Cloud Search Service(CSS)
1.1.8
User Guide

Issue	01
Date	2024-10-30

HUAWEI CLOUD COMPUTING TECHNOLOGIES CO., LTD.

Copyright © Huawei Cloud Computing Technologies Co., Ltd. 2024. All rights reserved.

No part of this document may be reproduced or transmitted in any form or by any means without prior written consent of Huawei Cloud Computing Technologies Co., Ltd.

Trademarks and Permissions


and other Huawei trademarks are the property of Huawei Technologies Co., Ltd.
All other trademarks and trade names mentioned in this document are the property of their respective holders.

Notice

The purchased products, services and features are stipulated by the contract made between Huawei Cloud and the customer. All or part of the products, services and features described in this document may not be within the purchase scope or the usage scope. Unless otherwise specified in the contract, all statements, information, and recommendations in this document are provided "AS IS" without warranties, guarantees or representations of any kind, either express or implied.

The information in this document is subject to change without notice. Every effort has been made in the preparation of this document to ensure accuracy of the contents, but all statements, information, and recommendations in this document do not constitute a warranty of any kind, express or implied.

Huawei Cloud Computing Technologies Co., Ltd.

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

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

Contents

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

1 Product Overview

1.1 What Is Cloud Search Service?

CSS

Cloud Search Service (CSS) is a fully hosted distributed search service based on Elasticsearch. You can use it for structured and unstructured data search, and use AI vectors for combine search, statistics, and reports. CSS is a fully managed cloud service of the ELK Stack and is compatible with open-source Elasticsearch, Kibana, and Cerebro.

Elasticsearch is a distributed search engine that can be deployed in standalone or cluster mode and provides hosted distributed search engine services. As the heart of the ELK Stack, Elasticsearch clusters support multi-condition search, statistical analysis, and create visualized reports of structured and unstructured text. For details about Elasticsearch, see the **Elasticsearch: The Definitive Guide**.

CSS can be automatically deployed, allowing you to quickly create Elasticsearch clusters. It provides the search engine optimization practices and does not require your O&M. Additionally, it has a robust monitoring system to present you key metrics, including clusters and query performance so that you can focus on the business logic.

Functions

- Compatible with Elasticsearch

  Freely use native Elasticsearch APIs and other software in the ecosystem, such as Beats and Kibana.
- Support various data sources

  A few simple configurations can allow you to smoothly connect to multiple data sources, such as FTP, OBS, HBase, and Kafka. No extra coding is required.
- One-click operation

  One-click cluster application, capacity expansion, and restart from small-scale testing to large-scale rollout
- User-defined snapshot policies

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

Trigger backup snapshots manually or configure an automated schedule.

1.2 Advantages

CSS has the following features and advantages.

Efficient and Ease of Use

You can get insights from terabyte-scale data in milliseconds. In addition, you can use the visualized platform for data display and analysis.

Flexible and Scalable

You can request resources as needed and perform capacity expansion online with zero service interruption.

Easy O&M

CSS is a fully-managed, out-of-the-box service. You can start using it with several clicks, instead of managing clusters.

Kernel Enhancement

Speed up data import, decouple storage and computing resources, isolate read and write requests, and take advantage of our high-performance vector search engine – all these at an affordable price.

High Reliability

You can choose to trigger snapshots manually or on a periodic basis for backup and restore snapshots to the current or other clusters. Snapshots of a cluster can be restored to another cluster to implement cluster data migration.

- Automatic backup using snapshots

  CSS provides the backup function. You can enable the automatic backup function on the CSS management console and set the backup period based on the actual requirements.

  Automatic backup is to back up the index data of a cluster. Index backup is implemented by creating cluster snapshots. For backup of the first time, you are advised to back up all index data.

  CSS allows you to store the snapshot data of Elasticsearch instances to OBS, thereby achieving cross-region backup with the cross-region replication function of OBS.
- Restoring data using snapshots

  If data loss occurs or you want to retrieve data of a certain period, click **Restore** in the **Operation** column in the **Snapshots** area to restore the backup index data to the specified cluster by using existing snapshots.

High Security

CSS ensures secure running of data and services from the following aspects:

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

- Network isolation

  The network is divided into two planes, service plane and management plane. The two planes are deployed and isolated physically to ensure the security of the service and management networks.

  - Service plane: refers to the network plane of the cluster. It provides service channels for users and delivers data definition, index, and search capabilities.
  - Management plane: refers to the management console. It is used to manage CSS.
  - VPC security groups or isolated networks ensure the security of hosts.
- Access control
  - Using the network access control list (ACL), you can permit or deny the network traffic entering and exiting the subnets.
  - Internal security infrastructure (including the network firewall, intrusion detection system, and protection system) can monitor all network traffic that enters or exits the VPC through the IPsec VPN.
  - User authentication and index-level authentication are supported. CSS also supports interconnection with third-party user management systems.
- Data security
  - In CSS, the multi-replica mechanism is used to ensure user data security.
  - Communication between the client and server can be encrypted using SSL.
- Operation audit

  Cloud Trace Service (CTS) can be used to perform auditing on key logs and operations.

High Availability

To prevent data loss and minimize the cluster downtime in case of service interruption, CSS supports cross-AZ cluster deployment. When creating a cluster, you can select two or three AZs in the same region. The system will automatically allocate nodes to these AZs. If an AZ is faulty, the remaining AZs can still run properly, significantly enhancing cluster availability and improving service stability.

1.3 Product Components

CSS supports Kibana and Cerebro.

Kibana

Kibana is an open-source data analytics and visualization platform that works with Elasticsearch. You can use Kibana to search for and view data stored in Elasticsearch indexes and display data in charts and maps. For details about Kibana, visit **https://www.elastic.co/guide/en/kibana/current/index.html**.

By default, the Elasticsearch cluster of CSS provides the access channel to Kibana. You can quickly access Kibana without installing it. CSS is compatible with Kibana visualizations and Elasticsearch statistical and analysis capabilities.

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

- Over 10 data presentation modes
- Nearly 20 data statistics methods
- Classification in various dimensions, such as time and tag

Cerebro

Cerebro is an open-source Elasticsearch web admin tool built using Scala, Play Framework, AngularJS, and Bootstrap. Cerebro allows you to manage clusters on a visualized page, such as executing REST requests, modifying Elasticsearch configurations, monitoring real-time disks, cluster loads, and memory usage.

By default, the Elasticsearch cluster of CSS provides the access channel to Cerebro. You can quickly access Cerebro without installing it. CSS is fully compatible with the open-source Cerebro and adapts to the latest 0.8.4 version.

- Elasticsearch visualized and real-time load monitoring
- Elasticsearch visualized data management

1.4 Scenarios

CSS can be used to build search boxes for websites and apps to improve user experience. You can also build a log analysis platform with it, facilitating data-driven O&M and business operations. CSS vector search can help you quickly build smart applications, such as AI-based image search, recommendation, and semantic search.

Site Search

CSS can be used to search for website content by keyword as well as search for and recommend commodities on e-commerce sites.

- Real-time search: When site content is updated, you can find the updated content in your search within minutes, or even just seconds.
- Categorized statistics: You can apply search filters to sort products by category.
- Custom highlight style: You can define how the search results are highlighted.

All-Scenario Log Analysis

Analyze the logs of Elastic Load Balance (ELB), servers, containers, and applications. In CSS, the Kafka message buffer queue is used to balance loads in peak and off-peak hours. Logstash is used for data extract, transform and load (ETL). Elasticsearch retrieves and analyzes data. The analysis results are visualized by Kibana and presented to you.

- High cost-effectiveness: CSS uses the Kunpeng computing power, separates cold and hot storage, and decouples computing and storage resources, achieving high performance and reducing costs by over 30%.
- Ease of use: Perform queries in a GUI editor. Easily create reports using drag-and-drop components.
- Powerful processing capability: CSS can import hundreds of terabytes of data per day, and can process petabytes of data.

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

Database Query Acceleration

CSS can be used to accelerate database queries. E-commerce and logistics companies have to respond to a huge number of concurrent order queries within a short period of time. Relational databases, although having good transaction atomicity, are weak in transaction processing, and can rely on CSS to enhance OLTP and OLAP capabilities.

- High performance: Retrieve data from hundreds of millions of records within milliseconds. Text, time, numeric, and spatial data types are supported.
- High scalability: CSS can be scaled to have over 200 data nodes and over 1000 columns.
- Zero service interruption: The rolling restart and dual-copy mechanisms can avoid service interruption in case of specifications change or configuration update.

Vector Search

When you search for unstructured data, such as images, videos, and corpuses, the nearest neighbors or approximate nearest neighbors are searched based on feature vectors. This has the following advantages:

- Efficient and reliable: The vector search engine provides optimal search performance and distributed DR capabilities.
- Abundant indexes: Multiple indexing algorithms and similarity measurement methods are available and can meet diverse needs.
- Easy learning: CSS is fully compatible with the open-source Elasticsearch ecosystem.

1.5 Permissions Management

If you need to assign different permissions to employees in your organization to access your CSS resources, IAM is a good choice for fine-grained permissions management. IAM provides identity authentication, permissions management, and access control.

If the current account has met your requirements, you do not need to create an independent IAM user for permission management. Then you can skip this section. This will not affect other functions of CSS.

With IAM, you can use your account to create IAM users for your employees and assign permissions to the users to control their access to your resources. IAM is free of charge. You pay only for the resources you purchase. For more information about IAM, see the section "Service Overview" in *Identity and Access Management User Guide*.

Permissions Management

New IAM users do not have any permissions assigned by default. You need to add them to one or more groups and attach policies or roles to these groups. The users then inherit permissions from the groups and can perform specified operations on cloud services based on the permissions they have been assigned.

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

CSS is a project-level service deployed in specific physical regions. Therefore, CSS permissions are assigned to projects in specific regions and only take effect in these regions. If you want the permissions to take effect in all regions, you need to assign the permissions to projects in each region. When accessing CSS, the users need to switch to a region where they have been authorized to use cloud services.

You can use roles and policies to grant users permissions.

- Roles: A type of coarse-grained authorization mechanism that defines permissions related to user responsibilities. This mechanism provides only a limited number of service-level roles for authorization. When using roles to grant permissions, you need to also assign dependency roles. Roles are not ideal for fine-grained authorization and secure access control.
- Policies: A type of fine-grained authorization mechanism that defines the permissions for performing operations on specific cloud resources under certain conditions. This mechanism allows for more flexible authorization. Policies allow you to meet requirements for more secure access control. For example, CSS administrators can only grant CSS users the permissions needed for managing a particular type of CSS resources. Most policies define permissions based on APIs. For the API actions supported by CSS, see the section "Permissions Policies and Supported Actions".

Table 1-1 lists all the system-defined roles and policies supported by CSS.

- **CSS Administrator** depends on the roles of other services to execute its permissions. Therefore, if you assign the **Elasticsearch Administrator** role to a user, assign its dependency roles at the same time.
- **CSS FullAccess** and **CSS ReadOnlyAccess** can be used to control the resources that users can access. For example, if you want your software developers to use CSS resources but not delete them or perform any high-risk operations, you can create IAM users for these software developers and assign them only the permissions required for using CSS resources.

Table 1-1 CSS system permission

Role/Policy Name	Type	Description	Dependency
CSS Administrator	System-defined role	Full permissions for CSS. This role depends on the **Tenant Guest** and **Server Administrator** roles in the same project.	• **Tenant Guest**: A global role, which must be assigned in the global project. • **Server Administrator**: A project-level role, which must be assigned in the same project.
CSS FullAccess	System-defined policy	Full CSS permissions granted through policies. Users with these permissions can perform all operations on CSS.	None

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

Role/Policy Name	Type	Description	Dependency
CSS ReadOnlyAccess	System-defined policy	Read-only permissions for CSS. Users with these permissions can only view CSS data.	None

Table 1-2 lists the common operations supported by each system permission of CSS. Please choose proper system permissions according to this table.

Table 1-2 Common operations supported by each system-defined policy

Operation	CSS FullAccess	CSS ReadOnlyAccess	CSS Administrator	Remarks
Creating a cluster	√	x	√	-
Querying the cluster list	√	√	√	-
Querying the cluster details	√	√	√	-
Deleting a cluster	√	x	√	-
Restarting a cluster	√	x	√	-
Expanding cluster capacity	√	x	√	-
Adding instances and expanding instance storage capacity	√	x	√	-
Querying tags of a specified cluster	√	√	√	-
Querying all tags	√	√	√	-

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

Operation	CSS FullAccess	CSS ReadOnlyAccess	CSS Administrator	Remarks
Automatically setting basic configurations of a cluster snapshot	√	x	√	Depends on OBS and IAM permissions
Modifying basic configurations of a cluster snapshot	√	x	√	Depends on OBS and IAM permissions
Setting the automatic snapshot creation policy	√	x	√	-
Querying the automatic snapshot creation policy	√	√	√	-
Manually creating a snapshot	√	x	√	-
Querying the snapshot list	√	√	√	-
Restoring a snapshot	√	x	√	-
Deleting a snapshot	√	x	√	-
Disabling the snapshot function	√	x	√	-
Modifying specifications	√	x	√	-

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

Operation	CSS FullAccess	CSS ReadOnlyAccess	CSS Administrator	Remarks
Scaling in clusters	√	x	√	-

1.6 Constraints

Restrictions on Clusters and Nodes

The following table describes restrictions on clusters and nodes in CSS.

Table 1-3 Restrictions on Elasticsearch clusters and nodes

Cluster and Node	Restriction
Maximum number of nodes in a cluster	Default: **32**. Maximum: 200. To change the default value, contact technical support.
Minimum number of nodes in a cluster	1

Restrictions on Browsers

- You are advised to use the following browsers to access the CSS management console:
  - Google Chrome 36.0 or later
  - Mozilla Firefox 35.0 or later
- You are advised to use the following browsers to access Kibana and Cerebro integrated in CSS:
  - Google Chrome 36.0 or later
  - Mozilla Firefox 35.0 or later

1.7 Quotas

CSS uses the following resources:

- Instance
- CPU
- Memory (GB)
- Disk quantity
- Disk size (GB)

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

1.8 Related Services

Figure 1-1 shows the relationships between CSS and other services.

Figure 1-1 Relationships between CSS and other services

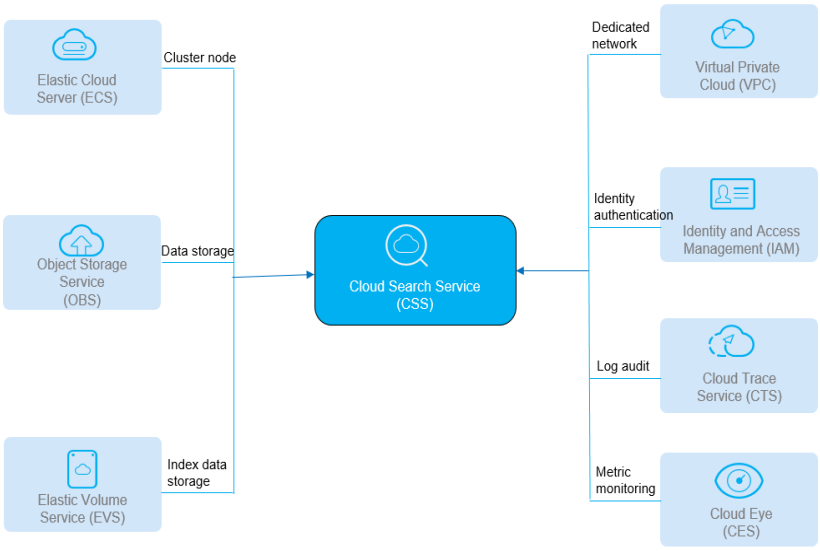


Table 1-4 Relationships between CSS and other services

Service	Description
Virtual Private Cloud (VPC)	CSS clusters are created in the subnets of a VPC. VPCs provide a secure, isolated, and logical network environment for your clusters.
Elastic Cloud Server (ECS)	In a CSS cluster, each node represents an ECS. When you create a cluster, ECSs are automatically created.
Elastic Volume Service (EVS)	CSS uses EVS to store index data. When you create a cluster, EVSs are automatically created for cluster data storage.
Object Storage Service (OBS)	Snapshots of CSS clusters are stored in OBS buckets.
Identity and Access Management (IAM)	IAM authenticates access to CSS.
Cloud Eye	CSS uses Cloud Eye to monitor cluster metrics in real time. The supported CSS metrics include the disk usage and cluster health status. You can learn about the disk usage of the cluster based on the disk usage metric. You can learn about the health status of a cluster based on the cluster health status metric.

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

Service	Description
Cloud Trace Service (CTS)	With CTS, you can record operations associated with CSS for query, audit, and backtracking operations.

1.9 Basic Concepts

Cluster

CSS provides functions on a per cluster basis. A cluster represents an independent search service that consists of multiple nodes.

Index

An index stores Elasticsearch data. It is a logical space in which one or more shards are grouped.

Shard

An index can potentially store a large amount of data that can exceed the hardware limits of a single node. To solve this problem, Elasticsearch provides the ability to subdivide your index into multiple pieces called shards. When you create an index, you can simply define the number of shards that you want. Each shard is in itself a fully-functional and independent "index" that can be hosted on any node in the cluster.

You need to specify the number of shards before creating an index and cannot change the number after the index is successfully created.

Replica

A replica is a copy of the actual storage index in a shard. It can be understood as a backup of the shard. Replicas help prevent single point of failures (SPOFs). You can increase or decrease the number of replicas based on your service requirements.

Document

An entity for Elasticsearch storage. Equivalent to the row in the RDB, the document is the basic unit that can be indexed.

Document Type

Similar to a table in the RDB, type is used to distinguish between different data.

In versions earlier than Elasticsearch 7.*x*, each index can contain multiple document types. Elasticsearch defines a type for each document.

Elasticsearch 7.*x* and later versions only support documents of the .doc type.

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

Mapping

A mapping is used to restrict the type of a field and can be automatically created based on data. It is similar to the schema in the database.

Field

The field is the minimum unit of a document. It is similar to the column in the database.

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

2 Getting Started

2.1 Getting Started with Elasticsearch

This section describes how to use Elasticsearch for product search. You can use the Elasticsearch search engine of CSS to search for data based on the scenario example. The basic operation process is as follows:

- **Step 1: Create a Cluster**
- **Step 2: Import Data**
- **Step 3: Search for Data**
- **Step 4: Delete the Cluster**

Scenario Description

A women's clothing brand builds an e-commerce website. It uses traditional databases to provide a product search function for users. However, due to an increase in the number of users and business growth, the traditional databases have slow response and low accuracy. To improve user experience and user retention, the e-commerce website plans to use Elasticsearch to provide the product search function for users.

This section describes how to use Elasticsearch to provide the search function for users.

Assume that the e-commerce website provides the following data:

```
{
"products":[
{"productName":"Latest art shirts for women in autumn 2017","size":"L"}
{"productName":"Latest art shirts for women in autumn 2017","size":"M"}
{"productName":"Latest art shirts for women in autumn 2017","size":"S"}
{"productName":"Latest jeans for women in spring 2018","size":"M"}
{"productName":"Latest jeans for women in spring 2018","size":"S"}
{"productName":"Latest jeans for women in spring 2017","size":"L"}
{"productName":"Latest casual pants for women in spring 2017","size":"S"}
]
}
```

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

Step 1: Create a Cluster

Create a cluster using Elasticsearch as the search engine. In this example, suppose that you create a cluster named **Sample-ESCluster**. This cluster is used only for getting started with Elasticsearch. For this cluster, you are advised to select **ess.spec-kc1.xlarge.2** for **Node Specifications**, **High I/O** for **Node Storage Type**, and **40 GB** for **Node Storage Capacity**. For details, see **Creating an Elasticsearch Cluster in Non-Security Mode**.

After you create the cluster, switch to the cluster list to view the created cluster. If the **Status** of the cluster is **Available**, the cluster is created successfully.

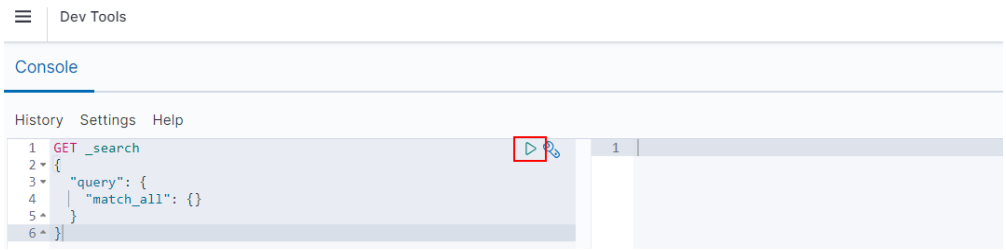
Step 2: Import Data

CSS supports importing data to Elasticsearch using Logstash, Kibana, or APIs. Kibana lets you visualize your Elasticsearch data. The following procedure illustrates how to import data to Elasticsearch using Kibana.

1. On the **Clusters** page, locate the target cluster and click **Access Kibana** in the **Operation** column to go to the Kibana login page.
   - Non-security cluster: The Kibana console is displayed.
   - Security cluster: Enter the username and password on the login page and click **Log In** to go to the Kibana console. The default username is **admin** and the password is the one specified during cluster creation.
2. In the navigation pane of Kibana on the left, choose **Dev Tools**, as shown in **Figure 2-1**.

   The text box on the left is the input box. The triangle icon in the upper right corner of the input box is the command execution button. The text box on the right area is the result output box.

   Figure 2-1 Console page

   NOTE

   The Kibana UI varies depending on the Kibana version.
3. On the **Console** page, run the following command to create index named **my_store**:

   (Versions earlier than 7.*x*)

```
PUT /my_store
{
 "settings": {
  "number_of_shards": 1
 },
 "mappings": {
  "products": {
```

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

```
    "properties": {
      "productName": {
        "type": "text",
        "analyzer": "ik_smart"
      },
      "size": {
        "type": "keyword"
      }
    }
  }
 }
}
```

(Versions later than 7.*x*)

```
PUT /my_store
{
  "settings": {
    "number_of_shards": 1
  },
  "mappings": {
          "properties": {
            "productName": {
              "type": "text",
              "analyzer": "ik_smart"
        },
        "size": {
          "type": "keyword"
        }
      }
    }
  }
```

The command output is similar to the following:

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

4. On the **Console** page, run the following command to import data to index named **my_store**:

   (Versions earlier than 7.*x*)

```
POST /my_store/products/_bulk
{"index":{}}
{"productName":"Latest art shirts for women in autumn 2017","size":"L"}
{"index":{}}
{"productName":"Latest art shirts for women in autumn 2017","size":"M"}
{"index":{}}
{"productName":"Latest art shirts for women in autumn 2017","size":"S"}
{"index":{}}
{"productName":"Latest jeans for women in spring 2018","size":"M"}
{"index":{}}
{"productName":"Latest jeans for women in spring 2018","size":"S"}
{"index":{}}
{"productName":"Latest casual pants for women in spring 2017","size":"L"}
{"index":{}}
{"productName":"Latest casual pants for women in spring 2017","size":"S"}
```

   (Versions later than 7.*x*)

```
POST /my_store/_doc/_bulk
{"index":{}}
{"productName":"Latest art shirts for women in autumn 2017","size":"L"}
{"index":{}}
{"productName":"Latest art shirts for women in autumn 2017","size":"M"}
{"index":{}}
{"productName":"Latest art shirts for women in autumn 2017","size":"S"}
{"index":{}}
{"productName":"Latest jeans for women in spring 2018","size":"M"}
```

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

```
{"index":{}}
{"productName":"Latest jeans for women in spring 2018","size":"S"}
{"index":{}}
{"productName":"Latest casual pants for women in spring 2017","size":"L"}
{"index":{}}
{"productName":"Latest casual pants for women in spring 2017","size":"S"}
```

If the value of the **errors** field in the command output is **false**, the data is imported successfully.

Step 3: Search for Data

- **Full-text search**

  If you access the e-commerce website and want to search for commodities whose names include "spring jeans", enter "spring jeans" to begin your search. The following example shows the command to be executed on Kibana and the command output.

  Command to be executed on Kibana:

  (Versions earlier than 7.*x*)

  ```
  GET /my_store/products/_search
  {
    "query": {"match": {
      "productName": "spring jeans"
    }}
  }
  ```

  (Versions later than 7.*x*)

  ```
  GET /my_store/_search
  {
    "query": {"match": {
      "productName": "spring jeans"
    }}
  }
  ```

  The command output is similar to the following:

  ```
  {
    "took" : 3,
    "timed_out" : false,
    "_shards" : {
      "total" : 1,
      "successful" : 1,
      "skipped" : 0,
      "failed" : 0
    },
    "hits" : {
      "total" : {
        "value" : 4,
        "relation" : "eq"
      },
      "max_score" : 1.7965372,
      "hits" : [
        {
          "_index" : "my_store",
          "_type" : "_doc",
          "_id" : "9xf6VHIBfClt6SDjw7H5",
          "_score" : 1.7965372,
          "_source" : {
            "productName": "Latest jeans for women in spring 2018",
            "size" : "M"
          }
        },
        {
          "_index" : "my_store",
          "_type" : "_doc",
          "_id" : "-Bf6VHIBfClt6SDjw7H5",
  ```

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

```
      "_score" : 1.7965372,
      "_source" : {
        "productName": "Latest jeans for women in spring 2018",
        "size" : "S"
      }
    },
    {
      "_index" : "my_store",
      "_type" : "_doc",
      "_id" : "-Rf6VHIBfClt6SDjw7H5",
      "_score" : 0.5945667,
      "_source" : {
        "productName": "Latest casual pants for women in spring 2017",
        "size" : "L"
      }
    },
    {
      "_index" : "my_store",
      "_type" : "_doc",
      "_id" : "-hf6VHIBfClt6SDjw7H5",
      "_score" : 0.5945667,
      "_source" : {
        "productName": "Latest casual pants for women in spring 2017",
        "size" : "S"
      }
    }
  ]
 }
}
```

  - Elasticsearch supports word segmentation. The preceding command segments "spring jeans" into "spring" and "jeans".
  - Elasticsearch supports full-text search. The preceding command searches for the information about all commodities whose names include "spring" or "jeans".
  - Unlike traditional databases, Elasticsearch can return results in milliseconds by using inverted indexes.
  - Elasticsearch supports sorting by score. In the command output, information about the first two commodities contains both "spring" and "jeans", while that about the last two products contain only "spring". Therefore, the first two commodities rank prior to the last two due to high keyword match.

- **Aggregation result display**

  The e-commerce website provides the function of displaying aggregation results. For example, it classifies commodities corresponding to "spring" based on the size so that you can collect the number of products of different sizes. The following example shows the command to be executed on Kibana and the command output.

  Command to be executed on Kibana:

  (Versions earlier than 7.*x*)

```
GET /my_store/products/_search
{
"query": {
"match": { "productName": "spring" }
},
"size": 0,
"aggs": {
"sizes": {
"terms": { "field": "size" }
}
```

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

```
}
}
```

(Versions later than 7.*x*)

```
GET /my_store/_search
{
"query": {
"match": { "productName": "spring" }
},
"size": 0,
"aggs": {
"sizes": {
"terms": { "field": "size" }
}
}
}
```

The command output is similar to the following:

(Versions earlier than 7.*x*)

```
{
  "took" : 31,
  "timed_out" : false,
  "_shards" : {
    "total" : 1,
    "successful" : 1,
    "skipped" : 0,
    "failed" : 0
  },
  "hits" : {
    "total" : 4,
    "max_score" : 0.0,
    "hits" : [ ]
  },
  "aggregations" : {
    "sizes" : {
      "doc_count_error_upper_bound" : 0,
      "sum_other_doc_count" : 0,
      "buckets" : [
        {
          "key" : "S",
          "doc_count" : 2
        },
        {
          "key" : "L",
          "doc_count" : 1
        },
        {
          "key" : "M",
          "doc_count" : 1
        }
      ]
    }
  }
}
```

(Versions later than 7.*x*)

```
{
  "took" : 3,
  "timed_out" : false,
  "_shards" : {
    "total" : 1,
    "successful" : 1,
    "skipped" : 0,
    "failed" : 0
  },
  "hits" : {
    "total" : {
      "value" : 4,
      "relation" : "eq"
```

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

```
  },
  "max_score" : null,
  "hits" : [ ]
 },
 "aggregations" : {
  "sizes" : {
   "doc_count_error_upper_bound" : 0,
   "sum_other_doc_count" : 0,
   "buckets" : [
    {
     "key" : "S",
     "doc_count" : 2
    },
    {
     "key" : "L",
     "doc_count" : 1
    },
    {
     "key" : "M",
     "doc_count" : 1
    }
   ]
  }
 }
}
```

Step 4: Delete the Cluster

Once you understand the process and method of using Elasticsearch, you can perform the following steps to delete the cluster you created for the example and its data to avoid resource wastage.

NOTE

After you delete a cluster, its data cannot be restored. Exercise caution when deleting a cluster.

1. Log in to the CSS management console. In the navigation pane on the left, choose **Clusters** > **Elasticsearch**.
2. Locate the row that contains cluster **Sample-ESCluster** and click **More** > **Delete** in the **Operation** column.
3. In the displayed dialog box, enter the name of the cluster to be deleted and click **OK**.

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

3 Permissions Management

3.1 Creating an IAM User and Granting Permissions

Contact the O&M administrator to obtain an account.

This section describes how to use a group to grant permissions to an IAM user. **Figure 3-1** shows the process for granting permissions.

CSS has two types of user permissions: CSS administrator permission and read-only permission.

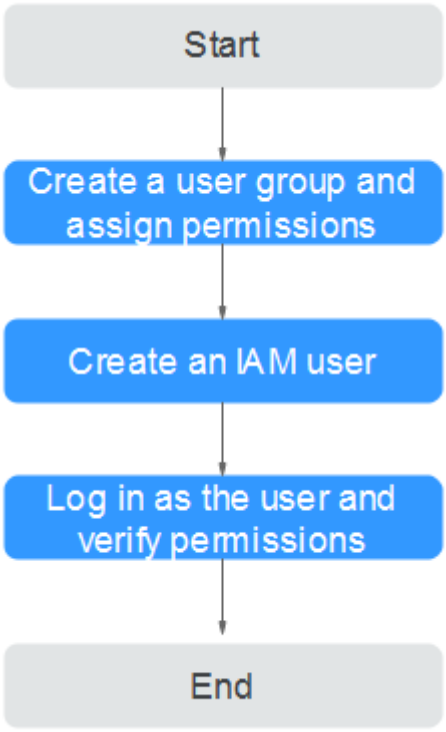
Prerequisites

Before assigning permissions to user groups, you have learned about the system policies listed in Permissions Management.

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

Process Flow

Figure 3-1 Process of granting CSS permissions

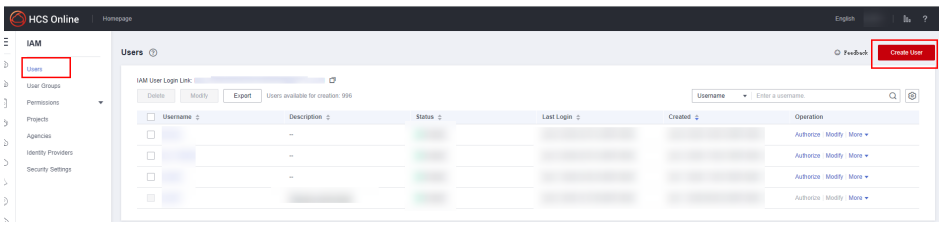


Step 1 Create a user group and grant the permission to it.

1. Log in to the IAM console.
2. Choose **User Groups** from the navigation pane, and click **Create User Group** in the upper right corner.
3. On the displayed page, enter a user group name.
4. Click **OK**. An IAM Identity Center group is created and displayed in the group list.
5. In the user group list, click **Authorize** in the row containing the newly created user group.
6. On the **Authorize User Group** page, select the permissions to be assigned to the user group and click **Next**.
7. Click **OK**.

Step 2 Create a user and add the user to the user group.

1. Choose **Users** from the navigation pane, and click **Create User** in the upper right corner.



 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

2. Specify the user information on the **Create User** page. To create more users, click **Add User**. A maximum of 10 users can be created at a time.
3. Enter the user information and set access mode, credential type, and login protection.
4. Click **Next**. Select the user groups to add the user. The user will inherit the permissions assigned to the user groups.
5. Click **Create**.

Step 3 Log in to the CSS console as the created user, and verify that it has CSS permissions.

----End

3.2 CSS Custom Policies

Custom policies can be created to supplement the system-defined policies of CSS. For the actions supported for custom policies, see section "Permissions Policies and Supported Actions" in the *Cloud Search Service API Reference*.

You can create custom policies in either of the following ways:

- Visual editor: Select cloud services, actions, resources, and request conditions. You do not need to have knowledge of the policy syntax.
- JSON: Create a JSON policy or edit based on an existing policy.

For details about how to create custom policies, see section "Creating a Custom Policy" in the *Identity and Access Management User Guide*. The following section provides examples of common CSS custom policies.

NOTE

IAM permissions and data plane cluster permissions are managed separately. To enable the security capability of the data plane, you need to use the security mode.

Example Custom Policies

NOTE

To let an IAM user access an OBS bucket, you need to grant the **GetBucketStoragePolicy**, **GetBucketLocation**, **ListBucket**, and **ListAllMyBuckets** permissions to the user.

Example 1: Allowing users to create a CSS cluster

```
{
   "Version": "1.1",
   "Statement": [
      {
         "Action": [
            "css:cluster:create",
            "vpc:securityGroups:get",
            "vpc:securityGroups:create",
            "vpc:securityGroups:delete",
            "vpc:securityGroupRules:get",
            "vpc:securityGroupRules:create",
            "vpc:securityGroupRules:delete",
            "vpc:vpcs:list",
            "vpc:privateIps:list",
            "vpc:ports:get",
            "vpc:ports:create",
```

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

```
            "vpc:ports:update",
            "vpc:ports:delete",
            "vpc:quotas:list",
            "vpc:subnets:get",
            "ecs:cloudServerFlavors:get",
            "ecs:serverInterfaces:use",
            "ecs:cloudServers:addNics",
            "ecs:quotas:get",
            "evs:types:get",
            "evs:quotas:get"
        ],
        "Effect": "Allow"
    }
  ]
}
```

Example 2: Denying cluster deletion

A policy with only **Deny** permissions must be used in conjunction with other policies for it to take effect. If the permissions assigned to a user contain both **Allow** and **Deny**, the **Deny** permissions take precedence over the **Allow** permissions.

The following method can be used if you need to assign permissions of the **CSS Admin** policy to a user but you want to prevent the user from deleting clusters. Create a custom policy for denying cluster deletion, and attach both policies to the group to which the user belongs. Then, the user can perform all operations on CSS except deleting clusters. The following is an example of a deny policy:

```
{
    "Version": "1.1",
    "Statement": [
        {
       "Effect": "Deny",
            "Action": [
                "css:cluster:delete"
            ]
        }
    ]
}
```

Example 3: Defining permissions for multiple services in a policy

A custom policy can contain the actions of multiple services that are of the global or project-level type. The following is an example policy containing actions of multiple services:

```
{
  "Version": "1.1",
  "Statement": [
    {
      "Action": [
        "ecs:cloudServers:resize",
        "ecs:cloudServers:delete",
        "ecs:cloudServers:delete",
        "css:cluster:restart",
        "css:*:get*",
        "css:*:list*"
      ],
      "Effect": "Allow"
    }
  ]
}
```

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

4 Viewing the Cluster Runtime Status and Storage Capacity Status

On the **Dashboard** page of the CSS management console, you can view information about the status and storage capacity of existing clusters.

Table 4-1 Cluster status description

Status	Description
Available	The cluster is running properly and is providing services.
Creating	The cluster is being created.
Processing	The cluster is being restarted, scaled, backed up, or recovered.
Abnormal	The cluster creation failed or the cluster is unavailable. If a cluster is in the **Abnormal** status, you can delete the cluster or use snapshots created when the cluster is available to restore data to other clusters. However, operations such as expanding cluster capacity, accessing Kibana, creating snapshots, and restoring snapshots to the cluster are not allowed. When a cluster is in the unavailable status, data importing is not recommended to avoid data loss. You can view the cluster metrics or restart the cluster. However, the operations may fail. If the operations fail, contact technical support in a timely manner.

Table 4-2 Cluster storage capacity status description

Status	Description
Normal	The storage capacity usage of all nodes in a cluster is less than 50%.

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

Status	Description
Warning	The storage capacity usage of any node in a cluster is greater than or equal to 50% and less than 80%.
Danger	The storage capacity usage of any node in a cluster is greater than or equal to 80%. You are advised to increase the storage space of the cluster to achieve normal data search or analysis.
Abnormal	The cluster storage capacity usage is unknown. For example, if the status of a cluster is **Abnormal** due to faults, the storage space status of the cluster will be **Abnormal**.

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

5 Cluster List Overview

The cluster list displays all CSS clusters. If there are a large number of clusters, these clusters will be displayed on multiple pages. You can view clusters of all statuses from the cluster list.

Clusters are listed in chronological order by default in the cluster list, with the most recent cluster displayed at the top. **Table 5-1** shows the cluster parameters.

In the upper right corner of the cluster list, you can enter the name or ID of a cluster and click  to search for a cluster. You can also click  in the upper right corner to refresh the cluster list. Click  to download the cluster list.

Table 5-1 Cluster list parameter description

Parameter	Description
Name/ID	Name and ID of a cluster. You can click a cluster name to switch to the **Basic Information** page. The cluster ID is automatically generated by the system and uniquely identifies a cluster.
Cluster Status	Status of a cluster. For details about the cluster status, see **Viewing the Cluster Runtime Status and Storage Capacity Status**.
Task Status	Status of a task, such as cluster restart, cluster capacity expansion, cluster backup, and cluster restoration.
Version	Elasticsearch version of the cluster.
Created	Time when the cluster is created.
Enterprise Project	Enterprise project that a cluster belongs to.
Private Network Address	Private network address and port number of the cluster. You can use these parameters to access the cluster. If the cluster has multiple nodes, the private network addresses and port numbers of all nodes are displayed.
Billing Mode	Billing mode of a cluster.

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

Parameter	Description
Operation	Operations that can be performed on a cluster, including accessing Kibana, checking metrics, restarting a cluster, and deleting a cluster. If an operation is not allowed, the button is gray.

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

6 Deploying a Cross-AZ Cluster

To prevent data loss and minimize the cluster downtime in case of service interruption, CSS supports cross-AZ cluster deployment. When creating a cluster, you can select two or three AZs in the same region. The system will automatically allocate nodes to these AZs.

Allocating Nodes

If you select two or three AZs when creating a cluster, CSS automatically enables the cross-AZ HA function and properly allocates nodes to different AZs. **Table 6-1** describes how the nodes are allocated.

NOTE

- When creating a cluster, ensure that the number of selected nodes is no less than the number of AZs. Otherwise, cross-AZ deployment is not supported.
- If you enable master nodes when deploying a cross-AZ cluster, the master nodes will also be distributed to different AZs.
- The node quantity difference between any two AZs is no more than one.

Table 6-1 Number of nodes and AZ distribution

Nodes	Single AZ	Two AZs		Three AZs		
	AZ1	AZ1	AZ2	AZ1	AZ2	AZ3
1	1	Not supported		Not supported		
2	2	1	1	Not supported		
3	3	2	1	1	1	1
4	4	2	2	2	1	1
...	...	...	...	...	...	...

Setting Replicas

Setting replicas enables clusters effectively use the HA capability of AZs.

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

- In two-AZ deployment, if one AZ becomes unavailable, the other AZ continues to provide services. In this case, at least one replica is required. Elasticsearch has one replica by default. You can retain the default value if you do not require higher read performance.
- In three-AZ deployment, if one AZ becomes unavailable, the other AZs continue to provide services. In this case, at least one replica is required. Elasticsearch has one replica by default. If you need more replicas to improve the cluster's ability to handle queries, modify **settings** to change the number of replicas.

  You can run the following command to modify the number of index replicas:

  curl -XPUT http://ip:9200/{index_name}/_settings -d '{"number_of_replicas":2}'

  Alternatively, run the following command to specify the number of replicas in the template:

  curl -XPUT http://ip:9200/ _template/templatename -d '{ "template": "*", "settings": {"number_of_replicas": 2}}'

NOTE

- **ip**: private network address
- **index_name**: index name
- **number_of_replicas**: number of replicas after modification. The value in the preceding command indicates that two replicas are required.

Possible Service Interruptions

Table 6-2 describes the possible service interruptions when an AZ of a two- or three-AZ cluster is faulty.

Table 6-2 Possible service interruptions

AZs	Master Nodes	Service Interruption Analysis
2	0	• When the number of nodes is an even number: – If half of data nodes are faulty, replace one node in the faulty AZ before you select the master node. • When the number of nodes is an odd number: – If the faulty AZ contains one more node than the normal AZ, you need to replace one node in the faulty AZ before you select the master node. For details about how to replace nodes, contact technical support. – If the faulty AZ contains one less node than the normal AZ, services will not be interrupted and you can select the master node.

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

AZs	Master Nodes	Service Interruption Analysis
2	3	There is a 50% possibility for service interruption. When two dedicated master nodes are allocated to one AZ and another master node is allocated to the other AZ: • If service interruption happens in the AZ with one master node, you can select a master node from the AZ that has two dedicated master nodes. • If service interruption happens in the AZ with two dedicated master nodes, you have no choice in the remaining AZ, because it has only one dedicated master node. In this case, services will be interrupted and you need to contact technical support.
3	0	If you configure four nodes in three AZs, each AZ will have at least one node. If the AZ with two nodes is faulty, the services will be interrupted. You are not advised to configure four nodes when selecting three AZs. Generally, service interruption will not occur.
3	3	Service interruption does not occur.

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

7 Elasticsearch

7.1 Managing Elasticsearch Clusters

7.1.1 Creating an Elasticsearch Cluster

7.1.1.1 Clusters in Security Mode

When creating an Elasticsearch cluster, you can enable the security mode for it. Identity authentication is required when users access a security cluster. You can also authorize and encrypt security clusters.

Identity Verification

To access a security cluster, you need to enter the username and password. The identity verification is required for the following two types of users:

- Administrator: The default administrator username is **admin**, and the password is the one specified during cluster creation.
- Users: Enter the username and password created through Kibana.

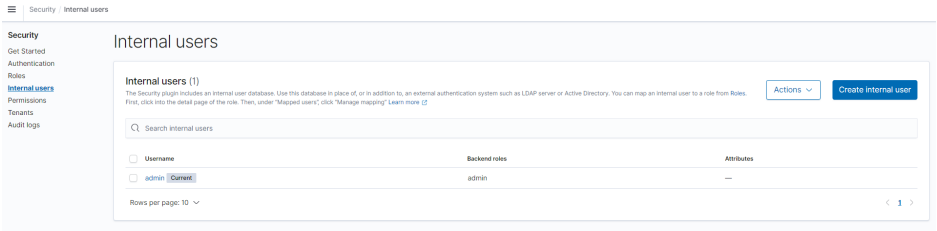
Authorization

On the **Kibana** console, click **Security** to control user permissions in Elasticsearch clusters. You can configure hierarchical user permissions by cluster, index, document, and field. For details, see **Creating a User and Granting Permissions by Using Kibana**.

You can add or delete users, and map users to different roles for permissions control.

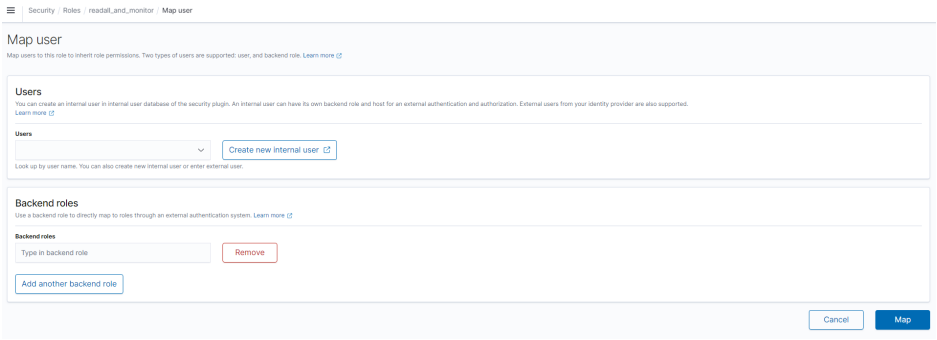
 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

Figure 7-1 Configuring users



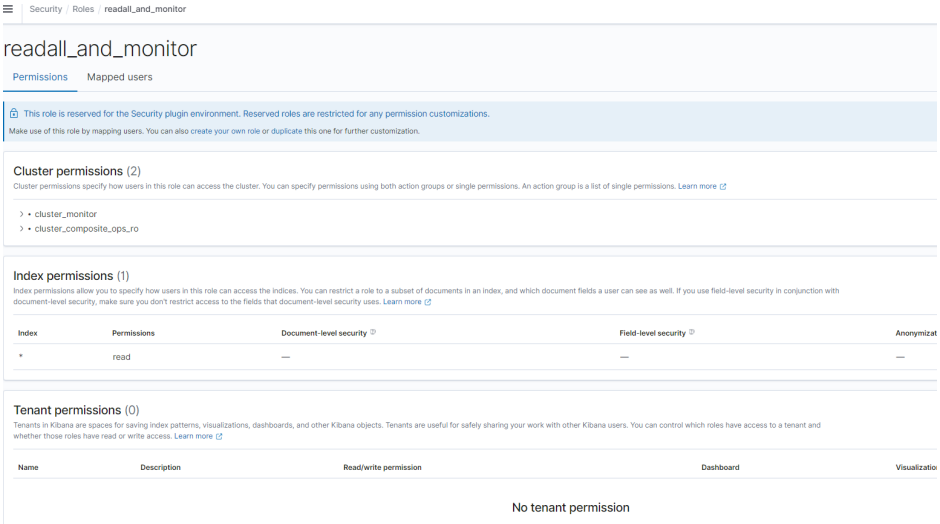
You can use role mapping to configure roles and map a user, backend role, and host name to a role.

Figure 7-2 Role mapping



You can set permissions for each role to access clusters, indexes and documents and assign Kibana tenants different roles.

Figure 7-3 Configuring role permissions

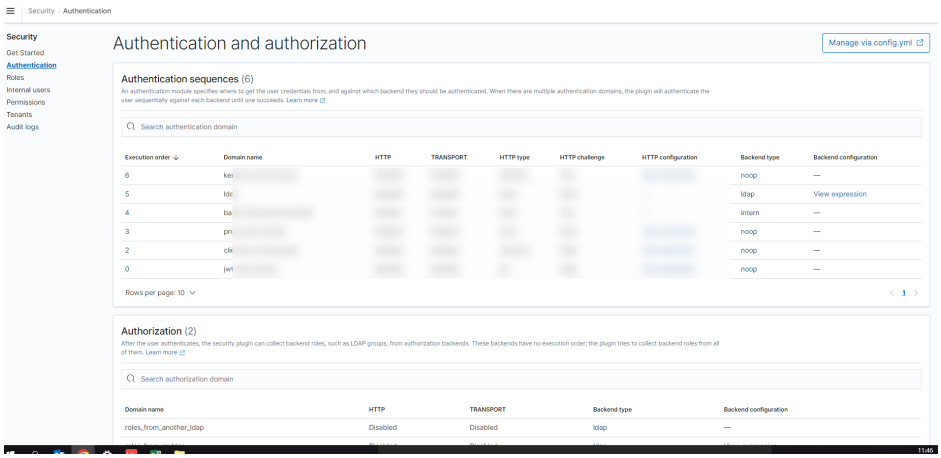


You can set action groups, assign the groups to roles, and configure the roles' permission for accessing indexes and documents.

You can view the parameters of authentication and authorization for the current cluster. You can also run the **securityadmin** command to modify the configuration.

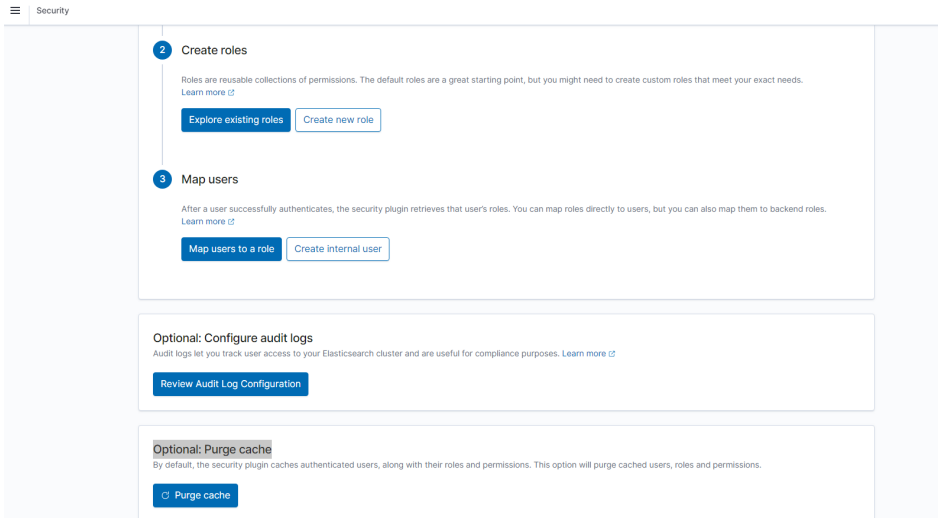
 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

Figure 7-4 Viewing cluster parameters



You can also clear the security cache.

Figure 7-5 Clearing the security cache



Encryption

When key data is transferred between nodes or through the HTTP protocol, SSL/TLS encryption is used to ensure data security.

You can perform the preceding functions on Kibana, using **.yml** files (not recommended), or by calling RESTful APIs. For more information about the security mode, see **Security**.

Resetting the Administrator Password

If you want to change the administrator password of a security cluster or you have forgotten the password, reset the password.

1. On the **Clusters** page, locate the target cluster whose password you want to reset and click the cluster name. The **Cluster Information** page is displayed.

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

2. In the **Configuration** area, click **Reset** next to **Reset Password**.

   NOTE

   - The password can contain 8 to 32 characters.
   - The password must contain at least three of the following character types: uppercase letters, lowercase letters, digits, and special characters. The following special characters are supported: ~!@#$%^&*()-_=+\|[{}];:,<.>/?
   - Do not use the administrator name, or the administrator name spelled backwards.
   - You are advised to change the password periodically.

7.1.1.2 Creating an Elasticsearch Cluster in Security Mode

This section describes how to create an Elasticsearch cluster in security mode.

NOTE

Public IP address access and Kibana public access can be used only after security mode is enabled.

Context

- When creating a cluster, the number of nodes that can be added varies according to the node type. For details, see **Table 7-1**.

Table 7-1 Number of nodes in different types

Node Type	Number
ess	ess: 1-32
ess, ess-master	ess: 1-200 ess-master: an odd number ranging from 3 to 9
ess, ess-client	ess: 1-32 ess-client: 1-32
ess, ess-cold	ess: 1-32 ess-cold: 1-32
ess, ess-master, ess-client	ess: 1-200 ess-master: an odd number ranging from 3 to 9 ess-client: 1-32
ess, ess-master, ess-cold	ess: 1-200 ess-master: an odd number ranging from 3 to 9 ess-cold: 1-32

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

Node Type	Number
ess, ess-client, ess-cold	ess: 1-32 ess-client: 1-32 ess-cold: 1-32
ess, ess-master, ess-client, ess-cold	ess: 1-200 ess-master: an odd number ranging from 3 to 9 ess-client: 1-32 ess-cold: 1-32
Details about the four node types: • **ess**: the default node type that is mandatory for cluster creation. The other three node types are optional. • **ess-master**: master node • **ess-client**: client node • **ess-cold**: cold data node	

Procedure

1. Log in to the CSS management console.
2. On the **Dashboard** page, click **Create Cluster** in the upper right corner. The **Create** page is displayed.

   Alternatively, choose **Clusters** > **Elasticsearch** in the navigation tree on the left. Click **Create Cluster** in the upper right corner. The **Create** page is displayed.
3. Specify **Region** and **AZ**.

   Table 7-2 Parameter description for Region and AZ

Parameter	Description
Region	Select a region for the cluster from the drop-down list on the right.
AZ	Select AZs associated with the cluster region. You can select a maximum of three AZs. For details, see **Deploying a Cross-AZ Cluster**.

4. Configure basic cluster information.

   Table 7-3 Description of basic parameters

Parameter	Description
Version	Select a cluster version from the drop-down list box.

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

Parameter	Description
Name	Cluster name, which contains 4 to 32 characters. Only letters, numbers, hyphens (-), and underscores (_) are allowed and the value must start with a letter. **NOTE** After a cluster is created, you can modify the cluster name as required. Click the name of a cluster to be modified. On the displayed **Basic Information** page, click next to the cluster name. After the modification is completed, click to save the modification. If you want to cancel the modification, click  .

5. Configure cluster specifications.

Table 7-4 Parameter description

Parameter	Description
Nodes	Number of nodes in a cluster. Select a number from 1 to 32. You are advised to configure three or more nodes to ensure high availability of the cluster. • If neither a master node nor client node is enabled, the nodes specified by this parameter are used to serve as both the master node and client node. Nodes provide the cluster management, data storage, cluster access, and data analysis functions. To ensure data stability in the cluster, it is recommended that you set this parameter to a value no less than 3. • If only the master node function is enabled, nodes specified by this parameter are used to store data and provide functions of client nodes. • If both the master and client node functions are enabled, the nodes specified by this parameter are only used for storing data. • If only the client node function is enabled, nodes specified by this parameter are used to store data and provide functions of the master node.
CPU Architecture	Currently, **x86** and **Kunpeng** are supported. The supported type is determined by the actual regional environment.
Node Specifications	Specifications of nodes in a cluster. You can select a specified specification based on your needs. Each cluster supports only one specification.
Node Storage Type	In the current version, the following options are available: **Common I/O**, **High I/O**, and **Ultra-high I/O**.

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

Parameter	Description
Node Storage Capacity	Storage space. Its value varies with node specifications. The node storage capacity must be a multiple of 20.
Master node	The master node manages all nodes in the cluster. If more than 20 nodes are required to store and analyze a large amount of data, you are advised to enable the master node to ensure cluster stability. Otherwise, you are advised to set only the **Nodes** parameter and use the nodes as both master and client nodes. After enabling the master node, specify **Node Specifications**, **Nodes**, and **Node Storage Type**. The value of **Nodes** must be an odd number equal to or greater than 3. Up to nine nodes are supported. The value of **Node Storage Capacity** is fixed. You can select a storage type based on your needs.
Client node	The client node allows clients to access clusters and analyze data. If more than 20 nodes are required to store and analyze a large amount of data, you are advised to enable the client node to ensure cluster stability. Otherwise, you are advised to set only the **Nodes** parameter and use the nodes as both master and client nodes. After enabling the client node, specify **Node Specifications**, **Nodes** and **Node Storage Type**. The value of **Nodes** ranges from 1 to 32. The value of **Node Storage Capacity** is fixed. You can select a storage type based on your needs.
Cold data node	The cold data node is used to store historical data, for which query responses can be returned in minutes. If you do not quire a quick query response, store historical data on cold data nodes to reduce costs. After enabling cold data node, configure **Node Specifications**, **Nodes**, **Node Storage Type**, and **Node Storage Capacity**. The value of **Nodes** ranges from 1 to 32. Select **Node Storage Type** and **Node Storage Capacity** as requirement. After the cold data node is enabled, CSS automatically adds cold and hot tags to related nodes.

6. Set the enterprise project.

   When creating a CSS cluster, you can bind an enterprise project to the cluster if you have enabled the enterprise project function. You can select an enterprise project created by the current user from the drop-down list on the right or click **View Project Management** to go to the **Enterprise Project Management** console and create a new project or view existing projects.

7. Click **Next: Configure Network**. Configure the cluster network.

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

Table 7-5 Network configuration parameters

Parameter	Description
VPC	A VPC is a secure, isolated, and logical network environment. Select the target VPC. Click **View VPC** to enter the VPC management console and view the created VPC names and IDs. If no VPCs are available, create one. **NOTE** The VPC must contain CIDRs. Otherwise, cluster creation will fail. By default, a VPC will contain CIDRs.
Subnet	A subnet provides dedicated network resources that are isolated from other networks, improving network security. Select the target subnet. You can access the VPC management console to view the names and IDs of the existing subnets in the VPC.
Security Group	A security group is a collection of access control rules for ECSs that have the same security protection requirements and are mutually trusted in a VPC. To view more details about the security group, click **View Security Group**. **NOTE** • For cluster access purposes, ensure that the security group contains port 9200. • If your cluster version is 7.6.2 or later, ensure that all the ports used for communication between nodes in the same security group are allowed. If such settings cannot be configured, ensure at least the access to port 9300 is allowed. • After the port 9300 is enabled, if the cluster disk usage is high, delete expired data to release the disk storage space.
Security Mode	After the security mode is enabled, communication will be encrypted and authentication required for the cluster. • The default administrator account is **admin**. • Set and confirm the **Administrator Password**. This password will be required when you access this cluster.

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

Parameter	Description
HTTPS Access	HTTPS access can be enabled only after the security mode of the cluster is enabled. After HTTPS access is enabled, communication is encrypted when you access the cluster. **NOTE** Security clusters use HTTPS for communication, which is much slower than non-security clusters that use HTTP for communication. If you want fast read performance and the permission provided by the security mode to isolate resources (such as indexes, documents, and fields), you can disable the **HTTPS Access** function. After **HTTPS Access** is disabled, HTTP protocol is used for cluster communication. In this case, data security cannot be ensured and public IP address cannot be used.
Public IP Address	If **HTTPS Access** is enabled, you can configure **Public Network Access** and obtain an IP address for public network access. This IP address can be used to access this security cluster through the public network. For details, see **Accessing a Cluster from a Public Network**.

8. Click **Next: Configure Advanced Settings**. Configure the automatic snapshot creation and other functions.
   a. Configure **Cluster Snapshot**. Set basic configuration and snapshot configuration.

      The cluster snapshot function is enabled by default. You can also disable this function as required. To store automatic snapshots in OBS, an agency will be created to access OBS. Additional cost will be incurred if snapshots are stored in standard storage.

      Table 7-6 Cluster snapshot parameter

Parameter	Description
OBS bucket	Select an OBS bucket for storing snapshots from the drop-down list box. You can also click **Create Bucket** on the right to create an OBS bucket. The created or existing OBS bucket must meet the following requirements: • **Storage Class** is **Standard**. • **Region** must be the same as that of the created cluster.

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

Parameter	Description
Backup Path	Storage path of the snapshot in the OBS bucket. The backup path configuration rules are as follows: • The backup path cannot contain the following characters: \:*?"<>| • The backup path cannot start with a slash (/). • The backup path cannot start or end with a period (.). • The backup path cannot contain more than 1,023 characters.
IAM Agency	IAM agency authorized by the current account for CSS to access or maintain data stored in OBS You can also click **Create IAM Agency** on the right to create an IAM agency. The created or existing IAM agency must meet the following requirements: • **Agency Type** must be **Cloud service**. • Set **Cloud Service** to **Elasticsearch** or **CSS**. • The agency must have the **Tenant Administrator** permission for the **OBS** project in **Global service**.

Table 7-7 Automatic snapshot creation parameter

Parameter	Description
Snapshot Name Prefix	The snapshot name prefix contains 1 to 32 characters and must start with a lowercase letter. Only lowercase letters, digits, hyphens (-), and underscores (_) are allowed. A snapshot name consists of a snapshot name prefix and a timestamp, for example, **snapshot-1566921603720**.
Time Zone	Time zone for the backup time, which cannot be changed. Specify **Backup Started Time** based on the time zone.
Backup Start Time	Time when the backup starts automatically every hour, every day, or on a specified day of a week. In the latter two cases, you must specify the beginning of an hour, for example, **00:00** or **01:00**. The value ranges from 00:00 to 23:00. Select a time from the drop-down list.

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

Parameter	Description
Retained Snapshots	Number of automatic snapshots to be retained. The value ranges from 1 to 90. The system automatically deletes excess snapshots every half hour. (The expiration deletion policy applies only to the snapshots that were automatically created at the same frequency as the current automatic snapshot creation policy.) **NOTE** If the snapshot execution interval is short or if the sizes of indexes are large, the number of retained automated snapshots may not reach the preset value. Set this parameter after careful consideration of all factors.

b. Configure advanced settings for the cluster.

- **Default**: The **Kibana Public Access** and **Tag** functions are disabled by default. You can manually enable these functions after the cluster is created.
- **Custom**: You can enable the **Kibana Public Access** and **Tag** functions as required.

Table 7-8 Parameters for advanced settings

Parameter	Description
Kibana Public Access	You can configure this parameter only when security mode is enabled for a cluster. After enabling this function, you can obtain a public IP address for accessing Kibana. For details, see **Accessing a Cluster from a Kibana Public Network**.
Tag	Adding tags to clusters can help you identify and manage your cluster resources. You can customize tags or use tags predefined by Tag Management Service (TMS). For details, see **Managing Tags**.

9. Click **Next: Confirm**. Check the configuration and click **Next** to create a cluster.
10. Click **Back to Cluster List** to switch to the **Clusters** page. The cluster you created is listed on the displayed page and its status is **Creating**. If the cluster is successfully created, its status will change to **Available**.

    If the cluster creation fails, create the cluster again.

7.1.1.3 Creating an Elasticsearch Cluster in Non-Security Mode

This section describes how to create an Elasticsearch cluster in non-security mode.

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

Procedure

1. Log in to the CSS management console.
2. On the **Dashboard** page, click **Create Cluster** in the upper right corner. The **Create** page is displayed.

   Alternatively, choose **Clusters** > **Elasticsearch** in the navigation tree on the left. Click **Create Cluster** in the upper right corner. The **Create** page is displayed.
3. Specify **Region** and **AZ**.

   Table 7-9 Parameter description for Region and AZ

Parameter	Description
Region	Select a region for the cluster from the drop-down list on the right.
AZ	Select AZs associated with the cluster region. You can select a maximum of three AZs. For details, see **Deploying a Cross-AZ Cluster**.

4. Configure basic cluster information.

   Table 7-10 Description of basic parameters

Parameter	Description
Version	Select a cluster version from the drop-down list box.
Name	Cluster name, which contains 4 to 32 characters. Only letters, numbers, hyphens (-), and underscores (_) are allowed and the value must start with a letter. **NOTE** After a cluster is created, you can modify the cluster name as required. Click the name of a cluster to be modified. On the displayed **Basic Information** page, click  next to the cluster name. After the modification is completed, click  to save the modification. If you want to cancel the modification, click  .

5. Configure cluster specifications.

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

Table 7-11 Parameter description

Parameter	Description
Nodes	Number of nodes in a cluster. Select a number from 1 to 32. You are advised to configure three or more nodes to ensure high availability of the cluster. • If neither a master node nor client node is enabled, the nodes specified by this parameter are used to serve as both the master node and client node. Nodes provide the cluster management, data storage, cluster access, and data analysis functions. To ensure data stability in the cluster, it is recommended that you set this parameter to a value no less than 3. • If only the master node function is enabled, nodes specified by this parameter are used to store data and provide functions of client nodes. • If both the master and client node functions are enabled, the nodes specified by this parameter are only used for storing data. • If only the client node function is enabled, nodes specified by this parameter are used to store data and provide functions of the master node.
CPU Architecture	Currently, **x86** and **Kunpeng** are supported. The supported type is determined by the actual regional environment.
Node Specifications	Specifications of nodes in a cluster. You can select a specified specification based on your needs. Each cluster supports only one specification.
Node Storage Type	In the current version, the following options are available: **Common I/O**, **High I/O**, and **Ultra-high I/O**.
Node Storage Capacity	Storage space. Its value varies with node specifications. The node storage capacity must be a multiple of 20.
Master node	The master node manages all nodes in the cluster. If more than 20 nodes are required to store and analyze a large amount of data, you are advised to enable the master node to ensure cluster stability. Otherwise, you are advised to set only the **Nodes** parameter and use the nodes as both master and client nodes. After enabling the master node, specify **Node Specifications**, **Nodes**, and **Node Storage Type**. The value of **Nodes** must be an odd number equal to or greater than 3. Up to nine nodes are supported. The value of **Node Storage Capacity** is fixed. You can select a storage type based on your needs.

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

Parameter	Description
Client node	The client node allows clients to access clusters and analyze data. If more than 20 nodes are required to store and analyze a large amount of data, you are advised to enable the client node to ensure cluster stability. Otherwise, you are advised to set only the **Nodes** parameter and use the nodes as both master and client nodes. After enabling the client node, specify **Node Specifications**, **Nodes** and **Node Storage Type**. The value of **Nodes** ranges from 1 to 32. The value of **Node Storage Capacity** is fixed. You can select a storage type based on your needs.
Cold data node	The cold data node is used to store historical data, for which query responses can be returned in minutes. If you do not quire a quick query response, store historical data on cold data nodes to reduce costs. After enabling cold data node, configure **Node Specifications**, **Nodes**, **Node Storage Type**, and **Node Storage Capacity**. The value of **Nodes** ranges from 1 to 32. Select **Node Storage Type** and **Node Storage Capacity** as requirement. After the cold data node is enabled, CSS automatically adds cold and hot tags to related nodes.

6. Set the enterprise project.

   When creating a CSS cluster, you can bind an enterprise project to the cluster if you have enabled the enterprise project function. You can select an enterprise project created by the current user from the drop-down list on the right or click **View Project Management** to go to the **Enterprise Project Management** console and create a new project or view existing projects.

7. Set network specifications of the cluster.

   Table 7-12 Parameter description

Parameter	Description
VPC	A VPC is a secure, isolated, and logical network environment. Select the target VPC. Click **View VPC** to enter the VPC management console and view the created VPC names and IDs. If no VPCs are available, create one. **NOTE** The VPC must contain CIDRs. Otherwise, cluster creation will fail. By default, a VPC will contain CIDRs.

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

Parameter	Description
Subnet	A subnet provides dedicated network resources that are isolated from other networks, improving network security. Select the target subnet. You can access the VPC management console to view the names and IDs of the existing subnets in the VPC.
Security Group	A security group is a collection of access control rules for ECSs that have the same security protection requirements and are mutually trusted in a VPC. To view more details about the security group, click **View Security Group**. **NOTE** • For cluster access purposes, ensure that the security group contains port 9200. • If your cluster version is 7.6.2 or later, ensure that all the ports used for communication between nodes in the same security group are allowed. If such settings cannot be configured, ensure at least the access to port 9300 is allowed. • After the port 9300 is enabled, if the cluster disk usage is high, delete expired data to release the disk storage space.
Security Mode	Security mode is disabled.

8. Click **Next: Configure Advanced Settings**. Configure the automatic snapshot creation and other functions.
   a. Configure **Cluster Snapshot**. Set basic configuration and snapshot configuration.
      The cluster snapshot function is enabled by default. You can also disable this function as required. To store automatic snapshots in OBS, an agency will be created to access OBS. Additional cost will be incurred if snapshots are stored in standard storage.

      Table 7-13 Cluster snapshot parameter

Parameter	Description
OBS bucket	Select an OBS bucket for storing snapshots from the drop-down list box. You can also click **Create Bucket** on the right to create an OBS bucket. The created or existing OBS bucket must meet the following requirements: • **Storage Class** is **Standard**. • **Region** must be the same as that of the created cluster.

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

Parameter	Description
Backup Path	Storage path of the snapshot in the OBS bucket. The backup path configuration rules are as follows: • The backup path cannot contain the following characters: \:*?"<>| • The backup path cannot start with a slash (/). • The backup path cannot start or end with a period (.). • The backup path cannot contain more than 1,023 characters.
IAM Agency	IAM agency authorized by the current account for CSS to access or maintain data stored in OBS You can also click **Create IAM Agency** on the right to create an IAM agency. The created or existing IAM agency must meet the following requirements: • **Agency Type** must be **Cloud service**. • Set **Cloud Service** to **Elasticsearch** or **CSS**. • The agency must have the **Tenant Administrator** permission for the **OBS** project in **Global service**.

Table 7-14 Automatic snapshot creation parameter

Parameter	Description
Snapshot Name Prefix	The snapshot name prefix contains 1 to 32 characters and must start with a lowercase letter. Only lowercase letters, digits, hyphens (-), and underscores (_) are allowed. A snapshot name consists of a snapshot name prefix and a timestamp, for example, **snapshot-1566921603720**.
Time Zone	Time zone for the backup time, which cannot be changed. Specify **Backup Started Time** based on the time zone.
Backup Start Time	Time when the backup starts automatically every hour, every day, or on a specified day of a week. In the latter two cases, you must specify the beginning of an hour, for example, **00:00** or **01:00**. The value ranges from 00:00 to 23:00. Select a time from the drop-down list.

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

Parameter	Description
Retained Snapshots	Number of automatic snapshots to be retained. The value ranges from 1 to 90. The system automatically deletes excess snapshots every half hour. (The expiration deletion policy applies only to the snapshots that were automatically created at the same frequency as the current automatic snapshot creation policy.) **NOTE** If the snapshot execution interval is short or if the sizes of indexes are large, the number of retained automated snapshots may not reach the preset value. Set this parameter after careful consideration of all factors.

b. Configure advanced settings for the cluster.

- **Default**: The **Kibana Public Access** and **Tag** functions are disabled by default. You can manually enable these functions after the cluster is created.
- **Custom**: You can enable the **VPC Endpoint Service** and **Tag** functions as required.

Table 7-15 Parameters for advanced settings

Parameter	Description
Kibana Public Access	Clusters in non-security mode cannot access Kibana through the Internet.
Tag	Adding tags to clusters can help you identify and manage your cluster resources. You can customize tags or use tags predefined by Tag Management Service (TMS). For details, see **Managing Tags**.

9. Click **Next: Confirm**. Check the configuration and click **Next** to create a cluster.
10. Click **Back to Cluster List** to switch to the **Clusters** page. The cluster you created is listed on the displayed page and its status is **Creating**. If the cluster is successfully created, its status will change to **Available**.

    If the cluster creation fails, create the cluster again.

7.1.2 Viewing Basic Information About an Elasticsearch Cluster

On the **Cluster Information** page, you can view the information about a cluster, including the private network address, public IP address, version, and node.

1. Log in to the CSS management console.
2. Choose **Clusters** > **Elasticsearch**. The cluster list page is displayed.

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

3. Click a cluster name to go to the **Cluster Information** page and view the basic information about the cluster.

Table 7-16 Parameters for configuring basic information

Type	Parameter	Description
Cluster Information	Name	Cluster name. The name can be customized. You can click  on the right to change the cluster name.
	ID	Unique ID of a cluster, which is automatically generated by the system. Each cluster in the same region has a unique ID.
	Version	Cluster version information.
	Cluster Status	Current status of a cluster
	Task Status	Current task status of a cluster. If no task is in progress, -- is displayed.
	Created	Time when a cluster was created
	Cluster Storage Capacity (GB)	Storage capacity of a cluster
	Used Cluster Storage (GB)	Used storage capacity of a cluster
Configuration	Region	Region where a cluster is located
	AZ	AZ where a cluster is located
	VPC	VPC to which the cluster belongs
	Subnet	Subnet to which the cluster belongs
	Security Group	Security group to which a cluster belongs. To change the security group of a cluster, click **Change Security Group** on the right. **NOTICE** Before changing the security group, ensure that the port 9200 required for service access has been enabled. Incorrect security group configuration may cause service access failures. Exercise caution when performing this operation.

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

Type	Parameter	Description
	Security Mode	Security mode of a cluster. • Enabled: The current cluster is a security cluster. • Disabled: The current cluster is a non-security cluster.
	Reset Password	This parameter is displayed only for security clusters. Click **Reset** to change the password of the administrator account **admin** of the security cluster. **NOTE** Requirements for administrator passwords: • The password can contain 8 to 32 characters. • The password must contain at least three of the following character types: uppercase letters, lowercase letters, digits, and special characters. The following special characters are supported: ~!@#$%^&*()-_=+|[{}];:,<.>/? • Do not use the administrator name, or the administrator name spelled backwards. • You are advised to change the password periodically.
	Enterprise Project	Enterprise project to which a cluster belongs. You can click the project name to view the basic information about the enterprise project.

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

Type	Parameter	Description
	Public IP Address	Public network access information, which is displayed only for clusters in security mode. • For a security cluster with public network access enabled, the configured public network address is displayed. You can use this address to access the security cluster from the public network. • For a security cluster with public network access disabled, **--** is displayed. When using a public IP address to access a cluster, you are advised to enable access control and configure an access whitelist to improve cluster security.
	Access Control	Whether to set access control for a cluster. This parameter is displayed only for clusters with public network access enabled. • **Enabled**: Only IP addresses in the whitelist can access the cluster through the public network. • **Disabled**: Any IP address can access the cluster through the public network. Click **Set** to configure the access control and the whitelist.
	Bandwidth	The bandwidth for public network access. This parameter is displayed only for clusters with public network access enabled. Click **Edit** to change the bandwidth size.

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

Type	Parameter	Description
	HTTPS Access	Whether to enable the HTTPS access protocol for a cluster. • **Disabled**: The HTTP protocol is used for cluster access. • **Enabled**: The HTTPS protocol is used for cluster access. Only security clusters can enable this function. If **HTTPS Access** is enabled, you can click **Download Certificate** to obtain the CER security certificate for accessing the security cluster. Currently, the security certificate cannot be used in the public network environment.
	Private Network Address	Private IP address and port number of a cluster, which can be used to access the cluster. If the cluster has only one node, the IP address and port number of only one node are displayed, for example, **10.62.179.32:9200**. If the cluster has multiple nodes, the IP addresses and port numbers of all nodes are displayed, for example, **10.62.179.32:9200,10.62.179.33:9200**.
Node	Node Specifications	Specifications of nodes in a cluster
	Node Storage Type	Storage capacity and storage type of nodes in a cluster
	Nodes	Number of nodes in a cluster

7.1.3 Managing Tags

Tags are cluster identifiers. Adding tags to clusters can help you identify and manage your cluster resources.

You can add tags to a cluster when creating the cluster or add them on the details page of the created cluster.

Managing Tags of a New Cluster

1. Log in to the CSS management console.
2. Click **Create Cluster** in the upper right corner. The **Create Cluster** page is displayed.

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

3. On the **Create Cluster** page, set **Advanced Settings** to **Custom**. Add tags for a cluster.

   You can select a predefined tag and set **Tag value** for the tag. You can click **View Predefined Tag** to switch to the TMS management console and view existing tags.

   You can also create new tags by specifying **Tag key** and **Tag value**.

   You can add a maximum of 20 tags for a CSS cluster. If the entered tag is incorrect, you can click **Delete** on the right of the tag to delete the tag.

   Table 7-17 Naming rules for a tag key and value

Parameter	Description
Tag key	Must be unique in a cluster. The value cannot contain more than 64 characters. It can contain only numbers, letters, and the following special characters: _.:=+-@ The value cannot start or end with a space. Cannot be left blank.
Tag value	The value cannot contain more than 64 characters. It can contain only numbers, letters, and the following special characters: _.:=+-@ The value cannot start or end with a space. Cannot be left blank.

Managing Tags of Existing Clusters

You can modify, delete, or add tags for a cluster.

1. Log in to the CSS management console.
2. On the **Clusters** page, click the name of a cluster for which you want to manage tags.

   The **Basic Information** page is displayed.
3. In the navigation pane on the left, choose the **Tags** tab. You can add, modify, or delete tags.
   - View

     On the **Tags** page, you can view details about tags of the cluster, including the number of tags and the key and value of each tag.
   - Add

     Click **Add** in the upper left corner. In the displayed **Add Tag** dialog box, enter the key and value of the tag to be added, and click **OK**.
   - Modify

     You can only change the value of an existing tag.

     In the **Operation** column of a tag, click **Edit**. In the displayed **Edit Tag** page, enter a new tag value and click **OK**.

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

- Delete

  In the **Operation** column of a tag, click **Delete**. After confirmation, click **Yes** on the displayed **Delete Tag** page.

Searching for Clusters by Tag

1. Log in to the CSS management console.
2. On the **Clusters** page, click the search box above the cluster list and select tag keys and values to search for clusters.

   You can select a tag key or tag value from their drop-down lists. The system returns a list of clusters that exactly match the tag key or tag value. If you enter multiple tags, the cluster that meets requirements of all the tags will be filtered.

   You can add a maximum of 10 tags at one time.

7.1.4 Binding an Enterprise Project

You can create enterprise projects based on your organizational structure. Then you can manage resources across different regions by enterprise project, add users and user groups to enterprise projects, and grant different permissions to the users and user groups. This section describes how to bind a CSS cluster to an enterprise project and how to modify an enterprise project.

Prerequisites

Before binding an enterprise project, you have created an enterprise project.

Binding an Enterprise Project

When creating a cluster, you can bind an existing enterprise project to the cluster, or click **View Enterprise Project** to go to the enterprise project management console and create a new project or view existing projects.

Modifying an Enterprise Project

For a cluster that has been created, you can modify its enterprise project based on the site requirements.

1. Log in to the CSS management console.
2. In the navigation pane on the left, select a cluster type. The cluster management page is displayed.
3. In the cluster list on the displayed page, click the target cluster name to switch to the **Cluster Information** page.
4. On the **Cluster Information** page, click the enterprise project name on the right of **Enterprise Project**. The project management page is displayed.
5. On the **Resources** tab page, select the region of the current cluster, and select **CSS** for **Service**. In this case, the corresponding CSS cluster is displayed in the resource list.
6. Select the cluster whose enterprise project you want to modify and click **Remove**.

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

7. On the **Remove Resource** page, specify **Mode** and select **Destination Enterprise Project**, and click **OK**.
8. After the resource is removed, you can view the modified enterprise project information on the **Clusters** page.

7.1.5 Restarting a Cluster

If a cluster becomes faulty, you can restart it to check if it can run normally.

Prerequisites

- The target cluster is not frozen and has no task in progress.
- If a cluster is available, ensure that it has stopped processing service requests (such as importing data and searching for data). Otherwise, data may be lost when the cluster is restarted. You are advised to perform this operation during off-peak hours.

Context

CSS supports quick restart and rolling restart.

Quick Restart

- All clusters support this function.
- If you select a node type for quick restart, all nodes of the selected type will be restarted together.
- If you select a node name for quick restart, only the specified node will be restarted.
- The cluster is unavailable during quick restart.

Rolling Restart

- Rolling restart is supported only when a cluster has at least three nodes (including master nodes, client nodes, and cold data nodes).
- Rolling restart can be performed only by specifying node types. If you select a node type for rolling restart, the nodes of the selected type will be restarted in sequence.
- During the rolling restart, only the nodes that are being restarted are unavailable and other nodes can run normally.
- When the data volume is large, rolling restart will take a long time.

Quick Restart

1. Log in to the CSS management console.
2. In the navigation tree on the left, select a cluster type. The cluster management list page is displayed.
3. In the **Operation** column of the target cluster, choose **More** > **Restart**.
4. On the **Restart Cluster** page, select **Quick Restart**.

   You can quick restart nodes by **Node type** or **Node name**. If you select **Node type**, then you can select multiple node types and perform quick restart at the time. If you select **Node name**, you can perform quick restart only on one node at a time.

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

5. Refresh the page and check the cluster status. During the restart, the cluster status is **Processing**, and the task status is **Restarting**. If the cluster status changes to **Available**, the cluster has been restarted successfully.

Rolling Restart

1. Log in to the CSS management console.
2. In the navigation tree on the left, select a cluster type. The cluster management list page is displayed.
3. In the **Operation** column of the target cluster, choose **More** > **Restart**.
4. On the **Restart Cluster** page, select **Rolling Restart**.

   You can perform rolling restart by **Node type**. Select specific node types for restart.
5. Refresh the page and check the cluster status. During the restart, the cluster status is **Processing**, and the task status is **Restarting**. If the cluster status changes to **Available**, the cluster has been restarted successfully.

7.1.6 Deleting a Cluster

You can delete clusters that you no longer need.

NOTE

- If you delete a cluster, the cluster service data will be cleared. Exercise caution when performing this operation.
- The snapshots of a cluster stored in OBS are not deleted with the cluster. You can restore a deleted cluster using its snapshots stored in the OBS bucket.

Procedure

1. Log in to the CSS management console.
2. In the navigation tree on the left, select a cluster type. The cluster list page is displayed.
3. Locate the target cluster and click **More** > **Delete** in the **Operation** column.
4. In the displayed dialog box, enter the name of the cluster to be deleted and click **OK**.

7.2 Accessing an Elasticsearch Cluster

7.2.1 Accessing an Elasticsearch Cluster

Elasticsearch clusters have built-in Kibana and Cerebro components. You can quickly access an Elasticsearch cluster through Kibana and Cerebro.

Access a Cluster Through Kibana

1. Log in to the CSS management console.
2. On the **Clusters** page, locate the target cluster and click **Access Kibana** in the **Operation** column to go to the Kibana login page.

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

- Non-security cluster: The Kibana console is displayed.
- Security cluster: Enter the username and password on the login page and click **Log In** to go to the Kibana console. The default username is **admin** and the password is the one specified during cluster creation.

3. After the login is successful, you can access clusters through Kibana.

Accessing a Cluster Through Cerebro

1. Log in to the CSS management console.
2. On the **Clusters** page, locate the target cluster and click **More** > **Cerebro** in the **Operation** column to go to the Cerebro login page.
   - Non-security cluster: Click the cluster name on the Cerebro login page to go to the Cerebro console.
   - Security cluster: Click the cluster name on the Cerebro login page, enter the username and password, and click **Authenticate** to go to the Cerebro console. The default username is **admin** and the password is the one specified during cluster creation.
3. After the login is successful, you can access clusters through Cerebro.

7.2.2 Accessing a Cluster from a Public Network

You can access a security cluster (Elasticsearch clusters in version 6.5.4 or later support the security mode) that has the HTTPS access enabled through the public IP address provided by the system.

NOTE

If public network access is enabled for CSS, then EIP and bandwidth resources will be used and billed.

Configuring Public Network Access

1. Log in to the CSS management console.
2. On the **Create Cluster** page, enable **Security Mode**. Set the administrator password and enable HTTPS access.
3. Select **Automatically assign** for **Public IP Address** and set related parameters.

Table 7-18 Public network access parameters

Parameter	Description
Bandwidth	Bandwidth for accessing Kibana with the public IP address
Access Control	If you disable this function, all IP addresses can access the cluster through the public IP address. If you enable access control, only IP addresses in the whitelist can access the cluster through the public IP address.

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

Parameter	Description
Whitelist	IP address or IP address range allowed to access a cluster. Use commas (,) to separate multiple addresses. This parameter can be configured only when **Access Control** is enabled.

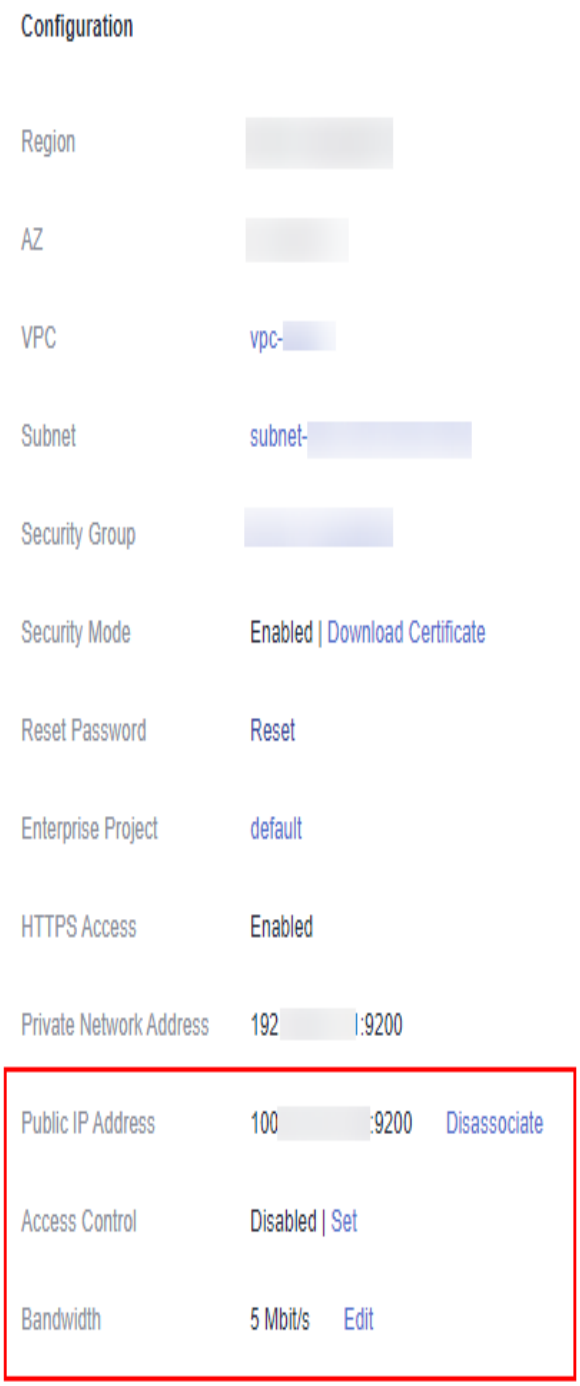
Managing Public Network Access

You can configure, modify, view the public network access of, or disassociate the public IP address from a cluster.

1. Log in to the CSS management console.
2. On the **Clusters** page, click the name of the target cluster. On the **Basic Information** page that is displayed, manage the public network access configurations.

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

Figure 7-6 Modifying public network access configurations



- Configuring public network access

  If you enabled HTTPS but did not configure the public network access during security cluster creation, you can configure it on the **Basic Information** page after configuring the cluster.

  Click **Associate** next to **Public IP Address**, set the access bandwidth, and click **OK**.

  If the association fails, wait for several minutes and try again.

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

- Modifying public network access

  For a cluster for which you have configured public network access, you can click **Edit** next to **Bandwidth** to modify the bandwidth, or you can click **Set** next to **Access Control** to set the access control function and the whitelist for access.

- Viewing public network access

  On the **Basic Information** page, you can view the public IP address associated with the current cluster.

- Disassociating a public IP address from a cluster

  To disassociate the public IP address, click **Disassociate** next to **Public IP Address**.

Accessing a Cluster Through the Public IP Address

After configuring the public IP address, you can use it to access the cluster.

For example, run the following cURL commands to view the index information in the cluster. In this example, the public access IP address of one node in the cluster is **10.62.179.32** and the port number is **9200**.

- If the cluster you access does not have the security mode enabled, run the following command:
  ```
  curl 'http://10.62.179.32:9200/_cat/indices'
  ```
- If the cluster you access has the security mode enabled, access the cluster using HTTPS and add the username, password and **-u** to the cURL command.
  ```
  curl -u username:password -k 'https://10.62.179.32:9200/_cat/indices'
  ```

7.3 Index Backup and Restoration

7.3.1 Backup and Restoration Overview

You can back up index data in clusters. If data loss occurs or you want to retrieve data of a specified duration, you can restore the index data. Index backup is implemented by creating cluster snapshots. When creating a backup for the first time, you are advised to back up data of all indexes.

- **Managing Automatic Snapshot Creation**: Snapshots are automatically created at a specified time each day according to the rules you created. You can enable or disable the automatic snapshot creation function and set the automatic snapshot creation policy.
- **Manually Creating a Snapshot**: You can manually create a snapshot at any time to back up all data or data of specified indexes.
- **Restoring Data**: You can use existing snapshots to restore indexes backed up earlier to a specified cluster.
- **Deleting a Snapshot**: Delete snapshots you do not require and release resources.

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

7.3.2 Managing Automatic Snapshot Creation

Snapshots are automatically created at a specified time according to the rules you create. You can enable or disable the automatic snapshot creation function and set the automatic snapshot creation policy.

Prerequisites

To use the function of creating or restoring snapshots, the account or IAM user logging in to the CSS management console must have both of the following permissions:

- **Tenant Administrator** for project **OBS** in region **Global service**
- **Elasticsearch Administrator** in the current region

Precautions

- When creating a backup for the first time, you are advised to back up data of all indexes.
- Cluster snapshots will increase the CPU usage and disk I/O. You are advised to take cluster snapshots during off-peak hours.
- Before creating a snapshot, you need to perform basic configurations, including configuring the OBS bucket for storing snapshots, the snapshot backup path, and IAM agency used for security authentication.
- If there are available snapshots in the snapshot list when you configure the OBS bucket for storing cluster snapshots for the first time, you cannot change the bucket for snapshots that are subsequently created automatically or manually. Exercise caution when you configure the OBS bucket.
- If snapshots have been stored in the OBS bucket, the OBS bucket cannot be changed. You can disable the snapshot function, enable the snapshot function, and specify a new OBS bucket. After you disable the snapshot function, you cannot use previously created snapshots to restore the cluster.
- If a cluster is in the **Unavailable** status, you can use the cluster snapshot function only to restore clusters and view existing snapshot information.
- During backup and restoration of a cluster, you can perform only certain operations, including scaling out, accessing Kibana, viewing metric, and deleting other snapshots of clusters. However, you cannot perform the following operations: restarting or deleting the cluster, deleting a snapshot that is in the **Creating** or **Restoring** status, and creating or restoring another snapshot. If a snapshot is being created or restored for a cluster, any automatic snapshot creation task initiated for the cluster will be canceled.
- The first snapshot of a cluster is a full snapshot, and subsequent snapshots are incremental snapshots. CSS snapshot files depend on each other.

Managing Automatic Snapshot Creation

1. In the CSS navigation pane on the left, click **Clusters**.
2. On the **Clusters** page that is displayed, click the name of the target cluster. In the navigation pane on the left, choose **Cluster Snapshots**.
3. On the displayed **Cluster Snapshots** page, click the icon to the right of **Cluster Snapshot** to enable the cluster snapshot function.

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

4. Enable the cluster snapshot function. OBS buckets and IAM agencies are automatically created to store snapshots. The automatically created OBS bucket and IAM agency are displayed on the page. You can also click on the right of **Basic Configuration** to edit the configuration.

   Table 7-19 Cluster snapshot parameter

Parameter	Description
OBS bucket	Select an OBS bucket for storing snapshots from the drop-down list box. You can also click **Create Bucket** on the right to create an OBS bucket. The created or existing OBS bucket must meet the following requirements: • **Storage Class** is **Standard**. • **Region** must be the same as that of the created cluster.
Backup Path	Storage path of the snapshot in the OBS bucket. The backup path configuration rules are as follows: • The backup path cannot contain the following characters: \:*?"<>| • The backup path cannot start with a slash (/). • The backup path cannot start or end with a period (.). • The backup path cannot contain more than 1,023 characters.
IAM Agency	IAM agency authorized by the current account for CSS to access or maintain data stored in OBS You can also click **Create IAM Agency** on the right to create an IAM agency. The created or existing IAM agency must meet the following requirements: • **Agency Type** must be **Cloud service**. • Set **Cloud Service** to **Elasticsearch** or **CSS**. • The agency must have the **Tenant Administrator** permission for the **OBS** project in **Global service**.

5. Enable the automatic snapshot creation function. The **Configure Automatic Snapshot Creation** dialog box is displayed. If the automatic snapshot creation function is enabled, you can click on the right of **Automatic Snapshot Creation** to modify the snapshot policy.
   - **Snapshot Name Prefix**: Enter a maximum of 32 characters starting with a lowercase letter. Only lowercase letters, digits, hyphens (-), and underscores (_) are allowed. A snapshot name consists of a snapshot name prefix and a timestamp, for example, **snapshot-2018022405925**.
   - **Time Zone**: indicates the time zone for the backup time. Specify backup start time based on the time zone.
   - **Backup Start Time**: indicates the time when the backup starts automatically every hour, every day, or a specified day of a week. You can

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

specify this parameter at the top of the hour every day or on a specified day of a week, for example, **00:00** or **01:00**. The value ranges from 00:00 to 23:00. Select the backup time from the drop-down list box.

- **Index**: Enter an index name. You can select an index for backup. Use commas (,) to separate multiple indexes. Uppercase letters, spaces, and special characters "\<|>/? are not allowed. If you do not specify this parameter, all indexes in the cluster are backed up by default. You can use the asterisk (*) to back up data of certain indexes. For example, if you enter **index***, then data of indexes with the name prefix of **index** will be backed up.

  Run the **GET /_cat/indices** command in Kibana to query the names of all indexes in the cluster.

6. Click **OK** to save the snapshot policy.

   Snapshots that are automatically created according to the snapshot policy are displayed in the snapshot list, along with manually created snapshots. You can distinguish them by the **Snapshot Type** setting. In the upper right corner of the snapshot list, enter the keyword of the snapshot name or snapshot ID to search for the desired snapshots.

7. (Optional) Disable the automatic snapshot creation function.

   After you disable the automatic snapshot creation function, the system stops automatic creation of snapshots. If the system is creating a snapshot based on the automatic snapshot creation policy and the snapshot is not yet displayed in the snapshot list, you cannot disable the automatic snapshot creation function. In this case, if you click the button next to **Automatic Snapshot Creation**, a message is displayed, indicating that you cannot disable the function. You are advised to disable the function after the system completes automatic creation of the snapshot, and the created snapshot is displayed in the snapshot list.

   When disabling the automatic snapshot creation function, you can choose whether to delete the snapshots that have been automatically created by selecting **Delete automated snapshots** in the displayed dialog box. By default, automatically created snapshots are not deleted.

   - If you do not select **Delete automated snapshots**, automatically created snapshots are not deleted when you disable the automatic snapshot creation function. You can manually delete them later. For details, see **Deleting a Snapshot**. If you do not manually delete the automatically created snapshots and enable the automatic snapshot creation function again, then all snapshots with **Snapshot Type** set to **Automated** in the snapshot list of the cluster can only be automatically deleted by the system. The system automatically deletes snapshots based on the policy configured when the automatic snapshot creation function is enabled. For example, if the number of retained snapshots is set to **10** in this policy and more than 10 snapshots are created, the system automatically deletes the excess snapshots on the half hour.
   - If you select **Delete automated snapshots**, all snapshots with **Snapshot Type** set to **Automated** in the snapshot list will be deleted when you disable the automatic snapshot creation function.

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

NOTE

If snapshots are disabled, existing snapshots will not be automatically deleted. If you need to delete the snapshots, manage the bucket that stores snapshots on the OBS console.

7.3.3 Manually Creating a Snapshot

You can manually create a snapshot at any time to back up all data or data of specified indexes.

Prerequisites

To use the function of creating or restoring snapshots, the account or IAM user logging in to the CSS management console must have both of the following permissions:

- **Tenant Administrator** for project **OBS** in region **Global service**
- **Elasticsearch Administrator** in the current region

Precautions

- When creating a backup for the first time, you are advised to back up data of all indexes.
- Cluster snapshots will increase the CPU usage and disk I/O. You are advised to take cluster snapshots during off-peak hours.
- Before creating a snapshot, you need to perform basic configurations, including configuring the OBS bucket for storing snapshots, the snapshot backup path, and IAM agency used for security authentication.
- If there are available snapshots in the snapshot list when you configure the OBS bucket for storing cluster snapshots for the first time, you cannot change the bucket for snapshots that are subsequently created automatically or manually. Exercise caution when you configure the OBS bucket.
- If snapshots have been stored in the OBS bucket, the OBS bucket cannot be changed. You can disable the snapshot function, enable the snapshot function, and specify a new OBS bucket. After you disable the snapshot function, you cannot use previously created snapshots to restore the cluster.
- If a cluster is in the **Unavailable** status, you can use the cluster snapshot function only to restore clusters and view existing snapshot information.
- During backup and restoration of a cluster, you can perform only certain operations, including scaling out, accessing Kibana, viewing metric, and deleting other snapshots of clusters. However, you cannot perform the following operations: restarting or deleting the cluster, deleting a snapshot that is in the **Creating** or **Restoring** status, and creating or restoring another snapshot. If a snapshot is being created or restored for a cluster, any automatic snapshot creation task initiated for the cluster will be canceled.
- The first snapshot of a cluster is a full snapshot, and subsequent snapshots are incremental snapshots. CSS snapshot files depend on each other.

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

Manually Creating a Snapshot

1. In the CSS navigation pane on the left, click **Clusters**.
2. On the **Clusters** page that is displayed, click the name of the target cluster. In the navigation pane on the left, choose **Cluster Snapshots**.
3. On the displayed **Cluster Snapshots** page, click the icon to the right of **Cluster Snapshot** to enable the cluster snapshot function.
4. Enable the cluster snapshot function. OBS buckets and IAM agencies are automatically created to store snapshots. The automatically created OBS bucket and IAM agency are displayed on the page. You can also click on the right of **Basic Configuration** to edit the configuration.

   Table 7-20 Cluster snapshot parameter

Parameter	Description
OBS bucket	Select an OBS bucket for storing snapshots from the drop-down list box. You can also click **Create Bucket** on the right to create an OBS bucket. The created or existing OBS bucket must meet the following requirements: • **Storage Class** is **Standard**.
Backup Path	Storage path of the snapshot in the OBS bucket. The backup path configuration rules are as follows: • The backup path cannot contain the following characters: \:*?"<>| • The backup path cannot start with a slash (/). • The backup path cannot start or end with a period (.). • The backup path cannot contain more than 1,023 characters.
IAM Agency	IAM agency authorized by the current account for CSS to access or maintain data stored in OBS You can also click **Create IAM Agency** on the right to create an IAM agency. The created or existing IAM agency must meet the following requirements: • **Agency Type** must be **Cloud service**. • Set **Cloud Service** to **Elasticsearch** or **CSS**. • The agency must have the **Tenant Administrator** permission for the **OBS** project in **Global service**.

5. After basic configurations are completed, click **Create**.
   - **Name** indicates the name of the manually created snapshot, which can contain 4 to 64 characters and must start with a lowercase letter. Only lowercase letters, digits, hyphens (-), and underscores (_) are allowed. For snapshots you create manually, you can specify the snapshot name. The system will not automatically add the time information to the snapshot name.

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

- **Index**: Enter an index name. You can select an index for backup. Use commas (,) to separate multiple indexes. Uppercase letters, spaces, and the following special characters are not allowed: "\<|>/? If you do not specify this parameter, data of all indexes in the cluster is backed up by default. You can use the asterisk (*) to back up data of certain indices. For example, if you enter **index***, then data of indices with the name prefix of **index** will be backed up.

  Run the **GET /_cat/indices** command in Kibana to query the names of all indexes in the cluster.
- **Description**: indicates the description of the created snapshot. The value contains 0 to 256 characters, and certain special characters (**<>**) are not allowed.

6. Click **OK**.

   After the snapshot is created, it will be displayed in the snapshot list. The status **Available** indicates that the snapshot is created successfully. along with manually created snapshots. You can distinguish them by the **Snapshot Type** setting. In the upper right corner of the snapshot list, enter the keyword of the snapshot name or snapshot ID to search for the desired snapshots.

7.3.4 Restoring Data

You can use existing snapshots to restore the backup index data to a specified cluster.

Prerequisites

To use the function of creating or restoring snapshots, the account or IAM user logging in to the CSS management console must have both of the following permissions:

- **Tenant Administrator** for project **OBS** in region **Global service**
- **Elasticsearch Administrator** in the current region

Precautions

- Cluster snapshots will increase the CPU usage and disk I/O. You are advised to take cluster snapshots during off-peak hours.
- If snapshots have been stored in the OBS bucket, the OBS bucket cannot be changed. You can disable the snapshot function, enable the snapshot function, and specify a new OBS bucket. After you disable the snapshot function, you cannot use previously created snapshots to restore the cluster.
- If a cluster is in the **Unavailable** status, you can use the cluster snapshot function only to restore clusters and view existing snapshot information.
- During backup and restoration of a cluster, you can perform only certain operations, including scaling out, accessing Kibana, viewing metric, and deleting other snapshots of clusters. However, you cannot perform the following operations: restarting or deleting the cluster, deleting a snapshot that is in the **Creating** or **Restoring** status, and creating or restoring another snapshot. If a snapshot is being created or restored for a cluster, any automatic snapshot creation task initiated for the cluster will be canceled.
- Cluster data cannot be queried during snapshot restoration.

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

- If you restore a CSS cluster snapshot to another cluster, indexes with the same name in the destination cluster will be overwritten. If the snapshot and the destination cluster use different shards, the indexes with the same name will not be overwritten.
- The version of the destination cluster used for restoration must be the same as or higher than that of the source cluster.

Restoring Data

You can use snapshots whose **Snapshot Status** is **Available** to restore cluster data. The stored snapshot data can be restored to other clusters.

Restoring data will overwrite current data in clusters. Therefore, exercise caution when restoring data.

1. In the **Snapshots** area, locate the row that contains the snapshot you want to restore and click **Restore** in the **Operation** column.
2. On the **Restore** page, set restoration parameters.

   Index: Enter the name of the index you want to restore. If you do not specify any index name, data of all indexes will be restored. The value can contain 0 to 1,024 characters. Uppercase letters, spaces, and certain special characters (including "\<|>/?) are not allowed. You can use the asterisk (*) to match multiple indexes. For example, **index*** indicates that all indexes with the prefix **index** in snapshots are restored.

   Rename Pattern: Enter a regular expression. Indexes that match the regular expression are restored. The default value **index_(.+)** indicates restoring data of all indexes. The value contains 0 to 1,024 characters. Uppercase letters, spaces, and certain special characters (including "\<|>/?,) are not allowed.

   Rename Replacement: Enter the index renaming rule. The default value **restored_index_$1** indicates that **restored_** is added in front of the names of all restored indexes. The value contains 0 to 1,024 characters. Uppercase letters, spaces, and certain special characters (including "\<|>/?,) are not allowed.

   NOTE

   The **Rename Pattern** and **Rename Replacement** take effect only when they are configured at the same time.

   Cluster: Select the cluster that you want to restore. You can select the current cluster or others. However, you can only restore the snapshot to clusters whose status is **Available**. If the status of the current cluster is **Unavailable**, you cannot restore the snapshot to the current cluster. When you restore data to another cluster, the version of the target cluster must be later than or equal to that of the current cluster.

   By default, the same-name, same-shard structure indexes in the target cluster will not be overwritten. Data is restored using snapshots by overwriting the snapshot files. After the index with the same name in the target cluster is overwritten, the index data in the target cluster may be lost. Exercise caution when performing this operation.
3. Click **OK**. If restoration succeeds, **Task Status** of the snapshot in the snapshot list will change to **Restoration succeeded**, and the index data is generated again according to the snapshot information.

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

7.3.5 Deleting a Snapshot

If you no longer need a snapshot, delete it to release storage resources. If the automatic snapshot creation function is enabled, snapshots that are automatically created cannot be deleted manually, and the system automatically deletes these snapshots on the half hour after the time specified by **Retention Period (days)**. If you disable the automatic snapshot creation function while retaining the automated snapshots, then you can manually delete them later. If you do not manually delete the automatically created snapshots and enable the automatic snapshot creation function again, then all snapshots with **Snapshot Type** set to **Automated** in the snapshot list of the cluster can only be automatically deleted by the system.

NOTE

After a snapshot is deleted, its data cannot be restored. Exercise caution when deleting a snapshot.

1. In the snapshot list, locate the snapshot that you want to delete.
2. Click **Delete** in the **Operation** column. In the dialog box that is displayed, confirm the snapshot information and click **OK**.

7.4 Changing the Elasticsearch Cluster Form

7.4.1 Overview

You can scale in or out a cluster and change cluster specifications. In this way, you can improve cluster efficiency and reduce O&M costs.

Scaling Out a Cluster

- If a data node (ess) processes many data writing and querying requests and responds slowly, you can expand its storage capacity to improve its efficiency. If some nodes turn unavailable due to the excessive data volume or misoperations, you can add new nodes to ensure the cluster availability.
- **Cold data nodes** (ess-cold) are used to share the workload of data nodes. To prevent cold data loss, you can expand the storage capacity of the cold data node or add new ones.

Changing Specifications

- If the allocation of new indexes or shards takes too long or the node coordination and scheduling are inefficient, you can change the master node (ess-master) specifications.
- If too many tasks need to be distributed or too many results have been aggregated, you can change the client node (ess-client) specifications.
- If the speed of data writing and query decreases suddenly, you can change the data node (ess) specifications.
- If cold data query becomes slow, you can change the cold node (ess-cold) specifications.

Scaling in a Cluster

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

- If a cluster can process existing data without fully using its resources, you can scale in the cluster to reduce costs.

Removing Specified Nodes

- If a cluster can process existing data without fully using its nodes, you can remove one or more specified nodes from the cluster to reduce costs.

Replacing a Specified Node

- If a node in the cluster is faulty, you can create a new node with the same specifications to replace it.

Adding Master/Client Nodes

- If the workloads on the data plane of a cluster increase, you can dynamically scale the cluster by adding master/client nodes.

Changing the Security Mode

After a cluster is created, its security mode can be changed using the following methods:

- Change a non-security cluster to a security cluster that uses HTTP or HTTPS protocol.
- Change a security cluster that uses HTTP or HTTPS protocol to a non-security cluster.
- Change the protocol of a security cluster.

7.4.2 Scaling Out a Cluster

If the workloads on the data plane of a cluster change, you can scale out the cluster by increasing the number or capacity of its nodes. Services are not interrupted during cluster scale-out.

Prerequisites

- The target cluster is available and has no tasks in progress.
- The target cluster has sufficient quotas available.

Constraints

- The **Node Specifications** cannot be modified during scale-out. You can modify **Node Specifications** by referring to **Changing Specifications**.
- If you change the number and storage capacity of a specified type of node, nodes in other types will not be changed.
- The quota of nodes in different types varies. For details, see **Table 7-21**.

Table 7-21 Number of nodes in different types

Node Type	Number
ess	ess: 1-32

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

Node Type	Number
ess, ess-master	ess: 1-200 ess-master: an odd number ranging from 3 to 9
ess, ess-client	ess: 1-32 ess-client: 1-32
ess, ess-cold	ess: 1-32 ess-cold: 1-32
ess, ess-master, ess-client	ess: 1-200 ess-master: an odd number ranging from 3 to 9 ess-client: 1-32
ess, ess-master, ess-cold	ess: 1-200 ess-master: an odd number ranging from 3 to 9 ess-cold: 1-32
ess, ess-client, ess-cold	ess: 1-32 ess-client: 1-32 ess-cold: 1-32
ess, ess-master, ess-client, ess-cold	ess: 1-200 ess-master: an odd number ranging from 3 to 9 ess-client: 1-32 ess-cold: 1-32
Details about the four node types: • **ess**: the default node type that is mandatory for cluster creation. The other three node types are optional. • **ess-master**: master node • **ess-client**: client node • **ess-cold**: cold data node	

Procedure

1. Log in to the CSS management console.
2. In the navigation pane, choose a cluster type. The cluster management page is displayed.
3. Choose **More** > **Modify Configuration** in the **Operation** column of the target cluster. The **Modify Configuration** page is displayed.
4. On the **Modify Configuration** page, choose the **Scale Cluster** tab and click **Scale out** to set parameters.

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

- **Action**: Select **Scale out**.
- **Resource**: The changed amount of resources.
- **Nodes**: The number of nodes and node storage capacity of the default data node.
  - **Nodes**: For details, see **Table 7-21**.
  - The value range of **Node Storage Type** depends on the **Node Specifications**. The value must be a multiple of 20.

5. Click **Next**.
6. Confirm the information and click **Submit**.
7. Click **Back to Cluster List** to switch to the **Clusters** page. The **Task Status** is **Scaling out**. When **Cluster Status** changes to **Available**, the cluster has been successfully scaled out.

7.4.3 Changing Specifications

If the workloads on the data plane of a cluster change, you can change its node specifications as needed.

Prerequisites

- The target cluster is available and has no tasks in progress.
- The target cluster has sufficient quotas available.
- When changing the node specifications, ensure that all service data has replicas so the services will not be interrupted.

  Run the **GET _cat/indices?v** command in Kibana. If the returned **rep** value is greater than **0**, the data has replicas. If the returned **rep** value is **0**, the data has no replicas. In this case, create snapshots for the cluster by referring to **Manually Creating a Snapshot**.
- If the data volume is large, it may take long to modify the node specifications. You are advised to modify specifications during off-peak hours.

Constraints

- The number of nodes and the capacity of node storage cannot be changed. You can add nodes and increase the node storage capacity by referring to **Scaling Out a Cluster**. For details about how to reduce the number of nodes, see **Scaling in a Cluster**.
- After decreasing cluster specifications, the cluster performance will deteriorate and service capabilities will be affected. Exercise caution when performing this operation.
- If a cluster has multiple node types, you can change the specifications of only one type at a time. After the change, nodes in other types still maintain their original specifications.
- Kibana is unavailable during specification change.
- During the specification modification, the nodes are stopped and restarted in sequence. It is a rolling process.

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

Procedure

1. Log in to the CSS management console.
2. In the navigation pane, choose a cluster type. The cluster management page is displayed.
3. Choose **More** > **Modify Configuration** in the **Operation** column of the target cluster. The **Modify Configuration** page is displayed.
4. On the **Modify Configuration** page, choose the **Scale Cluster** tab and click **Change Specifications** to set parameters.
   - **Action**: select **Change specifications**.
   - **Resources**: The changed amount of resources.
   - **Nodes**: Specifications of the default data nodes. Select the required specifications from the **Node Specifications** drop-down list and select the node that you want to change the specifications.
   - If a cluster has master nodes, client nodes, or cold data nodes, you can change their specifications.
5. Click **Next**.
6. Confirm the information and click **Submit**.
7. In the displayed **Verify Index Copy** dialog box, select **Verify index copies** if you need. Click **OK**.
   - Check items for node specifications change: **Verify index copies** and **Cluster status check**.

   Table 7-22 Check item description

Check Item	Description
Verify index copies	By default, CSS checks for indexes that do not have any replicas created for them. You can skip this step, but the lack of index replicas may impact service availability during a cluster specifications change. • If you select **Verify index copies** and the cluster has no master node, each index is required to have at least one replica and the cluster must have at least three nodes. • If you select **Verify index copies** and the cluster has a master node, each index is required to have at least one replica, but there is no requirement on how many nodes the cluster must have.

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

Check Item	Description
Cluster status check	By default, the cluster status is checked before specifications change. The specifications of nodes are changed one by one to ensure service availability and data security. If a cluster is overloaded and services are faulty, the request for a specifications change may fail to be delivered. In this case, you can skip cluster status check, but doing so may cause the cluster to become faulty and services become interrupted. Exercise caution when performing this operation.

8. Click **Back to Cluster List** to switch to the **Clusters** page. The **Cluster Status** is **Configuration modified**. When **Cluster Status** changes to **Available**, the cluster specifications have been successfully modified.

7.4.4 Scaling in a Cluster

If a cluster can process existing data without fully using its resources, you can scale in the cluster to reduce costs. Services are not interrupted during cluster scale-in.

Prerequisites

The target cluster is available and has no tasks in progress.

Constraints

- Only the number of nodes can be modified during cluster scale-in. The node specifications and node storage capacity cannot be modified. You can modify **Node Specifications** by referring to **Changing Specifications**. You can modify node storage capacity by referring to **Scaling Out a Cluster**.
- If you change the number and storage capacity of a specified type of node, nodes in other types will not be changed.
- Ensure that the disk usage after scale-in is less than 80% and each AZ of each node type has at least one node.
- When scaling in a cluster, the data in the node to be deleted is migrated to other nodes. The timeout threshold for data migration is five hours. If data migration is not complete within 5 hours, the cluster scale-in fails. You are advised to perform scale-in for multiple times when the cluster has huge amounts of data.
- For a cluster without master nodes, the number of remaining data nodes (including cold data nodes and other types of nodes) after scale-in must be greater than half of the original node number, and greater than the maximum number of index replicas.
- For a cluster with master nodes, the number of removed master nodes in a scale-in must be fewer than half of the original master node number. After scale-in, there has to be an odd number of master nodes, and there has to be at least three of them.

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

- A cluster with two nodes cannot be scaled in. You can create a cluster using a single node and then .
- The quota of nodes in different types varies. For details, see **Table 7-23**.

Table 7-23 Number of nodes in different types

Node Type	Number
ess	ess: 1-32
ess, ess-master	ess: 1-200 ess-master: an odd number ranging from 3 to 9
ess, ess-client	ess: 1-32 ess-client: 1-32
ess, ess-cold	ess: 1-32 ess-cold: 1-32
ess, ess-master, ess-client	ess: 1-200 ess-master: an odd number ranging from 3 to 9 ess-client: 1-32
ess, ess-master, ess-cold	ess: 1-200 ess-master: an odd number ranging from 3 to 9 ess-cold: 1-32
ess, ess-client, ess-cold	ess: 1-32 ess-client: 1-32 ess-cold: 1-32
ess, ess-master, ess-client, ess-cold	ess: 1-200 ess-master: an odd number ranging from 3 to 9 ess-client: 1-32 ess-cold: 1-32
Details about the four node types: • **ess**: the default node type that is mandatory for cluster creation. The other three node types are optional. • **ess-master**: master node • **ess-client**: client node • **ess-cold**: cold data node	

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

Procedure

1. Log in to the CSS management console.
2. In the navigation pane, choose a cluster type. The cluster management page is displayed.
3. Choose **More** > **Modify Configuration** in the **Operation** column of the target cluster. The **Modify Configuration** page is displayed.
4. On the **Modify Configuration** page, choose the **Scale Cluster** tab and click **Scale in** to set parameters.
   - **Action**: Select **Scale in**.
   - **Resources**: The changed amount of resources.
   - **Agency**: Select an IAM agency to grant the current account the permission to switch AZs.

     If no agency is available, click **Create IAM Agency** to go to the IAM console and create an agency.

     NOTE

     The selected agency must be assigned the **Tenant Administrator** or **VPC Administrator** policy.
   - **Nodes**: The number of the default data nodes. For details about the value range that can be changed, see **Table 7-23**.
5. Click **Next**.
6. Confirm the information and click **Submit**.
7. Click **Back to Cluster List** to switch to the **Clusters** page. The **Task Status** is **Scaling in**. When **Cluster Status** changes to **Available**, the cluster has been successfully scaled in.

7.4.5 Removing Specified Nodes

If a cluster can process existing data without fully using its nodes, you can remove one or more specified nodes from the cluster to reduce costs. Services will not be interrupted during the removal of specified nodes.

Prerequisites

The target cluster is available and has no tasks in progress.

Constraints

- Ensure that the disk usage after scale-in is less than 80% and each AZ of each node type has at least one node.
- In a cross-AZ cluster, the difference between the numbers of the same type nodes in different AZs cannot exceed 1.
- For a cluster without master nodes, the number of removed data nodes and cold data nodes in a scale-in must be fewer than half of the original number of data nodes and cold data nodes, and the number of remaining data nodes and cold data nodes after a scale-in must be greater than the maximum number of index replicas.
- For a cluster with master nodes, the number of removed master nodes in a scale-in must be fewer than half of the original master node number. After

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

scale-in, there has to be an odd number of master nodes, and there has to be at least three of them.

Procedure

1. Log in to the CSS management console.
2. In the navigation pane, choose a cluster type. The cluster management page is displayed.
3. Choose **More** > **Modify Configuration** in the **Operation** column of the target cluster. The **Modify Configuration** page is displayed.
4. On the **Modify Configuration** page, click the **Scale In** tab.
5. On the **Scale In** tab page, set the following parameters:
   - **Whether to perform data migration**: If this option is selected, data migration is performed. If the target node contains disabled indexes or indexes that have no replicas, this option must be selected.
   - In the data node table, select the node to be scaled in.
6. Click **Next**.
7. Confirm the information and click **Submit**.
8. Click **Back to Cluster List** to switch to the **Clusters** page. The **Task Status** is **Scaling in**. When **Cluster Status** changes to **Available**, the cluster has been successfully scaled in.

7.4.6 Replacing a Specified Node

If a node in the cluster is faulty, you can create a new node with the same specifications to replace it.

Prerequisites

The target cluster is available and has no tasks in progress.

Constraints

- Only one node can be replaced at a time.
- The ID, IP address, specifications, and AZ of the new node will be the same as those of the original one.
- The configurations you modified manually will not be retained after node replacement. For example, if you have manually added a return route to the original node, you need to add it to the new node again after the node replacement is complete.
- If the node you want to replace is a data node (ess) or cold data node (ess-cold), pay attention to the following precautions:
  a. Before a data node or cold data node is replaced, its data needs to be migrated to other nodes. To properly store the data, ensure the maximum sum of replicas and primary shards of an index is smaller than the total number of data nodes (ess and ess-cold nodes) in the cluster. The node replacement duration depends heavily on the migration speed.
  b. Clusters whose version is earlier than 7.6.2 cannot have closed indexes. Otherwise, data nodes or cold data nodes cannot be replaced.

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

c. The AZ of the node to be replaced must have two or more data nodes (including ess and ess-cold).
d. If the cluster of the node to be replaced does not have a master node (ess-master), the number of available data nodes (including ess and ess-cold) in the cluster must be greater than or equal to 3.
e. The preceding precautions do not apply if you are replacing a master node (ess-master) or client node (ess-client).
f. The precautions 1 to 4 do not apply if you are replacing a faulty node, regardless of its type. Faulty nodes are not included in **_cat/nodes**.

Procedure

1. Log in to the CSS management console.
2. In the navigation pane, choose a cluster type. The cluster management page is displayed.
3. Choose **More** > **Modify Configuration** in the **Operation** column of the target cluster. The **Modify Configuration** page is displayed.
4. On the **Modify Configuration** page, click the **Replace Node** tab.
5. On the **Replace Node** tab page, set the following parameters:
   - **Whether to perform data migration**: If this option is selected, data migration is performed. If the target node has disabled indexes or indexes that have no replicas, this option must be selected.
   - Select the node to be replaced in the data node table.
6. Click **Submit**.
7. Click **Back to Cluster List** to switch to the **Clusters** page. The **Task Status** is **Upgrading**. When **Cluster Status** changes to **Available**, the node has been successfully replaced.

7.4.7 Adding Master/Client Nodes

If workloads on the data plane of a cluster increase, you can add master or client nodes as needed. Services are not interrupted while they are added.

Prerequisites

The target cluster is available and has no tasks in progress.

Constraints

- If a cluster already has master and client nodes, the **Add Master/Client Node** tab is not displayed on the **Modify Configuration** page. In this case, you need to add the master or client nodes by referring to **Scaling Out a Cluster**.
- When you add master or client nodes, the number of nodes that can be configured varies depending on the node type. For details, see **Table 7-24**.

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

Table 7-24 Number of nodes in different types

Node Flavor	Number
Master node	An odd number ranging from 3 to 9
Client node	1 to 32

Procedure

1. Log in to the CSS management console.
2. In the navigation pane, choose a cluster type. The cluster management page is displayed.
3. Choose **More** > **Modify Configuration** in the **Operation** column of the target cluster. The **Modify Configuration** page is displayed.
4. On the **Modify Configuration** page, choose the **Add Master/Client Node** tab.
5. Select the target node type and set the node specifications, quantity, and storage.
   - Master and client nodes cannot be added at the same time.
   - If a cluster already has a master or client node, you can only add nodes of the other type.

   Figure 7-7 Adding a master or client node

6. Click **Next**.
7. Confirm the information and click **Submit**.

   Return to the cluster list page. The **Task Status** of the cluster is **Scaling out**.
   - If you added a master node and **Cluster Status** changed to **Available**, the master node has been successfully added.

     NOTICE

     If the cluster version is earlier than 7.*x*, when the **Cluster Status** changes to **Available**, you need to restart all data nodes and cold data nodes in the cluster to make the new node take effect. Before the restart, the cluster may be unavailable. For details, see **Restarting a Cluster**.
   - If you added a client node and **Cluster Status** changed to **Available**, the client node has been added. You can restart data nodes and cold data nodes to shut down Cerebro and Kibana processes on the nodes.

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

7.4.8 Changing the Security Mode

After a cluster is created, its security mode can be changed using the following methods:

- **Switching from the Non-Security Mode to Security Mode**
- **Switching from the Security to Non-Security Mode**
- **Switching the Protocol of Security Clusters**

Context

You can create clusters in multiple security modes. For details about the differences between security modes, see **Table 7-25**.

Table 7-25 Cluster security modes

Security Mode	Scenario	Advantage	Disadvantage
Non-Security Mode	Intranet services and test scenarios	Simple. Easy to access.	Poor security. Anyone can access such clusters.
Security Mode + HTTP Protocol	User permissions can be isolated, which is applicable to scenarios sensitive to cluster performance.	Security authentication is required for accessing such clusters, which improves cluster security. Accessing a cluster through HTTP protocol can retain the high performance of the cluster.	Cannot be accessed from the public network.
Security Mode + HTTPS Protocol	Scenarios that require high security and public network access.	Security authentication is required for accessing such clusters, which improves cluster security. HTTPS protocol allows public network to access such clusters.	The performance of clusters using HTTPS is 20% lower than that of using HTTP.

Prerequisites

- You are advised to back up data before changing the cluster security mode.

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

- The target cluster is available and has no tasks in progress.

Constraints

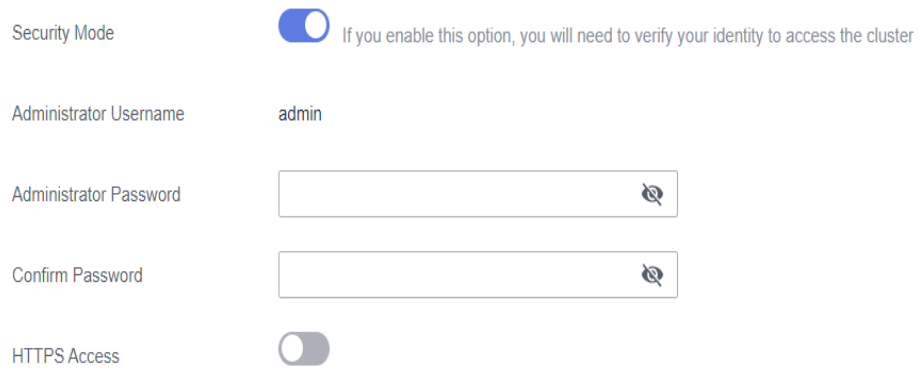
- Only clusters (whose version is 6.5.4 or later) created after November 2022 support security mode switching.
- A cluster automatically restarts when its security mode is being changed. Services are interrupted during the restart. The authentication mode for invoking the cluster will change after the restart, and client configurations need to be adjusted accordingly.
- If a cluster has already opened the Kibana session box, a session error message will be displayed after you change the cluster security mode. In this case, clear the cache and open Kibana again.

Switching from the Non-Security Mode to Security Mode

You can change a non-security cluster to a security cluster that uses HTTP or HTTPS. After a cluster's security mode is enabled, security authentication is required for accessing the cluster.

1. Log in to the CSS management console.
2. In the navigation pane, choose a cluster type. The cluster management page is displayed.
3. Choose **More** > **Modify Configuration** in the **Operation** column of the target cluster. The **Modify Configuration** page is displayed.
4. Choose the **Configure Security Mode** tab.
5. Enable the security mode. Enter and confirm the administrator password of the cluster.

   Figure 7-8 Enabling the security mode

6. Enable or disable **HTTPS Access**.
   - If you enable **HTTPS Access**: The HTTPS protocol is used to encrypt cluster communication and you can configure public networks to access the cluster.
   - If you disable **HTTPS Access**: The HTTP protocol is used and you cannot configure public networks to access the cluster.
7. Click **Submit**. Confirm the information and the cluster list page is displayed.

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

The **Task Status** of the cluster is **The security mode is changing**. When the cluster status changes to **Available**, the security mode has been successfully changed.

Switching from the Security to Non-Security Mode

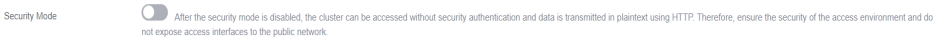
You can change a security cluster that uses HTTP or HTTPS to a non-security cluster. After a cluster's security mode is disabled, security authentication is no longer required for accessing the cluster.

NOTICE

- Clusters in non-security mode can be accessed without security authentication, and HTTP protocol is used to transmit data. Ensure the security of the cluster access environment and do not expose the access interface to the public network.
- During the switchover from the security mode to the non-security mode, the indexes of the original security cluster will be deleted. Back up data before disabling the security mode.
- If a security cluster has been bound to a public IP address, unbind it before changing the security mode.
- If a security cluster has enabled Kibana public network access, disable it before changing the security mode.

1. Log in to the CSS management console.
2. In the navigation pane on the left, choose **Clusters**. On the displayed **Clusters** page, locate the target cluster and choose **More** > **Modify Configuration** in the **Operation** column.
3. Choose the **Configure Security Mode** tab.
4. Disable the security mode.

   Figure 7-9 Disabling the security mode

5. Click **Submit**. Confirm the information and the cluster list page is displayed.

   The **Task Status** of the cluster is **The security mode is changing**. When the cluster status changes to **Available**, the security mode has been successfully changed.

Switching the Protocol of Security Clusters

You can change the protocol of a security cluster.

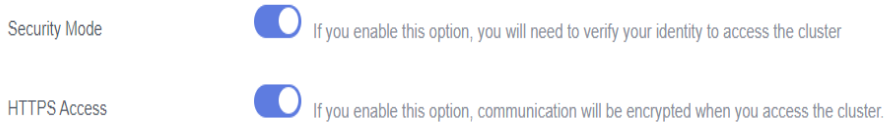
NOTICE

If a security cluster has been bound to a public IP address, you need to unbind it before changing HTTPS protocol to HTTP.

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

1. Log in to the CSS management console.
2. In the navigation pane on the left, choose **Clusters**. On the displayed **Clusters** page, locate the target cluster and choose **More** > **Modify Configuration** in the **Operation** column.
3. Choose the **Configure Security Mode** tab.
4. Enable or disable **HTTPS Access**.

   Figure 7-10 Configuring the protocol

   - If you enable **HTTPS Access**:

     HTTPS protocol is used to encrypt cluster communication and you can configure public network access.
   - If you disable **HTTPS Access**: An alarm message is displayed. Click **OK** to disable the function.

     Cluster communication is no longer encrypted and the public network access function cannot be enabled.
5. Click **Submit**. Confirm the information and the cluster list page is displayed.

   The **Task Status** of the cluster is **The security mode is changing**. When the cluster status changes to **Available**, the security mode has been successfully changed.

7.5 Configuring an Elasticsearch Cluster

7.5.1 Customizing Word Dictionaries

7.5.1.1 Managing Word Dictionaries

You can configure custom word dictionaries to support word segmentation. This gives the search engine improved performance when searching by keywords such as company names and buzzwords from social media.

NOTE

- You cannot use the custom word dictionary function for clusters created before the function was launched .
- Custom word dictionaries are used for Chinese word segmentation. They can also be used to segment English words based on special characters other than #&+-.@_

Context

Custom word dictionaries use the IK and synonym analyzers.

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

The IK analyzer has a main word dictionary and a stop word dictionary. The synonym analyzer has a synonym word dictionary. Before configuring a custom word dictionary, upload the prepared word dictionary file to OBS. For details, see **Uploading a Word Dictionary File to OBS**.

The IK analyzer uses the **ik_max_word** and **ik_smart** word segmentation policies. The synonym analyzer uses the ik_synonym word segmentation policy.

- ik_max_word: splits the text at a fine granularity.
- ik_smart: splits the text at a coarse granularity.

Prerequisites

To use the custom word dictionary, the account or IAM user used for logging in to the CSS management console must have both of the following permissions:

- **Tenant Administrator** for project **OBS** in region **Global service**
- **Elasticsearch Administrator** in the current region

Uploading a Word Dictionary File to OBS

Before configuring a custom word dictionary, upload the word dictionary file to an OBS bucket.

1. Prepare the word dictionary file according to **Table 7-26**.

Table 7-26 Word dictionary description

Word Dictionary Type	Description	Requirement
Main word dictionary	Main words are words that users want to use as basis for performing word segmentation. A main word dictionary is a custom collection of such words.	The word dictionary file must be a text file encoded using UTF-8 without BOM, with one word per row. The maximum size of a main word dictionary file is 100 MB. Letters must be in lowercase.
Stop word dictionary	Stop words are words that users wish to be ignored during word segmentation. The stop word dictionary is a custom collection of stop words.	The word dictionary file must be a text file encoded using UTF-8 without BOM, with one word per row. The maximum size of a stop word dictionary file is 100 MB.

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

Word Dictionary Type	Description	Requirement
Synonym dictionary	Synonyms are words with the same or similar meanings. A synonym dictionary is a custom collection of synonyms.	The word dictionary file must be a text file encoded using UTF-8 without BOM. Each row contains a group of synonyms separated by commas (,). The maximum size of a synonym dictionary file is 100 MB.
Static main word dictionary	The static main word dictionary is a collection of common main words preconfigured in CSS. To view the static main word dictionary, visit **https://github.com/infinilabs/analysis-ik/blob/master/config/main.dic**.	The word dictionary file must be a text file encoded using UTF-8 without BOM, with one word per row. The maximum size of a static main word dictionary file is 100 MB.
Static stop word dictionary	The static stop word dictionary is a collection of common stop words preconfigured in CSS. To view the static stop word dictionary, visit **https://github.com/infinilabs/analysis-ik/blob/master/config/stopword.dic**.	The word dictionary file must be a text file encoded using UTF-8 without BOM, with one word per row. The maximum size of a static stop word dictionary file is 100 MB.
Extra main word dictionary	The Extra main word dictionary is a collection of uncommon main words preconfigured in CSS. To access the Extra main word dictionary, visit **https://github.com/infinilabs/analysis-ik/blob/master/config/extra_main.dic**.	The word dictionary file must be a text file encoded using UTF-8 without BOM, with one word per row. The maximum size of an Extra main word dictionary file is 100 MB.
Extra stop word dictionary	The Extra stop word dictionary is a collection of uncommon stop words preconfigured in CSS. To access the Extra stop word dictionary, visit **https://github.com/infinilabs/analysis-ik/blob/master/config/extra_stopword.dic**.	The word dictionary file must be a text file encoded using UTF-8 without BOM, with one word per row. The maximum size of an Extra stop word dictionary file is 100 MB.

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

2. Upload the word dictionary file to an OBS bucket. For details, see *Object Storage Service User Guide*. The OBS bucket to which data is uploaded must be in the same region as the cluster.

Managing Word Dictionaries

1. Log in to the CSS management console.
2. In the navigation pane, choose **Clusters** > **Elasticsearch**.
3. On the **Clusters** page, click the name of the target cluster.
4. Click the **Word Dictionaries** tab.
5. On the **Word Dictionaries** page, configure custom word dictionaries for the cluster or modify preset ones.
   a. To configure custom word dictionaries, configure the required settings by referring to **Table 7-27**.

Table 7-27 Configuring custom word dictionaries

Parameter	Description
OBS Bucket	Select the OBS location for storing the word dictionary file. You can click **Create Bucket** to create an OBS bucket. The new OBS bucket must be in the same region as the cluster, and **Default Storage Class** must be **Standard** or **Infrequent Access**.
Main Word Dictionary	The main word dictionary is a custom word dictionary. Its initial status is empty. By default, **No Update** is selected, meaning not to configure this word dictionary. • To add a custom main word dictionary, click **Update** and select a .txt word dictionary file. • If you do not want to use a certain word dictionary, click **Do Not Use**.
Stop Word Dictionary	Stop Word Dictionary is a custom word dictionary. Its initial status is empty. By default, **No Update** is selected, meaning not to configure this word dictionary. • To add a custom stop word dictionary, click **Update** and select a .txt word dictionary file. • If you do not want to use a certain word dictionary, click **Do Not Use**.

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

Parameter	Description
Synonym Dictionary	Synonym Dictionary is a custom word dictionary. Its initial status is empty. By default, **No Update** is selected, meaning not to configure this word dictionary. • To add a custom synonym dictionary, click **Update** and select a .txt word dictionary file. • If you do not want to use a certain word dictionary, click **Do Not Use**.

b. To modify a preset word dictionary, toggle on **Modify Preset Word Dictionary**, and modify the word dictionary by referring to **Table 7-28**.

NOTE

If the four preset word dictionaries (static & extra main word and stop word) are not displayed, the current cluster version does not support the deletion or modification of them. To use this function, you are advised to upgrade the cluster version, or create a new cluster and migrate data to it.

Table 7-28 Configuring preset word dictionaries

Parameter	Description
Static Main Word Dictionary	Static Main Word Dictionary is a preset collection of common main words. By default, **No Update** is selected, meaning this word dictionary will be used. • To modify Static Main Word Dictionary, click **Update** and select a .txt word dictionary file. • If you do not want to use Static Main Word Dictionary, click **Do Not Use**.
Static Stop Word Dictionary	Static Stop Word Dictionary is a preset collection of common stop words. By default, **No Update** is selected, meaning this word dictionary will be used. • To modify Static Stop Word Dictionary, click **Update** and select a .txt word dictionary file. • If you do not want to use Static Stop Word Dictionary, click **Do Not Use**.
Extra Main Word Dictionary	Extra Main Word Dictionary is a preset collection of uncommon main words. By default, **No Update** is selected, meaning this word dictionary will be used. • To modify Extra Main Word Dictionary, click **Update** and select a .txt word dictionary file. • If you do not want to use Extra Main Word Dictionary, click **Do Not Use**.

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

Parameter	Description
Extra Stop Word Dictionary	Extra Stop Word Dictionary is a preset collection of uncommon stop words. By default, **No Update** is selected, meaning this word dictionary will be used. • To modify Extra Stop Word Dictionary, click **Update** and select a .txt word dictionary file. • If you do not want to use Extra Stop Word Dictionary, click **Do Not Use**.

6. Click **Save**. In the dialog box that is displayed, click **OK**. The word dictionary information is displayed in the lower part of the page. The word dictionary status is **Updating**. In approximately 1 minute, the word dictionaries are configured, and **Word Dictionary Status** changes to **Successful**.
7. The deletion or update of the four preset dictionaries (static main word, static stop word, extra main word, and extra stop word) requires a cluster restart to take effect. The update of other word dictionaries happens dynamically, and there is no need for cluster restart. For details about how to restart a cluster, see **Restarting a Cluster**.

7.5.1.2 Example

Application Scenarios

Configure a custom word dictionary for the cluster, set main words, stop words, and synonyms. Search for the target text by keyword and synonym and view the search results.

Step 1: Configure a Custom Word Dictionary

1. Prepare a word dictionary file (a text file encoded using UTF-8 without BOM) and upload it to the target OBS path.

   Set the main word dictionary file, stop word dictionary file, and synonym word dictionary file.

   NOTE

   The built-in static stop word dictionary contains common stop words such as **are** and **the**. If the built-in stop word dictionary was never deleted or updated, you do not need to upload such stop words.
2. In the navigation pane on the left, choose **Clusters**.
3. On the **Clusters** page, click the name of the target cluster.
4. Click the **Word Dictionaries** tab. Configure the word dictionary file for the step 1 by referring to **Managing Word Dictionaries**.
5. After the word dictionary takes effect, return to the cluster list. Locate the target cluster and click **Kibana** in the **Operation** column to access the cluster.
6. On the Kibana page, click **Dev Tools** in the navigation tree on the left. The operation page is displayed.
7. Run the following commands to check the performance of different word segmentation policies.

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

- Use the ik_smart word segmentation policy to split the target text.

  Example code:

  ```
  POST /_analyze
  {
    "analyzer":"ik_smart",
    "text":"Text used for word segmentation"
  }
  ```

  After the operation is completed, view the word segmentation result.

  ```
  {
    "tokens": [
      {
        "token": "word-1",
        "start_offset": 0,
        "end_offset": 4,
        "type": "CN_WORD",
        "position": 0
      },
      {
        "token": "word-2",
        "start_offset": 5,
        "end_offset": 8,
        "type": "CN_WORD",
        "position": 1
      }
    ]
  }
  ```

- Use the ik_max_word word segmentation policy to split the target text.

  Example code:

  ```
  POST /_analyze
  {
    "analyzer":"ik_max_word",
    "text":"Text used for word segmentation"
  }
  ```

  After the operation is completed, view the word segmentation result.

  ```
  {
    "tokens" : [
      {
        "token": "word-1",
        "start_offset" : 0,
        "end_offset" : 4,
        "type" : "CN_WORD",
        "position" : 0
      },
      {
        "token" : "word-3",
        "start_offset" : 0,
        "end_offset" : 2,
        "type" : "CN_WORD",
        "position" : 1
      },
      {
        "token" : "word-4",
        "start_offset" : 0,
        "end_offset" : 1,
        "type" : "CN_WORD",
        "position" : 2
      },
      {
        "token" : "word-5",
        "start_offset" : 1,
        "end_offset" : 3,
        "type" : "CN_WORD",
        "position" : 3
      },
      {
  ```

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

```
    "token" : "word-6",
    "start_offset" : 2,
    "end_offset" : 4,
    "type" : "CN_WORD",
    "position" : 4
  },
  {
    "token" : "word-7",
    "start_offset" : 3,
    "end_offset" : 4,
    "type" : "CN_WORD",
    "position" : 5
  },
  {
    "token" : "word-2",
    "start_offset" : 5,
    "end_offset" : 8,
    "type" : "CN_WORD",
    "position" : 6
  },
  {
    "token" : "word-8",
    "start_offset" : 5,
    "end_offset" : 7,
    "type" : "CN_WORD",
    "position" : 7
  },
  {
    "token" : "word-9",
    "start_offset" : 6,
    "end_offset" : 8,
    "type" : "CN_WORD",
    "position" : 8
  },
  {
    "token" : "word-10",
    "start_offset" : 7,
    "end_offset" : 8,
    "type" : "CN_WORD",
    "position" : 9
  }
 ]
}
```

Step 2: Use Keywords for Search

The commands for versions earlier than Elasticsearch 7.*x* are different from those for versions later than Elasticsearch 7.*x*. Examples are as follows.

- **Versions earlier than 7.*x***
  a. Create the **book** index and configure the word segmentation policy.

     In this example, both **analyzer** and **search_analyzer** are set to **ik_max_word**. You can also use **ik_smart**.

     ```
     PUT /book
     {
        "settings": {
           "number_of_shards": 2,
           "number_of_replicas": 1
        },
        "mappings": {
           "type1": {
              "properties": {
                 "content": {
                    "type": "text",
                    "analyzer": "ik_max_word",
                    "search_analyzer": "ik_max_word"
     ```

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

```
            }
          }
        }
      }
}
```

b. Import the text information to the **book** index.

```
PUT /book/type1/1
{
  "content":"Imported text"
}
```

c. Use a keyword to search for the text and view the search results.

```
GET /book/type1/_search
{
  "query": {
    "match": {
      "content": "Keyword"
    }
  }
}
```

Search result

```
{
  "took" : 20,
  "timed_out" : false,
  "_shards" : {
    "total" : 2,
    "successful" : 2,
    "skipped" : 0,
    "failed" : 0
  },
  "hits" : {
    "total" : 1,
    "max_score" : 1.1507283,
    "hits" : [
      {
        "_index" : "book",
        "_type" : "type1",
        "_id" : "1",
        "_score" : 1.1507283,
        "_source" : {
          "content" : "Imported text"
        }
      }
    ]
  }
}
```

- **7.*x* and later versions**

a. Create the **book** index and configure the word segmentation policy.

In this example, both **analyzer** and **search_analyzer** are set to **ik_max_word**. You can also use **ik_smart**.

```
PUT /book
{
    "settings": {
        "number_of_shards": 2,
        "number_of_replicas": 1
    },
    "mappings": {
        "properties": {
            "content": {
                "type": "text",
                "analyzer": "ik_max_word",
                "search_analyzer": "ik_max_word"
            }
        }
    }
}
```

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

b. Import the text information to the **book** index.

```
PUT /book/_doc/1
{
  "content":"Imported text"
}
```

c. Use a keyword to search for the text and view the search results.

```
GET /book/_doc/_search
{
  "query": {
    "match": {
      "content": "Keyword"
    }
  }
}
```

Search result

```
{
  "took" : 16,
  "timed_out" : false,
  "_shards" : {
    "total" : 2,
    "successful" : 2,
    "skipped" : 0,
    "failed" : 0
  },
  "hits" : {
    "total" : {
      "value" : 1,
      "relation" : "eq"
    },
    "max_score" : 1.7260926,
    "hits" : [
      {
        "_index" : "book",
        "_type" : "_doc",
        "_id" : "1",
        "_score" : 1.7260926,
        "_source" : {
          "content" : "Imported text"
        }
      }
    ]
  }
}
```

Step 3: Use Synonyms for Search

The commands for versions earlier than Elasticsearch 7.*x* are different from those for versions later than Elasticsearch 7.*x*. Examples are as follows.

- **Versions earlier than 7.*x***

  a. Create the **myindex** index and configure the word segmentation policy.

```
PUT myindex
{
  "settings": {
    "analysis": {
      "filter": {
        "my_synonym": {
          "type": "dynamic_synonym"
        }
      },
      "analyzer": {
        "ik_synonym": {
          "filter": [
            "my_synonym"
          ],
```

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

```
        "type": "custom",
        "tokenizer": "ik_smart"
      }
    }
  }
 },
 "mappings": {
  "mytype" :{
    "properties": {
     "desc": {
       "type": "text",
       "analyzer": "ik_synonym"
     }
    }
  }
 }
}
```

b. Import the text information to the **myindex** index.

```
PUT /myindex/mytype/1
{
 "desc": "Imported text"
}
```

c. Conduct search based on the synonym and view the search results.

```
GET /myindex/_search
{
 "query": {
  "match": {
    "desc": "Keyword"
  }
 }
}
```

Search result

```
{
 "took" : 2,
 "timed_out" : false,
 "_shards" : {
  "total" : 5,
  "successful" : 5,
  "skipped" : 0,
  "failed" : 0
 },
 "hits" : {
  "total" : 1,
  "max_score" : 0.49445358,
  "hits" : [
   {
     "_index" : "myindex",
     "_type" : "mytype",
     "_id" : "1",
     "_score" : 0.49445358,
     "_source" : {
       "desc" : "Imported text"
     }
   }
  ]
 }
}
```

- **7.*x* and later versions**

a. Create the **myindex** index and configure the word segmentation policy.

```
PUT myindex
{
    "settings": {
        "analysis": {
            "filter": {
                "my_synonym": {
                    "type": "dynamic_synonym"
```

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

```
            }
          },
          "analyzer": {
            "ik_synonym": {
              "filter": [
                "my_synonym"
              ],
              "type": "custom",
              "tokenizer": "ik_smart"
            }
          }
        }
    },
    "mappings": {
      "properties": {
        "desc": {
          "type": "text",
          "analyzer": "ik_synonym"
        }
      }
    }
}
```

b. Import the text information to the **myindex** index.

```
PUT /myindex/_doc/1
{
  "desc": "Imported text"
}
```

c. Conduct search based on the synonym and view the search results.

```
GET /myindex/_search
{
  "query": {
    "match": {
      "desc": "Keyword"
    }
  }
}
```

Search result

```
{
  "took" : 1,
  "timed_out" : false,
  "_shards" : {
    "total" : 1,
    "successful" : 1,
    "skipped" : 0,
    "failed" : 0
  },
  "hits" : {
    "total" : {
      "value" : 1,
      "relation" : "eq"
    },
    "max_score" : 0.1519955,
    "hits" : [
      {
        "_index" : "myindex",
        "_type" : "_doc",
        "_id" : "1",
        "_score" : 0.1519955,
        "_source" : {
          "desc" : "Imported text"
        }
      }
    ]
  }
}
```

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

7.5.2 Configuring YML Parameters

You can modify the **elasticsearch.yml** file.

Modifying Parameter Configurations

1. Log in to the CSS management console.
2. Choose **Clusters** in the navigation pane. On the **Clusters** page, click the name of the target cluster. The cluster information page is displayed.
3. Click **Parameter Configurations** and click **Edit** to modify module parameters as required.

Table 7-29 Module parameters

Module Name	Parameter	Description
Cross-domain Access	http.cors.allow-credentials	Whether to return the Access-Control-Allow-Credentials of the header during cross-domain access Value: **true** or **false** Default value: **false**
	http.cors.allow-origin	Origin IP address allowed for cross-domain access, for example, **122.122.122.122:9200**
	http.cors.max-age	Cache duration of the browser. The cache is automatically cleared after the time range you specify. Unit: s Default value: **1,728,000**
	http.cors.allow-headers	Headers allowed for cross-domain access, including **X-Requested-With**, **Content-Type**, and **Content-Length**. Use commas (,) and spaces to separate headers.
	http.cors.enabled	Whether to allow cross-domain access Value: **true** or **false** Default value: **false**
	http.cors.allow-methods	Methods allowed for cross-domain access, including **OPTIONS**, **HEAD**, **GET**, **POST**, **PUT**, and **DELETE**. Use commas (,) and spaces to separate methods.
Reindexing	reindex.remote.whitelist	Configured for migrating data from the current cluster to the target cluster through the reindex API. The example value is 122.122.122.122:9200.

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

Module Name	Parameter	Description
Custom Cache	indices.queries.cache.size	Cache size in the query phase Value range: 1 to 100 Unit: % Default value: **10%**
Queue Size in a Thread Pool	thread_pool.bulk.queue_size	Queue size in the bulk thread pool. The value is an integer. You need to customize this parameter. Default value: **200**
	thread_pool.write.queue_size	Queue size in the write thread pool. The value is an integer. You need to customize this parameter. Default value: **200**
	thread_pool.force_merge.size	Queue size in the force merge thread pool. The value is an integer. Default value: **1**
Customize	You can add parameters based on your needs.	Customized parameters **NOTE** • Enter multiple values in the format as **[value1, value2, value3...]**. • Separate values by commas (,) and spaces. • Colons (:) are not allowed.

4. After the modification is complete, click **Submit**.In the displayed **Submit Configuration** dialog box, select the box indicating "I understand that the modification will take effect after the cluster is restarted." and click **Yes**.

   If the **Status** is **Succeeded** in the parameter modification list, the modification has been saved. Up to 20 modification records can be displayed.

5. Return to the cluster list and choose **More** > **Restart** in the **Operation** column to restart the cluster and make the modification take effect.
   - You need to restart the cluster after modification, or **Configuration unupdated** will be displayed in the **Task Status** column on the **Clusters** page.
   - If you restart the cluster after the modification, and **Task Status** displays **Configuration error**, the parameter configuration file fails to be modified.

7.5.3 Hot and Cold Data Node Switchover

CSS provides you with cold data nodes. You can store data that requires query response in seconds on high-performance nodes and store data that requires query response in minutes on cold data nodes with large capacity and low specifications.

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

NOTE

- When creating a cluster, you need to configure nodes as data nodes. When you enable the cold data node function, data nodes become hot nodes.
- You can enable the cold data node, master node, and client node functions at the same time.
- You can increase nodes and expand storage capacity of cold data nodes. The maximum storage capacity is determined by the node specifications. Local disks do not support storage capacity expansion.

Hot and Cold Data Node Switchover

If you enable cold data nodes when creating a cluster, the cold data nodes are labeled with **cold**. Other data nodes become hot nodes and are labeled with **hot**. You can specify indexes to allocate data to cold or hot nodes.

You can configure a template to store indices on the specified cold or hot node.

The following figure shows this process. Log in to the **Kibana Console** page of the cluster, modify the template by configuring the index starting with **myindex**, and store the indexes on the cold node. In this case, the **myindex*** date is stored on the cold data node by modifying the template.

- For the 5.*x* version, run the following command to create a template:

```
PUT _template/test
{
    "order": 1,
    "template": "myindex*",
    "settings": {
        "index": {
            "refresh_interval": "30s",
            "number_of_shards": "3",
            "number_of_replicas": "1",
            "routing.allocation.require.box_type": "cold"
        }
    }
}
```

- For 6.*x* or later versions, run the following command to create a template:

```
PUT _template/test
{
  "order": 1,
  "index_patterns": "myindex*",
  "settings": {
    "refresh_interval": "30s",
    "number_of_shards": "3",
    "number_of_replicas": "1",
    "routing.allocation.require.box_type": "cold"
  }
}
```

You can perform operations on the created index.

```
PUT myindex/_settings
 {
       "index.routing.allocation.require.box_type": "cold"
   }
```

You can cancel the configurations of hot and cold data nodes.

```
PUT myindex/_settings
{
       "index.routing.allocation.require.box_type": null
   }
```

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

7.5.4 Managing Indexes

7.5.4.1 Creating and Managing Indexes

Clusters of version 7.6.2 or later support index status management. ISM is a plugin that allows you to automate periodic and administrative operations based on changes on the index age, index size, or number of documents. When using the ISM plug-in, you can define policies that automatically handle index rollovers or deletions based on your needs.

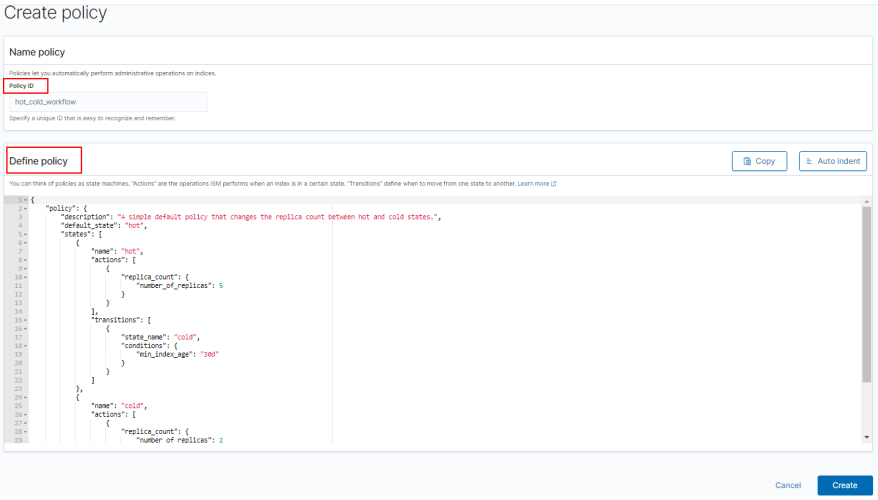
NOTE

The following procedure uses Elasticsearch 7.6.2 as an example. The Kibana UI varies depending on the Kibana version, but their operations are similar.

Creating an Index Policy

1. Log in to Kibana and choose **IM** or **Index Management** on the left. The **Index Management** page is displayed.
2. Click **Create policy** to create an index policy.
3. Enter a policy ID in the **Policy ID** text box and enter your policy in the **Define policy** text box.

   Figure 7-11 Configuring a policy

4. Click **Create**.

Attaching a Policy to an Index

You can attach a policy to one or more indexes and add the policy ID to an index template. When you create indexes using that index template pattern, the policy will be attached to all created indexes.

- **Method 1: Kibana commands**

  On the **Dev Tools** page of Kibana, run the following command to associate a policy ID with an index template:

  ```
  PUT _template/<template_name>
  {
  ```

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

```
  "index_patterns": ["index_name-*"],
  "settings": {
     "opendistro.index_state_management.policy_id": "policy_id"
  }
}
```

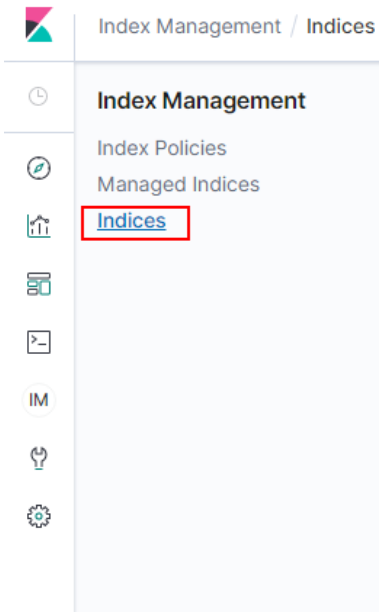
- *<template_name>*: Replace it with the name of a created index template.
- *policy_id*: Replace it with a custom policy ID.

For details about how to create an index template, see **Index Template**.

- **Method 2: Kibana console**

  a. On the **Index Management** page of Kibana, choose **Indices**.

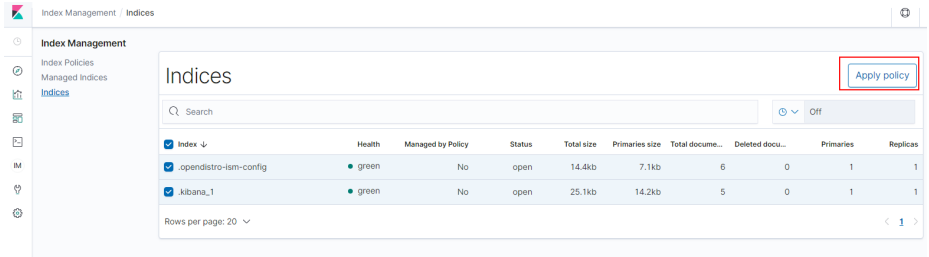
  Figure 7-12 Choosing Indexes

  b. In the **Indices** list, select the target index to which you want to attach a policy.

  c. Click **Apply policy** in the upper right corner.

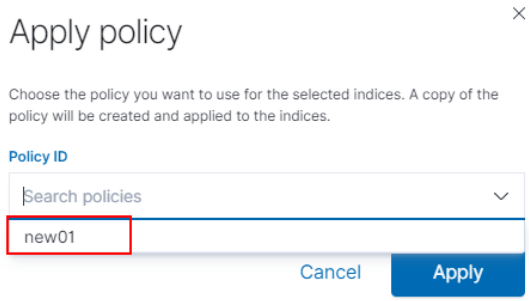
  Figure 7-13 Adding a policy

  d. Select the policy you created from the **Policy ID** drop-down list.

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

Figure 7-14 Selecting a policy



e. Click **Apply**.

After you attach a policy to an index, ISM creates a job that runs every 5 minutes by default, to execute the policy, check conditions, and convert the index to different statuses.

Managing Index Policies

1. Click **Managed Indices**.
2. If you want to change the policy, click **Change policy**. For details, see **Changing Policies**.
3. To delete a policy, select your policy, and click **Remove policy**.
4. To retry a policy, select your policy, and click **Retry policy**.

For details, see **Index State Management**.

7.5.4.2 Changing Policies

You can change any managed index policy. ISM has constraints to ensure that policy changes do not break indexes.

If an index is stuck in its current status, never proceeding, and you want to update its policy immediately, make sure that the new policy includes the same status (same name, action, and order) as the old policy. In this case, ISM applies the new policy even if the policy is being executed.

If you update the policy without including an identical status, ISM updates the policy only after all actions in the current status finish executing. Alternatively, you can select a specific status in the old policy and have the new policy take effect.

To change a policy using Kibana, do the following:

1. Under **Managed Indices**, select the indexes to which you want to attach the new policy.
2. Click **Change policy** in the upper right corner. The **Choose managed indices** page is displayed. Configure parameters required for changing a policy.

Table 7-30 Parameters required for changing a policy

Parameter	Description
Managed indices	Select the indexes to which you want to attach the new policy. Multiple indexes can be selected.

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

Parameter	Description
State filters	Select an index status. When a status is selected, the new policy is attached to an index in this status.
New policy	Select a new policy.

3. After configuration is complete, click **Change**.

7.6 Managing Logs

CSS provides log backup and search functions to help you locate faults. You can back up cluster logs to OBS buckets and download required log files to analyze and locate faults.

Log Query

1. Log in to the CSS management console.
2. Choose **Clusters** in the navigation pane. On the **Clusters** page, click the name of the target cluster. The cluster information page is displayed.
3. In the navigation pane on the left, choose **Log Management**.
4. Query logs on the log management page.

   Select the node, log type, and log level you want to query, and then click  . The query result is displayed.

   When you search for logs, the latest 10,000 logs are matched. A maximum of 100 logs are displayed.

Enabling Log Backup

1. Log in to the CSS management console.
2. Choose **Clusters** in the navigation pane. On the **Clusters** page, click the name of the target cluster. The cluster information page is displayed.
3. Click the **Logs** tab and toggle on the **Log Management** switch.
4. In the **Edit Log Backup Configuration** dialog box, set the parameters.

   In the displayed dialog box, **OBS Bucket**, **Backup Path**, and **IAM Agency** are automatically created for log backup. You can change the default value by referring to **Table 7-31**.

   If the **Log Management** function has been enabled for the cluster, you can click  on the right of **Log Backup Configuration** and modify the configuration in the displayed **Edit Log Backup Configuration** dialog box. For details, see **Table 7-31**.

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

Table 7-31 Parameters for configuring log backup

Parameter	Description	Remarks
OBS Bucket	Select an OBS bucket from the drop-down list for storing logs. You can also click **Create Bucket** on the right to create an OBS bucket.	The OBS bucket and the cluster must be in the same region. **NOTE** To let an IAM user access an OBS bucket, you need to grant the **GetBucketStoragePolicy**, **GetBucketLocation**, **ListBucket**, and **ListAllMyBuckets** permissions to the user.
Backup Path	Storage path of logs in the OBS bucket	The backup path configuration rules are as follows: • The backup path cannot contain the following characters: \:*?"<>| • The backup path cannot start with a slash (/). • The backup path cannot start or end with a period (.). • The total length of the backup path cannot exceed 1,023 characters.
IAM Agency	IAM agency authorized by the current account for CSS to access or maintain data stored in the OBS bucket. You can also click **Create IAM Agency** on the right to create an IAM agency.	The IAM agency must meet the following requirements: • **Agency Type** must be **Cloud service**. • Set **Cloud Service** to **Elasticsearch** or **CSS**. • Mandatory policies: **Tenant Administrator**

5. Back up logs.
   - Automatically backing up logs

     Click the icon on the right of **Auto Backup** to enable the auto backup function.

     After the automatic backup function is enabled, set the backup start time in the **Configure Auto Backup** dialog box. When the scheduled time arrives, the system will back up logs automatically.

     After the **Automatic Snapshot Creation** function is enabled, you can click  on the right of the parameter to change the backup start time.
   - Manually backing up logs

     On the **Log Backup** tab page, click **Back Up**. On the displayed page, click **Yes** to start backup.

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

If **Task Status** in the log backup list is **Successful**, the backup is successful.

NOTE

All logs in the cluster are copied to a specified OBS path. You can view or download log files from the path of the OBS bucket.

6. Search for logs.

   On the **Log Search** page, select the target node, log type, and log level, and click  . The search results are displayed.

   When you search for logs, the latest 10,000 logs are matched. A maximum of 100 logs are displayed.

Viewing Logs

After backing up logs, you can click **Backup Path** to go to the OBS console and view the logs.

Backed up logs mainly include deprecation logs, run logs, index slow logs, and search slow logs. **Table 7-32** lists the storage types of the OBS bucket.

Table 7-32 Log types

Log Name	Description
clustername_deprecation.log	Deprecation log
clustername_index_indexing_slowlog.log	Search slow log
clustername_index_search_slowlog.log	Index slow log
clustername.log	Elasticsearch run log
clustername_access.log	Access log
clustername_audit.log	Audit log

7.7 Managing Plugins

7.7.1 Viewing the Default Plugin List

CSS clusters have default plugins. You can view the default plugin information on the console or Kibana.

Viewing Plugins on the Console

1. Log in to the CSS management console.
2. In the navigation pane, choose **Clusters**. Click the target cluster name and go to the **Basic Information** page of the cluster.

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

3. Click the **Plugins** tab.
4. On the **Default** page, view default plugins supported by the current version.

Viewing Plugins on the Kibana

1. Log in to the CSS management console.
2. In the navigation pane, choose **Clusters**. Locate the target cluster and click **Access Kibana** in the **Operation** column to log in to Kibana.
3. Go to **Dev Tools** and run the following command to view the cluster plugin information:

```
GET _cat/plugins?v
```

The following is an example of the response body:

```
name                 component                      version
css-test-ess-esn-1-1 analysis-dynamic-synonym       7.6.2-xxxx-ei-css-v1.0.1
css-test-ess-esn-1-1 analysis-icu                   7.6.2-xxxx-ei-css-v1.1.6
css-test-ess-esn-1-1 analysis-ik                    7.6.2-xxxx-ei-css-v1.0.1
......
```

name indicates the cluster node name, **component** indicates the plugin name, and **version** indicates the plugin version.

7.7.2 Using the Open Distro SQL Plugin

For Elasticsearch 6.5.4 and later versions, Open Distro for Elasticsearch SQL lets you write queries in SQL rather than in the Elasticsearch query domain-specific language (DSL).

If you are already familiar with SQL and do not want to learn query DSL, this feature is a great option.

Basic Operations

- Kibana (recommended)
  - Log in to Kibana and send requests using request parameters or request body to **_opendistro/_sqlURI** in the **Dev Tools** page.

    ```
    GET _opendistro/_sql?sql=select * from my-index limit 50
    POST _opendistro/_sql
    {
      "query": "SELECT * FROM my-index LIMIT 50"
    }
    ```

  - By default, the result is returned in the JSON structure. If you want the result to be returned in the CSV format, run the following command:

    ```
    POST _opendistro/_sql?format=csv
    {
      "query": "SELECT * FROM my-index LIMIT 50"
    }
    ```

    When data is returned in the CSV format, each row corresponds to a document and each column corresponds to a field.
- cURL commands

  You can also run cURL commands in ECS to execute SQL statements.

  ```
  curl -XPOST https://localhost:9200/_opendistro/_sql -u username:password -k -d '{"query": "SELECT * FROM kibana_sample_data_flights LIMIT 10"}' -H 'Content-Type: application/json'
  ```

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

Supported Operations

Open Distro for Elasticsearch supports the following SQL operations: statements, conditions, aggregations, include and exclude fields, common functions, joins, and show.

- Statements

  Table 7-33 Statements

Statement	Example
Select	SELECT * FROM my-index
Delete	DELETE FROM my-index WHERE _id=1
Where	SELECT * FROM my-index WHERE ['field']='value'
Order by	SELECT * FROM my-index ORDER BY _id asc
Group by	SELECT * FROM my-index GROUP BY range(age, 20,30,39)
Limit	SELECT * FROM my-index LIMIT 50 (default is 200)
Union	SELECT * FROM my-index1 UNION SELECT * FROM my-index2
Minus	SELECT * FROM my-index1 MINUS SELECT * FROM my-index2

  NOTE

  As with any complex query, large UNION and MINUS statements can strain or even crash your cluster.

- Conditions

  Table 7-34 Conditions

Condition	Example
Like	SELECT * FROM my-index WHERE name LIKE 'j%'
And	SELECT * FROM my-index WHERE name LIKE 'j%' AND age > 21
Or	SELECT * FROM my-index WHERE name LIKE 'j%' OR age > 21
Count distinct	SELECT count(distinct age) FROM my-index
In	SELECT * FROM my-index WHERE name IN ('alejandro', 'carolina')
Not	SELECT * FROM my-index WHERE name NOT IN ('jane')

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

Condition	Example
Between	SELECT * FROM my-index WHERE age BETWEEN 20 AND 30
Aliases	SELECT avg(age) AS Average_Age FROM my-index
Date	SELECT * FROM my-index WHERE birthday='1990-11-15'
Null	SELECT * FROM my-index WHERE name IS NULL

- Aggregations

  Table 7-35 Aggregations

Aggregation	Example
avg()	SELECT avg(age) FROM my-index
count()	SELECT count(age) FROM my-index
max()	SELECT max(age) AS Highest_Age FROM my-index
min()	SELECT min(age) AS Lowest_Age FROM my-index
sum()	SELECT sum(age) AS Age_Sum FROM my-index

- Include and exclude fields

  Table 7-36 Include and exclude fields

Pattern	Example
include()	SELECT include('a*'), exclude('age') FROM my-index
exclude()	SELECT exclude('*name') FROM my-index

- Functions

  Table 7-37 Functions

Function	Example
floor	SELECT floor(number) AS Rounded_Down FROM my-index
trim	SELECT trim(name) FROM my-index

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

Function	Example
log	SELECT log(number) FROM my-index
log10	SELECT log10(number) FROM my-index
substring	SELECT substring(name, 2,5) FROM my-index
round	SELECT round(number) FROM my-index
sqrt	SELECT sqrt(number) FROM my-index
concat_ws	SELECT concat_ws(' ', age, height) AS combined FROM my-index
/	SELECT number / 100 FROM my-index
%	SELECT number % 100 FROM my-index
date_format	SELECT date_format(date, 'Y') FROM my-index

NOTE

You must enable fielddata in the document mapping for most string functions to work properly.

- Joins

Table 7-38 Joins

Join	Example
Inner join	SELECT s.firstname, s.lastname, s.gender, sc.name FROM student s JOIN school sc ON sc.name = s.school_name WHERE s.age > 20
Left outer join	SELECT s.firstname, s.lastname, s.gender, sc.name FROM student s LEFT JOIN school sc ON sc.name = s.school_name
Cross join	SELECT s.firstname, s.lastname, s.gender, sc.name FROM student s CROSS JOIN school sc

For details about the restrictions, see **Joins**.

- Show

  Show commands display indexes and mappings that match an index pattern. You can use ***** or **%** for wildcards.

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

Table 7-39 Show

Show	Example
Show tables like	SHOW TABLES LIKE logs-*

Joins

Open Distro for Elasticsearch SQL supports inner joins, left outer joins and cross joins. Joins have the following constraints:

- You can only join two indexes.
- You must use an alias for an index (for example, people p).
- In an ON clause, you can only use the AND conditions.
- In a WHERE statement, do not combine trees that contain multiple indexes. For example, the following statement will work:
  ```
  WHERE (a.type1 > 3 OR a.type1 < 0) AND (b.type2 > 4 OR b.type2 < -1)
  ```
  The following statement will not work:
  ```
  WHERE (a.type1 > 3 OR b.type2 < 0) AND (a.type1 > 4 OR b.type2 < -1)
  ```
- You cannot use GROUP BY or ORDER BY to obtain results.
- LIMIT with OFFSET (for example, LIMIT 25 OFFSET 25) is not supported.

JDBC Driver

The Java Database Connectivity (JDBC) driver allows you to integrate Open Distro for Elasticsearch with your business intelligence (BI) applications.

For details about how to download and use JAR files, see **GitHub Repositories**.

7.8 Kibana Platform

7.8.1 Constraints on Kibana Usage

Kibana has the following constraints:

You can customize the username, role name, and tenant name in Kibana.

7.8.2 Logging In to Kibana

Prerequisites

A CSS cluster has been created.

Procedure

Logging in to the console

1. Log in to the CSS management console.

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

2. On the **Clusters** page, locate the target cluster and click **Access Kibana** in the **Operation** column to go to the Kibana login page.
   - Non-security cluster: The Kibana console is displayed.
   - Security cluster: Enter the username and password on the login page and click **Log In** to go to the Kibana console. The default username is **admin** and the password is the one specified during cluster creation.
3. After the login is successful, you can access the Elasticsearch cluster through Kibana.

7.8.3 Accessing a Cluster from a Kibana Public Network

For CSS clusters that have security mode enabled, you can enable Kibana public access. After the configuration is complete, an IP address will be provided to access Kibana of this cluster over the Internet.

You can configure Kibana public access during cluster creation, or after a cluster in security mode is created.

NOTE

- You can enable **Security Mode** for clusters of version 6.5.4 and later versions.
- Kibana public access cannot be configured for Elasticsearch clusters created in security mode before this function was rolled out (before June 2020).
- The whitelist for Kibana public network access depends on the ELB whitelist. After you updated the whitelist, the new settings take effect immediately for new connections. For existing persistent connections using the IP addresses that have been removed from the whitelist, the new settings take effect about 1 minute after these connections are stopped.

Configuring Kibana Public Access When Creating a Cluster

1. Log in to the CSS management console.
2. Click **Create Cluster** in the upper right corner. The **Create Cluster** page is displayed.
3. On the **Create Cluster** page, enable **Security Mode**.
4. Set **Advanced Settings** to **Custom**, enable **Kibana Public Access**, and set parameters.

Table 7-40 Kibana public access parameters

Parameter	Description
Bandwidth	Bandwidth for accessing Kibana with the public IP address Value range: 1 to 100 Unit: Mbit/s
Access Control	If you disable this function, all IP addresses can access Kibana through the public IP address. If you enable this function, only IP addresses or IP address in the whitelist can access Kibana through the public IP address.

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

Parameter	Description
Whitelist	IP address or IP address range allowed to access a cluster. Use commas (,) to separate multiple addresses. This parameter can be configured only when **Access Control** is enabled. You are advised to enable this function.

After the cluster is created, click the cluster name to go to the **Basic Information** page. On the **Kibana Public Access** page, you can view the Kibana public IP address.

Configuring Kibana Public Access for an Existing Cluster

You can enable, disable, modify, and view Kibana public access for an existing cluster that has security mode enabled.

1. Log in to the CSS management console.
2. Choose **Clusters** in the navigation pane. On the **Clusters** page, click the name of the target cluster.
3. Click the **Kibana Public Access** tab. Turn on the **Kibana Public Access** switch to enable the Kibana public access function.
4. On the displayed page, set parameters.

   Table 7-41 Kibana public access parameters

Parameter	Description
Bandwidth	Bandwidth for accessing Kibana with the public IP address Value range: 1 to 100 Unit: Mbit/s
Access Control	If you disable this function, all IP addresses can access Kibana through the public IP address. If you enable this function, only IP addresses or IP address in the whitelist can access Kibana through the public IP address.
Whitelist	IP address or IP address range allowed to access a cluster. Use commas (,) to separate multiple addresses. This parameter can be configured only when **Access Control** is enabled. You are advised to enable this function.

5. After you set the parameters, click **OK**.

Modifying Kibana Public Access

For clusters configured Kibana public access, you can modify its bandwidth and access control or disable this function.

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

1. Log in to the CSS management console.
2. Choose **Clusters** in the navigation pane. On the **Clusters** page, click the name of the target cluster.
3. Click the **Kibana Public Access** tab to modify the Kibana public access configuration.
   - Modifying bandwidth

     Click **Modify** on the right of **Bandwidth**. On the **Modify Bandwidth** page, modify the bandwidth and click **OK**.
   - Modifying access control

     Click **Modify** on the right of **Access Control**. On the **Modify Access Control** page, set **Access Control** and **Whitelist**, and click **OK**.
   - Disabling Kibana public access

     Toggle off the **Kibana Public Access** switch.

Accessing Kibana with the Public IP Address

After configuring Kibana public access, you will obtain a public IP address that you can use to access Kibana of this cluster.

1. Log in to the CSS management console.
2. Choose **Clusters** in the navigation pane. On the **Clusters** page, click the name of the target cluster.
3. Click the **Kibana Public Access** tab to obtain the Kibana public IP address.
4. Use this IP address to access Kibana of this cluster through the Internet.

7.8.4 Creating a User and Granting Permissions by Using Kibana

Prerequisites

The security mode has been enabled for the cluster.

Parameter Description

Table 7-42 Parameters for creating a user and assigning permissions on Kibana

Parameter	Description
Permission	Single action, for example, creating an index (for example, **indices:admin/create**)
Action group	A group of permissions. For example, the predefined **SEARCH** action group grants roles permissions to use **_search** and **_msearchAPI**.
Role	A role is a combination of permissions and action groups, including operation permissions on clusters, indexes, documents, or fields.

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

Parameter	Description
Backend role	(Optional) Other external roles from the backend such as LDAP/Active Directory
User	A user can send operation requests to the Elasticsearch cluster. The user has credentials such as username and password, zero or more backend roles, and zero or more custom attributes.
Role mapping	A user will be assigned a role after successful authentication. Role mapping is to map a role to a user (or a backend role). For example, the mapping from **kibana_user** (role) to **jdoe** (user) means that John Doe obtains all permissions of **kibana_user** after being authenticated by **kibana_user**. Similarly, the mapping from **all_access** (role) to **admin** (backend role) means that any user with the backend role **admin** (from the LDAP/ Active Directory server) has all the permissions of role **all_access** after being authenticated. You can map each role to multiple users or backend roles.

Procedure

NOTE

- The Kibana UI varies depending on the Kibana version, but their operations are similar. This section takes Kibana 7.6.2 as an example to describe the procedure.
- You can customize the username, role name, and tenant name in Kibana.

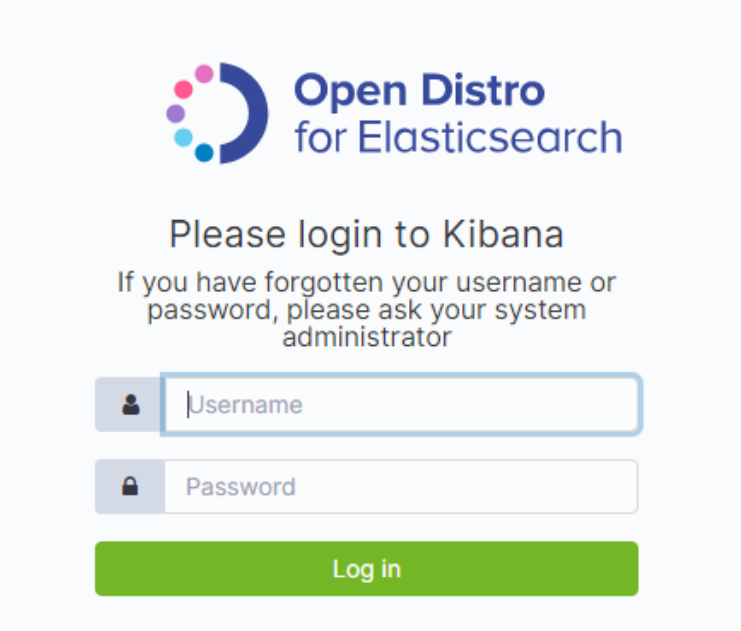
1. Log in to the CSS management console.
2. Choose **Clusters** in the navigation pane. On the **Clusters** page, locate the target cluster and click **Access Kibana** in the **Operation** column.

   Enter the administrator username and password to log in to Kibana.

   – Username: **admin** (default administrator account name)
   – Password: Enter the administrator password you set when creating the cluster in security mode.

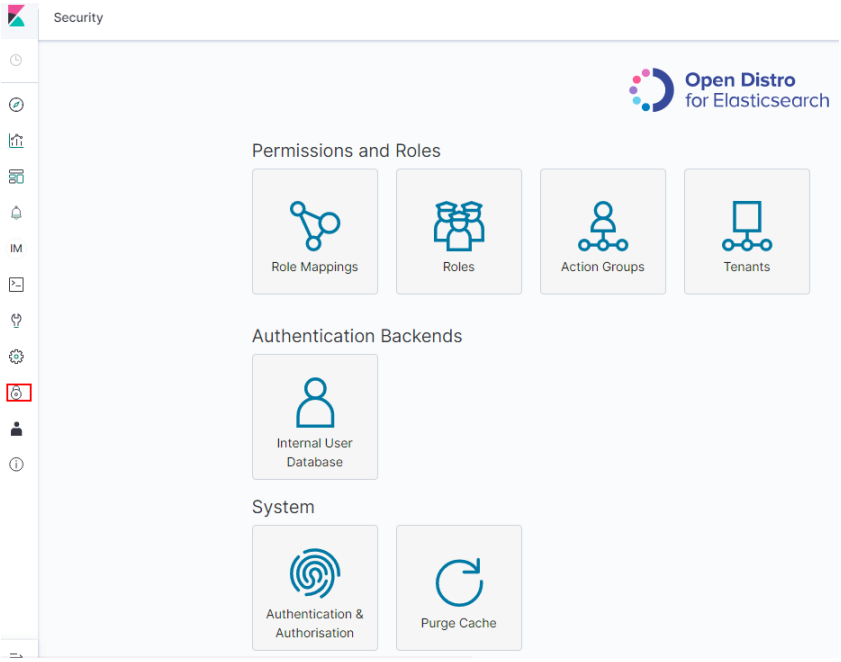
 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

Figure 7-15 Login page



3. Click the **Security** icon on the Kibana operation page.

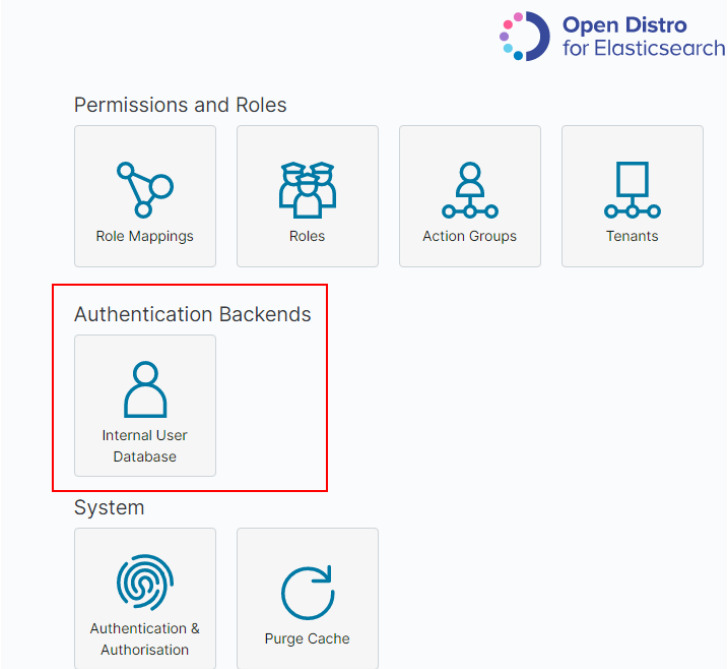
Figure 7-16 Going to the Security page



4. Create a user.
   a. Choose **Authentication Backends** > **Internal Users Database**.

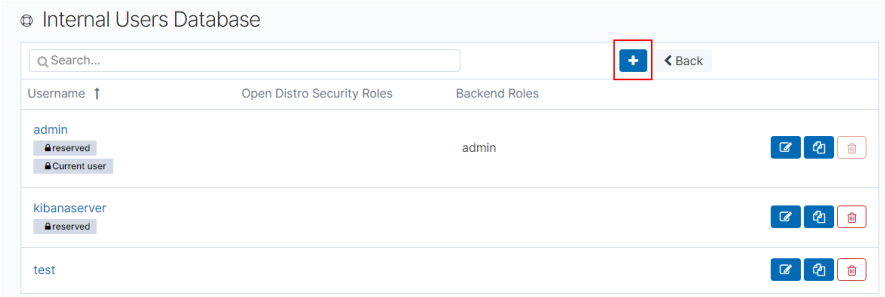
 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

Figure 7-17 Adding a user (1)



b. On the **Internal Users Database** page, choose  . The page for adding user information is displayed.

Figure 7-18 Adding a user (2)



c. On the user creation page, configure **Username** and **Password**, and click **Submit**.

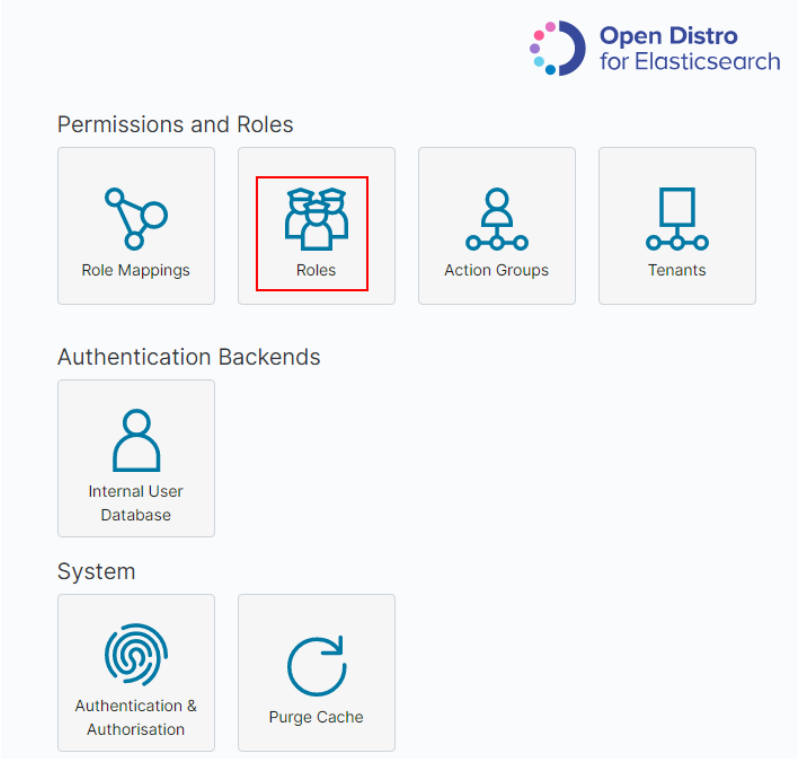
The user will be displayed in the user list.

5. Configure roles and permissions for the user.

   a. Click **Roles**.

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

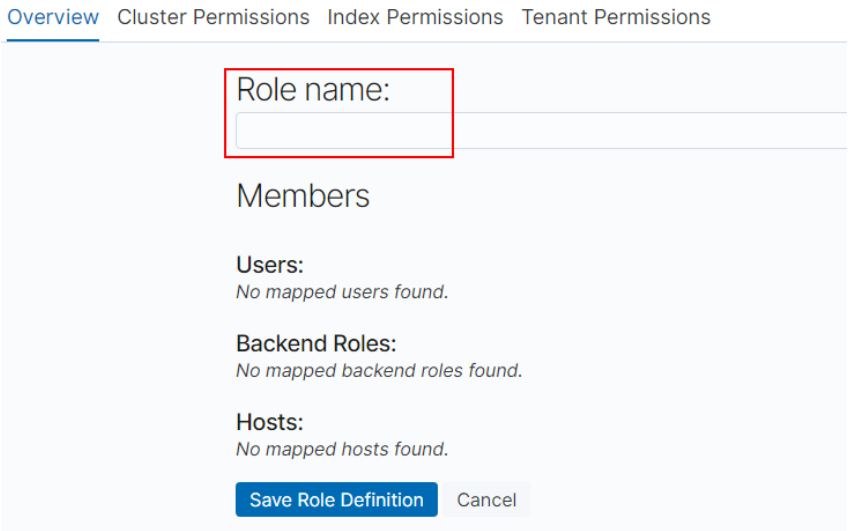
Figure 7-19 Adding a role



b. On the **Open Distro Security Roles** page, click  .

i. Enter a role name on the **Overview** page.

Figure 7-20 Entering a role name



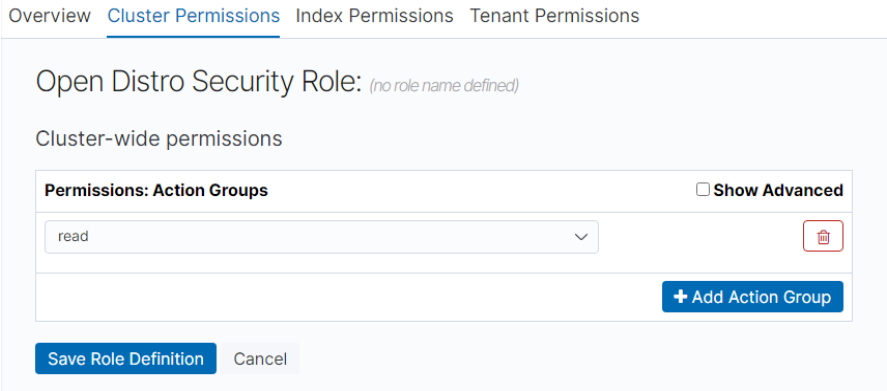
ii. Configure CSS cluster permissions on the **Cluster Permissions** page. Set cluster permissions based on service requirements. If this parameter is not specified for a role, the role has no cluster-level permissions.

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

You can set cluster permissions in **Permissions: Action Groups**. For example, if you select the **read** permission for a cluster, you can only view information such as the cluster status and cluster nodes.

After selecting **Show Advanced**, you can also configure **Permissions: Single Permissions**. For example, if the single permission is set to **indices:data/read**, the indexes can only be read.

Figure 7-21 Cluster Permissions tab



iii. Configure index permissions on the **Index Permissions** page.

Index patterns: Set this parameter to the name of the index whose permission needs to be configured. For example, **my_store**.

NOTE

Use different names for the index and the user.

Configure **Permissions: Action Groups** as required, for example, select the read-only permission **Search**.

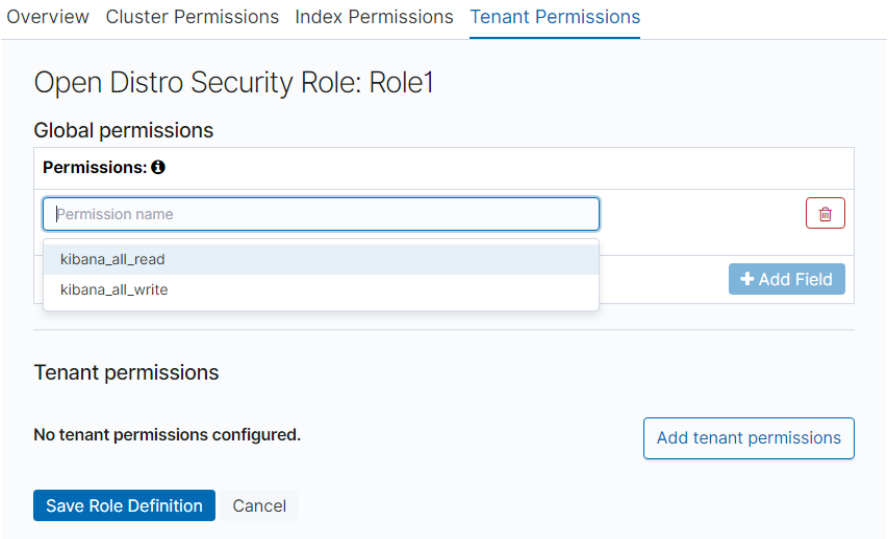
iv. On the **Tenant Permissions** page, set role permissions based on service requirements.

In **Global permissions**, you can select **kibana_all_read** or **kibana_all_write** to set the Kibana read and write permissions for the role.

You can add **tenant patterns** in the **Tenant Permissions** tab and set the **kibana_all_read** or **kibana_all_write** permission for the new tenant patterns.

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

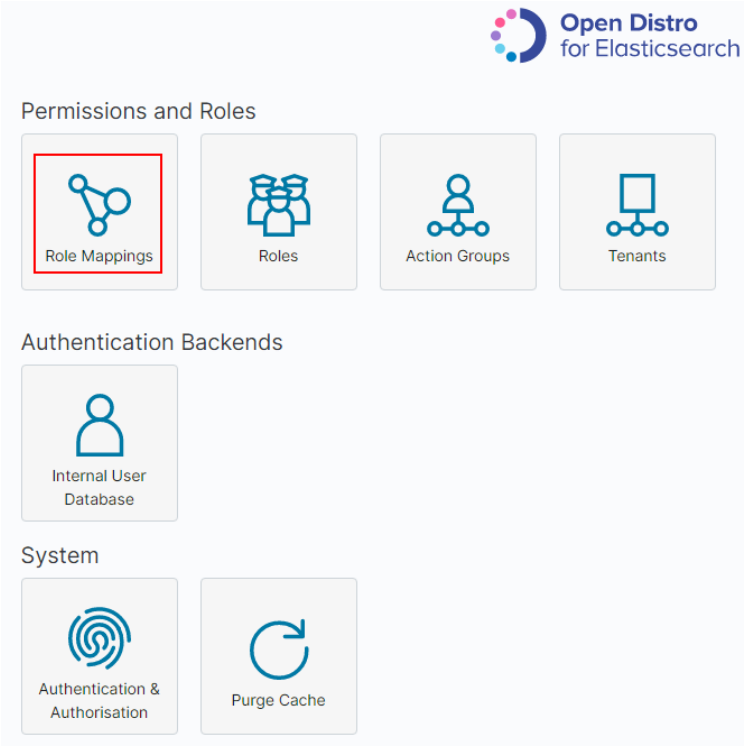
Figure 7-22 Tenant Permissions tab



After the configuration is complete, the role will be displayed.

6. Assign a role to the user.
   a. Click **Role Mappings**.

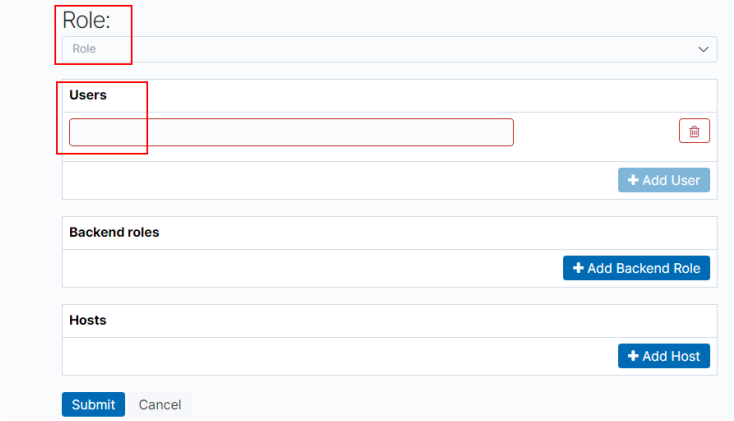
Figure 7-23 Role mapping



   b. Click  to add the mapping between users and roles.

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

Figure 7-24 Users and roles



c. Click **Submit**.

7. Verify that the configuration takes effect in Kibana.

7.8.5 Connecting User-Built Kibana to an Elasticsearch Cluster

To interconnect user-built Kibana with CSS Elasticsearch clusters, the following conditions must be met:

- The local environment must support access from external networks.
- Kibana is built using ECS in the same VPC as Elasticsearch. Kibana can be accessed from the local public network.
- Only Kibana images of the OSS version can be connected to Elasticsearch on CSS.

Example of a Kibana configuration file:

- Security mode:

```
elasticsearch.username: "***"
elasticsearch.password: "***"
elasticsearch.ssl.verificationMode: none
server.ssl.enabled: false
server.rewriteBasePath: false
server.port: 5601
logging.dest: /home/Ruby/log/kibana.log
pid.file: /home/Ruby/run/kibana.pid
server.host: 192.168.xxx.xxx
elasticsearch.hosts: https://10.0.0.xxx:9200
elasticsearch.requestHeadersWhitelist: ["securitytenant","Authorization"]
opendistro_security.multitenancy.enabled: true
opendistro_security.multitenancy.tenants.enable_global: true
opendistro_security.multitenancy.tenants.enable_private: true
opendistro_security.multitenancy.tenants.preferred: ["Private", "Global"]
opendistro_security.multitenancy.enable_filter: false
```

NOTE

  - In security mode, the **opendistro_security_kibana** plug-in must be installed. For details, see **https://github.com/opendistro-for-elasticsearch/security-kibana-plugin/tags?after=v1.3.0.0**.
  - The version of the installed plug-in must be the same as that of the cluster. To check the version of the plug-in version, run the **GET _cat/plugins** command.

- Non-security mode

```
server.port: 5601
logging.dest: /home/Ruby/log/kibana.log
```

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

```
pid.file: /home/Ruby/run/kibana.pid
server.host: 192.168.xxx.xxx
elasticsearch.hosts: http://10.0.0.xxx:9200
```

7.9 Enhanced Cluster Features

7.9.1 Vector Retrieval

7.9.1.1 Description

Image recognition and retrieval, video search, and personalized recommendation impose high requirements on the latency and accuracy of high-dimensional space vector retrieval. To facilitate large-scale vector search, CSS integrates the vector search feature powered by vector search engine and the Elasticsearch plug-in mechanism.

Principles

Vector search works in a way similar to traditional search. To improve vector search performance, we need to:

- **Narrow down the matched scope**

  Similar to traditional text search, vector search use indexes to accelerate the search instead of going through all data. Traditional text search uses inverted indexes to filter out irrelevant documents, whereas vector search creates indexes for vectors to bypass irrelevant vectors, narrowing down the search scope.

- **Reduce the complexity of calculating a single vector**

  The vector search method can quantize and approximate high dimensional vectors first. By doing this, you can acquire a smaller and more relevant data set. Then more sophisticated algorithms are applied to this smaller data set to perform computation and sorting. This way, complex computation is performed on only part of the vectors, and efficiency is improved.

Vector search means to retrieve the k-nearest neighbors (KNN) to the query vector in a given vector data set by using a specific measurement method. Generally, CSS only focuses on Approximate Nearest Neighbor (ANN), because a KNN search requires excessive computational resources.

Functions

The engine integrates a variety of vector indexes, such as brute-force search, Hierarchical Navigable Small World (HNSW) graphs, product quantization, and IVF-HNSW. It also supports multiple similarity calculation methods, such as Euclidean, inner product, cosine, and Hamming. The recall rate and retrieval performance of the engine are better than those of open-source engines. It can meet the requirements for high performance, high precision, low costs, and multi-modal computation.

The search engine also supports all the capabilities of the native Elasticsearch, including distribution, multi-replica, error recovery, snapshot, and permission

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

control. The engine is compatible with the native Elasticsearch ecosystem, including the cluster monitoring tool Cerebro, the visualization tool Kibana, and the real-time data ingestion tool Logstash. Several client languages, such as Python, Java, Go, and C++, are supported.

Constraints

- Only Elasticsearch 7.10.2 clusters support the vector search engine of CSS.
- The vector search plug-in performs in-memory computing and requires more memory than common indexes do. You are advised to use memory-optimized compute resources.
- To use the vector search plug-in of CSS, the memory capacity of a data node or cold data node must be greater than 16 GB. To enable the vector search plug-in, contact technical support.

7.9.1.2 Cluster Planning for Vector Retrieval

Off-heap memory is used for index construction and query in vector retrieval. Therefore, the required cluster capacity is related to the index type and off-heap memory size. You can estimate the off-heap memory required by full indexing to select proper cluster specifications. Due to the high memory usage of vector search, CSS disables the vector search plug-in by default for clusters whose memory is 8 GB or less.

There are different methods for estimating the size of off-heap memory required by different types of indexes. The calculation formulas are as follows:

- **Graph Index**

  $$mem\_needs = (dim \times dim\_size + neighbros \times 4) \times num + delta$$

  NOTE

  If you need to update indexes in real time, consider the off-heap memory overhead required for vector index construction and automatic merge. The actual size of required **mem_needs** is at least 1.5 to 2 times of the original estimation.

- **PQ Index**

  $$mem\_needs = frag\_num \times frag\_size \times num + delta$$

- **FALT and IVF Indexes**

  $$mem\_needs = dim \times dim\_size \times num + delta$$

Table 7-43 Parameter description

Parameter	Description
dim	Vector dimensions
neighbors	Number of neighbors of a graph node. The default value is **64**.

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

Parameter	Description
dim_size	Number of bytes required by each dimension. The default value is four bytes in the float type.
num	Total number of vectors
delta	Metadata size. This parameter can be left blank.

frag_num	Number of vector segments during quantization and coding. If this parameter is not specified when an index is created, the value is determined by the vector dimension **dim**.
	`if dim <= 256:`
	`  frag_num = dim / 4`
	`elif dim <= 512:`
	`  frag_num = dim / 8`
	`else :`
	`  frag_num = 64`

frag_size	Size of the center point during quantization and coding. The default value is **1**. If the value of **frag_num** is greater than **256**, the value of **frag_size** is **2**.

These calculation methods can estimate the size of off-heap memory required by a complete vector index. To determine cluster specifications, you also need to consider the heap memory overhead of each node.

Heap memory allocation policy: The size of the heap memory of each node is half of the node physical memory, and the maximum size is **31 GB**.

For example, if you create a Graph index for the SIFT10M dataset, set **dim** to **128**, **dim_size** to **4**, **neighbors** to default value **64**, and **num** to **10 million**, the off-heap memory required by the Graph index is as follows:

$$mem\_needs = (128 \times 4 + 64 \times 4) \times 10000000 \approx 7.5GB$$

Considering the overhead of heap memory, a single server with **8 vCPUs** and **16 GB memory** is recommended. If real-time write or update is required, you need to apply for larger memory.

7.9.1.3 Creating a Vector Index

Prerequisites

- A cluster of version 7.6.2 or 7.10.2 has been created by referring to **Cluster Planning for Vector Retrieval**.
- Cluster advanced settings have been configured as required by referring to **Advanced Cluster Configurations**.

Creating a Vector Index

1. Log in to the CSS management console.

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

2. Choose **Clusters** in the navigation pane. On the **Clusters** page, locate the target cluster and click **Access Kibana** in the **Operation** column.
3. Click **Dev Tools** in the navigation tree on the left and run the following command to create a vector index.

   Create an index named **my_index** that contains a vector field **my_vector** and a text field **my_label**. The vector field creates the graph index and uses Euclidean distance to measure similarity.

```
PUT my_index
{
  "settings": {
    "index": {
      "vector": true
    }
  },
  "mappings": {
    "properties": {
      "my_vector": {
        "type": "vector",
        "dimension": 2,
        "indexing": true,
        "algorithm": "GRAPH",
        "metric": "euclidean"
      },
      "my_label": {
        "type": "text"
      }
    }
  }
}
```

Table 7-44 Parameters for creating an index

Type	Parameter	Description
Index settings parameters	vector	To use a vector index, set this parameter to **true**.
Field mappings parameters	type	Field type, for example, **vector**.
	dimension	Vector dimension. The default value is **768** and cannot be changed. Value range: [1, 4096]

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

Type	Parameter	Description
	indexing	Whether to enable vector index acceleration. The value can be: • **false**: disables vector index acceleration. If this parameter is set to **false**, vector data is written only to docvalues, and only **ScriptScore** and **Rescore** can be used for vector query. • **true**: enables vector index acceleration. If this parameter is set to **true**, an extra vector index is created. The index algorithm is specified by the **algorithm** field and **VectorQuery** can be used for data query. Default value: **false**

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

Type	Parameter	Description
	algorithm	Index algorithm. This parameter is valid only when **indexing** is set to **true**. The value can be: • **FLAT**: brute-force algorithm that calculates the distance between the target vector and all vectors in sequence. The algorithm relies on sheer computing power and its recall rate reaches 100%. You can use this algorithm if you require high recall accuracy. • **GRAPH**: Hierarchical Navigable Small Worlds (HNSW) algorithm for graph indexes. This algorithm is mainly used in scenarios where high performance and precision are required and the data records of a single shard is fewer than 10 million. • **GRAPH_PQ**: combination of the HNSW algorithm and the PQ algorithm. The PQ algorithm reduces the storage overhead of original vectors, so that HNSW can easily search for data among hundreds of millions of records. • **IVF_GRAPH**: combination of IVF and HNSW. The entire space is divided into multiple cluster centroids, which makes search much faster but slightly inaccurate. You can use this algorithm if you require high performance when searching for data among hundreds of millions of records. • **IVF_GRAPH_PQ**: combination of the PQ algorithm with the IVF or HNSW algorithm to further improve the system capacity and reduce the system overhead. This algorithm is applicable to scenarios where there are more than 1 billion files in shards and high retrieval performance is required. Default value: **GRAPH** **NOTE** If **IVF_GRAPH** or **IVF_GRAPH_PQ** is specified, you need to pre-build and register a central point index. For details, see **(Optional) Pre-Building and Registering a Center Point Vector**.
	Table 7-45	If **Indexing** is set to **true**, CSS provides optional parameters for vector search to achieve higher query performance or precision.

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

Type	Parameter	Description
	metric	Method of calculating the distance between vectors. The value can be: • **euclidean**: Euclidean distance • **inner_product**: inner product distance • **cosine**: cosine distance • **hamming**: Hamming distance, which can be used only when **dim_type** is set to **binary**. Default value: **euclidean**
	dim_type	Type of the vector dimension value. The value can be **binary** and **float** (default).

Table 7-45 Optional parameters

Type	Parameter	Description
Graph index configuration parameters	neighbors	Number of neighbors of each vector in a graph index. The default value is **64**. A larger value indicates higher query precision. A larger index results in a slower build and query speed. Value range: [10, 255]
	shrink	Cropping coefficient during HNSW build. The default value is **1.0f**. Value range: (0.1, 10)
	scaling	Scaling ratio of the upper-layer graph nodes during HNSW build. The default value is **50**. Value range: (0, 128]
	efc	Queue size of the neighboring node during HNSW build. The default value is **200**. A larger value indicates a higher precision and slower build speed. Value range: (0, 100000]
	max_scan_num	Maximum number of nodes that can be scanned. The default value is **10000**. A larger value indicates a higher precision and slower indexing speed. Value range: (0, 1000000]

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

Type	Parameter	Description
PQ index configuration parameters	centroid_num	Number of cluster centroids of each fragment. The default value is **255**. Value range: (0, 65535]
	fragment_num	Number of fragments. The default value is **0**. The plug-in automatically sets the number of fragments based on the vector length. Value range: [0, 4096]

Importing Vector Data

Run the following command to import vector data. When writing vector data to the **my_index** index, you need to specify the vector field name and vector data.

- If the input vector data is an array of floating-point numbers separated by commas (,):

```
POST my_index/_doc
{
  "my_vector": [1.0, 2.0]
}
```

- If the input vector data is a Base64 string encoded using little endian:

  When writing binary vectors or high dimensional vectors that have a large number of valid bits, the Base64 encoding format is efficient for data transmission and parsing.

```
POST my_index/_doc
{
  "my_vector": "AACAPwAAAEA="
}
```

- To write a large amount of data, bulk operations are recommended.

```
POST my_index/_bulk
{"index": {}}
{"my_vector": [1.0, 2.0], "my_label": "red"}
{"index": {}}
{"my_vector": [2.0, 2.0], "my_label": "green"}
{"index": {}}
{"my_vector": [2.0, 3.0], "my_label": "red"}
```

Advanced Cluster Configurations

- When importing data offline, you are advised to set **refresh_interval** of indexes to **-1** to disable automatic index refreshing and improve batch write performance.
- You are advised to set **number_of_replicas** to **0**. After the offline data import is complete, you can modify the parameter value as needed.
- The parameters of other advanced functions as follows:

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

Table 7-46 Cluster parameters

Parameter	Description
native.cache.circuit_breaker.enabled	Whether to enable the circuit breaker for off-heap memory. Default value: **true**
native.cache.circuit_breaker.cpu.limit	Upper limit of off-heap memory usage of the vector index. For example, if the overall memory of a host is 128 GB and the heap memory occupies 31 GB, the default upper limit of the off-heap memory usage is 43.65 GB, that is, (128 - 31) x 45%. If the off-heap memory usage exceeds its upper limit, the circuit breaker will be triggered. Default value: **45%**
native.cache.expire.enabled	Whether to enable the cache expiration policy. If this parameter is set to **true**, some cache items that have not been accessed for a long time will be cleared. Value: **true** or **false** Default value: **false**
native.cache.expire.time	Expiration time. Default value: **24h**
native.vector.index_threads	Number of threads used for creating underlying indexes. Each shard uses multiple threads. Set a relatively small value to avoid resource preemption caused by the build queries of too many threads. Default value: **4**

7.9.1.4 Querying Vectors

Standard Query

Standard vector query syntax is provided for vector fields with vector indexes. The following command will return *n* (specified by **size**/**topk**) data records that are most close to the query vector.

```
POST my_index/_search
{
  "size":2,
  "_source": false,
  "query": {
    "vector": {
      "my_vector": {
        "vector": [1, 1],
        "topk":2
      }
    }
  }
}
```

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

Table 7-47 Parameters for standard query

Parameter	Description
vector (the first one)	Indicates that the query type is **VectorQuery**.
my_vector	Indicates the name of the vector field you want to query.
vector (the second one)	Indicates the vector value you want to query, which can be an array or a Base64 string
topk	Same as the value of **size** generally.
Table 7-48	Indicates optional query parameters. You can adjust the vector index parameters to achieve higher query performance or precision.

Table 7-48 Optional query parameters

Type	Parameter	Description
Graph index configuration parameters	ef	Queue size of the neighboring node during the query. A larger value indicates a higher query precision and slower query speed. The default value is **200**. Value range: (0, 100000]
	max_scan_num	Maximum number of scanned nodes. A larger value indicates a higher query precision and slower query speed. The default value is **10000**. Value range: (0, 1000000]
IVF index configuration parameters	nprobe	Number of center points. A larger value indicates a higher query precision and slower query speed. The default value is **100**. Value range: (0, 100000]

Compound Query

Vector search can be used together with other Elasticsearch subqueries, such as Boolean query and post-filtering, for compound query.

In the following two examples, top 10 (**topk**) results closest to the query vector are queried first. **filter** retains only the results whose **my_label** field is **red**.

- Example of a Boolean query

```
POST my_index/_search
{
  "size": 10,
  "query": {
    "bool": {
      "must": {
```

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

```
      "vector": {
        "my_vector": {
          "vector": [1, 2],
          "topk": 10
        }
      }
    },
    "filter": {
      "term": { "my_label": "red" }
    }
  }
 }
}
```

- Example of post-filtering

```
GET my_index/_search
{
  "size": 10,
  "query": {
    "vector": {
      "my_vector": {
        "vector": [1, 2],
        "topk": 10
      }
    }
  },
  "post_filter": {
    "term": { "my_label": "red" }
  }
}
```

ScriptScore Query

You can use **script_score** to perform Nearest Neighbor Search (NSS) on vectors. The query syntax is provided below.

The pre-filtering condition can be any query. **script_score** traverses only the pre-filtered results, calculates the vector similarity, and sorts and returns the results. The performance of this query depends on the size of the intermediate result set after the pre-filtering. If the pre-filtering condition is set to **match_all**, brute-force search is performed on all data.

```
POST my_index/_search
 {
   "size":2,
   "query": {
   "script_score": {
      "query": {
        "match_all": {}
      },
      "script": {
        "source": "vector_score",
        "lang": "vector",
        "params": {
          "field": "my_vector",
          "vector": [1.0, 2.0],
          "metric": "euclidean"
        }
      }
    }
   }
 }
```

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

Table 7-49 script_score parameters

Parameter	Description
source	Script description. Its value is **vector_score** if the vector similarity is used for scoring.
lang	Script syntax description. Its value is **vector**.
field	Vector field name
vector	Vector data to be queried
metric	Measurement method, which can be **euclidean**, **inner_product**, **cosine**, and **hamming**. Default value: **euclidean**

Re-Score Query

If the **GRAPH_PQ** or **IVF_GRAPH_PQ** index is used, the query results are sorted based on the asymmetric distance calculated by PQ. CSS supports re-scoring and ranking of query results to improve the recall rate.

Assuming that **my_index** is a PQ index, an example of re-scoring the query results is as follows:

```
GET my_index/_search
{
  "size": 10,
  "query": {
    "vector": {
      "my_vector": {
        "vector": [1.0, 2.0],
        "topk": 100
      }
    }
  },
  "rescore": {
    "window_size": 100,
    "vector_rescore": {
      "field": "my_vector",
      "vector": [1.0, 2.0],
      "metric": "euclidean"
    }
  }
}
```

Table 7-50 Rescore parameter description

Parameter	Description
window_size	Vector retrieval returns *topk* search results and ranks the first *window_size* results.
field	Vector field name
vector	Vector data to be queried

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

Parameter	Description
metric	Measurement method, which can be **euclidean**, **inner_product**, **cosine**, and **hamming**. Default value: **euclidean**

Painless Syntax Extension

CSS extension supports multiple vector distance calculation functions, which can be directly used in customized painless scripts to build flexible re-score formulas.

The following is an example:

```
POST my_index/_search
{
  "size": 10,
  "query": {
    "script_score": {
      "query": {
        "match_all": {}
      },
      "script": {
        "source": "1 / (1 + euclidean(params.vector, doc[params.field]))",
        "params": {
          "field": "my_vector",
          "vector": [1, 2]
        }
      }
    }
  }
}
```

The following table lists the distance calculation functions supported by the CSS.

Function Signature	Description
euclidean(Float[], DocValues)	Euclidean distance function
cosine(Float[], DocValues)	Cosine similarity function
innerproduct(Float[], DocValues)	Inner product function
hamming(String, DocValues)	Hamming distance function Only vectors whose **dim_type** is **binary** are supported. The input query vector must be a Base64-encoded character string.

7.9.1.5 Optimizing the Performance of Vector Retrieval

Optimizing Write Performance

- To reduce the cost of backup, disable the backup function before data import and enable it afterwards.

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

- Set **refresh_interval** to **120s** or a larger value. Larger segments can reduce the vector index build overhead caused by merging.
- Increase the value of **native.vector.index_threads** (the default value is **4**) to increase the number of threads for vector index build.

```
PUT _cluster/settings
{
  "persistent": {
    "native.vector.index_threads": 8
  }
}
```

Optimizing Query Performance

- After importing data in batches, you can run the **forcemerge** command to improve the query efficiency.

```
POST index_name/_forcemerge?max_num_segments=1
```

- If the off-heap memory required by the vector index exceeds the circuit breaker limit, index entry swap-in and swap-out occur, which affects the query performance. In this case, you can increase the circuit breaker threshold of off-heap memory.

```
PUT _cluster/settings
{
  "persistent": {
    "native.cache.circuit_breaker.cpu.limit": "75%"
  }
}
```

- If the end-to-end latency is greater than the **took** value in the returned result, you can configure **_source** to reduce the **fdt** file size and reduce the **fetch** overhead.

```
PUT my_index
{
  "settings": {
    "index": {
      "vector": "true"
    },
    "index.soft_deletes.enabled": false
  },
  "mappings": {
    "_source": {
      "excludes": ["my_vector"]
    },
    "properties": {
      "my_vector": {
        "type": "vector",
        "dimension": 128,
        "indexing": true,
        "algorithm": "GRAPH",
        "metric": "euclidean"
      }
    }
  }
}
```

7.9.1.6 (Optional) Pre-Building and Registering a Center Point Vector

When you perform operations in **Creating a Vector Index**, if **IVF_GRAPH** and **IVF_GRAPH_PQ** index algorithms are selected, you need to pre-build and register the center point vector.

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

Context

The vector index acceleration algorithms **IVF_GRAPH** and **IVF_GRAPH_PQ** are suitable for ultra-large-scale computing. These two algorithms allow you to narrow down the query range by dividing a vector space into subspaces through clustering or random sampling. Before pre-build, you need to obtain all center point vectors by clustering or random sampling.

Then, pre-construct and register the center point vectors to create the **GRAPH** or **GRAPH_PQ** index and register them with the Elasticsearch cluster. All nodes in the cluster can share the index file. Reuse of the center index among shards can effectively reduce the training overhead and the number of center index queries, improving the write and query performance.

Procedure

1. On the **Clusters** page, locate the target cluster, and click **Access Kibana** in the **Operation** column.
2. Click **Dev Tools** in the navigation tree on the left.
3. Create a center point index table.
   - For example, if the created index is named **my_dict**, **number_of_shards** of the index must be set to **1**. Otherwise, the index cannot be registered.
   - If you want to use the **IVF_GRAPH** index, set **algorithm** of the center point index to **GRAPH**.
   - If you want to use the **IVF_GRAPH_PQ** index, set **algorithm** of the center point index to **GRAPH_PQ**.

```
PUT my_dict
{
  "settings": {
    "index": {
      "vector": true
    },
    "number_of_shards": 1,
    "number_of_replicas": 0
  },
  "mappings": {
    "properties": {
      "my_vector": {
        "type": "vector",
        "dimension": 2,
        "indexing": true,
        "algorithm": "GRAPH",
        "metric": "euclidean"
      }
    }
  }
}
```

4. Write the center point vector to the created index.

   Write the center point vector obtained through sampling or clustering into the created **my_dict** index by referring to **Importing Vector Data**.

5. Call the registration API.

   Register the created **my_dict** index with a **Dict** object with a globally unique identifier name (**dict_name**).

```
PUT _vector/register/my_dict
{
  "dict_name": "my_dict"
}
```

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

6. Create an **IVF_GRAPH** or **IVF_GRAPH_PQ** index.

   You do not need to specify the dimension and metric information. Simply specify the registered dictionary name.

```
PUT my_index
{
  "settings": {
    "index": {
      "vector": true
    }
  },
  "mappings": {
    "properties": {
      "my_vector": {
        "type": "vector",
        "indexing": true,
        "algorithm": "IVF_GRAPH",
        "dict_name": "my_dict",
        "offload_ivf": false
      }
    }
  }
}
```

Table 7-51 Field mappings parameters

Parameter	Description
dict_name	Specifies the name of the depended central point index. The vector dimension and measurement metric of the index are the same as those of the Dict index.
offload_ivf	Unloads the IVF inverted index implemented by the underlying index to Elasticsearch. In this way, the use of non-heap memory and the overhead of write and merge operations are reduced. However, the query performance also deteriorates. You can use the default value. Value: **true** or **false** Default value: **false**

7.9.1.7 Managing the Vector Index Cache

The vector retrieval engine is developed in C++ and uses off-heap memory. You can use the following APIs to manage the index cache.

- **View cache statistics.**

```
GET /_vector/stats
```

  In the implementation of the vector plug-in, the vector index is the same as other types of Lucene indexes. Each segment constructs and stores an index file. During query, the index file is loaded to the non-heap memory. The plug-in uses the cache mechanism to manage the non-heap memory. You can use this API to query the non-heap memory usage, number of cache hits, and number of loading times.

- **Preload the vector index.**

```
PUT /_vector/warmup/{index_name}
```

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

You can use this API to preload the vector index specified by **index_name** to the off-heap memory for query.

- **Clear the cache.**

```
PUT /_vector/clear/cache
PUT /_vector/clear/cache/index_name
```

The caching mechanism limits the non-heap memory usage when vector indexes are used. When the total index size exceeds the cache size limit, index entry swap-in and swap-out occur, which affects the query performance. You can use this API to clear unnecessary index cache to ensure the query performance of hot data indexes.

7.9.1.8 Sample Code for Vector Search on a Client

Elasticsearch provides standard REST APIs and clients developed using Java, Python, and Go.

Based on the open-source dataset **SIFT1M** (http://corpus-texmex.irisa.fr/) and Python Elasticsearch client, this section provides a code snippet for creating a vector index, importing vector data, and querying vector data on the client.

Prerequisites

The Python dependency package has been installed on the client. If it is not installed, run the following commands to install it:

```
pip install numpy
pip install elasticsearch==7.6.0
```

Sample Code

```
import numpy as np
import time
import json

from concurrent.futures import ThreadPoolExecutor, wait
from elasticsearch import Elasticsearch
from elasticsearch import helpers

endpoint = 'http://xxx.xxx.xxx.xxx:9200/'

# Construct an Elasticsearch client object
es = Elasticsearch(endpoint)

# Index mapping information
index_mapping = '''
{
  "settings": {
    "index": {
      "vector": "true"
    }
  },
  "mappings": {
    "properties": {
      "my_vector": {
        "type": "vector",
        "dimension": 128,
        "indexing": true,
        "algorithm": "GRAPH",
        "metric": "euclidean"
      }
    }
  }
```

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

```
}
'''

# Create an index.
def create_index(index_name, mapping):
    res = es.indices.create(index=index_name, ignore=400, body=mapping)
    print(res)

# Delete an index.
def delete_index(index_name):
    res = es.indices.delete(index=index_name)
    print(res)


# Refresh indexes.
def refresh_index(index_name):
    res = es.indices.refresh(index=index_name)
    print(res)


# Merge index segments.
def merge_index(index_name, seg_cnt=1):
    start = time.time()
    es.indices.forcemerge(index=index_name, max_num_segments=seg_cnt, request_timeout=36000)
    print(f" Complete the merge within {time.time() - start} seconds")


# Load vector data.
def load_vectors(file_name):
    fv = np.fromfile(file_name, dtype=np.float32)
    dim = fv.view(np.int32)[0]
    vectors = fv.reshape(-1, 1 + dim)[:, 1:]
    return vectors


# Load the ground_truth data.
def load_gts(file_name):
    fv = np.fromfile(file_name, dtype=np.int32)
    dim = fv.view(np.int32)[0]
    gts = fv.reshape(-1, 1 + dim)[:, 1:]
    return gts


def partition(ls, size):
    return [ls[i:i + size] for i in range(0, len(ls), size)]


# Write vector data.
def write_index(index_name, vec_file):
    pool = ThreadPoolExecutor(max_workers=8)
    tasks = []

    vectors = load_vectors(vec_file)
    bulk_size = 1000
    partitions = partition(vectors, bulk_size)

    start = time.time()
    start_id = 0
    for vecs in partitions:
        tasks.append(pool.submit(write_bulk, index_name, vecs, start_id))
        start_id += len(vecs)
    wait(tasks)
    print(f" Complete the writing within {time.time() - start} seconds")


def write_bulk(index_name, vecs, start_id):
    actions = [
        {
            "_index": index_name,
```

```
            "my_vector": vecs[j].tolist(),
            "_id": str(j + start_id)
        }
        for j in range(len(vecs))
    ]
    helpers.bulk(es, actions, request_timeout=3600)


# Query an index.
def search_index(index_name, query_file, gt_file, k):
    print("Start query! Index name: " + index_name)

    queries = load_vectors(query_file)
    gt = load_gts(gt_file)

    took = 0
    precision = []
    for idx, query in enumerate(queries):
        hits = set()
        query_json = {
                "size": k,
                "_source": False,
                "query": {
                  "vector": {
                    "my_vector": {
                      "vector": query.tolist(),
                      "topk": k
                    }
                  }
                }
              }
        res = es.search(index=index_name, body=json.dumps(query_json))

        for hit in res['hits']['hits']:
            hits.add(int(hit['_id']))
        precision.append(len(hits.intersection(set(gt[idx, :k])))) / k)
        took += res['took']

    print("precision: " + str(sum(precision) / len(precision)))
    print(f" Complete the retrieval within {took / 1000:.2f} seconds; average took size is {took /
len(queries):.2f} ms")


if __name__ == "__main__":
    vec_file = r"./data/sift/sift_base.fvecs"
    qry_file = r"./data/sift/sift_query.fvecs"
    gt_file = r"./data/sift/sift_groundtruth.ivecs"

    index = "test"
    create_index(index, index_mapping)
    write_index(index, vec_file)
    merge_index(index)
    refresh_index(index)

    search_index(index, qry_file, gt_file, 10)
```

7.9.2 Storage-Compute Decoupling

7.9.2.1 Context

You can store hot data on SSD to achieve the optimal query performance, and store historical data in OBS to reduce data storage costs.

Application Scenarios

A large volume of data is written to and stored in SSDs. If historical data is no longer updated (is turned into cold data) and its QPS decreases, you can call CSS APIs to dump hot data from SSDs to OBS buckets. This operation freezes indexes, decoupling compute from storage.

Constraints

- Currently, only Elasticsearch clusters of the versions 7.6.2 and 7.10.2 support decoupled storage and computing.
- The storage-compute decoupling feature depends on OBS. Therefore, you must comply with the restrictions on OBS bandwidth and QPS. If these restrictions are violated, the performance of queries on OBS will deteriorate. For example, the speed of restoring shards and querying data will become slow.

7.9.2.2 Freezing an Index

Precautions

- Before freezing an index, ensure no data is being written to it. The index will be set to read only before being frozen, and data write will fail.
- After an index is frozen:
  - It becomes read-only.
  - The index data will be dumped to OBS. This process occupies network bandwidth.
  - The query latency of a dumped index will increase. During aggregation, the latency of processing complex queries and reading a large volume of data is long.
  - It cannot be unfrozen. That is, a read-only index cannot be changed to writable.
  - After the freezing is complete, the index data in your local disks will be deleted.

Procedure

1. Log in to the CSS management console.
2. Choose **Clusters** in the navigation pane. On the **Clusters** page, locate the target cluster and click **Access Kibana** in the **Operation** column.
3. Click **Dev Tools** in the navigation tree on the left.
4. Run the following command to freeze a specified index and dump it to OBS:

```
POST ${index_name}/_freeze_low_cost
```

Table 7-52 Parameter description

Parameter	Description
index_name	Name of the index to be frozen.

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

Information similar to the following is displayed:

```
{
   "freeze_uuid": "pdsRgUtSTymVDWR_HoTGFw"
}
```

Table 7-53 Response parameter

Parameter	Description
freeze_uuid	After an index freezing request is submitted, an asynchronous job will be started. The request returns the asynchronous job ID, which can be used to query the progress of the asynchronous job.

NOTE

After an index freezing request is submitted, data cannot be written to the index. During the index freezing, query requests are not affected. After the freezing is complete, the index is closed and then opened. During this period, the index cannot be queried, and the cluster may be in the **red** status for a short time. The index is restored after being opened.

5. Run the following command to check the freezing task progress:

```
GET _freeze_low_cost_progress/${freeze_uuid}
```

Table 7-54 Parameter description

Parameter	Description
freeze_uuid	Asynchronous task ID, which is obtained in **4**.

Information similar to the following is displayed:

```
{

"stage" : "STARTED",
"shards_stats" : {
  "INIT" : 0,
  "FAILURE" : 0,
  "DONE" : 0,
  "STARTED" : 3,
  "ABORTED" : 0
},
"indices" : {
  "data1" : [
    {
      "uuid" : "7OS-G1-tRke2jHZPlckexg",
      "index" : {
        "name" : "data1",
        "index_id" : "4b5PHXJITLaS6AurImfQ9A",
        "shard" : 2
      },
      "start_ms" : 1611972010852,
      "end_ms" : -1,
      "total_time" : "10.5s",
      "total_time_in_millis" : 10505,
      "stage" : "STARTED",
      "failure" : null,
      "size" : {
```

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

```
      "total_bytes" : 3211446689,
      "finished_bytes" : 222491269,
      "percent" : "6.0%"
    },
    "file" : {
      "total_files" : 271,
      "finished_files" : 12,
      "percent" : "4.0%"
    },
    "rate_limit" : {
      "paused_times" : 1,
      "paused_nanos" : 946460970
    }
  },
  {
    "uuid" : "7OS-G1-tRke2jHZPlckexg",
    "index" : {
      "name" : "data1",
      "index_id" : "4b5PHXJITLaS6AurImfQ9A",
      "shard" : 0
    },
    "start_ms" : 1611972010998,
    "end_ms" : -1,
    "total_time" : "10.3s",
    "total_time_in_millis" : 10359,
    "stage" : "STARTED",
    "failure" : null,
    "size" : {
      "total_bytes" : 3221418186,
      "finished_bytes" : 272347118,
      "percent" : "8.0%"
    },
    "file" : {
      "total_files" : 372,
      "finished_files" : 16,
      "percent" : "4.0%"
    },
    "rate_limit" : {
      "paused_times" : 5,
      "paused_nanos" : 8269016764
    }
  },
  {
    "uuid" : "7OS-G1-tRke2jHZPlckexg",
    "index" : {
      "name" : "data1",
      "index_id" : "4b5PHXJITLaS6AurImfQ9A",
      "shard" : 1
    },
    "start_ms" : 1611972011021,
    "end_ms" : -1,
    "total_time" : "10.3s",
    "total_time_in_millis" : 10336,
    "stage" : "STARTED",
    "failure" : null,
    "size" : {
      "total_bytes" : 3220787498,
      "finished_bytes" : 305789614,
      "percent" : "9.0%"
    },
    "file" : {
      "total_files" : 323,
      "finished_files" : 14,
      "percent" : "4.0%"
    },
    "rate_limit" : {
      "paused_times" : 3,
      "paused_nanos" : 6057933087
    }
```


 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

```
    }
   ]
  }
}
```

Table 7-55 Response parameters

Parameter	Description
stage	Status. Its value can be: • **INIT**: The instance has just started or is being initialized. • **FAILURE**: failed • **DONE**: complete • **STARTED**: started • **ABORTED**: Canceled. This field is reserved.
shards_stats	Numbers of shards in each state.
indices	Index status details.

Table 7-56 Return values of **indices**

Parameter	Description
uuid	UUID of the freezing operation
index	Index and shard information
start_ms	Start time
end_ms	End time. If no end time is specified, the value **-1** is displayed.
total_time	Time spent
total_time_in_millis	Time spent, in milliseconds
stage	Status of the current shard.
failure	Failure cause. If no failure occurs, **null** is displayed.
size.total_bytes	Size of files to be frozen, in bytes
size.finished_bytes	Frozen bytes
size.percent	Percentage of frozen bytes
file.total_bytes	Number of files to be frozen
file.finished_bytes	Number of frozen files
file.percent	Percentage of frozen files
rate_limit.paused_times	Number of times that freezing is suspended due to rate limit

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

Parameter	Description
rate_limit.paused_nanos	Duration of freezing task suspension due to rate limit, in nanoseconds

The following parameters are added to a frozen index. For details, see **Table 7-57**.

Table 7-57 Frozen index parameters

Parameter	Description
index.frozen_low_cost	Whether an index is frozen. The value is **true**.
index.blocks.write	Whether data writing is denied in a frozen index. The value is **true**.
index.store.type	Storage type of an index. The value is **obs**.

6. After an index is frozen, its data will be cached. Run the following command to check the current cache status: For details about the cache, see **Configuring Cache**.

```
GET _frozen_stats
GET _frozen_stats/${node_id}
```

Table 7-58 Parameter description

Parameter	Description
node_id	Node ID, which can be used to obtain the cache status of a node.

Information similar to the following is displayed:

```
{
  "_nodes" : {
    "total" : 3,
    "successful" : 3,
    "failed" : 0
  },
  "cluster_name" : "css-zzz1",
  "nodes" : {
    "7uwKO38RRoaON37YsXhCYw" : {
      "name" : "css-zzz1-ess-esn-2-1",
      "transport_address" : "10.0.0.247:9300",
      "host" : "10.0.0.247",
      "ip" : "10.0.0.247",
      "block_cache" : {
        "default" : {
          "type" : "memory",
          "block_cache_capacity" : 8192,
          "block_cache_blocksize" : 8192,
          "block_cache_size" : 12,
          "block_cache_hit" : 14,
          "block_cache_miss" : 0,
          "block_cache_eviction" : 0,
```

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

```
      "block_cache_store_fail" : 0
    }
  },
  "obs_stats" : {
    "list" : {
      "obs_list_count" : 17,
      "obs_list_ms" : 265,
      "obs_list_avg_ms" : 15
    },
    "get_meta" : {
      "obs_get_meta_count" : 79,
      "obs_get_meta_ms" : 183,
      "obs_get_meta_avg_ms" : 2
    },
    "get_obj" : {
      "obs_get_obj_count" : 12,
      "obs_get_obj_ms" : 123,
      "obs_get_obj_avg_ms" : 10
    },
    "put_obj" : {
      "obs_put_obj_count" : 12,
      "obs_put_obj_ms" : 2451,
      "obs_put_obj_avg_ms" : 204
    },
    "obs_op_total" : {
      "obs_op_total_ms" : 3022,
      "obs_op_total_count" : 120,
      "obs_op_avg_ms" : 25
    }
  },
  "reader_cache" : {
    "hit_count" : 0,
    "miss_count" : 1,
    "load_success_count" : 1,
    "load_exception_count" : 0,
    "total_load_time" : 291194714,
    "eviction_count" : 0
  }
},
"73EDpEqoQES749umJqxOzQ" : {
  "name" : "css-zzz1-ess-esn-3-1",
  "transport_address" : "10.0.0.201:9300",
  "host" : "10.0.0.201",
  "ip" : "10.0.0.201",
  "block_cache" : {
    "default" : {
      "type" : "memory",
      "block_cache_capacity" : 8192,
      "block_cache_blocksize" : 8192,
      "block_cache_size" : 12,
      "block_cache_hit" : 14,
      "block_cache_miss" : 0,
      "block_cache_eviction" : 0,
      "block_cache_store_fail" : 0
    }
  },
  "obs_stats" : {
    "list" : {
      "obs_list_count" : 17,
      "obs_list_ms" : 309,
      "obs_list_avg_ms" : 18
    },
    "get_meta" : {
      "obs_get_meta_count" : 79,
      "obs_get_meta_ms" : 216,
      "obs_get_meta_avg_ms" : 2
    },
    "get_obj" : {
      "obs_get_obj_count" : 12,
```


 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

```
      "obs_get_obj_ms" : 140,
      "obs_get_obj_avg_ms" : 11
    },
    "put_obj" : {
      "obs_put_obj_count" : 12,
      "obs_put_obj_ms" : 1081,
      "obs_put_obj_avg_ms" : 90
    },
    "obs_op_total" : {
      "obs_op_total_ms" : 1746,
      "obs_op_total_count" : 120,
      "obs_op_avg_ms" : 14
    }
  },
  "reader_cache" : {
    "hit_count" : 0,
    "miss_count" : 1,
    "load_success_count" : 1,
    "load_exception_count" : 0,
    "total_load_time" : 367179751,
    "eviction_count" : 0
  }
},
"EF8WoLCUQbqJl1Pkqo9-OA" : {
  "name" : "css-zzz1-ess-esn-1-1",
  "transport_address" : "10.0.0.18:9300",
  "host" : "10.0.0.18",
  "ip" : "10.0.0.18",
  "block_cache" : {
    "default" : {
      "type" : "memory",
      "block_cache_capacity" : 8192,
      "block_cache_blocksize" : 8192,
      "block_cache_size" : 12,
      "block_cache_hit" : 14,
      "block_cache_miss" : 0,
      "block_cache_eviction" : 0,
      "block_cache_store_fail" : 0
    }
  },
  "obs_stats" : {
    "list" : {
      "obs_list_count" : 17,
      "obs_list_ms" : 220,
      "obs_list_avg_ms" : 12
    },
    "get_meta" : {
      "obs_get_meta_count" : 79,
      "obs_get_meta_ms" : 139,
      "obs_get_meta_avg_ms" : 1
    },
    "get_obj" : {
      "obs_get_obj_count" : 12,
      "obs_get_obj_ms" : 82,
      "obs_get_obj_avg_ms" : 6
    },
    "put_obj" : {
      "obs_put_obj_count" : 12,
      "obs_put_obj_ms" : 879,
      "obs_put_obj_avg_ms" : 73
    },
    "obs_op_total" : {
      "obs_op_total_ms" : 1320,
      "obs_op_total_count" : 120,
      "obs_op_avg_ms" : 11
    }
  },
  "reader_cache" : {
    "hit_count" : 0,
```

```
      "miss_count" : 1,
      "load_success_count" : 1,
      "load_exception_count" : 0,
      "total_load_time" : 235706838,
      "eviction_count" : 0
    }
  }
 }
}
```

7. Run the following command to reset the cache status:

```
POST _frozen_stats/reset
```

Information similar to the following is displayed:

```
{
 "_nodes" : {
  "total" : 1,
  "successful" : 1,
  "failed" : 0
 },
 "cluster_name" : "Es-0325-007_01",
 "nodes" : {
  "mqTdk2YRSPyOSXfesREFSg" : {
   "result" : "ok"
  }
 }
}
```

NOTE

This command is used to debug performance issues. If you reset the cache status and run this command, you can check the cache command status. You do not need to run this command during service running.

8. Run the following command to check all the frozen indexes:

```
GET _cat/freeze_indices
```

Information similar to the following is displayed:

```
green open data2 0bNtxWDtRbOSkS4JYaUgMQ 3 0  5 0  7.9kb  7.9kb
green open data3 oYMLvw31QnyasqUNuyP6RA 3 0 51 0 23.5kb 23.5kb
```

NOTE

The parameters and return values of this command are the same as those of **_cat/indices** of Elasticsearch.

7.9.2.3 Configuring Cache

After data is dumped to OBS, some data is cached to reduce access to OBS and improve Elasticsearch query performance. Data that is requested for the first time is obtained from OBS. The obtained data is cached in the memory. In subsequent queries, the system searches for data in the cache first. Data can be cached in memory or files.

Elasticsearch accesses different files in different modes. The cache system supports multi-level cache and uses blocks of different sizes to cache different files. For example, a large number of small blocks are used to cache .fdx and .tip files, and a small number of large blocks are used to cache .fdt files.

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

Table 7-59 Cache configurations

Parameter	Type	Description
low_cost.obs.blockcache.names	Array	The cache system supports multi-level cache for data of different access granularities. This configuration lists the names of all caches. If this parameter is not set, the system has a cache named **default**. To customize the configuration, ensure there is a cache named **default**. Default value: **default**
low_cost.obs.blockcache.<NAME>.type	ENUM	Cache type, which can be **memory** or **file**. If it is set to **memory**, certain memory will be occupied. If it is set to **file**, cache will be stored in disks. You are advised to use ultra-high I/O disks to improve cache performance. Default value: **memory**
low_cost.obs.blockcache.<NAME>.blockshift	Integer	Size of each block in the cache. Its value is the number of bytes shifted left. For example, if this parameter is set to **16**, the block size is 2^{16} bytes, that is, 65536 bytes (64 KB). Default value: **13** (8 KB)
low_cost.obs.blockcache.<NAME>.bank.count	Integer	Number of cache partitions. Default value: **1**
low_cost.obs.blockcache.<NAME>.number.blocks.perbank	Integer	Number of blocks included in each cache partition. Default value: **8192**
low_cost.obs.blockcache.<NAME>.exclude.file.types	Array	Extensions of files that are not cached. If the extensions of certain files are neither in the **exclude** list nor in the **include** list, they are stored in the default cache.
low_cost.obs.blockcache.<NAME>.file.types	Array	Extensions of cached files. If the extensions of certain files are neither in the **exclude** list nor in the **include** list, they are stored in the default cache.

The following is a common cache configuration. It uses two levels of caches, **default** and **large**. The **default** cache uses 64 KB blocks and has a total of 30 x 4096 blocks. It is used to cache files except .fdt files. The **large** cache uses 2 MB blocks and contains 5 x 1000 blocks. It is used to cache .fdx, .dvd, and .tip files.

```
low_cost.obs.blockcache.names: ["default", "large"]
low_cost.obs.blockcache.default.type: file
low_cost.obs.blockcache.default.blockshift: 16
```

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

```
low_cost.obs.blockcache.default.number.blocks.perbank: 4096
low_cost.obs.blockcache.default.bank.count: 30
low_cost.obs.blockcache.default.exclude.file.types: ["fdt"]

low_cost.obs.blockcache.large.type: file
low_cost.obs.blockcache.large.blockshift: 21
low_cost.obs.blockcache.large.number.blocks.perbank: 1000
low_cost.obs.blockcache.large.bank.count: 5
low_cost.obs.blockcache.large.file.types: ["fdx", "dvd", "tip"]
```

Table 7-60 Other parameters

Parameter	Type	Description
index.frozen.obs.max_bytes_per_sec	String	Maximum rate of uploading files to OBS during freezing. It takes effect immediately after you complete configuration. Default value: **150MB**
low_cost.obs.index.upload.threshold.use.multipart	String	If the file size exceeds the value of this parameter during freezing, the multipart upload function of OBS is used. Default value: **1GB**
index.frozen.reader.cache.expire.duration.seconds	Integer	Timeout duration. To reduce the heap memory occupied by frozen indexes, the reader caches data for a period of time after the index shard is started, and stops caching after it times out. Default value: **300s**
index.frozen.reader.cache.max.size	Integer	Maximum cache size. Default value: **100**

7.9.2.4 Enhanced Cold Data Query Performance

Context

When you query data on the **Discover** page of Kibana for the first time, all data needs to be obtained from OBS because there is no cache. If a large number of documents are returned, it takes a long time to obtain the corresponding time fields and file metadata from OBS. To accelerate queries the first time they run on the **Discover** page, you can cache data locally.

Prerequisites

This feature is available in Elasticsearch clusters of versions 7.6.2 and 7.10.2 and Opensearch clusters created after February 2023.

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

API for Querying Cold Data from Local Cache

This API can be used to query the cold data from local cache.

Example request:

```
GET /_frozen_stats/local_cache
GET /_frozen_stats/local_cache/{nodeId}
```

Response example:

```
{
  "_nodes" : {
    "total" : 1,
    "successful" : 1,
    "failed" : 0
  },
  "cluster_name" : "elasticsearch",
  "nodes" : {
    "6by3lPy1R3m55Dcq3liK8Q" : {
      "name" : "node-1",
      "transport_address" : "127.0.0.1:9300",
      "host" : "127.0.0.1",
      "ip" : "127.0.0.1",
      "local_cache" : {
        "get_stats" : {
          "get_total_count" : 562,                    //Total number of times data was retrieved from the local
cold data cache.
          "get_hit_count" : 562,                      //Total number of hits in the local cold data cache.
          "get_miss_count" : 0,                       //Total number of local cold data cache misses.
          "get_total_ns" : 43849200,                    //Total duration for retrieving data from the local cold
data cache.
          "get_avg_ns" : 78023                        //Average duration for retrieving data from the local cold
data cache.
        },
        "load_stats" : {
          "load_count" : 2,                          //Number of times cold data was loaded from the local
cache
          "load_total_ms" : 29,                       //Total duration for loading cold data from the local cache
          "load_avg_ms" : 14,                         //Average duration for loading cold data from the local
cache
          "load_fail_count" : 0,                      //Number of failure times for loading cold data from the
local cache
          "load_overflow_count" : 0                     //Number of times the local cold data cache exceeds
the cache pool size.
        },
        "reload_stats" : {
          "reload_count" : 0,                        //Number of times the local cold data cache was
regenerated.
          "reload_total_ms" : 0,                      //Total duration for regenerating the local cold data
cache.
          "reload_avg_ms" : 0,                        //Average duration for regenerating the local cold data
cache.
          "reload_fail_count" : 0                     //Number of failures in regenerating the local cold data
cache.
        },
        "init_stats" : {
          "init_count" : 0,                          //Number of times the local cold data cache was initialized.
          "init_total_ms" : 0,                        //Total duration for initializing the local cold data cache.
          "init_avg_ms" : 0,                          //Average duration for initializing the local cold data
cache.
          "init_fail_count" : 0                       //Number of failures in initializing the local cold data
cache.
        }
      }
    }
  }
}
```

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

Configuring Parameters

Configuration Item	Type	Unit	Value Range	Scope	Can Be Dynamically Modified	Description
low_cost.local_cache.max.capacity	Integer	-	The value ranges from 10 to 5000. The default value is **500**.	node	Yes	Maximum number of available cold data caches on a node. Each shard corresponds to a cache object. **NOTE** • If the heap memory usage remains high, decrease the value. • If the value of **load_overflow_count** keeps increasing rapidly, increase the value.
index.low_cost.local_cache.threshold	Integer	%	The value ranges from 0 to 100. The default value is **50**.	index	Yes	Threshold for enabling the local cache of cold data. **NOTE** • If the percentage of date fields is less than the value of this parameter, the cold data of the date type will be cached locally. Otherwise, this parameter is not used. • If the date fields of the current index occupy most of the data volume of the current index, you are not advised to use this function.

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

Configuration Item	Type	Unit	Value Range	Scope	Can Be Dynamically Modified	Description
index.low_cost.local_cache.evict_time	String	Days	The value ranges from 1d to 365d. The default value is **30d**.	index	Yes	Wait time before cold data is deleted from local cache. The value is determined based on **index.frozen_date** (time when the freezing is successful). **NOTE** • For indexes that have been frozen in old clusters and do not have **index.frozen_date** specified, the value of this parameter is determined based on the index creation time. • You are advised to adjust the deletion time based on the disk usage to avoid high disk usage.

Modifying Parameters

- Run the following command to modify **low_cost.local_cache.max.capacity**:
```
PUT _cluster/settings
{
  "persistent": {
    "low_cost.local_cache.max.capacity":1000
  }
}
```
- Run the following command to modify **index.low_cost.local_cache.threshold**:
```
PUT es_write_pref2-00000000021/_settings
{
"index.low_cost.local_cache.threshold":20
}
```
- Run the following command to modify **index.low_cost.local_cache.evict_time**:
```
PUT es_write_pref2-00000000021/_settings
{
"index.low_cost.local_cache.evict_time":"7d"
}
```

7.9.2.5 Monitoring OBS Operations

To clearly display the operations of the storage and compute decoupling plugin in OBS, the real-time OBS rate metric is added to CSS and recorded in the system index.

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

Prerequisite

This feature is available at Elasticsearch of versions 7.6.2 and 7.10.2 and Opensearch clusters created after March 2023.

Description

- The **GET _frozen_stats/obs_rate** API is used to query the real-time rate of OBS operations.
- The system index **.freeze_obs_rate-YYYY.mm.dd** is added to store the real-time OBS operation rate and OBS operation data, helping you monitor the OBS operations.
- The **low_cost.obs_rate_index.evict_time** parameter is added to control the storage duration of the **.freeze_obs_rate-YYYY.mm.dd** index

GET _frozen_stats/obs_rate API

- Calculation method: The average OBS operation rate in the last 5 seconds is calculated every 5 seconds.
- Example request:

```
GET _frozen_stats/obs_rate
GET _frozen_stats/obs_rate/{nodeId}
```

  {nodeId} indicates the ID of the node whose OBS operation rate you want to query.

- Example response:

```
{
  "_nodes" : {
    "total" : 1,
    "successful" : 1,
    "failed" : 0
  },
  "cluster_name" : "elasticsearch",
  "nodes" : {
    "dflDvcSwTJ-fkillT2zE3A" : {
      "name" : "node-1",
      "transport_address" : "127.0.0.1:9300",
      "host" : "127.0.0.1",
      "ip" : "127.0.0.1",
      "update_time" : 1671777600482,                    // Time when the current statistics are updated.
      "obs_rate" : {
        "list_op_rate" : 0.0,                           // Rate of OBS list operations. Unit: times/s.
        "get_meta_op_rate" : 0.0,                       // Rate of OBS get meta operations. Unit: times/s.
        "get_obj_op_rate" : 0.0,                        // Rate of OBS get operations. Unit: times/s.
        "put_op_rate" : 0.0,                            // Rate of OBS put operations. Unit: times/s.
        "obs_total_op_rate" : 0.0,                      // Rate of all OBS operations. The unit is times/s.
        "obs_upload_rate" : "0.0 MB/s",                 // Data upload rate of OBS, in MB/s.
        "obs_download_rate" : "0.0 MB/s"                // Data download rate of OBS, in MB/s.
      }
    }
  }
}
```

System Index

- System index name: **.freeze_obs_rate-YYYY.mm.dd**.
- Example: **.freeze_obs_rate-2023.01.23**

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

NOTE

The default retention period of indexes is 30 days.

Configuration Item

Configuration Item	Type	Scope	Can Be Dynamically Modified	Description
low_cost.obs_rate_index.evict_time	String	node	Yes	The retention period of the **.freeze_obs_rate-YYYY.mm.dd** index. • Value range: 1d to 365d • Default value: **30d** • Unit: day

For example, run the following command to modify the retention period of the **.freeze_obs_rate-YYYY.mm.dd** index:

```
PUT _cluster/settings
{
  "persistent": {
    "low_cost.obs_rate_index.evict_time":  "7d"
  }
}
```

7.9.3 Flow Control

7.9.3.1 Flow Control 2.0

7.9.3.1.1 Context

Feature Description

CSS can control traffic at the node level. You can configure the blacklist and whitelist, the maximum concurrent HTTP connections, and the maximum HTTP connections for a node. You can also configure backpressure based on client traffic in the node memory and block access in one click. CSS can also collect statistics on node access IP addresses and URIs. Each function has an independent control switch, which is disabled by default. To restore default values of parameters, set them to **null**.

After the client write traffic backpressure and control is enabled, large requests will be rejected when too much node heap memory has been occupied. This function prevents nodes from being suspended and reduces the risk of node unavailability.

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

- **HTTP/HTTPS flow control:**
  - You can control client IP address access by setting IP addresses and subnets in HTTP/HTTPS blacklist or whitelist. If an IP address is in the blacklist, the client is disconnected and all its request are rejected. Whitelist rules take precedence over blacklist rules. If a client IP address exists in both the blacklist and whitelist, the client request will not be rejected.
  - HTTP/HTTPS concurrent connection flow control limits the total number of HTTP connections to a node per second.
  - HTTP/HTTPS new connection flow control limits the number of new connections to a node.
- Memory flow control limits the write traffic based on the node heap memory. You can back pressure requests to the client, trigger resource recycling as much as possible, and then accept requests based on the available heap memory.
- Request sampling can record the access of client IP addresses and the type of requests from the client. Based on the statistics, you can identify the access traffic of client IP addresses and analyze the client write and query requests.
- One-click access blocking can block all the access traffic of a node, excluding the traffic from Kibana and CSS O&M and monitoring APIs.
- Flow control provides an independent API for viewing traffic statistics and records the number of current client connections and client backpressure connections. You can evaluate the flow control threshold and analyze the cluster loads based on the statistics.
- Access logs record the URLs and bodies of HTTP/HTTPS requests received by nodes within a period of time. You can analyze the current traffic pressure based on the access logs.

Constraints

- Currently, only versions 7.6.2 and 7.10.2 support the flow control feature.
- Clusters of versions 7.6.2 and 7.10.2 created after January 2023 support only traffic control version 2.0. Clusters created before January 2023 support only traffic control version 1.0.

7.9.3.1.2 HTTP/HTTPS Flow Control

Context

You can run commands in Kibana to enable or disable HTTP/HTTPS flow control for your cluster. The command parameters are as follows.

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

Table 7-61 HTTP/HTTPS flow control parameters

Parameter	Type	Description
flowcontrol.http.enabled	Boolean	Whether to enable HTTP/HTTPS flow control. This function is disabled by default. Enabling it may affect node access performance. Value: **true** or **false** Default value: **false**
flowcontrol.http.allow	List<String>	IP address whitelist. It can contain multiple IP addresses and masks, or an IP address list. Use commas (,) to separate multiple values. Example: *xx.xx.xx.xx***/24,***xx.xx.xx.xx***/24**, or *xx.xx.xx.xx.xx***,***xx.xx.xx*. The default value is null.
flowcontrol.http.deny	List<String>	IP address blacklist. Multiple IP addresses and masks or an IP address list can be configured. Use commas (,) to separate multiple IP addresses and masks. The default value is null.
flowcontrol.http.concurrent	Integer	Maximum concurrent HTTP/HTTPS connections. Default value: Number of available cores on a node x 400
flowcontrol.http.newconnect	Integer	Maximum new connections that can be created for HTTP/HTTPS requests per second. Default value: Number of available cores on a node x 200
flowcontrol.http.warmup_period	Integer	Time required for the HTTP/HTTPS connection setup speed to reach the maximum. If **flowcontrol.http.newconnect** is set to **100** and **flowcontrol.http.warmup_period** is set to **5000ms**, it indicates the system can set up 100 connections per second in 5 seconds. Value range: 0–10000 Unit: ms Default value: **0**

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

Procedure

1. Log in to the CSS management console.
2. Choose **Clusters** in the navigation pane. On the **Clusters** page, locate the target cluster and click **Access Kibana** in the **Operation** column.
3. In the navigation pane on the left, choose **Dev Tools** and run commands to enable or disable HTTP/HTTPS flow control.
   - Enabling HTTP/HTTPS flow control for a node

     ```
     PUT /_cluster/settings
     {
       "persistent": {
         "flowcontrol.http.enabled": true,
         "flowcontrol.http.allow": ["192.168.0.1/24", "192.168.2.1/24"],
         "flowcontrol.http.deny": "192.168.1.1/24",
         "flowcontrol.http.concurrent": 1000,
         "flowcontrol.http.newconnect": 1000,
         "flowcontrol.http.warmup_period": 0
       }
     }
     ```

     NOTE

     If all parameters are set to **null**, they will be restored to default values.
   - Disabling HTTP/HTTPS flow control for a node

     ```
     PUT /_cluster/settings
     {
       "persistent": {
         "flowcontrol.http.enabled": false
       }
     }
     ```

7.9.3.1.3 Memory Flow Control

Context

Elasticsearch provides a circuit breaker, which will terminate requests or return the error code **429** if the memory usage exceeds its threshold. However, the circuit breaker rejects a request only after the node reads the entire request, which occupies heap memory. To prevent a request from being fully received by a node before the request is rejected, you can control the client traffic based on the real-time status of the node heap memory.

Configuring Parameters

The following table describes memory flow control parameters.

Table 7-62 Memory flow control parameters

Parameter	Type	Description
flowcontrol.memory.enabled	Boolean	Whether to enable memory flow control. After this function is enabled, the memory usage is continuously monitored. The value can be: • **true** • **false** (default value)

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

Parameter	Type	Description
flowcontrol.memory.heap_limit	String	Maximum global heap memory usage of a node. If the value of this parameter is exceeded, traffic backpressure is performed. Value range: 10%–100% Default value: **90%**
flowcontrol.holding.in_flight_factor	Float	Backpressure release factor. The principle is similar to that of the circuit breaker parameter **network.breaker.inflight_requests.overhead**. When the memory usage reaches the limit, a larger value indicates stronger backpressure. The write traffic will be limited. Value range: ≥ 0.5 Default value: **1.0**
flowcontrol.holding.max	TimeValue	Maximum delay of each request. If the delay exceeds the value of this parameter, you can disconnect the request backpressure or disconnect the request link. For details, see the configuration of **flowcontrol.holding.max_strategy**. Value range: ≥ 15s Default value: **60s**
flowcontrol.holding.max_strategy	String	Policy after the maximum delay time is exceeded. The value can be: • **keep** (default value): If the heap memory is still high, continue the backpressure. The server determines when to execute the request based on the real-time memory. • **soft**: The requests will be executed even if the heap memory is still high. The **inFlight** circuit breaker will determine whether to execute or reject the requests. • **hard**: If the heap memory is still high, requests will be discarded and the client connection of the requests will be disconnected.

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

Parameter	Type	Description
flowcontrol.memory.once_free_max	String	Maximum memory that can be opened at a time for a suspended request queue. This parameter is used to prevent a cluster from being entirely suspended due to temporary low memory under high pressure. Value range: 1 to 50 Default value: **10%**
flowcontrol.memory.nudges_gc	Boolean	Whether to trigger garbage collection to ensure write stability when the write pressure is too high. (The backpressure connection pool is checked every second. The write pressure is regarded high if all the existing connections are blocked and new write requests cannot be released.) The value can be: • **true** (default value) • **false**

NOTE

- **flowcontrol.memory.enabled** and **flowcontrol.memory.heap_limit** are the most important parameters. *enabled* indicates the memory flow control switch, and *heap_limit* indicates the heap memory threshold of a node.
- The default value 90% of **flowcontrol.memory.heap_limit** is a conservative threshold. When the heap memory usage is greater than 90%, the system stops reading large requests that exceed 64 KB from the client until the heap memory decreases. If the heap memory decreases to 85%, the maximum client data that can be read is 5% of the maximum heap memory. If the heap memory usage has been higher than 90% for a long time, client connection requests cannot be read. In this case, the GC algorithm is triggered to perform garbage collection until the heap memory usage is lower than the threshold.
- Generally, you can set the **flowcontrol.memory.heap_limit** threshold to 80% or less to ensure that the node has certain heap memory for operations besides data writing, such as Elasticsearch query and segment merge.

Procedure

1. Log in to the CSS management console.
2. Choose **Clusters** in the navigation pane. On the **Clusters** page, locate the target cluster and click **Access Kibana** in the **Operation** column.
3. In the navigation pane on the left, choose **Dev Tools** and run commands to enable or disable memory flow control.
   - Enable memory flow control

```
PUT /_cluster/settings
{
  "persistent": {
    "flowcontrol.memory.enabled": true,
    "flowcontrol.memory.heap_limit": "80%"
```

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

```
  }
}
```

- Disable cluster memory flow control

```
PUT /_cluster/settings
{
  "persistent": {
    "flowcontrol.memory.enabled": false
  }
}
```

7.9.3.1.4 Request Sampling

Context

Request sampling can record the access of client IP addresses and the type of requests from the client. Based on the statistics, you can identify the access traffic of client IP addresses and analyze the client write and query requests.

Table 7-63 Request statistics parameters

Parameter	Type	Description
flowcontrol.log.access.enabled	Boolean	Whether to collect statistics on the IP addresses of clients that accessed the ES cluster recently and the number of requests. The value can be: • **true** • **false** (default value)
flowcontrol.log.access.count	Integer	Number of client IP addresses that accessed a cluster recently. Value range: 0–100 Default value: **10**
flowcontrol.log.file.enabled	Boolean	Whether to record the log details of each request to the background log file. The value can be: • **true** • **false** (default value)

NOTE

- IP address statistics switches control whether to collect request type statistics and whether to enable logging.
- **flowcontrol.log.access.enabled** controls whether to collect statistics on client requests, including bulk write, search, and msearch requests.
- **flowcontrol.log.file.enabled** is the log access switch. Request details can be directly recorded in log files for audit analysis.

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

7.9.3.1.5 One-click Traffic Blocking

You can block all connections in one click, except the connections that passes through O&M APIs, to handle unexpected traffic burst and quickly recover your cluster.

Procedure

1. Log in to the CSS management console.
2. Choose **Clusters** in the navigation pane. On the **Clusters** page, locate the target cluster and click **Access Kibana** in the **Operation** column.
3. In the navigation pane on the left, choose **Dev Tools** and run commands to enable or disable one-click traffic blocking.
   - Enable one-click traffic blocking
     ```
     PUT /_cluster/settings
     {
       "persistent": {
         "flowcontrol.break.enabled": true
       }
     }
     ```
   - Disable one-click traffic blocking
     ```
     PUT /_cluster/settings
     {
       "persistent": {
         "flowcontrol.break.enabled": false
       }
     }
     ```

7.9.3.1.6 Access Statistics and Traffic Control Information Query

Flow control can be implemented via an independent API.

Procedure

1. Log in to the CSS management console.
2. Choose **Clusters** in the navigation pane. On the **Clusters** page, locate the target cluster and click **Access Kibana** in the **Operation** column.
3. In the navigation pane on the left, choose **Dev Tools** and run the commands to query traffic control information.
   - Check the traffic control status of all nodes.
     ```
     GET /_nodes/stats/filter/v2
     ```
   - View traffic control details of all nodes.
     ```
     GET /_nodes/stats/filter/v2?detail
     ```
   - View the traffic control status of a specific node.
     ```
     GET /_nodes/{nodeId}/stats/filter/v2
     ```
     {nodeId} indicates the ID of the node you want to check.

     Example response:
     ```
     {
       "_nodes" : {
         "total" : 1,
         "successful" : 1,
         "failed" : 0
       },
       "cluster_name" : "css-xxxx",
       "nodes" : {
         "d3qnVIpPTtSoadkV0LQEkA" : {
     ```

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

```
    "name" : "css-xxxx-ess-esn-1-1",
    "host" : "192.168.x.x",
    "timestamp" : 1672236425112,
    "flow_control" : {
      "http" : {
        "current_connect" : 52,
        "rejected_concurrent" : 0,
        "rejected_rate" : 0,
        "rejected_black" : 0,
        "rejected_breaker" : 0
      },
      "access_items" : [
        {
          "remote_address" : "10.0.0.x",
          "search_count" : 0,
          "bulk_count" : 0,
          "other_count" : 4
        }
      ],
      "holding_requests" : 0
    }
  }
 }
}
```

Table 7-64 Response parameters

Parameter	Description
current_connect	Number of HTTP connections of a node, which is recorded even if flow control is disabled. This value is equal to the **current_open** value of **GET /_nodes/stats/http** API. It includes the current client connections of nodes.
rejected_concurrent	Number of concurrent connections rejected during HTTP flow control. This value is not cleared when HTTP flow control is disabled.
rejected_rate	Number of new connections rejected during HTTP flow control. This value is not cleared when HTTP flow control is disabled.
rejected_black	Number of requests rejected based on the blacklist during HTTP flow control. This value is not cleared when HTTP flow control is disabled.
rejected_breaker	Number of rejected new connections after one-click traffic blocking is enabled.
remote_address	IP addresses and the number of requests.
search_count	Number of times that a client accessed a database using **_search** and **_msearch**.
bulk_count	Number of times that a client accessed a database using **_bulk**.
other_count	Number of times that a client accessed a database using other requests.

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

7.9.3.1.7 Temporary Access Statistics Logs

Context

You can check access logs in either of the following ways:

- Enable and check access logs via an independent API. Configure the API parameters to record the access log time and size. The access log content is returned through a REST API.
- Print access logs. Your access logs are printed as files in backend logs. To enable this method, set the **flowcontrol.log.file.enabled** configuration item in **16.5.4**.

The following table describes access log parameters.

Table 7-65 Access log parameters

Parameter	Type	Description
duration_limit	String	Duration recorded in an access log. Value range: 10 to 120 Unit: s Default value: **30**
capacity_limit	String	Size of an access log. After access logging is enabled, the size of recorded requests is checked. If the size exceeds the value of this parameter, the access logging stops. Value range: 1 to 5 Unit: MB Default value: **1**

NOTE

Access logging stops if either **duration_limit** or **capacity_limit** reaches the threshold.

Procedure

1. Log in to the CSS management console.
2. Choose **Clusters** in the navigation pane. On the **Clusters** page, locate the target cluster and click **Access Kibana** in the **Operation** column.
3. In the navigation pane on the left, choose **Dev Tools** and run commands to enable or disable access logs.
   - Enable access logs for all nodes in a cluster.
     ```
     PUT /_access_log?duration_limit=30s&capacity_limit=1mb
     ```
   - Enable access logs for a node in a cluster.
     ```
     PUT /_access_log/{nodeId}?duration_limit=30s&capacity_limit=1mb
     ```
     {nodeId} indicates the ID of the node where you want to enable access logs.

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

4. View access logs.
   - Check the access logs of all nodes in a cluster.
     ```
     GET /_access_log
     ```
   - Check the access logs of a node in a cluster.
     ```
     GET /_access_log/{nodeId}
     ```
     {nodeId} indicates the ID of the node where you want to enable access logs.

     Example response:
     ```
     {
       "_nodes" : {
         "total" : 1,
         "successful" : 1,
         "failed" : 0
       },
       "cluster_name" : "css-flowcontroller",
       "nodes" : {
         "8x-ZHu-wTemBQwpcGivFKg" : {
           "name" : "css-flowcontroller-ess-esn-1-1",
           "host" : "10.0.0.98",
           "count" : 2,
           "access" : [
             {
               "time" : "2021-02-23 02:09:50",
               "remote_address" : "/10.0.0.98:28191",
               "url" : "/_access/security/log?pretty",
               "method" : "GET",
               "content" : ""
             },
             {
               "time" : "2021-02-23 02:09:52",
               "remote_address" : "/10.0.0.98:28193",
               "url" : "/_access/security/log?pretty",
               "method" : "GET",
               "content" : ""
             }
           ]
         }
       }
     }
     ```

Table 7-66 Response parameters

Parameter	Description
name	Node name
host	Node IP address
count	Number of node access requests in a statistical period
access	Details about node access requests in a statistical period. For details, see **Table 7-67**.

Table 7-67 access

Parameter	Description
time	Request time

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

Parameter	Description
remote_address	Source IP address and port number of the request
url	Original URL of the request
method	Method corresponding to the request path
content	Request content

5. Run the following commands to delete access logs.
   - Delete access logs of all nodes in a cluster.
     ```
     DELETE /_access_log
     ```
   - Delete access logs of a specified node in a cluster.
     ```
     DELETE /_access_log/{nodeId}
     ```
     {nodeId} indicates the ID of the node where you want to enable access logs.

7.9.3.2 Flow Control 1.0

7.9.3.2.1 Context

Feature Description

CSS can control traffic at the node level. You can configure the blacklist and whitelist, the maximum concurrent HTTP connections, and the maximum HTTP connections for a node. You can also configure the maximum heap memory used by specific request paths, the maximum CPU usage, and block access in one click, and collect statistics on node access IP addresses and URIs. Each function has an independent control switch, which is disabled by default. To restore default values of parameters, set them to **null**.

If flow control is enabled, requests will be blocked at the entry, which relieves the cluster pressure in high-concurrency scenario and avoids unavailability issues.

- **HTTP/HTTPS flow control:**
  - You can control client IP address access by setting IP addresses and subnets in HTTP/HTTPS blacklist or whitelist. If an IP address is in the blacklist, the client is disconnected and all its request are rejected. Whitelist rules take precedence over blacklist rules. If a client IP address exists in both the blacklist and whitelist, the client request will not be rejected.
  - HTTP/HTTPS concurrent connection flow control limits the total number of HTTP connections to a node per second.
  - HTTP/HTTPS new connection flow control limits the number of new connections to a node.
- Memory flow control limits request paths based on the node heap memory. You can configure memory flow control whitelist, global memory usage threshold, and heap memory threshold for a single path. Global memory flow control threshold takes precedence over the memory threshold of a single path. Paths in the whitelist will not be blocked in memory flow control.

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

- You can configure the global path whitelist for flow control as required when you need to use custom plug-ins.
- Request sampling can record the number of access requests from client IP addresses and the request paths of sampled users. Based on the statistics, you can identify the access traffic of client IP addresses and analyze the access traffic of request paths.
- Flow control provides an independent API for viewing traffic statistics and records the number of times the API is triggered. You can evaluate the flow control threshold and analyze the cluster load based on the statistics.
- Access logs record the URLs and bodies of HTTP/HTTPS requests received by nodes within a period of time. You can analyze the current traffic pressure based on the access logs.
- You can configure the node CPU usage threshold to limit the accessed traffic on a single node.
- One-click access blocking can block all the access traffic of a node, excluding the traffic from Kibana and Elasticsearch monitor APIs.

Constraints

- Currently, only versions 7.6.2 and 7.10.2 support the flow control feature.
- Flow control may affect the performance of some nodes.
- If flow control is enabled, user requests that exceed the flow control threshold will be rejected.
- Memory flow control and CPU flow control are based on request paths. The length and number of paths cannot be too large, or the cluster performance will be affected.

7.9.3.2.2 HTTP/HTTPS Flow Control

Context

You can run commands in Kibana to enable or disable HTTP/HTTPS flow control for your cluster. The command parameters are as follows.

Table 7-68 HTTP/HTTPS flow control parameters

Parameter	Type	Description
flowcontrol.http.enabled	Boolean	Whether to enable HTTP/HTTPS flow control. This function is disabled by default. Enabling it may affect node access performance. Value: **true** or **false** Default value: **false**

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

Parameter	Type	Description
flowcontrol.http.allow	List<String>	IP address whitelist. It can contain multiple IP addresses and masks, or an IP address list. Use commas (,) to separate multiple values. Example: *xx.xx.xx.xx***/24,***xx.xx.xx.xx***/24**, or *xx.xx.xx.xx.xx***,***xx.xx.xx*. The default value is null.
flowcontrol.http.deny	List<String>	IP address blacklist. Multiple IP addresses and masks or an IP address list can be configured. Use commas (,) to separate multiple IP addresses and masks. The default value is null.
flowcontrol.http.concurrent	Integer	Maximum concurrent HTTP/HTTPS connections. Default value: Number of available cores on a node x 400
flowcontrol.http.newconnect	Integer	Maximum new connections that can be created for HTTP/HTTPS requests per second. Default value: Number of available cores on a node x 200
flowcontrol.http.warmup_period	Integer	Time required for the HTTP/HTTPS connection setup speed to reach the maximum. If **flowcontrol.http.newconnect** is set to **100** and **flowcontrol.http.warmup_period** is set to **5000ms**, it indicates the system can set up 100 connections per second in 5 seconds. Value range: 0–10000 Unit: ms Default value: **0**

Procedure

1. Log in to the CSS management console.
2. Choose **Clusters** in the navigation pane. On the **Clusters** page, locate the target cluster and click **Access Kibana** in the **Operation** column.
3. In the navigation pane on the left, choose **Dev Tools** and run commands to enable or disable HTTP/HTTPS flow control.

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

- Enabling HTTP/HTTPS flow control for a node

```
PUT /_cluster/settings
{
  "persistent": {
    "flowcontrol.http.enabled": true,
    "flowcontrol.http.allow": ["192.168.0.1/24", "192.168.2.1/24"],
    "flowcontrol.http.deny": "192.168.1.1/24",
    "flowcontrol.http.concurrent": 1000,
    "flowcontrol.http.newconnect": 1000,
    "flowcontrol.http.warmup_period": 0
  }
}
```

NOTE

If all parameters are set to **null**, they will be restored to default values.

- Disabling HTTP/HTTPS flow control for a node

```
PUT /_cluster/settings
{
  "persistent": {
    "flowcontrol.http.enabled": false
  }
}
```

7.9.3.2.3 Memory Flow Control

Context

Elasticsearch provides a circuit breaker, which will terminate requests if the memory usage exceeds its threshold. However, Elasticsearch does not check the heap memory usage when an API is called, and does not allow users to configure the threshold for a single request. In this case, memory usage can only be calculated during request processing, which may lead to frequent circuit breaking and cannot avoid heap memory waste. To solve this problem, CSS checks the heap memory usage when receiving REST requests, blocking excess API requests and protecting nodes. You can configure global memory flow control, or configure the request path and heap memory threshold for a specific request path. Before a request is processed, the system checks the configured heap memory threshold. If the threshold is exceeded, the request path will be blocked.

NOTE

- Memory flow control may affect request processing performance.
- If the memory flow control is enabled, some Kibana search requests may fail.
- If memory flow control is enabled in Elasticsearch 5.5.1, _mget requests will be blocked and Kibana access will be abnormal. You can add _mget requests to the request whitelist to avoid this problem.

The following table describes memory flow control parameters.

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

Table 7-69 Memory flow control parameters

Parameter	Type	Description
flowcontrol.memory.enabled	Boolean	Whether to enable memory flow control. This function is disabled by default. Enabling memory flow control may slightly affect node performance. Value: **true** or **false** Default value: **false**
flowcontrol.memory.allow_path	List<String>	Request path whitelist for memory flow control. Whitelisted paths are blocked in memory flow control. Wildcard characters are supported. By default, query APIs controlled by the cluster are not blocked in memory flow control. This prevents the failure to query cluster information when the memory usage reaches the threshold. Example: ● "flowcontrol.memory.allow_path": "/index/_search", ● "flowcontrol.memory.allow_path": "/index*/_search", ● "flowcontrol.memory.allow_path": ["/index/_search", "/index1/_bulk"], A maximum of 10 paths can be configured. A path can contain up to 32 characters. The default value is null.

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

Parameter	Type	Description
flowcontrol.memory.heap_limit	String	Maximum global heap memory usage of a node. The value cannot be less than 10% of the heap memory. Value range: 10%–100% Default value: **90%**
flowcontrol.memory.*.filter_path	String	Paths under memory flow control. The default value is ******, indicating all paths. If **flowcontrol.memory.heap_limit** is configured and **flowcontrol.memory.*.filter_path** is not, it indicates that all the paths, except those in the whitelist, are under control. The whitelist takes precedence over the single-path rule. If a path is specified in both **flowcontrol.memory.allow_path** and **flowcontrol.memory.*.filter_path**, the requests from the path will be allowed. For example, if **flowcontrol.memory.allow_path** and **flowcontrol.memory.*.filter_path** are both set to **abc/_search**, then **abc/_search** will not be under flow control. Maximum length: **32** characters
flowcontrol.memory.*.heap_limit	String	Heap memory usage threshold of request paths. If the heap memory usage exceeds the threshold, flow control will be triggered. Value range: 0–100% Default value: **90%**

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

Procedure

1. Log in to the CSS management console.
2. Choose **Clusters** in the navigation pane. On the **Clusters** page, locate the target cluster and click **Access Kibana** in the **Operation** column.
3. In the navigation pane on the left, choose **Dev Tools** and run commands to enable or disable memory flow control.
   - Enabling memory flow control

     ```
     PUT /_cluster/settings
     {
       "persistent": {
         "flowcontrol.memory.enabled": true,
         "flowcontrol.memory.allow_path": "/index/_search",
         "flowcontrol.memory.heap_limit": "85%"
       }
     }
     ```

   - Enabling memory flow control for a request path

     Configure the heap memory usage threshold for a request path. You can configure the priorities of such threshold rules.

     ```
     PUT /_cluster/settings
     {
       "persistent": {
         "flowcontrol.memory.enabled": true,
         "flowcontrol.memory": {
           "flowcontrol_search": {
             "filter_path": "index1/_search",
             "heap_limit": "50%"
           },
           "flowcontrol_bulk": {
             "filter_path": "index*/_bulk",
             "heap_limit": "50%"
           }
         }
       }
     }
     ```

   - Deleting the memory flow control configuration of a request path

     ```
     PUT /_cluster/settings
     {
       "persistent": {
         "flowcontrol.memory.enabled": true,
         "flowcontrol.memory": {
           "flowcontrol_search": {
             "filter_path": null,
             "heap_limit": null
           }
         }
       }
     }
     ```

   - Disabling cluster memory flow control

     ```
     PUT /_cluster/settings
     {
       "persistent": {
         "flowcontrol.memory.enabled": false
       }
     }
     ```

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

7.9.3.2.4 Global Path Whitelist for Flow Control

Context

The following table describes the global path whitelist parameters for flow control.

Table 7-70 Global path whitelist parameters for flow control

Parameter	Type	Description
flowcontrol.path.white_list	List<String>	Paths that are not under flow control. These paths are not affected by memory flow control, CPU flow control, or one-click blocking; but are under IP address-based flow control. A maximum of 10 paths can be configured. A path can contain up to 32 characters. The default value is null. **NOTE** You are advised not to configure this parameter, unless required by plug-ins.

Procedure

1. Log in to the CSS management console.
2. Choose **Clusters** in the navigation pane. On the **Clusters** page, locate the target cluster and click **Access Kibana** in the **Operation** column.
3. In the navigation tree on the left, choose **Dev Tools**. Run the following command to configure the global path whitelist for flow control:

```
PUT _cluster/settings
{
  "persistent": {
    "flowcontrol.path.white_list": "xxxx"
  }
}
```

7.9.3.2.5 Request Sampling

Context

Request sampling can record the access IP addresses, the number of accessed nodes, request paths, request URLs, and request bodies, which can be used to obtain the IP addresses and paths of clients that have sent a large number of access requests.

The following table describes request sampling parameters.

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

Table 7-71 Request sampling parameters

Parameter	Type	Description
flowcontrol.statics.enabled	Boolean	Whether to enable request sampling. Request sampling may affect node performance. Value: **true** or **false** Default value: **false**
flowcontrol.statics.threshold	Integer	Number of recent access requests whose statistics are collected. The value **100** indicates that statistics will be collected on the 100 IP addresses and 100 URLs that are most frequently accessed. Minimum value: **10** Maximum value: **1000** Default value: **100**
flowcontrol.statics.sample_frequency	Integer	Path sampling frequency. If this parameter is set to **100**, samples are collected from every 100 requests. Minimum value: **50** Default value: **100**

NOTE

- The IP address statistics and URL sampling statistics are cached based on their access time. If the cache space reaches the threshold (**flowcontrol.statics.threshold**), the records of the earliest access will be deleted.
- In URL sampling, an access path is uniquely identified by its URL hash.

Procedure

1. Log in to the CSS management console.
2. Choose **Clusters** in the navigation pane. On the **Clusters** page, locate the target cluster and click **Access Kibana** in the **Operation** column.
3. In the navigation pane on the left, choose **Dev Tools** and run commands to enable or disable sampling.
   - Enabling sampling

     ```
     PUT /_cluster/settings
     {
       "persistent": {
         "flowcontrol.statics.enabled": true,
         "flowcontrol.statics.threshold": 100,
         "flowcontrol.statics.sample_frequency": 50
       }
     }
     ```

   - Disabling sampling

     ```
     PUT /_cluster/settings
     {
     ```

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

```
 "persistent": {
   "flowcontrol.statics.enabled": false
 }
}
```

7.9.3.2.6 Flow Control

Flow control can be implemented via an independent API.

1. Log in to the CSS management console.
2. Choose **Clusters** in the navigation pane. On the **Clusters** page, locate the target cluster and click **Access Kibana** in the **Operation** column.
3. In the navigation pane on the left, choose **Dev Tools** and run the commands to query traffic control information.
   - Check the traffic control status of all nodes.
     ```
     GET /_nodes/stats/filter
     ```
   - View the traffic control status of a specific node.
     ```
     GET /_nodes/{nodeId}/stats/filter
     ```
     {nodeId} indicates the ID of the node you want to check.

   Example response:

```
{
  "_nodes" : {
    "total" : 1,
    "successful" : 1,
    "failed" : 0
  },
  "cluster_name" : "css-flowcontroller",
  "nodes" : {
    "ElBRNCMbTj6L1C-Wke-Dnw" : {
      "name" : "css-flowcontroller-ess-esn-1-1",
      "host" : "10.0.0.133",
      "timestamp" : 1613979513747,
      "flow_control" : {
        "transport" : {
          "concurrent_req" : 0,
          "rejected_concurrent" : 0,
          "rejected_new" : 0,
          "rejected_deny" : 0
        },
        "http" : {
          "concurrent_req" : 0,
          "rejected_concurrent" : 0,
          "rejected_new" : 0,
          "rejected_deny" : 0
        },
        "memory" : {
          "memory_allow" : 41,
          "memory_rejected" : 0
        },
        "cpu": {
          "rejected_cpu" : 0
        }
        "ip_address" : [
          {
            "ip" : "/10.0.0.198",
            "count" : 453
          },
          {
            "ip" : "/198.19.49.1",
            "count" : 42
          }
        ],
        "url_sample" : [
```

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

```
        {
          "url" : "/*/_search?pretty=true",
          "method" : "GET",
          "remote_address" : "/10.0.0.198:16763",
          "count" : 1
        }
      ]
    }
  }
}
```

In the response, the information of each node is separated. The **http** field records the numbers of concurrent connections and new connections. The **memory** records memory flow control statistics. The **ip_address** field records the recent client IP addresses that are accessed most recently. The **url_sample** field records the recent URLs that are requested most frequently. The **cpu** field records CPU flow control statistics.

Table 7-72 Response parameters

Parameter	Description
concurrent_req	Number of TCP connections of a node, which is recorded no matter whether flow control is enabled. This value is similar to the value of **current_open** of the **GET /_nodes/stats/http** API but is smaller, because whitelisted IP addresses and internal node IP addresses are not counted.
rejected_concurrent	Number of concurrent connections rejected during HTTP flow control. This value is not cleared when HTTP flow control is disabled.
rejected_new	Number of new connections rejected during HTTP flow control. This value is not cleared when HTTP flow control is disabled.
rejected_deny	Number of requests rejected based on the blacklist during HTTP flow control. This value is not cleared when HTTP flow control is disabled.
memory_allow	Number of allowed requests during memory flow control. This parameter takes effect when memory flow control is enabled, and its value is not cleared after memory flow control is disabled. The requests from the paths in the **allow_path** whitelist are not recorded. If **allow_path** is set to ******, no requests are recorded.
memory_rejected	Number of rejected requests during memory flow control. This parameter takes effect when memory flow control is enabled, and its value is not cleared after memory flow control is disabled. The requests from the paths in the **allow_path** whitelist are not recorded. If **allow_path** is set to ******, no requests are recorded.

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

Parameter	Description
rejected_cpu	Number of requests rejected when the CPU flow control threshold is exceeded. This parameter takes effect when CPU flow control is enabled, and its value is not cleared after CPU flow control is disabled.
ip_address	IP addresses and the number of requests. For details, see **Table 7-73**.
url_sample	Request path sampling. The number of URLs of a request are collected based on the configured time and sampling interval. For details, see **Table 7-74**.

Table 7-73 ip_address

Parameter	Description
ip	Source IP address for accessing the node.
method	Number of access requests from an IP address.

Table 7-74 url_sample

Parameter	Description
url	Request URL
method	Method corresponding to the request path
remote_address	Source IP address and port number of the request
count	How many times a path is sampled

7.9.3.2.7 Access Logs

Context

You can check access logs in either of the following ways:

- Enable and check access logs via an independent API. Configure the API parameters to record the access log time and size. The access log content is returned through a REST API.
- Print access logs. Your access logs are printed as files in backend logs.

Enabling the access log function may affect cluster performance.

The following table describes access log parameters.

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

Table 7-75 Access log parameters

Parameter	Type	Description
duration_limit	String	Duration recorded in an access log. Value range: 10 to 120 Unit: s Default value: **30**
capacity_limit	String	Size of an access log. After access logging is enabled, the size of recorded requests is checked. If the size exceeds the value of this parameter, the access logging stops. Value range: 1 to 5 Unit: MB Default value: **1**

NOTE

Access logging stops if either **duration_limit** or **capacity_limit** reaches the threshold.

Procedure

1. Log in to the CSS management console.
2. Choose **Clusters** in the navigation pane. On the **Clusters** page, locate the target cluster and click **Access Kibana** in the **Operation** column.
3. In the navigation pane on the left, choose **Dev Tools** and run commands to enable or disable access logs.
   - Enabling access logs for all nodes in a cluster
     ```
     PUT /_access_log?duration_limit=30s&capacity_limit=1mb
     ```
   - Enabling access logs for a node in a cluster
     ```
     PUT /_access_log/{nodeId}?duration_limit=30s&capacity_limit=1mb
     ```
     {nodeId} indicates the ID of the node where you want to enable access logs.
4. Use APIs to check access logs.
   - API for checking the access logs of all nodes in a cluster
     ```
     GET /_access_log
     ```
   - API for checking the access logs of a node in a cluster
     ```
     GET /_access_log/{nodeId}
     ```
     {nodeId} indicates the ID of the node where you want to enable access logs.

   Example response:
   ```
   {
     "_nodes" : {
       "total" : 1,
       "successful" : 1,
       "failed" : 0
     },
     "cluster_name" : "css-flowcontroller",
     "nodes" : {
   ```

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

```
  "8x-ZHu-wTemBQwpcGivFKg" : {
    "name" : "css-flowcontroller-ess-esn-1-1",
    "host" : "10.0.0.98",
    "count" : 2,
    "access" : [
      {
        "time" : "2021-02-23 02:09:50",
        "remote_address" : "/10.0.0.98:28191",
        "url" : "/_access/security/log?pretty",
        "method" : "GET",
        "content" : ""
      },
      {
        "time" : "2021-02-23 02:09:52",
        "remote_address" : "/10.0.0.98:28193",
        "url" : "/_access/security/log?pretty",
        "method" : "GET",
        "content" : ""
      }
    ]
  }
 }
}
```


Table 7-76 Response parameters

Parameter	Description
name	Node name
host	Node IP address
count	Number of node access requests in a statistical period
access	Details about node access requests in a statistical period For details, see **Table 7-77**.

Table 7-77 access

Parameter	Description
time	Request time
remote_address	Source IP address and port number of the request
url	Original URL of the request
method	Method corresponding to the request path
content	Request content

5. Enable or disable the access log function.

   All user access operation can be logged. By default, logs are recorded in the **acces_log.log** file in the background. The maximum size of a log file is 250 MB, and there can be a maximum of five log files. You can back up access log files to OBS.

   – Enabling access logs

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

```
PUT /_cluster/settings
{
  "persistent": {
    "flowcontrol.accesslog.enabled": true
  }
}
```

- Disabling access logs

```
PUT /_cluster/settings
{
  "persistent": {
    "flowcontrol.accesslog.enabled": false
  }
}
```

7.9.3.2.8 CPU Flow Control

Context

CPU flow control can be implemented based on the CPU usage of a node.

You can configure the CPU usage threshold of a node to prevent the node from breaking down due to heavy traffic. You can determine the CPU usage threshold based on the traffic threshold. If the CPU usage of a node exceeds the configured threshold, CPU flow control discards excess node requests to protect the cluster. Traffic within the node or passing through Elasticsearch monitoring APIs are not affected.

The following table describes CPU flow control parameters.

Table 7-78 CPU flow control parameters

Parameter	Type	Description
flowcontrol.cpu.enabled	Boolean	Whether to enable CPU flow control. If this function is enabled, the node access performance may be affected. Value: **true** or **false** Default value: **false**
flowcontrol.cpu.percent_limit	Integer	Maximum CPU usage of a node. Value range: 0–100 Default value: **90**
flowcontrol.cpu.allow_path	List	Path whitelist for CPU flow control. The paths specified in the **allow_path** whitelist are not under CPU flow control. The default value is null. A path can contain up to 32 characters. A maximum of 10 request paths can be configured. Wildcard characters are supported. For example, if this parameter is set to **auto_*/_search**, all the search requests of the indexes prefixed with **auto_** are not under the flow control.

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

Parameter	Type	Description
flowcontrol.cpu.*.filter_path	String	Paths under CPU flow control. Maximum length: 32 characters Example: "flowcontrol.cpu.search.filter_path": "/index/_search", "flowcontrol.cpu.search.limit": 60, The default value is ******, indicating all paths. If **limit** is configured and **filter_path** is not, it indicates that all the paths, except those in the whitelist, are under control. The whitelist takes precedence over the single-path rule. If a path is specified in both **allow_path** and **filter_path**, the requests from the path will be allowed. For example, if both **filter_path** and **allow_path** both set to **abc/_search**, then **abc/_search** will not be under flow control.
flowcontrol.cpu.*.limit	Integer	CPU threshold of request paths. If the CPU usage exceeds the threshold, flow control will be triggered. Value range: 0–100 Default value: **90**

Procedure

1. Log in to the CSS management console.
2. Choose **Clusters** in the navigation pane. On the **Clusters** page, locate the target cluster and click **Access Kibana** in the **Operation** column.
3. In the navigation pane on the left, choose **Dev Tools** and run commands to enable or disable memory flow control.
   - Enabling CPU flow control

     ```
     PUT /_cluster/settings
     {
       "persistent": {
         "flowcontrol.cpu.enabled": true,
         "flowcontrol.cpu.percent_limit": 80,
         "flowcontrol.cpu.allow_path": ["index/_search"]
       }
     }
     ```

   - Disabling CPU flow control

     ```
     PUT /_cluster/settings
     {
       "persistent": {
         "flowcontrol.cpu.enabled": false
       }
     }
     ```

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

7.9.3.2.9 One-click Traffic Blocking

You can block all traffic in one click, except the traffic that passes through O&M APIs, to handle unexpected traffic burst and quickly recover your cluster.

1. Log in to the CSS management console.
2. Choose **Clusters** in the navigation pane. On the **Clusters** page, locate the target cluster and click **Access Kibana** in the **Operation** column.
3. In the navigation pane on the left, choose **Dev Tools** and run commands to enable or disable one-click traffic blocking.
   - Enabling one-click traffic blocking
     ```
     PUT /_cluster/settings
     {
       "persistent": {
         "flowcontrol.break.enabled": true
       }
     }
     ```
   - Disabling one-click traffic blocking
     ```
     PUT /_cluster/settings
     {
       "persistent": {
         "flowcontrol.break.enabled": false
       }
     }
     ```

7.9.4 Large Query Isolation

7.9.4.1 Context

The large query isolation feature allows you to separately manage large queries. You can isolate query requests that consume a large amount of memory or take a long period of time. If the heap memory usage of a node is too high, the interrupt control program will be triggered. The program will interrupt a large query based on the policies you configured and cancel the running query tasks of the query.

You can also configure a global query timeout duration. Long queries will be intercepted.

NOTE

Currently, only versions 7.6.2 and 7.10.2 support large query isolation.

7.9.4.2 Procedure

The large query isolation and global timeout features are disabled by default. If you enable them, the configuration will take effect immediately. Perform the following steps to configure the features:

1. Log in to the CSS management console.
2. Choose **Clusters** in the navigation pane. On the **Clusters** page, locate the target cluster, and click **Access Kibana** in the **Operation** column.
3. In the navigation pane of Kibana on the left, choose **Dev Tools**. Run the following command to enable large query isolation and global timeout features:
   ```
   PUT _cluster/settings
   {
   ```

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

```
  "persistent": {
    "search.isolator.enabled": true,
    "search.isolator.time.enabled": true
  }
}
```

The two features each has an independent switch and the following parameters.

Table 7-79 Parameters for large query isolation and global timeout duration

Switch	Parameter	Description
search.isolator.enabled	search.isolator.memory.task.limit search.isolator.time.management	Thresholds of a shard query task. A query task exceeding one of these thresholds is regarded as a large query task.
	search.isolator.memory.pool.limit search.isolator.memory.heap.limit search.isolator.count.limit	Resource usage thresholds in the isolation pool. If the resource usage of a query task exceeds one of these thresholds, the task will be intercepted. **NOTE** **search.isolator.memory.heap.limit** defines the limit on the heap memory consumed by write, query, and other operations of a node. If the limit is exceeded, large query tasks in the isolation pool will be interrupted.
	search.isolator.strategy search.isolator.strategy.ratio	Policy for selecting a query task in the isolation pool.
search.isolator.time.enabled	search.isolator.time.limit	Global timeout interval of query tasks.

4. Configure the large query isolation and global timeout duration separately.
   - Configure the thresholds of a shard query task. A query task exceeding one of these thresholds is regarded as a large query task.

     ```
     PUT _cluster/settings
     {
       "persistent": {
         "search.isolator.memory.task.limit": "50MB",
         "search.isolator.time.management": "10s"
       }
     }
     ```

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

Table 7-80 Parameter description

Parameter	Data Type	Description
search.isolator.memory.task.limit	String	Threshold of the memory requested by a query task to perform aggregation or other operations. If the requested memory exceeds the threshold, the task will be isolated and observed. Value range: **0b** to the maximum heap memory of a node Default value: **50MB** **NOTE** You can run the following command to query the current heap memory and the maximum heap memory of a cluster: **GET _cat/nodes?&h=id,ip,port,r,ramPercent,ramCurrent,heapMax,heapCurrent**
search.isolator.time.management	String	Threshold of the duration of a query. (started when cluster resources are used for query). If the duration of a query exceeds the threshold, it will be isolated and observed. Value range: ≥ **0ms** Default value: **10s**

- Configure the resource usage thresholds in the isolation pool. If the resource usage of a query task exceeds one of these thresholds, the task will be intercepted.

```
PUT _cluster/settings
{
  "persistent": {
    "search.isolator.memory.pool.limit": "50%",
    "search.isolator.memory.heap.limit": "90%",
    "search.isolator.count.limit": 1000
  }
}
```

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

Table 7-81 Parameter description

Parameter	Data Type	Description
search.isolator.memory.pool.limit	String	Threshold of the heap memory percentage of the current node. If the total memory requested by large query tasks in the isolation pool exceeds the threshold, the interrupt control program will be triggered to cancel one of the tasks. Value range: **0.0** to **100.0%** Default value: **50%**
search.isolator.memory.heap.limit	String	Heap memory threshold of the current node. If the heap memory of the node exceeds the threshold, the interrupt control program will be triggered to cancel a large query task in the isolation pool. Value range: **0.0** to **100.0%** Default value: **90%**
search.isolator.count.limit	Integer	Threshold of the number of large query tasks in the current node isolation pool. If the number of observed query tasks exceeds the threshold, the interrupt control program will be triggered to stop accepting new large queries. New large query requests will be directly canceled. Value range: **10**–**50000** Default value: **1000**

NOTE

In addition to **search.isolator.memory.pool.limit** and **search.isolator.count.limit** parameters, you can configure **search.isolator.memory.task.limit** and **search.isolator.time.management** to control the number of query tasks that enter the isolation pool.

- Policy for selecting a query task in the isolation pool.

```
PUT _cluster/settings
{
  "persistent": {
    "search.isolator.strategy": "fair",
    "search.isolator.strategy.ratio": "0.5%"
  }
}
```

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

Parameter	Data Type	Description
search.isolator.strategy	String	Policy for selecting large queries when the interrupt control program is triggered. The selected query will be interrupted. **NOTE** The large query isolation pool is checked every second until the heap memory is within the safe range. Values: **fair**, **mem-first**, or **time-first** ● **mem-first**: The query task that uses the most heap memory in the isolation pool is interrupted. ● **time-first**: The query task that has been running for the longest time in the isolation pool is interrupted. ● **fair**: If the difference between the heap memory of shard queries is smaller than *Maximum_heap_memory* x **search.isolator.strategy.ratio**, the query that takes the longest time should be interrupted. Otherwise, the query that uses the most heap memory is interrupted. Default value: **fair**
search.isolator.strategy.ratio	String	Threshold of the **fair** policy. This parameter takes effect only if **search.isolator.strategy** is set to **fair**. If the difference between the memory usage of large query tasks does not exceed the threshold, the query that takes the longest time should be interrupted. If the difference between the memory usage of large query tasks exceeds the threshold, the query that uses the most memory is interrupted. Value range: **0.0** to **100.0%** Default value: **1%**

– Configure the global timeout duration of query tasks.

```
PUT _cluster/settings
{
  "persistent": {
    "search.isolator.time.limit": "120s"
  }
}
```

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

Parameter	Data Type	Description
search.isolator.time.limit	String	Global query timeout duration. If this function is enabled, all the query tasks that exceed the specified duration will be canceled. Value range: ≥ **0ms** Default value: **120s**

7.9.5 Index Monitoring

7.9.5.1 Context

CSS monitors various metrics of the running status and change trend of cluster indexes to measure service usage and handle potential risks in a timely manner, ensuring that clusters can run stably.

During index monitoring, the **stats** information about indexes is collected and saved to the monitoring index (**monitoring-eye-css-***[yyyy-mm-dd]*) of the cluster, and retained for one week by default.

Currently, only clusters of the version 7.6.2 and 7.10.2 support index monitoring.

7.9.5.2 Enabling Index Monitoring

1. Log in to the CSS management console.
2. Choose **Clusters** in the navigation pane. On the **Clusters** page, locate the target cluster and click **Access Kibana** in the **Operation** column.
3. Choose **Dev Tools** in the navigation pane on the left and run the following command to enable index monitoring:

```
PUT _cluster/settings
{
  "persistent": {
    "css.monitoring.index.enabled": "true"
  }
}
```

4. (Optional) To monitor a specific index, run the following command on the **Dev Tools** page of Kibana:

```
PUT _cluster/settings
{
  "persistent": {
    "css.monitoring.index.enabled": "true",
    "css.monitoring.index.interval": "30s",
    "css.monitoring.index.indices": ["index_name"],
    "css.monitoring.history.duration": "3d"
  }
}
```

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

Table 7-82 Parameter description

Parameter	Data Type	Description
css.monitoring.index.enabled	Boolean	Whether to enable index monitoring. If this parameter is set to **true**, the monitoring will be enabled. Default value: **false**
css.monitoring.index.interval	Time	Interval for collecting index monitoring data. Minimum value: **1s** Default value: **10s**
css.monitoring.index.indices	String	Name of an index to be monitored. By default, all indexes are monitored. You can configure specific indexes or a type of indexes to monitor. Example: • **""css.monitoring.index.indices": ["***index_name***"]"** indicates only *index_name* is monitored. • **"css.monitoring.index.indices": ["log_*"]** indicates that only indexes starting with **log_** are monitored. • **"css.monitoring.index.indices": ["index1", "index2"]** indicates that **index1** and **index2** are monitored. Default value: ***** (indicating that all indexes are monitored)
css.monitoring.history.duration	Time	Retention period of monitoring data storage. The default period is a week. Minimum value: **1d** Default value: **7d**

NOTICE

Indexes starting with **monitoring-eye-css-*** are regarded as monitoring indexes and will not be monitored.

7.9.5.3 Checking the Index Read and Write Traffic

You can call an API to query the index read and write traffic within a period of time.

Prerequisites

A cluster has been created and **index monitoring** has been enabled.

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

Procedure

1. Log in to the CSS management console.
2. Choose **Clusters** in the navigation pane. On the **Clusters** page, locate the target cluster, and click **Access Kibana** in the **Operation** column.
3. Choose **Dev Tools** in the navigation pane on the left and run the following commands to query the index read and write traffic:
   - Check read and write traffic of all the indexes.
     ```
     GET  /_cat/monitoring
     ```
   - Check read and write traffic of a specific index.
     ```
     GET  /_cat/monitoring/{indexName}
     ```
     {indexName} indicates the name of the index whose read and write traffic you want to check.
   - Check the read and write traffic of indexes for different periods.
     ```
     GET _cat/monitoring?begin=1650099461000
     GET _cat/monitoring?begin=2022-04-16T08:57:41
     GET _cat/monitoring?begin=2022-04-16T08:57:41&end=2022-04-17T08:57:41
     ```

     Table 7-83 Parameter description

Parameter	Mandatory	Description
begin	No	Start time (UTC time) of the monitoring data you want to view. Time format: strict_date_optional_time|epoch_millis The default start time is five minutes before the current time.
end	No	End time (UTC time) of the monitoring data you want to view. Time format: strict_date_optional_time|epoch_millis The default end time is the current time.

     NOTE

     These parameters cannot be used for system indexes, whose names start with a dot (.).

     Information similar to the following is displayed:

     ```
     index  begin                  end                    status pri rep init unassign docs.count docs.deleted store.size
     pri.store.size delete.rate indexing.rate search.rate
     test 2022-03-25T09:46:53.765Z 2022-03-25T09:51:43.767Z yellow  1   1  0    1     9          0
     5.9kb          5.9kb          0/s           0/s           0/s
     ```

     Table 7-84 Parameters in the returned information

Parameter	Description
index	Index name

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

Parameter	Description
begin	Start time of the monitoring data you queried.
end	End time of the monitoring data you queried.
status	Index status within the queried monitoring interval.
pri	The number of index shards within the queried monitoring interval.
rep	The number of index replicas within the queried monitoring interval.
init	The number of initialized indexes within the queried monitoring interval.
unassign	The number of unallocated indexes within the queried monitoring interval.
docs.count	The number of documents within the queried monitoring interval.
docs.deleted	The number of deleted documents within the queried monitoring interval.
store.size	Index storage size within the queried monitoring interval.
pri.store.size	Size of the primary index shard within the queried monitoring interval.
delete.rate	Number of indexes deleted per second within the queried monitoring interval.
indexing.rate	Number of indexes wrote per second within the queried monitoring interval.
search.rate	Number of indexes queried per second within the queried monitoring interval.

7.9.5.4 Checking Index Monitoring Information

You can check preconfigured index monitoring visualizations on the **Dashboard** and **Visualizations** pages of Kibana. You can also customize tables and charts.

Prerequisites

A cluster has been created and **index monitoring** has been enabled.

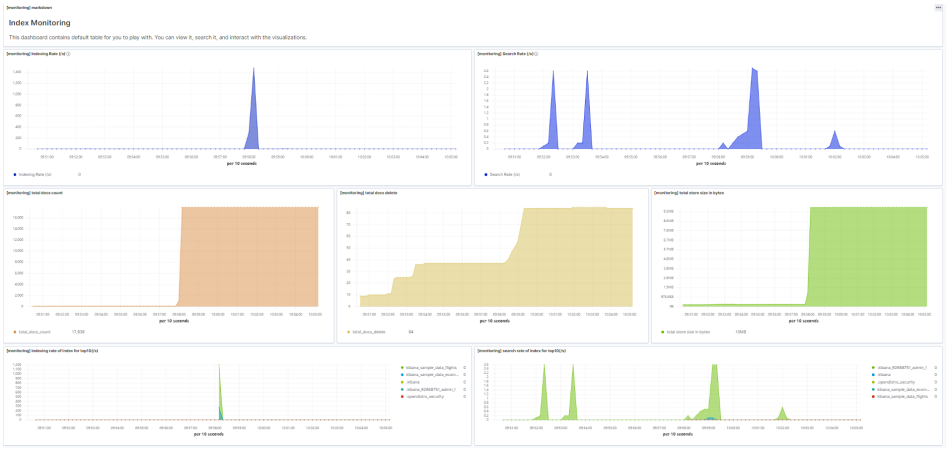
Checking Dashboard Charts

1. Log in to the CSS management console.
2. Choose **Clusters** in the navigation pane. On the **Clusters** page, locate the target cluster, and click **Access Kibana** in the **Operation** column.

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

3. In the navigation tree on the left, click **Dashboard**.
4. Click **[Monitoring] Index Monitoring Dashboard** to view the preconfigured dashboard.

Figure 7-25 Preconfigured dashboard charts



The preconfigured dashboard displays the number of read and write operations per second in the cluster and the top 10 indexes with the most read and write operations per second.

Table 7-85 Preconfigured charts

Chart Name	Description
[monitoring] markdown	Markdown chart, which briefly describes the dashboard content.
[monitoring] Indexing Rate (/s)	Number of documents written to a cluster per second.
[monitoring] Search Rate (/s)	Average number of queries per second in a cluster.
[monitoring] indexing rate of index for top10	Top 10 indexes with the most documents written per second.
[monitoring] search rate of index for top10	Top 10 indexes with the most queries per second.
[monitoring] total docs count	Total number of documents in a cluster.
[monitoring] total docs delete	Total number of deleted documents in a cluster.
[monitoring] total store size in bytes	Total storage occupied by documents in a cluster.
[monitoring] indices store_size for top10	Top 10 indexes that occupy the largest storage space.

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

Chart Name	Description
[monitoring] indices docs_count for top10	Top 10 indexes with the largest number of documents.
[monitoring] indexing time in millis of index for top10(ms)	Top 10 indexes with the longest document write latency in a unit time (ms).
[monitoring] search query time in millis of index for top10(ms)	Top 10 indexes with the longest index query time in a unit time (ms).
[monitoring] segment count of index for top10	Top 10 indexes with the largest number of index segments.
[monitoring] segment memory in bytes of index for top10	Top 10 indexes with the largest heap memory usage of index segments.

NOTICE

The index pattern of **monitoring-eye-css-*** cannot be deleted during index monitoring. Otherwise, the monitoring chart will be abnormal.

Customizing Visualizations Charts

The index monitoring module periodically stores the index/stats information in the **monitoring-eys-css** index. You can use the Kibana chart function to draw customized charts.

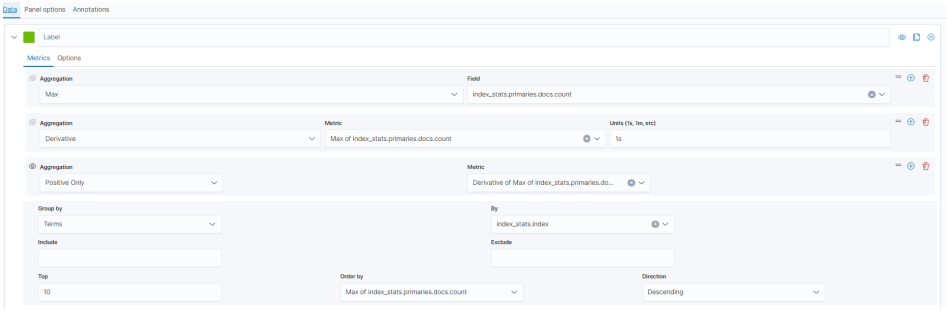
The following procedure describes how to check the trend of the document quantity in a chart as an example.

1. Log in to the CSS management console.
2. Choose **Clusters** in the navigation pane. On the **Clusters** page, locate the target cluster and click **Access Kibana** in the **Operation** column.
3. Choose **Visualize**.
4. Click **Create visualization** and select **TSVB**.
5. Set chart parameters and view the visualizations.

   On the **Data** tab page, **index_stats.primaries.docs.count** indicates the number of documents in the primary shard. **Derivative** indicates the difference between aggregation buckets. Set **Unit** to **1s**, visualizing network rates as "per second". Select **Positive only** to prevent negative numbers after resetting. To sort statistics by index, set **Group by** to **Terms** and **By** to **index_stats.index**. Statistics will be grouped by index name.

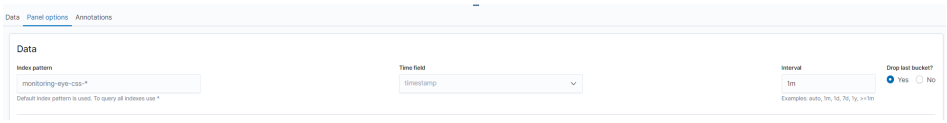
 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

Figure 7-26 TSVB page



To view data in different time segments, set the aggregation interval, or the displayed data will be incomplete. On the **Panel options** tab page, set **Interval** to **1m** or **30m** to adjust the interval of **timestamp**.

Figure 7-27 Setting the interval



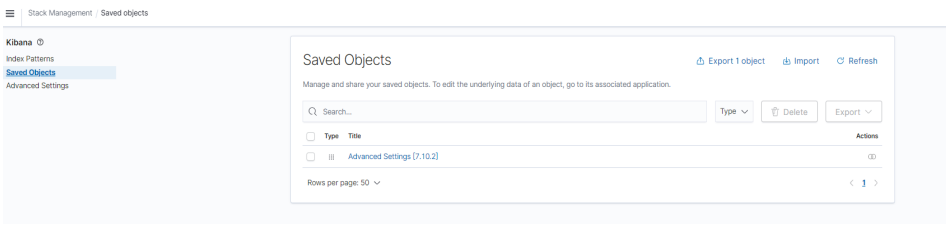
Importing Index Monitoring Charts

You can import or export charts on Kibana. If the index monitoring charts are not displayed, you can import the charts to Kibana again to load the monitoring view.

The following describes how to import a chart to Kibana:

1. Create the **monitoring-kibana.ndjson** file by referring to **kibana-monitor**.
2. Log in to Kibana and choose **Management** > **Stack Management** > **Saved objects**.

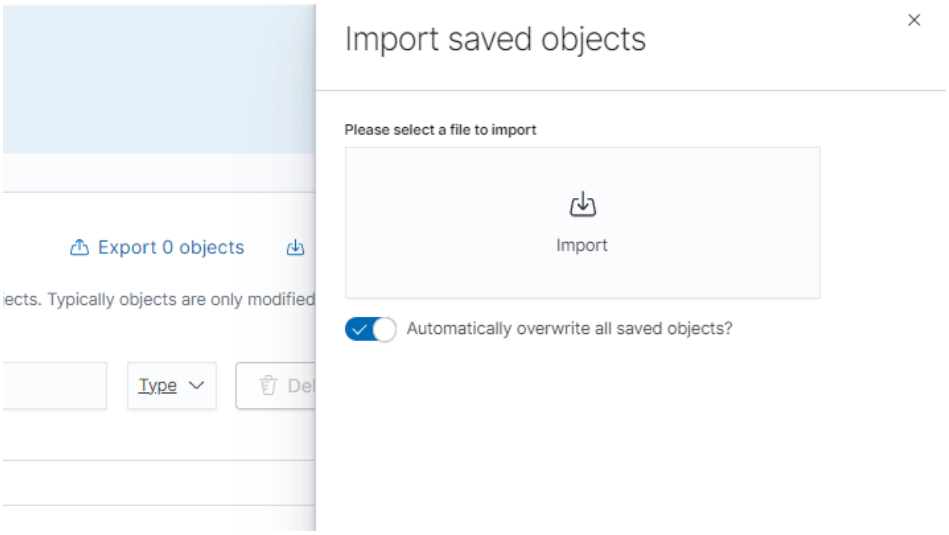
Figure 7-28 Selecting saved objects



3. Click **Import** and upload the **monitoring-kibana.ndjson** file created in step 1.

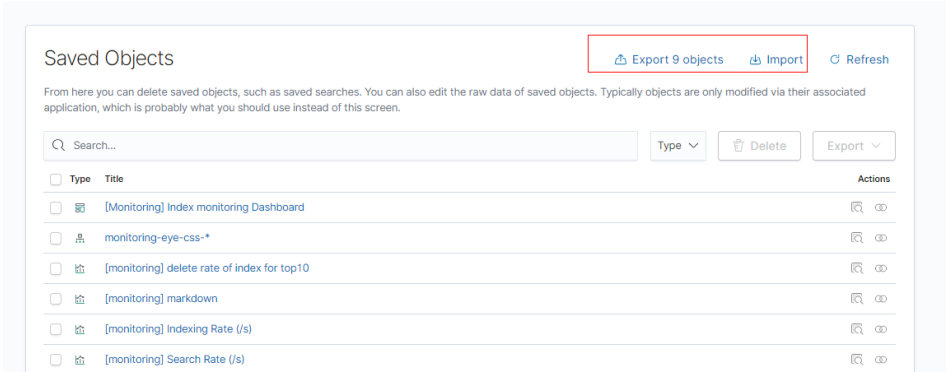
 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

Figure 7-29 Uploading a file



4. After the upload is complete, click **Done**. The index monitoring chart is successfully imported.

Figure 7-30 Successfully importing index monitoring charts



7.9.5.5 kibana-monitor

The content of the kibana-monitor configuration file is as follows:

You are advised to save the file as **monitoring-kibana.ndjson**.

{"attributes":{"description":"","kibanaSavedObjectMeta":{"searchSourceJSON":"{}"},"title":"[monitoring] segment memory in bytes of index for top10","uiStateJSON":"{}","version":1,"visState":"{\"title\":\"[monitoring] segment memory in bytes of index for top10\",\"type\":\"metrics\",\"aggs\":[],\"params\":{\"id\":\"61ca57f0-469d-11e7-af02-69e470af7417\",\"type\":\"timeseries\",\"series\":[{\"id\":\"61ca57f1-469d-11e7-af02-69e470af7417\",\"color\":\"#68BC00\",\"split_mode\":\"terms\",\"split_color_mode\":\"kibana\",\"metrics\":[{\"id\":\"61ca57f2-469d-11e7-af02-69e470af7417\",\"type\":\"max\",\"field\":\"index_stats.total.segments.memory_in_bytes\"}],\"separate_axis\":0,\"axis_position\":\"right\",\"formatter\":\"bytes\",\"chart_type\":\"line\",\"line_width\":1,\"point_size\":1,\"fill\":0.5,\"stacked\":\"none\",\"label\":\"segments memory in bytes \",\"type\":\"timeseries\",\"terms_field\":\"index_stats.index\",\"terms_order_by\":\"61ca57f2-469d-11e7-af02-69e470af7417\"}],\"time_field\":\"timestamp\",\"index_pattern\":\"monitoring-eye-css-*\",\"interval\":\"\",\"axis_position\":\"left\",\"axis_formatter\":\"number\",\"axis_scale\":\"normal\",\"show_legend\":1,\"show_grid\":1,\"tooltip_mode\":\"show_all\",\"default_index_pattern\":\"monitoring-eye-css-*\",\"default_timefield\":\"timestamp\",\"isModelInvalid\":false}}"},"id":"3ae5d820-6628-11ed-8cd7-973626cf6f70","references":[],"type":"visualization","updated_at":"2022-12-01T12:41:01.165Z","version":"WzIwNiwyXQ=="}
{"attributes":{"description":"","kibanaSavedObjectMeta":{"searchSourceJSON":"{}"},"title":"[monitoring] segment count of index for top10","uiStateJSON":"{}","version":1,"visState":"{\"aggs\":[],\"params\":{\"axis_formatter\":\"number\",\"axis_position\":\"left\",\"axis_scale\":\"normal\",\"default_index_pattern

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

\":\"monitoring-eye-css-*\",\"default_timefield\":\"timestamp\",\"filter\":{\"language\":\"kuery\",\"query
\":\"\"},\"id\":\"61ca57f0-469d-11e7-af02-69e470af7417\",\"index_pattern\":\"monitoring-eye-css-*
\",\"interval\":\"\",\"isModelInvalid\":false,\"series\":[{\"axis_position\":\"right\",\"chart_type\":\"line\",\"color
\":\"rgba(231,102,76,1)\",\"fill\":0.5,\"formatter\":\"number\",\"id\":\"61ca57f1-469d-11e7-
af02-69e470af7417\",\"label\":\"segment count of index for top10\",\"line_width\":1,\"metrics\":[{\"field
\":\"index_stats.total.segments.count\",\"id\":\"61ca57f2-469d-11e7-af02-69e470af7417\",\"type\":\"max
\"}],\"point_size\":1,\"separate_axis\":0,\"split_color_mode\":\"kibana\",\"split_mode\":\"terms\",\"stacked
\":\"none\",\"terms_field\":\"index_stats.index\",\"terms_order_by\":\"61ca57f2-469d-11e7-
af02-69e470af7417\",\"type\":\"timeseries\"}],\"show_grid\":1,\"show_legend\":1,\"time_field\":\"timestamp
\",\"tooltip_mode\":\"show_all\",\"type\":\"timeseries\"},\"title\":\"[monitoring] segment count of index for
top10\",\"type\":\"metrics\"}"},"id":"45d571c0-6626-11ed-8cd7-973626cf6f70","references":
[],"type":"visualization","updated_at":"2022-12-01T12:41:01.165Z","version":"WzIwNywyXQ=="}
{"attributes":{"description":"","kibanaSavedObjectMeta":{"searchSourceJSON":"{}"},"title":"[monitoring]
markdown","uiStateJSON":"{}","version":1,"visState":"{\"title\":\"[monitoring] markdown\",\"type
\":\"markdown\",\"params\":{\"fontSize\":12,\"openLinksInNewTab\":false,\"markdown\":\"### Index
Monitoring \\nThis dashboard contains default table for you to play with. You can view it, search it, and
interact with the visualizations.\"},\"aggs\":[]}"},"id":"b2811c70-a5f1-11ec-9a68-ada9d754c566","references":
[],"type":"visualization","updated_at":"2022-12-01T12:41:01.165Z","version":"WzIwOCwyXQ=="}
{"attributes":{"description":"number of document being indexing for primary and replica
shards","kibanaSavedObjectMeta":{"searchSourceJSON":"{}"},"title":"[monitoring] Indexing Rate (/
s)","uiStateJSON":"{}","version":1,"visState":"{\"title\":\"[monitoring] Indexing Rate (/s)\",\"type\":\"metrics
\",\"params\":{\"id\":\"61ca57f0-469d-11e7-af02-69e470af7417\",\"type\":\"timeseries\",\"series\":[{\"id
\":\"61ca57f1-469d-11e7-af02-69e470af7417\",\"color\":\"rgba(0,32,188,1)\",\"split_mode\":\"everything
\",\"metrics\":[{\"id\":\"61ca57f2-469d-11e7-af02-69e470af7417\",\"type\":\"max\",\"field
\":\"indices_stats._all.total.indexing.index_total\"},{\"unit\":\"1s\",\"id\":\"fed72db0-a5f8-11ec-
aa10-992297d21a2e\",\"type\":\"derivative\",\"field\":\"61ca57f2-469d-11e7-af02-69e470af7417\"},{\"unit
\":\"\",\"id\":\"14b66420-a5f9-11ec-aa10-992297d21a2e\",\"type\":\"positive_only\",\"field\":\"fed72db0-
a5f8-11ec-aa10-992297d21a2e\"}],\"separate_axis\":0,\"axis_position\":\"right\",\"formatter\":\"number
\",\"chart_type\":\"line\",\"line_width\":1,\"point_size\":1,\"fill\":0.5,\"stacked\":\"none\",\"label\":\"Indexing
Rate (/s)\",\"type\":\"timeseries\",\"split_color_mode\":\"rainbow\",\"hidden\":false}],\"time_field
\":\"timestamp\",\"index_pattern\":\"monitoring-eye-css-*\",\"interval\":\"\",\"axis_position\":\"left
\",\"axis_formatter\":\"number\",\"axis_scale\":\"normal\",\"show_legend\":1,\"show_grid
\":1,\"default_index_pattern\":\"monitoring-eye-css-*\",\"default_timefield\":\"timestamp\",\"isModelInvalid
\":false,\"legend_position\":\"bottom\"},\"aggs\":[]}"},"id":"de4f8ab0-a5f8-11ec-9a68-
ada9d754c566","references":
[],"type":"visualization","updated_at":"2022-12-01T12:41:01.165Z","version":"WzIwOSwyXQ=="}
{"attributes":{"description":"number of search request being executed in primary and replica
shards","kibanaSavedObjectMeta":{"searchSourceJSON":"{}"},"title":"[monitoring] Search Rate (/
s)","uiStateJSON":"{}","version":1,"visState":"{\"title\":\"[monitoring] Search Rate (/s)\",\"type\":\"metrics
\",\"params\":{\"id\":\"61ca57f0-469d-11e7-af02-69e470af7417\",\"type\":\"timeseries\",\"series\":[{\"id
\":\"61ca57f1-469d-11e7-af02-69e470af7417\",\"color\":\"rgba(0,33,224,1)\",\"split_mode\":\"everything
\",\"metrics\":[{\"id\":\"61ca57f2-469d-11e7-af02-69e470af7417\",\"type\":\"max\",\"field
\":\"indices_stats._all.total.search.query_total\"},{\"unit\":\"1s\",\"id\":\"b1093ac0-a5f7-11ec-
aa10-992297d21a2e\",\"type\":\"derivative\",\"field\":\"61ca57f2-469d-11e7-af02-69e470af7417\"},{\"unit
\":\"\",\"id\":\"c17db930-a5f7-11ec-aa10-992297d21a2e\",\"type\":\"positive_only\",\"field\":\"b1093ac0-
a5f7-11ec-aa10-992297d21a2e\"}],\"separate_axis\":0,\"axis_position\":\"right\",\"formatter\":\"number
\",\"chart_type\":\"line\",\"line_width\":1,\"point_size\":1,\"fill\":0.5,\"stacked\":\"none\",\"split_color_mode
\":\"rainbow\",\"label\":\"Search Rate (/s)\",\"type\":\"timeseries\",\"filter\":{\"query\":\"\",\"language
\":\"kuery\"}}],\"time_field\":\"timestamp\",\"index_pattern\":\"monitoring-eye-css-*\",\"interval
\":\"\",\"axis_position\":\"left\",\"axis_formatter\":\"number\",\"axis_scale\":\"normal\",\"show_legend
\":1,\"show_grid\":1,\"default_index_pattern\":\"monitoring-eye-css-*\",\"default_timefield\":\"timestamp
\",\"isModelInvalid\":false,\"legend_position\":\"bottom\"},\"aggs\":[]}"},"id":"811df7a0-a5f8-11ec-9a68-
ada9d754c566","references":
[],"type":"visualization","updated_at":"2022-12-01T12:41:01.165Z","version":"WzIxMCwyXQ=="}
{"attributes":{"description":"","kibanaSavedObjectMeta":{"searchSourceJSON":"{}"},"title":"[monitoring]
total docs count","uiStateJSON":"{}","version":1,"visState":"{\"title\":\"[monitoring] total docs count\",\"type
\":\"metrics\",\"aggs\":[],\"params\":{\"id\":\"61ca57f0-469d-11e7-af02-69e470af7417\",\"type\":\"timeseries
\",\"series\":[{\"id\":\"61ca57f1-469d-11e7-af02-69e470af7417\",\"color
\":\"rgba(218,139,69,1)\",\"split_mode\":\"everything\",\"split_color_mode\":\"kibana\",\"metrics\":[{\"unit
\":\"\",\"id\":\"61ca57f2-469d-11e7-af02-69e470af7417\",\"type\":\"max\",\"field
\":\"indices_stats._all.total.docs.count\"}],\"separate_axis\":0,\"axis_position\":\"right\",\"formatter
\":\"number\",\"chart_type\":\"line\",\"line_width\":1,\"point_size\":1,\"fill\":0.5,\"stacked\":\"none\",\"label
\":\"total_docs_count\",\"type\":\"timeseries\"}],\"time_field\":\"timestamp\",\"index_pattern\":\"monitoring-
eye-css-*\",\"interval\":\"\",\"axis_position\":\"left\",\"axis_formatter\":\"number\",\"axis_scale\":\"normal
\",\"show_legend\":1,\"show_grid\":1,\"tooltip_mode\":\"show_all\",\"default_index_pattern\":\"monitoring-
eye-css-*\",\"default_timefield\":\"timestamp\",\"isModelInvalid\":false,\"legend_position\":\"bottom
\"}}"},"id":"eea89780-664b-11ed-8cd7-973626cf6f70","references":
[],"type":"visualization","updated_at":"2022-12-01T12:41:01.165Z","version":"WzIxMSwyXQ=="}
{"attributes":{"description":"","kibanaSavedObjectMeta":{"searchSourceJSON":"{}"},"title":"[monitoring]

total docs delete","uiStateJSON":"{}","version":1,"visState":"{\"title\":\"[monitoring] total docs delete\",\"type\":\"metrics\",\"aggs\":[],\"params\":{\"id\":\"61ca57f0-469d-11e7-af02-69e470af7417\",\"type\":\"timeseries\",\"series\":[{\"id\":\"61ca57f1-469d-11e7-af02-69e470af7417\",\"color\":\"rgba(214,191,87,1)\",\"split_mode\":\"everything\",\"split_color_mode\":\"kibana\",\"metrics\":[{\"id\":\"61ca57f2-469d-11e7-af02-69e470af7417\",\"type\":\"max\",\"field\":\"indices_stats._all.total.docs.deleted\"}],\"separate_axis\":0,\"axis_position\":\"right\",\"formatter\":\"number\",\"chart_type\":\"line\",\"line_width\":1,\"point_size\":1,\"fill\":0.5,\"stacked\":\"none\",\"label\":\"totol_docs_delete\",\"type\":\"timeseries\",\"hidden\":false}],\"time_field\":\"timestamp\",\"index_pattern\":\"monitoring-eye-css-*\",\"interval\":\"\",\"axis_position\":\"left\",\"axis_formatter\":\"number\",\"axis_scale\":\"normal\",\"show_legend\":1,\"show_grid\":1,\"tooltip_mode\":\"show_all\",\"default_index_pattern\":\"monitoring-eye-css-*\",\"default_timefield\":\"timestamp\",\"isModelInvalid\":false,\"drop_last_bucket\":1,\"legend_position\":\"bottom\"}}"},"id":"cfbb4e20-664c-11ed-8cd7-973626cf6f70","references":[],"type":"visualization","updated_at":"2022-12-01T12:41:01.165Z","version":"WzIxMiwyXQ=="}
{"attributes":{"description":"","kibanaSavedObjectMeta":{"searchSourceJSON":"{}"},"title":"[monitoring] total store size in bytes","uiStateJSON":"{}","version":1,"visState":"{\"title\":\"[monitoring] total store size in bytes\",\"type\":\"metrics\",\"aggs\":[],\"params\":{\"id\":\"61ca57f0-469d-11e7-af02-69e470af7417\",\"type\":\"timeseries\",\"series\":[{\"id\":\"61ca57f1-469d-11e7-af02-69e470af7417\",\"color\":\"#68BC00\",\"split_mode\":\"everything\",\"split_color_mode\":\"kibana\",\"metrics\":[{\"id\":\"61ca57f2-469d-11e7-af02-69e470af7417\",\"type\":\"max\",\"field\":\"indices_stats._all.total.store.size_in_bytes\"}],\"separate_axis\":0,\"axis_position\":\"right\",\"formatter\":\"bytes\",\"chart_type\":\"line\",\"line_width\":1,\"point_size\":1,\"fill\":0.5,\"stacked\":\"none\",\"label\":\"total store size in bytes\",\"type\":\"timeseries\"}],\"time_field\":\"timestamp\",\"index_pattern\":\"monitoring-eye-css-*\",\"interval\":\"\",\"axis_position\":\"left\",\"axis_formatter\":\"number\",\"axis_scale\":\"normal\",\"show_legend\":1,\"show_grid\":1,\"tooltip_mode\":\"show_all\",\"default_index_pattern\":\"monitoring-eye-css-*\",\"default_timefield\":\"timestamp\",\"isModelInvalid\":false,\"legend_position\":\"bottom\",\"background_color_rules\":[{\"id\":\"7712e550-664f-11ed-8b5d-8db37e5b4cc4\"}],\"bar_color_rules\":[{\"id\":\"77680a30-664f-11ed-8b5d-8db37e5b4cc4\"}]}}"},"id":"c7f72ae0-664e-11ed-8cd7-973626cf6f70","references":[],"type":"visualization","updated_at":"2022-12-01T12:41:01.165Z","version":"WzIxMywyXQ=="}
{"attributes":{"description":"","kibanaSavedObjectMeta":{"searchSourceJSON":"{}"},"title":"[monitoring] indexing rate of index for top10(/s)","uiStateJSON":"{}","version":1,"visState":"{\"title\":\"[monitoring] indexing rate of index for top10(/s)\",\"type\":\"metrics\",\"aggs\":[],\"params\":{\"id\":\"61ca57f0-469d-11e7-af02-69e470af7417\",\"type\":\"timeseries\",\"series\":[{\"id\":\"61ca57f1-469d-11e7-af02-69e470af7417\",\"color\":\"#68BC00\",\"split_mode\":\"terms\",\"metrics\":[{\"id\":\"61ca57f2-469d-11e7-af02-69e470af7417\",\"type\":\"max\",\"field\":\"index_stats.total.indexing.index_total\"},{\"unit\":\"1s\",\"id\":\"541ed8f0-a5ee-11ec-aa10-992297d21a2e\",\"type\":\"derivative\",\"field\":\"61ca57f2-469d-11e7-af02-69e470af7417\"},{\"unit\":\"\",\"id\":\"67ec1f50-a5ee-11ec-aa10-992297d21a2e\",\"type\":\"positive_only\",\"field\":\"541ed8f0-a5ee-11ec-aa10-992297d21a2e\"}],\"separate_axis\":0,\"axis_position\":\"right\",\"formatter\":\"number\",\"chart_type\":\"line\",\"line_width\":1,\"point_size\":1,\"fill\":0.5,\"stacked\":\"none\",\"label\":\"indexing_rate\",\"type\":\"timeseries\",\"split_filters\":[{\"color\":\"#68BC00\",\"id\":\"81004200-a5ee-11ec-aa10-992297d21a2e\",\"filter\":{\"query\":\"\",\"language\":\"kuery\"}}],\"filter\":{\"query\":\"\",\"language\":\"kuery\"},\"terms_field\":\"index_stats.index\",\"terms_order_by\":\"61ca57f2-469d-11e7-af02-69e470af7417\",\"terms_size\":\"10\",\"terms_direction\":\"desc\",\"split_color_mode\":\"rainbow\"}],\"time_field\":\"timestamp\",\"index_pattern\":\"monitoring-eye-css-*\",\"interval\":\"\",\"axis_position\":\"left\",\"axis_formatter\":\"number\",\"axis_scale\":\"normal\",\"show_legend\":1,\"show_grid\":1,\"default_index_pattern\":\"monitoring-eye-css-*\",\"default_timefield\":\"timestamp\",\"isModelInvalid\":false,\"tooltip_mode\":\"show_all\"}}"},"id":"943b3e00-a5ef-11ec-9a68-ada9d754c566","references":[],"type":"visualization","updated_at":"2022-12-01T12:41:01.165Z","version":"WzIxNCwyXQ=="}
{"attributes":{"description":"","kibanaSavedObjectMeta":{"searchSourceJSON":"{}"},"title":"[monitoring] search rate of index for top10(/s)","uiStateJSON":"{}","version":1,"visState":"{\"title\":\"[monitoring] search rate of index for top10(/s)\",\"type\":\"metrics\",\"aggs\":[],\"params\":{\"id\":\"61ca57f0-469d-11e7-af02-69e470af7417\",\"type\":\"timeseries\",\"series\":[{\"id\":\"61ca57f1-469d-11e7-af02-69e470af7417\",\"color\":\"rgba(99,157,12,1)\",\"split_mode\":\"terms\",\"metrics\":[{\"id\":\"61ca57f2-469d-11e7-af02-69e470af7417\",\"type\":\"max\",\"field\":\"index_stats.total.search.query_total\"},{\"unit\":\"1s\",\"id\":\"fdfdfad0-a5ef-11ec-aa10-992297d21a2e\",\"type\":\"derivative\",\"field\":\"61ca57f2-469d-11e7-af02-69e470af7417\"},{\"unit\":\"\",\"id\":\"0aaa26a0-a5f0-11ec-aa10-992297d21a2e\",\"type\":\"positive_only\",\"field\":\"fdfdfad0-a5ef-11ec-aa10-992297d21a2e\"}],\"separate_axis\":0,\"axis_position\":\"right\",\"formatter\":\"number\",\"chart_type\":\"line\",\"line_width\":1,\"point_size\":1,\"fill\":0.5,\"stacked\":\"none\",\"label\":\"search rate\",\"type\":\"timeseries\",\"terms_field\":\"index_stats.index\",\"terms_order_by\":\"61ca57f2-469d-11e7-af02-69e470af7417\",\"split_color_mode\":\"rainbow\"}],\"time_field\":\"timestamp\",\"index_pattern\":\"monitoring-eye-css-*\",\"interval\":\"\",\"axis_position\":\"left\",\"axis_formatter\":\"number\",\"axis_scale\":\"normal\",\"show_legend\":1,\"show_grid\":1,\"default_index_pattern\":\"monitoring-eye-css-*\",\"default_timefield\":\"timestamp\",\"isModelInvalid\":false,\"tooltip_mode\":\"show_all\"}}"},"id":"ab503550-a5ef-11ec-9a68-ada9d754c566","references":[],"type":"visualization","updated_at":"2022-12-01T12:41:01.165Z","version":"WzIxNSwyXQ=="}
{"attributes":{"description":"","kibanaSavedObjectMeta":{"searchSourceJSON":"{}"},"title":"[monitoring]

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

indices store_size for top10","uiStateJSON":"{}","version":1,"visState":"{\"title\":\"[monitoring] indices store_size for top10\",\"type\":\"metrics\",\"aggs\":[],\"params\":{\"id\":\"61ca57f0-469d-11e7-af02-69e470af7417\",\"type\":\"timeseries\",\"series\":[{\"id\":\"38474c50-a5f5-11ec-aa10-992297d21a2e\",\"color\":\"#68BC00\",\"split_mode\":\"terms\",\"metrics\":[{\"id\":\"38474c51-a5f5-11ec-aa10-992297d21a2e\",\"type\":\"max\",\"field\":\"index_stats.total.store.size_in_bytes\"}],\"separate_axis\":0,\"axis_position\":\"right\",\"formatter\":\"bytes\",\"chart_type\":\"line\",\"line_width\":1,\"point_size\":1,\"fill\":0.5,\"stacked\":\"none\",\"label\":\"store_size for index\",\"type\":\"timeseries\",\"terms_field\":\"index_stats.index\",\"terms_order_by\":\"38474c51-a5f5-11ec-aa10-992297d21a2e\",\"filter\":{\"query\":\"\",\"language\":\"kuery\"},\"split_color_mode\":\"rainbow\"}],\"time_field\":\"timestamp\",\"index_pattern\":\"monitoring-eye-css-*\",\"interval\":\"\",\"axis_position\":\"left\",\"axis_formatter\":\"number\",\"axis_scale\":\"normal\",\"show_legend\":1,\"show_grid\":1,\"default_index_pattern\":\"monitoring-eye-css-*\",\"default_timefield\":\"timestamp\",\"isModelInvalid\":false,\"filter\":{\"query\":\"\",\"language\":\"kuery\"},\"bar_color_rules\":[{\"id\":\"7d9d3cb0-a5f5-11ec-aa10-992297d21a2e\"}],\"tooltip_mode\":\"show_all\"}}","id":"c78119a0-a5f5-11ec-9a68-ada9d754c566","references":[],"type":"visualization","updated_at":"2022-12-01T12:41:01.165Z","version":"WzIxNiwyXQ=="}
{"attributes":{"description":"","kibanaSavedObjectMeta":{"searchSourceJSON":"{}"},"title":"[monitoring] search query time in millis of index for top10(ms)","uiStateJSON":"{}","version":1,"visState":"{\"title\":\"[monitoring] search query time in millis of index for top10(ms)\",\"type\":\"metrics\",\"aggs\":[],\"params\":{\"axis_formatter\":\"number\",\"axis_max\":\"\",\"axis_min\":\"\",\"axis_position\":\"left\",\"axis_scale\":\"normal\",\"default_index_pattern\":\"monitoring-eye-css-*\",\"default_timefield\":\"timestamp\",\"id\":\"61ca57f0-469d-11e7-af02-69e470af7417\",\"index_pattern\":\"monitoring-eye-css-*\",\"interval\":\"\",\"isModelInvalid\":false,\"series\":[{\"axis_position\":\"right\",\"chart_type\":\"line\",\"color\":\"#68BC00\",\"fill\":0.5,\"formatter\":\"number\",\"id\":\"61ca57f1-469d-11e7-af02-69e470af7417\",\"label\":\"index_query_time_in_millis\",\"line_width\":1,\"metrics\":[{\"field\":\"index_stats.total.search.query_time_in_millis\",\"id\":\"61ca57f2-469d-11e7-af02-69e470af7417\",\"type\":\"max\"},{\"unit\":\"1s\",\"id\":\"42c92b10-6645-11ed-925a-6de90846447d\",\"type\":\"derivative\",\"field\":\"61ca57f2-469d-11e7-af02-69e470af7417\"}],\"point_size\":1,\"separate_axis\":0,\"split_color_mode\":\"kibana\",\"split_mode\":\"terms\",\"stacked\":\"none\",\"terms_field\":\"index_stats.index\",\"terms_order_by\":\"61ca57f2-469d-11e7-af02-69e470af7417\",\"type\":\"timeseries\"}],\"show_grid\":1,\"show_legend\":1,\"time_field\":\"timestamp\",\"tooltip_mode\":\"show_all\",\"type\":\"timeseries\",\"background_color\":null,\"filter\":{\"query\":\"\",\"language\":\"kuery\"},\"legend_position\":\"right\"}}"},"id":"c8109100-6627-11ed-8cd7-973626cf6f70","references":[],"type":"visualization","updated_at":"2022-12-01T12:41:01.165Z","version":"WzIxNywyXQ=="}
{"attributes":{"description":"","hits":0,"kibanaSavedObjectMeta":{"searchSourceJSON":"{\"query\":{\"language\":\"kuery\",\"query\":\"\"},\"filter\":[]}"},"optionsJSON":"{\"hidePanelTitles\":false,\"useMargins\":true}","panelsJSON":"[{\"gridData\":{\"x\":0,\"y\":0,\"w\":48,\"h\":5,\"i\":\"971ed6c6-81b9-491b-9f08-e3ae9c382abd\"},\"panelIndex\":\"971ed6c6-81b9-491b-9f08-e3ae9c382abd\",\"embeddableConfig\":{},\"panelRefName\":\"panel_0\"},{\"gridData\":{\"x\":0,\"y\":5,\"w\":24,\"h\":15,\"i\":\"5a6982e7-0c6c-4733-8a2d-e4c57cdf7397\"},\"panelIndex\":\"5a6982e7-0c6c-4733-8a2d-e4c57cdf7397\",\"embeddableConfig\":{},\"panelRefName\":\"panel_1\"},{\"gridData\":{\"x\":24,\"y\":5,\"w\":24,\"h\":15,\"i\":\"662476f4-739c-4a05-858c-2ee8230cf410\"},\"panelIndex\":\"662476f4-739c-4a05-858c-2ee8230cf410\",\"embeddableConfig\":{},\"panelRefName\":\"panel_2\"},{\"gridData\":{\"x\":0,\"y\":20,\"w\":16,\"h\":15,\"i\":\"d89c38e2-33f3-4592-b503-20460a6a7a57\"},\"panelIndex\":\"d89c38e2-33f3-4592-b503-20460a6a7a57\",\"embeddableConfig\":{},\"panelRefName\":\"panel_3\"},{\"gridData\":{\"x\":16,\"y\":20,\"w\":16,\"h\":15,\"i\":\"1f693b49-79fa-4807-94e8-0c12f51e54f8\"},\"panelIndex\":\"1f693b49-79fa-4807-94e8-0c12f51e54f8\",\"embeddableConfig\":{},\"panelRefName\":\"panel_4\"},{\"gridData\":{\"x\":32,\"y\":20,\"w\":16,\"h\":15,\"i\":\"616b143d-74e9-4dac-98ba-5849536f0fba\"},\"panelIndex\":\"616b143d-74e9-4dac-98ba-5849536f0fba\",\"embeddableConfig\":{},\"panelRefName\":\"panel_5\"},{\"gridData\":{\"x\":0,\"y\":35,\"w\":24,\"h\":11,\"i\":\"cfa82f27-1b8d-49ba-a7b9-d8809d3b258c\"},\"panelIndex\":\"cfa82f27-1b8d-49ba-a7b9-d8809d3b258c\",\"embeddableConfig\":{},\"panelRefName\":\"panel_6\"},{\"gridData\":{\"x\":24,\"y\":35,\"w\":24,\"h\":11,\"i\":\"135d13eb-aab6-43ca-9029-7d26e91d90e3\"},\"panelIndex\":\"135d13eb-aab6-43ca-9029-7d26e91d90e3\",\"embeddableConfig\":{},\"panelRefName\":\"panel_7\"},{\"gridData\":{\"x\":0,\"y\":46,\"w\":24,\"h\":11,\"i\":\"28a77de1-9110-49e8-b273-724f880b1653\"},\"panelIndex\":\"28a77de1-9110-49e8-b273-724f880b1653\",\"embeddableConfig\":{},\"panelRefName\":\"panel_8\"},{\"gridData\":{\"x\":24,\"y\":46,\"w\":24,\"h\":11,\"i\":\"80ece867-cf23-4935-bfbc-430afa51bcca\"},\"panelIndex\":\"80ece867-cf23-4935-bfbc-430afa51bcca\",\"embeddableConfig\":{},\"panelRefName\":\"panel_9\"},{\"gridData\":{\"x\":0,\"y\":57,\"w\":24,\"h\":11,\"i\":\"2ba970aa-c9c4-491b-bdd3-c1b1ee9bc8d3\"},\"panelIndex\":\"2ba970aa-c9c4-491b-bdd3-c1b1ee9bc8d3\",\"embeddableConfig\":{},\"panelRefName\":\"panel_10\"},{\"gridData\":{\"x\":24,\"y\":57,\"w\":24,\"h\":11,\"i\":\"f2e1b6ab-ddf7-492e-aaca-9460f11aa4aa\"},\"panelIndex\":\"f2e1b6ab-ddf7-492e-aaca-9460f11aa4aa\",\"embeddableConfig\":{},\"panelRefName\":\"panel_11\"},{\"gridData\":{\"x\":0,\"y\":68,\"w\":24,\"h\":11,\"i\":\"dd14182d-d8b9-47f2-bf36-6cba3b09586c\"},\"panelIndex\":\"dd14182d-d8b9-47f2-bf36-6cba3b09586c\",\"embeddableConfig\":{},\"panelRefName\":\"panel_12\"},{\"gridData\":{\"x\":24,\"y\":68,\"w\":24,\"h\":11,\"i\":\"a47f9333-52b7-49b7-8cac-f470cf405131\"},\"panelIndex\":\"a47f9333-52b7-49b7-8cac-f470cf405131\",\"embeddableConfig\":{},\"panelRefName\":\"panel_13\"}]","timeRestore":false,"title":"[Monitoring] Index monitoring Dashboard","version":1},"id":"524eb000-a5f2-11ec-9a68-ada9d754c566","references":[{"id":"b2811c70-

a5f1-11ec-9a68-ada9d754c566","name":"panel_0","type":"visualization"},{"id":"de4f8ab0-a5f8-11ec-9a68-ada9d754c566","name":"panel_1","type":"visualization"},{"id":"811df7a0-a5f8-11ec-9a68-ada9d754c566","name":"panel_2","type":"visualization"},
{"id":"eea89780-664b-11ed-8cd7-973626cf6f70","name":"panel_3","type":"visualization"},
{"id":"cfbb4e20-664c-11ed-8cd7-973626cf6f70","name":"panel_4","type":"visualization"},
{"id":"c7f72ae0-664e-11ed-8cd7-973626cf6f70","name":"panel_5","type":"visualization"},{"id":"943b3e00-a5ef-11ec-9a68-ada9d754c566","name":"panel_6","type":"visualization"},{"id":"ab503550-a5ef-11ec-9a68-ada9d754c566","name":"panel_7","type":"visualization"},{"id":"c78119a0-a5f5-11ec-9a68-ada9d754c566","name":"panel_8","type":"visualization"},{"id":"225f6020-a5f1-11ec-9a68-ada9d754c566","name":"panel_9","type":"visualization"},
{"id":"17d49220-662a-11ed-8cd7-973626cf6f70","name":"panel_10","type":"visualization"},
{"id":"c8109100-6627-11ed-8cd7-973626cf6f70","name":"panel_11","type":"visualization"},
{"id":"45d571c0-6626-11ed-8cd7-973626cf6f70","name":"panel_12","type":"visualization"},
{"id":"3ae5d820-6628-11ed-8cd7-973626cf6f70","name":"panel_13","type":"visualization"}],"type":"dashboard","updated_at":"2022-12-01T12:41:01.165Z","version":"WzIxOCwyXQ=="}
{"exportedCount":16,"missingRefCount":0,"missingReferences":[]}

7.9.6 Enhanced Cluster Monitoring

7.9.6.1 P99 Latency Monitoring

Context

The Elasticsearch community only discusses how to monitor the average latency of search requests, which cannot reflect the actual search performance of a cluster. To enhance monitoring, CSS allows you to monitor the P99 latency of search requests in clusters.

Prerequisites

Currently, only clusters of version 7.6.2 and 7.10.2 support P99 latency monitoring.

Obtaining Monitoring Information

1. Log in to the CSS management console.
2. Choose **Clusters** in the navigation pane. On the **Clusters** page, locate the target cluster and click **Access Kibana** in the **Operation** column.
3. In the navigation tree on the left, choose **Dev Tools** and run the following command to check the P99 latency of the current cluster:

   ```
   GET /search/stats/percentile
   ```

   Example response:

   ```
   {
     "overall" : {
       "1.0" : 2.0,
       "5.0" : 2.0,
       "25.0" : 6.5,
       "50.0" : 19.5,
       "75.0" : 111.0,
       "95.0" : 169.0,
       "99.0" : 169.0,
       "max" : 169.0,
       "min" : 2.0
     },
     "last_one_day" : {
       "1.0" : 2.0,
       "5.0" : 2.0,
       "25.0" : 6.5,
       "50.0" : 19.5,
       "75.0" : 111.0,
   ```

```
    "95.0" : 169.0,
    "99.0" : 169.0,
    "max" : 169.0,
    "min" : 2.0
  },
  "latest" : {
    "1.0" : 26.0,
    "5.0" : 26.0,
    "25.0" : 26.0,
    "50.0" : 26.0,
    "75.0" : 26.0,
    "95.0" : 26.0,
    "99.0" : 26.0,
    "max" : 26.0,
    "min" : 26.0
  }
}
```

NOTE

- In the response, **overall** indicates all the statistics that have been collected since the cluster startup, **last_one_day** indicates the statistics collected in the last day, and **latest** indicates the statistics that have been collected since the last reset.
- The calculated P99 latency is an estimation. It is more precise than the P50 latency.
- The P99 latency of a cluster is cleared and recalculated if the cluster is restarted.

Other Operations

- Define percentage.

  You can run the following command to specify the percentage:

  ```
  GET /search/stats/percentile
  {
    "percents": [1, 50, 90]
  }
  ```

- Reset the **latest** statistics.

  You can run the following command to reset the **latest** statistics:

  ```
  POST /search/stats/reset
  ```

  Example response:

  ```
  {
    "nodes" : {
      "css-c9c8-ess-esn-1-1" : "ok"
    }
  }
  ```

7.9.6.2 HTTP Status Code Monitoring

Context

When an external system accesses Elasticsearch through the HTTP protocol, a response and the corresponding status code are returned. The open-source Elasticsearch server does not collect the status code, so users cannot monitor Elasticsearch APIs status or cluster request status. CSS allows you to monitor the HTTP status codes of clusters.

Prerequisites

Currently, only clusters of versions 7.6.2 and 7.10.2 support HTTP status code monitoring.

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

Obtaining Status Codes

1. Log in to the CSS management console.
2. Choose **Clusters** in the navigation pane. On the **Clusters** page, locate the target cluster and click **Access Kibana** in the **Operation** column.
3. In the navigation tree on the left, choose **Dev Tools**.
4. On the console page of **Dev Tools**, run commands based on the cluster version.
   - For clusters of version 7.6.2, run the following command to obtain the status code statistics:

     ```
     GET /_nodes/http_stats
     ```

     Example response:

     ```
     {
       "_nodes" : {
         "total" : 1,
         "successful" : 1,
         "failed" : 0    },
        "cluster_name" : "css-8362",
        "nodes" : {
         "F9IFdQPARaOJI7oL7HOXtQ" : {
           "http_code" : {
             "200" : 114,
             "201" : 5,
             "429" : 0,
             "400" : 7,
             "404" : 0,
             "405" : 0
           }
         }
       }
     }
     ```

   - For clusters of version 7.10.2, run the following command to obtain the status code statistics:

     ```
     GET _nodes/stats/http
     ```

     Example response:

     ```
     {
     // ...
       "cluster_name" : "css-2985",
       "nodes" : {
     // ...
         "omvR9_W-TsGApraMApREjA" : {

     // ...
           "http" : {
             "current_open" : 4,
             "total_opened" : 37,
             "http_code" : {
               "200" : 25,
               "201" : 7,
               "429" : 0,
               "400" : 3,
               "404" : 0,
               "405" : 0
             }
           }
         }
       }
     }
     ```

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

7.10 Intelligent O&M

7.10.1 Overview of Intelligent O&M

CSS provides intelligent O&M to detect potential cluster risks and provide risk handling suggestions.

Currently, the Elasticsearch clusters support intelligent O&M.

Intelligent O&M supports the following functions:

- **Creating a Scan Task**

  Before using the intelligent O&M function, you need to create a scan task.
- **Viewing Cluster Risk Items**

  After a scan task is started, you can view details about cluster risk items in the intelligent O&M list.
- **Deleting a Scan Task**

  After processing all risk items found in a scan task, you can delete the scan task.

7.10.2 Creating a Scan Task

If the intelligent O&M function is enabled for CSS, you need to start a scan task.

Prerequisites

A CSS cluster has been created. For details, see **Creating an Elasticsearch Cluster in Security Mode**.

Procedure

1. Log in to the CSS management console.
2. On the cluster management page, click the name of the cluster for which you want to perform intelligent O&M. The basic information page of the cluster is displayed.
3. Choose **Intelligent O&M** from the navigation pane.
4. On the Intelligent O&M page, click **Scan** in the upper left corner.
5. In the dialog box, enter the basic information about the scan task and click **OK**.

Table 7-86 Detection task information

Parameter	Description
Name	Name of a scan task.
Description	Brief description of a scan task.

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

After a scan task is created, you can view it in the intelligent O&M list.

Follow-up Operations

View cluster risks and diagnose the cluster health status. For details, see **Viewing Cluster Risk Items**.

7.10.3 Viewing Cluster Risk Items

After a scan task is started, you can view details about cluster risk items in the intelligent O&M list.

Prerequisites

A scan task has been started. For details, see **Creating a Scan Task**.

Check Items

The following items will be checked and the detected risks will be displayed in the intelligent O&M list:

- Check the current health status of the cluster. Red: Some primary shards are not allocated. Yellow: Some secondary shards are not allocated. Green: that all shards are allocated.
- Check the number of nodes in the cluster and the number of AZs to evaluate the high availability status of the distributed Elasticsearch cluster.
- Check whether index replicas are enabled. If replicas are not enabled and a fault occurs, an index may be unavailable, and the data in a cluster using local disks may be lost.
- Check for .kibana index conflicts in the cluster.
- Check disk usage. If the disk usage of a node is too high, new index shards may fail to be allocated to the node and the cluster performance may be affected.
- Check whether the storage usage of cluster data nodes or cold data nodes is balanced. Unbalanced storage distribution may result in unbalanced cluster loads and increase read/write latency.
- Check whether any node in the current cluster is disconnected or unavailable for 5 consecutive minutes.
- Check for nodes with too many shards. A large number of shards will consume too many node resources, increasing read/write latency and slowing down metadata update.
- Check the size of all shards. A large shard may affect performance deterioration, occupy too much node memory, and slow down shard restoration during scaling or fault recovery.
- Check whether the current cluster has an available new version.
- Check for snapshot creation failures and snapshot records in the cluster in the last seven days.

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

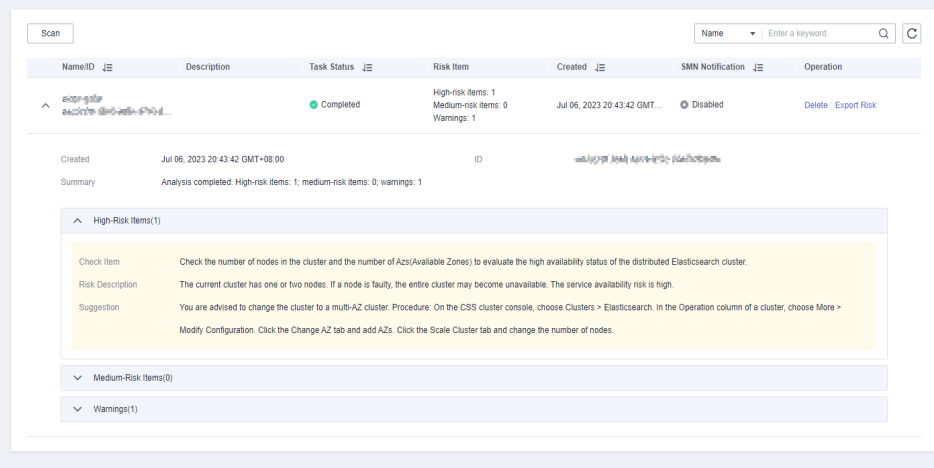
Procedure

1. Log in to the CSS management console.
2. On the cluster management page, click a cluster name to go to the basic information page of the cluster.
3. Choose **Intelligent O&M** from the navigation pane.
4. On the intelligent O&M list page, select a started scan task. Click ⌄ on the left of the task name to view its creation time, abstract, ID, and risk items.

   Click ⌄ on the left of a risk item to view its details, including the check item, risk description, and risk suggestion.

   You can handle cluster risks in a timely manner based on the suggestions.

Figure 7-31 Risk items



7.10.4 Deleting a Scan Task

After processing all risk items found in a scan task, you can delete the scan task. After a scan task is deleted, the system deletes all diagnosis information corresponding to the scan task.

Prerequisites

A scan task has been started. For details, see **Creating a Scan Task**.

Procedure

1. Log in to the CSS management console.
2. On the cluster management page, click a cluster name to go to the basic information page of the cluster.
3. Choose **Intelligent O&M** from the navigation pane.
4. Locate a scan task you want to delete and click **Delete** in the **Operation** column.
5. In the dialog box, click **OK**.

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

8 Importing Data to Elasticsearch

8.1 Using Logstash to Import Data to Elasticsearch

You can use Logstash to collect data and migrate collected data to Elasticsearch in CSS. This method helps you effectively obtain and manage data through Elasticsearch. Data files can be in the JSON or CSV format.

Logstash is an open-source, server-side data processing pipeline that ingests data from a multitude of sources simultaneously, transforms it, and then sends it to Elasticsearch. For details about Logstash, visit the following website: **https://www.elastic.co/guide/en/logstash/current/getting-started-with-logstash.html**

The following two scenarios are involved depending on the Logstash deployment:

- **Importing Data When Logstash Is Deployed on the External Network**
- **Importing Data When Logstash Is Deployed on an ECS**

Prerequisites

- To facilitate operations, you are advised to deploy Logstash on a host that runs the Linux operating system (OS).
- To download Logstash, visit the following website: **https://www.elastic.co/downloads/logstash-oss**

  NOTE

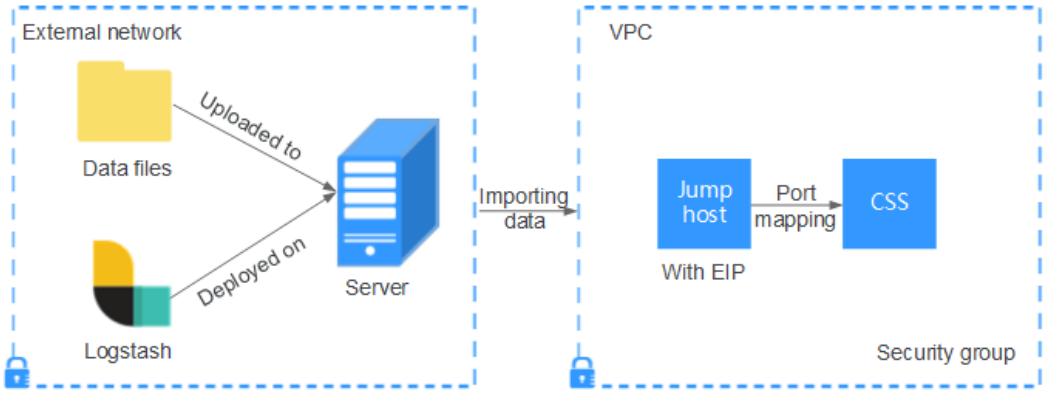
  Logstash requires an OSS version same as the CSS version.
- After installing Logstash, perform the following steps to import data. For details about how to install Logstash, visit the following website: **https://www.elastic.co/guide/en/logstash/current/installing-logstash.html**
- The JDK must be installed before Logstash is installed. In Linux OS, you can run the **yum -y install java-1.8.0** command to install JDK 1.8.0. In Windows OS, you can download the required JDK version from the **official website of JDK**, and install it by following the installation guide.
- In the **Importing Data When Logstash Is Deployed on an ECS** scenario, ensure that the ECS and the Elasticsearch cluster to which data is imported reside in the same VPC.

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

Importing Data When Logstash Is Deployed on the External Network

Figure 8-1 illustrates how data is imported when Logstash is deployed on an external network.

Figure 8-1 Importing data when Logstash is deployed on an external network



1. Create a jump host and configure it as follows:
   - The jump host is an ECS running the Linux OS and has been bound with an EIP.
   - The jump host resides in the same VPC as the CSS cluster.
   - SSH local port forwarding is configured for the jump host to forward requests from a chosen local port to port **9200** on one node of the CSS cluster.
   - Refer to **SSH documentation** for the local port forwarding configuration.
2. Use PuTTY to log in to the created jump host with the EIP.
3. Run the following command to perform port mapping and transfer the request sent to the port on the jump host to the target cluster:

   ```
   ssh -g -L <Local port of the jump host.Private network address and port number of a node> -N -f root@<Private IP address of the jump host>
   ```

   NOTE

   - In the preceding command, *<Local port of the jump host>* refers to the port obtained in 1.
   - In the preceding command, *<Private network address and port number of a node>* refers to the private network address and port number of a node in the cluster. If the node is faulty, the command execution will fail. If the cluster contains multiple nodes, you can replace the value of **<private network address and port number of a node>** with the private network address and port number of any available node in the cluster. If the cluster contains only one node, restore the node and execute the command again.
   - Replace <Private IP address of the *jump host*> in the preceding command with the IP address (with **Private IP**) of the created jump host in the **IP Address** column in the ECS list on the ECS management console.

   For example, port **9200** on the jump host is assigned external network access permissions, the private network address and port number of the node are **192.168.0.81** and **9200**, respectively, and the private IP address of the jump host is **192.168.0.227**. You need to run the following command to perform port mapping:

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

```
ssh -g -L 9200:192.168.0.81:9200 -N -f root@192.168.0.227
```

4. Log in to the server where Logstash is deployed and store the data files to be imported on the server.

   For example, data file **access_20181029_log** needs to be imported, the file storage path is **/tmp/access_log/**, and the data file includes the following data:

   NOTE

   Create the **access_log** folder if it does not exist.

```
|   All |             Heap used for segments |                    |    18.6403 |      MB |
|   All |          Heap used for doc values |                    |   0.119289 |      MB |
|   All |               Heap used for terms |                    |    17.4095 |      MB |
|   All |               Heap used for norms |                    |  0.0767822 |      MB |
|   All |              Heap used for points |                    |   0.225246 |      MB |
|   All |       Heap used for stored fields |                    |   0.809448 |      MB |
|   All |                     Segment count |                    |        101 |         |
|   All |                  Min Throughput |         index-append |    66232.6 |  docs/s |
|   All |               Median Throughput |         index-append |    66735.3 |  docs/s |
|   All |                  Max Throughput |         index-append |    67745.6 |  docs/s |
|   All |        50th percentile latency |         index-append |    510.261 |      ms |
```

5. In the server where Logstash is deployed, run the following command to create configuration file **logstash-simple.conf** in the Logstash installation directory:

```
cd /<Logstash installation directory>/
vi logstash-simple.conf
```

6. Input the following content in **logstash-simple.conf**:

```
input {
Location of data
}
filter {
Related data processing
}
output {
    elasticsearch {
        hosts => "<EIP of the jump host>:<Number of the port assigned external network access permissions on the jump host>"
    }
}
```

   - The **input** parameter indicates the data source. Set this parameter based on the actual conditions. For details about the **input** parameter and parameter usage, visit the following website: **https://www.elastic.co/guide/en/logstash/current/input-plugins.html**
   - The **filter** parameter specifies the mode in which data is processed. For example, extract and process logs to convert unstructured information into structured information. For details about the **filter** parameter and parameter usage, visit the following website: **https://www.elastic.co/guide/en/logstash/current/filter-plugins.html**
   - The **output** parameter indicates the destination address of the data. For details about the **output** parameter and parameter usage, visit **https://www.elastic.co/guide/en/logstash/current/output-plugins.html**. Replace *<EIP address of the jump host>* with the IP address (with **EIP**) of the created jump host in the **IP Address** column in the ECS list on the ECS management console. *<Number of the port assigned external network access permissions on the jump host>* is the number of the port obtained in **1**, for example, **9200**.

   Consider the data files in the **/tmp/access_log/** path mentioned in **4** as an example. Assume that data import starts from data in the first row of the

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

data file, the filtering condition is left unspecified (indicating no data processing operations are performed), the public IP address and port number of the jump host are **192.168.0.227** and **9200**, respectively, and the name of the target index is **myindex**. Edit the configuration file as follows, and enter **:wq** to save the configuration file and exit.

```
input {
    file{
      path => "/tmp/access_log/*"
      start_position => "beginning"
    }
}
filter {
}
output {
    elasticsearch {
      hosts => "192.168.0.227:9200"
      index => "myindex"

    }
}
```

NOTE

If a license error is reported, set **ilm_enabled** to **false**.

If the cluster has the security mode enabled, you need to download a certificate first.

a. Download a certificate on the **Basic Information** page of the cluster.
b. Store the certificate to the server where Logstash is deployed.
c. Modify the **logstash-simple.conf** configuration file.

   Consider the data files in the **/tmp/access_log/** path mentioned in **4** as an example. Assume that data import starts from data in the first row of the data file, the filtering condition is left unspecified (indicating no data processing operations are performed), and the public IP address and port number of the jump host are **192.168.0.227** and **9200**, respectively. The name of the index for importing data is **myindex**, and the certificate is stored in **/logstash/logstash6.8/config/CloudSearchService.cer**. Edit the configuration file as follows, and enter **:wq** to save the configuration file and exit.

```
input{
    file {
        path => "/tmp/access_log/*"
        start_position => "beginning"
    }
}
filter {
    }
output{
    elasticsearch{
        hosts => ["https://192.168.0.227:9200"]
        index => "myindex"
        user => "admin"
        password => "******"
        cacert => "/logstash/logstash6.8/config/CloudSearchService.cer"
    }
}
```

NOTE

password: password for logging in to the cluster

7. Run the following command to import the data collected by Logstash to the cluster:

```
./bin/logstash -f logstash-simple.conf
```

NOTE

This command must be executed in the directory where the **logstash-simple.conf** file is stored. For example, if the **logstash-simple.conf** file is stored in **/root/logstash-7.1.1/**, go to the directory before running the command.

8. Log in to the CSS management console.
9. In the navigation pane on the left, choose **Clusters** > **Elasticsearch** to switch to the **Clusters** page.
10. From the cluster list, locate the row that contains the cluster to which you want to import data and click **Access Kibana** in the **Operation** column.
11. In the Kibana navigation pane on the left, choose **Dev Tools**.
12. On the **Console** page of Kibana, search for the imported data.

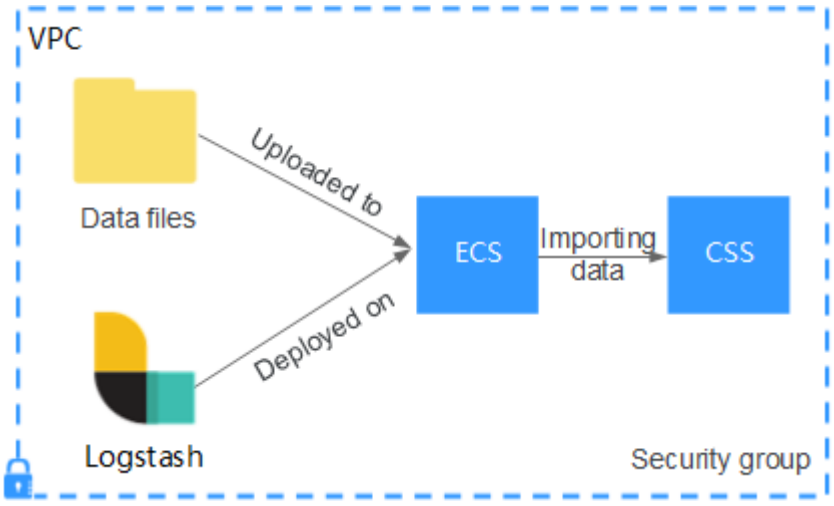
    On the **Console** page of Kibana, run the following command to search for data. View the search results. If the searched data is consistent with the imported data, the data has been imported successfully.

```
GET myindex/_search
```

Importing Data When Logstash Is Deployed on an ECS

Figure 8-2 illustrates how data is imported when Logstash is deployed on an ECS that resides in the same VPC as the cluster to which data is to be imported.

Figure 8-2 Importing data when Logstash is deployed on an ECS



1. Ensure that the ECS where Logstash is deployed and the cluster to which data is to be imported reside in the same VPC, port **9200** of the ECS security group has been assigned external network access permissions, and an EIP has been bound to the ECS.

   NOTE

   - If there are multiple servers in a VPC, you do not need to associate EIPs to other servers as long as one server is associated with an EIP. Switch to the node where Logstash is deployed from the node with which the EIP is associated.
   - If a private line or VPN is available, you do not need to associate an EIP.

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

2. Use PuTTY to log in to the ECS.

   For example, data file **access_20181029_log** is stored in the **/tmp/access_log/** path of the ECS, and the data file includes the following data:

```
|   All |                    Heap used for segments |                       |     18.6403 |      MB |
|   All |                  Heap used for doc values |                       |    0.119289 |      MB |
|   All |                       Heap used for terms |                       |     17.4095 |      MB |
|   All |                       Heap used for norms |                       |   0.0767822 |      MB |
|   All |                      Heap used for points |                       |    0.225246 |      MB |
|   All |            Heap used for stored fields |                       |    0.809448 |      MB |
|   All |                          Segment count |                       |         101 |         |
|   All |                         Min Throughput |         index-append |     66232.6 |  docs/s |
|   All |                      Median Throughput |         index-append |     66735.3 |  docs/s |
|   All |                         Max Throughput |         index-append |     67745.6 |  docs/s |
|   All |                50th percentile latency |         index-append |     510.261 |      ms |
```

3. Run the following command to create configuration file **logstash-simple.conf** in the Logstash installation directory:

```
cd /<Logstash installation directory>/
vi logstash-simple.conf
```

   Input the following content in **logstash-simple.conf**:

```
input {
Location of data
}
filter {
Related data processing
}
output {
    elasticsearch{
        hosts => "<Private network address and port number of the node>"}
}
```

   - The **input** parameter indicates the data source. Set this parameter based on the actual conditions. For details about the **input** parameter and parameter usage, visit the following website: **https://www.elastic.co/guide/en/logstash/current/input-plugins.html**
   - The **filter** parameter specifies the mode in which data is processed. For example, extract and process logs to convert unstructured information into structured information. For details about the **filter** parameter and parameter usage, visit the following website: **https://www.elastic.co/guide/en/logstash/current/filter-plugins.html**
   - The **output** parameter indicates the destination address of the data. For details about the **output** parameter and parameter usage, visit **https://www.elastic.co/guide/en/logstash/current/output-plugins.html**. *<private network address and port number of a node>* refers to the private network address and port number of a node in the cluster.

     If the cluster contains multiple nodes, you are advised to replace the value of *<Private network address and port number of a node>* with the private network addresses and port numbers of all nodes in the cluster to prevent node faults. Use commas (,) to separate the nodes' private network addresses and port numbers. The following is an example:

     ```
     hosts => ["192.168.0.81:9200","192.168.0.24:9200"]
     ```

     If the cluster contains only one node, the format is as follows:

     ```
     hosts => "192.168.0.81:9200"
     ```

   Consider the data files in the **/tmp/access_log/** path mentioned in **2** as an example. Assume that data import starts from data in the first row of the data file, the filtering condition is left unspecified (indicating no data processing operations are performed), the private network address and port number of the node in the cluster where data is to be imported are

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

192.168.0.81 and **9200**, respectively, and the name of the target index is **myindex**. Edit the configuration file as follows, and enter **:wq** to save the configuration file and exit.

```
input {
    file{
      path => "/tmp/access_log/*"
      start_position => "beginning"
    }
}
filter {
}
output {
    elasticsearch {
      hosts => "192.168.0.81:9200"
      index => "myindex"

    }
}
```

If the cluster has the security mode enabled, you need to download a certificate first.

a. Download a certificate on the **Basic Information** page of the cluster.

b. Store the certificate to the server where Logstash is deployed.

c. Modify the **logstash-simple.conf** configuration file.

Consider the data files in the **/tmp/access_log/** path mentioned in step **2** as an example. Assume that data import starts from data in the first row of the data file, the filtering condition is left unspecified (indicating no data processing operations are performed), the public IP address and port number of the jump host are **192.168.0.227** and **9200**, respectively. The name of the index for importing data is **myindex**, and the certificate is stored in **/logstash/logstash6.8/config/CloudSearchService.cer**. Edit the configuration file as follows, and enter **:wq** to save the configuration file and exit.

```
input{
    file {
        path => "/tmp/access_log/*"
        start_position => "beginning"
    }
}
filter {
    }
output{
    elasticsearch{
        hosts => ["https://192.168.0.227:9200"]
        index => "myindex"
        user => "admin"
        password => "******"
        cacert => "/logstash/logstash6.8/config/CloudSearchService.cer"
    }
}
```

NOTE

password: password for logging in to the cluster

4. Run the following command to import the ECS data collected by Logstash to the cluster:

```
./bin/logstash -f logstash-simple.conf
```

5. Log in to the CSS management console.

6. In the navigation pane on the left, choose **Clusters** > **Elasticsearch** to switch to the **Clusters** page.

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

7. From the cluster list, locate the row that contains the cluster to which you want to import data and click **Access Kibana** in the **Operation** column.
8. In the Kibana navigation pane on the left, choose **Dev Tools**.
9. On the **Console** page of Kibana, search for the imported data.

   On the **Console** page of Kibana, run the following command to search for data. View the search results. If the searched data is consistent with the imported data, the data has been imported successfully.

   ```
   GET myindex/_search
   ```

8.2 Using Kibana or APIs to Import Data to Elasticsearch

You can import data in various formats, such as JSON, to Elasticsearch in CSS by using Kibana or APIs.

Importing Data Using Kibana

Before importing data, ensure that you can use Kibana to access the cluster. The following procedure illustrates how to use the **POST** command to import data.

1. Log in to the CSS management console.
2. In the navigation pane on the left, choose **Clusters** > **Elasticsearch** to switch to the **Clusters** page.
3. Choose **Clusters** in the navigation pane. Locate the target cluster and click **Access Kibana** in the **Operation** column to log in to Kibana.
4. Click **Dev Tools** in the navigation tree on the left.
5. (Optional) On the **Console** page, run the related command to create an index for storing data and specify a custom mapping to define the data type.

   If there is an available index in the cluster where you want to import data, skip this step. If there are no available indexes, create one by referring to the following sample code.

   For example, on the **Console** page of Kibana, run the following command to create an index named **my_store** and specify a user-defined mapping to define the data type:

   Versions earlier than 7.*x*

   ```
   PUT /my_store
   {
     "settings": {
       "number_of_shards": 1
     },
     "mappings": {
       "products": {
         "properties": {
           "productName": {
             "type": "text"
           },
           "size": {
             "type": "keyword"
           }
         }
       }
     }
   }
   ```

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

Versions later than 7.*x*

```
PUT /my_store
{
   "settings": {
      "number_of_shards": 1
   },
   "mappings": {
      "properties": {
         "productName": {
            "type": "text"
         },
         "size": {
            "type": "keyword"
         }
      }
   }
}
```

6. Run commands to import data. For example, run the following command to import a piece of data:

   Versions earlier than 7.*x*

   ```
   POST /my_store/products/_bulk
   {"index":{}}
   {"productName":"Latest art shirts for women in 2017 autumn","size":"L"}
   ```

   Versions later than 7.*x*

   ```
   POST /my_store/_bulk
   {"index":{}}
   {"productName":"Latest art shirts for women in 2017 autumn","size":"L"}
   ```

   The command output is similar to that shown in **Figure 8-3**. If the value of the **errors** field in the result is **false**, the data is successfully imported.

   Figure 8-3 Response message


```
{
  "took": 42,
  "errors": false,
  "items": [
    {
      "index": {
        "_index": "my_store",
        "_type": "products",
        "_id": "AWTGbHt7BwpN-hb3LKau",
        "_version": 1,
        "result": "created",
        "_shards": {
          "total": 2,
          "successful": 2,
          "failed": 0
        },
        "created": true,
        "status": 201
      }
    }
  ]
}
```

Importing Data Using APIs

You can call the bulk API using the cURL command to import a JSON data file.

NOTE

You are advised to import a file smaller than 50 MB.

1. Log in to the ECS that you use to access the cluster.

   For details about how to log in to an ECS, see .

2. Run the following command to import JSON data:

   In the command, replace the value of {*Private network address and port number of the node*} with the private network address and port number of a node in the cluster. If the node fails to work, the command will fail to be executed. If the cluster contains multiple nodes, you can replace the value of {*Private network address and port number of the node*} with the private network address and port number of any available node in the cluster. If the cluster contains only one node, restore the node and execute the command again. **test.json** indicates the JSON file whose data is to be imported.

   ```
   curl -X PUT "http://{Private network address and port number of the node} /_bulk" -H 'Content-Type: application/json' --data-binary @test.json
   ```

   If communication encryption has been enabled on the cluster where you will import data, you need to send HTTPS requests and add **-k** to the cURL command.

   ```
   curl -X PUT -k "https://{Private network address and port number of the node} /_bulk" -H 'Content-Type: application/json' --data-binary @test.json
   ```

   NOTE

   The value of the **-X** parameter is a command and that of the **-H** parameter is a message header. In the preceding command, **PUT** is the value of the **-X** parameter and **'Content-Type: application/json' --data-binary @test.json** is the value of the **-H** parameter. Do not add **-k** between a parameter and its value.

   Example 1: In this example, assume that you need to import data in the **testdata.json** file to an Elasticsearch cluster, where communication encryption is disabled and the private network address and port number of one node are **192.168.0.90** and **9200** respectively. The data in the **testdata.json** file is as follows:

   Versions earlier than 7.*x*

   ```
   {"index": {"_index":"my_store","_type":"products"}}
   {"productName":"Autumn new woman blouses 2019","size":"M"}
   {"index": {"_index":"my_store","_type":"products"}}
   {"productName":"Autumn new woman blouses 2019","size":"L"}
   ```

   Versions later than 7.*x*

   ```
   {"index": {"_index":"my_store"}}
   {"productName":"Autumn new woman blouse 2019","size":"M"}
   {"index": {"_index":"my_store"}}
   {"productName":"Autumn new woman blouse 2019","size":"L"}
   ```

   Perform the following steps to import the data:

   a. Run the following command to create an index named **my_store**:

      Versions earlier than 7.*x*

      ```
      curl -X PUT http://192.168.0.90:9200/my_store -H 'Content-Type: application/json' -d '
      {
        "settings": {
          "number_of_shards": 1
      ```

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

```
  },
  "mappings": {
    "products": {
      "properties": {
        "productName": {
          "type": "text"
        },
        "size": {
          "type": "keyword"
        }
      }
    }
  }
}'
```

Versions later than 7.*x*

```
curl -X PUT http://192.168.0.90:9200/my_store -H 'Content-Type: application/json' -d '
{
    "settings": {
        "number_of_shards": 1
    },
    "mappings": {
        "properties": {
            "productName": {
                "type": "text"
            },
            "size": {
                "type": "keyword"
            }
        }
    }
}'
```

b. Run the following command to import the data in the **testdata.json** file:

```
curl -X PUT "http://192.168.0.90:9200/_bulk" -H 'Content-Type: application/json' --data-binary
@testdata.json
```

Example 2: In this example, assume that you need to import data in the **testdata.json** file to an Elasticsearch cluster, where communication encryption has been enabled and the node access address and content in the **testdata.json** are the same as those in example 1. Perform the following steps to import the data:

a. Run the following command to create an index named **my_store**:

```
curl -X PUT -k https://192.168.0.90:9200/my_store -H 'Content-Type: application/json' -d '
{
  "settings": {
    "number_of_shards": 1
  },
  "mappings": {
    "products": {
      "properties": {
        "productName": {
          "type": "text"
        },
        "size": {
          "type": "keyword"
        }
      }
    }
  }
}'
```

b. Run the following command to import the data in the **testdata.json** file:

```
curl -X PUT -k "https://192.168.0.90:9200/_bulk" -H 'Content-Type: application/json' --data-
binary @testdata.json
```

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

9 Monitoring

9.1 Monitoring Metrics of Clusters

Function

This topic describes CSS metrics that can be monitored by Cloud Eye as well as their namespaces and dimensions.

Namespace

SYS.ES

Monitoring Metrics

- **Table 9-1** describes the monitoring metrics of CSS clusters.
- Monitored object: CSS cluster
- Monitoring period (original metric): 1 minute

NOTE

Accumulated value: The value is accumulated from the time when a node is started. After the node is restarted, the value is reset to zero and accumulated again.

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

Table 9-1 CSS metrics

Metric ID	Metric	Description	Value Range	Monitored Object	Monitoring Interval (Raw Data)
status	Cluster Health Status	Health status of the monitored object	0,1,2,3 **0**: The cluster is 100% available. **1**: The data is complete while some replicas are missing. Exceptions may occur because the high availability is compromised. This is a warning that should prompt investigation. **2**: Data is missing and the cluster fails to work. **3**: The cluster status is not obtained.	CSS cluster	1 minute
disk_util	Disk Usage	Disk usage of the monitored object. Unit: %	0-100%	CSS cluster	1 minute
max_jvm_heap_usage	Max. JVM Heap Usage	Maximum JVM heap usage of nodes in a CSS cluster. Unit: %	0-100%	CSS cluster	1 minute
max_jvm_young_gc_time	Max. JVM Young GC Duration	Maximum accumulated JVM Young GC duration of nodes in a CSS cluster. Unit: ms	≥ 0 ms	CSS cluster	1 minute

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

Metric ID	Metric	Description	Value Range	Monitored Object	Monitoring Interval (Raw Data)
max_jvm_young_gc_count	Max. JVM Young GC Count	Maximum accumulated JVM Young GC count of nodes in a CSS cluster.	≥ 0	CSS cluster	1 minute
max_jvm_old_gc_time	Max. JVM Old GC Duration	Maximum accumulated JVM Old GC duration of nodes in a CSS cluster. Unit: ms	≥ 0 ms	CSS cluster	1 minute
max_jvm_old_gc_count	Max. JVM Old GC Count	Maximum accumulated JVM Old GC count of nodes in a CSS cluster.	≥ 0	CSS cluster	1 minute
total_fs_size	Total Size of File Systems	Total size of file systems in a CSS cluster. Unit: byte	≥ 0 bytes	CSS cluster	1 minute
free_fs_size	Available Size of File Systems	Available size of file systems in a CSS cluster. Unit: byte	≥ 0 bytes	CSS cluster	1 minute
max_cpu_usage	Max. CPU Usage	Maximum node CPU usage in a CSS cluster. Unit: %	0-100%	CSS cluster	1 minute

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

Metric ID	Metric	Description	Value Range	Monitored Object	Monitoring Interval (Raw Data)
max_cpu_time_of_jvm_process	Max. CPU Time of JVM Process	Maximum accumulated CPU usage duration of node JVM processes in a CSS cluster. Unit: ms	≥ 0 ms	CSS cluster	1 minute
max_virtual_memory_size_of_jvm_process	Max. Virtual Memory Size of JVM Process	Maximum virtual memory size of node JVM processes in a CSS cluster. Unit: byte	≥ 0 bytes	CSS cluster	1 minute
max_current_opened_http_count	Current Max. Opened HTTP Connections	Maximum number of HTTP connections that are currently open for nodes in a CSS cluster.	≥ 0	CSS cluster	1 minute
max_total_opened_http_count	Total Max. Opened HTTP Connections	Maximum number of HTTP connections that were open for nodes in a CSS cluster.	≥ 0	CSS cluster	1 minute
indices_count	Indexes	Number of indexes in a CSS cluster	≥ 0	CSS cluster	1 minute
total_shards_count	Shards	Number of shards in a CSS cluster	≥ 0	CSS cluster	1 minute

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

Metric ID	Metric	Description	Value Range	Monitored Object	Monitoring Interval (Raw Data)
primary_shards_count	Primary Shards	Number of primary shards in a CSS cluster	≥ 0	CSS cluster	1 minute
docs_count	Documents	Number of documents in a CSS cluster	≥ 0	CSS cluster	1 minute
docs_deleted_count	Deleted Documents	Number of documents deleted in a CSS cluster	≥ 0	CSS cluster	1 minute
nodes_count	Nodes	Number of nodes in a CSS cluster	≥ 0	CSS cluster	1 minute
data_nodes_count	Data Nodes	Number of data nodes in a CSS cluster	≥ 0	CSS cluster	1 minute
coordinating_nodes_count	Coordinating Nodes	Number of coordinating nodes in a CSS cluster	≥ 0	CSS cluster	1 minute
master_nodes_count	Master Nodes	Number of master nodes in a CSS cluster	≥ 0	CSS cluster	1 minute
ingest_nodes_count	Client Nodes	Number of client nodes in a CSS cluster	≥ 0	CSS cluster	1 minute
max_load_average	Max. Node Load	Maximum number of average queuing tasks per minute on nodes in a cluster.	≥ 0	CSS cluster	1 minute

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

Metric ID	Metric	Description	Value Range	Monitored Object	Monitoring Interval (Raw Data)
avg_cpu_usage	Avg. CPU Usage	Average node CPU usage in a CSS cluster. Unit: %	0-100%	CSS cluster	1 minute
avg_load_average	Avg. Node Load	Average number of queuing tasks per minute on nodes in a CSS cluster.	≥ 0	CSS cluster	1 minute
avg_jvm_heap_usage	Avg. JVM Heap Usage	Average node JVM heap usage in a CSS cluster. Unit: %	0-100%	CSS cluster	1 minute
max_open_file_descriptors	Max. Open File Descriptors	Maximum number of node file descriptors that are currently open in a CSS cluster.	≥ 0	CSS cluster	1 minute
avg_open_file_descriptors	Avg. Open File Descriptors	Average number of node file descriptors that are currently open in a CSS cluster.	≥ 0	CSS cluster	1 minute
sum_max_file_descriptors	Max. Allowed File Descriptors	Maximum number of allowed node file descriptors in a CSS cluster.	≥ 0	CSS cluster	1 minute

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

Metric ID	Metric	Description	Value Range	Monitored Object	Monitoring Interval (Raw Data)
sum_open_file_descriptors	Open File Descriptors	Number of node file descriptors that are currently open in a cluster.	≥ 0	CSS cluster	1 minute
sum_thread_pool_write_queue	Tasks in Write Queue	Number of job queues in a write thread pool	≥ 0	CSS cluster	1 minute
sum_thread_pool_search_queue	Tasks in Search Queue	Total number of queuing tasks in the search thread pools of nodes in the CSS cluster.	≥ 0	CSS cluster	1 minute
sum_thread_pool_force_merge_queue	Tasks in ForceMerge Queue	Total number of queuing tasks in the force merge thread pools of nodes in the CSS cluster.	≥ 0	CSS cluster	1 minute
sum_thread_pool_write_rejected	Rejected Tasks in Write Queue	Total number of rejected tasks in the write thread pools of nodes in the CSS cluster.	≥ 0	CSS cluster	1 minute

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

Metric ID	Metric	Description	Value Range	Monitored Object	Monitoring Interval (Raw Data)
sum_thread_pool_search_rejected	Rejected Tasks in Search Queue	Total number of rejected tasks in the search thread pools of nodes in the CSS cluster.	≥ 0	CSS cluster	1 minute
sum_thread_pool_force_merge_rejected	Rejected Tasks in ForceMerge Queue	Total number of rejected tasks in the force merge thread pools of nodes in the CSS cluster.	≥ 0	CSS cluster	1 minute
max_thread_pool_search_queue	Max. Tasks in Search Queue	Maximum number of queuing tasks in the search thread pools of nodes in a CSS cluster.	≥ 0	CSS cluster	1 minute
max_thread_pool_force_merge_queue	Max. Tasks in ForceMerge Queue	Maximum number of queuing tasks in the force merge thread pools of nodes in a CSS cluster.	≥ 0	CSS cluster	1 minute
sum_thread_pool_write_threads	Size of Write Thread Pool	Total size of the write thread pools of nodes in the CSS cluster.	≥ 0	CSS cluster	1 minute

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

Metric ID	Metric	Description	Value Range	Monitored Object	Monitoring Interval (Raw Data)
sum_thread_pool_search_threads	Size of Search Thread Pool	Total size of the search thread pools of nodes in the CSS cluster.	≥ 0	CSS cluster	1 minute
sum_thread_pool_force_merge_threads	Size of ForceMerge Thread Pool	Total size of the force merge thread pools of nodes in the CSS cluster.	≥ 0	CSS cluster	1 minute
avg_thread_pool_write_queue	Avg. Tasks in Write Queue	Average number of queuing tasks in the write thread pools of nodes in a CSS cluster.	≥ 0	CSS cluster	1 minute
avg_thread_pool_search_queue	Avg. Tasks in Search Queue	Average number of queuing tasks in the search thread pools of nodes in a CSS cluster.	≥ 0	CSS cluster	1 minute
avg_thread_pool_force_merge_queue	Avg. Tasks in ForceMerge Queue	Average number of queuing tasks in the force merge thread pools of nodes in the CSS cluster.	≥ 0	CSS cluster	1 minute

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

Metric ID	Metric	Description	Value Range	Monitored Object	Monitoring Interval (Raw Data)
avg_thread_pool_search_threads	Avg. Size of Search Thread Pool	Average size of the search thread pool of a node in a CSS cluster.	≥ 0	CSS cluster	1 minute
avg_thread_pool_write_threads	Avg. Size of Write Thread Pool	Average size of the write thread pool of a node in a CSS cluster.	≥ 0	CSS cluster	1 minute
avg_thread_pool_force_merge_threads	Avg. Size of ForceMerge Thread Pool	Average size of the force merge thread pool of a node in a CSS cluster.	≥ 0	CSS cluster	1 minute
avg_thread_pool_write_rejected	Avg. Rejected Tasks in Write Queue	Average number of rejected tasks in the write thread pool of a node in a CSS cluster.	≥ 0	CSS cluster	1 minute
min_free_fs_size	Min. Available Storage Space	Minimum available storage space of nodes in a CSS cluster. Unit: byte	≥ 0 bytes	CSS cluster	1 minute
avg_jvm_old_gc_count	Avg. GCs of Old-Generation JVM	Average number of old-generation garbage collections of nodes in a CSS cluster.	≥ 0	CSS cluster	1 minute

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

Metric ID	Metric	Description	Value Range	Monitored Object	Monitoring Interval (Raw Data)
avg_jvm_old_gc_time	Avg. GC Duration of Old-Generation JVM	Average old-generation garbage collection duration of nodes in a CSS cluster. Unit: ms	≥ 0 ms	CSS cluster	1 minute
avg_jvm_young_gc_count	Avg. GCs of Young-Generation JVM	Average number of young-generation garbage collections of nodes in a CSS cluster.	≥ 0	CSS cluster	1 minute
avg_jvm_young_gc_time	Avg. GC Duration of Young-Generation JVM	Average young-generation garbage collection duration of nodes in a CSS cluster. Unit: ms	≥ 0 ms	CSS cluster	1 minute
avg_max_file_descriptors	Avg. Maximum Allowed File Descriptors	Average value of the maximum number of allowed file descriptors on each node in a CSS cluster.	≥ 0	CSS cluster	1 minute
avg_mem_free_in_bytes	Avg. Available Memory	Average unused memory capacity of nodes in a CSS cluster. Unit: byte	≥ 0 bytes	CSS cluster	1 minute

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

Metric ID	Metric	Description	Value Range	Monitored Object	Monitoring Interval (Raw Data)
avg_mem_free_percent	Avg. Available Memory Percentage	Average percentage of unused memory of nodes in a CSS cluster. Unit: %	0-100%	CSS cluster	1 minute
avg_mem_used_in_bytes	Avg. Used Memory	Average used memory of nodes in a CSS cluster. Unit: byte	≥ 0 bytes	CSS cluster	1 minute
avg_mem_used_percent	Avg. Used Memory Percentage	Average percentage of used memory of nodes in a CSS cluster. Unit: %	0-100%	CSS cluster	1 minute
max_mem_free_in_bytes	Max. Available Memory	Maximum unused memory of nodes in a CSS cluster. Unit: byte	≥ 0 bytes	CSS cluster	1 minute
max_mem_free_percent	Max. Available Memory Percentage	Maximum percentage of unused memory of nodes in a CSS cluster. Unit: %	0-100%	CSS cluster	1 minute
max_mem_used_in_bytes	Max. Used Memory	Maximum used memory of nodes in a CSS cluster. Unit: byte	≥ 0 bytes	CSS cluster	1 minute

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

Metric ID	Metric	Description	Value Range	Monitored Object	Monitoring Interval (Raw Data)
max_mem_used_percent	Max. Used Memory Percentage	Maximum percentage of used memory of nodes in a CSS cluster. Unit: %	0-100%	CSS cluster	1 minute
sum_jvm_old_gc_count	Total GCs of Old-Generation JVM	Number of old-generation garbage collections of nodes in a CSS cluster.	≥ 0	CSS cluster	1 minute
sum_jvm_old_gc_time	Total GC Duration of Old-Generation JVM	Total old-generation garbage collection duration of nodes in the CSS cluster. Unit: ms	≥ 0ms	CSS cluster	1 minute
sum_jvm_young_gc_count	Total GCs of Young-Generation JVM	Number of young-generation garbage collections of nodes in a CSS cluster.	≥ 0	CSS cluster	1 minute
sum_jvm_young_gc_time	Total GC Duration of Young-Generation JVM	Total young-generation garbage collection duration of nodes in the CSS cluster. Unit: ms	≥ 0 ms	CSS cluster	1 minute

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

Metric ID	Metric	Description	Value Range	Monitored Object	Monitoring Interval (Raw Data)
sum_current_opened_http_count	Currently Open HTTP Connections	Number of HTTP connections that are open on nodes in a CSS cluster.	≥ 0	CSS cluster	1 minute
sum_total_opened_http_count	Historical Open HTTP Connections	Number of HTTP connections that were open on nodes in a CSS cluster.	≥ 0	CSS cluster	1 minute
IndexingLatency	Average Index Latency	Average time required for a shard to complete an index operation. Unit: ms	≥ 0 ms	CSS cluster	1 minute
IndexingRate	Average Index Rate	Average number of index operations per second in a cluster.	≥ 0	CSS cluster	1 minute
SearchLatency	Average Search Latency	Average time required for a segment to complete the search operation. Unit: ms	≥ 0 ms	CSS cluster	1 minute
SearchRate	Average QPS	Average queries per second (QPS) in a cluster.	≥ 0	CSS cluster	1 minute

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

Dimension

Table 9-2 Dimension description

Key	Value
cluster_id	CSS cluster

9.2 Monitoring Metrics of Nodes

Function

This topic describes CSS metrics that can be monitored by Cloud Eye as well as their namespaces and dimensions. You can search for the monitoring metrics and alarms generated for CSS on the Cloud Eye console or by calling APIs.

Namespace

SYS.ES

Monitoring Metrics

- Table 9-3 describes the monitoring metrics of CSS clusters.
- Monitored object: Cloud service nodes of CSS clusters
- Monitoring period (original metric): 1 minute

NOTE

Accumulated value: The value is accumulated from the time when a node is started. After the node is restarted, the value is reset to zero and accumulated again.

Table 9-3 CSS metrics

Metric ID	Metric	Description	Value Range	Monitored Object	Monitoring Interval (Raw Data)
jvm_heap_usage	JVM Heap Usage	JVM heap memory usage of a node. Unit: %	0-100 %	CSS cluster - cloud service node	1 minute

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

Metric ID	Metric	Description	Value Range	Monitored Object	Monitoring Interval (Raw Data)
cpu_usage	CPU Usage	CPU usage. Unit: %	0-100 %	CSS cluster - cloud service node	1 minute
load_average	Average Load	Average number of queuing tasks per minute on a node	≥ 0	CSS cluster - cloud service node	1 minute
open_file_descriptors	Open File Descriptors	Number of opened file descriptors on a node	≥ 0	CSS cluster - cloud service node	1 minute
max_file_descriptors	Max. Allowed File Descriptors	Maximum number of allowed file descriptors	≥ 0	CSS cluster - cloud service node	1 minute
thread_pool_write_queue	Tasks in Write Queue	Number of job queues in a write thread pool	≥ 0	CSS cluster - cloud service node	1 minute
thread_pool_search_queue	Tasks in Search Queue	Number of job queues in a search thread pool	≥ 0	CSS cluster - cloud service node	1 minute
thread_pool_force_merge_queue	Tasks in ForceMerge Queue	Number of job queues in a force merge thread pool	≥ 0	CSS cluster - cloud service node	1 minute

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

Metric ID	Metric	Description	Value Range	Monitored Object	Monitoring Interval (Raw Data)
thread_pool_write_rejected	Rejected Tasks in Write Queue	Number of rejected jobs in a write thread pool	≥ 0	CSS cluster - cloud service node	1 minute
thread_pool_search_rejected	Rejected Tasks in Search Queue	Number of rejected jobs in a search thread pool	≥ 0	CSS cluster - cloud service node	1 minute
thread_pool_force_merge_rejected	Rejected Tasks in ForceMerge Queue	Number of rejected jobs in a force merge thread pool	≥ 0	CSS cluster - cloud service node	1 minute
thread_pool_write_threads	Size of Write Thread Pool	Size of a write thread pool	≥ 0	CSS cluster - cloud service node	1 minute
thread_pool_search_threads	Size of Search Thread Pool	Size of a search thread pool	≥ 0	CSS cluster - cloud service node	1 minute
thread_pool_force_merge_threads	Size of ForceMerge Thread Pool	Size of a force merge thread pool	≥ 0	CSS cluster - cloud service node	1 minute
free_fs_size	Available Size of File Systems	Available size of file systems in a CSS cluster	≥ 0	CSS cluster - cloud service node	1 minute

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

Metric ID	Metric	Description	Value Range	Monitored Object	Monitoring Interval (Raw Data)
total_fs_size	Total Size of File Systems	Total size of file systems in a CSS cluster	≥ 0	CSS cluster - cloud service node	1 minute
jvm_old_gc_count	Total GCs of Old-Generation JVM	Number of old-generation garbage collection times	≥ 0	CSS cluster - cloud service node	1 minute
jvm_old_gc_time	Total GC Duration of Old-Generation JVM	Old-generation garbage collection duration. Unit: ms	≥ 0 ms	CSS cluster - cloud service node	1 minute
jvm_young_gc_count	Total GCs of Young-Generation JVM	Number of young-generation garbage collection times	≥ 0	CSS cluster - cloud service node	1 minute
jvm_young_gc_time	GC Duration of Young-Generation JVM	Young-generation garbage collection duration. Unit: ms	≥ 0 ms	CSS cluster - cloud service node	1 minute
mem_free_in_bytes	Available Memory	Unused memory space of a node. Unit: byte	≥ 0 bytes	CSS cluster - cloud service node	1 minute

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

Metric ID	Metric	Description	Value Range	Monitored Object	Monitoring Interval (Raw Data)
mem_free_percent	Available Memory Percentage	Percentage of unused memory space on a node. Unit: byte	≥ 0 bytes	CSS cluster - cloud service node	1 minute
mem_used_in_bytes	Used Memory	Used memory space of a node. Unit: byte	≥ 0 bytes	CSS cluster - cloud service node	1 minute
current_opened_http_count	Currently Open HTTP Connections	Number of HTTP connections on a node	≥ 0	CSS cluster - cloud service node	1 minute
total_opened_http_count	Total Open HTTP Connections	Total number of HTTP connections on a node	≥ 0	CSS cluster - cloud service node	1 minute

Dimension

Table 9-4 Dimension description

Key	Value
cluster_id	CSS cluster - cloud service node

9.3 Configuring Cluster Monitoring

You can use Cloud Eye to monitor the created clusters. After configuring the cluster monitoring, you can log in to the Cloud Eye management console to view cluster metrics.

The procedure for configuring cluster monitoring:

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

1. **Creating Alarm Rules**: Customize alarm rules for the monitoring metrics. Once a metric exceeds the threshold, the system will notify you by sending emails or HTTP/HTTPS requests.
2. **Configuring Monitoring Metrics**: Configure monitoring metrics for a cluster or a node in the cluster.
3. **Viewing Monitoring Metrics**: View the statistics of the monitoring metrics in specific periods.

Prerequisites

- The cluster is in the **Available** or **Processing** status.
- The cluster has been running properly for more than 10 minutes.

Recommended Monitoring Metrics

- Cluster CPU and JVM usage. You are advised to configure the following monitoring metrics: average JVM heap usage, maximum JVM heap usage, average CPU usage, and maximum CPU usage.
- Cluster write and query latency and throughput. You are advised to configure the following monitoring metrics: average index latency, average index rate, average search latency, and average QPS.
- Cluster write and query queue and rejected tasks. You are advised to configure the following monitoring metrics: tasks in write queue, tasks in search queue, rejected tasks in write queue, and rejected tasks in search queue.

Creating Alarm Rules

1. Log in to the Cloud Eye console.
2. In the navigation pane on the left, choose **Alarm Management** > **Alarm Rules**.
3. In the **Resource Type** column, select **Cloud Search Service** as criteria to search for alarm rules that meet the requirements.

   If no alarm rules are available, create one by referring to the "Creating an Alarm Rule" section. For details about how to set **Resource Type** and **Dimension**, see **Table 9-5**.

Table 9-5 Alarm rule configuration parameter

Parameter	Description	Remark
Resource Type	Type of the resource that the alarm rule is created for	Select **Cloud Search Service**.

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

Parameter	Description	Remark
Dimension	Metric dimension of the selected resource type	CSS supports two dimensions. Select a dimension as required. • **CSS Clusters**: Alarm rules are specified by cluster. • **CSS Clusters - CSS Instances**: Alarm rules are specified by node in a cluster.

Configuring Monitoring Metrics

1. Create a monitoring panel by referring to the "Creating a Dashboard" section. If an available monitoring panel has been created, skip this step.
2. Add CSS monitoring graphs by referring to the "Adding a Graph" section.

   For details about how to set **Resource Type** and **Dimension**, see **Table 9-6**.

Table 9-6 Graph configuration parameter

Parameter	Description	Remark
Resource Type	Type of the resource to be monitored	Select **Cloud Search Service**.
Dimension	Metric dimension	CSS supports two dimensions. Select a dimension as required. • **CSS Clusters**: Monitoring is executed by cluster. • **CSS Clusters - CSS Instances**: Monitoring is executed by node in a cluster.

Viewing Monitoring Metrics

1. Log in to the CSS management console.
2. Choose **Clusters**. Locate the target cluster and choose **More** > **View Metric** in the **Operation** column.
3. Select a time range.
4. View the monitoring metrics.

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

10 Auditing

10.1 Key Operations Recorded by CTS

With CTS, you can record operations associated with CSS for future query, audit, and backtracking.

Prerequisites

CTS has been enabled.

Key Operations Recorded by CTS

Table 10-1 Key operations that can be recorded by CTS

Operation	Resource Type	Event Name
Creating a cluster	cluster	createCluster
Querying the cluster details	cluster	showClusterDetail
Deleting a cluster	cluster	deleteCluster
Changing a cluster name	cluster	updateClusterName
Querying the cluster list	cluster	listClusters
Changing the password of a cluster	cluster	resetPassword
Restarting a cluster	cluster	restartCluster
Expanding the cluster capacity	cluster	updateExtendCluster

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

Operation	Resource Type	Event Name
Adding instances and expanding instance storage capacity	cluster	updateExtendInstanceStorage
Changing specifications	cluster	updateFlavor
Obtaining the instance specifications list	cluster	listFlavors
Querying all tags	cluster	listClustersTags
Querying tags of a specified cluster	cluster	showClusterTag
Adding tags for a cluster	cluster	createClustersTags
Deleting the tags of a cluster	cluster	deleteClustersTags
Adding or deleting cluster tags in batches	cluster	updateBatchClustersTags
Deleting specified nodes	cluster	updateShrinkNodes
Changing the flavor of a specified node type	cluster	updateFlavorByType
Scaling in nodes of a specific type	cluster	updateShrinkCluster
Downloading the security certificate	cluster	downloadCert
Replacing nodes	cluster	updateInstance
Changing the security mode	cluster	changeMode
Adding a dedicated Master or Client node	cluster	addIndependentNode
Changing the security group	cluster	changeSecurityGroup
Creating a V2 cluster	cluster	createClusterV2
Restarting a V2 cluster	cluster	restartCluster
Rolling restart	cluster	rollingRestart
Loading a custom word dictionary	cluster	createLoadIkThesaurus
Viewing custom word dictionary configurations	cluster	showIkThesaurus

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

Operation	Resource Type	Event Name
Enabling Kibana public access	cluster	startKibanaPublic
Disabling Kibana public access	cluster	updateCloseKibana
Modifying the Kibana public network bandwidth	cluster	updateAlterKibana
Modifying Kibana public network access control	cluster	updatePublicKibanaWhitelist
Disabling Kibana public network access control	cluster	stopPublicKibanaWhitelist
Enabling logging	cluster	startLogs
Disabling logging	cluster	stopLogs
Querying the job list	cluster	listLogsJob
Querying basic log configurations	cluster	showGetLogSetting
Modifying basic log configurations	cluster	updateLogSetting
Enabling automatic log backup	cluster	startLogAutoBackupPolicy
Disabling automatic log backup	cluster	stopLogAutoBackupPolicy
Backing up logs	cluster	createLogBackup
Querying logs	cluster	showLogBackup
Enabling public network access	cluster	createBindPublic
Disabling public network access	cluster	updateUnbindPublic
Modifying public network access bandwidth	cluster	updatePublicBandWidth
Enabling the public network access whitelist	cluster	startPublicWhitelist
Disabling the public network access whitelist	cluster	stopPublicWhitelist

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

Operation	Resource Type	Event Name
Automatically setting basic configurations of a cluster snapshot	cluster	startAutoSetting
Modifying basic configurations of a cluster snapshot	cluster	updateSnapshotSetting
Manually creating a snapshot	snapshot	createSnapshot
Restoring a snapshot	snapshot	restoreSnapshot
Deleting a specified snapshot	snapshot	deleteSnapshot
Setting the automatic snapshot creation policy	cluster	createAutoCreatePolicy
Querying automatic snapshot creation policies	cluster	showAutoCreatePolicy
Querying the snapshot list	cluster	listSnapshots
Disabling the snapshot function	cluster	stopSnapshot
Enabling automatic snapshot creation	cluster	startAutoCreateSnapshots
Disabling automatic snapshot creation	cluster	stopAutoCreateSnapshots
Modifying parameter settings	cluster	listYmls
Obtaining the task list of parameter settings	cluster	listYmlsJob
Obtaining the parameter settings list	cluster	updateYmls
Obtaining the intelligent O&M task list and details	cluster	listAiOps
Creating a cluster detection task	cluster	createAiOps
Deleting a detection task record	cluster	deleteAiOps

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

Querying Real-Time Traces

After a management tracker is created on the CTS console, the system starts recording operations performed on cloud service resources. After a data tracker is created, the system starts recording operations performed on data in OBS buckets. CTS retains operation records generated in the latest seven days.

To view or export operation records of the last seven days on the CTS console, see *CTS User Guide*.

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

11 FAQs

11.1 General Consulting

11.1.1 What Are Regions and AZs?

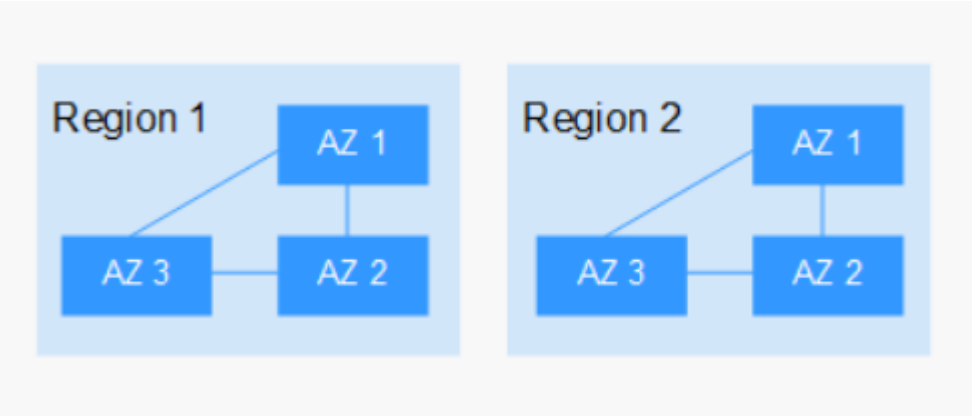
Regions and AZs

A region or an availability zone (AZ) identifies the location of a data center. You can create resources in a specific region or an AZ.

- A region is a geographic area where physical data centers are located. Each region is completely independent, improving the fault tolerance capability and stability. After a resource is created, its region cannot be changed.
- An AZ is a physical location with independent power supplies and network in a region. A region can contain multiple AZs, which are physically isolated but interconnected through internal networks. This ensures the independence of AZs and provides low-cost and low-latency network connections.

Figure 11-1 illustrates the relationship between regions and AZs.

Figure 11-1 Regions and AZs



 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

Selecting a Region

You are advised to select a region close to you or your target users. This reduces network latency and improves access rate.

Selecting an AZ

When deploying resources, consider your applications' requirements on disaster recovery (DR) and network latency.

- For high DR capability, deploy resources in different AZs within the same region.
- If your applications require low latency between instances, you are advised to deploy resources in the same AZ.

Regions and Endpoints

Before using an API to call resources, you will need to specify the resource region and endpoint. For details, see "Endpoints" in *Cloud Search Service API Reference*.

11.1.2 How Does CSS Ensure Data and Service Security?

CSS uses network isolation, in addition to various host and data security measures.

- Network isolation

  The entire network is divided into two planes: service plane and management plane. The two planes are deployed and isolated physically to ensure the security of the service and management networks.

  - Service plane: This is the network plane of the cluster. It provides service channels for users and delivers data definitions, indexing, and search capabilities.
  - Management plane: This is the management console, where you manage CSS.

- Host security

  CSS provides the following security measures:

  - The VPC security group ensures the security of the hosts in a VPC.
  - Network access control lists (ACLs) allow you to control what data can enter or exit your network.
  - The internal security infrastructure (including the network firewall, intrusion detection system, and protection system) monitors all network traffic that enters or exits the VPC through an IPsec VPN.

- Data security

  Multiple replicas, cross-AZ deployment of clusters, and third-party (OBS) backup of index data ensure the security of user data.

11.1.3 Which CSS Metrics Should I Focus On?

Disk usage and cluster health status are two key metrics that you need to focus on. You can log in to Cloud Eye and configure alarm rules for these metrics. If alarms are reported, handle them by taking appropriate measures.

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

Configuration examples:

- Alarms are reported if the disk usage is higher than or equal to a specified value (for example, 85%) and has reached this value multiple times (for example, 5 times) within a specified time period (for example, 5 minutes).
- Alarms are reported if the value of the cluster health status metric exceeds 0 for multiple times (for example, 5 times) within a specified time period (for example, 5 minutes).

Measures:

- If disk usage alarms are reported, view available disk space, check whether data can be deleted from cluster nodes or archived to other systems to free up space, or check if you can expand the disk capacity.
- If cluster health status alarms are reported, check whether shard allocation is normal, whether shards have been lost, and check whether the process has been restarted on Cerebro.

11.1.4 What Storage Options Does CSS Provide?

CSS uses EVS and local disks to store your indices. During cluster creation, you can specify the EVS disk type and specifications (the EVS disk size).

- Supported EVS disk types include common I/O, high I/O, and ultra-high I/O.
- The EVS disk size varies depending on the node specifications selected when you create a cluster.

11.1.5 What Is the Maximum Storage Capacity of CSS?

You can configure up to 200 nodes for a cluster (each node corresponds to an ECS). The maximum storage capacity of an ECS is the total capacity of EVS disks attached to the ECS.

You can calculate the total storage capacity of CSS based on the sizes of EVS disks attached to different ECSs. The EVS disk size is determined by the node specifications selected when you create the cluster.

11.1.6 Which Tools Can I Adopt to Use Cloud Search Service?

You can use any of the following three methods to manage CSS or to use search engine APIs. You can initiate requests based on constructed request messages.

- curl

  curl is a command-line tool used to transfer data to or from a given URL. It serves as an HTTP client that can send HTTP requests to the HTTP server and receive response messages. You can also use curl to debug APIs. For more information about curl, visit **https://curl.haxx.se/**.
- Encoding

  You can call APIs through code to assemble, send, and process request messages.
- REST client

  Both Mozilla Firefox and Google Chrome provide a graphical browser plugin, the REST client, which you can use to send and process requests.

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

- For Mozilla Firefox, see **Firefox REST Client**.
- For Google Chrome, see **Postman**.

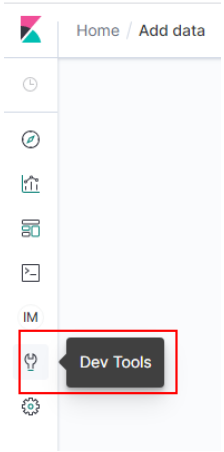
11.1.7 What Can the Disk Space of a CSS Cluster Be Used For?

You can store the following logs and files:

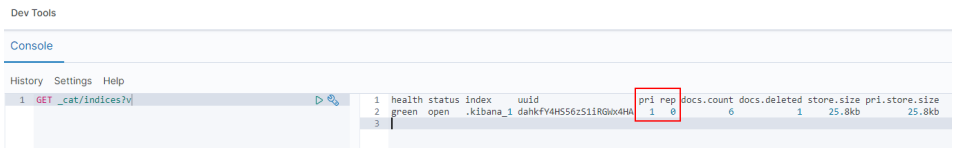
- Log files: Elasticsearch logs
- Data files: Elasticsearch index files
- Other files: cluster configuration files
- OS: 5% storage space reserved for the OS by default

11.1.8 How Do I Check the Numbers of Shards and Replicas in a Cluster on the CSS Console?

1. Log in to the console.
2. On the **Clusters** page, click **Access Kibana** in the **Operation** column of a cluster.
3. Log in to Kibana and choose **Dev Tools**.



4. On the **Console** page, run the **GET _cat/indices?v** command query the number of shards and replicas in a cluster. In the following figure, the **pri** column indicates the number of index shards, and the **rep** column indicates the number of replicas. After an index is created, its **pri** value cannot be modified. Its **rep** value can be modified.



11.1.9 What Data Compression Algorithms Does CSS Use?

CSS supports two data compression algorithms: LZ4 (by default) and best_compression.

- **LZ4 algorithm**

  LZ4 is the default compression algorithm of Elasticsearch. This algorithm can compress and decompress data quickly, but its compression ratio is low.

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

LZ4 scans data with a 4-byte window, which slides 1 byte forward at a time. Duplicate data is compressed. This algorithm applies to scenarios where a large amount of data to be read while a small amount of data to be written.

- **best_compression algorithm**

  In addition to the default LZ4 algorithm, CSS also supports the custom best_compression algorithm. This algorithm can be used when a large amount of data is written and the index storage cost is high, such as log scenarios and time sequence analysis scenarios. This algorithm can greatly reduce the index storage cost.

Run the following command to switch the default compression algorithm (LZ4) to best_compression:

```
PUT index-1
{
   "settings": {
      "index": {
         "codec": "best_compression"
      }
   }
}
```

The LZ4 algorithm can quickly compress and decompress data while the best_compression algorithm has a higher compression and decompression ratio.

11.2 Functions

11.2.1 Can Elasticsearch Data Be Migrated Between VPCs?

Elasticsearch does not support direct data migration between different VPCs. You can use either of the following methods to migrate data.

Method 1

Use the backup and restoration function to migrate cluster data.

Method 2

1. Connect the VPC network and establish a VPC peering connection.
2. After the network is connected, use Logstash to migrate data.

11.2.2 How Do I Migrate a CSS Cluster Across Regions?

CSS clusters cannot be directly migrated. You can back up a cluster to an OBS bucket and restore it to a new region.

- If the OBS bucket is in the same region as your CSS cluster, migrate the cluster by following the instructions in the section "Index Backup and Restoration".
- If the OBS bucket is not in the same region as your CSS cluster, back up the cluster to the bucket by referring to the section "Configuring Cross-Region Replication" and migrate the cluster by referring to the section "Index Backup and Restoration" of *Object Storage Service User Guide*.

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

NOTE

- Before cross-region replication, ensure the snapshot folder of the destination cluster is empty. Otherwise, the snapshot information cannot be updated to the snapshot list of the destination cluster.
- Before every migration, ensure the folder is empty.

11.2.3 How Do I Configure the Threshold for CSS Slow Query Logs?

The slow query log settings of CSS are the same as those of Elasticsearch. You can configure slow query logs via the _settings API. For example, you can run the following command in Kibana to set the index level:

```
PUT /my_index/_settings
{
   "index.search.slowlog.threshold.query.warn": "10s",
   "index.search.slowlog.threshold.fetch.debug": "500ms",
   "index.indexing.slowlog.threshold.index.info": "5s"
}
```

- If a query takes longer than 10 seconds, a WARN log will be generated.
- If retrieval takes longer than 500 milliseconds, a DEBUG log will be generated.
- If an index takes longer than 5 seconds, an INFO log will be generated.

For details, visit the official website: https://www.elastic.co/guide/cn/elasticsearch/guide/current/logging.html

11.2.4 How Do I Update the CSS Lifecycle Policy?

The CSS lifecycle is implemented using the Index State Management (ISM) of Open Distro. For details about how to configure policies related to the ISM template, see the **Open Distro documentation**.

1. When a policy is created, the system writes a record to the **.opendistro-ism-config** index. In the record, **_id** is the policy name, and the content is the policy definition.

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

Figure 11-2 Writing a data record

```
{
  "_index" : ".opendistro-ism-config",
  "_type" : "_doc",
  "_id" : "policy1",
  "_score" : 1.0,
  "_source" : {
    "policy" : {
      "policy_id" : "policy1",
      "description" : "A simple default policy that changes the replica count between hot and cold states.",
      "last_updated_time" : 1641432150329,
      "schema_version" : 1,
      "error_notification" : null,
      "default_state" : "hot",
      "states" : [
        {
          "name" : "hot",
          "actions" : [ ],
          "transitions" : [
            {
              "state_name" : "delete",
              "conditions" : {
                "min_index_age" : "2d"
              }
            }
          ]
        },
        {
          "name" : "delete",
          "actions" : [
            {
              "delete" : { }
            }
          ],
          "transitions" : [ ]
        }
      ]
    }
  }
]
```


2. After a policy is bound to an index, the system writes another record to the **.opendistro-ism-config** index. The following figure shows the initial status of a record.

Figure 11-3 Initial data status

```
{
  "_index" : ".opendistro-ism-config",
  "_type" : "_doc",
  "_id" : "FABkSF5GSTCmR0QkW41HVw",
  "_score" : 1.0,
  "_source" : {
    "managed_index" : {
      "name" : "data1",
      "enabled" : true,
      "index" : "data1",
      "index_uuid" : "FABkSF5GSTCmR0QkW41HVw",
      "schedule" : {
        "interval" : {
          "start_time" : 1641432652693,
          "period" : 1,
          "unit" : "Minutes"
        }
      },
      "last_updated_time" : 1641432652694,
      "enabled_time" : 1641432652694,
      "policy_id" : "policy1",
      "policy_seq_no" : null,
      "policy_primary_term" : null,
      "policy" : null,
      "change_policy" : null
    }
  }
}
]
```


3. Run the **explain** command. Only a policy ID will be returned.

```
GET _opendistro/_ism/explain/data2
{
  "data2" : {
```

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

```
  "index.opendistro.index_state_management.policy_id" : "policy1"
 }
}
```

Open Distro will execute an initialization process to fill the policy content in the record. The following figure shows the initialized data.

Figure 11-4 Initialized data

After the initialization, **min_index_age** in the policy will be copied.

NOTE

The initialized index uses a copy of this policy. The policy update will not take effect on the index.

4. After the policy is modified, call the **change_policy** API to update the policy.

```
POST _opendistro/_ism/change_policy/data1
{
 "policy_id": "policy1"
}
```

11.2.5 How Do I Set the Numbers of Index Copies to 0 in Batches?

1. Log in to the Kibana page of the cluster. In the navigation pane, choose **Dev Tools**.
2. Modify and run the **PUT /*/_settings{"number_of_replicas":0}** command.

NOTE

Do not directly run the preceding command, because the asterisk (*) may match security indexes. You are advised to specify the index required for the batch operation. Example: **PUT /test*/_settings{"number_of_replicas":0}**

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

11.2.6 Why All New Index Shards Are Allocated to the Same Node?

Possible Cause

The possible causes are as follows:

- Shards were unevenly distributed in previous index allocations, and the predominate parameter in the latest indexed shard allocation was **balance.shard**. To balance the shard distribution across nodes, the new shards were allocated to the node with only a small number of shards.
- After a new node was added to a cluster and before the automatic cluster rebalancing completes, the predominate parameter was **balance.shard**. The shards of a new index are allocated to the new node, where there are no shards yet.

The following two parameters are used to balance the shard allocation in a cluster:

cluster.routing.allocation.balance.index (default value: **0.45f**)

cluster.routing.allocation.balance.shard (default value: **0.55f**)

NOTE

- **balance.index**: A larger value indicates that all the shards of an index are more evenly distributed across nodes. For example, if an index has six shards and there are three data nodes, two shards will be distributed on each node.
- **balance.shard**: A larger value indicates that all the shards of all the indexes are more evenly distributed across nodes. For example, if index **a** has two shards, index **b** has four, and there are three data nodes, two shards will be distributed on each node.
- You can specify both **balance.index** and **balance.shard** to balance the shard allocation.

Solution

To prevent the all the shards of an index from being allocated to a single node, use either of the following methods:

1. To create an index during cluster scale-out, configure the following parameter:
   ```
   "index.routing.allocation.total_shards_per_node": 2
   ```
   That is, allow no more than two shards of an index to be allocated on each node. Determine the maximum number of shards allocated to each node based on the number of data nodes in your cluster and the number of index shards (both primary and secondary).
2. If too many shards are distributed on only a few nodes, you can move some of the shards to other nodes to balance the distribution. Run the **move** command of **POST _cluster/reroute**. The rebalance module will automatically exchange the shard with a shard on the destination node. Determine the values of **balance.index** and **balance.shard** as needed.

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

11.2.7 How Do I Query Snapshot Information?

Prerequisites

The snapshot function has been enabled for the cluster and snapshot information has been configured.

Querying a Snapshot

1. Log in to the CSS management console, and click **Clusters** in the navigation pane. On the displayed **Clusters** page, locate the target cluster and click **Access Kibana** in the **Operation** column.
2. In the left navigation pane of the Kibana page, click **Dev Tools**. Click **Get to work** to switch to the **Console** page.

   Enter the code as required in the left pane, click  to execute the command, and view the result in the right pane.
3. Run the **GET _snapshot/_all** command to query information about all repositories.

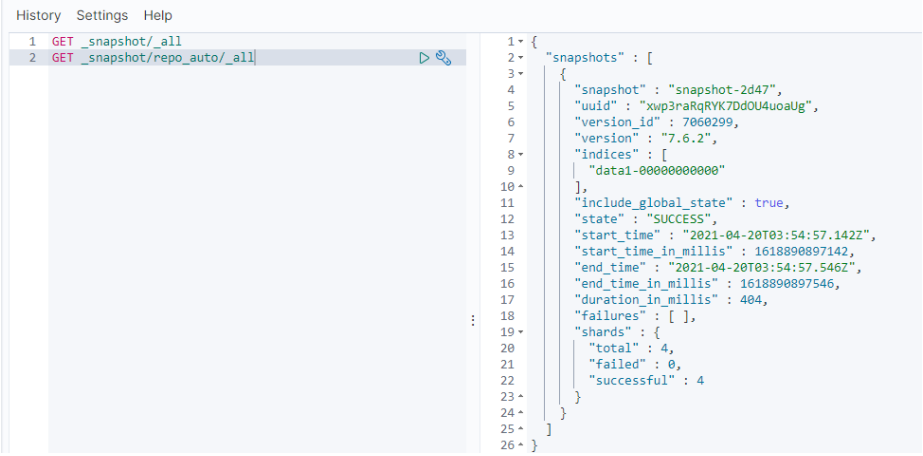
   Figure 11-5 Querying information about all repositories

   - **bucket**: OBS bucket name
   - **base_path**: Path. It consists of a fixed prefix and a cluster name.
   - **endpoint**: OBS domain name
   - **region**: your region
4. Query snapshot information.
   a. Run the **GET _snapshot/repo_auto/_all** command to query the list of all the snapshots in the current repository.

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

Figure 11-6 Snapshot information



- **snapshot**: snapshot name
- **state**: snapshot status
- **start_time**, **start_time_in_millis**, **end_time**, and **end_time_in_millis**: snapshot time
- **shards**: the number of shards. **total** indicates the total number of shards. **failed** indicates the number of failures. **successful** indicates the number of successes.

b. Run the **GET _snapshot/repo_auto/$snapshot-xxx** command to query information about a specified snapshot.

- Replace **$snapshot-xxx** with the actual snapshot name.
- **repo_auto** is followed by a snapshot name or wildcard characters.

5. (Optional) Delete information about a specified snapshot.

   To delete a specific snapshot, run the **DELETE _snapshot/ repo_auto/ $snapshot-xxx** command.

   Replace **$snapshot-xxx** with the actual snapshot name.

11.2.8 Can I Upgrade a Cluster from an Earlier Version to a Later Version?

A cluster cannot be directly upgraded. You can purchase a cluster of a later version and migrate your data to it.

1. Creating a cluster: Create a cluster of a later version in the same region.
2. Migrating a cluster: You can use the index backup and restoration function to migrate a cluster.

11.2.9 Can I Restore a Deleted Cluster?

Yes. You can use a snapshot stored in OBS to restore a cluster. A deleted cluster that has no snapshots in OBS cannot be restored. Exercise caution when deleting a cluster.

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

To restore a deleted cluster, perform the following steps:

1. Log in to the CSS management console.
2. In the navigation pane on the left, choose **Clusters**. On the displayed **Clusters** page, click **Create Cluster** in the upper right corner to create a cluster and enable the snapshot function. Set the OBS bucket and backup path to those of the cluster to be restored.

   To restore a deleted cluster to an existing cluster, set the OBS bucket and backup path to those of the deleted cluster.

   NOTICE

   To restore a deleted cluster to a new cluster, ensure they are in the same region. The new cluster version must be the same as or later than that of the deleted cluster. The number of nodes in the new cluster must be greater than half of that in the deleted cluster. Otherwise, the cluster may fail to be restored.

3. If the status of the new cluster changes to **Available**, click the cluster name to go to the **Cluster Information** page.
4. In the navigation pane on the left, choose **Cluster Snapshots**.

   In the snapshot management list, you can view the snapshot information of the deleted cluster. If no information is displayed, wait for several minutes and refresh the page.
5. Locate the target snapshot and click **Restore** in the **Operation** column. The **Restore** page is displayed.
6. On the **Restore** page, set restoration parameters.

   Index: Enter the name of the index you want to restore. If you do not specify any index name, data of all indexes will be restored. The value can contain 0 to 1,024 characters. Uppercase letters, spaces, and certain special characters (including "\<|>/?) are not allowed.

   Rename Pattern: Enter a regular expression. Indexes that match the regular expression are restored. The default value **index_(.+)** indicates restoring data of all indexes. The value contains 0 to 1,024 characters. Uppercase letters, spaces, and certain special characters (including "\<|>/?,) are not allowed.

   Rename Replacement: Enter the index renaming rule. The default value **restored_index_$1** indicates that **restored_** is added in front of the names of all restored indexes. The value contains 0 to 1,024 characters. Uppercase letters, spaces, and certain special characters (including "\<|>/?,) are not allowed. You can set **Rename Replacement** only if you have specified **Rename Pattern**.

   Cluster: Select the cluster that you want to restore. You can select the current cluster or others. However, you can only restore the snapshot to clusters whose status is **Available**. If the status of the current cluster is **Unavailable**, you cannot restore the snapshot to the current cluster. If you choose to restore the snapshot to another cluster, ensure that the target cluster runs an Elasticsearch version that is not earlier than that of the current cluster. If you select another cluster and two or more indexes in the cluster have the same name, data of all indexes with the same name as the name you specify will be overwritten. Therefore, exercise caution when you set the parameters.

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

7. Click **OK**. If restoration succeeds, **Task Status** of the snapshot in the snapshot list will change to **Restoration succeeded**, and the index data is generated again according to the snapshot information.

11.2.10 Can I Modify the TLS Algorithm of an Elasticsearch Cluster?

You can modify TLS algorithms in CSS 7.6.2 and later versions.

1. Log in to the CSS management console.
2. In the navigation pane, choose **Clusters**. The cluster list is displayed.
3. Click the name of the target cluster to go to the cluster details page.
4. Select **Parameter Configurations**, click **Edit**, expand the **Customize** parameter, and click **Add**.

   Add the **opendistro_security.ssl.http.enabled_ciphers** parameter and set it to **['TLS_ECDHE_RSA_WITH_AES_128_GCM_SHA256', 'TLS_ECDHE_RSA_WITH_AES_256_GCM_SHA384']**.

   NOTE

   If the parameter value contains multiple algorithm protocols, enclose the value with a pair of square brackets ([]). If the parameter value is a single algorithm protocol, enclose the value with a pair of single quotation marks(' ').

5. After the modification is complete, click **Submit**.In the displayed **Submit Configuration** dialog box, select the box indicating "I understand that the modification will take effect after the cluster is restarted." and click **Yes**.

   If the **Status** is **Succeeded** in the parameter modification list, the modification has been saved.

6. Return to the cluster list and choose **More** > **Restart** in the **Operation** column to restart the cluster and make the modification take effect.

11.2.11 How Do I Set the search.max_buckets Parameter for an Elasticsearch Cluster?

Issue

If the query results on shards exceed the upper limit of records that can be returned (default value: **10000**), you need to increase the limit by changing the value of **search.max_buckets**.

Solution

Run the following command on the **Dev Tools** page of Kibana:

```
PUT _cluster/settings
{
   "persistent": {
      "search.max_buckets": 20000
   }
}
```

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

11.2.12 Does the Value i of node.roles Indicate an Ingest Node?

Issue

If the value of **node.roles** of a client node is **i**, then is this client node an ingest node?

- Are there coordinating only nodes in clusters? Are the client requests distributed to coordinating nodes?
- Are ingest nodes in the idle state when there are no ingest requests?

Solution

If the value of **node.roles** of a client node is **i**, the ingest node mode is enabled.

- The coordinating only nodes of Elasticsearch are called client nodes in CSS. If a cluster has no client nodes, client requests will be distributed to all nodes.
- An ingest node functions as a set of ELK for data conversion. If there is no ingest requests, ingest nodes are not in the idle state.

11.2.13 How Do I Create a Type Under an Index in an Elasticsearch 7.*x* Cluster?

In Elasticsearch 7.*x* and later versions, types cannot be created for indexes.

If you need to use types, add **include_type_name=true** to the command. For example:

```
PUT _template/urldialinfo_template?include_type_name=true
```

After the command is executed, the following information is displayed:

```
"#! Deprecation: [types removal] Specifying include_type_name in put index template requests is deprecated. The parameter will be removed in the next major version. "
```

11.3 Clusters in Security Mode

11.3.1 What Is the Relationship Between the Filebeat Version and Cluster Version?

- Non-security mode: no restrictions.
- Cluster in security mode: The Filebeat OSS version must match the cluster version. For details on how to download the Filebeat OSS version, see **Past Releases of Elastic Stack Software**.

11.3.2 How Do I Obtain the Security Certificate of CSS?

The security certificate (**CloudSearchService.cer**) can be downloaded only for security clusters that have enabled HTTPS access.

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

1. Log in to the CSS management console.
2. In the navigation pane, choose **Clusters**. The cluster list is displayed.
3. Click the name of a cluster to go to the cluster details page.
4. On the **Configuration** page, click **Download Certificate** next to **Security Mode**.

11.3.3 How Do I Convert the Format of a CER Security Certificate?

The security certificate (**CloudSearchService.cer**) can be downloaded only for security clusters that have enabled HTTPS access. Most software supports certificates in the **.pem** or **.jks** format. You need to convert the format of the CSS security certificate.

- Run the following command to convert the security certificate from **.cer** to **.pem**:
  ```
  openssl x509 -inform der -in CloudSearchService.cer –out newname.pem
  ```
- Run the following command to convert the security certificate from **.cer** to **.jks**:
  ```
  keytool -import -alias newname -keystore ./truststore.jks -file ./CloudSearchService.cer
  ```

In the preceding commands, *newname* indicates the user-defined certificate name.

After the command is executed, set the certificate password and confirm the password as prompted. Securely store the password. It will be used for accessing the cluster.

11.4 Resource Usage and Change

11.4.1 How Do I Clear Expired Data to Release Storage Space?

- Run the following command to delete a single index data record.

  curl -XDELETE http://IP:9200/*Index_name*

  NOTE

  IP: the IP address of any node in the cluster
- Run the following command to delete all Logstash data of a day. For example, delete all data on June 19, 2017:

  For a cluster in non-security mode: **curl -XDELETE 'http://IP:9200/logstash-2017.06.19*'**

  For a cluster in security mode: **curl -XDELETE -u username:password 'https://IP:9200/logstash-2017.06.19' -k**

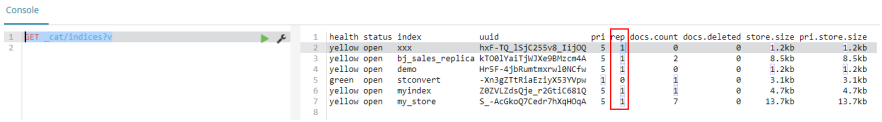
  NOTE

  - **username**: username of the administrator. The default value is **admin**.
  - **password**: the password set during cluster creation
  - **IP**: the IP address of any node in the cluster

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

11.4.2 How Do I Configure a Two-Replica CSS Cluster?

1. Run **GET _cat/indices?v** in Kibana to check the number of cluster replicas. If the value of **rep** is **1**, the cluster has two replicas.



2. If the value of **rep** is not **1**, run the following command to set the number of replicas:

   PUT /index/_settings

   {

   "number_of_replicas" : 1 //Number of replicas

   }

   NOTE

   index specifies the index name. Set this parameter based on site requirements.

11.4.3 How Do I Delete Index Data?

- Manually: Run the **DELETE /my_index** command in Kibana.
- Automatically: Create scheduled tasks to call the index deletion request and periodically execute the tasks. CSS supports Open Distro Index State Management. For details, see: https://opendistro.github.io/for-elasticsearch-docs/docs/im/ism/

11.4.4 Can I Change the Number of Shards to Four with Two Replicas When There Is One Shard Set in the JSON File?

Once an index is created, the number of primary shards cannot be changed.

You can run the following command in Kibana to change the number of replicas:

```
PUT /indexname/_settings
{
"number_of_replicas" :1       //Number of replicas
}
```

NOTE

index specifies the index name. Set this parameter based on site requirements.

11.4.5 What Are the Impacts If an Elasticsearch Cluster Has Too Many Shards?

1. A large number of shards in a cluster slows down shard creation.
2. If automatic index creation is enabled, slow index creation may cause a large number of write requests to be stacked in the memory or result in a cluster break down.
3. If there are too many shards and you cannot properly monitor workloads, the number of records in a single shard may exceed the threshold, and write requests may be denied.

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

11.4.6 How Do I Set the Default Maximum Number of Records Displayed on a Page for an Elasticsearch Cluster

Solution

- Method 1

  Open Kibana and run the following commands on the **Dev Tools** page:

  ```
  PUT _all/_settings?preserve_existing=true
  {
  "index.max_result_window" : "10000000"
  }
  ```

- Method 2

  Run the following commands in the background:

  ```
  curl –XPUT 'http://localhost:9200/_all/_setting?preserve_existing=true'-d
  {
  "index.max_result_window":"1000000"
  }
  ```

CAUTION

This configuration consumes memory and CPU resources. Exercise caution when setting this parameter.

11.4.7 Why Does the Disk Usage Increase After the delete_by_query Command Was Executed to Delete Data?

Running the **delete_by_query** command can only add a deletion mark to the target data instead of really deleting it. When you search for data, all data is searched and the data with the deletion mark is filtered out.

The space occupied by an index with the deletion mark will not be released immediately after you call the disk deletion API. The disk space is released only when the segment merge is performed next time.

Querying the data with deletion mark occupies disk space. In this case, the disk usage increases when you run the disk deletion commands.

11.4.8 How Do I Clear the Cache of a CSS Cluster?

- **Clear the fielddata**

  During aggregation and sorting, data are converted to the fielddata structure, which occupies a large amount of memory.

  a. Run the following commands on Kibana to check the memory occupied by index fielddata:

  ```
  DELETE /_search/scroll
  {
  "scroll_id" :
  "DXF1ZXJ5QW5kRmV0Y2gBAAAAAAAAAD4WYm9laVYtZndUQlNsdDcwakFMNjU1QQ=="
  }
  ```

  b. If the memory usage of fielddata is too high, you can run the following command to clear fielddata:

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

```
POST /test/_cache/clear?fielddata=true
```

In the preceding command, *test* indicates the name of the index whose fielddata occupies a large amount of memory.

- **Clear segments**

  The FST structure of each segment is loaded to the memory and will not be cleared. If the number of index segments is too large, the memory usage will be high. You are advised to periodically clear the segments.

  a. Run the following command on Kibana to check the number of segments and their memory usage on each node:
  ```
  GET /_cat/nodes?v&h=segments.count,segments.memory&s=segments.memory:desc
  ```
  b. If the memory usage of segments is too high, you can delete or disable unnecessary indexes, or periodically combine indexes that are not updated.

- **Clear the cache**

  Run the following command on Kibana to clear the cache:
  ```
  POST _cache/clear
  ```

11.4.9 The Average Memory Usage of an Elasticsearch Cluster Reaches 98%

Symptom

The cluster monitoring result shows that the average memory usage of a cluster is 98%. Does it affect cluster performance?

Possible Cause

In an ES cluster, 50% of the memory is occupied by Elasticsearch and the other 50% is used by Lucene to cache files. It is normal that the average memory usage reaches 98%.

Solution

You can monitor the cluster memory usage by checking the maximum JVM heap usage and average JVM heap usage.

11.5 Component Usage

11.5.1 Can I Install Search Guard on CSS?

CSS does not currently support installation of Search Guard.

CSS provides clusters in security mode, which have the same functions as Search Guard.

11.6 Kibana Usage

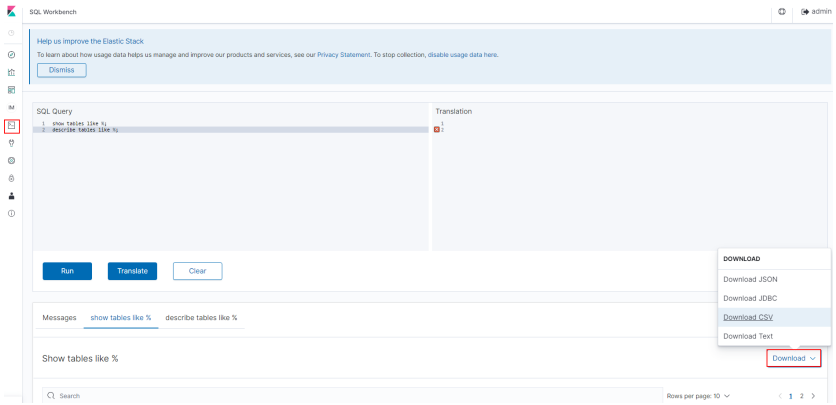
 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

11.6.1 Can I Export Data from Kibana?

Exporting data from Kibana requires the SQL Workbench plugin. Currently, you can only export data from Kibana 7.6.2 or later.

In SQL Workbench of Kibana, you can enter Elasticsearch SQL statements to query data or click **Download** to export data. You can export 1 to 200 data records. By default, 200 data records are exported.

Figure 11-7 SQL Workbench



11.6.2 How Do I Query Index Data on Kibana in an ES Cluster?

Run the following command to query index data through an API on Kibana:

```
GET indexname/_search
```

The returned data is shown in the following figure.

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

Figure 11-8 Returned data


```
{
  "took": 5,
  "timed_out": false,
  "_shards": {
    "total": 5,
    "successful": 5,
    "skipped": 0,
    "failed": 0
  },
  "hits": {
    "total": 3,
    "max_score": 2.0794415,
    "hits": [
      {
        "_index": "book",
        "_type": "novel",
        "_id": "7",
        "_score": 2.0794415,
        "_source": {
          "author": "[illegible]",
          "title": "Elasticsea[illegible]",
          "word_count": 3000,
          "publish_date": "2017-10-01"
        }
      },
      {
```


NOTE

- **took**: How many milliseconds the query cost.
- **time_out**: Whether a timeout occurred.
- **_shard**: Data is split into five shards. All of the five shards have been searched and data is returned successfully. No query result fails to be returned. No data is skipped.
- **hits.total**: Number of query results. Three documents are returned in this example.
- **max_score**: Score of the returned documents. The document that is more relevant to your search criteria would have a higher score.
- **hits.hits**: Detailed information of the returned documents.

11.7 Cluster Access

11.7.1 Why Does My ECS Fail to Connect to a Cluster?

Perform the following steps to troubleshoot this problem:

1. Check whether the ECS instance and cluster are in the same VPC.
   - If they are, go to 2.
   - If they are not, create an ECS instance and ensure that the ECS instance is in the same VPC as the cluster.
2. View the security group rule setting of the cluster to check whether port **9200** (TCP protocol) is allowed or port **9200** is included in the port range allowed in both the outbound and inbound directions.

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

- If it is allowed, go to 3.
- If it is not allowed, switch to the VPC management console and configure the security group rule of the cluster to allow port **9200** in both the outbound and inbound directions.

3. Check whether the ECS instance has been added to a security group.
   - If the instance has been added to a security group, check whether the security group configuration rules are appropriate. You can view the **Security Group** information on the **Basic Information** tab page of the cluster. Then, go to step 4.
   - If the instance has not been added to the security group, go to the VPC page from the ECS instance details page, select a security group, and add the ECS to the group.
4. Check whether the ECS instance can connect to the cluster.

   ssh *<Private network address and port number of a node>*

   NOTE

   If the cluster contains multiple nodes, check whether the ECS can be connected to each node in the cluster.

   - If the connection is normal, the network is running properly.
   - If the connection still fails, contact technical support.

11.7.2 Can a New Cluster Use the IP Address of an Old Cluster?

Question: Can a new cluster use the IP address of an old cluster?

Answer: No.

11.7.3 Can I Associate My EIP If I Want to Make a Cluster Accessible from the Internet?

Question: Can I associate my EIP if I want to make a cluster accessible from the Internet?

Answer: No. If you need to access a cluster through a public network, see section "Public IP Address Access".

11.7.4 Can I Use x-pack-sql-jdbc to Access CSS Clusters and Query Data?

Question: Can I use x-pack-sql-jdbc to access CSS clusters and query data?

Answer: No. Currently, CSS has not integrated the x-pack component.

11.8 Port Usage

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

11.8.1 Do Ports 9200 and 9300 Both Open?

Yes. Port 9200 is used by external systems to access CSS clusters, and port 9300 is used for communication between nodes.

The methods for accessing port 9300 are as follows:

- If your client is in the same VPC and subnet with the CSS cluster, you can access it directly.
- If your client is in the same VPC with but different subnet from the CSS cluster, apply for a route separately.
- If your client is in the different VPCs and subnets from the CSS cluster, create a VPC peering connection to enable communication between the two VPCs, and then apply for routes to connect the two subnets.

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

A Change History

Released On	Description
2024-10-30	Modified: • **Index Backup and Restoration** • **Customizing Word Dictionaries**
2024-04-30	Changed incorrect words. Deleted: • 10.2.4-(Optional) Interconnecting with a Dedicated Load Balancer • 10.6.9-Changing AZs
2023-10-30	Added the following topics: • 10.2.4-(Optional) Interconnecting with a Dedicated Load Balancer • **Changing the Security Mode** • 10.6.9-Changing AZs • **Using the Open Distro SQL Plugin** • **Monitoring Metrics of Nodes**
2023-05-30	Added the following topics: • Product Components • Quotas • **Creating an IAM User and Granting Permissions** • **Deploying a Cross-AZ Cluster** • **Overview** • **Scaling Out a Cluster** • **Scaling in a Cluster** • **Adding Master/Client Nodes** • **Vector Retrieval** • **Kibana Platform**

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.

Released On	Description
2023-01-30	Deleted the description about word dictionary management API.
2022-10-12	This issue is the fourth official release.
2022-05-30	This issue is the third official release. • Added **Removing Specified Nodes**. • Added **Storage-Compute Decoupling**. • Added **Flow Control**. • Added **Large Query Isolation**. • Added **Index Monitoring**.
2021-10-09	This issue is the second official release.
2021-01-30	This issue is the first official release.

 Copyright © Huawei Cloud Computing Technologies Co., Ltd.