ALIBABA CLOUD

Alibaba Cloud

Elasticsearch

FAQ

Document Version: 20220128

Alibaba Cloud

Legal disclaimer

Alibaba Cloud reminds you to carefully read and fully understand the terms and conditions of this legal disclaimer before you read or use this document. If you have read or used this document, it shall be deemed as your total acceptance of this legal disclaimer.

1. You shall download and obtain this document from the Alibaba Cloud website or other Alibaba Cloud-authorized channels, and use this document for your own legal business activities only. The content of this document is considered confidential information of Alibaba Cloud. You shall strictly abide by the confidentiality obligations. No part of this document shall be disclosed or provided to any third party for use without the prior written consent of Alibaba Cloud.

2. No part of this document shall be excerpted, translated, reproduced, transmitted, or disseminated by any organization, company or individual in any form or by any means without the prior written consent of Alibaba Cloud.

3. The content of this document may be changed because of product version upgrade, adjustment, or other reasons. Alibaba Cloud reserves the right to modify the content of this document without notice and an updated version of this document will be released through Alibaba Cloud-authorized channels from time to time. You should pay attention to the version changes of this document as they occur and download and obtain the most up-to-date version of this document from Alibaba Cloud-authorized channels.

4. This document serves only as a reference guide for your use of Alibaba Cloud products and services. Alibaba Cloud provides this document based on the "status quo", "being defective", and "existing functions" of its products and services. Alibaba Cloud makes every effort to provide relevant operational guidance based on existing technologies. However, Alibaba Cloud hereby makes a clear statement that it in no way guarantees the accuracy, integrity, applicability, and reliability of the content of this document, either explicitly or implicitly. Alibaba Cloud shall not take legal responsibility for any errors or lost profits incurred by any organization, company, or individual arising from download, use, or trust in this document. Alibaba Cloud shall not, under any circumstances, take responsibility for any indirect, consequential, punitive, contingent, special, or punitive damages, including lost profits arising from the use or trust in this document (even if Alibaba Cloud has been notified of the possibility of such a loss).

5. By law, all the contents in Alibaba Cloud documents, including but not limited to pictures, architecture design, page layout, and text description, are intellectual property of Alibaba Cloud and/or its affiliates. This intellectual property includes, but is not limited to, trademark rights, patent rights, copyrights, and trade secrets. No part of this document shall be used, modified, reproduced, publicly transmitted, changed, disseminated, distributed, or published without the prior written consent of Alibaba Cloud and/or its affiliates. The names owned by Alibaba Cloud shall not be used, published, or reproduced for marketing, advertising, promotion, or other purposes without the prior written consent of Alibaba Cloud. The names owned by Alibaba Cloud include, but are not limited to, "Alibaba Cloud", "Aliyun", "HiChina", and other brands of Alibaba Cloud and/or its affiliates, which appear separately or in combination, as well as the auxiliary signs and patterns of the preceding brands, or anything similar to the company names, trade names, trademarks, product or service names, domain names, patterns, logos, marks, signs, or special descriptions that third parties identify as Alibaba Cloud and/or its affiliates.

6. Please directly contact Alibaba Cloud for any errors of this document.

Document conventions

Style	Description	Example
Danger	A danger notice indicates a situation that will cause major system changes, faults, physical injuries, and other adverse results.	**Danger:** Resetting will result in the loss of user configuration data.
Warning	A warning notice indicates a situation that may cause major system changes, faults, physical injuries, and other adverse results.	**Warning:** Restarting will cause business interruption. About 10 minutes are required to restart an instance.
Notice	A caution notice indicates warning information, supplementary instructions, and other content that the user must understand.	**Notice:** If the weight is set to 0, the server no longer receives new requests.
Note	A note indicates supplemental instructions, best practices, tips, and other content.	**Note:** You can use Ctrl + A to select all files.
>	Closing angle brackets are used to indicate a multi-level menu cascade.	Click **Settings**> **Network**> **Set network type**.
Bold	Bold formatting is used for buttons , menus, page names, and other UI elements.	Click **OK**.
`Courier font`	Courier font is used for commands	Run the `cd /d C:/window` command to enter the Windows system folder.
Italic	Italic formatting is used for parameters and variables.	`bae log list --instanceid` *Instance_ID*
[] or [a|b]	This format is used for an optional value, where only one item can be selected.	`ipconfig [-all|-t]`
{} or {a|b}	This format is used for a required value, where only one item can be selected.	`switch {active|stand}`

Table of Contents

1.Imbalanced loads on a cluster

Imbalanced loads on an Elasticsearch cluster

Imbalanced loads on an Alibaba Cloud Elasticsearch cluster may be caused by several reasons. These reasons include inappropriate shard settings, uneven segment sizes, unseparated hot and cold data, and persistent connections that are used for Service Load Balancer (SLB) instances and multi-zone architecture. This topic describes the analysis and solutions for imbalanced loads on an Elasticsearch cluster.

Problem description

- Disk usage is similar between nodes, but the values of the NodeCPUUtilization or NodeLoad_1m metric indicate imbalanced loads.
- Disk usage is significantly different between nodes, and the values of the NodeCPUUtilization or NodeLoad_1m metric indicate imbalanced loads.

Causes

- Shard allocation is inappropriate.

  > **Notice** In most cases, imbalanced loads are caused by inappropriate shard allocation. We recommended that you first check shard allocation.

- Segment sizes are uneven.
- Hot data and cold data are not separated on nodes.

  > **Notice** For example, if you add the routing parameter in queries or query hot data, imbalanced loads occur.

- Persistent connections are used for SLB instances and multi-zone architecture. In this case, traffic may be unevenly distributed to nodes. However, this rarely occurs. For more information, see Multi-zone architecture.

> **Notice** If imbalanced loads occur due to other reasons, contact Alibaba Cloud technical support engineers for troubleshooting.

Inappropriate shard allocation

- Scenario

  Company A has purchased an Alibaba Cloud Elasticsearch cluster. The cluster contains three dedicated master nodes and nine data nodes. Each dedicated master node offers 16 vCPUs and 32 GiB of memory. Each data node offers 32 vCPUs and 64 GiB of memory. Major data is stored in the test index. During peak hours (16:21 to 18:00), the read performance is about 2,000 QPS and the write performance is 1,000 QPS. Both cold and hot data are queried. In addition, the CPU utilization of two nodes reaches 100%, which affects the Elasticsearch service.

- Analysis

  i. Check the network and Elastic Compute Service (ECS) instances. If ECS instances are normal, view network monitoring data.

The network monitoring data shows that the number of network requests increases during the peak hours. At the same time, the QPS of read operations increases, and the CPU utilization of related nodes significantly increases. Based on the preceding information, you can conclude that the nodes with high loads are mainly used to process query requests.

ii. Run the `GET _cat/shards?v` command to query the shards of the index.

The command output shows that nodes with high loads contain a large number of shards. This indicates that shards are not evenly allocated to nodes. In addition, the monitoring data for disk usage shows that the disk usage of nodes with high loads is greater than that of other nodes. You can conclude that the uneven allocation of shards results in uneven storage. When you query or write data, nodes with high storage handle major workloads.

iii. Run the `GET _cat/indices?v` command to query information of the index.

The command output shows that the index has five primary shards and one replica shard for each primary shard. In addition, cluster configurations indicate that shards are not evenly allocated and specific documents are deleted. When Elasticsearch searches for data, it also searches for and filters documents marked with .del. This significantly reduces search efficiency and consumes additional resources. We recommend that you call the force merge operation during off-peak hours.

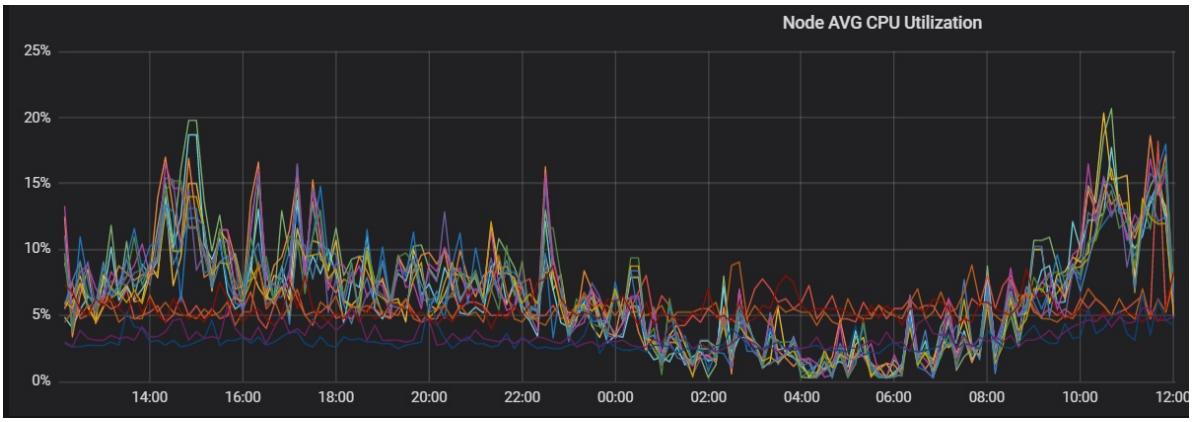
pri	rep	docs.count	docs.deleted	store.size	pri.store.size
1	1	8640	0	6.7mb	3.3mb
1	1	8639	0	6.7mb	3.3mb
1	1	8639	0	6.7mb	3.3mb
1	1	297847	488	688.7mb	344.8mb
1	1	93230	364	222.1mb	111mb
1	1	297803	588	693.5mb	346.1mb
1	1	297756	540	697.5mb	348.8mb
1	1	297852	536	687.9mb	343.4mb
1	1	297750	588	693.1mb	344.2mb
1	1	8639	0	6.7mb	3.3mb
1	1	8639	0	6.7mb	3.3mb
1	1	2	0	19.3kb	8.3kb
3	1	0	0	1.1kb	576b
3	1	71888299	15614684	124gb	64.6gb
3	1	0	0	1.1kb	576b
1	1	2661	0	2.2mb	1.1mb
1	1	8639	0	6.7mb	3.3mb
1	1	8639	0	6.6mb	3.3mb
1	1	297732	540	700.5mb	349.9mb
1	1	297795	540	691.6mb	344.7mb

iv. View cluster logs and search slow logs.

The logs show that the queries are all normal term queries, and the cluster logs do not indicate that an error has occurred. Therefore, the Elasticsearch cluster does not encounter errors and query statements that consume CPU resources.

- Summary

The preceding analysis indicates that the uneven CPU utilization is mainly caused by uneven shard allocation. You must re-allocate shards for the index. Make sure that the total number of primary shards and replica shards is a multiple of data nodes in the cluster. After the optimization, CPU utilization is not significantly different between nodes. The following figure shows the CPU utilization.



- Solution

  Plan shards properly when you create indexes. For more information, see Shard evaluation guidelines.

Shard evaluation guidelines

Both the number of shards and size of each shard contribute to the stability and performance of an Elasticsearch cluster. You must properly plan shards for each index of an Elasticsearch cluster. This prevents numerous shards from affecting cluster performance when it is difficult to define business scenarios. This section provides the following guidelines:

> **Note** In versions earlier than Elasticsearch V7.X, one index has five primary shards and one replica shard for each primary shard by default. In Elasticsearch V7.X and later, one index has one primary shard and one replica shard by default.

- For nodes with low specifications, the size of each shard is no more than 30 GB. For nodes with high specifications, the size of each shard is no more than 50 GB.
- For log analysis or extremely large indexes, the size of each shard is no more than 100 GB.
- The total number of primary shards and replica shards is the same as or a multiple of the number of data nodes.

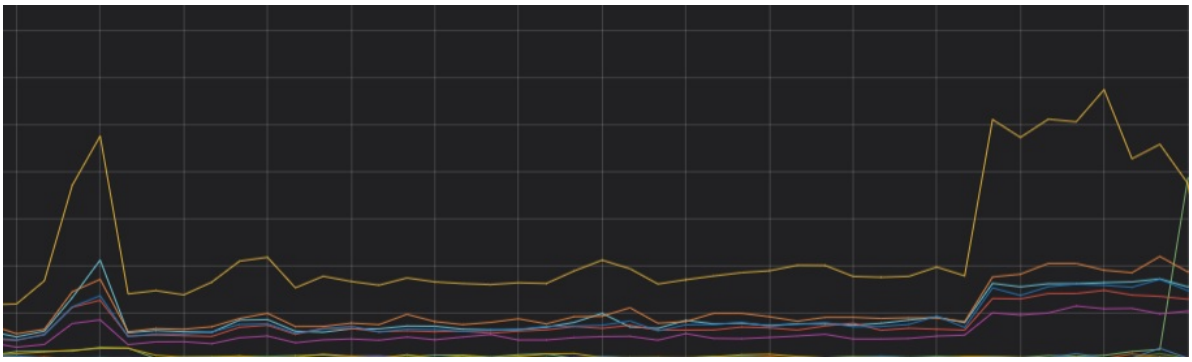
  > **Note** The more shards, the more performance overheads of your Elasticsearch cluster.

- You can determine the number of shards on a single node based on the memory size multiplied by 30. If a large number of shards are planned, file handle exhaustion can easily occur and result in cluster failures.
- You can configure a maximum of five primary shards for an index on a node.

Uneven segment sizes

- Scenario

A node in the Elasticsearch cluster of Company A experiences an abrupt increase in CPU utilization. This affects query performance. Queries are mainly performed on the test index. The index has three primary shards and one replica shard for each primary shard. The shards are evenly allocated to nodes. The index contains a large number of documents that are marked with docs.deleted. In addition, you have checked that ECS instances are normal.



- Analysis
  i. Add `"profile": true` to the query body.

     The query results show that Elasticsearch requires a longer time to query Shard 1 of the test index than other shards.

  ii. Send a query request with `preference` set to _primary and `"profile": true` added to the query body, and view the time required to query the primary shard. Then, send another query request with `preference` set to _replica and "profile": true added to the query body, and view the time required to query the replica shard.

     The time required to query Shard 1 (primary shard) is longer than that required to query its replica shard. This indicates that imbalanced loads are caused by Shard 1.

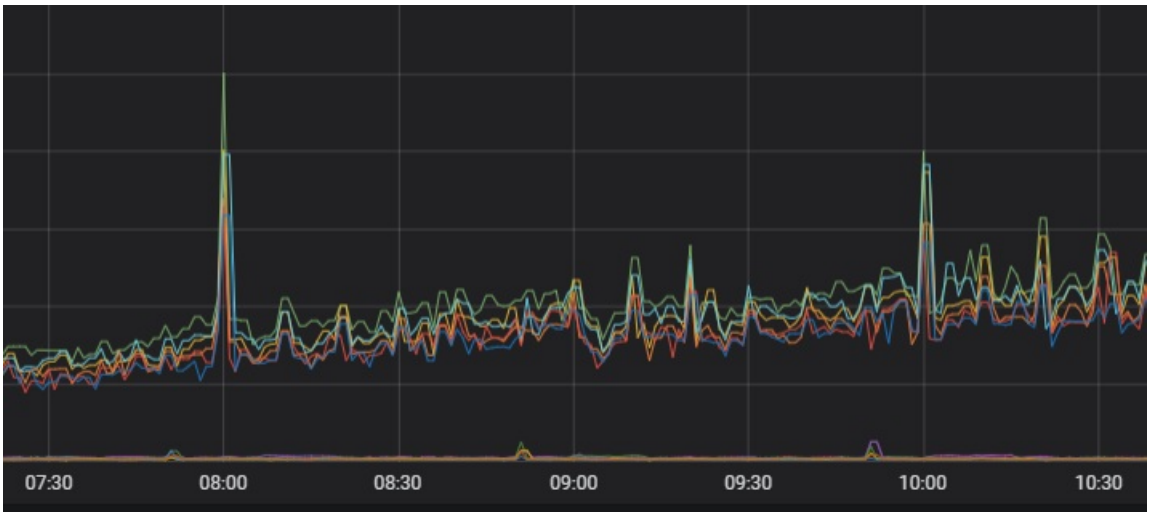
  iii. Run the `GET _cat/segments/index?v&h=shard,segment,size,size.momery,ip` and `GET _cat/shards?v` commands to query the information of Shard 1.

     The command outputs show that Shard 1 contains large segments and the number of documents in the shard is greater than that in its replica shard. Based on the preceding information, you can determine that imbalanced loads are caused by uneven segment sizes.

> **Note** The inconsistency in the number of documents is caused by many different reasons. Examples:
>
> - A latency exists in data synchronization between primary and replica shards. If documents are continuously written to the primary shard, data inconsistency may occur. However, after you stop writing documents, the number of documents is the same between the primary shard and its replica shard.
> - A primary shard forwards requests to its replica shards after data is written to the primary shard. If you use automatically generated document IDs to write documents to a primary shard, you cannot perform delete operations on the primary shard during the write operation. If you perform a delete operation (such as sending a Delete by Query request to delete the document that you have just written), the operation is also performed on the replica shard. Then, the primary shard forwards the write request to the replica shard. As the document ID is automatically generated by the system, the document is written to the replica shard without verification. As a result, the number of documents in the replica shard differs from that in the primary shard. In addition, the primary shard includes a large number of documents that are marked with `docs.deleted` .

- Solution
  - Solution 1: During off-peak hours, call the force merge operation to merge small segments and remove documents that are marked with docs.deleted.
  - Solution 2: Restart the node where the primary shard resides to promote the replica shard to a primary shard. Use the new primary shard to generate a new replica shard. This ensures that the segments in the new primary and replica shards are the same.

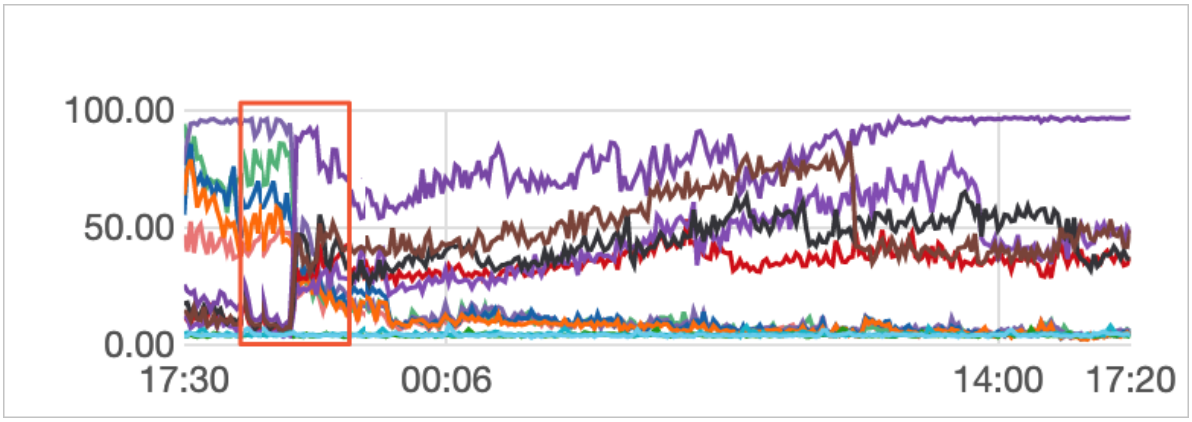
  The following figure shows the loads after optimization.



Multi-zone architecture

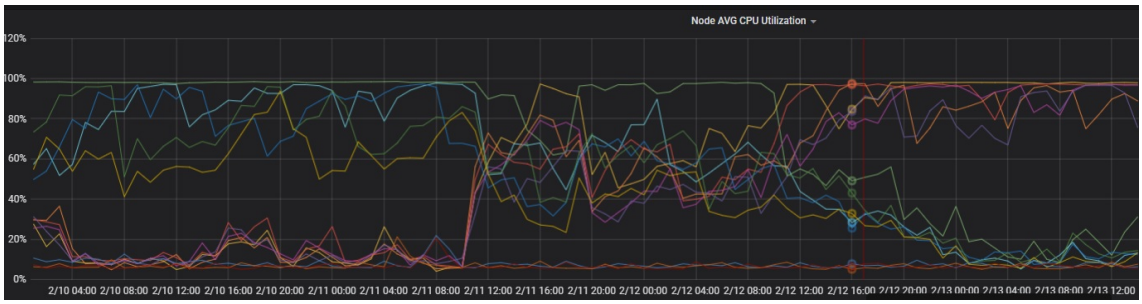
- Scenario

To deploy an Elasticsearch cluster across zones, Company A deploys the multi-zone architecture in both Zone B and Zone C. When the cluster provides services, the loads of nodes in Zone C are higher than the loads of nodes in Zone B. You have checked that the imbalanced loads are not caused by hardware or uneven data distribution.



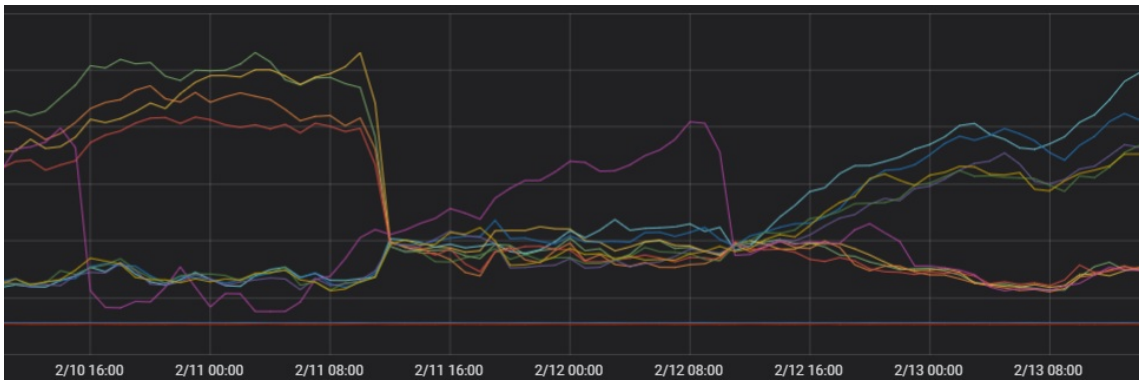
- Analysis
  i. View the CPU utilization of nodes in the two zones within the last four days.

     The monitoring data shows that the CPU utilization of the nodes significantly changed.



  ii. View the TCP connections to the nodes.

     The monitoring data shows that the number of TCP connections in the two regions significantly differs. This indicates that the imbalanced loads are caused by network connections.

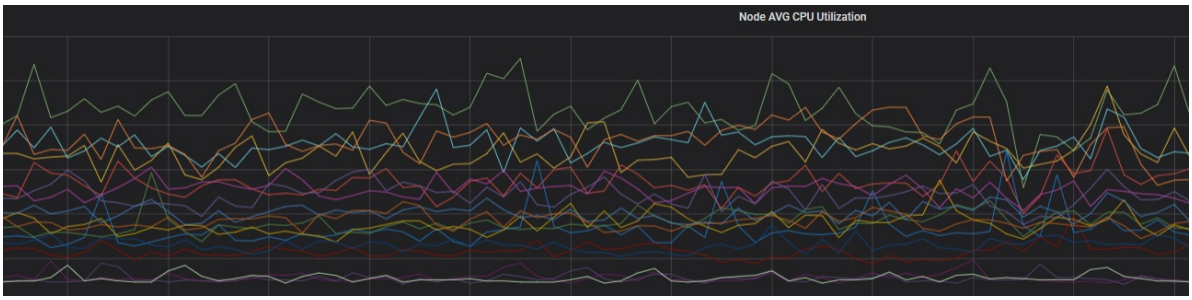


  iii. Check client connections.

The client uses persistent connections and establishes a small number of new connections. This scenario is at risk of independent scheduling for a multi-zone network. Network services are independently scheduled based on the number of connections. Each scheduling unit selects the optimal node to create a connection. Independent scheduling provides higher performance. However, if the number of new connections is small, most scheduling units may choose the same node to establish a connection. A client node of an Elasticsearch cluster first forwards requests to another node that resides in the same zone. This causes imbalanced loads between zones.

- Solution
  - Solution 1: Use independent client nodes to forward complex traffic. In this case, data nodes are not affected even though the client nodes are heavily loaded. This reduces the risk of load imbalance.
  - Solution 2: Configure two independent domain names on the client to ensure balanced traffic on the client. This solution cannot ensure high availability. When you upgrade the configuration of an Elasticsearch cluster, access failures may occur because nodes are not removed from the SLB instance.

The following figure shows the loads after optimization.



2.High disk usage and read-only indexes

If the disk usage of your Alibaba Cloud Elasticsearch cluster exceeds 85%, the cluster or Kibana may not provide services. This topic describes how to resolve this issue.

> **Notice** Disclaimer: This topic may contain information about third-party products. Such information is only for reference. Alibaba Cloud does not make any guarantee, express or implied, with respect to the performance and reliability of third-party products, as well as potential impacts of operations on the products.

Problem description

- After the system receives an index request, it returns an error message similar to `index read_only`, such as `FORBIDDEN/12/index read-only / allow delete (api)];]`.
- The cluster is in a state that is indicated by the color red. In severe cases, some nodes do not join the cluster. You can run the `GET _cat/nodes?` command to view the nodes in the cluster. In addition, some shards are not allocated to nodes. You can run the `GET _cat/allocation?v` command to view the allocation of shards.

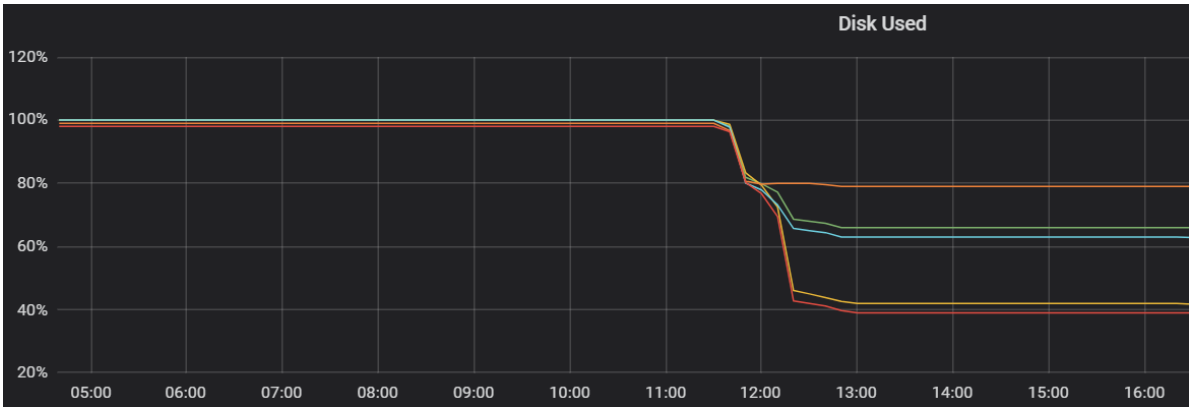
  > **Note** If a cluster is in a state that is indicated by the color red, the primary shards of the cluster are unavailable, and data on the cluster may be lost.

- When a pipeline is created or a Beat is enrolled in the Kibana console, the `internal server error` message is returned.
- On the Cluster Monitoring page of the cluster or the Monitoring page in the Kibana console of the cluster, the disk usage has reached 100% recently.

Cause

The preceding issues are caused by high disk usage. The disk usage of nodes has the following thresholds:

- 85%: If the disk usage of a node exceeds 85%, the system no longer allocates new shards to the node.
- 90%: If the disk usage of a node exceeds 90%, the system migrates the shards on the node to other data nodes with low disk usage.
- 95%: If the disk usage of a node exceeds 95%, the system forcibly adds the `read_only_allow_delete` attribute to all indexes in the cluster. As a result, data cannot be written to the indexes, and you can only read data from the indexes or delete the indexes.



Solution

1. Run the following command to delete data:

   > **Warning** Deleted data cannot be restored. Proceed with caution. You can also retain the data, but you must resize disks. For more information, see Upgrade the configuration of a cluster.

   ```
   curl -u <username>:<password> -XDELETE  http://<host>:<port>
   ```

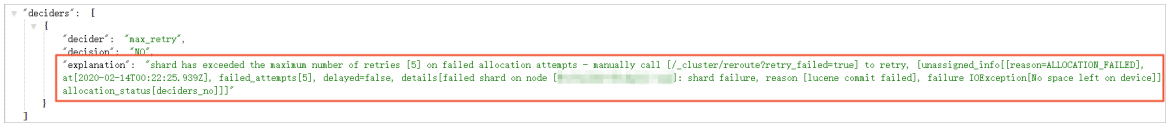
   - Set `<host>` to the internal or public endpoint of the cluster. We recommend that you configure the related whitelist before you run this command.
   - If the cluster has no response after you run the preceding command, we recommend that you trigger a forced restart and try to run this command during the restart.

2. Check whether indexes are still read-only. If they are, run the following command to set the `index.blocks.read_only_allow_delete` attribute to `null` for all indexes to ensure that all indexes on the cluster are not read-only:

   ```
   PUT _settings
   {
      "index.blocks.read_only_allow_delete": null
   }
   ```

3. Check whether the cluster is still in a state that is indicated by the color red. If it is, run the `_cat/allocation?v` command to check whether the cluster contains shards that are not allocated.

4. If the cluster contains shards that are not allocated, run the `GET _cluster/allocation/explain` command to view the reason. If the reason is similar to that shown in the following figure, run the `POST /_cluster/reroute?retry_failed=true` command.

5. After shards are allocated, view the cluster status. If the cluster is still in a state that is indicated by the color red, contact Alibaba Cloud technical support engineers.

Additional information

To avoid the impact of high disk usage on Alibaba Cloud Elasticsearch, we recommend that you enable disk usage monitoring and alerting. In addition, you must view the alerting text message in time and take appropriate measures in advance. For more information, see Configure the monitoring and alerting feature in Cloud Monitor.

3.Installation failures of Beats shippers

This topic describes how to troubleshoot installation failures and abnormal heartbeats of a Beats shipper.

Procedure

1. Check whether the Elastic Compute Service (ECS) instance on which Beats is installed runs Aliyun Linux, Red Hat Enterprise Linux (RHEL), or CentOS.
2. Check whether the ECS instance on which Beats is installed resides in the same virtual private cloud (VPC) as your Alibaba Cloud Elasticsearch or Logstash cluster.
3. Check whether Cloud Assistant and Docker are installed on the ECS instance on which Beats is installed.

   To perform the check, connect to the ECS instance and run the following commands:

   - Check the status of Cloud Assistant

     ```
     systemctl status aliyun.service
     ```

     If Cloud Assistant is in a normal state, the result shown in the following figure is returned.

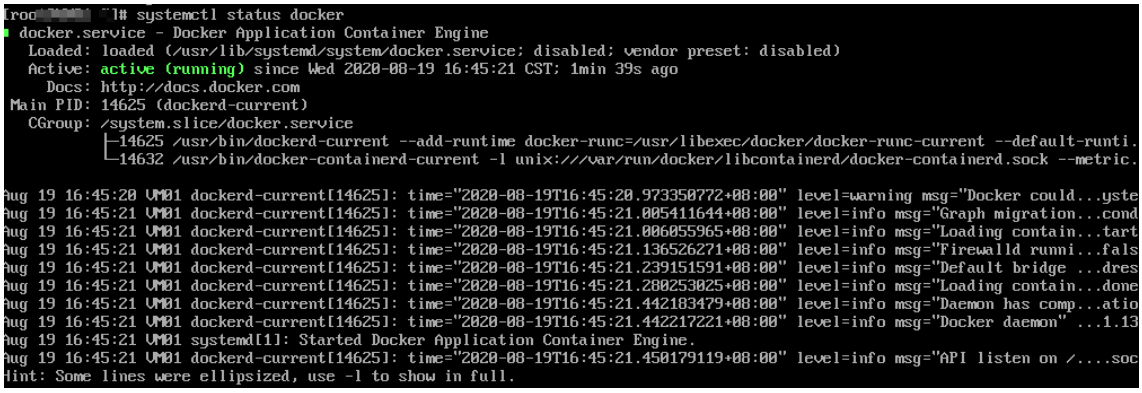
     

     For more information about how to install Cloud Assistant, see Install the Cloud Assistant client.

   - Check the status of Docker

     ```
     systemctl status docker
     ```

     If Docker is in a normal state, the result shown in the following figure is returned.

     For more information about how to install Docker, see Deploy and use Docker on Alibaba Cloud Linux 2 instances.

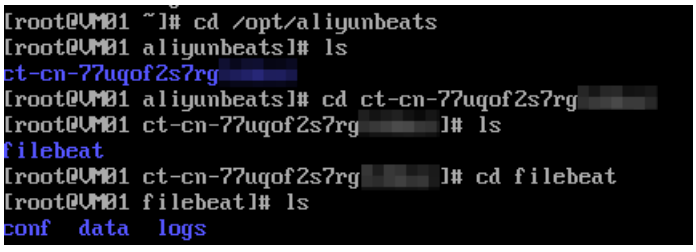
4. Check whether the following parameters are configured in the YML configuration file of the shipper:

```
- type: log
  # Change to true to enable this input configuration.
     enabled: true
  # Paths that should be crawled and fetched. Glob based paths.
  paths:
    - /var/log/*.log
```

Parameter	Description
enabled	Set this parameter to true when you use Beats. Default value: false.
paths	You can use a wildcard to specify this parameter, such as *.log. This facilitates the query of the log file.

> **Notice**
>
> - The value of paths is different from Filebeat Log File Path that you specified on the configuration page. For Filebeat, collected data can be stored in the file indicated by paths only after the path specified by paths is mapped by Docker. However, Docker maps the path specified by Filebeat Log File Path. Therefore, we recommend that you set paths and Filebeat Log File Path to the same value.
> - If you already specify **Output** on the configuration page, you cannot specify it again in the YML configuration file. Otherwise, the system prompts an installation error.
> - Exercise caution when you modify the parameters that are commented out by using number signs (#) in the YML configuration file of the shipper, such as X-Pack-related parameters. Otherwise, the shipper fails to be installed.

5. Connect to the ECS instance, check whether a Beats instance is created under the */opt/aliyunbeats/* path, and check whether the conf, data, and logs folders exist.



You can also query Beats logs in the logs folder to locate problems.

6. Check the status of the Docker container in which the Beats service is provided. Then, locate

problems based on logs.

i. Check the status of the Docker container.

```
docker ps -a | grep filebeat
```

ii. If the Docker container is in the exited state, check the logs that the container generates.

```
docker logs -f Container ID
```

4.Uneven distribution of hot data on nodes

Elasticsearch evenly allocates shards to data nodes and uses hash maps to evenly route documents to shards. In this case, some nodes may store more hot data than others, which results in high loads on these nodes. To address this issue, you can restart your cluster or manually migrate shards to re-allocate shards. This topic describes how to manually migrate shards.

Problem description

- A node with high loads stores a large number of shards that have the same attribute. For example, a node stores only primary shards.
- A node with high loads stores more shards for business indexes than other nodes.

Precautions

- Manual shard migration is only a temporary solution to the high loads of nodes. If some nodes are absent from the cluster for a short time and shards are re-allocated, this issue may occur again. Therefore, before you migrate shards, we recommend that you optimize shard allocation based on the instructions provided in Imbalanced loads on a cluster.
- If the shards that store hot data are evenly allocated, but the cluster runs with high loads, we recommend that you upgrade the configuration of the cluster.

Solution

1. Disable shard allocation.

```
PUT /_cluster/settings
{
  "transient" : {
     "cluster.routing.allocation.enable" : "none"
  }
}
```

> **Notice** The preceding command is used only to temporarily disable shard allocation. After shards are migrated, you must enable shard allocation again. To enable shard allocation, run the PUT /_cluster/settings command with cluster.routing.allocation.enable set to all.

2. Manually migrate shards.

   In this example, Shard 3 of the index_parkingorder_v1 index is migrated from the node whose IP address is 192.168.130.77 to the node whose IP address is 192.168.130.78.

```
POST /_cluster/reroute
{
 "commands" : [
  {
   "move" : {
    "index" : "index_parkingorder_v1",
    "shard" : 3,
    "from_node" : "192.168.130.77",
    "to_node" : "192.168.130.78"
   }
  }
 ]
}
```

> **Note**
>
> - When you migrate a shard of an index to a node, make sure that the node does not store shards that belong to the same index or have the same sequence number. For example, Replica Shard 3 of an index exists on Node A. In this case, Primary Shard 3 of the index cannot be migrated to Node A.
> - For more information about how to manually migrate shards, see cluster-reroute.

3. Check the shard allocation status.

```
GET _cat/shards?v
```

If the command is successfully run, the result shown in the following figure is returned.

```
index      shard prirep state         docs   store ip         node
qosgeoname 4     r      STARTED    2279054 623.8mb 10.6.      y3m8a_2
qosgeoname 4     p      STARTED    2279054 628.3mb 10.6.      Xwhpjaf
qosgeoname 3     r      RELOCATING 2279002 627.3mb 10.6.      4Mh39az -> 10.6.      y3m8a_2tTdi_16TAy7B91w y3m8a_2
qosgeoname 3     p      STARTED    2279002 626.4mb 10.6.      Xwhpjaf
qosgeoname 1     r      STARTED    2278983 622.3mb 10.6.      I8SRFdj
qosgeoname 1     p      STARTED    2278983 621.6mb 10.6.      Qn3qSgY
qosgeoname 2     p      STARTED    2280694   623mb 10.6.      Qn3qSgY
qosgeoname 2     r      STARTED    2280694 622.7mb 10.6.      4Mh39az
qosgeoname 0     p      STARTED    2278770 624.7mb 10.6.      I8SRFdj
qosgeoname 0     r      STARTED    2278770 624.3mb 10.6.      y3m8a_2
```

4. After the shard is migrated, enable shard allocation.

```
PUT /_cluster/settings
{
 "transient" : {
  "cluster.routing.allocation.enable" : "all"
 }
}
```