CloudTable Service

FAQs

Issue 01

Date 2025-08-14

HUAWEI CLOUD COMPUTING TECHNOLOGIES CO., LTD.

Trademarks and Permissions


and other Huawei trademarks are the property 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 Cloud 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 Cloud Computing Technologies Co., Ltd.

Address: Huawei Cloud Data Center Jiaoxinggong Road
Qianzhong Avenue
Gui'an New District
Gui Zhou 550029
People's Republic of China

Website: https://www.huaweicloud.com/intl/en-us/

Contents

1 Product Consulting

1.1 What Services Does a CloudTable Cluster Provide?

CloudTable Service (CloudTable) provides you with dedicated clusters, which are available upon subscription. It is suitable for users with high service throughput and low delay requirements.

- CloudTable offers HBase-based fully managed NoSQL services, delivering lightning-fast random read/write capabilities within milliseconds. It is an excellent choice for storing vast amounts of structured, semi-structured, spatiotemporal, and time series data. Whether in IoT, IoV, finance, smart cities, or meteorology, CloudTable finds versatile applications.
- Based on Doris, CloudTable provides fully managed real-time data warehouse services and query results of mass data can be returned in subseconds. Doris supports high-concurrency point queries and high-throughput complex analysis. All this makes Doris an ideal tool for report analysis, ad-hoc query, unified data warehouse, and data lake query acceleration. On Doris, users can build various applications, such as user behavior analysis, AB test platform, log retrieval analysis, user portrait analysis, and order analysis.
- ClickHouse is an open-source columnar database oriented to online analysis and processing. It is independent of the Hadoop big data system and features compression rate and fast query performance. In addition, ClickHouse supports SQL query and provides good query performance, especially the aggregation analysis and query performance based on large and wide tables. The query speed is one order of magnitude faster than that of other analytical databases.

1.2 Why Do I Choose CloudTable Service?

CloudTable is a fully managed service based on the public cloud. It provides you with dedicated clusters, which are available upon subscription. It is suitable for users who require high service throughput and low latency. CloudTable is O&M-free and features effortless deployment. It allows you to create clusters on demand to reduce costs and offers professional tutorial.

1.3 What Should I Pay Attention to When Using CloudTable Service?

- On the CloudTable management console, check cluster status and task status. If the cluster status is **Creation failed** or **Sub-health**, or a task fails to be executed, submit a service ticket to technical support.
- On the CloudTable management console, click **Monitor** on the right of the **Cluster Management** page to access the **CloudTable Service Monitoring** page on the Cloud Eye management console. View metric monitoring information and set alarm rules for key metrics. If there is an alarm, submit a service ticket to technical support. If the cluster CPU or memory usage exceeds the threshold for a long time, you need to increase the number of computing units to balance loads.
- On the CloudTable management console, click **Cluster Management** to open the cluster list. Check whether the used storage capacity of a cluster exceeds the threshold. If it exceeds the threshold, increase storage quotas and storage capacities.

2 Connection and Access

2.1 How Do I Access a CloudTable Cluster?

CloudTable is based on an elastic cloud server (ECS). You need to create an ECS in the same subnet as the CloudTable cluster, synchronize the cluster IP address, and access the cluster using the client.

2.2 What Should I Do If a CloudTable Cluster Cannot Be Connected?

Table 2-1 Possible causes and solutions

Possible Causes	Solutions
Creation failed, **Sub-health**, or **Frozen** is displayed in the **Status** column.	• If **Creation failed** is displayed in the **Status** column, delete the cluster that fails to be created, and create a new one, and connect to it when the creation is complete. • If **Subhealthy** is displayed in the **Status** column, contact technical support engineers. • If **Frozen** is displayed in the **Status** column, top up the account. Ensure that the account balance is not **0** to guarantee the normal use of the cluster. **NOTE** If you want to unfreeze a cloud service that is frozen in violation of regulations or by the public security department, contact the administrator.

Possible Causes	Solutions
The connection command, username, password, IP address, or port number is incorrect.	• Modes for connecting to an HBase client: – To connect to an HBase normal cluster, see **Connecting to an HBase Normal Cluster Using HBase Shell**. – To connect to an HBase security cluster, see **Connecting to an HBase Security Cluster Using the HBase Shell**. • Modes for connecting to a Doris client: – To connect to a Doris normal cluster, see **Using the MySQL Client to Connect to a Doris Normal Cluster**. – To connect to a Doris security cluster, see **Using the MySQL Client to Connect to a Doris Security Cluster**. • Modes for connecting to a ClickHouse client: – To connect to a ClickHouse normal cluster, see **Using a Client to Connect to a ClickHouse Normal Cluster** – To connect to a ClickHouse security cluster, see **Using a Client to Connect to a ClickHouse Security Cluster**.
The OS type or the version of the client is correct.	• For HBase 1.*x* clusters, use the hbase1.*x* client, and for HBase 2.*x* clusters, use the hbase2.*x* client. • For normal Doris clusters, use the open-source MySQL client, and for security Doris clusters, use the MySQL 8 client or later. • Use the 23.*x* ClickHouse client for ClickHouse clusters. • For normal StarRocks clusters, use the open-source MySQL client.
The client is incorrectly installed.	• To install the HBase client, see **Connecting to an HBase Normal Cluster Using HBase Shell**. • To install the Doris client, see **Using the MySQL Client to Connect to a Doris Normal Cluster**. • To install the ClickHouse client, see **Using a Client to Connect to a ClickHouse Normal Cluster**.
The ECSs are not in the same AZ, VPC, subnet, and security group as the cluster.	Check the AZ, VPC, subnet, and security group displayed on the ECS details page and the cluster details page.

Possible Causes	Solutions
The inbound and outbound rules of the security group are incorrect.	On the cluster details page, click the security group name. On the displayed **Security Groups** page, click **Manage Rules** to check the inbound and outbound rules. If there is no inbound rule, press **Windows**+**R** to display the **Run** dialog box, enter **cmd** to open the **cmd** window, and enter **ipconfig**. The IP address of the local host is displayed. Click **Add Rule**, enter the obtained IP address, and try to connect to the cluster again. For details, see **Configuring Security Group Rules**.
Your network is not connected to the Internet.	Press **Windows**+**R** to display the **Run** dialog box, enter **cmd** to open the **cmd** window, and enter **ping *URL***.
The firewall blocks the access.	Modify the firewall policy or temporarily disable the firewall.
The proxy for accessing the Internet is not enabled.	Enable the proxy.

NOTE

If you cannot determine the cause and rectify the fault, submit a service ticket to report it. Log in to the CloudTable console, choose **More** > **Service Tickets** and click **Create Service Ticket**. Create a ticket and submit it.

2.3 Cross-VPC/Public Network Access Scenarios Supported by CloudTable Clusters

Scenario

- Connecting a CloudTable cluster to the peer cluster in another VPC in the same region
- Connecting a CloudTable cluster to the peer cluster in another region
- Accessing a CloudTable cluster from the public network
- Connecting a CloudTable cluster to an on-premises data center

Table 2-2 CloudTable cluster usage scenarios

Scenario	Component	Solution	Constraints and Limitations
Connecting a CloudTable cluster to the peer cluster in another VPC in the same region	ClickHouse	VPC peering connections: connect VPCs in a region in the same account or different accounts. For more information, see **VPC Peering Connection Overview**.	To enable the CloudTable cluster to access the peer cluster, you need to add a route for the cluster. **NOTE** To add a route, contact technical support.
Connecting a CloudTable cluster to the peer cluster in another region	ClickHouse	Cloud Connect: connects VPCs in different regions to quickly build cross-region networks. For more information, see **What Is Cloud Connect?**	To enable the CloudTable cluster to access the peer cluster, you need to add a route for the cluster. **NOTE** To add a route, contact technical support.
Accessing a CloudTable cluster from the public network	ClickHouse	Bind an ELB. Binding an EIP to the ELB provides public network access and load balancing capabilities for clusters. For more information, see **What Is ELB?**	The peer cluster cannot be accessed.

Scenario	Component	Solution	Constraints and Limitations
Connecting a CloudTable cluster to an on-premises data center	ClickHouse	• Direct Connect: Direct Connect provides a high-speed, low-latency, stable, and secure dedicated network connection between an on-premises data center and a cloud-based VPC. With Direct Connect, you can establish a large-scale hybrid cloud network. For more information, see **What Is Direct Connect?** • Virtual Private Network (VPN): A VPN establishes a secure, encrypted communication tunnel between your on-premises data center and your VPC. For more information, see **What Is Virtual Private Network?**	To enable the CloudTable cluster to access the peer cluster, you need to add a route for the cluster. **NOTE** To add a route, contact technical support.

NOTE

HBase public network access configuration:

- You can assign a public IP address to each HMaster and RegionServer node in the HBase cluster, bind the NAT gateway to the cluster's VPC, and set DNAT to full port mapping. For more information, see **Configuring a NAT Gateway**.
- Configure the host mapping on the client.
- Obtain the host names mapped to **hosts** and the private IP addresses of HMaster and RegionServer from technical support.

2.4 How Do I View the IP Address Corresponding to a Domain Name in a CloudTable Link?

Submit a service ticket to technical support to obtain the IP address corresponding to a domain name, because you cannot view it by yourself.

To submit a service ticket, log in to the CloudTable management console, and choose **More** > **Service Tickets** > **Create Service Ticket** in the upper right corner.

Fill in the service ticket to obtain the mapping between the intranet domain name of the CloudTable link and the IP address.

2.5 Can I Use SSH to Access Computing Nodes of HBase?

No. Underlying computing units of CloudTable HBase run on a Virtual Private Cloud (VPC) network. VPC isolates networks logically and provides a secure and isolated network environment to protect your data. You can only use a client tool or application to connect to the ZooKeeper link of HBase (HBase connection address) to access the HBase cluster.

NOTE

The OpenTSDB function applies only to HBase 1.3.1.

2.6 Why Is the Error "Will not attempt to authenticate using SASL (unknown error)" Reported When Connecting to HBase?

Symptom

CloudTable HBase fails to be connected and the following error information is displayed:

```
Opening socket connection to server 192.168.0.107/192.168.0.107:2181. Will not attempt to authenticate using SASL (unknown error)
```

Possible Cause

The possible cause is as follows:

Network access failure

The HBase connection address is an internal network address, not a public network address. Consequently, direct connections to the HBase cluster from the public internet are not supported. You can ping the ZooKeeper link of HBase on a client host. If it cannot be pinged, the network connection is abnormal.

Solution

You need to apply for an ECS as the client host to install the CloudTable client. Then, you can use the client on the ECS to access the CloudTable cluster. The ECS to be applied for and the CloudTable cluster must be in the same region, VPC, subnet, and security group. In addition, you need to configure a correct DNS server address for the ECS. For details about how to prepare an ECS, see **Creating an ECS** .

2.7 What Should I Do If It Takes a Long Time to Concurrently Access an HBase Cluster Through the Thrift Interface Using Python?

Symptom

When a CloudTable HBase data table is accessed, concurrent tasks run slowly.

Possible Causes

GC memory leakage occurs on cluster nodes. Perform gray restart on the nodes.

NOTE

In addition, check whether the load is too high. If the nodes are overloaded, contact related personnel.

Solutions

Contact maintenance personnel to restart the RegionServer instance.

3 Data Read/Write

3.1 Is Raw Data Stored in CloudTable HBase?

CloudTable HBase stores raw data and supports compressed storage. The SNAPPY and GZ compression algorithms are supported.

CloudTable HBase does not encrypt or decrypt raw data. If data needs to be encrypted or decrypted, use service code on an application to implement data encryption and decryption.

3.2 Why Can't I Write Data to HBase?

Symptom

Data cannot be written to HBase. As a result, service data on an application cannot be updated in time.

Possible Cause

There are too many ZooKeeper connections.

Solution

When service code is connected to the same CloudTable HBase cluster, you are advised to create one Connection and reuse it for multiple threads. You do not need to create a Connection for every thread. Connection is a connector for connecting to a CloudTable HBase cluster. Excessive Connections will increase loads on ZooKeeper and deteriorate service read/write performance. Multiple client threads can share one Connection. Typically, client threads of a client program share one Connection.

3.3 What Is the Maximum Size of Data Written to the HBase Cluster?

Symptom

When a large amount of data is concurrently written to an HBase cluster, data cannot be fully written.

Possible Causes

The data written at a time is too large and an error occurs.

Solutions

The size of data written at a time should not exceed 2 MB and the size of a single data record should not exceed 200 KB.

3.4 How Do I Check the Daily Incremental Data in HBase Tables?

Symptom

This section describes how you can check the daily incremental data volume in HBase tables.

Solution

Incremental data in HBase tables cannot be directly checked. You need to query the incremental data in the background. Alternatively, you can view the daily incremental data on the console, view the used storage capacity on the details page, and calculate the incremental data by recording the storage capacity.

3.5 How Do I View the List in the Database Specified by Catalog?

Catalog is an external data directory. The following describes how to connect the JDBC Catalog to other data sources through the standard JDBC protocol.

Syntax

```
CREATE CATALOG [IF NOT EXISTS] catalog_name
   PROPERTIES ("key"="value", ...);
```


Table 3-1 Parameter description

Parameter	Mandatory	Default Value	Description
user	Yes	-	Username of the corresponding database.
password	Yes	-	Password of the database.
jdbc_url	Yes	-	JDBC URL.
driver_url	Yes	-	Name of the JDBC Driver JAR package.
driver_class	Yes	-	Name of the JDBC driver class.
lower_case_table_names	No	"false"	Whether to synchronize the database name and table name of the JDBC external data source in lowercase.
only_specified_database	No	"false"	Whether to synchronize only the specified database.
include_database_list	No	""	This parameter is used to synchronize certain databases when **only_specified_database** is set to **true**. Use commas (,) to separate databases. The database name is case sensitive.
exclude_database_list	No	""	This parameter is used to not synchronize certain databases when **only_specified_database** is set to **true**. Use commas (,) to separate databases. The database name is case sensitive.

NOTE

- **driver_url** can be set to:
  - File name, for example, **mysql-connector-java-5.1.47.jar**. Place the JAR package in the **jdbc_drivers/** directory under the FE and BE deployment directories in advance and the system can locate the file.
  - Local absolute path, for example, **file:///path/to/mysql-connector-java-5.1.47.jar**. Place the JAR package in the specified paths of all FE and BE nodes in advance.
  - HTTP address, for example, **https://doris-community-test-1308700295.cos.ap-hongkong.myqcloud.com/jdbc_driver/mysql-connector-java-5.1.47.jar**. The system downloads the driver file from the HTTP address. This only supports HTTP services with no authentication requirements.
- **only_specified_database**: You can specify the database or schema to be connected during JDBC connection. For example, you can specify a database in **jdbc_url** of MySQL and **currentSchema** in **jdbc_url** of PostgreSQL.
- **include_database_list**: This parameter is used to synchronize certain databases when **only_specified_database** is set to **true**. Use commas (,) to separate databases. The database names are case sensitive.
- **exclude_database_list**: This parameter is used to not synchronize certain databases when **only_specified_database** is set to **true**. Use commas (,) to separate databases. The database names are case sensitive.
- When **include_database_list** and **exclude_database_list** specify overlapping databases, **exclude_database_list** would take effect with higher privilege over **include_database_list**.

  If you need to connect to the Oracle database when using these parameters, use the version of the JAR package no less than 8 (such as **jdbc8.jar**).

Procedure

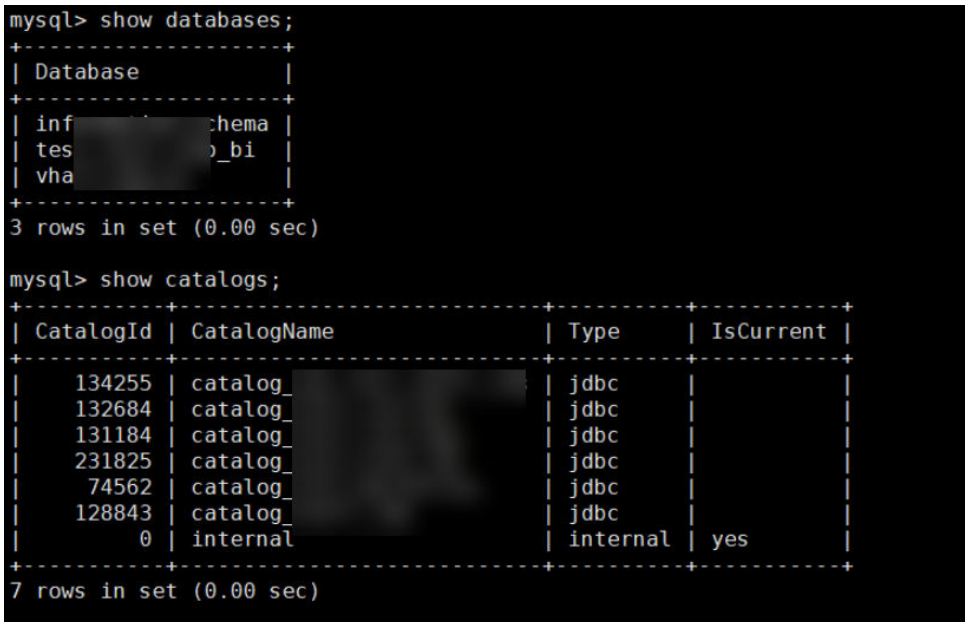
Step 1 Create a catalog.

```
CREATE CATALOG jdbc_mysql PROPERTIES (
    "type"="jdbc",
    "user"="root",
    "password"="******",
    "jdbc_url" = "JDBC URL",
    "driver_url" = "mysql-connector-java-8.0.25.jar",
    "driver_class" = "com.mysql.cj.jdbc.Driver"
);
```

Step 2 View the Catalog.

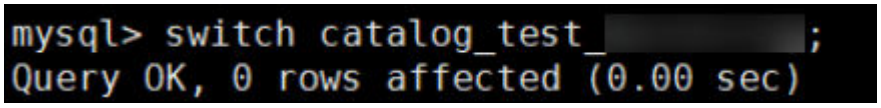
```
show catalogs;
```


Figure 3-1 Viewing the Catalog



Step 3 Run the **switch** *Catalog name* command to switch to the created catalog.

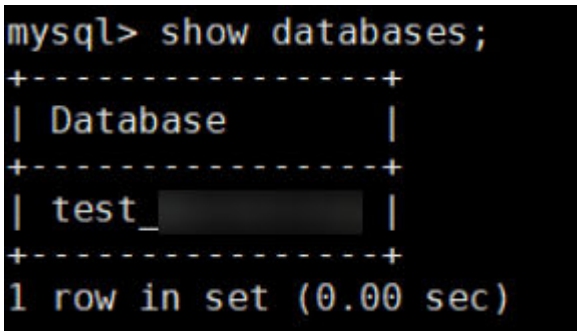
Figure 3-2 Switching to the created catalog



Step 4 View the mapped database.

```
show databases;
```

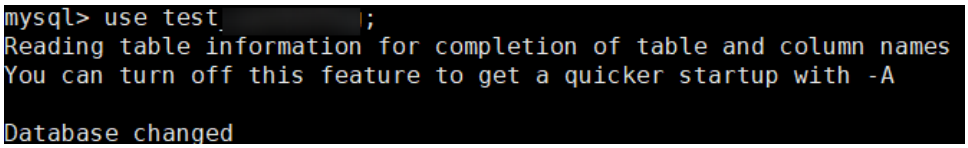
Figure 3-3 Viewing the mapped database



Step 5 Use the database.

```
use Database name;
```

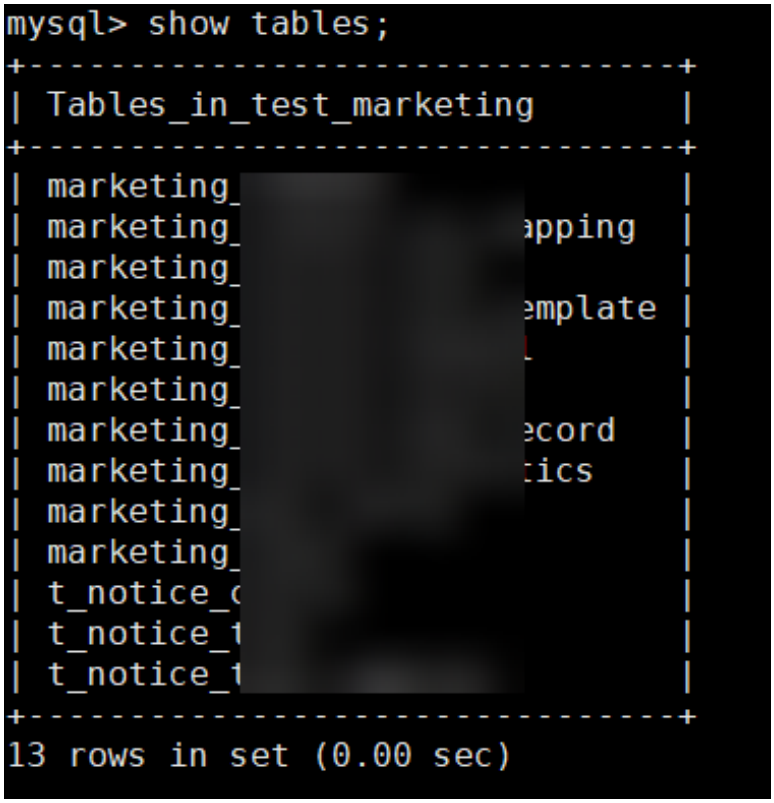
Figure 3-4 Using the database



Step 6 Check the database table.

```
show tables;
```

Figure 3-5 Checking the database table



----End

3.6 What Should I Do If an Error Is Reported When I Access the CloudTable HBase Cluster?

Symptom

An error is reported when the CloudTable HBase cluster is accessed, indicating that the hmaster domain name cannot be identified.

Possible Causes

The mapping between the private domain name and IP address of the CloudTable HBase instance nodes is not configured on the server where the program is running. As a result, the private domain name of the hmaster node cannot be identified, and a link error occurs.

Solutions

In the **host** file on the server where the program is running, configure the mapping between the private domain name and IP address of CloudTable HBase instance nodes. For details, see **How Do I Configure DNS for an ECS?**

3.7 What Can I Do If the Index Table Does Not Match the Data Table After HBase Cluster Data Is Deleted?

Symptom

When you query data in the HBase table, the value of **count** is inconsistent with the number of data records in **list**. Data loss frequently occurs, and many links time out during project running.

Possible Causes

After data is deleted, the index table does not match the data table. As a result, an exception occurs.

Solutions

- Short-term solution: Perform offline batch processing analysis on data streams and supplement the missing data on the next day.
- Long-term solutions: Check whether the TTL enhanced index can be removed in the CBC data aging scenario.

4 HBase

4.1 What Compression Algorithms Are Supported by CloudTable HBase Clusters?

Symptom

This section describes what compression algorithms CloudTable HBase clusters support.

Solution

CloudTable HBase clusters support the SNAPPY, LZ4, ZSTD, BZIP2, LZO, and GZ compression algorithms.

4.2 Which Programming Languages Are Supported by HBase External APIs?

Currently, external APIs of HBase 2.4.14 only support the Java programming language.

For more information about HBase APIs, see **Apache HBase External APIs** in the HBase official documentation.

4.3 How do I view the TTL attribute of HBase shell?

- If the TTL of the cell is set when the data is inserted, the TTL attribute cannot be viewed. You can check whether the TTL configuration is successful.
- If the TTL of the cell is not set when data is inserted, the system automatically inserts the current time as the timestamp.

4.4 How Do I Determine the Number of Faulty RegionServers?

Symptom

This section describes how to determine the number of faulty RegionServers.

Solution

The number of faulty RegionServers is the number of faulty RegionServer instances in the cluster where the monitored object is located.

4.5 Can I Configure the hbase-site.xml File?

- hbase-site.xml file on a client

  If you have deployed an HBase shell client, you can configure the **hbase/conf/hbase-site.xml** file in the client directory.

- hbase-site.xml file on a server

  You cannot directly edit the **hbase-site.xml** file on the server. However, you can modify some parameters in the **hbase-site.xml** file on the CloudTable management console, for details, see **Modifying HBase Parameters to Optimize Cluster Performance**. If modification of these parameters cannot meet your requirements and you want to modify other parameters, contact technical support.

4.6 How Do I Query the Creation Time of a Table in CloudTable HBase?

You can use either of the following methods to query the table creation time in CloudTable:

Check the table creation time in the output of the **create** command.

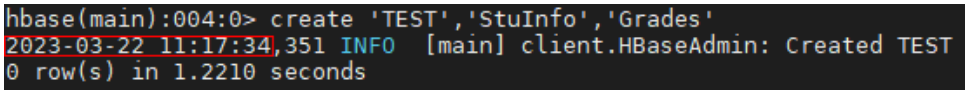
Step 1 Connect to the cluster and use the HBase shell to access the cluster. For details, see **Using HBase Shell to Access a Cluster** in the *User Guide*.

Step 2 Run the **create** command to create a table.

```
create 'TEST','StuInfo','Grades'
```

Step 3 In the command output, the first row indicates the table creation time and table name, and the second row indicates the table creation duration.

Figure 4-1 Table creation time



----End

5 Doris

5.1 What Factors Influence Data Balancing?

During Doris's operation, the frontend (FE) node continuously monitors each disk's load through the metadata. Should a data imbalance arise, the FE node promptly redistributes data from the overloaded disk to the underutilized one until an equilibrium is achieved across all disks.

What Factors Influence Data Balancing?

- Frequent data writing can lead to constant changes in the load experienced by each node and disk, thereby initiating data balancing.
- Moreover, if new data is written amidst an ongoing data balancing process, it may interfere with the current balance.
- Disk load calculations are influenced by the presence of junk files. Therefore, mass deletion of these files can lead to disk load imbalances.

Symptom

Unchecked data disk space in Doris can lead to a full disk, preventing further write processes. To prevent this, it is crucial to manage Doris system operations to circumvent a full disk scenario.

Cause Analysis

The imbalance of disks often stems from excessive data writing coupled with the deletion of numerous junk files.

Solutions

- To mitigate frequent shifts in disk load, it is advisable to schedule data writing sessions.
- Additionally, manually removing junk files and shortening the junk file expiration period can help. Once the data is balanced and the volume of junk files is minimized, the original junk file expiration period can be reinstated. For guidance on adjusting the recycle bin duration, see **Setting the Recycle Bin Duration**.

5.2 How to Adjust the Sensitivity of Data Equalization? What Are the Impacts After the Adjustment?

To ensure smooth operation within the Doris system, it is essential to manage data disk space effectively to prevent full disks that halt processes. This involves monitoring disk usage and available space, setting various alert levels to maintain control over system operations, and avoiding scenarios where disk space is completely utilized.

Fine-Tuning Data Equalization Sensitivity

The backend (BE) regularly reports disk usage statistics to the FE every minute. The FE then uses these statistics to regulate different operational requests.

Two critical thresholds are established within the FE: the High Watermark and the Flood Stage. The Flood Stage is a more critical level than the High Watermark. When disk usage surpasses the High Watermark, Doris limits certain tasks, such as backup balancing. Should usage exceed the Flood Stage, more severe restrictions are applied, and some processes, like imports, are halted.

Additionally, the BE sets its own Flood Stage. Since the FE cannot instantly detect the BE's disk usage or control its ongoing operations, such as Compaction, the BE uses its Flood Stage to autonomously halt certain actions to safeguard itself.

Impacts

- Setting the sensitivity too low can result in significant data volume discrepancies between disks, adversely affecting concurrent performance.
- If the interval is set too long, an excessive number of junk files may accumulate, consuming valuable disk space. In such cases, it is advisable to modify the timeout interval for junk files.

5.3 How Can I View and Restore Recycle Bin Data in the Doris Cluster?

Scenarios Where Recycle Bin Data Is Generated

- During data balancing, tablets from high-load disks are copied to low-load disks. The original tablets are placed in the recycle bin without physical deletion, resulting in additional junk files.
- Operations like Delete, Drop, and Truncate only remove data logically, not physically, resulting in additional junk files.
- Post-merging of data files, the outdated data remains unerased, further contributing to junk file accumulation.

Junk File Implications

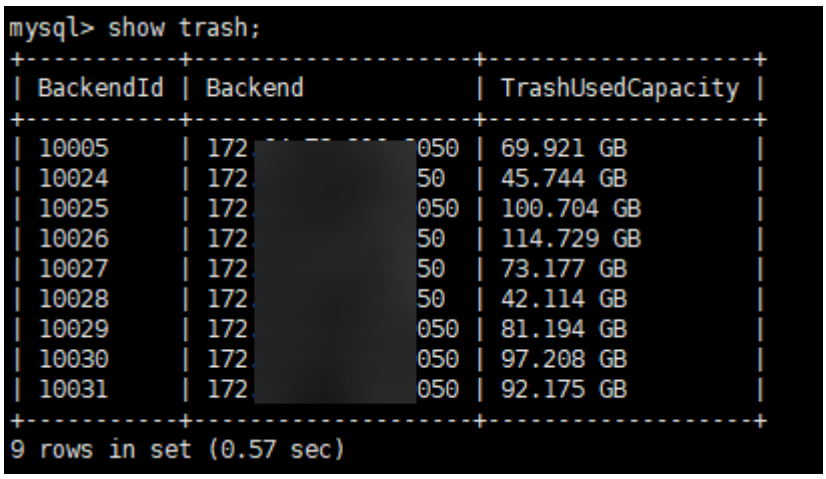
- An overabundance of junk files can consume disk space, limiting availability for valid data and potentially causing data loss.

- These files also serve a protective function against accidental deletion, allowing Doris to recover data. A minimal number of junk files could compromise this safeguard.

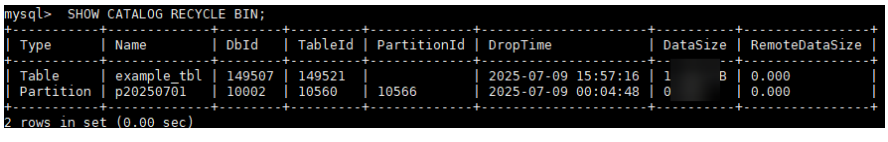
Viewing the Data in the Recycle Bin

1. Log in to the CloudTable console.
2. **Create a Doris cluster.**
3. **Connect to a Doris cluster.**
4. Run this command to view and restore the recycle bin data.
   - For version 2.0.x, run the following commands to view and restore data in the recycle bin:
     - View the data in the recycle bin.
       ```
       show trash;
       ```

       Figure 5-1 Data in the recycle bin

     - Restore data in the recycle bin.
       ```
       curl -X POST http://{be_host}:{be_webserver_port} /api/restore_tablet?tablet_id={tablet_id}
       \&schema_hash={schema_hash}
       ```
       - **be_host**: node address.
       - **be_webserver_port**: port number of the node.
   - For versions 2.1.X, run the following commands to view and restore data in the recycle bin:
     - View the data in the recycle bin.
       ```
       SHOW CATALOG RECYCLE BIN;
       ```

     - Restore data in the recycle bin.
       ```
       RECOVER TABLE demo.example_tbl;
       ```


Setting the Recycle Bin Duration

- The principle behind the recycle bin is that deleted data is not immediately removed from the disk. It is first moved to the recycle bin. Once the designated timeout period lapses, the data is then permanently erased from the disk.
- When setting the recycle bin duration, consider the following:
  - A prolonged recycle bin duration leads to junk file buildup, occupying disk space.
  - If the recycle bin timeout is extensive, calling the **admin clean trash** command might cause data imbalances, triggering further data balancing and junk file generation.
  - A brief recycle bin duration risks accidental tablet deletions or the inability to recover data in exceptional cases. You are advised to gauge the average disk space the recycle bin occupies based on actual service usage and set an appropriate timeout period that balances space utilization with a safety window to prevent accidental deletions.

    ```
    curl -X POST http://{be_ip}:{be_http_port}/api/update_config?trash_file_expire_time_sec={value}
    \&persist=true
    ```

    - **be_host**: node address.
    - **be_webserver_port**: port number of the node.
    - **trash_file_expire_time_sec**: interval for clearing the recycle bin, which is 72 hours. When the disk space is insufficient, the storage period of files in the trash directory does not comply with this parameter. The default value is **259200**.

5.4 What Should I Do If the Error Message "Failed to create partition [xxx] . Timeout" Is Displayed When I Create a Doris Data Table?

Symptom

The error message "Failed to create partition [xxx] . Timeout" is displayed during Doris data table creation.

Solution

Doris tables are created in sequence based on the partition granularity. This error message may appear when a partition creation task fails, but it could also appear in table creation tasks with no partitioning operations, because, as mentioned earlier, Doris will create an unmodifiable default partition for tables with no partitions specified.

This error usually pops up when the data fragment creation goes wrong in the BE nodes. You can follow the steps below for troubleshooting:

- In fe.log, search for the **Failed to create partition** log at the corresponding time point. The log file contains a series of number pairs that look like

{10001-10010}. The first number of the pair is the Backend ID and the second number is the Tablet ID. As for **{10001-10010}**, it means that on Backend whose ID is **10001**, the creation of Tablet whose ID is **10010** failed.

- Find the **be.INFO** log of the corresponding backend and search for the tablet ID-related records in the corresponding time period. The error information can be found.
- Common tablet creation failures include:
  - If the BE does not receive the creation task, no tablet ID-related records can be found in the **be.INFO** log or the table is successfully created on the BE but reported as failed.
  - Pre-allocated memory failure. The byte length of a row in the table may exceed 100 KB.
  - Too many open files. The number of opened file handles exceeds the Linux system limit. You need to modify the number of handles allowed by the Linux system.

  If the data fragment creation times out, you can also extend the timeout interval by setting **tablet_create_timeout_second=xxx** and **max_create_table_timeout_second=xxx** in **fe.conf**. The default value of **tablet_create_timeout_second** is 1 second, and that of **max_create_table_timeout_second** is 60 seconds. The overall timeout would be **min(tablet_create_timeout_second * replication_num, max_create_table_timeout_second)**.

5.5 What Can I Do If No Result Is Returned When I Create a Doris Table?

Symptom

No result is returned during Doris table creation.

Possible Causes

The Doris table creation command is a synchronous command. The timeout interval of this command is simple, that is, *tablet num* × *replication num*, in seconds. If a large number of data fragments are created and some fragments fail to be created, an error may be returned after a long timeout.

Solution

In normal cases, table creation statements return results within a few seconds or a dozen or so seconds. If the timeout interval is longer than one minute, you are advised to cancel the operation and check the errors in the logs of the FE or BE node.

6 ClickHouse

6.1 How Do I Delete the Copy Table of the ZooKeeper Node in a ClickHouse Cluster with Coupled Storage and Compute?

Symptom

When you connect a ClickHouse cluster and create a table, the ZooKeeper node does not get fully removed after you delete the table.

Figure 6-1 Execution results

```
ENGINE = MergeTree
ORDER BY id

Query id: 9765507e-bd39-4409-9810-1d8ca82fda4d

┌─host──────────┬─port─┬─status─┬─error──────────────────────────────────────────────────────────────────────────────────────────────────────┬─num_hosts_remaining─┬─num_hosts_active─┐
│ 192.168.5.11  │ 9000 │      0 │                                                                                                        │                   1 │                0 │
│ 192.168.5.117 │ 9000 │     57 │ Code: 57. DB::Exception: Table tt.table1 already exists. (TABLE_ALREADY_EXISTS) (version 22.3.2.1)     │                   0 │                0 │
└───────────────┴──────┴────────┴────────────────────────────────────────────────────────────────────────────────────────────────────────┴─────────────────────┴──────────────────┘

↘ Progress: 0.00 rows, 0.00 B (0.00 rows/s., 0.00 B/s.)  0%
2 rows in set. Elapsed: 0.113 sec.
```

Possible Causes

The command for creating a table contains **ON CLUSTER default_cluster**. Therefore, a local table is created on each node.

Solutions

Delete the tables and stored data on other nodes.

Step 1 Create a database.

```
create database demo ON CLUSTER default_cluster;
```

Step 2 Create tables in the database.

```
use demo;
```

Create the **test** table.

```
CREATE TABLE demo.test ON CLUSTER default_cluster(`EventDate` DateTime, `id` UInt64)ENGINE = ReplicatedMergeTree('/clickhouse/tables/{shard}/default/test', '{replica}') PARTITION BY toYYYYMM(EventDate) ORDER BY id;
```


Step 3 Delete the table.

```
drop table test SYNC;
```

Step 4 The following error message is displayed if a table with the same name as the deleted one is created:

```
┌─host─────────────┬─port─┬─status─┬─error──────────────────────────────────────────────────────────────────┬─num_hosts_remaining─┬─num_hosts_active─┐
│ 192.168.2.185 │ 9000 │     57 │ Code: 57. DB::Exception: Table demo.test already exists. (TABLE_ALREADY_EXISTS) (version 22.3.2.1) │                   1 │                1 │
└──────────────┴──────┴────────┴────────────────────────────────────────────────────────────────────────┴─────────────────────┴──────────────────┘
┌─host──────────┬─port─┬─status─┬─error─┬─num_hosts_remaining─┬─num_hosts_active─┐
│ 192.168.2.16 │ 9000 │      0 │       │                   0 │                0 │
└──────────────┴──────┴────────┴───────┴─────────────────────┴──────────────────┘
```

Step 5 Method 1: Run the following command to delete the tables and related data on each node:

```
drop table test ON CLUSTER default_cluster SYNC;
```

No error occurs if a table with the same name as the deleted one is created.

Step 6 Method 2: Delete the table on other nodes.

```
drop table test SYNC;
```

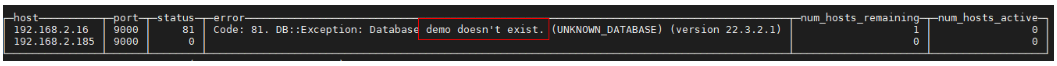
----End

6.2 What Should I Do If a Database Missing Error Occurs When a Table Is Created in the ClickHouse Cluster with Coupled Storage and Compute?

Symptom

When a table is created, the system displays a message indicating that the database does not exist on another node.

Figure 6-2 Execution results



Possible Causes

A database is created only on one node when the cluster is created.

Solutions

Create a database on another code or use another command to create one on each node.

Method 1:

Step 1 Create a database.

```
create database demo;
```


Use the **demo** database.

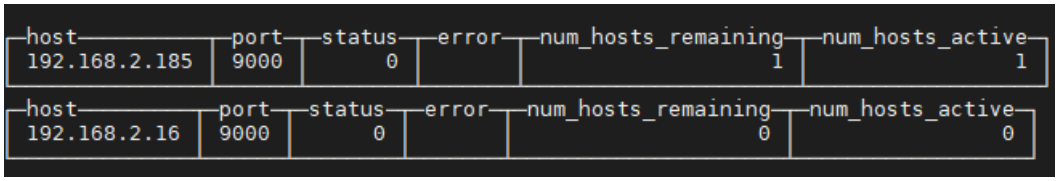
```
use demo;
```

Step 2 Create the **test** table. The result shown in **Figure 1** is displayed. The **demo** database does not exist on other nodes.

```
CREATE TABLE demo.test ON CLUSTER default_cluster(`EventDate` DateTime, `id` UInt64)ENGINE = ReplicatedMergeTree('/clickhouse/tables/{shard}/default/test', '{replica}') PARTITION BY toYYYYMM(EventDate) ORDER BY id;
```

Step 3 Create a database **demo** on another node and create a table. If the information shown in the following figure is displayed, no error occurs.

Figure 6-3 Successful table creation



----End

Method 2:

Step 1 Delete the existing database and use the following command to create a database on each node.

```
create database demo ON CLUSTER default_cluster;
```

Step 2 Create a table and the error message shown in the **Figure** above is not displayed.

----End

