Tencent Kubernetes Engine

Cloud Native Service Guide

Product Documentation

Contents

Cloud Native Service Guide
Cloud Service for etcd
Overview

Last updated：2025-03-19 18:30:11

Product Introduction

Cloud Service for etcd is an etcd management solution based on open-source etcd and is tailored to cloud-native service scenarios. It is provided by the TKE team and fully compatible with distributed storage capabilities of open-source etcd. It provides highly stable, observable, Ops-free cloud-native etcd services.

> **Notes:**
> The beta testing of Cloud Service for etcd has ended on June 8, 2022. The service is now officially available with commercial billing. For pricing details, see Purchase Guide.

Use Cases

etcd is a distributed and highly reliable key-value store that can tolerate failures of some nodes in a cluster. It can provide services with more than half of the nodes being alive. It is mainly used for metadata storage, service discovery, distributed election, and other purposes. Features such as release and subscription can be easily implemented based on the Watch mechanism provided by etcd.

Reasons for Using Cloud Service for etcd

- Users lack sufficient understanding of etcd, making it difficult to use etcd.
- Users lack usage experience in Ops of self-built etcd clusters. They cannot quickly locate issues during the use of them.
- Self-built etcd clusters often requires maintaining a monitoring and alarm system, as well as a backup and recovery mechanism, which increases the users' Ops burden.

The TKE team currently maintains tens of thousands of Kubernetes clusters online and uses thousands of etcd clusters for storage at the backend. The team has accumulated a wealth of practical experience in ensuring the stable operation of etcd, which can help users reduce the Ops burden of etcd and focus more on business development.

Product Features

- One-click etcd cluster deployment: Support high availability deployment, HTTPS access, and automatic data compression for clusters.

- Integration with cloud-native monitoring capabilities: Provide a comprehensive monitoring and alarm mechanism.
- Routine Ops management: Support backup and recovery, node scaling, version upgrade, and other features.

Product Strengths

Complete Cluster Protection Mechanism

- Secure data access: Support enabling HTTPS two-way authentication for more secure data access. Support restricting access sources through security groups.
- Data backup and recovery: Support enabling cluster backup policies to regularly back up data to Tencent Cloud Object Storage (COS), and support recovering faulty clusters from backup data.
- High availability deployment: Support deploying cluster nodes to different AZs. If 3 different AZs are used, failures of a single AZ do not affect the normal service of the cluster. Support quick self-healing of faulty nodes to minimize cluster downtime.
- Reliable version update mechanism: Carry out comprehensive internal testing and large-scale verification before a new version is released. Conduct fault experiment on Chaotic Fault Generator (CFG) to ensure the stability of a new version.

Powerful Monitoring and Alarm Capabilities

Cloud Service for etcd can be seamlessly integrated with TencentCloud Managed Service for Prometheus (TMP) to provide default performance and availability metrics you want to monitor. You can also aggregate the required monitoring metrics and panels to better monitor the etcd cluster status.

Full-Process Ops Service

Users do not need to worry about the installation, deployment, version updates, and failure handling of Cloud Service for etcd.

Product Pricing

The beta testing of Cloud Service for etcd has ended on June 8, 2022. The service is now officially available with commercial billing. For pricing details, see Purchase Guide. Resources of other products such as Cloud Load Balancer (CLB) and COS will be created automatically during the use of Cloud Service for etcd. For billing details of these products, see the corresponding billing overview, such as CLB Purchase Overview and COS Billing Overview.

Cloud Resource	Applicable Scenarios
CLB instance	Private network access of clusters
COS instance	Data backup

Related Services

TMP is a monitoring and alarm solution for cloud-native service scenarios. It fully supports the monitoring capabilities of open-source Prometheus and provides lightweight, stable, and highly available cloud-native Prometheus monitoring services for users.

Purchase Guide

Last updated: 2025-03-19 18:30:11

Billing Mode

Cloud Service for etcd supports the following 2 billing modes:

Billing Mode	Payment Method	Billing Cycle	Applicable Scenario
Monthly subscription	Prepayment. Fees are paid when an instance is created.	Month	Apply to the businesses of stable scale with long-term resource requirements. The fees are lower than that of the pay-as-you-go mode. The longer the purchase duration, the higher the discount.
Pay-as-you-go	Postpayied. Resources are applied on-demand and are charged based on the actual usage upon settlement.	Hour	Apply to short-term businesses with on-demand usage needs. Instances can be released immediately after use to save costs.

Instance Specifications

Specifications of Cloud Service for etcd instances are defined in dimensions of CPU, memory, and number of replicas.

Pay-as-You-Go

If the usage duration is less than 1 hour, fees of 1 hour are charged.

Node Specifications	Number of Nodes	USD/Hour
1C2G	3	0.33 USD
	5	0.41 USD
2C4G	3	0.49 USD
	5	0.65 USD
4C8G	3	0.73 USD

	5	1.05 USD
8C16G	3	1.21 USD
	5	1.85 USD
8C32G	3	1.65 USD
	5	2.58 USD
16C64G	3	3.04 USD
	5	4.91 USD
32C128G	3	5.84 USD
	5	9.56 USD

Monthly Subscription

Node Specifications	Number of Nodes	USD/Month
1C2G	3	217.06 USD
	5	269.38 USD
2C4G	3	321.7 USD
	5	426.34 USD
4C8G	3	478.66 USD
	5	687.94 USD
8C16G	3	792.57 USD
	5	1,211.13 USD
8C32G	3	1,078.68 USD
	5	1,687.97 USD
16C64G	3	1,992.61 USD
	5	3,211.19 USD
32C128G	3	3,820.48 USD
	5	6,257.65 USD

Payment Overdue

Notifications of Payment Overdue and Resource Expiration

The platform will send notifications via SMS messages, emails, and Message Center before instance resources expire or overdue payment occurs.

Recycling Mechanism

- If pay-as-you-go is adopted, instances can still be used within 24 hours after overdue payment occurs. **After 24 hours**, instances will be isolated, and billing will stop. Isolated instance cannot be accessed. Users can top up to recover instances within 7 days after the isolation. If instances are not recovered **7 days later**, they will be released, and resources will be recycled, which cannot be restored.
- If monthly subscription is adopted, instances can still be used within 3 days after overdue payment occurs. **After 3 days**, instances will be isolated, and billing will stop. Isolated instance cannot be accessed. Users can top up to recover instances within 7 days after the isolation. If instances are not recovered **7 days later**, they will be released, and resources will be recycled, which cannot be restored.

Refund Instructions for Monthly Subscription Instances

Last updated：2025-03-19 18:30:11

To facilitate your use of Cloud Service for etcd clusters, self-service return and refund are supported if you are in any way not satisfied with monthly subscription clusters after purchase. Each entity is entitled to an unconditional return of one etcd cluster within five days, and the valid amount paid will be refunded to your Tencent Cloud account. You can also proceed with a standard self-service return. After the fees for resources you have already used are deducted, the refund amount is returned to your Tencent Cloud account based on the proportion of cash and free credits used for the purchase. All these operations can be performed in the console.

Five-Day Unconditional Self-Service Return

Cloud Service for etcd follows Tencent Cloud refund rules stated in Refunds. If you are not satisfied after purchase, five-day unconditional self-service return is supported. The specific rules are as follows:

- Each entity is entitled to an unconditional return of one monthly subscription etcd cluster **within five days (inclusive) from the date of purchase** by default.
- For orders that meet the five-day unconditional return requirements, the refund amount is the total amount paid at the time of purchase, including the amounts paid by cash, transferred , and free credits.
- If you want to return a cluster purchased during a campaign but the return rules conflict with the campaign rules, the campaign rules shall prevail.
- In case of suspected abnormal or malicious returns, Tencent Cloud reserves the right to reject your return request.

> **Note:**
> - Vouchers are not refundable.
> - The refund amount will be returned to the Tencent Cloud account.

Standard Self-Service Return

The refund amount will be returned to your Tencent Cloud account in the proportion of the cash and free credits used at the time of purchase.

Standard Self-Service Return Rules

Refund amount = amount of currently effective orders + amount of unstarted orders - amount of fees for used resources.

- **Amount of currently effective orders**: The amount paid for currently effective orders, excluding discounts and vouchers.

- **Amount of unstarted orders**: The amount paid for future orders, excluding vouchers.
- **Amount of fees for used resources**: If resources are used for complete months by the day the refund is initiated, deduct the fees based on the monthly subscription prices with corresponding discounts. Otherwise, deduct the fees based on the monthly subscription prices by the usage duration (accurate to seconds).

Self-Service Return Instructions for Monthly Subscription Instances

- After a self-service return is performed for a monthly subscription instance, **disks of the instance are recycled along with the instance**, and data stored on these disks will not be retained.
- After a self-service return is performed for a monthly subscription instance, services running on the instance are completely interrupted, relevant resources are released, which **cannot be restored by renewal**.
- If data backup is enabled for a monthly subscription instance, **you can choose whether to release the backup resources (snapshots) in the COS bucket when the instance is terminated**.
- If public network access is enabled for a monthly subscription instance, **CLB resources are release when the instance is terminated**.
- In case of suspected abnormal or malicious returns, Tencent Cloud reserves the right to reject your return request.

Version Maintenance

Version Maintenance Mechanism

Last updated：2025-03-19 18:30:11

This document introduces the versions supported by Cloud Service for etcd of Tencent Cloud and etcd of the community edition. This can help you choose the version that can best meet your business needs when using Cloud Service for etcd.

Version Definition

- **Major version**: etcd versions are in the format of x.y.z. x indicates the major version, y indicates the minor version, and z indicates the patch version. They follow the requirements of Semantic Versioning. For example, the current versions are 3.5.4 and 3.4.13.
- **Minor version**: The `tke.x` suffix indicates the minor version corresponding to the current major version. For example, the current version is v3.4.13-tke.5.

> **Note:**
> If a version has no suffix, it is an open-source version in the community.

Version Maintenance Strategy

Cluster Creation

The platform supports creating clusters of the latest 2 major versions of etcd, such as **3.5.4** and **3.4.13**. When a new version is released and it supports upgrades, the earlier version no longer supports cluster creation. For example, if the version **3.6.x** is released and it supports upgrade from **3.5.4** to **3.6.x**, **3.4.13** will no longer support cluster creation. If version **3.6.x** is released but does not support upgrades, **3.4.13** still supports cluster creation.

> **Note:**
> Only the latest minor versions of the current major versions can be selected for cluster creation.

Upgrade and Ops Guarantee

The platform focuses on guaranteeing the normal upgrade and stable operation of the latest two major versions of etcd. For example, if the current latest version is **3.5.4**, the platform focuses on guaranteeing the upgrade feature of versions **3.5.4** and **3.4.13**, meanwhile supports troubleshooting, fault recovery, and bug fixing. Clusters of outdated versions may face the risks of unstable operation and upgrade failures. Please upgrade the version in time.

> **Note:**
> Cluster upgrades include major version upgrades and minor version upgrades.

Ticket Submission

Cloud Service for etcd provides technical support for the latest 2 major versions of etcd, including Q&A, online guidance, troubleshooting, and problem solving. However, for clusters with outdated versions, Cloud Service for etcd does not guarantee the quality and effectiveness of provided technical support.

Suggestions on etcd Major Version Selection

This document provides version selection suggestions from the perspectives such as stability, performance, and usability.

- **Stability**: **3.4** series versions have undergone large-scale usage and verification by global users for 2-3 years, and the overall stability is better than that of **3.5** series versions. The **3.5.4** and later versions fixed the critical bugs in earlier **3.5** versions. The open-source project Kubernetes has also used **3.5.4** in its version 1.25, and the stability in core scenarios has met the production environment requirements.
- **Performance**: Version **3.5.4** has merged a series of performance optimization PRs, significantly reduced the cluster startup time, and improved read/write QPS.
- **Usability**: **3.5** series versions have solved the dependency management problems in **3.4** series versions and earlier versions through package modularization, which reduces burden for the developers. The use of **3.5** series versions is good for subsequent version upgrades.

In summary, version 3.5.4 is recommended for creating clusters. For migrating existing clusters, you can choose the version same as the source clusters.

Version Change Description

Last updated：2025-03-19 18:30:11

v3.4.13 TKE Revisions

Time	Version	Update
September 2, 2022	v3.4.13-tke.5	Supported recording request error code in audit logs.
August 1, 2022	v3.4.13-tke.4	• Enhanced data security, added the blocking feature for full deletion requests, and supported viewing details of the blocked requests via metrics and audit logs. The blocking feature is disabled by default and can be enabled by setting the environment variable `ETCD_FORBID_DELETING_ALL_KEYS=true` . • Enhanced data security, added the blocking feature for batch deletion requests, and supported viewing details of the blocked requests via metrics and audit logs. The blocking feature is disabled by default and can be enabled by setting the environment variable `ETCD_MAX_DELETE_KEY_NUM=1000` . (Deletion will fail if the number of keys to be deleted exceeds 1,000.) • Improved Ops capabilities and added the audit feature for deletion requests (delete/txn), high-latency requests, and other key scenarios.
July 16, 2022	v3.4.13-tke.3	Optimized the performance and supported setting the parameters for maximum concurrent streams.
November 2, 2020	v3.4.13-tke.2	• Optimized the performance in the count-only scenario. • Improved the fault diagnosis capability and supported outputting the detailed information on high-latency requests. • Optimized the performance and pushed down the query limit parameter to the index layer.

v3.5.4 TKE Revisions

Time	Version	Update
November 24, 2022	v3.5.4-tke.4	Merged PR14733 to fix the issue of possible data inconsistency during defrag operations.
September 2, 2022	v3.5.4-tke.3	Supported recording request error code in audit logs.

August 25, 2022	v3.5.4-tke.2	• Enhanced data security, added the feature of blocking the requests for deleting all data, and supported viewing details of blocked requests via metrics and audit logs. The blocking feature is disabled by default and can be enabled by setting the environment variable `ETCD_FORBID_DELETING_ALL_KEYS=true` . • Enhanced data security, added the blocking feature for batch deletion requests, and supported viewing details of the blocked requests via metrics and audit logs. The blocking feature is disabled by default and can be enabled by setting the environment variable `ETCD_MAX_DELETE_KEY_NUM=1000` . (Deletion will fail if the number of keys to be deleted exceeds 1,000.) • Improved Ops capabilities and added the audit feature for deletion requests (delete/txn), high-latency requests, and other key scenarios.
July 19, 2022	v3.5.4-tke.1	Optimized the performance, merged PR14219, and supported setting the parameters of maximum concurrent streams.

Account Authorization

Last updated：2025-03-19 18:30:11

Authorizing Accounts

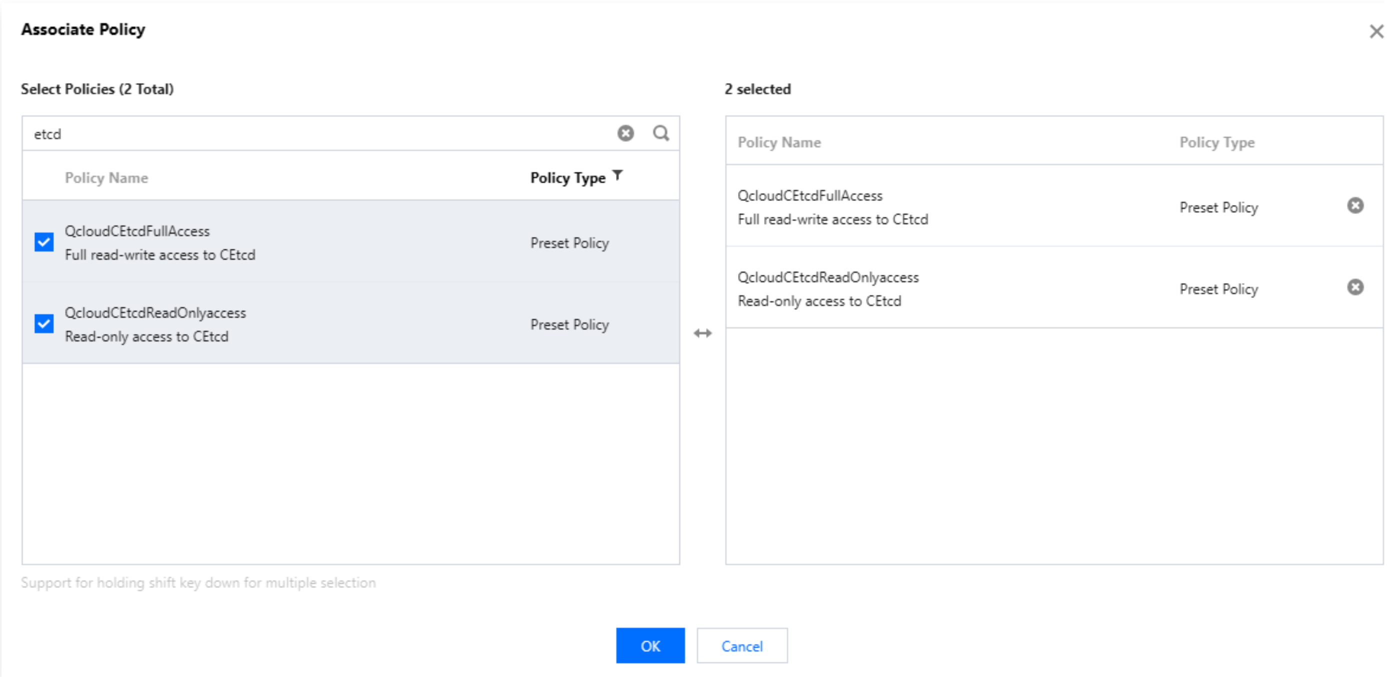
To perform more granular permission control for businesses, Cloud Service for etcd was integrated with Cloud Access Management (CAM) on May 5, 2022. After the integration, both users and backend services need to be authorized to access the console and call APIs. You can follow the steps below to authorize sub-users and the service to operate other cloud resources.

Directions

Authorizing Sub-users

Cloud Service for etcd has predefined the QcloudCEtcdFullAccess and QcloudCEtcdReadOnlyAccess policies on CAM for full access and read-only access scenarios respectively. If a sub-user needs to access the etcd console and call APIs, use an account with administrator permissions to authorize the sub-user, so that the sub-user can use the etcd service. The steps are as follows:

1. Log in to the CAM console and choose **Users > User List** from the left sidebar.
2. Click **Authorize** on the right of a sub-account's username in the user list.
3. Select the following policies in the Associate Policy pop-up window:
    - QcloudCEtcdFullAccess
    - QcloudCEtcdReadOnlyaccess

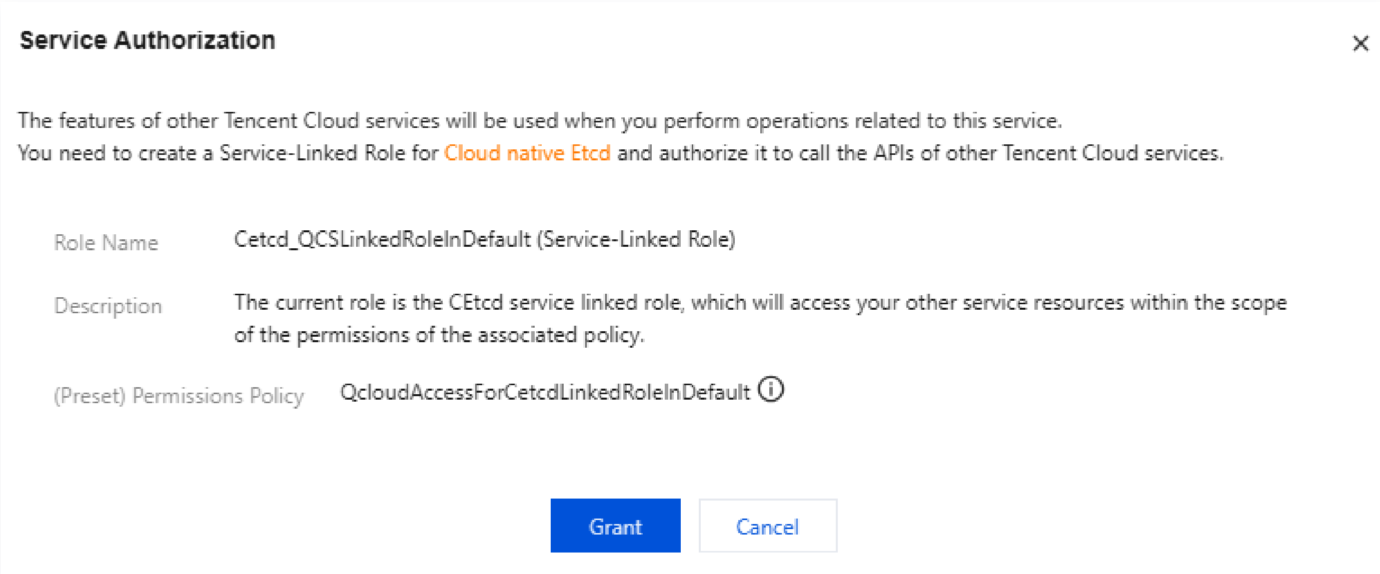


4. Click **OK**. After the authorization is completed, the sub-account can access the etcd service normally. The administrator can create custom policies to set more specific permissions for users.

Authorizing the Service

Service authorization refers to granting permissions for the current service to operate other cloud resources. For example, when a user creates an etcd cluster, Cloud Service for etcd should have the corresponding operation permissions if the user chooses to create TMP resources and tags. In this case, the administrator needs to authorize Cloud Service for etcd.

1. Log in to the console of Cloud Service for etcd. The **Service Authorization** window will pop up if no authorization is performed previously, as shown in the figure below:



2. Click **Grant** and complete authentication for authorization.

Cluster Creation

Last updated：2025-03-19 18:30:11

Overview

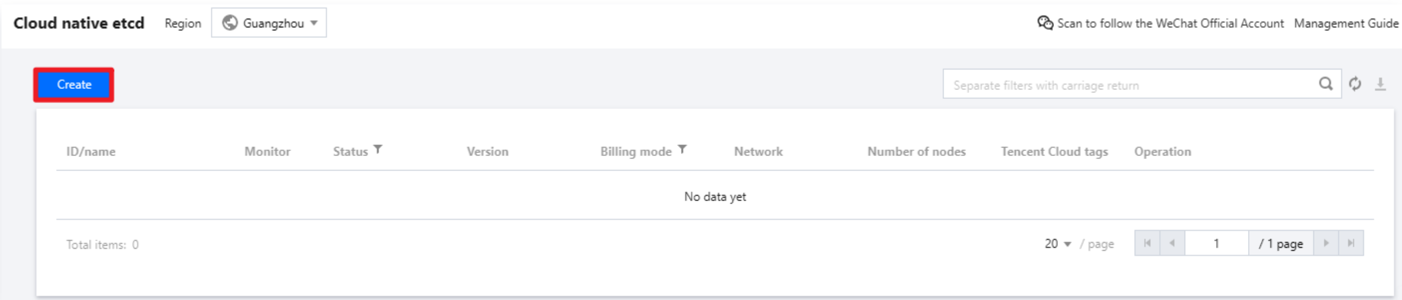
This document describes the creation of a managed etcd cluster in the Tencent Cloud TKE console. You can follow the steps below to create an etcd cluster.

Prerequisites

1. The service-related role TKE_QCSLinkedRoleInEtcdService has been authorized if you are using Cloud Service for etcd for the first time. For details, see Account Authorization.
2. Cloud Service for etcd uses various cloud resources of cloud products such as CBS and COS. Make sure that you have obtained the permissions to use the relevant cloud resources before you use the service. In addition, resources such as VPC, subnets, and security groups will also be used during cluster creation.

Directions

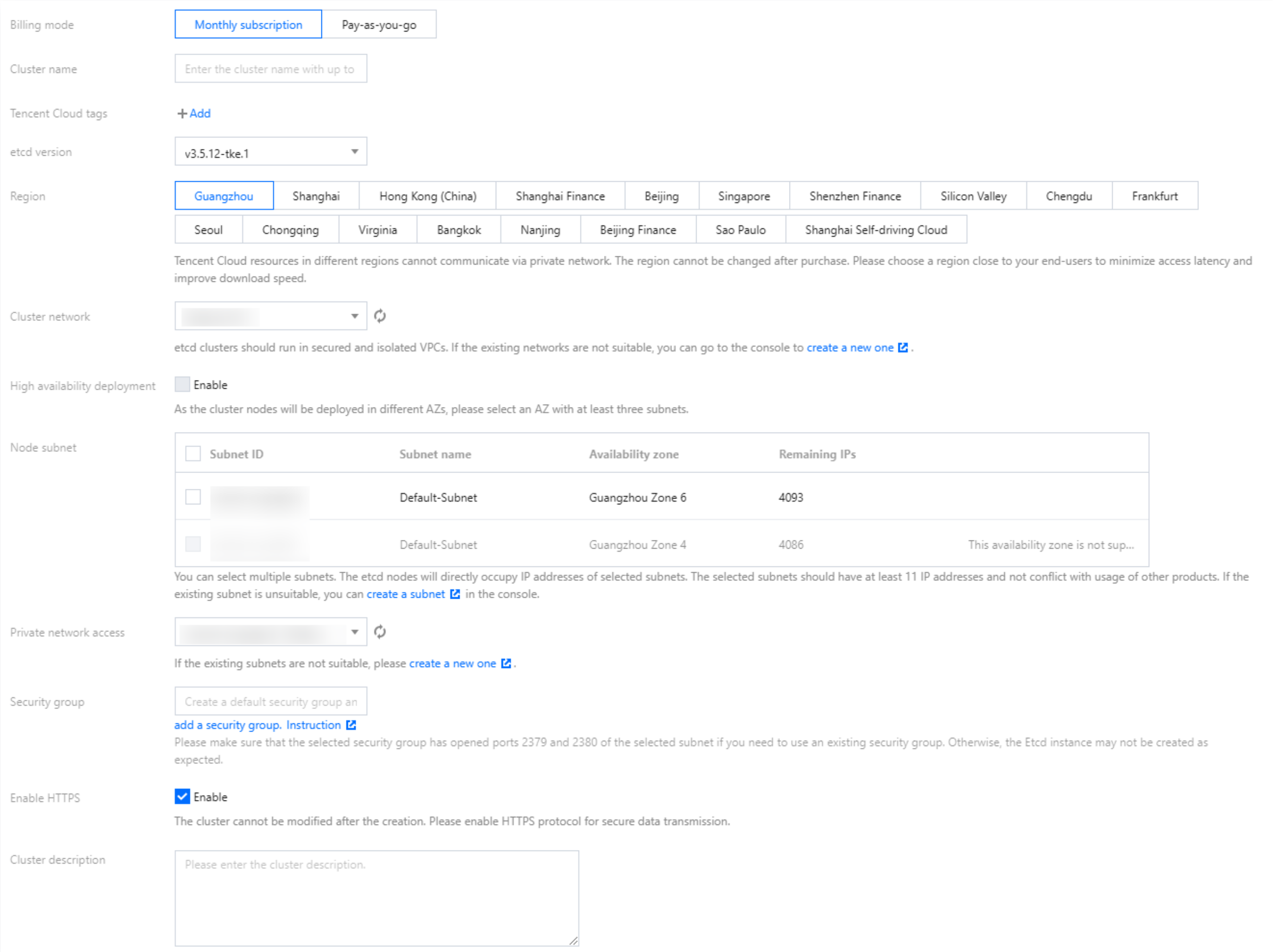
Log in to the console of Cloud Service for etcd, go to the etcd cluster list page, select the region where you want to create the etcd cluster at the top of the page, and click **Create**, as shown in the figure below:



The specific steps to create an etcd cluster are as follows:

1. Specifying the Cluster Information

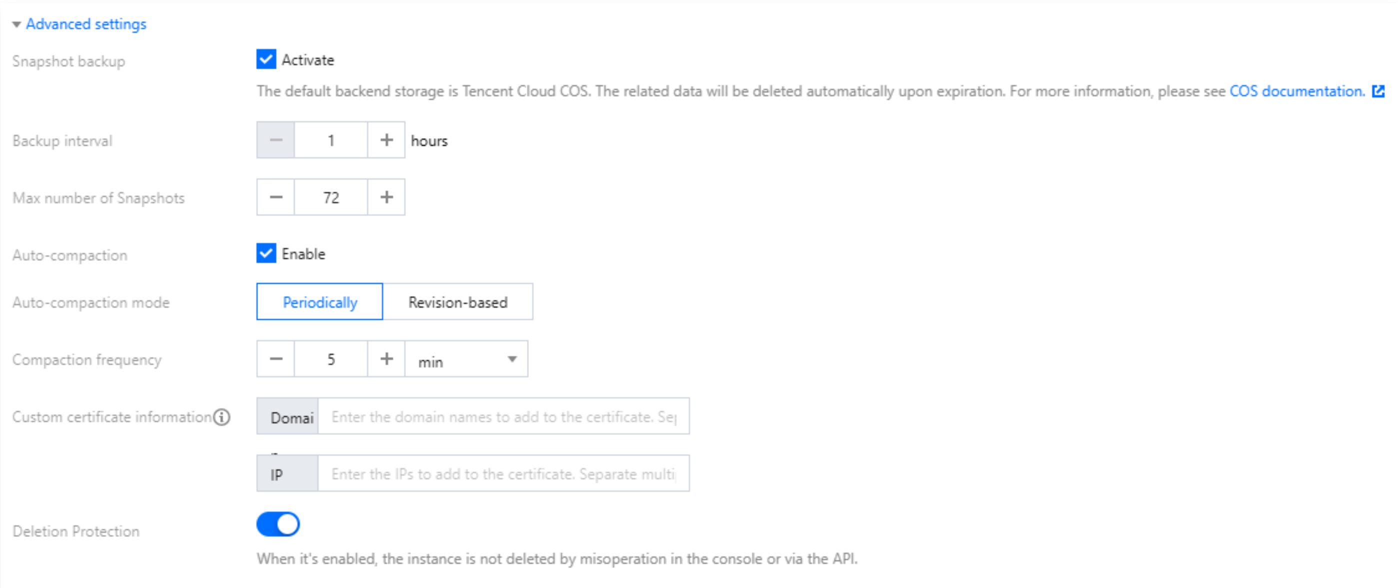
Go to the **Cluster information** page and complete the cluster information, as shown in the figure below:



- **Cluster name**: Enter a name for the etcd cluster with no more than 60 characters.
- **etcd version**: Select the etcd version.
- **Region**: Select the region where you want to deploy the cluster. Currently, you can only select certain regions such as Beijing, Shanghai, and Guangzhou. The region cannot be modified after the cluster is created. You are suggested to select a region closest to your businesses according to the geolocation to reduce access latency and improve data transfer speed.
- **Cluster network**: Select an existing VPC in the selected region. If no VPC is available in the region, go to the VPC console to create one. For details, see Virtual Private Cloud.
- **High availability deployment**: To ensure the high availability of the cluster, nodes in the cluster can be deployed randomly in different availability zones. If you enable high availability deployment, subnets in at least 3 AZs are required.
- **Node subnet**: Select the subnets to assign IP addresses to etcd cluster nodes. You should select subnets that have sufficient IP addresses and do not conflict with other products.

> **Note:**
> High availability deployment requires at least 3 etcd cluster nodes. If you enable **high availability deployment**, make sure that the available subnets are distributed in at least 3 different availability zones.

- **Private network access**: The external access address of the etcd cluster is in the selected subnets by default. You need to select subnets that have sufficient IP addresses and do not conflict with other products.
- **Security group**: Security groups can serve as firewalls and are used to control network access to nodes. The following operations are supported:
  - Create and bind a default security group. The inbound/outbound rules allow all ports of 0.0.0.0/0 and ::/0.
  - If you want to use an existing security group, **make sure that the security group allows ports 2379 and 2380 of the selected subnets**. Otherwise, the etcd cluster creation may fail.
- **Enable HTTPS**: Support enabling HTTPS two-way authentication. **Settings of this feature cannot be modified after the cluster is created**. You are suggested to enable it to ensure secure data transmission of the cluster.
  - **Client certificate verification**: Client certificate verification can be enabled if HTTPS is enabled.
- **Cluster description**: Enter information related to the cluster. The information will be displayed on the "Basic information" page of the cluster.



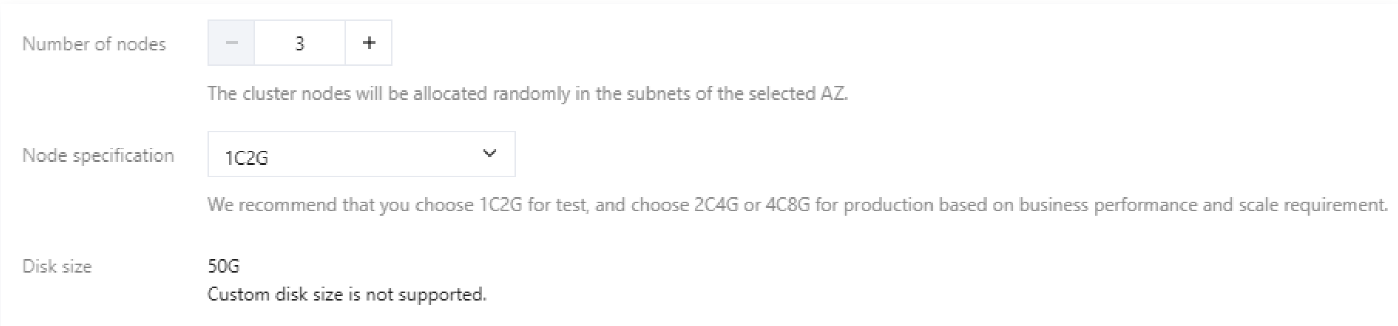
- **Snapshot backup:** Backup settings can effectively protect cluster data security. You can set the backup interval and the maximum number of snapshots. After the cluster is created, a COS bucket will be automatically created and billed according to actual resource usage. For billing modes, see COS Billing Overview.

  - Backup interval: Set the backup interval in hours.
  - **Max number of snapshots**: Set the maximum number of snapshots (maximum value: 1000). If this number is exceeded, new snapshots cannot be uploaded to the COS bucket.

- **Monitoring and alarm**: Cloud Service for etcd provides monitoring and alarm features by default. It also supports integration with TMP for cluster monitoring, analysis, and alarm. For details, see Prometheus Monitoring Overview.
  - **Existing Prometheus**: Support selecting an existing TMP instance, which should be in the same VPC as the etcd cluster.
- **Auto-compaction**: To prevent cluster performance degradation or space exhaustion, you can enable the automatic compaction feature to periodically clear historical data. Currently, it supports two compaction modes: **periodic compaction** and **revision-based compaction**.
  - **Periodically**: Set the compaction cycle in "hour, minute, or second".
  - **Revision-based**: Set the compaction cycle based on the number of retained revisions.

2. Specifying the Instance Specification

Complete the node information on the **Instance specification** page, as shown in the figure below:

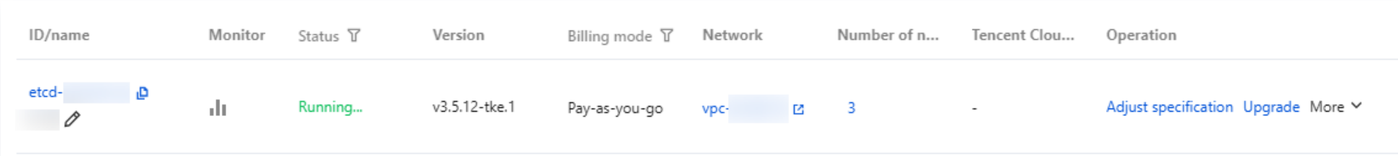


- **Number of nodes**: Specify the number of nodes in the cluster. At least 3 nodes are required to ensure high availability of the cluster.
- **Node specification**: Select the node specification based on business needs.
- **Disk size**: The disk size is fixed at 50 GB currently and cannot be modified. This size can meet the needs of most scenarios. If you have special disk size requirements, you can submit a ticket for application.

> **Note:**
> You are suggested to select 1C2G as the node specification in a test environment. For a production environment, 2C4G or 4C8G is recommended based on the business performance and scale requirements.

3. Confirming the Information

Check and confirm the cluster configuration and then click **Create**.
You can view the cluster creation progress on the **Cloud native etcd** page. When the cluster status is "Running", the cluster is successfully created and can be used, as shown in the figure below:



Monitoring and Alarm Configuration

Last updated：2025-03-19 18:30:12

Overview

Cloud Service for etcd provides monitoring data of node resource metrics, application metrics, instance-level metrics, and instance API metrics by default. Alarm rules can be set for these metrics. In addition, if you want to display custom monitoring metrics, you can use TMP as an additional monitoring service and configure custom monitoring metrics.

Prerequisites

etcd clusters have been created.

Directions

1. Log in to the console of Cloud Service for etcd.
2. Go to the etcd cluster list page, select the region at the top of the page, and find the target etcd cluster in the cluster list below.

Viewing Monitoring Metrics

Click the icon of the corresponding cluster in the cluster list to enter the **details page**. Then, go to the **Instance monitoring** page.

Metric Description

Aggregation method:

1. Time range: Time range of the data displayed in monitoring charts.
2. Granularity: Aggregation method for each data point in charts. Raw data is collected every 15 seconds by default. The console supports displaying the average value of data aggregated by 1 minute or 5 minutes.
3. Automatic refresh: Interval for automatically refreshing the charts, which can be 30 seconds, 5 minutes, 30 minutes, or 1 hour. This feature is disabled by default.

Metric overview:

Metric Dimension	Metric Name	Unit	Metric Description
Node resource	CPU usage percentage	Percentage	Proportion of the current node CPU usage to Pod CPU request value of the current node (the CPU specification selected during instance creation).

Node resource	File system read rate	MiBytes/s	Volume of data read from the node data disk per second.
Node resource	File system write rate	MiBytes/s	Volume of data written to the node data disk per second.
Node resource	Memory usage percentage	Percentage	Proportion of the current node memory usage to Pod MEM request value of the current node (the memory specification selected during instance creation).
Node resource	Total memory usage	MiBytes	Memory usage of the node.
Node resource	Network receiving rate	MiBytes	Data receiving rate of the node NIC.
Node resource	Network sending rate	MiBytes	Data sending rate of the node NIC.
Application metric	Number of database keys	Count	Number of keys on the node. The data is obtained from etcd metrics. Calculation formula: etcd_debugging_mvcc_keys_total{job="$job"}.
Application metric	Database MVCC writes	Times	Number of data writes on the node. Calculation formula: etcd_mvcc_put_total{job="$job"}.
Application metric	Database size	MiBytes	Database size calculated on the node. Calculation formula: etcd_debugging_mvcc_db_total_size_in_bytes{job="$job"}.
Application metric	Consensus proposal application rate	Time/s	The value is usually small. (Only a few thousand proposals are applied even under high load). If the value keeps increasing, it indicates that the etcd server is overloaded, which may be caused by queries with high consumption (such as large-range queries or large txn operations). Calculation formula: rate(etcd_server_proposals_applied_total{job="$job"}[5m]).
Application metric	Consensus proposal commit rate	Time/s	The value usually increases over time. If the delay between a single member and the leader is high for a long time, it indicates that the member is running slowly or is not healthy. Calculation formula: rate(etcd_server_proposals_committed_total{job="$job"}[5m]).

Application metric	Total number of queueing consensus proposals	Count	An increase of the value indicates that the client load is high or the member cannot submit proposals. Calculation formula: etcd_server_proposals_pending{job="$job"}.
Application metric	Growth rate of failed consensus proposals	Time/s	This metric is usually related to two issues: temporary failures related to leader election or longer-term failures due to the lack of the arbitration node in the cluster. Calculation formula: rate(etcd_server_proposals_failed_total{job="$job"}[5m]).
Instance-level metric	Cluster leader existence	Boolean value	If there is no leader, the instance is unavailable. Calculation formula: max(etcd_server_has_leader{job="$job"}).
Instance-level metric	Total number of leader switches	Times	If there is no leader, frequent leader changes will great affect the etcd performance, which may be caused by network connection issues or high load on the etcd cluster. Calculation formula: max(etcd_server_leader_changes_seen_total{job="$job"}). **Note:** The data collected for this metric is the aggregated data after the cluster is successfully created. It is irrelevant to the alarm cycle.
Instance API	gRPC call rate	Time/s	gRPC call rate of operations using the specific method. Calculation formula: sum(rate(grpc_server_handled_total{job="$job"}[1m])) by (job,grpc_method,instance).

Node Addition and Configuration Adjustment

Last updated: 2025-03-19 18:30:11

Overview

This document introduces node scaling out and node scaling up/in for etcd clusters in the console.

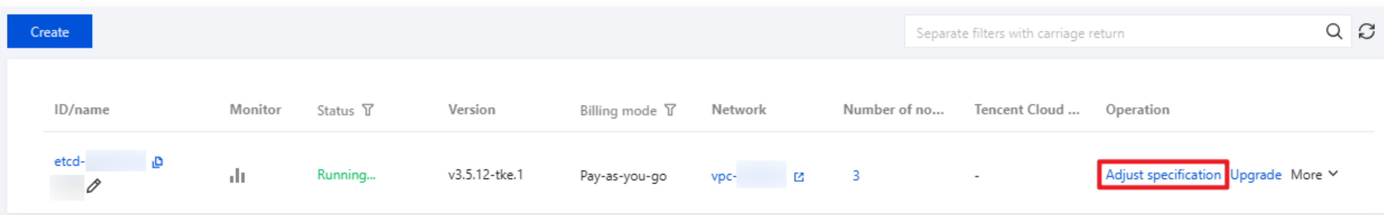
> **Notes:**
> - Node scaling out (by adjusting the number of nodes) does not affect normal business operation.
> - Node scaling up/in (by adjusting the node specifications) will trigger rolling updates. It is recommended to perform the operation at off-peak hours.

Prerequisites

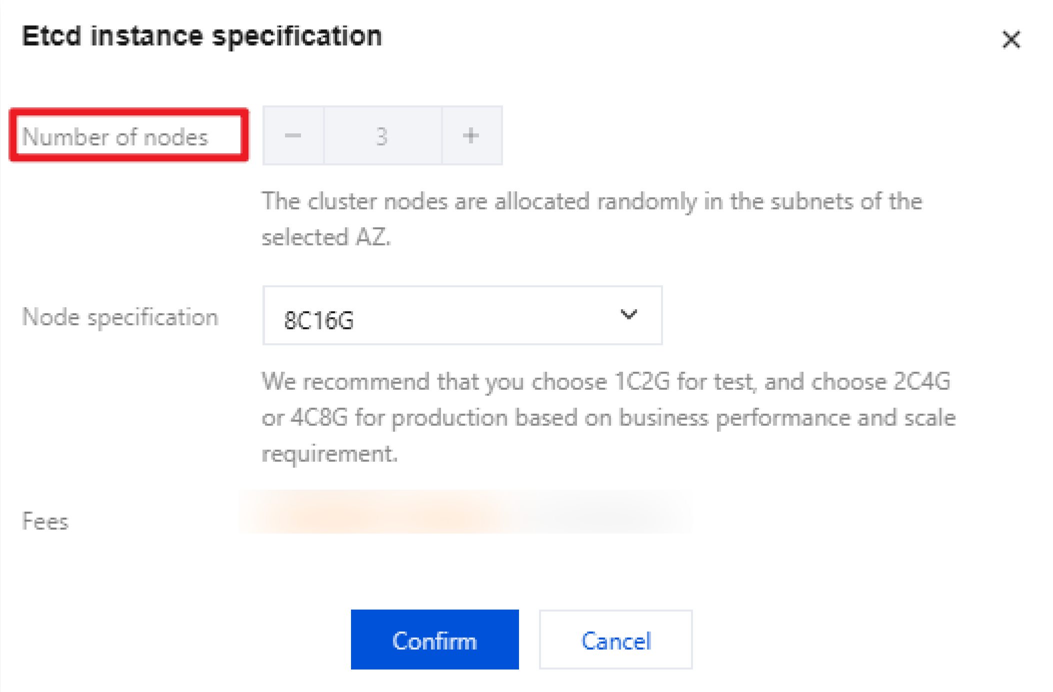
etcd clusters have been created.

Directions

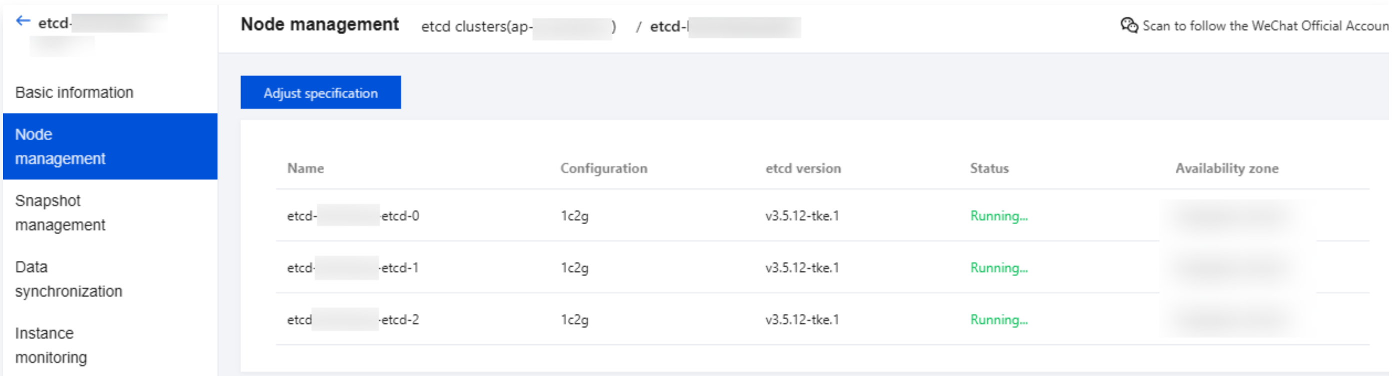
1. Log in to the console of Cloud Service for etcd.
2. Go to the etcd cluster list page, select the region at the top of the page, and find the target etcd cluster in the cluster list below.
3. Click **Adjust specification** in the operation column on the right, as shown in the figure below:



4. Specify the number of nodes after addition or the node specifications in the popup window.



5. Click **Confirm** to start creating nodes.
6. Click the target cluster to enter the cluster details page and view node details on the **Node management** page.



Multi-AZ Adjustment for Nodes

Last updated：2025-03-19 18:30:13

Overview

This document describes multi-AZ adjustment for nodes of etcd clusters in the console.

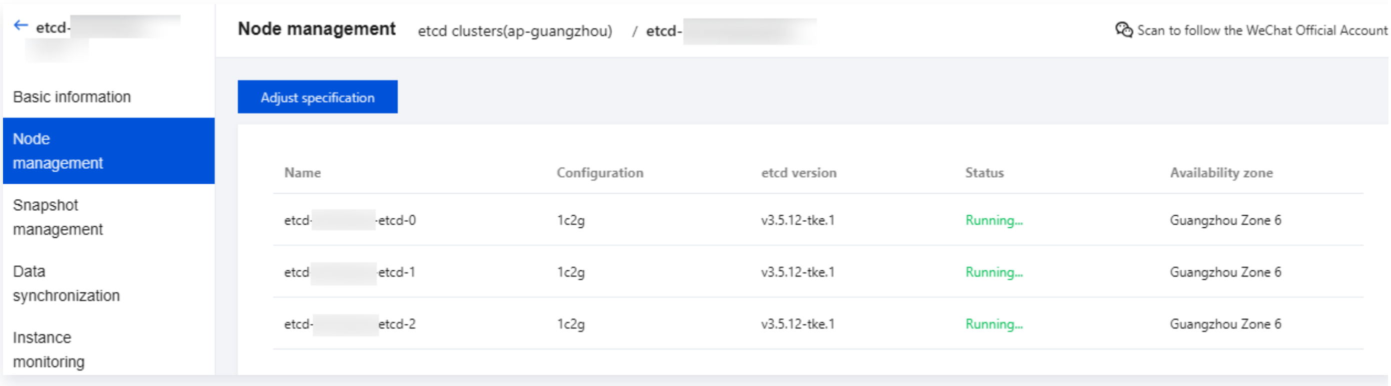
> **Note:**
> - Node addition (increasing the number of nodes) does not affect normal business operation.
> - Node removal (decreasing the number of nodes) may trigger a leader election in the cluster. It is recommended to perform this operation at off-peak hours.

Prerequisites

etcd cluster have been created.

Viewing AZ Distribution of Cluster Nodes

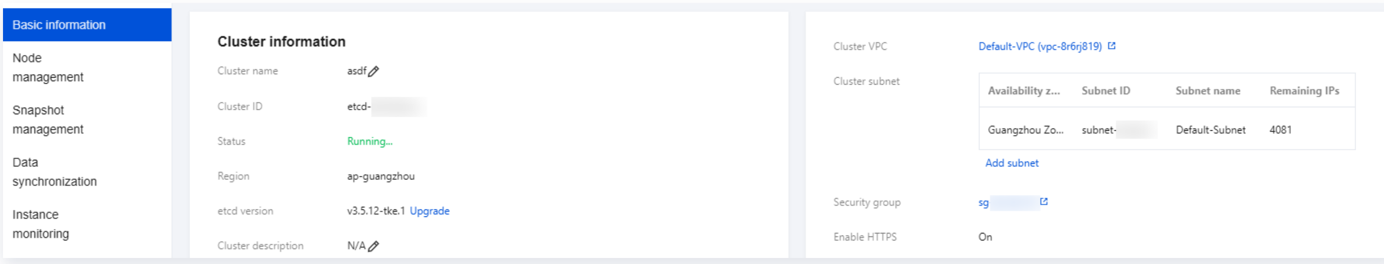
1. Log in to the console of Cloud Service for etcd.
2. Go to the etcd cluster list page, select the region at the top of the page, and find the target etcd cluster in the cluster list below.
3. View the AZ distribution of nodes on the node management page, as shown in the figure below:



Viewing AZ Distribution of Subnets and Adding Subnets in Different AZs

1. Log in to the console of Cloud Service for etcd.
2. Go to the etcd cluster list page, select the region at the top of the page, and find the target etcd cluster in the cluster list below.
3. Go to the basic information page and view the AZ distribution of current subnets and the button for

adding subnets, as shown in the figure below:



4. Click **Add subnet**, select subnets in different AZs, and click **Confirm**. If high availability deployment is enabled, subnets in at least three AZs are required.

Multi-AZ Adjustment Methods

Number of Cluster Nodes	Actual Number of AZs in the Region	Number of AZs for Current Cluster Subnets	Actual Number of AZs for Cluster Nodes	Multi-AZ Adjustment Method
1 node	1 AZ	1 AZ	1 AZ	N/A
	2 AZs	1 AZ	1 AZ	Add 1 subnet in a different AZ, and increase the number of nodes to 3.
		2 AZs	1 AZ	Increase the number of nodes to 3.
	⩾ 3 AZs	1 AZ	1 AZ	Add 2 subnets in different AZs, and increase the number of nodes to 3.
		2 AZs	1 AZ	Add 1 subnet in a different AZ, and increase the number of nodes to 3.
		⩾ 3 AZs	1 AZ	Increase the number of nodes to 3.
3 nodes	1 AZ	1 AZ	1 AZ	N/A
	2 AZs	1 AZ	1 AZ	Add 1 subnet in a different AZ, and increase the number of nodes to 5.

		2 AZs	1 AZ	Increase the number of nodes to 5.
		2 AZs	2 AZs	N/A
	⩾ 3 AZs	1 AZ	1 AZ	Add 2 subnets in different AZs, and increase the number of nodes to 5.
		2 AZs	1 AZ	Add 1 subnet in a different AZ, and increase the number of nodes to 5.
		2 AZs	2 AZs	Add 1 subnet in a different AZ, and increase the number of nodes to 5.
		⩾ 3 AZs	1 AZ	Increase the number of nodes to 5.
		⩾ 3 AZs	2 AZs	Increase the number of nodes to 5.
		⩾ 3 AZs	⩾ 3 AZs	N/A
5 nodes	1 AZ	1 AZ	1 AZ	N/A
	2 AZs	1 AZ	1 AZ	Add 1 subnet in a different AZ, decrease the number of nodes to 3, and then increase the number of nodes to 5.
		2 AZs	1 AZ	Decrease the number of nodes to 3, and then increase the number of nodes to 5.
		2 AZs	2 AZs	N/A
	⩾ 3 AZs	1 AZ	1 AZ	Add 2 subnets in different AZs, decrease the number of nodes to 3, and then increase the number of nodes to 5.
		2 AZs	1 AZ	Add 1 subnet in a different

				AZ, decrease the number of nodes to 3, and then increase the number of nodes to 5.
		2 AZs	2 AZs	Add 1 subnet in a different AZ, decrease the number of nodes to 3, and then increase the number of nodes to 5.
		⩾ 3 AZs	1 AZ	Decrease the number of nodes to 3, and then increase the number of nodes to 5.
		⩾ 3 AZs	2 AZs	Decrease the number of nodes to 3, and then increase the number of nodes to 5.
		⩾ 3 AZs	⩾ 3 AZs	N/A

Automatic Compression Configuration

Last updated：2025-03-19 18:30:12

Overview

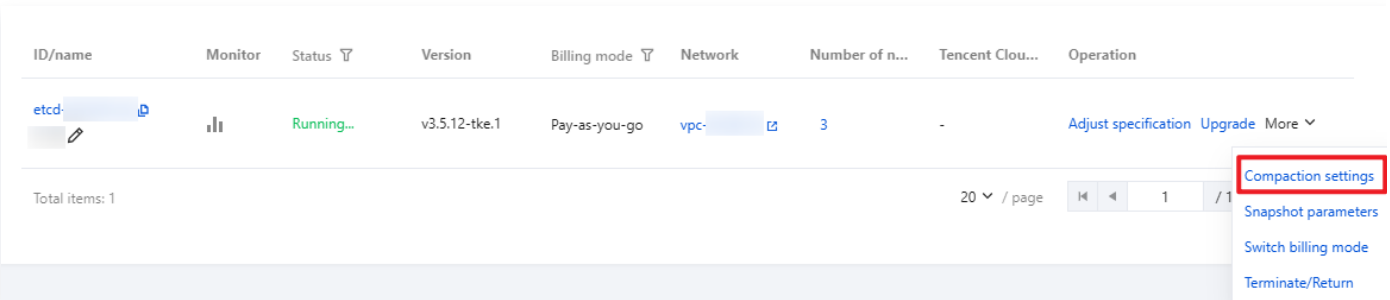
During the cluster operation, all update operations on etcd key-value pairs will be recorded. To prevent cluster performance degradation or storage space exhaustion, Cloud Service for etcd supports scheduled data cleanup by using the automatic compaction feature. Currently, etcd clusters support two compaction modes: **periodic compaction** and **revision-based compaction**. This document describes automatic data compaction settings for etcd clusters in the console.

Prerequisites

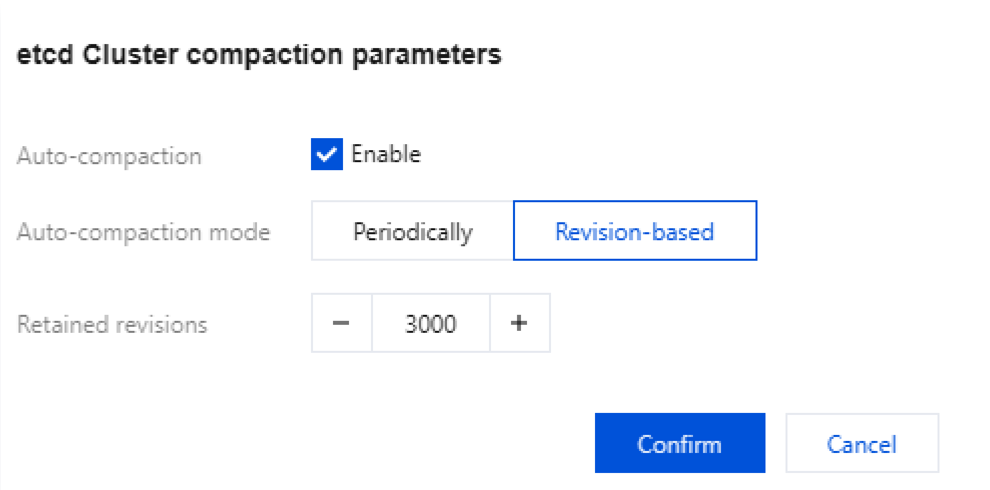
etcd clusters have been created.

Directions

1. Log in to the console of Cloud Service for etcd.
2. Go to the etcd cluster list page, select the region at the top of the page, and find the target etcd cluster in the cluster list below.
3. Choose **More** > **Compaction settings** in the operation column on the right of the target cluster, as shown in the figure below:



4. Check the Enable option to enable the automatic compaction feature in the pop-up window and select the automatic compaction mode, as shown in the figure below:



   - **Periodically**: Set the compaction cycle in "hour, minute, or second".
   - **Revision-based**: Set the compression cycle based on the number of retained revisions.

5. Click **Confirm** to complete the automatic compaction configuration.

Snapshot Management

Last updated：2025-03-19 18:30:12

Overview

Cloud Service for etcd supports both automatic backup and manual backup. The result of each backup operation is displayed as a snapshot in the snapshot list. You can recover a cluster to a previous state by using a snapshot. This document introduces etcd cluster snapshot management.

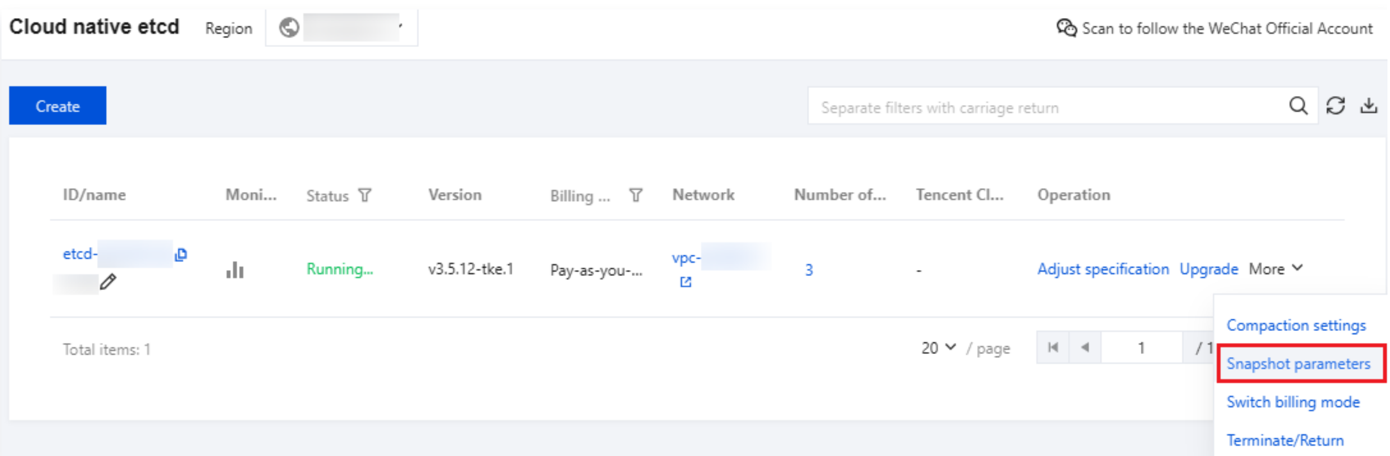
> **Notes:**
> - After an etcd cluster is created, a COS bucket will be automatically created and billed based on actual resource usage. Snapshots will be stored in this bucket. For COS billing details, see COS Billing Overview.
> - Snapshots of etcd clusters in the same region are stored in the same bucket (named in the format of "tencentcloud-tke-etcd-backups-region abbreviation-account AppID"). These snapshots are isolated by cluster ID, being used as subdirectories.

Prerequisites

etcd clusters have been created.

Directions

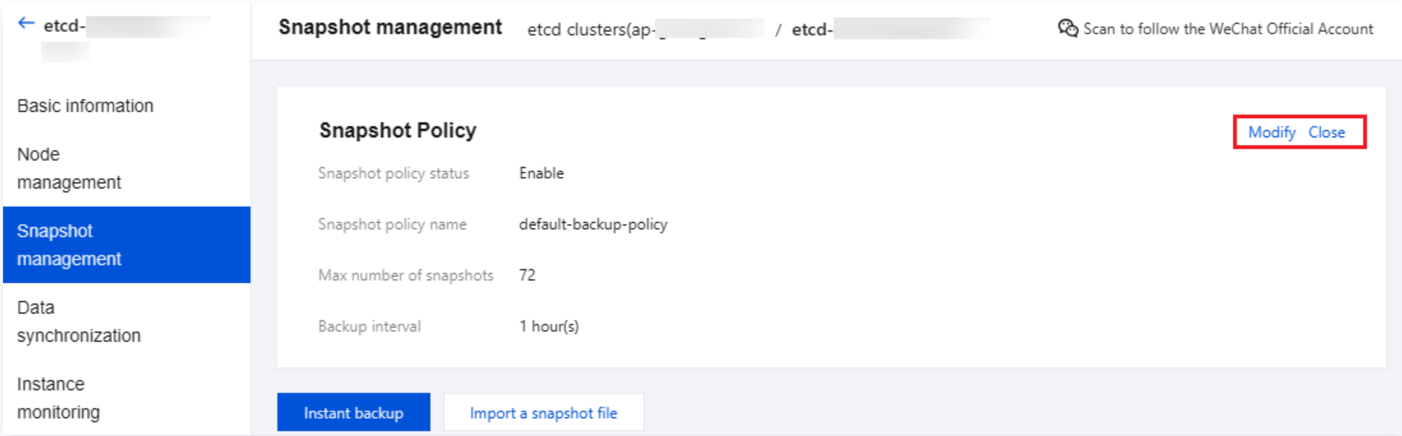
1. Log in to the console of Cloud Service for etcd.
2. Select the region on the Cloud native etcd page, and choose the target etcd cluster from the cluster list below.
3. Choose **More** > **Snapshot parameters** in the operation column on the list page, or click the target cluster to enter the **Snapshot management** page, as shown in the figure below:



Snapshot Policy Management

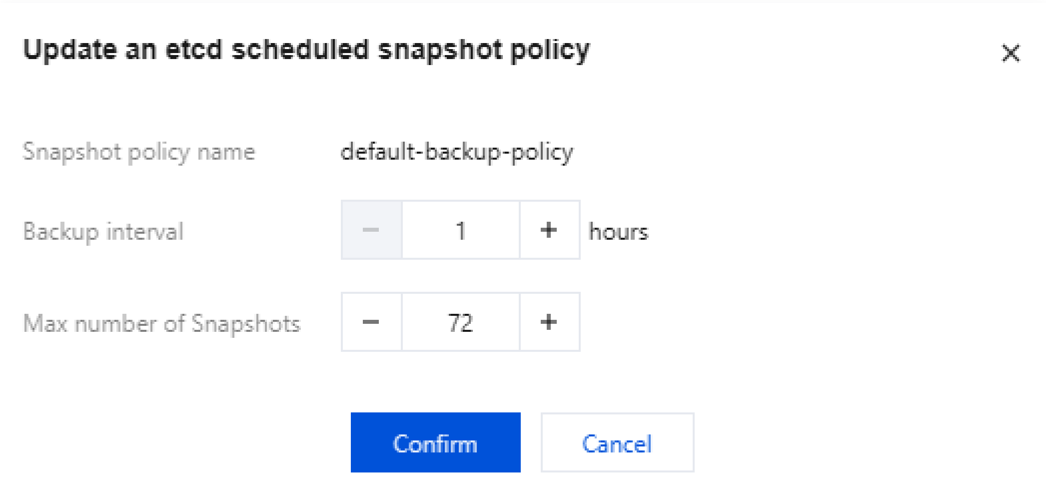
Viewing a Snapshot Policy

The **data backup** feature is enabled for etcd clusters by default during cluster creation. The system will generate a scheduled snapshot policy based on the backup parameters. You can view the policy details on the **Snapshot management** page, as shown in the figure below:



Modifying a Snapshot Policy

You can click **Modify** in the upper right corner of the **Snapshot Policy** area to update the scheduled snapshot policy, as shown in the figure below:



- **Snapshot policy name**: Retain the current policy name.
- **Backup interval**: Set the backup interval (in hours).
- **Max number of snapshots**: Set the maximum number of snapshots (maximum value is 1000). If this number is exceeded, new snapshots cannot be uploaded to the COS bucket.

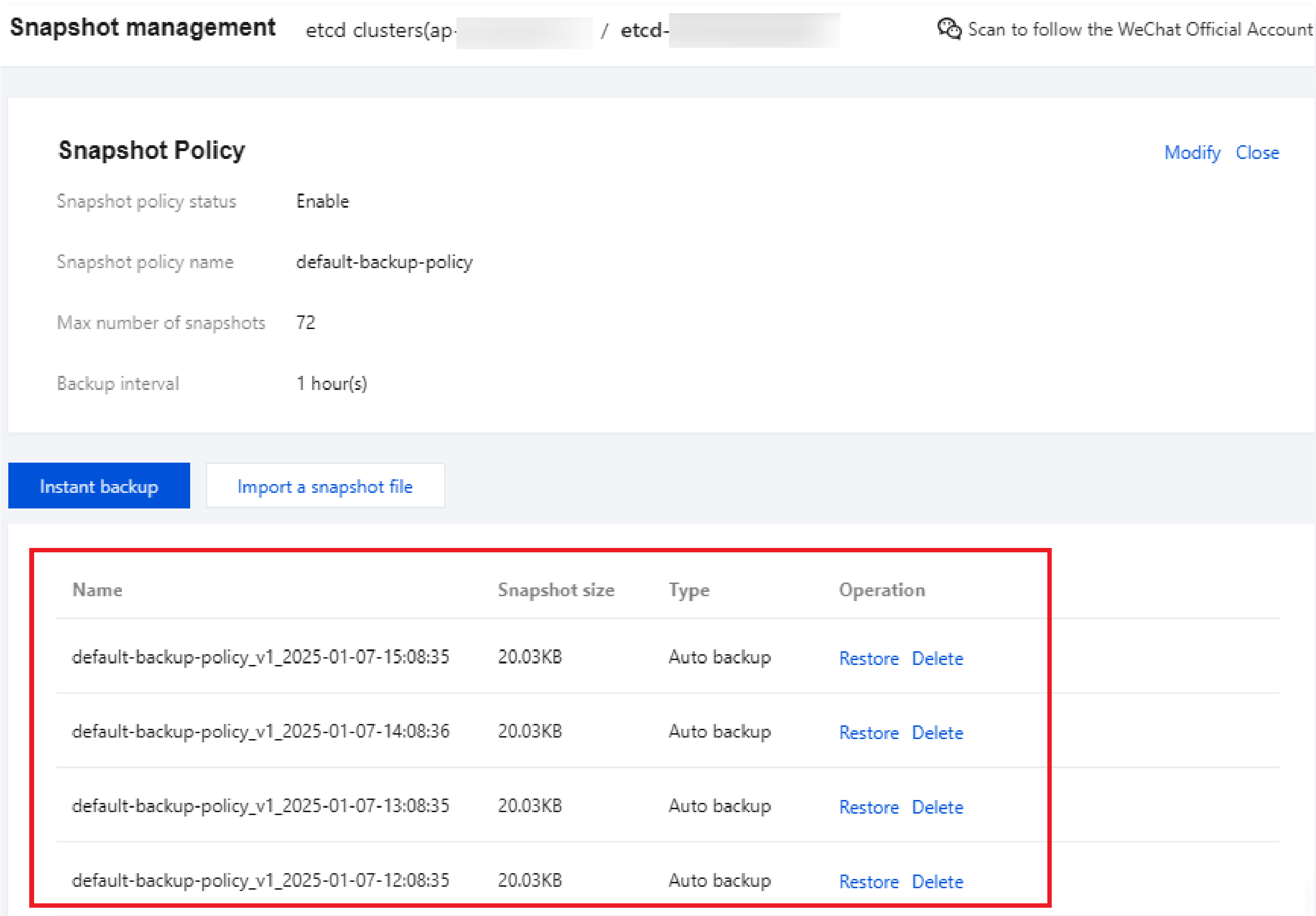
Disabling a Snapshot Policy

You can click **Close** in the upper right corner of the **Snapshot Policy** area to disable the scheduled snapshot policy. After the scheduled snapshot policy is disabled, the generated snapshots will not be deleted.

Backup Operations

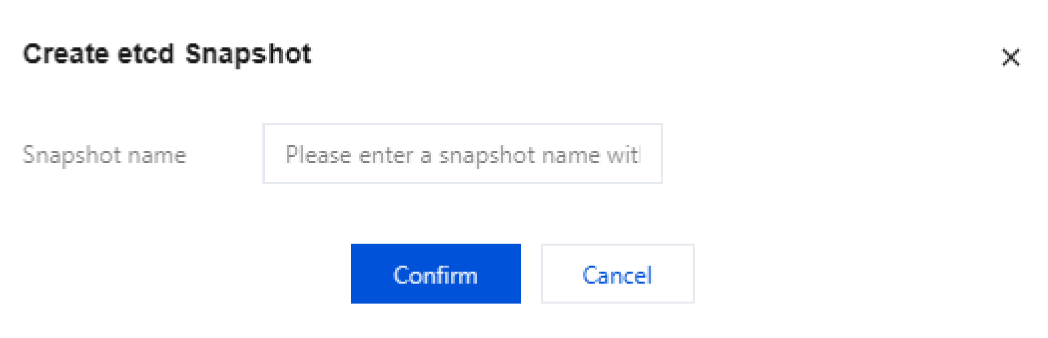
Performing Scheduled Backup Operations

The system performs scheduled backup operations on the cluster based on the snapshot policy. You can go to the **Snapshot management** page to view the scheduled snapshot list, as shown in the figure below:



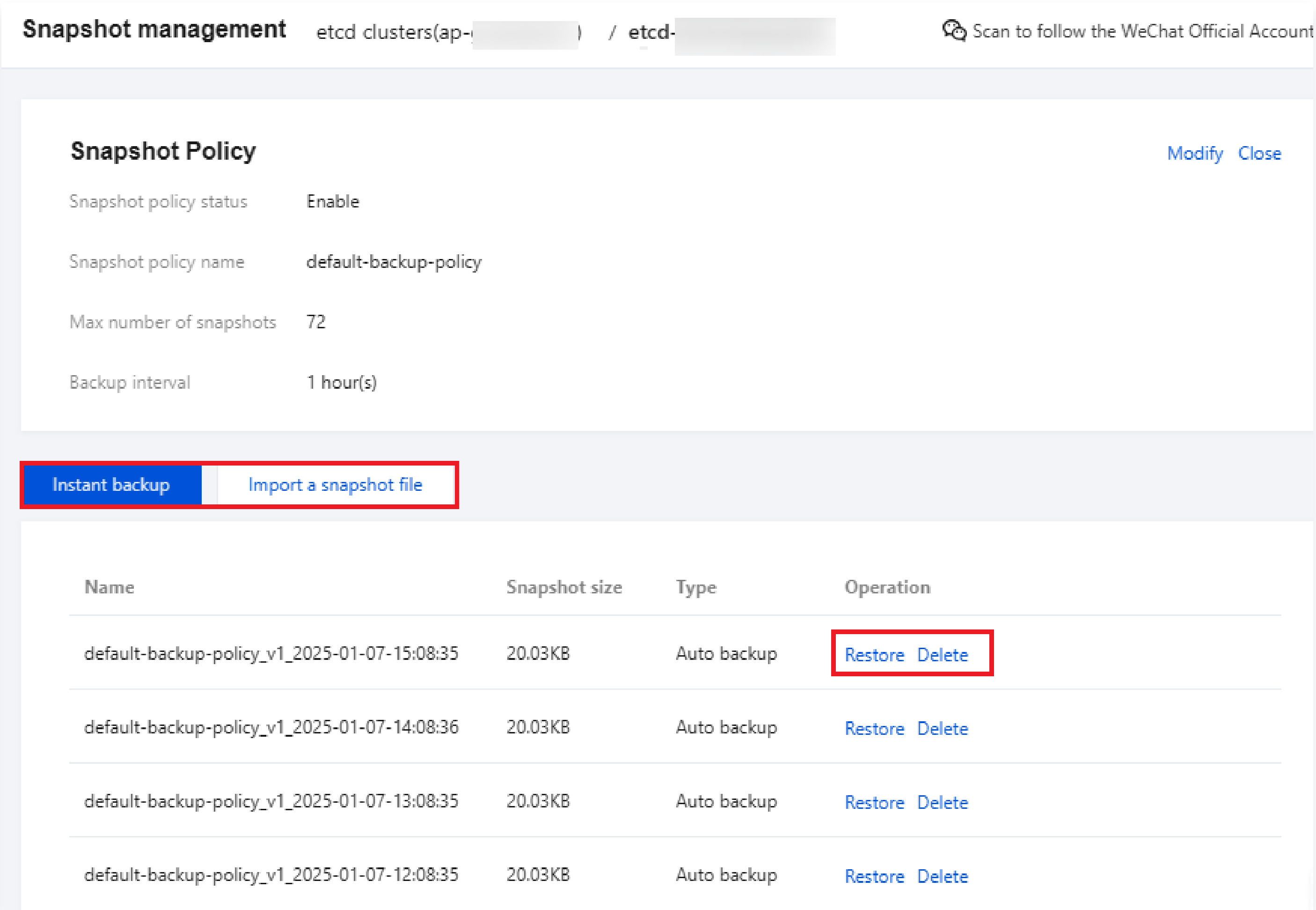
Performing Instant Backup Operations

The system supports instant backup operations. You can go to the **Snapshot management** page, click **Instant backup**, and enter the snapshot name in the pop-up window. The snapshot will be displayed in the snapshot list. The operation is shown in the figure below:



Snapshot Operations

The etcd cluster snapshots support snapshot file **import**, **restoration**, and **deletion** operations, as shown in the figure below:



Importing a Snapshot File

You can use this method to synchronize the data from the source etcd cluster to a cloud etcd cluster. It is recommended to stop access to the source etcd cluster on the business side first, and then click **Import a snapshot file**, import a local snapshot file containing the data of the source etcd cluster, and synchronize the data by clicking **Restore**.

Warning:
Do not directly copy the db file of etcd clusters as the snapshot file. Continuous updates to the source file during file copying can cause data corruption, and data integrity cannot be guaranteed. You can execute the `etcdctl snapshot save` command to generate a snapshot and the `etcdctl snapshot status` command to verify the snapshot status.

Restoring from a Snapshot

You can click **Restore** in the operation column to restore a cluster to the state recorded by the snapshot.

Note:
The existing cluster data will be overwritten by the data contained in the snapshot during restoration, and the cluster will be temporarily unavailable. Operate with caution.

Deleting a Snapshot

You can click **Delete** in the operation column to delete a snapshot from the snapshot list and the bucket. **Once a snapshot is deleted, it cannot be restored**.

Data Synchronization

Last updated：2025-03-19 18:30:12

Overview

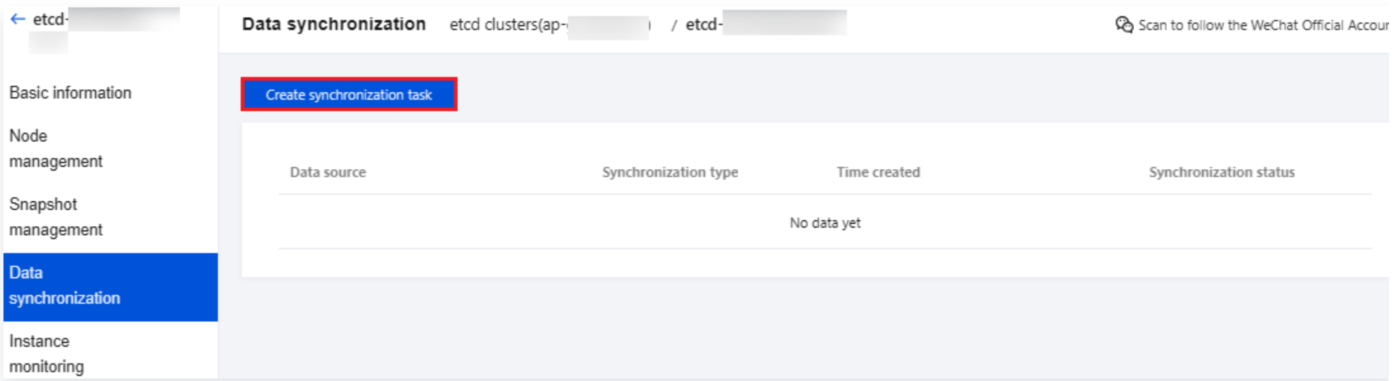
This document describes data synchronization from the source etcd cluster to a Cloud Service for etcd cluster in the console for unified management.

Prerequisites

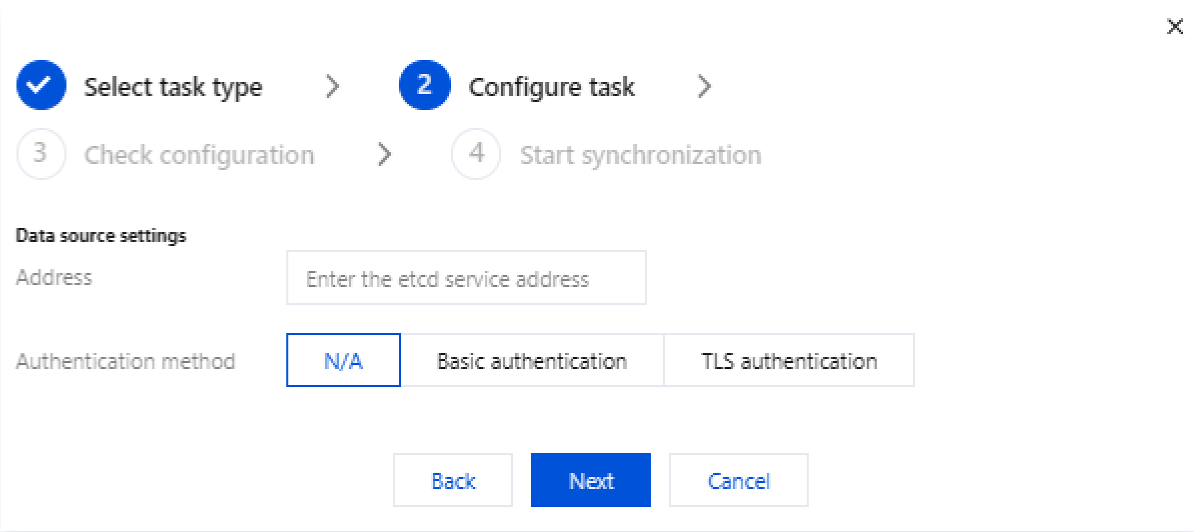
etcd clusters have been created.

Directions

1. Log in to the console of Cloud Service for etcd.
2. Go to the etcd cluster list page, select the region at the top of the page, and click the name of the target etcd cluster in the cluster list below.
3. Go to the **Data synchronization** page and click **Create synchronization task**, as shown in the figure below:



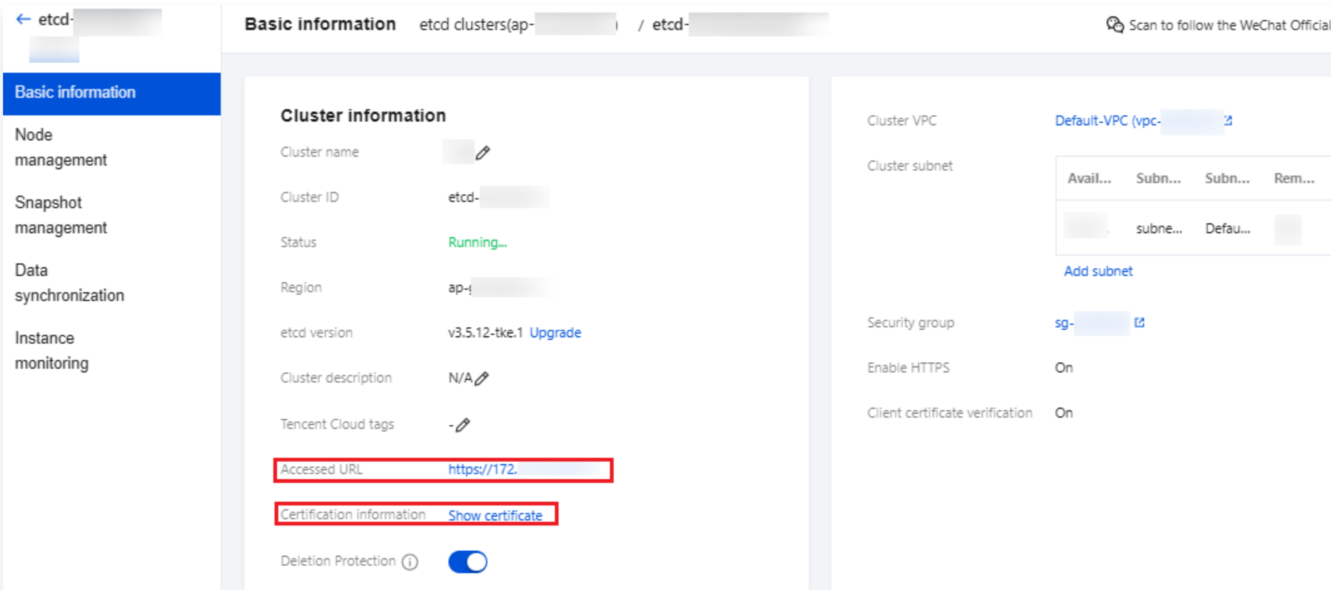
4. Select the task type.
   - Type 1: Automatically import all data from the source etcd cluster to the target cluster at one time by using a snapshot.
   - Type 2: Synchronize data from the source etcd cluster to the target cluster in real time. (This type is not available currently.)
5. Click **Next** to configure the task, as shown in the figure below:



5.1 Fill in the service address of the source etcd cluster. For an existing Cloud Service for etcd cluster, you can obtain the **access address** of the cluster from the **Basic information** page.

5.2 Complete the key or certificate information based on the authentication method of the source etcd cluster. For an existing Cloud Service for etcd cluster, you can obtain the certificate information on the cluster from the **Basic information** page, select **TLS authentication**, and fill in the certificate information, as shown in the figure below:

> **Note:**
> A normal network connection is required between the source etcd cluster and the target cluster, which means that they should be deployed in the same VPC. Otherwise, you are suggested to use the snapshot management feature to synchronize data.



6. Click **Next** to check the configuration. After the task configuration is completed, the backend will conduct a connection test based on the access address and authentication information of the source etcd cluster. The test covers network connection, permission settings, version support, and data volume.
7. Click **Confirm** to start synchronization. The synchronization between the source etcd cluster and the target cluster is expected to take about 5 minutes.

Cluster Deletion

Last updated：2025-03-19 18:30:12

Cloud Service for etcd clusters do not support cascade deletion. Before you delete a cluster, make sure that all stored data in the cluster has been removed. Cluster deletion cannot be restored. Operate with caution.

Note:
- Before you delete a cluster, it is recommended that you check the instance monitoring data to confirm if client businesses are still using the cluster.
- For clusters with password authentication enabled, you cannot check the data storage status. Therefore, they are not subject to the cascade deletion restriction. Operate with caution.

Directions

Clearing the Stored Data

To clear all key-value data in an etcd cluster, you can use the etcdctl command-line tool to perform the following operations:

1. Ensure that etcdctl is installed. If not, download and install the latest version from the GitHub release page of etcd.
2. Set environment variables for etcdctl, typically including ETCDCTL_API, ETCDCTL_ENDPOINTS, ETCDCTL_CACERT, ETCDCTL_CERT, and ETCDCTL_KEY.
3. Before you delete the data, you are suggested to use the etcdctl tool to create a snapshot of the etcd cluster and save it as a file for backup and recovery. Execute the following command:

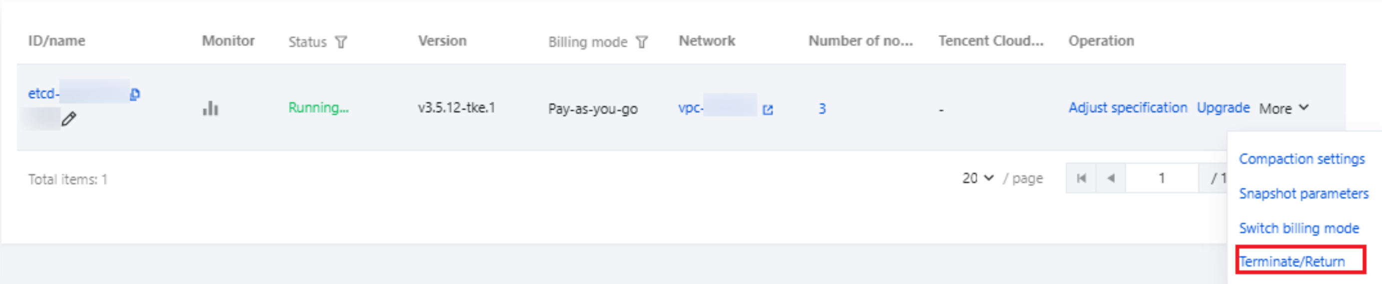
```
etcdctl --command-timeout=30s snapshot save $etcdID-final.db
```

   - `--command-timeout=30s` is an optional parameter that sets the command timeout to 30 seconds. If the command is not executed successfully within this time, it will be terminated.
   - `snapshot save $etcdID-final.db` is the actually executed command. It creates a snapshot of the etcd cluster and saves it as a file. `$etcdID` is a variable. You need to replace it with the actual etcd node ID or any other name you want to use.
4. Use etcdctl to delete all key-value data. The following command will delete all key-value pairs under the root directory:

```
etcdctl del "" --prefix
```

Deleting the Cluster

1. Log in to the console of Cloud Service for etcd.
2. Choose **More > Terminate/Return** on the right of the cluster on the **Cloud native etcd** page, as shown in the figure below:



3. Select other cloud resources to be terminated on the **Terminate/return instances** page.
   - Snapshot data stored in the COS bucket will be retained by default. If the option of COS snapshot data deletion is selected, the snapshot data of the cluster in the bucket will be deleted, but the bucket will be retained.
   - CLB resources generated after public network access is enabled will be deleted along with the Cloud Service for etcd cluster.
4. Click **Confirm**.

Cluster Troubleshooting
Snapshot Generation Failure

Last updated：2025-03-19 18:30:12

Description

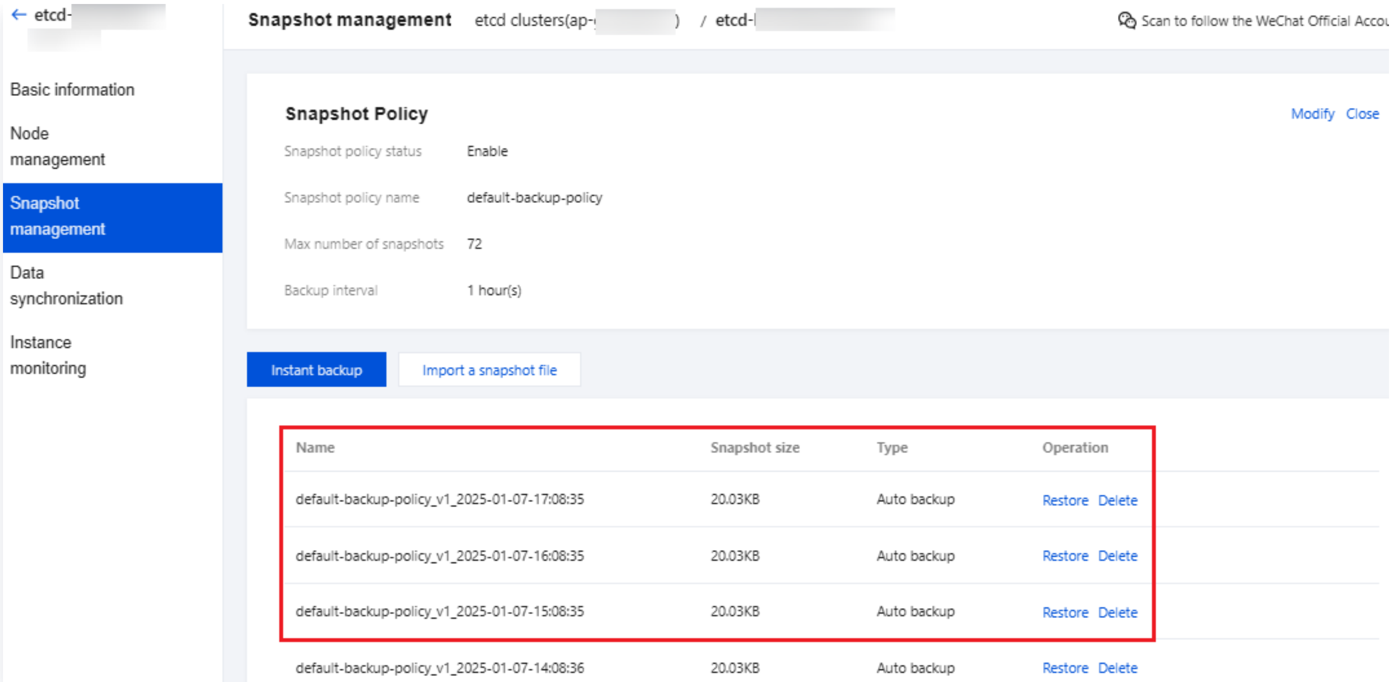
You have enabled the snapshot policy, but snapshot files are not generated as expected. It is possibly because that you have changed the etcd password. You can submit a ticket to solve the problem of snapshot file generation failure.

Directions

Viewing the Snapshot Policy and Snapshot Files

When an etcd cluster is created, the system enables the data backup feature by default and generates the corresponding scheduled snapshot policy based on the backup parameters. You can go to the snapshot management page to view the policy details and snapshot file list. The specific steps are as follows:

1. Log in to the console of Cloud Service for etcd.
2. Go to the etcd cluster list page, select the region at the top of the page, and find the target etcd cluster in the cluster list below.
3. Go to the snapshot management page, view the policy details and the snapshot file list, check whether the snapshot policy is configured properly, and make sure that the snapshot generation interval and retention period meet your requirements, as shown in the figure below:



4. If you change the etcd password, snapshot files may fail to be generated as expected. In this case, you can submit a ticket to solve this problem.

Abnormal Data Volume

Last updated: 2025-03-19 18:30:12

Description

A data volume error occurred in your etcd cluster. The number of keys or leases may not match expectations, leading to increased operation risk.

Directions

Data Volume Error Troubleshooting

1. If the number of keys in your cluster is large, you can view the statistics of the key count in the cluster and check if it meets business expectations. If there is a significant difference, the business logic may not clear keys as expected, or accidental business writes are performed. Use the command-line tool etcdctl to export all keys for detailed comparison and troubleshooting.

```
etcdctl get "" --from-key --keys-only
```

2. If the number of leases in your cluster is large, you can use the command-line tool etcdctl to view the number of leases and check if it meets business expectations. If there is a significant difference, you can troubleshoot from the business side.

```
etcdctl lease list
```

Suggestions in Case of a Large Data Volume

1. If the number of keys in your cluster exceeds 1,000,000, the operation risk is high. You are suggested to divide them on the business side and distribute them to multiple clusters.
2. If the number of leases in your cluster exceeds 100,000, the operation risk is high. You are suggested to divide them on the business side and distribute them to multiple clusters.

Viewing Statistics of the Cluster Key Count

1. Log in to the console of Cloud Service for etcd.
2. Go to the etcd cluster list page, select the region at the top of the page, and find the target etcd cluster in the cluster list below.
3. Click the icon of the corresponding cluster in the cluster list to enter the **instance detail page**, and then click **Instance monitoring.**

TMP

TMP Overview

Last updated：2024-12-18 17:46:07

Product Overview

TencentCloud Managed Service for Prometheus (TMP) is a highly available, fully managed service built on open-source Prometheus. It is highly integrated with Tencent Kubernetes Engine (TKE) and compatible with various application components in the open-source ecosystem. Combined with the Tencent Cloud Observability Platform (TCOP) alarms and the Prometheus Alertmanager capabilities, TMP provides you with high-efficiency operation capabilities without setup, reducing the development and operation costs. For details, see TMP overview.

Purchase Guide

When you are using TMP, fees of the Prometheus service itself and your actual cloud resource usage will be charged. For details, see Billing Mode and Resource Usage.

Operation Guide

Quick Start

Getting Started: Quickly start using TMP through simple steps.

Cluster Monitoring Instance Management

- Creating a monitoring instance: Create a Prometheus monitoring instance in a TKE cluster.
- Associating with Cluster: Associate an existing TKE cluster with a Prometheus monitoring instance.
- Terminating Instance: Delete the unneeded Prometheus monitoring instances.

Data Collection Configurations

- Data Collection Configurations: Configure Prometheus data collection rules to ensure accuracy and completeness of monitoring data.
- Streamlining Monitoring Metrics: Optimize and streamline the monitoring metrics to reduce unnecessary data collection.
- Free Metrics in Pay-as-You-Go Mode: Understand which monitoring metrics are charged in pay-as-you-go mode and which are free.
- Recommended Common Metrics for TKE: Recommend the common container monitoring metrics to help you better monitor containerized applications.

Non-Container Scenario Monitoring

- Integration Center: Integrate the monitoring data of non-containerized applications to extend the Prometheus monitoring capabilities.

Recording Rule

- Rule Management: Manage the Prometheus recording rules to address the inconvenience of needing to modify the configuration files in native Prometheus.

Alarm Configuration

- Creating Alerting Rule: Create alarm policies to promptly detect and handle the anomalies.
- Disabling Alerting Rule: Disable the unnecessary alarm policies to avoid unwanted alarm notifications.
- Rule Type Description: Understand the different types of alarm policies and their use cases.

Create a monitoring instance

Last updated：2024-12-18 17:46:07

Overview

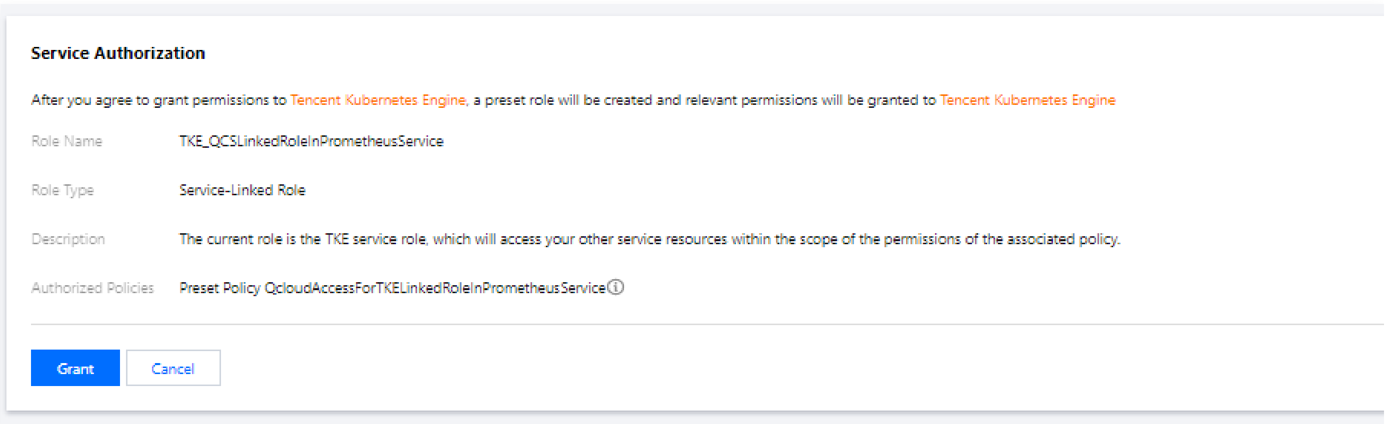
You can easily create a TMP instance and associate it with a cluster in the current region. Clusters associated with the same TMP instance share the same monitoring metrics and alarming policies. Currently, TMP supports managed clusters, self-deployed clusters, serverless clusters and edge clusters. This document describes how to create and manage TMP instances in the TKE console.

Directions

Service authorization

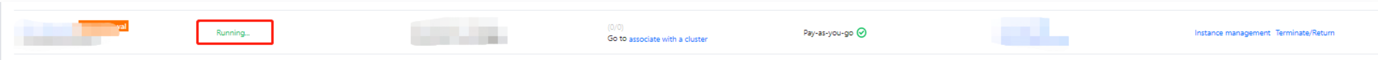
When using TMP for the first time, you need to assign the `TKE_QCSLinkedRoleInPrometheusService` role to the service, which is used to authorize the TMP to access the COS bucket.

1. Log in to the TKE console and click **TMP** in the left sidebar to pop up the **Service authorization** window.
2. Click **Go to Cloud Access Management** to enter the **Role management** page.
3. Click **Grant** to complete authentication.



Creating TMP instance

1. Log in to the TKE console and click **TMP** in the left sidebar.
2. Click **Create** at the top of the instance list page.
3. You will be redirected to the **Tencent Managed Service for Prometheus** page.
4. Purchase an instance as needed. For parameter details, see Creating Instance.
5. Click **Complete**. Now you can click **Associate with TKE** to see the list of TMP instances.
6. You can check the instance creation progress on the page. If the instance status changes to "Running", the instance was successfully created and is running properly.



Note:

If it takes too long to create an instance or the displayed status is abnormal, submit a ticket.

Associating with Cluster

Last updated：2024-12-18 17:46:07

Overview

This document describes how to associate clusters with TMP instances. When the association is established, you can edit configurations such as data collection rules. Currently, the service supports cross-VPC associations, allowing you to monitor clusters in multiple VPCs in different regions within the same TMP instance.

Prerequisites

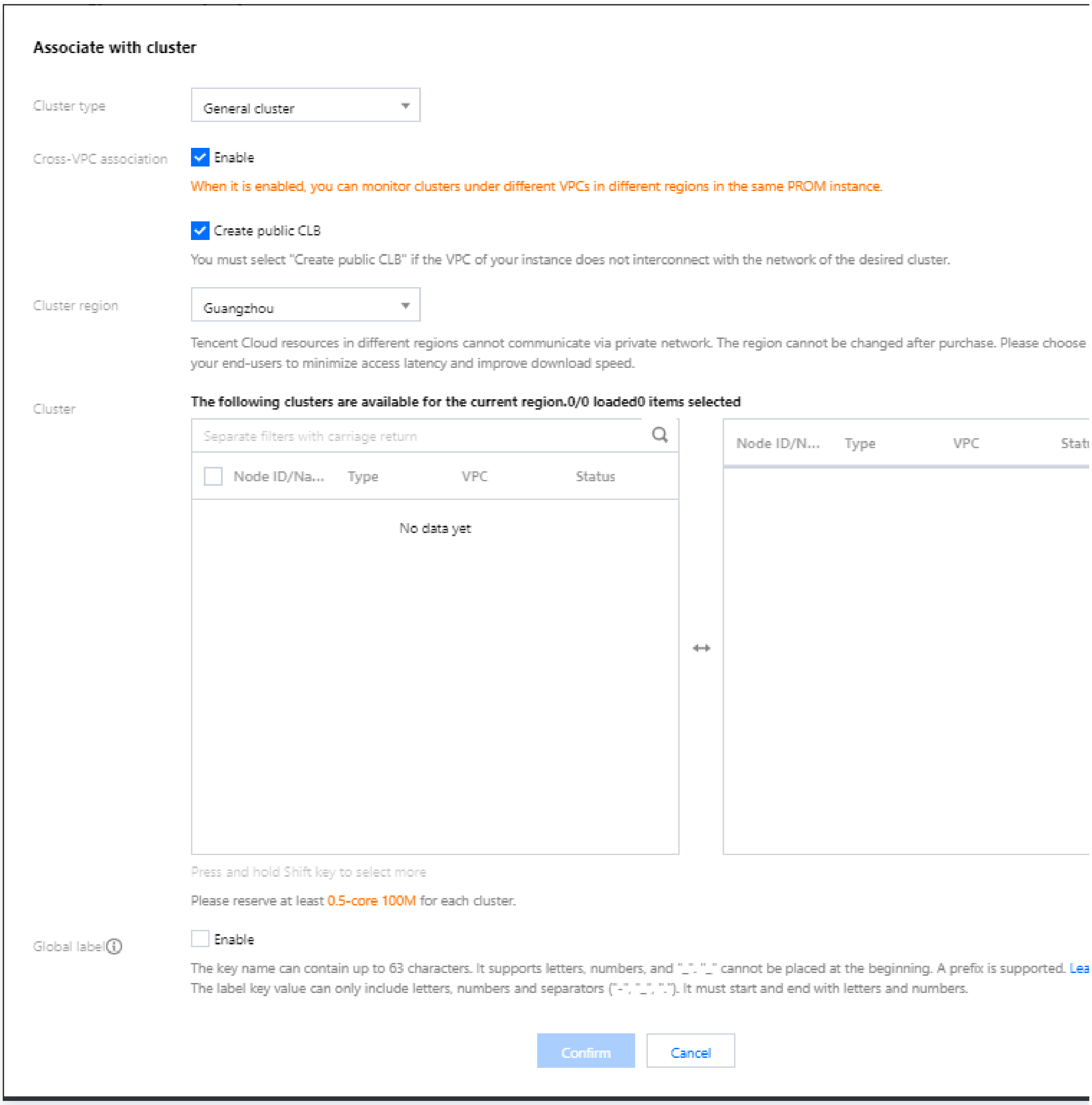
- You have logged in to the TKE console and created a cluster.
- You have created a TMP instance.

Directions

Associating with cluster

> **Note:**
> After the cluster is successfully associated, the monitoring data collection add-on will be installed in the cluster, which will be deleted when the cluster is disassociated.

1. Log in to the TKE console and click **TMP** in the left sidebar.
2. On the instance list page, select the target instance to enter its details page.
3. On the **Cluster monitoring** page, click **Associate with Cluster**.
4. In the **Associate with cluster** pop-up window, select the target cluster.



- **Cluster type**: TKE general clusters, serverless clusters, edge clusters, and registered clusters are supported.
- **Cross-VPC association**: When it is enabled, you can monitor clusters in multiple VPCs in different regions within the same TMP instance.
    - Public CLB: You don't need to create a public CLB instance if your monitoring instance's VPC is connected to the VPC of the cluster you want to associate with; if not, you must select

Create public CLB, otherwise, you cannot collect cluster data across VPCs. For example, if your instance VPC is already connected to the cluster VPC through Cloud Connect Network, you don't need to create a public CLB instance.

- **Region**: Select the region where the cluster resides.
- **Cluster**: Select one or multiple clusters to be associated with.
- **Global tag**: It is used to tag each monitoring metric with the same key-value pair.

5. Click **OK**.

Canceling association

1. Log in to the TKE console and click **TMP** in the left sidebar.
2. On the instance list page, select the target instance to enter its details page.
3. On the **Associate with cluster** page, click **Cancel association** on the right side of the instance.
4. Click **OK** in the **Disassociate cluster** pop-up window.

Data Collection Configurations

Last updated：2024-12-18 17:46:07

Scenarios

This document describes how to configure monitoring collection items for the associated cluster.

Prerequisites

Before configuring monitoring collection items, you need to perform the following operations:

- Create a TMP instance.
- Associate the desired clusters with the TMP instance.

How It Works

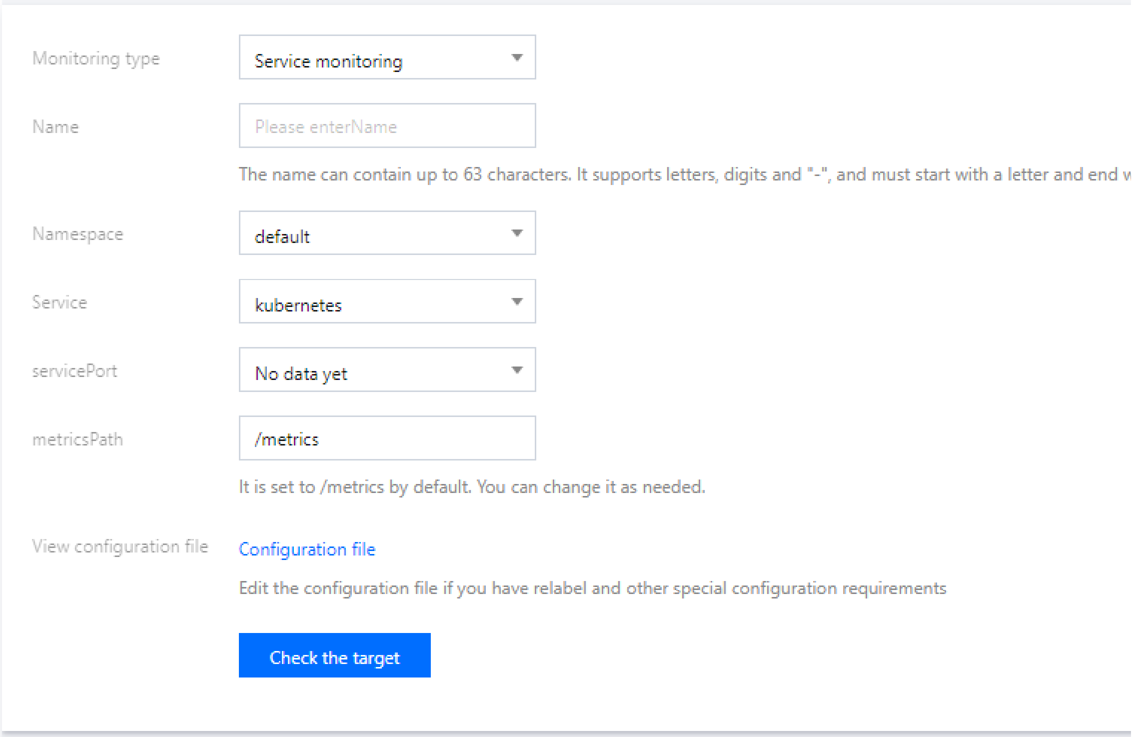
Configuring data collection

1. Log in to the TKE console and click **TMP** in the left sidebar.
2. On the instance list page, select an instance name that needs to configure data collection rules to go to its details page.
3. On the **Cluster monitoring** page, click **Data collection** on the right of the instance to enter the collection configuration list page.
4. On the **Data collection** page, click **Custom monitoring** and add the data collection configuration. TMP has preset some collection configuration files to collect regular monitoring data. You can configure new data collection rules to monitor your business data by using the following two methods:

Adding data collection configuration via the console

Monitoring a service

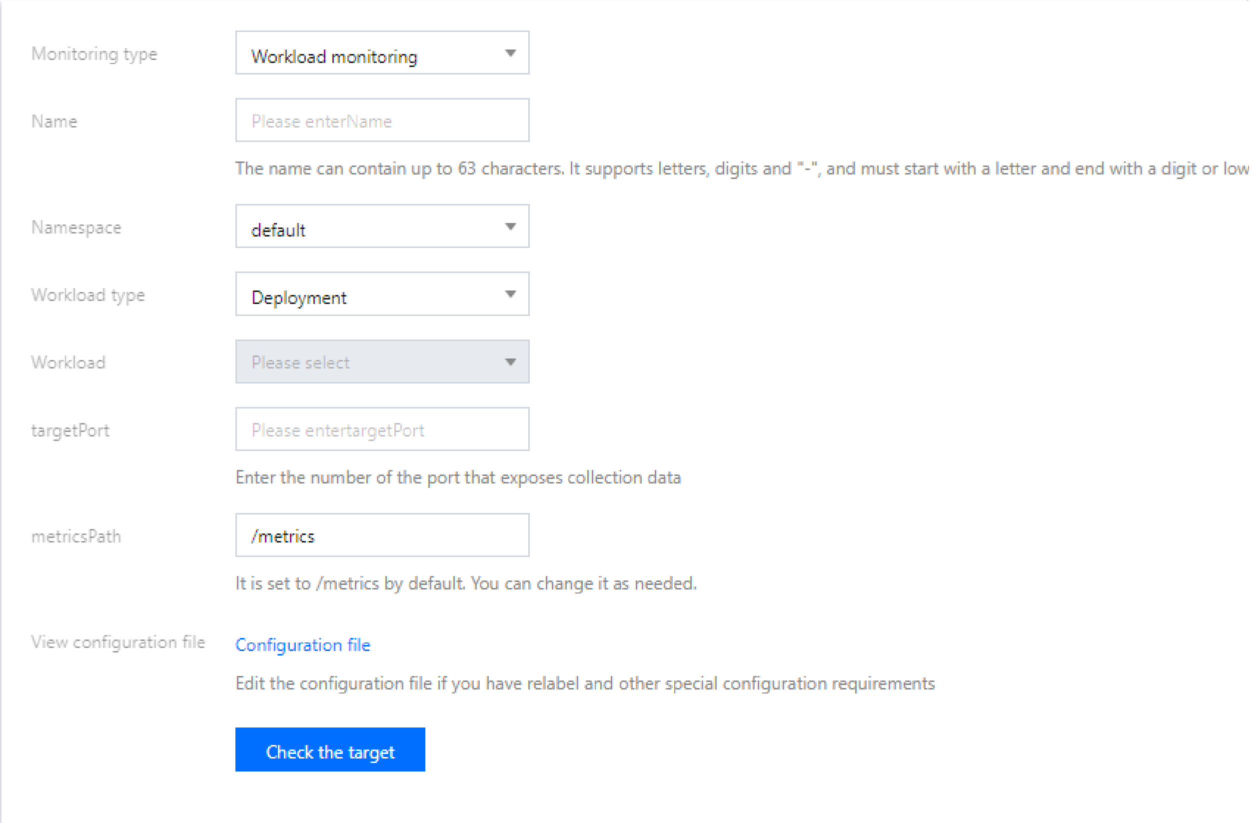
1. Click **Add**.
2. In the **Create collection policy** pop-up window, enter the configuration information.



- **Monitoring type**: Select **Service monitoring**.
- **Name**: enter the rule name.
- **Namespace**: select the namespace to which the Service belongs.
- **Service**: select the service to be monitored.
- **ServicePort**: select the corresponding port.
- **MetricsPath**: defaults to `/metrics`. You can directly enter the collection API as needed.
- **View configuration file**: Click **Configuration file** to view the current configuration file. If you have special configuration requirements such as relabel, you can edit them in the configuration file.
- **Check the target**: Click **Check the target** to view a list of all targets that can be collected under the current collection policy, and confirm whether the collection policy meets your expectations.

Monitoring a workload

1. Click **Add**.
2. In the **Create collection policy** pop-up window, enter the configuration information.

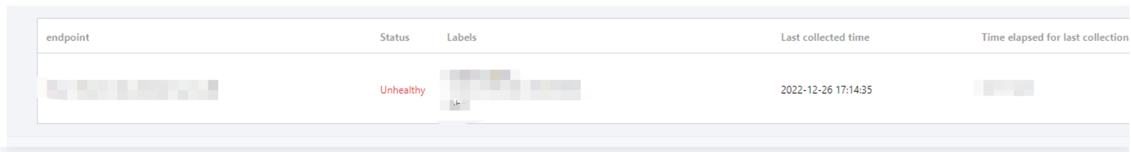


- **Monitoring type**: Select **Workload monitoring**.
- **Name**: enter the rule name.
- **Namespace**: select the namespace to which the workload belongs.
- **Workload type**: Select the workload type to be monitored.
- **Workload**: select the workload to be monitored.
- **targetPort**: enter the target port that exposes the collection metrics through which the collection target can be found. If the port is incorrect, the collection target will not be obtained correctly.
- **MetricsPath**: defaults to `/metrics` . You can directly enter the collection API as needed.
- **View configuration file**: Click **Configuration file** to view the current configuration file. If you have special configuration requirements such as relabel, you can edit them in the configuration file.
- **Check the target**: Click **Check the target** to view a list of all targets that can be collected under the current collection policy, and confirm whether the collection policy meets your expectations.

Adding data collection configuration via yaml

1. Click **Add via YAML**.
2. In the pop-up window, select the monitoring type and enter the relevant configurations. Here takes **Add PodMonitors** as an example.
   You can refer to the use methods in the community to submit the corresponding yaml to complete the data collection configuration.
   - **Workload monitoring**: The corresponding configuration is `PodMonitors` .
   - **Service monitoring**: The corresponding configuration is `ServiceMonitors` .
   - **RawJobs monitoring**: The corresponding configuration is `RawJobs` .

5. Click **OK**.
6. You can view the status of the collection target on the **Data collection** page of the instance.
   targets (3/3) indicates three actually captured targets/three checked collection targets. When the numbe the number of checked targets, the status will be "up", which means that the current capture is normal. W targets is less than the number of checked targets, the status will be "down", which means that some end Click the field value (3/3) to view the details of the collection target. The "down" status is as follows:



   On the **Cluster monitoring** tab of the instance, click **More** > **Target Jobs** on the right of the cluster name cluster.

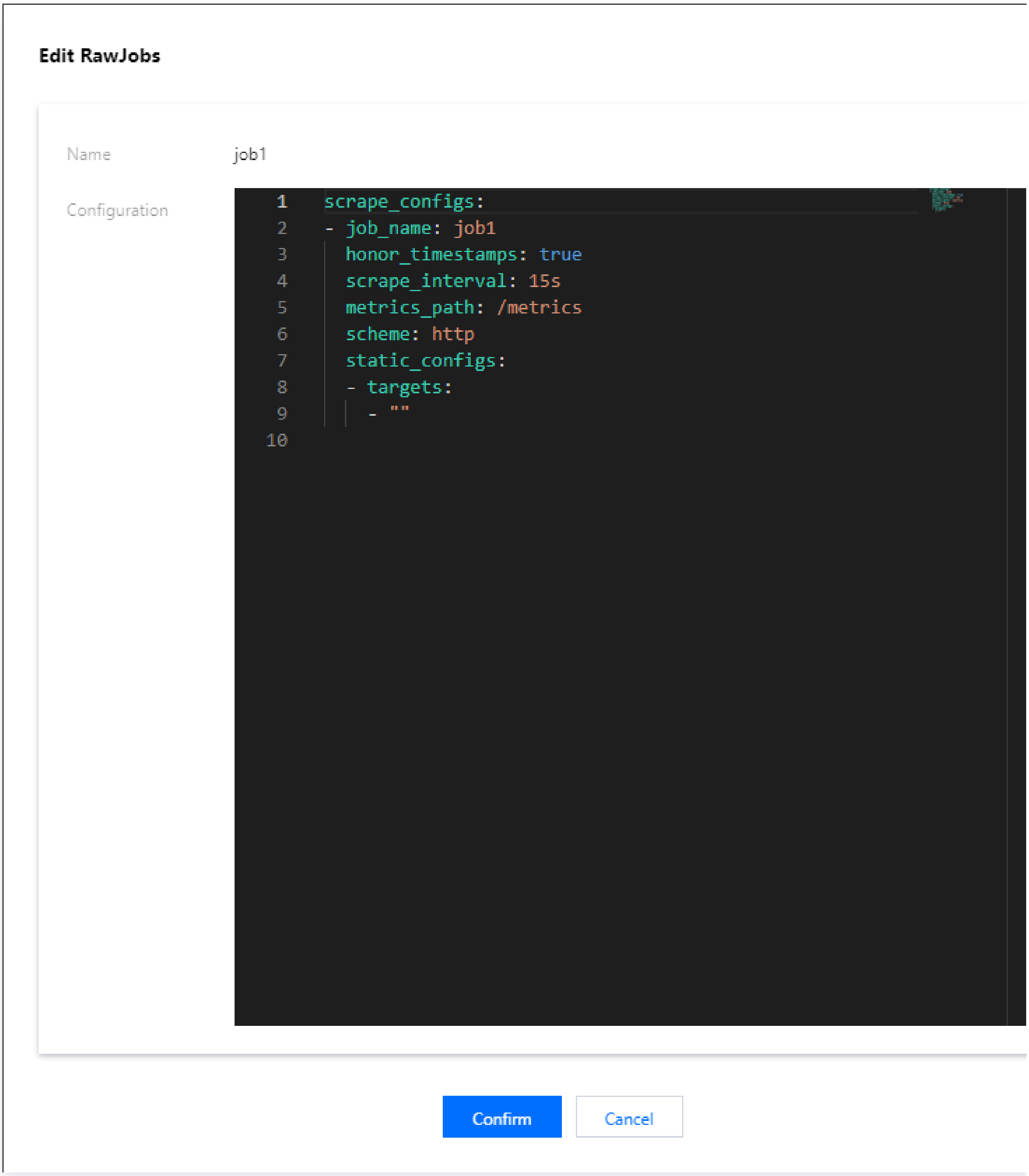
Viewing configuration

> **Note:**
> To view the configured YAML files, you can use only **Custom Monitoring** but not **Basic Monitoring**. All data collection configurations of the basic monitoring are productized, and you only need to **click/select** metrics to add/remove them.

1. Log in to the TKE console and click **TMP** in the left sidebar.
2. On the instance list page, select an instance name that needs to configure data collection rules to go to its details page.
3. On the **Cluster Monitoring** page, select the **Associate with Cluster** tab and click **Data Collection** on the right of the instance to enter the data collection page. Select **Custom Monitoring** and click **Edit** on the

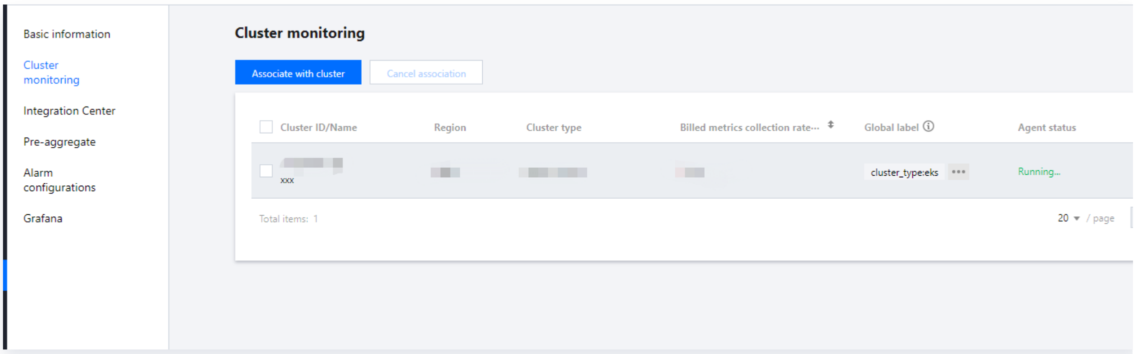
right.

4. In the **Edit RawJobs** pop-up window, view all monitoring metrics configured in the YAML file.

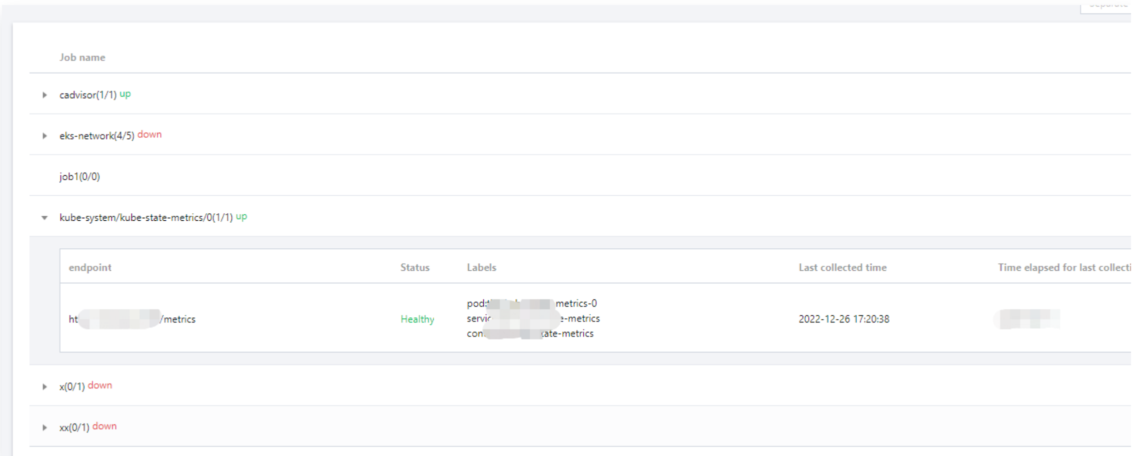


Viewing collection targets

1. Log in to the TKE console and click **TMP** in the left sidebar.
2. On the instance list page, select an instance name that needs to view Targets and go to its details page.
3. On the **Cluster monitoring** page, click **More** > **Target Jobs** in the **Operation** column.



4. On the targets list page, view the collection status of current data.



> **Note:**
> - The endpoints in the status of "Unhealthy" are displayed at the top of the list by default.
> - You can search the target by the resource attributes in the collection target page.

More

Mounting the file to the collector

When configuring the collection item, if you need to provide some files for the configuration, such as a certificate, you can mount the file to the collector in the following way, and the update of the file will be synchronized to the collector in real time.

- **prometheus.tke.tencent.cloud.com/scrape-mount = "true"**
  Add the above label to the configmap under the `prom-xxx` namespace, and all the keys will be mounted to the collector path `/etc/prometheus/configmaps/[configmap-name]/` .
- **prometheus.tke.tencent.cloud.com/scrape-mount = "true"**
  Add the above label to the secret under the `prom-xxx` namespace, and all the keys will be mounted to the collector path `/etc/prometheus/secrets/[secret-name]/` .

Streamlining Monitoring Metrics

Last updated: 2024-12-18 17:46:07

> **Note**
> TMP has adjusted the free storage period for free metrics to 15 days on October 27, 2022. For instances with a storage period of more than 15 days, storage fees for their free metrics will be charged based on the excessive storage period. For more information on the billing rules, see Billing Rules for Free Metrics Exceeding Free Storage Period.

This document describes how to streamline the TMP **collection metrics** to avoid unnecessary expenses.

Prerequisites

Before configuring monitoring collection items, you need to perform the following operations:

- Create a TMP instance.
- Associate the desired clusters with the TMP instance.

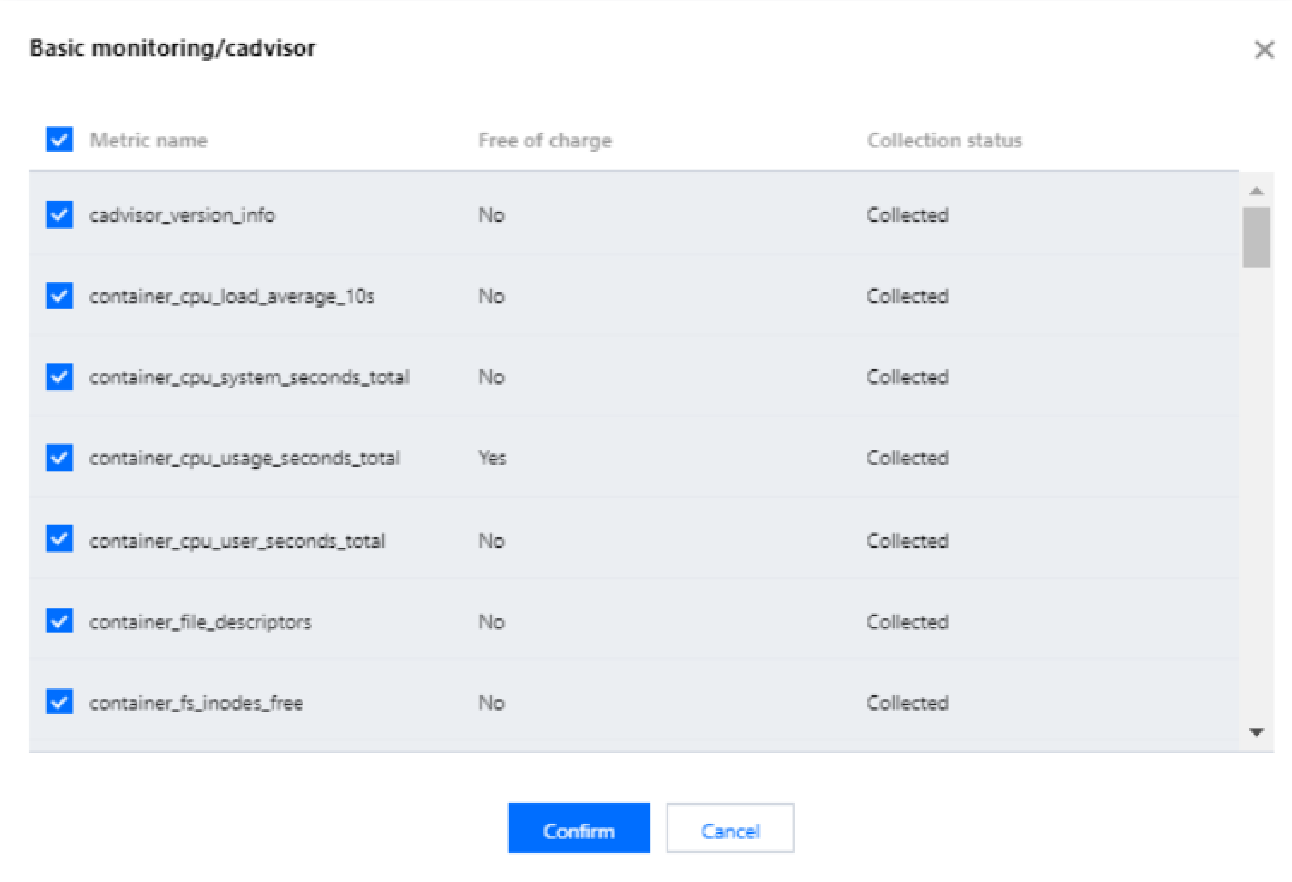
Streamlining Metrics

Streamlining metrics in the console

TMP offers more than 100 free basic monitoring metrics as listed in Free Metrics in Pay-as-You-Go Mode.

1. Log in to the TKE console and click TMP in the left sidebar.
2. On the instance list page, select an instance name that needs to configure data collection rules to go to its details page.
3. On the **Cluster monitoring** page, click **Data collection** on the right of the cluster to enter the collection configuration list page.
4. You can add or remove the basic metrics to be collected by selecting/unselecting the metrics. Click **Metric details** on the right.
5. The following shows whether the metrics are free. If you select a metric, it will be collected. We recommend you deselect paid metrics to avoid additional costs. Only metrics for basic monitoring are free of charge. For more information on free metrics, see Free Metrics in Pay-as-You-Go Mode. For

more information on paid metrics, see Pay-as-You-Go.



Streamlining metrics through YAML

Currently, TMP is billed by the number of monitoring data points. We recommend you optimize your collection configuration to collect only required metrics and filter out unnecessary ones. This will save costs and reduce the overall reported data volume. For more information on the billing mode and cloud resource usage, see here.

The following describes how to add filtering configurations to ServiceMonitors, PodMonitors, and RawJobs to streamline custom metrics.

1. Log in to the TKE console and click TMP in the left sidebar.
2. On the instance list page, select an instance name that needs to configure data collection rules to go to its details page.
3. On the **Cluster monitoring** page, click **Data collection** on the right of the cluster to enter the collection configuration list page.
4. Click **Edit** on the right of the instance to view the metric details.

ServiceMonitor and PodMonitor

A ServiceMonitor and a PodMonitor use the same filtering fields, and this document uses a ServiceMonitor as an example.
Sample for ServiceMonitor:

```
apiVersion: monitoring.coreos.com/v1
kind: ServiceMonitor
metadata:
  labels:
    app.kubernetes.io/name: kube-state-metrics
    app.kubernetes.io/version: 1.9.7
  name: kube-state-metrics
  namespace: kube-system
spec:
  endpoints:
  - bearerTokenSecret:
      key: ""
    interval: 15s # It indicates the collection frequency. You can
increase it to reduce the data storage costs. For example, set it to
`300s` for less important metrics, which can reduce the amount of
monitoring data collected by 20 times.
    port: http-metrics
    scrapeTimeout: 15s # It indicates the collection timeout period.
TMP configuration requires that this value does not exceed the
collection interval, i.e., `scrapeTimeout` <= `interval`.
  jobLabel: app.kubernetes.io/name
  namespaceSelector: {}
  selector:
    matchLabels:
      app.kubernetes.io/name: kube-state-metrics
```

To collect `kube_node_info` and `kube_node_role` metrics, you need to add the `metricRelabelings` field to the endpoint list of the ServiceMonitor. Note that it is **`metricRelabelings`** but not `relabelings` .
Sample for adding `metricRelabelings` :

```
apiVersion: monitoring.coreos.com/v1
kind: ServiceMonitor
metadata:
```



```
  labels:
    app.kubernetes.io/name: kube-state-metrics
    app.kubernetes.io/version: 1.9.7
  name: kube-state-metrics
  namespace: kube-system
spec:
  endpoints:
  - bearerTokenSecret:
      key: ""
    interval: 15s # It indicates the collection frequency. You can increase it to reduce the data storage costs. For example, set it to `300s` for less important metrics, which can reduce the amount of monitoring data collected by 20 times.
    port: http-metrics
    scrapeTimeout: 15s
    # The following four lines are added:
    metricRelabelings: # Each collected item is subject to the following processing.
    - sourceLabels: ["__name__"] # The name of the label to be detected. `__name__` indicates the name of the metric or any label that comes with the item.
      regex: kube_node_info|kube_node_role # Whether the above label satisfies this regex. Here, `__name__` should satisfy the requirements of `kube_node_info` or `kube_node_role`.
      action:  keep # Keep the item if it meets the above conditions, or drop it otherwise.
  jobLabel: app.kubernetes.io/name
  namespaceSelector: {}
  selector:
```

RawJob

If Prometheus' RawJob is used, see the following method for metric filtering.
Sample job:

```
scrape_configs:
  - job_name: job1
```

```
    scrape_interval: 15s # It indicates the collection frequency.
You can increase it to reduce the data storage costs. For example,
set it to `300s` for less important metrics, which can reduce the
amount of monitoring data collected by 20 times.
    static_configs:
      - targets:
          - '1.1.1.1'
```

If you only need to collect `kube_node_info` and `kube_node_role` metrics, add the `metric_relabel_configs` field. Note that it is **`metric_relabel_configs`** but not `relabel_configs`.

Sample for adding `metric_relabel_configs`:

```
scrape_configs:
  - job_name: job1
    scrape_interval: 15s # It indicates the collection frequency.
You can increase it to reduce the data storage costs. For example,
set it to `300s` for less important metrics, which can reduce the
amount of monitoring data collected by 20 times.
    static_configs:
    - targets:
      - '1.1.1.1'
    # The following four lines are added:
    metric_relabel_configs: # Each collected item is subject to the
following processing.
    - source_labels: ["__name__"] # The name of the label to be
detected. `__name__` indicates the name of the metric or any label
that comes with the item.
      regex: kube_node_info|kube_node_role # Whether the above label
satisfies this regex. Here, `__name__` should satisfy the
requirements of `kube_node_info` or `kube_node_role`.
      action: keep # Keep the item if it meets the above conditions,
or drop it otherwise.
```

5. Click **OK**.

Blocking collection targets

Blocking the monitoring of the entire namespace

TMP will monitor all the ServiceMonitors and PodMonitors in a cluster by default after the cluster is associated. If you want to block the monitoring of a namespace, you can add the label of `tps-skip-monitor: "true"` as instructed in Labels and Selectors.

Blocking certain targets

TMP collects monitoring data by creating CRD resources of ServiceMonitor and PodMonitor types in your cluster. If you want to block the collection of the specified ServiceMonitor and PodMonitor resources, you can add the label of `tps-skip-monitor: "true"` to these CRD resources as instructed in Labels and Selectors.

Alarm records

Last updated：2024-12-18 17:46:07

Overview

This document describes how to query the alarm history in cloud native monitoring.

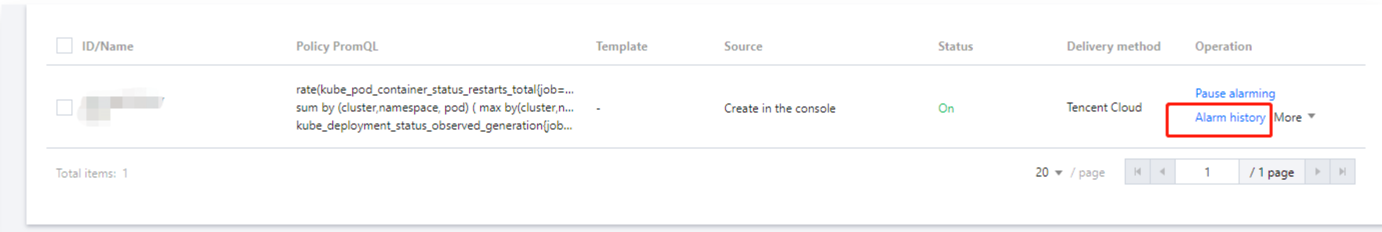
Prerequisites

Before querying alarm history, you need to perform the following operations:

- Create a TMP instance.
- Associate the desired clusters with the TMP instance.
- Configure the information to be collected.
- Configure alarm rules.

Directions

1. Log in to the TKE console and select **TMP** on the left sidebar.
2. On the instance list page, select an instance name that needs to query alarm history to go to its details page.
3. On the **Alarm Configuration** page, select **Alarm history**.



Billing Mode and Resource Usage

Last updated：2024-12-18 17:46:07

Currently, when you use the TMP service, TKE clusters and CLB resources will be created under your account. These resources and TMP are pay-as-you-go. This document describes resource usage details when you use TMP.

Resource List

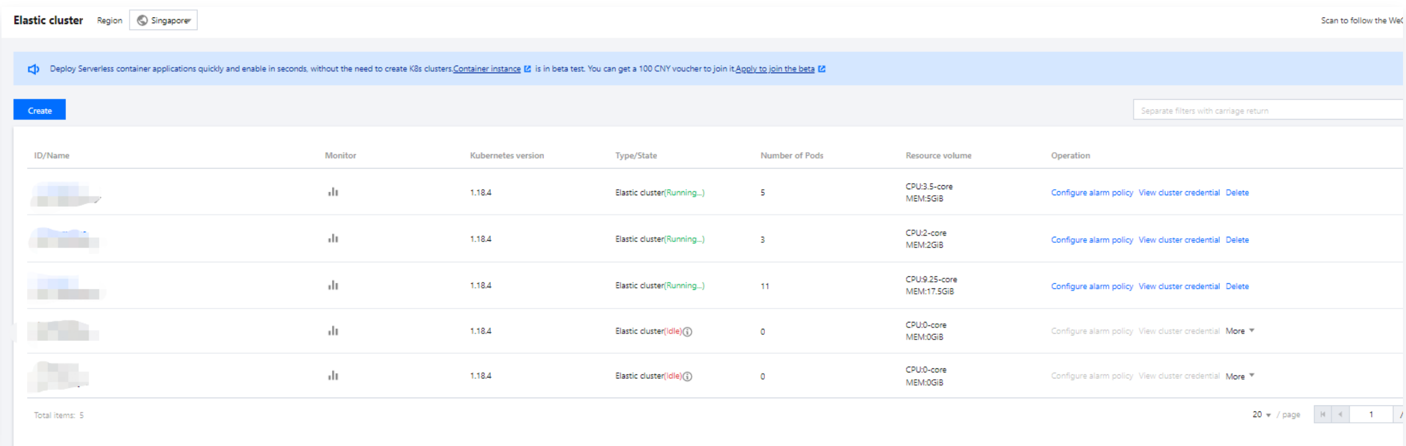
TMP instance

The capability of **Billable Metric Collection Rate** has been launched for TMP, which helps you estimate the cost of monitoring by instance, cluster, target, and metric.

1. Log in to the TKE console and select **TMP** on the left sidebar.
2. In the TMP instance list, view the **Billable Metric Collection Rate**, which indicates the collection rate of billable metrics of a TMP instance and is estimated based on your reported metric data volume and the collection frequency. This value multiplied by 86400 is the number of monitoring data points per day, and you can calculate the estimated published price as instructed in Pay-as-You-Go.
   You can also view the **Billable Metric Collection Rate** under different dimensions on various pages such as **Associate with Cluster**, **Data Collection Configuration**, and **Metric Details**.

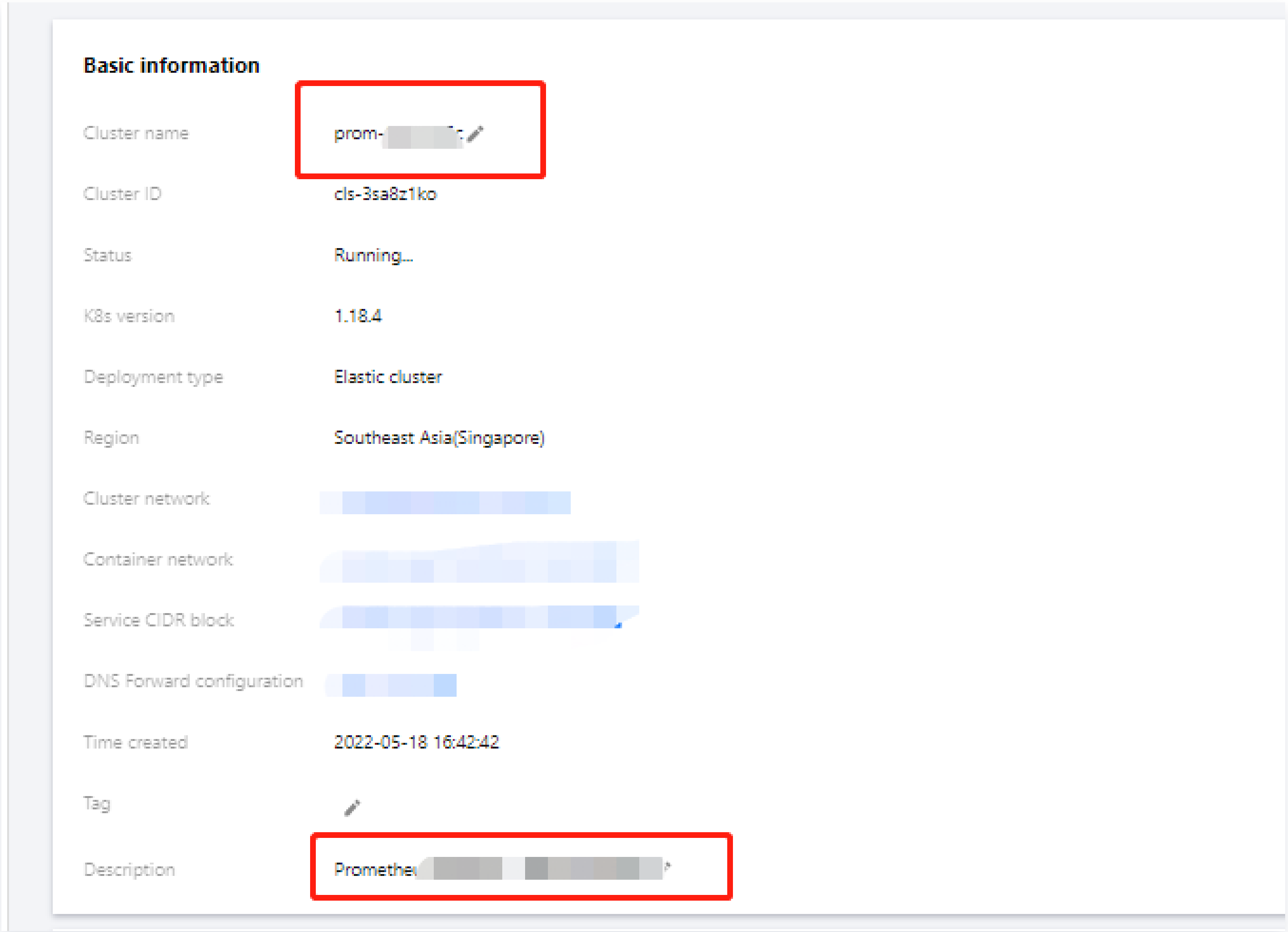
TKE cluster

After each TMP instance is created, a pay-as-you-go TKE cluster will be created under your account for data collection. View the resource information on the elastic cluster list page as shown below:



Notes

The name of the TKE cluster is the TMP **instance ID**, and the cluster description states that "For TMP use only. Do not modify or delete".



Billing

The billing mode is **pay-as-you-go**. For more information, see Product Pricing.
The TKE cluster automatically scales according to the monitoring size. The relationship between the monitoring size and the TKE cluster cost is as shown below:

Reported Instantaneous Series	Estimated TKE Resources Required	Published Price/Day
<500,000	1.25 cores, 1.6 GiB	0.35 USD
1 million	0.5 core, 1.5 GiB*2	1.46 USD
5 million	1 core, 3 GiB*3	2.93 USD
20 million	1 core, 6 GiB*5	7.98 USD
30 million	1 core, 6 GiB*8	12.77 USD

Sample TKE cluster costs are as follows:
The TKE cluster used for a newly initialized TMP instance consumes 1.25 CPU cores and 1.5 GiB memory.
The estimated published price per day is 0.0319 x 24 + 0.0132 x 24 = 1.0824 USD.

CLB

When you use TMP to associate the cluster monitoring container service, a private network CLB instance will be created under your account for network connectivity between the collector and the cluster.
If you associate an edge cluster or another cluster that is not connected, a public network CLB will be created for network connectivity.
To access the Grafana service over the public network, you need to create a public network CLB instance.
These CLB resources will be charged. You can view the resource information of the created public network CLB instances in the CLB console.
This resource is pay-as-you-go. For more information, see Billing for Bill-by-IP Accounts.

Resource Termination

Currently, you cannot delete resources in their respective consoles. For example, when you terminate TMP instances in the TMP console, all relevant resources will also be terminated. Tencent Cloud does not repossess TMP instances proactively. If you no longer use TMP, you need to delete the instances promptly to avoid additional charges.

Terminating Instance

Last updated：2024-12-18 17:46:07

Overview

If you no longer need to use TMP to monitor clusters, you can delete all monitoring instances in the TMP console. The system will automatically uninstall the monitoring add-on and terminate relevant resources.

Directions

1. Log in to the TKE console and select TMP on the left sidebar.
2. On the instance list page, find the target instance and click **Terminate/Return** on the right.
3. In the **Terminate/Return** pop-up window, confirm the instance information and click **OK**.

> **Note:**
> - After an instance is deleted, the TMP console will no longer display its information.
> - After an instance is deleted, its resources (such as monitoring add-on) and configurations will be deleted, the associated cluster will be automatically disassociated and no longer be monitored, and the associated TKE serverless cluster will be deleted.
> - Note that the instance data cannot be recovered after the deletion. Proceed with caution.