复记录。在没有使用syslog情况下，推荐使用`log_line_prefix`记录PID或会话ID，方便将当前语句消息连接到最后的持续时间消息。

**取值范围：**整型，-1 ~ INT\_MAX，单位为毫秒。

- 设置为250，所有运行时间不短于250ms的SQL语句都会被记录。
- 设置为0，输出所有语句的持续时间。
- 设置为-1，关闭此功能。

**默认值：**30min

## backtrace\_min\_messages

**参数说明：**控制当产生该设置参数级别相等或更高级别的信息时，会打印函数的堆栈信息到服务器日志文件中。

**参数类型：**SUSET

### 须知

该参数作为客户现场问题定位手段使用，且由于频繁的打印函数栈会对系统的开销及稳定性有一定的影响，因此如果需要进行问题定位时，建议避免将backtrace\_min\_messages的值设置为fatal及panic以外的级别。

**取值范围：**枚举类型

有效值有debug5、debug4、debug3、debug2、debug1、info、log、notice、warning、error、fatal、panic。参数的详细信息请参见[表17-1](#)。

**默认值：**panic

**表17-1**解释GaussDB(DWS)中使用的消息安全级别。当日志输出到syslog或者eventlog时，GaussDB(DWS)进行如表中的转换。

**表 17-1** 信息严重程度分类

信息严重程度类型	详细说明	系统日志	事件日志
debug[1-5]	报告详细调试信息。	DEBUG	INFORMATION
log	报告对数据库管理员有用的信息，比如检查点操作统计信息。	INFO	INFORMATION
info	报告用户可能需求的信息，比如在VACUUM VERBOSE过程中的信息。	INFO	INFORMATION
notice	报告可能对用户有帮助的信息，比如，长标识符的截断，作为主键一部分创建的索引等。	NOTICE	INFORMATION
warning	报告警告信息，比如在事务块范围之外的COMMIT。	NOTICE	WARNING
error	报告导致当前命令退出的错误。	WARNING	ERROR
fatal	报告导致当前会话终止的原因。	ERR	ERROR
panic	报告导致整个数据库被关闭的原因。	CRIT	ERROR

## plog\_merge\_age

**参数说明：**该参数用于控制性能日志数据输出的周期。

**参数类型：**SUSET

### 须知

该参数以毫秒为单位的，建议在使用过程中设置值为1000的整数倍，即设置值以秒为最小单位。该参数所控制的性能日志文件以prf为扩展名，文件放置在\$GAUSSLOG/gs\_profile/<node\_name> 目录下面，其中node\_name是由postgres.conf文件中的pgxc\_node\_name的值，不建议外部使用该参数。

**取值范围：**0~INT\_MAX，单位为毫秒（ms）。

- 当设置为0时，当前会话不再输出性能日志数据。
- 当设置为非0时，当前会话按照指定的时间周期进行输出性能日志数据。该参数设置得越小，输出的日志数据越多，对性能的负面影响越大。

**默认值：**3s

## profile\_logging\_module

**参数说明：**用于设置记录性能日志的类型，使用该参数时需确保plog\_merge\_age参数值非0。该参数属于会话级参数，不建议通过gs\_guc工具来设置。仅8.1.3及以上集群版本支持。

**参数类型：**USERSET

**取值范围：**字符串

**默认值：**默认打开OBS,HADOOP,REMOTE\_DATANODE，关闭MD。可由SHOW profile\_logging\_module查看。

**设置方法：**首先，可以通过SHOW profile\_logging\_module来查看哪些模块是支持可控制的。例如，查询输出结果为：

```
show profile_logging_module;
profile_logging_module
-----
ALL,on(OBS,HADOOP,REMOTE_DATANODE),off(MD)(1 row)
```

打开MD性能日志，并查看设置结果。其中ALL标识是相当于一个快捷操作，即对所有模块的日志可输出进行开启或关闭。

```
set profile_logging_module='on(md)';
SET

show profile_logging_module;
profile_logging_module
-----
ALL,on(MD,OBS,HADOOP,REMOTE_DATANODE),off()(1 row)
```

## 17.10.3 记录日志的内容

### debug\_print\_parse

**参数说明：**用于控制打印解析树结果。

**参数类型：** SIGHUP

**取值范围：** 布尔型

- on表示开启打印结果的功能。
- off表示关闭打印结果的功能。

**默认值：** off

## debug\_print\_rewritten

**参数说明：** 用于控制打印查询重写结果。

**参数类型：** SIGHUP

**取值范围：** 布尔型

- on表示开启打印结果的功能。
- off表示关闭打印结果的功能。

**默认值：** off

## debug\_print\_plan

**参数说明：** 用于控制打印查询执行结果。

**参数类型：** SIGHUP

**取值范围：** 布尔型

- on表示开启打印结果的功能。
- off表示关闭打印结果的功能。

**默认值：** off

---

### 须知

- 只有当日志的级别为log及以上时，debug\_print\_parse、debug\_print\_rewritten和debug\_print\_plan的调试信息才会输出。当这些选项打开时，调试信息只会记录在服务器的日志中，而不会输出到客户端的日志中。通过设置[client\\_min\\_messages](#)和[log\\_min\\_messages](#)参数可以改变日志级别。
  - 在打开debug\_print\_plan开关的情况下需尽量避免调用gs\_encrypt\_aes128及gs\_decrypt\_aes128函数，避免敏感参数信息在日志中泄露的风险。同时建议用户在打开debug\_print\_plan开关生成的日志中对gs\_encrypt\_aes128及gs\_decrypt\_aes128函数的参数信息进行过滤后再提供给外部维护人员定位，日志使用完成后请及时删除。
- 

## debug\_pretty\_print

**参数说明：** 设置此选项对debug\_print\_parse、debug\_print\_rewritten和debug\_print\_plan产生的日志进行缩进，会生成易读但比设置为off时更长的输出格式。

**参数类型：** USERSET

**取值范围：**布尔型

- on表示进行缩进。
- off表示不进行缩进。

**默认值：**on

## log\_checkpoints

**参数说明：**控制在服务器日志中记录检查点和重启点的信息。打开此参数时，服务器日志消息包含涉及检查点和重启点的统计量，其中包含需要写的缓存区的数量及写入所花费的时间等。

**参数类型：**SIGHUP

**取值范围：**布尔型

- on表示打开此参数时，服务器日志消息包含涉及检查点和重启点的统计量。
- off表示关闭此参数时，服务器日志消息包含不涉及检查点和重启点的统计量。

**默认值：**off

## log\_connections

**参数说明：**控制记录客户端的连接请求信息。

**参数类型：**BACKEND

---

### 须知

- 会话连接参数，不建议用户设置。
- 有些客户端程序（例如gsql），在判断是否需要口令的时候会尝试连接两次，因此日志消息中重复的“connection receive”（收到连接请求）并不意味着一定是问题。

---

**取值范围：**布尔型

- on表示记录信息。
- off表示不记录信息。

**默认值：**off

## log\_disconnections

**参数说明：**控制记录客户端结束连接信息。

**参数类型：**BACKEND

**取值范围：**布尔型

- on表示记录信息。
- off表示不记录信息。

**默认值：**off

### 📖 说明

会话连接参数，不建议用户设置。

## log\_duration

**参数说明：**控制记录每个已完成SQL语句的执行时间。对使用扩展查询协议的客户端、会话记录语法分析、绑定和执行每一步所花费的时间。

**参数类型：**SUSET

**取值范围：**布尔型

- 设置为off，该选项与[log\\_min\\_duration\\_statement](#)的不同之处在于log\_min\_duration\_statement强制记录查询文本。
- 设置为on并且log\_min\_duration\_statement大于零，记录所有持续时间，但是仅记录超过阈值的语句。这可用于在高负载情况下搜集统计信息。

**默认值：**on

## log\_error\_verbosity

**参数说明：**控制服务器日志中每条记录的消息写入的详细度。

**参数类型：**SUSET

**取值范围：**枚举类型

- terse输出不包括DETAIL、HINT、QUERY及CONTEXT错误信息的记录。
- verbose输出包括SQLSTATE错误代码、源代码文件名、函数名及产生错误所在的行号。
- default输出包括DETAIL、HINT、QUERY及CONTEXT错误信息的记录，不包括SQLSTATE错误代码、源代码文件名、函数名及产生错误所在的行号。

**默认值：**default

## log\_hostname

**参数说明：**默认状态下，连接消息日志只显示正在连接主机的IP地址。打开此选项同时可以记录主机名。由于解析主机名可能需要一定的时间，可能影响数据库的性能。

**参数类型：**SIGHUP

**取值范围：**布尔型

- on表示可以同时记录主机名。
- off表示不可以同时记录主机名。

**默认值：**off

## log\_lock\_waits

**参数说明：**当一个会话的等待获得一个锁的时间超过[deadlock\\_timeout](#)的值时，此选项控制在数据库日志中记录此消息。这对于决定锁等待是否会产生一个坏的行为是非常有用的。

**参数类型：** SUSET

**取值范围：** 布尔型

- on表示记录此信息。
- off表示不记录此信息。

**默认值：** off

## log\_statement

**参数说明：** 控制记录SQL语句。对于使用扩展查询协议的客户端，记录接收到执行消息的事件和绑定参数的值（内置单引号要双写）。

**参数类型：** SUSET

### 须知

即使log\_statement设置为all，包含简单语法错误的语句也不会被记录，因为仅在完成基本的语法分析并确定了语句类型之后才记录日志。在使用扩展查询协议的情况下，在执行阶段之前（语法分析或规划阶段）同样不会记录。将log\_min\_error\_statement设为ERROR或更低才能记录这些语句。

**取值范围：** 枚举类型

- none表示不记录语句。
- ddl表示记录所有的数据定义语句，比如CREATE、ALTER和DROP语句。
- mod表示记录所有DDL语句，还包括数据修改语句INSERT、UPDATE、DELETE、TRUNCATE和COPY FROM。
- all表示记录所有语句，PREPARE、EXECUTE和EXPLAIN ANALYZE语句也同样被记录。

**默认值：** none

## log\_temp\_files

**参数说明：** 控制记录临时文件的删除信息。临时文件可以用来排序、哈希及临时查询结果。当一个临时文件被删除时，将会产生一条日志消息。

**参数类型：** SUSET

**取值范围：** 整型，-1~INT\_MAX，单位为KB

- 正整数表示只记录比log\_temp\_files设定值大的临时文件的删除信息。
- 0表示记录所有的临时文件的删除信息。
- -1表示不记录任何临时文件的删除信息。

**默认值：** -1

## log\_timezone

**参数说明：** 设置服务器写日志文件时使用的时区。与TimeZone不同，这个值是数据库范围的，针对所有连接到本数据库的会话生效。



**参数类型:** SIGHUP

**取值范围:** 字符串

**默认值:** PRC

 **说明**

gs\_initdb进行相应系统环境设置时会对默认值进行修改。

## logging\_module

**参数说明:** 用于设置或者显示模块日志在服务端的可输出性。该参数属于会话级参数，不建议通过gs\_guc工具来设置。

**参数类型:** USERSET

**取值范围:** 字符串

**默认值:** 所有模块日志在服务端是不输出的，可由SHOW logging\_module查看。

**设置方法:** 首先，可以通过SHOW logging\_module来查看哪些模块是支持可控制的。例如，查询输出结果为：

```
show logging_module;
logging_module
-----
-----
-----
ALL,on(),off(DFS,GUC,HDFS,ORC,SLRU,MEM_CTL,AUTOVAC,ANALYZE,CACHE,ADIO,SSL,GDS,TBLSPC,WLM,SPACE,OBS,EXECUTOR,VEC_EXECUTOR,STREAM,LLVM,OPT,OPT_REWRITE,OPT_JOIN,OPT_AGG,OPT_SUBPLAN,OPT_SETOP,OPT_CARD,OPT_SKEW,SMP,UDF,COOP_ANALYZE,WLMCP,ACCELERATE,PLANHINT,PARQUET,CARBONDATA,SNAPSHOT,XACT,HANDLE,CLOG,TQUAL,EC,REMOTE,CN_RETRY,PLSQL,TEXTSEARCH,SEQ,INSTR,COMM_IPC,COMM_PARAM,CSTORE,JOB,STREAMPOOL,STREAM_CTESCAN)
(1 row)
```

支持可控制的模块使用大写来标识，特殊标识ALL用于对所有模块日志进行设置。可以使用on/off来控制模块日志的输出。设置SSL模块日志为可输出，使用如下命令：

```
set logging_module='on(SSL)';
SET
show
logging_module;
logging_module
-----
-----
-----
ALL,on(SSL),off(DFS,GUC,HDFS,ORC,SLRU,MEM_CTL,AUTOVAC,ANALYZE,CACHE,ADIO,GDS,TBLSPC,WLM,SPACE,OBS,EXECUTOR,VEC_EXECUTOR,STREAM,LLVM,OPT,OPT_REWRITE,OPT_JOIN,OPT_AGG,OPT_SUBPLAN,OPT_SETOP,OPT_CARD,OPT_SKEW,SMP,UDF,COOP_ANALYZE,WLMCP,ACCELERATE,PLANHINT,PARQUET,CARBONDATA,SNAPSHOT,XACT,HANDLE,CLOG,TQUAL,EC,REMOTE,CN_RETRY,PLSQL,TEXTSEARCH,SEQ,INSTR,COMM_IPC,COMM_PARAM,CSTORE,JOB,STREAMPOOL,STREAM_CTESCAN)
(1 row)
```

可以看到模块SSL的日志输出被打开。

ALL标识是相当于一个快捷操作，即对所有模块的日志可输出进行开启或关闭。

```
set logging_module='off(ALL)';
SET
show
```

```

logging_module;

      logging_module
-----
-----
-----
ALL,on(),off(DFS,GUC,HDFS,ORC,SLRU,MEM_CTL,AUTOVAC,ANALYZE,CACHE,ADIO,SSL,GDS,TBLSPC,WLM,SPACE,OBS,EXECUTOR,VEC_EXECUTOR,STREAM,LLVM,OPT,OPT_REWRITE,OPT_JOIN,OPT_AGG,OPT_SUBPLAN,OPT_SETOP,OPT_CARD,OPT_SKEW,SMP,UDF,COOP_ANALYZE,WLMCP,ACCELERATE,PLANHINT,PARQUET,CARBONDATA,SNAPSHOT,XACT,HANDLE,CLOG,TQUAL,EC,REMOTE,CN_RETRY,PLSQL,TEXTSEARCH,SEQ,INSTR,COMM_IPC,COMM_PARAM,CSTORE,JOB,STREAMPOOL,STREAM_CTESCAN)
(1 row)

set logging_module='on(ALL)';
SET
show
logging_module;

      logging_module
-----
-----
-----
ALL,on(DFS,GUC,HDFS,ORC,SLRU,MEM_CTL,AUTOVAC,ANALYZE,CACHE,ADIO,SSL,GDS,TBLSPC,WLM,SPACE,OBS,EXECUTOR,VEC_EXECUTOR,STREAM,LLVM,OPT,OPT_REWRITE,OPT_JOIN,OPT_AGG,OPT_SUBPLAN,OPT_SETOP,OPT_CARD,OPT_SKEW,SMP,UDF,COOP_ANALYZE,WLMCP,ACCELERATE,PLANHINT,PARQUET,CARBONDATA,SNAPSHOT,XACT,HANDLE,CLOG,TQUAL,EC,REMOTE,CN_RETRY,PLSQL,TEXTSEARCH,SEQ,INSTR,COMM_IPC,COMM_PARAM,CSTORE,JOB,STREAMPOOL,STREAM_CTESCAN),off()
(1 row)

```

所有模块中，COMM\_IPC必须显式的打开/关闭，执行以下命令都可以将该模块的日志打开：

```

set logging_module='on(ALL)';
SET
set logging_module='on(COMM_IPC)';
SET

```

设置成功后，COMM\_IPC模块日志不会自动关闭，关闭COMM\_IPC模块的日志，必须手动执行关闭命令，以下两条命令都可以将该模块日志关闭：

```

set logging_module='off(ALL)';
SET
set logging_module='off(COMM_IPC)';
SET

```

**依赖关系：**该参数依赖于[log\\_min\\_messages](#)参数的设置。

## enable\_unshipping\_log

**参数说明：**用于控制是否打印语句不下推的日志，主要用于帮助用户定位不下推语句可能导致的性能问题。

**参数类型：**SUSET

**取值范围：**布尔型

- on表示打印日志。
- off表示不打印日志。

**默认值：**on

## 17.11 告警检测

在集群运行的过程中，会对数据库中的错误场景进行检测，便于用户及早感知到数据库集群的错误。

### alarm\_report\_interval

**参数说明：**指定告警上报的时间间隔。

**参数类型：**SIGHUP

**取值范围：**非负整型，单位为秒。

**默认值：**10

## 17.12 运行时统计

### 17.12.1 查询和索引统计收集器

查询和索引统计收集器负责收集数据库系统运行中的统计数据，如在一个表和索引上进行了多少次插入与更新操作、磁盘块的数量和元组的数量、每个表上最近一次执行清理和分析操作的时间等。可以通过查询系统视图pg\_stats和pg\_statistic查看统计数据。下面的参数设置服务器范围内的统计收集特性。

### track\_activities

**参数说明：**控制收集每个会话中当前正在执行命令的统计数据。

**参数类型：**SUSET

**取值范围：**布尔型

- on表示开启收集功能。
- off表示关闭收集功能。

**默认值：**on

### track\_counts

**参数说明：**控制收集数据库活动的统计数据。

**参数类型：**SUSET

**取值范围：**布尔型

- on表示开启收集功能。
- off表示关闭收集功能。

#### 说明

在AutoVacuum自动清理进程中选择清理的数据库时，需要数据库的统计数据，故默认值设为on。

**默认值：**on

## track\_io\_timing

**参数说明：**控制收集数据库I/O调用时序的统计数据。I/O时序统计数据可以在pg\_stat\_database中查询。

**参数类型：**SUSET

**取值范围：**布尔型

- on表示开启收集功能，开启时，收集器会在重复地去查询当前时间的操作系统，这可能会引起某些平台的重大开销，故默认值设置为off。
- off表示关闭收集功能。

**默认值：**off

## track\_functions

**参数说明：**控制收集函数的调用次数和调用耗时的统计数据。

**参数类型：**SUSET

---

### 须知

当SQL语言函数设置为调用查询的“内联”函数时，不管是否设置此选项，这些SQL语言函数无法被追踪到。

---

**取值范围：**枚举类型

- pl表示只追踪过程语言函数。
- all表示追踪SQL和C语言函数。
- none表示关闭函数追踪功能。

**默认值：**none

## track\_activity\_query\_size

**参数说明：**设置用于跟踪每一个活动会话的当前正在执行命令的字节数。

**参数类型：**POSTMASTER

**取值范围：**整型，100~102400

**默认值：**1024

## update\_process\_title

**参数说明：**控制收集因每次服务器接收到一个新的SQL语句时而产生的进程名称更新的统计数据。

进程名称可以通过ps命令进行查看，在Windows下通过任务管理器查看。

**参数类型：**SUSET

**取值范围：**布尔型

- on表示开启收集功能。
- off表示关闭收集功能。

默认值: off

## track\_thread\_wait\_status\_interval

**参数说明:** 用来定期收集thread状态信息的时间间隔。

**参数类型:** SUSET

**取值范围:** 整型, 0~1440, 单位为min。

默认值: 30min

## enable\_save\_datachanged\_timestamp

**参数说明:** 确定是否收集insert/update/delete, exchange/truncate/drop partition操作对表数据改动的时间。

**参数类型:** USERSET

**取值范围:** 布尔型

- on表示允许收集相关操作对表数据改动的时间。
- off表示禁止收集相关操作对表数据改动的时间。

默认值: on

## instr\_unique\_sql\_count

**参数说明:** 控制是否收集Unique SQL, 以及收集数量限制。

**参数类型:** SIGHUP

**取值范围:** 整型, 0~INT\_MAX

- 值为0时, 表示不收集Unique SQL统计信息;
- 值大于0时, 在CN节点上, 将会控制收集的Unique SQL数量不超过该设置值。当收集数量达到限制时, 不再收集新的Unique SQL, 此时可通过reload调大设置值, 继续收集新的Unique SQL。

默认值: 0

---

### 注意

通过reload加载新的设置值时, 如果新设置值小于原设置值, 将会清空对应CN节点已收集的Unique SQL统计信息。需特别注意该清理操作将由资源管理后台线程完成, 若GUC参数`use_workload_manager`为off时清理操作可能失败, 可直接使用函数`reset_instr_unique_sql`进行清理。

---

## instr\_unique\_sql\_timeout

**参数说明：**控制Unique SQL的存在时间。StatCollector后台线程每小时对所有的Unique SQL做一次检查，如果发现某个Unique SQL超过instr\_unique\_sql\_timeout小时未被执行，则将其删除。（该参数在8.1.2及以上版本支持。）

**参数类型：**SIGHUP

**取值范围：**整型，0~ INT\_MAX，单位为小时。

- 值为0时，表示不删除过期的Unique SQL；
- 值大于0时，发现超过instr\_unique\_sql\_timeout小时未执行Unique SQL则删除。

**默认值：**24

## track\_sql\_count

**参数说明：**控制对每个会话中当前正在执行的SELECT、INSERT、UPDATE、DELETE、MERGE INTO语句是否进行计数统计，对SELECT、INSERT、UPDATE、DELETE语句进行响应时间的统计，以及对DDL、DML、DCL语句进行计数的统计。

**参数类型：**SUSET

**取值范围：**布尔型

- on表示开启统计功能。
- off表示关闭统计功能。

**默认值：**on

### 说明

- track\_sql\_count参数受track\_activities约束：
  - track\_activities开启而track\_sql\_count关闭时，如果查询了gs\_sql\_count、pgxc\_sql\_count、gs\_workload\_sql\_count、pgxc\_workload\_sql\_count、global\_workload\_sql\_count、gs\_workload\_sql\_elapse\_time、pgxc\_workload\_sql\_elapse\_time、或global\_workload\_sql\_elapse\_time视图，将会有LOG提示track\_sql\_count是关闭的；
  - track\_activities和track\_sql\_count同时关闭，那么此时将会有两条LOG，分别提示track\_activities是关闭的和track\_sql\_count是关闭的；
  - track\_activities关闭而track\_sql\_count开启，此时将仅有LOG提示track\_activities是关闭。
- 当参数关闭时，查询视图的结果为0行。

## enable\_track\_wait\_event

**参数说明：**控制是否对各类等待事件的发生次数、失败次数、持续时间、最大、最小和平均等待时间等信息进行统计。

**参数类型：**SIGHUP

**取值范围：**布尔型

- on表示开启统计功能。
- off表示关闭统计功能。

**默认值：**off

### 📖 说明

- enable\_track\_wait\_event参数受track\_activities约束，如果track\_activities关闭，即使开启enable\_track\_wait\_event也不启用相关功能。
- track\_activities或enable\_track\_wait\_event关闭时，如果查询get\_instr\_wait\_event函数、gs\_wait\_events视图或pgxc\_wait\_events视图将会提示GUC参数关闭，查询结果为0行；
- 在集群运行过程中关闭track\_activities或enable\_track\_wait\_event，GaussDB(DWS)不再对等待事件相关信息进行统计，但已统计记录的数据不受影响。

## enable\_wdr\_snapshot

**参数说明：**控制是否启用性能视图快照功能。开启后，GaussDB(DWS)会定期对部分系统性能视图创建快照并持久化保存，并接受手动创建快照请求。

**参数类型：**SIGHUP

**取值范围：**布尔型

- on表示开启快照功能。
- off表示关闭快照功能。

**默认值：**off

### 📖 说明

- 如果enable\_wdr\_snapshot参数关闭，执行create\_wdr\_snapshot函数手动创建视图会提示GUC参数未打开。
- 如果在快照的过程中修改enable\_wdr\_snapshot参数，正在创建的快照不受影响，下次定期或手动创建快照时生效。

## wdr\_snapshot\_interval

**参数说明：**设置自动创建性能视图快照的时间间隔。

**参数类型：**SIGHUP

**取值范围：**整型，10~180，单位为分钟。

**默认值：**60

### 📖 说明

- 该参数取值应与集群负载相对应，不宜取值过小，建议大于创建一次快照所需时间。
- 如果wdr\_snapshot\_interval小于创建一次快照所需时间，也就是到了快照的时间，发现上一个快照还没执行完，则跳过本次快照。

## wdr\_snapshot\_retention\_days

**参数说明：**设置性能快照数据保留的最大天数。

**参数类型：**SIGHUP

**取值范围：**整型，1~15，单位为天。

**默认值：**8

### 说明

- 开启enable\_wdr\_snapshot的情况下，保存wdr\_snapshot\_retention\_days天数的快照数据会被自动清除。
- 该参数取值应与可用磁盘空间相对应，取值越大，需要的磁盘空间越大。
- 对该参数的修改不会立即生效，等到下一次自动创建快照时才会清除过期快照数据。

## 17.12.2 性能统计

在数据库在运行过程中，会涉及到锁的访问、磁盘IO操作、无效消息的处理，这些操作都可能是数据库的性能瓶颈，通过GaussDB(DWS)提供的性能统计方法，可以方便定位性能问题。

### 输出性能统计日志

**参数说明：**对每条查询，以下4个选项控制在服务器日志里记录相应模块的性能统计数据，具体含义如下：

- log\_parser\_stats控制在服务器日志里记录解析器的性能统计数据。
- log\_planner\_stats控制在服务器日志里记录查询优化器的性能统计数据。
- log\_executor\_stats控制在服务器日志里记录执行器的性能统计数据。
- log\_statement\_stats控制在服务器日志里记录整个语句的性能统计数据。

这些参数只能辅助管理员进行粗略分析，类似Linux中的操作系统工具getrusage()。

**参数类型：** SUSET

#### 须知

- log\_statement\_stats记录总的语句统计数据，而其他的只记录针对每个模块的统计数据。
- log\_statement\_stats不能和其他任何针对每个模块统计的选项一起打开。

**取值范围：** 布尔型

- on表示开启记录性能统计数据的功能。
- off表示关闭记录性能统计数据的功能。

**默认值：** off

## 17.13 资源管理

未对数据库资源做控制时，容易出现并发任务抢占资源导致操作系统过载甚至最终崩溃。操作系统过载时，其响应用户任务的速度会变慢甚至无响应；操作系统崩溃时，整个系统将无法对用户提供任何服务。GaussDB(DWS)的负载管理功能能够基于可用资源的多少均衡数据库的负载，以避免数据库系统过载。

### use\_workload\_manager

**参数说明：** 是否开启资源管理功能。此参数需在CN和DN同时应用。



**参数类型：** SIGHUP

**取值范围：** 布尔型

- on表示打开资源管理。
- off表示关闭资源管理。

#### 📖 说明

- 当使用**GUC参数设置**来修改参数值时，新参数值只能对更改操作执行后启动的线程生效。此外，对于后台线程以及线程复用执行的新作业，该参数值的改动不会生效。如果希望这类线程即时识别参数变化，可以使用kill session或重启节点的方式来实现。
- use\_workload\_manager参数由off变为on状态后，资源管理视图变为可用，并且可以查询off状态下统计的存储资源使用情况。若存在些许误差的情况下，需要矫正用户使用的存储资源，可数据库中执行如下命令，在执行该命令的过程中，如果对表中插入数据，可能会出现统计不够准确的情况：

```
select gs_wlm_readjust_user_space(0);
```

**默认值：** on

## enable\_control\_group

**参数说明：** 是否开启Cgroups功能。此参数需在CN和DN同时应用。

**参数类型：** SIGHUP

**取值范围：** 布尔型

- on表示打开Cgroups管理。
- off表示关闭Cgroups管理。

**默认值：** on

#### 📖 说明

当使用**GUC参数设置**来修改参数值时，新参数值只能对更改操作执行后启动的线程生效。此外，对于后台线程以及线程复用执行的新作业，该值的改动不会生效。如果希望这类线程即时识别参数变化，可以使用kill session或重启节点的方式来实现。

## enable\_backend\_control

**参数说明：** 是否控制数据库常驻线程到DefaultBackend控制组。此参数需在CN和DN同时应用。

**参数类型：** POSTMASTER

**取值范围：** 布尔型

- on表示控制常驻线程到DefaultBackend控制组。
- off表示不控制常驻线程到DefaultBackend控制组。

**默认值：** on

## enable\_vacuum\_control

**参数说明：** 是否控制数据库常驻线程autoVacuumWorker到Vacuum控制组。此参数需在CN和DN同时应用。

**参数类型：** POSTMASTER

**取值范围：**布尔型

- on表示控制数据库常驻线程autoVacuumWorker到Vacuum控制组。
- off表示不控制数据库常驻线程autoVacuumWorker到Vacuum控制组。

**默认值：**on

## enable\_perm\_space

**参数说明：**是否开启perm space功能。此参数需在CN和DN同时应用。

**参数类型：**POSTMASTER

**取值范围：**布尔型

- on表示打开perm space管理。
- off表示关闭perm space管理。

**默认值：**on

## space\_once\_adjust\_num

**参数说明：**空间管控和空间统计功能中，控制慢速构建与细粒度校准操作中每次处理的文件个数阈值。该参数8.1.3及以上集群版本支持。

**参数类型：**SIGHUP

**取值范围：**整型，0~INT\_MAX

- 0表示不启用慢速构建和细粒度校准功能。

**默认值：**300

### 说明

文件个数阈值影响数据库资源，建议合理设置。

## space\_readjust\_schedule

**参数说明：**空间管控和空间统计功能中，控制是否触发自动校准以及校准空间误差阈值。该参数8.1.3及以上集群版本支持。

**参数类型：**SIGHUP

**取值范围：**字符串

- off表示关闭自动校准功能。
- auto表示打开自动校准功能，并且触发自动校准的误差阈值为1GB。
- auto(空间大小+K/M/G)表示打开自动校准功能，并且触发自动校准的误差阈值为自定义的空间大小KB/MB/GB。例如，auto(200M)表示打开自动校准功能，且触发自动校准的误差阈值为200MB。

**默认值：**auto

## enable\_verify\_active\_statements

**参数说明：**在静态自适应负载场景下，是否开启后台校准功能。此参数需在CN上应用。

**参数类型：**SIGHUP

**取值范围：**布尔型

- on表示打开后台校准功能。
- off表示关闭后台校准功能。

**默认值：**on

## max\_active\_statements

**参数说明：**设置全局的最大并发数量。此参数只应用到CN，且针对一个CN上的执行作业。

数据库管理员需根据系统资源（如CPU资源、IO资源和内存资源）情况，调整此数值大小，使得系统支持最大限度的并发作业，且防止并发执行作业过多，引起系统崩溃。

**参数类型：**SIGHUP

**取值范围：**整型，-1 ~ INT\_MAX。设置为-1和0表示对最大并发数不做限制。

**默认值：**60

## parctl\_min\_cost

**参数说明：**静态资源管理场景下，复杂作业最小估算代价。用于简单作业和复杂作业划分的阈值，估算代价小于该值的作业为简单作业，估算代价大于等于该值的作业为复杂作业。

**参数类型：**SIGHUP

**取值范围：**整型，-1 ~ INT\_MAX

- parctl\_min\_cost等于-1时，所有作业都是简单作业；
- 估算代价小于10的作业任何场景下都是简单作业。

**默认值：**100000

## cgroup\_name

**参数说明：**设置当前使用的Cgroups的名字或者调整当前group下排队的优先级。

即如果先设置cgroup\_name，再设置session\_respool，那么session\_respool关联的控制组起作用，如果再切换cgroup\_name，那么新切换的cgroup\_name起作用。

切换cgroup\_name的过程中如果指定到Workload控制组级别，数据库不对级别进行验证。级别的范围只要在1-10范围内都可以。

**参数类型：**USERSET

建议尽量不要混合使用cgroup\_name和session\_respool。

**取值范围：**字符串

**默认值:** DefaultClass:Medium

 **说明**

DefaultClass:Medium表示DefaultClass下Timeshare控制组中的Medium控制组。

## cpu\_collect\_timer

**参数说明:** 设置语句执行时在DN上收集CPU时间的周期。

数据库管理员需根据系统资源（如CPU资源、IO资源和内存资源）情况，调整此数值大小，使得系统支持较合适的收集周期，太小会影响执行效率，太大会影响异常处理的精确度。

**参数类型:** SIGHUP

**取值范围:** 整型，1 ~ INT\_MAX，单位为秒。

**默认值:** 30

## enable\_cgroup\_switch

**参数说明:** 是否控制数据库执行语句时根据类型自动切换到TopWD组。

**参数类型:** USERSET

**取值范围:** 布尔型

- on表示控制数据库执行语句时根据类型自动切换到TopWD组。
- off表示控制数据库执行语句时根据类型不自动切换到TopWD组。

**默认值:** off

## memory\_tracking\_mode

**参数说明:** 设置记录内存信息的模式。

**参数类型:** USERSET

**取值范围:**

- none，不启动内存统计功能。
- normal，仅做内存实时统计，不生成文件。
- executor，生成统计文件，包含执行层使用过的所有已分配内存的上下文信息。
- fullexec，生成文件包含执行层申请过的所有内存上下文信息。

**默认值:** none

## memory\_detail\_tracking

**参数说明:** 设置需要的线程内分配内存上下文的顺序号以及当前线程所在query的plannodeid。

**参数类型:** USERSET

**取值范围:** 字符型

**默认值:** 空

**须知**

该参数不允许用户进行设置，建议保持默认值。

## enable\_resource\_track

**参数说明：**设置是否开启资源实时监控功能。此参数需在CN和DN同时应用。

**参数类型：**SIGHUP

**取值范围：**布尔型

- on表示打开资源监控。
- off表示关闭资源监控。

**默认值：**on

## enable\_resource\_record

**参数说明：**设置是否开启资源监控记录归档功能。开启时，对于执行结束的记录，会分别被归档到相应的INFO视图（[GS\\_WLM\\_SESSION\\_INFO](#)和[GS\\_WLM\\_OPERATOR\\_INFO](#)）。此参数需在CN和DN同时应用。

**参数类型：**SIGHUP

**取值范围：**布尔型

- on表示开启资源监控记录归档功能。
- off表示关闭资源监控记录归档功能。

**默认值：**on

**说明**

新建集群默认值为on，升级场景该参数的默认值为保持前向兼容维持原值。

## enable\_track\_record\_subsql

**参数说明：**设置是否开启子语句记录归档功能。开启时，存储过程、匿名块内部的子语句会被纪录归档到相应的INFO表（[GS\\_WLM\\_SESSION\\_INFO](#)）。此参数为会话级参数，可在与CN的连接会话中设置生效，仅影响该会话连接中的语句；也可在CN和DN上同时设置，能全局生效。

**参数类型：**USERSET

**取值范围：**布尔型

- on表示开启子语句资源监控记录归档功能。
- off表示关闭子语句资源监控记录归档功能。

**默认值：**off

## enable\_user\_metric\_persistent

**参数说明：**设置是否开启用户/资源池历史资源监控转存功能。开启时，对于[PG\\_TOTAL\\_USER\\_RESOURCE\\_INFO](#)视图中数据，会定期采样保存到

[GS\\_WLM\\_USER\\_RESOURCE\\_HISTORY](#)系统表中；对于[GS\\_RESPOOL\\_RESOURCE\\_INFO](#)视图中数据，会定期采样保存到[GS\\_RESPOOL\\_RESOURCE\\_HISTORY](#)系统表中。

**参数类型：**SIGHUP

**取值范围：**布尔型

- on表示开启用户/资源池历史资源监控转存功能。
- off表示关闭用户/资源池历史资源监控转存功能。

**默认值：**on

## user\_metric\_retention\_time

**参数说明：**设置用户历史资源监控数据的保存天数。该参数仅在enable\_user\_metric\_persistent为on时有效。

**参数类型：**SIGHUP

**取值范围：**整型，0~3650，单位为天。

- 值等于0时，用户历史资源监控数据将永久保存。
- 值大于0时，用户历史资源监控数据将保存对应天数。

**默认值：**7

## enable\_instance\_metric\_persistent

**参数说明：**设置是否开启实例资源监控转存功能。开启时，对实例的监控数据会保存到[GS\\_WLM\\_INSTANCE\\_HISTORY](#)系统表中。

**参数类型：**SIGHUP

**取值范围：**布尔型

- on表示开启实例资源监控转存功能。
- off表示关闭实例资源监控转存功能。

**默认值：**on

## instance\_metric\_retention\_time

**参数说明：**设置实例历史资源监控数据的保存天数。该参数仅在enable\_instance\_metric\_persistent为on时有效。

**参数类型：**SIGHUP

**取值范围：**整型，0~3650，单位为天。

- 值等于0时，实例历史资源监控数据将永久保存。
- 值大于0时，实例历史资源监控数据将保存对应设置天数。

**默认值：**7

## resource\_track\_level

**参数说明：**设置当前会话的资源监控的等级。该参数只有当参数enable\_resource\_track为on时才有效。

**参数类型：**USERSET

**取值范围：**枚举型

- none，不开启资源监控功能。
- query，开启query级别资源监控功能，开启此功能会把SQL语句的计划信息（类似explain输出信息）记录到top SQL中。
- perf，开启perf级别资源监控功能，开启此功能会把包含实际执行时间和执行行数的计划信息（类似explain analyze输出信息）记录到top SQL中。
- operator，开启operator级别资源监控功能，开启此功能不仅会把包含实际执行时间和执行行数的信息记录到top SQL中，还会把算子级别执行信息刷新到top SQL中。

**默认值：**query

## resource\_track\_cost

**参数说明：**设置对当前会话的语句进行资源监控的最小执行代价。该参数只有当参数enable\_resource\_track为on时才有效。

**参数类型：**USERSET

**取值范围：**整型，-1 ~ INT\_MAX

- 值为-1时，不进行资源监控。
- 值大于或等于0时，对执行代价超过该参数值的语句进行资源监控。

**默认值：**0

### 说明

新建集群默认值为0，升级场景该参数的默认值为保持前向兼容维持原值。

## resource\_track\_duration

**参数说明：**设置资源监控实时视图（参见表12-1）中记录的语句执行结束后进行历史信息转存的最小执行时间。当执行完成的作业，其执行时间不小于此参数值时，作业信息会从实时视图（以statistics为后缀的视图）转存到相应的历史视图（以history为后缀的视图）中。该参数只有当enable\_resource\_track为on时才有效。

**参数类型：**USERSET

**取值范围：**整型，0 ~ INT\_MAX，单位为秒。

- 值为0时，资源监控实时视图（表12-1）中记录的所有语句都进行历史信息归档。
- 值大于0时，资源监控实时视图（表12-1）中记录的语句的执行时间超过这个值就会进行历史信息归档。

**默认值：**60s

## query\_exception\_count\_limit

**参数说明：**设置作业触发异常规则次数的上限。作业触发异常规则次数达到上限后该作业会被自动加入黑名单且禁止再执行，只有移除黑名单后才能恢复运行。

**参数类型：**USERSET

**取值范围：**整型，-1 ~ INT\_MAX。

- 值为-1时，表示对作业触发异常规则的次数不做限制，即作业触发异常规则次数再多也不会自动将该作业加入黑名单。
- 值大于等于0时，表示作业触发异常规则次数达到阈值时该作业立即被加入黑名单。其中，0和1含义相同，表示只要作业触发异常规则即会将作业加入黑名单。

**默认值：**-1

## dynamic\_memory\_quota

**参数说明：**自适应负载场景下，设置内存控制的比重，即可以使用系统最大可用内存的比例。

**参数类型：**SIGHUP

**取值范围：**整型，1 ~ 100

**默认值：**80

## disable\_memory\_protect

**参数说明：**禁止内存保护功能。当系统内存不足时如果需要查询系统视图，可以先将此参数置为on，禁止内存保护功能，保证视图可以正常查询。该参数只适用于在系统内存不足时进行系统诊断和调试，正常运行时请保持该参数配置为off。

**参数类型：**USERSET

**取值范围：**布尔型

- on表示禁止内存保护功能。
- off表示启动内存保护功能。

**默认值：**off

## query\_band

**参数说明：**用于标示当前会话的作业类型，由用户自定义。

**参数类型：**USERSET

**取值范围：**字符型

**默认值：**空

## enable\_bbox\_dump

**参数说明：**是否开启黑匣子功能，在系统不配置core机制的时候仍可产生core文件。此功能需要在CN和DN同时应用。

**参数类型：**SIGHUP



**取值范围：**布尔型

- on表示打开黑匣子功能。
- off表示关闭黑匣子功能。

**默认值：**off

## enable\_dynamic\_workload

**参数说明：**是否开启动态负载管理功能。

**参数类型：**POSTMASTER

**取值范围：**布尔型

- on表示打开动态负载管理功能。
- off表示关闭动态负载管理功能。

**默认值：**on

### 须知

- 开启内存自适应后，收集统计信息后不再需要使用work\_mem进行算子内存使用调优，由系统根据当前负载情况，为每个语句生成计划，并估算每个算子的内存使用量和整个语句的内存使用量。系统根据负载情况和整个语句内存使用量进行队列调度，所以多并发场景会出现语句排队的情况。
- 由于优化器行数估算不准现象的存在，会出现语句内存使用量低估或高估的情况。低估时，执行时内存会自动扩展。高估时，会导致系统内存利用不足，排队语句增多，可能导致性能非最优。此时需要识别语句估算内存远大于实际DN峰值内存的语句，通过设置query\_max\_mem进行调优，详见[SQL调优关键参数调整](#)。

## enable\_wlm\_internal\_memory\_limit

**参数说明：**是否开启负载管理对语句估算内存的内置限制功能。（该参数仅8.2.0及以上集群版本支持）

负载管理中的内存管理模块中，会对语句的估算内存做一些内置限制，如：

- 语句的估算内存无法超过关联资源池内存上限的80%。
- 当资源池的并发控制参数active\_statements不为1时，语句的估算内存无法超过关联资源池内存上限的40%。
- 内核中语句的估算内存开始时是一个范围值，可以理解为[最小估算内存，最大估算内存]，其中，最大估算内存为语句运行性能最佳所需要的内存资源，最小估算内存为通过结果集下盘而保证语句能运行所需要的内存资源；最终的语句估算内存会在这个范围内取一个合理值，这个范围值中的最大估算内存也无法超过关联资源池内存上限的90%；

这些内置限制的目的是为了防止语句内存过高估算场景，高估场景下，语句抢占内存资源过大，会引发后续作业排队，导致系统资源利用率下降，因此，内核限制了单个语句的估算内存上限值。内置限制可能会导致语句执行计划不是最优，可能在一定程度上影响单个语句性能；从8.2.0集群版本开始增加了内存负反馈机制，从一定程度上缓解了语句严重高估导致系统资源利用率下降的问题。为此，对应的从8.2.0版本开

始，增加了enable\_wlm\_internal\_memory\_limit参数，由用户自由控制是否开启内置限制功能。

**参数类型：** SIGHUP

**取值范围：** 布尔型

- on表示打开负载管理对语句估算内存的内置限制。
- off表示关闭负载管理对语句估算内存的内置限制。

**默认值：** on

## enable\_strict\_memory\_expansion

**参数说明：** 是否开启对语句内存扩展的严格控制。（该参数仅8.2.0及以上集群版本支持）

语句在CN节点上会计算出估算内存，用该值来预占内存资源，在DN节点上会根据系统负载情况，在语句内存不足时尝试做内存扩展，以提升语句的执行效率。当打开此参数时，将会严格限制语句内存扩展的范围，尽可能保证语句的内存扩展不会超过语句的最大估算内存。语句的算子内存每次扩展是按比例扩展，因此，可能出现最后一次扩展内存之后一定幅度超过限制的现象，但能实现限制语句内存扩展在一定范围内。

**参数类型：** SIGHUP

**取值范围：** 布尔型

- on表示打开对语句内存扩展的严格控制。
- off表示关闭对语句内存扩展的严格控制。

**默认值：** off

## allow\_zero\_estimate\_memory

**参数说明：** 是否允许语句估算内存为0。（该参数仅8.2.0及以上集群版本支持）

当语句查询的表没有统计信息时，该语句在CN节点上的估算内存可能设置为0，当估算内存为0时，语句的内存将会变化为算子内存受到work\_mem限制（work\_mem来限制算子内存的方式已经不推荐使用），如果work\_mem设置较大，而语句的算子数量较多时，语句的实际内存值也能会较大，该参数设置为off时，将不允许未做analyze的查询出现估算内存为0，避免不可预期的问题。

**参数类型：** SIGHUP

**取值范围：** 布尔型

- on表示允许语句估算内存为0。
- off表示不允许语句估算内存为0。

**默认值：** on

## wlm\_memory\_feedback\_adjust

**参数说明：** 是否开启动态负载管理中的内存负反馈功能。（该参数仅8.2.0及以上集群版本支持）

语句在CN节点上会计算出估算内存，用该值来预占内存资源，语句内存过高估算场景时，语句预占内存资源过大，会引发后续作业排队，导致系统资源利用率下降，内存负反馈机制，会基于DN上的实际使用内存情况，判断如果集群连续一段时间处于高估场景时，会由CCN节点动态降低语句在CN预占的内存资源，将高估的内存腾让出来，给后续作业使用，从一定程度上缓解了语句严重高估导致系统资源利用率下降的问题。

**参数类型：** SIGHUP

**取值范围：** 字符串

- on表示启用内存负反馈功能。
- off表示关闭内存负反馈功能。
- on(触发负反馈需要的时间,触发负反馈需要的估算内存百分比)表示启用内存负反馈功能，并且指定触发负反馈需要的时间以及触发负反馈需要的估算内存百分比参数。例如：on(60,50)表示启用内存负反馈功能，触发负反馈机制生效需要连续60秒都是高估场景，还需要语句预占的估算内存总值要超过系统可用内存的50%。默认触发负反馈机制生效的时间长度为50秒，默认触发负反馈生效的最低估算内存总值要超过系统可用内存的40%。

**默认值：** on

## bbox\_dump\_count

**参数说明：** 在bbox\_dump\_path定义的路径下，允许存储的GaussDB(DWS)所产生core文件最大数。超过此数量，旧的core文件会被删除。此参数只有当enable\_bbox\_dump为on时才生效。

**参数类型：** USERSET

**取值范围：** 整型，1~20

**默认值：** 8

### 说明

在并发产生core文件时，core文件的产生个数可能大于bbox\_dump\_count。

## io\_limits

**参数说明：** 该参数8.1.2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。

**参数类型：** USERSET

**取值范围：** 整型，0~1073741823

**默认值：** 0

## io\_priority

**参数说明：** 该参数8.1.2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。

**参数类型：** USERSET

**取值范围：** 枚举型

- None
- Low
- Medium
- High

默认值: None

## session\_respool

**参数说明:** 当前的session关联的resource pool。

**参数类型:** USERSET

即如果先设置cgroup\_name, 再设置session\_respool, 那么session\_respool关联的控制组起作用, 如果再切换cgroup\_name, 那么新切换的cgroup\_name起作用。

切换cgroup\_name的过程中如果指定到Workload控制组级别, 数据库不对级别进行验证。级别的范围只要在1-10范围内都可以。

建议尽量不要混合使用cgroup\_name和session\_respool。

**取值范围:** 字符串, 通过create resource pool所设置的资源池。

默认值: invalid\_pool

## enable\_transaction\_parctl

**参数说明:** 是否管控事务块语句和存储过程语句。

**参数类型:** USERSET

**取值范围:** 布尔型

- on表示对事务块及存储过程语句进行管控。
- off表示对事务块及存储过程语句不进行管控。

默认值: on

## session\_statistics\_memory

**参数说明:** 该参数8.1.2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。

**参数类型:** SIGHUP

**取值范围:** 整型, 5MB ~ max\_process\_memory的50%。

默认值: 5MB

## session\_history\_memory

**参数说明:** 设置历史查询视图的内存大小。

**参数类型:** SIGHUP

**取值范围:** 整型, 10MB ~ max\_process\_memory的50%。

默认值: 100MB

## topsql\_retention\_time

**参数说明：**设置历史TopSQL中gs\_wlm\_session\_info和gs\_wlm\_operator\_info表中数据的保存时间。

**参数类型：**SIGHUP

**取值范围：**整型，0~3650，单位为天。

- 值为0时，表示数据永久保存。
- 值大于0时，表示数据能够保存的对应天数。

**默认值：**30

---

### 注意

- 设置此GUC参数启用数据保存功能前，请先清理gs\_wlm\_session\_info和gs\_wlm\_operator\_info表中的数据。
  - 新建集群默认值为30，升级场景该参数的默认值为保持前向兼容维持原值。
- 

## transaction\_pending\_time

**参数说明：**当enable\_transaction\_parctl为on时，事务块语句和存储过程语句排队的最大时间。

**参数类型：**USERSET

**取值范围：**整型，-1 ~ INT\_MAX，单位为秒。

- 值为-1或0：事务块语句和存储过程语句无超时判断，排队至资源满足可执行条件。
- 值大于0：事务块语句和存储过程语句排队超过所设数值的时间后，无视当前资源情况强制执行。

**默认值：**0

---

### 须知

此参数仅对存储过程及事务块的内部语句有效，即**PG\_SESSION\_WLMSTAT**中enqueue字段显示为Transaction或StoredProc的语句才会生效。

---

## wlm\_sql\_allow\_list

**参数说明：**用于指定资源管理SQL白名单语句，SQL白名单语句不受资源管理监控。

**参数类型：**SIGHUP

**取值范围：**字符串

**默认值：**空

**须知**

- wlm\_sql\_allow\_list中可指定一条或多条SQL白名单语句，指定多条时，通过“;”进行分隔。
- 系统通过前置匹配判断SQL语句是否受监控，不区分大小写，例如：  
wlm\_sql\_allow\_list='SELECT'，则所有select语句均不受资源管理监控。
- 识别参数值白名单字符串头部的空格，例如：'SELECT'与' SELECT'的含义是不一致的，' SELECT'只过滤头部带空格的SELECT语句。
- 系统默认部分SQL语句为白名单语句，默认白名单语句不可修改；可以通过系统视图gs\_wlm\_sql\_allow查询默认和已经通过GUC设置成功的SQL白名单语句。
- 通过wlm\_sql\_allow\_list指定的SQL语句不可追加，只能通过覆盖的方式设置；若需追加SQL语句，需要先查出原先指定的GUC值，在原值后面加补上新增的语句，以“;”分隔后重新设置。

## 17.14 自动清理

系统自动清理进程（autovacuum）自动执行VACUUM和ANALYZE命令，回收被标识为删除状态的记录空间，并更新表的统计数据。

### autovacuum

**参数说明：**控制是否启动数据库自动清理进程（autovacuum）。自动清理进程运行的前提是将**track\_counts**设置为on。

**参数类型：**SIGHUP

**取值范围：**布尔型

- on表示开启数据库自动清理进程。
- off表示关闭数据库自动清理进程。

**默认值：**on

#### 说明

如系统在故障恢复后，需具备自动清理两阶段事务的功能，请将autovacuum设置为on。

- 当设置autovacuum为on，**autovacuum\_max\_workers**为0时，表示系统不会自动进行autovacuum，只会在故障恢复后，自动清理两阶段事务。
- 当设置autovacuum为on，**autovacuum\_max\_workers**大于0时，表示系统不仅在故障恢复后，自动清理两阶段事务，并且还可以自动清理进程。

**须知**

即使此参数设置为off，数据库也会在需要防止事务ID回卷时发起清理进程。对于CREATE/DROP DATABASE发生异常时，可能有的节点提交或回滚，有的节点未提交（prepared状态），此时系统不能自动修复，需要手动修复，修复步骤：

1. 使用gs\_clean工具（-N参数）查询出异常两阶段事务的xid以及处于prepared的节点。
2. 登录事务处于prepared状态的节点，系统管理员连接一个可用的数据库（如gaussdb），执行语句set xc\_maintenance\_mode = on。
3. 根据事务全局状态提交或者回滚此两阶段事务（如提交语句、回滚语句）。

**autovacuum\_mode**

**参数说明：**该参数仅在autovacuum设置为on的场景下生效，它控制autoanalyze或autovacuum的打开情况。

**参数类型：**SIGHUP

**取值范围：**枚举类型

- analyze表示只执行autoanalyze。
- vacuum表示只执行autovacuum。
- mix表示autoanalyze和autovacuum都执行。
- none表示二者都不执行。

**默认值：**mix

**autoanalyze\_mode**

**参数说明：**设置autoanalyze的模式。该参数仅8.2.0及以上集群版本支持。

**参数类型：**USERSET

**取值范围：**枚举类型

- normal表示普通的autoanalyze。
- light表示轻量化的autoanalyze。

**默认值：**

- 若当前集群为低版本升级到8.2.0及以上集群版本，为保持和前向兼容，默认值为normal。
- 若当前集群为新装的8.2.0及以上集群版本，默认值为light。

**autoanalyze\_cache\_num**

**参数说明：**设置轻量化autoanalyze最多缓存多少张表的统计信息，超过该值后会触发淘汰，淘汰清理前100张表统计信息。该参数仅8.2.0及以上版本支持。

**参数类型：**SIGHUP

**取值范围：**整型，100 ~ INT\_MAX

**默认值：**10000

## autoanalyze\_timeout

**参数说明：**设置autoanalyze的超时时间。在对某张表做autoanalyze时，如果该表的analyze时长超过了autoanalyze\_timeout，则自动取消该表此次analyze。

**参数类型：**SIGHUP

**取值范围：**整型，0~2147483，单位为秒（s）。

**默认值：**5min

## analyze\_stats\_mode

**参数说明：**设置analyze计算统计信息的模式。

**参数类型：**USERSET

**取值范围：**枚举类型

- memory表示强制使用内存计算统计信息，不计算多列统计信息。
- sample\_table表示强制使用临时采样表计算统计信息，临时表不支持使用该模式。
- dynamic表示按内存maintenance\_work\_mem大小自适应选择统计信息计算模式，若maintenance\_work\_mem可放下样本，则使用内存方式，否则使用临时采样表方式。

**默认值：**

- 若当前集群为低版本升级到8.2.0.100及以上集群版本，为保持和前向兼容，默认值为memory。
- 若当前集群为新装的8.2.0.100及以上集群版本，默认值为dynamic。

## analyze\_sample\_mode

**参数说明：**设置analyze时使用的采样模型。

**参数类型：**USERSET

**取值范围：**整型，0~2。

- 0 表示使用默认蓄水池采样模型。
- 1 表示使用优化的蓄水池采样模型。
- 2 表示使用range采样模型。

**默认值：**0

## autovacuum\_io\_limits

**参数说明：**控制autovacuum进程每秒触发IO的上限。该参数8.1.2版本中已废弃，为兼容历史版本功能保留该函数，当前版本设置无效。

**参数类型：**SIGHUP

**取值范围：**整型，-1~1073741823。其中-1表示不控制，而是使用系统默认控制组。

**默认值：**-1



## log\_autovacuum\_min\_duration

**参数说明：**当自动清理的执行时间大于或者等于某个特定的值时，向服务器日志中记录自动清理执行的每一步操作。设置此选项有助于追踪自动清理的行为。

**参数类型：**SIGHUP

例如：将log\_autovacuum\_min\_duration设置为250ms，记录所有运行大于或者等于250ms的自动清理命令的相关信息。

**取值范围：**整型，-1~INT\_MAX，单位为毫秒（ms）。

- 当参数设置为0时，表示所有的自动清理操作都记录到日志中。
- 当参数设置为-1时，表示所有的自动清理操作都不记录到日志中。
- 当参数设置为非-1时，当由于锁冲突的存在导致一个自动清理操作被跳过，记录一条消息。

**默认值：**-1

## autovacuum\_max\_workers

**参数说明：**设置能同时运行的自动清理线程的最大数量。

**参数类型：**SIGHUP

**取值范围：**整型，0~128。其中0表示不会自动进行autovacuum。

**默认值：**3

### 📖 说明

该参数与**autovacuum**共同发挥作用，对系统表和用户表的清理规则如下：

- autovacuum\_max\_workers = 0时，autovacuum被彻底关闭，不会对任何表做清理。
- autovacuum\_max\_workers > 0 和 autovacuum = off，只对系统表和开了delta表的列存表做清理（如vacuum delta表，vacuum cudesc表和delta merge）。
- autovacuum\_max\_workers > 0 和 autovacuum = on，会对所有表做清理。

## autovacuum\_naptime

**参数说明：**设置两次自动清理操作的时间间隔。

**参数类型：**SIGHUP

**取值范围：**整型，1~2147483，单位为秒（s）。

**默认值：**60s

## autovacuum\_vacuum\_threshold

**参数说明：**设置触发VACUUM的阈值。当表上被删除或更新的记录数超过设定的阈值时才会对这个表执行VACUUM操作。

**参数类型：**SIGHUP

**取值范围：**整型，0~INT\_MAX

**默认值：**50

## autovacuum\_analyze\_threshold

**参数说明：**设置触发ANALYZE操作的阈值。当表上被删除、插入或更新的记录数超过设定的阈值时才会对这个表执行ANALYZE操作。

**参数类型：**SIGHUP

**取值范围：**整型，0~INT\_MAX

**默认值：**

- 若当前集群为低版本升级到8.1.3及以上集群版本，为保持和前向兼容，默认值为10000。
- 若当前集群为新装的8.1.3及以上集群版本，默认值为50。

## autovacuum\_vacuum\_scale\_factor

**参数说明：**设置触发一个VACUUM时增加到autovacuum\_vacuum\_threshold的表大小的缩放系数。

**参数类型：**SIGHUP

**取值范围：**浮点型，0.0~100.0

**默认值：**0.2

## autovacuum\_analyze\_scale\_factor

**参数说明：**设置触发一个ANALYZE时增加到autovacuum\_analyze\_threshold的表大小的缩放系数。

**参数类型：**SIGHUP

**取值范围：**浮点型，0.0~100.0

**默认值：**

- 若当前集群为低版本升级到8.1.3及以上集群版本，为保持和前向兼容，默认值为0.25。
- 若当前集群为新装的8.1.3及以上集群版本，默认值为0.1。

## autovacuum\_freeze\_max\_age

**参数说明：**设置事务内的最大时间，使得表的pg\_class.relfrozensid字段在VACUUM操作执行之前被写入。

VACUUM也可以删除pg\_clog/子目录中的旧文件；即使自动清理进程被禁止，系统也会调用自动清理进程来防止循环重复。

**参数类型：**SIGHUP

**取值范围：**整型，100 000~576 460 752 303 423 487

**默认值：**4000000000

## autovacuum\_vacuum\_cost\_delay

**参数说明：**设置在自动VACUUM操作里使用的开销延迟数值。

**参数类型:** SIGHUP

**取值范围:** 整型, -1 ~ 100, 单位为毫秒 (ms)。其中-1表示使用常规的vacuum\_cost\_delay。

**默认值:** 2ms

## autovacuum\_vacuum\_cost\_limit

**参数说明:** 设置在自动VACUUM操作里使用的开销限制数值。

**参数类型:** SIGHUP

**取值范围:** 整型, -1 ~ 10000。其中-1表示使用常规的vacuum\_cost\_limit。

**默认值:** -1

## colvacuum\_threshold\_scale\_factor

**参数说明:** 控制列存vacuum重写中, 重写文件最低dead tuple的比例值。当文件中的dead tuple占(all\_tuple - null\_tuple)的比例大于此值, 该文件才会被重写。

**参数类型:** SIGHUP

**取值范围:** 整型, -2 ~ 100。

- -2表示不会执行vacuum重写, 也不会执行vacuum清理;
- -1表示不会执行vacuum重写, 只会执行vacuum清理;
- 0-100为dead tuple的比例值。

**默认值:** 70

# 17.15 客户端连接缺省设置

## 17.15.1 语句行为

介绍SQL语句执行过程的相关默认参数。

### search\_path

**参数说明:** 当一个被引用对象没有指定模式时, 此参数设置模式搜索顺序。它的值由一个或多个模式名构成, 不同的模式名用逗号隔开。

**参数类型:** USERSET

- 当前会话如果存放临时表的模式时, 可以使用别名pg\_temp将它列在搜索路径中, 如'pg\_temp, public'。存放临时表的模式始终会作为第一个被搜索的对象, 排在pg\_catalog和search\_path中所有模式的前面, 即具有第一搜索优先级。建议用户不要在search\_path中显示设置pg\_temp。如果在search\_path中指定了pg\_temp, 但不是在最前面, 系统会提示设置无效, pg\_temp仍被优先搜索。通过使用别名pg\_temp, 系统只会在存放临时表的模式中搜索表、视图和数据类型这样的数据库对象, 不会在里面搜索函数或运算符这样的数据库对象。
- 系统表所在的模式pg\_catalog, 总是排在search\_path中指定的所有模式前面被搜索, 即具有第二搜索优先级 (pg\_temp具有第一搜索优先级)。建议用户不要在

search\_path中显式设置pg\_catalog。如果在search\_path中指定了pg\_catalog，但不是在最前面，系统会提示设置无效，pg\_catalog仍被第二优先搜索。

- 当没有指定一个特定模式而创建一个对象时，它们被放置到以search\_path为命名的第一个有效模式中。当搜索路径为空时，会报错误。
- 通过SQL函数current\_schema可以检测当前搜索路径的有效值。这和检测search\_path的值不尽相同，因为current\_schema显示search\_path中首位有效的模式名称。

**取值范围：**字符串

#### 说明

- 设置为"\$user", public时，支持共享数据库（没有用户具有私有模式和所有共享使用public），用户私有模式和这些功能的组合使用。可以通过改变默认搜索路径来获得其他效果，无论是全局化的还是私有化的。
- 设置为空串（"）的时候，系统会自动转换成一对双引号。
- 设置的内容中包含双引号，系统会认为是不安全字符，会将每个双引号转换成一对双引号。

**默认值：**"\$user",public

#### 说明

\$user表示与当前会话用户名同名的模式名，如果这样的模式不存在，\$user将被忽略。

## current\_schema

**参数说明：**设置当前的模式。

**参数类型：**USERSET

**取值范围：**字符串

**默认值：**"\$user",public

#### 说明

\$user表示与当前会话用户名同名的模式名，如果这样的模式不存在，\$user将被忽略。

## default\_tablespace

**参数说明：**当CREATE命令没有明确声明表空间时，所创建对象(表和索引等)的缺省表空间。

- 值是一个表空间的名字或者一个表示使用当前数据库缺省表空间的空字符串。若指定的是一个非默认表空间，用户必须具有它的CREATE权限，否则尝试创建会失败。
- 临时表不使用此参数，可以用temp\_tablespaces代替。
- 创建数据库时不使用此参数。默认情况下，一个新的数据库从模板数据库继承表空间配置。

**参数类型：**USERSET

**取值范围：**字符串，其中空表示使用默认表空间。

**默认值：**空

## default\_storage\_nodegroup

**参数说明：**设置当前的默认建表所在的Node Group，目前只适用普通表。

**参数类型：**USERSET

**取值范围：**字符串

- installation，表示默认建表在安装的Node Group上。
- random\_node\_group，表示默认建表在随机选择的NodeGroup上，该配置8.1.2及以上版本支持，仅用于测试环境。
- roach\_group，表示默认建表在所有节点上，该值为roach工具预留，不能用于其他场景。
- 取值为其他字符串，表示默认建表在设置的Node Group上。

**默认值：**installation

## default\_colversion

**参数说明：**设置当前默认创建列存表的存储格式版本。

**参数类型：**SIGHUP

**取值范围：**枚举类型

- 1.0表示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表示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

## temp\_tablespaces

**参数说明：**当一个CREATE命令没有明确指定一个表空间时，temp\_tablespaces指定了创建临时对象（临时表和临时表的索引）所在的表空间。在这些表空间中创建临时文件用来做大型数据的排序工作。

其值是一系列表空间名的列表。如果列表中有多个表空间时，每次临时对象的创建，GaussDB(DWS)会在列表中随机选择一个表空间；如果在事务中，连续创建的临时对象被放置在列表里连续的表空间中。如果选择的列表中的元素是一个空串，GaussDB(DWS)将自动将当前的数据库设为默认的表空间。

**参数类型：**USERSET

**取值范围：**字符串。空字符串表示所有的临时对象仅在当前数据库默认的表空间中创建，请参见[default\\_tablespace](#)。

**默认值：**空

## check\_function\_bodies

**参数说明：**设置是否在CREATE FUNCTION执行过程中进行函数体字符串的合法性验证。为了避免产生问题（比如避免从转储中恢复函数定义时向前引用的问题），偶尔会禁用验证。

**参数类型：**USERSET

**取值范围：**布尔型

- on表示在CREATE FUNCTION执行过程中进行函数体字符串的合法性验证。
- off表示在CREATE FUNCTION执行过程中不进行函数体字符串的合法性验证。

**默认值：**on

## default\_transaction\_isolation

**参数说明：**设置默认的事务隔离级别。

**参数类型：**USERSET

**取值范围：**枚举型

- 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。这是缺省值。
- read uncommitted: 读未提交隔离级别，GaussDB(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它并发事务提交的修改。
- serializable: 事务可序列化，GaussDB(DWS)不支持SERIALIZABLE，如果设置了serializable，实际上使用的是repeatable read。

**默认值：**read committed

## default\_transaction\_read\_only

**参数说明：**设置每个新创建事务是否是只读状态。

**参数类型：**SIGHUP

**取值范围：**布尔型

- on表示只读状态。
- off表示非只读状态。

**默认值：**off

## default\_transaction\_deferrable

**参数说明：**控制每个新事务的默认延迟状态。只读事务或者那些比序列化更加低的隔离级别的事务除外。

GaussDB(DWS)不支持可串行化的隔离级别，因此，该参数无实际意义。

**参数类型：**USERSET

**取值范围：**布尔型

- on表示默认延迟。
- off表示默认不延迟。

**默认值：**off

## session\_replication\_role

**参数说明：**控制当前会话与复制相关的触发器和规则的行为。

**参数类型：**USERSET

### 须知

设置此参数会丢弃之前任何缓存的执行计划。

**取值范围：**枚举型

- origin表示从当前会话中复制插入、删除、更新等操作。
- replica表示从其他地方复制插入、删除、更新等操作到当前会话。
- local表示函数执行复制时会检测当前登录数据库的角色并采取相应的操作。

**默认值：**origin

## statement\_timeout

**参数说明：**当语句执行时间超过该参数设置的时间（从服务器收到命令时开始计时）时，该语句将会报错并退出执行。

**参数类型：**USERSET

**取值范围：**整型，0~2147483647，单位为毫秒（ms）。

**默认值：**

- 若当前集群为低版本升级到8.2.0版本，继承升级前参数，默认值为0。
- 若当前集群为新装的8.2.0版本，默认值为24h。

## vacuum\_freeze\_min\_age

**参数说明：**指定VACUUM在扫描一个表时用于判断是否用FrozenXID替换事务ID的中断寿命(在同一个事务中)。

**参数类型：**USERSET

**取值范围：**整型，0~576 460 752 303 423 487

### 说明

尽管随时可以将此参数设为0到10亿之间的任意值，但是，VACUUM将默认其有效值范围限制在 `autovacuum_freeze_max_age` 的50%以内。

**默认值：**5000000000

## vacuum\_freeze\_table\_age

**参数说明：**指定VACUUM对全表的扫描冻结元组的时间。如果表的 `pg_class.relfrozensid` 字段的值已经达到了参数指定的时间，VACUUM对全表进行扫描。

**参数类型：**USERSET

**取值范围：** 整型，0 ~ 576 460 752 303 423 487

#### 📖 说明

尽管随时可以将此参数设为零到20亿之间的值，但是，VACUUM将默认其有效值范围限制在 `autovacuum_freeze_max_age` 的95%以内。定期的手动VACUUM可以在对此表的反重叠自动清理启动之前运行。

**默认值：** 15000000000

## bytea\_output

**参数说明：** 设置bytea类型值的输出格式。

**参数类型：** USERSET

**取值范围：** 枚举型

- hex：将二进制数据编码为每字节2位十六进制数字。
- escape：传统化的PostgreSQL格式。采用以ASCII字符序列表示二进制串的方法，同时将那些无法表示成ASCII字符的二进制串转换成特殊的转义序列。

**默认值：** hex

## xmlbinary

**参数说明：** 设置二进制值是如何在XML中进行编码的。

**参数类型：** USERSET

**取值范围：** 枚举型

- base64
- hex

**默认值：** base64

## xmloption

**参数说明：** 当XML和字符串值之间进行转换时，设置document或content是否是隐含的。

**参数类型：** USERSET

**取值范围：** 枚举型

- document：表示HTML格式的文档。
- content：普通的字符串。

**默认值：** content

## max\_compile\_functions

**参数说明：** 设置服务器存储的函数编译结果的最大数量。存储过多的函数和存储过程的编译结果可能占用很大内存。将此参数设置为一个合理的值，有助于减少内存占用，提升系统性能。

**参数类型：** POSTMASTER



**取值范围：**整型，1 ~ INT\_MAX。

**默认值：**1000

## gin\_pending\_list\_limit

**参数说明：**设置当GIN索引启用fastupdate时，pending list容量的最大值。当pending list的容量大于设置值时，会把pending list中数据批量移动到GIN索引数据结构中进行清理。单个GIN索引可通过更改索引存储参数覆盖此设置值。

**参数类型：**USERSET

**取值范围：**整型，64 ~ INT\_MAX，单位为KB。

**默认值：**4MB

## 17.15.2 区域和格式化

介绍时间格式设置的相关参数。

### DateStyle

**参数说明：**设置日期和时间值的显示格式，以及有歧义的输入值的解析规则。

这个变量包含两个独立的部分：输出格式声明（ISO、Postgres、SQL、German）和输入输出的年/月/日顺序（DMY、MDY、YMD）。这两个可以独立设置或者一起设置。关键字Euro和European等价于DMY；关键字US、NonEuro、NonEuropean等价于MDY。

**参数类型：**USERSET

**取值范围：**字符串

**默认值：**ISO, MDY

#### 说明

gs\_initdb会将这个参数初始化成与lc\_time一致的值。

**设置建议：**优先推荐使用ISO格式。Postgres、SQL和German均采用字母缩写的形式来表示时区，例如“EST、WST、CST”等。

### IntervalStyle

**参数说明：**设置区间值的显示格式。

**参数类型：**USERSET

**取值范围：**枚举型

- sql\_standard表示产生与SQL标准规定匹配的输出生。
- postgres表示产生与PostgreSQL 8.4版本相匹配的输出，当DateStyle参数被设为ISO时。
- postgres\_verbose表示产生与PostgreSQL 8.4版本相匹配的输出，当DateStyle参数被设为non\_ISO时。
- iso\_8601表示产生与在ISO 8601中定义的“格式与代号”相匹配的输出。

- oracle表示产生于Oracle中与numtodsinterval函数相匹配的输出结果，详细请参考numtodsinterval。

#### 须知

IntervalStyle参数也会影响不明确的间隔输入の説明。

默认值: postgres

## TimeZone

**参数说明:** 设置显示和解释时间类型数值时使用的时区。

**参数类型:** USERSET

**取值范围:** 字符串，可查询视图pg\_timezone\_names获得。

**默认值:** UTC

#### 说明

gs\_initdb将设置一个与其系统环境一致的时区值。

## timezone\_abbreviations

**参数说明:** 设置服务器接受的时区缩写值。

**参数类型:** USERSET

**取值范围:** 字符串，可查询视图pg\_timezone\_names获得。

**默认值:** Default

#### 说明

Default表示通用时区的缩写。但也有其他诸如 'Australia' 和 'India' 等用来定义特定的安装。

## extra\_float\_digits

**参数说明:** 调整浮点值显示的数据位数，浮点类型包括float4、float8 以及几何数据类型。参数值加在标准的数据位数上（FLT\_DIG或DBL\_DIG中合适的）。

**参数类型:** USERSET

**取值范围:** 整型，-15 ~ 3

#### 说明

- 设置为3，表示包括部分有效的数据位。对转储需要精确恢复的浮点数据尤其有用。
- 设置为负数，表示摒弃不需要的数据位。

**默认值:** 0

## client\_encoding

**参数说明:** 设置客户端的字符编码类型。

请根据前端业务的情况确定。尽量客户端编码和服务器端编码一致，提高效率。

**参数类型：**USERSET

**取值范围：**兼容PostgreSQL所有的字符编码类型。其中UTF8表示使用数据库的字符编码类型。

#### 说明

- 使用命令locale -a查看当前系统支持的区域和相应的编码格式，并可以选择进行设置。
- 默认情况下，gs\_initdb会根据当前的系统环境初始化此参数，通过locale命令可以查看当前的配置环境。
- 参数建议保持默认值，不建议通过gs\_guc工具或其他方式直接在postgresql.conf文件中设置client\_encoding参数，即使设置也不会生效，以保证集群内部通信编码格式一致。

**默认值：**UTF8

**推荐值：**SQL\_ASCII/UTF8

## lc\_messages

**参数说明：**设置信息显示的语言。

可接受的值是与系统相关的；在一些系统上，这个区域范畴并不存在，不过仍然允许设置这个变量，只是不会有任何效果。同样，也有可能是所期望的语言的翻译信息不存在。在这种情况下，用户仍然能看到英文信息。

**参数类型：**SUSET

**取值范围：**字符串

#### 说明

- 使用命令locale -a查看当前系统支持的区域和相应的编码格式，并可以选择进行设置。
- 默认情况下，gs\_initdb会根据当前的系统环境初始化此参数，通过locale命令可以查看当前的配置环境。

**默认值：**C

## lc\_monetary

**参数说明：**设置货币值的显示格式，影响to\_char之类的函数的输出。可接受的值是系统相关的。

**参数类型：**USERSET

**取值范围：**字符串

#### 说明

- 使用命令locale -a查看当前系统支持的区域和相应的编码格式，并可以选择进行设置。
- 默认情况下，gs\_initdb会根据当前的系统环境初始化此参数，通过locale命令可以查看当前的配置环境。

**默认值：**C

## lc\_numeric

**参数说明：**设置数值的显示格式，影响to\_char之类的函数的输出。可接受的值是系统相关的。

**参数类型：**USERSET

**取值范围：**字符串

### 说明

- 使用命令locale -a查看当前系统支持的区域和相应的编码格式，并可以选择进行设置。
- 默认情况下，gs\_initdb会根据当前的系统环境初始化此参数，通过locale命令可以查看当前的配置环境。

**默认值：**C

## lc\_time

**参数说明：**设置时间和区域的显示格式，影响to\_char之类的函数的输出。可接受的值是系统相关的。

**参数类型：**USERSET

**取值范围：**字符串

### 说明

- 使用命令locale -a查看当前系统支持的区域和相应的编码格式，并可以选择进行设置。
- 默认情况下，gs\_initdb会根据当前的系统环境初始化此参数，通过locale命令可以查看当前的配置环境。

**默认值：**C

## default\_text\_search\_config

**参数说明：**设置全文检索的配置信息。

如果设置为不存在的文本搜索配置时将会报错。如果default\_text\_search\_config对应的文本搜索配置被删除，需要重新设置default\_text\_search\_config，否则会报设置错误。

- 其被文本搜索函数使用，这些函数并没有一个明确指定的配置。
- 当与环境相匹配的配置文件确定时，gs\_initdb会选择一个与环境相对应的设置来初始化配置文件。

**参数类型：**USERSET

**取值范围：**字符串

### 说明

GaussDB(DWS)支持pg\_catalog.english，pg\_catalog.simple两种配置。

**默认值：**pg\_catalog.english

### 17.15.3 其他缺省

主要介绍数据库系统默认的库加载参数。

#### dynamic\_library\_path

**参数说明：**设置数据查找动态加载的共享库文件的路径。当需要打开一个可以动态加载的模块并且在CREATE FUNCTION或LOAD命令里面声明的名字没有目录部分时，系统将搜索这个目录以查找声明的文件。

用于dynamic\_library\_path的数值必须是一个冒号分隔（Windows下是分号分隔）的绝对路径列表。当一个路径名字以特殊变量\$libdir为开头时，会替换为GaussDB(DWS)发布提供的模块安装路径。例如：

```
dynamic_library_path = '/usr/local/lib/postgresql:/opt/testtgs/lib:$libdir'
```

**参数类型：**SUSET

**取值范围：**字符串

#### 📖 说明

设置为空字符串，表示关闭自动路径搜索。

**默认值：**\$libdir

#### gin\_fuzzy\_search\_limit

**参数说明：**设置GIN索引返回的集合大小的上限。

**参数类型：**USERSET

**取值范围：**整型，0~INT\_MAX，0表示没有限制。

**默认值：**0

## 17.16 锁管理

在GaussDB(DWS)中，并发执行的事务由于竞争资源会导致死锁。本节介绍的参数主要管理事务锁的机制。

#### deadlock\_timeout

**参数说明：**设置死锁超时检测时间，以毫秒为单位。当申请的锁超过设定值时，系统会检查是否产生了死锁。

- 死锁的检查代价是比较高的，服务器不会在每次等待锁的时候都运行这个过程。在系统运行过程中死锁是不经常出现的，因此在检查死锁前只需等待一个相对较短的时间。增加这个值就减少了无用的死锁检查浪费的时间，但是会减慢真正的死锁错误报告的速度。在一个负载过重的服务器上，用户可能需要增大它。这个值的设置应该超过事务持续时间，这样就可以减少在锁释放之前就开始死锁检查的问题。
- 设置**log\_lock\_waits**时，这个选项也决定了在一个日志消息发出关于锁等待以前要等待的时间。当需要调查锁延迟时，请设置比正常deadlock\_timeout更小的值。

**参数类型：**SUSET

**取值范围：**整型，1~2147483647，单位为毫秒（ms）。

**默认值：**1s

## ddl\_lock\_timeout

**参数说明：**通过该参数单独指定阻塞DDL语句锁等待的时间，当申请的锁等待时间超过设定值时，系统会报错。该参数仅8.1.3.200及以上版本支持。

**参数类型：**SUSET

**取值范围：**整型，0 ~ INT\_MAX，单位为毫秒（ms）。

该参数优先级高于lockwait\_timeout，只针对AccessExclusiveLock生效。

- 如果该参数的值等于0，表示该参数不生效。
- 如果该参数的值大于0，DDL锁阻塞时间为该参数的值，其它锁等待时间为lockwait\_timeout参数值。

**默认值：**0

## ddl\_select\_concurrent\_mode

**参数说明：**通过该参数控制ddl语句和select语句并发的模式。该参数仅8.2.0.100及以上集群版本支持。

**参数类型：**SUSET

**取值范围：**枚举型

- off：表示该参数不生效，DDL语句和select语句不能并发，保持锁等待状态。
- truncate：表示允许truncate语句和select语句并发，其它DDL语句和select语句保持锁等待状态。

**默认值：**off

### 📖 说明

当前版本中设置的ddl\_lock\_timeout的值不足1秒时按照1s处理，原因是为了给select语句预留响应信号的时间。

## lockwait\_timeout

**参数说明：**控制单个锁的最长等待时间。当申请的锁等待时间超过设定值时，系统会报错。

**参数类型：**SUSET

**取值范围：**整型，0 ~ INT\_MAX，单位为毫秒（ms）。

**默认值：**20min

## update\_lockwait\_timeout

**参数说明：**允许并发更新参数开启情况下，该参数控制并发更新同一行时单个锁的最长等待时间。当申请的锁等待时间超过设定值时，系统会报错。

**参数类型：**SUSET

**取值范围：**整型，0 ~ INT\_MAX，单位为毫秒（ms）。

**默认值：**2min

## max\_locks\_per\_transaction

**参数说明：**控制每个事务能够得到的平均的对象锁的数量。

- 共享的锁表的大小是以假设任意时刻最多只有  $\text{max\_locks\_per\_transaction} * (\text{max\_connections} + \text{max\_prepared\_transactions})$  个独立的对象需要被锁住为基础进行计算的。不超过设定数量的多个对象可以在任一时刻同时被锁定。当在一个事务里面修改很多不同的表时，可能需要提高这个默认数值。只能在数据库启动的时候设置。
- 增大这个参数可能导致GaussDB(DWS)请求更多的System V共享内存，有可能超过操作系统的缺省配置。
- 当运行备机时，请将此参数设置不小于主机上的值，否则，在备机上查询操作不会被允许。

**参数类型：**POSTMASTER

**取值范围：**整型，10 ~ INT\_MAX

**默认值：**256

## max\_pred\_locks\_per\_transaction

**参数说明：**控制每个事务允许断定锁的最大数量，是一个平均值。

- 共享的断定锁表的大小是以假设任意时刻最多只有  $\text{max\_pred\_locks\_per\_transaction} * (\text{max\_connections} + \text{max\_prepared\_transactions})$  个独立的对象需要被锁住为基础进行计算的。不超过设定数量的多个对象可以在任一时刻同时被锁定。当在一个事务里面修改很多不同的表时，可能需要提高这个默认数值。只能在服务器启动的时候设置。
- 增大这个参数可能导致GaussDB(DWS)请求更多的System V共享内存，有可能超过操作系统的缺省配置。

**参数类型：**POSTMASTER

**取值范围：**整型，10 ~ INT\_MAX

**默认值：**64

## partition\_lock\_upgrade\_timeout

**参数说明：**分区上的锁级别由允许读的ExclusiveLock升级到读写阻塞的AccessExclusiveLock时，会进行尝试性的锁升级，partition\_lock\_upgrade\_timeout指示了尝试锁升级的超时时间。

- 在分区表上进行MERGE PARTITION和CLUSTER PARTITION操作时，都利用了临时表进行数据重排和文件交换，为了最大程度提高分区上的操作并发度，在数据重排阶段给相关分区加锁ExclusiveLock，在文件交换阶段加锁AccessExclusiveLock。
- 常规加锁方式是等待加锁，直到加锁成功，或者等待时间超过[lockwait\\_timeout](#)发生超时失败。

- 在分区表上进行MERGE PARTITION或CLUSTER PARTITION操作时，进入文件交换阶段需要申请加锁AccessExclusiveLock，加锁方式是尝试性加锁，加锁成功了则立即返回，不成功则等待50ms后继续下次尝试，加锁超时时间使用会话级设置参数partition\_lock\_upgrade\_timeout。
- 特殊值：若partition\_lock\_upgrade\_timeout取值-1，表示无限等待，即不停的尝试锁升级，直到加锁成功。

**参数类型：**USERSET

**取值范围：**整型，-1 ~ 3000，单位为秒（s）。

**默认值：**1800

## enable\_online\_ddl\_waitlock

**参数说明：**控制DDL是否会阻塞等待pg\_advisory\_lock/pgxc\_lock\_for\_backup等集群锁。主要用于OM在线操作场景，不建议用户设置。

**参数类型：**SIGHUP

**取值范围：**布尔型

- on表示开启。
- off表示关闭。

**默认值：**off

## 17.17 版本和平台兼容性

### 17.17.1 历史版本兼容性

GaussDB(DWS)介绍数据库的向下兼容性和对外兼容性特性的参数控制。数据库系统的向后兼容性能够为对旧版本的数据库应用提供支持。本节介绍的参数主要控制数据库的向后兼容性。

## array\_nulls

**参数说明：**控制数组输入解析器是否将未用引用的NULL识别为数组的一个NULL元素。

**参数类型：**USERSET

**取值范围：**布尔型

- on表示允许向数组中输入空元素。
- off表示向下兼容旧式模式。仍然能够创建包含NULL值的数组。

**默认值：**on

## backslash\_quote

**参数说明：**控制字符串文本中的单引号是否够用\表示。

**参数类型：**USERSET



**须知**

在字符串文本符合SQL标准的情况下，\没有任何其他含义。这个参数影响的是如何处理不符合标准的字符串文本，包括明确的字符串转义语法是（E'...'）。

**取值范围：**枚举类型

- on表示一直允许使用\表示。
- off表示拒绝使用\表示。
- safe\_encoding表示仅在客户端字符集编码不会在多字节字符末尾包含\的ASCII值时允许。

**默认值：**safe\_encoding

## default\_with\_oids

**参数说明：**在没有声明WITH OIDS和WITHOUT OIDS的情况下，这个选项控制在新创建的表中CREATE TABLE和CREATE TABLE AS是否包含一个OID字段。它还决定SELECT INTO创建的表里面是否包含OID。

不推荐在用户表中使用OID，故默认设置为off。需要带有OID字段的表应该在创建时声明WITH OIDS。

**参数类型：**USERSET

**取值范围：**布尔型

- on表示在新创建的表中CREATE TABLE和CREATE TABLE AS可以包含一个OID字段。
- off表示在新创建的表中CREATE TABLE和CREATE TABLE AS不可以包含一个OID字段。

**默认值：**off

## escape\_string\_warning

**参数说明：**警告在普通字符串中直接使用反斜杠转义。

- 如果需要使用反斜杠作为转义，可以调整为使用转义字符串语法(E'...')来做转义，因为在每个SQL标准中，普通字符串的默认行为现在将反斜杠作为一个普通字符。
- 这个变量可以帮助定位需要改变的代码。

**参数类型：**USERSET

**取值范围：**布尔型

**默认值：**on

## lo\_compat\_privileges

**参数说明：**控制是否启动对大对象权限检查的向后兼容模式。

**参数类型：**SUSET

**取值范围：**布尔型

on表示当读取或修改大对象时禁用权限检查，与PostgreSQL 9.0以前的版本兼容。

**默认值：** off

## quote\_all\_identifiers

**参数说明：** 当数据库生成SQL时，此选项强制引用所有的标识符（包括非关键字）。这将影响到EXPLAIN的输出及函数的结果，例如pg\_get\_viewdef。详细说明请参见gs\_dump的--quote-all-identifiers选项。

**参数类型：** USERSET

**取值范围：** 布尔型

- on表示打开强制引用。
- off表示关闭强制引用。

**默认值：** off

## sql\_inheritance

**参数说明：** 控制继承语义。

**参数类型：** USERSET

**取值范围：** 布尔型

off表示各种命令不能访问子表，即默认使用ONLY关键字。这是为了兼容7.1之前版本而设置的。

**默认值：** on

## standard\_conforming\_strings

**参数说明：** 控制普通字符串文本 ('...') 中是否按照SQL标准把反斜杠当普通文本。

- 应用程序通过检查这个参数可以判断字符串文本的处理方式。
- 建议明确使用转义字符串语法 (E'...') 来转义字符。

**参数类型：** USERSET

**取值范围：** 布尔型

- on表示打开控制功能。
- off表示关闭控制功能。

**默认值：** on

## synchronize\_seqscans

**参数说明：** 控制启动同步的顺序扫描。在大约相同的时间内并行扫描读取相同的数据块，共享I/O负载。

**参数类型：** USERSET

**取值范围：** 布尔型

- on表示扫描可能从表的中间开始，然后选择"环绕"方式来覆盖所有的行，为了与已经在进行中的扫描活动同步。这可能会造成没有用ORDER BY子句的查询得到行排序造成不可预测的后果。
- off表示确保顺序扫描是从表头开始的。

默认值: on

## enable\_beta\_features

**参数说明:** 控制开启某些功能受限的特性，例如GDS表关联操作。这些特性在历史版本未明确禁止，但某些场景功能上存在问题，不建议用户使用。

**参数类型:** USERSET

**取值范围:** 布尔型

- on表示开启这些功能受限的特性，保持前向兼容。但某些场景可能存在功能上的问题。
- off表示禁止使用这些特性。

默认值: off

## 17.17.2 平台和客户端兼容性

很多平台都使用数据库系统，数据库系统的对外兼容性给平台提供了很大的方便。

### transform\_null\_equals

**参数说明:** 控制表达式 $expr = NULL$ （或 $NULL = expr$ ）当做 $expr IS NULL$ 处理。如果 $expr$ 得出NULL值则返回真，否则返回假。

- 正确的SQL标准兼容的 $expr = NULL$ 总是返回NULL（未知）。
- Microsoft Access里的过滤表单生成的查询使用 $expr = NULL$ 来测试空值。打开这个选项，可以使用该接口来访问数据库。

**参数类型:** USERSET

**取值范围:** 布尔型

- on表示控制表达式 $expr = NULL$ （或 $NULL = expr$ ）当做 $expr IS NULL$ 处理。
- off表示不控制，即 $expr = NULL$ 总是返回NULL（未知）。

默认值: off

#### 说明

新用户经常在涉及NULL的表达式上语义混淆，故默认值设为off。

### td\_compatible\_truncation

**参数说明:** 控制是否开启与Teradata数据库相应兼容的特征。该参数在用户连接上与TD兼容的数据库时，可以将参数设置成为on（即超长字符串自动截断功能启用），该功能启用后，在后续的insert语句中，对目标表中char和varchar类型的列插入超长字符串时，会按照目标表中相应列定义的最大长度对超长字符串进行自动截断。保证数据都能插入目标表中，而不是报错。

**说明**

- 超长字符串自动截断功能不适用于insert语句包含外表的场景。
- 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。

**参数类型：**USERSET

**取值范围：**布尔型

- on表示启动超长字符串自动截断功能。
- off表示停止超长字符串自动截断功能。

**默认值：**off

## 17.18 容错性

当数据库系统发生错误时，以下参数控制服务器处理错误的方式。

### exit\_on\_error

**参数说明：**控制终止会话。

**参数类型：**SUSET

**取值范围：**布尔型

- on表示任何错误都会终止当前的会话。
- off表示只有FATAL级别的错误才会终止会话。

**默认值：**off

### omit\_encoding\_error

**参数说明：**设置为on，数据库的客户端字符集编码为UTF-8时，出现的字符编码转换错误将打印在日志中，有转换错误的被转换字符会被忽略，以"?"代替。

**参数类型：**USERSET

**取值范围：**布尔型

- on表示有转换错误的字符将被忽略，以"?"代替，打印错误信息到日志中。
- off表示有转换错误的字符不能被转换，打印错误信息到终端。

**默认值：**off

### max\_query\_retry\_times

**参数说明：**指定SQL语句出错自动重试功能的最大重跑次数（目前支持重跑的错误类型为“Connection reset by peer”、“Lock wait timeout”和“Connection timed out”等，设定为0时关闭重跑功能。

**参数类型：**USERSET

**取值范围：**整型，0~20

**默认值:** 6

## cn\_send\_buffer\_size

**参数说明:** 指定CN端数据发送数据缓存区的大小。

**参数类型:** POSTMASTER

**取值范围:** 整型, 8~128, 单位为KB。

**默认值:** 8KB

## max\_cn\_temp\_file\_size

**参数说明:** 指定SQL语句出错自动重试功能中CN端使用临时文件的最大值, 设定为0表示不使用临时文件。

**参数类型:** SIGHUP

**取值范围:** 整型, 0~10485760, 单位为KB。

**默认值:** 5GB

## retry\_ecode\_list

**参数说明:** 指定SQL语句出错自动重试功能支持的错误类型列表。

**参数类型:** USERSET

**取值范围:** 字符串

**默认值:** YY001 YY002 YY003 YY004 YY005 YY006 YY007 YY008 YY009 YY010  
YY011 YY012 YY013 YY014 YY015 53200 08006 08000 57P01 XX003 XX009 YY016  
CG003 CG004 F0011 F0012 45003

## data\_sync\_retry

**参数说明:** 控制当fsync到磁盘失败后是否继续运行数据库。由于在某些操作系统的场景下, fsync失败后重试阶段即使再次fsync失败也不会报错, 从而导致数据丢失。

**参数类型:** POSTMASTER

**取值范围:** 布尔型

- on表示当fsync同步到磁盘失败后采取重试机制, 数据库继续运行。
- off表示当fsync同步到磁盘失败后直接报panic, 停止数据库。

**默认值:** off

## 17.19 连接池参数

当使用连接池访问数据库时, 在系统运行过程中, 数据库连接是被当作对象存储在内存中的, 当用户需要访问数据库时, 并非建立一个新的连接, 而是从连接池中取出一个已建立的空闲连接来使用。用户使用完毕后, 数据库并非将连接关闭, 而是将连接放回连接池中, 以供下一个请求访问使用。

## min\_pool\_size

**参数说明：** CN的连接池与其它某个CN/DN的最小连接数。

**参数类型：** POSTMASTER

**取值范围：** 整型， 1~65535

**默认值：** 1

## max\_pool\_size

**参数说明：** CN的连接池与其它某个CN/DN的最大连接数。

**参数类型：** POSTMASTER

**取值范围：** 整型， 1~65535

**默认值：** CN为800， DN为5000

## persistent\_datanode\_connections

**参数说明：** 会话是否会释放获得的连接。

**参数类型：** USERSET

**取值范围：** 布尔型

- off表示会释放获得连接。
- on表示不会释放获得连接。

---

### 须知

打开此开关后，会存在会话持有连接但并未运行查询的情况，导致其他查询申请不到连接报错。出现此问题时，需约束会话数量小于等于max\_active\_statements。

---

**默认值：** off

## max\_coordinators

**参数说明：** 集群中CN的最大数目。

**参数类型：** POSTMASTER

**取值范围：** 整型， 2~40

**默认值：** 40

## max\_datanodes

**参数说明：** 集群中DN的最大数目。

**参数类型：** POSTMASTER

**取值范围：** 整型， 2~65535

**默认值：** 4096

## cache\_connection

**参数说明：** 是否回收连接池的连接。

**参数类型：** SIGHUP

**取值范围：** 布尔型

- on表示回收连接池的连接。
- off表示不回收连接池的连接。

**默认值：** on

## enable\_force\_reuse\_connections

**参数说明：** 会话是否强制重用新的连接。

**参数类型：** BACKEND

**取值范围：** 布尔型

- on表示强制使用新连接。
- off表示使用现有连接。

**默认值：** off

### 说明

会话连接参数，不建议用户设置。

## enable\_pooler\_parallel

**参数说明：** CN的连接池是否可以在并行的模式下进行连接。

**参数类型：** SIGHUP

**取值范围：** 布尔型

- on表示CN的连接池可以在并行的模式下进行连接。
- off表示CN的连接池不可以在并行的模式下进行连接。

**默认值：** on

## 17.20 集群事务

介绍集群事务隔离、事务只读、最大prepared事务数、集群维护模式目的参数设置及取值范围等内容。

## transaction\_isolation

**参数说明：** 设置当前事务的隔离级别。

**参数类型：** USERSET

**取值范围：**

- 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。这是缺省值。
- read uncommitted: 读未提交隔离级别, GaussDB(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它并发事务提交的修改。
- serializable: 事务可序列化, GaussDB(DWS)不支持SERIALIZABLE, 如果设置了serializable, 实际上使用的是repeatable read。

**默认值:** read committed

## transaction\_read\_only

**参数说明:** 设置当前事务是只读事务。

**参数类型:** USERSET

**取值范围:** 布尔型

- on表示设置当前事务为只读事务。
- off表示该事务可以是非只读事务。

**默认值:** CN节点为off, DN节点为on。

## xc\_maintenance\_mode

**参数说明:** 设置系统进入维护模式。

**参数类型:** SUSERSET

**取值范围:** 布尔型

- on表示该功能启用。
- off表示该功能被禁用。

**默认值:** off

---

### 须知

谨慎打开这个开关, 避免引起集群数据不一致。

---

## allow\_concurrent\_tuple\_update

**参数说明:** 设置是否允许并发更新。

**参数类型:** USERSET

**取值范围:** 布尔型

- on表示该功能启用。
- off表示该功能被禁用。

**默认值:** on



## gtm\_backup\_barrier

**参数说明：**指定是否为GTM启动点创建还原点。

**参数类型：**SUSET

**取值范围：**布尔型

- on表示创建还原点。
- off表示不创建还原点。

**默认值：**off

## gtm\_conn\_check\_interval

**参数说明：**设置CN检查本地线程与主GTM连接是否正常时间。

**参数类型：**SUSET

**取值范围：**整型，0 ~ INT\_MAX / 1000，单位为秒。

**默认值：**10s

## transaction\_deferrable

**参数说明：**指定是否允许一个只读串行事务延迟执行，使其不会执行失败。该参数设置为on时，当一个只读事务发现读取的元组正在被其他事务修改，则延迟该只读事务直到其他事务修改完成。目前，GaussDB(DWS)暂时未用到这个参数。与该参数类似的还有一个[default\\_transaction\\_deferrable](#)，设置它来指定一个事务是否允许延迟。

**参数类型：**USERSET

**取值范围：**布尔型

- on表示允许执行。
- off表示不允许执行。

**默认值：**off

## enforce\_two\_phase\_commit

**参数说明：**为了兼容历史版本功能保留该参数，当前版本设置无效。

## enable\_show\_any\_tuples

**参数说明：**该参数只有在只读事务中可用，用于分析。当这个参数被置为on/true时，表中元组的所有版本都会可见。

**参数类型：**USERSET

**取值范围：**布尔型

- on/true表示表中元组的所有版本都会可见。
- off/false表示表中元组的所有版本都不可见。

**默认值：**off

## gtm\_connect\_retries

**参数说明：**控制GTM连接重试的次数。

**参数类型：**SIGHUP

**取值范围：**int，最小值为1，最大值为2147483647。

**默认值：**30

## enable\_redistribute

**参数说明：**节点不匹配时是否重新分配。

**参数类型：**SUSET

**取值范围：**布尔型

- on表示节点不匹配时重新分配。
- off表示节点不匹配时不重新分配。

**默认值：**off

## enable\_gtm\_free

**参数说明：**大并发场景下同一时刻存在活跃事务较多，GTM下发的快照变大且快照请求变多的情况下，瓶颈卡在GTM与CN通讯的网络上。为消除该瓶颈，引入GTM-FREE模式。取消CN和GTM的交互，取消CN下发GTM获取的事务信息给DN。CN只向各个DN发送query，各个DN由本地产生快照及xid等信息，开启该参数支持分布式事务读最终一致性，即分布式事务只有写外部一致性，不具有读外部一致性。

对于要求强一致性读的OLTP场景或OLAP场景，建议不要开启该参数。GaussDB(DWS)不支持该特性，设置后无法生效。

**参数类型：**SUSET

**取值范围：**布尔型

- on表示开启GTM-FREE模式，集群状态为读最终一致性。
- off表示非GTM-FREE模式。

**默认值：**off

## 17.21 开发人员选项

### enable\_light\_colupdate

**参数说明：**控制是否使用列存轻量化UPDATE。

**参数类型：**USERSET

**取值范围：**布尔型

- on表示开启列存轻量化UPDATE。
- off表示关闭列存轻量化UPDATE。

**默认值：**off

## enable\_fast\_query\_shipping

**参数说明：**控制查询优化器是否使用分布式框架。

**参数类型：**USERSET

**取值范围：**布尔型

- on表示执行计划在CN和DN上各自生成。
- off表示使用分布式框架，即执行计划在CN上生成，然后发送到DN中执行。

**默认值：**on

## enable\_trigger\_shipping

**参数说明：**控制触发器场景是否允许将触发器下推到DN执行。

**参数类型：**USERSET

**取值范围：**布尔型

- on表示允许将触发器下推到DN执行。
- off表示不允许将触发器下推到DN执行，在CN执行。

**默认值：**on

## enable\_remotejoin

**参数说明：**设置是否允许连接操作计划下推到DN执行。

**参数类型：**USERSET

**取值范围：**布尔型

- on表示允许连接操作计划下推到DN执行。
- off表示不允许连接操作计划下推到DN执行。

**默认值：**on

## enable\_remotegroup

**参数说明：**设置是否允许group by与aggregates执行计划下推到DN执行。

**参数类型：**USERSET

**取值范围：**布尔型

- on表示允许group by与aggregates执行计划下推到DN执行。
- off表示不允许group by与aggregates执行计划下推到DN执行。

**默认值：**on

## enable\_remotelimit

**参数说明：**设置是否允许LIMIT子句执行计划下推到DN执行。

**参数类型：**USERSET

**取值范围：**布尔型

- on表示允许LIMIT子句执行计划下推到DN执行。
- off表示不允许LIMIT子句执行计划下推到DN执行。

**默认值：**on

## enable\_remoresort

**参数说明：**设置是否允许ORDER BY子句操作计划下推到DN执行。

**参数类型：**USERSET

**取值范围：**布尔型

- on表示允许ORDER BY子句操作计划下推到DN执行。
- off表示不允许ORDER BY子句操作计划下推到DN执行。

**默认值：**on

## enable\_join\_pseudoconst

**参数说明：**设置是否允许与伪常数进行join。伪常数是指join两侧的变量都等于同一个常量。

**参数类型：**USERSET

**取值范围：**布尔型

- on表示允许与伪常数进行join。
- off表示不允许与伪常数进行join。

**默认值：**off

## cost\_model\_version

**参数说明：**控制应用场景中估算时cost使用的模型。该参数的影响范围主要涵盖：表达式distinct估算、HashJoin代价模型、行数估算、重分布时分布键的选择及Aggregate的行数估算等。

**参数类型：**USERSET

**取值范围：**0、1、2

- 0表示使用原始的cost估算模型。
- 1表示在0的基础上，使用增强的表达式distinct估算、HashJoin代价模型、行数估算、重分布时分布键的选择及Aggregate的行数估算。
- 2表示在1的基础上，使用随机性更优的analyze采样算法，以提高统计信息准确性。

**默认值：**1

## debug\_assertions

**参数说明：**控制打开各种断言检查。能够协助调试，当遇到奇怪的问题或者崩溃，请把此参数打开，因为它能暴露编程的错误。要使用这个参数，必须在编译

GaussDB(DWS)的时候定义宏USE\_ASSERT\_CHECKING（通过configure选项 --enable-cassert完成）。

**参数类型：**USERSET

**取值范围：**布尔型

- on表示打开断言检查。
- off表示不打开断言检查。

#### 说明

当启用断言选项编译GaussDB(DWS)时，debug\_assertions缺省值为on。

**默认值：**off

## distributed\_test\_param

**参数说明：**控制分布式测试框架打桩点是否生效。通常开发人员在故障注入测试时会在代码中预埋一些打桩点，使用唯一的名称进行标识，使用此参数可以控制代码中预埋的打桩点是否生效。参数采用逗号分隔的三元组形式，分别指定线程级别、测试桩名称和注入故障的错误级别。

**参数类型：**USERSET

**取值范围：**字符串，任一个已预埋的测试桩名称。

**默认值：**-1, default, default

## ignore\_checksum\_failure

**参数说明：**设置读取数据时是否忽略校验信息检查失败（但仍然会告警），继续执行。该参数仅在enable\_crc\_check为on时有效。继续执行可能导致崩溃，传播或隐藏损坏数据，无法从远程节点恢复数据及其他严重问题。不建议用户修改设置。

**参数类型：**SUSET

**取值范围：**布尔型

- on表示忽略数据校验错误。
- off表示数据校验错误正常报错。

**默认值：**off

## default\_orientation

**参数说明：**创建表时，当不指定存储方式时，根据该GUC参数的值创建对应类型的表。使用时，各节点值需要保持一致。

**参数类型：**SUSET

**取值范围：**row, column, column enabledelta

- row表示创建行存表。
- column表示创建列存表。
- column enabledelta表示创建开启delta表的列存表。

**默认值：**row

## enable\_colstore

**参数说明：**创建表时，当不指定存储方式时，默认创建为列存表。使用时，各节点值需要保持一致。属于测试参数，禁止用户启用。

**参数类型：**SUSET

**取值范围：**布尔型

**默认值：**off

## enable\_force\_vector\_engine

**参数说明：**对于支持向量化的执行器算子，如果其子节点是非向量化的算子，通过设置此参数为on，强制生成向量化的执行计划。当打开enable\_force\_vector\_engine开关时，无论是行存表、列存表或者是行列混存，如果plantee中不包含不支持向量化的场景，则强制走向量化执行引擎。

**参数类型：**USERSET

**取值范围：**布尔型

**默认值：**off

## enable\_csqual\_pushdown

**参数说明：**进行查询时，是否要将过滤条件下推，进行Rough Check。

**参数类型：**USERSET

**取值范围：**布尔型

- on表示进行查询时，要将过滤条件下推，进行Rough Check。
- off表示进行查询时，不要将过滤条件下推，进行Rough Check。

**默认值：**on

## explain\_dna\_file

**参数说明：**指定**explain\_perf\_mode**为run，导出的csv信息的目标文件。

**参数类型：**USERSET

---

### 须知

这个参数的取值必须是绝对路径加上.csv格式的文件名。

---

**取值范围：**字符串

**默认值：**NULL

## explain\_perf\_mode

**参数说明：**此参数用来指定explain的显示格式。

**参数类型：**USERSET

**取值范围:** normal、pretty、summary、run

- normal: 代表使用默认的打印格式。
- pretty: 代表使用GaussDB(DWS)改进后的新显示格式。新的格式层次清晰, 计划包含了plan node id, 性能分析简单直接。
- summary: 是在pretty的基础上增加了对打印信息的分析。
- run: 在summary的基础上, 将统计的信息输出到csv格式的文件中, 以便于进一步分析。

**默认值:** pretty

## join\_num\_distinct

**参数说明:** 控制应用场景中Join列或表达式的默认distinct值。

**参数类型:** USERSET

**取值范围:** 双精度浮点型, 大于或等于-100, 客户端显示小数时可能会有截断。

- 值大于0时, 表示使用该值作为默认distinct值。
- 值大于等于-100且小于0时, 表示估算默认distinct时使用的百分比。
- 值为0时, 表示使用200作为默认distinct值。

**默认值:** -20

## outer\_join\_max\_rows\_multipler

**参数说明:** 控制应用场景中外连接估算行数的最大值。

**参数类型:** USERSET

**取值范围:** 双精度浮点型, 0或大于等于1, 客户端显示小数时可能会有截断。

- 值为0时, 表示不限制outer join的估算行数。
- 值大于等于1时, 表示限制估算行数不超过外连接中外表行数的倍数。

**默认值:** -20

## qual\_num\_distinct

**参数说明:** 控制应用场景中过滤列或表达式的默认distinct值。

**参数类型:** USERSET

**取值范围:** 双精度浮点型, 大于或等于-100, 客户端显示小数时可能会有截断。

- 值大于0时, 表示使用该值作为默认distinct值。
- 值大于等于-100且小于0时, 表示估算默认distinct时使用的百分比。
- 值为0时, 表示使用200作为默认distinct值。

**默认值:** 200

## trace\_notify

**参数说明：**为LISTEN和NOTIFY命令生成大量调试输出。[client\\_min\\_messages](#)或[log\\_min\\_messages](#)级别必须是DEBUG1或者更低时，才能把这些输出分别发送到客户端或者服务器日志。

**参数类型：**USERSET

**取值范围：**布尔型

- on表示打开输出功能。
- off表示关闭输出功能。

**默认值：**off

## trace\_recovery\_messages

**参数说明：**启用恢复相关调试输出的日志录，否则将不会被记录。该参数允许覆盖正常设置的[log\\_min\\_messages](#)，但是仅限于特定的消息，这是为了在调试备机中使用。

**参数类型：**SIGHUP

**取值范围：**枚举类型，有效值有debug5、debug4、debug3、debug2、debug1、log，取值的详细信息请参见[log\\_min\\_messages](#)。

**默认值：**log

### 📖 说明

- 默认值log表示不影响记录决策。
- 除默认值外，其他值会导致优先级更高的恢复相关调试信息被记录，因为它们有log优先权。对于常见的log\_min\_messages设置，这会导致无条件地将它们记录到服务器日志上。

## trace\_sort

**参数说明：**控制是否在日志中打印排序操作中的资源使用相关信息。这个选项只有在编译GaussDB(DWS)的时候定义了TRACE\_SORT宏的时候才可用，不过目前TRACE\_SORT是由缺省定义的。

**参数类型：**USERSET

**取值范围：**布尔型

- on表示打开控制功能。
- off表示关闭控制功能。

**默认值：**off

## zero\_damaged\_pages

**参数说明：**控制检测导致GaussDB(DWS)报告错误的损坏的页头，中止当前事务。

**参数类型：**SUSET

**取值范围：**布尔型



- on表示打开控制功能。
- off表示关闭控制功能。

#### 📖 说明

- 设置为on时，系统报告一个警告，把损坏的页面填充为零然后继续处理。该行为会破坏数据，即被损坏页面上的所有行。但是它允许绕开损坏页面然后从表中存在的未损坏页面上继续检索数据行。因此该参数在硬件或者软件错误导致的数据损坏中进行恢复是有作用的。通常不建议该参数设置为on，除非不需要从损坏的页面中恢复数据。
- 对于列存表，会将整个CU跳过然后继续处理。支持的场景包括crc校验失败、magic校验失败以及读取的CU长度错误。

**默认值：** off

## string\_hash\_compatible

**参数说明：** 该参数用来说明char类型和varchar/text类型的hash值计算方式是否相同，以此来判断进行分布列从char类型到相同值的varchar/text类型转换，数据分布变化时，是否需要进行重分布。

**参数类型：** POSTMASTER

**取值范围：** 布尔型

- on表示计算方式相同，不需要进行重分布。
- off表示计算方式不同，需要进行重分布。

#### 📖 说明

计算方式的不同主要体现在字符串计算hash值时传入的字节长度上。（如果为char，则会忽略字符串后面空格的长度，如果为text或varchar，则会保留字符串后面空格的长度。）hash值的计算会影响到查询的计算结果，因此此参数一旦设置后，在整个数据库使用过程中不能再对其进行修改，以避免查询错误。

**默认值：** off

## replication\_test

**参数说明：** 此参数用于数据复制的内部功能测试。

**参数类型：** USERSET

**取值范围：** 布尔型

- on表示启用功能测试。
- off表示关闭功能测试。

**默认值：** off

## cost\_param

**参数说明：** 该参数用于控制在特定的客户场景中，使用不同的估算方法使得估算值与真实值更接近。此参数可以同时控制多种方法，与某一方法对应的位做与操作，不为0表示该方法被选择。

当cost\_param & 1 不为0，表示对于求不等值连接选择率时选择一种改良机制，此方法在自连接（两个相同的表之间连接）的估算中更加准确，V300R002C00版本开始，已弃用cost\_param & 1 不为0时的路径，默认选择更优的估算公式；

当cost\_param & 2 不为0，表示求多个过滤条件（Filter）的选择率时，选择最小的作为总的选择率，而非两者乘积，此方法在过滤条件的列之间关联性较强时估算更加准确；

当cost\_param & 4 不为0，表示在进行stream节点估算时，选用调试模型，该模型不推荐用户使用。

当cost\_param & 16不为0，表示在计算两个及以上过滤条件或Join条件的综合选择率时，采用介于完全相关和完全不相关之间的一种模型，过滤条件较多时倾向于相关性较强的模型。

**参数类型：**USERSET

**取值范围：**整型，1 ~ INT\_MAX

**默认值：**16

## convert\_string\_to\_digit

**参数说明：**设置隐式转换优先级，是否优先将字符串转为数字。

**参数类型：**USERSET

**取值范围：**布尔型

- on表示优先将字符串转为数字。
- off表示不优先将字符串转为数字。

**默认值：**on

---

### 须知

请谨慎调整该参数，调整该参数会修改内部数据类型转换规则并可能导致不可预期的行为。

---

## nls\_timestamp\_format

**参数说明：**设置时间戳默认格式。

**参数类型：**USERSET

**取值范围：**字符串

**默认值：**DD-Mon-YYYY HH:MI:SS.FF AM

## enable\_partitionwise

**参数说明：**分区表连接操作是否选择智能算法。

**参数类型：**USERSET

**取值范围：**布尔型

- on表示选择智能算法。
- off表示不选择智能算法。

**默认值：**off

## enable\_partition\_dynamic\_pruning

**参数说明：**分区表扫描是否支持动态剪枝。

**参数类型：**USERSET

**取值范围：**布尔型

- on表示分区表扫描开启动态剪枝。
- off表示分区表扫描关闭动态剪枝。

**默认值：**on

## max\_user\_defined\_exception

**参数说明：**异常最大个数，默认值不可更改。

**参数类型：**USERSET

**取值范围：**整型

**默认值：**1000

## datanode\_strong\_sync

**参数说明：**目前该参数已废弃，无实际意义。

**参数类型：**USERSET

**取值范围：**布尔型

- on表示启用stream节点间强同步。
- off表示不启用stream节点间强同步。

**默认值：**off

## enable\_debug\_vacuum

**参数说明：**允许输出一些与VACUUM相关的日志，便于定位VACUUM相关问题。开发人员专用，不建议普通用户使用。

**参数类型：**SIGHUP

**取值范围：**布尔型

- on/true表示开启此日志开关。
- off/false表示关闭此日志开关。

**默认值：**off

## enable\_global\_stats

**参数说明：**标识当前统计信息模式，为便于进行全局统计信息生成计划和单DN统计信息计划对比使用，默认创建为全局统计信息模式。属于测试参数，禁止用户启用。

**参数类型：**SUSET

**取值范围：**布尔型

- on/true表示全局统计信息。
- off/false表示单DN统计信息。

默认值: on

## enable\_fast\_numeric

**参数说明:** 标识是否开启Numeric类型数据运算优化。Numeric数据运算是较为耗时的操作之一，通过将Numeric转化为int64/int128类型，提高Numeric运算的性能。

**参数类型:** USERSET

**取值范围:** 布尔型

- on/true表示开启Numeric优化。
- off/false表示关闭Numeric优化。

默认值: on

## enable\_row\_fast\_numeric

**参数说明:** 标识行存表numeric数据落盘的格式。

**参数类型:** USERSET

**取值范围:** 布尔型

- on/true表示行存表numeric落盘格式为bigint格式。
- off/false表示行存表numeric落盘格式为原始格式。

### 须知

参数值设置为on时，建议同步打开enable\_force\_vector\_engine，可提升大数据集query的查询性能。但相比于原始格式，大概率会占用更多磁盘空间。以TPC-H测试集为例，大约多占7%空间（不同环境参考值可能有差异）。

默认值: off

## rewrite\_rule

**参数说明:** 标识开启的可选查询重写规则。有部分查询重写规则是可选的，开启它们并不能总是对查询效率有提升效果。在特定的客户场景中，通过此GUC参数对查询重写规则进行设置，使得查询效率最优。

此参数可以控制查询重写规则的组合，比如有多个重写规则：rule1、rule2、rule3、rule4。可以设置：

```
set rewrite_rule=rule1;      --启用查询重写规则rule1
set rewrite_rule=rule2,rule3; --启用查询重写规则rule2和rule3
set rewrite_rule=none;      --关闭所有可选查询重写规则
```

**参数类型:** USERSET

**取值范围:** 字符串

- none: 不使用任何可选查询重写规则
- lazyagg: 使用Lazy Agg查询重写规则（消除子查询中的聚集运算）
- magicset: 使用Magic Set查询重写规则（从主查询中下推条件到提升的子链接）
- uniquecheck: 使用Unique Check重写规则（允许目标列不含聚集函数的表达式子链接场景提升，需在子链接按关联列聚集后目标列值唯一才能开启，建议专业调优人员使用）。
- disablerep: 使用禁止复制表的子链接提升规则（针对复制表禁止子链接提升）。
- projection\_pushdown: 使用Projection Pushdown重写规则（子查询中消除父查询不使用的列）。
- or\_conversion: 使用OR转换重写规则（消除执行效率低下的关联OR条件）。
- plain\_lazyagg: 使用Plain Lazy Agg查询重写规则（消除单子查询中的聚集操作）。该选项仅8.1.3.100及以上集群版本支持。
- eager\_magicset: 使用eager\_magicset查询重写规则（从主查询中下推条件到子查询）。该选项仅8.2.0及以上集群版本支持。

**默认值:** magicset, or\_conversion, projection\_pushdown, plain\_lazyagg

## enable\_compress\_spill

**参数说明:** 标识是否开启下盘压缩功能。

**参数类型:** USERSET

**取值范围:** 布尔型

- on/true表示开启下盘优化。
- off/false表示关闭下盘优化。

**默认值:** on

## analysis\_options

**参数说明:** 通过开启对应选项中所对应的功能选项使用相应的定位功能，包括数据校验，性能统计等，参见取值范围中的选项说明。

**参数类型:** USERSET

**取值范围:** 字符串

- LLVM\_COMPILE表示在explain performance显示界面中显示每个线程的codegen编译时间。
- HASH\_CONFLICT表示在DN进程的pg\_log目录中的log日志中显示hash表的统计信息，包括hash表大小，hash链长，hash冲突情况。
- STREAM\_DATA\_CHECK表示对网络传输前后的数据进行CRC校验。

**默认值:** off(ALL)，不开启任何定位功能。

## resource\_track\_log

**参数说明:** 控制自诊断的日志级别。目前仅对多列统计信息进行控制。

**参数类型:** USERSET

**取值范围：**字符串

- summary：显示简略的诊断信息。
- detail：显示详细的诊断信息。

目前这两个参数值只在显示多列统计信息未收集的告警的情况下有差别，summary不显示未收集多列统计信息的告警，detail会显示这类告警。

**默认值：**summary

## hll\_default\_log2m

**参数说明：**该参数可以指定hll数据结构桶的个数。桶的个数会影响hll计算distinct值的精度，桶的个数越多，误差越小。误差范围为： $[-1.04/2^{\log_2 m^{1/2}}, +1.04/2^{\log_2 m^{1/2}}]$ 。

**参数类型：**USERSET

**取值范围：**整型，10~16。

**默认值：**11

## hll\_default\_regwidth

**参数说明：**该参数可以指定hll数据结构每个桶的位数，该值越大，hll所占内存越高。hll\_default\_regwidth和hll\_default\_log2m可以决定当前hll能够计算的最大distinct value。具体对应关系可以参见[表1 hll\\_default\\_log2m和hll\\_default\\_regwidth与当前能计算的最大distinct value值的关系](#)。

**参数类型：**USERSET

**取值范围：**整型，1~5。

**默认值：**5

**表 17-2** hll\_default\_log2m 和 hll\_default\_regwidth 与当前能计算的最大 distinct value 值的关系

log2m	regwidth = 1	regwidth = 2	regwidth = 3	regwidth = 4	regwidth = 5
10	7.4e+02	3.0e+03	4.7e+04	1.2e+07	7.9e+11
11	1.5e+03	5.9e+03	9.5e+04	2.4e+07	1.6e+12
12	3.0e+03	1.2e+04	1.9e+05	4.8e+07	3.2e+12
13	5.9e+03	2.4e+04	3.8e+05	9.7e+07	6.3e+12
14	1.2e+04	4.7e+04	7.6e+05	1.9e+08	1.3e+13
15	2.4e+04	9.5e+04	1.5e+06	3.9e+08	2.5e+13

## hll\_default\_expthresh

**参数说明：**该参数可以用来设置从Explicit模式到Sparse模式的默认阈值大小。

**参数类型：**USERSET

**取值范围：** 整型，-1~7。-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 $2^{\text{hll\_default\_expthresh}}$ 时切换模式。

**默认值：** -1

## hll\_default\_sparseon

**参数说明：** 该参数可用来指定是否默认开启Sparse模式。

**参数类型：** USERSET

**取值范围：** 0，1。0表示默认关闭，1表示默认开启。

**默认值：** 1

## hll\_max\_sparse

**参数说明：** 该参数可以用来指定max\_sparse的大小。

**参数类型：** USERSET

**取值范围：** 整型，-1~INT\_MAX

**默认值：** -1

## enable\_compress\_hll

**参数说明：** 该参数可以用来指定是否对hll开启内存优化模式。

**参数类型：** USERSET

**取值范围：** 布尔型

- on/true表示对hll开启内存优化模式。
- off/false表示不开启内存优化模式。

**默认值：** off

## udf\_memory\_limit

**参数说明：** 控制每个CN、DN执行UDF时可用的最大物理内存量。

**参数类型：** POSTMASTER

**取值范围：** 整型， $200 \times 1024 \sim \text{max\_process\_memory}$ ，单位为KB。

**默认值：**  $0.05 * \text{max\_process\_memory}$

## FencedUDFMemoryLimit

**参数说明：** 控制每个fenced udf worker进程使用的虚拟内存。

**参数类型：** USERSET

**设置建议：** 不建议设置此参数，可用`udf_memory_limit`代替。

**取值范围：** 整数，可带单位（KB，MB，GB）。其中0表示不做内存控制。

**默认值：** 0

## UDFWorkerMemHardLimit

**参数说明：**控制fencedUDFMemoryLimit的最大值。

**参数类型：**POSTMASTER

**设置建议：**不建议设置此参数，可用**udf\_memory\_limit**代替。

**取值范围：**整数，可带单位（KB，MB，GB）。

**默认值：**1GB

## pljava\_vmoptions

**参数说明：**用户自定义设置PL/Java函数所使用的JVM虚拟机的启动参数。

**参数类型：**SUSET

**取值范围：**字符串，支持：

- JDK8 JVM启动参数
- JDK8 JVM系统属性参数（以-D开头如-Djava.ext.dirs）
- 用户自定义参数（以-D开头，如-Duser.defined.option）

---

### 须知

如果用户在pljava\_vmoptions中设置参数不满足上述取值范围，会在使用PL/Java语言函数时报错。

---

**默认值：**空

## javaudf\_disable\_feature

**参数说明：**该参数用于控制javaudf行为的细粒度。

**参数类型：**SIGHUP

**取值范围：**字符串

- none，不禁用其他细粒度参数中指定的任何行为，当和其他参数一起设置时，none失效。
- all，禁止执行所有javaudf函数，该选项设置后优先级最高。
- extdir，使在第三方路径放置依赖jar包的功能失效。
- hadoop，使hadoop相关功能失效。
- reflection，javaudf函数执行过程中禁用反射(ReflectPermission权限)。
- loadlibrary，javaudf函数执行过程中禁止加载动态库(loadLibrary权限)。
- net，javaudf函数执行过程中禁用网络权限(NetPermission权限)。
- socket，javaudf函数执行过程中禁用socket套接字(SocketPermission权限)。
- security，javaudf函数执行过程中禁止Security配置修改(SecurityPermission权限)。
- classloader，javaudf函数执行过程中禁止自定义classLoder(createClassLoader权限)。



- `access_declared_members`, `javaudf`函数执行过程中禁止读取其他类的内部成员 (`accessDeclaredMembers`权限)。

**默认值:**

`extdir,hadoop,reflection,loadlibrary,net,socket,security,classloader,access_declared_members`

## `enable_pbe_optimization`

**参数说明:** 设置优化器是否对以PBE ( Parse Bind Execute ) 形式执行的语句进行查询计划的优化。

**参数类型:** USERSET

**取值范围:** 布尔型。

- `on`表示优化器将优化PBE语句的查询计划。
- `off`表示不使用优化。

**默认值:** `on`

## `enable_light_proxy`

**参数说明:** 设置优化器是否对简单查询在CN上优化执行。

**参数类型:** USERSET

**取值范围:** 布尔型。

- `on`表示优化器将优化CN上简单查询的执行。
- `off`表示不使用优化。

**默认值:** `on`

## `checkpoint_flush_after`

**参数说明:** 设置`checkpointer`线程在连续写多少个磁盘页后会进行异步刷盘操作。GaussDB(DWS)中, 磁盘页大小为8KB。

**参数类型:** SIGHUP

**取值范围:** 整型, 0~256 ( 0表示关闭异步刷盘功能 )。例如, 取值32, 表示`checkpointer`线程连续写32个磁盘页, 即 $32*8=256$ KB磁盘空间后会进行异步刷盘。

**默认值:** 32

## `enable_parallel_ddl`

**参数说明:** 控制多CN对同一数据库对象是否能安全的并发执行DDL操作。

**参数类型:** USERSET

**取值范围:** 布尔型

- `on`表示可以安全的并发执行DDL操作, 不会出现分布式死锁。
- `off`表示不能安全的并发执行DDL操作, 可能会出现分布式死锁。

**默认值:** `on`

## gc\_fdw\_verify\_option

**参数说明：**在协同分析特性中，控制是否启用结果集行数校验逻辑。该参数仅8.1.3.310及以上集群版本支持。

**参数类型：**USERSET

**取值范围：**布尔型

- on表示开启结果集行数校验逻辑，额外发送select count语句获取预期行数，与实际获取结果集进行比对。
- off表示关闭结果集行数校验逻辑，仅获取所需结果集。

**默认值：**on

### 说明

- 该参数开启时性能会有轻微劣化，性能敏感场景可通过关闭该参数提升性能。
- 结果集行数校验失败时会抛出异常，可通过设置参数log\_min\_messages=debug1和logging\_module='on(COOP\_ANALYZE)'来打开协同分析日志。

## show\_acce\_estimate\_detail

**参数说明：**在GaussDB(DWS)集群使用加速集群场景下（即[acceleration\\_with\\_compute\\_pool](#)设置为on），控制explain命令是否显示用于评估执行计划下推到加速集群的评估信息。评估信息一般用于运维人员在维护工作中使用，因此该参数默认关闭，此外为了避免这些信息干扰正常的explain信息显示，只有在explain命令的verbose选项打开的情况下才显示评估信息。

**参数类型：**USERSET

**取值范围：**布尔型

- on表示可以在explain命令的输出中显示评估信息。
- off表示不在explain命令的输出中显示评估信息。

**默认值：**off

## support\_batch\_bind

**参数说明：**控制是否允许通过JDBC、ODBC、Libpq等接口批量绑定和执行PBE形式的语句。

**参数类型：**SIGHUP

**取值范围：**布尔型

- on表示使用批量绑定和执行。
- off表示不使用批量绑定和执行。

**默认值：**on

## enable\_immediate\_interrupt

**参数说明：**控制是否允许在信号处理函数中立即中断当前语句或会话的执行。

**参数类型：**SIGHUP

**取值范围：**布尔型

- on表示允许信号处理函数中立即中断语句或会话的执行。
- off表示不允许信号处理函数中立即中断语句或会话的执行。

**默认值：**off

#### 说明

请谨慎开启为on，因为允许在信号函数中立即中断语句或会话的执行可能会导致某些关键流程执行被中断，且导致系统内部全局锁无法释放。建议该参数仅在系统调试过程中或规避故障时临时开启。

## 17.22 审计

### 17.22.1 审计开关

#### audit\_enabled

**参数说明：**控制审计进程的开启和关闭。审计进程开启后，将从管道读取后台进程写入的审计信息，并写入审计文件。

**参数类型：**SIGHUP

**取值范围：**布尔型

- on表示启动审计功能。
- off表示关闭审计功能。

**默认值：**on

#### audit\_data\_format

**参数说明：**审计日志文件的格式。当前仅支持二进制格式。

**参数类型：**POSTMASTER

**取值范围：**字符串

**默认值：**binary

#### audit\_rotation\_interval

**参数说明：**指定创建一个新审计日志文件的时间间隔。当现在的时间减去上次创建一个审计日志的时间超过了此参数值时，服务器将生成一个新的审计日志文件。

**参数类型：**SIGHUP

**取值范围：**整型，1~INT\_MAX/60，单位为min。

**默认值：**1d

### 须知

请不要随意调整此参数，否则可能会导致[audit\\_resource\\_policy](#)无法生效，如果需要控制审计日志的存储空间和时间，请使用[audit\\_resource\\_policy](#)、[audit\\_space\\_limit](#)和[audit\\_file\\_remain\\_time](#)参数进行控制。

## audit\_rotation\_size

**参数说明：**指定审计日志文件的最大容量。当审计日志消息的总量超过此参数值时，服务器将生成一个新的审计日志文件。

**参数类型：**SIGHUP

**取值范围：**整型，1~1024，单位为MB。

**默认值：**10MB

### 须知

请不要随意调整此参数，否则可能会导致[audit\\_resource\\_policy](#)无法生效，如果需要控制审计日志的存储空间和时间，请使用[audit\\_resource\\_policy](#)、[audit\\_space\\_limit](#)和[audit\\_file\\_remain\\_time](#)参数进行控制。

## audit\_resource\_policy

**参数说明：**控制审计日志的保存策略，以空间还是时间限制为优先策略。

**参数类型：**SIGHUP

**取值范围：**布尔型

- on表示采用空间优先策略，最多存储[audit\\_space\\_limit](#)大小的日志。
- off表示采用时间优先策略，最少存储[audit\\_file\\_remain\\_time](#)长度时间的日志。如果[audit\\_file\\_remain\\_time](#)设置过大，导致存储的审计日志所占磁盘空间达到[audit\\_space\\_limit](#)，此时将删除最早的审计文件。

**默认值：**on

## audit\_file\_remain\_time

**参数说明：**表示需记录审计日志的最短时间要求，该参数在[audit\\_resource\\_policy](#)为off时生效。

**参数类型：**SIGHUP

**取值范围：**整型，0~730，单位为day，0表示无时间限制。

**默认值：**90

## audit\_space\_limit

**参数说明：**审计文件占用的磁盘空间总量。

**参数类型：**SIGHUP

**取值范围：**整型，1024KB~1024GB，单位为KB。

**默认值：**1GB

## audit\_object\_name\_format

**参数说明：**控制审计日志object\_name字段所显示的对象名的格式。

**参数类型：**USERSET

**取值范围：**枚举类型

- single，表示object\_name字段显示单个对象名，且为目标对象的名称。
- all，表示object\_name字段显示多个对象名称。

**默认值：**single

### 📖 说明

当默认值设置为all时，显示多个对象名的场景有：SELECT、DELETE、UPDATE、INSERT、MERGE、CREATE TABLE AS、CREATE VIEW AS。

## audit\_file\_remain\_threshold

**参数说明：**审计目录下审计文件个数的最大值。

**参数类型：**SIGHUP

**取值范围：**整型，1~1048576

**默认值：**1048576

### 须知

请尽量确保此参数取值为1048576，不要随意调整此参数，否则可能会导致audit\_resource\_policy无法生效。如果需要控制审计日志的存储空间和时间，请使用audit\_resource\_policy、audit\_space\_limit和audit\_file\_remain\_time参数进行控制。

## 17.22.2 操作审计

### audit\_operation\_exec

**参数说明：**该参数决定是否审计GaussDB(DWS)中各类执行成功的操作，由用户根据实际需求进行配置。

**参数类型：**SIGHUP

**取值范围：**字符串

- none：表示未配置审计项，如果同时配置了其他任何审计项，则none失效。
- all：表示对所有操作成功的场景进行审计。如果同时配置了其他任何审计项，则覆盖所有其他审计项的配置。注意，即使配置为all，也不表示对所有的DDL操作进行审计，仍然需要结合[audit\\_system\\_object](#)，对DDL操作的对象级别进行控制。

- login: 表示对用户登录成功的场景进行审计。
- logout: 表示对用户退出进行审计。
- database\_process: 表示对数据库启动、停止、切换、恢复操作进行审计。
- user\_lock: 表示对用户锁定和解锁成功的场景进行审计。
- grant\_revoke: 表示对用户权限授予和回收成功的场景进行审计。
- ddl: 表示对DDL操作成功的场景进行审计，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）。
- select: 表示对select操作成功的场景进行审计。
- copy: 表示对copy操作成功的场景进行审计。
- userfunc: 表示对用户自定义函数、存储过程、匿名块操作成功的场景进行审计。
- set: 表示对set操作成功的场景进行审计。
- transaction: 表示对事务操作成功的场景进行审计。
- vacuum: 表示对vacuum操作成功的场景进行审计。
- analyze: 表示对analyze操作成功的场景进行审计。
- explain: 表示对explain操作成功的场景进行审计。
- specialfunc: 表示对特殊函数调用操作成功的场景进行审计，特殊函数包括：pg\_terminate\_backend、pg\_cancel\_backend。
- insert: 表示对insert操作成功的场景进行审计。
- update: 表示对update操作成功的场景进行审计。
- delete: 表示对delete操作成功的场景进行审计。
- merge: 表示对merge操作成功的场景进行审计。
- show: 表示对show操作成功的场景进行审计。
- checkpoint: 表示对checkpoint操作成功的场景进行审计。
- barrier: 表示对barrier操作成功的场景进行审计。
- cluster: 表示对cluster操作成功的场景进行审计。
- comment: 表示对comment操作成功的场景进行审计。
- cleanconn: 表示对cleanconnection操作成功的场景进行审计。
- prepare: 表示对PREPARE、EXECUTE、DEALLOCATE操作成功的场景进行审计。
- constraints: 表示对constraints操作成功的场景进行审计。
- cursor: 表示对游标操作成功的场景进行审计。
- discard: 表示对清理当前会话的全局临时表信息操作成功的场景进行审计。

**默认值:** login, logout, database\_process, user\_lock, grant\_revoke, set, transaction, cursor

**须知**

- 建议transaction审计项保留，否则事务内语句都不会被审计。
- 建议cursor审计项保留，否则cursor内select语句不会被审计。
- 需注意Data Studio客户端会自动给select语句封装cursor。

**audit\_operation\_error**

**参数说明：**该参数决定是否审计GaussDB(DWS)中各类执行失败的操作，由用户根据实际需求进行配置。

**参数类型：**SIGHUP

**取值范围：**字符串

- none：表示未配置审计项，如果同时配置了其他任何审计项，则none失效。
- syn\_success：表示同步[audit\\_operation\\_exec](#)的配置，即配置了某操作执行成功场景的审计，则对应的执行失败场景也记入审计。需注意，配置了syn\_success后，仍可以继续配置其他操作执行失败场景的审计；如果audit\_operation\_exec配置为all，则所有的失败场景均记入审计；如果audit\_operation\_exec配置为none，则syn\_success等同于none，即未配置审计项。
- parse：表示对用户输入命令解析失败场景进行审计，包含等待命令超时的失败场景。
- login：表示对用户登录失败的场景进行审计。
- user\_lock：表示对用户锁定和解锁失败的场景进行审计。
- violation：表示对用户访问存在越权的场景进行审计。
- grant\_revoke：表示对用户权限授予和回收失败的场景进行审计。
- ddl：表示对DDL操作失败的场景进行审计，因为DDL操作由会根据操作对象进行更细粒度控制，仍然需要结合[audit\\_system\\_object](#)的配置情况，所以此处配置ddl后，将对[audit\\_system\\_object](#)指定类型的DDL失败场景进行审计。
- select：表示对SELECT操作失败的场景进行审计。
- copy：表示对COPY操作失败的场景进行审计。
- userfunc：表示对用户自定义函数、存储过程、匿名块操作失败的场景进行审计。
- set：表示对set操作失败的场景进行审计。
- transaction：表示对事务操作失败的场景进行审计。
- vacuum：表示对VACUUM操作失败的场景进行审计。
- analyze：表示对ANALYZE操作失败的场景进行审计。
- explain：表示对EXPLAIN操作失败的场景进行审计。
- specialfunc：表示对特殊函数调用操作失败的场景进行审计，特殊函数包括：pg\_terminate\_backend、pg\_cancel\_backend。
- insert：表示对INSERT操作失败的场景进行审计。
- update：表示对UPDATE操作失败的场景进行审计。
- delete：表示对DELETE操作失败的场景进行审计。

- merge: 表示对MERGE操作失败的场景进行审计。
- show: 表示对SHOW操作失败的场景进行审计。
- checkpoint: 表示对CHECKPOINT操作失败的场景进行审计。
- barrier: 表示对BARRIER操作失败的场景进行审计。
- cluster: 表示对CLUSTER操作失败的场景进行审计。
- comment: 表示对COMMENT操作失败的场景进行审计。
- cleanconn: 表示对CLEANCONNECTION操作失败的场景进行审计。
- prepare: 表示对PREPARE、EXECUTE、DEALLOCATE操作失败的场景进行审计。
- constraints: 表示对CONSTRAINTS操作失败的场景进行审计。
- cursor: 表示对游标操作失败的场景进行审计。
- blacklist: 表示对黑名单执行失败进行审计。
- discard: 表示对清理当前会话的全局临时表信息执行失败的场景进行审计。

**默认值:** login

## audit\_inner\_tool

**参数说明:** 该参数决定是否审计GaussDB(DWS)中内部维护工具的各类操作。

**参数类型:** SIGHUP

**取值范围:** 布尔型

- on表示审计来自内部维护工具的各类操作。
- off表示不审计来自内部维护工具的各类操作。

**默认值:** off

## audit\_system\_object

**参数说明:** 该参数决定是否对GaussDB(DWS)数据库对象的CREATE、DROP、ALTER操作进行审计。GaussDB(DWS)数据库对象包括DATABASE、USER、schema、TABLE等。通过修改该配置参数的值，可以只审计需要的数据库对象的操作。

**参数类型:** SIGHUP

**取值范围:** 整型，0~4194303

- 0代表关闭GaussDB(DWS)数据库对象的CREATE、DROP、ALTER操作审计功能。
- 非0代表只审计GaussDB(DWS)的某类或者某些数据库对象的CREATE、DROP、ALTER操作。

**取值说明:**

该参数的值由22个二进制位的组合求出，这22个二进制位分别代表GaussDB(DWS)的22类数据库对象。如果对应的二进制位取值为0，表示不审计对应的数据库对象的CREATE、DROP、ALTER操作；取值为1，表示审计对应的数据库对象的CREATE、DROP、ALTER操作。这22个二进制位代表的具体审计内容请参见[表17-3](#)。

**默认值:** 12303



表 17-3 audit\_system\_object 取值含义说明

二进制位	含义	取值说明
第0位	是否审计DATABASE对象的CREATE、DROP、ALTER操作。	<ul style="list-style-type: none"> <li>0表示不审计该对象的CREATE、DROP、ALTER操作；</li> <li>1表示审计该对象的CREATE、DROP、ALTER操作。</li> </ul>
第1位	是否审计SCHEMA对象的CREATE、DROP、ALTER操作。	<ul style="list-style-type: none"> <li>0表示不审计该对象的CREATE、DROP、ALTER操作；</li> <li>1表示审计该对象的CREATE、DROP、ALTER操作。</li> </ul>
第2位	是否审计USER对象的CREATE、DROP、ALTER操作。	<ul style="list-style-type: none"> <li>0表示不审计该对象的CREATE、DROP、ALTER操作；</li> <li>1表示审计该对象的CREATE、DROP、ALTER操作。</li> </ul>
第3位	是否审计TABLE对象的CREATE、DROP、ALTER、TRUNCATE操作。	<ul style="list-style-type: none"> <li>0表示不审计该对象的CREATE、DROP、ALTER、TRUNCATE操作；</li> <li>1表示审计该对象的CREATE、DROP、ALTER、TRUNCATE操作。</li> </ul>
第4位	是否审计INDEX对象的CREATE、DROP、ALTER操作。	<ul style="list-style-type: none"> <li>0表示不审计该对象的CREATE、DROP、ALTER操作；</li> <li>1表示审计该对象的CREATE、DROP、ALTER操作。</li> </ul>
第5位	是否审计VIEW对象的CREATE、DROP、ALTER操作。	<ul style="list-style-type: none"> <li>0表示不审计该对象的CREATE、DROP操作；</li> <li>1表示审计该对象的CREATE、DROP操作。</li> </ul>
第6位	是否审计TRIGGER对象的CREATE、DROP、ALTER操作。	<ul style="list-style-type: none"> <li>0表示不审计该对象的CREATE、DROP、ALTER操作；</li> <li>1表示审计该对象的CREATE、DROP、ALTER操作。</li> </ul>
第7位	是否审计PROCEDURE/FUNCTION对象的CREATE、DROP、ALTER操作。	<ul style="list-style-type: none"> <li>0表示不审计该对象的CREATE、DROP、ALTER操作；</li> <li>1表示审计该对象的CREATE、DROP、ALTER操作。</li> </ul>
第8位	是否审计TABLESPACE对象的CREATE、DROP、ALTER操作。	<ul style="list-style-type: none"> <li>0表示不审计该对象的CREATE、DROP、ALTER操作；</li> <li>1表示审计该对象的CREATE、DROP、ALTER操作。</li> </ul>

二进制位	含义	取值说明
第9位	是否审计RESOURCE POOL对象的CREATE、DROP、ALTER操作。	<ul style="list-style-type: none"> <li>0表示不审计该对象的CREATE、DROP、ALTER操作；</li> <li>1表示审计该对象的CREATE、DROP、ALTER操作</li> </ul>
第10位	是否审计WORKLOAD对象的CREATE、DROP、ALTER操作。	<ul style="list-style-type: none"> <li>0表示不审计该对象的CREATE、DROP、ALTER操作；</li> <li>1表示审计该对象的CREATE、DROP、ALTER操作</li> </ul>
第11位	是否审计SERVER FOR HADOOP对象的CREATE、DROP、ALTER操作。	<ul style="list-style-type: none"> <li>0表示不审计该对象的CREATE、DROP、ALTER操作；</li> <li>1表示审计该对象的CREATE、DROP、ALTER操作</li> </ul>
第12位	是否审计DATA SOURCE对象的CREATE、DROP、ALTER操作。	<ul style="list-style-type: none"> <li>0表示不审计该对象的CREATE、DROP、ALTER操作；</li> <li>1表示审计该对象的CREATE、DROP、ALTER操作。</li> </ul>
第13位	是否审计NODE GROUP对象的CREATE、DROP、ALTER操作。	<ul style="list-style-type: none"> <li>0表示不审计该对象的CREATE、DROP操作；</li> <li>1表示审计该对象的CREATE、DROP操作。</li> </ul>
第14位	是否审计ROW LEVEL SECURITY对象的CREATE、DROP、ALTER操作。	<ul style="list-style-type: none"> <li>0表示不审计该对象的CREATE、DROP、ALTER操作；</li> <li>1表示审计该对象的CREATE、DROP、ALTER操作。</li> </ul>
第15位	是否审计TYPE对象的CREATE、DROP、ALTER操作。	<ul style="list-style-type: none"> <li>0表示不审计TYPE对象的CREATE、DROP、ALTER操作；</li> <li>1表示审计TYPE对象的CREATE、DROP、ALTER操作。</li> </ul>
第16位	是否审计TEXT SEARCH对象（CONFIGURATION和DICTIONARY）的CREATE、DROP、ALTER操作。	<ul style="list-style-type: none"> <li>0表示不审计TEXT SEARCH对象的CREATE、DROP、ALTER操作；</li> <li>1表示审计TEXT SEARCH对象的CREATE、DROP、ALTER操作。</li> </ul>
第17位	是否审计DIRECTORY对象的CREATE、DROP、ALTER操作。	<ul style="list-style-type: none"> <li>0表示不审计DIRECTORY对象的CREATE、DROP、ALTER操作；</li> <li>1表示审计DIRECTORY对象的CREATE、DROP、ALTER操作。</li> </ul>

二进制位	含义	取值说明
第18位	是否审计SYNONYM对象的CREATE、DROP、ALTER操作。	<ul style="list-style-type: none"> <li>0表示不审计SYNONYM对象的CREATE、DROP、ALTER操作；</li> <li>1表示审计SYNONYM对象的CREATE、DROP、ALTER操作。</li> </ul>
第19位	是否审计REDACTION POLICY对象的CREATE、DROP、ALTER操作。	<ul style="list-style-type: none"> <li>0表示不审计REDACTION POLICY对象的CREATE、DROP、ALTER操作；</li> <li>1表示审计REDACTION POLICY对象的CREATE、DROP、ALTER操作。</li> </ul>
第20位	是否审计SEQUENCE对象的CREATE、DROP、ALTER操作。	<ul style="list-style-type: none"> <li>0表示不审计SEQUENCE对象的CREATE、DROP、ALTER操作；</li> <li>1表示审计SEQUENCE对象的CREATE、DROP、ALTER操作。</li> </ul>
第21位	是否审计NODE对象的CREATE、DROP、ALTER操作。	<ul style="list-style-type: none"> <li>0表示不审计NODE对象的CREATE、DROP、ALTER操作；</li> <li>1表示审计NODE对象的CREATE、DROP、ALTER操作。</li> </ul>

## enableSeparationOfDuty

**参数说明：**是否开启三权分立选项。

**参数类型：**POSTMASTER

**取值范围：**布尔型

- on表示开启三权分立。
- off表示不开启三权分立。

**默认值：**off

## security\_enable\_options

**参数说明：**该参数决定是否允许安全模式下使用grant\_to\_public、grant\_with\_grant\_option和foreign\_table\_options三种功能，可根据实际需求进行配置。（该参数仅8.2.0及以上集群版本支持）

**参数类型：**SIGHUP

**取值范围：**字符串

- grant\_to\_public，表示允许安全模式下使用grant to public功能。
- grant\_with\_grant\_option，表示允许安全模式下使用with grant option功能。
- foreign\_table\_options，表示允许安全模式下使用外表操作功能，不需要显式赋予用户useft权限。

**默认值：**空

#### 📖 说明

- 新安装集群场景下，该参数默认值为空，表示安全模式下不允许使用grant\_to\_public、grant\_with\_grant\_option和foreign\_table\_options中任何一种功能。
- 升级场景下，该参数的默认值保持前向兼容，若用户升级前原版本中GUC参数enable\_grant\_public和enable\_grant\_option默认为ON，那么升级后security\_enable\_options参数的默认值为“grant\_to\_public, grant\_with\_grant\_option”。

## enable\_copy\_server\_files

**参数说明：**是否开启copy服务器端文件的权限。

**参数类型：**POSTMASTER

**取值范围：**布尔型

- on表示开启copy服务端文件的权限。
- off表示不开启copy服务端文件的权限。

**默认值：**true

#### 须知

copy from/to file要求具有系统管理员权限的用户才能使用，但是，在三权分立开启的状态下，系统管理员与初始用户的权限不同，可以通过使用enable\_copy\_server\_file控制系统管理员的copy权限，避免系统管理员权限升级。

## 17.23 事务监控

通过设置事务超时预警，可以监控自动回滚的事务并定位其中的语句问题，并且也可以监控执行时间过长的语句。

## transaction\_sync\_naptime

**参数说明：**为保证数据一致性，当本地事务与GTM上snapshot中状态不一样时会阻塞其他事务的运行，需要等待本地节点上事务状态与GTM状态一致后再运行。当CN上等待时长超过transaction\_sync\_naptime时会主动触发gs\_clean进行清理，缩短不一致时的阻塞时长。

**参数类型：**USERSET

**取值范围：**整型，最小值为0，单位为秒。

**默认值：**5s

#### 📖 说明

若该值设为0，则不会在阻塞达到时长时主动调用gs\_clean进行清理，而是靠gs\_clean\_timeout间隔来调用gs\_clean，默认是5分钟。

## transaction\_sync\_timeout

**参数说明：**为保证数据一致性，当本地事务与GTM上snapshot中状态不一样时会阻塞其他事务的运行，需要等待本地节点上事务状态与GTM状态一致后再运行。当CN上等待时长超过transaction\_sync\_timeout时会报错，回滚事务，避免由于sync lock等其他情况长时间进程停止响应造成对系统的阻塞。

**参数类型：**USERSET

**取值范围：**整型，最小值为0，单位为秒。

**默认值：**10min

### 📖 说明

- 若该值设为0，则不会在阻塞超时报错，回滚事务。
- 该值必须大于gs\_clean\_timeout，避免DN上由于还未被gs\_clean清理的残留事务阻塞超时引起的不必要的事务回滚。

## 17.24 GTM 相关参数

### log\_min\_messages

**参数说明：**控制写到服务器日志文件中的消息级别。每个级别都包含排在它后面的所有级别中的信息。级别越低，服务器运行日志中记录的消息就越少。

#### 须知

当client\_min\_messages和log\_min\_messages取值相同时，其值所代表的级别不同。

**参数类型：**SUSET

**取值范围：**枚举类型，有效值有debug、debug5、debug4、debug3、debug2、debug1、info、log、notice、warning、error、fatal、panic。参数的详细信息请参见表17-1。

**默认值：**warning

### enable\_alarm

**参数说明：**是否允许打开告警检测线程，检测数据库中可能的错误场景。

**参数类型：**POSTMASTER

**取值范围：**布尔型

- on表示允许打开告警检测线程。
- off表示不允许打开告警检测线程。

**默认值：**on

## 17.25 其它选项

### enable\_cluster\_resize

**参数说明：**对于sql语句中涉及多个表，并且属于不同group，打开此开关可以支持此语句执行计划下推来提高性能。

**参数类型：**SUSET

**取值范围：**布尔型

- on表示支持此语句执行计划下推来提高性能。
- off表示不支持此语句执行计划下推来提高性能。

**默认值：**off

#### 说明

此参数用于内部运维场景，请勿随意开启。

### dfs\_partition\_directory\_length

**参数说明：**在HDFS文件系统中，构造HDFS VALUE分区表的分区目录时，目录名长度的上限值。

**参数类型：**USERSET

**取值范围：**92 ~ 7999

**默认值：**512

### enable\_hadoop\_env

**参数说明：**设置使用Hadoop特性时，是否允许在数据库中创建本地行存表和列存表。GaussDB(DWS)集群中，集群安装好后，该参数默认设为off。以支持本地行列存储和跨集群访问Hadoop特性。不推荐用户调整enable\_hadoop\_env的值。

**参数类型：**USERSET

**取值范围：**布尔型

- on/true，表示使用Hadoop特性时，不允许在数据库中创建本地行存表和列存表。
- off/false，表示使用Hadoop特性时，可以在数据库中创建本地行存表和列存表。

**默认值：**off

### remote\_read\_mode

**参数说明：**设置当开启enable\_crc\_check为on，主DN读取的数据校验失败后是否进行远程读的开关，以及是否采用安全认证方式连接。设置后需要重启集群才能生效。

**参数类型：**POSTMASTER

**取值范围：**off, non\_authentication, authentication

- off, 表示关闭远程读功能。
- non\_authentication, 表示采用非认证的方式连接备DN并获取数据。
- authentication, 表示采用认证方式连接备DN并获取数据, 重启集群前在\$GAUSSHOME/share/sslcert/grpc/目录下必须存在证书, 否则无法启动集群。

**默认值:** non\_authentication

## enable\_upgrade\_merge\_lock\_mode

**参数说明:** 当该参数设置为on时, 通过提升deltamerge内部实现的锁级别, 避免和update/delete并发操作时的报错。

**参数类型:** USERSET

**取值范围:** 布尔型

- on, 提升deltamerge内部实现的锁级别, 并发执行deltamerge和update/delete操作时, 一个操作先执行, 另一个操作被阻塞, 在前一个操作完成后, 后一个操作再执行。
- off, 在对HDFS表的delta table的同一行并发执行deltamerge和update/delete操作时, 后一个对同一行数据更新的操作会报错退出。

**默认值:** off

## job\_queue\_processes

**参数说明:** 表示系统可以并发执行的job数目。

**参数类型:** POSTMASTER

**取值范围:** 0~1000

**功能:**

- 当job\_queue\_processes设置为0值, 表示不启用定时任务功能, 任何job都不会被执行 (因为开启定时任务的功能会对系统的性能有影响, 有些局点可能不需要定时任务的功能, 可以通过设置为0不启用定时任务功能)。
- 当job\_queue\_processes为大于0时, 表示启用定时任务功能且系统能够并发处理的最大任务数。

启用定时任务功能后, job\_scheduler线程会在定时时间间隔轮询pg\_jobs系统表, 系统设置定时任务检查周期默认为1s。

由于并行运行的任务数太多会消耗更多的系统资源, 因此需要设置系统并发处理的任務数, 当前并发的任务数达到job\_queue\_processes时, 且此时又有任务到期, 那么这些任务本次得不到执行而延期到下一轮询周期。因此, 建议用户需要根据每个任务的执行时长合理的设置任务的时间间隔 (即submit接口中的interval参数), 来避免由于任务执行时间太长而导致下个轮询周期无法正常执行。

注: 如果同一时间内并行的job数很多, 过小的参数值会导致job等待。而过大的参数值则消耗更多的系统资源, 建议设置此参数为100, 用户可以根据系统资源情况合理调整。

**默认值:** 10

## ngram\_gram\_size

**参数说明：**ngram解析器分词的长度。

**参数类型：**USERSET

**取值范围：**整型，1~4

**默认值：**2

## ngram\_grapsymbol\_ignore

**参数说明：**ngram解析器是否忽略图形化字符。

**参数类型：**USERSET

**取值范围：**布尔型

- on表示忽略图形化字符。
- off表示不忽略图形化字符。

**默认值：**off

## ngram\_punctuation\_ignore

**参数说明：**ngram解析器是否忽略标点符号。

**参数类型：**USERSET

**取值范围：**布尔型

- on表示忽略标点符号。
- off表示不忽略标点符号。

**默认值：**on

## zhparser\_dict\_in\_memory

**参数说明：**Zhparser解析器是否将字典加载到内存中。

**参数类型：**POSTMASTER

**取值范围：**布尔型

- on表示将字典加载到内存中
- off表示不将字典加载到内存中

**默认值：**on

## zhparser\_multi\_duality

**参数说明：**Zhparser解析器设定是否将长词内的文字自动以二字分词法聚合。

**参数类型：**USERSET

**取值范围：**布尔型

- on表示将长词内的文字自动以二字分词法聚合。



- off表示不将长词内的文字自动以二字分词法聚合。

**默认值：** off

## zhparser\_multi\_short

**参数说明：** Zhparser解析器分词执行时是否执行针对长词复合切分。

**参数类型：** USERSET

**取值范围：** 布尔型

- on表示执行针对长词复合切分。
- off表示不执行针对长词复合切分。

**默认值：** on

## zhparser\_multi\_zall

**参数说明：** Zhparser解析器是否将全部单字单独显示。

**参数类型：** USERSET

**取值范围：** 布尔型

- on表示将全部单字单独显示。
- off表示不将全部单字单独显示。

**默认值：** off

## zhparser\_multi\_zmain

**参数说明：** Zhparser解析器是否将重要单字单独显示。

**参数类型：** USERSET

**取值范围：** 布尔型

- on表示将重要单字单独显示。
- off表示不将重要单字单独显示。

**默认值：** off

## zhparser\_punctuation\_ignore

**参数说明：** Zhparser解析器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）。

**参数类型：** USERSET

**取值范围：** 布尔型

- on：忽略所有的标点等特殊符号。
- off：不忽略所有的标点等特殊符号。

**默认值：** on

## zhparser\_seg\_with\_duality

**参数说明：**Zhparser解析器是否将闲散文字自动以二字分词法聚合。

**参数类型：**USERSET

**取值范围：**布尔型

- on表示将闲散文字自动以二字分词法聚合。
- off表示不将闲散文字自动以二字分词法聚合。

**默认值：**off

## acceleration\_with\_compute\_pool

**参数说明：**在查询包含OBS时，通过该参数决定查询是否通过计算资源池进行加速。

**参数类型：**USERSET

**取值范围：**布尔型

- on表示包含有OBS的查询在计算资源池可用时，会根据代价评估决定是否通过计算资源池对查询加速。
- off表示任何查询都不会通过计算资源池进行加速。

**默认值：**off

## behavior\_compat\_options

**参数说明：**数据库兼容性行为配置项，该参数的值由若干个配置项用逗号隔开构成。

**参数类型：**USERSET

**取值范围：**字符串

**默认值：**升级场景下保持前向兼容，即与升级前的集群中该参数的默认值保持一致。新安装集群场景下，该参数默认值为check\_function\_conflicts，以防止用户定义错误的函数属性导致严重的问题。

### 说明

- 当前只支持表17-4。
- 配置多个兼容性配置项时，相邻配置项用逗号隔开，例如：set behavior\_compat\_options='end\_month\_calculate,display\_leading\_zero';
- 此参数选项中strict\_concat\_functions和strict\_text\_concat\_td不能同时设置。

表 17-4 兼容性配置项

兼容性配置项	兼容性行为控制	适用兼容模式
display_leading_zero	<p>浮点数显示配置项。</p> <ul style="list-style-type: none"> <li>不设置此配置项时，对于-1~0和0~1之间的小数，不显示小数点前的0。比如，0.25显示为.25。</li> <li>设置此配置项时，对于-1~0和0~1之间的小数，显示小数点前的0。比如，0.25显示为0.25。</li> </ul>	ORA TD
end_month_calculate	<p>add_months函数计算逻辑配置项。</p> <p>假定函数add_months的两个参数分别为param1和param2，param1的月份和param2的月份和为result。</p> <ul style="list-style-type: none"> <li>不设置此配置项时，如果param1的日期（Day字段）为月末，并且param1的日期（Day字段）比result月份的月末日期小，计算结果中的日期字段（Day字段）和param1的日期字段保持一致。比如，</li> </ul> <pre>select add_months('2018-02-28',3) from dual; add_months ----- 2018-05-28 00:00:00 (1 row)</pre> <ul style="list-style-type: none"> <li>设置此配置项时，如果param1的日期（Day字段）为月末，并且param1的日期（Day字段）比result月份的月末日期比小，计算结果中的日期字段（Day字段）和result的月末日期保持一致。比如，</li> </ul> <pre>select add_months('2018-02-28',3) from dual; add_months ----- 2018-05-31 00:00:00 (1 row)</pre>	ORA TD
compat_analyze_sample	<p>analyze采样行为配置项。</p> <p>设置此配置项时，会优化analyze的采样行为，主要体现在analyze时全局采样会更精确的控制3万条左右，更好的控制analyze时Coordinator端的内存消耗，保证analyze性能的稳定性。</p>	ORA TD MySQL
bind_schema_tablespace	<p>绑定模式与同名表空间配置项。</p> <p>如果存在与模式名sche_name相同的表空间名，那么如果设置search_path为sche_name，default_tablespace也会同步切换到sche_name。</p>	ORA TD MySQL

兼容性配置项	兼容性行为控制	适用兼容模式
bind_procedure_searchpath	<p>未指定模式名的数据库对象的搜索路径配置项。</p> <p>在存储过程中如果不显示指定模式名，会优先在存储过程所属的模式下搜索。</p> <p>如果找不到，则有两种情况：</p> <ul style="list-style-type: none"> <li>若不设置此参数，报错退出。</li> <li>若设置此参数，按照search_path中指定的顺序继续搜索。如果还是找不到，报错退出。</li> </ul>	ORA TD MySQL
correct_to_number	<p>控制to_number()结果兼容性的配置项。</p> <p>若设置此配置项，则to_number()函数结果与PG11保持一致，否则默认与Oracle保持一致。</p>	ORA
unbind_divide_bound	<p>控制对整数除法的结果进行范围校验。</p> <ul style="list-style-type: none"> <li>不设置此配置项时，将对除法结果进行校验，超出范围则报错。例如，示例中INT_MIN/(-1)会因为超过结果大于INT_MAX而报越界错误：  <pre>SELECT (-2147483648)::int / (-1)::int; ERROR: integer out of range</pre> </li> <li>若设置此配置项，则不需要对除法结果进行范围校验。例如，示例中INT_MIN/(-1)可以得到输出结果INT_MAX+1：  <pre>SELECT (-2147483648)::int / (-1)::int; ?column? ----- 2147483648 (1 row)</pre> </li> </ul>	ORA TD
merge_update_multi	<p>控制行存表merge into匹配多行时是否进行update操作。</p> <p>若设置此配置项，匹配多行时update不报错，否则默认与Oracle保持一致，报错。</p>	ORA TD
disable_row_update_multi	<p>控制行存表update匹配多行时是否进行update操作。</p> <p>若设置此配置项，匹配多行时update报错，否则默认可以进行多行匹配更新。</p>	ORA TD

兼容性配置项	兼容性行为控制	适用兼容模式
return_null_string	<p>控制函数lpad()、rpad()、repeat()、regexp_split_to_table()和split_part()的结果为空字符串"的显示配置项。</p> <ul style="list-style-type: none"> <li>不设置此配置项时，空字符串显示为NULL。</li> </ul> <pre>select length(lpad('123',0,'*')) from dual; length ----- (1 row)</pre> <ul style="list-style-type: none"> <li>设置此配置项时，空字符串显示为"。</li> </ul> <pre>select length(lpad('123',0,'*')) from dual; length ----- 0 (1 row)</pre>	ORA
compat_concat_variadic	<p>控制函数concat()和concat_ws()对variadic类型结果兼容性的配置项。</p> <p>若设置此配置项，当concat函数参数为variadic类型时，保留Oracle和Teradata兼容模式下不同的结果形式；否则默认Oracle和Teradata兼容模式下结果相同，且与Oracle保持一致。</p>	ORA TD
convert_string_digit_to_numeric	<p>控制CHAR类型和INT类型进行二元BOOL运算时类型转换优先级的配置项。</p> <ul style="list-style-type: none"> <li>不设置此配置项时，类型转换优先级与PG9.6一致。</li> <li>设置此配置项时，所有CHAR类型和INT类型的二元BOOL运算均强制转换为NUMERIC类型进行计算。</li> </ul> <p>设置此配置项后会被影响的CHAR类型包括BPCHAR、VARCHAR、NVARCHAR2、TEXT四种类型，会被影响的INT类型包括INT1、INT2、INT4、INT8四种类型。</p> <p><b>注意</b> 此配置项只对二元BOOL运算生效，例如，INT2&gt;TEXT、INT4=BPCHAR，非BOOL运算不会受到影响，该配置项暂不支持INT&gt;'1.1'这类UNKNOWN类型运算的转换。由于该配置项开启后，CHAR类型与INT类型的BOOL运算会优先转换为NUMERIC类型进行计算，因此会影响数据库计算性能，当JOIN列为受影响的类型组合时，还会影响执行计划。</p>	ORA TD MySQL

兼容性配置项	兼容性行为控制	适用兼容模式
check_function_conflicts	<p>控制是否检查自定义plpgsql/SQL函数的属性。</p> <ul style="list-style-type: none"> <li>不设置此配置项时，不检查自定义函数的IMMUTABLE/STABLE/VOLATILE属性。</li> <li>设置此配置项时，会检查自定义函数的IMMUTABLE属性，如果函数中含有表，或者是有STABLE/VOLATILE函数时，在执行时会报错。因为函数中如果有表或者STABLE/VOLATILE函数时，与函数定义中的IMMUTABLE属性冲突，即这种场景下，函数的行为非IMMUTABLE。</li> </ul> <p>例如：设置此参数时，以下场景下会执行报错：            CREATE OR replace FUNCTION sql_immutable (INTEGER)            RETURNS INTEGER AS 'SELECT a+\$1 from shipping_schema.t4            where a=1;'            LANGUAGE SQL IMMUTABLE            RETURNS NULL            ON NULL INPUT;            select sql_immutable(1);            ERROR: IMMUTABLE function cannot contain SQL statements            with relation or Non-IMMUTABLE function.            CONTEXT: SQL function "sql_immutable" during startup            referenced column: sql_immutable</p>	ORA TD MySQL
varray_verification	<p>控制是否校验数组长度以及数组类型长度。用于兼容GaussDB(DWS) 8.1.0之前的版本。</p> <p>若设置此配置项，不会校验数组长度以及数组类型长度。</p> <pre>-- 场景1 CREATE OR REPLACE PROCEDURE varray_verification AS     TYPE org_varray_type IS varray(5) OF VARCHAR2(2);     v_org_varray org_varray_type; BEGIN     v_org_varray(1) := '111'; --例如赋值已经超过了VARCHAR2(2)的限制，配置该选项后将和历史版本保持一致不进行校验 END; / --场景2 CREATE OR REPLACE PROCEDURE varray_verification_i3_1 AS     TYPE org_varray_type IS varray(2) OF NUMBER(2);     v_org_varray org_varray_type; BEGIN     v_org_varray(3) := 1; --例如赋值已经超过了varray(2)的数组长度限制，配置该选项后将和历史版本保持一致不进行校验 END; /</pre>	ORA TD

兼容性配置项	兼容性行为控制	适用兼容模式
strict_concat_functions	<p>控制函数textanycat()和anytextcat()在参数存在空值时，对返回值兼容性的配置项。此参数不能和strict_text_concat_td同时设置。</p> <p>MySQL兼容模式下，此参数无影响。</p> <ul style="list-style-type: none"> <li>不设置此配置项时，函数textanycat()和anytextcat()的返回值默认与Oracle保持一致。</li> <li>设置此配置项时，若函数textanycat()和anytextcat()的参数存在空值，则返回值也为空值，保留与Oracle和Teradata兼容模式下不同的结果。</li> </ul> <p>例如，不设置此配置项时，函数textanycat()和anytextcat()的返回值与Oracle保持一致：</p> <pre>SELECT textanycat('gauss', cast(NULL as BOOLEAN)); textanycat ----- gauss (1 row)</pre> <p>SELECT 'gauss'    cast(NULL as BOOLEAN); --这种情况下，  运算符会被转换为函数textanycat ?column? ----- gauss (1 row)</p> <p>设置此配置项时，保留与Oracle和Teradata兼容模式下不同的结果：</p> <pre>SELECT textanycat('gauss', cast(NULL as BOOLEAN)); textanycat ----- (1 row)</pre> <p>SELECT 'gauss'    cast(NULL as BOOLEAN); --这种情况下，  运算符会被转换为函数textanycat ?column? ----- (1 row)</p>	ORATD

兼容性配置项	兼容性行为控制	适用兼容模式
<p>strict_text_concat_td</p>	<p>Teradata兼容模式下，控制函数textcat()、textanycat()和anytextcat()在参数存在空值时，对返回值兼容性的配置项。此参数不能和strict_concat_functions同时设置。</p> <ul style="list-style-type: none"> <li>不设置此配置项时，Teradata兼容模式下函数textcat()、textanycat()和anytextcat()的返回值与GaussDB(DWS)一致。</li> <li>设置此配置项时，若Teradata兼容模式下函数textcat()、textanycat()和anytextcat()的参数存在空值，则返回值为空值。</li> </ul> <p>例如，不设置此配置项时，函数textcat()、textanycat()和anytextcat()的返回值与GaussDB(DWS)保持一致：  td_compatibility_db=# SELECT textcat('abc', NULL);  textcat  -----  abc  (1 row)</p> <p>td_compatibility_db=# SELECT 'abc'    NULL; --这种情况下，  运算符会被转换为函数textcat()  ?column?  -----  abc  (1 row)</p> <p>设置此配置项时，若函数textcat()、textanycat()和anytextcat()的返回值有空值，则返回NULL：  td_compatibility_db=# SELECT textcat('abc', NULL);  textcat  -----    (1 row)</p> <p>td_compatibility_db=# SELECT 'abc'    NULL;  ?column?  -----    (1 row)</p>	<p>TD</p>
<p>compat_display_ref_table</p>	<p>设置视图中列的显示格式。</p> <ul style="list-style-type: none"> <li>不设置该选项时默认带前缀，即tab.col的格式。</li> <li>设置该选项时与原始定义一致，原始定义带前缀则显示，否则不显示。</li> </ul> <pre>SET behavior_compat_options='compat_display_ref_table'; CREATE OR REPLACE VIEW viewtest2 AS SELECT a.c1, c2, a.c3, 0 AS c4 FROM viewtest_tbl a; SELECT pg_get_viewdef('viewtest2'); pg_get_viewdef ----- SELECT a.c1, c2, a.c3, 0 AS c4 FROM viewtest_tbl a; (1 row)</pre>	<p>ORA TD</p>



兼容性配置项	兼容性行为控制	适用兼容模式
para_support_set_func	<p>列存表中控制函数COALESCE()、NVL()、GREATEST()、LEAST()入参是否支持多结果集表达式。</p> <ul style="list-style-type: none"> <li>不设置此配置项时，函数入参包含多结果集表达式时，直接报错不支持。</li> </ul> <pre>SELECT COALESCE(regexp_split_to_table(c3,'#'), regexp_split_to_table(c3,'#')) FROM regexp_ext2_tb1 ORDER BY 1 LIMIT 5; ERROR: set-valued function called in context that cannot accept a set</pre> <ul style="list-style-type: none"> <li>设置此配置项时，支持函数入参包含多结果集表达式。</li> </ul> <pre>SELECT COALESCE(regexp_split_to_table(c3,'#'), regexp_split_to_table(c3,'#')) FROM regexp_ext2_tb1 ORDER BY 1 LIMIT 5; coalesce ----- a a a a a (5 rows)</pre>	ORA TD
disable_select_truncate_parallel	<p>控制分区表的truncate等ddl的锁等级。</p> <ul style="list-style-type: none"> <li>设置此配置项时，将禁止分区表的不同分区上truncate与DML(如select)的并发，允许分区表上select的FQS(快速下发)。在OLTP场景下分区表上的简单查询较多，并且没有分区表不同分区truncate与DML并发的需求，可以考虑设置此配置项。</li> <li>不设置此配置项时，分区表上不同分区的select与truncate可以并发进行，同时关闭分区表的FQS（快速下发）来避免可能的不一致问题。</li> </ul>	ORA TD MySQL
bpchar_text_without_rtrim	<p>Teradata兼容模式下，设置此参数时，控制bpchar到text转换保留右侧空格，如果实际长度不足bpchar指定的长度，对其进行补空格操作，兼容Teradata对bpchar字符串的处理风格。</p> <p>当前不支持“比较字符串时忽略尾部空格”，拼接后结果如果存在尾部空格，进行比较时会对空格敏感。</p> <p>例如，设置参数时：</p> <pre>td_compatibility_db=# select length('a'::char(10)::text); length ----- 10 (1 row)  td_compatibility_db=# select length('a'::char(10)); length ----- 11 (1 row)</pre>	TD

兼容性配置项	兼容性行为控制	适用兼容模式
<p>convert_empty_str_to_null_t_d</p>	<p>Teradata兼容模式下，设置此参数时，控制to_date, to_timestamp和to_number类型转换函数处理空串时，返回null；同时控制to_char函数处理date类型入参时返回的格式。</p> <p>例如：</p> <p>未设置此参数时：</p> <pre>td_compatibility_db=# select to_number(""); to_number ----- 0 (1 row)</pre> <pre>td_compatibility_db=# select to_date(""); ERROR: the format is not correct DETAIL: invalid date length "0", must between 8 and 10. CONTEXT: referenced column: to_date</pre> <pre>td_compatibility_db=# select to_timestamp(""); to_timestamp ----- 0001-01-01 00:00:00 BC (1 row)</pre> <pre>td_compatibility_db=# select to_char(date '2020-11-16'); to_char ----- 2020-11-16 00:00:00+08 (1 row)</pre> <p>设置此参数，若to_number, to_date, to_timestamp函数的参数有空串时：</p> <pre>td_compatibility_db=# select to_number(""); to_number ----- (1 row)</pre> <pre>td_compatibility_db=# select to_date(""); to_date ----- (1 row)</pre> <pre>td_compatibility_db=# select to_timestamp(""); to_timestamp ----- (1 row)</pre> <pre>td_compatibility_db=# select to_char(date '2020-11-16'); to_char ----- 2020/11/16 (1 row)</pre>	<p>TD</p>

兼容性配置项	兼容性行为控制	适用兼容模式
<p>disable_case_specific</p>	<p>控制字符类型匹配时是否忽略大小写。仅在Teradata兼容模式下生效。</p> <ul style="list-style-type: none"> <li>不设置此配置项时，字符类型匹配时，字符的大小写敏感。</li> <li>设置此配置项时，字符类型匹配时，字符的大小写不敏感。</li> </ul> <p>设置此配置项后会影响的字符类型包括CHAR、TEXT、BPCHAR、VARCHAR、NVARCHAR五种类型，会被影响的操作符包括&lt;、&gt;、=、&gt;=、&lt;=、!=、&lt;&gt;、!=、like、not like、in、not in共12种操作符以及case when、decode 表达式。</p> <p><b>注意</b> 由于该配置项开启后，字符类型前会增加UPPER函数进而会影响估算逻辑，需要使用增强的估算模型。（建议设置：cost_param=16、cost_model_version = 1、join_num_distinct=-20、qual_num_distinct=200）</p>	<p>TD</p>
<p>enable_interval_to_text</p>	<p>控制interval到text类型的隐式转换功能。</p> <ul style="list-style-type: none"> <li>设置此选项时，支持interval类型到text类型的隐式转换。  <pre>SELECT TO_DATE('20200923', 'yyyymmdd') - TO_DATE('20200920', 'yyyymmdd') = '3'::text; ?column? ----- f (1 row)</pre> </li> <li>不设置此选项时，不支持interval类型到text类型的隐式转换。  <pre>SELECT TO_DATE('20200923', 'yyyymmdd') - TO_DATE('20200920', 'yyyymmdd') = '3'::text; ?column? ----- t (1 row)</pre> </li> </ul>	<p>ORA TD MySQL</p>
<p>case_insensitive</p>	<p>MySQL兼容模式下，设置此参数，控制locate，strpos，instr字符串函数入参大小写不敏感。目前默认未设置该参数，即入参大小写敏感。例如：</p> <ul style="list-style-type: none"> <li>未设置此选项时，入参大小写敏感。  <pre>mysql_compatibility_db=# SELECT LOCATE('sub', 'Substr'); locate ----- 0 (1 row)</pre> </li> <li>设置此选项时，入参大小写不敏感。  <pre>mysql_compatibility_db=# SELECT LOCATE('sub', 'Substr'); locate ----- 1 (1 row)</pre> </li> </ul>	<p>MySQL</p>

兼容性配置项	兼容性行为控制	适用兼容模式
inherit_not_null_strict_func	<p>控制函数原有的strict属性，参数为1个的函数可以传递NOT NULL属性的行为。即：对于func(x)，如果func()为strict属性，且x包含NOT NULL约束，则认为func(x)也是包含NOT NULL约束的。</p> <p>该兼容配置项在某些优化场景，例如：NOT IN优化、COUNT(DISTINCT)优化，会有特定的优化效果，但特定场景可能导致结果错误。</p> <p>目前默认未设置该参数，保证结果正确，但可能导致性能回退，如果出现问题可设置该参数回退到历史版本行为。</p>	ORA TD MySQL
disable_compatibility_minmax_expr_mysql	<p>MySQL兼容模式下，控制greatest/least表达式对null入参的处理方式。</p> <p>默认兼容MySQL。可通过设置此参数，回退到历史版本行为。</p> <ul style="list-style-type: none"> <li>不设置此选项时，兼容MySQL行为，入参为null时返回null。 mysql_compatibility_db=# SELECT greatest(1, 2, null), least(1, 2, null); greatest   least -----+-----             (1 row)</li> <li>设置此选项时，返回非null参数中的最大/小值。 mysql_compatibility_db=# SELECT greatest(1, 2, null), least(1, 2, null); greatest   least -----+-----           2   1 (1 row)</li> </ul>	MySQL

兼容性配置项	兼容性行为控制	适用兼容模式
<p>disable_compatibility_substr_mysql</p>	<p>MySQL兼容模式下，控制substr/substring函数在起始位置pos &lt;= 0时的行为。</p> <p>默认兼容MySQL。可通过设置此参数，回退到历史版本行为。</p> <ul style="list-style-type: none"> <li>不设置此选项时，兼容MySQL行为，即pos = 0时返回空串，pos &lt; 0时从倒数第  pos  个位置开始截取字符。  <pre>mysql_compatibility_db=# SELECT substr('helloworld',0); substr ----- (1 row) mysql_compatibility_db=# SELECT substring('helloworld',0),substring('helloworld',-2,4); substring   substring -----+-----             ld (1 row)</pre> </li> <li>设置此选项时，pos &lt;= 0时仍然从左侧开始截取字符。  <pre>mysql_compatibility_db=# SELECT substr('helloworld',0); substr ----- helloworld (1 row) mysql_compatibility_db=# SELECT substring('helloworld',0),substring('helloworld',-2,4); substring   substring -----+----- helloworld   h (1 row)</pre> </li> </ul>	<p>MySQL</p>
<p>disable_compatibility_trim_mysql</p>	<p>MySQL兼容模式下，控制trim/ltrim/rtrim函数对入参的处理方式。</p> <p>默认兼容MySQL。可通过设置此参数，回退到历史版本行为。</p> <ul style="list-style-type: none"> <li>不设置此选项时，兼容MySQL行为，匹配完整子串。  <pre>mysql_compatibility_db=# SELECT trim('{{name}}','{}'),trim('xyznamezyx','xyz'); btrim   btrim -----+----- {name}   namezyx (1 row)</pre> </li> <li>设置此选项时，匹配字符集合中的单个字符。  <pre>mysql_compatibility_db=# SELECT trim('{{name}}','{}'),trim('xyznamezyx','xyz'); btrim   btrim -----+----- name   name (1 row)</pre> </li> </ul>	<p>MySQL</p>

兼容性配置项	兼容性行为控制	适用兼容模式
light_object_mtime	<p>控制pg_object系统表mtime字段是否会记录对象行为的操作。</p> <ul style="list-style-type: none"> <li>• 设置此选项时，GRANT/REVOKE/TRUNCATE操作不被mtime记录即不更新mtime字段。</li> <li>• 不设置此选项时（默认行为），ALTER操作、COMMENT、GRANT/REVOKE和TRUNCATE均会被mtime记录即更新mtime字段。</li> </ul>	ORA TD MySQL

兼容性配置项	兼容性行为控制	适用兼容模式																																								
<p>disable_including_all_mysql</p>	<p>MySQL兼容模式下，控制CREATE TABLE ... LIKE语法是否为INCLUDING_ALL模式。</p> <p>默认不设置此参数，即MySQL兼容模式下，CREATE TABLE ... LIKE语法默认为INCLUDING_ALL模式。</p> <p>可通过设置此参数，回退到历史版本行为。</p> <ul style="list-style-type: none"> <li>不设置此选项，MySQL兼容模式下，CREATE TABLE ... LIKE语法为INCLUDING_ALL模式。 mysql_compatibility_db=# create table mysql_like(id int, name varchar(10), score int) distribute by hash(id) COMMENT 'mysql_like'; CREATE TABLE mysql_compatibility_db=# create index index_like on mysql_like(name); CREATE INDEX mysql_compatibility_db=# \d+ mysql_like; Table "public.mysql_like"  <table border="1"> <thead> <tr> <th>Column</th> <th>Type</th> <th>Modifiers</th> <th>Storage</th> <th>Stats target</th> </tr> </thead> <tbody> <tr> <td>id</td> <td>integer</td> <td></td> <td>plain</td> <td></td> </tr> <tr> <td>name</td> <td>character varying(10)</td> <td></td> <td>extended</td> <td></td> </tr> <tr> <td>score</td> <td>integer</td> <td></td> <td>plain</td> <td></td> </tr> </tbody> </table> <p>Indexes: "index_like" btree (name) TABLESPACE pg_default Has OIDs: no Distribute By: HASH(id) Location Nodes: ALL DATANODES Options: orientation=row, compression=no</p>   <p>mysql_compatibility_db=# create table copy_like like mysql_like; CREATE TABLE mysql_compatibility_db=# \d+ copy_like; Table "public.copy_like"  <table border="1"> <thead> <tr> <th>Column</th> <th>Type</th> <th>Modifiers</th> <th>Storage</th> <th>Stats target</th> </tr> </thead> <tbody> <tr> <td>id</td> <td>integer</td> <td></td> <td>plain</td> <td></td> </tr> <tr> <td>name</td> <td>character varying(10)</td> <td></td> <td>extended</td> <td></td> </tr> <tr> <td>score</td> <td>integer</td> <td></td> <td>plain</td> <td></td> </tr> </tbody> </table> <p>Indexes: "copy_like_name_idx" btree (name) TABLESPACE pg_default Has OIDs: no Distribute By: HASH(id) Location Nodes: ALL DATANODES Options: orientation=row, compression=no</p> </p></li> <li>设置此选项，MySQL兼容模式下，CREATE TABLE ... LIKE语法为空模式。 mysql_compatibility_db=# set behavior_compat_options = 'disable_including_all_mysql'; SET mysql_compatibility_db=# create table mysql_copy like mysql_like; 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</li> </ul>	Column	Type	Modifiers	Storage	Stats target	id	integer		plain		name	character varying(10)		extended		score	integer		plain		Column	Type	Modifiers	Storage	Stats target	id	integer		plain		name	character varying(10)		extended		score	integer		plain		<p>MySQL</p>
Column	Type	Modifiers	Storage	Stats target																																						
id	integer		plain																																							
name	character varying(10)		extended																																							
score	integer		plain																																							
Column	Type	Modifiers	Storage	Stats target																																						
id	integer		plain																																							
name	character varying(10)		extended																																							
score	integer		plain																																							

兼容性配置项	兼容性行为控制	适用兼容模式
	<pre>mysql_db=# \d+ mysql_copy;           Table "public.mysql_copy"   Column        Type        Modifiers   Storage   Stats target     Description -----+-----+-----+-----+-----+----- id        integer                      plain                    name      character varying(10)             extended                  score     integer                      plain                    Has OIDs: no Distribute By: ROUND ROBIN Location Nodes: ALL DATANODES Options: orientation=row, compression=no</pre>	
cte_onetime_inline	<p>控制非stream计划是否执行inline。</p> <ul style="list-style-type: none"> <li>● 设置此选项时，非stream计划且只被引用一次的CTE执行inline。</li> <li>● 不设置此选项时，非stream计划且只被引用一次的CTE不执行inline。</li> </ul>	ORA TD MySQL
skip_first_after_mysql	<p>MySQL兼容模式下，控制是否在ALTER TABLE ADD/MODIFY/CHANGE COLUMN中忽略FIRST/AFTER colname语法。</p> <ul style="list-style-type: none"> <li>● 设置此选项时，将忽略FIRST/AFTER colname语法，执行不报错。  <pre>mysql_compatibility_db=# set behavior_compat_options = 'skip_first_after_mysql'; mysql_compatibility_db=# alter table t1 add column b text after a; ALTER TABLE</pre> </li> <li>● 不设置此选项时，将不支持FIRST/AFTER colname语法，执行报错。  <pre>mysql_compatibility_db=# set behavior_compat_options = ""; mysql_compatibility_db=# alter table t1 add column b text after a; ERROR: FIRST/AFTER is not yet supported.</pre> </li> </ul>	MySQL
enable_division_by_zero_mysql	<p>MySQL兼容模式下，除数为0时，控制除法或取余操作是否报错。（该配置项仅8.1.3.110及以上集群版本支持。）</p> <ul style="list-style-type: none"> <li>● 设置此选项时，除法或取余操作除数为0时，返回NULL。  <pre>compatible_mysql_db=# SET behavior_compat_options = 'mysql_enable_division_by_zero'; SET compatible_mysql_db=# select 1/0 as test; test ----- (1 row)</pre> </li> <li>● 不设置此选项时，除法或取余操作除数为0时，执行报错。  <pre>compatible_mysql_db=# select 1/0; ERROR: division by zero</pre> </li> </ul>	MySQL



兼容性配置项	兼容性行为控制	适用兼容模式
normal_session_id	<p>控制是否生成normal格式的会话ID。</p> <ul style="list-style-type: none"> <li>设置此选项时，生成normal格式的会话ID，兼容8.1.3及之前集群版本的会话ID。  <pre>SET behavior_compat_options='normal_session_id'; SELECT pg_current_sessionid(); pg_current_sessionid ----- 1660268184.140594655524608 (1 row)</pre> </li> <li>不设置此选项时，生成pretty格式的会话ID。  <pre>SET behavior_compat_options=""; SELECT pg_current_sessionid(); pg_current_sessionid ----- 1660268184.140594655524608.coordinator1 (1 row)</pre> </li> </ul>	ORA TD MySQL
disable_jsonb_exact_match	<p>控制操作符匹配规则，在对二元操作符进行模糊匹配时是否考虑jsonb类型。（该参数仅8.2.0及以上集群版本支持）</p> <ul style="list-style-type: none"> <li>设置此选项时，当对操作符进行模糊匹配时，会在所有备选（即不区分jsonb类型）中匹配，兼容8.1.1到8.1.3集群版本的匹配规则。  <pre>SET behavior_compat_options='disable_jsonb_exact_match'; select '2022' - '2'::text; ERROR: cannot delete from scalar</pre> </li> <li>不设置此选项时，当对操作符进行模糊匹配时，会在不含jsonb类型的备选匹配，兼容8.1.1之前集群版本的匹配规则。  <pre>SET behavior_compat_options=""; select '2022' - '2'::text; ?column? ----- 2020 (1 row)</pre> </li> </ul>	ORA TD MySQL
merge_into_with_trigger	<p>控制是否支持对有触发器的表执行MERGE INTO操作。（该参数仅8.1.3.200及以上集群版本支持）</p> <ul style="list-style-type: none"> <li>设置此选项时，可以对有触发器的表执行MERGE INTO操作。注意在MERGE INTO操作执行时，表上的触发器不会被触发执行。</li> <li>不设置此选项时，对有触发器的表执行MERGE INTO操作时报错。</li> </ul>	ORA TD MySQL

兼容性配置项	兼容性行为控制	适用兼容模式
add_column_default_v_function	<p>控制alter table add column default expression中expression是否支持volatile类型的函数。（该参数仅8.1.3.200及以上集群版本支持）。</p> <ul style="list-style-type: none"> <li>• 设置此选项时，alter table add column default expression中的expression支持volatile类型的函数。</li> <li>• 不设置此选项时，alter table add column default expression中expression不支持volatile类型的函数，如果expression中有volatile的函数，该语句执行会报错。</li> </ul>	ORA TD MySQL

兼容性配置项	兼容性行为控制	适用兼容模式
<p>disable_gc_fdw_filter_partial_pushdown</p>	<p>协同分析外表（类型为gc_fdw）场景，控制使用过滤条件查询外表数据时过滤条件的下推情况。（该参数仅8.2.0及以上集群版本支持）</p> <ul style="list-style-type: none"> <li> <p>设置此选项时，过滤条件中若存在不满足下推条件的因素（如非immutable函数），为了保证结果集文档，则全部过滤条件不下推，此行为兼容8.2.0版本之前的行为。</p> <pre> --源端集群建表 CREATE TABLE t1(c1 INT, c2 INT, c3 INT) DISTRIBUTE BY HASH(c1); --本地集群建相同结构外表 CREATE SERVER server_remote FOREIGN DATA WRAPPER gc_fdw options(ADDRESS 'address', DBNAME 'dbname', USERNAME 'username', PASSWORD 'password'); CREATE FOREIGN TABLE t1(c1 INT, c2 INT, c3 INT) SERVER server_remote; --打开参数时，条件下推情况 SET behavior_compat_options = 'disable_gc_fdw_filter_partial_pushdown'; EXPLAIN (verbose on,costs off) SELECT * FROM t1 WHERE c1&gt;3 AND c2 &lt;100 AND now() - '20230101' &lt; c3;                                 QUERY PLAN ----- Streaming (type: GATHER)   Output: c1, c2, c3   Node/s: All datanodes   -&gt; Foreign Scan on ca_schema.t1       Output: c1, c2, c3       Filter: ((t1.c1 &gt; 3) AND (t1.c2 &lt; 100) AND ((now() - '2023-01-01 00:00:00-08'::timestamp with time zone) &lt; (t1.c3)::interval))       Remote SQL: SELECT c1, c2, c3 FROM ca_schema.t1 (7 rows) </pre> </li> <li> <p>不设置此选项时，过滤条件中可下推的部分将下推到源端集群执行，不可下推部分将在本地集群执行，如此可提升外表查询效率。</p> <pre> --关闭参数时，条件下推情况 SET behavior_compat_options = ''; EXPLAIN (verbose on,costs off) SELECT * FROM t1 WHERE c1&gt;3 AND c2 &lt;100 AND now() - '20230101' &lt; c3;                                 QUERY PLAN ----- Streaming (type: GATHER)   Output: c1, c2, c3   Node/s: All datanodes   -&gt; Foreign Scan on ca_schema.t1       Output: c1, c2, c3       Filter: ((now() - '2023-01-01 00:00:00-08'::timestamp with time zone) &lt; (t1.c3)::interval)       Remote SQL: SELECT c1, c2, c3 FROM ca_schema.t1 WHERE ((c1 &gt; 3)) AND ((c2 &lt; 100)) (7 rows) </pre> </li> </ul>	<p>ORA TD MySQL</p>

## redact\_compat\_options

**参数说明：**设置数据脱敏可算不可见兼容性行为配置项。该参数仅8.1.3.310及以上集群版本支持。

**参数类型：**USERSET

**取值范围：**字符串

- none表示未配置兼容项。
- disable\_comparison\_operator\_mask表示允许不存在暴露原始数据风险的比较操作符绕过脱敏检查，输出实际计算结果。

**默认值：**none

## table\_skewness\_warning\_threshold

**参数说明：**设置用于表倾斜告警的阈值。

**参数类型：**SUSET

**取值范围：**浮点型，0~1

**默认值：**1

## table\_skewness\_warning\_rows

**参数说明：**设置用于表倾斜告警的行数。

**参数类型：**SUSET

**取值范围：**整型，0~INT\_MAX

**默认值：**100000

## max\_cache\_partition\_num

**参数说明：**设置扩容重分布过程中列存节省内存模式的分区数目。如果超过分区数据目，则最早缓存的分区将直接写入列存文件中。

**参数类型：**SIGHUP

**取值范围：**整型，最小值为0，最大值为32767。

- 0表示关闭列存节省内存模式。
- 1~32767表示存分区表最多缓存的分区数目。

**默认值：**0

### 说明

该参数用于扩容重分布，合理设置可以缓解列存分区表重分布过程中的内存消耗。但某些分区数据分布非常不均衡的表在重分布完成后，可能会产生较多小CU。如果出现较多小CU，需要通过VACUUM FULL来合并小CU。

## enable\_prevent\_job\_task\_startup

**参数说明：**设置用于阻止job线程的启动。该参数属于系统内部参数，不建议用户修改设置。

**参数类型：**SIGHUP

**取值范围：**布尔型

- on表示阻止启动job线程。当job周期到来时，不会启动job执行线程。
- off表示允许启动job线程。当job周期到来时，会启动job执行线程，完成job中规定的操作。

**默认值：** off

### 说明

该参数只需在CN上设置。

## auto\_process\_residualfile

**参数说明：**控制残留文件记录功能的开关。

**参数类型：**SIGHUP

**取值范围：**布尔型

- on表示打开残留文件记录功能。
- off表示关闭残留文件记录功能。

**默认值：** off

## enable\_view\_update

**参数说明：**用于设置是否开启视图更新功能。

**参数类型：**POSTMASTER

**取值范围：**布尔型

- on表示启用视图更新功能。
- off表示关闭视图更新功能。

**默认值：** off

## view\_independent

**参数说明：**用于设置是否开启视图与表、函数、同义词的解耦功能。基表恢复后目前已支持自动关联重建。

**参数类型：**SIGHUP

**取值范围：**布尔型

- on表示启用视图解耦功能，存在视图依赖的表、函数、同义词及其他视图可以单独删除（临时表及临时视图除外），关联视图保留但不可用。
- off表示关闭视图解耦功能，存在视图依赖的表、函数、同义词及其他视图不可以单独删除，仅可使用cascade级联删除。

**默认值:** off

## bulkload\_report\_threshold

**参数说明:** 设置导入导出统计信息上报阈值。

**参数类型:** SIGHUP

**取值范围:** 整型, 0~INT\_MAX

**默认值:** 50

## assign\_abort\_xid

**参数说明:** 查询时将指定的xid判断为需要abort的事务。

**参数类型:** USERSET

**取值范围:** 指定xid的字符串

---

### 注意

此参数只用于用户误删数据(delete操作)后进行快速恢复。其他场景禁止使用, 否则造成事务可见性错误问题。

---

## default\_distribution\_mode

**参数说明:** 用于设置表的默认分布方式。该参数仅8.1.2及以上版本支持。

**参数类型:** USERSET

**取值范围:** 枚举类型

- roundrobin, 创建表不指定分布方式时, 按如下规则选取默认分布方式:
  - 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。
- hash, 创建表不指定分布方式时, 按如下规则选取默认分布方式:
  - 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。

**默认值:** roundrobin

### 说明

新建8.1.2集群版本默认值为roundrobin, 升级到8.1.2集群版本场景该参数的默认值为hash。

## object\_mtime\_record\_mode

**参数说明：**用于设置PG\_OBJECT系统表中mtime字段的更新行为。

**参数类型：**SIGHUP

**取值范围：**字符串

- default, 表示默认行为包括ALTER、COMMENT、GRANT/REVOKE和TRUNCATE操作会更新mtime字段。
- none, 表示不更新mtime字段。
- disable\_acl, 表示GRANT/REVOKE操作不更新mtime字段。
- disable\_truncate, 表示TRUNCATE操作不更新mtime字段。
- disable\_partition, 表示分区表相关ALTER操作不更新mtime字段。

**默认值：**default

## enable\_redactcol\_computable

**参数说明：**用于设置是否开启数据脱敏可算功能。该参数仅8.2.0及以上集群版本支持。

**参数类型：**POSTMASTER

**取值范围：**布尔型

- on表示启用数据脱敏可算功能。
- off表示关闭数据脱敏可算功能。

**默认值：**off

## enable\_redactcol\_equal\_const

**参数说明：**用于设置数据脱敏可算场景下是否允许脱敏列和常量做等值比较。该参数仅8.2.0及以上集群版本支持。

**参数类型：**SIGHUP

**取值范围：**布尔型

- on, 表示允许数据脱敏可算场景下脱敏列和常量做等值比较。
- off, 表示禁止数据脱敏可算场景下脱敏列和常量做等值比较。

**默认值：**off

## max\_volatile\_tables

**参数说明：**指定每个session创建的volatile临时表最大个数, 其中包括volatile临时表自身及其附属表。该参数仅8.2.0及以上集群版本支持。

**参数类型：**USERSET

**取值范围：**整型, 0~INT\_MAX

**默认值：**300

# 18 术语表

术语	解释
<b>A - E</b>	
ACID	在可靠数据库管理系统（DBMS）中，事务（transaction）所应该具有四个特性：原子性（Atomicity）、一致性（Consistency）、隔离性（Isolation）、和持久性（Durability）。
安全环	每一个环都由若干物理机组成，环内的DN形成主、备、从备关系，不向环外延伸。也就是说，环内的任何一个节点的主，或者备，或从备，依然在环内。扩容与缩容时，是以环为最小单位进行的。
Bgwriter	数据库启动时创建的一个后台写线程，此线程用于将数据库中脏页面写入到持久性设备（例如磁盘）中。
bit	比特。计算机处理的最小的信息单位。比特用来表示二进制数字1或0，或者一种逻辑条件真或假。在物理上，比特表示一个电路上高或低的电压点或者磁盘上的磁化单程或其它。一个单独的比特位所传达的信息很少有意义的。然而，一个8位组却构成了一个字节，可用于表示如一个英文字母，十进制数字，或其它字符等多种类型的信息。
Bloom Filter	布隆过滤器。由Howard Bloom在1970年提出的二进制向量数据结构，它具有很好的空间和时间效率，被用来检测一个元素是不是集合中的一个成员，这种检测只会对在集合内的数据错判，而不会对不是集合内的数据进行错判，这样每个检测请求返回有“在集合内（可能错误）”和“不在集合内（绝对不在集合内）”两种情况，可见Bloom filter是牺牲了正确率换取时间和空间。
CCN	Central Coordinator， GaussDB(DWS)动态负载管理中心协调节点。负责进行各CN中复杂作业是否可以执行的中心判断、排队和调度，以实现动态负载管理。



术语	解释
CIDR	Classless Inter-Domain Routing, 无类域间路由IP编址方案。CIDR摒弃传统的基于类 (A类: 8, B类: 16, C类: 24) 的地址分配方式, 允许使用任意长度的地址前缀, 有效提高地址空间的利用率。CIDR表示方法: IP地址/网络ID的位数。比如192.168.23.35/21, 其中“21”表示前面地址中的前21位代表网络部分, 其余位代表主机部分。
Cgroups	Control Groups, 控制组 ( GaussDB(DWS)中也称之为优先级组)。SUSE Linux和RedHat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的机制。
CLI	Command-line Interface, 命令行界面。应用程序和用户交互的一种方式, 完全基于文本输入和输出。命令通过键盘或类似装置输入, 由程序编译并执行。结果是以文本或图形的方式呈现在终端界面。
CM	Cluster Manager, 集群管理模块。管理和监控分布式系统中各个功能单元和物理资源的运行情况, 确保整个系统的稳定运行。
CMS	Cluster Management Service, 集群管理服务。是用于管理集群状态的部件。
CN	Coordinator, 负责数据库系统元数据存储、查询任务的分解和部分执行, 以及将DN中查询结果汇聚在一起。
CU	Compression Unit, 压缩单元。列存表的最小存储单位。
core文件	当程序出现内存越界、断言失败或者访问非法内存时, 操作系统会中止进程, 并将当前内存状态导出到core文件中, 以便进一步分析。 core文件包含内存转储, 支持全二进制和指定端口格式。core文件名称由字符串core以及操作系统进程ID组成。 core文件不依赖于任何平台。
Core Dump	通常在程序异常终止时, 核心转储 ( Core Dump )、内存转储或系统转储用于记录特定时间计算机程序工作内存的状态。实际上, 其它关键程序的状态经常在同一时间进行转储, 例如处理器寄存器, 包括程序指标和栈指针、内存管理信息、其它处理器和操作系统标记及信息。Core Dump经常用于辅助诊断和纠错计算机程序问题。
DBA	Database Administrator, 数据库管理员。指导或执行所有和维护数据库环境相关的操作。
DBLINK	DBLINK是定义一个数据库到另一个数据库路径的对象, 通过它可以查询远程数据库对象。
DBMS	Database Management System, 数据库管理系统。数据库管理系统是为了访问数据库中的信息而使用的一个管理系统软件。它包含一组程序使用户可以进入、管理、查询数据库中数据。基于真实数据的位置, 可以分为内存数据库管理系统和磁盘数据库管理系统。
DCL	Data Control Language, 数据控制语言。
DDL	Data Definition Language, 数据定义语言。

术语	解释
DML	Data Manipulation Language，数据操纵语言。
DN	Data Node，和CN对应的概念。负责实际执行表数据的存储、查询操作。
ETCD	Editable Text Configuration Daemon，分布式键值存储系统，用于共享配置和服务发现（服务注册和查找）。
ETL	Extract-Transform-Load，描述将数据从来源端经过抽取（extract）、转换（transform）、加载（load）至目的端的过程。
Extension Connector	Extension Connector是GaussDB(DWS)提供的功能模块，使用它可以 将SQL语句发送到集群外部的Spark，并在当前库中返回执行结果，实现跨集群处理数据。
备份	备份件或者备份过程。指复制并归档计算机数据，当发生数据丢失事件时，可以用该复制并归档的数据来恢复原始数据。
备份和恢复	保护数据库防止由于媒介失效或人为错误造成的数据丢失过程中涉及的一组概念、过程及策略。
备机	GaussDB(DWS)双机方案中的一个节点，用于作为主机的备份，在主机异常时，备机会切换到主机状态，以确保能正常提供数据服务。
崩溃	崩溃（或系统崩溃）指计算机或程序（例如软件应用程序或操作系统）异常终止的事件。出现错误后，通常会 自动退出。有时出现恶意程序冻结或挂起直到崩溃上报服务记录崩溃的详细信息。对于操作系统内核关键部分的程序，整个计算机可能瘫痪（可能造成致命的系统错误）。
编码	编码是指用代码来表示各组数据资料，使其成为可利用计算机进行处理和分析的信息。用预先规定的方法将文字、数字或其它对象编成数码，或将信息、数据转换成规定的电脉冲信号。
编码技术	呈现计算机软硬件识别的特定字符集数据的技术。
表	表是由行与列组合成的。每一列被当作是一个字段。每个字段中的值代表一种类型的数据。例如，一个表可能有3个字段：姓名、城市和国家。这个表就会有3列：一列代表姓名，一列代表城市，一列代表国家。表中的每一行包含3个字段的内容，姓名字段包含姓名，城市字段包含城市，国家字段包含国家。
表空间	包含表、索引、大对象、长数据等数据的逻辑存储结构。表空间在物理数据和逻辑数据间提供了抽象的一层，为所有的数据库对象分配存储空间。表空间创建好后，创建数据库对象时可以指定该对象所属的表空间。
并发控制	在多用户环境下同时执行多个事务并保证数据完整性的一个DBMS服务。并发控制是GaussDB(DWS)提供的一种多线程管理机制，用来保证多线程环境下在数据库中执行的操作是安全的和一致的。
查询	向数据库发出的信息请求，包含更新、修改、查询或删除信息的请求。

术语	解释
查询操作符	Query Operator，也称为查询迭代算子（Iterator）或查询节点（Query Tree Node）。一个查询的执行可以分解为一个或多个查询操作符，是构成一个查询执行的最基本单位。常见的查询操作符包括表扫描（Scan），表关联（Join），表聚集（Aggregation）等。
查询片段	每一个查询任务都可以分解成为一个或者多个查询片段。每个查询片段由一个或多个查询操作符构成，可独立在节点上运行。通过数据流操作符与其它查询片段块交换数据。
持久性	数据库事务的ACID特性之一。在事务完成以后，该事务对数据库所作的更改便持久的保存在数据库之中，并不会被回滚。
存储过程	存储过程（Stored Procedure）是在大型数据库系统中，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称并设置参数（如果该存储过程带有参数）来执行它。
操作系统	操作系统OS（operating system）由引导程序加载到计算中，对计算机中其它程序进行管理。其它程序叫做应用或应用程序。
从备	Secondary，为了保证集群的高可靠性，主、备间无法正常同步数据时，主节点会将日志同步到从备。如果主节点突然故障不可用，备节点会升主，并且升主成功后从从备节点上同步之前异常期间的日志。
大对象	大对象（Blob）在数据库中是指使用二进制方式存储的数据。它通常可以用于存储视频、音频和图像等多媒体数据。
动态负载	GaussDB(DWS)动态负载是指基于系统中CPU、I/O、内存等资源的使用情况，自动调节并发作业的运行数量，避免因为系统资源过载导致业务报错或无响应。
段	数据库中，一段指包含一个或多个区域的数据中的一部分。区域是数据库的最小范围，由单元调用块组成。一个或多个段组成一个表空间。
<b>F - J</b>	
Failover	指当某个节点出现故障时，自动切换到备节点上的过程。反之，从备节点上切换回来的过程称为Failback。
FDW	Foreign Data Wrapper，外部数据封装器。是Postgres提供的一个SQL接口，用于访问远程数据存储中的大数据对象，使DBA可以整合来自不相关数据源的数据，将它们存入数据库中的一个公共模型。

术语	解释
Freeze	在事务ID耗尽时由AutoVacuum Worker进程自动执行的操作。GaussDB(DWS)会把事务ID记在行头，在一个事务取得一行时，通过比较行头的事务ID和事务本身的ID判断这行是否可见，而事务ID是一个无符号整数，如果事务ID耗尽，事务ID会跨过整数的界限重新计算，此时原先可见的行就会变成不可见的行，为了避免这个问题，Freeze操作会将行头的事务标记为一个特殊的事务ID，标记了这个特殊的事务ID的行将对所有事务可见，以此避免事务ID耗尽产生的问题。
GDB	GNU工程调试器，可以监控其它程序运行时的内部情况，或者其它程序要崩溃时发生了什么。GDB支持如下四种主要操作（使PDK功能更加强大），辅助查找缺陷。 <ul style="list-style-type: none"> <li>• 启动程序，指定可能影响行为的任何因素。</li> <li>• 特定条件下，停止程序。</li> <li>• 程序停止时，检查发生了什么。</li> <li>• 修改程序内容，尝试纠正一个缺陷并继续下一个。</li> </ul>
GDS	General Data Service，数据并行加载工具。向GaussDB(DWS)导入数据时，需要将此工具部署到源数据所在的服务器上，使DN可以通过该工具获取数据。
GIN索引	Generalized Inverted Index，通用倒排索引。作用为处理索引项为组合值的情况，查询时需要通过索引搜索出出现在组合值中的特定元素值。
GNU	GNU计划，又称革奴计划，是由RichardStallman在1983年9月27日公开发起的。它的目标是创建一套完全自由的操作系统。GNU是“GNU's NotUnix”的递归缩写。Stallman宣布GNU应当发音为Guh-NOO以避免与new这个单词混淆（注：Gnu在英文中原意为非洲牛羚，发音与new相同）。Unix是一种广泛使用的商业操作系统的名称。技术上讲，GNU类似Unix。但是GNU却给了用户自由。
gsql	GaussDB(DWS)交互终端。通过gsql能够以交互的方式输入查询，下发查询到GaussDB(DWS)，然后查看查询结果。或者，也可以从文件中输入。此外，gsql还提供许多元命令和各种类似shell命令，协助脚本编写及自动化各种任务。
GTM	Global Transaction Manager，全局事务管理器。用于管理事务状态的部件。
GUC	Grand Unified Configuration，数据库运行参数。配置这些参数可以影响数据库系统的行为。
HA	高可用性（HighAvailability），通过尽量缩短因日常维护操作（计划）和突发的系统崩溃（非计划）所导致的停机时间，以提高系统和应用的可用性。
HBA	host-based authentication，主机认证。主机鉴权允许主机鉴权部分或全部系统用户。适用于系统所有用户或者使用Match指令的子集。该类型鉴权对于管理计算集群以及其它完全同质设备非常有用。总之，服务器上的三个文件以及客户端上的一个文件必须修改，为主机鉴权做准备。

术语	解释
HDFS	Hadoop Distributed File System, Apache Hadoop项目的一个子项目。一个高度容错的分布式文件系统, 设计用于在低成本硬件上运行。HDFS提供高吞吐量应用程序数据访问功能, 适合带有大型数据集的应用程序。
服务器	为客户端提供服务的软硬件的组合。单独使用时, 指运行服务器操作系统的计算机, 也可以指提供服务的软件或者专用硬件。
高级包	GaussDB(DWS)提供的具有一定逻辑和功能的存储过程、函数, 这些具备功能的存储过程、函数统称为高级包。
隔离性	数据库事务的ACID特性之一。它是指一个事务内部的操作及使用的数据对其它并发事务是隔离的, 并发执行的各个事务之间不能互相干扰。
关系型数据库	创建在关系模型基础上的数据库。关系型数据库借助于集合代数等数学概念和方法来处理数据库中的数据。
归档线程	数据库打开归档功能时启动的一个线程, 此线程用于将数据库日志归档到指定的路径。
故障接管	功能对等的系统部件对于故障部件的自动替换过程。系统部件包含处理器、服务器、网络、数据库等。
环境变量	定义进程操作环境某一方面的变量。例如, 环境变量可以为主目录, 命令搜索路径, 使用终端或当前时区。
检查点	将数据库内存中某一时刻的数据存到磁盘的机制。GaussDB(DWS)定期将已提交的事务数据和未提交的事务数据存到磁盘, 这些数据用来和Redo日志一起在数据库重启和崩溃时恢复数据库。
加密	用于传输数据的功能。通过该功能, 可以隐藏信息内容, 防止非法使用。
节点	将构成GaussDB(DWS)集群环境的各台服务器(物理机或虚拟机)称为集群节点, 简称节点。
纠错	系统自动识别软件和数据流上的错误并自动修正错误的能力, 提升系统的稳定性和可靠性。
进程	在单个计算机上执行程序的实例。一个进程由一个或多个线程组成。其它进程不能接入某个进程已占用的线程。
基于时间点恢复	PITR(Point-In-Time Recovery), 基于时间点恢复是GaussDB(DWS)备份恢复的一个特性, 是指在备份数据和WAL日志正常的情况下, 数据可以恢复到指定时间点。
记录	在关系型数据库中, 每一条记录对应表中的每一行数据。
集群	集群是由一组服务器和其它资源组成的一个单独的系统, 可以实现高可用性。有的情况下, 可以实现负载均衡及并行处理。
<b>K - O</b>	

术语	解释
LLVM	LLVM命名最早源自于底层虚拟机（Low Level Virtual Machine）的缩写。LLVM是构架编译器（compiler）的框架系统，以C++编写而成，用于优化以任意程序语言编写的程序的编译时间（compile-time）、链接时间（link-time）、运行时间（run-time）以及空闲时间（idle-time），对开发者保持开放，并兼容已有脚本。 GaussDB(DWS) LLVM动态编译技术可以为每个查询生成定制化的机器码用于替换原本的通用函数。通过减少实际查询时冗余的条件逻辑判断、虚函数调用并提高数据局域性，从而达到提升查询整体性能的目的。
LVS	Linux Virtual Server，虚拟服务器集群系统，用于负责集群的负载均衡。
MPP	Massive Parallel Processing，大规模并行处理。指利用多个机器构成集群的架构方式，也称为集群（Cluster）系统。
MVCC	Multi-Version Concurrency Control，多版本并发控制。数据库并发控制协议的一种，它的基本算法是一个元组可以有多个版本，不同的查询可以工作在不同的版本上。一个基本的好处是读和写可以不冲突。
NameNode	NameNode是Hadoop系统中的一个中心服务器，负责管理文件系统的名字空间(namespace)以及客户端对文件的访问。
OLAP	Online Analytical Processing，联机分析处理。是数据仓库系统最主要的应用，专门设计用于支持复杂的分析操作，侧重对决策人员和高层管理人员的决策支持，可以根据分析人员的要求快速、灵活地进行大数据量的复杂查询处理，并且以一种直观而易懂的形式将查询结果提供给决策人员，以便准确掌握企业（公司）的经营状况，了解对象的需求，制定正确的方案。
OM	Operations Management，运维管理模块。提供集群日常运维、配置管理的管理接口、工具。
ORC	Optimized Row Columnar，一种广泛使用的Hadoop系统结构化数据的文件格式，由Hadoop HIVE项目引入。
客户端	连接或请求其它计算机或程序服务的计算机或程序。
空闲空间管理	管理表内空闲空间的机制，通过记录每个表内空闲空间信息，并建立易于查找的数据结构，可以加速对空闲空间进行的操作（例如INSERT）。
跨集群	GaussDB(DWS)支持通过外表和Extension Connector访问当前GaussDB(DWS)集群外的其他DBMS中的数据，这种行为称为跨集群。
垃圾元组	是指使用DELETE和UPDATE语句删除的元组，GaussDB(DWS)在删除元组时只是打个删除标记，由Vacuum线程清理这种垃圾元组。
列	字段的等效概念。在数据库中，表由一列或多列组成。
逻辑节点	一个物理节点上可以安装多个逻辑节点。一个逻辑节点是一个数据库实例。

术语	解释
模式	数据库对象集，包括逻辑结构，例如表、视图、序、存储过程、同义名、索引、集群及数据库链接。
模式文件	用于决定数据库结构的SQL文件。
<b>P - T</b>	
Page	GaussDB(DWS)数据库关系对象结构中行存的最小存储单元。一个Page大小为默认为8KB。
PostgreSQL	PostgreSQL是一个开源的关系数据库管理系统(DBMS)，由全球志愿者团队开发。PostgreSQL不受任何公司或个体所控制，源代码免费使用。
Postgres-XC	一款多节点同步，读写可扩展的PostgreSQL集群数据库。
Postmaster	数据库服务启动时启动的一个线程。用于监听来自集群其它节点或客户端的连接请求。 主机上监听到备机连接请求，并接受后，就会创建一个WAL Sender线程，用于处理与备机的交互。
RHEL	Red Hat Enterprise Linux，红帽企业Linux。
REDO日志	记录对数据库进行操作的日志，这些日志包含重新执行这些操作所需要的信息。当数据库故障时，可以利用REDO日志将数据库恢复到故障前的状态。
SCTP	Stream Control Transmission Protocol，流控制传输协议。是IETF于2000年新定义的一个传输层协议。是提供基于不可靠传输业务的协议之上的可靠的数据报传输协议。SCTP的设计用于通过IP网传输SCN窄带信令消息。
Savepoint	保存点。是一种在关系数据库管理系统中实现子事务（也称为嵌套事务）的方法。在一个长事务中，可以把操作过程分成几部分，前面部分执行成功后，可以建一个保存点，若后面的执行失败，则回滚到这个保存点即可，无需回滚整个事务。保存点对于在数据库应用程序中实现复杂错误恢复很有用。如果在多语句事务中发生错误，则应用程序可能能够从错误中恢复（通过回滚到保存点）而无需中止整个事务。
Session	数据库系统在接收到应用程序的连接请求时，为该连接创建的一个任务。它被Session Manager管理，完成一些初始化任务，执行用户的所有操作。
Shared-nothing architecture	无共享架构是一种分布式计算架构，这种架构中不存在集中共享CPU、存储的状态，这种架构具有非常强的扩展性。
SLES	SUSE Linux Enterprise Server，由SUSE提供的企业级Linux操作系统。

术语	解释
SMP	Symmetric Multi-Processing, 对称多处理技术, 是指在一台计算机上汇集了一组处理器(多CPU), 各CPU之间共享内存子系统以及总线结构。操作系统必须支持多任务和多线程处理, 以使得SMP系统发挥高效的性能。数据库领域的SMP并行技术, 一般指利用多线程技术实现查询的并行执行, 以充分利用CPU资源, 从而提升查询性能。
SQL	Structure Query Language, 结构化查询语言。数据库的标准查询语言。它可以分为数据定义语言(DDL), 数据操纵语言(DML)和数据控制语言(DCL)。
SSL	Secure Socket Layer, 安全套接层。SSL是Netscape公司率先采用的网络安全协议。它是在传输通信协议(TCP/IP)上实现的一种安全协议, 采用公开密钥技术。SSL广泛支持各种类型的网络, 同时提供三种基本的安全服务, 它们都使用公开密钥技术。SSL支持服务通过网络进行通信而不损害安全性。它在客户端和服务器之间创建一个安全连接。然后通过该连接安全地发送任意数据量。
收敛比	交换机下行带宽与上行带宽的比值。收敛比越高, 流量收敛程度越大, 丢包越严重。
TCP	Transmission Control Protocol, 传输控制协议。用于将数据信息分解成信息包, 使之经过IP协议发送; 并对利用IP协议接收来的信息包进行校验并将其重新装配成完整的信息。TCP是面向连接的可靠协议, 能够确保信息的无误发送。
trace	一种特殊的日志记录方法, 用来记录程序执行的信息。程序员使用该信息进行纠错。另外, 根据trace日志中信息的类型和内容, 有经验的系统管理员或技术支持人员以及软件监控工具诊断软件常见问题。
全备份	备份整个数据库集群。
全量同步	GaussDB(DWS)双机方案中的一种数据同步机制, 是指把主机中的所有数据同步给备机。
日志文件	计算机记录自身活动的记录。
事务	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, 事务必须满足ACID原则。
数据	事实或指令的一种表达形式, 适用于人为或自动的通信、解释或处理。数据包含常量、变量、阵列和字符串。
数据重分布	用户改变数据的分布方式后, 数据表在节点间重新分布的过程。
数据分布	表数据在分布式环境中的分布方式(Distribution), 即数据表以何种方式打散存储到各个数据库实例上去。具体的分布方式可以有: 散列(Hash)方式, 复制方式(Replication)和随机方式(Random)。散列方式根据元组中指定字段的取值算得哈希值, 根据节点与哈希值的映射关系获得该元组的目标存储位置。复制方式将元组复制到所有节点上。随机方式将数据随机分布到各节点。



术语	解释
数据分区	数据分区是指在一个数据库实例内部，将表按照划分为多个数据互不重叠的部分（Partition）。具体的分区方式可以有：范围分区（Range），它根据元组中指定字段的取值所处的范围映射到目标存储位置。
数据库	数据库是存储在一起的相关数据的集合，这些数据可以被访问，管理以及更新。同一视图中，数据库可以根据存储内容类型分为以下几类：数目类、全文本类、数字类及图像类。
数据库实例	一个数据库实例是一个GaussDB(DWS)进程以及它控制的数据库文件。GaussDB(DWS)在一个物理节点上安装多个数据库实例，集群各节点上所安装的GTM、CM、CN、DN统称为实例。一个数据库实例也被称为一个逻辑节点。
数据库双机	GaussDB(DWS)提供的高可靠性双机方案。在此方案中，每个GaussDB(DWS)逻辑节点标识为主机或备机。在同一时间内，只有一个GaussDB(DWS)被标识为主机。双机初次建立时，主机会对每个备机数据做全量同步，然后做增量同步。双机建立之后的运行过程中，主机能接受数据读和写的操作请求，备机只做日志同步。
数据库文件	保存用户数据和数据库系统内部数据的二进制文件。
数据流操作符	负责查询片段间交换数据的操作符。根据数据流的输入、输出关系，可以细分为聚合流（Gather）、广播流（Broadcast）和重分布流（Redistribution）。聚合流将数据从多个查询片段聚合到一个。广播流将数据从一个查询片段向多个传输。重分布流则将多个查询片段的数据，按照一定规则重组后向多个传输。
数据字典	数据字典是一系列只读的表，用来提供数据库的信息。这些信息包括：数据库设计信息、存储过程信息、用户权限、用户统计数据、数据库进程信息、数据库增长统计数据和数据库性能统计数据。
死锁	为使用同一资源而产生的无法解决的争用状态。
索引	数据库索引，是数据库管理系统中一个排序的数据结构，以协助快速查询、更新数据库表中数据。
统计信息	数据库使用统计信息估算查询代价，以查找代价最小的执行计划，统计信息一般是数据库自动收集的，包括表级信息（元组数、页面数等）和列级信息（列的值域分布直方图）。
停用词	在信息检索中，为节省存储空间和提高搜索效率，在处理自然语言数据（或文本）之前或之后会自动过滤掉某些字或词，这些字或词即被称为Stop Words（停用词）。
<b>U - Z</b>	
Vacuum	数据库定期启动的清理垃圾元组的线程，根据配置参数可以同时启动多个。
verbose	verbose选项指定显示在屏幕上的处理信息。
WAL	Write-Ahead Logging，预写日志系统。实现事务日志的标准方法，是指对数据文件（表和索引的载体）持久化修改之前必须先持久化相应的日志。

术语	解释
WAL Receiver	数据库复制时备机创建的一个线程的名称。此线程用于从主机接收数据、命令，并反馈确认信息至主机。一个备机只有一个WAL Receiver线程。
WAL Sender	数据库复制过程中，主机接受到备机的连接请求后创建的一个线程的名称。此线程用于发送命令、数据到备机，并从备机接收信息。一个主机可能会有多个WAL Sender线程，每一个WAL Sender线程对应一个备机的一个连接请求。
WAL Writer	数据库启动时创建的一个写Redo日志的线程，用于将内存中的日志写入到持久性设备（如：磁盘）。
WLM	WorkLoad Manager，负载管理。GaussDB(DWS)中系统资源控制和分配的模块。
Xlog	表示事务日志，一个逻辑节点中只有一个，不允许创建多个Xlog文件。
xDR	详单。用户面和信令面详单的统称，包括CDR和UFDR、TDR和SDR。
网络备份	网络备份为各种平台提供一套完整的、灵活的数据保护方案。平台包含Microsoft Windows、UNIX及Linux。网络备份支持备份、归档、恢复计算机上的文件、文件夹或目录、卷或分区。
物理节点	一个物理机器称为一个物理节点。
系统表	存储数据库元信息的表，元信息包括数据库中的用户表、索引、列、函数和数据类型等。
下推	GaussDB(DWS)是分布式数据库，其可以利用多DN分布式并行执行查询计划，即将CN中的查询计划下发到各DN中并行执行。这种行为称为下推。与将数据抽取到CN上执行查询的方式相比，下推可以大幅提升查询性能。
压缩	数据压缩，信源编码，或比特率降低涉及使用相比原来较少比特的编码信息。压缩可以是有损或无损。无损压缩通过识别和消除统计冗余降低比特位。无损压缩中没有信息丢失。有损压缩识别并删除次要信息，减少了比特位。减少数据文件大小的方法被普遍称为数据压缩，尽管其正式名称为源编码（数据源的编码，然后将其存储或传输）。
一致性	数据库事务的ACID特性之一。在事务开始之前和事务结束以后，数据库的完整性约束没有被破坏。
元数据	用来定义数据的数据。主要是描述数据自身信息，包含源、大小、格式或其它数据特征。数据库字段中，元数据用于理解以及诠释数据仓库的内容。
原子性	数据库事务的ACID特性之一。整个事务中的所有操作，要么全部完成，要么全部不完成，不可能停滞在中间某个环节。事务在执行过程中发生错误，会被回滚到事务开始前的状态，就像这个事务从来没有执行过一样。

术语	解释
在线扩容	在线扩容是指GaussDB(DWS)扩容重分布过程中支持数据持续入库、查询业务不中断。
脏页面	已经被修改且未写入持久性设备的页面。
增量备份	基于上次有效备份之后对文件修改的备份。
增量同步	GaussDB(DWS)双机方案中的一种数据同步机制，是指把主机中数据增量同步给备机，即只同步主备间有差异的数据。
主机	GaussDB(DWS)数据库双机系统中接受数据读写操作的节点，和所有备机一起协同工作。在同一时间内，双机系统中只有一个节点被标识为主机。
主题词	在标引和检索中用以表达文献主题的规范化的词或词组。
转储文件	转储文件是一种特定类型的trace文件。转储文件为响应事件过程中一次性输出的诊断数据，trace文件指诊断数据的连续输出。
资源池	资源池是GaussDB(DWS)提供的一种资源划分的配置机制。通过将用户绑定到资源池，来限定其所执行作业的优先级及能够利用到的资源。
租户	数据库业务用户在给定的计算资源（cpu，内存和io）和存储资源下执行业务，通过资源管理和隔离，达成业务的服务等级协定（SLA）。
最小恢复点	最小恢复点是GaussDB(DWS)提供的数据库一致性保障手段之一。最小恢复点特性可以在GaussDB(DWS)启动时检查出WAL日志和持久化到磁盘的数据的不一致性，并提示用户进行处理。