

CloudTable Service

User Guide

Issue 01
Date 2024-06-20



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 HBase User Guide.....	1
1.1 HBase Cluster Management.....	1
1.1.1 Overview.....	1
1.1.2 Managing Clusters.....	1
1.1.2.1 Creating an HBase Cluster.....	1
1.1.2.2 Introduction to HBase Cluster Management.....	6
1.1.2.3 Viewing Basic HBase Cluster Information.....	9
1.1.2.4 Configuring HBase Parameters.....	11
1.1.2.5 HBase Cluster Scale-Out.....	16
1.1.2.6 Restarting a Cluster.....	19
1.1.2.7 Deleting a Cluster.....	20
1.1.3 Preparing an ECS.....	21
1.1.4 Using HBase.....	26
1.1.4.1 Using HBase Shell to Access a Cluster.....	26
1.1.4.2 Introducing HBase Shell Commands.....	28
1.1.5 HBase Thrift Server.....	30
1.1.5.1 Thrift Server Overview.....	30
1.1.5.2 Using Thrift Server.....	31
1.1.6 HBase Security Channel Encryption.....	31
1.2 HBase Batch Data Import.....	33
1.2.1 Using CDM to Migrate Data to CloudTable.....	33
1.2.2 Using the Import Tool to Importing Data.....	37
1.2.3 Using CopyTable to Import Data.....	38
1.2.4 Sample Code for Copying a Table.....	39
1.3 HBase Monitoring Clusters.....	42
1.3.1 HBase Cluster Supported Metrics.....	42
1.3.2 Setting Alarm Rules.....	49
1.3.3 Viewing Cluster Monitoring Information.....	54
1.3.4 HBase Monitoring Page.....	56
1.4 HBase Audit Logs.....	56
1.4.1 Supported Key Audit Log Operations.....	57
1.4.2 Viewing Audit Logs.....	57
1.5 Self-Healing from HBase Hotspotting.....	59

1.6 Global Secondary Indexes.....	60
1.6.1 Introduction.....	60
1.6.2 Restrictions.....	61
1.6.3 Using the GSI Tool.....	62
1.6.3.1 Creating Indexes.....	62
1.6.3.2 Querying Indexes.....	63
1.6.3.3 Deleting Indexes.....	64
1.6.3.4 Changing Index Status.....	64
1.6.4 GSI APIs.....	66
1.6.5 Querying Data with Indexes.....	67
2 ClickHouse User Guide.....	69
2.1 ClickHouse Cluster Management.....	69
2.1.1 Overview.....	69
2.1.2 Managing ClickHouse Clusters.....	69
2.1.2.1 Creating a ClickHouse Cluster.....	69
2.1.2.2 Introduction to ClickHouse Cluster Management.....	75
2.1.2.3 Viewing the Basic Information of a ClickHouse Cluster.....	77
2.1.2.4 Restarting a ClickHouse Cluster.....	80
2.1.2.5 Deleting a ClickHouse Cluster.....	81
2.1.2.6 Scaling Out a ClickHouse Cluster.....	82
2.1.2.7 ClickHouse Disk Capacity Expansion.....	83
2.1.2.8 ClickHouse Specification Change.....	85
2.1.2.9 Changing the ClickHouse Billing Method to Yearly/Monthly.....	87
2.1.3 Using ClickHouse.....	88
2.1.3.1 Preparing an ECS.....	88
2.1.3.2 Using a Client to Access a ClickHouse Cluster.....	90
2.1.3.3 Using Clickhouse Commands.....	92
2.1.4 Migrating ClickHouse Data.....	93
2.1.5 Configuring Secure Channel Encryption for ClickHouse Clusters.....	96
2.1.6 Application of ClickHouse Cold and Hot Data Separation.....	98
2.1.6.1 Overview.....	98
2.1.6.2 Using ClickHouse Cold and Hot Data Separation.....	99
2.1.6.3 Commands Related to ClickHouse Cold and Hot Data Separation	100
2.2 ClickHouse Monitoring Clusters.....	101
2.2.1 ClickHouse Cluster Supported Metrics.....	101
2.2.2 Setting Alarm Rules.....	108
2.2.3 Viewing Monitoring Information of a Cluster.....	113
2.2.4 Checking ClickHouse Monitoring Metrics.....	114
3 Permissions Management.....	116
3.1 Creating a User and Granting Permissions.....	116
A Change History.....	119

1 HBase User Guide

1.1 HBase Cluster Management

1.1.1 Overview

The CloudTable cluster mode provides a distributed, scalable, 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.

In CloudTable HBase cluster management, you need to create a dedicated cluster and can use it on demand. Dedicated clusters are physically isolated and are not affected by other services, facilitating user management. The CloudTable HBase cluster management is applicable to users with high service throughput and low delay requirements.

After a CloudTable HBase cluster is created, you can access it using the following methods:

- Using a client to access the cluster
For details, see [Using HBase](#). This section describes how to use HBase Shell to access a cluster.
- Call native open source APIs to develop HBase applications, and access the cluster through a CloudTable link.
For details, see CloudTable Service Developer Guide.

1.1.2 Managing Clusters

1.1.2.1 Creating an HBase Cluster

You can centrally manage clusters with CloudTable. A cluster is necessary for using CloudTable. This section describes how to create a cluster on the CloudTable console.

HBase clusters support two billing modes: pay-per-use and yearly/monthly. The pay-per-use billing mode is used by default when you create a cluster. In pay-per-

use mode, compute units are charged by the duration you use them, with a billing cycle of one hour. With this mode, you can start or stop a cluster at any time and pay what you use. The yearly/monthly billing mode is recommended for long-term users to enjoy a larger discount. You can also customize a CloudTable HBase cluster with specified computing capabilities and storage space to meet your business needs.

 **NOTE**

Created HBase clusters can be accessed without passing Kerberos authentication. If you have requirements on cluster access security, you are advised to use the HBase component on MRS.

Creating a Cluster


- Step 1** Log in to the CloudTable console.
- Step 2** Click  in the upper left corner to select a region.
- Step 3** Click **Purchase Cluster**. The **Purchase Cluster** page is displayed.
- Step 4** Configure basic cluster information by referring to the following table.

Figure 1-1 Configuring basic cluster information

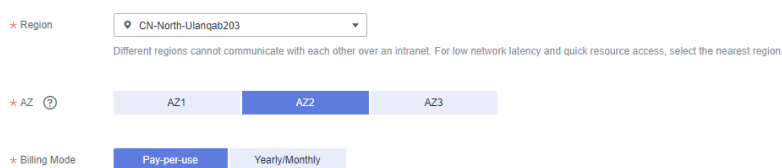


Table 1-1 Parameters

Parameter	Description
Region	Working zone of the cluster. For more information about regions, see Regions and Endpoints .
AZ	Select the AZ associated with the cluster's region. For more information, see Regions and AZs .
Billing Mode	Select Pay-per-use or Yearly/Monthly .
Required Duration	This option is available only when Billing Mode is set to Yearly/Monthly . Configure this parameter based on your service requirements.
Auto-renew	If you select Auto-renew when creating a cluster, the system will automatically renew your subscription before it expires.

Figure 1-2 Cluster configuration

* Name ?

* VPC ? View VPC

* Subnet ? View Subnet

* Security Group ? View Security Group

* Database Engine

Doris
 HBase
 ClickHouse

* HBase Version

Table 1-2 Network configuration parameters

Parameter	Description
Name	Name of a cluster. A cluster name must start with a letter and contain 4 to 32 letters, digits, and hyphens (-). It cannot contain other special characters.
VPC	A VPC is a secure, isolated, and logical network environment. You can select an existing VPC or click View VPC to create a new one. For details about how to create a VPC, see User Guide > VPC and Subnet > Creating a VPC in Virtual Private Cloud .
Subnet	A subnet provides dedicated network resources that are logically isolated from other networks, improving network security. A subnet is automatically created when a VPC is created. If you want to create a subnet, you can refer to Virtual Private Cloud User Guide > VPC and Subnet > Creating a Subnet for the VPC . NOTE When creating a cluster, ensure that the network segment of the subnet is within the VPC network segment. For details, see VPC and Subnet Planning Suggestions .

Parameter	Description
Security Group	<p>A security group is used to control ECS access within a security group or between security groups by defining access rules. You can define different access control rules for a security group. These rules can specify which ECS ports or protocols are accessible and can be used to control inbound and outbound network traffic of VMs. After an ECS is added to the security group, it is protected by these access control rules. ECSs that do not belong to the security group cannot communicate with ECSs in the security group.</p> <p>The underlying compute unit of a CloudTable cluster is ECS. In terms of security and stable service running, the ECSs must be added to the same security group in the same VPC. VPCs isolate networks and security groups specify which ports and protocols can be opened in the VPC.</p> <p>You can use an existing security group or click View Security Group to create a new one.</p> <p>For more information about security groups, see Security Group in the <i>Virtual Private Cloud User Guide</i>.</p>
Database Engine	Select the type of cluster to be created.

Figure 1-3 Master/Core nodes

Master Nodes

* Compute 4*CU₄ (One CU₄=1U4GB)

The cluster is configured in HA mode by default. Two master nodes are used. One master node is provided for free for cluster monitoring.

Core Nodes

* Compute 4*CU₄ (One CU₄=1U4GB)

* Storage GB / Nodes

* Nodes

Dual copies are used by default. The storage capacity is the capacity needed for both copies.

Advanced Feature Enable Hot/Cold ?

Table 1-3 Master node configuration

Parameter	Description
Compute	<p>Select proper compute specifications based on service requirements. Each specification displays the CPU, memory, and recommended application scenarios of a single node.</p> <p>NOTE</p> <ul style="list-style-type: none"> Available compute specifications: <ul style="list-style-type: none"> 4U16G 8U16G 8U32G 16U32G 16U64G 32U64G
Storage	Ultra-high I/O is recommended.
Nodes	Use the default value 2 .

Table 1-4 Core node configuration

Parameter	Description
Compute	<p>Select proper compute specifications based on service requirements. Each specification displays the CPU, memory, and recommended application scenarios of a single node.</p> <p>NOTE</p> <ul style="list-style-type: none"> Available compute specifications: <ul style="list-style-type: none"> 4U16G 8U16G 8U32G 16U32G 16U64G 32U64G
Storage	<p>Select the disk specifications and capacity of the HBase compute node.</p> <p>NOTE</p> <ul style="list-style-type: none"> Available storage specifications: <ul style="list-style-type: none"> Common I/O High I/O Ultra-high I/O The capacity ranges from 400 GB to 30,000 GB per node.
Nodes	Set the number of nodes in a cluster. The value ranges from 2 to 20 .

Parameter	Description
Advanced Feature	Whether to enable the cold and hot separation feature. You can toggle it on to enable this feature.
Enterprise Project	<p>You can group related cloud resources (for example, resources used for the same purpose) and manage them by enterprise project.</p> <p>NOTE</p> <ul style="list-style-type: none"> You can delete a user that was authorized to manage an enterprise project if the enterprise's business has changed and the user should not have the permissions to manage the enterprise project. You can delete a user or multiple users. After a user that was authorized to manage an enterprise project is deleted, the user cannot manage the enterprise project. If the user needs to manage the enterprise project, authorize the user to manage the enterprise project again.

Step 5 Click **Next**.

Step 6 Confirm the details of the order and click **Submit**. The cluster creation task is submitted successfully.

Step 7 Click **Back to Cluster List** to view the cluster status.

The cluster creation task takes some time. The initial status of the cluster is **Creating**. After the cluster is created, the cluster status changes to **In service**.

Step 8 Submit the creation task of a yearly/monthly cluster.

Click **Pay**. On the displayed purchase page, confirm the information, select a proper payment method, and confirm the payment.

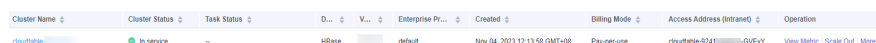
Return to the console and check the cluster status. Cluster creation takes some time. The initial status of the cluster is **Creating**. After the cluster is created, the cluster status changes to **In service**.

----End

1.1.2.2 Introduction to HBase Cluster Management

Log in to the CloudTable console. In the left navigation pane, click **Cluster Management**. On the **Cluster Management** page, a cluster list is displayed. If there are a large number of clusters, you can turn pages to view clusters in any state.

Figure 1-4 Cluster list







Clusters are listed in chronological order by default, with the latest cluster displayed at the top. [Table 1-5](#) and [Table 1-6](#) provides parameters and icons in the cluster list.

Table 1-5 Cluster management parameters

Parameter	Description
Cluster Name	Name of a cluster. Set this parameter when creating a cluster.
Cluster Status	Cluster status, including Creating , In service , Sub-health , and Creation failed , and Frozen .
Task Status	Task status of a cluster For details, see Table 1-8 .
Database Engine	HBase and ClickHouse
Cluster Version	Cluster kernel version
Enterprise Project	<p>You can group related cloud resources (for example, resources used for the same purpose) and manage them by enterprise project.</p> <p>NOTE</p> <ul style="list-style-type: none"> You can delete a user that was authorized to manage an enterprise project if the enterprise's business has changed and the user should not have the permissions to manage the enterprise project. You can delete a user or multiple users. After a user that was authorized to manage an enterprise project is deleted, the user cannot manage the enterprise project. If the user needs to manage the enterprise project, authorize the user to manage the enterprise project again.
Created	Time when a cluster is created
Billing Mode	Cluster billing mode.
ZK Link (Intranet)	ZooKeeper address
Operation	<ul style="list-style-type: none"> Click Monitor to access the CloudTable service monitoring page. Scale Out: Increase compute units in the cluster. For details, see HBase Cluster Scale-Out. More <ul style="list-style-type: none"> Restart: Click Restart to restart a cluster. For details, see Restarting a Cluster. Delete: You can click Delete to delete a cluster. For details, see Deleting a Cluster.

Table 1-6 Icon description

Icon	Description
▼	Click ▼ to view all projects.

Icon	Description
	Enter a cluster name in the search box and click  to search for the cluster.
	Click  to manually refresh the cluster list.

Cluster Status

Table 1-7 Cluster status description

Status	Description
Creating	Indicates that a cluster is being created.
In service	If a cluster is successfully created and can provide services, the cluster status is In service .
Sub-health	If the cluster status cannot be monitored within the specified time, the cluster status changes to Sub-health . Manual intervention is required to recover a cluster that is in Sub-health status. For example, you can restart the cluster to recover the cluster.
Creation failed	Indicates that a cluster fails to be created.
Frozen	If the balance is insufficient for renewing a cluster, the cluster status is Frozen . To unfreeze the cluster, you need to recharge your account to ensure that the account balance is not 0 . NOTE A frozen cluster is unavailable and its all ECSs are shut down. After being unfrozen, the cluster recovers to the In service state. If you do not renew the cluster before the freeze period ends, the cluster will be deleted.

Task Status

Table 1-8 Task status description

Status	Description
Deleting	Indicates that a cluster is being deleted.
Restarting	Indicates that a cluster is being restarted.

1.1.2.3 Viewing Basic HBase Cluster Information

You can monitor and manage the clusters you create. On the CloudTable console, click **Cluster Management**. In the cluster list, locate the cluster to be viewed and click the cluster name to access the basic information page. You can view the cluster information and network configurations.

Table 1-9 and **Table 1-10** describe parameters about basic cluster information.

Table 1-9 Cluster information


Parameter	Description
Cluster Name	Name of a cluster. Set this parameter when creating a cluster.
Cluster ID	Cluster ID
Cluster Status	Cluster status information
Used Storage Capacity	Used storage capacity
Billing Mode	Billing mode of the cluster
Version	Cluster kernel version
Created	Time when a cluster is created
Enterprise Project	<p>You can group related cloud resources (for example, resources used for the same purpose) and manage them by enterprise project.</p> <p>NOTE</p> <ul style="list-style-type: none"> You can delete a user that was authorized to manage an enterprise project if the enterprise's business has changed and the user should not have the permissions to manage the enterprise project. You can delete a user or multiple users. After a user that was authorized to manage an enterprise project is deleted, the user cannot manage the enterprise project. If the user needs to manage the enterprise project, authorize the user to manage the enterprise project again.
ZK Link (Intranet)	<p>ZooKeeper address.</p> <p>You can click  to copy the ZooKeeper link to the clipboard.</p>
Database Engine	The value is HBase .

Table 1-10 Network configuration

Parameter	Description
Region	Working area of the cluster. Set this parameter when creating a cluster.

Parameter	Description
AZ	AZ you select during cluster creation
VPC	VPC you select during cluster creation A Virtual Private Cloud (VPC) is a secure, isolated, logical network environment.
Subnet	Subnet you select during cluster creation A subnet provides dedicated network resources that are logically isolated from other networks, improving network security.
Security Group	Security group you select during cluster creation

Table 1-11 Master node configuration

Parameter	Parameters
Compute	Select proper compute specifications based on service requirements. Each specification displays the CPU, memory, and recommended application scenarios of a single node. NOTE <ul style="list-style-type: none"> • Available compute specifications: <ul style="list-style-type: none"> • 4U16G • 8U16G • 8U32G • 16U32G • 16U64G • 32U64G
Storage	Ultra-high I/O is recommended.
Nodes	Use the default value 2 .

Table 1-12 Core node configuration

Parameter	Description
Compute	Select proper compute specifications based on service requirements. Each specification displays the CPU, memory, and recommended application scenarios of a single node. NOTE <ul style="list-style-type: none">Available compute specifications:<ul style="list-style-type: none">4U16G8U16G8U32G16U32G16U64G32U64G
Storage	Select the disk specifications and capacity of the HBase compute node. NOTE <ul style="list-style-type: none">Available storage specifications:<ul style="list-style-type: none">Common I/OHigh I/OUltra-high I/OThe capacity ranges from 400 GB to 30,000 GB per node.
Nodes	Set the number of nodes in a cluster. The value ranges from 2 to 20 .

1.1.2.4 Configuring HBase Parameters

Scenario

A CloudTable HBase cluster has many HBase parameters related to read/write performance. You can modify the HBase parameters to tune cluster performance under different read/write request loads. You need to restart the cluster to make changes take effect.

Prerequisites


No task is running in the cluster.

Procedure



NOTE

- Restart the CloudTable HBase cluster after parameter configuration modification. Otherwise, services will be interrupted.
- Do not modify cluster parameters when the CloudTable HBase cluster is being restarted.

Step 1 Log in to the CloudTable management console.

- Step 2** Click  in the upper left corner to select a region.
- Step 3** In the navigation tree, click **Cluster Management**.
- Step 4** Click the name of a cluster for which you want to modify HBase parameters to access the basic cluster information page.
- Step 5** In the **Parameter Configuration** area, click the **Parameters** tab to modify HBase parameters.

For details about the HBase parameters you can modify, see [HBase Parameters](#).

1. Select the target parameter and click  in the **Value** column.
2. Enter a new value in the text box, and then click . The parameter value has been successfully changed if the system prompts "The parameter changed to xx successfully. Save the modified value." The new parameter value is marked with a red asterisk (*).


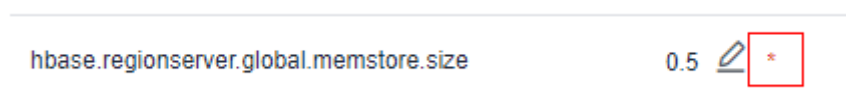
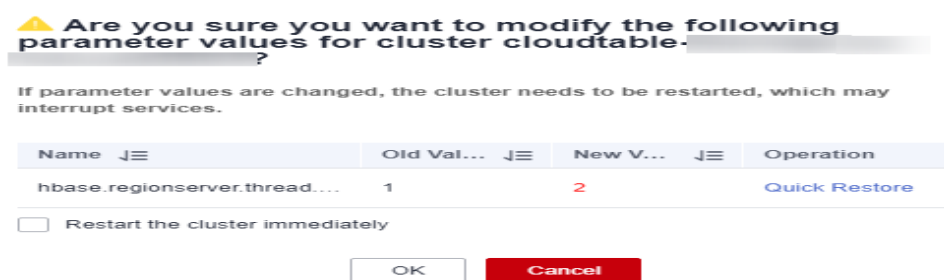
If you want to cancel it, click .


Figure 1-5 Modifying HBase parameters



3. Click **Save Changes** in the upper left corner of the parameter list. The **Save Changes** dialog box is displayed.
4. In the **Save Changes** dialog box, verify whether the parameter settings are correct. Select **Restart the cluster immediately** and click **OK**.

Figure 1-6 Save changes



- After you select **Restart the cluster immediately** and click **OK**, the cluster restarts immediately. After the cluster restart is complete, the new parameter configurations take effect. Click  above the parameter list. The system prompts "The new value is applied" in the upper left corner of the parameter list.
- If you do not select **Restart the cluster immediately** and click **OK** in the **Save Changes** dialog box, the system prompts "The new value is not applied" above the parameter list. In this case, you need to restart the

cluster for the changes to take effect. For details about how to restart a cluster, see [Restarting a Cluster](#).

- If the new parameter value is incorrect, click **Quick Restore** to cancel the modification.

Step 6 After modifying parameters, you can click the **Change History** tab to view the change history.

On the **Change History** tab page, you can view the following information.

- **Name:** Name of the changed parameter.
- **Old Value:** old parameter value
- **New Value:** new parameter value
- **Modification Time:** time when you modify a parameter value

----End

HBase Parameters

Table 1-13 lists HBase parameters you can modify currently.

NOTE

The sum of the values of `hbase.regionserver.global.memstore.size` and `hfile.block.cache.size` cannot exceed 0.8.

Table 1-13 HBase parameters

Parameter	Default Value	Unit	Value Range	Default Value	Description
<code>hbase.regionserver.thread.compaction.small</code>	1	--	[1,20]	1	Indicates the number of HFile compaction threads. You can increase the parameter value in heavy-put-load scenarios.
<code>hbase.regionserver.global.memstore.size</code>	0.4	--	(0,0.8)	0.4	It is recommended that you set this parameter to "hbase.hregion.memstore.flush.size x Number of regions with active writes/ RegionServer GC -Xmx)." The default value is 0.4 , indicating that 40% of RegionServer GC -Xmx is used.

Parameter	Default Value	Unit	Value Range	Default Value	Description
hbase.hstore.blockingStoreFiles	60	--	[1,2147 483647]	60	When the HFile number in the column cluster reaches this threshold, all operations in the region are blocked until the compaction is complete. You can increase the parameter value in heavy-put-load scenarios.
hbase.client.scanner.timeout.period	60000	ms	[1,2147 483647]	60000	A client and RegionServer parameter, indicating the scan lease period. It is recommended that you set this parameter to an integral multiple of 60000 ms, and increase the parameter value in heavy-read-load scenarios. Unit: millisecond
hfile.block.cache.size	0.2	--	(0,0.8)	0.2	Indicates the data cache percentage in the RegionServer GC -Xmx. You can increase the parameter value in heavy-read-load scenarios, in order to improve cache hit ratio and performance. The default value is 0.2, indicating that 20% of RegionServer GC -Xmx is used.
hbase.regionserver.handler.count	100	--	[1,300]	100	Indicates the number of RPC server instances on the RegionServer. The recommended value ranges from 100 to 300.
hbase.regionserver.metahandler.count	50	--	[1,100]	50	Indicates the number of program instances for processing prioritized requests. The recommended value ranges from 20 to 100.
hbase.hstore.flusher.count	2	--	[1,10]	2	Indicates the number of memstore flush threads. You can increase the parameter value in heavy-put-load scenarios.

Parameter	Default Value	Unit	Value Range	Default Value	Description
hbase.ipc.server.callqueue.read.ratio	0.5	--	[0,1]	0.5	<p>When used under different load models, it controls the ratio between the numbers of read and write RPC queues. The value ranges from 0 to 1.0, and the default value is 0.5.</p> <p>0 indicates the total RPC queues of read and write operations.</p> <p>If the value is less than 0.5, it indicates that the read load is less than the write load.</p> <p>0.5 indicates that the read load equals to the write load.</p> <p>If the value is greater than 0.5, it indicates that the read load is greater than the write load.</p> <p>1.0 indicates that all RPC queues except one are used for read operations.</p>
hbase.regionserver.hotregion.handler.count	66	--	[1,65535]	66	Number of RPC listener instances started on RegionServers for hotspot regions.
hbase.ipc.server.hotregion.max.callqueue.length	330	--	[1,65535]	330	Maximum length of the queue for RegionServers to process requests of hotspot regions. Upon receiving a new request, the system checks whether the length of the queue exceeds the threshold. If so, the request is discarded.
hbase.metric.controller.analysis.period	60	Second	[1,2147483647]	60	Hotspot analysis period of MetricController, in seconds.
hbase.metric.controller.analysis.threads.max	10	--	[1,100]	10	Maximum number of threads for hotspot analysis in the thread pool.

Parameter	Default Value	Unit	Value Range	Default Value	Description
hbase.metric.controller.collect.threads.max	16	--	[1,100]	16	Maximum number of threads for hotspot analysis in the traffic collection thread pool.
hbase.metric.regionserver.hotspot.threshold	20000	--	[1,2147483647]	20000	Hotspot threshold of a RegionServer. Unit: requests per second.
hbase.metric.region.hotspot.threshold	10000	--	[1,2147483647]	10000	Hotspot threshold of a single region. Unit: requests per second.
hbase.hotspot.enable	false	--	[true,false]	false	Whether to enable hotspot self-healing. The value true means to enable it, and false means to disable it. After this function is enabled, access hotspots will be automatically processed.
hbase.tries.cache.enabled	false	--	[true,false]	false	If this parameter is set to true, LruBlockCache is used to cache index blocks and data blocks.
hbase.write.tries	false	--	[true,false]	false	If this parameter is set to true, the succinct tries feature is enabled. In this case, a new data structure is used to improve the utilization of index blocks.

1.1.2.5 HBase Cluster Scale-Out

You can perform a scale-out to expand cluster capacity and improve system computing and storage capabilities. Scale-out operations incur charges. You can dynamically add nodes in response to site needs or service conditions. The cluster seamlessly manages load balancing to maintain uninterrupted service and facilitate smooth expansion of capacity.


Precautions

- During cluster capacity expansion, the system does not automatically restart the cluster, ensuring service continuity and smooth capacity expansion.
- If the scale-out fails, **Scale-out failed** is displayed in the **Task Status** column, and the cluster automatically rolls back to the state before the scale-out. You can try again. If you have any questions, contact technical support.

- Only core nodes can be scaled-out.

Increasing Computing Units

Step 1 Log in to the CloudTable console.

Step 2 Click  in the upper left corner to select a region.

Step 3 In the navigation tree, click **Cluster Management**.

Step 4 In the cluster list, locate the row that contains the target cluster, and choose **More** > **Scale Out**. The **Scale Out** page is displayed.

Figure 1-7 Scaling out

Scale Out

1 1. Perform scale-out during off-peak hours to avoid intermittent service interruptions

Cluster ID/Name	85e65565-b32e-44f4-8938-c14...
Region/Availability Zone	CN-North-Ulanqab203/AZ3
Billing Mode	Pay-per-use
Master Node Specification	4 vCPUs 16 GB
Master Storage Specifications	Ultra-high I/O 200 GB
Master Nodes	3
Core Node Specification	4 vCPUs 16 GB
Core Storage Specifications	High I/O 400 GB
Core Nodes	3
Configured Node	<input type="radio"/> Master Node <input checked="" type="radio"/> Core Node
Core Nodes Increased To	<input type="button" value="-"/> <input type="text" value="4"/> <input type="button" value="+"/>
Total Fee ?	¥8.57/hour

 NOTE

You can also click **Scale Out** on the cluster details page to go to the **Scale Out** page.

Figure 1-8 Accessing the Scale Out page

Core Nodes		
Compute	4 vCPUs 16 GB	4*CU4 (One CU4 =1U4GB)
Storage	High I/O 400GB	
Nodes	3	Scale Out

Step 5 Set **Configured Node** to **Core Node**.

Step 6 Click + to add nodes.

Step 7 Confirm the fee and resource quota, and click **OK**.

----End

1.1.2.6 Restarting a Cluster

If a cluster is in the unbalanced state or cannot work properly, you may need to restart it for restoration. After modifying a cluster's configurations, such as security settings and parameters, restart the cluster to make the configurations take effect.

 NOTE


If your cluster is in arrears, this function may be unavailable. Please top up your account in time.

Impact on the System

- A cluster cannot provide services during the restart. Therefore, before the restart, ensure that no task is running and all data is saved.
- If a cluster is processing transactional data, for example, importing data, querying data, files may be damaged or the cluster may fail to be restarted once the cluster is restarted. You are advised to stop all cluster tasks before restarting a cluster.
- If the restart fails, the cluster may be unavailable. Try again later or contact technical support.

Procedure

Step 1 Log in to the CloudTable management console.

Step 2 Click  in the upper left corner to select a region.

Step 3 In the navigation tree, click **Cluster Management**.


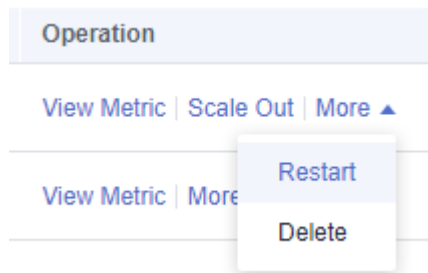
Step 4 In the upper right corner of the cluster list, enter the name of a cluster in the search box and click .

Figure 1-9 Searching for a cluster by name



Step 5 In the **Operation** column of the cluster, click **Restart**.

Figure 1-10 Restarting a cluster



Step 6 In the dialog box that is displayed, select the check box and click **OK** to restart the cluster.


----End

1.1.2.7 Deleting a Cluster

You can delete clusters you will no longer use. Deleting a CloudTable cluster will clear all resources and data related to the cluster. This operation cannot be undone. Exercise caution when deleting a cluster.

Procedure

Step 1 Log in to the CloudTable management console.

Step 2 Click  in the upper left corner to select a region.

Step 3 In the navigation tree, click **Cluster Management**.


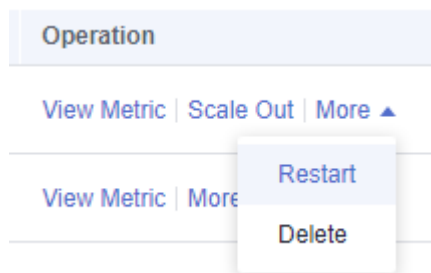
Step 4 In the upper right corner of the cluster list, enter the name of a cluster in the search box and click .

Figure 1-11 Searching for a cluster



Step 5 In the **Operation** column of the cluster, choose **More > Delete**.

Figure 1-12 Deleting a cluster



Step 6 In the displayed dialog box, enter **DELETE** or click **Auto Enter**, and click **OK** to delete the cluster.

Figure 1-13 Confirming the deletion



----End

1.1.3 Preparing an ECS

If the client tool runs on Linux, you need to prepare a Linux ECS that is in the VPC as a CloudTable cluster and the Linux ECS serves as a client host. If you use the following client tools to access a cluster, you are advised to use the HBase shell, a Linux ECS.

If the client tool runs on Windows, you need to prepare a Windows ECS that is in the VPC as a CloudTable cluster and the Windows ECS serves as a client host.

You can select an ECS of different images and functions to meet your needs.

Preparing an ECS

For details about how to purchase a Linux or Windows ECS, see how to [purchase an ECS](#) in the *Elastic Cloud Server Getting Started*.

The purchased ECS must meet the following requirements:

- The ECS must have the same region, AZ, VPC, and subnet as the CloudTable cluster.
For details about how to create a VPC, see **User Guide > VPC and Subnet in Virtual Private Cloud**.
- The ECS must have the same security group as the CloudTable cluster.

For more information about security groups, see [Security Group](#) in the *Virtual Private Cloud User Guide*.

 **NOTE**

When cross-VPC communication is used to access a CloudTable cluster, the network administrator needs to authorize the access to the VPC, security group, and subnet where the cluster resides.

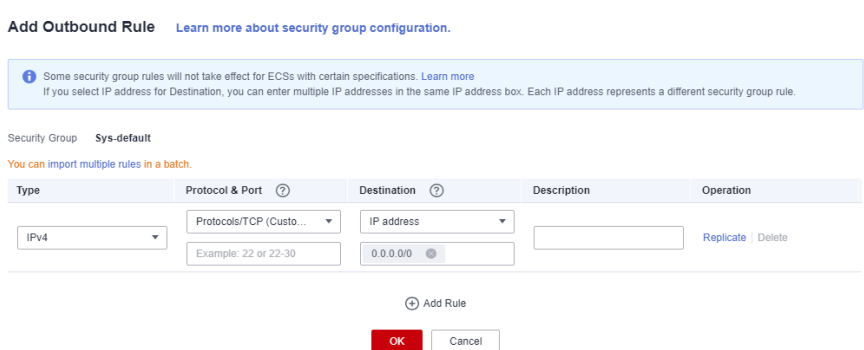
Check whether the outbound rule of the security group contains the following rule:

- Protocol & Port: All
- Type: IPv4
- Destination: 0.0.0.0/0

If the preceding rules do not exist, add the following outbound rule according to [Figure 1-14](#).

- Protocol: TCP
- Port: 80
- Destination: Select **IP address** and set it to **0.0.0.0/0**.

Figure 1-14 Adding an outbound rule



- When purchasing an ECS, you need to set **EIP** to **Automatically assign**. Alternatively, you can bind an EIP to an ECS after the ECS is created.
- To access a Linux ECS, you are advised to use an SSH password.
For details about how to log in to a Linux ECS, see [Logging In to a Linux ECS](#) in the *Elastic Cloud Server User Guide*.
- To access a Windows ECS, you are advised to use the MSTSC-based remote desktop connection tool.
For details, see [Logging In to a Windows ECS](#).

Configuring the DNS Address and hosts File for the Linux ECS

You do not need to perform this operation when you install the HBase shell by deploying a client in one click. In other cases, perform the following steps to configure the DNS address and the **hosts** file for the Linux ECS:

Step 1 Configure `/etc/hosts`.

Configure the host IP address in the `/etc/hosts` file to accelerate the access to the CloudTable cluster.

To configure /etc/hosts for a Linux ECS, perform the following steps:

1. Log in to the Linux ECS as user **root**.
2. Run the **hostname** command to query the host name.

```
[root@euler ~]# hostname  
euler.novalocal
```
3. Run the **ifconfig** command to query the IP address of the local host.

Figure 1-15 Querying the IP address

```
[root@euler ~]# ifconfig  
eth0: flags=4163<UP,BROADCAST,RUNNING,MULTICAST> mtu 1500  
    inet 192.168.0.58 netmask 255.255.255.0 broadcast 192.168.0.255  
    inet6 fe80::f816:3eff:fe8d:7acb prefixlen 64 scopeid 0x20<link>  
    ether fa:16:3e:8d:7a:cb txqueuelen 1000 (Ethernet)  
    RX packets 379443 bytes 522261185 (498.0 MiB)  
    RX errors 0 dropped 0 overruns 0 frame 0  
    TX packets 148302 bytes 10571485 (10.0 MiB)  
    TX errors 0 dropped 0 overruns 0 carrier 0 collisions 0  
  
lo: flags=73<UP,LOOPBACK,RUNNING> mtu 65536  
    inet 127.0.0.1 netmask 255.0.0.0  
    inet6 ::1 prefixlen 128 scopeid 0x10<host>  
    loop txqueuelen 0 (Local Loopback)  
    RX packets 23819 bytes 16629808 (15.8 MiB)  
    RX errors 0 dropped 0 overruns 0 frame 0  
    TX packets 23819 bytes 16629808 (15.8 MiB)  
    TX errors 0 dropped 0 overruns 0 carrier 0 collisions 0  
  
[root@euler ~]#
```

4. Run the **vi /etc/hosts** command to edit the file and add the host configuration.

```
192.168.0.58 euler.novalocal  
127.0.0.1 euler.novalocal
```
5. Press **Esc** and enter **:wq** to save the settings and exit.
6. Run the following command to check whether the IP address is successfully added:

```
cat /etc/hosts
```
7. Run the following command to check whether the host name can be resolved:

```
ping Host name
```

Step 2 Configure the DNS.

The DNS server is used to resolve the domain name in the CloudTable cluster link, for example, the ZooKeeper link. Set the private DNS server address based on the region where the CloudTable cluster is located. For details about the DNS server addresses in each region, see [What Are the Private DNS Server Addresses Provided by the DNS Service?](#) in the *Domain Name Service FAQs*.

To configure DNS for a Linux ECS, perform the following steps:

1. Log in to the Linux ECS as user **root**.
2. Run the **vi /etc/resolv.conf** command to edit the **/etc/resolv.conf** file and add the **nameserver** configuration.
Above the Internet IP address, add two lines to input the DNS server IP addresses in the following format.

```
nameserver 100.125.1.250
```
3. Press **Esc** and enter **:wq** to save the settings and exit.
4. Run the following command to check whether the IP address is successfully added:

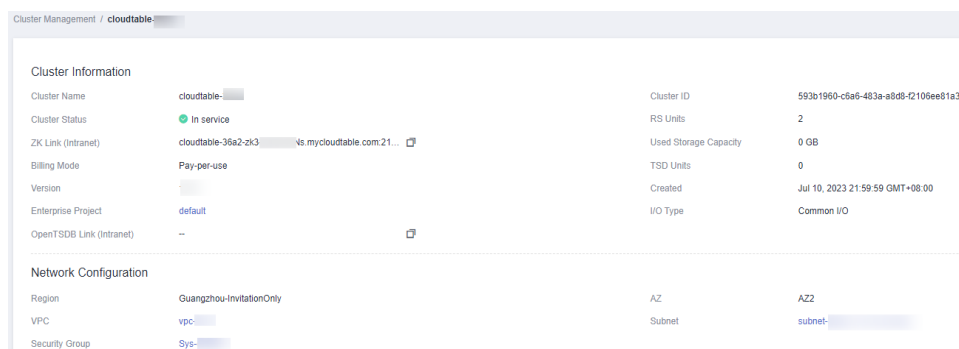
```
cat /etc/resolv.conf
```

5. Run the following command to check whether an IP address can be resolved from the access domain name.

```
ping Access domain name
```

Access domain name indicates the domain name in the CloudTable cluster link, for example, the ZooKeeper link. On the CloudTable management console, click the cluster name to go to the basic cluster information page and obtain the link of CloudTable. The obtained link contains three private domain names separated by commas (.). You can ping any of the private domain names.

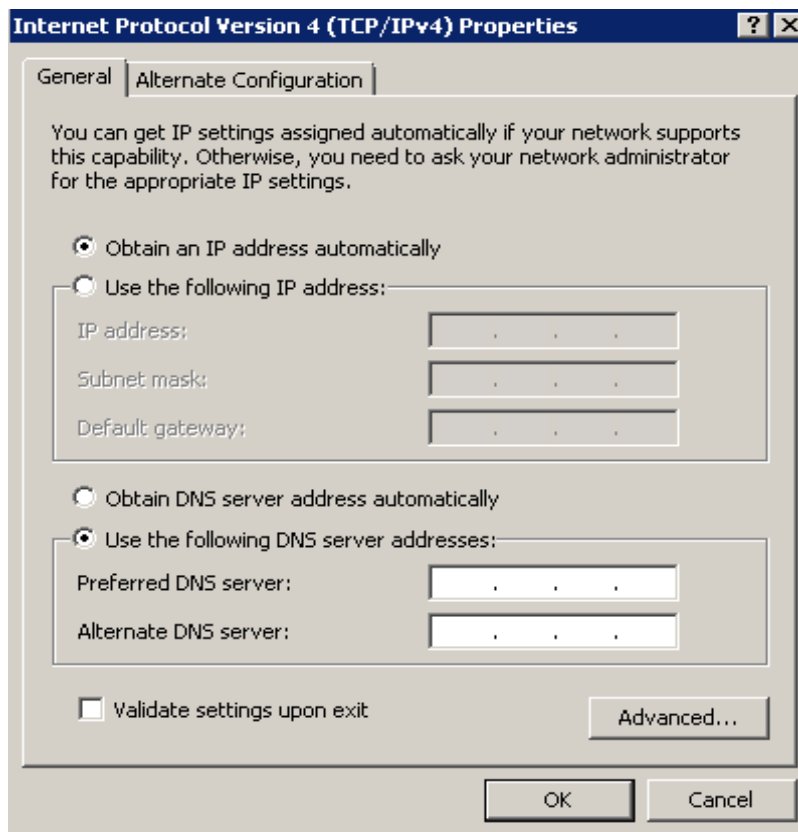
Figure 1-16 Viewing link addresses



----End

Configuring DNS Address for a Windows ECS

1. Log in to the Windows ECS as user **Administrator**.
2. On the Windows ECS, choose **Start > Control Panel > Network and Sharing Center**.
3. In the **Network and Sharing Center** window, click **Change adapter settings** on the right.
4. Right-click the network adapter name and choose **Properties** from the shortcut menu. The **Properties** dialog box is displayed.
5. Click **Internet Protocol Version 4 (TCP/IPv4)**. The **Internet Protocol Version 4 (TCP/IPv4) Properties** dialog box is displayed.

Figure 1-17 Internet protocol version 4 (TCP/IPv4) properties

- In the dialog box shown in [Figure 1-17](#), select **Use the following DNS server addresses**, enter the IP address of the DNS server in the **Preferred DNS server** textbox, and click **OK**.

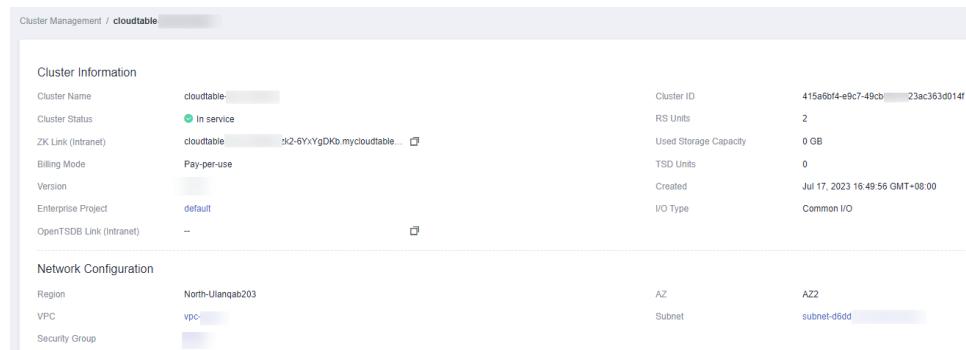
The DNS server is used to resolve the domain name in the CloudTable cluster link, for example, the ZooKeeper link. Set the private DNS server address based on the region where the CloudTable cluster is located. For details about the DNS server addresses in each region, see [What Are the Private DNS Server Addresses Provided by the DNS Service?](#) in the *Domain Name Service FAQs*.

- Click **Start**, enter **cmd** in the search box, and press **Enter**. The CLI is displayed.
- Run the following command on the CLI to check whether the DNS is successfully configured:

```
ping Access domain name
```

Access domain name indicates the domain name in the CloudTable cluster link, for example, the ZooKeeper link. On the CloudTable management console, click the cluster name to go to the basic cluster information page and obtain the link of CloudTable. The obtained links contain three private domain names separated by commas (.). You can ping any of the private domain names.

Figure 1-18 Checking link addresses



1.1.4 Using HBase

1.1.4.1 Using HBase Shell to Access a Cluster

You can use the HBase shell to access a cluster by **deploying a client in one click** or **manually installing a client** on an ECS. **Deploying a Client in One Click** is recommended.

Deploying a Client in One Click

Step 1 Prepare a Linux ECS.

Using a one-click client deployment tool. The recommended Linux ECS operating systems are EulerOS, CentOS, Ubuntu, and SUSE. For details, see **Preparing an ECS**.

Step 2 Download the one-click client deployment tool.

Use the SSH login tool (such as PuTTY) to remotely log in to the Linux ECS through the EIP and run the following command to obtain the one-click deployment tool.

```
curl -O -k "https://cloudtable-publish.obs.myhuaweicloud.com/cloudtable-client/quick_start_hbase_shell.sh"
```

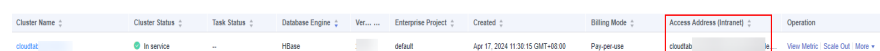
NOTE

- This command applies to HBase 2.x.
- The verification file is contained in the one-click deployment package.

Step 3 Obtain a cluster access address.

Log in to the CloudTable console and choose **Cluster Management**. In the cluster list, locate the required cluster and obtain the address in the **Access Address (Intranet)** column. The parameter value is the cluster access address, as shown in **Figure 1-19**.

Figure 1-19 Access address



Step 4 Use the tool to deploy the client.

Replace **\$zookeeper_address** in the following command with the ZK link you obtained in [Step 3](#). Then, run the command on the CLI of the ECS to deploy the client in one click.

```
source quick_start_hbase_shell.sh $zookeeper_address
```

Step 5 Start the shell to access the cluster.

After you run the source command to automatically deploy the client, the HBase shell is automatically started. You can also run the **bin/hbase shell** command to start the HBase shell to access the cluster.

----End

Manually Installing a Client

Step 1 Prepare a Linux ECS.

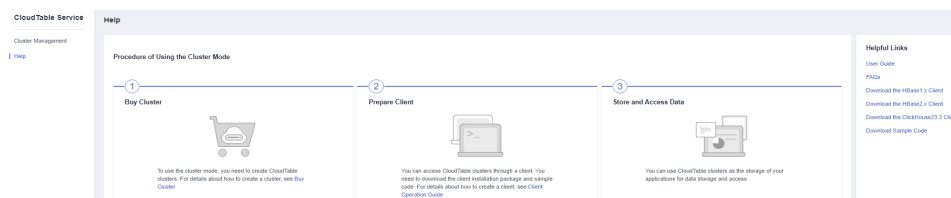
For details, see the following content in [Preparing an ECS](#).

1. [Preparing an ECS](#)
2. [Configuring the DNS Address and hosts File for the Linux ECS](#)

Step 2 Download the client and verification file.

Log in to the CloudTable management console. In the navigation pane on the left, choose **Help**. On the right of the page, click **Download Client** and **Client Verification File** to download the client installation package and client verification file.

Figure 1-20 Downloading the client



Step 3 Install the client and verify the client.

1. Use a file transfer tool (such as WinSCP) to upload the client installation package to the Linux ECS.
2. Use the SSH login tool (such as PuTTY) to log in to the Linux ECS through the EIP.

For details about how to log in to the Linux ECS, see [Logging In to a Linux ECS](#) > Login Using an SSH Password in the *Elastic Cloud Server User Guide*.

Run the following command to decompress the client installation package:

```
cd <Path of the client installation package>  
tar xzvf hbase-1.3.1-bin.tar.gz  
cd <Path of the client installation package>  
tar xzvf hbase-2.4.14-bin.tar.gz
```

<Path of the client installation package>: Replace it with the actual path.

3. Decompress the client verification file to the same directory as the client.
 - a. Decompress the client verification file.

```
cd <Path for storing the client verification file>  
tar xzvf Client_sha256.tar.gz
```


- b. Obtain the client verification code.

```
sha256sum HBase_Client_2.4.14.tar.gz
```
 - c. Check the verification code in the client verification file and compare it with the client verification code. A match indicates no tampering, while a mismatch suggests tampering.

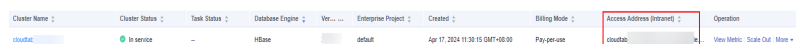
```
less HBase_Client_2.4.14.tar.gz.sha256
```
4. Configure the ZooKeeper address in a configuration file.

In the decompression directory of the client installation package, open the **hbase/conf/hbase-site.xml** file and set the following parameters:

- **hbase.zookeeper.quorum**: The value of this parameter is the access address (Intranet) of the cluster found in the cluster list.

Log in to the CloudTable console and choose **Cluster Management**. In the cluster list, locate the required cluster and obtain the address in the **Access Address (Intranet)** column. See the following figure.

Figure 1-21 ZK link



- **mapreduce.cluster.local.dir**: Check whether the configuration item exists. If the configuration item does not exist, add it.

The configuration example is as follows:

```
<configuration>
<property>
<name>hbase.zookeeper.quorum</name>
<value>xxx-zk1.cloudtable.com,xxx-zk2.cloudtable.com,xxx-zk3.cloudtable.com</value>
</property>

<property>
<name>mapreduce.cluster.local.dir</name>
<value>${hadoop.tmp.dir}/mapred/local</value>
</property>
</configuration>
```

Step 4 Start the shell to access the cluster.

Run the **bin/hbase shell** command to start the shell to access the cluster.

----End

1.1.4.2 Introducing HBase Shell Commands

This section describes common HBase shell commands.

1. Start the HBase shell.

Go to the HBase directory and run the following command to access the HBase shell:

```
./bin/hbase shell
```

2. Get help information.

After you run the **help** command in the HBase shell, all command information as well as common command instructions and use methods will be returned.

```
hbase(main):001:0> help
```

3. Create a table.

Run the **create** command to create a table. When creating a table, you must specify the table name and column family name.

```
hbase(main):007:0> create 'cloudtable','cf'  
0 row(s) in 1.5530 seconds  
=> Hbase::Table - cloudtable
```

4. Query a table:

Run the **list** command to query the specified table. The table name must be specified.

```
hbase(main):009:0> list 'cloudtable'  
TABLE  
cloudtable  
1 row(s) in 0.0060 seconds  
=> ["cloudtable"]
```

5. Insert a piece of record to a table.

Run the **put** command to insert a piece of record to the specified table. You need to specify the table name, primary key, customized column, and inserted value.

```
hbase(main):004:0> put 'cloudtable','row1','cf:a','value1'  
0 row(s) in 0.2720 seconds
```

The following describes parameters in the command:

- **cloudtable**: table name
- **row1**: primary key
- **cf:a**: customized column
- **value1**: inserted value

6. Scan records

Run the **scan** command to scan a table. You need to specify the table name and you can scan a full table or specify a scanning range.

```
hbase(main):001:0> scan 'cloudtable'  
ROW                                COLUMN+CELL  
row1                                column=cf:a, timestamp=1504866237162,  
value=value1  
1 row(s) in 0.2420 seconds
```

NOTE

- 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.

7. Query a single record.

Run the **get** command to query a single record. The name and primary key of the queried table must be specified.

```
hbase(main):001:0> get 'cloudtable','row1'  
COLUMN                                CELL  
cf:a                                timestamp=1504866237162, value=value1  
1 row(s) in 0.2280 seconds
```

8. Disable a table.

Before modifying or deleting a table, you need to disable the table. Run the **disable** command to disable the table. When you perform operations on a disabled table, "ERROR" will be reported, indicating that the table is disabled.

```
hbase(main):002:0> disable 'cloudtable'  
0 row(s) in 2.3550 seconds
```

9. Enable a table.

If you want to use a table that has been disabled, run the **enable** command to enable the table.

```
hbase(main):004:0> enable 'cloudtable'  
0 row(s) in 1.2500 seconds
```

10. Delete a table.

Run the drop command to delete a table that you do not require anymore. Before deleting a table, disable the table. Otherwise, "ERROR" will be displayed, indicating that the table is enabled. Deleting a table will cause data loss. Therefore, exercise caution when deleting a table.

```
hbase(main):007:0> disable 'cloudtable'  
0 row(s) in 2.2380 seconds
```

```
hbase(main):008:0> drop 'cloudtable'  
0 row(s) in 1.2600 seconds
```

11. Exit the HBase shell.

Run the **quit** command to exit the HBase shell.

```
hbase(main):009:0> quit
```

1.1.5 HBase Thrift Server

1.1.5.1 Thrift Server Overview

The HBase Thrift Server operates as a service within the HBase cluster, offering a gateway to the HBase database via the Thrift protocol. This server is adept at producing both client and server code, accommodating a multitude of programming languages for diverse application needs.

NOTE

The Thrift Serve feature is supported in only HBase 2.4.14 or later.

Background

- HBase needs to support multiple languages.

Use Restrictions

- The Thrift Server supports the following programming languages: C++, and Java.
- The Thrift Server does not support features such as cold and hot data separation, automatic hotspot self-healing, or secondary indexing.
- Only non-security clusters support Thrift Server.
- Thrift Server is deployed only on Core nodes.
- Scaling out the Thrift Server is directly dependent on the scaling capabilities of the Core nodes.
- Once user authentication is activated within the cluster, the Thrift Server feature becomes unavailable.

- If the Thrift Server encounters issues, services may be disrupted, especially when cluster nodes are accessed via IP addresses.
- Thrift Server cannot interconnect with ELB.

1.1.5.2 Using Thrift Server

The Thrift Server service is enabled or disabled on the CloudTable console.

Procedure for Enabling Thrift Server

- Step 1** Log in to the CloudTable console.
- Step 2** Select a region in the upper left corner.
- Step 3** In the navigation pane on the left, choose **Cluster Management**.
- Step 4** Click **Buy Cluster** in the upper right corner.
- Step 5** Go to the **Create Cluster** page. On the cluster purchase page, select the HBase database engine and enable Thrift Server.

NOTE

- For details about how to create a cluster, see [Creating an HBase Cluster](#).
- Thrift Server is disabled by default.
- You can also choose **More > Enable Thrift Server** on the cluster management page to enable Thrift Server.

Figure 1-22 Enabling Thrift Server



----End

Procedure for Disabling Thrift Server

- Step 1** Log in to the CloudTable console.
- Step 2** Select a region in the upper left corner.
- Step 3** In the navigation pane on the left, choose **Cluster Management**.
- Step 4** Select the cluster for which you want to disable Thrift Server and choose **More > Disable Thrift Server**.
- Step 5** In the displayed window for disabling the Thrift Server service, click **OK**.

----End

1.1.6 HBase Security Channel Encryption

You can enable security channel encryption to encrypt data transmission. This section describes how to enable the security channel for a HBase cluster.

Enabling Channel Encryption


- Step 1** Log in to the CloudTable console.
- Step 2** Click  in the upper left corner to select a region.
- Step 3** Click **Buy Cluster** in the upper right corner.
- Step 4** Check whether **Enable Channel Encryption** (which is toggled on by default) is toggled on after completing other configurations.


Figure 1-23 Enabling channel encryption



- Step 5** Completing the parameter setting and click **Next**.
- Step 6** Confirm the cluster specification order information on the displayed page and submit the order. After the cluster is created, go to its details page to view the channel status.

----End

Procedure for Connecting to an HBase Security Cluster

- Step 1** Log in to the CloudTable console.
- Step 2** Click  in the upper left corner to select a region.
- Step 3** Choose **Help > Download hbase2.x Client** to download the client.
- Step 4** Install the client and [prepare an ECS](#).
- Step 5** Use the SSH login tool (such as PuTTY) to log in to the Linux ECS through the EIP.

For details about how to log in to the ECS, see "Remotely Logging In to a Linux ECS (Using an SSH Password)" in [Logging In to a Linux ECS](#) of the *Elastic Cloud Server User Guide*.

- Step 6** Install the client downloaded in [Step 3](#). For details, see [Manually Installing a Client](#). (You can also use [One-Click Deployment](#) to install the client.)
- Step 7** Modify the **hbase-env.sh** file. Go to the directory where the client is installed and run the **ll** command to check whether the HBase client exists.
 - Go to the hbase-client folder and view the **conf** folder.


```
cd hbase-client
```
 - Go to the **conf** folder and edit the **hbase-env.sh** file.
 - Go to the **conf** folder.


```
cd conf/
```
 - Edit the **hbase-env.sh** file.


```
vi hbase-env.sh
```
 - Add the following content to the end of the **hbase-env.sh** file:


```
CLIENT_JVMFLAGS="-
Dzookeeper.clientCnxnSocket=org.apache.zookeeper.ClientCnxnSocketNetty -
Dzookeeper.client.secure=true"
export HBASE_OPTS="${CLIENT_JVMFLAGS} ${HBASE_OPTS}"
```

- Click Ecs to exit the editing page.
- Save the added content.
:wq

Step 8 Go to the client installation directory to configure the client. Add the following configurations to the **hbase-site.xml** file in the **conf** directory on the client:

- Edit the **hbase-site.xml** file.
vi hbase-site.xml
- Add the following content to the end of the **hbase-site.xml** file:

```
<property>  
  <name>hbase.rpc.protection</name>  
  <value>privacy</value>  
</property>
```
- Click Ecs to exit the editing page.
- Save the added content.
:wq

Step 9 Connect to the HBase cluster.

- Configure the ZooKeeper address in the configuration file. For details, see [3.2.5.1-c. Configuring the ZooKeeper Address](#).
- Start the shell to access the cluster.
Run the **bin/hbase shell** command to start the shell to access the cluster.

----End

1.2 HBase Batch Data Import

1.2.1 Using CDM to Migrate Data to CloudTable

CloudTable uses 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 details about the supported data sources, see [Data Sources Supported by CDM](#) in the *Cloud Data Migration User Guide*.

The procedure of using CDM to migrate data to CloudTable is as follows:

1. [Creating a CloudTable Cluster](#)
2. [Using CDM to Migrate Data to CloudTable](#)
3. [Viewing the Imported Data on CloudTable](#)

Creating a CloudTable Cluster

Step 1 Log in to the CloudTable console.

Step 2 Select a region in the upper left corner.

Step 3 Create a [CloudTable HBase](#) cluster. If there is one available, you can skip this step.

For example, create a CloudTable cluster named **CloudTable-demo**.

----End

Using CDM to Migrate Data to CloudTable

Step 1 Log in to the Cloud Data Migration management console to create a CDM cluster.

The CDM cluster you create must have the same region, AZ, VPC, subnet, and security group as the CloudTable cluster.

For details about how to create a CDM cluster, see [Creating a Cluster](#) in the *Cloud Data Migration User Guide*.

For example, create a CDM cluster named **CDM-demo**.

Step 2 Create a source link in the new CDM cluster.

CDM enables you to migrate data from multiple data sources to CloudTable. Create a link as the source link based on the actual data source. The link is used for CDM to connect to the data source and read data from the data source.

For details about how to create a link, see [Creating a Link](#) in the *Cloud Data Migration User Guide*.

Step 3 Create a destination link in the new CDM cluster.

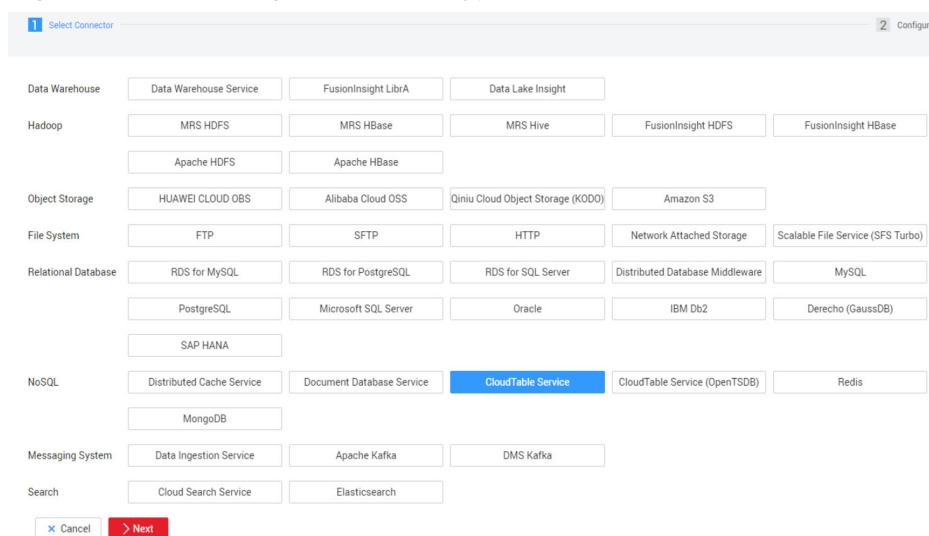
In the CDM cluster, create a link for which **Data Source Type** is set to **CloudTable** as a destination link. The link is used for CDM to connect to the CloudTable cluster, read data from the data source, and import data to the HBase table of the CloudTable cluster.

For details about how to create a link, see [Creating a Link](#) in the *Cloud Data Migration User Guide*.

The following is an example of destination link settings.

1. On the **Cluster Management** page, locate the newly created CDM cluster and click **Job Management** in the row of the CDM cluster. The job management page is displayed.
2. Click the **Link Management** tab and then click **Create Link**. On the **Select connector type** page that is displayed, select **CloudTable Service** and click **Next**.

Figure 1-24 Selecting a connector type



3. Configure link parameters.
 - **Name:** Enter a link name. You can create a link name based on the data source type for easy memorization. For example, enter **cloudtable_connect**.
 - **ZK Link:** Set this parameter to the access address (Intranet) of the CloudTable cluster. Log in to the CloudTable console and choose **Cluster Management**. In the cluster list, locate the required cluster and obtain the address in the **Access Address (Intranet)** column.
 - : Set this parameter to **No**.
 - **IAM Authentication:** CloudTable does not use IAM authentication. Therefore, set this parameter to **No**.
 - **Run Mode:** Retain the default value.

Figure 1-25 Link parameter configuration

The screenshot shows a configuration form with the following fields and values:

- Name:** cloudtable_connector
- Connector:** HBase
- HBase Type:** CloudTable
- ZK Link:** cloudtable-demo-zk1.cloudt (with a "Select" button to the right)
- IAM Authentication:** No (selected)
- Run Mode:** STANDALONE

At the bottom of the form, there are four buttons: "Cancel", "Previous", "Test", and "Save".

4. Click **Save**.

Step 4 In the CDM cluster, create a job on the **Table/File Migration** page.

For details about how to create a table/file migration job, see [Table/File Migration](#) in the *Cloud Data Migration User Guide*.

NOTE

When you create a table/file migration job, if the table or file of the source end does not have the corresponding HBase table in the destination CloudTable cluster, set a table name and field mapping in the job for the destination end (the field name can be copied from the source end). Therefore, when a job is running, the corresponding HBase table is automatically created at the destination end.

The following uses the SFTP data source on an ECS as an example to describe how to create a job.

1. On the **Cluster Management** page, locate the newly created CDM cluster and click **Job Management** in the row of the CDM cluster. The job management page is displayed.
2. Choose **Table/File Migration > Create Job**.
3. Set parameters as follows:

Figure 1-26 Configuring basic information

The screenshot shows the 'Job Configuration' page. At the top, there is a 'Job Name' field with the value 'cloudtable'. Below this, the page is divided into two main sections: 'Source Link Configuration' and 'Destination Link Configuration'.
Source Link Configuration:
 - Source Link Name: 'source' (dropdown menu) with a 'Create Link' button.
 - Source Directory/File: '/opt/abc.txt' (text input) with a file selection icon.
 - File Format: 'CSV_FILE' (dropdown menu).
 - A 'Show Advanced Attributes' link is present below.
Destination Link Configuration:
 - Destination Link Name: 'destination' (dropdown menu) with a 'Create Link' button.
 - Table Name: 'tableDemo' (text input) with a file selection icon.
 - Clear data before import: Radio buttons for 'Yes' and 'No', with 'No' selected.
 - A 'Show Advanced Attributes' link is present below.
 At the bottom of the configuration area, there are 'Cancel' and 'Next' buttons.

4. Set the field mapping as follows:

Figure 1-27 Field mapping

The screenshot shows the 'Field mapping' interface. It consists of two tables: 'Source Field' and 'Destination Field'.
Source Field Table:

Column ID	Example Value	Operation
1	i	[Refresh] [Delete]
2	b	[Refresh] [Delete]
3	c	[Refresh] [Delete]

Destination Field Table:


Column Family	Column Name	Row Key	Operation
info	c1	<input type="checkbox"/>	[Delete]
info	c2	<input type="checkbox"/>	[Delete]
info	c3	<input type="checkbox"/>	[Delete]

 Arrows indicate the mapping from source fields 1, 2, and 3 to destination columns c1, c2, and c3 respectively. Below the tables, there is an 'Add Field+' link and 'Cancel', 'Previous', and 'Next' buttons.

5. Configure the task as follows:

Figure 1-28 Configuring a task

Configure Task

Concurrent Extractors 

1

Schedule Execution [Show Advanced Attributes](#)

Cancel

Previous

Save

Save and Run

6. Click **Save and Run**.

----End

Viewing the Imported Data on CloudTable

Step 1 Prepare a Linux ECS.

Assume that the ECS name is `ecs_20170916`. For details, see [Preparing an ECS](#).

Step 2 Install the client and start the shell to access the CloudTable cluster.

For details about how to use the HBase shell to access the cluster, see [Using HBase Shell to Access a Cluster](#).

Step 3 On the CloudTable client, run a command to query the data migrated to CloudTable using CDM.

The following is an example of the command. Replace `table_name` with the name of the table specified when a job is created in the CDM cluster.

```
scan 'table_name'
```

----End

1.2.2 Using the Import Tool to Importing Data

Use the Import tool to import the **SequenceFile** data file of HBase to HBase of CloudTable. The Import tool is a tool contained in the installation package of the client tool.

NOTE

The **SequenceFile** file is a data file exported from HBase by the Export tool.

Step 1 Prepare a server where a CloudTable client is installed. For details about how to install a client, see [Using HBase Shell to Access a Cluster](#).

Step 2 Upload the **SequenceFile** file directory to the server where the CloudTable client is located. Assume that the directory is `/tmp/sequencefile`.

- Step 3** The Import tool must be executed on the shell interface of the operating system of the client host. Go to the HBase directory on the client host and run the Import tool. The command format of the Import tool is as follows:

```
./bin/hbase org.apache.hadoop.hbase.mapreduce.Import <TableName>  
<InputPath>
```

TableName: Name of the table to which data is to be imported

InputPath: Directory of the **SequenceFile** file In the local file system, the directory must start with file://.

Example:

```
./bin/hbase org.apache.hadoop.hbase.mapreduce.Import t1 file:///tmp/sequencefile
```

----End

1.2.3 Using CopyTable to Import Data

CopyTable is a utility provided by HBase. It can copy part or of all of a table, either to the same cluster or another cluster. The target table must exist first. The CloudTable client tool includes CopyTable. After deploying the client tool, you can use CopyTable to import data to a CloudTable cluster.

Using CopyTable to Import Data

- Step 1** Prepare a Linux ECS as the client host and deploy the CloudTable client tool on it.

For details, see [Using HBase Shell to Access a Cluster](#).

When deploying the client tool, set the ZK link to the access address (Intranet) of the CloudTable cluster where the source table resides.

- Step 2** (Optional) If you want to copy a table to another cluster, obtain the access address (Intranet) of the target CloudTable cluster.

Log in to the CloudTable console and choose **Cluster Management**. In the cluster list, locate the required cluster and obtain the address in the **Access Address (Intranet)** column.

- Step 3** Before using CopyTable to copy table data, ensure that the target table exists in the target CloudTable cluster. If the target table does not exist, create it first.

For details about how to create a table, see [Creating an HBase Cluster](#).

- Step 4** On the client host, open the CLI, access the **hbase** directory in the installation directory of the client tool, and run the CopyTable command to import data to the CloudTable cluster.

The following is an example of the command. In this example, the data in the specified 1 hour in **TestTable** is copied to the target cluster.

```
cd ${Installation directory of the client tool}/hbase  
./bin/hbase org.apache.hadoop.hbase.mapreduce.CopyTable --starttime=1265875194289 --  
endtime=1265878794289 --peer.adr=${ZK link of the target CloudTable cluster}/hbase --  
families=myOldCf:myNewCf,cf2,cf3 TestTable
```

----End

Overview of the CopyTable Command

The CopyTable command format is as follows:

```
CopyTable [general options] [--starttime=X] [--endtime=Y] [--new.name=NEW] [--peer.adr=ADR]
<tablename>
```

For details about the CopyTable command, see [CopyTable](#).

The following provides description about common options:

- **startrow**: the start row
- **stoprow**: the stop row
- **starttime**: beginning of the time range (unixtime in milliseconds). If **endtime** is not specified, it implies that the duration extends from the start time indefinitely.
- **endtime**: end of the time range. If no starttime is specified, ignore it.
- **versions**: number of cell versions to be copied
- **new.name**: name of a new table
- **peer.adr**: Address of the target cluster. The format is **hbase.zookeeper.quorum:hbase.zookeeper.client.port:zookeeper.znode.paren**. For the HBase clusters, the parameter value is **#{ZK link of the target CloudTable cluster}/hbase**.
- **families**: List of column families to be copied. Multiple column families are separated by commas (,).
If you want to copy from **sourceCfName** to **destCfName**, specify **sourceCfName:destCfName**.
If the column family name needs to remain unchanged after copying, you only need to specify **cfName**.
- **all.cells**: Deletion markers and the deleted cells are also copied.

The parameter description is as follows:

tablename: name of the table to be copied

1.2.4 Sample Code for Copying a Table

You can develop an application and call HBase APIs to copy some or all data in a table either to the same cluster or another cluster. The target table must exist first.

Environment Preparation

Step 1 If you want to copy a table to another cluster, obtain the access address (Intranet) of the target CloudTable cluster first.

Log in to the CloudTable console and choose **Cluster Management**. In the cluster list, locate the required cluster and obtain the address in the **Access Address (Intranet)** column.

Step 2 Before copying table data, ensure that the target table exists in the target CloudTable cluster. If the target table does not exist, create it first.

For details about how to create a table, see [Creating an HBase Cluster](#).

Step 3 Download the [sample code](#) and prepare the development environment by referring to the corresponding section.

If you do not use the sample code project, copy the JAR file in the **cloudtable-example\lib** directory in the downloaded sample code to your project and add the JAR file to the dependency path in the project.

----End

Sample Code for Copying a Table

You can develop applications based on HBase APIs to copy table data to clusters based on service requirements. The following sample code is for reference only.

The following sample code is used to copy table data from one cluster to another cluster.

```
package com.huawei.cloudtable.hbase.tool.client;

import java.io.IOException;
import java.util.ArrayList;
import java.util.Collections;
import java.util.List;
import java.util.concurrent.BlockingQueue;
import java.util.concurrent.LinkedBlockingQueue;
import java.util.concurrent.ThreadPoolExecutor;
import java.util.concurrent.TimeUnit;

import org.apache.commons.logging.Log;
import org.apache.commons.logging.LogFactory;
import org.apache.hadoop.conf.Configuration;
import org.apache.hadoop.hbase.Cell;
import org.apache.hadoop.hbase.HBaseConfiguration;
import org.apache.hadoop.hbase.HRegionInfo;
import org.apache.hadoop.hbase.TableName;
import org.apache.hadoop.hbase.client.Admin;
import org.apache.hadoop.hbase.client.Connection;
import org.apache.hadoop.hbase.client.ConnectionFactory;
import org.apache.hadoop.hbase.client.Put;
import org.apache.hadoop.hbase.client.Result;
import org.apache.hadoop.hbase.client.ResultScanner;
import org.apache.hadoop.hbase.client.Scan;
import org.apache.hadoop.hbase.client.Table;

public class DataCoper {
    public static final Log LOG = LogFactory.getLog(DataCoper.class);
    private static int batchSize;
    private static int threads;

    public static void main(String[] args) throws IOException {
        if (args.length != 5) {
            System.err.println(
                "Command : ./hbase com.huawei.cloudtable.hbase.tool.client.DataCoper [srcZK] [dstZK] [tableName] [batchSize] [threads]");
            System.exit(-1);
        }
        batchSize = Integer.valueOf(args[3]);
        threads = Integer.valueOf(args[4]);
        copyData(args[0], args[1], args[2]);
    }

    public static void copyData(String srcZk, String dstZk, String tableName) throws IOException {
        // Create a source HBase link.
        Configuration srcConf = HBaseConfiguration.create();
        srcConf.set("hbase.zookeeper.quorum", srcZk);
        //srcConf.set("zookeeper.znode.parent", "/hbase");
    }
}
```

```
Connection srcConnection = ConnectionFactory.createConnection(srcConf);

// Create a target HBase link.
Configuration dstConf = HBaseConfiguration.create();
dstConf.set("hbase.zookeeper.quorum", dstZk);
//dstConf.set("zookeeper.znode.parent", "/hbase_1.3.1");
Connection dstConnection = ConnectionFactory.createConnection(dstConf);

// Copy a table.
copyDataTable(srcConnection, dstConnection, TableName.valueOf(tableName));
}

private static void copyDataTable(final Connection srcConnection, final Connection dstConnection,
    final TableName tableName) throws IOException {
    try (Admin admin = srcConnection.getAdmin()) {
        // Obtain the region information of the data table.
        List<HRegionInfo> tableRegions = admin.getTableRegions(tableName);
        Collections.shuffle(tableRegions);

        // Create a thread pool and concurrently copy data.
        BlockingQueue<Runnable> workQueue = new LinkedBlockingQueue<Runnable>(10000);
        ThreadPoolExecutor executor = new ThreadPoolExecutor(threads, threads, 60, TimeUnit.SECONDS,
            workQueue);

        for (final HRegionInfo regionInfo : tableRegions) {
            executor.submit(new Runnable() {
                @Override
                public void run() {
                    try (final Table srcTable = srcConnection.getTable(tableName);
                        final Table dstTable = dstConnection.getTable(tableName)) {
                        LOG.info("Start to copy region " + regionInfo.toString());
                        Scan scan = new Scan();
                        scan.setStartRow(regionInfo.getStartKey());
                        scan.setStopRow(regionInfo.getEndKey());
                        scan.setCaching(batchSize);
                        copyOneRegion(srcTable, dstTable, scan, batchSize);
                    } catch (IOException e) {
                        LOG.error("CopyData failed .....", e);
                        System.exit(-1);
                    }
                    LOG.info("Stop to copy region " + regionInfo.toString());
                }
            });
        }
    }
}

private static void copyOneRegion(Table srcTable, Table dstTable, Scan scan, int batchSize)
    throws IOException {
    ResultScanner scanner = srcTable.getScanner(scan);
    Result result = null;
    List<Put> puts = new ArrayList<>();
    long counter = 0;
    while ((result = scanner.next()) != null) {
        Put put = new Put(result.getRow());
        for (Cell cell : result.rawCells()) {
            put.add(cell);
        }
        puts.add(put);

        counter += 1;
        if (puts.size() >= batchSize) {
            dstTable.put(puts);
            puts.clear();
        }
        if (counter % 10000 == 0) {
            LOG.info(srcTable + " has send: " + counter);
        }
    }
}
```

```
}  
if (puts.size() != 0) {  
    dstTable.put(puts);  
}  
}  
}
```

1.3 HBase Monitoring Clusters

1.3.1 HBase Cluster Supported Metrics

Description

Monitoring is critical to ensure CloudTable reliability, availability, and performance. You can monitor the running status of CloudTable servers.

This section describes the metrics that can be monitored by CES as well as their namespaces and dimensions. You can use the management console or [APIs](#) provided by Cloud Eye to query the metrics of the monitored objects and alarms generated for CloudTable.

Namespace

SYS.CloudTable

CloudTable HBase HMaster Instance Monitoring Metrics

Table 1-14 CloudTable HBase HMaster instance monitoring metrics

Metric ID	Name	Meaning	Value Range	Monitoring Interval (Raw Data)
cmdForIORead	Disks Read Rate	Volume of data read from the monitored object per second	≥ 0 bytes/s	1 min
cmdForIOWrite	Disks Write Rate	Volume of data written to the monitored object per second	≥ 0 bytes/s	1 min
cmdForNumberOfCpus	CPU Count	Number of CPUs of the monitored object	> 0	1 min
cmdForTotalMemory	Total Memory	Total memory size of the monitored object	> 0 Byte	1 min

Metric ID	Name	Meaning	Value Range	Monitoring Interval (Raw Data)
cmdProcessCPU	CPU Usage	CPU usage of the monitored object	0%–100%	1 min
cmdProcessMem	Memory Usage	Memory usage of the monitored object	0%–100%	1 min
hm_deadregionservernum	Faulty RegionServers	Number of faulty RegionServers in the cluster	≥ 0	1 min
hm_regionservernum	Normal RegionServers	Number of normal RegionServers in the cluster	≥ 0	1 min
hm_ritCount	RIT Count	Number of regions in the Region In Transaction (RIT) state in the cluster where the monitored object is located	≥ 0	1 min
hm_ritCountOverThreshold	RIT Count Over Threshold	Number of regions in the RIT state and reach the threshold in the cluster where the monitored object is running	≥ 0	1 min
rs_queuecalltime_max	RPC Queue Call Time (Max)	Maximum RPC queue call time	≥ 0 ms	1 min
rs_queuecalltime_mean	RPC Queue Call Time (Mean)	Mean RPC queue call time	≥ 0 ms	1 min
nn_percentallused	Disk Utilization Rate	Disk space usage of the cluster	0%–100%	1 min

Metric ID	Name	Meaning	Value Range	Monitoring Interval (Raw Data)
nn_capacityremaining	Disk capacity remaining of cluster	Remaining disk space of the cluster	Depends on the cluster disk capacity.	1 min
nn_capacityused	Disk capacity used of cluster	Disk space used in the cluster	Depends on the cluster disk capacity.	1 min

 NOTE

hmaster instances include **hmaster-standby** (standby) and **hmaster-active** (active). When **hmaster-active** becomes faulty, **hmaster-standby** becomes active to provide services.

CloudTable HBase RegionServer Instance Monitoring Metrics

Table 1-15 lists the monitoring metrics supported by CloudTable HBase RegionServer instances.

Table 1-15 Monitored CloudTable metrics

Metric ID	Metric	Meaning	Value Range	Monitoring Period (Raw Data)
cmdForNumberOfCpus	CPU Number	Number of CPUs of the instance	> 0	1 minute
cmdProcessCPU	CPU Usage	CPU usage of the monitored object Unit: %	0%–100%	1 minute
cmdForTotalMemory	Total Memory	Total memory size of the monitored object Unit: byte	> 0 byte	1 minute

Metric ID	Metric	Meaning	Value Range	Monitoring Period (Raw Data)
cmdProcessMem	Memory Usage	Memory usage of the monitored object Unit: %	0%–100%	1 minute
cmdForIOWrite	Disks Write Rate	Volume of data written to the monitored object per second Unit: byte/s	≥ 0 bytes/s	1 minute
cmdForIORead	Disks Read Rate	Volume of data read from the monitored object per second Unit: byte/s	≥ 0 bytes/s	1 minute
hm_regionservernum	Normal RegionServers	Number of normal RegionServers	≥ 0	1 minute
hm_deadregionservernum	Faulty RegionServers	Number of faulty RegionServers	≥ 0	1 minute
hm_ritCountOverThreshold	RIT Count Over Threshold	Region in transaction count over threshold	≥ 0	1 minute
hm_ritCount	RIT Count	Region in transaction count	≥ 0	1 minute
rs_requests	Requests Per Second	Number of requests of a RegionServer per second Unit: Request/s	≥ 0 requests/s	1 minute
rs_regions	Regions	Number of regions of a RegionServer	≥ 0	1 minute
rs_writerequestscout	Write Requests	Number of write requests of a RegionServer	≥ 0	1 minute

Metric ID	Metric	Meaning	Value Range	Monitoring Period (Raw Data)
rs_readrequestcount	Read Requests	Number of read requests of a RegionServer	≥ 0	1 minute
rs_blockcachehitcachingratio	Hit Cache Block Caching Ratio	Block cache hit caching ratio Unit: %	0%–100%	1 minute
rs_blockCacheCountHitPercent	Hit Cache Block Ratio	Block cache hit ratio Unit: %	0%–100%	1 minute
rs_getavgtime	Get Delay (Avg)	Average Get operation delay of the RegionServer per unit time Unit: millisecond	≥ 0 ms	1 minute
rs_putavgtime	Put Delay (Avg)	Average Put operation delay of the RegionServer per unit time Unit: millisecond	≥ 0 ms	1 minute
rs_deleteavgtime	Delete Delay (Avg)	Average Delete operation delay of the RegionServer per unit time Unit: millisecond	≥ 0 ms	1 minute
rs_getnumops	Get Operations	Number of Get operations of the RegionServer per unit time	≥ 0	1 minute
rs_putnumops	Put Operations	Number of Put operations of the RegionServer per unit time	≥ 0	1 minute
rs_deletenumops	Delete Operations	Number of Delete operations of the RegionServer per unit time	≥ 0	1 minute

Metric ID	Metric	Meaning	Value Range	Monitoring Period (Raw Data)
rs_queuecalltime_max	RPC Queue Call Time (Max)	Maximum RPC queue call time Unit: millisecond	≥ 0 ms	1 minute
rs_queuecalltime_mean	RPC Queue Call Time (Mean)	Mean RPC queue call time Unit: millisecond	≥ 0 ms	1 minute
rs_flush_time_mean	Flush Time (Mean)	Mean time of flush Unit: millisecond	≥ 0 ms	1 minute
rs_compaction_queue_size	Compaction Queue Size	Point in time length of the compaction queue. The number of Stores for compaction in the RegionServer.	≥ 0	1 minute
rs_flush_queue_size	Flush Queue Size	Flush queue size	≥ 0	1 minute
rs_compactions_completed_count	Compaction Count	Count of compaction	≥ 0	1 minute
rs_flush_operations_num	Flush Operation Count	Count of flush operation	≥ 0	1 minute
rs_blockcache_evicted_count	Discarded Cache Blocks	Block cache evict count	≥ 0	1 minute
rs_sync_time_max	Sync WAL Time (Max)	Maximum time it took to sync the WAL to HDFS Unit: millisecond	≥ 0 ms	1 minute
rs_sync_time_mean	Sync WAL Time (Mean)	Mean time it took to sync the WAL to HDFS Unit: millisecond	≥ 0 ms	1 minute
dn_bytes_written_speed	Bytes written per second	Bytes written per second of the node	≥ 0 byte	1 min

Metric ID	Metric	Meaning	Value Range	Monitoring Period (Raw Data)
dn_bytesread_speed	Bytes read per second	Bytes read per second of the node	≥ 0 byte	1 min
rs_numActiveHandler	Number of RegionServer Active Handlers	Number of active RegionServer handlers (total number of handlers for processing user table requests, meta table requests, and replication requests)	≥ 0	1 min
rs_numActiveGeneralHandler	Number of RegionServer Active Handlers for Processing User Table Requests	Number of active RegionServer handlers for processing user table requests	≥ 0	1 min
rs_scanTime_p999	99.9th Percentile of the Scan Operation Delay	99.9th percentile of the RegionServer Scan operation delay	≥ 0 ms	1 min
rs_syncTime_p999	99.9th Percentile of the WAL Sync Operation Delay	99.9th percentile of the RegionServer WAL Sync operation delay	≥ 0 ms	1 min
rs_Get_99th_percentile	99th Percentile of the Get Operation Delay	99th percentile of the RegionServer Get operation delay	≥ 0 ms	1 min
rs_Put_99th_percentile	99th Percentile of the Put Operation Delay	99th percentile of the RegionServer Put operation delay	≥ 0 ms	1 min
rs_Delete_99th_percentile	99th Percentile of the Delete Operation Delay	99th percentile of the RegionServer Delete operation delay	≥ 0 ms	1 min

Metric ID	Metric	Meaning	Value Range	Monitoring Period (Raw Data)
rs_Get_999th_percentile	99.9th Percentile of the Get Operation Delay	99.9th percentile of the RegionServer Get operation delay	≥ 0 ms	1 min
rs_Put_999th_percentile	99.9th Percentile of the Put Operation Delay	99.9th percentile of the RegionServer Put operation delay	≥ 0 ms	1 min
rs_Delete_999th_percentile	99.9th Percentile of the Delete Operation Delay	99.9th percentile of the RegionServer Delete operation delay	≥ 0 ms	1 min

Dimension

Key	Value
cluster_id	CloudTable cluster ID.
instance_name	Name of a CloudTable cluster node.

1.3.2 Setting Alarm Rules


You can set CloudTable alarm rules to customize the monitored objects and notification policies. Then, you can learn about CloudTable running status in a timely manner.

The CloudTable alarm rules include alarm rule name, instance, metric, threshold, monitoring interval and whether to send notification. This section describes how to set CloudTable alarm rules.

Procedure

Step 1 Log in to the CloudTable console.

Step 2 In the navigation tree, click **Cluster Management**.

Step 3 In the upper right corner of the cluster list, enter the name of a cluster in the search box and click .

Step 4 In the cluster list, locate the row where the target cluster resides and click **View Metric** in the **Operation** column. The CES console is displayed.

The status of the cluster whose monitoring information you want to view must be **In service**.

Figure 1-29 Viewing the monitoring metrics

ID	Permanent Data Storage	Operation
<input type="checkbox"/> zookeeper-3-3	--	View Metric Create Alarm Rule Configure Storage View Alarm Rule
<input type="checkbox"/> zookeeper-2-2	--	View Metric Create Alarm Rule Configure Storage View Alarm Rule
<input type="checkbox"/> zookeeper-1-1	--	View Metric Create Alarm Rule Configure Storage View Alarm Rule
<input type="checkbox"/> server-1-1	--	View Metric Create Alarm Rule Configure Storage View Alarm Rule

Step 5 Click **Create Alarm Rule** to add an alarm rule.

1. Enter an alarm rule name and description.

Figure 1-30 Configuring an alarm rule name

★ Name

Description

0/256

2. Configuring the parameters related to the monitoring function.

Figure 1-31 Parameters related to the monitoring function

★ Alarm Type	Metric
★ Resource Type	CloudTable Service
★ Dimension	Compute Units
★ Monitoring Scope	Specific resources
★ Monitored Objects	hmaster-active

Table 1-16 Parameter description

Parameter	Description	Example Value
Alarm Type	Default instance or the selected cluster	Metric
Resource Type	Name of the cloud service resource for which the alarm rule is configured	CloudTable Service
Dimension	Metric dimension of the alarm rule. Only the Compute Units option is supported.	Computing Units
Monitoring Scope	Clusters to be monitored	Specific resources
Monitored Object	Nodes to be monitored	-

3. Configure alarm policies as prompted.

Figure 1-32 Configuring an alarm policy

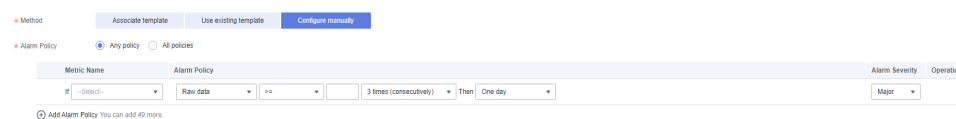


Table 1-17 Parameters for configuring an alarm policy

Parameter	Description	Example Value
Method	<p>Select Use existing template or Configure manually as required.</p> <ul style="list-style-type: none"> - If no alarm template is available, set Method to Configure manually and configure related parameters to create an alarm rule. - If you have available alarm rule templates, set Method to Use existing template or Associate template, so that you can use a template to quickly create alarm rules. 	Create manually
Template	<p>This parameter is valid only when Use existing template is selected.</p> <p>Select the template to be imported. If no alarm template is available, click Create Custom Template to create one that meets your requirements.</p>	Use Existing Template

Parameter	Description	Example Value
Alarm Policy	This parameter is valid only when Configure manually is selected. For details about the monitoring metrics supported by CloudTable HBase, see HBase Cluster Supported Metrics .	-
Alarm Severity	Severity of an alarm. Valid values are Critical, Major, Minor, and Informational .	Major

4. Configure the alarm notification parameters as prompted.

Figure 1-33 Alarm notification

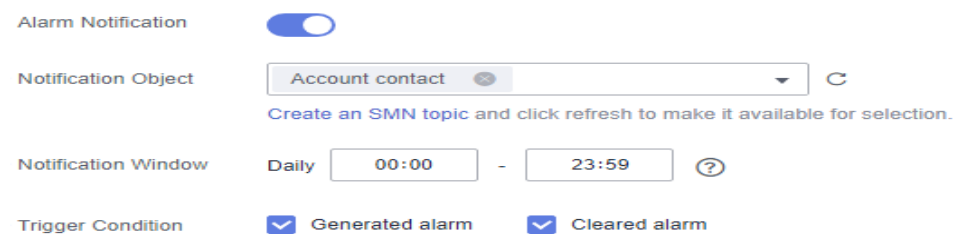


Table 1-18 Alarm notification parameters

Parameter	Description	Configuration
Alarm Notification	Whether to notify users when alarms are triggered. Notifications can be sent by emails, text messages, or HTTP/HTTPS requests. If this function is enabled (recommended), notifications will be sent. Otherwise, no notification is sent.	Toggle it on.
Notification Recipient	Notification group or topic subscription.	-

Parameter	Description	Configuration
Notification Group/Object	<ul style="list-style-type: none"> Notification group that needs to send alarm notifications. If you enable Send Notification, you need to select an existing notification group. If you have not created a notification group, click Create Notification Group to access the CES page to create one. Topic to which the alarm notification is to be sent. If you enable Notification, you need to select an existing topic. If you have not created a topic, click Create an SMN Topic to go to the SMN management console to create a topic. 	-
Validity Period	<p>CES sends notifications only within the validity period specified in the alarm rule. If Validity Period is set to 00:00-23:59, CES send notifications only within 00:00-23:59.</p>	The value ranges from 00:00 to 23:59.
Trigger Condition	Condition for triggering an alarm notification.	Select all options.

5. Configuring advanced settings.

Figure 1-34 Configuring advanced settings

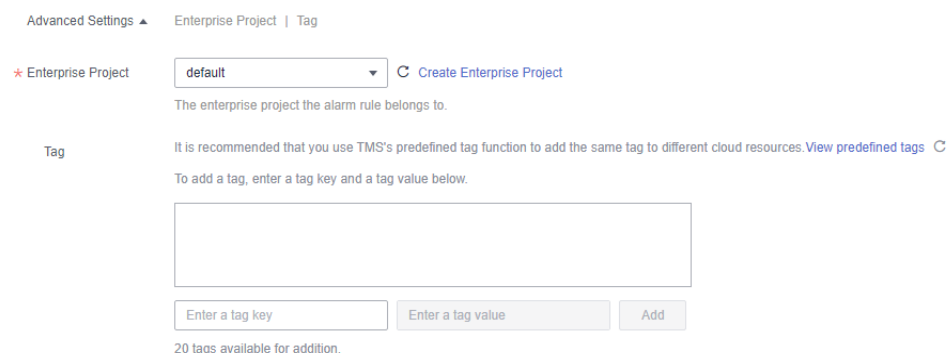


Table 1-19 Advanced configuration parameters

Parameter	Description
Enterprise Project	You can group related cloud resources (for example, resources used for the same purpose) and manage them by enterprise project.

Parameter	Description
Tag	It is recommended that you use TMS's predefined tag function to add the same tag to different cloud resources.

Step 6 After the configuration is complete, click **Next**. After the alarm rule is created, if the metric data reaches the specified threshold, CES will immediately inform you that an exception has occurred.

 **NOTE**

If you want to set an alarm rule for a single metric, go to the page for viewing monitoring information.

- View the monitoring metrics and click the metric graph.
- Click + in the upper right corner.
- On the page for setting alarm rules, set the rule.

----End

1.3.3 Viewing Cluster Monitoring Information

Scenario


CES on the public cloud monitors CloudTable cluster running status. You can view CloudTable monitoring metrics on the management console. According to the monitoring information, you can quickly learn about cluster health status and key system information.

Background Information

- Monitoring metrics of an unavailable CloudTable cluster node cannot be displayed on the CES page. You can view the monitoring information only after the CloudTable cluster node is restarted or recovered.
- CES will delete a CloudTable cluster node that becomes faulty for 1 hour from the monitoring list and will not monitor it anymore. However, you need to manually clear its alarm rules.

Viewing Monitoring Information

Step 1 Log in to the CloudTable management console.

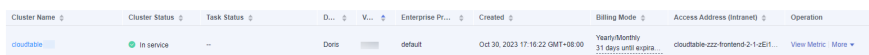
Step 2 Click  in the upper left corner to select a region.

Step 3 In the navigation tree on the left, click **Cluster Management**.

Step 4 In the cluster list, locate the row where the target cluster resides, click **View Metric** in the **Operation** column. The CES console is displayed.

The status of the cluster to be viewed must be **In service**.

Figure 1-35 Viewing monitoring information




Step 5 On the **Cloud Service Monitoring** page, click  on the left of the cluster ID to expand the compute unit list, and select the corresponding node to view the monitoring information.

Figure 1-36 Monitoring page

ID	Permanent Data Storage	Operation
<input type="checkbox"/> backend-2-2	--	View Metric Create Alarm Rule Configure Storage View Alarm Rule
<input type="checkbox"/> backend-3-3	--	View Metric Create Alarm Rule Configure Storage View Alarm Rule
<input type="checkbox"/> backend-1-1	--	View Metric Create Alarm Rule Configure Storage View Alarm Rule
<input type="checkbox"/> frontend-2-2	--	View Metric Create Alarm Rule Configure Storage View Alarm Rule
<input type="checkbox"/> frontend-3-3	--	View Metric Create Alarm Rule Configure Storage View Alarm Rule
<input type="checkbox"/> frontend-1-1	--	View Metric Create Alarm Rule Configure Storage View Alarm Rule

- **ID:** ID of the monitored instance, that is, the CloudTable cluster ID
- **View Metric:** In the compute unit list shown in [Figure 1-36](#), find the compute unit to be viewed and click **View Metric** to go to the monitoring metric page of the compute unit.

Step 6 Set the metrics to be viewed if there are too many metrics on the monitoring page.

1. If there are too many metrics, delete them on the **Select Metric** page.
2. If the metrics displayed on the page do not contain the desired metrics, add the metrics on the **Select Metric** page.
3. Select at least one metric. You can drag a selected metric and drop it to a desired location to sort the metrics.

Figure 1-37 Selecting metrics

Select Metric

Stream Loading... ✕

Rows Imported ... ✕

Bytes Received ... ✕

Maximum Base ... ✕

Maximum Cumu... ✕

Total Heartbeat ... ✕

Total Backend S... ✕

Current Heartbe... ✕

Current Backen... ✕

Fragment Instan... ✕

Stream Loading... ✕

Successful Stre... ✕

Failed Stream L... ✕

Discarded Rows... ✕

Total Rowset Data ✕

Messages Sent ... ✕

Rows Sent via T... ✕

Queue Length o... ✕

All Metrics 82

🔍

<input checked="" type="checkbox"/>	Metric	Description
<input checked="" type="checkbox"/>	CPU Usage	CPU usage of the instance
<input checked="" type="checkbox"/>	Memory Usage	Memory usage of the instance
<input checked="" type="checkbox"/>	Ratio of Used Storage ...	Ratio of the used storage space to the total storage space in the cluster
<input checked="" type="checkbox"/>	Scanners Enabled Ext...	--
<input checked="" type="checkbox"/>	Created hmc stubs (B...	These stubs are used for interactions between backend nodes

OK

Cancel

----End

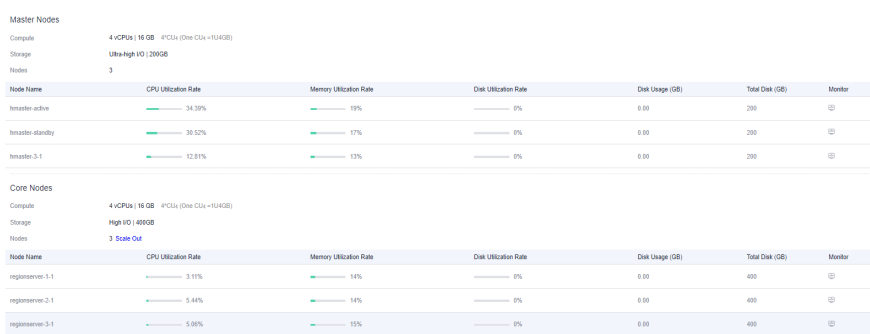
1.3.4 HBase Monitoring Page


The cluster details page displays the CPU, memory, disk utilization rate, as well as the disk usage and total disk capacity of each HBase node.

Accessing the HBase Cluster Details Page

- Step 1** Log in to the CloudTable console.
- Step 2** Click the cluster list page and select an HBase cluster.
- Step 3** View the monitoring metrics on the cluster details page.

Figure 1-38 HBase monitoring page



- Step 4** Click  on the details page to go to the CES console and view detailed monitoring metrics of the HBase cluster. For details, see [Viewing Cluster Monitoring Information](#).

----End

Monitoring Metrics

Table 1-20 Monitoring metrics

Metric	Description
CPU Utilization Rate	CPU utilization rate of the HBase cluster
Memory Utilization Rate	Memory utilization rate of the HBase cluster
Disk Utilization Rate	Data utilization rate of the HBase cluster

1.4 HBase Audit Logs

1.4.1 Supported Key Audit Log Operations

CloudTable uses CTS to record operations associated with CloudTable for later query, audit, and backtrack operations.

The following key operation traces of CloudTable are recorded in audit logs. For details, see [Table 1-21](#).

Table 1-21 CloudTable operation traces supported by CTS

Operation	Trace Name	Resource Type
Creating a cluster	createCluster	cluster
Deleting a cluster	deleteCloudTableCluster	cluster
Expanding cluster capacity	growCloudTableCluster	cluster
Restarting a cluster	rebootCloudTableCluster	cluster
Setting the storage quota	storageClusterAction	cluster
Modifying a feature	modifyClusterFeatures	cluster
Configuring parameters	modifyClusterSetting	cluster

1.4.2 Viewing Audit Logs

After you enable CTS, the system starts recording CloudTable operation logs. Operation records of the last 7 days can be viewed on the CTS management console.

This section includes the following content:

- [Enabling CTS](#)
- [Disabling the Audit Log Function](#)
- [Viewing CTS Logs of CloudTable](#)

Enabling CTS

A tracker will be automatically created after CTS is enabled. All traces recorded by CTS are associated with a tracker. Currently, only one tracker can be created for each account.

Step 1 On the CloudTable management console, choose **Service List > Management & Deployment > Cloud Trace Service**. The CTS management console is displayed.

Step 2 Enabling CTS

If you are a first-time CTS user and do not have any created trackers in the tracker list, enable CTS first. For details, see [Enabling CTS](#) in the [Cloud Trace Service Getting Started](#).

If you have enabled CTS, the system has automatically created a management tracker. Only one management tracker can be created and it cannot be deleted. You can also manually create a data tracker. For details, see **Managing Trackers > Creating a Tracker** in the [Cloud Trace Service User Guide](#).

----End

Disabling the Audit Log Function

If you want to disable the audit log function, disable the tracker in CTS.

Step 1 On the CloudTable management console, choose **Service List > Management & Deployment > Cloud Trace Service**. The CTS management console is displayed.

Step 2 Disable the audit log function by disabling the tracker. To enable the audit log function again, you only need to enable the tracker.

For details about how to enable or disable a tracker, see [Disabling or Enabling a Tracker](#) in the *Cloud Trace Service Getting Started*.

----End

Viewing CTS Logs of CloudTable

Step 1 Log in to the management console.

Step 2 Click **Service List** on the upper part of the page and choose **Management & Deployment > Cloud Trace Service**.

Step 3 In the navigation pane on the left, choose **Trace List**.

Step 4 In the upper right corner of the trace list, click **Filter** to set the search criteria.

The following four filter criteria are available:

- **Trace Source, Resource Type, and Search By**
 - **Trace Source:** Select **CloudTable**.
 - **Resource Type:** Select **All resource types** or specify a resource type.
 - **Search By:** Select **All** or any of the following options:
 - **Trace name:** If you select this option, you also need to select a specific trace name.
 - **Resource ID:** If you select this option, you also need to select or enter a specific resource ID.
 - **Resource name:** If you select this option, you also need to select or enter a specific resource name.
- **Operator:** Select a specific operator (at user level rather than tenant level).
- **Trace Status:** Available options include **All trace statuses, normal, warning, and incident**. You can only select one of them.
- **Start Date and End Date:** You can specify the time period to query traces.

Step 5 Click **Query**.

Step 6 Click  on the left of the trace to be queried to extend its details.

Figure 1-39 Trace

Trace Name	Resource Ty...	Trace So...	Resource ID ⓘ	Resource Name...	Trace Stat...	Operator ⓘ	Operation Time	Operation
rebootCloud...	cluster	CloudTab...	689ae304-086...	cloudtable-93...	normal		05/29/2018 14:50:16 GMT+...	View Trace

Trace ID	8b253930-630c-11e8-82ae-286ed488cbe3	Source IP Address	
Trace Type	ConsoleAction	Generated	05/29/2018 14:50:16 GMT+08:00

Step 7 Locate the row containing the target trace and click **View Trace** in the **Operation** column.

Figure 1-40 Viewing a trace

View Trace ×

```

{
  "service_type": "CloudTable",
  "user": {
    "name": " ",
    "id": "2f210e62179a425c96e5b57f25d1d492",
    "domain": {
      "name": " ",
      "id": "760dfba8429649e7a477445e1033a839"
    }
  },
  "time": "05/29/2018 14:50:16 GMT+08:00",
  "code": 200,
  "resource_type": "cluster",
  "resource_name": "cloudtable-9300",
  "resource_id": "689ae304-086b-4fa9-984e-16c58e3d75bf",
  "source_ip": " ",
  "trace_name": "rebootCloudTableCluster",
  "trace_type": "ConsoleAction",
  "api_version": "v1.0",
  "record_time": "05/29/2018 14:50:16 GMT+08:00",
  "trace_id": "8b253930-630c-11e8-82ae-286ed488cbe3",
  "trace_status": "normal"
}

```

For details about key fields in the CTS trace structure, see the [Trace Structure](#) in the *Cloud Trace Service User Guide*.

----End

1.5 Self-Healing from HBase Hotspotting

HBase is a distributed key-value database. Regions are the smallest units HBase data management. Poorly designed row keys make requests directed at a few fixed regions. As a result, the service pressure is high on a node, causing performance deterioration or even request failures.

MetricController instances are added to HBase. After hotspotting detection is enabled, the request traffic of each RegionServer node is monitored. Through aggregation analysis, the nodes and regions with excessive requests can be identified, accelerating hotspotting detection. In addition, this self-healing function transfers workload or performs region splitting. For hotspotting caused by a single rowkey and sequential writes where the self-healing function does not work, traffic limiting is provided instead to minimize the impact on other services on this node.

1.6 Global Secondary Indexes

1.6.1 Introduction

Scenario

HBase secondary indexes can accelerate conditional queries with filters. There are local secondary indexes (LSIs, also called HIndexes) and global secondary indexes (GSIs). Compared with LSIs, GSIs have better query performance and are suitable for scenarios that require low read latency.

HBase GSIs use independent index tables to store index data. When a given query condition hits an index, a full table query is converted into an exact range query on an index table. This way, query speed is greatly improved. You do not need to modify your application code to enable HBase GSIs.

NOTE

HBase GSIs cannot be created for a table that contains data.

Key features of HBase GSIs are as follows:

- **Composite index**
Multiple columns of different column families can be specified as index columns.
- **Covering index**
Multiple columns or column families can be stored in the index table in redundancy to cover all data needed for a query. With covering indexes, you can quickly query non-index columns in index query.
- **Index TTL**
Index table TTL takes effect if data table TTL is enabled. To ensure consistency with the data table, the index table TTL is automatically inherited from the index column and the column to overwrite an index of the data table and cannot be specified.
- **Online index change**
Indexes can be created, deleted, and their status can be modified online without affecting data table read and write.
- **Online index repair**
If the index data hit by a query is invalid, index data rebuilding is triggered to ensure that the final query result is correct.
- **Index tool**
The index tool helps you to check consistency, repair, create, and delete indexes, modify index status, and rebuild index data.

1.6.2 Restrictions

Application Scenarios

- GSIs cannot be used together with HIndexes. That is, they cannot be created in the same data table.
- Index tables do not support DR.
- DISABLE, DROP, MODIFY, and TRUNCATE cannot be directly performed on index tables.
- Index definition cannot be modified. You need to delete definitions and create indexes again. Other DDL operations on indexes are allowed, for example, modify index status, and delete and create indexes.

Creating Indexes

- An index name must comply with the regular expression requirements and does not support other characters. Regular expressions support the characters of [a-zA-Z_0-9-].
- The data table specified for index creation must exist. An index cannot be repeatedly created.
- The index table cannot have multiple versions.

Indexes cannot be created on data tables with multiple versions (**VERSION>1**). The **VERSION=1** setting is a must.

- The number of indexes in a single data table cannot exceed five.

Do not create too many indexes for a data table. Otherwise, bigger storage is required and write operations become slow. If more than five indexes need to be created, add the **hbase.gsi.max.index.count.per.table** parameter to the custom configuration **hbase.hmaster.config.expandor** of HMaster and set the parameter to a value greater than **5**. Restart HMaster to make the configuration take effect.

- The index name can contain a maximum of 18 characters.

Do not use long index names. If you have to, add the **hbase.gsi.max.index.name.length** parameter to the custom configuration **hbase.hmaster.config.expandor** of HMaster, set the parameter to a value greater than **18**, and restart HMaster to make the configuration take effect.

- Indexes cannot be created for index tables.

Indexes cannot be nested. Index tables are used only to accelerate queries and do not provide data table functions.

- Indexes that can be covered by existing indexes cannot be created.

If indexes you want to create are a subset of the existing indexes, they cannot be created. Duplicate indexes waste storage space. In the following example, index 2 cannot be created:

Creating a table

```
create 't1','cf1'
```

Creating index 1

```
hbase org.apache.hadoop.hbase.hindex.global.mapreduce.GlobalTableIndexer -  
Dtablename.to.index='t1' -Dindexspecs.to.add='idx1=>cf1:[q1],[q2]'
```

Creating index 2

```
hbase org.apache.hadoop.hbase.hindex.global.mapreduce.GlobalTableIndexer -
Dtablename.to.index='t1' -Dindexspecs.to.add='idx2=>cf1:[q1]'
```

- Indexes with the same name cannot be created in the same data table, but can be created in different data tables.
- The TTL of a column family in an index table is inherited from the original table, and must be the same as that of the original table.

The TTLs of all column families in an index table are the same and are inherited from a data table. The TTLs of associated column families in the data table must be the same. Otherwise, associated indexes cannot be created.

- Properties of user-defined index tables are not supported.

Writing Indexes

- Only the Put/Delete interface can be used to generate index data. If data is written to a data table with other methods (such as Increment and Append), the corresponding index will not be generated.
- When the index column data is defined as the string type, do not write special characters `\x00` and `\x01` (special invisible characters).
- Do not write data to index columns by specifying timestamps.

Index Query

- The index status must be **ACTIVE** during an index query.
- Index queries do not support **specified timestamp ranges**. If you need to query data within a time range by index, add a time column to store data timestamps. Otherwise, the data table will be used for query.
- Index query supports only **SingleColumnValueFilter**. Index acceleration cannot be triggered when other filters are used or no filter condition is used.

1.6.3 Using the GSI Tool

1.6.3.1 Creating Indexes

Scenario

- If a large amount of data exists in a table, you can add an index on a column to accelerate data queries.
- For a table that does not have indexes, this tool allows you to add and create indexes.

Usage

Run the following command on the HBase client to add or create indexes to a table (the added or created indexes will be in the ACTIVE state):

```
hbase org.apache.hadoop.hbase.hindex.global.mapreduce.GlobalTableIndexer -
Dtablename.to.index='table' -Dindexspecs.to.add='idx1=>cf1:[c1->string],[c2]#idx2=>cf2:[c1->string],
[c2]#idx3=>cf1:[c1];cf2:[c1]' -Dindexspecs.covered.family.to.add='idx2=>cf1' -
Dindexspecs.covered.to.add='idx1=>cf1:[c3],[c4]' -Dindexspecs.coveredallcolumn.to.add='idx3=>true' -
Dindexspecs.splitkeys.to.set='idx1=>[\x010,\x011,\x012]#idx2=>[\x01a,\x01b,\x01c]#idx3=>[\x01d,\x01e,\x0
1f]'
```

The parameters are described as follows:

- **tablename.to.index**: name of the data table for which an index is created
- **indexspecs.to.add**: mapping between the index name and the index column in the data table (definition of index column)
- (Optional) **indexspecs.covered.to.add**: column of the data table that is redundantly stored in an index table (definition of covering index column)
- (Optional) **indexspecs.covered.family.to.add**: column family of the data table that is redundantly stored in an index table (definition of covering index column family)
- (Optional) **indexspecs.coveredallcolumn.to.add**: all data in a data table that is redundantly stored in an index table (definition of all covering index columns)
- (Optional) **indexspecs.splitkeys.to.set**: pre-partition split keys of an index table. **Specify this parameter** in case hotspotting occurs in the region of the index table. You can configure pre-partitioning using the following characters:
 - '#' separates indexes.
 - '[' contains **splitkeys**.
 - ',' separates **splitkeys**.

 NOTE

Each **splitkey** set for per-partitioning must start with `\x01`.

- **indexspecs.to.addandbuild** (optional): Index data will be generated during data table creation. **If the data table is large, do not enable this parameter.** Use an index data generation tool instead.

The parameters in the preceding command are described as follows:

- **idx1**, **idx2**, and **idx3** are index names.
- **cf1** and **cf2** are column family names.
- **c1**, **c2**, **c3**, and **c4** are column names.
- **string** indicates a data type. The value can be **STRING**, **INTEGER**, **FLOAT**, **LONG**, **DOUBLE**, **SHORT**, **BYTE**, or **CHAR**.

 NOTE

- '#' separates indexes, ';' separates column families, and '.' separates column qualifiers.
- The column name and its data type must be included in '['.
- Column names and their data types are separated by ' - >'.
- If the data type of a column is not specified, the default data type (**string**) will be used.

1.6.3.2 Querying Indexes

Scenario

You can use the GSI tool to view the definition and status of indexes of a data table in batches.

Usage

Run the following command on the HBase client to view the definition and status of an index:

```
hbase org.apache.hadoop.hbase.hindex.global.mapreduce.GlobalTableIndexer -
Dtablename.to.show='table'
```

The related parameters are described as follows:

tablename.to.show: indicates the name of the table where the index to be queried is.

Figure 1-41 shows the query result. The index column definition, covering column definition, TTL, pre-partition information, and index status are displayed.

Figure 1-41 Index query result

```
Scbb negotiated timeout = 60000
2023-08-18 10:47:59.784 INFO [main] client.GlobalIndexTracker: GlobalIndexCacheTracker started successfully
IndexName : idx1, IndexColumns : [cf1:c1 -> type:STRING, cf1:c2 -> type:STRING], CoveredColumns : [cf1:c3 -> type:STRING, cf1:c4 -> type:STRING], CoveredFamilies : [], CoveredAllColumns : false, TTL : 2147483647,
SplitKeys : [\x01d,\x011,\x012], IndexState : ACTIVE
IndexName : idx2, IndexColumns : [cf2:c1 -> type:STRING, cf2:c2 -> type:STRING], CoveredColumns : [], CoveredFamilies : [cf1], CoveredAllColumns : false, TTL : 2147483647, SplitKeys : [\x01a,\x01b,\x01c], IndexSta
te : ACTIVE
IndexName : idx3, IndexColumns : [cf1:c1 -> type:STRING, cf2:c1 -> type:STRING], CoveredColumns : [], CoveredFamilies : [], CoveredAllColumns : true, TTL : 2147483647, SplitKeys : [\x01d,\x01e,\x01f], IndexState
ACTIVE
```

1.6.3.3 Deleting Indexes

Scenario

You can use the GSI tool to delete an index.

Usage

Run the following command on the HBase client to delete an index:

```
hbase org.apache.hadoop.hbase.hindex.global.mapreduce.GlobalTableIndexer -
Dtablename.to.index='table' -Dindexnames.to.drop='idx1#idx2'
```

The parameters are described as follows:

- **tablename.to.index:** indicates the name of the table where the index to be deleted is.
- **indexnames.to.drop:** indicates the name of the index to be deleted. You can specify multiple indexes and separate them with number signs (#).

1.6.3.4 Changing Index Status

Index Status

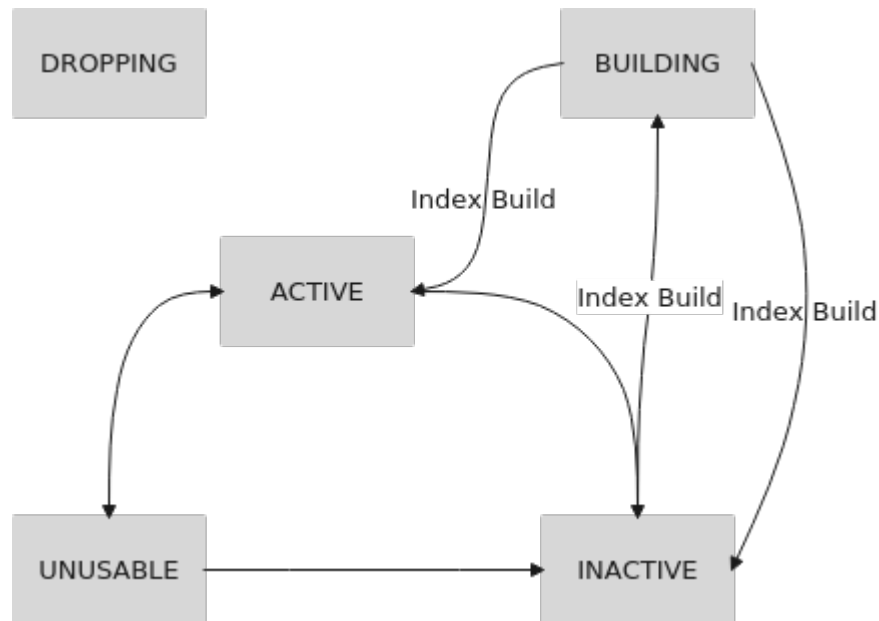
A GSI has the following states:

- **ACTIVE:** The index can be read and written normally.
- **UNUSABLE:** The index is disabled. Index data can be written normally but cannot be used for query.
- **INACTIVE:** The index is abnormal. The index data is inconsistent with that in the data table. The indexed data is skipped and the index cannot be used during data query.

- **BUILDING**: Index data is being generated in batches. After the generation is complete, the index is automatically switched to the **ACTIVE** state. In this state, data can be read and written properly.
- **DROPPING**: The index is being deleted. The indexed data is skipped, and the index cannot be used during data query.

You can change index status with the GSI tool. [Figure 1-42](#) describes the states and transitions between them.

Figure 1-42 State transitions



Scenario

You can use the GSI tool to disable or enable an index.

Usage

Run the following command on the HBase client to disable or enable an index:

```
hbase org.apache.hadoop.hbase.hindex.global.mapreduce.GlobalTableIndexer -
Dtablename.to.index='table' -D[idx_state_opt]='idx1'
```

The related parameters are described as follows:

- **tablename.to.index**: indicates the name of the data table whose index status needs to be changed.
- **idx_state_opt**: indicates the target status of the index to be modified. The options are as follows:
 - **indexnames.to.inactive**: disables a specified index (**INACTIVE**).
 - **indexnames.to.active**: enables a specified index (**ACTIVE**).
 - **indexnames.to.unusable**: switches the specified index to **UNUSABLE**.

The following example changes the state of **idx1** of **table** to **UNUSABLE**:

```
hbase org.apache.hadoop.hbase.hindex.global.mapreduce.GlobalTableIndexer -
Dtablename.to.index='table' -Dindexnames.to.unusable='idx1'
```

After the command is executed, check the index information.

```
hbase org.apache.hadoop.hbase.hindex.global.mapreduce.GlobalTableIndexer -
Dtablename.to.show='table'
```

As shown in [Figure 1-43](#), the status of index **idx1** is changed.

Figure 1-43 idx1 status

```
hbase> org.apache.hadoop.hbase.hindex.global.mapreduce.GlobalTableIndexer started successfully
IndexName = idx1, IndexColumns : [cf1:c1 -> type:STRING, cf1:c2 -> type:STRING], CoveredColumns : [cf1:c3 -> type:STRING, cf1:c4 -> type:STRING], CoveredFamilies : [], CoveredAllColumns : false, TTL : 2147483647,
SplitKeys : [\x01a,\x01b,\x01c], IndexState : DISABLE
IndexName = idx2, IndexColumns : [cf2:c1 -> type:STRING, cf2:c2 -> type:STRING], CoveredColumns : [], CoveredFamilies : [cf1], CoveredAllColumns : false, TTL : 2147483647, SplitKeys : [\x01a,\x01b,\x01c], IndexSt
ate : ACTIVE
IndexName = idx3, IndexColumns : [cf1:c1 -> type:STRING, cf2:c1 -> type:STRING], CoveredColumns : [], CoveredFamilies : [], CoveredAllColumns : true, TTL : 2147483647, SplitKeys : [\x01d,\x01e,\x01f], IndexState
: ACTIVE
```

1.6.4 GSI APIs

APIs that use global indexes are packed in the **org.apache.hadoop.hbase.hindex.global.GlobalIndexAdmin** class. Related APIs are listed in the following table.

Operation	API	Description
Adding an index	addIndices()	Adds a specified index to an empty table but skips index data generation. You can use this API to add indexes in batches to a table that contains a large amount of data and use the GlobalTableIndexer tool to build index data.
	addIndicesWithData()	Adds a specified index to a table that contains data and generates index data for existing table data. You can call this API to generate an index and then generate index data when table data is being stored. This API is not recommended when a data table contains a large amount of data.
Deleting an index	dropIndices()	Deletes both index metadata and index data. After deletion, the index cannot be used for scan or filter operations.
Changing index status	alterGlobalIndicesUnusable()	Disables a specified index so that it cannot be used for the scan or filter operation.
	alterGlobalIndicesActive()	Enables an index specified by the user so that it can be used for the scan or filter operation.
	alterGlobalIndicesInactive()	Disables a specified index and skips index data generation so that it cannot be used for the scan or filter operation. This API is usually used for index data re-building.
Viewing created indexes	listIndices()	Lists all indexes in a given table.

1.6.5 Querying Data with Indexes

Index-based Query

You can use **SingleColumnValueFilter** to query data in a table with indexes. When the query condition hits an index, the query speed is much faster than that of an ordinary table query.

Typical index query conditions are as follows:

- Query by multiple AND conditions
 - When the columns used for a query contain at least the first indexed column, the query performance is optimized.

Assume that a composite index is created for C1, C2, and C3.

The index takes effect in the following queries:

```
Filter_Condition (IndexCol1) AND Filter_Condition (IndexCol2) AND Filter_Condition (IndexCol3)
Filter_Condition (IndexCol1) AND Filter_Condition (IndexCol2)
Filter_Condition (IndexCol1) AND Filter_Condition (IndexCol3)
Filter_Condition (IndexCol1)
```

The index does not take effect in the following queries:

```
Filter_Condition (IndexCol2) AND Filter_Condition (IndexCol3)
Filter_Condition (IndexCol2)
Filter_Condition (IndexCol3)
```

- When you use "Index Column AND Non-Index Column" as a query condition, the index can improve query performance. If a non-index column hits a covering column, the query performance is optimal. If a non-index column needs to be frequently queried, you are advised to define it as a covering column. The following statement is an example:

```
Filter_Condition (IndexCol1) AND Filter_Condition (NonIndexCol1)
Filter_Condition (IndexCol1) AND Filter_Condition (IndexCol2) AND Filter_Condition (NonIndexCol1)
```
- When multiple columns are used for query, you can specify a value range for only the last column in the composite index and set other columns to specified values.
Assume that a composite index is created for C1, C2, and C3. In a range query, only the value range of C3 can be set. The search criteria are "C1 = XXX, C2 = XXX, and C3 = Value range."

- Query by multiple OR conditions

Assume that a composite index is created for C1, C2, and C3.

- If only the first field in the index column is searched (range filtering is supported), indexing improves the query performance.

```
Filter_Condition (IndexCol1) OR Filter_Condition (IndexCol1) OR Filter_Condition (IndexCol1)
```

- When non-index and non-index columns are searched, indexes cannot be hit, and query performance is not improved.

```
Filter_Condition (IndexCol1) OR Filter_Condition (NonIndexCol1)
```

- During a combined query, if the outermost layer contains the OR condition, the index cannot be hit, and the query performance is not improved.

```
Filter_Condition (IndexCol1) OR Filter_Condition (NonIndexCol1)
(Filter_Condition (IndexCol1) AND Filter_Condition (IndexCol2)) OR (Filter_Condition (NonIndexCol1))
```


 **NOTE**

Reduce the use of OR conditions, especially an OR condition used together with a range condition. Otherwise, large-scale data is queried in slow speed when indexes are hit.

2 ClickHouse User Guide

2.1 ClickHouse Cluster Management

2.1.1 Overview

ClickHouse is a column-oriented database for online analytical processing. It supports SQL query and provides good query performance. The aggregation analysis and query performance based on large and wide tables is excellent, which is one order of magnitude faster than other analytical databases. CloudTable ClickHouse cluster management also provides the following functions:

In ClickHouse cluster management, you need to create a dedicated cluster and can use it on demand. Dedicated clusters are physically isolated and are not affected by other services, facilitating user management.

After the ClickHouse cluster is created, use the client to access the cluster. For details, see the following sections.

- [Preparing an ECS.](#)
- [Using a Client to Access a ClickHouse Cluster.](#)
- [ClickHouse Commands.](#)

2.1.2 Managing ClickHouse Clusters

2.1.2.1 Creating a ClickHouse Cluster

You can centrally manage clusters with CloudTable. A cluster is necessary for using CloudTable. This section describes how to create a cluster on the CloudTable console.

ClickHouse clusters support two billing modes: pay-per-use and yearly/monthly. The pay-per-use billing mode is used by default when you create a cluster. In pay-per-use mode, compute units are charged by the duration you use them, with a billing cycle of one hour. With this mode, you can start or stop a cluster at any time and pay what you use. The yearly/monthly billing mode is recommended for long-term users to enjoy a larger discount. You can also customize a ClickHouse

cluster with specified computing capabilities and storage space to meet your business needs.

Procedure


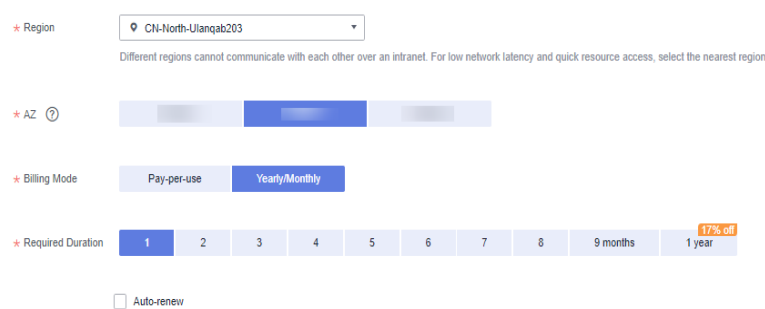
- Step 1** Log in to the CloudTable console.
- Step 2** Click  in the upper left corner to select a region.
- Step 3** Click **Buy Cluster** in the upper right corner.
- Step 4** Configure basic cluster information by referring to the following table.

Figure 2-1 Setting parameters



* Region: CN-North-Ulanqab203
Different regions cannot communicate with each other over an intranet. For low network latency and quick resource access, select the nearest region.

* AZ: [selected]

* Billing Mode: Pay-per-use | Yearly/Monthly

* Required Duration: 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 months | 1 year 17% off

Auto-renew

Table 2-1 Parameter description

Parameter	Description
Region	Current working region of the cluster. <ul style="list-style-type: none"> Select the region for the cluster nodes to run. For details about regions, see Regions and Endpoints.
AZ	Select the AZ associated with the cluster's region. For details about the AZs supported by ClickHouse, see Region and AZ .
Billing Mode	Select Pay-per-use or Yearly/Monthly .
Required Duration	This option is available only when Billing Mode is set to Yearly/Monthly . Configure this parameter based on your service requirements.
Auto-renew	If you select Auto-renew when creating a cluster, the system will automatically renew your subscription before it expires.

Figure 2-2 Configuring cluster networks

* Name [?](#)

* VPC [?](#) [View VPC](#)

* Subnet [?](#) [View Subnet](#)

* Security Group [?](#) [View Security Group](#)

* Database Engine Doris HBase ClickHouse

* ClickHouse Kernel

* Cluster HA Enabled

Enable: Two replicas are maintained for each shard on a single node, with three ZooKeeper nodes as the default.
Disable: Only one single ClickHouse node with a single shard and replica is provided for testing purposes. This configuration does not support scale-out and cannot be used in any production environments.

Table 2-2 Cluster and network configuration

Parameter	Description
Name	Name of a cluster. A cluster name must start with a letter and contain 4 to 32 letters, digits, and hyphens (-). It cannot contain other special characters.
VPC	A VPC is a secure, isolated, and logical network environment. Retain the default settings. If there is no available VPC, click View VPC to access the VPC console and create a new VPC.
Subnet	Specify a VPC subnet. A subnet provides dedicated network resources that are isolated from other networks, improving network security.
Security Group	A security group is used to control ECS access within a security group or between security groups by defining access rules. You can define different access control rules for a security group. These rules can specify which ECS ports or protocols are accessible and can be used to control inbound and outbound network traffic of VMs. After an ECS is added to the security group, it is protected by these access control rules. ECSs that do not belong to the security group cannot communicate with ECSs in the security group. You can use an existing security group or click View Security Group to create a new one. For more information about security groups, see Security Group in the <i>Virtual Private Cloud User Guide</i> .
Database Engine	Select the type of cluster to be created.

Parameter	Description
Clickhouse Kernel Version	Retain the default settings.
Cluster HA	<p>Enable: Two replicas are maintained for each shard on a single node, with three ZooKeeper nodes as the default.</p> <p>Disable: Only one single ClickHouse node with a single shard and replica is provided for testing purposes. This configuration does not support scale-out and cannot be used in any production environments.</p> <p>NOTE</p> <ul style="list-style-type: none"> • A single node does not support scale-out or data migration. • An environment with just a single node is unsuitable for genuine production tasks or customer-facing activities. • Users of a single-node system must establish a non-replicated table. <p>Example of creating a table: <pre>CREATE TABLE test_p (`EventDate` DateTime,`CounterID` UInt32,`UserID` UInt32,`ver` UInt16)ENGINE = ReplacingMergeTree()PARTITION BY sipHash64(CounterID) ORDER BY (CounterID, EventDate, intHash32(UserID))SAMPLE BY intHash32(UserID)SETTINGS index_granularity = 8192;</pre></p>

Table 2-3 Compute node parameters

Parameter	Description
Compute	<p>Select proper compute specifications based on service requirements. Each specification displays the CPU, memory, and recommended application scenarios of a single node.</p> <p>NOTE</p> <p>Available compute specifications:</p> <ul style="list-style-type: none"> • 8U32G • 16U64 • 32U128G • 64U256G • 128U512G
Storage	<p>Select the disk specifications and capacity of the Clickhouse compute node.</p> <ul style="list-style-type: none"> • Available storage specifications: <ul style="list-style-type: none"> - High I/O - Ultra-high I/O - Extreme SSD • The capacity ranges from 500 GB to 10,000 GB per node.

Parameter	Description
Nodes	Set the number of nodes in a cluster. The value ranges from 2 to 20 .

Figure 2-3 ZooKeeper node configuration

ZooKeeper Node Configuration ?

* Compute

* Storage GB / Nodes

* Nodes

Table 2-4 ZooKeeper node parameters

Parameter	Description
Compute	Select proper compute specifications based on service requirements. Each specification displays the CPU, memory, and recommended application scenarios of a single node. NOTE Available compute specifications: <ul style="list-style-type: none"> • 4U8G (This specification is not available for production environments.) • 4U16G • 8U32G • 16U64G • 32U128G
Storage	Ultra-high I/O is recommended. Set the disk capacity to 100 GB per node.
Nodes	Number of nodes in a cluster. The default value is 3 .

Parameter	Description
Enterprise Project	<p>You can group related cloud resources (for example, resources used for the same purpose) and manage them by enterprise project.</p> <p>NOTE</p> <ul style="list-style-type: none"> You can delete a user that was authorized to manage an enterprise project if the enterprise's business has changed and the user should not have the permissions to manage the enterprise project. You can delete a user or multiple users. After a user that was authorized to manage an enterprise project is deleted, the user cannot manage the enterprise project. If the user needs to manage the enterprise project, authorize the user to manage the enterprise project again. For details, see Enterprise Management User Guide.

Figure 2-4 Setting passwords

* Username

* Password

* Confirm Password

* Broker Process

* Enable Https After Https is enabled, you need to use the MySQL 8 or later client to log in.

* Enable API Authentication Enabling this feature will require account and password authentication when calling the cluster's REST APIs, which may slightly reduce the cluster performance.

* Enable Hot/Cold

* Enterprise Project [View Enterprise Project](#)

Step 5 Set the parameters and click **Buy Now**.

Step 6 On the displayed page, confirm the cluster specification order information and click **Submit**. The cluster creation task is submitted.

Step 7 Click **Back to Cluster List** to view the cluster status.

The cluster creation task takes some time. The initial status of the cluster is **Creating**. After the cluster is created, the cluster status changes to **In service**.

Step 8 Submit the creation task of a yearly/monthly cluster.

Click **Pay**. On the displayed purchase page, confirm the information, select a proper payment method, and confirm the payment.

Return to the console and check the cluster status. Cluster creation takes some time. The initial status of the cluster is **Creating**. After the cluster is created, the cluster status changes to **In service**.

----End

2.1.2.2 Introduction to ClickHouse Cluster Management

Log in to the CloudTable console. In the left navigation pane, click **Cluster Management**. On the **Cluster Management** page, a cluster list is displayed. If there are a large number of clusters, you can turn pages to view clusters in any state.







Figure 2-5 Cluster list

Table 2-5 Cluster management list parameters

Parameter	Description
Cluster Name	Cluster name specified when a cluster is created.
Cluster Status	Cluster statuses include Creating , In service , Sub-health , Creation failed , and Frozen .
Task Status	Task status of a cluster.
Database Engine	The values include HBase and ClickHouse.
Version	Cluster version.
Enterprise Project	You can group related cloud resources (for example, resources used for the same purpose) and manage them by enterprise project. NOTE <ul style="list-style-type: none"> You can delete a user that was authorized to manage an enterprise project if the enterprise's business has changed and the user should not have the permissions to manage the enterprise project. You can delete a user or multiple users. After a user that was authorized to manage an enterprise project is deleted, the user cannot manage the enterprise project. If the user needs to manage the enterprise project, authorize the user to manage the enterprise project again.
Created	Time when a cluster is created.
Billing Mode	Cluster billing mode.
Access Address (Intranet)	Private network link address.

Parameter	Description
Operation	<ul style="list-style-type: none"> Click Monitor to access the CloudTable service monitoring page. Click Delete to delete a cluster. Click Restart to restart the cluster. Choose More > Change to Yearly/Monthly to change the billing mode to yearly/monthly. Click Scale Out to add compute nodes. Click More <Unsubscribe in the Operation column to unsubscribe from the yearly/monthly cluster.

Table 2-6 Icon description

Icon	Description
	Click  to view all projects.
	Click  to refresh the cluster list.
	Enter a cluster name in the search box and click  to search for it.

Cluster Status

Table 2-7 Cluster status description

Status	Description
Creating	Indicates that a cluster is being created.
In service	If a cluster is successfully created and can provide services, the cluster status is In service .
Sub-health	If the cluster status cannot be monitored within the specified time, the cluster status changes to Sub-health . Manual intervention is required to recover a cluster that is in Sub-health status. For example, you can restart the cluster to recover the cluster.
Creation failed	Indicates that a cluster fails to be created.

Status	Description
Frozen	<p>If the balance is insufficient for renewing a cluster, the cluster status is Frozen.</p> <p>To unfreeze the cluster, you need to recharge your account to ensure that the account balance is not 0..</p> <p>NOTE A frozen cluster is unavailable and its all ECSs are shut down. After being unfrozen, the cluster recovers to the In service state. If you do not renew the cluster before the freeze period ends, the cluster will be deleted.</p>

Task Status


Table 2-8 Task status description

Status	Description
Deleting	Indicates that a cluster is being deleted.
Restarting	Indicates that a cluster is being restarted.
Scaling out	Indicates that a cluster is being scaled out.

2.1.2.3 Viewing the Basic Information of a ClickHouse Cluster

You can monitor and manage the clusters you create. On the CloudTable management console, click **Cluster Management**. In the cluster list, locate the cluster to be viewed and click the cluster name to access the basic information page. You can view the cluster information and network configurations.

Table 2-9 Cluster information

Parameter	Description
Cluster Name	Cluster name specified when a cluster is created.
Cluster ID	Cluster ID.
Cluster Status	Cluster status information.
Version	Kernel version of the cluster.
Access Address	Address for accessing the cluster. You can click  to copy the access link to the clipboard.

Parameter	Description
Access Address Port	Port for accessing the cluster. <ul style="list-style-type: none"> • HTTP port: 8123 • TCP port: 9000 • HTTPS port: 8443 • Secure TCP port: 9440
Billing Mode	Billing mode of the cluster. <ul style="list-style-type: none"> • Pay-per-use • Yearly/monthly
Created	Time when a cluster is created.
Enterprise Project	You can group related cloud resources (for example, resources used for the same purpose) and manage them by enterprise project. NOTE <ul style="list-style-type: none"> • You can delete a user that was authorized to manage an enterprise project if the enterprise's business has changed and the user should not have the permissions to manage the enterprise project. • You can delete a user or multiple users. • After a user that was authorized to manage an enterprise project is deleted, the user cannot manage the enterprise project. If the user needs to manage the enterprise project, authorize the user to manage the enterprise project again.
Database Engine	A core service used to store, process, and protect data.
Channel Status	Whether secure channels are enabled.
Enable Secure and Non-secure Channels	This parameter is displayed if secure channels are not enabled.

Table 2-10 Network configuration

Parameter	Description
Region	Working area of the cluster. Set this parameter when creating a cluster.
AZ	AZ you select during cluster creation.
VPC	VPC you select during cluster creation. A VPC is a secure, isolated, logical network environment.

Parameter	Description
Subnet	Subnet you select during cluster creation. A subnet provides dedicated network resources that are logically isolated from other networks, improving network security.
Security Group	Security group you select during cluster creation.

Table 2-11 Compute node parameters

Parameter	Description
Compute	Select proper compute specifications based on service requirements. Each specification displays the CPU, memory, and recommended application scenarios of a single node. NOTE Available compute specifications: <ul style="list-style-type: none"> • 8U32G • 16U64 • 32U128G • 64U256G • 128U512G
Storage	Select the disk specifications and capacity of the Clickhouse compute node. <ul style="list-style-type: none"> • Available storage specifications: <ul style="list-style-type: none"> - High I/O - Ultra-high I/O - Extreme SSD • The capacity ranges from 500 GB to 10,000 GB per node.
Nodes	Set the number of nodes in a cluster. The value ranges from 2 to 20 .

Table 2-12 ZooKeeper node parameters

Parameter	Description
Compute	<p>Select proper compute specifications based on service requirements. Each specification displays the CPU, memory, and recommended application scenarios of a single node.</p> <p>NOTE Available compute specifications:</p> <ul style="list-style-type: none"> • 4U16G • 8U32G • 16U64G • 32U128G
Storage	<p>Select the disk specifications and capacity of the Clickhouse compute node.</p> <ul style="list-style-type: none"> • Available storage specifications: <ul style="list-style-type: none"> - General-purpose SSD - Ultra-high I/O - Extreme SSD • The capacity ranges from 100 to 1000.
Nodes	Number of nodes in a cluster. The default value is 3 .

2.1.2.4 Restarting a ClickHouse Cluster

If a cluster is in the unbalanced state or cannot work properly, you may need to restart it for restoration.

 **NOTE**

If your cluster is in arrears, this function may be unavailable. Please top up your account in time.

Impact on the System

- A cluster cannot provide services during the restart. Therefore, before the restart, ensure that no task is running and all data is saved.
- If a cluster is processing transactional data, for example, importing data, querying data, files may be damaged or the cluster may fail to be restarted once the cluster is restarted. You are advised to stop all cluster tasks before restarting a cluster.
- If the restart fails, the cluster may be unavailable. Try again later or contact technical support.

Procedure

Step 1 Log in to the CloudTable console.



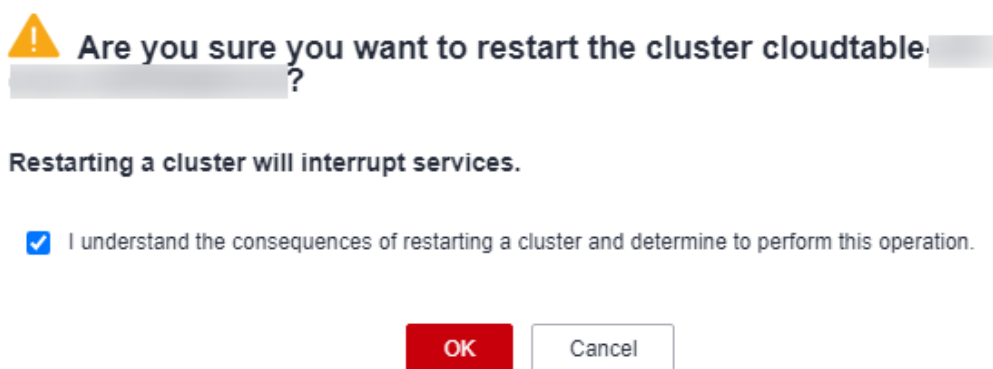
- Step 2** Click  in the upper left corner of the console and select a region.
- Step 3** In the navigation tree, click **Cluster Management**.
- Step 4** In the upper right corner of the cluster list, enter the name of a cluster in the search box and click .
- Step 5** In the **Operation** column of the cluster, click **More > Restart**.
- Step 6** In the dialog box that is displayed, select the check box and click **OK** to restart the cluster.

Figure 2-6 Confirming the restart



----End

2.1.2.5 Deleting a ClickHouse Cluster

You can delete clusters you will no longer use. Deleting a CloudTable ClickHouse cluster will clear all resources and data related to the cluster. This operation cannot be undone. Exercise caution when deleting a cluster.

Procedure



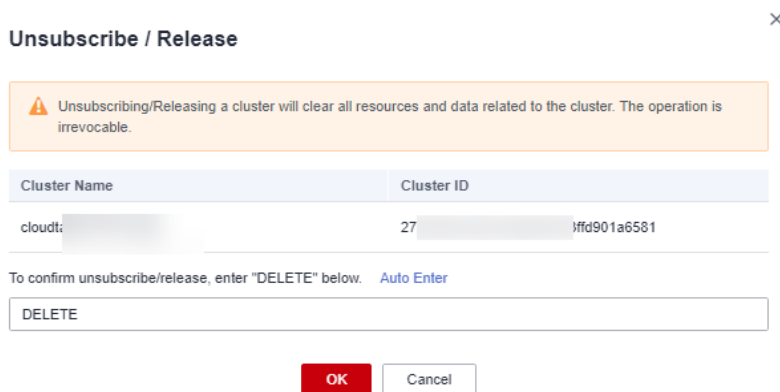
- Step 1** Log in to the CloudTable console.
- Step 2** Click  in the upper left corner to select a region.
- Step 3** In the navigation tree, click **Cluster Management**.
- Step 4** In the upper right corner of the cluster list, enter the name of a cluster in the search box and click .
- Step 5** In the **Operation** column of the cluster, click **More > Delete**.
- Step 6** In the displayed dialog box, enter **DELETE** or click **Auto Enter**, and click **OK** to delete the cluster.

Figure 2-7 Confirming the deletion



----End

2.1.2.6 Scaling Out a ClickHouse Cluster

Add compute nodes in a ClickHouse cluster. The scale-out does not affect system running.

NOTE

- Scale-out cannot be performed if your account balance is insufficient.
- Scale-out cannot be performed if your quota is insufficient.
- ZooKeeper nodes can be added for scale-out.

Procedure

- Step 1** Log in to the CloudTable console.
- Step 2** Click **Cluster Management**.
- Step 3** In the cluster list, locate the row that contains the target cluster and click **Scale Out** in the **Operation** column.

NOTE

You can also click **Scale Out** on the cluster details page to go to the **Scale Out** page.

- Step 4** Click + to add nodes.
- Step 5** Confirm the fee and resource quota, and click **OK**.

----End

ClickHouse Scale-Out Statuses

Table 2-13 Scale-out statuses

Task Status	Description
Scaling out	Indicates that a cluster is being scaled out.

Task Status	Description
Scale-out failed	Indicates that the cluster fails to be scaled out.

Viewing Scaling Details

- After you click **OK**, the cluster task status changes to **Scaling out**. After the cluster scale-out is complete, the cluster status changes to **In service**.
- After you click **OK**, the cluster task status changes to **Scaling out**. If the scale-out fails, the cluster status is **In service** and the task status is **Scale-out failed**.

2.1.2.7 ClickHouse Disk Capacity Expansion

The vertical capacity expansion is performed by expanding disk capacity. Data is usually stored on the compute nodes, and disk expansion is required when the disk capacity of the compute nodes is insufficient.

NOTE

- During disk capacity expansion, cluster services may jitter.
- You can perform disk capacity expansion on a cluster when the cluster is in the **In service** state and no task is being performed (such as scale-out, parameter modification, and specification change).

Procedure

- Step 1** Log in to the CloudTable console.
- Step 2** Click **Cluster Management**.
- Step 3** In the cluster list, locate the row that contains the target cluster and click **Disk Expansion** in the **Operation** column. If you fail to find **Disk Expansion**, click **More** and choose **Disk Expansion** in the **Operation** column.

Figure 2-8 Disk Expansion page

Disk Expansion

1. Perform scale-out during off-peak hours to avoid intermittent service interruptions

2. The fee for changing the on-demand billing method refers to the new fee for this change

Cluster ID/Name	d8de1221-d97d-49cc-ac2c-0f4a8a071...
Region/Availability Zone	CN-North-Ulanqab203/AZ3
Billing Mode	Pay-per-use
Compute Node Specification	4 vCPUs 16 GB
Storage Specifications	High I/O 500 GB
Compute Nodes	2
Zookeeper Node Specification	4 vCPUs 8 GB
Zookeeper Storage Specifications	Ultra-high I/O 100 GB
Zookeeper Nodes	3
Configured Node	<input checked="" type="radio"/> Compute Node Configuration <input type="radio"/> ZooKeeper Node Configuration
Compute Node Increased To	<input type="text" value="600"/> GB / Nodes

OK
Cancel

NOTE

You can also click **Disk Expansion** on the cluster details page to go to the page for expanding disk capacity.

Figure 2-9 Accessing the Disk Expansion page

Compute Node Configuration						
4 vCPUs 16 GB <small>4 vCPU (One CPU + 16GB)</small> Specification Expansion						
High I/O 500GB Disk Expansion						
2 Scale Out						
Node Name	CPU Utilization Rate	Memory Utilization Rate	Disk Utilization Rate	Disk Usage (GB)	Total Disk (GB)	Monitor
server-1-1	5.42%	6%	0.05%	0.45	500	
server-2-2	6.53%	6%	0.05%	0.45	500	

ZooKeeper Node Configuration						
4 vCPUs 8 GB <small>4 vCPU (One CPU + 8GB)</small> Specification Expansion						
Ultra-high I/O 100GB Disk Expansion						
3						
Node Name	CPU Utilization Rate	Memory Utilization Rate	Disk Utilization Rate	Disk Usage (GB)	Total Disk (GB)	Monitor
zookeeper-1-1	9.21%	14%	0.05%	0.06	100	
zookeeper-2-2	9.84%	14%	0.05%	0.06	100	
zookeeper-3-3	7.47%	15%	0.05%	0.06	100	

Step 4 Select the target node type.

Step 5 Press + to increase nodes.

Step 6 Confirm the fee and resource quota, and click **OK**.

----End

Disk Expansion Status

Table 2-14 Disk expansion status description

Status	Description
Disk expanding	Indicates that a cluster is being scaled out.
In service	This status is shown in Cluster Status , indicating that the scale-out is complete and the cluster can provide services.
Scale-out failed	This status is shown in Task Status , indicating that the cluster fails to be scaled out.

2.1.2.8 ClickHouse Specification Change

Prerequisites

You can perform specification expansion on a cluster when the cluster is in the **In service** state and no task is being performed (such as scale-out and disk expansion).

 **NOTE**

- You can only increase cluster specifications. To decrease cluster specifications, create a cluster with smaller specifications and migrate the data from the current cluster to it.
- You can change the specifications of only one node type (the compute or ZooKeeper node) at a time. After the change, nodes in other types still maintain their original specifications.
- During the specification change, the cluster is unavailable and the change cannot be initiated when a change is in progress.
- The system is unavailable during the specification change.

Procedure

Step 1 Log in to the CloudTable console.

Step 2 Click **Cluster Management**.

Step 3 In the cluster list, locate the row that contains the target cluster and choose **More > Specification Expansion** in the **Operation** column.

Figure 2-10 Specification expansion

Specification Expansion

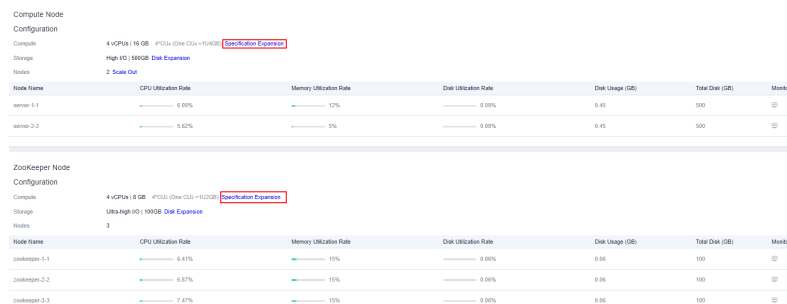
1 1. The system is unavailable during specification changes
2. The fee for changing the on-demand billing method refers to the new fee for this change

Cluster ID/Name	4e91f35c-704d-43f2-8391-de0fc3d937fe/cloudtab...
Region/Availability Zone	CN-North-Ulanqab203/AZ3
Billing Method	Pay-per-use
Compute Node Specification	<input type="text"/>
Compute Nodes	2
Zookeeper Node Specification	<input type="text"/>
ZooKeeper Nodes	3
* Configured Node	<input checked="" type="radio"/> Compute Node Configuration <input type="radio"/> ZooKeeper Node Configuration
* Calculation Specifications	<input type="text"/> 4*CU ₂ (One CU ₂ =1U2GB)
Total Fee ?	¥7.91/hour
Expansion Fee ?	¥0.00/hour

 **NOTE**

You can also click **Specification Expansion** on the cluster details page to switch to the **Specification Expansion** page.

Figure 2-11 Accessing the Specification Expansion page



Step 4 Select the target node type.

Step 5 Select the CPU and memory specifications from the drop-down list.

Step 6 Confirm the fee and resource quota, and click **OK**.

After you click **OK**, the cluster status changes to **Sub-health** and the task status changes to **Resizing Flavor**. After the cluster specifications are changed, the cluster status changes to **In service** and the task status is cleared.

----End

Specification Expansion Status

Table 2-15 Specification expansion status description

Status	Description
Resizing Flavor	Indicates that the specifications of the target cluster are being changed.
In service	Indicates that the specifications of the target cluster are changed and the cluster can provide services.
Resizing flavor failed	Indicates that the specifications of the target cluster failed to be changed.

2.1.2.9 Changing the ClickHouse Billing Method to Yearly/Monthly

If you want to use a pay-per-use cluster for a long time, you can change its billing mode to yearly/monthly. This section describes how to change the cluster billing mode to yearly-monthly.

Changing the Billing Mode from Pay-per-Use to Yearly/Monthly

- Prerequisites: The cluster billing mode is **Pay-per-use**.

- Procedure for changing the billing mode
 - a. Log in to the CloudTable console.
 - b. Click **Cluster Management**.
 - c. In the **Operation** column of a cluster, choose **More > Change to Yearly/Monthly**.

 **NOTE**

You can also click **Change to Yearly/Monthly** on the cluster details page to change the billing mode.

- d. On the displayed page, select the required duration and click **Pay**.
- e. On the payment page, select a payment method and click **Pay**.
- f. Return to the cluster list, click the target cluster to check the billing mode on its details page.

Yearly/Monthly unsubscription

- Prerequisites: The cluster billing mode is **Yearly/Monthly**.
- Unsubscribing from a yearly/monthly cluster
 - a. Log in to the CloudTable console.
 - b. Click **Cluster Management**.
 - c. In the **Operation** column of a cluster, choose **More > Unsubscribe**.
 - d. In the displayed dialog box, confirm the warning message, and click **OK**.
 - e. On the displayed page, select the unsubscription reason and confirm the prompt message. After the unsubscription, resources that are not moved to the recycle bin will be immediately deleted and cannot be restored. Back up the data you may use in the future and click **Confirm**.
 - f. Confirm the information shown in the displayed dialog box to complete the unsubscription.

2.1.3 Using ClickHouse

2.1.3.1 Preparing an ECS

Procedure

Step 1 Buy an ECS.


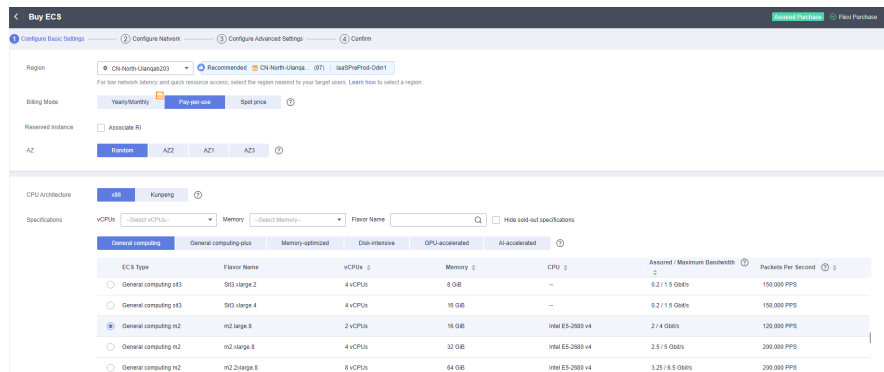
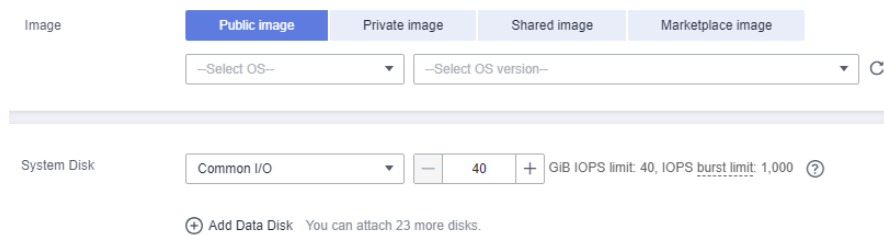
1. Log in to the Huawei Cloud console.
2. Click  in the upper left corner to select a region.
3. Click **Service List** and choose **Compute > Elastic Cloud Server**.
4. Click **Buy ECS**.
5. Configure ECS parameters.
 - a. Select the CPU architecture and specifications.

Figure 2-12 Configuring basic settings



b. Select the image and disk specifications.

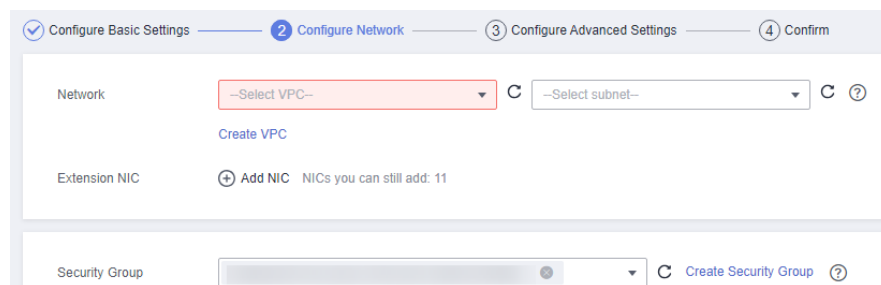
Figure 2-13 Selecting the image and disk specifications



Step 2 Click **Next: Configure Network**.

1. Select a VPC and a security group.

Figure 2-14 Configuring Network



NOTE

- Select the existing VPCs and security groups.
 - If no VPC is available, click **Create VPC** to create one.
 - If no security group is available, click **Create Security Group** to create one.
 - For details about how to create an ECS, see [Elastic Cloud Server User Guide](#).
2. Select **Auto assign** and set **Billed By** to **Traffic**.

Figure 2-15 Configuring EIP type and bandwidth billing type

Step 3 Click **Next: Configure Advanced Settings**.

Configure the ECS name and password.

Figure 2-16 Configuring advanced settings

Step 4 Click **Next: Confirm** to confirm the configuration information.

Figure 2-17 Confirming the configuration information

Step 5 Select an enterprise project, select the sentence in the **Agreement** area, confirm the configuration information, and click **Submit**.

Step 6 Return to the ECS list page and view the ECS creation progress.

When the ECS status changes to **Running**, the ECS is created.

----End

2.1.3.2 Using a Client to Access a ClickHouse Cluster

You can use SQL to access a cluster on an ECS. For details about how to install the client, see [Installing the Client](#).

 NOTE

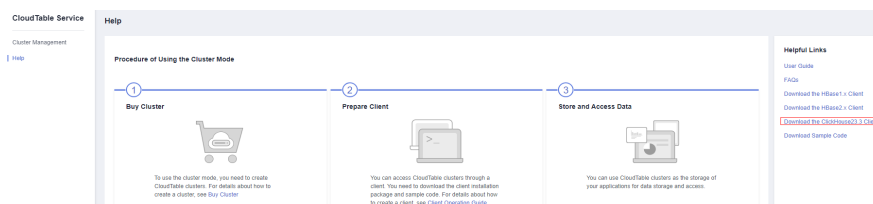
The VPC and security group of the cluster to be created must be the same as those of the ECS on the public network. Otherwise, the client cannot access the cluster.

Installing the Client

Step 1 Prepare a Linux ECS. For details, see [Preparing an ECS](#).

Step 2 Log in to the CloudTable management console, click **Help** in the navigation pane on the left, and click **Download Client** and **Client Verification File** on the right of the page to download the client installation package and client verification file.

Figure 2-18 Downloading the ClickHouse client



Step 3 Install the client and connect to the cluster.

1. Use the SSH login tool to remotely log in to the Linux ECS through the EIP. For details, see [Login Using an SSH Password](#) in the *Elastic Cloud Server User Guide*.

2. Go to the root directory of the SSH login tool.
`cd /`

3. Create a folder in the root directory.
`mkdir Folder name`

4. Go to the directory of the created folder.
`cd /Folder name/`

5. Place the client in the directory.

6. Decompress the client package.
`tar -xzf Client package name`

7. Decompress the client verification file to the same directory as the client.

a. Decompress the client verification file.

```
cd <Path for storing the client verification file >  
tar xzvf Client_sha256.tar.gz
```

b. Obtain the client verification code.

```
sha256sum ClickHouse_Client_23.3.tar.gz
```

c. Check the verification code in the client verification file and compare it with the client verification code. A match indicates no tampering, while a mismatch suggests tampering.

```
less ClickHouse_Client_23.3.tar.gz.sha256
```

8. Load the **.so** file.
`sh install.sh`

9. Go to the **bin** directory.
`cd bin/`

Grant the 700 permission to the directory.

```
chmod 700 clickhouse
```


10. Connect to the ClickHouse cluster. For details about the port, see [Table 2-16](#).

Use the following command to connect to a normal cluster.

```
./clickhouse client --host Private IP address of the cluster --port Port --user admin --password Password
```

For details about the security cluster connection command. For details, see [Connecting to a Security Cluster](#).

```
./clickhouse client --host Private IP address of the cluster --port port --user admin --password Password --secure --config-file /root/config.xml
```

 **NOTE**

- *Private IP address of the cluster*: cluster access address on the cluster details page. Replace it with the access address of the cluster you purchased.
- *Password*: the password set when you purchase the cluster. If there are special characters, use backslashes (\) to escape them. If the password is enclosed in single quotation marks ('), the special characters do not need to be escaped.

Table 2-16 Custom security rules

Direction	Action	Port/Range	Type	Destination/Source Address	Usage	
Outbound	Allow	All	IPv4/IPv6	0.0.0.0/0	Permit in the outbound direction	
Inbound	Allow	8123		Security group of the CloudTable ClickHouse cluster		ClickHouse HTTP port number
	Allow	9000				ClickHouse TCP port number
	Allow	8443				ClickHouse HTTPS port number
	Allow	9440				Secure TCP security port of ClickHouse

----End

2.1.3.3 Using Clickhouse Commands

This section describes common ClickHouse commands. After installing the ClickHouse client and loading environment variables, run the following commands to perform operations:

1. Connect to the ClickHouse cluster.

Use the following command to connect to a normal cluster.

```
./clickhouse client --host Private IP address of the cluster --port Port --user admin --password Password
```

For details about the security cluster connection command. For details, see [Connecting to a Security Cluster](#).

```
./clickhouse client --host Private IP address of the cluster --port port --user admin --password Password --secure --config-file /root/config.xml
```

2. Create a database.

```
create database demo;
```
 3. Use the database.

```
use demo;
```
 4. Check the database in use.

```
select currentDatabase();
```
 5. Create a table.

```
create table demo_t(uid Int32,name String,age UInt32,gender String)engine = TinyLog;
```
 6. View the table structure.


```
desc demo_t;
```
 7. Insert data.

```
insert into demo_t values(1,'Candy','23','M'),(2,'cici','33','F');
```
 8. View the table.

```
select * from demo_t;
```
 9. View the database and table.
 - Viewing the database

```
show databases;
```
 - Viewing the table

```
show tables;
```
 10. Delete the database and table.
 - Deleting the table

```
drop table demo_t;
```
-  **NOTE**
- Before deleting a data table, check whether the data table is in use to avoid unnecessary troubles.
 - After a data table is deleted, it can be restored within 24 hours. The restoration command is as follows:

```
set allow_experimental_undrop_table_query = 1;  
UNDROP TABLE Data table name;
```
 - Deleting the database

```
drop database demo;
```

2.1.4 Migrating ClickHouse Data

This section describes the data migration capability.

Application Scenarios

Once you have performed a scale-out in ClickHouse, a data migration is necessary.


Precautions

- Each data table is limited to a single task. A cluster can only execute one task at a time.
- The local table's database must be either atomic (the default) or ordinary, and the table itself should be from the Mergetree family of engine tables, which includes both replicated and non-replicated types. Materialized view charts are not compatible.
- The backup relationship for local tables mirrors that of clusters. In the context of shards, distributed tables are utilized.

- By default, the original table becomes read-only during data migration.
- Initially, data is moved to a temporary table. Subsequently, the original table is swapped with the table containing the migrated data. This transition, which may result in reading incorrect data, is completed within seconds.
- Should cluster issues arise, the data migration process can be paused. After addressing the reported error in the cluster, the migration task can resume.
- The source and redistribution nodes involved in data migration must share identical tables.
- Data migration is not supported on a single node system.

Creating a Data Migration Task

Step 1 Log in to the CloudTable console.

Step 2 Click  in the upper left corner to select a region.

Step 3 Click **Cluster Management** and click a cluster name to go to the cluster details page.

Step 4 In the navigation pane, choose **Data Migration**.

Table 2-17 Data migration parameters

Parameter	Description
Task ID/Name	ID or name of the new migration task.
Logical Cluster	Name of the selected logical cluster.
Source Nodes	Node where data is stored.
Distribution Nodes	Node where data is distributed.
Status/Progress	Status/Progress of data distribution. The task can be in the initializing, running, or completed state.
Created	Task creation time.
Start Time	Task start time.
Update Time	Task modification time.
Operation	<ul style="list-style-type: none">• Start: Start the task.• Edit: Edit task information.• Cancel: Cancel the task.• Details: View task details.• Delete: Delete the task.

Step 5 Click **New Task** in the upper left corner.

1. Enter a task name (starting with a letter).

2. Select a logical cluster.
3. Select the migration percentage.
4. Select the source node.
5. Select a redistribution node.
6. Select the data table to be migrated.

Step 6 Click **OK** to create the task.

Step 7 Click **Start** in the **Operation** column to start the created task.

----End

Modifying a Data Migration Task

Step 1 Log in to the CloudTable console.

Step 2 Click  in the upper left corner to select a region.

Step 3 Click **Cluster Management** and click a cluster name to go to the cluster details page.

Step 4 Choose **Data Migration**.


Step 5 Click **Edit** in the **Operation** column.

Step 6 After modifying the parameters, click **OK**.

----End

Viewing Migration Task Details

Step 1 Log in to the CloudTable console.

Step 2 Click  in the upper left corner to select a region.

Step 3 Click **Cluster Management** and click a cluster name to go to the cluster details page.

Step 4 Choose **Data Migration**.

Step 5 Click **Details** in the **Operation** column to access the task details page.

Step 6 View task information.

----End

Deleting a Migration Task

Step 1 Log in to the CloudTable console.

Step 2 Click  in the upper left corner to select a region.

Step 3 Click **Cluster Management** and click a cluster name to go to the cluster details page.

Step 4 Choose **Data Migration**.

Step 5 Click **Delete** in the **Operation** column. In the displayed dialog box, click **OK** to delete the task.


----End

2.1.5 Configuring Secure Channel Encryption for ClickHouse Clusters

You can enable secure channel encryption to encrypt data transmission. This section describes how to enable HTTPS for a ClickHouse cluster.

Enabling Channel Encryption

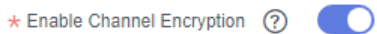
Step 1 Log in to the CloudTable console.

Step 2 Click  in the upper left corner to select a region.

Step 3 Click **Buy Cluster** in the upper right corner.

Step 4 Check whether **Enable Channel Encryption** (which is toggled on by default) is toggled on after completing other configurations.

Figure 2-19 Enabling channel encryption

 * Enable Channel Encryption  

NOTE

Disabling HTTPS will pose risks to enterprise services.

Step 5 Completing the parameter setting and click **Next**.

Step 6 Confirm the cluster specification order information on the displayed page and submit the order. After the cluster is created, go to its details page to view the channel status.

----End

Enabling Secure and Non-Secure Channels

Step 1 Log in to the CloudTable console.

Step 2 Click  in the upper left corner to select a region.

Step 3 Click **Buy Cluster** in the upper right corner.

Step 4 Completing the parameter setting and click **Next**.

Step 5 On the displayed page, confirm the cluster specification order information and click **Submit**. The cluster creation task is submitted.

Step 6 After the cluster is created, go to its details page and toggle **Enable Secure and Non-secure Channels** on. In the displayed dialog box, confirm the information and click **OK**.

 NOTE

- The cluster restarts after the secure channels are enabled on the cluster details page.
- The secure and non-secure channels cannot be disabled after being enabled concurrently.

----End

Connecting to a Security Cluster

Step 1 Click the name of the target security cluster to download the certificates on its details page.

Step 2 Connect to the client and specify the configuration file.

```
./clickhouse client --host Private IP address of the cluster --port port --user admin --password Password --secure --config-file /root/config.xml
```

Configure files.

```
<config>
  <secure>true</secure>
  <openSSL>
    <client>
      <caConfig>/etc/ssl/certificate.crt</caConfig>
    </client>
  </openSSL>
</config>
```

 NOTE

- `<caConfig>/etc/ssl/certificate.crt</caConfig>` indicates the path where certificates are stored.
- `root` indicates the path for storing the configuration file.
- Certificates can be downloaded only once per minute.

----End

HTTPS Connection

Step 1 Click the name of the target security cluster to download the certificates on its details page.

Step 2 Specify the path for storing the certificates.

Step 3 Execute the sample SQL statement through HTTPS.

```
echo 'select 1' | curl -H 'X-ClickHouse-User: user' -H 'X-ClickHouse-Key: password' --cacert /clickhouse/client/client/bin/certificate.crt 'https://host:port/?' --data-binary @-
```

- `select 1` indicates the executed SQL statement.
- `user` indicates the username.
- `password` indicates the password created during cluster creation.
- `/clickhouse/client/client/bin/certificate.crt` indicates the path for storing the certificates.
- `host` indicates the private IP address and `port` indicates the HTTPS port.

----End

JDBC Connection

```
public void run()
    throws InterruptedException {
    final ClickHouseProperties clickHouseProperties = new ClickHouseProperties();
    // There will be security risks if the password used for authentication is directly written into code. Encrypt
the password in the configuration file or environment variables for storage;
    // In this example, the password is stored in environment variables for identity authentication. Before
running the code in this example, configure environment variable CK_PASSWORD.
    String password = System.getenv("CK_PASSWORD");
    clickHouseProperties.setSslRootCertificate("/etc/ssl/certificate.crt");
    clickHouseProperties.setSsl(true);
    clickHouseProperties.setSslMode("strict");
    clickHouseProperties.setUser("test");
    clickHouseProperties.setPassword(password);
    clickHouseProperties.setSocketTimeout(2 * 3600 * 1000);
    final BalancedClickhouseDataSource dataSource = new
BalancedClickhouseDataSource("xxx.mycloudtable.com:8443/default?ssl=true", clickHouseProperties);
    try {
        final ClickHouseConnection conn = dataSource.getConnection();
        conn.createStatement().executeQuery("select now()");
    } catch (Throwable e) {
        e.printStackTrace();
    }
}
```

NOTE

`/etc/ssl/certificate.crt` in `clickHouseProperties.setSslRootCertificate("/etc/ssl/certificate.crt");` indicates the path for storing the certificates.

2.1.6 Application of ClickHouse Cold and Hot Data Separation

2.1.6.1 Overview

In big data scenarios, data storage and resource consumption increase rapidly with the growth of data. The need for data may vary in different time periods, therefore, data is managed as hot and cold data, improving data analysis performance and reducing service costs.

Background

In real-world data analysis, hot and cold data have distinct query frequencies and speed needs. Storing all data locally as historical data grows can lead to resource wastage. The cold and hot data separation feature allows for cost-effective storage by segregating data onto different media types. This not only accelerates query speeds and responsiveness for hot data but also cuts down on storage costs for cold data. Additionally, this feature offers adaptable configurations to align with the specific demands of various service scenarios. Hot and cold data is classified based on the data access frequency and update frequency.

- **Hot Data:** This type of data experiences frequent access and updates. It is likely to be needed in future operations and demands swift response times due to its active nature.
- **Cold Data:** In contrast, cold data is characterized by its static state; it is rarely updated or accessed and has minimal requirements for response speed.

You can define cold and hot management tables to switch cold data that meets the specified rules to OBS for storage. Cold and hot data can be automatically determined and migrated by partition.

Principles

- **Table Creation:** Create a data table with a cold-hot separation policy. The `storage_policy` must be set to `hot_to_cold`.
- **Data Writing:** Upon importing new data, ClickHouse generates a new part for each write operation, ensuring that new data is allocated to the cold part seamlessly, thus maintaining concurrent cold and hot data storage.
- **Data Separation:** Data initially resides in hot storage and transitions to cold storage based on capacity or elapsed time. ClickHouse's separation is part-based; when the threshold is reached, qualifying parts are transferred to OBS, and corresponding local data is deleted. Newly imported data, forming new parts, will also migrate to OBS upon reaching the set threshold.
- **Data Querying:** Data queries initiated by users are directed to the root directories of corresponding buckets according to the table's storage policy. ClickHouse retrieves necessary data from various table parts to the local system for processing.

2.1.6.2 Using ClickHouse Cold and Hot Data Separation

Enable the cold and hot separation feature on the CloudTable management plane.

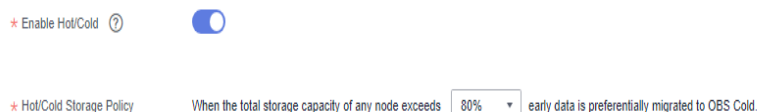
Procedure

- Step 1** Log in to the CloudTable console.
- Step 2** Select a region in the upper left corner.
- Step 3** In the navigation pane on the left, choose **Cluster Management**.
- Step 4** Click **Buy Cluster** in the upper right corner.
- Step 5** On the **Create Cluster** page, select **ClickHouse** on the cluster purchase page and enable the cold and hot data separation feature. After the cluster is created, the cold and hot separation feature will be enabled.

NOTE

- For details about how to create a cluster, see [Creating a ClickHouse Cluster](#).
- If the cold and hot data separation feature is not enabled during cluster creation, you can enable it on the cluster details page or in the cluster list.
- If the used capacity of a node surpasses the threshold defined in the cold and hot storage policy, data will be automatically transferred to the cold storage volume. For instance, as in [figure 1](#), the hot storage volume usage reaches 80%, data migration to the cold storage is initiated automatically.

Figure 2-20 Enabling cold and hot data separation



----End

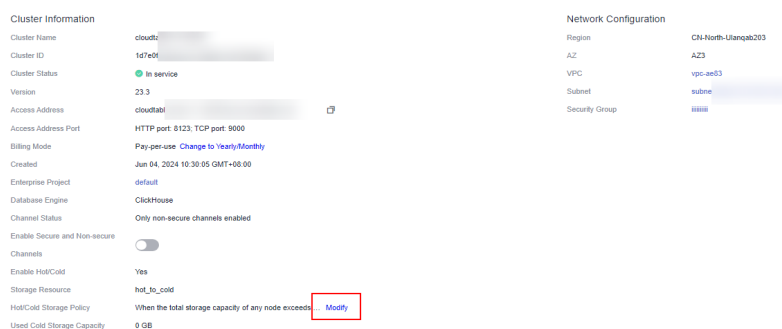
Modifying the Cold and Hot Data Separation Policy

- Step 1** Log in to the CloudTable console.
- Step 2** Select a region in the upper left corner.
- Step 3** In the navigation pane on the left, choose **Cluster Management**.
- Step 4** Click a ClickHouse cluster name to go to the cluster details page.
- Step 5** Choose **Cluster Information > Hot/Cold Storage Policy > Modify**. On the displayed page, modify the policy and click **OK**. The policy is modified.

NOTE

You can also create a hot/cold storage policy when creating a cluster. Once a policy is created, it cannot be deleted.

Figure 2-21 Modifying a hot/cold storage policy



----End

2.1.6.3 Commands Related to ClickHouse Cold and Hot Data Separation

This section describes the commands related to CloudTable ClickHouse cold and hot data separation. For details about how to use basic ClickHouse commands, see [Using Clickhouse Commands](#).

- Create a cold-hot separation table.

```
CREATE TABLE hot_cold_table
(
  `f1` String,
  `f2` String,
  `f3` Int64,
  `f4` Float64,
  `date` Date
)
ENGINE = MergeTree()
PARTITION BY date
ORDER BY f1
SETTINGS storage_policy = 'hot_to_cold';
```

- Set the cold data storage duration.

```
CREATE TABLE hot_cold_table
(
  `f1` String,
  `f2` String,
  `f3` Int64,
  `f4` Float64,
  `date` Date
)
```

```
ENGINE = MergeTree()
PARTITION BY date
ORDER BY f1
TTL date + INTERVAL 90 DAY TO DISK 'cold_disk'
SETTINGS storage_policy = 'hot_to_cold';
```

 **NOTE**

Parameters:

- storage_policy: Specifies the storage policy. The value is **hot_to_cold**. Once a storage policy is specified, it cannot be modified later.
- Update the TTL.
 - Modify metadata:
ALTER TABLE hot_cold_table MODIFY TTL toDate(_cw_raw_time/ 1000) +toIntervalDay(29) TO VOLUME 'volum1_cold' SETTINGS materialize_ttl_after_modify=0;
 - **Manually moving data:** You can run the **alter table hot_cold_table move partition '*partition_name*' to volume '*volume_name*'** command to move a specified partition to a specified volume or disk.

2.2 ClickHouse Monitoring Clusters

2.2.1 ClickHouse Cluster Supported Metrics

Description

Monitoring is critical to ensure CloudTable reliability, availability, and performance. You can monitor the running status of CloudTable servers. This section describes the metrics that can be monitored by CES as well as their namespaces and dimensions.

Namespace

SYS.CloudTable

Monitoring metrics

Table 2-18 ClickHouse monitoring metric description

Metric	Name	Description	Minimum Unit	Monitoring Period
cpu_usage	CPU Utilization Rate	CPU utilization rate of each node.	%	60
memory_usage	Memory Usage	Memory usage of each node.	Byte	60
memory_usage_ratio	Memory Utilization Rate	Memory utilization rate of each node.	%	60

Metric	Name	Description	Minimum Unit	Monitoring Period
disk_usage	Disk Usage	Disk space used by each node.	Byte	60
disk_usage_ratio	Disk Utilization Rate	Disk utilization rate of each node.	%	60
network_throughput_inbound_rate	Inbound Throughput	Inbound data volume over network of each node per second.	Byte/s	60
network_throughput_outgoing_rate	Outbound Throughput	Outbound data volume over network of each node per second.	Byte/s	60
clickhouse_process_survival	ClickHouse Process Survival Status	Heartbeat check of the ClickHouse process.	-	60
number_of_mutations	Mutations	Number of running modification tasks on each node.	Count	60
number_of_run_merges	Running Merges	Number of running merge tasks.	Count	60
number_of_merges_thread	Merge Threads	Number of threads being merged.	Count	60
number_of_zookeeper_requests	ZooKeeper Requests	Number of ZooKeeper requests on the node at the current time.	Count	60
number_of_zk_session	ZooKeeper Sessions	Number of connected ZooKeeper sessions on the node at the current time.	Count	60

Metric	Name	Description	Minimum Unit	Monitoring Period
number_of_zookeeper_watches	ZooKeeper Watches	Number of connected ZooKeeper watches on the node at the current time.	Count	60
number_of_zookeeper_nodes	ZooKeeper Nodes	Number of temporary nodes created in ZooKeeper.	Count	60
number_of_parts	Parts	Number of parts.	Count	60
clickhouse_zookeeper_disconnection	ZooKeeper Connection Status	ZooKeeper connection status.	-	60
number_of_threads_holding_read_locks	Threads Holding the Read Lock	Number of threads that hold read locks.	Count	60
number_of_threads_waiting_to_be_read	Threads Waiting to Be Read	Number of threads waiting to be read.	Count	60
number_of_threads_waiting_to_be_written	Number of Threads Waiting to Be Written	Number of threads waiting to be written.	Count	60
number_of_threads_holding_write_locks	Threads Holding the Write Lock	Number of threads that hold write locks.	Count	60
number_of_active_tasks_background_processing_pool	Active Tasks in Background Processing Pool	Number of active tasks in the background processing pool.	Count	60
maximum_number_of_active_blocks	Max. Active Blocks	Maximum number of active blocks of the ClickHouse instance.	Count	60
number_of_consistency_checkdata_blocks	Data Blocks for Consistency Check	Number of data blocks for consistency check on a ClickHouse instance.	Count	60

Metric	Name	Description	Minimum Unit	Monitoring Period
extracting_the_number_of_data_chunks_from_the_replica	Data Blocks Being Extracted from a Replica	Number of data blocks that are being extracted from a replica by a ClickHouse instance.	Count	60
number_of_data_chunks_being_sent_to_the_replica	Data Blocks Being Sent to a Replica	Number of data blocks on a ClickHouse instance that are being sent to a replica.	Count	60
number_of_readonly_replicas	Read-Only Replicas	Number of read-only replicas of a ClickHouse instance.	Count	60
number_of_inserted_into_all_tables	Inserted Rows	Number of rows inserted into all tables in a ClickHouse instance.	Count	60
number_of_open_read_files	Opened Read Files	Number of opened read files on a ClickHouse instance.	Count	60
number_of_open_write_files	Opened Write Files	Number of opened write files on a ClickHouse instance.	Count	60
total_file_opens	Opened Files	Number of opened files on a ClickHouse instance.	Count	60
number_of_read_file_descriptors	Read File Descriptors	Number of file descriptors that have been read.	Count	60
number_of_written_file_descriptors	Write File Descriptors	Number of file descriptors that have been written.	Count	60

Metric	Name	Description	Minimum Unit	Monitoring Period
instance_process_running_duration	Instance Process Running Duration	Running duration of the process on a ClickHouse instance.	s	60
number_of_threads_running_in_global_thread_pool	Threads of Running Tasks in Global Thread Pool	Number of threads of running tasks in the global thread pool of a ClickHouse instance.	Count	60
number_of_threads_in_global_thread_pool	Threads in Global Thread Pool	Number of threads in the global thread pool of a ClickHouse instance.	Count	60
number_of_threads_in_local_thread_pool	Threads in Local Thread Pool	Number of threads in the local thread pool of a ClickHouse instance.	Count	60
number_of_threads_waiting_locked	Threads to Lock in Context	Number of threads waiting to be locked in context on a ClickHouse instance.	Count	60
number_of_threads_running_in_local_thread_pool	Threads of Running Tasks in Local Thread Pool	Number of threads of running tasks in the local thread pool of a ClickHouse instance.	Count	60
number_of_tcp_connections	TCP Connections	Number of TCP connections.	Count	60
number_of_http_connections	HTTP Connections	Number of HTTP connections.	Count	60
number_of_databases	Databases	Number of databases on each node.	Count	60

Metric	Name	Description	Minimum Unit	Monitoring Period
number_of_tables	Tables	Number of local tables on each node.	Count	60
write_size_per_second	Write Size per Second	Amount of data written to each node per unit time.	Byte	60
number_of_running_queries	Running Queries	Number of running queries on each node.	Count	60
number_of_query_threads	Query Threads	Number of running queries on each node.	Count	60
number_of_distributed_ddls	Distributed DDLs	Number of distributed DDL statements.	Count	60
number_of_distributed_table_files	Distributed Tables	Number of tables to which data is to be inserted.	Count	60
number_of_concurrency	Concurrency of Each ClickHouse Node	Concurrency of each ClickHouse node.	Count	60
zookeeper_quota_percentage	ZooKeeper Quantity Percentage	Quota of the ZooKeeper nodes used by ClickHouse.	%	60
zookeeper_process_is_alive	ZooKeeper Process Survival Status	ZooKeeper heartbeat check. 1 indicates that the ZooKeeper heartbeat is normal, and 0 indicates that it is abnormal.	-	60
number_of_global_sessions	Global Sessions	Number of global sessions.	Count	60
number_of_rejected_connections	Refused Connections	Number of connections refused by Zookeeper.	Count	60

Metric	Name	Description	Minimum Unit	Monitoring Period
number_of_request_submission_queues	Requests to Submit in a Queue	Number of requests submitted in a queue.	Count	60
waiting_time_of_the_preprocessing_queue	Wait Time of Request in Preprocessing Queue	Time that a request in the preprocessing queue waits to be processed.	s	60
number_of_zk_watches	ZooKeeper Watches	Number of ZooKeeper watches at the current time.	Count	60
zk_heapmemory_usage	Heap Memory Used	Heap memory size used by ZooKeeper process.	Byte	60
zk_directmemory_usage	Used Direct Memory	Direct memory size used by ZooKeeper process.	Byte	60
zk_heapmemory_usage_ratio	Heap Memory Usage	Heap memory usage of the ZooKeeper service.	%	60
zk_directmemory_usage_ratio	Direct Memory Usage	Direct memory usage of the ZooKeeper service.	%	60
cluster_number_of_sent_packets	ZooKeeper Sent Packets	Number of packets sent by a ZooKeeper node.	Count	60
cluster_number_of_received_packets	ZooKeeper Received Packets	Number of packets received by a ZooKeeper node.	Count	60
cluster_preprocessing_time	Preprocessing Duration	Preprocessing time.	s	60
rows_written_per_second	Rows Written per Second	Number of rows written to each node per unit time.	Count	60
number_of_failed_queries	Failed Queries	Number of failed query statements.	Count	60

Metric	Name	Description	Minimum Unit	Monitoring Period
number_of_failed_insert_queries	Failed Insert Queries	Number of failed insert statements.	Count	60
number_of_failed_select_query_messages	Failed Select Queries	Number of failed select statements.	Count	60
number_of_delayed_inserts	Delayed Inserts	Number of delayed insertions on each node.	Count	60
number_of_rows	Rows	Number of mergetree engine fields on each node.	Count	60
disk_throughput_read_rate	Disk Read Throughput	Disk read throughput.	Byte	60
disk_throughput_write_rate	Disk Write Throughput	Disk write throughput.	Byte	60
cmdForUsedStorageRate	Used Storage Rate	Ratio of the used storage space to the total storage space in the cluster	%	60


Dimension

Key	Value
cluster_id	CloudTable cluster ID.
instance_name	Name of a CloudTable cluster node.

2.2.2 Setting Alarm Rules

You can set CloudTable ClickHouse alarm rules to customize the monitored objects and notification policies. Then, you can learn about ClickHouse running status in a timely manner. The ClickHouse alarm rules include alarm rule name, instance, metric, threshold, monitoring interval and whether to send notification. This section describes how to set alarm rules.

Creating an Alarm Rule for a ClickHouse Cluster

- Step 1** Log in to the CloudTable console.
- Step 2** In the navigation tree, click **Cluster Management**.
- Step 3** In the upper right corner of the cluster list, enter the name of a cluster in the search box and click .
- Step 4** In the cluster list, locate the row where the target cluster resides and click **More > View Metric** in the **Operation** column. The CES console is displayed.

The status of the cluster whose monitoring information you want to view must be **In service**.

Figure 2-22 Viewing the monitoring metrics

Compute Units		
ID	Permanent Data Storage	Operation
<input type="checkbox"/> zookeeper-2-2	--	View Metric Create Alarm Rule Configure Storage
<input type="checkbox"/> zookeeper-1-1	--	View Metric Create Alarm Rule Configure Storage
<input type="checkbox"/> zookeeper-3-3	--	View Metric Create Alarm Rule Configure Storage
<input type="checkbox"/> server-2-2	--	View Metric Create Alarm Rule Configure Storage
<input type="checkbox"/> server-1-1	--	View Metric Create Alarm Rule Configure Storage
<input type="checkbox"/> server-5-5	--	View Metric Create Alarm Rule Configure Storage

- Step 5** Click **Create Alarm Rule** to add an alarm rule.
 1. Enter an alarm rule name and description.

Figure 2-23 Configuring an alarm rule name

Name

Description

0/256

2. Specify the monitoring scope.

Figure 2-24 Specifying the monitored objects

- * Alarm Type Metric
- * Resource Type CloudTable Service
- * Dimension Compute Units
- * Monitoring Scope Specific resources
- * Monitored Objects server-2-2

Table 2-19 Parameters

Parameter	Description	Example Value
Alarm Type	Default instance or the selected cluster.	Metric
Resource Type	Name of the cloud service resource for which the alarm rule is configured.	CloudTable Service
Dimension	Metric dimension of the alarm rule. Only the Compute Units option is supported.	Computing Units
Monitoring Scope	Clusters to be monitored.	Specific resources
Monitored Objects	Nodes to be monitored.	-

3. Configure the alarm policies as prompted.

Figure 2-25 Configuring the alarm policy

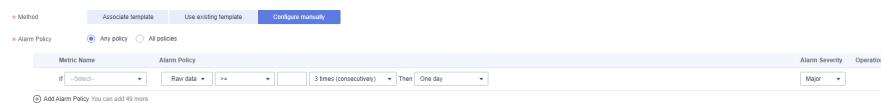


Table 2-20 Parameters for configuring an alarm policy

Parameter	Description	Example Value
Method	Select Use existing template or Configure manually as required. <ul style="list-style-type: none"> - If no alarm template is available, set Method to Configure manually and configure related parameters to create an alarm rule. - If you have available alarm rule templates, set Method to Use existing template or Associate template, so that you can use a template to quickly create alarm rules. 	Configure manually
Template	This parameter is valid only when Use existing template is selected. Select the template to be imported. If no alarm template is available, click Create Custom Template to create one that meets your requirements.	Use existing template

Parameter	Description	Example Value
Alarm Policy	This parameter is valid only when Configure manually is selected. For details about the monitoring metrics supported by CloudTable ClickHouse, see ClickHouse Cluster Supported Metrics .	-
Alarm Severity	Severity of an alarm. Valid values are Critical, Major, Minor, and Informational .	Major

- Configure the alarm notification parameters as prompted.

Figure 2-26 Alarm notification

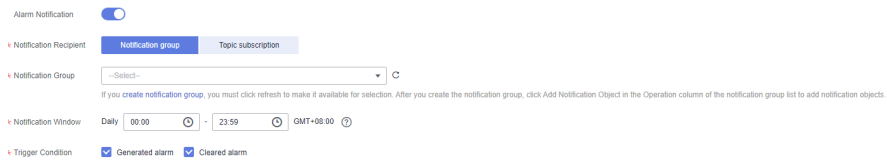


Table 2-21 Alarm notification parameters

Parameter	Description	Configuration
Alarm Notification	Whether to notify users when alarms are triggered. Notifications can be sent by emails, text messages, or HTTP/HTTPS requests. If this function is enabled (recommended), notifications will be sent. Otherwise, no notification is sent.	Enabled
Notification Recipient	Notification group or topic subscription.	-

Parameter	Description	Configuration
Notification Group/Object	<ul style="list-style-type: none"> Notification group that needs to send alarm notifications. If you enable Send Notification, you need to select an existing notification group. If you have not created a notification group, click Create Notification Group to access the CES page to create one. Topic to which the alarm notification is to be sent. If you enable Notification, you need to select an existing topic. If you have not created a topic, click Create an SMN Topic to go to the SMN management console to create a topic. 	-
Validity Period	<p>CES sends notifications only within the validity period specified in the alarm rule.</p> <p>If Validity Period is set to 00:00-23:59, CES send notifications only within 00:00-23:59.</p>	00:00 ~ 23:59
Trigger Condition	Condition for triggering an alarm notification.	Select all options.

5. Configure advanced settings.

Figure 2-27 Configuring advanced settings

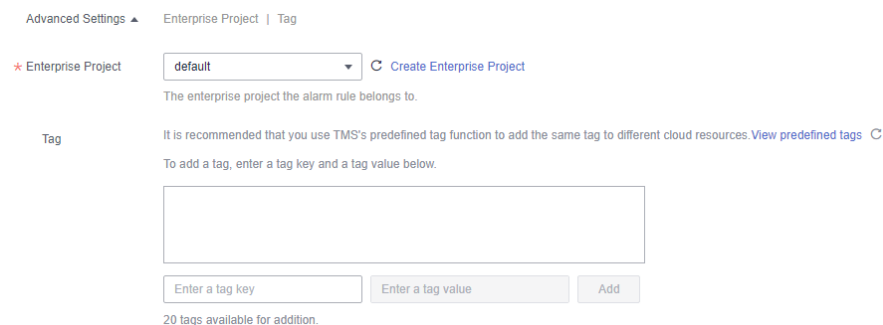


Table 2-22 Advanced configuration parameters

Parameter	Description
Enterprise Project	You can group related cloud resources (for example, resources used for the same purpose) and manage them by enterprise project.
Tag	It is recommended that you use TMS's predefined tag function to add the same tag to different cloud resources.

Step 6 After the configuration is complete, click **Next**. After the alarm rule is created, if the metric data reaches the specified threshold, CES will immediately inform you that an exception has occurred.

 **NOTE**

If you want to set an alarm rule for a single metric, go to the page for viewing monitoring information.

- View the monitoring metrics and click the metric graph.
- Click + in the upper right corner.
- Set the rule on the page for setting alarm rules.

----End

2.2.3 Viewing Monitoring Information of a Cluster

Scenario

CES on the public cloud monitors the running status of CloudTable ClickHouse clusters. You can view the monitoring metrics of ClickHouse on the management console. According to the monitoring information, you can quickly learn about cluster health status and key system information.


Monitoring Function

The cluster monitoring function consists of compute node and ZooKeeper metric monitoring. You can adjust the time range of monitoring data in each module to view historical data at different time. You can also adjust the time granularity of monitoring data to view data in different dimensions. If you want to view monitoring data in real time, enable automatic refresh. The platform can automatically refresh monitoring graphs at intervals of 1 hour, 3 hours, 12 hours, 24 hours, and 7 days.

If you are interested in a metric graph, you can click the zoom-in button to view the graph or export the graph.

Checking ClickHouse Cluster Monitoring Information


Step 1 Log in to the CloudTable console.

Step 2 Click  in the upper left corner to select a region.

Step 3 In the navigation tree, click **Cluster Management**.

Step 4 In the cluster list, locate the row where the target cluster resides, click **View Metric** in the **Operation** column. The CES console is displayed.


The status of the cluster to be viewed must be **In service**.

Step 5 On the **Cloud Service Monitoring** page, click  on the left of the cluster ID to expand the compute unit list, and select the corresponding node to view the monitoring information.

Step 6 Set the metrics to be viewed if there are too many metrics on the monitoring page.

1. If there are too many metrics, delete them on the **Select Metric** page.
2. If the metrics displayed on the page do not contain the desired metrics, add the metrics on the **Select Metric** page.
3. Select at least one metric. You can drag a selected metric and drop it to a desired location to sort the metrics.

 **NOTE**

- If you want to view monitoring data in real time, enable automatic refresh. The platform can automatically refresh monitoring graphs at intervals of 1 hour, 3 hours, 12 hours, 24 hours, and 7 days.
- If you want to zoom in on a single metric view, click  in the upper right corner of the metric view to view its details.

----End

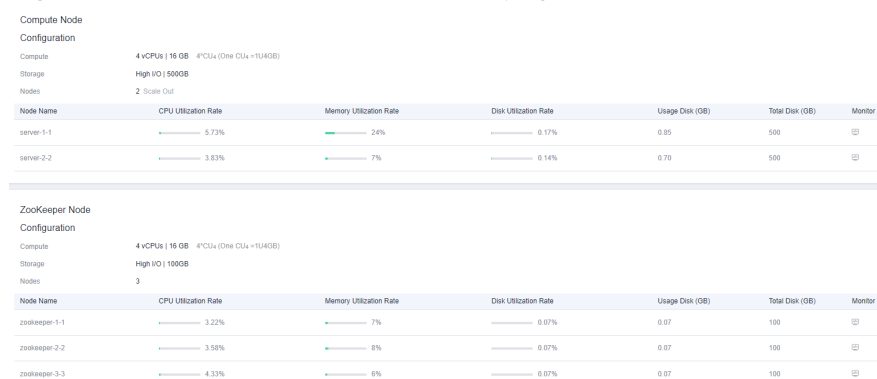
2.2.4 Checking ClickHouse Monitoring Metrics


The cluster details page displays the CPU, memory, disk utilization rate, as well as the disk usage and total disk capacity of each ClickHouse node.

Accessing the ClickHouse Cluster Details Page

- Step 1** Log in to the CloudTable console.
- Step 2** Click the cluster list page and select a ClickHouse cluster.
- Step 3** View the monitoring metrics on the cluster details page.

Figure 2-28 ClickHouse cluster details page



Step 4 Click  on the details page to go to the CES console and view detailed monitoring metrics of the ClickHouse cluster. For details, see [Viewing Monitoring Information of a Cluster](#).

----End

Monitoring Metrics

Table 2-23 Monitoring metrics

Metric	Description
CPU Utilization Rate	CPU utilization rate of the ClickHouse cluster
Memory Utilization Rate	Memory utilization rate of the ClickHouse cluster
Disk Utilization Rate	Data utilization rate of the ClickHouse cluster

3 Permissions Management

3.1 Creating a User and Granting Permissions

This chapter describes [IAM](#) fine-grained permissions management for your CloudTable. With IAM, you can:

- Create IAM users for employees based on your enterprise's organizational structure. Each IAM user will have their own security credentials for accessing CloudTable resources.
- Grant only the permissions required for users to perform a task.
- Entrust a HUAWEI CLOUD account or service to perform professional and efficient O&M on your CloudTable resources.

If your HUAWEI CLOUD account does not require individual IAM users, skip this chapter.

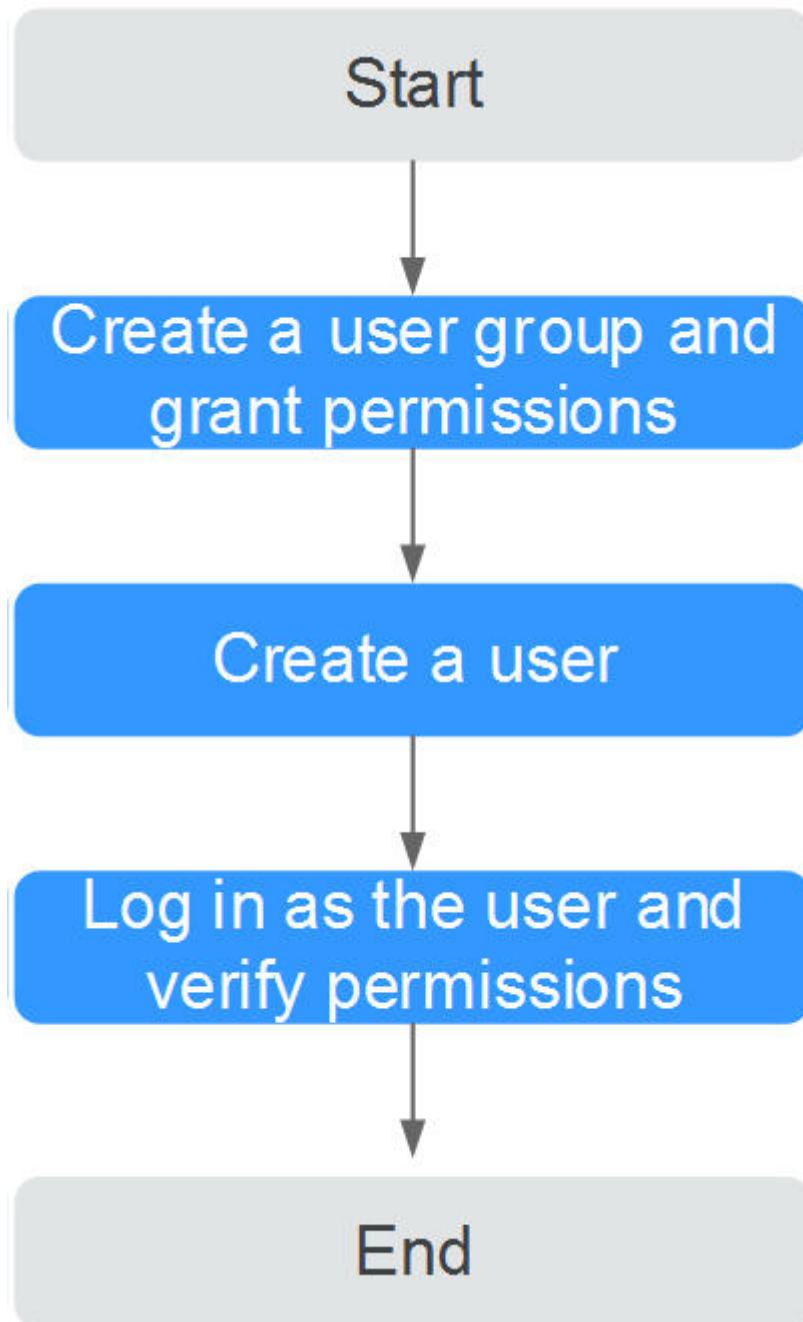
This section describes the procedure for granting permissions (see [Figure 3-1](#)).

Prerequisites

- CloudTable does not support fine-grained policies. It supports only the system-defined role **cloudtable Administrator**. The default system role **cloudtable Administrator** of CloudTable applies only to accounts that have not enabled the enterprise project function. If the enterprise project function has been enabled, use the IAM custom policy to grant permissions to users.
- Learn about the permissions (see CloudTable System-Defined Permissions) supported by CloudTable and choose policies or roles according to your requirements. For the system policies of other services, see [Permissions Policies](#).

Process Flow

Figure 3-1 Process for granting CloudTable permissions



1. **Create a user group and assign permissions** to it.
Log in to the IAM console, create a user group, and assign the **cloudtable Administrator** permission to it.
2. **Create a user and add it to a user group.**
Create a user on the IAM console and add the user to the group created in Step 1.

3. **Log in** and verify permissions.

Log in to the console by using the user created, and verify that the user has the granted permissions.

- Choose **Service List > CloudTable Service**. Then click **Buy Cluster** on the CloudTable console. If the cluster can be created (for example, there is only the **cloudtable Administrator** permission), the **cloudtable Administrator** permission has already taken effect
- Choose any other service in the **Service List** (for example, there is only the **cloudtable Administrator** policy). If a message appears indicating insufficient permissions to access the service, the **cloudtable Administrator** policy has already taken effect.

A Change History

Release Date	Description
2024-6-20	This issue is the second official release. Added the following sections: <ul style="list-style-type: none">• HBase Security Channel Encryption• HBase Thrift Server.
2024-04-30	This issue is the first official release.