

CloudTable Service

Service Overview

Issue 01
Date 2024-04-30



Copyright © Huawei Technologies Co., Ltd. 2024. All rights reserved.

No part of this document may be reproduced or transmitted in any form or by any means without prior written consent of Huawei Technologies Co., Ltd.

Trademarks and Permissions



HUAWEI and other Huawei trademarks are trademarks of Huawei Technologies Co., Ltd.

All other trademarks and trade names mentioned in this document are the property of their respective holders.

Notice

The purchased products, services and features are stipulated by the contract made between Huawei and the customer. All or part of the products, services and features described in this document may not be within the purchase scope or the usage scope. Unless otherwise specified in the contract, all statements, information, and recommendations in this document are provided "AS IS" without warranties, guarantees or representations of any kind, either express or implied.

The information in this document is subject to change without notice. Every effort has been made in the preparation of this document to ensure accuracy of the contents, but all statements, information, and recommendations in this document do not constitute a warranty of any kind, express or implied.

Huawei Technologies Co., Ltd.

Address: Huawei Industrial Base
Bantian, Longgang
Shenzhen 518129
People's Republic of China

Website: <https://www.huawei.com>

Email: support@huawei.com

Security Declaration

Vulnerability

Huawei's regulations on product vulnerability management are subject to the *Vul. Response Process*. For details about this process, visit the following web page:

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

For vulnerability information, enterprise customers can visit the following web page:

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

Contents

1 What Is CloudTable.....	1
2 Application Scenarios.....	3
2.1 HBase Application Scenarios.....	3
2.2 ClickHouse Application Scenarios.....	4
3 Cluster Management Functions.....	6
3.1 HBase Cluster Management.....	6
3.2 ClickHouse Cluster Management.....	8
4 Related Services.....	10
5 Permissions Management.....	12
6 Use Restrictions.....	14
7 Quota Description.....	15

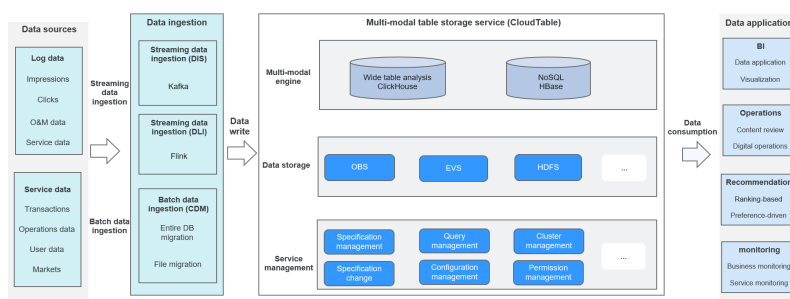
1 What Is CloudTable

CloudTable is a fully managed data storage and analysis service based on HBase and ClickHouse. It provides data storage and analysis capabilities ranging from gigabytes to petabytes, catering to scenarios such as online queries, data warehousing, data marts, and real-time analysis. Widely applicable across industries, CloudTable serves the Internet, Internet of Things (IoT), Internet of Vehicles (IoV), finance, government, logistics, manufacturing, and retail sectors.

Architecture

The following figure shows the architecture of CloudTable.

Figure 1-1 Architecture



- ClickHouse: supports multidimensional aggregation analysis for large and wide tables (with 10,000 columns), subsecond-level response, and full self-service analysis.
- HBase: supports high concurrency and millisecond-level query response.

Advantages

- Wide applicability: CloudTable is compatible with multiple engines (such as HBase, and ClickHouse).
- High reliability: The architecture is highly available and the kernel is optimized to improve system stability.
- Cost-effectiveness: Cold and hot data can be segregated, utilizing different compression algorithms to reduce storage costs.

- Easy to use: Analysis clusters can be swiftly established through the console within minutes. The platform offers comprehensive cluster management, monitoring, and alarm reporting functions. With robust SQL statement support, you can focus on extracting value from your data without concerning yourself with the underlying infrastructure.

New to CloudTable

If you are new to CloudTable, consider exploring the following information:

- Basic knowledge
This section covers the fundamental principles and scenarios of CloudTable components, along with specific concepts and functions related to CloudTable.
- Getting started
The Getting Started part provides detailed operational guidance using practical samples. You can create and utilize CloudTable clusters based on this guide.
- Advanced features and operation guides
If you are an existing CloudTable cluster user, consult the User Guide to create clusters, configure parameters, and monitor alarms.
Developers can refer to the Development Guide, which includes sample projects for building, running, and deploying custom applications. You can also call APIs to create or query clusters. For details, see API Reference.

2 Application Scenarios

2.1 HBase Application Scenarios

Storage and Query of Message Logs

Application Scenarios:

Structured and semi-structured key-value data can be stored and queried, including messages, reports, recommendation data, risk control data, logs, and orders.

Advantages:

- Mass storage
Offline and online storage of massive volumes of key-value data, and flexible capacity expansion
- High-performance read/write
100-million-level write throughput, millisecond-level query latency for presenting online applications and reports
- Enriched ecosystem
A large number of Hadoop ecosystem components, integrated with HUAWEI CLOUD products

Profile Storage and Query

Application Scenarios:

Labels are used to describe characteristics of people and objects. Each person or object has a set of labels that are uncertain because data is frequently updated. This type of data is widely used in marketing decision-making, recommendation, and advertising systems.

Advantages

- Sparse matrix

The sparse matrix model of HBase is suitable for storing unstructured data. No schema needs to be predefined for tables and no strict column definition is required among rows.

- Easy update

You can update any rows at any time without performance loss. HBase built-in versioning mechanism is used to save multiple historical versions of data.

Storage and Queries of Mass Key-Value Data

- Data types

Structured and semi-structured key-value data, including messages, reports, recommendation data, risk control data, logs, and orders.

- Application scenarios

CloudTable allows high-speed write of mass online and offline key-value data and low-latency data queries. It applies to online applications or report display. CloudTable can easily scaled to achieve HA and low-latency storage and queries of vast amounts of data.

2.2 ClickHouse Application Scenarios

ClickHouse, an abbreviation for "Click Stream + Data WareHouse", serves as a web traffic analysis tool initially designed for OLAP analysis within data warehouses based on page click event flows. ClickHouse has a large number of applications and practices worldwide. It is widely used in various fields such as Internet advertising, apps, web, telecommunications, finance, and IoT. It suits business intelligence ideally.

ClickHouse is widely used in the following scenarios:

- User behavior analysis

Collect use data such as user clicks and browsing duration from the Internet (websites, APPs, and games) and import the data to ClickHouse to construct a wide table for user feature analysis. With the excellent query performance of ClickHouse, multi-dimensional and multi-mode analysis requests can be responded within subseconds. This helps quickly analyze user behaviors for precision marketing and member conversion.

- Enterprise operations analysis

Large-scale transaction data is imported to ClickHouse to construct a large wide table with hundreds of millions of records and hundreds of dimensions. Queries are responded within subseconds. ClickHouse supports personalized statistics and uninterrupted analysis at any time, facilitating business decision-making.

- Visitor source analysis

User behaviors are associated in user access logs through batch offline computing and a wide table of user behavior paths is generated and then synchronized to ClickHouse. A visualization system is built based on ClickHouse to display the interactive visitor source analysis results.

- BI reports

Use ClickHouse to construct real-time interactive query reports to analyze core service metrics such as orders, revenues, and number of users in real time.

- User segmentation statistics

Construct a user information table, select user attribute, tag data, and filter criteria in real time, and perform people feature statistics analytics based on a large number of data records.

3 Cluster Management Functions

3.1 HBase Cluster Management

Introduction to HBase

HBase is a column-oriented distributed cloud storage system that features enhanced reliability, excellent performance, and elastic scalability. It applies to the storage of massive amounts of data and distributed computing. You can use HBase to build a storage system capable of storing terabytes to petabytes of data. With HBase, you can filter and analyze data with ease and get responses in milliseconds, rapidly mining data value.

HBase applies to the following scenarios:

- Mass data storage
HBase applies to TB- or even PB-level data storage and provides dynamic scaling capabilities so that you can adjust cluster resources to meet specific performance or capacity requirements.
- Real-time query
The columnar and key-value storage models apply to the ad-hoc query of enterprise user details. The primary key-based low-latency point query reduces the response latency to seconds or even milliseconds, facilitating real-time data analysis.

For details about HBase architecture and principles, visit <https://hbase.apache.org/book.html>.

NOTE

Currently, CloudTable HBase does not have a security authentication mechanism. If HBase with an authentication mechanism is required, you are advised to use the HBase component in the MapReduce Service (MRS).

HBase Cluster Management Functions

CloudTable is a distributed and scalable key-value data storage service provided by Huawei Cloud. CloudTable provides the following functions of HBase cluster management on the web-based console:

- **Creating a cluster:** You can create a cluster on the CloudTable console. It supports charging based on the number of compute units you selected when creating the cluster and the actual storage capacity. You can independently choose and install the advanced features, which are charged separately. You will be notified of renewal if your balance is insufficient for fee deduction. Cluster resources will be frozen during a retention period and unfrozen after your renewal. CloudTable helps you reduce costs as much as possible by adopting an architecture with computing isolated from storage and dynamically adjusting compute resources.
- **Expanding the cluster capacity:** Compute units of a cluster can be increased.
Increasing compute units: You can dynamically increase the number of compute units based on site requirements or service conditions to ensure read and write performance. The cluster automatically implements load balancing to ensure service continuity and smooth capacity expansion. An extra fee will be generated when you increase compute units.
- **Managing a cluster:** You can manage a created cluster.
 - **Metric monitoring:** The system collects monitoring data during cluster running, reports the data to Cloud Eye (CES), and displays the cluster running status in graphics. When a metric is spotted as abnormal, a message is sent for notification so that users and administrators can handle this problem in a timely manner.
 - **Deleting a cluster:** You can delete a cluster that is no longer needed. This is a high-risk operation. Deleting a cluster may cause data loss. Therefore, before deleting a cluster, ensure that no service is running and all data has been saved.
 - **Restarting a cluster:** You need to restart a cluster if HBase parameters of this cluster have been modified or the system runs slowly due to long-time running. Restart may cause data loss in running services. If you have to restart a cluster, ensure that there is no running service and all data has been saved.
 - **Querying alarms:** If either the system or a cluster is faulty, CloudTable will collect fault information and report it to the network management system. Maintenance personnel will then be able to locate the faults.
 - **Querying logs:** Cluster, job, and configuration operations are recorded, helping locate faults in case of cluster operating exceptions.

Advantages

- **Native HBase APIs:** CloudTable HBase is designed to be compatible with native HBase APIs, ensuring high availability of the architecture through the separation of computing and storage for enhanced reliability, along with in-depth kernel optimization.
- **Ease of use:** Secondary indexes are supported to meet non-primary key query requirements.
- **Low costs:** Cold and hot data can be segregated to fulfill the needs of data archiving and the storage of historical data with infrequent access, thereby minimizing storage expenses.
- **Stability and Reliability:** CloudTable HBase provides stable and reliable performance through hotspot diagnosis and self-healing mechanism.

- Visualized monitoring and O&M: CloudTable HBase offers visualized monitoring and user-defined alarm rules, simplifying system operation and maintenance.

3.2 ClickHouse Cluster Management

Introduction to ClickHouse

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

Cluster Management Functions

- Creating a cluster: You can create a cluster on the CloudTable console.
- Viewing a cluster: You can view cluster details on the CloudTable cluster management page.
- Restarting a cluster: You need to restart a cluster if the system runs slowly due to long-time running. Restart may cause data loss in running services. If you have to restart a cluster, ensure that there is no running service and all data has been saved.
- Deleting a cluster: You can delete a cluster that is no longer needed. This is a high-risk operation. Deleting a cluster may cause data loss. Therefore, before deleting a cluster, ensure that no service is running and all data has been saved.

Advantages

- High performance: ClickHouse employs column-oriented storage. This means data of the same type is stored into the same column, bringing a higher data compression ratio. Generally, the compression ratio can reach 10:1, significantly reducing storage costs and read overhead, and improving query performance.
- Replication mechanism: ClickHouse supports data replication using ZooKeeper and the ReplicatedMergeTree engine (of Replicated series). When creating a table, you can specify a storage engine and determine whether to replicate the table.
- Easy-of-use: You can create a ClickHouse analysis cluster in minutes on the console. No underlying infrastructure management is needed, helping you focus on analyzing data value with complete SQL statements.
- Superior performance: Queries are processed as quickly as possible by using distributed MPP architecture and all available hardware. The query efficiency is several times faster than traditional data warehouses and a single query can process up to terabytes of data per second.
- Security and reliability: Your clusters are independently deployed in isolated VPCs for more secure data access.

- Lower costs: Cost-effective devices on the cloud are used to build a cost-effective hosted ClickHouse cluster.

4 Related Services

IAM

CloudTable uses Identity and Access Management (IAM) for authentication.
For more information about IAM, see [Identity and Access Management](#).

ECS

CloudTable uses an Elastic Cloud Server (ECS) as a node in the cluster.
For more information about ECS, see [Elastic Cloud Server](#).

VPC

CloudTable uses Virtual Private Cloud (VPC) to provide a network topology for clusters to isolate clusters and control access
For more information about VPC, see [Virtual Private Cloud](#).

OBS

CloudTable uses Object Storage Service (OBS) to store data backups and snapshots, making storage secure, reliable, and cost-effective.
For more information about OBS, see [Object Storage Service](#).

CDM

CloudTable uses Cloud Data Migration (CDM) to migrate data from multiple data sources on the cloud and on-premises or third-party cloud to an HBase table of a CloudTable cluster.
For more information about the CDM, see [Cloud Data Migration](#).

CTS

CloudTable uses Cloud Trace Service (CTS) to provide users with operation records of CloudTable resource operation requests and request results for querying, auditing, and backtracking.

For more information about CTS, see [Cloud Trace Service](#).

CES

CloudTable uses CES to monitor cluster performance metrics, delivering status information in a concise and efficient manner. CES supports alarm customization so that you can keep track of all exceptions in real time.

For more information about CES, see [Cloud Eye](#).

5 Permissions Management

If you need to assign different permissions to employees in your enterprise to access your CloudTable resources, IAM is a good choice for fine-grained permissions management. IAM provides identity authentication, permissions management, and access control, helping you secure access to your HUAWEI CLOUD resources.

With IAM, you can use your HUAWEI CLOUD account to create IAM users for your employees, and assign permissions to the users to control their access to specific resource types. For example, some software developers in your enterprise need to use CloudTable resources but must not delete them or perform any high-risk operations. To achieve this result, you can create IAM users for the software developers and grant them only the permissions required for using CloudTable resources.

If your HUAWEI CLOUD account does not need individual IAM users for permissions management, you may skip over this chapter.

IAM can be used free of charge. You pay only for the resources in your account. For more information about IAM, see [IAM Service Overview](#).

CloudTable Permissions

By default, new IAM users do not have any permissions assigned. You need to add a user to one or more groups, and assign permissions policies or roles to these groups. The user then inherits permissions from the groups it is a member of. This process is called authorization. After authorization, the users can perform specified operations on CloudTable based on the permissions.

CloudTable is a project-level service deployed and accessed in specific physical regions. To assign CloudTable permissions to a user group, specify the scope as region-specific projects and select projects for the permissions to take effect. If **All projects** is selected, the permissions will take effect for the user group in all region-specific projects. When accessing CloudTable, the users need to switch to a region where they have been authorized to use cloud services.

Table 5-1 describes the system-defined role supported by CloudTable. Because HUAWEI CLOUD services interact with each other, the CloudTable role is dependent on the roles of other services to implement functions. When assigning a CloudTable role to users, you need to also assign dependent roles for the CloudTable permissions to take effect.

Table 5-1 CloudTable system-defined role

System-Defined Role	Description	Category	Dependencies
cloudtable Administrator	Administrator permissions for CloudTable	System-defined role	The Tenant Guest and Server Administrator roles need to be assigned in the same project.

Table 5-2 lists the common operations supported by each system policy of CloudTable. Please choose proper system policies according to this table.

Table 5-2 Common operations supported by each system policy

Mode	Operation	cloudtable Administrator
Cluster mode	Creating a cluster	√
	Restarting a cluster	√
	Expanding cluster capacity	√
	Deleting a cluster	√
	Configuring parameters	√
	Viewing the cluster list and cluster details of CloudTable	√
	Viewing monitoring information	√
	Viewing audit logs	√

Helpful Links

- [IAM Service Overview](#)
- [Creating a User and Granting Permissions](#)

6 Use Restrictions

Pay attention to the following restrictions before you use CloudTable.

- CloudTable clusters must be created in VPC subnets.
- You are advised to use one of the following browsers to log in to the CloudTable management console.
 - Google Chrome 36.0 or later
 - Mozilla Firefox 35.0 or later
 - Internet Explorer 9.0 or later

If you use Internet Explorer 9.0, you may fail to log in to the CloudTable management console because user **Administrator** is disabled by default in some Windows systems, such as Windows 7 Ultimate. The Internet Explorer automatically selects a system user for installation. As a result, the Internet Explorer cannot access the management console. Reinstall Internet Explorer 9.0 or later (recommended) or run Internet Explorer 9.0 as user **Administrator**.

 - Microsoft Edge: 13.0 or later
- When you create a CloudTable cluster, the system will create a default security group. Do not delete or change the default security group. Deleting or changing the default security group may cause a cluster exception and affect the use of CloudTable clusters.
- Do not assign the access permission for security groups used by CloudTable to prevent malicious access.
- CloudTable does not save the initial password you set for logging in to a node. Please set and keep the password. To prevent malicious attacks, you are advised to set a password with a high complexity.
- Cluster nodes are used only for running CloudTable clusters. You are advised to apply for an independent ECS to deploy your other client applications and business applications.
- Currently, CloudTable does not have a security authentication mechanism. If HBase with an authentication mechanism is required, you are advised to use the HBase component in MRS.
- When creating a table in ClickHouse, it is advised to employ the Replicated series engines to enhance data reliability. Again, cloud services disclaim liability for any data loss or damage arising from a single backup.

7 Quota Description

CloudTable users' default basic resources are as follows:

- Instances
- CPUs
- Memory capacity (GB)
- Disks
- Disk capacity (GB)

For details about how to view and modify quotas of the preceding resources, see [Quotas](#).