

Cloud Data Warehouse

Glossary

Product Documentation



Copyright Notice

©2013-2024 Tencent Cloud. All rights reserved.

Copyright in this document is exclusively owned by Tencent Cloud. You must not reproduce, modify, copy or distribute in any way, in whole or in part, the contents of this document without Tencent Cloud's the prior written consent.

Trademark Notice



All trademarks associated with Tencent Cloud and its services are owned by Tencent Cloud Computing (Beijing) Company Limited and its affiliated companies. Trademarks of third parties referred to in this document are owned by their respective proprietors.

Service Statement

This document is intended to provide users with general information about Tencent Cloud's products and services only and does not form part of Tencent Cloud's terms and conditions. Tencent Cloud's products or services are subject to change. Specific products and services and the standards applicable to them are exclusively provided for in Tencent Cloud's applicable terms and conditions.

Glossary

Last updated : 2024-01-19 16:38:27

Local table

The data in a local table is stored only on the node where it is currently written to.

CDW

For more information, see [Cloud Data Warehouse](#).

Non-replicated table

The data in a non-replicated table is stored only on the current node but not replicated to other nodes.

Distributed table

A distributed table abstracts multiple local tables into one with write and query features. When you write data to a distributed table, the data will be automatically distributed to the local tables it contains; when you query data in it, all its local tables will be queried to return an aggregated result. You can leverage the storage and computing resources of multiple servers for writes and queries in a distributed table and perform on-demand horizontal scaling.

Shard

In ES, a shard is a data container. Documents are stored in shards, each of which is an underlying work unit that stores only a part of the data. Shards are assigned to nodes in a cluster. If the cluster is scaled out or in, ES will automatically migrate shards among nodes to keep the data evenly distributed in the cluster.

A shard can be either a primary shard or a replica shard. Any document in an index belongs to a primary shard, so the number of primary shards determines the maximum volume of data that can be stored in the index. Technically, a primary shard can hold up to 128 documents (`Integer.MAX_VALUE`).

A replica shard is just a copy of a primary shard. It acts as a redundant backup that protects data from loss when hardware fails and serves read operations such as searching for and returning documents. The number of primary shards is configured during index creation, but the number of replica shards can be modified at any time.

In CDW, massive amounts of data are stored on multiple nodes, each of which only stores and processes a part of the data. Each node is called a shard.

Replica

In CKafka, a replica is a redundant message backup. Each partition can have multiple replicas that have the same messages (this is subject to the sync mechanism as replicas may not be exactly the same at the same time).

Each partition in CKafka has at least two replicas to ensure high service availability.

In CDW, to guarantee a high service availability, ClickHouse offers the replica mechanism to store the data of a single node on two or more nodes in a redundant manner.

Replicated table

The data in a replicated table is automatically replicated to multiple nodes to create multiple replicas. As long as one replica is normal, the service of the replicated table will not be interrupted.

High availability

High availability (HA) is a service attribute of a ClickHouse cluster and is enabled by default.

If HA is enabled, the ClickHouse cluster consists of an even number of (at least two) ClickHouse Server nodes and three ZooKeeper nodes, with data of a single server stored on two or more servers to ensure redundancy and service continuity in case of replica failures.

If HA is disabled, the ClickHouse cluster consists of one or more ClickHouse Server nodes and one ZooKeeper node. In this setup, data is stored in only one replica, so when a replica becomes unavailable, the entire cluster will fail.

Cloud Data Warehouse

Cloud Data Warehouse (CDW) offers easy-to-use, flexible, and stable ClickHouse hosting services in the cloud. A data warehouse can be created in minutes for massive real-time data query and analysis, thereby improving the overall efficiency of data value mining. By leveraging the massively parallel processing (MPP) architecture, it can query data several times faster than traditional data warehouses.