

腾讯云数据仓库 TCHouse-C

词汇表

产品文档



腾讯云

【版权声明】

©2013-2024 腾讯云版权所有

本文档著作权归腾讯云单独所有，未经腾讯云事先书面许可，任何主体不得以任何形式复制、修改、抄袭、传播全部或部分本文档内容。

【商标声明】

及其它腾讯云服务相关的商标均为腾讯云计算（北京）有限责任公司及其关联公司所有。本文档涉及的第三方主体的商标，依法由权利人所有。

【服务声明】

本文档意在向客户介绍腾讯云全部或部分产品、服务的当时的整体概况，部分产品、服务的内容可能有所调整。您所购买的腾讯云产品、服务的种类、服务标准等应由您与腾讯云之间的商业合同约定，除非双方另有约定，否则，腾讯云对本文档内容不做任何明示或模式的承诺或保证。

词汇表

最近更新时间：2024-01-19 16:39:32

本地表

本地表（Local Table）的数据，只会存储在当前写入的节点上，不会被分散到多台机器上。

CDW

参见 [云数据仓库](#)。

单机表

单机表（Non-Replicated Table）的数据，只会存储在当前节点上，不会被复制到其它节点。

分布式表

分布式表（Distributed Table）是本地表的集合，它将多个本地表抽象为一张统一的表，对外提供写入、查询功能。当写入分布式表时，数据会被自动分发到集合中的各个本地表中；当查询分布式表时，集合中的各个本地表会被分别查询，并且把最终结果汇总后返回。分布式表的写入和查询可以利用多台服务器的存储、计算资源，能按需横向扩展。

分片

在 Elasticsearch Service 中，分片（Shards）是数据的容器，文档保存在分片内，一个分片是一个底层的工作单元，它仅保存了全部数据中的一部分，分片又被分配到集群内的各个节点里，当您的集群规模扩大或者缩小时，Elasticsearch 会自动的在各节点中迁移分片，使得数据仍然均匀分布在集群里。

一个分片可以是主分片或者副本分片。索引内任意一个文档都归属于一个主分片，所以主分片的数目决定着索引能够保存的最大数据量（技术上来说，一个主分片最大能够存储 `Integer.MAX_VALUE - 128` 个文档）。

一个副本分片（Replicas）只是一个主分片的拷贝。副本分片作为硬件故障时保护数据不丢失的冗余备份，并为搜索和返回文档等读操作提供服务。在索引建立的时候就已经确定了主分片数，但是副本分片数可以随时修改。

在云数据仓库中，将海量数据分散存储到多个节点上，每个节点只存储和处理海量数据的一部分，每台节点被称为一个分片（Shard）。

副本

在消息队列 CKafka 中，副本（Replica）是消息的冗余备份，每个分区可以有多个副本，每个副本包含的消息是一样的（在同一时刻，副本之间并不完全一样，这依赖同步机制）。

在消息队列 CKafka 中每个分区至少有双副本，保障服务的高可用。

在云数据仓库中，为了保障服务的高可用性，ClickHouse 提供了副本机制，将单个节点的数据冗余存储在2个或多个节点上。

复制表

复制表（Replicated Table）的数据，会自动复制到多个节点上，形成多副本，复制表在至少1个正常副本情况下，仍可对外提供服务。

高可用

高可用（High Availability, HA）是 ClickHouse 集群的服务属性，默认为开启状态。

高可用属性开启时，ClickHouse 集群由偶数个（最少2个）ClickHouse Server 节点和3个 ZooKeeper 节点构成，单台服务器的数据冗余存储在2台或多台服务器上，当某个副本不可用时，另一个集群还可以继续服务。

高可用属性关闭时，ClickHouse 集群由多个（最少1个）ClickHouse Server 节点和1个 ZooKeeper 节点构成，数据只有1个副本，当某个副本不可用时会导致整个集群不可用。

云数据仓库

云数据仓库（Cloud Data Warehouse, CDW）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。云数据仓库 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。