

表格存储服务

产品介绍

文档版本 01
发布日期 2024-04-30



版权所有 © 华为技术有限公司 2024。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。

华为技术有限公司

地址： 深圳市龙岗区坂田华为总部办公楼 邮编： 518129

网址： <https://www.huawei.com>

客户服务邮箱： support@huawei.com

客户服务电话： 4008302118

安全声明

漏洞处理流程

华为公司对产品漏洞管理的规定以“漏洞处理流程”为准，该流程的详细内容请参见如下网址：

<https://www.huawei.com/cn/psirt/vul-response-process>

如企业客户须获取漏洞信息，请参见如下网址：

<https://securitybulletin.huawei.com/enterprise/cn/security-advisory>

目录

1 什么是 CloudTable.....	1
2 应用场景.....	3
2.1 HBase 应用场景.....	3
2.2 ClickHouse 应用场景.....	4
3 集群管理功能简介.....	5
3.1 HBase 集群管理.....	5
3.2 ClickHouse 集群管理.....	6
4 与其他云服务的关系.....	8
5 CloudTable 权限管理.....	10
6 使用限制.....	12
7 配额说明.....	13

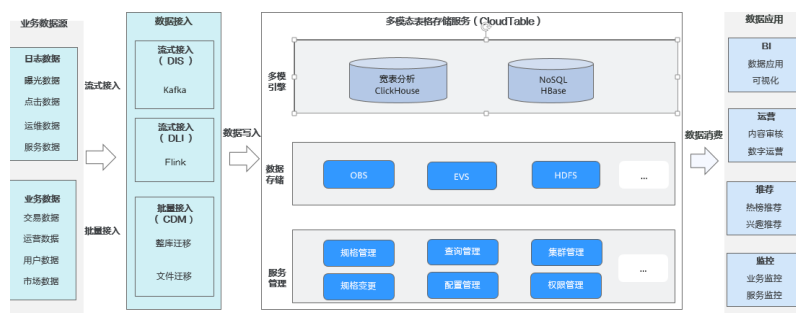
1 什么是 CloudTable

表格存储服务（CloudTable）是基于HBase、ClickHouse提供的全托管数据存储和分析服务。CloudTable提供GB~PB级数据存储和分析能力，用于在线查询、数据仓库、数据集市、实时分析等场景。可被广泛应用于互联网、物联网、车联网、金融、政府、物流、制造、零售等行业。

CloudTable 产品架构

CloudTable产品架构如下图所示：

图 1-1 产品结构



- ClickHouse：万列大宽表多维聚合分析，亚秒级响应，全自助分析。
- HBase：高并发，毫秒级查询响应。

产品优势

- 丰富场景：兼容HBase、ClickHouse等多种引擎。
- 高可靠：架构高可用，内核深度优化，提升系统稳定性。
- 高性价比：支持冷热分离，不同压缩算法，存储成本低。
- 简单易用：通过控制台分钟级构建分析集群，提供完善的集群运维管理、监控告警等功能，使您无需关注底层基础设施，利用完善的SQL语句支持，专注于数据价值的分析。

首次使用 CloudTable

如果您是首次使用CloudTable的用户，建议您学习并了解如下信息：

- 基础知识了解
通过CloudTable产品功能章节的内容，了解CloudTable相关的基础知识，包含CloudTable各组件的基本原理和场景介绍，以及CloudTable服务的特有概念和功能的详细介绍。
- 入门使用
您可以参考《快速入门》学习并上手使用CloudTable。《快速入门》提供了样例的详细操作指导，您可以基于此操作指导，创建和使用CloudTable集群。
- 使用更多的功能，并查看其相关操作指导
如果您是一个CloudTable集群使用人员，可以参考用户指南完成集群创建、参数配置、查看告警等操作。
如果您是一个开发者，可以参考CloudTable提供的开发指南操作指导及样例工程开发并运行调测自己的应用程序。您也可以通过API调用完成CloudTable集群创建/查询操作，您可以参考《API参考》获取详情。

2 应用场景

2.1 HBase 应用场景

消息日志类数据存储和查询

应用场景：

消息数据、报表数据、推荐类数据、风控类数据、日志数据、订单数据等结构化、半结构化的Key-Value数据均可以存储和查询。

优势：

- 海量存储
支持离线、在线海量Key-Value数据存储，存储容量可扩展。
- 高性能读写
亿级写入吞吐量、ms级查询，用于在线应用和报表展现。
- 生态丰富
基于Hadoop生态组件丰富，与华为云产品有高度的整合能力。

画像数据存储和查询

应用场景：

画像通常用一些标签来刻画自然人/物的特征，而每一个自然人/物所拥有的标签集合是不确定的，数据更新非常频繁，这类数据被广泛应用于市场决策、推荐以及广告系统中。

优势

- 稀疏矩阵
HBase的稀疏矩阵模型，天然适合非结构化数据的存储，数据表无需预先定义schema，行与行之间不需要严格的列定义。
- 支持任意更新
支持行的任意更新，无性能损耗。而且利用HBase自身的多版本机制，支持保存数据的多个历史版本。

通用海量 KeyValue 数据存储与查询

- 应用类型
消息数据、报表数据、推荐类数据、风控类数据、日志数据、订单数据等结构化、半结构化的KeyValue数据均可以存储和查询。
- 应用场景
支持离线、在线海量KeyValue数据高速写入，并提供低延时的查询服务，用于在线应用或者报表展现等。扩展简单，实现高可用、低时延的海量数据存储和查询。

2.2 ClickHouse 应用场景

ClickHouse是Click Stream+Data WareHouse的缩写，起初应用于一款Web流量分析工具，基于页面的单击事件流，面向数据仓库进行OLAP分析。当前ClickHouse被广泛的应用于互联网广告、App和Web流量、电信、金融、物联网等众多领域，非常适用于商业智能化应用场景，在全球有大量的应用和实践。

应用场景

- 用户行为分析。
在网站、App和游戏中，对用户的点击、浏览时长等使用数据进行收集，导入到云数据仓库ClickHouse中，构建用户特征分析大宽表。利用云ClickHouse的优异查询性能，分析系统进行多维度、多模式分析时，可以在亚秒级内响应，快速分析出用户行为特征和规律，为精准营销和会员转化等业务提供有效支持。
- 企业经营分析。
在企业经营分析中，把规模庞大的业务数据导入到云数据仓库ClickHouse，对数亿记录或更大规模的大宽表和数百维度的查询，都能在亚秒级内响应，得到查询结果。让客户随时进行个性化统计和不间断的分析，辅助商业决策。
- 访客来源分析展示。
通过批量离线计算对用户访问日志中的用户行为进行关联，生成用户行为路径大宽表同步到ClickHouse，基于ClickHouse构建交互式访客来源探索分析可视化系统。
- BI报表。
利用ClickHouse构建实时交互查询报表，实时分析订单、收入、用户数等核心业务指标。
- 用户分群统计。
构建用户信息表，实时选择用户属性标签数据和筛选条件，通过大量的数据记录进行人群特征统计分析。

3 集群管理功能简介

3.1 HBase 集群管理

HBase 产品简介

HBase是一个稳定可靠，性能卓越、可伸缩、面向列的分布式云存储系统，适用于海量数据存储以及分布式计算的场景，用户可以利用HBase搭建起TB至PB级数据规模的存储系统，对数据轻松进行过滤分析，毫秒级得到响应，快速发现数据价值。

HBase适用场景有：

- 海量数据存储。
适用于TB ~ PB级以上的数据存储，提供动态伸缩能力，方便用户在性能或容量需要改变时，改变集群资源，轻松构建企业海量数据存储系统。
- 实时查询。
HBase的列式KeyValue存储机制，适用于企业用户明细数据即时查询，基于主键的低时延点查，响应时延一般为秒级或毫秒级，方便用户对数据的实时分析。

HBase的架构和详细原理介绍，请参见：<https://hbase.apache.org/book.html>

📖 说明

当前CloudTable HBase暂无安全认证机制，如果需使用认证鉴权机制的HBase服务，建议使用华为云MapReduce服务。

集群管理功能

CloudTable服务是华为云提供的一项分布式、可扩展的KeyValue数据存储服务。CloudTable提供Web官网服务界面，CloudTable HBase集群管理的具体功能如下：

- 创建集群：在CloudTable服务界面完成集群的创建。支持按照用户创建集群时选择的计算单元个数、实际使用存储容量计费。用户可以自主地选择服务提供的高级特性，独立安装、独立收费。扣费时余额不足先提醒用户续费，在保留期冻结集群资源，续费后再解冻。通过计算存储分离架构和计算资源动态调整，尽可能的为用户降低成本。
- 扩容集群：支持集群扩容计算单元。

扩容计算单元：用户根据实际需要或业务情况，动态的增加计算单元的个数，保证读写性能。集群自适应的实现负载均衡，保证业务不中断，平滑扩容。扩容计算单元将会产生额外的费用。

- 管理集群：对创建的集群进行管理。
 - 指标监控：集群运行中，收集各项监控数据，上报云监控（Cloud Eye），向用户以图形化的方式呈现集群运行状况。当指标出现异常时，通过消息通知用户和管理员，及时人工介入。
 - 删除集群：当用户不再需要集群时，可选择删除集群。此操作为高危操作，删除集群可能导致数据丢失，因此在执行删除操作之前，请确认不存在正在运行的业务，所有数据都已经保存。
 - 重启集群：当修改完集群HBase参数后，或者因长时间不重启导致系统运行缓慢等情况下，用户需要重启集群。重启操作将有可能导致正在运行中的业务数据丢失，如果需要执行重启操作，请确定不存在正在运行的业务，所有数据都已经保存。
 - 查询告警：集群运行异常或系统故障时，CloudTable服务会收集故障信息并上报网管系统，维护人员可根据用户提供的告警信息定位问题原因。
 - 日志查询：记录用户对集群操作信息，便于集群运行异常时定位分析问题原因。

产品优势

- HBase原生接口：兼容原生HBase接口，架构高可用，计算和存储分离保证高可靠，内核深度优化。
- 易使用：支持二级索引，满足数据非主键查询需求，简单易用。
- 低成本：支持冷热分离，满足数据归档、访问频率较低历史数据存储需求，降低存储成本。
- 稳定可靠：支持热点诊断和自愈，提升系统稳定性。
- 可视化监控运维：提供常用的监控信息和自定义告警规则，简化系统运维。

3.2 ClickHouse 集群管理

ClickHouse 产品介绍

ClickHouse为您提供方便易用、灵活稳定的云端ClickHouse托管服务。只需要几分钟，便可完成海量数据查询数据仓库的搭建，简单轻松地完成对数据的实时查询分析，提升数据价值挖掘的整体效率。云数据仓库ClickHouse是一种基于MPP（大规模并行处理）架构的数仓服务，基于ClickHouse优异的查询性能，查询效率数倍于传统数据仓库。

ClickHouse 集群管理功能

- 集群创建：在CloudTable集群管理界面完成集群的创建。
- 集群查看：在CloudTable集群管理界面可以查看集群详情。
- 重启集群：因长时间不重启导致系统运行缓慢等情况下，用户需要重启集群。重启操作将有可能导致正在运行中的业务数据丢失，如果需要执行重启操作，请确定不存在正在运行的业务，所有数据都已经保存。

- 删除集群：当用户不再需要集群时，可选择删除集群。此操作为高危操作，删除集群可能导致数据丢失，因此在执行删除操作之前，请确认不存在正在运行的业务，所有数据都已经保存。

ClickHouse 产品优势

- 性能优势：ClickHouse采用列存储，相同列的数据属于同一类型，有利于获得更高的数据压缩比。通常可以达到10: 1的压缩比，大幅降低存储成本和读取开销，提高查询性能。
- 副本机制：ClickHouse利用Zookeeper，通过ReplicatedMergeTree引擎（Replicated系列引擎）实现了副本机制。用户在创建表时，可指定存储引擎，选择该表是否进行复制。
- 简单易用：通过控制台分钟级构建ClickHouse分析集群，使您无需关注底层基础设施，利用完善的SQL语句支持，专注于数据价值的分析。
- 极致性能：使用分布式大规模并行处理MPP框架，并充分利用所有可用的硬件，以尽可能快地处理每个查询。查询效率数倍于传统数据仓库，单个查询的峰值处理性能高达每秒数TB。
- 安全可靠：用户集群独立部署，支持VPC私有网络隔离，数据访问安全多重保障。
- 成本更低：利用云端高性价比设备，构建极具成本优势的托管ClickHouse集群。

4 与其他云服务的关系

统一身份认证 IAM

CloudTable使用统一身份认证服务（Identity and Access Management，简称IAM）实现认证功能。

了解更多IAM的相关信息，请参见[统一身份认证服务](#)。

弹性云服务器 ECS

CloudTable使用弹性云服务器（Elastic Cloud Server，简称ECS）作为服务实例的节点，每个弹性云服务器是服务实例中的一个节点。

了解更多ECS的相关信息，请参见[弹性云服务器](#)。

虚拟私有云 VPC

CloudTable使用虚拟私有云（Virtual Private Cloud，简称VPC）为集群提供网络拓扑，实现多个不同集群互相隔离并控制访问。

了解更多VPC的相关信息，请参见[虚拟私有云](#)。

对象存储服务 OBS

CloudTable使用对象存储服务（Object Storage Service，简称OBS）存储集群数据的备份和快照，实现安全、高可靠和低成本存储需求。

了解更多OBS的相关信息，请参见[对象存储服务](#)。

云数据迁移 CDM

CloudTable使用云数据迁移（Cloud Data Migration，简称CDM），可以将云上云下或第三方云上的多种数据源的数据迁移到CloudTable集群的HBase表中。

了解更多CDM的相关信息，请参见[数据迁移](#)。

云审计服务 CTS

CloudTable使用云审计服务（Cloud Trace Service，简称CTS）为用户提供CloudTable资源操作请求及请求结果的操作记录，供用户查询、审计和回溯使用。

了解更多CTS的相关信息，请参见[云审计服务](#)。

云监控服务 CES

CloudTable使用云监控服务CES监控集群中的多项性能指标，从而集中高效地呈现状态信息。云监控支持发送自定义告警，用户可第一时间获取异常通知。

了解更多CES的相关信息，请参见[云监控服务](#)。

5 CloudTable 权限管理

如果您需要对华为云上购买的CloudTable资源，为企业中的员工设置不同的访问权限，以达到不同员工之间的权限隔离，您可以使用统一身份认证服务（Identity and Access Management，简称IAM）进行精细的权限管理。该服务提供用户身份认证、权限分配、访问控制等功能，可以帮助您安全的控制华为云资源的访问。

通过IAM，您可以在华为云账号中给员工创建IAM用户，并授权控制他们对华为云资源的访问范围。例如您的员工中有负责软件开发的人员，您希望他们拥有CloudTable的使用权限，但是不希望他们拥有删除集群等高危操作的权限，那么您可以使用IAM为开发人员创建用户，通过授予仅能使用CloudTable，但是不允许删除集群的权限，控制他们对CloudTable资源的使用范围。

如果华为云账号已经能满足您的要求，不需要创建独立的IAM用户进行权限管理，您可以跳过本章节，不影响您使用CloudTable服务的其它功能。

IAM是华为云提供权限管理的基础服务，无需付费即可使用，您只需要为您账号中的资源进行付费。关于IAM的详细介绍，请参见《[IAM产品介绍](#)》。

CloudTable 权限

默认情况下，管理员创建的IAM用户没有任何权限，需要将其加入用户组，并给用户组授予策略或角色，才能使得用户组中的用户获得对应的权限，这一过程称为授权。授权后，用户就可以基于被授予的权限对云服务进行操作。

CloudTable部署时通过物理区域划分，为项目级服务。授权时，“作用范围”需要选择“区域级项目”，然后在指定区域对应的项目中设置相关权限，并且该权限仅对此项目生效；如果在“所有项目”中设置权限，则该权限在所有区域项目中都生效。访问CloudTable时，需要先切换至授权区域。

如表5-1所示，包括了CloudTable的所有系统角色。由于华为云各服务之间存在业务交互关系，表格存储服务的角色依赖其他服务的角色实现功能。因此给用户授予表格存储服务的角色时，需要同时授予依赖的角色，表格存储服务的权限才能生效。

表 5-1 CloudTable 系统角色

系统角色	描述	类别	依赖关系
cloudtable Administrator	表格存储服务的管理员权限。	系统角色	依赖Tenant Guest和Server Administrator角色，在同项目中勾选依赖的角色。

表5-2列出了CloudTable常用操作与系统权限的授权关系，您可以参照该表选择合适的系统权限。

表 5-2 CloudTable 操作与系统策略关系

模式	操作	cloudtable Administrator
集群模式	创建集群	√
	重启集群	√
	扩容集群	√
	删除集群	√
	参数配置	√
	查看CloudTable服务集群列表、 集群详细信息	√
	查看监控信息	√
	查看审计日志	√

相关链接

- [IAM产品介绍](#)
- [创建用户并授权使用CloudTable](#)

6 使用限制

使用CloudTable前，您需要认真阅读并了解以下使用限制。

- CloudTable集群必须创建在VPC子网内。
- 建议使用支持的浏览器登录表格存储服务。
 - Google Chrome: 36.0及更高版本
 - Mozilla FireFox: 35.0及更高版本
 - Internet Explorer: 9.0及更高版本

当使用Internet Explorer 9.0时可能无法登录CloudTable管理控制台，原因是某些Windows系统例如Win7旗舰版，默认禁止Administrator用户，Internet Explorer在安装时自动选择其他用户如System用户安装，从而导致Internet Explorer无法打开登录页面。请使用管理员身份重新安装Internet Explorer 9.0或更高版本（建议），或尝试使用管理员身份运行Internet Explorer 9.0。

 - Microsoft edge: 13.0及更高版本
- 创建CloudTable集群时，会为您创建一个默认安全组，请勿随意删除或更改默认安全组。删除或更改默认安全组可能导致集群异常，影响CloudTable集群的使用。
- CloudTable集群使用的安全组请勿随意放开权限，避免被恶意访问。
- CloudTable不会保存您设置的登录节点的初始密码，请您设置并保管好密码。为避免被恶意攻击，建议设置复杂度高的密码。
- 集群节点仅用于运行CloudTable集群，其他客户端应用程序、用户业务程序建议申请独立弹性云服务器部署。
- 当前CloudTable暂无安全认证机制，如果需使用认证鉴权机制的HBase服务，建议使用华为云MapReduce服务。
- ClickHouse创建表声明：为了保证ClickHouse数据的可靠性，建议使用Replicated系列引擎，因单副本引发的数据丢失、数据损坏等问题，云服务不承担任何责任。

7 配额说明

表格存储服务用户默认的基础资源如下：

- 实例数
- CPU数量
- 内存容量(GB)
- 磁盘数
- 磁盘容量(GB)

上述资源的配额查看及修改，请参见[关于配额](#)。