CloudTable Service

Product Description

Issue 06
Date 2019-02-27
## Contents

1 What Is CloudTable ................................................................. 1
2 Application Scenarios ............................................................. 3
3 Cluster Mode Functions ........................................................... 9
   3.1 Cluster Management ....................................................... 9
   3.2 HBase ........................................................................ 10
   3.3 OpenTSDB ................................................................. 10
   3.4 GeoMesa ................................................................. 11
   3.5 SQL ........................................................................ 12
4 Serverless Mode Functions ....................................................... 13
5 Related Services .................................................................... 14
6 Permissions Required for Using CloudTable ................................. 16
7 Use Restrictions ................................................................... 18
8 Service Quota ....................................................................... 19
What Is CloudTable

CloudTable Service (CloudTable for short) is a fully managed NoSQL data storage service based on Apache HBase. It offers strong consistency and single-digit millisecond latency, making it a great fit for storing massive amounts of structured, semi-structured, time series, and spatiotemporal data. It can be widely used in Internet of Things (IoT), Internet of Vehicles (IoV), finance, smart city, meteorology, and other industries.

CloudTable provides two modes:

- **Cluster Mode**: The cluster mode requires a dedicated cluster and is suitable for customers with high service throughput and low latency requirements.

- **Serverless Mode**: The serverless mode requires no resource application and is cost-efficient and available once subscribed. It is suitable for customers who have great service volume fluctuation. The serverless mode is in the open beta test, during which you can use it free of charge.

Architecture

The following figure shows the architecture of CloudTable.

**Figure 1-1 Architecture**

- **Open ecosystem**: CloudTable is compatible with the native APIs of HBase, OpenTSDB, and GeoMesa.
HBase: Key-value data models and high availability (HA) of the architecture are supported. Two HMaster nodes in active/standby mode realize real-time HA detection. Failover between regions in seconds can address unexpected computing unit failures and ensure service HA. Storage and computing are separated to ensure high data reliability. Multiple backup mechanisms are used for storage.

Time series database: OpenTSDB is integrated to provide efficient read, write, query, and computing of time series data. Read/write performance is improved by 30% to 60%. Interpolation, downsampling, and aggregation analysis capabilities are supported. High compression ratio (10:1) lowers costs.

Spatiotemporal big data: GeoMesa (an open source suite of tools that enables large-scale geospatial querying and analytics) is integrated to store and analyze spatiotemporal data and provide functions such as tracking query, region distribution statistics, region query, density analysis, aggregation, and origin-destination (OD) analysis, surpassing traditional spatial database functions.

Storage layer: HA is supported. Faulty disks do not affect data read/write and reliability. The computing and storage are separated, and storage is charged by usage. This costs less.

Cluster management: A visualized management platform provides functions such as cluster management, configuration parameter modification, and online computing unit scale-out. In addition, the platform supports visualized monitoring and O&M and provides common monitoring information (such as CPU usage, IOPS, and incoming and outgoing traffic) and custom alarm rules, simplifying system O&M.

Security: CloudTable uses Identity and Access Management (IAM) to implement identity authentication. Underlying computing units of CloudTable run on a Virtual Private Cloud (VPC) network. VPC isolates networks logically and provides a secure and isolated network environment. Security groups are used to implement access control in and between security groups. You can customize security group rules to control access to CloudTable.
IoT Device Monitoring

**Application scenarios:**

IoT devices such as IoE, gas, water, electricity, chemical devices, and Internet are connected to the cloud through IoT suite services. CloudTable keeps track of your devices, writes device data and analysis results to OpenTSDB in real time, and outputs time series results to the user's front-end monitoring system using OpenTSDB APIs. It implements real-time monitoring and analysis of IoT devices.

**Advantages:**

- **Easy access**
  CloudTable (OpenTSDB) supports open protocols that enable easy interconnection with the messaging system and real-time stream computing system, reducing development difficulties.

- **High-performance read/write**
  10-million-level write throughput of time series data, 3-second query latency for million-level data points; 30% to 60% read performance improvement and 60% write concurrency increase compared with open source OpenTSDB

- **Aggregation capabilities**
  Interpolation, downsampling, and powerful aggregation function capabilities

- **Low cost**
  On-demand storage charging and flexible capacity expansion helping cope with the uncertainty of services; high compression ratio (10:1), lowering costs

**Related Services:**

Cloud Stream Service (CS), Data Ingestion Service (DIS), Data Lake Insight (DLI), IoT Platform, and Object Storage Service (OBS)
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
  Various Hadoop ecosystem components, integrated with HUAWEI CLOUD products

Related Services:
DIS and CS
Location-based Big Data Applications in IoV

Application Scenarios:

In IoV scenarios, various types of data are generated, such as basic information about vehicles and drivers; monitoring data of vehicle status, batteries, and motors; and vehicle driving data. GeoMesa stores and analyzes spatiotemporal data and provides functions such as tracking query, region distribution statistics, region query, density analysis, aggregation, and origin-destination (OD) analysis.

Advantages:

- Multimodal Database Capabilities
  Different indexes provided for various types of data to deliver optimal performance, query, and analysis

- Abundant query and analysis functions
  GeoMesa provides functions such as tracking query, region distribution statistics, region query, density analysis, aggregation, and OD analysis.

- Seamless interconnection
  Spatiotemporal databases seamlessly interconnected with DLI to provide better analysis for time and space data, such as heat maps

Related Services:

CS, DIS, DLI, IoT Platform, and OBS
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.

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

**Related Services:**

DIS and CS
Serverless Web/Mobile App Back-ends

Application Scenarios:
CloudTable and FunctionGraph are collaborated to quickly build high available, auto-scaling web and mobile application back-ends.

Advantages:
- High availability
  CloudTable and OBS HA ensuring high reliability of website data, and API Gateway and FunctionGraph improving HA of website logic
- Superb performance
  A maximum of 20,000 IOPS per disk and 350 MB/s throughput
- Scalable
  If services surge, resources can be automatically allocated to run more function instances to meet processing requirements.
- Low cost
  You are only charged for the duration during which functions are processing files or data, and the storage capacity you use. Auto scaling enables you to avoid resource redundancy during off-peak hours.

Related Services:
OBS and FunctionGraph
Figure 2-5 Web/mobile app back-ends
3 Cluster Mode Functions

3.1 Cluster Management

CloudTable is a distributed and scalable key-value data storage service provided by HUAWEI CLOUD. CloudTable provides the following functions of the cluster mode on the web-based console.

- Cluster creation: You can create a cluster on the CloudTable Service console. It supports charging based on the number of computing 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 computing resources.

- Expanding the cluster capacity: Computing units of a cluster can be increased.
  - Increasing computing units: You can dynamically increase the number of computing units based on site requirements or service conditions to ensure read/write performance. Clusters can automatically implement load balancing to ensure service continuity and smooth capacity expansion. An extra fee will be generated when you increase computing units.

- Advanced feature: Currently, the CloudTable cluster mode supports the OpenTSDB, SQL and GeoMesa advanced features. You can choose to enable or disable advanced features. If the OpenTSDB and SQL advanced features are enabled, they will be charged.

- Cluster management: You can manage clusters you create.
  - Monitoring metrics: CloudTable collects monitored metric data from a running cluster and reports it to Cloud Eye so that you can view cluster running status displayed in graphics. If there is an abnormal metric, CloudTable sends a message to notify you and administrators for immediate manual intervention.
  
  - Deleting a cluster: You can delete a cluster when you do not need it. Deleting a cluster is risky and may cause data loss. Before deleting a cluster, ensure that there is no running service and all data has been saved.
  
  - Restarting a cluster: After the HBase parameters of the cluster are modified or the system runs slowly due to running for a long time, you need to restart the cluster.
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: In case of a cluster operating exception or system failure, CloudTable will collect fault information and report it to the network management system for maintenance personnel to locate faults.
- Querying logs: CloudTable records your operations on clusters to help you locate causes of a cluster running exception.

### 3.2 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 TB- or even PB-level 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:

- **Storage of massive amounts of data**
  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.


### 3.3 OpenTSDB

The CloudTable cluster mode also provides the time series database capabilities based on OpenTSDB.

OpenTSDB is a distributed, scalable time series database based on HBase. It stores time series data. Time series data refers to the data collected at different time points. This type of data reflects the change status or degree of an object over time.

**Product Advantages**

- **Low cost**
  Timestamps are compressed with delta encoding and data values are compressed using XOR.
  Computing and storage are decoupled to cope with massive sets of data and dynamic hot data in IoT scenarios and facilitate independent capacity expansion by concurrency and storage volumes on demand.

- **Enterprise-level**
  Distributed architecture and horizontal expansion are supported.
  Using algorithms with a high compression ratio reduces costs and improves query speed.
Compatibility

Compatible with OpenTSDB 2.3.0 of the community
Compatible with native OpenTSDB APIs, allowing no modification on applications during service migration

Time series data computing

Interpolation: Linear interpolation can be implemented to interpolate missing data points.
Downsampling: Pre-downsampling and real-time downsampling are supported, meeting requirements on efficient query.
Space aggregation: Space aggregation and "group by" calculation can be implemented based on different tags.
A wide variety of aggregation functions are provided, such as AVG, SUM, MAX, and MIN.

Application Scenarios of OpenTSDB

OpenTSDB supports storage, index, and queries of massive amounts of time series data. It can be used for system monitoring and measurement as well as monitoring and analysis of IoT data, financial data, and scientific experimental results.

- OpenTSDB can help O&M engineers obtain the real-time status of infrastructure and services, and display software and hardware faults, performance changes, and performance bottlenecks of clusters.
- OpenTSDB can help managers store and analyze system monitoring data, measure system SLA, and understand interactions between complicated systems, as well as display resource consumption. It also provides the overall job status of the cluster to help managers with budget and cluster resource coordination.
- OpenTSDB assists developers in finding major cluster performance bottlenecks by storing and analyzing system monitoring data.

When to Enable OpenTSDB

When you want to store and query time series data, you can enable OpenTSDB.

3.4 GeoMesa

The CloudTable cluster mode also provides the big data query and analysis capabilities based on GeoMesa.

GeoMesa is a distributed, scalable, open source spatio-temporal data spatial-temporal database based on HBase. GeoMesa uses highly parallel index policies to provide HBase-based spatio-temporal data query and processing capabilities.

Application Scenarios of GeoMesa

GeoMesa supports spatio-temporal queries in milliseconds and can store, index, and query various kinds of user GIS data.

GeoMesa applies to the following scenarios:

- Large-scale geospatial data storage
GeoMesa integrates with HBase storage capabilities and can store TB-level or PB-level geospatial data.

- Real-time query
  GeoMesa provides quick spatio-temporal data queries based on an especially designed index architecture.

When to Enable GeoMesa

When you want to store and query spatio-temporal and geospatial data, you can enable GeoMesa.

3.5 SQL

The CloudTable cluster mode enables you to access clusters using SQL.

SQL integrates with Apache Phoenix to enable HBase in CloudTable clusters to be accessed through SQL.

Application Scenarios of SQL

Users use third-party data analysis tools, such as Tableau, SQuirreL, and SQLLine, or SQL clients to access CloudTable clusters.

In the users' service code, ODBC or JDBC APIs are called to execute SQL to access HBase in CloudTable clusters.

SQL Syntax Reference

The CloudTable serverless mode provides a distributed and fully managed NoSQL data storage system based on Apache HBase. It provides strong consistency and single-digit millisecond latency so it is optimal for storage and queries of massive amounts of structured and semi-structured data.

Without applying for resources, you can manage tables and namespaces, as well as add, delete, modify, and query data in CloudTable serverless mode. You can manage your own table and namespace resources and other users have no permission to access your resources.

The CloudTable serverless mode provides the following functions:

- **Table management**
  
  On the CloudTable management console, you can perform the following operations: creating tables, managing data in tables, viewing basic information and monitoring information of tables, modifying read and write capacity of tables, and deleting tables.

- **Namespace management**
  
  By default, the CloudTable serverless mode supports the `default` namespace. You can also create and delete namespaces on the management console.

- **REST API**
  
  The CloudTable serverless mode provides REST APIs for you to call them to perform DDL and DML operations on tables or namespaces to add, delete, modify, and query table data.
5 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.

CSS
CloudTable uses the Elasticsearch search engine of Cloud Search Service (CSS) to improve full-text search capability. In an HBase table creation statement, you can use METADATA to define the Elasticsearch schema of an index field and specify an address to access a CSS cluster so that the CSS cluster can be automatically accessed and an index can be created in...
the Elasticsearch during table creation. Your source data is stored in HBase of CloudTable, and the index data is stored in the Elasticsearch of CSS.

For more information about CSS, see Cloud Search Service.

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.

Cloud Eye

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

For more information about Cloud Eye, see Cloud Eye.
Permissions Required for Using CloudTable

This section describes permissions required for using CloudTable.

Permission List

Table 6-1 lists the permissions for using CloudTable.

For more information about permissions, see Permission Description.
<table>
<thead>
<tr>
<th>Node Name</th>
<th>Permission Name</th>
<th>Managed Cloud Service Resource</th>
<th>Permission Description</th>
</tr>
</thead>
</table>
| CloudTable| cloudtable Administrator   | CloudTable Service             | CloudTable Administrator permission. Users with this permission and the Tenant Guest and Server Administrator permissions can perform all operations on CloudTable resources. CloudTable cannot run properly if either Tenant Guest or Server Administrator permission is unavailable. If you have the CloudTable Administrator permission only, you can perform the following operations:  
- Create, restart, expand, and delete a CloudTable cluster.  
- Enable advanced features such as OpenTSDB and SQL.  
- View and configure CloudTable cluster parameters.  
- View the cluster list and cluster details of CloudTable.  
- In serverless mode, create and delete namespaces, create tables, update tables, view table monitoring information, manage data in tables, and delete tables.  
- View the monitoring information and alarm list of CloudTable.  
- Query operation logs of CloudTable. |

Table 6-1 Permission description
7 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.
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.