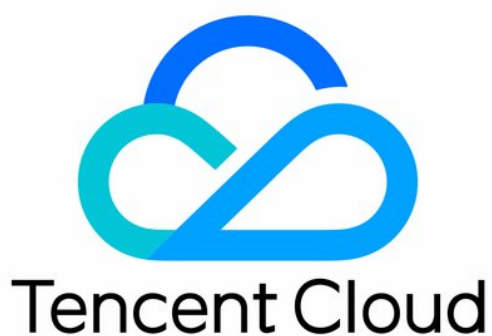


Elasticsearch Service

Elasticsearch Guide

Product Documentation



Copyright Notice

©2013-2024 Tencent Cloud. All rights reserved.

Copyright in this document is exclusively owned by Tencent Cloud. You must not reproduce, modify, copy or distribute in any way, in whole or in part, the contents of this document without Tencent Cloud's the prior written consent.

Trademark Notice



All trademarks associated with Tencent Cloud and its services are owned by Tencent Cloud Computing (Beijing) Company Limited and its affiliated companies. Trademarks of third parties referred to in this document are owned by their respective proprietors.

Service Statement

This document is intended to provide users with general information about Tencent Cloud's products and services only and does not form part of Tencent Cloud's terms and conditions. Tencent Cloud's products or services are subject to change. Specific products and services and the standards applicable to them are exclusively provided for in Tencent Cloud's applicable terms and conditions.

Contents

Elasticsearch Guide

Managing Clusters

- Cluster Status

- Restarting Clusters

- Terminating Clusters

- Advanced Configuration

Access Control

- CAM-Based Access Control Configuration

- ES Cluster

- LDAP Authentication

Multi-AZ Cluster Deployment

Cluster Scaling

- Adjusting Configuration

- Suggestions and Principles for Cluster Specification Adjustment

Cluster Configuration

- Synonym Configuration

- YML File Configuration

- Scenario-based Cluster Template Configuration

- Kona JDK

Plugin Configuration

- Plugin List

- IK Analysis Plugin

- QQ Analysis Plugin

Monitoring and Alarming

- Viewing Monitoring Information

- Configuring Alarms

- Suggestions for Configuring Monitors and Alarms

Log Query

- Querying Cluster Logs

Data Backup

- Automatic Snapshot Backup

- Using COS for Backup and Restoration

Upgrade

- ES Version Upgrade Check

- Upgrading ES Clusters

Elasticsearch Guide

Managing Clusters

Cluster Status

Last updated : 2020-05-13 21:10:01

On the cluster list page and in the basic information section on the cluster details page, you can view the status information of the cluster.





Cluster ListElasticsearch Service User Guide

Guangzhou(2)Shanghai(0)Nanjing(0)Beijing(0)Chengdu(0)Chongqing(0)Hong Kong (China)(0)Singapore(0)Mumbai(0)Seoul(0)

Silicon Valley(0)Toronto(0)Frankfurt(0)

CreateEdit Tag

Separate each search term by pressing the Enter key and separate each keyword with a vertical bar "|". The keywords of cluster tags should be

ID/Name	Status	Cluster Configuration	Health Status ⓘ	AZ	Network	ES Version	Billing Mode	Operation
	Normal	Standard 2 core 4G, 3 100GB SSD cloud disk	Green	Guangzhou Zone 3		6.8.2 Basic edition	Pay-as-you-go Created on 2020-04-13 19:51:53	Kibana CM More
	Normal	Standard 2 core 4G, 3 100GB SSD cloud disk	Green	Guangzhou Zone 3		7.5.1 Platinum edition	Pay-as-you-go Created on 2020-04-13 19:47:47	Kibana CM More

Total 2 Lines

10 Lines per page

1 / 1 page

es-q37swjrf

Kibana CM Upgrade More Help

Basic Configuration

Cluster Monitoring

Cluster Logs

Advanced Configuration

Cluster Change History

Basic Info

Cluster Name

Cluster ID

Cluster Status

Health Status

Region

Network

Elasticsearch Version

X-Pack

AZ and Subnet

Creation Time

Billing Mode

Tag Info

Modify

Cluster Configuration

Adjust Configuration

Node Type	Quantity	Specification	Node Storage	Total Storage
Data Node	3	Standard 2 core 4G ES.S1.MEDIUM4	100GB SSD cloud disk	300GB
Dedicated Master Node	3	Standard 2 core 4G ES.S1.MEDIUM4	50GB Premium Cloud Storage	150GB

Access Control

Username

Password

User Authentication

Elasticsearch Private IP Address

Elasticsearch Public IP Address

Kibana Private IP Address

Kibana Public IP Address

Kibana Public Access Policy

The cluster status is a metric that reflects whether the cluster is being changed or running normally as described below:

Status	Description
Normal	The cluster has been created and can be accessed and used normally with no operations such as configuration change or cluster restart in progress.
Processing	This is the status when an operation such as cluster creation, cluster configuration change, or cluster restart that takes time to complete is being processed. During this period, access to some services will be affected, such as Kibana, data storage, and query.

The health status is one of the most important monitoring metrics in the ES cluster, which is used to indicate whether the cluster is working normally. Health status divides into the following:

Color	Health Status
Green	All master and replica shards are running normally.

Color	Health Status
Yellow	All master shards but not all replica shards are running normally.
Red	Some master shards are not running normally.

For more information, please see [Cluster Health](#).

Restarting Clusters

Last updated : 2019-11-15 14:57:24

Operation scenarios

ES provides a cluster restart feature that allows you to restart your cluster as needed.

Notes

1. Restarting a cluster is time-consuming; therefore, you are not recommended to do so unless necessary.
2. During the restart process, all nodes will be restarted one by one. Therefore, you should ensure that your cluster has at least one replica; otherwise, your cluster will change to RED health status or even incur a risk of data loss.
3. If the health status of your cluster is YELLOW or RED, or there is an index that has no replicas, you will be prompted to force restart your cluster when performing a restart operation, which involves a high risk of data loss. In this case, you are recommended to restore the status of your cluster to GREEN first and create at least one replica for all indices before proceeding. If the status of your cluster cannot be restored, and you still want to restart it after becoming aware of the risk involved in a force restart, you can check the force restart option to restart your cluster.

Directions

1. Log in to the [ES Console](#).
2. You can restart your cluster either on the cluster list page or the cluster details page.
 - On the cluster list page, select a cluster and then select **Operation > More > Restart**.
 - Click the cluster name to enter the cluster details page, and select **More > Restart** in the top-right corner.
3. The following screen will appear after you click **Restart**. After you click **OK**, the cluster status will change to "processing", and the restart operation will be finished once the cluster status is restored to normal.

Cluster restart ×

Cluster ID	Cluster Name
es-l3vgjt9	brown_test

As the current health status of the cluster is not in green or the index does not have replicas, restart is not recommended and may lead to unstable services and even data loss. Please restart after the cluster status returns to normal.

☐ Aware of the risks but still want to force restart

Restart

Cancel

4. If the cluster status is YELLOW or RED, or there is an index in your cluster that has no replicas, a force restart prompt will appear as shown below, and you can only initiate a restart operation by checking "Force Restart". In this case, you can refer to the description in item 3 of [Notes](#). After you initiate the restart operation, the cluster status will change to "processing", and the restart operation will be finished once the cluster status is restored to normal.

Are you sure you want to modify the configuration? ×

The new configuration requires cluster restart to take effect, during which the cluster service will be affected.

- Current state of clusters; restart not recommended, but if necessary, check "Force restart".
- Restart may lead to unstable service or data loss

☐ Force restart

OK

Cancel

Terminating Clusters

Last updated : 2021-07-13 17:13:27

Operation Scenario

If an ES cluster is no longer needed as your business changes, you can terminate it in the console to avoid further fees. If the cluster cannot meet your needs, you can also scale it to the appropriate specifications by adjusting its configuration. For more information, see [Adjusting Cluster Configuration](#).

Note on Fees

- Pay-as-you-go clusters are billed by the actual usage and can be terminated at any time. After the termination, no more fees will be incurred.

Directions

1. Log in to the [ES Console](#).
2. You can terminate a cluster either on the cluster list page or the cluster details page.

- On the cluster list page, select a cluster and select **More > Terminate** in the "Action" column.

Cluster List Elasticsearch Service User Guide

Guangzhou(0) Shanghai(0) Beijing(0) Chengdu(0) Shenzhen Finance(0) Shanghai Finance(0) **Hong Kong(2)** Toronto(0)

Silicon Valley(0) Singapore(0) Mumbai(0) Seoul(0)

[Create](#) Q ↻

ID/N...	Status	Node S...	Nodes	Health	Availabi...	Network	ES Versi...	Billing Type	Operation
es-a0ffht... den...	Normal	1 core 2... 100GB S...	3	Green	Hong Ko...	vpc-ek2... Default...	6.4.3 Platinum...	Pay-as-you-go Created on 2019-06-25 16:13:51	Kibana Cloud Monitor More ▼
es-fgor7... es-clu.	Normal	2 core 4... 100GB S...	3	Green Configur...	Hong Ko...	vpc-rsrp... vpc_doni...	6.4.3 Platinum...	Pay-as-you-go Created on 2019-06-20 23:41:53	Kibana Cl More ▼

共 2 项 Lines per page 10 1 / 1 页

- Click the cluster name to enter the cluster details page and select **More > Terminate** in the top-right corner.

[← es-a0ffht...](#) Kibana Cloud Monitor More ▼ Help

[Basic Configuration](#) [Cluster Monitoring](#) [Cluster Logs](#) [Advanced Configuration](#) [Cluster Change History](#)

Basic Info

Name: dengbuntu_test

Cluster Configuration Name: dengbo

ID: es-a0ffht...

Private Network Address: 172.10.0.129200

Cluster Status: Normal

Health Status: Green

Cluster Configuration

Node Model: ES.S1.SMALL2 (1 core 2G)

Node Storage Type: SSD Cloud Storage

Node Storage: 100GB

Nodes: 3

Total Cluster Storage: 300GB

Dedicated Master Node: Not configured

Restart
Adjust configuration
Terminate

- In the cluster termination dialog box, click **OK** and the system will clear the cluster data and repossess the resources. **Data can not be recovered once cleared.**

Terminate cluster

ID	Name
es-a-0ffbtvg	dengbavu.test

Are you sure you want to terminate this cluster? After termination, the cluster data will be purged and the resources will be reclaimed.

OK

Cancel

Advanced Configuration

Last updated : 2020-05-12 15:01:53

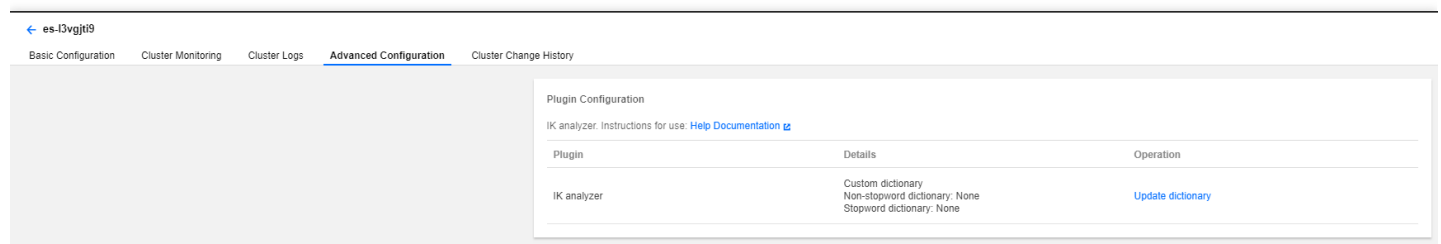
This section describes how to manage the configurations of certain advanced cluster features, such as customizing dictionaries for IK analysis plugin and modifying the `elasticsearch.yml` parameter.

On the cluster list page, click a cluster ID to enter the cluster details page, and click **Advanced Configuration** to enter the advanced configuration management page.

Plugin Configuration

IK analysis plugin

In the plugin configuration section, you can see the IK analysis plugin pre-installed in the cluster. For more information on this plugin, please see [IK Analysis Plugin for Elasticsearch](#). You can use it to create an index of Chinese keywords for your data stored in the ES cluster to implement searches.



The screenshot shows the 'Advanced Configuration' tab selected in a navigation bar. Below the navigation bar, there is a 'Plugin Configuration' section. It includes a link to 'IK analyzer. Instructions for use: [Help Documentation](#)'. Below this, there is a table with three columns: 'Plugin', 'Details', and 'Operation'.

Plugin	Details	Operation
IK analyzer	Custom dictionary Non-stopword dictionary: None Stopword dictionary: None	Update dictionary

Click **Update Dictionary** to enter the dictionary update page, where you can see two options: non-stopwords and stopwords. After selecting the dictionary file to be updated, click **Save** to hot update it (with no cluster restart needed).

- Dictionary file requirements: One word per line in UTF-8 encoding and with .dic extension; one single dictionary file cannot exceed 10 MB in size; up to 10 dictionary file are supported; the rules for stopwords and non-stopwords must be the same, but they cannot use the same filename.
- Dictionary update process: Upload the dictionary file, be sure to click "Save" after upload succeeds, and the dictionary will be updated into the configuration of the IK analyzer and take effect in a short while. Please verify and make sure the dictionary is in effect.
- Viewing dictionary list: "In effect" indicates the dictionary has been uploaded and taken effect in the IK analyzer; "Successfully uploaded" indicates the dictionary has been newly uploaded and will take effect after "Save" is clicked; "Pending upload" indicates the waiting state before uploading and "Upload failed" indicates upload failure due to network or other issues.

Non-stopwords

Words required for full-text index creation and word segmentation

Upload

Filename	Size	Status	Opera...
Click the "Upload" button above or drag and drop files to this area			

Stopwords

Words to be filtered out for full-text index creation and word segmentation such as is, ah and of

Upload

Filename	Size	Status	Opera...
Click the "Upload" button above or drag and drop files to this area			

Save

Cancel

Requirements for dictionary files:

- **Dictionary type:** There are two categories of words: "non-stopwords" and "stopwords". The "non-stopwords" dictionary specifies IK as the analyzer when data is stored to the ES cluster for index creation. If the data stored contains a word in this category, an index will be created and can be queried and found using keywords. "Stopwords" will be deliberately avoided and will not be indexed.
- **Restrictions and requirements:** There are certain restrictions and requirements for dictionary files. For example, a dictionary file should contain one word per line and be encoded in UTF-8. For avoidance of confusion, the name of a non-stopword file should not be the same as that of a stopword file. In addition, as dictionary files will be loaded into the memory, you are allowed to upload a maximum of 10 files of up to 10 MB in size each.
- **Update process:** The list displays the dictionaries that have been uploaded and updated. A new dictionary will be blocked during upload if it does not meet the requirements. After a file is uploaded, it will be displayed in "pending activation" status. After all dictionaries to be updated are uploaded, click **Save**, and they will be saved to your cluster and take effect. If there is a file that fails to be uploaded or is not in UTF-8 format, a failure will be prompted, and you need to delete the failing file before you can click Save for other ones to take effect.

YML configuration

Viewing configuration items

In the cluster details, click **Advanced Configuration** to view the configuration items.

YML Configuration

On-demand advanced configuration options of YML. For the meanings of specific configuration items, see the official documentation of Elasticsearch [Help Documentation](#)

[Batch Modify](#)

Parameter Name	Note	Current value	Value Range Description
action.destructive_requires_name	Deletes indices via wildcards or specifying a...		true or false; default value is true
indices.fielddata.cache.size	Specifies the percentage of java heap spac...		Percentage; format: 1%-100%; default value...
indices.query.bool.max_clause_count	Specifies the maximum number of clauses a...		Positive integer; default value is 1024

Configuration Item	Description	Valid Values
action.destructive_requires_name	Whether it is required to specify the index name when deleting an index	true or false. Default value: true
indices.fielddata.cache.size	Specifies the percentage of Java heap space allocated to the field data	Percentage in the format of 1-100%. Default value: 15%
indices.query.bool.max_clause_count	Specifies the maximum number of clauses allowed in a Lucene Boolean query	Positive integer. Default value: 1024

For more information on the configuration items, please see [Elasticsearch's official documentation](#).

Modifying configuration item

Click **Batch Modify** to modify configuration items. Applicable restrictions are as described above.

YML Configuration

On-demand advanced configuration options of YML. For the meanings of specific configuration items, see the official documentation of Elasticsearch [Help Documentation](#)

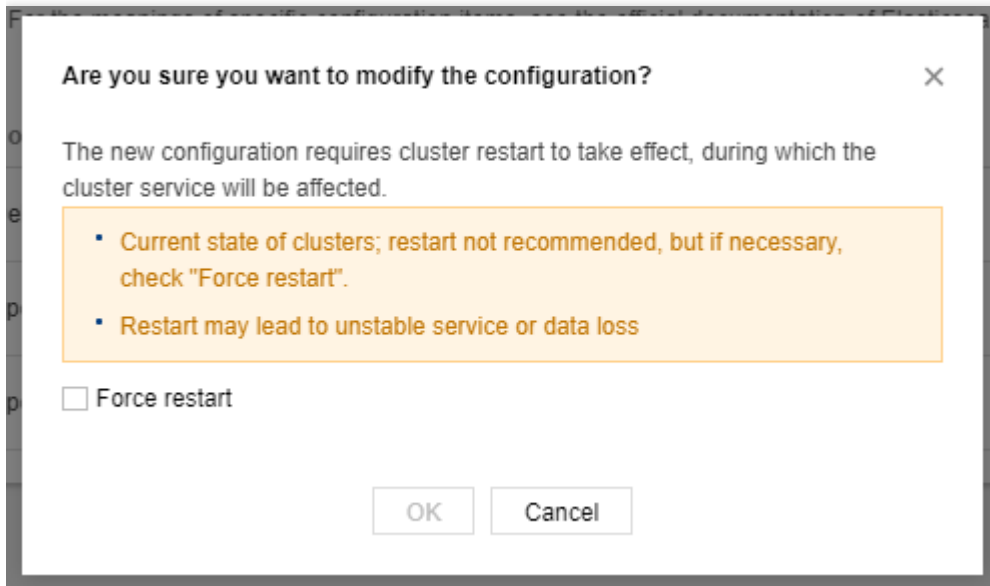
[Batch Modify](#)

Parameter Name	Note	Current value	Value Range Description
action.destructive_requires_name	Deletes indices via wildcards or specifying a...		true or false; default value is true
indices.fielddata.cache.size	Specifies the percentage of java heap spac...		Percentage; format: 1%-100%; default value...
indices.query.bool.max_clause_count	Specifies the maximum number of clauses a...		Positive integer; default value is 1024

If the cluster health status is YELLOW or RED, or there is an index in the cluster that has no replicas, you will be prompted to force restart if you try to modify the configuration items. In this case, there will be a high risk if

you update the configuration, and you are recommended to repair the cluster status first by adding replicas to all indices.

If you are aware of the risk and still want to update the configuration, you can check the "Force Restart" box and click **OK** to restart. For more information, please see the figure below:



Click **Batch Modify** to modify the corresponding configuration items. Applicable restrictions are as described above.

YML Configuration

On-demand advanced configuration options of YML. For the meanings of specific configuration items, see the official documentation of Elasticsearch [Help Documentation](#)

Modify Cancel

Parameter Name	Note	Current value	Value Range Description
action.destructive_requires_name	Deletes indices via wildcards or specifying a...	<input type="text" value="true"/>	true or false; default value is true
indices.fielddata.cache.size	Specifies the percentage of java heap spac...	<input type="text"/> %	Percentage; format: 1%-100%; default value...
indices.query.bool.max_clause_count	Specifies the maximum number of clauses a...	<input type="text"/>	Positive integer; default value is 1024

If you want to customize other configuration items, please [submit a ticket](#).

Access Control

CAM-Based Access Control Configuration

Last updated : 2022-11-09 17:48:39

ES CAM Overview

Cloud Access Management (CAM) is a web-based Tencent Cloud service that helps you securely manage and control access permissions to resources under your Tencent Cloud account. With CAM, you can create, manage, and terminate users (user groups) and use identities and policies to control user access to Tencent Cloud resources. For more information on CAM policies and usage, please see [CAM Policy](#).

ES CAM Policies

General permission policy

ES provides two general policies by default:

- Full access policy (QcloudElasticsearchServiceFullAccess), which grants a user permission to create and manage all ES cluster instances.
- Read-only access policy (QcloudElasticsearchServiceReadOnlyAccess), which grants a user permission to view ES cluster instances but not create, update, or delete them.

You can log in to the [Policy Management](#) page, select "Elasticsearch Service" in "Service Type", and bind the default policies displayed in the list to accounts as needed.

Bind users or user groups with the policy to assign them related permissions.

Create Custom Policy Delete

Elasticsearch

<input type="checkbox"/>	Policy Name	Description	Service Type	Operation
Search "Elasticsearch", 2 results are found. Back to Original List				
<input type="checkbox"/>	QcloudElasticsearchServiceFullAccess	QcloudElasticsearchServiceFullAccess	Elasticsearch Service	Bind User/Group
<input type="checkbox"/>	QcloudElasticsearchServiceReadOnlyAccess	QcloudElasticsearchServiceReadOnlyAccess	Elasticsearch Service	Bind User/Group

Selected 0 items, Total 2 items

Lines per page: 20 1/1

If the default policies cannot meet your requirements, you can click **Create Custom Policy** to customize the authorization.

Custom permission policy

Types of resources that can be authorized in ES include:

Resource Type	Resource Description
instance	<code>qcs::es:\$region:\$account:instance/*</code>

Below describes the details of resource-level access control supported by each API:

API Name	Description	Associated with Resource	Resource Description
Getting cluster list and information of individual clusters	DescribeInstances	Yes	<code>qcs::es:\${Region}:uin/\${ownerUin}:instance/</code>
Creating cluster	CreateInstance	No	<code>*</code>
Updating cluster	UpdateInstance	Yes	<code>qcs::es:\${Region}:uin/\${ownerUin}:instance/</code>
Restarting cluster	RestartInstance	Yes	<code>qcs::es:\${Region}:uin/\${ownerUin}:instance/</code>
Deleting cluster	DeleteInstance	Yes	<code>qcs::es:\${Region}:uin/\${ownerUin}:instance/</code>
Updating plugin	UpdatePlugins	Yes	<code>qcs::es:\${Region}:uin/\${ownerUin}:instance/</code>

Supported regions include:

Region	Name	Region ID
South China	Guangzhou	<code>ap-guangzhou</code>
East China	Shanghai	<code>ap-shanghai</code>
	Nanjing	<code>ap-nanjing</code>

Region	Name	Region ID
North China	Beijing	ap-beijing
Southwest China	Chengdu	ap-chengdu
	Chongqing	ap-chongqing
Hong Kong/Macao/Taiwan	Hong Kong (China)	ap-hongkong
Southeast Asia Pacific	Singapore	ap-singapore
South Asia Pacific	Mumbai	ap-mumbai
Northeast Asia Pacific	Seoul	ap-seoul
	Tokyo	ap-tokyo
West US	Silicon Valley	na-siliconvalley
East US	Virginia	na-ashburn
North America	Toronto	na-toronto
Europe	Frankfurt	eu-frankfurt

The syntax of a custom policy is as follows:

```
{
  "version": "2.0",
  "statement": [
    {
      "action": [
        "Action"
      ],
      "resource": "Resource",
      "effect": "Effect"
    }
  ]
}
```

- Action: replace it with the operation to be allowed or denied.
- Resource: replace it with the resources that you want to authorize the user to manipulate.
- Effect: replace it with "allow" or "deny".

ES currently supports access control management for all APIs except `DescribeInstances` . You can authorize a sub-account to perform various operations on a cluster under your account such as updating, restarting, and deleting.

Custom permission sample

To grant an account permission to update the specified cluster, use the following policy syntax:

```
{
  "version": "2.0",
  "statement": [
    {
      "action": [
        "es:Describe*"
      ],
      "resource": [
        "qcs::es:ap-guangzhou:uin/$uin:instance/$instanceID"
      ],
      "effect": "allow"
    },
    {
      "action": [
        "vpc:Describe*",
        "vpc:Inquiry*",
        "vpc:Get*"
      ],
      "resource": "*",
      "effect": "allow"
    },
    {
      "action": [
        "monitor:*",
        "cam:ListUsersForGroup",
        "cam:ListGroups",
        "cam:GetGroup"
      ],
      "resource": "*",
      "effect": "allow"
    },
    {
      "action": [
        "es:Update*"
      ],
      "resource": [
        "qcs::es:ap-guangzhou:uin/$uin:instance/$instanceID"
      ],
      "effect": "allow"
    }
  ]
}
```

```
]
}
```

ES Cluster

Last updated : 2020-08-03 11:27:33

ES clusters are deployed in logically isolated VPCs, giving you full control over your environment configuration and the ability to customize network access control lists (ACLs) and security groups. In addition, to help ensure the security of your resources in the cloud, a wide variety of security capabilities are provided, including:

- CAM for resources under Tencent Cloud account (for more information, please see [CAM-Based Access Control Configuration](#))
- ES cluster access password/user authentication
- IP blocklist/allowlist for public network access to Kibana (you can also enable only private network access to Kibana)
- Control over public network access to ES clusters and IP allowlist
- Role-based access control (RBAC)

Setting ES Cluster Access Password

When creating an ES cluster, you will be asked to set a password for the default user `elastic`. The account and password will be used to log in to the Kibana page. If [ES cluster user authentication](#) has been enabled for your cluster, then they will be used for ES cluster login authentication for stricter security protection as show below:

Username	<code>elastic</code> <small>Used for Kibana login and user authentication. (Note: the open source edition does not has the user login authentication feature; basic edition above V6.8 can enable the feature if needed, and platinum edition enables this feature by default.)</small>
Password	<input type="password" value="Enter the password"/> <small>8-16 characters, including at least three out of the following four types of characters: [a-z], [A-Z], [0-9] and [-!@#\$\$%^&*+=_.,:~?]</small>
Confirm Password	<input type="password" value="Enter the password again"/>

Resetting ES Cluster Access Password

You can use the password resetting feature on the cluster details page to reset the password of the `elastic` account for your ES cluster as shown below:

The screenshot displays the Tencent Cloud Elasticsearch Service console. At the top, there are tabs for 'Basic Configuration', 'Cluster Monitoring', 'Cluster Logs', 'Advanced Configuration', and 'Cluster Change History'. The 'Basic Configuration' tab is active, showing a 'Basic Info' section with details like Cluster Name, ID, Status (Normal), Health Status (Green), Region (South China(Guangzhou)), Network, Elasticsearch Version (7.5.1), X-Pack (Platinum edition), and AZ and Subnet (Guangzhou Zone 3). To the right, the 'Cluster Configuration' section shows a table of node types and their specifications. Below this, the 'Access Control' section includes fields for Username (elastic) and Password, with a 'Reset' button next to the password field.

Node Type	Quantity	Specification	Node Storage	Total Storage
Data Node	3	Standard 2 core 4G ES.S1.MEDIUM4	100GB SSD cloud disk	300GB
Dedicated Master Node	3	Standard 2 core 4G ES.S1.MEDIUM4	50GB Premium Cloud Storage	150GB

Access Control	
Username	elastic
Password	Reset

Setting IP Blocklist/Allowlist for Public Network Access to Kibana

If the Kibana page can be accessed over the public network, ES provides IP blocklist/allowlist in addition to password-based authentication for Kibana access, further enhancing the access security of you clusters.

- Configuration rule: up to 10 IPs in the format of `192.168.0.1` or `192.168.0.0/24` separated by commas are supported.
- Blocklist/allowlist settings: you can set either of them. If both are configured, the allowlist shall prevail. The configuration items are as shown below:

Access Control

Username	elastic
Password	Reset
User Authentication?	Enabled
Elasticsearch Private IP Address	192.168.15.100
Elasticsearch Public IP Address	<input type="checkbox"/>
Kibana Private IP Address	<input type="checkbox"/>
Kibana Public IP Address	<input checked="" type="checkbox"/> https://es-

Enabling Only Private Network Access to Kibana

If you have concerns over the security of public network access, you can disable it and enable only private network access.

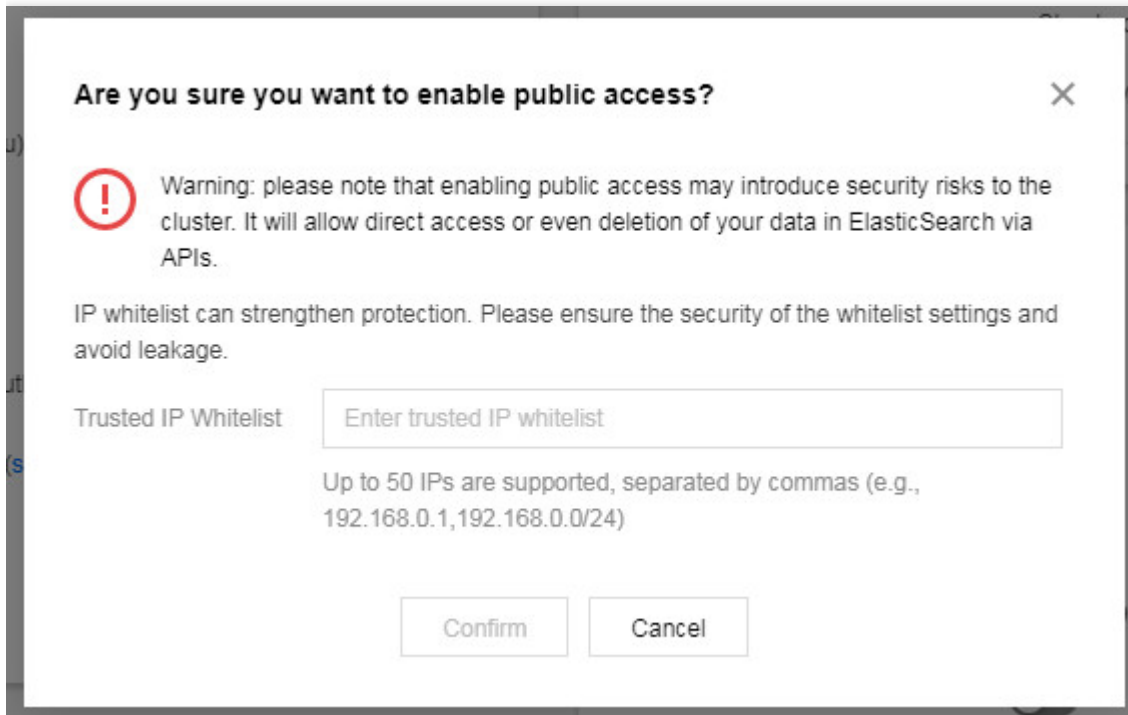
Access Control

Username	elastic
Password	Reset
User Authentication?	Enabled
Elasticsearch Private IP Address	192.168.19.100
Elasticsearch Public IP Address	<input type="checkbox"/>
Kibana Private IP Address	<input checked="" type="checkbox"/>
Kibana Public IP Address	<input checked="" type="checkbox"/> https://es-xxxxx.us-east-1.amazonaws.com
Kibana Public Access Policy	No data ✎

Enabling Limited Public Network Access to ES Cluster and Setting IP Allowlist

For the sake of security, access to ES clusters over the public network is disabled by default. For clusters having [ES cluster user authentication](#) enabled, you can enable access over the public network for convenience, but you need to

set the IP allowlist for security protection.

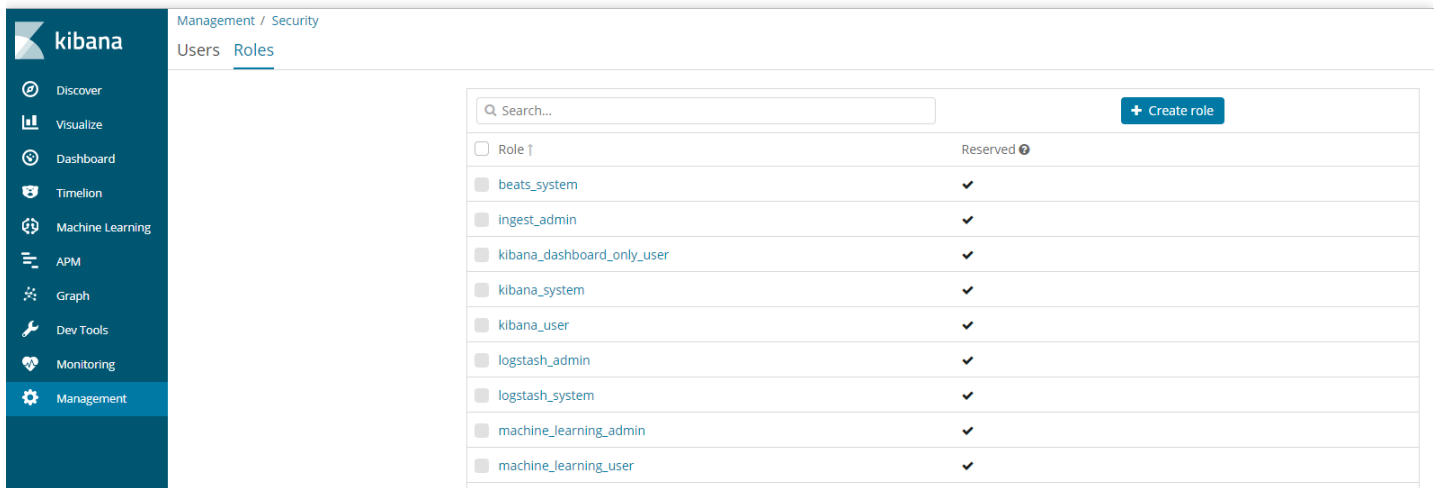


Role-Based Access Control (RBAC)

For clusters having [ES cluster user authentication](#) enabled, you can use more security management features. In addition, the Platinum Edition offers more refined access control by document or field. For more information, please see [Role-based access control](#) at Elasticsearch official website.

Role management

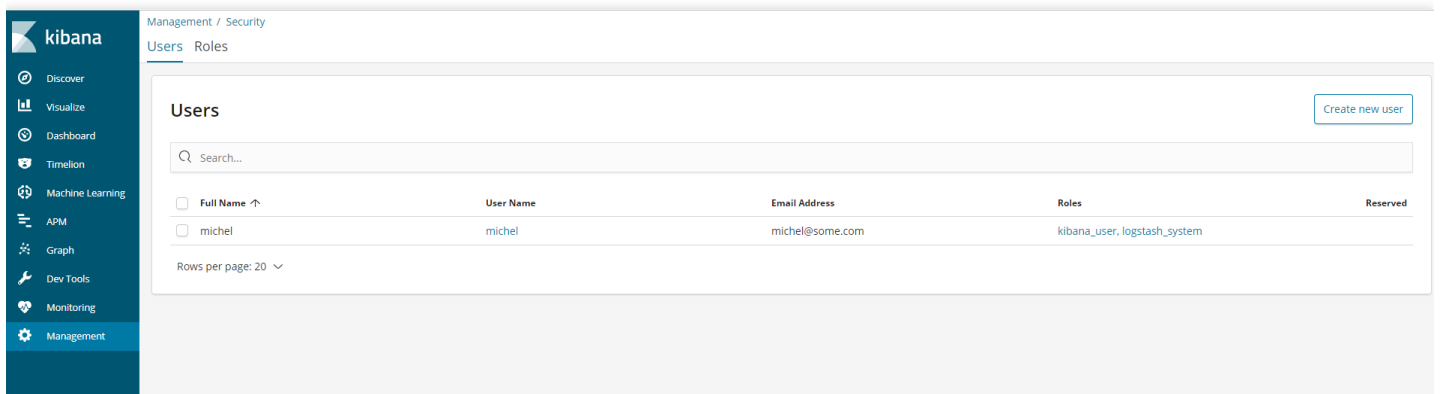
You can create, modify, and delete roles with different permissions in **Management > Security > Roles** on the Kibana page as shown below:



User management

You can create, modify (information, password, etc.), and delete users with multiple roles in **Management > Security > Users** on the Kibana page as shown below:

The password of the default ES user `elastic` can be reset only in the console on the official website.



For more information on how to use relevant security features, please see the following:

- [Protect your data in the Elastic Stack](#)
- [Kibana X-Pack Security](#)
- [Elasticsearch Security APIs](#)

LDAP Authentication

Last updated : 2022-06-27 14:17:44

This document describes how to configure Lightweight Directory Access Protocol (LDAP) authentication in ES, so that users with LDAP roles can access ES resources.

Use Limits

LDAP authentication is an advanced feature provided by Elasticsearch [X-Pack](#) and is currently only supported for Platinum Edition clusters. If you want to use this feature for other clusters, upgrade them to Platinum Edition first.

Configuring LDAP Authentication

1. Log in to the [ES console](#) and click **Cluster Name** of the target cluster to enter its **Basic Configuration** page.
2. In the **Access Control** module, click the edit icon after **Authentication** to enter the authentication configuration page.

Access control

Username	elastic
Password	Reset
User authentication ⓘ	Enabled
Identity verification ⓘ	Not set
Private access address	<input type="checkbox"/> <input type="checkbox"/>
Security group settings	No data
Public access address	<input type="checkbox"/>

3. Set configuration items.
 - url: LDAP server address, which must start with "ldap://" followed by the domain name or IP address. Make sure that the entered URL is accessible over the private network in your VPC; otherwise, this configuration will not

take effect.

- `bind_dn`: Member DN used for LDAP server authentication, which must be in the DN hierarchical syntax structure (e.g., `cn=admin,dc=husor,dc=com`) and can contain up to 200 characters.
- `bind_password`: LDAP server connection password, which can contain 6–63 letters, digits, or special symbols `-!@#$%&^*+=_~:;, .? .`.
- `user_search.base_dn`: Base DN used to search for bound LDAP members, which must be in the DN hierarchical syntax structure (e.g., `ou=HusorSSO,ou=People,dc=husor,dc=com`) and can contain up to 200 characters.
- `user_search.filter`: Search filter, by which the system will filter eligible binding relationships, such as `(uid={0})`.
- `group_search.base_dn`: Base DN used to search for bound LDAP user groups, which must be in the DN hierarchical syntax structure (e.g., `ou=HusorSSO,ou=People,dc=husor,dc=com`) and can contain up to

200 characters.

Select an identity verification method LDAP ✕

url *

ldap://

Supports only the IP + port format, such as 192.168.1.1:8080

Ensure the provided URL is accessible over the private network in your VPC. Otherwise, the configuration will not take effect.

bind_dn *

cn=admin,dc=husor,dc=com

LDAP bind_dn, such as cn=admin,dc=husor,dc=com

bind_password *

LDAP password

It must contain 6–63 characters in a combination of letters, digits, and symbols (-!@#\$\$%^*+=_.,;,.?).

user_search.base_dn *

ou=HusorSSO,ou=People,dc=husor,dc=com

LDAP user_search.base_dn, such as ou=HusorSSO,ou=People,dc=husor,dc=com

user_search.filter *

User search filter, such as (uid={0})

An LDAP filter, such as (uid={0})

group_search.base_dn *

ou=HusorSSO,ou=People,dc=husor,dc=com

LDAP group_search.base_dn, such as ou=HusorSSO,ou=People,dc=husor,dc=com

Confirm

Cancel

- After confirming that everything is correct, click **OK**. In the pop-up window, read the notes and click **OK**. Then, the cluster will be restarted. You can view the change progress in **Cluster Change History**.

Note**Modify configuration will restart the cluster. Continue now?**☐

This operation will restart the cluster. It is recommended to operate when the cluster load is low.

[Confirm](#)[Cancel](#)

5. After successful configuration, you will see **LDAP enabled** after **Authentication**. To modify the LDAP authentication configuration, click **LDAP enabled** to make changes.

Access control

Username elastic

Password [Reset](#)

User authentication ⓘ Enabled

Identity verification ⓘ [LDAP enabled](#) [Close](#)

Private access address

Security group settings No data

Public access address

Public access policy

Configuring LDAP Role Permission

1. After LDAP authentication is configured, an LDAP user has not been granted any permissions and cannot access the ES cluster/Kibana via LDAP. Therefore, you need to perform role mapping for the LDAP user in Kibana.
2. Log in to the [ES console](#), go to the Kibana homepage of the cluster, and enter the **Dev Tools** page of Kibana.
3. Create the mapping between the LDAP user and role as instructed in [Create or update role mappings API](#). The example below grants the LDAP user **zhangsan** the permissions of the **superuser** role.

```
POST _xpack/security/role_mapping/ldap_role_mapping_zhangsan?pretty
{
  "roles": [ "superuser" ],
  "enabled": true,
  "rules": {
    "any": [
      {
        "field": {
          "username": "zhangsan"
        }
      }
    ]
  }
}
```

Disabling LDAP Authentication

1. Log in to the [ES console](#) and click **Cluster Name** of the target cluster to enter its **Basic Configuration** page.
2. In the **Access Control** module, click **Disable** after **Authentication**. In the pop-up window, read the notes and click **OK**. Then, the LDAP disabling operation will start, and the cluster will be restarted. You can view the change progress in **Cluster Change History**.

Notes

- If the cluster health status is YELLOW or RED, there will be a greater risk in modifying the configuration. You must fix the cluster status problem first before proceeding.
- If there is an index with no replicas in the cluster, you will be prompted to force restart the cluster when you try to modify the configuration. In this case, there will be a greater risk in the operation, and some data may be temporarily inaccessible. We recommend you add replicas for all indexes first before proceeding.

Multi-AZ Cluster Deployment

Last updated : 2021-08-11 11:17:22

Creating a Cluster That Supports Multiple AZs

Go to the [ES purchase page](#), select **AZ Deployment Mode**, and set up a multi-AZ network.

Note :

- To enable multi-AZ disaster recovery, a cluster must first have at least three dedicated master nodes enabled; therefore, there must be at least three data centers selected where multi-AZ disaster recovery is supported. Currently, this feature is only supported in certain major regions such as Beijing, Shanghai, and Guangzhou, and it will be gradually rolled out in other regions as new Tencent Cloud data centers are constructed.
- Other parameter settings are generally the same as those for single-AZ deployment.

Taking the Shanghai region as an example, in **AZ Deployment Mode**, you can create 2-AZ and 3-AZ clusters, and you need to select AZs and subnets equal to the number of AZs.

Network	<div>Select a network</div>	<div>⌵</div>	<div>↻</div>	
	If the existing VPCs do not meet your requirements, you can go to create a VPC .			
AZ Deployment Mode	<div>Single-AZ</div>	<div>Multi-AZ</div>	<div>?</div>	
AZ and Subnet	<div>Select AZ</div>	<div>⌵</div>	<div>Select subnet</div>	<div>⌵</div>
	<div>↻</div> There is no valid subnet in the current AZ. Please change to another AZ or create one.			
	<div>Select AZ</div>	<div>⌵</div>	<div>Select subnet</div>	<div>⌵</div>
	<div>↻</div> There is no valid subnet in the current AZ. Please change to another AZ or create one.			
	If the existing subnets do not meet your requirements, you can go to the console to create a subnet .			

The **number of data nodes** will be automatically adjusted proportionally to the number of AZs. In order to ensure the stability and reliability of the cluster, **Dedicated Master Node** is **enabled** by default when you select multiple AZs, and you can select 3 or 5 **dedicated master nodes**. Such nodes will be evenly distributed among the three AZs to ensure that when an AZ becomes unavailable, the circumstance where more than half of the nodes are unavailable will never happen. Plus, this guarantees that the cluster always has electable nodes which form a quorum for the

election of a master node, thus ensuring cluster reliability.

Dedicated Master Node	<input checked="" type="checkbox"/> Enable Insufficient resources. Please submit a ticket to give feedback. Dedicated master node is a type of node in Elasticsearch clusters. It does not store data and is designed to guarantee cluster stability. Details
Node Specs	<div>No data</div> ⓘ
Node Qty	<div>3</div>
Kibana Node	Standard ES.S1.SMALL2-1 core 2 GB Free of charge and modification is currently not supported

How Multi-AZ Disaster Recovery Works

Data node

To ensure that multi-AZ disaster recovery takes effect, you need to follow the principles below:

1. The number of data nodes in the cluster you purchase should be equal to a multiple of the number of AZs. For example, if you select two AZs for disaster recovery, the number of data nodes should be 2, 4, 6, 8, and so on.
2. At least one replica should be set for an index shard to ensure that the cluster always has more than two replicas of data.

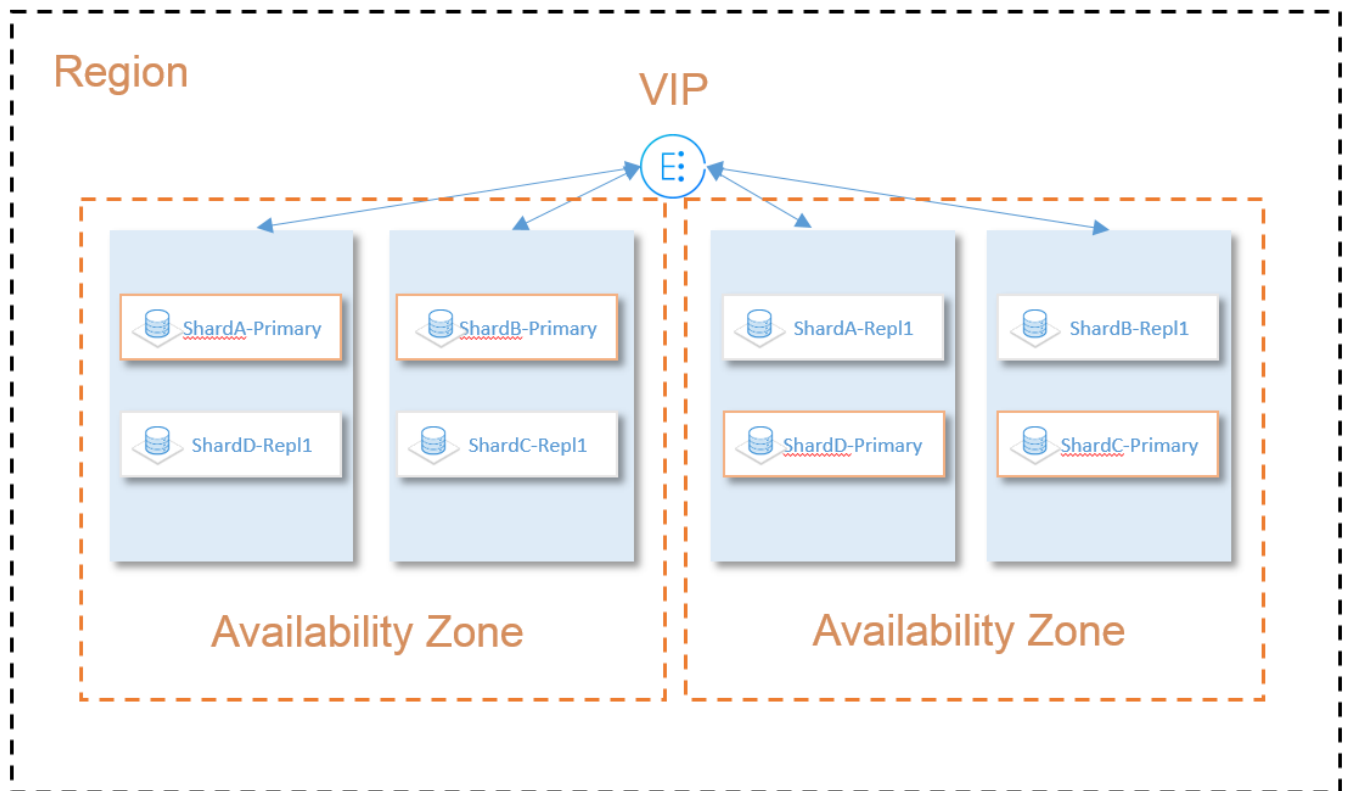
ES automatically deploys the purchased data nodes evenly among the selected AZs, and the deployed data nodes can perceive and identify the AZs. This feature distributes the replicas of your data across multiple AZs so as to make sure that there is only one replica in a single AZ.

ES offers load balancing within VPC, which allows you to connect to the cluster through the provided VIP for reading and writing data and controlling the cluster through ES APIs.

- This VIP is bound to all data nodes in the cluster and offers the load balancing capability, so that all requests initiated by you will be evenly distributed across all the data nodes.
- This VIP also features health check. If it is determined that a node is not responding in multiple checks within a certain period, the health checker will temporarily remove the problematic node from the list of nodes bound to the VIP until it returns to normal.

This makes sure that when a node is down, or an AZ of a data center is unavailable, the problematic node will be automatically removed, so that it will not be requested by the client. This helps implement imperceptible failover in

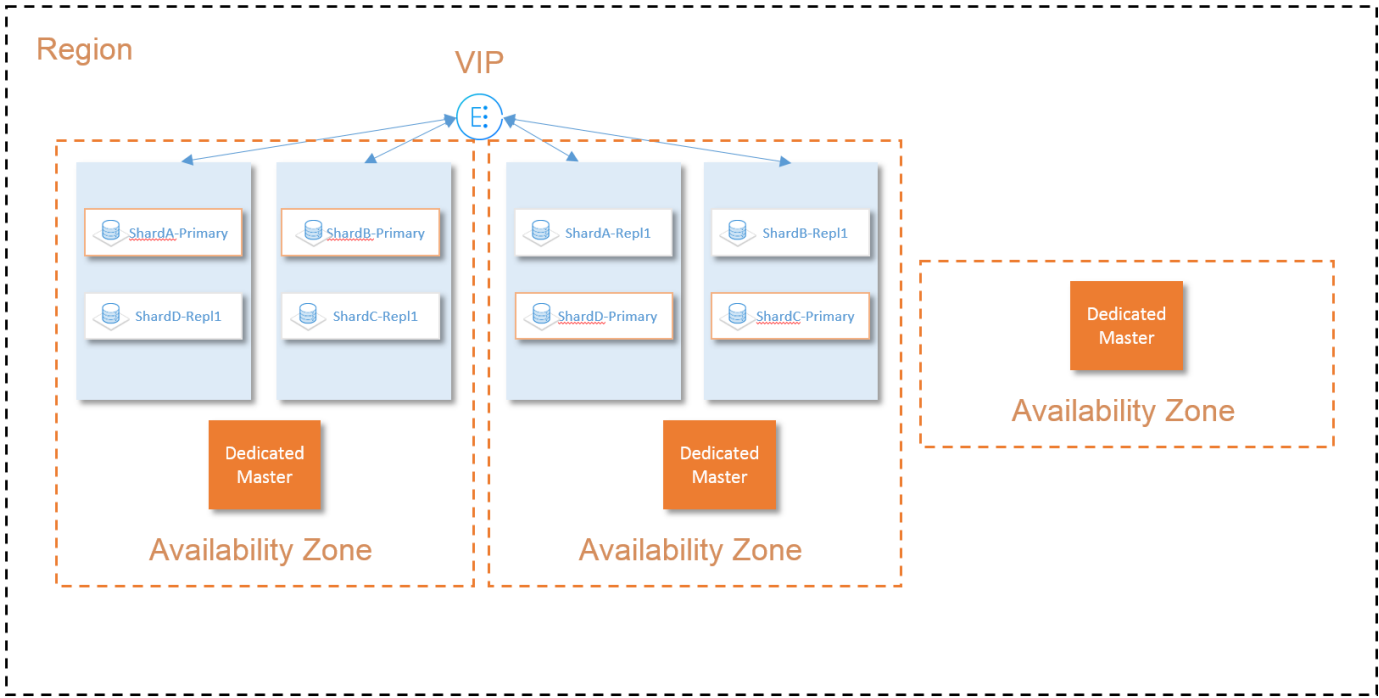
case of an AZ failure, thus improving the stability of your business.



Dedicated master node

In order to improve the reliability of your cluster, you must create at least three dedicated master nodes when using multi-AZ disaster recovery, and they must be distributed in three different AZs. Even if you select two AZs to deploy data nodes, an additional one will be automatically chosen to deploy a dedicated master node. This deployment scheme helps ensure that when an AZ becomes unavailable, your cluster still has electable nodes which form a

quorum (majority) for the election of a master node.



Cluster Scaling

Adjusting Configuration

Last updated : 2021-07-20 17:05:27

Use Cases

As your business grows, the amounts of data and access requests in a cluster are ever-changing and may increase significantly. When the size of the cluster fails to meet your actual business needs, you can dynamically adjust the cluster configuration to scale it out. You can also temporarily scale it in if the amounts of data and access requests get smaller. ES supports adjusting configuration items such as number of nodes, node types, and storage capacity of individual data nodes. In addition, it also supports adjusting dedicated master nodes as detailed below.

Note :

Scale-in is currently restricted and you can [submit a ticket](#) for application.

Directions

1. Log in to the [ES Console](#) and enter the cluster list page.
2. You can **adjust the cluster configuration** on either the cluster list page or cluster details page.
 - On the cluster list page, select the cluster to be scaled out and select **More > Adjust Configuration** in the "Operation" column.
 - Click the cluster name to enter the cluster details page and select **More > Adjust Configuration** in the top-right corner.
3. In the **Adjust Configuration** dialog box that pops up, adjust the cluster node model or quantity according to your business needs and click **OK**.

Note :

Either node type (node model and storage capacity) or node quantity can be adjusted at a time.

- iv. After the configuration adjustment starts, the cluster status will change to **processing**. After the status changes to **normal**, the cluster can be used normally.

- You can view the progress of cluster adjustment on the cluster details page.

Data Node (Warm)

Node Model	Standard		
Node Specs	ES.S1.MEDIUM4-2 cores 4 GB ▾		
Single-node Data Disk	Premium cloud disk	100	GB ?
Node Qty	- 3 +		

Dedicated Master Node ☒ Dedicated master node is a type of node in Elasticsearch clusters. It does not store data and is designed to guarantee cluster stability. [Details](#)

Node Specs ES.S1.MEDIUM4-2 cores 4 GB ▾

Node Qty 3 ▾

Original Configuration Fees

New Configuration Fees

Confirm

Cancel

- View the progress of cluster adjustment.

es-ouagghha

Kibana Cloud Monitor

Basic Configuration Cluster Monitoring Node Monitoring Cluster Logs Advanced configuration Plugin List **Cluster Change History**

You haven't set an alarm recipient for the Cloud Monitor alarming policy of the current cluster, therefore the alarming feature hasn't taken effect. To keep track of the running status of the cluster and ensure stability of your business, please set an alarm recipient in time, which only takes a few simple steps [Configure now](#) or [view tutorial](#)

All Last 24 hours Last 7 days Last 30 days 2018-01-01 00:00:00 ~ 2020-08-20 12:17:40

Time	Operation	Details	Progress
	Change configuration	Original: Dedicated Master Node Model: None; Dedicated Master Nodes: None New: Dedicated Master Node Model: ES.S1.SMALL2; Dedicated Master Nodes: 3	13%Progress: / Collapse Prepare resources Progress 66% Change cluster nodes Progress 0%
2020-08-20 11:56:10	Create	--	100% Show More

Configuration Adjustment Fees

- For postpaid pay-as-you-go clusters, fees are charged by the minute. After the configuration adjustment is completed, fees for the next minute will be calculated based on the price of the new configuration.

Suggestions and Principles for Cluster Specification Adjustment

Last updated : 2021-07-13 17:14:55

As your business develops, the data volume and access traffic of your ES cluster grow. When the cluster size and configuration become insufficient to satisfy your actual business needs, you can expand your ES cluster as needed.

Configuration Adjustment Suggestion Overview

If your business hits a bottleneck, you need to consider which configuration adjustment method should be used. You can choose a method based on the table below and your actual business needs (the rolling mode and blue-green mode mentioned in the table will be detailed later in this document):

Configuration Adjustment Method	Use Case	How It Works	Characteristics
Expanding disk capacity	The computing resources are sufficient, but the disk storage space is insufficient.	The configuration is directly adjusted to increase the capacity of disks mounted to the cluster nodes one by one (the blue-green mode is used for encrypted cloud disks).	The cluster can be smoothly expanded without interrupting your business, and the expansion is quick (about 30 seconds for each node).
Increasing disk quantity	The computing resources are sufficient, but the disk storage space, IOPS, or throughput is insufficient.	You can choose the rolling mode or the blue-green mode.	The rolling mode is quicker, while the blue-green mode is slower but has no impact on the business in the production environment.
Adding more nodes	The node specification is high, but the overall computing resources of the cluster are insufficient.	The configuration is directly adjusted to add more nodes in the same specification as that of existing ones to the cluster.	The cluster can be smoothly expanded without interrupting your business, and the expansion is quick, which is not subject to the number of nodes and generally takes 5–10 minutes.

Configuration Adjustment Method	Use Case	How It Works	Characteristics
Upgrading node specification	The node specification is low, and the overall computing resources of the cluster are insufficient.	You can choose the rolling mode or the blue-green mode.	The rolling mode is quicker, while the blue-green mode is slower but has no impact on the business in the production environment.

Note :

- The table above focuses on the mappings between different scenarios and configuration adjustment methods. The description of how it works is applicable to the scenarios where only the corresponding >configuration item is adjusted. If multiple configuration items (such as node specification, disk capacity, and disk quantity) are adjusted at the same time, the actual adjustment method may be different.
- An expansion operation can adjust either vertical (node specification, disk capacity, and disk quantity) or horizontal (node quantity) configuration items at a time. You cannot adjust vertical and horizontal >configuration items at the same time.

Basic Configuration Adjustment Concepts

Common cluster configuration adjustment modes include direct adjustment mode, rolling mode, blue-green mode, etc. as detailed below:

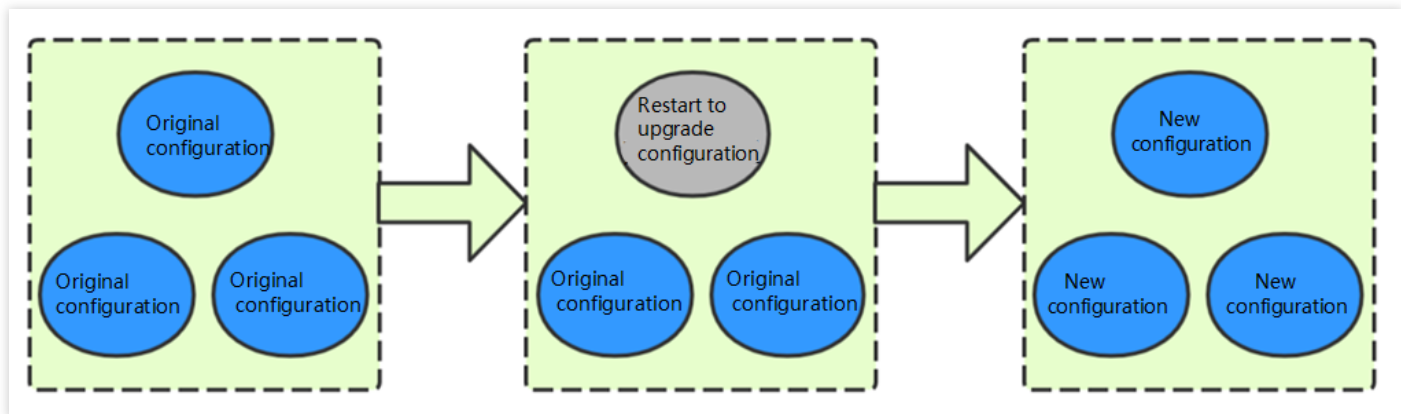
Direct adjustment

As direct adjustment does not involve heavy change operations on the cluster (such as restart or data replication), it generally has no impact on the business in the production environment and is relatively quick. For example, both horizontally adding more nodes to the cluster or vertically expanding the disk capacity are direct adjustments.

Rolling mode

In this mode, nodes in the cluster are restarted one by one on a rolling basis, and the system service is not interrupted, but the access performance in the production environment may be affected.

- **Advantage:** as no data migration is required, the expansion duration is not subject to the cluster data volume, and the configuration adjustment can be completed quickly. Therefore, this mode is suitable for scenarios where your cluster hits a bottleneck and you want to quickly complete the expansion and configuration adjustment.
- **Disadvantage:** as all nodes in the cluster need to be restarted on a rolling basis, if the cluster has no replicas, the data may be lost. Moreover, there are nodes consecutively leaving and joining the cluster during the configuration adjustment, which may compromise the cluster performance. Therefore, we recommend you not use this scheme during peak hours of your business.

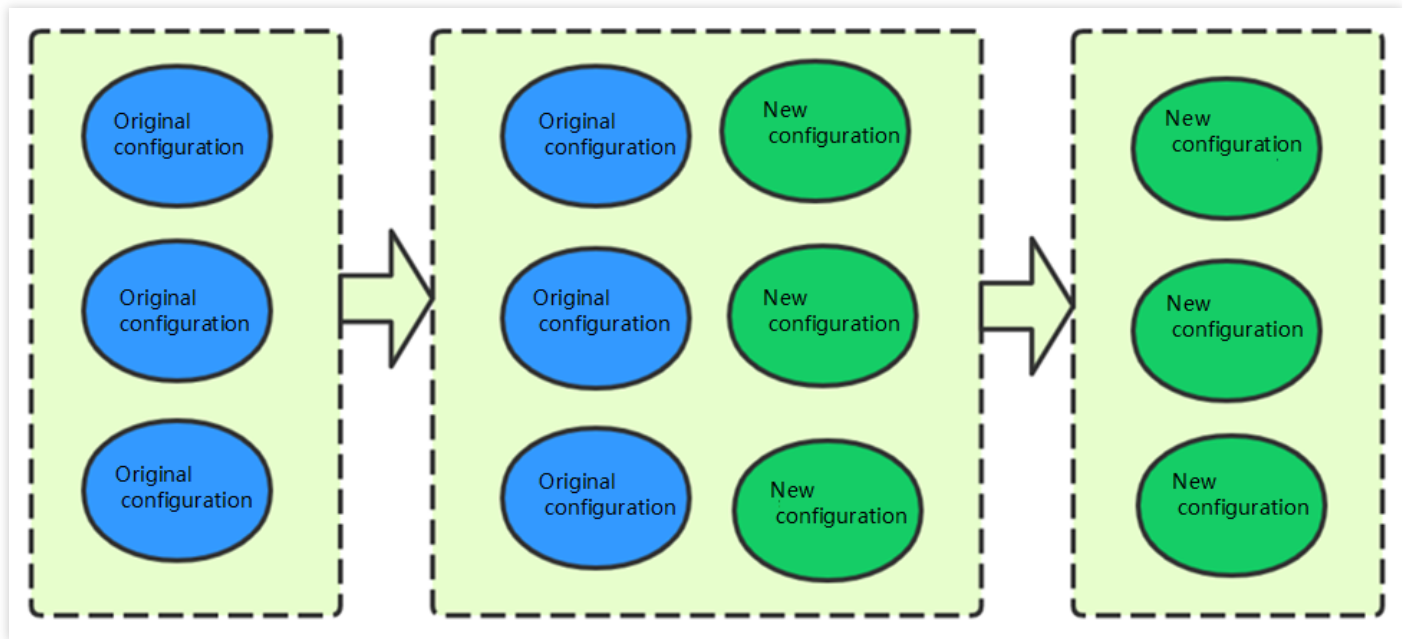


Blue-Green mode

The same number of new nodes as the existing nodes are added to the original cluster with no cluster restart required (the new nodes need to be configured as the desired type, and data on the existing nodes needs to be replicated to the new nodes). After the configuration, you can seamlessly switch to the new nodes and remove the legacy ones.

The node IPs will change after the configuration adjustment is completed.

- **Advantage:** the business does not need to be stopped during the configuration adjustment, and the expansion is very smooth; therefore, this mode is applicable to scenarios that have high requirements for cluster availability.
- **Disadvantage:** as data replication is involved, the configuration adjustment duration is directly proportional to the cluster data volume and can range from several minutes to several days or even longer.

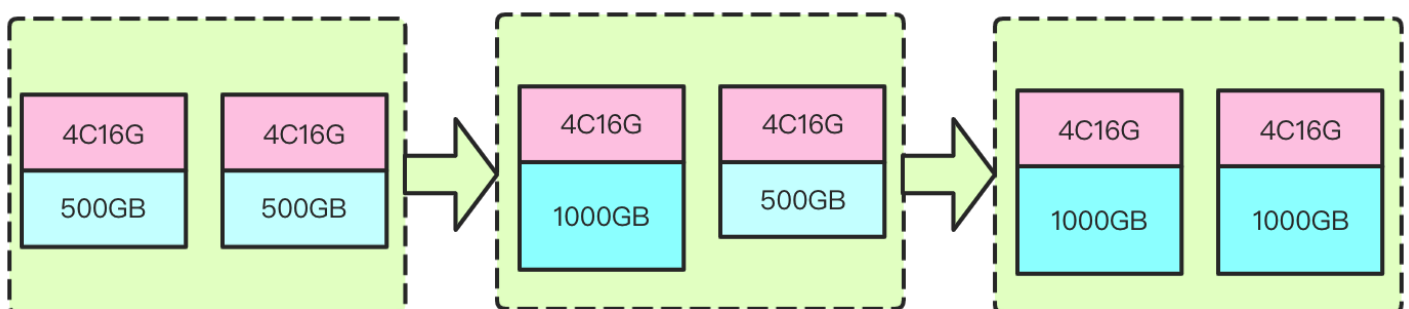


Principles of Main Configuration Adjustment Scenarios

1. Expanding disk capacity

It refers to increasing the disk capacity of each node so as to increase the overall storage capacity of the cluster without changing the computing resource configuration.

How it works: during disk capacity expansion, vertical expansion is performed on disks mounted to each data node in the cluster one by one to increase the overall disk storage capacity of the cluster; for example, the original disk capacity is 1,000 GB, and you can use this expansion method to increase the capacity to 5,000 GB. This method does not require node restart; therefore, the use of the cluster by your business will not be affected. As data nodes are expanded on a rolling basis, the more the nodes in the cluster, the longer the scaling process. If it takes about 10 seconds to expand a node, then it will take about 2 minutes to expand a cluster with 10 data nodes. The process and principle are as shown below:



Use cases: it is suitable for scenarios where the computing resources are sufficient, but the disk capacity is insufficient; for example, the CPU load and memory utilization of your cluster are relatively low, but the average disk utilization is high, and all data is important, which cannot be deleted to release disk space. In this case, you can expand the cluster by only expanding the disk capacity.

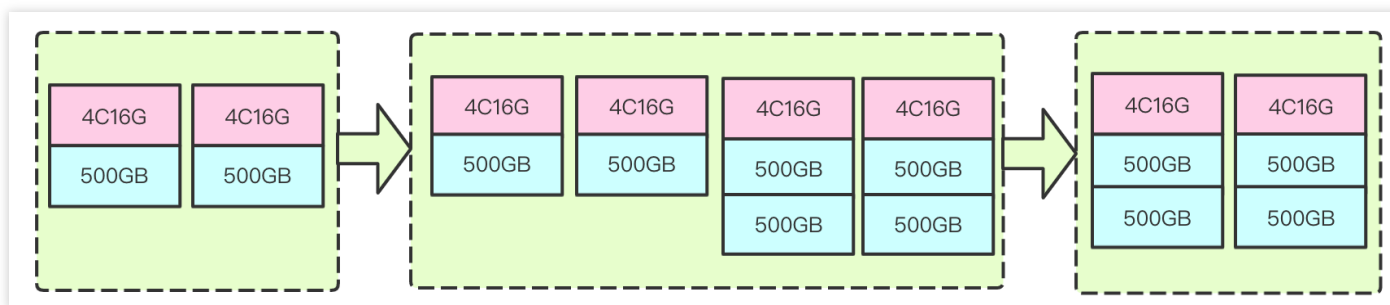
2. Increasing disk quantity

It refers to increasing the disk quantity of each node so as to increase the overall storage capacity, IOPS, and throughput of the cluster without changing the computing resource configuration.

How it works: the number of cloud disks mounted to a node is increased; for example, only one disk is mounted to each data node in the original cluster, but after this process is completed, multiple disks can be mounted to each data node. Currently, the blue-green mode and rolling mode are supported.

- **Blue-green mode:**

This mode is to add the same number of new nodes as existing nodes to the cluster with multiple cloud disks mounted to each new node, migrate data in the existing nodes to the new nodes, and then remove the existing nodes from the cluster. Its process and principle are as shown below:

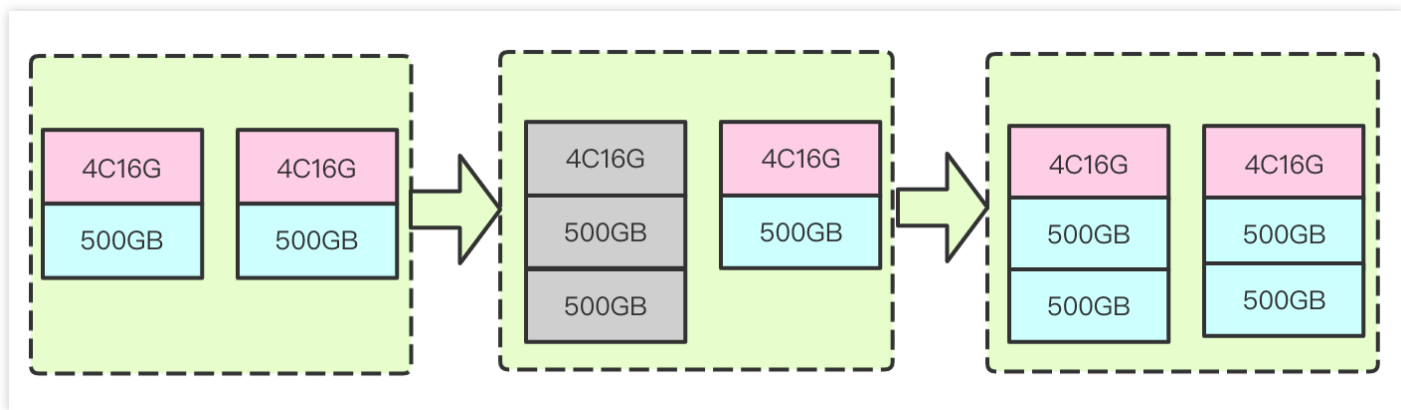


- **Advantage:** the entire cluster expansion process is smooth, which is applicable to scenarios with high requirements for business stability and availability.
- **Disadvantage:** the configuration adjustment duration is subject to the cluster data volume. The higher the volume, the longer the expansion duration.

- **Rolling mode:**

This mode is to directly add more disks to the original nodes so as to mount multiple disks. As the configuration of the cluster nodes needs to be modified, they need to be restarted for the change to take effect. To minimize the impact on the cluster performance, rolling restart is used, that is, disk addition and restart are performed on only one node at a time. Only after a node joins the cluster again will the next node be processed. The process and

principle are as shown below (gray indicates the restarting status):



- **Advantage:** as the rolling restart expansion mode involves no data migration, the expansion duration is not subject to the cluster data volume, and the configuration adjustment can be completed quickly. Therefore, this mode is suitable for scenarios where your cluster hits a bottleneck and you want to quickly complete the expansion and configuration adjustment.
- **Disadvantage:** as all nodes in the cluster need to be restarted on a rolling basis in this mode, if the cluster has no replicas, the data may be lost. Moreover, there are nodes consecutively leaving and joining the cluster during the configuration adjustment, which may compromise the cluster performance. Therefore, we recommend you not use this scheme during peak hours of your business.

In addition, the ES cluster does not rebalance the shards on individual nodes, that is, ES does not balance the shard assignment between different data paths; therefore, if you choose the rolling restart mode to directly add disks to existing nodes, ES will assign shards to the disk with most available storage space first, which cannot take advantage of multiple disks in a short time and may even result in performance problems such as excessive I/O load. This problem cannot be alleviated until the number of new disks reaches the number of existing ones. Moreover, when the utilization of a single disk in the existing cluster exceeds the threshold configured in ES, the entire node will also reach the threshold. In this case, even if there is available space on the new disks and you add more disks to the node, the problem will still persist. To solve it, you can migrate the data to other nodes or delete some data.

Comparison between different disk quantity increasing schemes:

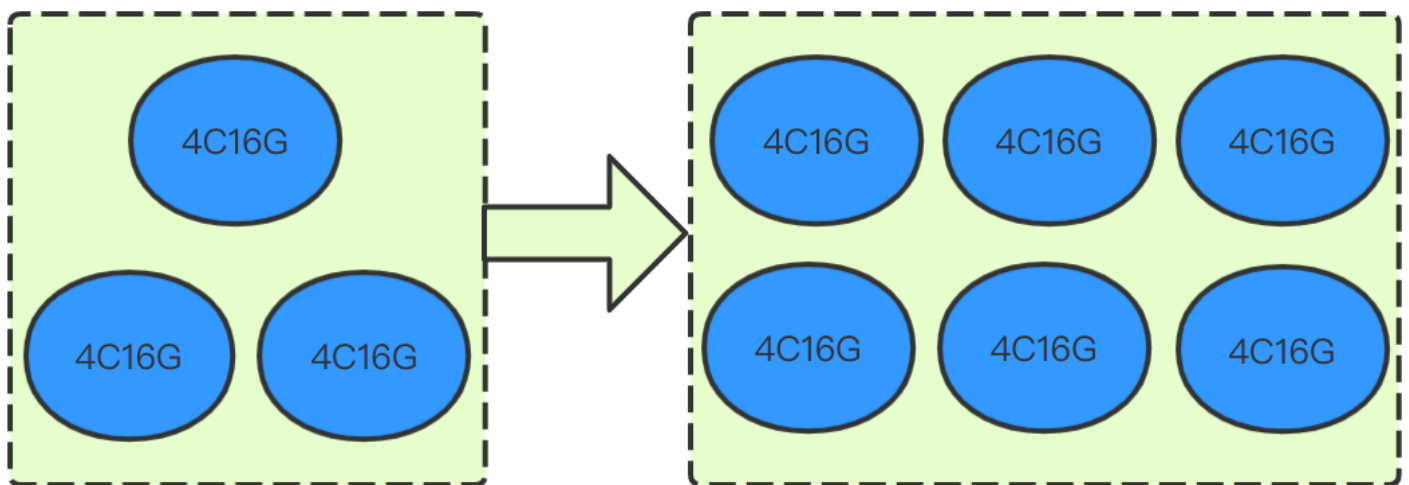
Disk Quantity Increasing Scheme	Advantage	Disadvantage
Blue-green mode	The configuration adjustment is smooth and imperceptible to the business	<ul style="list-style-type: none"> • The process duration is uncontrollable • The node IPs change

Disk Quantity Increasing Scheme	Advantage	Disadvantage
Rolling mode	<ul style="list-style-type: none"> The configuration adjustment duration is short and not subject to the cluster data volume The node IPs do not change 	<ul style="list-style-type: none"> Nodes are restarted on a rolling basis, affecting the business continuity You cannot take advantage of multiple disks in a short time Data in indices with no replicas may get lost

3. Adding more nodes

It is suitable for scenarios where the cluster computing resources hit the bottleneck; for example, the CPU load is continuously high, or the memory utilization stays high and repeatedly triggers memory issues or even OOM. In this case, you can expand the cluster nodes to improve the cluster's overall performance.

How it works: adding cluster nodes is an expansion method in which more nodes are added to the cluster without changing the cluster node model configuration. Its process and principle are as shown below:



The major advantage of this expansion method is smoothness, so that your business will not be interrupted during the expansion. As no data migration between new and existing nodes is involved, the expansion will not be affected by the node quantity and can be completed generally in 5–10 minutes.

Use cases: the node configuration is already very high, but you want to further improve the cluster's overall read/write performance, and your business requires high cluster stability during the expansion. In this case, you can expand the cluster by adding nodes to it.

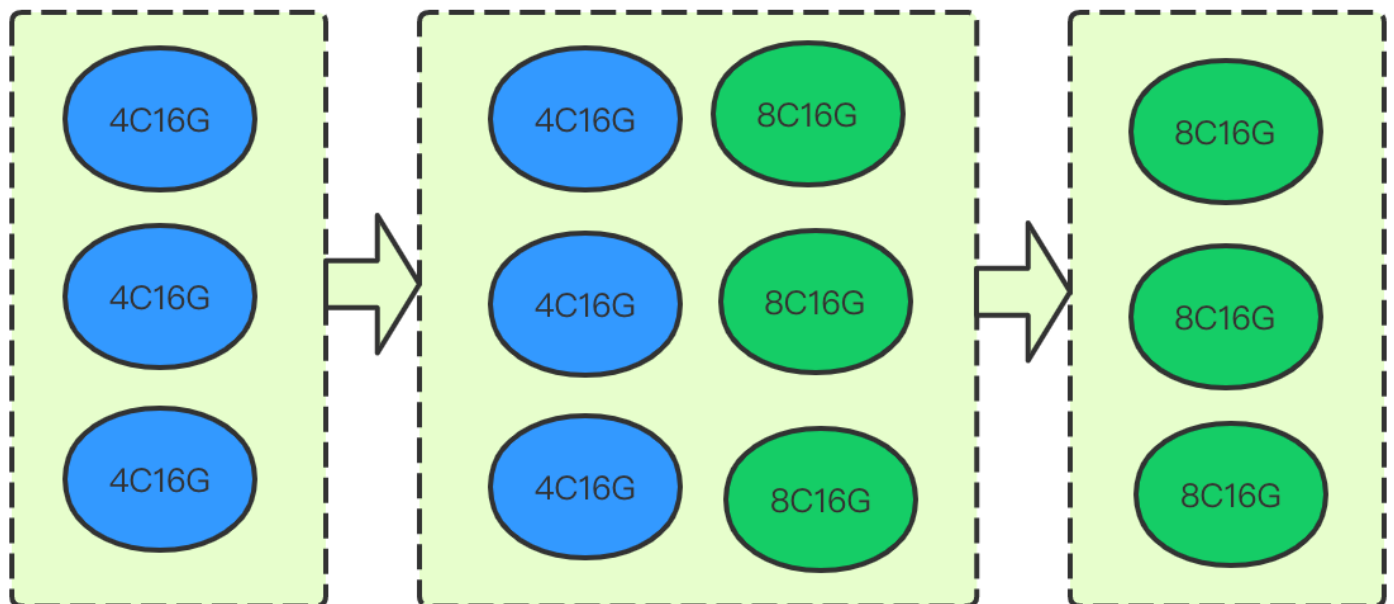
4. Upgrading node specification

It is suitable for scenarios where the cluster computing resources hit the bottleneck; for example, the CPU load is continuously high, or the memory utilization stays high and repeatedly triggers memory issues or even OOM. In this case, you can upgrade the specification of the cluster nodes to improve the cluster's overall performance.

How it works: it refers to upgrading the computing resource configuration of data nodes in the cluster without changing the number of nodes, such as upgrading the data node configuration from 4-core 16 GB MEM to 8-core 16 GB MEM. Currently, blue-green and rolling configuration adjustment modes are supported.

Blue-Green mode

This mode is to first add the same number of new nodes as existing nodes in a higher specification in the cluster, migrate data in the existing nodes to the new nodes, and then remove the existing nodes. Its process and principle are as shown below:



The major advantage of this expansion method is smoothness, so that the cluster use and availability will not be affected during the expansion. However, as the data in existing nodes needs to be migrated to the new nodes, the migration duration is greatly subject to the data volume. Therefore, if the cluster data volume is over 1 TB and you want to complete expansion as soon as possible, you can choose the rolling mode or set the following attributes in Kibana to speed up the data migration:

```
PUT _cluster/settings { "persistent": { "cluster.routing.allocation.node_concurrent_recoveries": "8", "indices.recovery.max_bytes_per_sec": "80mb" } }
```

Here:

- The **cluster.routing.allocation.node_concurrent_recoveries** attribute specifies the number of shards concurrently restored on each node in the cluster, which is 2 by default and cannot exceed 50. You can determine

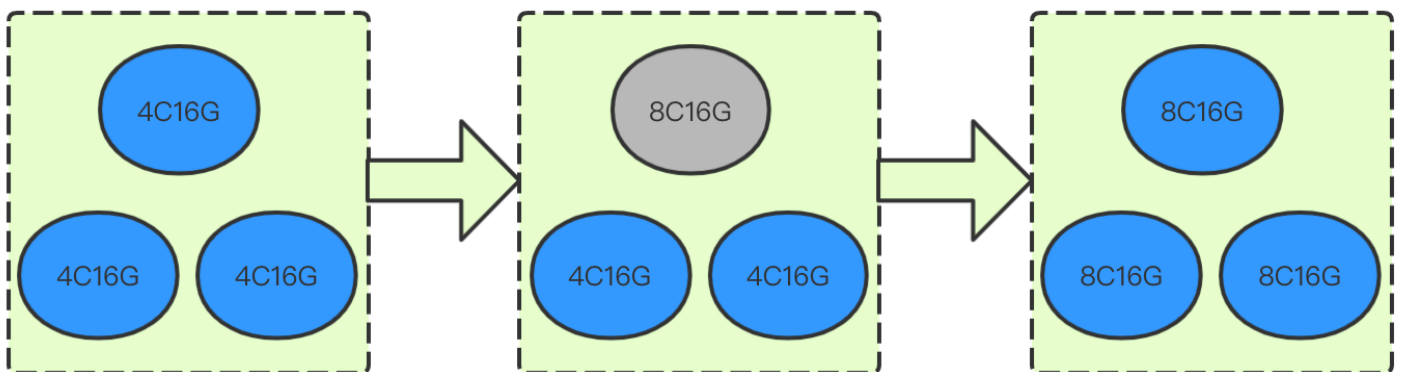
the specific value by multiplying the number of CPU cores in an existing data node by 4; for example, you are recommended to set this attribute to 16 if the existing data node specification is 4-core 16 GB MEM or to 50 if the specification is 16-core 64 GB MEM. If you find that the cluster stability is compromised after you increase the attribute value, you can appropriately set a smaller value.

- The **indices.recovery.max_bytes_per_sec** attribute specifies the maximum bandwidth for data transfer between nodes, which is `40mb` by default. The bandwidth limit should not be too high; otherwise, the cluster stability may be compromised. You can adjust the limit by an increment of `5mb` while checking the cluster stability so as to determine an appropriate value.

Use cases: the node configuration is relatively low, you want to further improve the cluster's overall read/write performance, your business requires high cluster stability during the expansion, and you have ample time for the expansion. In this case, you can expand the cluster through data migration.

Rolling mode

This mode is to upgrade the node specification of the cluster by restarting the nodes one by one. As it does not involve data migration, the specification adjustment duration is not subject to the cluster data volume but directly proportional to the number of nodes. It takes around 3–5 minutes for each node. The process and principle are as shown below (gray indicates the restarting status):



Use cases: it is suitable for scenarios where the cluster node configuration is low and hits a performance bottleneck, the requirement for cluster stability is low, and you want to quickly improve the cluster performance. As the nodes are restarted on a rolling basis during the configuration adjustment, there are nodes consecutively leaving and joining the cluster, and we recommend you use this mode during off-peak hours of your business.

Comparison between different node specification upgrade schemes:

Node Specification Upgrade Scheme	Advantage	Disadvantage
-----------------------------------	-----------	--------------

Node Specification Upgrade Scheme	Advantage	Disadvantage
Blue-green mode	The configuration adjustment is smooth and imperceptible to the business	<ul style="list-style-type: none">• The process duration is uncontrollable• The node IPs change
Rolling mode	<ul style="list-style-type: none">• The configuration adjustment duration is short and not subject to the cluster data volume• The node IPs do not change	<ul style="list-style-type: none">• Nodes are restarted on a rolling basis, affecting the business continuity• Data in indices with no replicas may get lost

Cluster Configuration

Synonym Configuration

Last updated : 2020-07-29 13:34:39

Tencent Cloud Elasticsearch Service (ES) allows you to configure synonyms in the following two ways: uploading a synonym file or directly referencing synonyms.

Method 1. Upload a synonym file

Notes

- Rolling restart of the cluster will be triggered after you upload a synonym file.
- A newly uploaded/modified synonym file does not take effect for legacy indices, so you need to create a new index accordingly. For example, if the existing index `myindex` uses the `synonym.txt` synonym file, but the file is modified and uploaded again, then the existing index `myindex` will not dynamically load the updated synonyms. You need to reindex the existing index; otherwise, the updated synonym file will take effect only for new indices.
- A synonym file must contain one synonym expression per line (the expressions support [Solr rules](#) and [WordNet rules](#)), be encoded in UTF-8, and have the `.txt` file extension; for example:


```
coke,cola => cola,coke  
elasticsearch,es => es
```

- You are allowed to upload a maximum of 10 synonym files of up to 10 MB in size each.

Directions

- Log in to the Tencent Cloud ES Console.
- On the cluster list page, click a cluster ID to enter the cluster details page.
- Click **Advanced Configuration** and enter the **Synonym Configuration** page.

Synonym Configuration

UTF-8 encoded files in .txt format are supported. Up to 10 MB. For the setting method, see [Help](#) 

Update Dictionary

File Name
None

4. Click **Update Dictionary** and upload a synonym file on the synonym update page.

Synonym dictionary

Upload

File	Size	Status	Operation
Click the "Upload" button above or drag and drop files to this area			

Save

Cancel

5. After the upload, click **Save**.

Using synonym dictionary

The following example uses the `filter` filter to configure synonyms and the `synonym.txt` as the testing file, and the file content is `elasticsearch,es => es`.

1. Log in to the Kibana Console of the cluster to which the synonym file has been uploaded. For detailed directions, please see [Accessing Cluster Through Kibana](#).
2. Click "Dev Tools" on the left sidebar.
3. Run the following command in the console to create an index:

```
PUT /my_index
{
  "settings": {
    "index": {
      "analysis": {
        "analyzer": {
          "my_ik": {
            "type": "custom",
            "tokenizer": "ik_smart",
            "filter": [
              "my_synonym"
            ]
          }
        }
      },
      "filter": {
        "my_synonym": {
          "type": "synonym",
          "synonyms_path": "analysis/synonym.txt"
        }
      }
    }
  },
  "mappings": {
    "_doc": {
      "properties": {
        "content": {
          "type": "text",
          "analyzer": "my_ik",
          "search_analyzer": "my_ik"
        }
      }
    }
  }
}
```

4. Run the following command to verify the synonym configuration:

```
GET /my_index/_analyze
{
  "analyzer": "my_ik",
  "text": "tencent elasticsearch service"
}
```

If the command is successfully executed, the following result will be returned:

```
{
  "tokens": [
    {
      "token": "tencent",
      "start_offset": 0,
      "end_offset": 6,
      "type": "ENGLISH",
      "position": 0
    },
    {
      "token": "es",
      "start_offset": 7,
      "end_offset": 20,
      "type": "SYNONYM",
      "position": 1
    },
    {
      "token": "service",
      "start_offset": 21,
      "end_offset": 28,
      "type": "ENGLISH",
      "position": 2
    }
  ]
}
```

In the output result, `token` and `es` are in `SYNONYM` type.

5. Run the following command to add some documents:

```
POST /my_index/_doc/1
{
  "content": "tencet elasticsearch service"
}

POST /my_index/_doc/2
{
  "content": "hello es"
}
```

6. Run the following command to search for synonyms:

```
GET my_index/_search
{
  "query" : { "match" : { "content" : "es" } },
  "highlight" : {
    "pre_tags" : [ "<tag1>", "<tag2>" ],
    "post_tags" : [ "</tag1>", "</tag2>" ],
    "fields" : { "content": {} }
  }
}
```

After the command is successfully executed, the following result will be returned:

```
{
  "took": 4,
  "timed_out": false,
  "_shards": {
    "total": 5,
    "successful": 5,
    "skipped": 0,
    "failed": 0
  },
  "hits": {
    "total": 2,
    "max_score": 0.25811607,
    "hits": [
      {
        "_index": "my_index",
        "_type": "_doc",
        "_id": "2",
        "_score": 0.25811607,
        "_source": {
```

```
"content": "hello es"
},
"highlight": {
  "content": [
    "hello <tag1>es</tag1>"
  ]
},
},
{
  "_index": "my_index",
  "_type": "_doc",
  "_id": "1",
  "_score": 0.25316024,
  "_source": {
    "content": "tencet elasticsearch service"
  },
  "highlight": {
    "content": [
      "tencet <tag1>elasticsearch</tag1> service"
    ]
  }
}
]
```

Method 2. Directly reference synonyms

1. Log in to the Kibana Console of the target cluster. For detailed directions, please see [Accessing Cluster Through Kibana](#).
2. Click "Dev Tools" on the left sidebar.
3. Run the following command in the console to create an index:

```
PUT /my_index
{
  "settings": {
    "index": {
      "analysis": {
        "analyzer": {
          "my_ik": {
            "type": "custom",
```

```
"tokenizer": "ik_smart",
"filter": [
  "my_synonym"
]
},
"filter": {
  "my_synonym": {
    "type": "synonym",
    "synonyms": [
      "elasticsearch,es => es"
    ],
  }
}
},
"mappings": {
  "_doc": {
    "properties": {
      "content": {
        "type": "text",
        "analyzer": "my_ik",
        "search_analyzer": "my_ik"
      }
    }
  }
}
```

Here, different from the method of using a synonym file, when synonyms are defined in `filter`, synonyms instead of a synonym file are directly referenced: `"synonyms": ["elasticsearch, es => es"]`

4. Run the following command to verify the synonym configuration:

```
GET /my_index/_analyze
{
  "analyzer": "my_ik",
  "text": "tencent elasticsearch service"
}
```

If the command is successfully executed, the following result will be returned:

```
{
  "tokens": [
    {
      "token": "tencet",
      "start_offset": 0,
      "end_offset": 6,
      "type": "ENGLISH",
      "position": 0
    },
    {
      "token": "es",
      "start_offset": 7,
      "end_offset": 20,
      "type": "SYNONYM",
      "position": 1
    },
    {
      "token": "service",
      "start_offset": 21,
      "end_offset": 28,
      "type": "ENGLISH",
      "position": 2
    }
  ]
}
```

In the output result, `token` and `es` are in `SYNONYM` type.

5. Run the following command to add some documents:

```
POST /my_index/_doc/1
{
  "content": "tencet elasticsearch service"
}

POST /my_index/_doc/2
{
  "content": "hello es"
}
```

6. Run the following command to search for synonyms:


```
GET my_index/_search
{
  "query" : { "match" : { "content" : "es" } },
  "highlight" : {
    "pre_tags" : [ "<tag1>", "<tag2>" ],
    "post_tags" : [ "</tag1>", "</tag2>" ],
    "fields" : { "content" : {} }
  }
}
```

After the command is successfully executed, the following result will be returned:

```
{
  "took": 4,
  "timed_out": false,
  "_shards": {
    "total": 5,
    "successful": 5,
    "skipped": 0,
    "failed": 0
  },
  "hits": {
    "total": 2,
    "max_score": 0.25811607,
    "hits": [
      {
        "_index": "my_index",
        "_type": "_doc",
        "_id": "2",
        "_score": 0.25811607,
        "_source": {
          "content": "hello es"
        },
        "highlight": {
          "content": [
            "hello <tag1>es</tag1>"
          ]
        }
      },
      {
        "_index": "my_index",
        "_type": "_doc",
        "_id": "1",
        "_score": 0.25316024,
        "_source": {
```

```
"content": "tencet elasticsearch service"
},
"highlight": {
  "content": [
    "tencet <tag1>elasticsearch</tag1> service"
  ]
}
}
```

YML File Configuration

Last updated : 2021-10-29 15:02:06

You can configure common parameters by modifying the YML parameter configurations of your Tencent Cloud Elasticsearch Service (ES) instance.

Directions

Viewing configuration items

Log in to the [ES Console](#) and click the **ID/name** of the cluster you want to modify to enter the details page. Click **Advanced configuration** to view configuration items.

YML Configuration

You need to restart the cluster for changes made to the following YML configuration to take effect. For configuration items supported in More Configurations, see the help documentation.[Help](#)

[Modify Configuration](#)

Parameter name	Parameter Description	Current Value	Value Description
indices.fielddata.cache.size	Specifies the percentage of java heap space tha...	Not configured	Percentage; format: 1%-100%; default value is ...
indices.query.bool.max_clause_count	Specifies the maximum number of clauses allo...	Not configured	An integer from the range [1, 2147483647]; the ...

Modifying configuration items

Click **Modify Configuration** in the **YML Configuration** section to modify the configuration items. Please see the "Value Description" column for descriptions of the specific input. You can customize other configuration items under **More Configurations** with YML syntax support.

YML Configuration

On-demand advanced configuration options of YML. For the meanings of specific configuration items, see the official documentation of Elasticsearch [Help](#)

[Confirm](#) [Cancel](#)

Parameter name	Note	Current Value	Value Range Description
action.destructive_requires_name	Deletes indices via wildcards or specifying an in...	<input type="text"/>	true or false; default value is true
indices.fielddata.cache.size	Specifies the percentage of java heap space tha...	<input type="text"/> %	Percentage; format: 1%-100%; default value is ...
indices.query.bool.max_clause_count	Specifies the maximum number of clauses allo...	<input type="text"/>	An integer from the range [1, 2147483647]; the ...

Click **Confirm** and the configured parameter will apply to your cluster. Check "Got It" on the pop-up window and click

Restart to restart the cluster. We recommend you perform the restart during off-peak hours if there are replicas of the indices in your cluster, although the operation will not affect your business.

Note :

If the cluster health status is Yellow or Red, or if there is an index without replicas in the cluster, you will be prompted to force restart if you try to modify the configuration items. In such cases, there is a higher risk of data loss or service interruptions if you update the configuration. We recommend repairing the cluster status first by adding replicas to all indices.

If you are aware of the risk and still want to update the configuration, you can check the "Force Restart" box and click **Confirm** to restart. For more information, please see the figure below:

The Tencent Cloud ES instance will restart upon confirmation. You can check the progress in the cluster change history during the restart and your YML file configuration will be completed once the restart is successful.

Supported Parameters

Parameter	Description	Default Value
indices fielddata.cache.size	Specifies the percentage of Java heap space allocated to the field data	15%
indices.query.bool.max_clause_count	Specifies the maximum number of clauses allowed in a Lucene Boolean query	1024

Supported Hotfix Parameters

Parameter	Description	Valid Values
action.destructive_requires_name	Specifies whether it is required to define the index name when deleting an index	true or false. Default value: true

Reindex Allowlist Configuration

Parameter	Description	Default Value
reindex.remote.whitelist	Allowlist of remote ES cluster access addresses	[]

Custom Queue Sizes

Parameter	Description	Default Value
thread_pool.bulk.queue_size	Document write queue size, applicable to v5.6.4	1024
thread_pool.write.queue_size	Document write queue size, applicable to v6.4.3 and above	1024
thread_pool.search.queue_size	Document search queue size	1024

Custom CORS Access Configurations

Parameter	Description	Default Value
-----------	-------------	---------------

Parameter	Description	Default Value
http.cors.enabled	Cross-origin resource sharing (CORS) configuration item. <code>true</code> indicates to enable CORS, while <code>false</code> indicates to disable it	false
http.cors.allow-origin	Origin resource configuration item, which indicates the domain names allowed for access requests	" "
http.cors.max-age	Cache period of the obtained CORS configuration information in the browser	1728000 (20 days)
http.cors.allow-methods	Allowed request methods for CORS	OPTIONS , HEAD , GET , POST , PUT , DELETE
http.cors.allow-headers	Request header information allowed for CORS	X-Requested-With , Content-Type , Content-Length
http.cors.allow-credentials	Specifies whether to allow <code>Access-Control-Allow-Credentials</code> information to be returned in the response header	false

Watcher Configuration

Parameter	Description	Default Value
xpack.watcher.enabled	<code>true</code> indicates to enable the Watcher feature of X-Pack	true

For more information on the configuration items, please see [Elasticsearch's documentation](#). If you want to customize other configuration items, please [submit a ticket](#).

Scenario-based Cluster Template Configuration

Last updated : 2020-10-15 11:15:32

Tencent Cloud Elasticsearch Service (ES) provides the scenario-based template configuration feature. You can select an appropriate scenario-based configuration index template based on your business needs to optimize the cluster performance in the corresponding scenario and reduce the performance problems caused by the use of an incorrect index template. General, search, and log scenarios are supported for such a template. This document describes how to select a template and the template parameters.

Notes

Before using a scenario-based index template, please pay attention to the following:

- There is a [default index template](#) named `default@template` that has been optimized for most scenarios in your ES cluster. You can run the `GET _template/default@template` command in **Dev Tools** of Kibana to view the template.
- The index template optimized for specific scenarios is named `scene@template`. You can run the `GET _template/scene@template` command in **Dev Tools** of Kibana to view the template.
- When purchasing an instance, you can select a scenario on the [purchase page](#). The general scenario is selected by default. After the purchase is completed, the configuration of the corresponding index template will be automatically applied to the cluster. You can change the scenario in the console.

Directions

Initializing scenario configuration

Enter the [ES purchase page](#) and select an initial scenario configuration template in the scenario-based configuration section. After the ES cluster is successfully purchased, the corresponding index configuration will be applied to it.

Default scenario configuration

☒ **General scenario** Contains general optimization configuration items, suitable for a variety of scenarios.

☐ **Log scenario** Scenarios where there are more writes than reads and that do not require high real-time performance

☐ **Search scenario** Nearly real-time scenarios that require high query performance

☐ **Not now**

The above scenario configuration is an initial index template provided based on different business characteristics and experience in application. It can be adjusted at any time after the cluster is created. [Learn More](#)

Modifying scenario configuration

1. Log in to the [ES Console](#), click a cluster ID in the **cluster list** to enter the cluster details page, and then enter the advanced configuration page.

Cluster Optimization Template

- ☒ **General scenario** Contains general optimization configuration items, suitable for a variety of scenarios.
- ☐ **Log scenario** Scenarios where there are more writes than reads and that do not require high real-time performance
- ☐ **Search scenario** Nearly real-time scenarios that require high query performance
- ☐ **Not now**

Default scenario configuration is an initial index template provided based on different business characteristics and experience in application. The corresponding template will be applied to the incremental index of your cluster after you switch the scenario configuration. [Learn More](#)

2. In the scenario-based configuration section, select the target scenario configuration template to be modified. In the pop-up window, if you click **Cancel**, the existing template will stay unchanged; if you click **OK**, the corresponding index configuration will be applied to the cluster immediately.

Scenario Switch Configuration ×

i The corresponding template will be applied to the incremental index of your cluster immediately after you switch the scenario configuration.

Confirm

Cancel

Scenario-based Index Template Description

The scenario-based index template parameters are detailed as below:

Parameter	Description
order	Template priority. The higher the value, the higher the priority
index_patterns	Index mode matched by index template. Wildcard is supported, which is <code>*</code> by default
index.refresh_interval	Index refreshing interval. Once indexed, a document can be queried only after the specified interval elapses. If you have a high requirement for query real-timeliness, you can slightly decrease the value; however, if the value is too small, the write performance will be affected

index.translog.durability	<p>Translog data storage method.</p> <ul style="list-style-type: none"><code>request</code> : stores translog data after each update to ensure data reliability and that no translog data will be lost when a node is exceptional<code>async</code> : stores translog data asynchronously on a regular basis. It increases the write performance at the cost of data reliability
index.translog.sync_interval	<p>Async refreshing interval which takes effect when the translog storage method is <code>async</code></p>

Kona JDK

Last updated : 2022-06-29 12:22:32

Kona JDK Overview

Kona JDK overview

Tencent Kona JDK is a JDK version developed by Tencent based on open-source JDK. It widely serves Tencent's internal businesses and customers in Tencent Cloud. Well proven by complex business scenarios such as big data and AI, it provides professional and continuous support for Tencent's Java ecosystem, with high stability, security, and performance.

Kona JDK strengths

- High performance at low costs: Compared with OpenJDK 8, Kona JDK 8 has an 8% increase in throughput and an around 10% reduction in both CPU and memory utilization, as proven by tens of thousands of production servers in Tencent's big data computing scenario.
- Out-of-the-box vector API support: This solves JVM crashes caused by vector instruction adaptation issues and is the first in the industry to stably support advertising training scenarios.
- Various GC optimizations: G1 GC memory overhead and parallel full GC algorithms are optimized, and production-grade ZGC is available for strongly real-time online services.
- KonaFiber coroutine: It has been implemented in Tencent IEG's TiMi game businesses. According to benchmark testing, its performance of creation, switch, and scheduling greatly exceeds that of Loom.
- Support for Chinese ARM-based CPUs.

Kona JDK Settings

1. Operation page: Cluster details page > advanced configuration page.
2. Restart: After the JDK is switched, the cluster will restart, and operations such as configuration adjustment and upgrade cannot be performed during the restart.

Switch JDKs

JDK replacement requires a cluster restart to take effect. [Learn more](#) 

- ☐ Open JDK Open-source JDK edition
- ☒ Kona JDK Tencent open-source edition

Data Comparison

This document describes the performance metrics of a 3-node ES Kona JDK + G1-GC cluster with 4 CPU cores and 16 GB memory.

Note :

The data comes from GeoNames and contains 11,396,503 entries of geographic location data in text, long, geo, and other types stored in columns and rows with a total size of around 3 GB.

The comparison between the 4-core 16 GB MEM 200 GB SSD 3-node ES Kona JDK (11.0.9.1-ga+1) + G1-GC cluster and an Oracle JDK (1.8.0_181-b13) + CMS-GC cluster with the same specification shows that ES has better performance in all aspects thanks to Tencent Cloud's proprietary JDK and GC parameter tuning.

Description	Metric	Task	ES (OpenJDK) CMS-GC
Total write time	Cumulative indexing time of primary shards		17.7745
Total GC count and time	Total Young Gen GC time		76.597
Total Young Gen GC count		4129	981
Total Old Gen GC time		0.175	0

Description	Metric	Task	ES (OpenJDK CMS-GC)
Total Old Gen GC count		2	0
Index size	Store size		2.885286
Translog size		3.59E-07	3.59E-07
Heap memory usage	Heap used for segments		0.045909
Heap used for doc values		0.000507355	0.000507
Heap used for terms		0.037261963	0.037261
Heap used for norms		0.003967285	0.003967
Heap used for points		0	0
Heap used for stored fields		0.004173279	0.004173
Total segment count	Segment count		7
Write throughput and time	Min Throughput	index-append	82341.03
Mean Throughput	index-append	85463.64521	87617.83
Median Throughput	index-append	85203.41999	87749.05
Max Throughput	index-append	88282.10166	90448.56
50th percentile latency	index-append	369.9228433	360.6725
90th percentile latency	index-append	582.2157889	521.3938
99th percentile latency	index-append	2566.001355	3331.056
99.9th percentile latency	index-append	3249.023346	4277.054
100th percentile latency	index-append	3677.799375	5966.865

Description	Metric	Task	ES (OpenJDK CMS-GC)
50th percentile service time	index-append	369.9228433	360.6725
90th percentile service time	index-append	582.2157889	521.3938
99th percentile service time	index-append	2566.001355	3331.056
99.9th percentile service time	index-append	3249.023346	4277.054
100th percentile service time	index-append	3677.799375	5966.865
error rate	index-append	0	0
Index metric statistics	Min Throughput	index-stats	90.00917
Mean Throughput	index-stats	90.01643728	90.02981
Median Throughput	index-stats	90.01545276	90.03117
Max Throughput	index-stats	90.0358266	90.04178
50th percentile latency	index-stats	2.71807611	2.706557
90th percentile latency	index-stats	3.542617336	3.530823
99th percentile latency	index-stats	4.209796032	4.047978
99.9th percentile latency	index-stats	10.97994745	4.319285
100th percentile latency	index-stats	18.87320075	4.494163
50th percentile service time	index-stats	1.521472819	1.515000
90th percentile service time	index-stats	1.832026243	1.815253

Description	Metric	Task	ES (OpenJDK CMS-GC)
99th percentile service time	index-stats	2.493222319	2.149661
99.9th percentile service time	index-stats	3.265745808	2.558897
100th percentile service time	index-stats	18.49333011	2.714292
error rate	index-stats	0	0
Node metric statistics	Min Throughput	node-stats	89.77465
Mean Throughput	node-stats	89.90322336	89.99453
Median Throughput	node-stats	89.92739222	89.99716
Max Throughput	node-stats	89.95708367	90.01224
50th percentile latency	node-stats	2.864421345	2.847921
90th percentile latency	node-stats	4.03423449	4.022879
99th percentile latency	node-stats	4.780995743	4.921176
99.9th percentile latency	node-stats	11.95199643	8.974571
100th percentile latency	node-stats	19.6932666	13.60371
50th percentile service time	node-stats	2.031991258	2.032643
90th percentile service time	node-stats	2.502718102	2.520979
99th percentile service time	node-stats	3.355697962	3.726954
99.9th percentile service time	node-stats	6.070044631	8.643416

Description	Metric	Task	ES (OpenJDK CMS-GC)
100th percentile service time	node-stats	19.13440693	11.63361
error rate	node-stats	0	0
Default query with all documents having a score of 1 (match_all)	Min Throughput	default	49.66239
Mean Throughput	default	49.80337019	49.93227
Median Throughput	default	49.8202478	49.93857
Max Throughput	default	49.87518669	49.95635
50th percentile latency	default	3.394149244	3.262675
90th percentile latency	default	4.578030575	4.408122
99th percentile latency	default	6.253439859	4.992298
99.9th percentile latency	default	19.66198959	8.490681
100th percentile latency	default	20.0197883	8.887056
50th percentile service time	default	2.622524276	2.356512
90th percentile service time	default	3.102052212	2.851721
99th percentile service time	default	4.579989612	3.174401
99.9th percentile service time	default	18.49881567	8.051926
100th percentile service time	default	18.62356346	8.214787
error rate	default	0	0

Description	Metric	Task	ES (OpenJDK CMS-GC)
Term conditional query	Min Throughput	term	98.93043
Mean Throughput	term	99.33413382	99.81852
Median Throughput	term	99.38459416	99.83007
Max Throughput	term	99.57176235	99.88279
50th percentile latency	term	3.250969574	3.228969
90th percentile latency	term	3.966032993	3.853681
99th percentile latency	term	10.50691157	4.505703
99.9th percentile latency	term	17.10123536	7.033703
100th percentile latency	term	19.53481138	9.737900
50th percentile service time	term	2.501523588	2.488659
90th percentile service time	term	3.069854062	2.982806
99th percentile service time	term	7.066733902	3.509562
99.9th percentile service time	term	16.17278317	6.230151
100th percentile service time	term	19.29396484	8.562799
error rate	term	0	0
Phrase query	Min Throughput	phrase	109.1666
Mean Throughput	phrase	109.4885892	109.7260
Median Throughput	phrase	109.531043	109.7627

Description	Metric	Task	ES (OpenJDK) CMS-GC
Max Throughput	phrase	109.6776159	109.8360
50th percentile latency	phrase	2.736125607	2.723197
90th percentile latency	phrase	3.2537736	3.277133
99th percentile latency	phrase	5.174562978	5.252283
99.9th percentile latency	phrase	17.94986153	11.65228
100th percentile latency	phrase	20.16797382	18.00533
50th percentile service time	phrase	1.983109396	1.964103
90th percentile service time	phrase	2.38582911	2.413296
99th percentile service time	phrase	4.028498856	3.426500
99.9th percentile service time	phrase	17.23640091	10.39778
100th percentile service time	phrase	19.54707783	17.02223
Aggregation query without cache	error rate	phrase	0
Min Throughput	country_agg_uncached	2.996727436	2.999315
Mean Throughput	country_agg_uncached	2.997330498	2.999446
Median Throughput	country_agg_uncached	2.997367399	2.999449
Max Throughput	country_agg_uncached	2.997811835	2.999560
50th percentile latency	country_agg_uncached	137.708772	138.3623
90th percentile latency	country_agg_uncached	162.6020876	162.0003

Description	Metric	Task	ES (OpenJDK CMS-GC)
99th percentile latency	country_agg_uncached	196.1384525	190.0452
100th percentile latency	country_agg_uncached	208.7042201	205.8009
50th percentile service time	country_agg_uncached	136.977708	137.0970
90th percentile service time	country_agg_uncached	161.4701347	160.9131
99th percentile service time	country_agg_uncached	195.4892302	188.7832
100th percentile service time	country_agg_uncached	208.3484558	204.5730
error rate	country_agg_uncached	0	0
Aggregation query with cache	Min Throughput	country_agg_cached	98.51641
Mean Throughput	country_agg_cached	98.94635299	99.03419
Median Throughput	country_agg_cached	98.99545733	99.08216
Max Throughput	country_agg_cached	99.24729184	99.31333
50th percentile latency	country_agg_cached	2.211962827	2.139798
90th percentile latency	country_agg_cached	3.517023474	3.494661
99th percentile latency	country_agg_cached	4.158023223	4.199306
99.9th percentile latency	country_agg_cached	9.866942695	8.245296
100th percentile latency	country_agg_cached	18.06280296	12.30363
50th percentile service time	country_agg_cached	1.467651688	1.393478

Description	Metric	Task	ES (OpenJDK CMS-GC)
90th percentile service time	country_agg_cached	1.777389366	1.689927
99th percentile service time	country_agg_cached	2.282693414	3.276122
99.9th percentile service time	country_agg_cached	4.195660669	7.769071
100th percentile service time	country_agg_cached	16.24826528	11.39958
error rate	country_agg_cached	0	0
Paged pull	Min Throughput	scroll	20.04421
Mean Throughput	scroll	20.05368445	20.05111
Median Throughput	scroll	20.05292541	20.05042
Max Throughput	scroll	20.0660563	20.06287
50th percentile latency	scroll	272.1138773	259.2352
90th percentile latency	scroll	290.9470227	265.0907
99th percentile latency	scroll	302.488716	284.5098
100th percentile latency	scroll	303.7193846	297.6893
50th percentile service time	scroll	270.1747189	257.1577
90th percentile service time	scroll	289.0668329	263.4378
99th percentile service time	scroll	300.3281443	282.0971
100th percentile service time	scroll	301.2135932	296.1045

Description	Metric	Task	ES (OpenJDK CMS-GC)
error rate	scroll	0	0
Script query (using expression script)	Min Throughput	expression	1.500956
Mean Throughput	expression	1.501160838	1.501741
Median Throughput	expression	1.501147998	1.501719
Max Throughput	expression	1.501414072	1.502131
50th percentile latency	expression	327.3259858	295.2159
90th percentile latency	expression	342.0345129	317.3502
99th percentile latency	expression	372.6446468	378.8509
100th percentile latency	expression	396.7165332	417.1186
50th percentile service time	expression	325.855901	293.9883
90th percentile service time	expression	340.6900207	316.3654
99th percentile service time	expression	370.736203	377.5147
100th percentile service time	expression	395.4625437	415.9661
error rate	expression	0	0
Script query (using painless static script without dynamically getting field values)	Min Throughput	painless_static	1.396916
Mean Throughput	painless_static	1.397478748	1.397943
Median Throughput	painless_static	1.397513941	1.397977

Description	Metric	Task	ES (OpenJDK CMS-GC)
Max Throughput	painless_static	1.397920498	1.398303
50th percentile latency	painless_static	431.2919118	371.3489
90th percentile latency	painless_static	465.1254796	391.1945
99th percentile latency	painless_static	512.2339443	437.3164
100th percentile latency	painless_static	538.9131764	465.5702
50th percentile service time	painless_static	429.9421017	369.7913
90th percentile service time	painless_static	463.2926716	390.1903
99th percentile service time	painless_static	511.3802825	434.9970
100th percentile service time	painless_static	537.7559569	464.3589
error rate	painless_static	0	0
Script query (using painless static script with dynamically getting field values)	Min Throughput	painless_dynamic	1.398724
Mean Throughput	painless_dynamic	1.398964022	1.396996
Median Throughput	painless_dynamic	1.398981831	1.397038
Max Throughput	painless_dynamic	1.399149307	1.397521
50th percentile latency	painless_dynamic	432.8310895	356.7619
90th percentile latency	painless_dynamic	462.9847418	383.0218
99th percentile latency	painless_dynamic	494.9476089	428.2430

Description	Metric	Task	ES (OpenJDK CMS-GC)
100th percentile latency	painless_dynamic	536.4017347	446.9218
50th percentile service time	painless_dynamic	431.5832192	355.6409
90th percentile service time	painless_dynamic	462.0900041	381.7875
99th percentile service time	painless_dynamic	494.3205597	425.9035
100th percentile service time	painless_dynamic	534.6057713	445.0034
error rate	painless_dynamic	0	0
Geographic range query (based on Gaussian decay function)	Min Throughput	decay_geo_gauss_function_score	1.001927
Mean Throughput	decay_geo_gauss_function_score	1.002340802	1.002568
Median Throughput	decay_geo_gauss_function_score	1.002308555	1.002535
Max Throughput	decay_geo_gauss_function_score	1.002881625	1.003158
50th percentile latency	decay_geo_gauss_function_score	387.5082242	332.3548
90th percentile latency	decay_geo_gauss_function_score	397.8741518	344.7444
99th percentile latency	decay_geo_gauss_function_score	407.4444408	356.9588
100th percentile latency	decay_geo_gauss_function_score	409.4531341	369.3594
50th percentile service time	decay_geo_gauss_function_score	386.1244814	331.0354
90th percentile service time	decay_geo_gauss_function_score	396.8515609	343.3262

Description	Metric	Task	ES (OpenJDK CMS-GC)
99th percentile service time	decay_geo_gauss_function_score	406.6675034	355.0055
100th percentile service time	decay_geo_gauss_function_score	407.7369291	368.1781
error rate	decay_geo_gauss_function_score	0	0
Geographic range query (based on Gaussian decay function with dynamically getting field values through script)	Min Throughput	decay_geo_gauss_script_score	1.001446
Mean Throughput	decay_geo_gauss_script_score	1.001755635	1.001885
Median Throughput	decay_geo_gauss_script_score	1.001731537	1.001860
Max Throughput	decay_geo_gauss_script_score	1.002160032	1.002318
50th percentile latency	decay_geo_gauss_script_score	411.4939715	334.8041
90th percentile latency	decay_geo_gauss_script_score	429.658707	345.1207
99th percentile latency	decay_geo_gauss_script_score	453.6645598	355.9493
100th percentile latency	decay_geo_gauss_script_score	454.430094	358.0469
50th percentile service time	decay_geo_gauss_script_score	409.7672189	333.3149
90th percentile service time	decay_geo_gauss_script_score	428.3069702	343.9684
99th percentile service time	decay_geo_gauss_script_score	451.8706388	354.2061
100th percentile service time	decay_geo_gauss_script_score	452.8327445	356.5891

Description	Metric	Task	ES (OpenJDK CMS-GC)
error rate	decay_geo_gauss_script_score	0	0
Custom scoring function query (defining function based on field value)	Min Throughput	field_value_function_score	1.503388
Mean Throughput	field_value_function_score	1.504118746	1.504719
Median Throughput	field_value_function_score	1.504074621	1.504659
Max Throughput	field_value_function_score	1.505051463	1.505800
50th percentile latency	field_value_function_score	194.2629726	134.5724
90th percentile latency	field_value_function_score	203.7090491	150.1895
99th percentile latency	field_value_function_score	214.6481861	166.6002
100th percentile latency	field_value_function_score	217.926288	184.5367
50th percentile service time	field_value_function_score	192.3880568	133.1520
90th percentile service time	field_value_function_score	202.3297952	148.4251
99th percentile service time	field_value_function_score	213.3810514	165.5014
100th percentile service time	field_value_function_score	215.2935127	183.1076
error rate	field_value_function_score	0	0
Custom scoring function query (dynamically getting field values through script to calculate scores)	Min Throughput	field_value_script_score	1.499697

Description	Metric	Task	ES (OpenJDK CMS-GC)
Mean Throughput	field_value_script_score	1.499757694	1.500311
Median Throughput	field_value_script_score	1.499759282	1.500306
Max Throughput	field_value_script_score	1.499799232	1.500380
50th percentile latency	field_value_script_score	240.0929602	174.8193
90th percentile latency	field_value_script_score	250.0571281	188.9238
99th percentile latency	field_value_script_score	270.1539508	215.9618
100th percentile latency	field_value_script_score	291.1372129	229.1083
50th percentile service time	field_value_script_score	238.8174967	173.5835
90th percentile service time	field_value_script_score	248.7244156	187.4786
99th percentile service time	field_value_script_score	268.9089779	214.8508
100th percentile service time	field_value_script_score	290.2693953	228.2811
error rate	field_value_script_score	0	0
Large terms query	Min Throughput	large_terms	1.101304
Mean Throughput	large_terms	1.101582867	1.100849
Median Throughput	large_terms	1.101561184	1.100839
Max Throughput	large_terms	1.101945785	1.101043
50th percentile latency	large_terms	211.8277326	242.0076
90th percentile latency	large_terms	231.8088979	252.8580
99th percentile latency	large_terms	251.2304624	265.9718

Description	Metric	Task	ES (OpenJDK CMS-GC)
100th percentile latency	large_terms	255.3527635	269.8639
50th percentile service time	large_terms	203.8265727	233.9178
90th percentile service time	large_terms	223.8224964	245.1530
99th percentile service time	large_terms	241.849935	258.2161
100th percentile service time	large_terms	246.3486325	262.1599
error rate	large_terms	0	0
Large filtered terms query	Min Throughput	large_filtered_terms	1.102296
Mean Throughput	large_filtered_terms	1.102784885	1.102979
Median Throughput	large_filtered_terms	1.102747397	1.102939
Max Throughput	large_filtered_terms	1.103436209	1.103668
50th percentile latency	large_filtered_terms	227.9210831	243.4361
90th percentile latency	large_filtered_terms	249.2253724	255.9631
99th percentile latency	large_filtered_terms	263.3142567	276.2566
100th percentile latency	large_filtered_terms	265.4732559	280.1711
50th percentile service time	large_filtered_terms	220.1946224	235.7397
90th percentile service time	large_filtered_terms	241.1826614	248.5632
99th percentile service time	large_filtered_terms	255.2141531	268.2613

Description	Metric	Task	ES (OpenJDK CMS-GC)
100th percentile service time	large_filtered_terms	256.941474	272.5524
error rate	large_filtered_terms	0	0
Large prohibited terms query	Min Throughput	large_prohibited_terms	1.102827
Mean Throughput	large_prohibited_terms	1.103422713	1.102860
Median Throughput	large_prohibited_terms	1.103376606	1.102821
Max Throughput	large_prohibited_terms	1.104211668	1.103525
50th percentile latency	large_prohibited_terms	202.5767318	232.1078
90th percentile latency	large_prohibited_terms	220.4174595	247.2566
99th percentile latency	large_prohibited_terms	234.3344817	266.7954
100th percentile latency	large_prohibited_terms	246.6131421	268.8084
50th percentile service time	large_prohibited_terms	193.9010601	224.5439
90th percentile service time	large_prohibited_terms	212.6108234	239.8692
99th percentile service time	large_prohibited_terms	226.4359237	258.9916
100th percentile service time	large_prohibited_terms	238.984541	260.8724
error rate	large_prohibited_terms	0	0
Descending order query	Min Throughput	desc_sort_population	1.504037
Mean Throughput	desc_sort_population	1.504907329	1.505087
Median Throughput	desc_sort_population	1.504841628	1.505025

Description	Metric	Task	ES (OpenJDK) CMS-GC
Max Throughput	desc_sort_population	1.506034196	1.506249
50th percentile latency	desc_sort_population	61.13778474	54.50106
90th percentile latency	desc_sort_population	73.92849587	69.35394
99th percentile latency	desc_sort_population	85.77715084	86.20061
100th percentile latency	desc_sort_population	85.84200498	87.74290
50th percentile service time	desc_sort_population	59.92501229	53.58439
90th percentile service time	desc_sort_population	72.30911367	68.09855
99th percentile service time	desc_sort_population	84.09957183	84.57749
100th percentile service time	desc_sort_population	84.19063408	85.95814
error rate	desc_sort_population	0	0
Ascending order query	Min Throughput	asc_sort_population	1.504247
Mean Throughput	asc_sort_population	1.505166062	1.505508
Median Throughput	asc_sort_population	1.505099341	1.505440
Max Throughput	asc_sort_population	1.506349989	1.506767
50th percentile latency	asc_sort_population	63.16776341	62.82690
90th percentile latency	asc_sort_population	78.09764324	75.61749
99th percentile latency	asc_sort_population	87.33172638	84.58683
100th percentile latency	asc_sort_population	87.89979294	85.19899

Description	Metric	Task	ES (OpenJDK CMS-GC)
50th percentile service time	asc_sort_population	61.90986466	61.71039
90th percentile service time	asc_sort_population	76.55056985	74.45018
99th percentile service time	asc_sort_population	85.81453795	83.30245
100th percentile service time	asc_sort_population	86.60695888	84.05557
error rate	asc_sort_population	0	0
search_after query with sorting in ascending order	Min Throughput	asc_sort_with_after_population	1.503017
Mean Throughput	asc_sort_with_after_population	1.503667166	1.504271
Median Throughput	asc_sort_with_after_population	1.503620246	1.504214
Max Throughput	asc_sort_with_after_population	1.504506304	1.505248
50th percentile latency	asc_sort_with_after_population	79.49512405	81.97531
90th percentile latency	asc_sort_with_after_population	94.07415418	97.05960
99th percentile latency	asc_sort_with_after_population	115.1407234	110.1778
100th percentile latency	asc_sort_with_after_population	117.4867153	115.6357
50th percentile service time	asc_sort_with_after_population	78.12653808	80.56232
90th percentile service time	asc_sort_with_after_population	92.57791536	96.00112
99th percentile service time	asc_sort_with_after_population	113.538067	108.2517

Description	Metric	Task	ES (OpenJDK CMS-GC)
100th percentile service time	asc_sort_with_after_population	116.0558164	114.5531
error rate	asc_sort_with_after_population	0	0
Query with sorting high base fields in descending order (quickly getting topK based on DistanceFeatureQuery)	Min Throughput	desc_sort_geonameid	6.011806
Mean Throughput	desc_sort_geonameid	6.014040004	6.016121
Median Throughput	desc_sort_geonameid	6.013860893	6.015844
Max Throughput	desc_sort_geonameid	6.016975785	6.019491
50th percentile latency	desc_sort_geonameid	8.037513588	6.896098
90th percentile latency	desc_sort_geonameid	8.790209144	7.481213
99th percentile latency	desc_sort_geonameid	20.16597	7.890859
100th percentile latency	desc_sort_geonameid	22.69194461	8.130467
50th percentile service time	desc_sort_geonameid	7.199986372	6.043605
90th percentile service time	desc_sort_geonameid	7.634483464	6.330675
99th percentile service time	desc_sort_geonameid	18.95111335	6.674837
100th percentile service time	desc_sort_geonameid	22.06934988	6.795545
error rate	desc_sort_geonameid	0	0

Description	Metric	Task	ES (OpenJDK CMS-GC)
search_after query with sorting high base fields in descending order	Min Throughput	desc_sort_with_after_geonameid	6.003999
Mean Throughput	desc_sort_with_after_geonameid	6.00483332	6.002715
Median Throughput	desc_sort_with_after_geonameid	6.004831591	6.002684
Max Throughput	desc_sort_with_after_geonameid	6.005864935	6.003257
50th percentile latency	desc_sort_with_after_geonameid	64.12782287	69.34804
90th percentile latency	desc_sort_with_after_geonameid	79.63361973	85.98741
99th percentile latency	desc_sort_with_after_geonameid	87.09606319	91.30932
100th percentile latency	desc_sort_with_after_geonameid	88.47462852	91.78488
50th percentile service time	desc_sort_with_after_geonameid	63.23770666	68.51645
90th percentile service time	desc_sort_with_after_geonameid	78.83979175	85.22403
99th percentile service time	desc_sort_with_after_geonameid	86.525729	90.76162
100th percentile service time	desc_sort_with_after_geonameid	87.29847241	91.37092
error rate	desc_sort_with_after_geonameid	0	0
Query with sorting high base fields in ascending order (quickly getting topK based on DistanceFeatureQuery)	Min Throughput	asc_sort_geonameid	6.018840
Mean Throughput	asc_sort_geonameid	6.022518134	6.022078
Median Throughput	asc_sort_geonameid	6.022245684	6.021816

Description	Metric	Task	ES (OpenJDK CMS-GC)
Max Throughput	asc_sort_geonameid	6.027178807	6.026594
50th percentile latency	asc_sort_geonameid	7.024060003	9.012220
90th percentile latency	asc_sort_geonameid	7.69297732	9.680523
99th percentile latency	asc_sort_geonameid	20.44826921	11.18117
100th percentile latency	asc_sort_geonameid	21.87036537	11.28741
50th percentile service time	asc_sort_geonameid	6.123304367	8.064015
90th percentile service time	asc_sort_geonameid	6.746383384	8.737695
99th percentile service time	asc_sort_geonameid	19.78318544	10.16213
100th percentile service time	asc_sort_geonameid	21.25467733	10.39039
error rate	asc_sort_geonameid	0	0
search_after query with sorting high base fields in ascending order	Min Throughput	asc_sort_with_after_geonameid	6.013236
Mean Throughput	asc_sort_with_after_geonameid	6.015849171	6.015858
Median Throughput	asc_sort_with_after_geonameid	6.015618744	6.015641
Max Throughput	asc_sort_with_after_geonameid	6.019167352	6.019115
50th percentile latency	asc_sort_with_after_geonameid	60.27546292	64.34633
90th percentile latency	asc_sort_with_after_geonameid	78.63363056	85.38805
99th percentile latency	asc_sort_with_after_geonameid	89.31191583	91.76640
100th percentile latency	asc_sort_with_after_geonameid	90.85853212	91.99177

Description	Metric	Task	ES (OpenJDK) CMS-GC
50th percentile service time	asc_sort_with_after_geonameid	59.692265	63.68059
90th percentile service time	asc_sort_with_after_geonameid	78.16235274	84.53184
99th percentile service time	asc_sort_with_after_geonameid	88.15484255	91.29356
100th percentile service time	asc_sort_with_after_geonameid	89.73695803	91.64701
error rate	asc_sort_with_after_geonameid	0	0

Plugin Configuration

Plugin List

Last updated : 2021-08-11 11:17:23

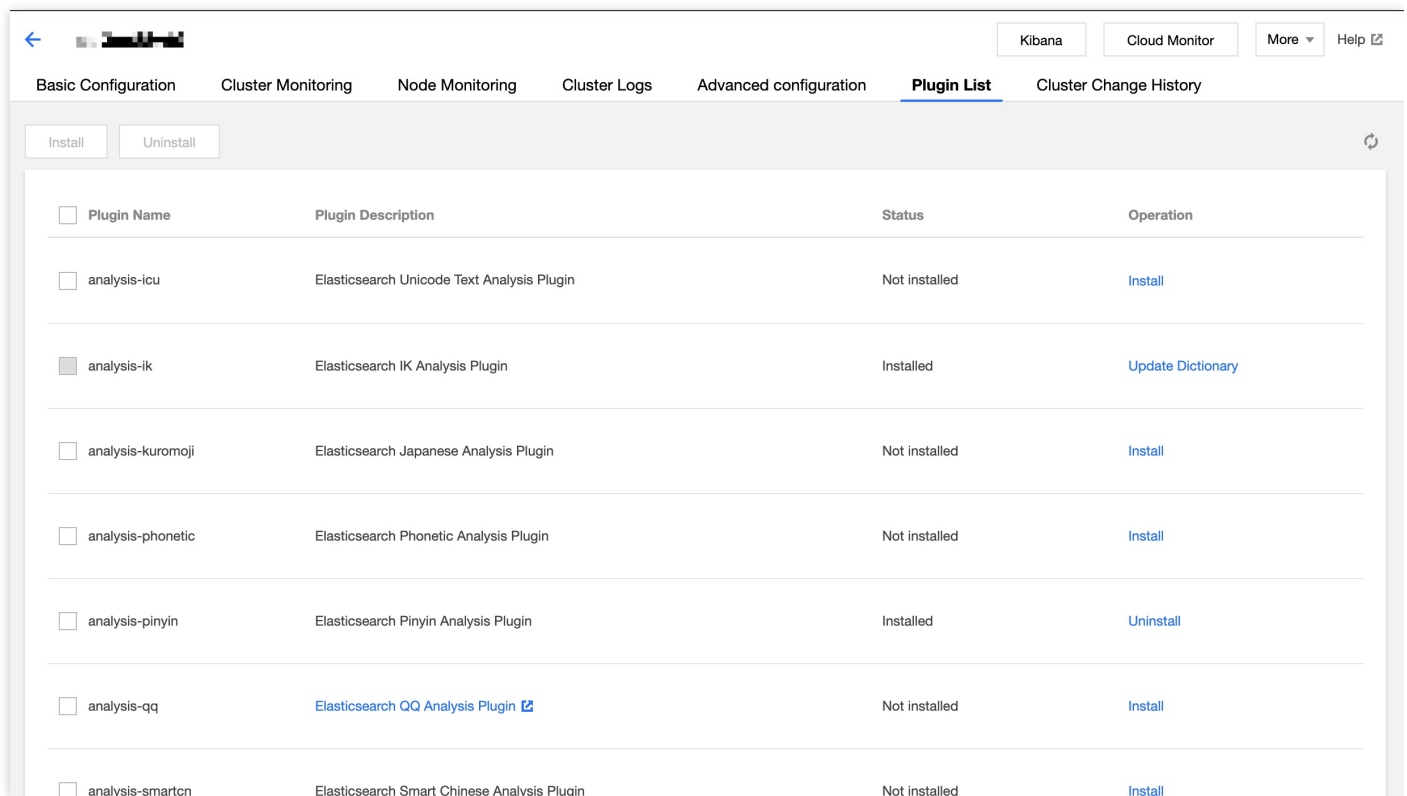
Tencent Cloud Elasticsearch Service (ES) provides over 10 plugins such as open-source Elasticsearch plugins and Tencent Cloud's proprietary plugins to deliver a wealth of plugin features. After purchasing an ES instance, you can install/uninstall these plugins on the plugin list page as needed. This document describes how to do so.

Note :

Rolling restart of the cluster will be triggered after you install or uninstall a plugin; therefore, please do so with caution.

Directions

1. Log in to the [ES console](#).
2. On the cluster list page, click a cluster ID to enter the cluster details page.
3. Click **Plugin List** to enter the plugin list management page.



Basic Configuration Cluster Monitoring Node Monitoring Cluster Logs Advanced configuration Plugin List Cluster Change History			
Install Uninstall			
<input type="checkbox"/> Plugin Name	Plugin Description	Status	Operation
<input type="checkbox"/> analysis-icu	Elasticsearch Unicode Text Analysis Plugin	Not installed	Install
<input checked="" type="checkbox"/> analysis-ik	Elasticsearch IK Analysis Plugin	Installed	Update Dictionary
<input type="checkbox"/> analysis-kuromoji	Elasticsearch Japanese Analysis Plugin	Not installed	Install
<input type="checkbox"/> analysis-phonetic	Elasticsearch Phonetic Analysis Plugin	Not installed	Install
<input type="checkbox"/> analysis-pinyin	Elasticsearch Pinyin Analysis Plugin	Installed	Uninstall
<input type="checkbox"/> analysis-qq	Elasticsearch QQ Analysis Plugin	Not installed	Install
<input type="checkbox"/> analysis-smartcn	Elasticsearch Smart Chinese Analysis Plugin	Not installed	Install

4. Click **Install** or **Uninstall** in the **Operation** column on the right of the target plugin.

5. In the pop-up dialog box, read the notes and click **OK** after confirming that everything is correct. Then, the plugin change operation will start, and the cluster will be restarted. You can view the change progress in **Cluster Change History**.

Plugin Information

ES supports the following plugins:

Plugin	Default Status	Description	Supported Operations
analysis-icu	Not installed	Elasticsearch Unicode text analysis plugin	Installation and uninstallation
analysis-ik	Installed	Elasticsearch IK analysis plugin, which cannot be uninstalled by default	Dictionary update
analysis-kuromoji	Not installed	Elasticsearch Japanese analysis plugin	Installation and uninstallation
analysis-nori	Not installed	Elasticsearch Korean analysis plugin	Installation and uninstallation
analysis-phonetic		Uninstalled	Elasticsearch phonetic symbol analysis plugin
analysis-pinyin	Installed	Elasticsearch Pinyin analysis plugin	Installation and uninstallation
analysis-qq	Not installed	Elasticsearch QQ analysis plugin	Installation, uninstallation, and dictionary update
analysis-smartcn	Not installed	Elasticsearch smart Chinese analysis plugin	Installation and uninstallation

Plugin	Default Status	Description	Supported Operations
analysis-stconvert	Installed	Elasticsearch Simplified and Traditional Chinese analysis plugin	Installation and uninstallation
ingest-attachment	Not installed	Apache Tika information extraction plugin	Installation and uninstallation
ingest-geoip	Not installed	IP address resolution plugin, which has already been integrated into ES 6.8.2 and above and does not need to be installed separately	Installation and uninstallation
ingest-user-agent	Not installed	Browser user agent information extraction plugin, which has already been integrated into ES 6.8.2 and above and does not need to be installed separately	Installation and uninstallation
mapper-murmur3	Not installed	This plugin is used to calculate and save the hash value of a field when you create an index	Installation and uninstallation
mapper-size	Not installed	This plugin is used to record the size of the document before compression when you create an index	Installation and uninstallation
repository-cos	Installed	This plugin is used to save Elasticsearch snapshots to Tencent Cloud COS, which cannot be uninstalled by default. For more information, please see Using COS for Backup and Restoration	None
repository-hdfs	Not installed	This plugin adds support for using HDFS as a repository for Snapshot/Restore	Installation and uninstallation
sql	Not installed	Open-source SQL parsing plugin. It does not need to be installed in Basic Edition and Platinum Edition ES clusters as they have already incorporated the SQL parsing feature	Installation and uninstallation

stored data contains a word in this dictionary, an index will be created and can be queried and found by using keywords. "Stopwords" will be deliberately avoided and will not be indexed.

- **Restrictions and requirements:** there are certain restrictions and requirements for dictionary files. For example, a dictionary file should contain one word per line and be encoded in UTF-8. For avoidance of confusion, the name of a customized dictionary file should not be the same as that of a stopwords dictionary file. In addition, as dictionary files will be loaded into the memory, you are allowed to upload a maximum of 10 files of up to 10 MB in size each.
- **Update process:** the list displays the dictionaries that have been uploaded and updated. A new dictionary will be blocked during upload if it does not meet the requirements. After a file is uploaded, it will be displayed in "pending activation" status. After all dictionaries to be updated are uploaded, click **Save**, and they will be saved to your cluster and take effect. If there is a file that fails to be uploaded or is not in UTF-8 format, a failure will be prompted, and you need to delete the failing file before you can click Save for other ones to take effect.

QQ Analysis Plugin

Last updated : 2021-07-01 10:02:56

Jointly developed by Tencent Cloud Elasticsearch Service (ES) team and Tencent Cloud NLP team, the QQ analysis plugin is widely used for Chinese text analysis among Tencent businesses such as QQ, WeChat, and QQ Browser. On the basis of traditional dictionary-based analysis, it supports features such as named-entity recognition (NER) and custom dictionaries. Through many years of application and continuous optimization, it has become industry-leading on key metrics such as analysis accuracy and speed. You can use it in Tencent Cloud ES to analyze and search for documents.

Notes

The QQ analysis plugin supports only clusters with data node specifications above 2-core 8 GB. If it is not installed in your cluster, please install it (`analysis-qq`) on the plugin list page.

The QQ analysis plugin provides the following analyzers and tokenizers:

- Analyzers: `qq_smart`, `qq_max`, `qq_smart_ner`, `qq_max_ner`
- Tokenizers: `qq_smart`, `qq_max`, `qq_smart_ner`, `qq_max_ner`

You can analyze and query documents by using the analyzers and tokenizers above. You can also use the dictionary configuration feature to customize and update the analysis dictionaries. For more information, please see dictionary configuration below.

Note :

- What is the difference between `qq_max` and `qq_smart` ?
 - `qq_max`: it splits text at the finest granularity; for example, it will split "tomato egg soup" into "tomato egg soup, tomato egg, egg soup, tomato, egg, soup".
 - `qq_smart`: it splits text at the roughest granularity; for example, it will split "tomato egg soup" into "tomato, egg, soup".
- What is NER? Why does it have an independent tokenizer?

NER (named-entity recognition) can recognize entities with specific meaning in text, such as person names, place names, institution names, and other proper nouns. You do not need to upload custom dictionaries for such proper nouns. The reason why the NER feature has a separate tokenizer is that a model needs to be loaded for NER, and the first loading takes much time.

Directions

1. Log in to the Kibana console of the cluster where the QQ analysis plugin has been installed. For detailed directions, please see [Accessing Cluster Through Kibana](#).
2. Click **Dev Tools** on the left sidebar.
3. Use an analyzer of the QQ analysis plugin in the console to create an index.

```
PUT /index
{
  "mappings": {
    "_doc": {
      "properties": {
        "content": {
          "type": "text",
          "analyzer": "qq_max",
          "search_analyzer": "qq_smart"
        }
      }
    }
  }
}
```

The statements above create an index named `index` in `_doc` type (for ES 7 or above, you need to add `?include_type_name=true` during index creation to support types). It contains the `content` attribute in `text` type and uses the `qq_max` and `qq_smart` analyzers. After the statements are successfully executed, the following result will be returned:

```
{
  "acknowledged": true,
  "shards_acknowledged": true,
  "index": "index"
}
```

4. Add some documents.

```
POST /index/_doc/1
{
  "content": "I downloaded the Honor of Kings from WeChat"
}
POST /index/_doc/2
{
  "content": "Ministry of Housing and Urban-Rural Development: to complete landscape resource registration of famous towns and villages by the end of September"
}
POST /index/_doc/3
{
  "content": "Latest weather forecast from China Meteorological Administration"
}
```



```
POST /index/_doc/4
{
  "content": "I live near ICOMOS China"
}
```

The statements above import four documents, and the `qq_max` analyzer will be used to analyze them.

5. Query the documents by highlighting keywords.

```
GET index/_search
{
  "query" : { "match" : { "content" : "China" } },
  "highlight" : {
    "pre_tags" : [ "<tag1>", "<tag2>" ],
    "post_tags" : [ "</tag1>", "</tag2>" ],
    "fields" : { "content": {} }
  }
}
```

The statements above are used to search for the documents in `_doc` type whose `content` field contains "China" by using the `qq_smart` analyzer. After the statements are successfully executed, the following result will be returned:

```
{
  "took" : 108,
  "timed_out" : false,
  "_shards" : {
    "total" : 1,
    "successful" : 1,
    "skipped" : 0,
    "failed" : 0
  },
  "hits" : {
    "total" : {
      "value" : 2,
      "relation" : "eq"
    },
    "max_score" : 0.7199211,
    "hits" : [
      {
        "_index" : "index",
        "_type" : "_doc",
        "_id" : "4",
        "_score" : 0.7199211,
        "_source" : {
          "content" : "I live near ICOMOS China"
        }
      },

```

```





"highlight" : {
"content" : [
"I live near ICOMOS <tag1>China</tag1>"
]
},
{
"_index" : "index",
"_type" : "_doc",
"_id" : "3",
"_score" : 0.6235748,
"_source" : {
"content" : "Latest weather forecast from China Meteorological Administration"
},
"highlight" : {
"content" : [
"Latest weather forecast from <tag1>China</tag1> Meteorological Administration"
]
}
}
]
}
}
}

```

Using Custom Dictionary

The QQ analysis plugin allows you to configure custom dictionaries. After being uploaded, a dictionary will trigger rolling restart of the cluster; therefore, please ensure that the cluster is in GREEN status and there are no single-replica indices.

1. Log in to the [ES console](#) and click a **cluster ID/name** on the cluster list page to enter the cluster details page.

<div>Create</div> <div>Separate each search term by pressing the Enter key and separate each keyword with a vertical bar " ". The keywords of cluster tags</div> <div>Q</div> <div></div>								
ID/Name	Running Status	Cluster Configuration	Health Status	AZ	Network	ES Version	Billing Mode	Operation
	Normal	StandardS1 2-core 4 GB, 3 nodes 100GB Cloud SSD x 1	Green	Guangzhou Zone 3	 new-vpc	7.5.1 Platinum edition	Pay-as-you-go Created on 2020-10-28 10:24:12	Kibana Cloud Monitor More
	Normal	StandardS1 2-core 4 GB, 3 nodes 100GB Cloud SSD x 1	Green	Guangzhou Zone 3	 new-vpc	7.5.1 Platinum edition	Pay-as-you-go Created on 2020-10-28 10:23:52	Kibana Cloud Monitor More

2. Click **Plugin List** to enter the plugin list management page.

Basic Configuration Cluster Monitoring Node Monitoring Log Advanced configuration Plugin List Visual Configuration Change History				
<input type="button" value="Install"/>		<input type="button" value="Uninstall"/>		
<input type="checkbox"/> Plugin Name	Plugin Description	Status	Operation	
<input type="checkbox"/> analysis-icu	Elasticsearch Unicode Text Analysis Plugin	Not installed	Install	
<input checked="" type="checkbox"/> analysis-ik	Elasticsearch IK Analysis Plugin	Installed	Update Dictionary	

3. Find the QQ analysis plugin (`analysis-qq`) and click **Update Dictionary** on the right.

4. The dictionary file must meet the following requirements:

- A dictionary file must be encoded in UTF-8, contain one word per line, and have the `.dic` extension.
- You can upload a maximum of 10 files of up to 10 MB each.

5. Click "Save". Cluster restart will not be triggered immediately, but cluster change will be triggered after several minutes for the dictionary file to take effect.

Troubleshooting and Testing

If the returned result of the QQ analysis plugin does not meet your expectations, you can run the following statements to troubleshoot and test the analyzers and tokenizers:

```
GET _analyze
{
  "text": "I live near ICOMOS China",
  "analyzer": "qq_max"
}
GET _analyze
{
  "text": "I live near ICOMOS China",
  "tokenizer": "qq_smart"
}
```

Monitoring and Alarming

Viewing Monitoring Information

Last updated : 2021-10-26 16:42:13

Overview

ES provides a number of monitoring metrics for running ES clusters to monitor cluster operations such as storage, I/O, CPU, and memory utilization. Based on these metrics, you can understand the cluster operations in real time and promptly handle possible risks to ensure stable cluster operations. This document describes how to view cluster monitoring information in the ES console.

Directions

1. Log in to the [ES console](#) and click a **cluster ID/name** on the cluster list page to enter the cluster details page.
2. Select the **Cluster Monitoring** tab to view the overall cluster running status. Select **Metric Group** to view the cluster monitoring metrics of data nodes, warm data nodes, and dedicated master nodes separately.
3. Select the **Node Monitoring** tab to view the operations and performance metrics of the nodes in the cluster.

Cluster monitoring

On the cluster monitoring page, you can set alarm policies and view the cluster monitoring data. You can view the overall cluster status and cluster performance metrics by time range, metric group, and time granularity.

Note :

You can also view all the ES cluster monitoring metrics in the [Cloud Monitor console](#).

[← es-13vgjt9](#)

Basic ConfigurationCluster MonitoringCluster LogsAdvanced ConfigurationCluster Change History

You haven't configured Cloud Monitor alarms for your clusters. It is recommended that you configure alarms for key metrics to ensure stable running of business. Preset alarm policy template is provided, so it only takes a few steps to finish configuration. [Configure Now](#) [View Tutorial](#)

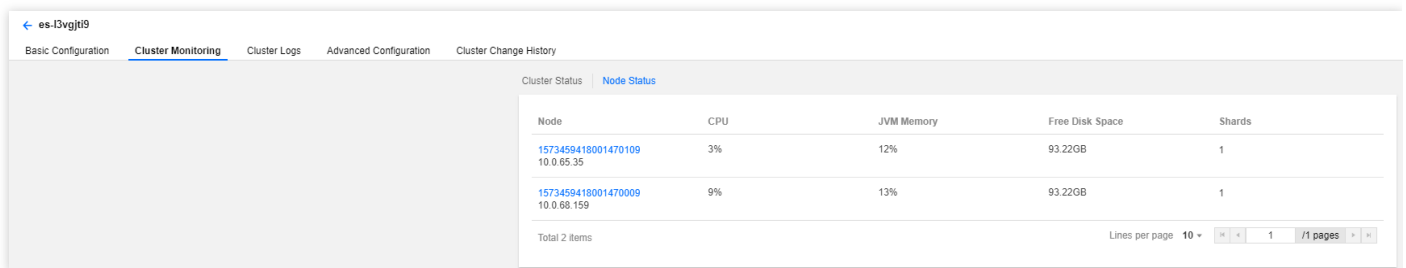
Cluster StatusNode Status

Nodes	Data Nodes	Indexes	Files	Storage
0	0	0	0	0 B
Shards	Primary Shards	Migrating Shards	Initializing Shards	Unassigned Shards
0	0	0	0	0

Node monitoring

• Node list

This section shows real-time health metrics of each node in the cluster.



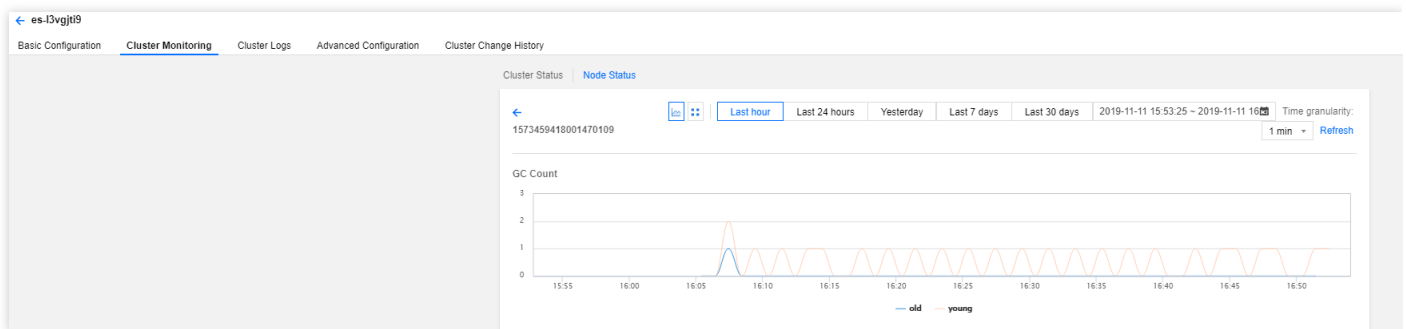
Node	CPU	JVM Memory	Free Disk Space	Shards
1573459418001470109 10.0.65.35	3%	12%	93.22GB	1
1573459418001470009 10.0.68.159	9%	13%	93.22GB	1

Total 2 items

Lines per page: 10 | 1 / 1 pages

• Single node status

This section shows detailed historical health status of each metric of each node.



Descriptions of certain metrics

An ES cluster is generally composed of multiple nodes. To reflect the overall health status of the cluster, certain monitoring metrics provide two types of values: average value and maximum value.

- The average value is the average of the metric's values of all nodes in the cluster.
- The maximum value is the maximum value of the metric of all nodes in the cluster.

The statistical period of each metric is 1 minute; that is, the cluster's metrics are collected once every minute. The metrics are as described below:

Monitoring Metric	Statistical Method	Details
Cluster health	ES cluster health status. 0: green (the cluster is normal); 1: yellow (alarm; some replica shards are unavailable);	<ul style="list-style-type: none"> • Green indicates that all primary and replica shards are available and the cluster is in the healthiest status. • Yellow indicates that all the primary shards are available, but

	2: red (exception; some primary shards are unavailable).	<p>some replica shards are unavailable. In this case, the search results are still complete; however, the high availability of the cluster is affected to some extent, and there are high risks with data loss. When the cluster health status changes to yellow, you should locate and troubleshoot the problem in a timely manner to prevent data loss.</p> <ul style="list-style-type: none"> Red indicates that at least one primary shard and all its replicas are unavailable. When the cluster health status changes to red, some data has already been lost, the search can only return partial data, and the write requests allocated to a lost shard will return an exception. In this case, you should locate and troubleshoot the exceptional shard as soon as possible.
Avg disk usage	The average of disk utilization values of all nodes in the cluster in one statistical period (1 minute).	If the disk utilization is too high, data cannot be written properly. Solution: Clean up useless indices promptly. Expand the cluster capacity by increasing the disk capacity of individual nodes or increasing the number of nodes.
Max disk utilization	The maximum disk utilization value of all nodes in the cluster in one statistical period (1 minute).	-
Avg JVM memory utilization	The average of JVM memory utilization values of all nodes in the cluster in one statistical period (1 minute).	<p>If this value is too high, frequent GC or even OOM will occur on cluster nodes.</p> <p>This happens generally because the tasks to be processed by ES exceed the load capacity of the nodes' JVMs. You need to pay attention to the tasks that are being executed by the cluster or adjust the cluster configuration.</p>
Max JVM memory	The maximum JVM memory utilization value of all nodes in the cluster in one statistical period (1	-

utilization	minute).	
Avg CPU utilization	The average of CPU utilization values of all nodes in the cluster in one statistical period (1 minute).	<p>When the read/write tasks processed by the nodes in the cluster exceed the load capacity of the nodes' CPUs, the value of this metric will become too high. In this case, the cluster nodes will experience a decrease in processing power or even crash. You can solve this problem in the following ways:</p> <ul style="list-style-type: none"> Observe whether the value of this metric is persistently or temporarily high. If it is temporarily soaring, determine whether there are temporary complex tasks in progress. If it is persistently high, analyze whether the read/write operations on the cluster by your business can be optimized, lower the read/write frequency, and decrease the amount of data so as to reduce the node load. If the node configuration cannot meet the throughput requirement of your business, you are recommended to perform vertical scaling of the cluster nodes to improve the load capacity of individual nodes.
Max CPU utilization	The maximum CPU utilization value of all nodes in the cluster in one statistical period (1 minute).	-
Avg cluster load per minute	The average load per minute (load_1m) of all nodes in the cluster. Source of the metric: ES node status API (<code>_nodes/stats/os/cpu/load_average/1m</code>).	If load_1m is too high, you are recommended to lower the cluster load or upgrade the cluster node specification.
Max cluster load per minute	The maximum load per minute (load_1m) of all nodes in the cluster.	-
Avg write	<ul style="list-style-type: none"> Write latency (index_latency) refers to the time 	Write latency is the average time it

latency	<p>taken by a single index request (ms/request). The average write latency of the cluster is the average of the time taken by a single index request of all nodes in one statistical period (1 minute).</p> <ul style="list-style-type: none"> Calculation rule for the single index request time of a node: two metrics are recorded once every statistical period (1 minute), i.e., total number of historical indices on a node (<code>_nodes/stats/indices/indexing/index_total</code>) and total time taken by historical indices (<code>_nodes/stats/indices/indexing/index_time_in_millis</code>), and the difference between two adjacent records (i.e., the absolute value in one statistical period) is taken for calculation (index time / number of indices) to get the average single index time in one statistical period (1 minute). 	<p>takes to write a single document. The average write latency of the cluster refers to the average of write time of all nodes in one statistical period.</p> <p>If the write latency is too high, you are recommended to upgrade the node specification or increase the number of nodes.</p>
Max write latency	<ul style="list-style-type: none"> Write latency (<code>index_latency</code>) refers to the time taken by a single index request (ms/request). The maximum write latency of the cluster is the maximum value of time taken by a single index request of all nodes in one statistical period (1 minute). Calculation rule for single index request time of a node: see the average write latency section. 	-
Avg query latency	<ul style="list-style-type: none"> Query latency (<code>search_latency</code>) refers to the time taken by a single query request (ms/request). The average query latency of the cluster is the average of the time taken by a single query request of all nodes in one statistical period (1 minute). Calculation rule for the single query request time of a node: two metrics are recorded once every statistical period (1 minute), i.e., total number of historical queries on a node (<code>_nodes/stats/indices/search/query_total</code>) and total time taken by historical queries (<code>_nodes/stats/indices/search/query_time_in_millis</code>), and the difference between two adjacent records (i.e., the absolute value in one statistical period) is taken for calculation (query time / number of queries) to get the average single query time in one statistical period (1 minute). 	<p>Query latency is the average time it takes to perform a single query. The average query latency of the cluster refers to the average of query time of all nodes in one statistical period. If the query latency is too high, you are recommended to upgrade the node specification or increase the number of nodes.</p>
Max query latency	<ul style="list-style-type: none"> Query latency (<code>search_latency</code>) refers to the time taken by a single query request (ms/request). The maximum query latency of the cluster is the 	-

	<p>maximum value of time taken by a single query request of all nodes in one statistical period (1 minute).</p> <ul style="list-style-type: none"> Calculation rule for single query request time of a node: see the average query latency section. 	
Avg number of writes per second	<p>The average of the number of index requests received by all nodes in the cluster per second. Calculation rule for the number of index requests per second of a node: the total number of historical indices on a node (<code>_nodes/stats/indices/indexing/index_total</code>) is recorded once every statistical period (1 minute), and the difference between two adjacent records (i.e., the absolute value in one statistical period) is taken for calculation (number of indices / 60 seconds) to get the average number of index requests per second in one statistical period.</p>	-
Avg number of queries per second	<p>The average of the number of query requests received by all nodes in the cluster per second. Calculation rule for the number of query requests per second of a node: the total number of historical queries on a node (<code>_nodes/stats/indices/search/query_total</code>) is recorded once every statistical period (1 minute), and the difference between two adjacent records (i.e., the absolute value in one statistical period) is taken for calculation (number of queries / 60 seconds) to get the average number of query requests per second in one statistical period.</p>	-
Write rejection rate	<p>This is the ratio calculated by dividing the number of write requests rejected by the cluster by the total number of write requests in one statistical period. Calculation rule: two metrics are collected once every statistical period, i.e., the number of historical write requests rejected (v5.6.4: <code>_nodes/stats/thread_pool/bulk/rejected</code>; v6.4.3 and above: <code>_nodes/stats/thread_pool/write/rejected</code>) and the total number of historical write requests (v5.6.4: <code>_nodes/stats/thread_pool/bulk/completed</code>; v6.4.3 and above: <code>_nodes/stats/thread_pool/write/completed</code>), and the difference between two adjacent records (i.e., the absolute value in one statistical period) is taken for</p>	<p>When the write QPS is too large or the CPU, memory, and disk utilization is too high, the cluster's write rejection rate may increase. Generally, this is because that the current configuration of the cluster cannot meet the requirements of write operations on the business side. For scenarios where the node configuration is too low, you can solve this problem by upgrading the node specification or reducing the number of write operations. For scenarios where the disk utilization</p>

	calculation (number of rejected write requests / total number of write requests).	is too high, you can solve this problem by expanding the cluster's disk capacity or deleting useless data.
Query rejection rate	This is the ratio calculated by dividing the number of query requests rejected by the cluster by the total number of query requests in one statistical period. Calculation rule: two metrics are collected once every statistical period, i.e., the number of historical query requests rejected (<code>_nodes/stats/thread_pool/search/rejected</code>) and the total number of historical query requests (<code>_nodes/stats/thread_pool/search/completed</code>), and the difference between two adjacent records (i.e., the absolute value in one statistical period) is taken for calculation (number of rejected query requests / total number of query requests).	When the write QPS is too large or the CPU and memory utilization is too high, the cluster's query rejection rate may increase. Generally, this is because that the current configuration of the cluster cannot meet the requirements of read operations on the business side. If this value is too high, you are recommended to upgrade the cluster node specification so as to improve the processing capabilities of the cluster nodes.
Total documents	Total number of documents written to the cluster. Calculation rule: ES cluster document quantity API (<code>_cluster/stats/indices/docs/count</code>).	-
Auto snapshot backup status	The backup result after auto snapshot backup is enabled for the cluster. 0: auto backup is not enabled; 1: auto backup is normal; -1: auto backup failed.	Auto snapshot backup will periodically back up the cluster data to COS, so that the data can be recovered when needed, thus more comprehensively ensuring data security. We recommend you enable it. For more information, please see Automatic Snapshot Backup .

Configuring Alarms

Last updated : 2021-11-22 16:56:06

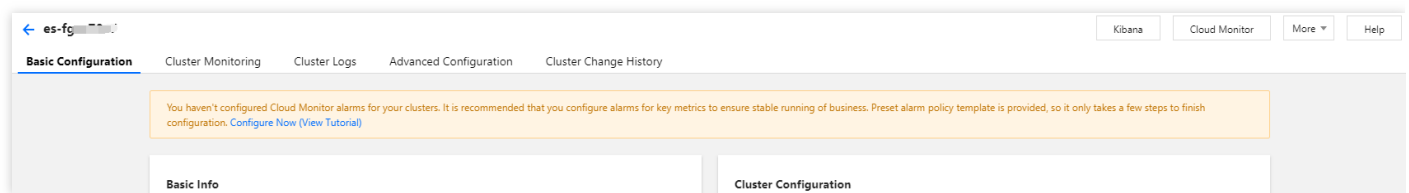
Operation Scenario

ES supports configuring alarms for key metrics, which can help you identify cluster problems and address them in a timely manner. This document describes how to configure alarms in the console.

Directions

Check whether alarms have been configured for a cluster

1. Log in to the [ES Console](#) and click a cluster ID/name in the cluster list to enter the cluster details page.
2. Select **Cluster Monitoring**. In the **Alarm Configuration** section, you can check whether alarms have been configured for the cluster. If not, you are recommended to configure an alarm policy to get notified of and address exceptions and risks in cluster running in a timely manner and ensure service stability.



Note :

You can also log in to the [Cloud Monitor Console](#) and check whether an alarm policy has been configured for a cluster in **Alarm Configuration > Alarm Policy**.

Customize alarm configuration

1. Log in to the Cloud Monitor Console and click **Create** on the **Alarm Policy** page.
2. On the policy creating page, configure the policy parameters.
 - **Policy Type**: Select **Elasticsearch Service**.
 - **Alarm Object**: Select the cluster for which to configure the alarm policy.
 - **Trigger**: **Trigger template** and **Configure a trigger** are supported. The latter is selected by default. For more information about custom configuration, see the description below. For more information about how to create a template, see the [Create a trigger template](#) section below.

Note :

Metric: For example, "CPU utilization". The statistical period is 1 minute or 5 minutes. All the metrics of an ES cluster are collected once per minute. If the statistical period is 1 minute, an alarm will be triggered as soon as a threshold is exceeded in the cluster. If 5 minutes is selected as the statistical period, an alarm will be triggered only when the threshold is continuously exceeded within 5 minutes.

Alarm frequency: For example, "Alarm once every 30 minutes" means that there will be only one alarm triggered every 30 minutes if a metric exceeds the threshold in several consecutive statistical periods.

Another alarm will be triggered only if the metric exceeds the threshold again in the next 30 minutes.

- **Alarm Channel:** Select the recipient group, valid time period, and receiving channel. For more information about the configuration method, see [Creating and Managing an Alarm Recipient Group](#).

3. After the configuration is completed, click **Finish**. Return to the **Alarm Policy** list and you can see the alarm policy just configured.

Note :

For more information about how to configure an alarm policy, see [Cloud Monitor Alarm Configuration](#).

Create Policy

Policy Name

1-20 Chinese, English chars or underlines

Remarks

1-100 Chinese and English characters or underlines

Policy Type

Elasticsearch Service

Existing: 2 item(s) and you can also create 298 policies

Alarm Object

☐ All Objects☒ Select some objects(0 selected)☐ Select instance group [Create instance group](#)

Region: Guangzhou



ID/Name

Private IP Address

Port



ID/Name

Private IP Address

9200

ID/Name

Private IP Addr...

Port

Press Shift to select multiple items

Trigger Condition

☐ Trigger Condition Template [Add Trigger Condition Template](#)☒ Configure trigger conditions☒ Indicator alarm

index_latency_avg

Measurement Pe...

>

0

ms

Continuous1

Alarm occurs every 1

[Add](#)

Alarm Channel

Recipient Object

Recipient Group

[Add Recipient Group](#)

User Group Name

User Name

full_resource_Access_g...
p

bkim



admin

kiyor, bkim

Valid Period

00:00:00

to

23:59:59

Receiving Channel



Email



SMS

Create a new trigger template

1. Click **Create a trigger template** in **Trigger**.
2. Click **Create** on the trigger template page.
3. On the template creating page, configure the policy type.
 - **Policy Type:** Select **Elasticsearch Service**.
 - **Use preset triggers:** Select this option and the system recommended alarm policy will appear.
4. After confirming everything is correct, click **Save**.

Trigger Condition Template

The trigger condition template function supports multiple use and unified modification of the same alarm rule. Click to view [Trigger Condition Template Document](#)

Create

Template Name: 1-20 Chinese, English chars or underlines

Remarks: 1-100 Chinese and English characters or underlines

Policy Type: Elasticsearch Service ☒ Use preset trigger conditions①

Trigger Condition ☒ Indicator alarm

cpu_usage_avg	Measurement Pe	>	90	%	Continuous5	Alarm occurs every 3	①✕
disk_usage_avg	Measurement Pe	>	80	%	Continuous5	Alarm occurs every 3	①✕
jvm_mem_usage_avg	Measurement Pe	>	85	%	Continuous5	Alarm occurs every 3	①✕
status	Measurement Pe	=	2		Continuous5	Alarm occurs every 3	①✕
status	Measurement Pe	=	1		Continuous5	Alarm occurs every 3	①✕

Add

Save Cancel

5. Return to alarm policy creating page and click **Refresh**. The alarm policy template just configured will appear.

Trigger Condition

☒ Trigger Condition Template [Add Trigger Condition Template](#)

ES default Alerting [Refresh](#)

☐ Indicator alarm

cpu_usage_avg	Measurement Pe	>	90	%	Continuous5	Alarm occurs every 3	①
disk_usage_avg	Measurement Pe	>	80	%	Continuous5	Alarm occurs every 3	①
jvm_mem_usage_avg	Measurement Pe	>	85	%	Continuous5	Alarm occurs every 3	①
status	Measurement Pe	=	2		Continuous5	Alarm occurs every 3	①
status	Measurement Pe	=	1		Continuous5	Alarm occurs every 3	①

Suggestions for Configuring Monitors and Alarms

Last updated : 2019-11-12 17:40:54

ES not only provides a number of monitoring metrics for running ES clusters to monitor their health, but also allows you to configure alarms for key metrics, so that you can identify cluster problems and address them in a timely manner. For more information, see [Viewing Monitoring Metrics](#) and [Configuring Alarms](#).

This document describes some metrics that require special attention during your use of an ES cluster, as well as recommended alarm configurations:

Metric	Suggested Alarm Configuration	Description
Cluster health status	The statistical period is 1 minute. If this value is ≥ 1 in 5 consecutive periods, an alarm will be triggered once every 30 minutes	<p>Value range:</p> <ul style="list-style-type: none">0: Green, which indicates that all primary and replica shards are available and the cluster is in the healthiest status.1: Yellow, which indicates that all the primary shards are available, but some replica shards are unavailable. In this case, the search results are still complete; however, the high availability of the cluster is affected to some extent, and there is a high risk of data loss.2: Red, which indicates that at least one primary shard and all its replicas are unavailable. When the cluster health status changes to red, some data has become unavailable, the search can only return partial data, and the requests allocated to a lost shard will return an exception. <p>The cluster health status is the most direct manifestation of the current health of the cluster. If it changes to yellow or red, you should troubleshoot and repair the problem in a timely manner to prevent data loss or service unavailability.</p>
Avg disk utilization	The statistical period is 1 minute. If this value is $> 80\%$ in 5 consecutive periods, an alarm will be triggered once every 30 minutes	<p>The avg disk utilization refers to the average of the disk utilization values of all nodes in the cluster. If the disk utilization of a node is too high, the node will not have sufficient disk capacity to accommodate the shards allocated to it, leading to failures in basic operations such as index creating and document adding. You are recommended to promptly clear the data or scale out your cluster when this value is above 75%.</p>

Metric	Suggested Alarm Configuration	Description
Avg JVM memory utilization	The statistical period is 1 minute. If this value is > 85% in 5 consecutive periods, an alarm will be triggered once every 30 minutes	The avg JVM memory utilization refers to the average of the JVM memory utilization values of all nodes in the cluster. A too high JVM memory utilization can lead to rejection of read and write operations, frequent GC, or even OOM. When this value exceeds the threshold, you are recommended to upgrade the node specification through vertical scaling.
Avg CPU utilization	The statistical period is 1 minute. If this value is > 90% in 5 consecutive periods, an alarm will be triggered once every 30 minutes	The avg CPU utilization refers to the average of the CPU utilization values of all nodes in the cluster. A too high average CPU utilization can lead to a decline in the processing capability of the cluster nodes or even downtime. If this value is too high, you should upgrade the node specification or reduce the number of requests based on the current node configuration of your cluster and your business.
Bulk rejection rate	The statistical period is 1 minute. If this value is > 0% in one period, an alarm will be triggered once every 30 minutes	The bulk rejection rate refers to the percentage of rejected bulk operations in all bulk operations performed by your cluster during a single period. When this value is greater than 0%, i.e., one or more bulk rejections have occurred, your cluster has reached the upper limit of its capability to process bulk operations, or an exception has occurred. In this case, you should troubleshoot and repair the problem in a timely manner; otherwise, bulk operations will be affected, or data loss will occur.
Query rejection rate	The statistical period is 1 minute. If this value is > 0% in one period, an alarm will be triggered once every 30 minutes	The query rejection rate refers to the percentage of rejected query operations in all query operations performed by your cluster during a single period. When this value is greater than 0%, i.e., one or more query rejections have occurred, your cluster has reached the upper limit of its capability to process query operations, or an exception has occurred. In this case, you should troubleshoot and repair the problem in a timely manner; otherwise, query operations will be affected.

Log Query

Querying Cluster Logs

Last updated : 2024-08-16 19:04:49

This document describes how to use ES cluster logs. You can use the logs to learn about the running status of the cluster, locate problems, and assist with cluster application development and OPS.

Querying Cluster Logs

1. Log in to the [ES Console](#) and click a cluster name to enter the cluster details page.
2. Select the **Cluster Logs** page to view the logs of the cluster.

ES has four types of logs: master log, slow search log, slow index log, and GC log. The log content includes log time, log level, and specific information.

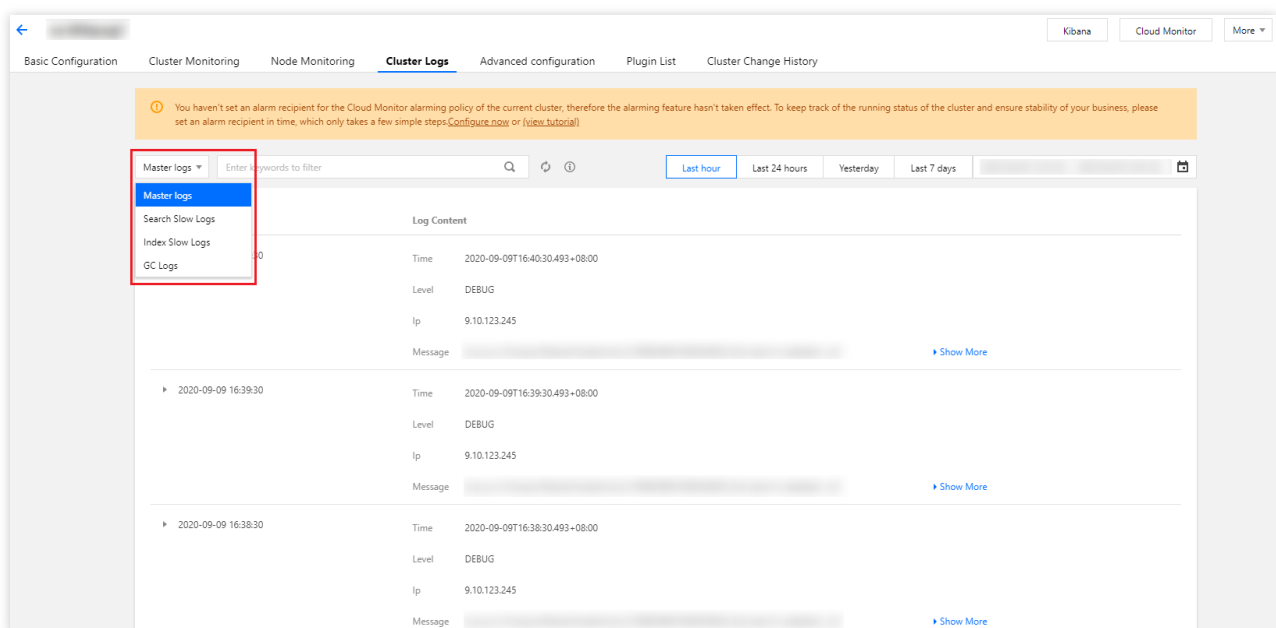
ES retains cluster logs of the last 7 days by default, which are displayed from newest to oldest and can be queried by time and keyword. In addition, you can also call ES APIs to adjust the log-related configuration. For example, for slow logs, you can set the time threshold that is considered as slow response during data querying or indexing.

3. In the search box on the log page, you can query related logs by time range and keyword. The keyword query syntax is the same as the Lucene query syntax.

Enter a keyword for querying, such as "YELLOW".

Specify a field-specific keyword, such as "message:YELLOW".

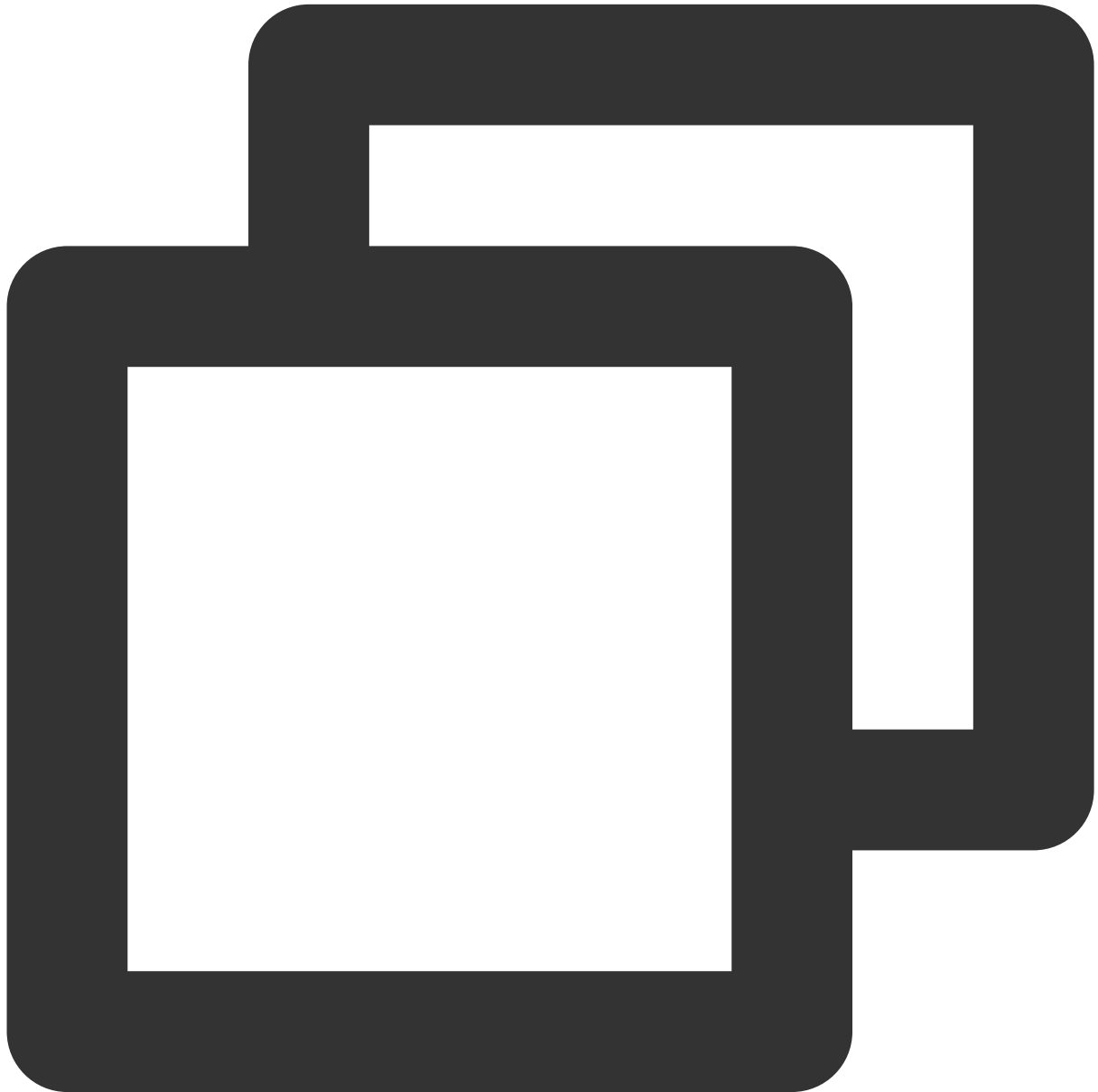
Combine multiple criteria for querying, such as `level: INFO and ip: 10.0.1.1`.



Log Description

Master log

It displays the time, level, and information of the log generated by the cluster and has different levels such as INFO, WARN, and DEBUG.



```
2019-2-14 08:00:00 10.0****  
INFO  
[o.e.c.r.a.AllocationService] [1550199698000783811] Cluster health status changed f
```

```
2019-2-14 02:30:02 10.0****
```

```
DEBUG
```

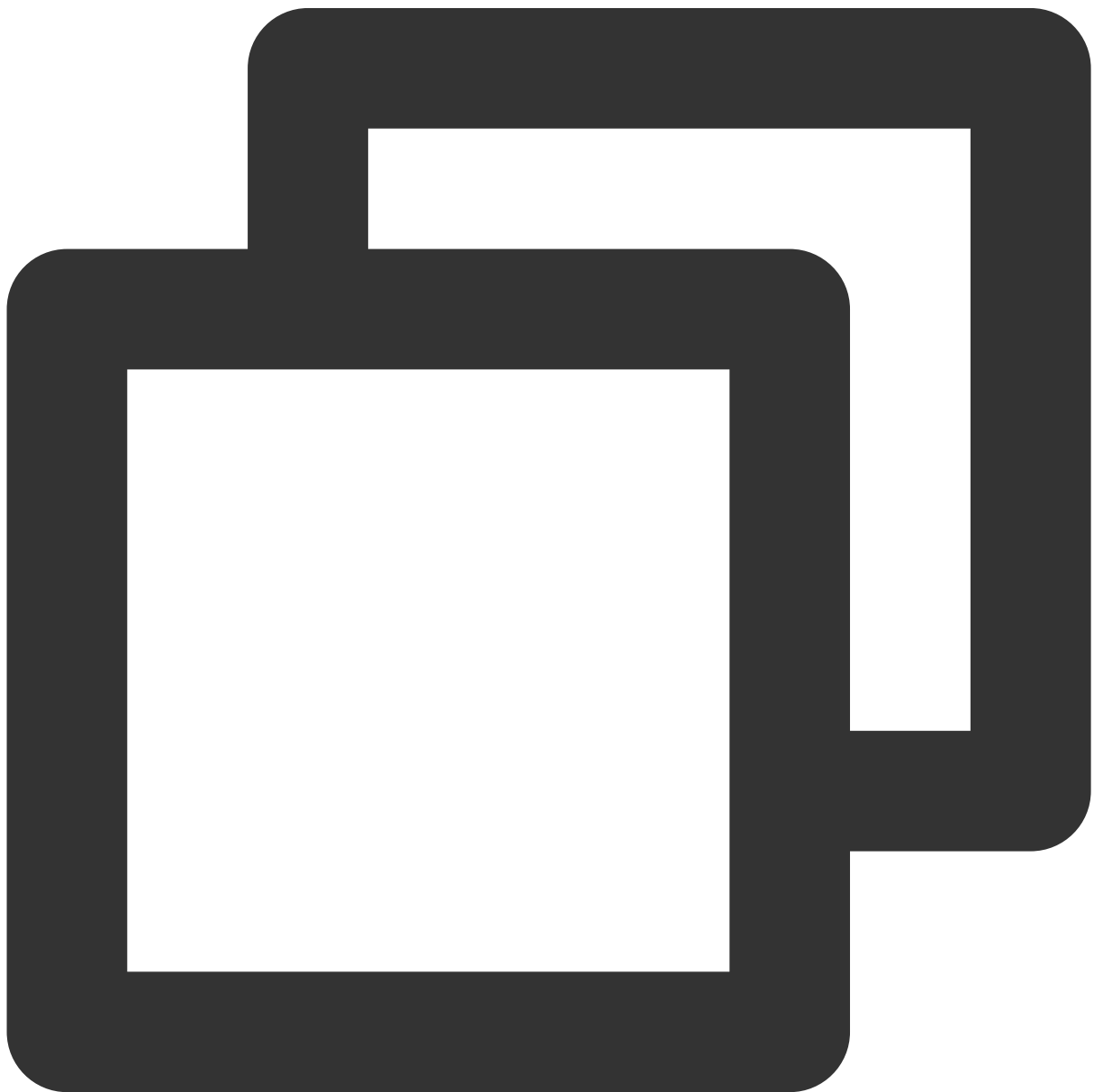
```
[o.e.a.a.i.d.TransportDeleteIndexAction] [1550199698000783811] failed to delete index  
org.elasticsearch.index.IndexNotFoundException: no such index
```

```
    at org.elasticsearch.cluster.metadata.Metadata.getIndexSafe(Metadata.java:475)
```

```
    at org.elasticsearch.cluster.metadata.MetadataDeleteIndexService.lambda$deleteIndex$0(MetadataDeleteIndexService.java:100)
```

Slow log

Slow log is used to capture querying and indexing requests that exceed the specified time threshold for tracking and analyzing very slow requests generated by user.



```
2018-10-28 12:04:17
```

```
WARN
```

```
[index.indexing.slowlog.index] [15402985020000001009] [pmc/wCALr6BfRm-sr3qOQuGX  
Xw] took[18.6ms], took_millis[18], type[articles], id[AWa41-J9c0s1mOPvR6F3], routin
```

Enable and adjust slow log

Slow log is disabled by default. To enable it, you need to define the specific action (query, fetch, or index), expected event record level (INFO, WARN, or DEBUG), and time threshold. You can enable and adjust related configuration based on your business needs.

To enable slow log, click **Kibana** in the top-right corner on the cluster details page to enter the Kibana page, call the Elasticsearch-related API via Dev Tools, or call the configuration modifying API through the client.

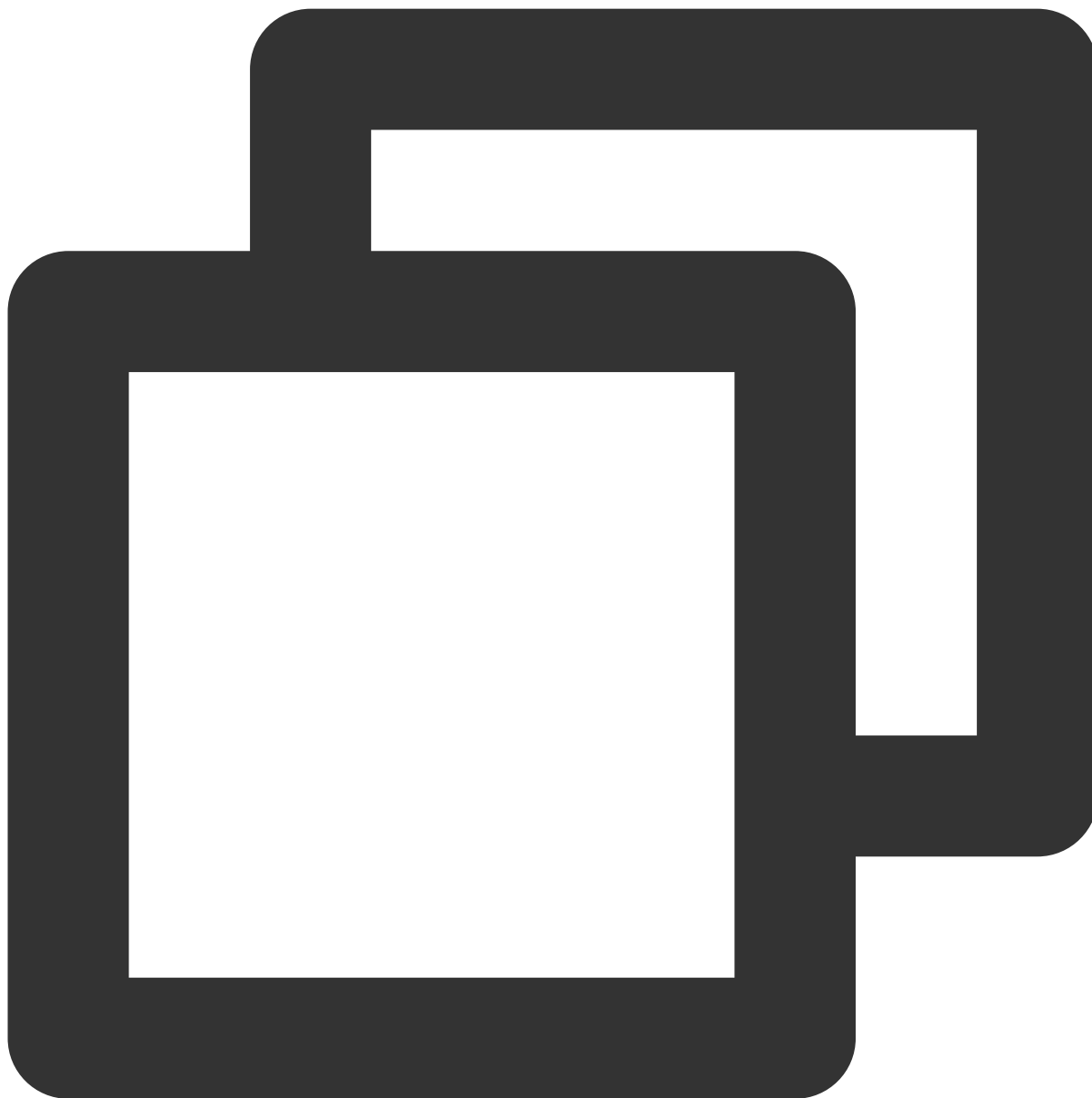
Configure all indices:



```
PUT */_settings
{
  "index.indexing.slowlog.threshold.index.debug" : "5ms",
  "index.indexing.slowlog.threshold.index.info" : "50ms",
  "index.indexing.slowlog.threshold.index.warn" : "100ms",
  "index.search.slowlog.threshold.fetch.debug" : "10ms",
  "index.search.slowlog.threshold.fetch.info" : "50ms",
  "index.search.slowlog.threshold.fetch.warn" : "100ms",
  "index.search.slowlog.threshold.query.debug" : "100ms",
  "index.search.slowlog.threshold.query.info" : "200ms",
  "index.search.slowlog.threshold.query.warn" : "1s"
```

```
}
```

Configure one single index:

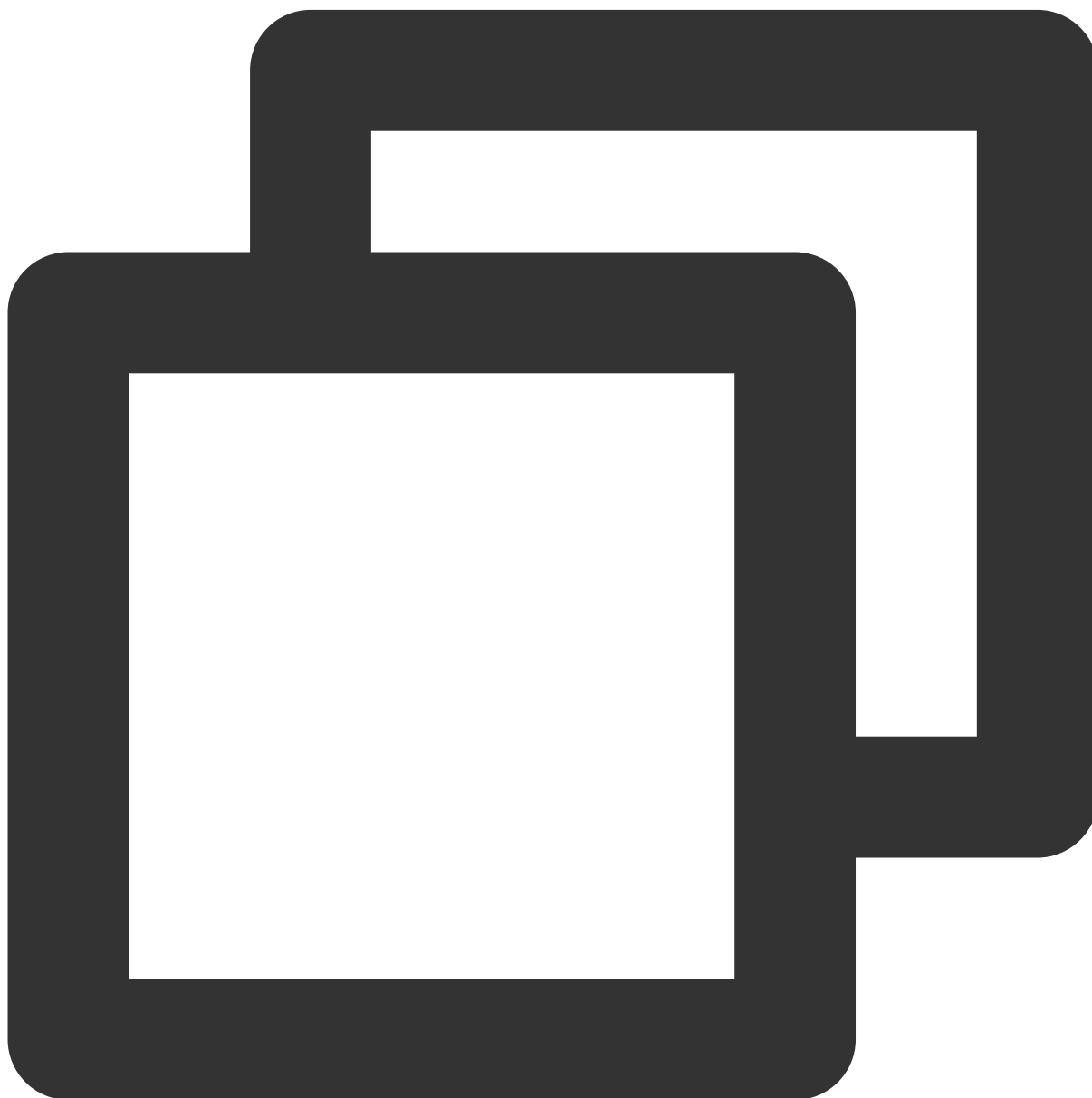


```
PUT /my_index/_settings
{
  "index.indexing.slowlog.threshold.index.debug" : "5ms",
  "index.indexing.slowlog.threshold.index.info" : "50ms",
  "index.indexing.slowlog.threshold.index.warn" : "100ms",
  "index.search.slowlog.threshold.fetch.debug" : "10ms",
  "index.search.slowlog.threshold.fetch.info" : "50ms",
  "index.search.slowlog.threshold.fetch.warn" : "100ms",
}
```

```
"index.search.slowlog.threshold.query.debug" : "100ms",  
"index.search.slowlog.threshold.query.info" : "200ms",  
"index.search.slowlog.threshold.query.warn" : "1s"  
}
```

GC log

GC log is enabled in ES by default. The following two specific GC logs display the time, node IP, and log level of the log.



2019-2-14 20:48:22

```
10.0.***
```

```
INFO
```

```
[o.e.m.j.JvmGcMonitorService] [1550199698000783711] [gc][380573] overhead, spent [3  
2019-2-14 10:04:09
```

```
10.0.***
```

```
WARN
```

```
[o.e.m.j.JvmGcMonitorService] [1550199698000784111] [gc][341943] overhead, spent [5
```


Data Backup

Automatic Snapshot Backup

Last updated : 2020-11-24 12:12:43

ES is capable of automatic backup, which involves automatically creating a snapshot of the primary index shards in a cluster and backing it up to COS. The snapshot can be restored to the cluster as needed.

Note :

This feature is currently free of charge.

Notes on Backup

- ES automatically performs snapshot backup and only retains the snapshot data for the past 7 days.
- Data generated from automatic snapshot backup can only be restored to the source cluster. For more information on how to restore to other regions, please see [Manual Snapshot](#).
- By default, automatic snapshot backup is performed at midnight, and you are recommended to select a time when the number of access requests to your cluster is small based on your business needs.
- The first snapshot of a cluster is a complete copy of all cluster data, and the time it takes to complete depends on the amount of data. Subsequent backups only retain the incremental difference between the saved snapshots and new data.
- When the health status of your cluster is RED, the automatic snapshot service will be suspended, so you are recommended to pay close attention to the health status of your cluster.

Directions

Enabling automatic snapshot backup

1. Log in to the [ES Console](#) and click a cluster name in the cluster list to enter the cluster details page.
2. On the "Advanced Configuration" page, enable automatic snapshot backup and configure the start time for automatic snapshot backup.

< es-l3vgjt9
Kibana
Cloud Monitor
Upgrade
More
Help

Basic Configuration
Cluster Monitoring
Cluster Logs
Advanced Configuration
Cluster Change History

Plugin Configuration

IK analyzer. Instructions for use: [Help Documentation](#)

Plugin	Details	Operation
IK analyzer	Custom dictionary Non-stopword dictionary: None Stopword dictionary: None	Update dictionary

Auto snapshot backup (free trial)

After enabled, Tencent Cloud ES will automatically create a snapshot for your cluster and back up data to COS once a day. You can view the snapshots and select one to recover the data. We recommend that you set the auto snapshot backup time to when the access traffic to the clusters is light. [Help Documentation](#)

Enable auto snapshot backup ☐

Viewing snapshot repository

Click **Kibana** on the cluster details page to enter the Kibana console where you can view the repository of all cluster snapshots using ES API on the "Dev Tools" page.

- View the cluster snapshot repository.

```
GET _snapshot?pretty
```

If only automatic snapshot is performed, the following message will be returned:

```
{
  "ES_AUTO_BACKUP": {
    "type": "cos",
    "settings": {
      "bucket": "es-ap-guangzhou",
      "base_path": "/es_backup/es-2s8x1b9u",
      "chunk_size": "500mb",
      "region": "ap-guangzhou",
      "compress": "true"
    }
  }
}
```

- View the snapshot information in the automatic snapshot repository.

```
GET _snapshot/ES_AUTO_BACKUP/_all?pretty
```

The returned result is as follows:

If there are no snapshots, the list will be empty.

```
{
  "snapshots": []
}
{
  "snapshots": [
    {
      "snapshot": "es-2s8x1b9u_20181220",
      "uuid": "gsXPYwb1SN0lTuj3eNs2gA",
      "version_id": 5060499,
      "version": "5.6.4",
      "indices": [
        ".kibana"
      ],
      "state": "SUCCESS",
      "start_time": "2018-12-20T08:00:12.336Z",
      "start_time_in_millis": 1545292812336,
      "end_time": "2018-12-20T08:00:12.945Z",
      "end_time_in_millis": 1545292812945,
      "duration_in_millis": 609,
      "failures": [],
      "shards": {
        "total": 1,
        "failed": 0,
        "successful": 1
      }
    }
  ]
}
```

Restoring data

```
POST _snapshot/ES_AUTO_BACKUP/es-2s8x1b9u_20181220/_restore
```

Using COS for Backup and Restoration

Last updated : 2022-05-05 16:58:45

Creating a Repository

You can create a repository by running the following command:

```
PUT _snapshot/my_cos_backup
{
  "type": "cos",
  "settings": {
    "app_id": "xxxxxxx",
    "access_key_id": "xxxxxx",
    "access_key_secret": "xxxxxxx",
    "bucket": "xxxxxx",
    "region": "ap-guangzhou",
    "compress": true,
    "chunk_size": "500mb",
    "base_path": "/"
  }
}
```

- app_id: APPID of your Tencent Cloud account.
- access_key_id: SecretId of your Tencent Cloud API key.
- access_key_secret: SecretKey of your Tencent Cloud API key.
- bucket: COS bucket name, **which cannot contain the** `-{appId}` **prefix.**
- region: COS bucket region, **which must be the same region as that of the ES cluster.** For more information about regions, see [Regions and Availability Zones](#).
- compress : true by default, which means the stored index metadata will be compressed.
- base_path: Backup directory.

Listing Repository Information

You can get repository information via `GET _snapshot` or get the information of a specified repository via `GET _snapshot/my_cos_backup` .

Creating a Snapshot Backup

Backing up all indices

Back up all indices in the ES cluster to the repository `my_cos_backup` and name it `snapshot_1` .

```
PUT _snapshot/my_cos_backup/snapshot_1
```

This command will return a response immediately and be executed asynchronously in the background until the end. If you want to wait for the execution to complete before returning, you can add the `wait_for_completion` parameter. **The duration of the command execution depends on the index size.**

```
PUT _snapshot/my_cos_backup/snapshot_1?wait_for_completion=true
```

Backing up specified indices

You can specify the indices to be backed up when creating a snapshot. **If the value of the `indices` parameter is multiple indices, they should be separated by `,` with no spaces.**

```
PUT _snapshot/my_cos_backup/snapshot_2
{
  "indices": "index_1,index_2"
}
```

Querying a Snapshot

Query the information of a single snapshot:

```
GET _snapshot/my_cos_backup/snapshot_1
```

This command returns the information about the snapshot:

Note :

When the value of the `state` field is `SUCCESS` , the snapshot backup is completed.

```
{
  "snapshots": [
    {
      "snapshot": "snapshot_1",
      "uuid": "zUSugNiGR-OzH0CCcgLmQ",
      "version_id": 5060499,
```

```
"version": "5.6.4",
"indices": [
  "index_1",
  "index_2"
],
"state": "SUCCESS",
"start_time": "2018-05-04T11:44:15.975Z",
"start_time_in_millis": 1525434255975,
"end_time": "2018-05-04T11:45:29.395Z",
"end_time_in_millis": 1525434329395,
"duration_in_millis": 73420,
"failures": [],
"shards": {
  "total": 3,
  "failed": 0,
  "successful": 3
}
}
```

Deleting a Snapshot

Delete a specified snapshot:

```
DELETE _snapshot/my_cos_backup/snapshot_1
```

Note :

If the creation of the snapshot hasn't been completed, the snapshot deletion command will still be executed and cancel the creation of the snapshot.

Restoring Indices from a Snapshot

1. Restore all indices backed up in a snapshot to the ES cluster:

```
POST _snapshot/my_cos_backup/snapshot_1/_restore
```

- If `snapshot_1` contains five indices, all of them will be restored to the ES cluster.
- You can also rename the indices through an additional option which allows you to match the index name by pattern and provide a new name through the restoration process. Use this option if you want to restore old data to verify the content or perform other operations without replacing existing data.

2. Restore a single index from a snapshot and provide an alternate name:

```
POST /_snapshot/my_cos_backup/snapshot_1/_restore
{
  "indices": "index_1",
  "rename_pattern": "index_(.+)",
  "rename_replacement": "restored_index_$1"
}
```

- `indices`: Only restores `index_1` and ignores other indices in the snapshot.
- `rename_pattern`: Finds the index being restored that can be matched by the specified pattern.
- `rename_replacement`: Renames the matching index to the name specified in this parameter.

Querying the Status of Snapshot Restoration

You can check the status of snapshot restoration and monitor the progress by running the `_recovery` command.

1. You can call the following API separately in the specified index to be restored:

```
GET index_1/_recovery
```

2. This command will return the restoration status of each shard of the specified index:

```
{
  "sonested": {
    "shards": [
      {
        "id": 1,
        "type": "SNAPSHOT",
        "stage": "INDEX",
        "primary": true,
        "start_time_in_millis": 1525766148333,
        "total_time_in_millis": 8718,
        "source": {
          "repository": "my_cos_backup",

```

```
"snapshot": "snapshot",
"version": "5.6.4",
"index": "index_1"
},
"target": {
  "id": "TlzmXJHwSqyv4rhyQfRkow",
  "host": "10.0.0.6",
  "transport_address": "10.0.0.6:9300",
  "ip": "10.0.0.6",
  "name": "node-1"
},
"index": {
  "size": {
    "total_in_bytes": 1374967573,
    "reused_in_bytes": 0,
    "recovered_in_bytes": 160467084,
    "percent": "11.7%"
  },
  "files": {
    "total": 132,
    "reused": 0,
    "recovered": 20,
    "percent": "15.2%"
  },
  "total_time_in_millis": 8716,
  "source_throttle_time_in_millis": 0,
  "target_throttle_time_in_millis": 0
},
"translog": {
  "recovered": 0,
  "total": 0,
  "percent": "100.0%",
  "total_on_start": 0,
  "total_time_in_millis": 0
},
"verify_index": {
  "check_index_time_in_millis": 0,
  "total_time_in_millis": 0
}
}
]
```

- **type**: Describes the nature of the restoration, which means this shard is restored from a snapshot.

- **source:** Describes the specific snapshot and repository as the source of restoration.
- **percent:** Describes the restoration status. 94% of the files in the particular shard has now been restored, so it will be completely restored soon.

The output lists all the indices that are being restored and all the shards in them. Each shard contains statistics such as the start/stop time, duration, percentage of restoration, and number of bytes transferred.

Canceling Snapshot Restoration

```
DELETE /restored_index_1
```

If `restored_index_1` is being restored, this deletion command will stop the restoration and delete all data that has been restored to the cluster.

Upgrade

ES Version Upgrade Check

Last updated : 2020-04-22 15:34:24

Different versions of Elasticsearch have certain incompatible configuration items, and if you have set such items, an upgrade may affect the use of your cluster. You can use the upgrade check feature to check whether there are any incompatible configuration items and adjust them accordingly. The following describes the check to be performed when upgrading the Elasticsearch version.

On the details page in the [console](#), click **Upgrade** in the top-right corner. For detailed directions, please see [Upgrading ES Clusters](#).

Configuration Check for Elasticsearch Upgrade from 5.x to 6.x

Checklist of configuration items

No.	Configuration Level	Configuration Information	Compatibility	Description
1	Cluster	Snapshot settings	CRITICAL	The <code>cluster.routing.allocation</code> setting has been disused since v6.0 Breaking changes in 6.0
2	Cluster	Store throttling settings	CRITICAL	The <code>indices.store.throttle</code> and <code>indices.store.throttle.max</code> settings have been disused since v6.0. For more information, please see Breaking changes in 6.0
3	Index	Similarity settings	WARNING	The <code>index.similarity.base</code> setting has been disused since v6.0. For more information, please see Breaking changes in 6.0
4	Index	Shadow replicas settings	CRITICAL	The <code>index.shared_filesystem</code> setting has been disused since v6.0. For more information, please see Breaking changes in 6.0
5	Index	Index store settings	CRITICAL	The <code>default</code> <code>index.store</code> settings have been disused since v6.0. For more information, please see Breaking changes in 6.0

No.	Configuration Level	Configuration Information	Compatibility	Description
6	Index	Index store throttling settings	CRITICAL	The <code>index.store.throttle.t</code> and <code>index.store.throttle.max_</code> settings have been disused since v6.0. For more information, see Breaking changes in 6.0 .
7	Index	<code>include_in_all</code> index mapping parameter	WARNING	The <code>include_in_all</code> mapping parameter is deprecated in v6.0 or above (indices created by v5.x are compatible after upgrade to v6.x). For more information, see Breaking changes in 6.0 .
8	Index	<code>index.version.created</code>	CRITICAL	The <code>index.version.created</code> setting is deprecated in v6.0 or above. For example, you cannot create an index with version 7.x; instead, you have to reindex the index before upgrading.
9	Index template	Similarity settings	CRITICAL	The <code>index.similarity.base</code> setting is deprecated in v6.0 or above. For more information, please see Breaking changes in 6.0 . If this item is in the template, the template cannot be used to create indices after the upgrade.
10	Index template	Shadow replicas settings	CRITICAL	The <code>index.shared_filesystem</code> setting is deprecated in v6.0 or above. For more information, please see Breaking changes in 6.0 . If this item is in the template, the template cannot be used to create indices after the upgrade.
11	Index template	Index store settings	CRITICAL	The <code>default</code> <code>index.store</code> settings are deprecated in v6.0 or above. For more information, please see Breaking changes in 6.0 . If this item is in the template, the template cannot be used to create indices after the upgrade.
12	Index template	Index store throttling settings	CRITICAL	The <code>index.store.throttle.t</code> and <code>index.store.throttle.max_</code> settings have been disused since v6.0. For more information, see Breaking changes in 6.0 . If this item is in the template, the template cannot be used to create indices after the upgrade.
13	Index template	<code>include_in_all</code> mapping parameter	CRITICAL	The <code>include_in_all</code> mapping parameter is deprecated in v6.0 or above. For more information, please see Breaking changes in 6.0 . If this item is in the template, the template cannot be used to create indices after the upgrade.
14	Index template	<code>_all</code> mapping metafield	CRITICAL	The <code>_all</code> mapping metafield is deprecated in v6.0 or above. For more information, please see Breaking changes in 6.0 . If this item is in the template, the template cannot be used to create indices after the upgrade.

No.	Configuration Level	Configuration Information	Compatibility	Description
15	Index template	Mapping types	CRITICAL	Multiple mapping types have been c please see Removal of mapping type template cannot be used to create ir

- WARNING: upgrade can still be performed even if the check fails. The settings for this type of check items will be ignored after the upgrade.
- CRITICAL: upgrade cannot be performed if the check fails. The settings for this type of check items are incompatible in the target version.

How to adjust incompatible configuration items

Cluster level

- Snapshot settings

You can cancel these settings (including persistent and transient) through the ES cluster settings update API

(`PUT _cluster/settings`):

```
PUT _cluster/settings
{
  "persistent": {
    "cluster.routing.allocation.snapshot.relocation_enabled": null
  },
  "transient": {
    "cluster.routing.allocation.snapshot.relocation_enabled": null
  }
}
```

- Store throttling settings

You can cancel these settings (including persistent and transient) through the ES cluster settings update API

(`PUT _cluster/settings`):

```
PUT _cluster/settings
{
  "persistent": {
    "indices.store.throttle.type": null,
    "indices.store.throttle.max_bytes_per_sec": null
  },
  "transient": {
    "indices.store.throttle.type": null,

```

```
"indices.store.throttle.max_bytes_per_sec": null
}
```

Index level

- Similarity settings

These settings will be ignored after upgrade to 6.x, which will not affect the upgrade though. To cancel them, follow the steps below:

- The index needs to be closed before you modify these settings, and the closed index cannot be read or written.

Close the index:

```
POST my_index/_close
```

- Cancel these settings through the ES index settings update API:

```
PUT my_index/_settings
{
  "index.similarity.base.*": null
}
```

- Then, open the index by running the following command:

```
POST my_index/_open
```

- Shadow replicas settings

- The index needs to be closed before you modify these settings, and the closed index cannot be read or written.

Close the index:

```
POST my_index/_close
```

- Cancel these settings through the ES index settings update API:

```
PUT my_index/_settings
{
  "index.shared_filesystem": null,
  "index.shadow_replicas": null
}
```

- Then, open the index by running the following command:

```
POST my_index/_open
```

- Index store settings

- The index needs to be closed before you modify these settings, and the closed index cannot be read or written.

Close the index:

```
POST my_index/_close
```

- Cancel these settings through the ES index settings update API:

```
PUT my_index/_settings
{
  "index.store.type": null
}
```

- Then, open the index by running the following command:

```
POST my_index/_open
```

- Index store throttling settings

Cancel these settings through the ES index settings update API:

```
PUT my_index/_settings
{
  "settings": {
    "index.store.throttle.type": null,
    "index.store.throttle.max_bytes_per_sec": null
  }
}
```

- The `include_in_all` index mapping parameter

A previously created index containing this parameter is compatible after the upgrade and does not need to be repaired.

Index template level

1. Use the `GET _template/my_template` API to get the incompatible template `my_template`. There are three incompatible items: index store settings, index template mapping parameter (`include_in_all`), and index template mapping metafield (`_all`).

```
{
  "my_template": {
    "order": 0,
    "template": "my_*",
    "settings": {
      "index": {
        "store": {
          "throttle": {
            "max_bytes_per_sec": "10m"
          }
        }
      }
    }
  }
}
```

```
}
}
},
"mappings": {
  "my_type": {
    "_all": {
      "enabled": true
    },
    "properties": {
      "my_field": {
        "type": "text",
        "include_in_all": true
      }
    }
  }
},
"aliases": {}
}
```

2. After copying and removing incompatible configuration items from the template, use the `PUT _template/my_template` API to update the template.

```
PUT _template/my_template
{
  "order": 0,
  "template": "my_*",
  "settings": {
  },
  "mappings": {
    "my_type": {
      "properties": {
        "my_field": {
          "type": "text"
        }
      }
    }
  },
  "aliases": {}
}
```

Upgrading ES Clusters

Last updated : 2020-09-04 16:47:53

Tencent Cloud ES offers an upgrade feature that allows you to upgrade ES and X-Pack. You can upgrade your cluster based on your business needs to achieve seamless business transition.

The upgrade feature is currently made available through an allowlist. If you want to enable it, please [submit a ticket](#) for application.

Supported Upgrade Types

ES supports the following two upgrade types.

1. Upgrade of the Elasticsearch version

Source Elasticsearch Version	Target Elasticsearch Version
Low version (such as v5.x)	High version (such as v6.x)

2. Upgrade of X-Pack

Source X-Pack Edition	Target X-Pack Edition
Open Source Edition	Basic or Platinum Edition
Basic Edition	Platinum Edition

X-Pack description:

The Basic Edition and Platinum Edition integrate Elasticsearch's official X-Pack plugin which including features such as security, SQL, machine learning, and monitoring. The former only comes with certain SQL features and monitoring, while the latter has all X-Pack features. For more information, please see [X-Pack](#).

- You can perform only one of the two types of upgrade above at a time. When upgrading from a lower version of the Open Source Edition, you can choose to upgrade to the Basic Edition at the same time.
- v5.x is available only in the Open Source but not the Basic or Platinum Edition.

Restart After Upgrade

1. As far as restart is concerned, when an upgrade is performed, the cluster may:

- Not restart.
- Restart on a rolling basis: the service can be accessed normally during upgrade, but the performance may be affected; therefore, you are recommended to upgrade during off-peak hours.
- Fully restart: the service is inaccessible during upgrade. This option should be used with caution.

2. In case of a full restart of the cluster, as the service will become inaccessible, please note the following:

Elasticsearch supports user authentication starting from a certain edition to improve the security of cluster access. When an Elasticsearch edition with this feature unavailable or not enabled is upgraded to an edition with this feature enabled, according to the official design requirements of Elasticsearch, the cluster needs to be fully restarted and will be inaccessible during the full restart; therefore, this option should be used with caution.

User authentication is supported as follows:

- Unavailable in the Open Source Edition.
- Supported in v6.8 or higher of the Basic Edition (you can choose to enable or disable it).
- Enabled by default in v6.4 or higher of the Platinum Edition.

Example:

- For upgrade from the Open Source Edition v6.4 (with user authentication unavailable) to the Basic Edition v6.8 (with user authentication not enabled), no full restart of the cluster is needed.
- For upgrade from the Open Source Edition v6.4 (with user authentication unavailable) to the Basic Edition v6.8 (with user authentication enabled), full restart of the cluster is needed.
- For upgrade from the Basic Edition v6.8 (with user authentication not enabled) to the Platinum Edition v6.8 (with user authentication enabled by default), full restart of the cluster is needed.
- For upgrade from the Basic Edition v6.8 (with user authentication enabled) to the Platinum Edition v6.8 (with user authentication enabled by default), no full restart of the cluster is needed.

Notes on Upgrade

Compatibility and use of Elasticsearch 5.x and 6.x

1. Multi-type index.

Starting from 6.x, Elasticsearch no longer supports multiple types within a single index. After you upgrade from v5.x

to v6.x, creating a multi-type index will cause an error. Multi-type indices previously created on v5.x will not be affected, and data can be written to them normally.

2. Accessing a cluster by using curl.

When accessing a cluster by using curl, you need to add the `header -H 'Content-Type: application/json'` request.

```
curl -XPUT http://10.0.0.2:9200/china/city/beijing -H 'Content-Type: application/json' -d '{
  "name": "Beijing",
  "province": "Beijing",
  "lat": 39.9031324643,
  "lon": 116.4010433787,
  "x": 6763,
  "level.range": 4,
  "level.level": 1,
  "level.name": "Tier-1 city",
  "y": 6381,
  "cityNo": 1
}'
```

3. Compatibility of configuration items.

Different versions of Elasticsearch have certain incompatible configuration items, and if you have set such items, an upgrade may affect the use of your cluster. The upgrade feature of ES comes with a process to check configuration items as well as instructions for adjustment, as described below in [Upgrade check](#).

4. For more information, please see [Breaking changes in 6.0](#).

Upgrade Process

To upgrade the Elasticsearch version, you need to complete upgrade check and data backup first. The upgrade process can be started only after those two steps are successfully completed.

1. Upgrade check

This is supported only when upgrading the Elasticsearch version.

Check whether the two versions have any incompatible configuration items. If the check fails, the process will terminate, and you can view the specific check items and corresponding solutions. For more information, please see [Upgrade Check](#). You can also choose to only perform an upgrade check before upgrade so as to see whether your cluster meets the conditions for upgrade.

2. Snapshot backup

This is supported only when upgrading the Elasticsearch version.

Before the upgrade, ES will perform snapshot backup of your cluster, so that you can restore it from the snapshot in case of an upgrade failure. Therefore, if the snapshot backup fails, the upgrade process will also terminate. **The time it takes to complete the snapshot backup depends on the amount of data in your cluster. If automatic snapshot backup is not enabled for your cluster, and the data volume is high, it will take longer to create a snapshot for the first time.**

3. Upgrade process and cluster restart

For a cluster on v6.x or higher, you can upgrade X-Pack (i.e., from the Open Source Edition to the Basic or Platinum Edition). During the upgrade, the service needs to be restarted as follows:

- To upgrade from the Open Source Edition to the Basic Edition (with user authentication unavailable or not enabled), your cluster must be restarted on a rolling basis, its health status must be green, and there can be no single-replica or closed indices in it. During the upgrade, the service is accessible, but its performance will be partially affected. As a result, you are recommended to perform the upgrade during off-peak hours.
- To upgrade from the Open Source Edition to the Basic Edition (with user authentication enabled) or from the Basic Edition (with user authentication unavailable or not enabled) to the Platinum Edition, your cluster must have user authentication enabled and be fully restarted. During the full restart, your cluster will be inaccessible until the upgrade is completed; therefore, this option should be used with caution.

How to Upgrade a Cluster

- Log in to the [ES Console](#) to enter the **Cluster List** page, and click **More** -> **Adjust Configuration** in the top-right corner.

The screenshot displays the Tencent Cloud Elasticsearch Service console interface. At the top, there's a navigation bar with tabs for 'Basic Configuration', 'Cluster Monitoring', 'Node Monitoring', 'Cluster Logs', 'Advanced configuration', 'Plugin List', and 'Cluster Change History'. On the right side of the navigation bar, there's a 'More' dropdown menu that is currently open, showing options: 'Upgrade', 'Restart', 'Adjust Configuration', and 'Terminate'. The 'Upgrade' option is highlighted with a red box. Below the navigation bar, there's a warning message in a yellow box stating: 'You haven't set an alarm recipient for the Cloud Monitor alarming policy of the current cluster, therefore the alarming feature hasn't taken effect. To keep track of the running status of the cluster and ensure stability of your business, please set an alarm recipient in time, which only takes a few simple steps. [Configure now](#) or [view tutorial](#).' The main content area is divided into two sections. The left section, titled 'Basic Info', shows 'Cluster Name: Demo', 'Cluster ID', and 'Cluster Status: Normal'. The right section, titled 'Cluster Configuration', shows a table with columns: 'Node Type', 'Quantity', 'Specification', 'Node Storage', and 'Total Storage'. The table contains one row: 'Data Node', '3', '2-core 4 GB Standard', '100GB x 1 SSD', and '300GB'. There is an 'Adjust Configuration' button next to the table.

- Choose to upgrade Elasticsearch or X-Pack in the upgrade dialog box.

Upgrading the Elasticsearch version

1. Select **Upgrade the Elasticsearch version** as the **Upgrade Type** in the upgrade dialog box.
2. Select the version that you want to upgrade to in the **Elasticsearch Version** drop-down list.

- If you do not choose to upgrade X-Pack at the same time, the original X-Pack Edition will be retained when upgrading to a higher version by default.
- Note: when upgrading the major version of Elasticsearch, such as from v5.x to v6.x, you can upgrade X-Pack from the Open Source Edition to the Basic Edition at the same time. You are recommended to select X-Pack **Basic Edition**, which contains features such as monitoring and SQL. When upgrading to the Basic Edition v6.8 or higher, you can also choose to enable user authentication; in this case, the cluster needs to be fully restarted. During the full restart, your cluster will be inaccessible until the upgrade is completed; therefore, this option should be used with caution.

3. Before the upgrade, the cluster status and configuration will be checked first to determine whether your cluster can be upgraded. You can select **Upgrade Check Only**. After you click **OK**, only an upgrade check will be performed, and the upgrade command will not be executed. You can view the check result in the **Cluster Change History** on the details page.

If you want to perform a major version upgrade for Elasticsearch (e.g., from v5.x to v6.x), as some configuration items at the cluster or index level are not compatible, you need to perform an **upgrade check** to determine whether your cluster can be upgraded. Incorrect configuration items that trigger alarms need to be adjusted as appropriate. For more information, please see [ES Version Upgrade Check](#)

4. Read and indicate your consent to the upgrade notice and click **OK** to start the upgrade (if you select **Upgrade Check Only**, an upgrade check will be started).

Upgrade

Version Upgrade Description

- To upgrade the cluster version, you need to check the cluster status and version compatibility to see if the current version can be upgraded to the target version.
- If upgrade is supported, a snapshot backup will be performed for the cluster first. We recommend that you perform the upgrade when the cluster load is low.
- After the upgrade starts, you can view the progress of the upgrade in Cluster Change History.

Cluster ID	Cluster Name	Current Elasticsearch vers...	Current X-Pack version
		6.8.2	Basic edition

Upgrade Type

☒ Upgrade Elasticsearch version
 ☐ Upgrade [X-Pack] Edition

Elasticsearch Version

7.5.1

X-Pack

☒ Basic edition
 When upgrading to V6.4 or above, you can also choose to upgrade the X-Pack to the basic version, which supports features such as monitoring and SQL.
 ☒ Enable user authentication
 Basic edition above V6.8 can enable this feature if needed. When it is enabled, the cluster needs to be full restarted. Service is not accessible during the restart. At the same time, any access to the ES cluster must be authenticated. You should modify the business code in advance to support authentication, otherwise, the business will be inaccessible. [Upgrade Intro](#)

Upgrade Check Reminder

☐ Elasticsearch API and SDK may have incompatibility issues, therefore, you should check the relevant business code in advance, especially the content that cannot be automatically checked by the system and needs to be checked manually. [Learn more](#)

Upgrade Check Only

☐ After you select this checkbox, only upgrade check is performed to determine whether the cluster can be upgraded without executing the upgrade command. The upgrade check does cover the content that needs manual check. The check results can be viewed in "Cluster Change History".

Operation Instruction

☐ During the upgrade check or upgrade, the process cannot be canceled. However, you can still write data to and read data from the cluster, but cannot modify other configurations, such as scaling out.

Confirm

Cancel

Upgrading X-Pack

1. Select **Upgrade Type > Upgrade [X-Pack] Edition** in the upgrade dialog box.

2. Select the X-Pack edition that you want to upgrade to in **X-Pack**.
3. Click **OK** to start the upgrade.

- Note on upgrade of X-Pack: currently, X-Pack can be upgraded for v6.x or higher but not v5.x (v5.x is available only in the Open Source but not the Basic or Platinum Edition).
- The upgrade process varies by edition. Please note the following:
- To upgrade to the Basic Edition (with user authentication unavailable or not enabled), your cluster needs to be restarted on a rolling basis, during which the service will be affected momentarily; therefore, you are recommended to perform the upgrade during off-peak hours.
- To upgrade from the Open Source Edition to the Basic Edition (with user authentication enabled) or from the Basic Edition (with user authentication unavailable or not enabled) to the Platinum Edition, your cluster must have user authentication enabled and be fully restarted. During the full restart, your cluster will be inaccessible until the upgrade is completed; therefore, this option should be used with caution.

Upgrade

Version Upgrade Description

- To upgrade the cluster version, you need to check the cluster status and version compatibility to see if the current version can be upgraded to the target version.
- If upgrade is supported, a snapshot backup will be performed for the cluster first. We recommend that you perform the upgrade when the cluster load is low.
- After the upgrade starts, you can view the progress of the upgrade in Cluster Change History.

Cluster ID	Cluster Name	Current Elasticsearch vers...	Current X-Pack version
		6.4.3	Open source edition

Upgrade Type

☐ Upgrade Elasticsearch version ☒ Upgrade [X-Pack] Edition

X-Pack

Open source edition(Current version)

Basic edition

Platinum edition

Original Configuration Fees

0.51 USD/hour

New Configuration Fees

0.51 USD/hour

Confirm

Cancel

4. After the upgrade starts, you can check the upgrade progress in the **Cluster Change History** on the cluster details page.

Kibana

Cloud Monitor

More

Basic Configuration

Cluster Monitoring

Node Monitoring

Cluster Logs

Advanced configuration

Plugin List

Cluster Change History

You haven't set an alarm recipient for the Cloud Monitor alarming policy of the current cluster, therefore the alarming feature hasn't taken effect. To keep track of the running status of the cluster and ensure stability of your business, please set an alarm recipient in time, which only takes a few simple steps [Configure now](#) or [view tutorial](#).

All

Last 24 hours

Last 7 days

Last 30 days

2018-01-01 00:00:00 ~ 2020-08-21 17:44:36

Time	Operation	Details	Progress
▼ 2020-08-21 17:23:10	Create	--	<div><div>100% Collapse</div><div><div>Prepare resources</div><div>Progress 100% 2020-08-21 17:30:32</div></div><div><div>Deploy ES Clusters</div><div>Progress 100% 2020-08-21 17:33:05</div></div><div><div>Deploy Kibana</div><div>Progress 100% 2020-08-21 17:35:51</div></div><div><div>Configure load balancing</div><div>Progress 100% 2020-08-21 17:35:55</div></div></div>