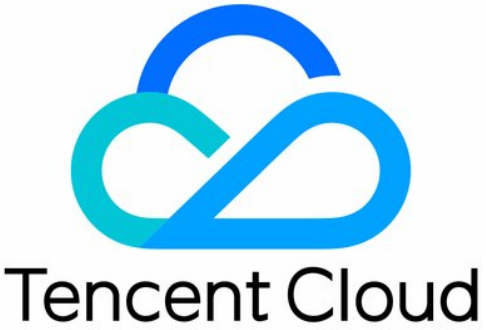
Elasticsearch Service

Elasticsearch Guide



Copyright Notice

Trademark Notice

Service Notice

This document provides an overview of the as-is details of Tencent Cloud's products and services in their entirety or part. The descriptions of certain products and services may be subject to adjustments from time to time.

The commercial contract concluded by you and Tencent Cloud will provide the specific types of Tencent Cloud products and services you purchase and the service standards. Unless otherwise agreed upon by both parties, Tencent Cloud does not make any explicit or implied commitments or warranties regarding the content of this document.

Contact Us

We are committed to providing personalized pre-sales consultation and technical after-sale support. Don't hesitate to contact us at 4009100100 or 95716 for any inquiries or concerns.

Contents

Elasticsearch Guide
Service Overview

Last updated: 2024-11-12 21:28:43

> **Note**
> Starting from August 1, 2023, new users who activate the ES Serverless Service can claim a free 50 Yuan Voucher to offset the fees incurred from using the Serverless Service. The validity period is 90 days from the date of claim. Any excess fees will be deducted according to the corresponding billing item pricing.

Industry Challenges

During log analysis using open-source Elasticsearch, users often need to estimate cluster configurations based on write traffic, peak write traffic, and storage duration. This includes CPU, memory, disk size, etc., to ensure smooth operation. However, based on extensive online operations practice, the following issues arise:

- Elasticity capability is not adaptable to business development. In scenarios like promotional activities and holidays, log data exhibits significant peak-valley effects, with high write throughput and high availability requirements. It is challenging to anticipate sudden read-write traffic and scale out a cluster, making it difficult to ensure Elasticsearch cluster stability.
- Resource costs are high. Insufficient resources affect traffic writing during peak periods, while cluster capacity planning based on peak period traffic leads to resource redundancy and waste during off-peak periods, increasing costs significantly.
- High operation and management costs. Businesses need to plan and configure clusters, index configurations, and build monitoring and alerting platforms themselves. In the context of cost reduction and efficiency improvement, companies have a strong demand for optimizing these costs and hope to further reduce them.

Overview

Elasticsearch Serverless is a cloud-based, fully managed Elasticsearch service created by Tencent Cloud based on its proprietary cloud-native Serverless technology architecture. It offers auto-scalability and a completely maintenance-free product capability, effectively solving issues like resource costs caused by fluctuating demand in business scenarios such as log analysis and metric monitoring. Additionally, it is fully compatible with the ELK ecosystem, featuring end-to-end data ingestion, data management, and data visualization product features, providing a ready-to-use product experience.

At the Enterprise Cloud Adoption and Computing–Cloud Convergence Industry Conference held on March 29, 2023, the Tencent Cloud Elasticsearch Serverless service was awarded the Outstanding Case Award by the China Academy of Information and Communications Technology "2022 Trusted Computing Power Service · Pioneer Program".

Benefits and features

- Automatic Elasticity: Automatic index–granularity AS to effortlessly handle sudden traffic increases, ensuring business continuity while reducing operations and management costs incurred during peak and trough periods in scenarios such as log analysis and observability.
- Completely zero maintenance: Built–in sharding, automatic tuning, intelligent lifecycle management, and self–healing capabilities. Users can create and use indexes on demand without worrying about underlying resource configuration, cluster scaling, or index settings. The entire process is completely maintenance–free.
- Ultimate cost: Self–developed low–cost, high–performance, high–availability separated storage and computing architecture. Billing is based on actual access and storage volume, achieving dynamic matching of business loads and resources with on–demand payment. This reduces redundant costs from idle resources, significantly lowering the costs.
- Flexible and easy to use: Provides end–to–end one–stop product capabilities from data access to data management and then to data analysis and exploration, greatly reducing the barrier for businesses to migrate to the cloud. Users can implement business operations at the minute level.
- Open integration: Fully compliant with the ELK ecosystem, retaining users' original usage habits and achieving seamless migration, facilitating rapid cloud migration for businesses. At the same time, it connects cloud data sources (such as cloud services CVM, TKE), lowering the data access barrier and enabling minute–level business implementation.
- Stable and Reliable: Cluster Configuration, read/write performance is optimized by the backend, reducing issues caused by improper use, enhancing stability, and ensuring business protection.

Contact Us

Scan the code to join Tencent Cloud Big Data Elasticsearch Serverless community group. Activities and exquisite gifts are distributed periodically.

Basic Concepts

Last updated: 2024-10-25 09:06:32

This article introduces the basic concepts related to project space and index in the ES Serverless service.

Project space

The project space is a basic resource unit in the ES Serverless service. You can create indexes under the same project space for the same business, making index management easier. When reading and writing data, you can access indexes in that space through the project space's access address and username password information.

Index

In the ES Serverless service, the index is the smallest unit of data storage and management. It uses Tencent Cloud ES's self-developed Autonomous Indexing capability, built-in Automatic Sharding Optimization, Intelligent Lifecycle Management, and Self-healing Faults. Compared to traditional methods, you don't need to worry about index rolling and shard size issues. You can focus on business data, write queries, and visual analysis.

Upgrade Notice:
ES Serverless service offers a brand-new experience upgrade, supporting unified access address and Kibana management, accessing multiple indexes, and better fitting original usage habits. The differences before and after the upgrade are as follows:

- For project spaces created before January 23, 2024, there is no Independent Access Control feature. Simultaneous access to multiple indices in Kibana is not supported. Writing and querying are done through the index's access address.
- For project spaces created after January 23, 2024, unified management and access of all indices under the space through the project's access address and Kibana are supported. Additionally, you can set Permission Types and Permission Scope via the visual user management feature, making it more consistent with the original ES Cluster usage habits and meeting various scenarios. Upgrading does not require changes to your business code; you only need to migrate existing indices to the new space. We strongly recommend migrating indices to the new space.

5-Minute Quick Experience

Last updated: 2024-10-25 09:06:52

Overview

The ES Serverless service is a cloud-based, one-stop, fully managed ES service built by Tencent Cloud based on its proprietary cloud-native Serverless architecture. It eliminates the concept of clusters, allowing users to create and use indexes as needed. It offers **automatic elasticity and is completely maintenance-free**, effectively solving issues such as high resource costs caused by peaks and troughs in business scenarios like **log analysis and metric monitoring**. Additionally, it is fully compatible with the ELK ecosystem and provides an end-to-end product feature set for data access, management, and visualization, delivering a **plug-and-play log analysis experience**.

Quick Experience

The ES Serverless service supports data write into the index via **ES native API, Logstash, Flink, Kafka,** and other methods. If you have CVM CVMs, , TKE TKES, and Cloud Data Warehouse TCHouse-C log collection requirements, it also supports one-stop visual configuration on the interface. By simply setting data sources and index information, logs can be collected into the index and quickly searched and analyzed. This document will introduce the full process operation of **creating an index > data write > search and analysis,** giving you a simple experience of the ES Serverless service in log analysis scenarios.

Basic Concepts

Before we officially start the experience, let's introduce some basic concepts that will be used in this experience:

Name	Introduction
Project Space	The Project Space is the basic resource unit in the ES Serverless service. You can create indices under the same business in the same project space for easier index management.
Index	The index is the minimum granularity unit of data storage and management, providing log storage and near real-time query capability. We can store the collected log data in the index.

Kibana	Kibana is a data analysis and visualization platform integrated with ES. We can perform log writing, retrieval, and chart construction (e.g., maps, line charts) in Kibana.
Log	Logs are records produced during the running of an application system, such as operation logs, access logs, error logs, etc.

Create a space

1. Receive 50 yuan voucher with no threshold.

> **Note:**
> Assuming 4GB of raw log data is written each day and stored for 7 days, this voucher supports up to 1 month of continuous use.

2. Log in to ES Serverless Console.
3. In the space list, click **Create a new space** to enter the Create New Space page.
4. On the Create New Space page, we need to set the following information:

- **Space Name**: This name is used to identify the space, following the naming conventions indicated on the page.
- **Network/Availability Zone and Subnet**: The project space is created under a VPC to ensure secure access. You can select the corresponding VPC, Availability Zone, and Subnet. If you need to create a new one, refer to Create VPC, Create New Subnet.

Search for the required CAM policy as needed, and click to complete policy association.

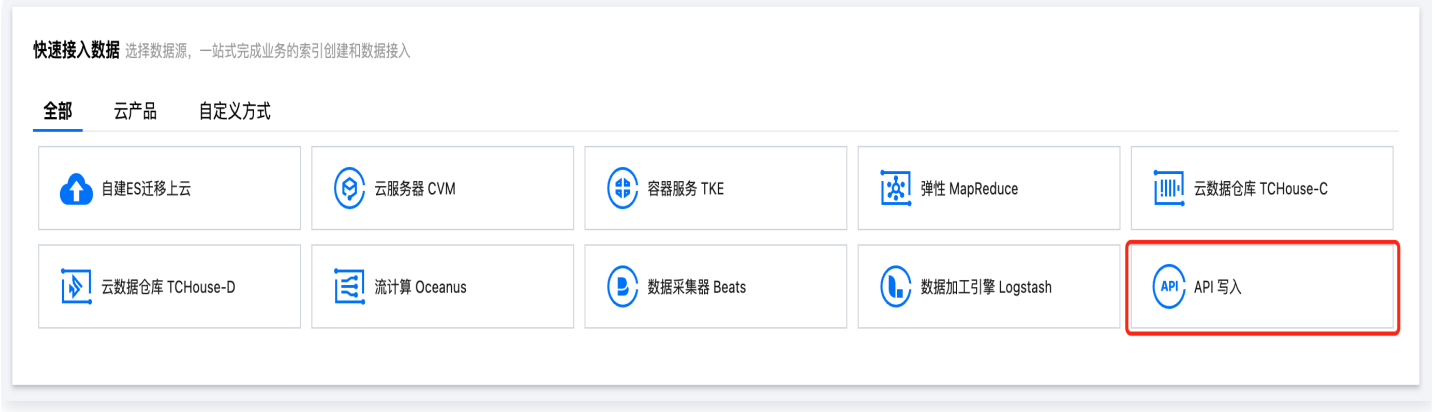
5. After filling in the information, click **Confirm Creation** to successfully create the space.

Create an index

There are two ways to create an index: either directly on the space list page or on the space basic page. The following describes the process of creating it from the space list page.

1. On the space list page, select the data source on the Quick Access Data page. Here, we take API Write as an example.

Search for the required CAM policy as needed, and click to complete policy association.



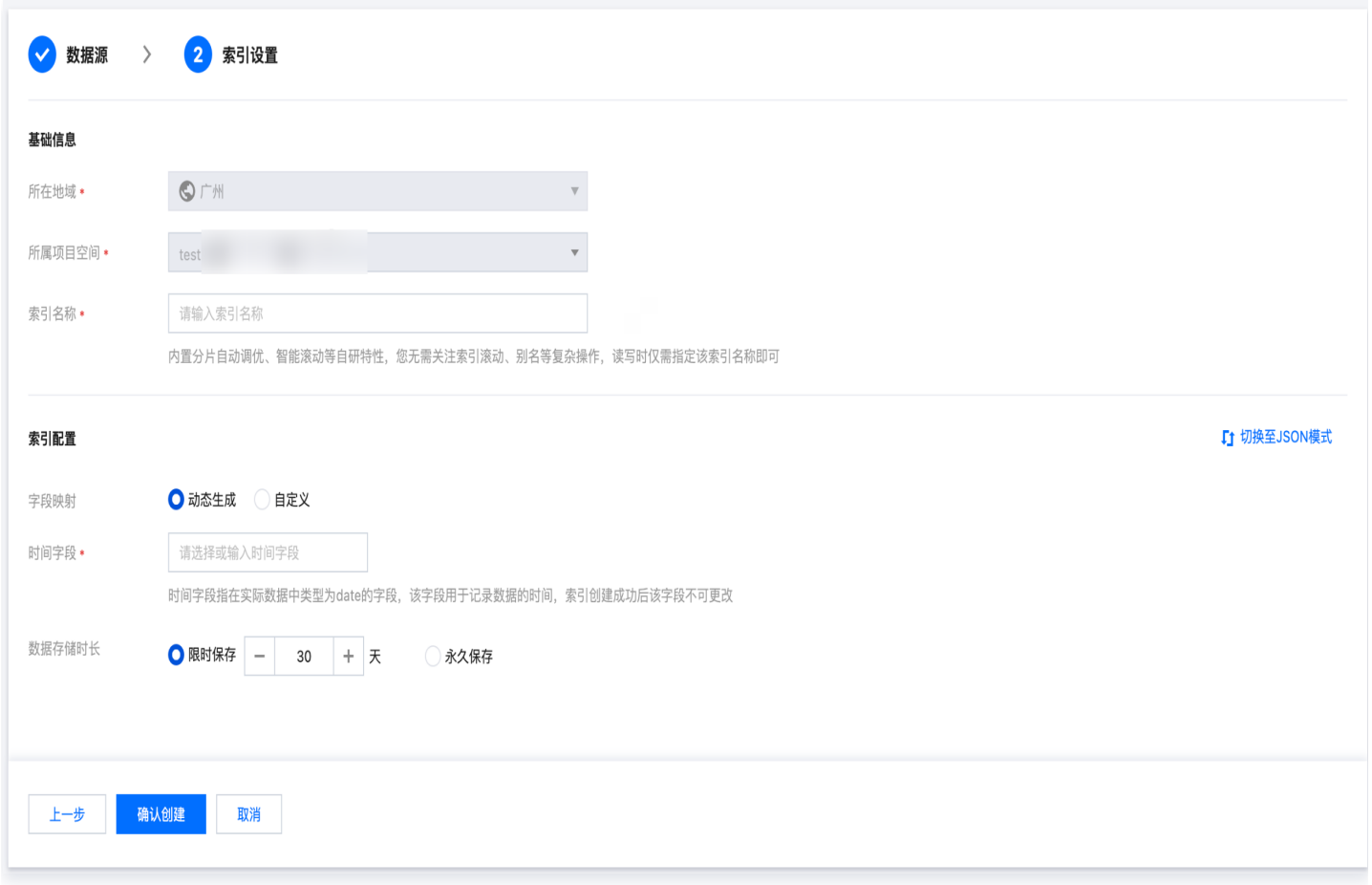
2. To view write prompts, click **Next** .

Search for the required CAM policy as needed, and click to complete policy association.

3. Enter the Index Settings page, fill in the basic information and index configuration, then click **Confirm creation**.

- **Region:** Select the region information from the dropdown list.
- **Project space:** You can create the index in the desired project space for management convenience. If not in the dropdown, click **Create a new project space** on the page and follow the instructions to complete the creation.
- **Index Name:** This name can be used for subsequent data writing and querying. Naming conventions follow the page prompts.
- **Field Mapping:** Used to set the field information of the data. Here, you can choose **Dynamic Generation** to auto-generate field settings based on your written data, or you can choose to set the field information by yourself.
- **Time Field:** Select or enter the date-type field in your data. Once the index is created, the time field cannot be modified.
- **Data Storage Duration:** The retention period of the data. For example, if you set **Timed saving for 30 days**, data will be deleted on the 30th day after being written into the index.

Search for the required CAM policy as needed, and click to complete policy association.



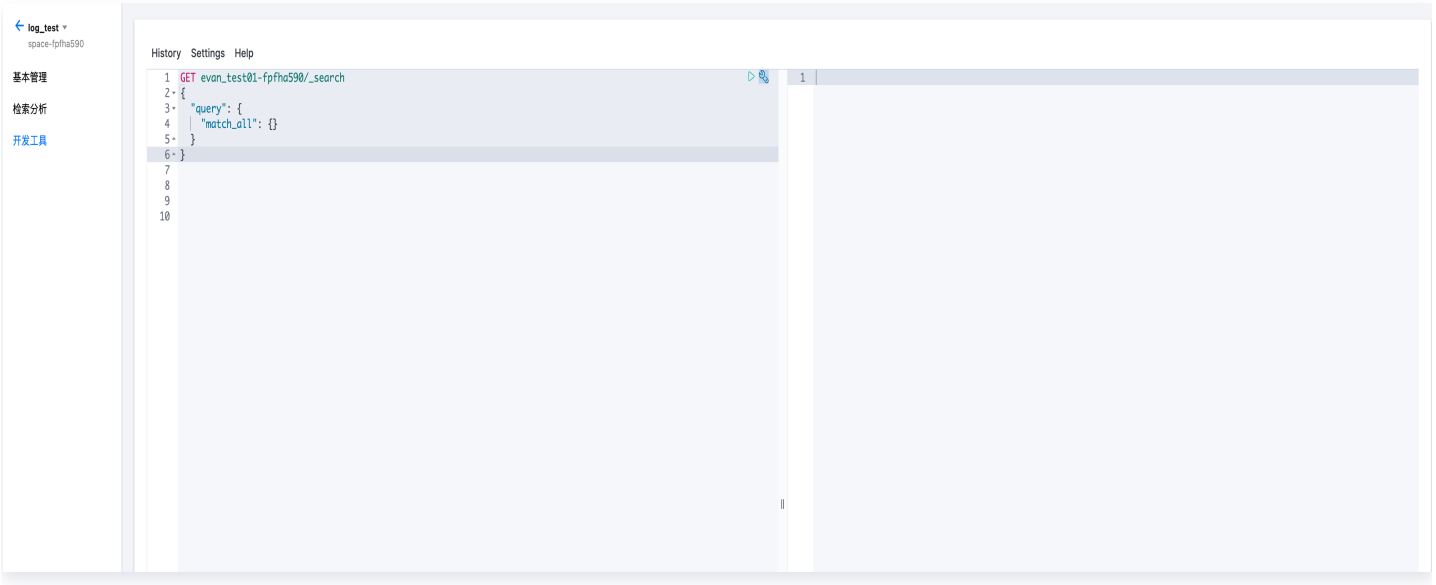
Writing Data

1. In the space list, click the name of the corresponding space to enter the space management page.

> **Note:**
> Kibana's related modules are embedded in the Tencent Cloud console, allowing us to directly use the search and analysis capabilities in the console. Among them, **Search and analysis** corresponds to **Discover**, and **Development tools** correspond to **Dev tools**. The embedded capabilities require browser support for third-party cookies. If it does not work properly, please try enabling third-party cookie settings in the browser. For external link access to Kibana for data writing, refer to Writing Data.

- Enter the **Development tools** page.

Search for the required CAM policy as needed, and click to complete policy association.



- Enter the following statement and click the triangle in the diagram to successfully write the data. Each click counts as writing one piece of data (the content in {} represents a complete log). You can click multiple times to facilitate the subsequent data retrieval demonstration.

Search for the required CAM policy as needed, and click to complete policy association.

The reference statement is as follows:

```
POST index_name/_doc
{
    "id":"090798",
    "routing_no":"4087",
    "region":"10002424",
    "user_name":"user-Ufa9Yee1P",
    "user_type":"01",
    "ip":"119.147.10.191",
    "now_local":"gz",
```

```
    "@timestamp":1705648983762
}
```

> **Note:**
> You need to change **index_name** in the statement to your own index name.
> If your time field is not **@timestamp**, you need to change **@timestamp** in the diagram to the time field you set yourself.
> For bulk data writing, please refer to Data Writing.

Search and Analysis

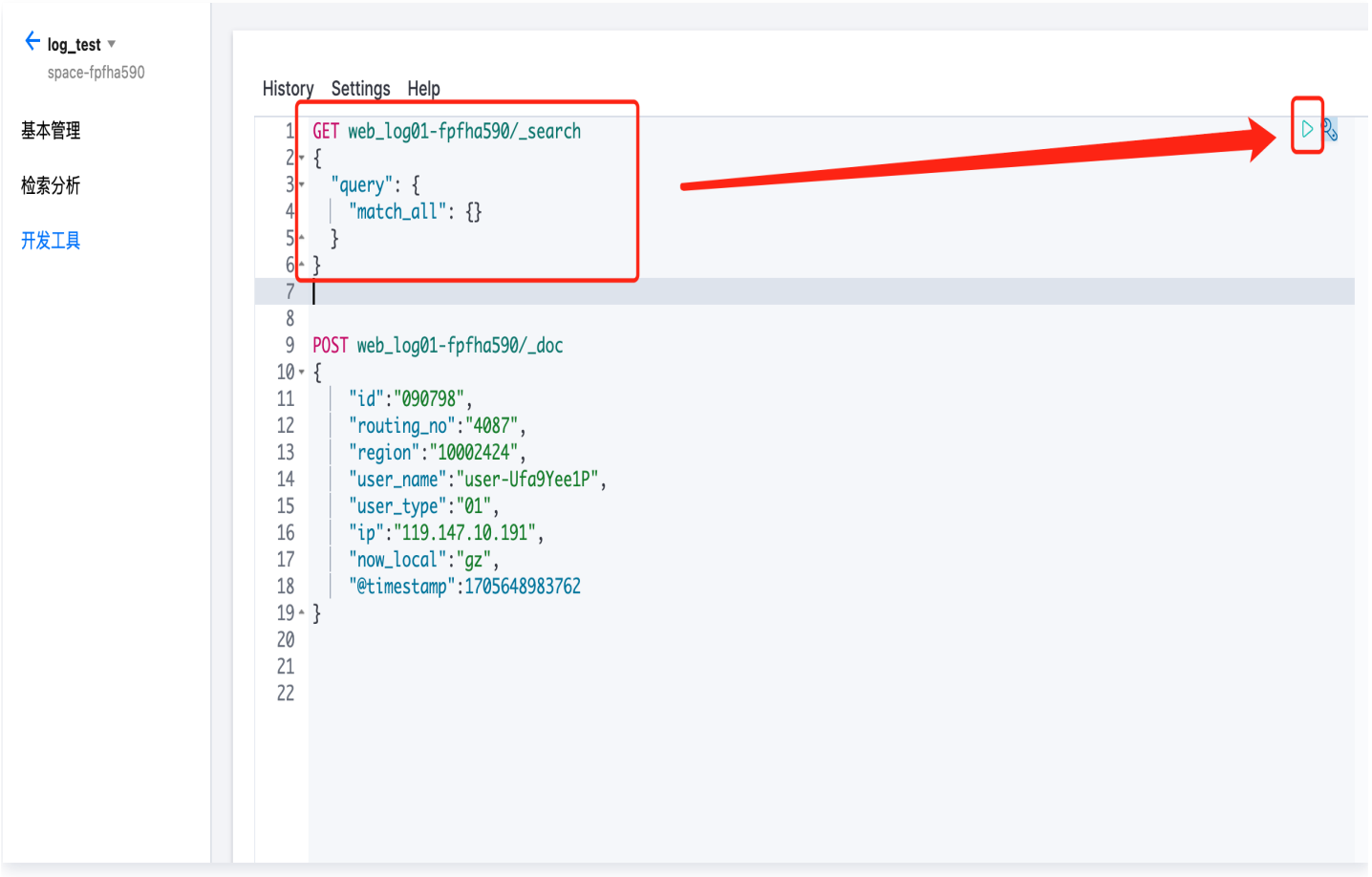
Through the above steps, we have successfully written the data into ES Serverless service. Next, we will demonstrate how to query this data.

Method 1: Using DSL

1. You can directly copy the example statement below and click the triangle in the diagram to query the written data.

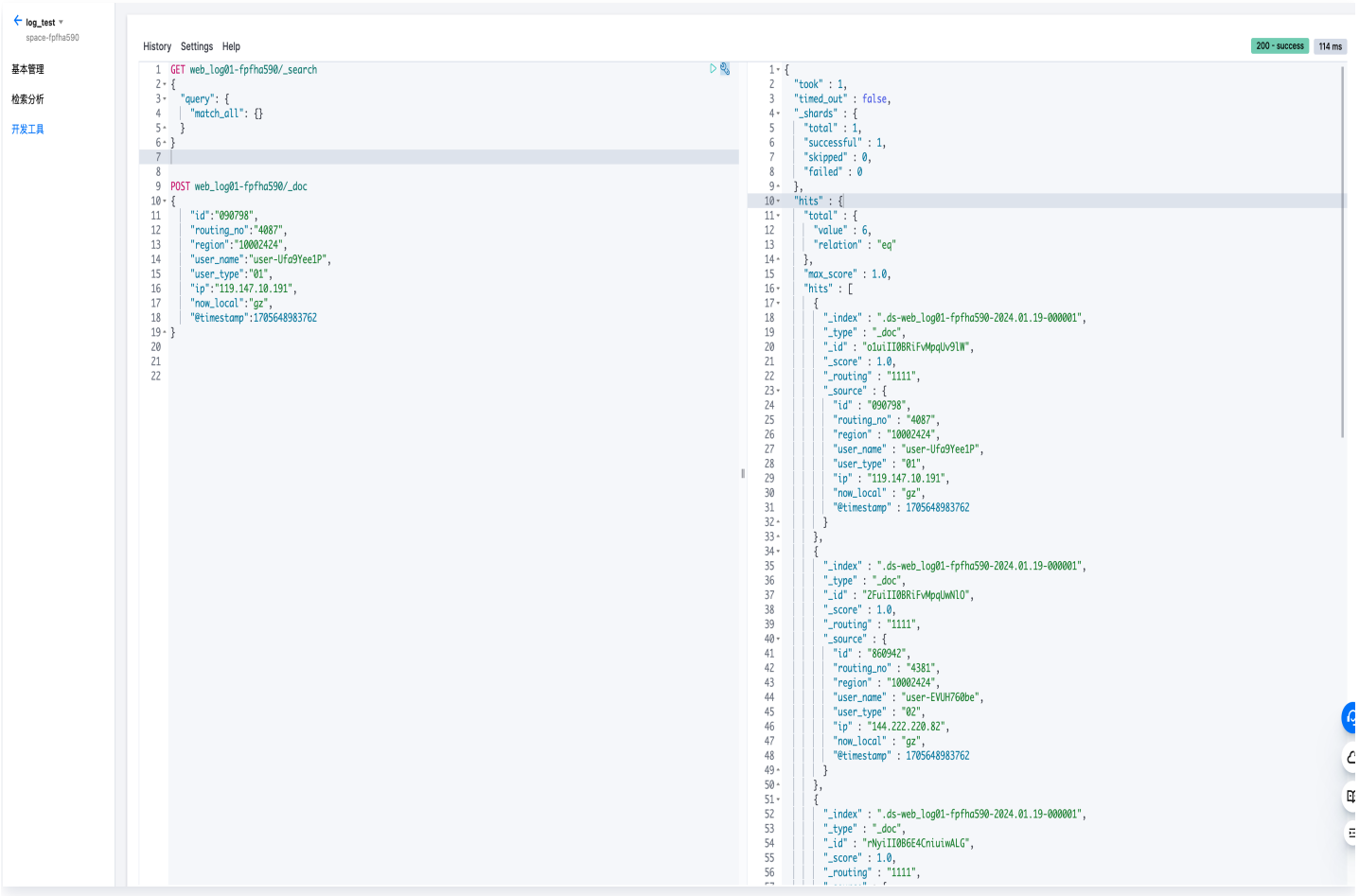
```
GET index_name/_search
{
    "query":{
      "match_all":{}
      }
}
```

Search for the required CAM policy as needed, and click to complete policy association.



2. The following results are returned, indicating that the data was successfully queried.

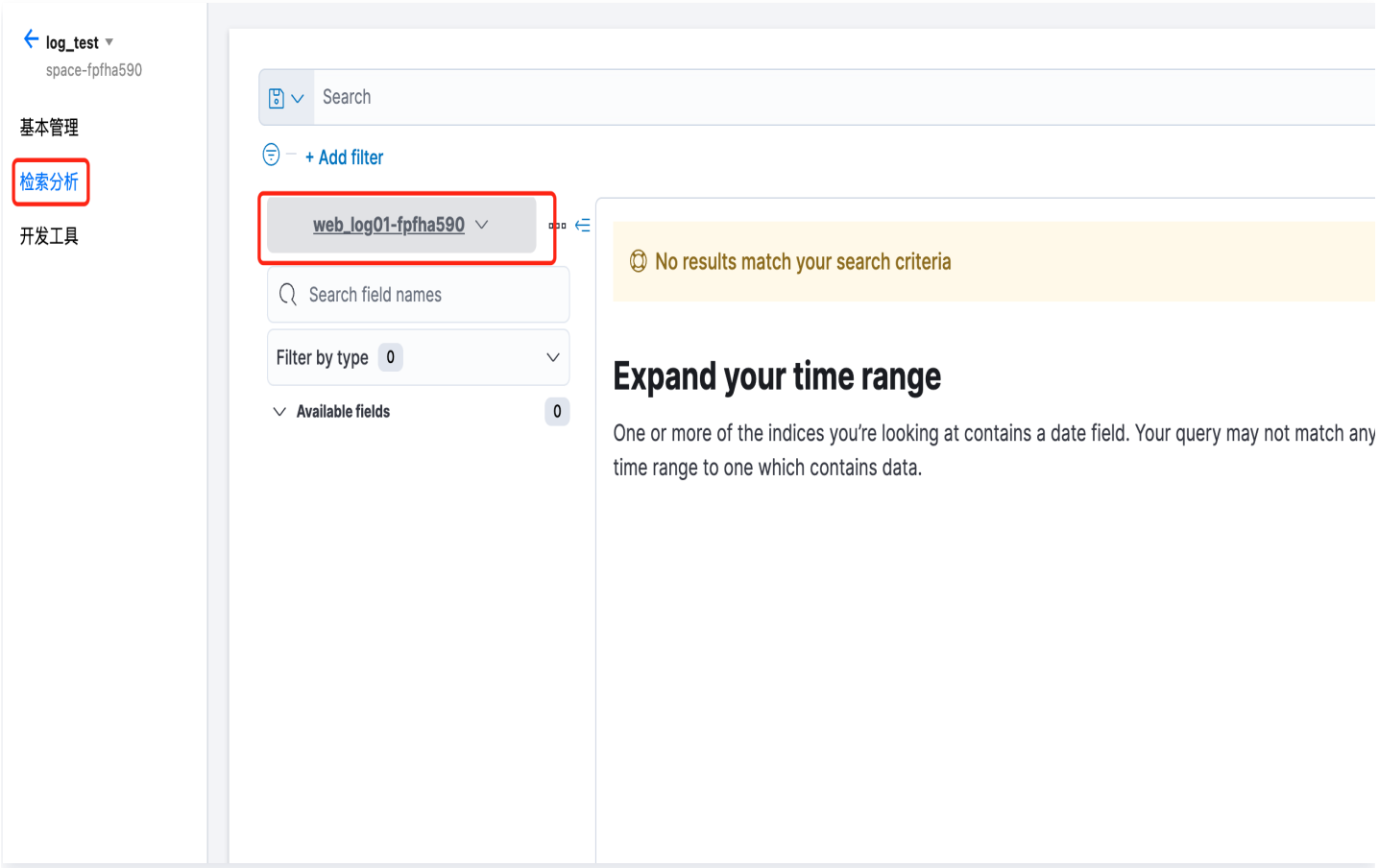
Search for the required CAM policy as needed, and click to complete policy association.



Method 2: via Discover

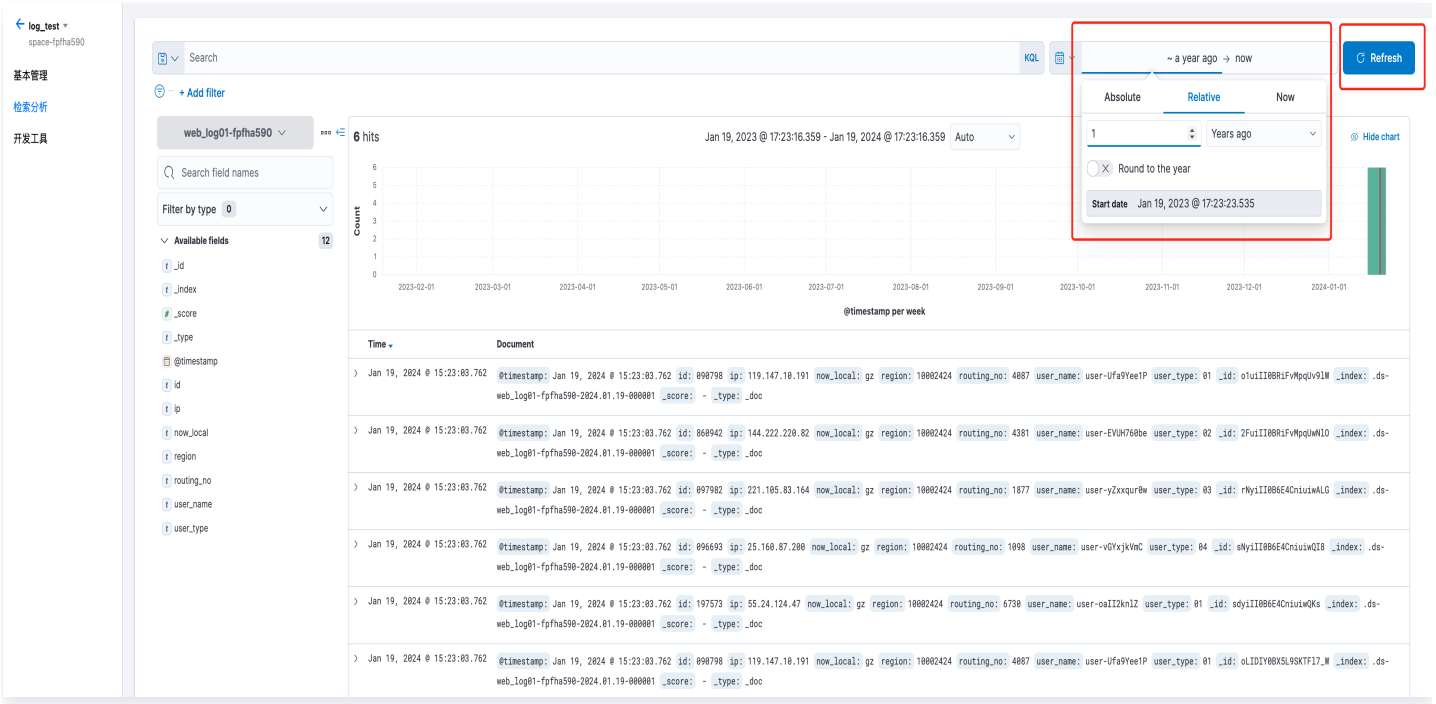
1. Click **Search and analysis**, and select the index you just wrote in from the drop-down list.

Search for the required CAM policy as needed, and click to complete policy association.



2. Filter by time. In the example, the data was written in January 2024, so we can select the past year, i.e., **last 1 year** to successfully retrieve the data from the past year.

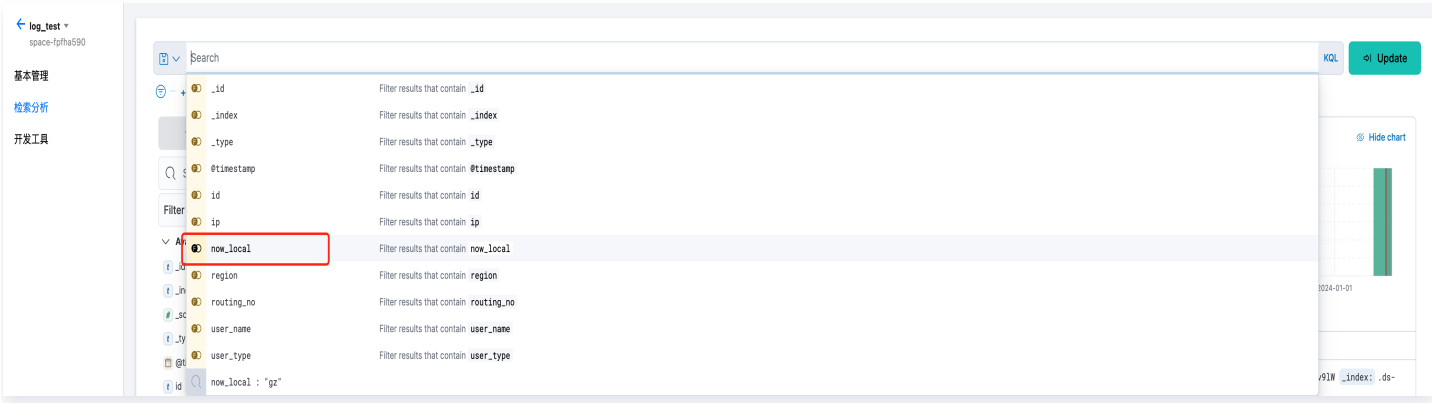
Search for the required CAM policy as needed, and click to complete policy association.



3. You can also enter keywords to search for the desired content. For example, you want to search for the field value "**gz**" in **now_local**.

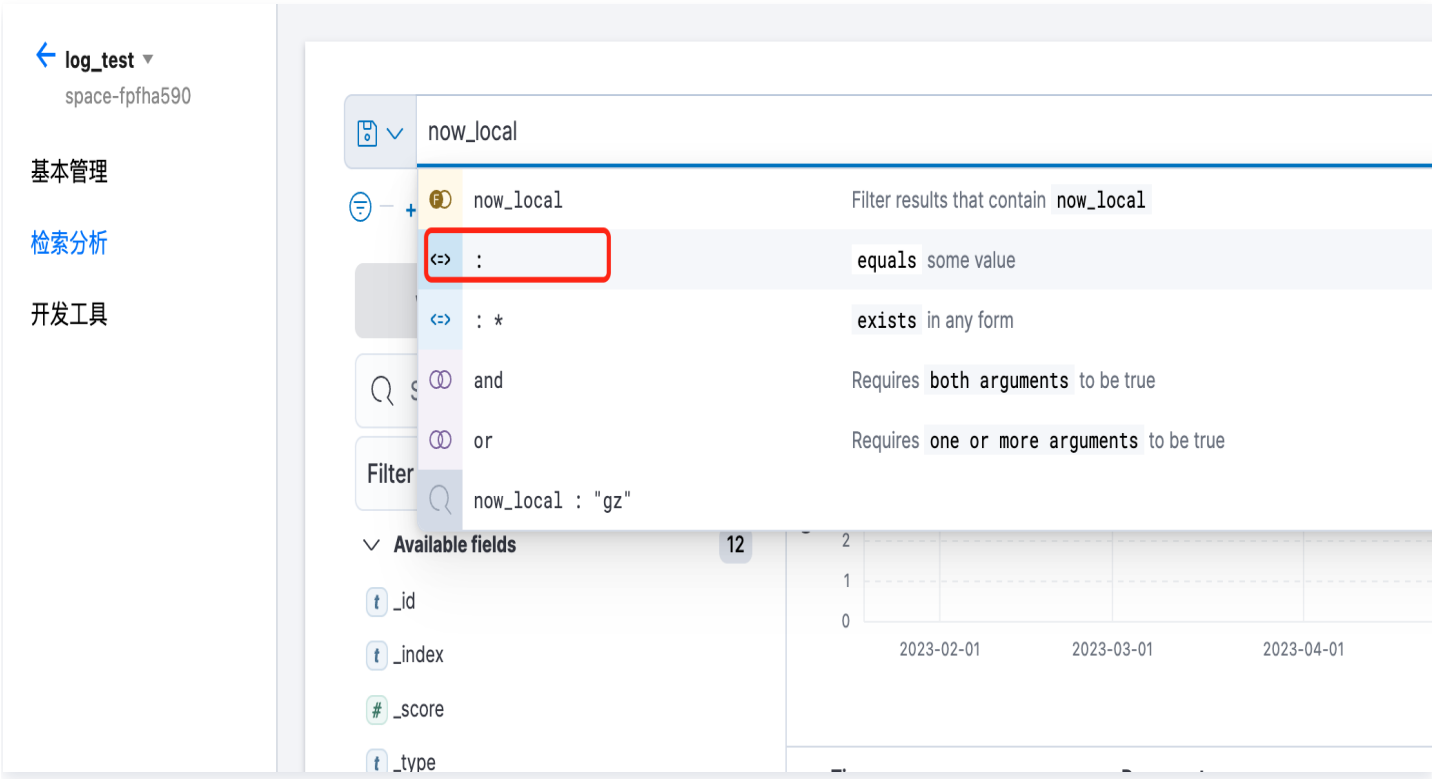
- Click **now_local**, as shown below:

Search for the required CAM policy as needed, and click to complete policy association.



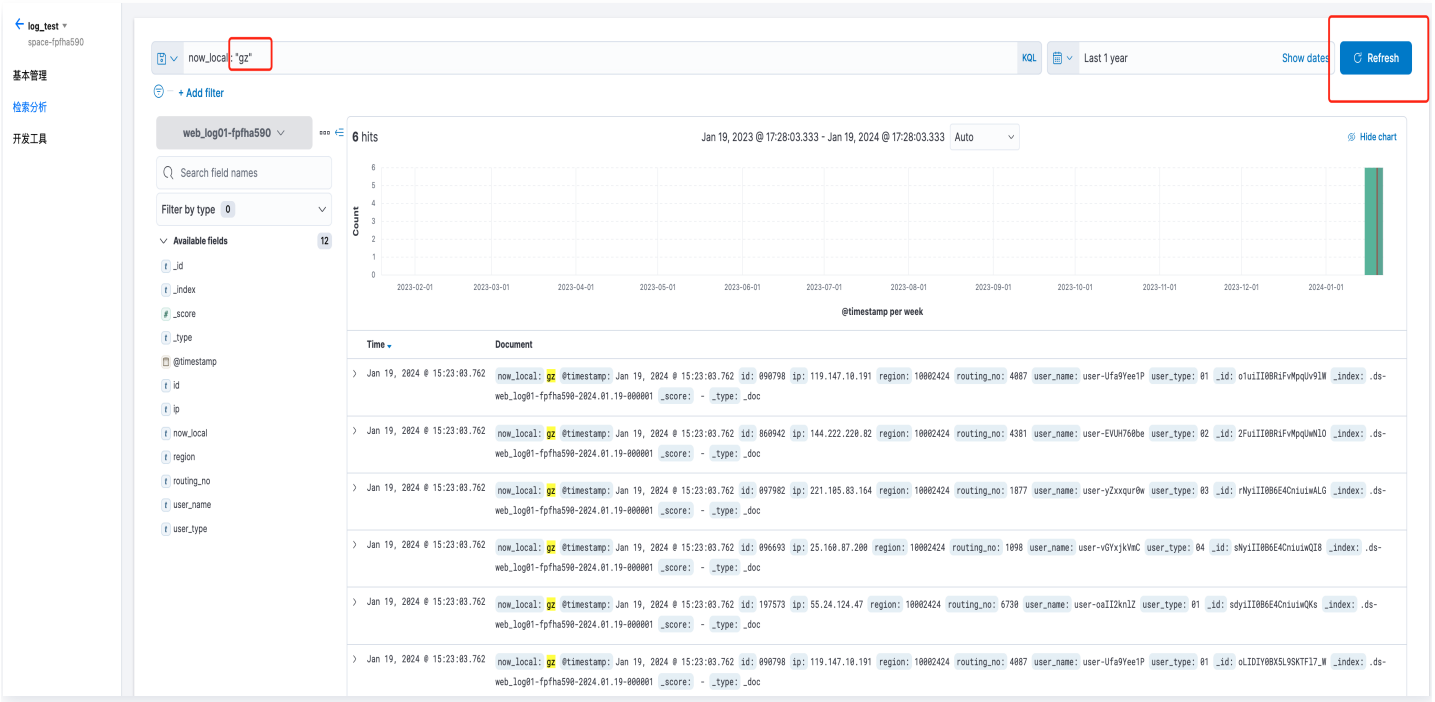
- Select **:**, as shown below:

Search for the required CAM policy as needed, and click to complete policy association.



- Enter "gz" and click the **Refresh** button to search for all contents with the value "**gz**" in **now_local** and highlight them, as shown below:

Search for the required CAM policy as needed, and click to complete policy association.



For more information on using search and analysis, see Data Query.

Quick Start
Create an index

Last updated: 2024-10-25 09:07:11

Prerequisites

- You have a Tencent Cloud account. For more information on how to create an account, please see Signing up for a Tencent Cloud Account.
- If using a sub-account to log in, please ensure the account has read and write permissions for ES.

Directions

Log in to the Console

1. Log in to the ES Console.
2. In the top menu bar, select a region. Currently, Beijing, Shanghai, Guangzhou, Nanjing, and Hong Kong (China) regions are supported.
3. In the left sidebar under Serverless mode, select **Log Analysis**.

Create Project Space

1. click **Create a new space**.
2. Enter the project space name, supporting 1 - 20 Chinese characters, English letters, numbers, underscores, or the separator "-".
3. click **Confirm**. If the verification is correct, the project space will be successfully created.

Search for the required CAM policy as needed, and click to complete policy association.

Note:
In Elasticsearch Serverless Log Analysis, you can only create indexes and then write data through APIs later, or you can access data sources such as CVM or TKE on the "Data Access" tab of the corresponding index. You can also access data while creating the index, completing the one-stop CVM log access, TKE log access, etc. Below is an introduction to the creation operation of indexes using the API writing method.

Create a new index

1. On the ES Serverless Log Analysis homepage, click **Project Space Name** to enter the Index List Page, and click **Create a new index.**

2. Enter the Create Index Page and select **API Writing Method.**

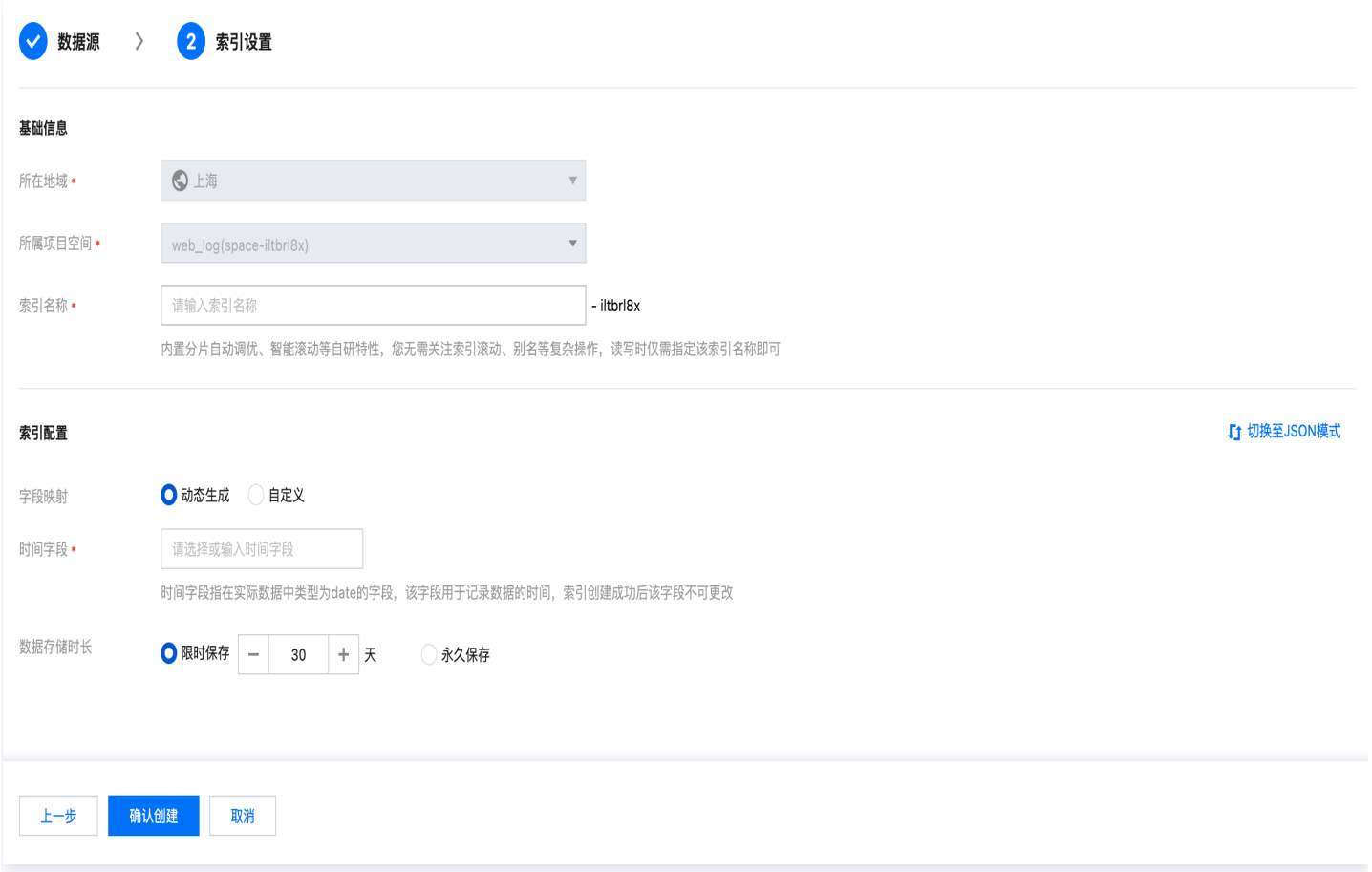
Search for the required CAM policy as needed, and click to complete policy association.

3. If you want to learn how to write data to ES Serverless through API, you can click to view the documentation. click **Next** .

Search for the required CAM policy as needed, and click to complete policy association.

4. Enter the index settings interface and fill in the basic information.
   - Region: Consistent with the region of the project space.
   - Project space: Created in the current project space by default.
   - Index Name: Length 1 – 100, supports lowercase letters, numbers, –, _, ;, @, &, =, !, ', %, $, ., +, (,).

Search for the required CAM policy as needed, and click to complete policy association.



5. Fill in the index configuration information.

- **Field Mapping**
  - Dynamic Generation: Enabled by default. Once enabled, it will automatically parse the written data and set the index fields.
  - Auto Configuration by Input Sample: After disabling **Dynamic Generation**, you can use **Auto Configuration by Input Sample** to generate the field mappings of the index. Enter a sample JSON data in the input field, and the platform will automatically validate it for you. After successful validation, the relevant fields will be mapped to the Field Mapping table.

Field Mapping splits the original data into multiple tokens by field (i.e., key:value) for index construction and performs retrieval based on this mapping. Details are as follows:

Parameter	Feature Description
Field Name	Field names in the written data
Field Type	Field data types supported in the interface: "text,date,boolean,keyword,long,double,integer,ip,geo_point"—a total of 9 types. More field types are supported in **JSON Edit Mode**. For details, refer to the Official Documentation
Allow Chinese Characters	This feature can be enabled when the field contains Chinese and you need to retrieve the Chinese content. After enabling, the ik_max_word tokenizer will be used by default for the text field
Enabling index	After enabling, an index will be built for the field to be used for retrieval
Enable Statistics	After enabling, statistical analysis can be performed on the field values, which will increase index storage

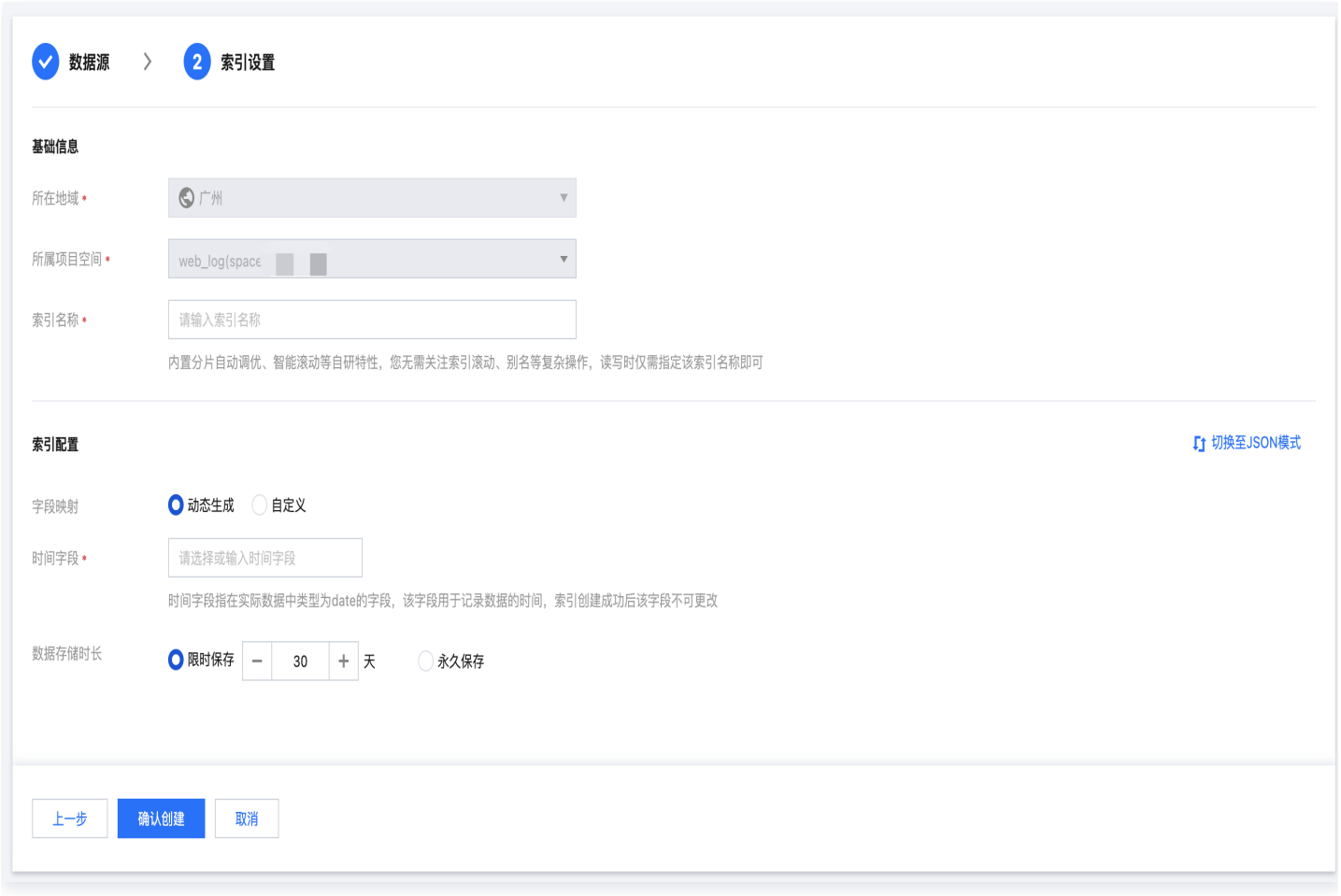
- **Time Field**
  The time field refers to the field of the type date in the actual data. Once the index is

successfully created, this field cannot be changed.

> **Note:**
> The time field is enabled by default with **Indexing Enabled** and **Statistics Enabled**, and cannot be turned off.

- **Storage Duration**
  You can set the storage duration for the data. The default option is 30 days, and permanent storage is also supported.

Search for the required CAM policy as needed, and click to complete policy association.



6. After ensuring that the information is error-free, click **Confirm creation** to complete the index creation. For data writing, please refer to the documentation.

CVM Log Access

Last updated: 2024-10-25 09:08:53

Prerequisites

- You have a Tencent Cloud account. For more information on how to create an account, please see Signing up for a Tencent Cloud Account.
- If using a sub-account to log in, please ensure the account has read and write permissions for ES.

Operation Steps

Logging in to the Console

1. Log in to the ES Console.
2. In the top menu bar, select **Region**. Currently, the Beijing, Shanghai, Guangzhou, Nanjing, and Hong Kong (China) regions are supported.
3. In the left sidebar under Serverless mode, select **Log Analysis**.

Create Project Space

1. click **Create a new space**.
2. Enter the project **space name** , supporting 1 - 20 Chinese characters, English letters, numbers, underscores, or the separator "-".
3. click **Confirm**. If the verification is correct, the project space will be successfully created.

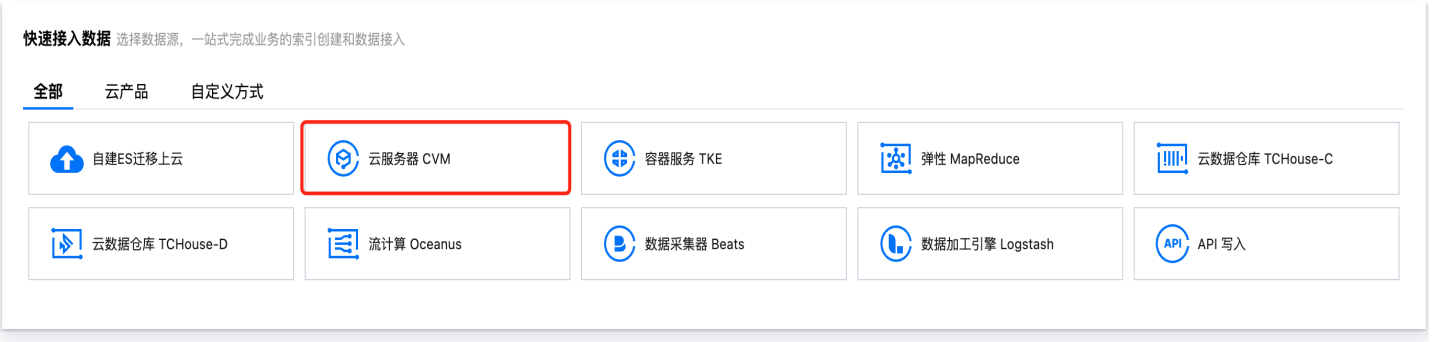
Search for the required CAM policy as needed, and click to complete policy association.

Note:
In ES Serverless Log Analysis, you can only create an index initially, then write data via API or access data sources like CVM or TKE through the "Data Access" tab of the corresponding index. You can also perform data access while creating indexes to achieve one-stop CVM log access and TKE log access. Below is a guide to one-stop CVM log access operations.

CVM Log Access

In the ES Serverless Log Analysis homepage, select **CVM Log Access** to enter the CVM Log Access page.

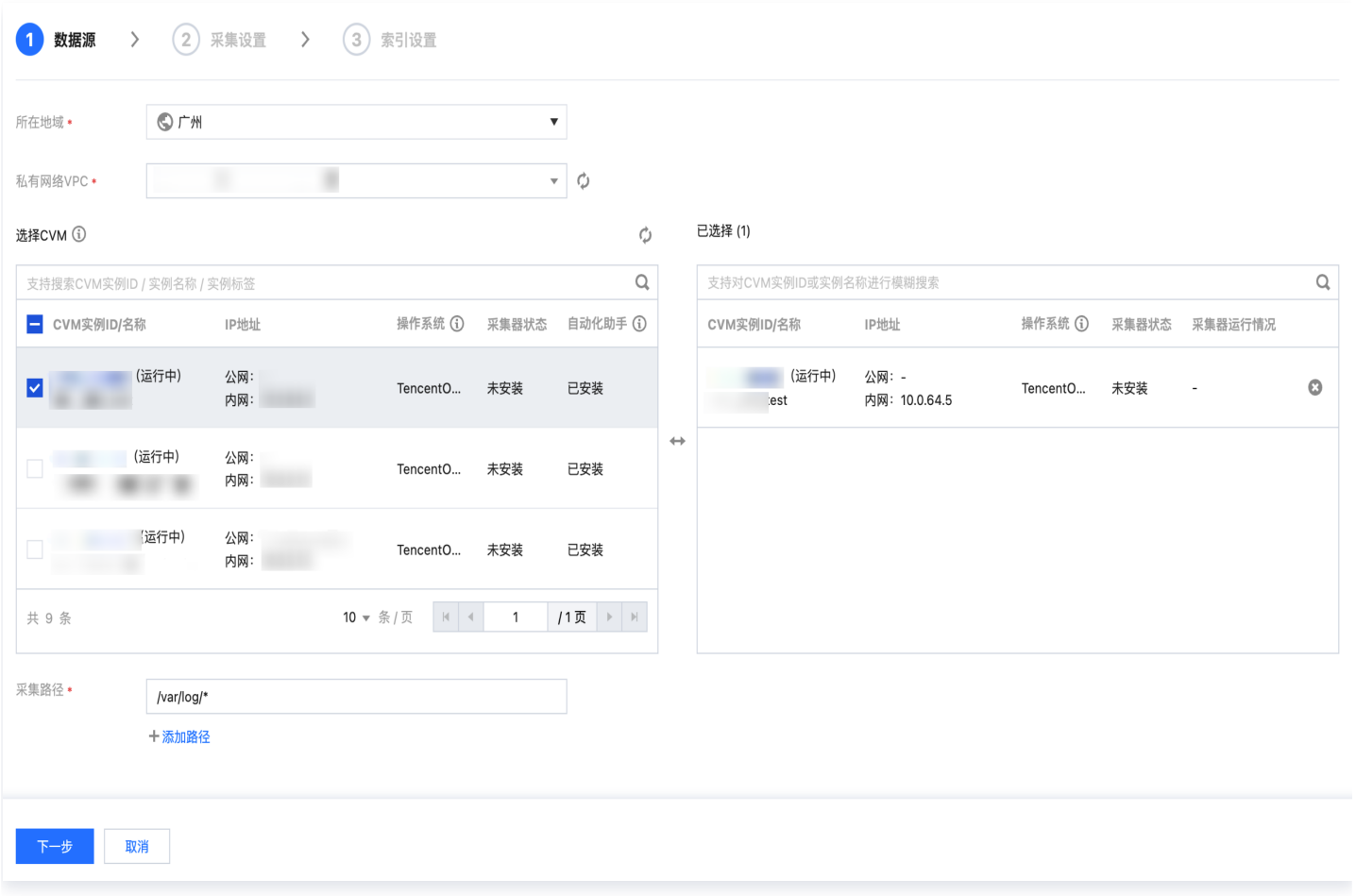
Search for the required CAM policy as needed, and click to complete policy association.



Data Source Settings

- **Region**: Required. The region where the CVM is located.
- **VPC**: Required. The VPC where the CVM is located. After confirmation, the servers under this VPC will be pulled.
- **Select CVM**: Select the CVM instances for log collection. Currently, only Linux system CVMs are supported, and data collection is possible only after TAT is installed.
- **Collection Path**: Set the log directory and file name based on the location of the log on the server, supporting one or multiple entries. Directory and file names support full names and wildcard patterns.

Search for the required CAM policy as needed, and click to complete policy association.



Collection Settings

Basic Settings

Collection Strategy: Supports full and incremental collections. Once created, the collection strategy cannot be modified. Full collection will collect historical logs as well as logs generated after the Filebeat configuration takes effect. Incremental collection will only collect logs generated after the Filebeat configuration takes effect.

Collection Parsing

Collection Template: If you need to set up quickly or experience, you can select the appropriate collection template based on your log's output format. After confirming, you can return to the interface, update the log sample to the actual log data, and quickly complete the collection parsing settings.

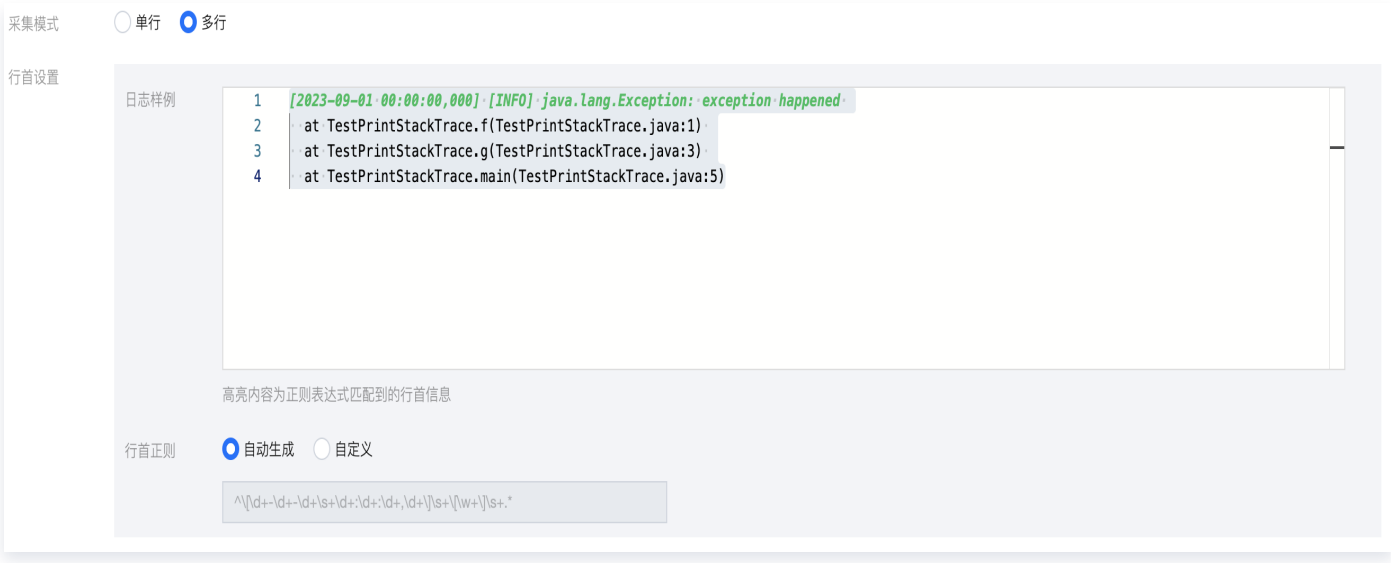
Search for the required CAM policy as needed, and click to complete policy association.

Collection Mode: Supports single-line and multi-line. Once created, the collection mode cannot be changed.

- Single Line Text Log: Each line of log content is a separate log entry, separated by line breaks in the log file.
- Multiline Text Log: A log entry consists of multiple lines, such as Java stack logs. In this mode, you need to configure log samples and regular expressions at the beginning of a line. Filebeat matches the beginning of a log entry using the regular expression for the line start, confirming the start of a log entry, and considers unmatched portions as part of the log entry until the next line start appears. After inputting the log sample, the system will automatically generate a regular expression for the line start. You can also choose to generate it from Definition. The highlighted content in the input box is the information matched by the regular expression for the line start.

> **Note:**
> Be sure to use logs from actual scenarios to facilitate the automatic extraction of regular expressions for the line start.

Search for the required CAM policy as needed, and click to complete policy association.



Extraction Settings: Support setting the extraction mode to full text log, JSON format, delimiter. Once created, the extraction mode cannot be modified. Details are as follows:

Full Text Log

No key-value extraction is performed on the log data; the log content will be stored in a field named "message", and you can perform search and analysis using automatic word segmentation capabilities.
For example, the raw data of a single-line log is as follows:

```
Tue Jan 01 00:00:00 CST 2023 Running: Content of processing
something
```

The data collected into the index is as follows:

```
massage:Tue Tue Jan 01 00:00:00 CST 2023 Running: Content of
processing something
```

JSON

For logs in standard JSON format, we can extract the corresponding fields according to the Key:Value pairs in the log.

Assume that one of your JSON log raw data is:

```
{"pid":321,"name":"App01","status":"WebServer is up and running"}
```

After structured processing, the log will become as follows:

```
{
    "pid":321,
    "name":"App01",
    "status":"WebServer is up and running"
}
```

Separator

For logs with fixed separator content, we can extract key-value pairs from the log using the specified separator. The separator can be a single character or a string, and can be selected or entered in the console.

Assume that one of your log raw data is:

```
321 - App01 - WebServer is up and running
```

If the separator is specified as "-", the log will be split into 3 fields. We can assign unique keys to these 3 fields in the extraction results, as shown below:

```
pid: pid
name: App01
status: WebServer is up and running
```

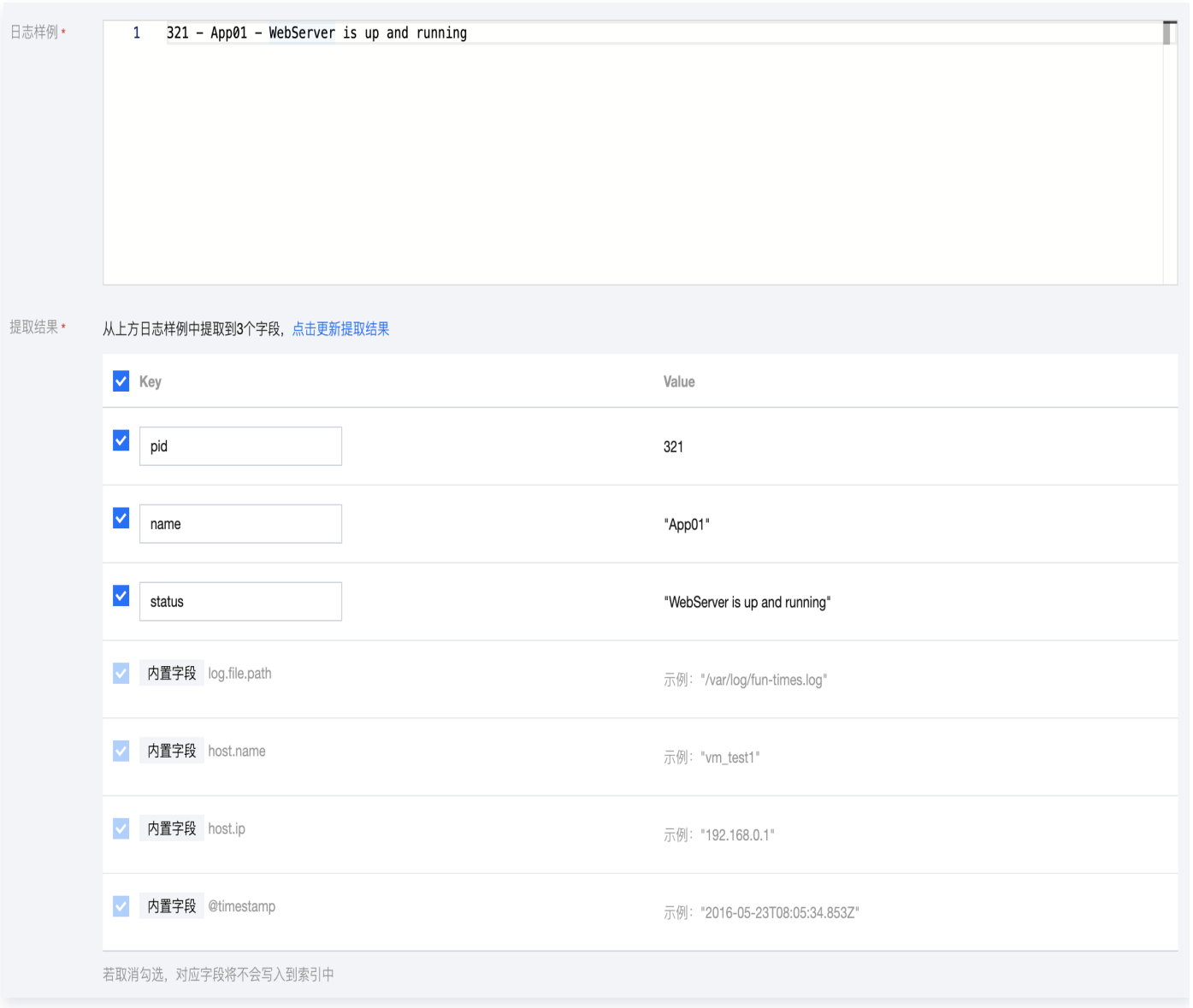
Extraction results: If the extraction mode is selected as "JSON format" or "Separator", we can input a log sample, and the system will automatically extract the sample:

- If the extraction mode is JSON format, the system will automatically fill in the extracted Key and Value. If unchecked, the corresponding fields will not be written into the index.
- If the extraction mode is delimiter, the system will default to filling the extracted data into Value. You can define a unique Key for each Value. If unchecked, the corresponding field will not be written into the index.

- Built-in Fields: When configuring CVM log collection in the console, Filebeat will write information such as log source and timestamp into the log in the form of Key-Value pairs. These fields are built-in fields. If the Key names in your business logs duplicate the built-in field names, the field content of the business logs will be collected first, and the corresponding built-in fields will not be written into the index. The meanings of the built-in fields are as follows:

Built-in Field Name	Meaning
log.file.path	Path where the logs are stored
host.name	Name of the server where the logs are located
host.ip	IP of the server where the logs are located
@timestamp	Time when the log was collected

Search for the required CAM policy as needed, and click to complete policy association.



Retain Raw Logs : If checked, the raw log content before parsing extraction will be retained in this field.

Record parsing error: If the extraction mode is "delimiter", you can choose whether to record parsing errors. When checked, if parsing fails, the error message will be uploaded as the value (Value) to this field.

Index Configuration

- **Project Space** : You can assign all indexes of the same business to a project space for easier management.
- **Index Name** : Length is 1 – 100, supports lowercase letters, numbers, –, _, ;, @, &, =, !, ', %, $, ., +, (,).
- **Field Mapping**

- Dynamic Generation: Enabled by default. Once enabled, it will automatically parse the written data and set the index fields.
- After disabling **Auto Generate** for input sample configuration, you can use **Input Sample Auto Configuration** to generate field mappings for the index. Enter a JSON formatted data sample in the input box. After confirmation, the platform will automatically validate it. Once validated correctly, the relevant fields will be mapped to the field mapping table.

Field Mapping splits the original data into multiple tokens by field (i.e., key:value) for index construction and performs retrieval based on this mapping. Details are as follows:

Parameter	Feature Description
Field Name	Field names in the written data
Field Type	Field data types supported in the interface: "text,date,boolean,keyword,long,double,integer,ip,geo_point"—a total of 9 types. More field types are supported in **JSON Edit Mode**. For details, refer to the Official Documentation
Allow Chinese Characters	This feature can be enabled when the field contains Chinese and you need to retrieve the Chinese content. After enabling, the ik_max_word tokenizer will be used by default for the text field
Enabling index	After enabling, an index will be built for the field to be used for retrieval
Enable Statistics	After enabling, statistical analysis can be performed on the field values, which will increase index storage

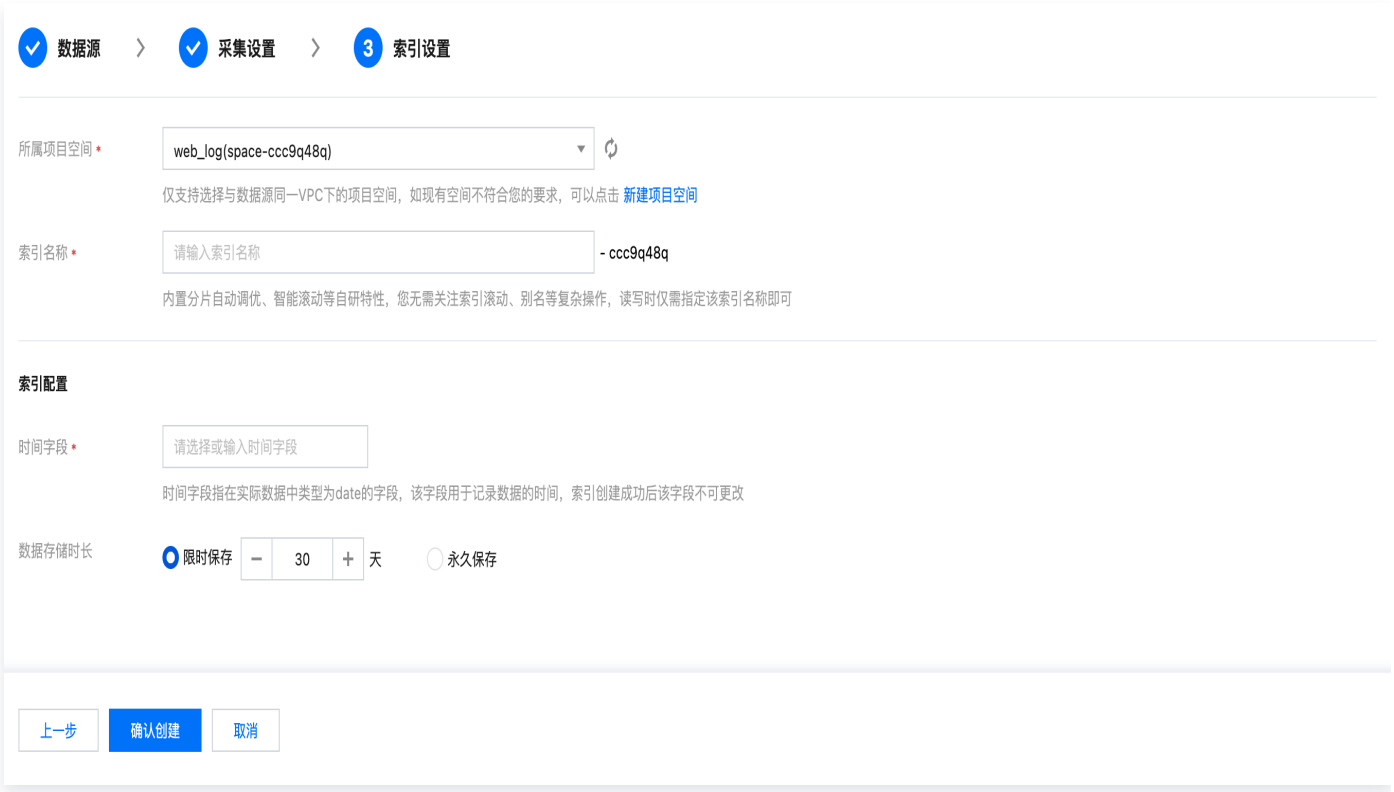
- **Time Field**
  The time field refers to the field of the type date in the actual data. Once the index is successfully created, this field cannot be changed.

> **Note:**
> The time field is enabled by default with **Indexing Enabled** and **Statistics Enabled**, and cannot be turned off.

- **Data Storage Duration:**

1.1 You can set the data storage duration. By default, it is set to 30 days, and permanent storage is also supported.

Search for the required CAM policy as needed, and click to complete policy association.



1.2 After filling in the information correctly, click **Confirm Create** to complete CVM log collection.

TKE TKE Log access

Last updated: 2024-10-25 09:09:28

Prerequisites

- You have a Tencent Cloud account. For more information on how to create an account, please see Signing up for a Tencent Cloud Account.
- If using a sub-account to log in, please ensure the account has read and write permissions for ES.

Directions

Logging in to the Console

1. Log in to the ES Console.
2. In the top menu bar, select **Region**. Currently, the Beijing, Shanghai, Guangzhou, Nanjing, and Hong Kong (China) regions are supported.
3. In the left sidebar under Serverless mode, select **Log Analysis**.

Create Project Space

1. click **Create a new space**.
2. Enter the project **space name** , supporting 1 - 20 Chinese characters, English letters, numbers, underscores, or the separator "-".
3. click **Confirm**. If the verification is correct, the project space will be successfully created.

Search for the required CAM policy as needed, and click to complete policy association.

Note:
In ES Serverless Log Analysis, you can only create index and then write data via API or access data sources such as CVM or TKE on the corresponding index's "Data Access" tab. You can also access data while creating the index, achieving one-stop CVM log access and TKE log access. Below is an introduction to one-stop TKE log access operations.

TKE TKE Log access

On the ES Serverless Log Analysis homepage, select **TKE TKE** to enter the TKE log access page.

Search for the required CAM policy as needed, and click to complete policy association.

Data Source Settings

- **Region**: The region where the TKE cluster is located.
- **VPC VPC**: Mandatory. The VPC where the TKE cluster is located.
- **Collecting TKE Cluster ID**: Mandatory. The ID of the TKE cluster to be collected. The TKE cluster must be in running status and be a standard cluster. If you need to collect logs from a Serverless Cluster (EKS), please contact us through a Ticket.
- **Based on namespace/host path**: Mandatory. **Namespace**: The first dropdown allows you to select **Include/Exclude,** the second dropdown allows you to select the namespace. Multiple selections are supported, but you cannot select to exclude all namespaces. Based on the host path, please enter the host's **absolute path**, for example, **/var/log/*.log**.
- **Pod Tag**: Optional. Supports creating multiple Pod Tags, with a logical AND relationship between Tags.
- **Container Name**: Optional. The filled container name must be under the target cluster and namespace. If empty, Filebeat will collect all containers under the namespace that meet the Pod Tag requirements.

Search for the required CAM policy as needed, and click to complete policy association.

Collection Settings

Basic Settings

Collection Strategy: Supports full and incremental collections. Once created, the collection strategy cannot be modified. Full collection will collect historical logs as well as logs generated after the Filebeat configuration takes effect. Incremental collection will only collect logs generated after the Filebeat configuration takes effect.

Collection Parsing

Collection Template: If you need to set up quickly or experience, you can select the appropriate collection template based on your log's output format. After confirming, you can return to the interface, update the log sample to the actual log data, and quickly complete the collection parsing settings.

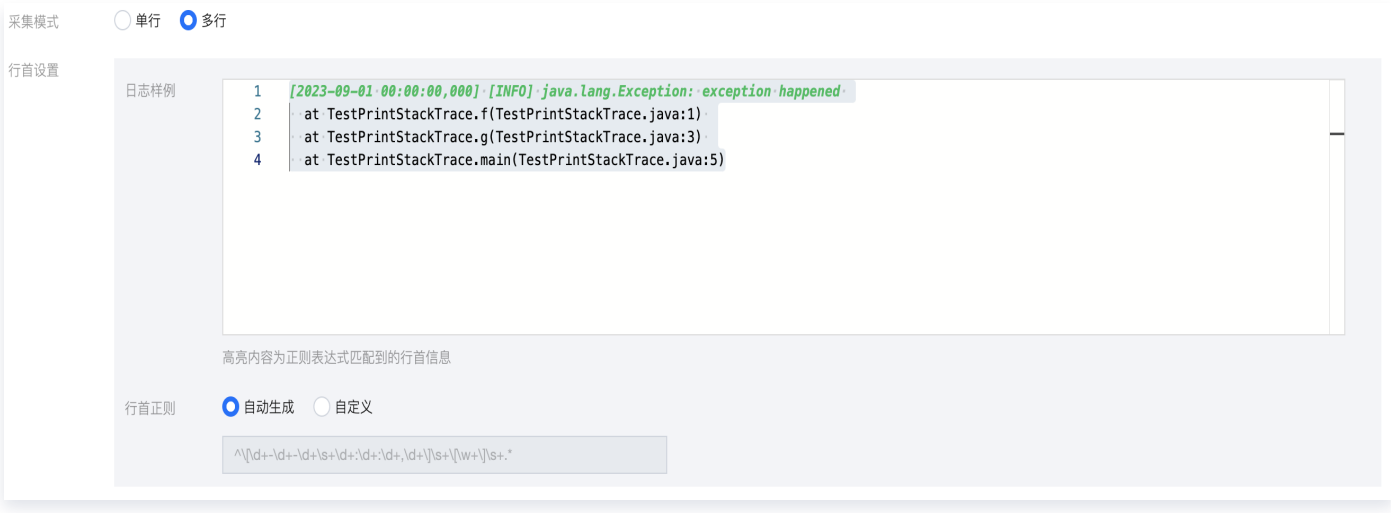
Search for the required CAM policy as needed, and click to complete policy association.

Collection Mode: Supports single-line and multi-line. Once created, the collection mode cannot be changed.

- Single Line Text Log: Each line of log content is a separate log entry, separated by line breaks in the log file.
- Multiline Text Log: A log entry consists of multiple lines, such as Java stack logs. In this mode, you need to configure log samples and regular expressions at the beginning of a line. Filebeat matches the beginning of a log entry using the regular expression for the line start, confirming the start of a log entry, and considers unmatched portions as part of the log entry until the next line start appears. After inputting the log sample, the system will automatically generate a regular expression for the line start. You can also choose to generate it from Definition. The highlighted content in the input box is the information matched by the regular expression for the line start.

> **Note:**
> Be sure to use logs from actual scenarios to facilitate the automatic extraction of regular expressions for the line start.

Search for the required CAM policy as needed, and click to complete policy association.



Extraction Settings: Support setting the extraction mode to full text log, JSON format, delimiter. Once created, the extraction mode cannot be modified. Details are as follows:

Full Text Log

No key-value extraction is performed on the log data; the log content will be stored in a field named "message", and you can perform search and analysis using automatic word segmentation capabilities.
For example, the raw data of a single-line log is as follows:

```
Tue Jan 01 00:00:00 CST 2023 Running: Content of processing
something
```

The data collected into the index is as follows:

```
massage:Tue Tue Jan 01 00:00:00 CST 2023 Running: Content of
processing something
```

JSON

For logs in standard JSON format, we can extract the corresponding fields according to the Key:Value pairs in the log.

Assume that one of your JSON log raw data is:

```
{"pid":321,"name":"App01","status":"WebServer is up and running"}
```

After structured processing, the log will become as follows:

```
{
    "pid":321,
    "name":"App01",
    "status":"WebServer is up and running"
}
```

Separator

For logs with fixed separator content, we can extract key-value pairs from the log using the specified separator. The separator can be a single character or a string, and can be selected or entered in the console.

Assume that one of your log raw data is:

```
321 - App01 - WebServer is up and running
```

If the separator is specified as "-", the log will be split into 3 fields. We can assign unique keys to these 3 fields in the extraction results, as shown below:

```
pid: pid
name: App01
status: WebServer is up and running
```

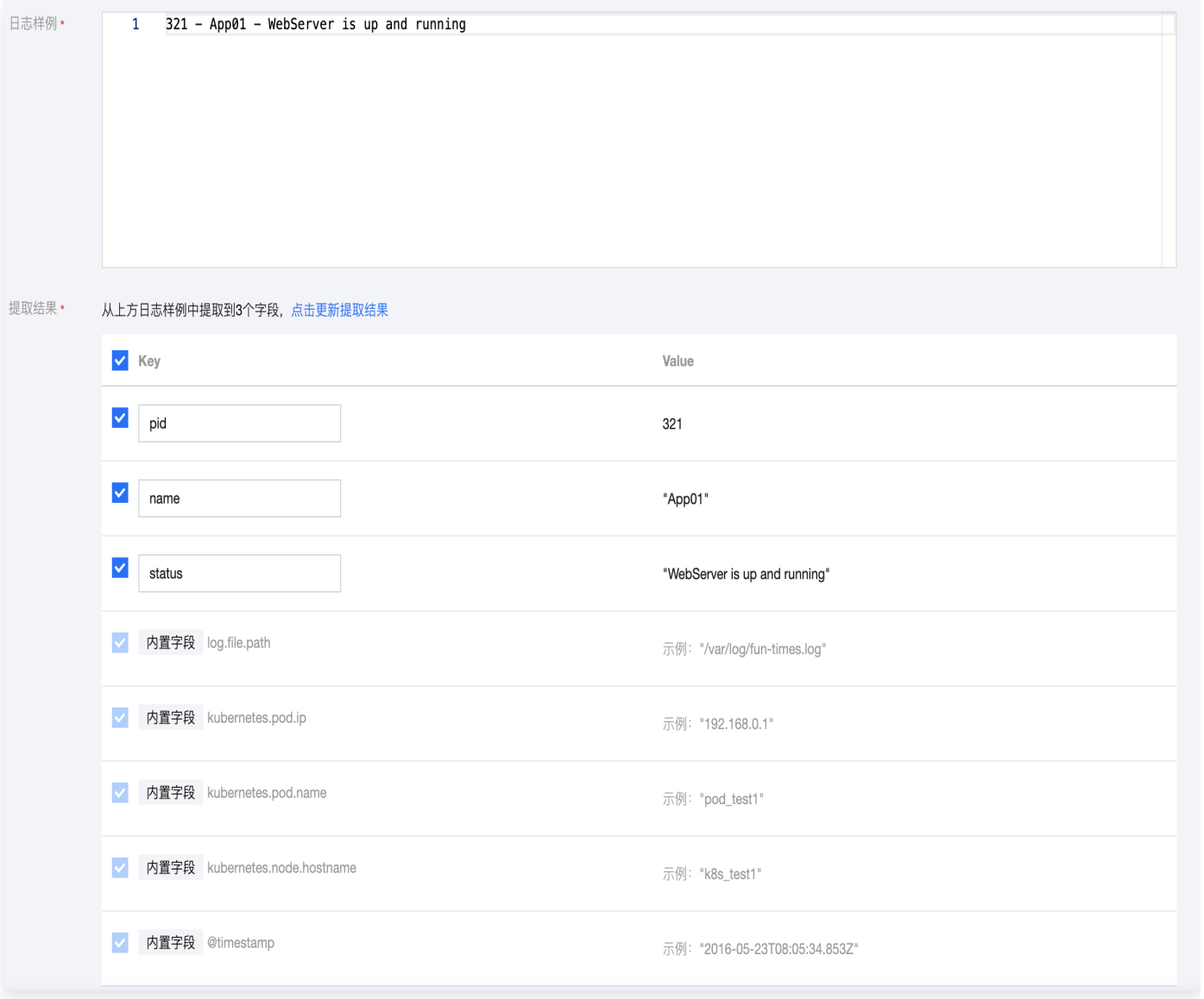
Extraction results: If the extraction mode is selected as "JSON format" or "Separator", we can input a log sample, and the system will automatically extract the sample:

- If the extraction mode is JSON format, the system will automatically fill in the extracted Key and Value. If unchecked, the corresponding fields will not be written into the index.
- If the extraction mode is Separator, the system will automatically fill in the extracted Value. You can assign unique keys to each Value. If unchecked, the corresponding fields will not be written into the index.

- Built-in Fields: When configuring TKE log collection in the console, Filebeat will write information such as log source, timestamp, etc., into the log in Key-Value pairs. These fields are built-in fields. If the Key names in your business logs overlap with the built-in field names, the content of the business log fields will take precedence, and the corresponding built-in fields will not be written into the index. The meanings of the built-in fields are as follows:

Built-in Field Name	Meaning
log.file.path	Path where the logs are stored
kubernetes.pod.ip	IP address of the pod where the log resides
kubernetes.pod.name	Name of the pod where the log resides
kubernetes.node.hostname	Name of the host where the log resides
@timestamp	Time when the log was collected

Search for the required CAM policy as needed, and click to complete policy association.



Retain Raw Logs : If checked, the raw log content before parsing extraction will be retained in this field.

Record parsing error: If the extraction mode is "delimiter", you can choose whether to record parsing errors. When checked, if parsing fails, the error message will be uploaded as the value (Value) to this field.

Index Configuration

- **Project Space** : You can assign all indexes of the same business to a project space for easier management.
- **Index Name** : Length is 1 – 100, supports lowercase letters, numbers, –, _, ;, @, &, =, !, ', %, $, ., +, (,).
- **Field Mapping**

- Dynamic Generation: Enabled by default. Once enabled, it will automatically parse the written data and set the index fields.
- Auto Configuration by Input Sample: After disabling **Dynamic Generation**, you can use **Auto Configuration by Input Sample** to generate the field mappings of the index. Enter a sample JSON data in the input field, and the platform will automatically validate it for you. After successful validation, the relevant fields will be mapped to the Field Mapping table.

Field Mapping splits the original data into multiple tokens by field (i.e., key:value) for index construction and performs retrieval based on this mapping. Details are as follows:

Parameter	Feature Description
Field Name	Field names in the written data
Field Type	Field data types supported in the interface: "text,date,boolean,keyword,long,double,integer,ip,geo_point"—a total of 9 types. More field types are supported in **JSON Edit Mode**. For details, refer to the Official Documentation
Allow Chinese Characters	This feature can be enabled when the field contains Chinese and you need to retrieve the Chinese content. After enabling, the ik_max_word tokenizer will be used by default for the text field
Enabling index	After enabling, an index will be built for the field to be used for retrieval
Enable Statistics	After enabling, statistical analysis can be performed on the field values, which will increase index storage

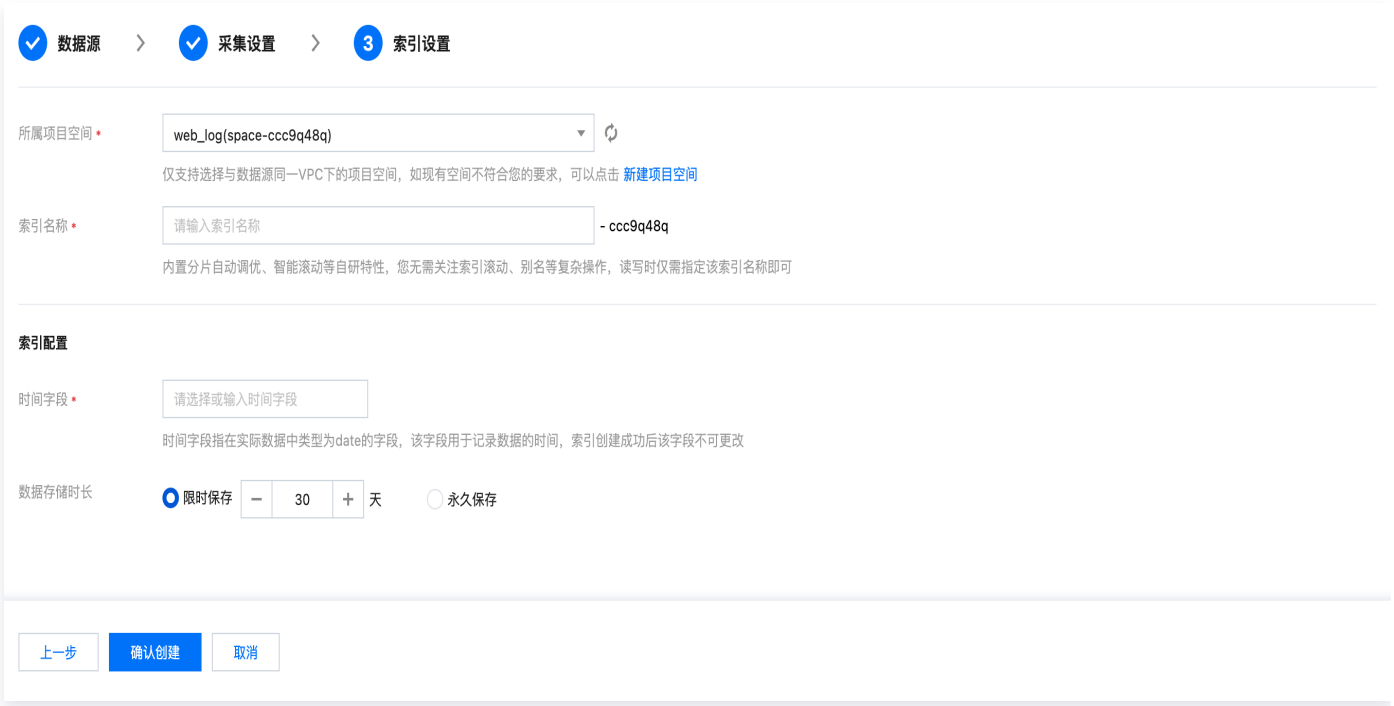
- **Time Field**
  The time field refers to the field of the type date in the actual data. Once the index is successfully created, this field cannot be changed.

> ⚠ **Note:**
> The time field is enabled by default with **Indexing Enabled** and **Statistics Enabled**, and cannot be turned off.

- **Data Storage Duration:**

1.1 You can set the data storage duration. By default, it is set to 30 days, and permanent storage is also supported.

Search for the required CAM policy as needed, and click to complete policy association.



1.2 After filling in the information correctly, click **Confirm Create** to complete TKE log collection.

EMR Log access

Last updated: 2024-10-25 09:09:53

Prerequisites

- You have a Tencent Cloud account. For more information on how to create an account, please see Signing up for a Tencent Cloud Account.
- If using a sub-account to log in, please ensure the account has read and write permissions for ES.

Operation Steps

Logging in to the Console

1. Log in to the ES Console.
2. In the left sidebar under Serverless mode, select **Log Analysis**.

Create Project Space

1. click **Create a new space**.
2. Enter the project **space name**, supporting 1 - 20 Chinese characters, English letters, numbers, underscores, or the separator "-".
3. click **Confirm**. If the verification is correct, the project space will be successfully created.

Search for the required CAM policy as needed, and click to complete policy association.

Note:
In ES Serverless log analysis, you can create an index, and then write data through the API or access data sources such as CVM or TKE on the corresponding index's "Data Access" Tab page. You can also access data while creating the index to accomplish one-stop CVM log access, TKE log access, and EMR (EMR) log access. Below is an introduction to the one-stop EMR log access operation.

EMR log access

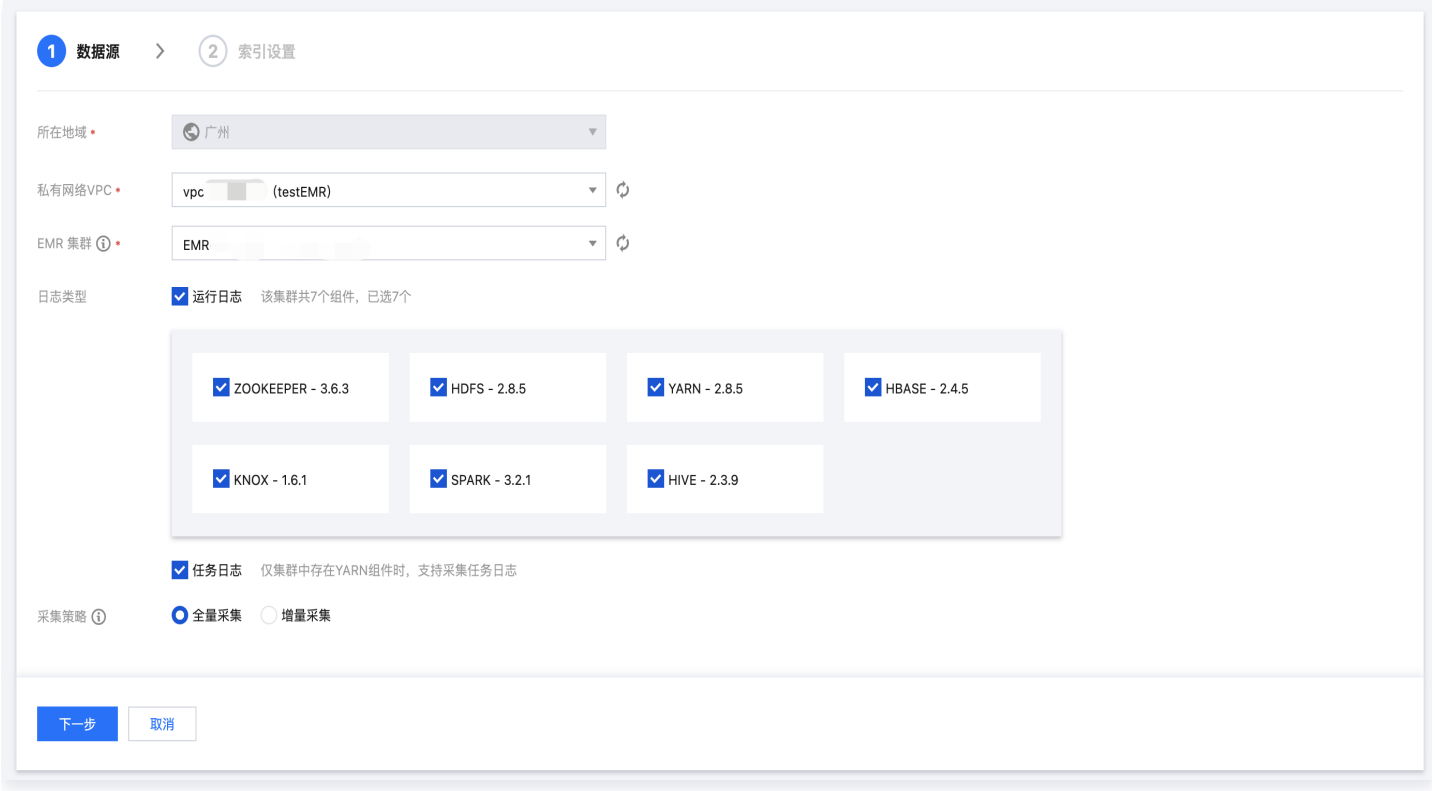
1. On the ES Serverless log analysis homepage, select **EMR** to go to the EMR log access page.

Search for the required CAM policy as needed, and click to complete policy association.

2. Go to the data source settings page, configure the data source, and after completing the configuration, click **Next**.

- **Region**: Select the region of the EMR cluster. If you enter this page from a project space details page, the default is the same as the project space region.
- **VPC VPC**: The VPC where the EMR cluster is located.
- **EMR Cluster** : The EMR cluster from which logs need to be collected.
- **Log Type** : You can specify the component's running logs to be collected. For instance, if there is a YARN component in the cluster, task logs can be collected.
- **Collection Strategy** : Supports both full and incremental collection. If incremental collection is selected, only logs generated after data access will be collected.

Search for the required CAM policy as needed, and click to complete policy association.



3. Enter the index settings page to set up the index.

- **Region Location** : The region where the project space is located, default is consistent with the EMR cluster.
- **Project Space** : You can assign all indexes of the same business to a project space for easier management.
- **Index Name** : Length is 1 – 100, supports lowercase letters, numbers, –, _, ;, @, &, =, !, ', %, $, ., +, (,).
- **Field Mapping**
  - Dynamic Generation: Enabled by default. Once enabled, it will automatically parse the written data and set the index fields.
  - After disabling **Auto Generate** for input sample configuration, you can use **Input Sample Auto Configuration** to generate field mappings for the index. Enter a JSON formatted data sample in the input box. After confirmation, the platform will automatically validate it. Once validated correctly, the relevant fields will be mapped to the field mapping table.

Field Mapping splits the original data into multiple tokens by field (i.e., key:value) for index construction and performs retrieval based on this mapping. Details are as follows:

Parameter	Feature Description

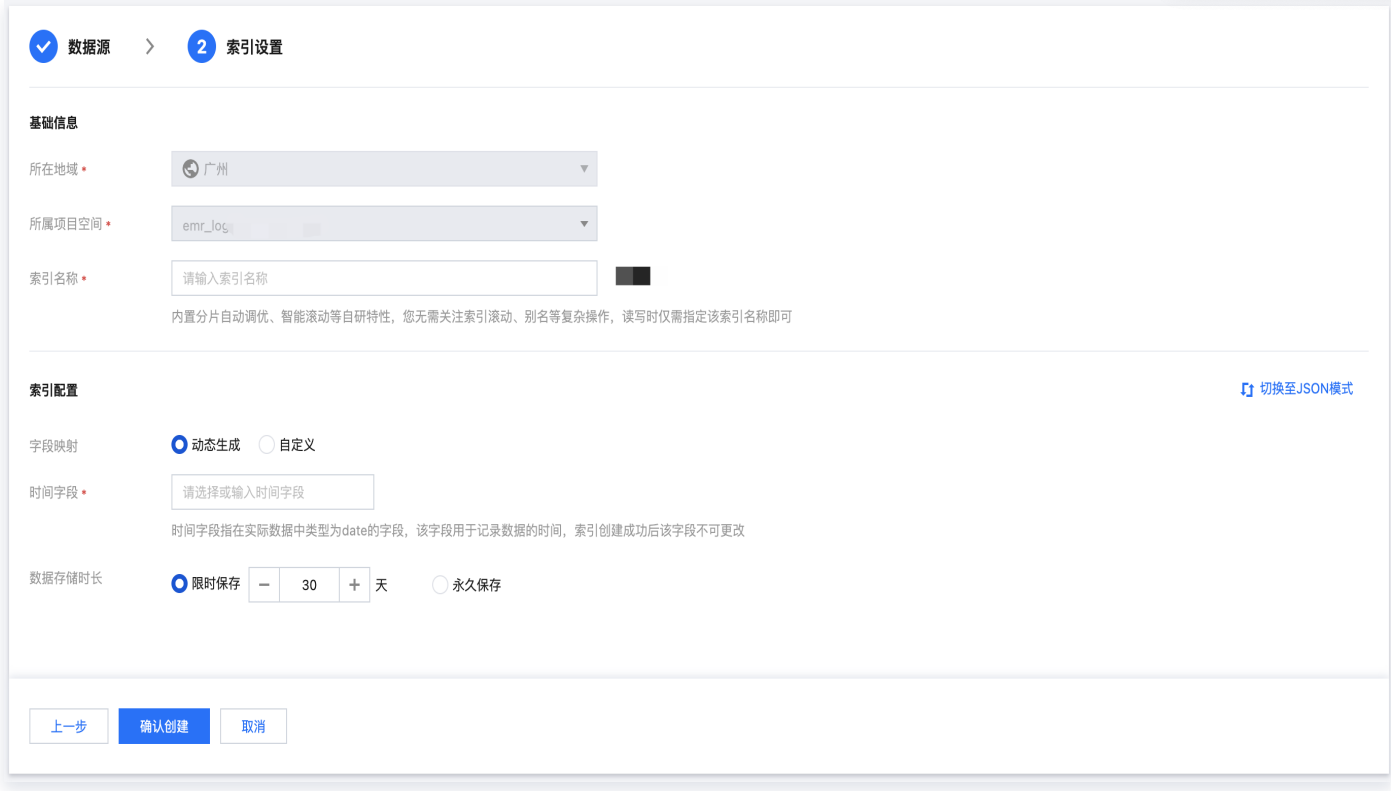
Field Name	Field names in the written data
Field Type	Field data types supported in the interface: "text,date,boolean,keyword,long,double,integer,ip,geo_point"—a total of 9 types. More field types are supported in **JSON Edit Mode**. For details, refer to the Official Documentation
Allow Chinese Characters	This feature can be enabled when the field contains Chinese and you need to retrieve the Chinese content. After enabling, the ik_max_word tokenizer will be used by default for the text field
Enabling index	After enabling, an index will be built for the field to be used for retrieval
Enable Statistics	After enabling, statistical analysis can be performed on the field values, which will increase index storage

- **Time Field**
  The time field refers to the field of the type date in the actual data. Once the index is successfully created, this field cannot be changed.

  > ⚠ **Note:**
  > The time field is enabled by default with **Indexing Enabled** and **Statistics Enabled**, and cannot be turned off.

- **Data Storage Duration:**
  You can set the data storage duration. By default, it is set to 30 days, and permanent storage is also supported.

Search for the required CAM policy as needed, and click to complete policy association.



TCHouse-C Cluster Log Access

Last updated: 2024-10-25 09:10:24

The ES Serverless service supports the collection of TCHouse-C node logs for easier problem localization and analysis. For more details, please refer to Log Search.

TCHouse-D Cluster Logs Access

Last updated: 2024-11-12 21:29:20

The ES Serverless service supports the collection of TCHouse-D node logs for problem identification and analysis. For more details, please refer to Log Search and Analysis.

Custom Definition Filebeat Data Access

Last updated: 2024-10-25 09:10:59

Self-built Filebeat Data Collection

Version Explanation

Only supports Filebeat versions 7.10.2 or 7.14.2.

Category	Parameter	Parameter Description	Instructions
Elasticsearch template setting	setup.template.enabled	Index Template	Boolean type, set to false, currently not supported
	setup.ilm.enabled	Index Lifecycle Management	Boolean type, set to false, currently not supported
	allow_older_versions	Version Compatibility with ES	Boolean type, can be set to "true" or "false"
output	protocol	Data transfer protocol	String Type, default is "http", can be set to "https"
	hosts	Index private network address	Array type, If protocol is set to "http", then the port number is 80, for example: ["http://index-xxx.qcloudes.com:80"] If protocol is set to "https", then the port number is 443, for example: ["https://index-xxx.qcloudes.com:443"]

Configuration Note

```
# ============================== Filebeat inputs
===============================

filebeat.inputs:
- 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
# ============================== Filebeat modules
==============================

filebeat.config.modules:
  # Glob pattern for configuration loading
  path: \${path.config}/modules.d/*.yml

  # Set to true to enable config reloading
  reload.enabled: false

  # Period on which files under path should be checked for changes
  #reload.period: 10s

# ======================== Elasticsearch template setting
========================
setup.template.enabled: false
setup.ilm.enabled: false
  #template setting's value is set to false by default. If you set it to
true, an error will be reported when the configuration is submitted


# ================================== General
==================================

# The name of the shipper that publishes the network data. It can be
used to group
# all the transactions sent by a single shipper in the web interface.
#name:

# The tags of the shipper are included in their own field with each
# transaction published.
#tags: ["service-X", "web-tier"]

# Optional fields that you can specify to add additional information to
the
# output.
#fields:
#  env: staging
```

```
# ================================== Processors
==================================
processors:
  - add_host_metadata:
      when.not.contains.tags: forwarded

# ================================== Logging
==================================

# Sets log level. The default log level is info.
# Available log levels are: error, warning, info, debug
#logging.level: debug

# At debug level, you can selectively enable logging only for some
components.
# To enable all selectors use ["*"]. Examples of other selectors are
"beat",
# "publisher", "service".
#logging.selectors: ["*"]
############################## output
#####################################
output.elasticsearch:
  # Array of hosts to connect to.
  allow_older_versions: true
  protocol: "http"
  hosts: ["Index intranet access address"]

  # Authentication credentials - either API key or username/password.
  username: "your index username"
  password: "your index password"
  indices:
  - index: The_index_name
    when.equals:
      fields.type: log
```

Logstash Data Delivery

Last updated: 2024-10-25 09:11:27

This document introduces the operations for delivering data to the Elasticsearch Serverless service using Logstash.

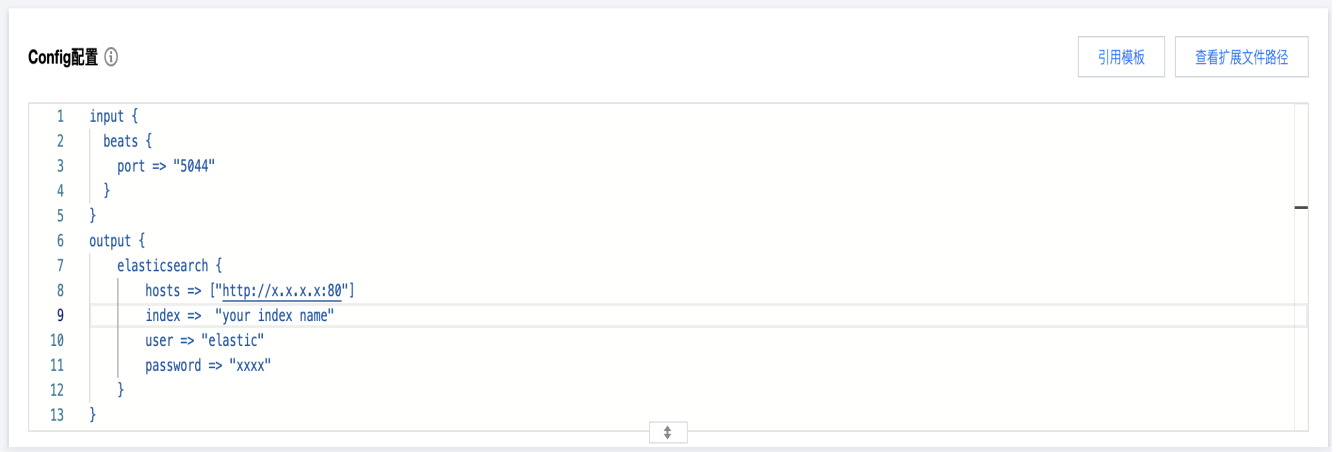
Prerequisites

- Log in to Elasticsearch Console, and go to the **Logstash Management** interface.
- If you already have a Logstash instance, you can directly go to Next. If you have not purchased a Logstash instance, click **New Instance** to go to the Logstash purchase page. Choose the same VPC as the index, select Logstash version 7.10.2 or 7.14.2, and for advanced features, select "X-Pack Edition", then click **Buy Now**.

Directions

1. In the console Logstash Management interface, click the corresponding Logstash name. After entering the sub-page, click **Pipeline Management > New Pipeline**.
2. In the Config configuration, enter the relevant information for the input and output ends. The configuration description for the output end is as follows:
   - hosts: Intranet Access Address for Index.
   - index: Index Name.
   - user: Index Username, you can get it from the access control module of the corresponding index.
   - password: Index Password, you can get it from the access control module of the corresponding index.

     In Parameter Configuration, Pipeline ID is a key item. You can choose whether to fill in other parameters based on your business scenario. After completing the entries, click **Save and Deploy**, and wait for the status to show **Running**.



Python SDK

Last updated: 2024-12-30 16:21:09

This article introduces how to use the Python SDK to send data to the Elasticsearch Serverless service.

Prerequisites

- The Serverless space and index have been created. Obtain the Intranet Access Address, Username, Password, Index Name, and other information.
- It is recommended to install the elasticsearch py library version 7.10.1 or above. The latest version can be installed as shown below:

```
pip3 install elasticsearch
```

Code demo

The following example code features writing to the Elasticsearch Serverless service through the Python SDK and performing a simple query.

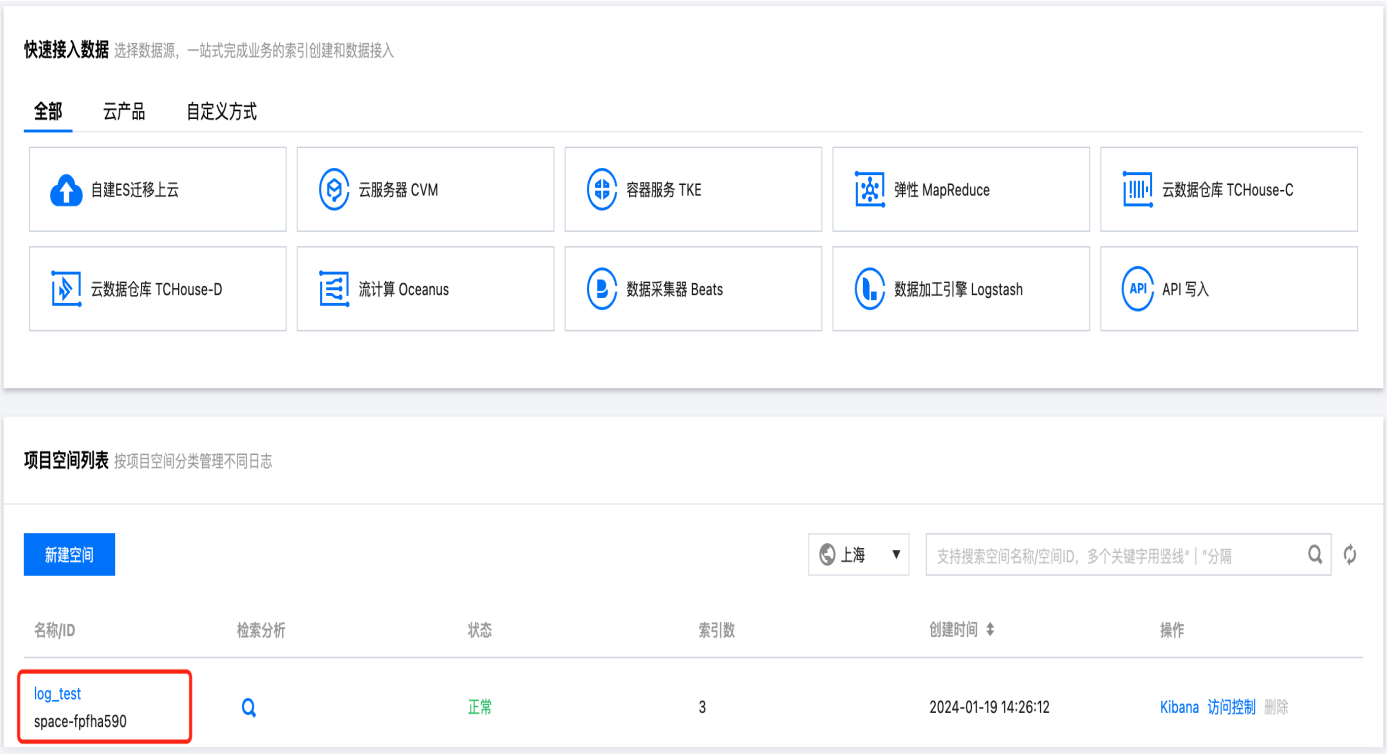
```
from datetime import datetime, timezone
from elasticsearch import Elasticsearch, helpers

# create an elasticsearch client instance, provide http authentication
information
es = Elasticsearch(
    hosts=[
        {
            'host': 'space-12345678.ap-shanghai.qcloudes.com',  #
replace it with your intranet access address
            'port':  80,  # replace it with the port number of your
intranet access address
            'scheme': 'http',  # specify the scheme here
        }
    ],
    http_auth=('elastic', 'xxxxxxxxxxx'), # replace it with your
username and password
)

# Define the document to index
docs = [
```

```
    {
        '_index': 'my_index',
        # '_id': i,
        '_source': {
            'author': 'user',
            'text': 'Hi! Elasticsearch Serverless!',
            '@timestamp': datetime.now(timezone.utc),
        }
    }
    for i in range(10)  # Create 10 documents
]

# Use bulk API to index documents
helpers.bulk(es, docs)

# Search documents, replace index with your index name, replace body
with your query statement
res = es.search(index="my_index", body={"query": {"match_all": {}}})
print("Got %d Hits:" % res['hits']['total']['value'])
for hit in res['hits']['hits']:
    print("%(@timestamp)s %(author)s: %(text)s" % hit["_source"])
```

Access Control

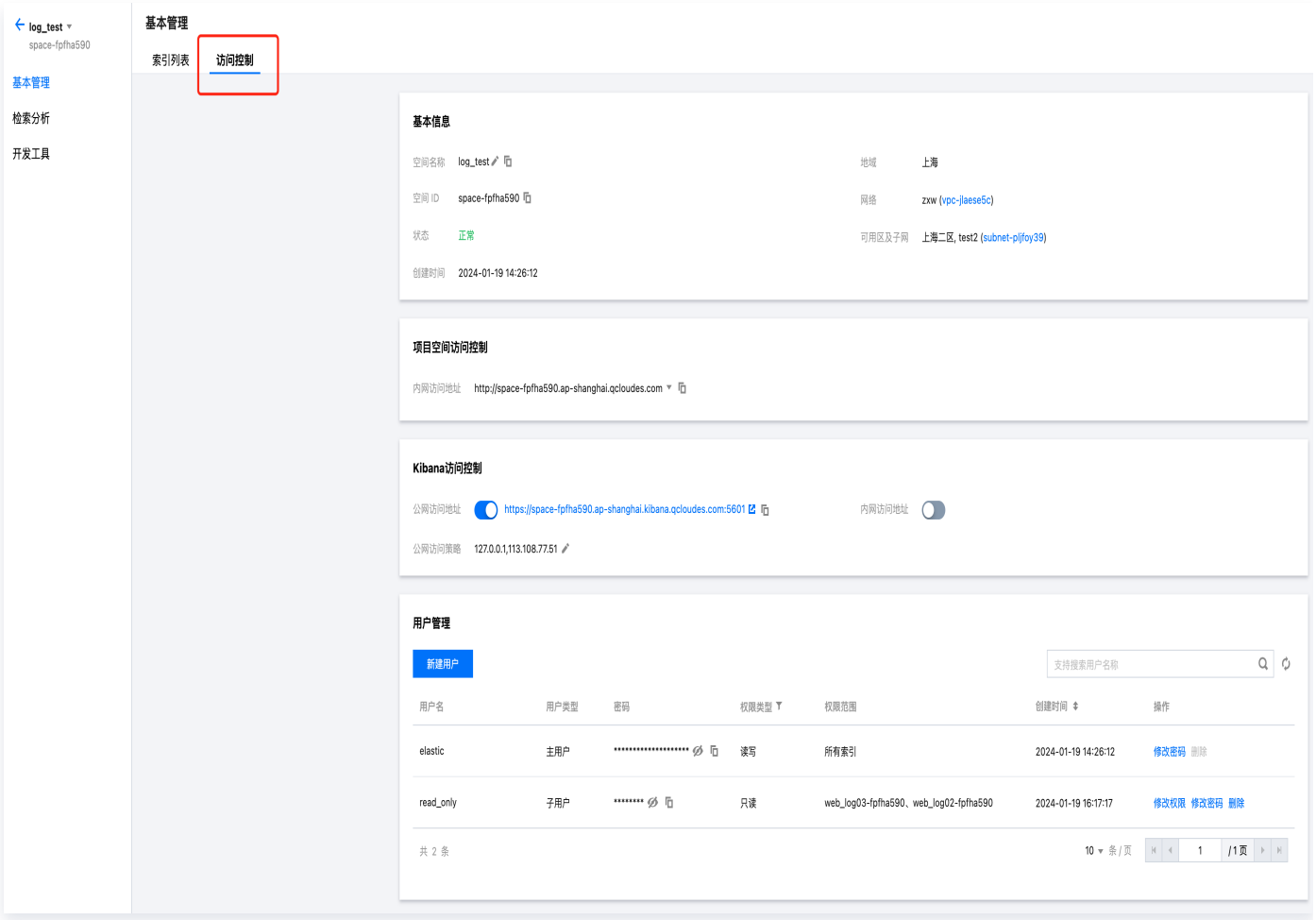
Last updated：2024-10-25 09:12:09

1. In the space list, click the space name to enter the Basic Information page.
   Search for the required CAM policy as needed, and click to complete policy association.



2. Then click **Access Control** to enter the Access Control page.

Search for the required CAM policy as needed, and click to complete policy association.



3. In the Access Control module, the following operations are supported:
   - View the project space's internal network access address, **This address can be used for data writing or querying.**
   - Enable/Disable Kibana intranet access or public network access.
   - Modify the Kibana public network access IP allowlist. It supports multiple IPs, which can be separated by commas, semicolons, or newline characters in English, in formats like 192.168.0.1, 192.168.0.0/24, with a maximum of 50 IPs supported. If you are not aware of your current IP address, you can click **click here to automatically obtain your current IP address** for automatic retrieval and entry.

> **Note:**
> Configuring 127.0.0.1 to forbid access from all IPv4 addresses. For security reasons, setting the IP allowlist to 0.0.0.0 is not allowed. If there are special requirements, you can submit a ticket for consultation.

4. Modify Master/Sub-user password. On the user management page, click **Modify password** to change the index access password.

Search for the required CAM policy as needed, and click to complete policy association.

5. Modify Sub-user permissions. On the user management page, click **Modify Permissions**, then select the permission type and scope. Read-only and read-write permission types are supported. You can choose from all indexes under this space from the dropdown menu.

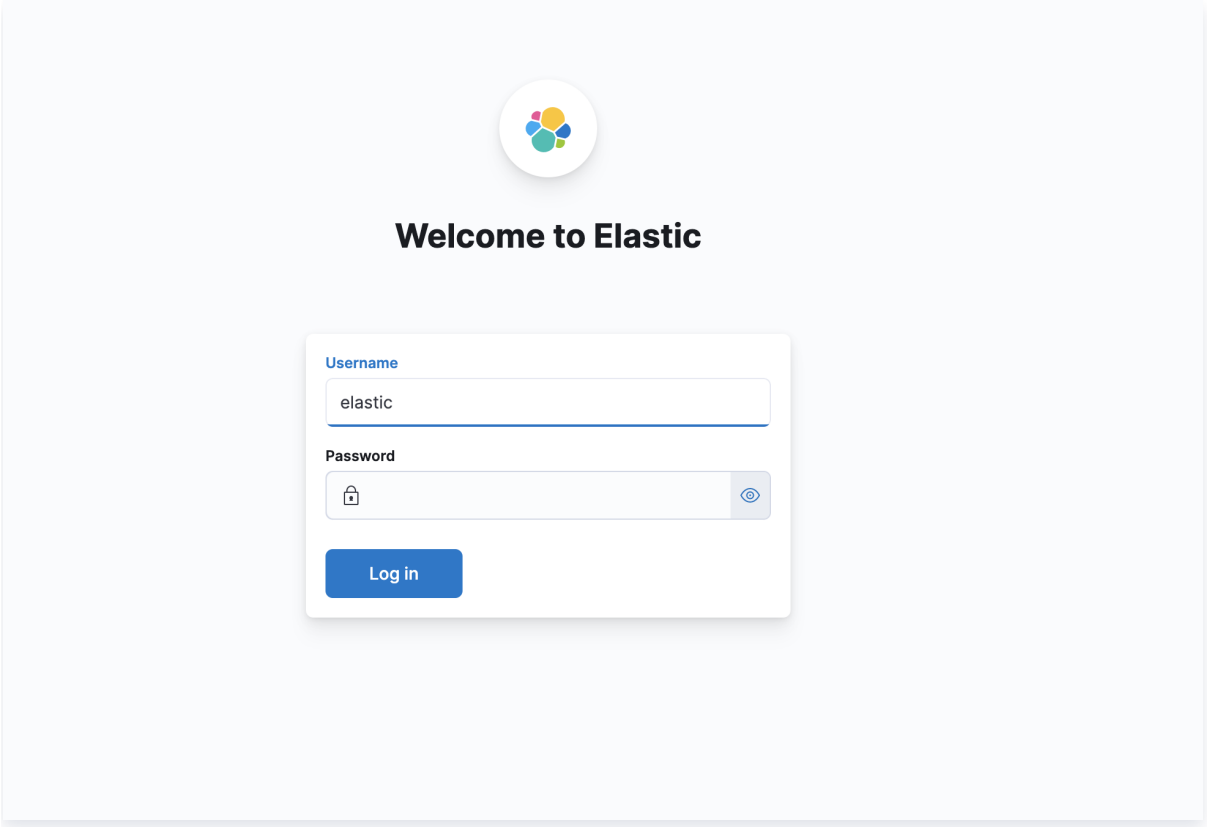
Search for the required CAM policy as needed, and click to complete policy association.

Search for the required CAM policy as needed, and click to complete policy association.

6. log in to Kibana

After enabling Kibana public network access and setting the Kibana public network access IP allowlist, click the Kibana public network access address to enter the Kibana login interface. Enter the Sub-user username and password for this space, and click **Log in** to successfully

log in to Kibana.



Writing Data

Last updated：2024-10-25 09:14:12

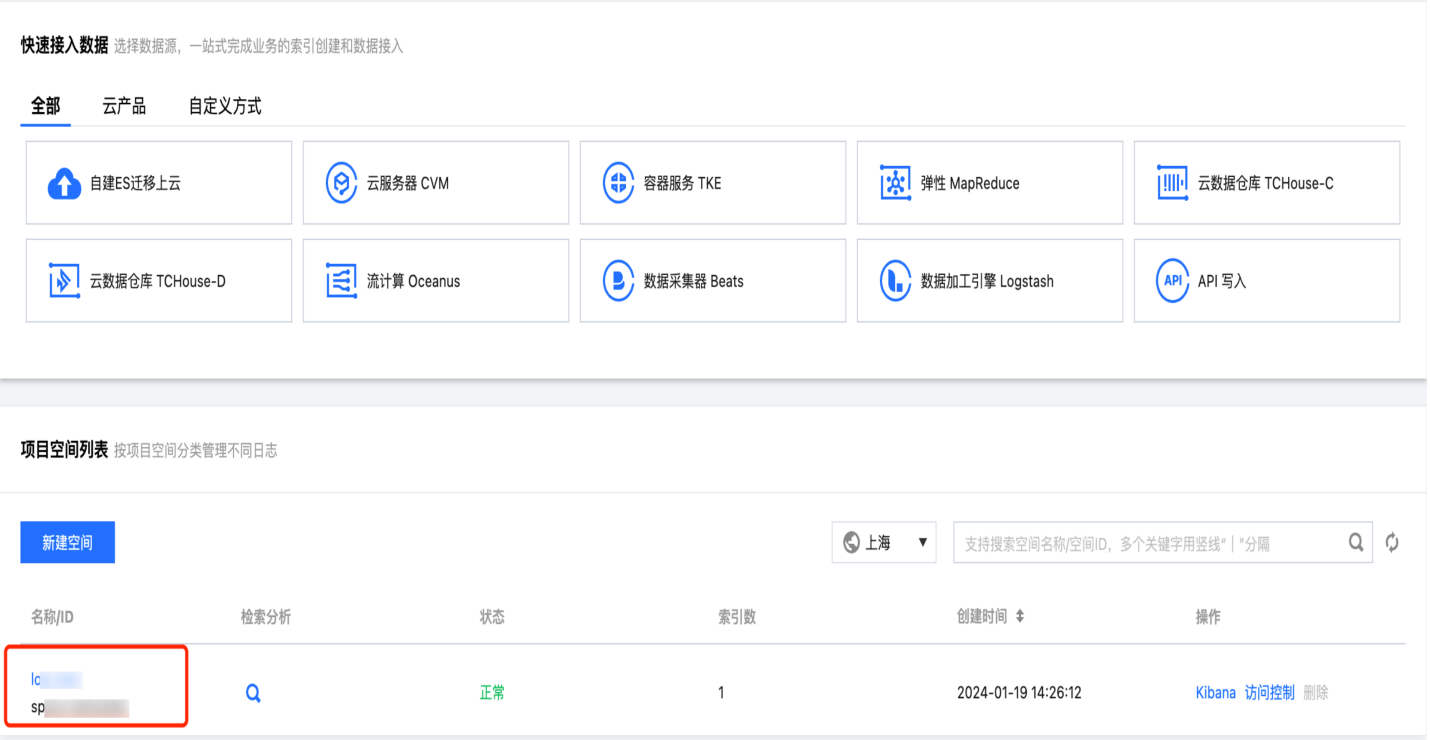
Overview

The ES Serverless Service supports writing data into indexes through **ES native API, Logstash, Flink, Kafka**, etc. If you have log collection needs for CVM CVM , , TKE TKE , Cloud Data Warehouse TCHouse-C , etc., it also supports one-stop visual configuration on the interface. You only need to set the data source and index information to collect logs into the index and perform search and analysis quickly. This article introduces the related operations to write a single document and bulk write documents through **Kibana** and **Curl commands**.

Access Control

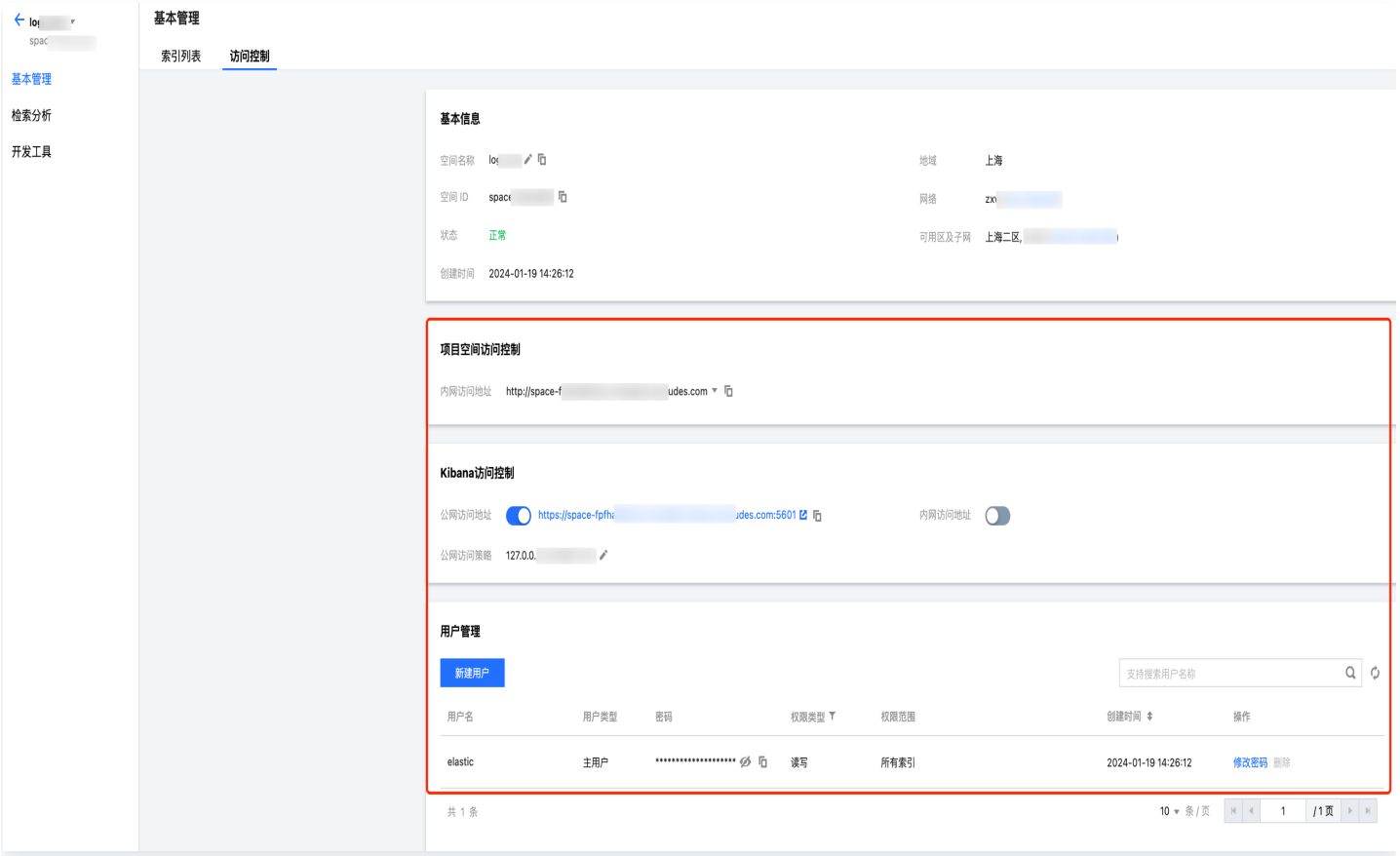
1. In the space list, click the corresponding space name to enter the basic space information page.

Search for the required CAM policy as needed, and click to complete policy association.



2. In the **Access Control** module, we can obtain the index's username and password, intranet access address, Kibana intranet access address, and Kibana public network access address. Meanwhile, we can set the Kibana public network access policy.

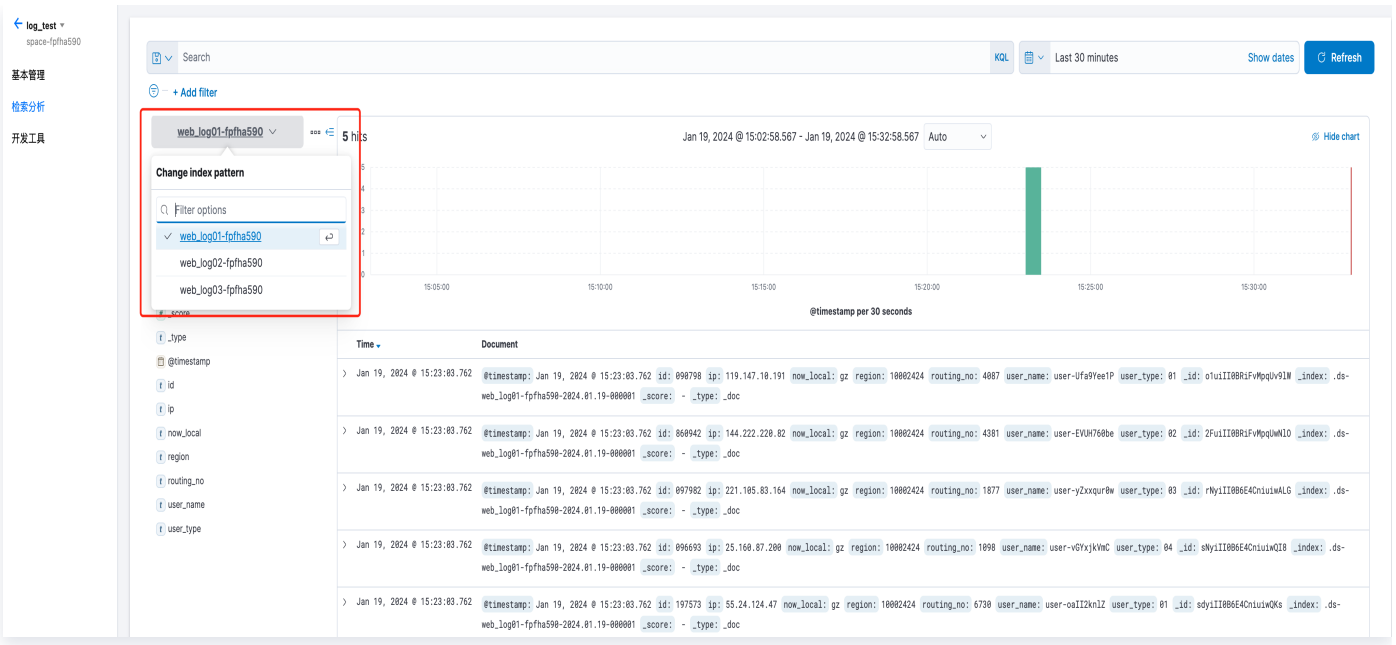
Search for the required CAM policy as needed, and click to complete policy association.



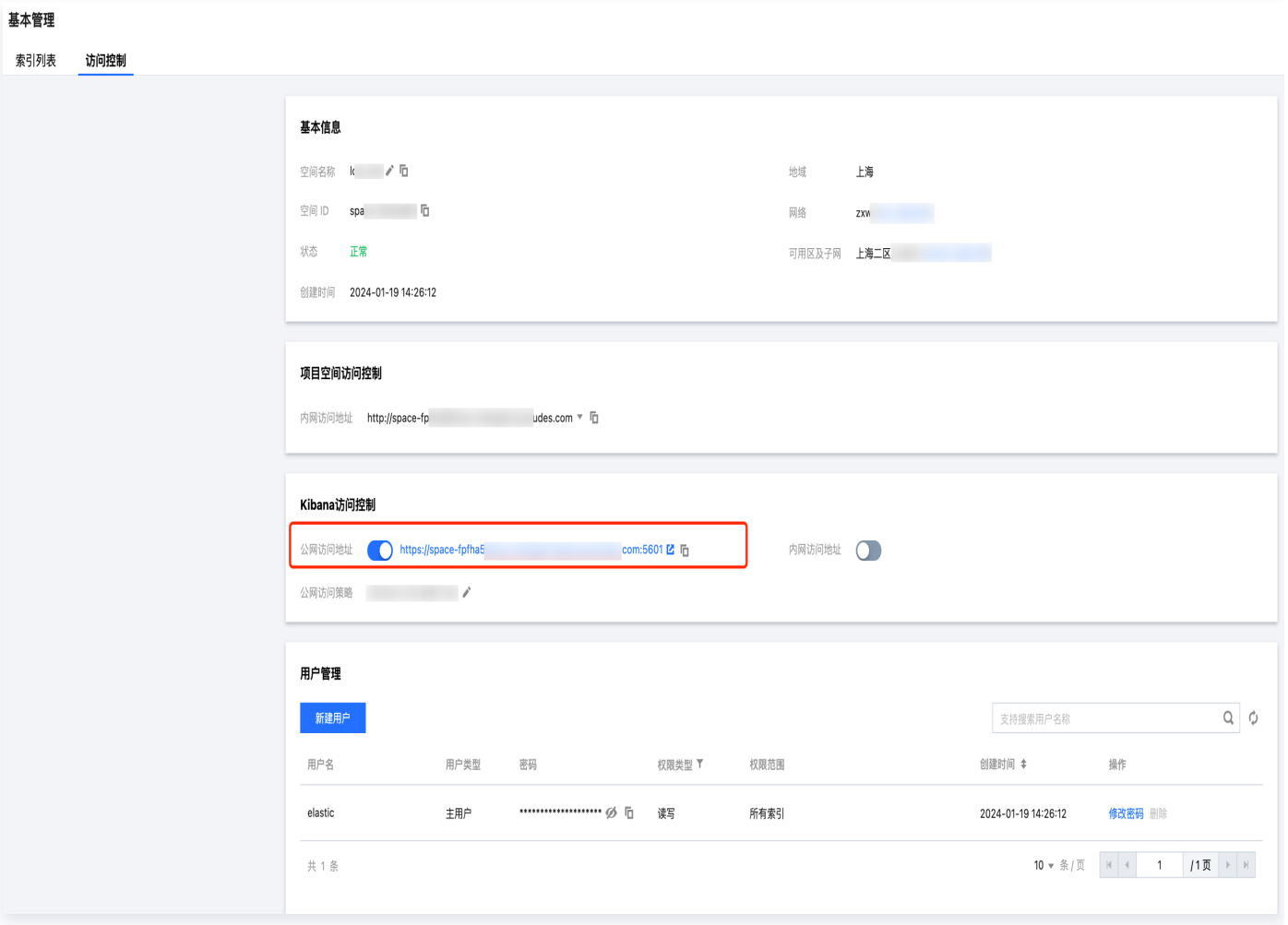
3. **Access Kibana**: The **Discover** and **Dev tools** features of Kibana are embedded in the Tencent Cloud Console. Therefore, we can use the search and analysis capabilities directly in the console or access Kibana via external links.

- **Through the Console**: In the space details page sidebar, click **Search and Analysis** to enter the related page. Click the index pattern drop-down list on the left side of Search and Analysis to switch different index views. **Log Search** corresponds to **Discover**, and **Development Tools** corresponds to **Dev tools**.

> **Note:**
> Embedded capability requires browser support for third-party cookies. If you encounter issues, please try enabling third-party cookies in your browser settings.

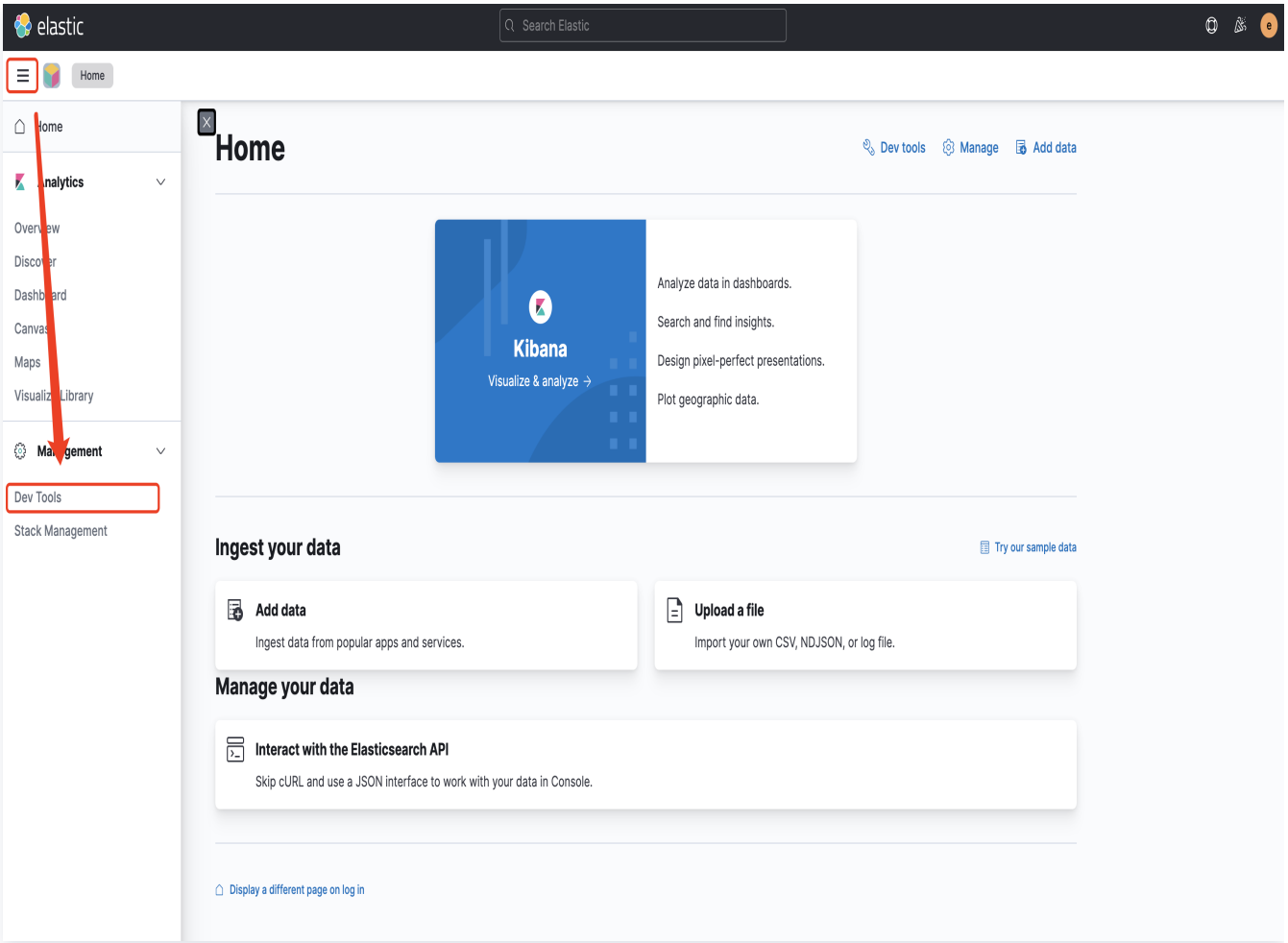


- **Through Kibana public network access address:** click **Kibana public network access address** to enter the Kibana page.



  - On the Kibana log in page, enter username and password. The username and password can be copied directly from the user management page.

- After entering the Kibana page, click the three bars on the upper right and click **Dev tools** to enter the Developer Tools page.



Note:
Kibana public network access has an allowlist access mechanism, meaning IPs not in the access policy cannot access Kibana to enhance access security. If the page prompts "Sorry, you do not have permission to access", click "**Kibana Public Network Access Policy**" in the picture above, and in the pop-up window

click "**Click to automatically obtain the current IP address**" to fill in the current IP address.

Write a single document

Through Kibana Dev Tools

```
POST /index_name/_doc
{
"@timestamp": "2023-09-28T11:06:07.000Z",
"user":{
    "id" : "8a4f500"
  },
"message": "Login successful"
}
```

Via command line

```
curl -X POST "project space access address/index name/_doc/?pretty" -H
'Content-Type: application/json' -d'
{
  "@timestamp": "2023-09-28T11:06:07.000Z",
  "user": {
```

```
    "id": "8a4f500d"
  },
  "message": "Login successful"
}
```

Search for the required CAM policy as needed, and click to complete policy association.

Note

- You cannot use `PUT /index_name/_doc/document_ID` format for the write request. To specify a document ID for writing, use `PUT /index_name/_create/document_ID` format.
- Please ensure that the data being written includes the **time field** set when creating the index.

Bulk Writing Documents

Through Kibana Dev Tools

```
PUT /index_name/_bulk?refresh
```

```
{"create":{ }}
{ "@timestamp": "2023-03-28T11:04:05.000Z", "user": { "id": "vlb44hny"
}, "message": "Login attempt failed" }
{"create":{ }}
{ "@timestamp": "2023-03-29T11:06:07.000Z", "user": { "id": "8a4f500d"
}, "message": "Login successful" }
{"create":{ }}
{ "@timestamp": "2023-03-30T11:07:08.000Z", "user": { "id": "l7gk7f82"
}, "message": "Logout successful" }
```

Via command line

```
curl -X PUT "project space access address/index name/_bulk?
refresh&pretty" -H 'Content-Type: application/json' -d'
{"create":{ }}
{ "@timestamp": "2023-03-28T11:04:05.000Z", "user": { "id": "vlb44hny"
}, "message": "Login attempt failed" }
{"create":{ }}
{ "@timestamp": "2023-03-29T11:06:07.000Z", "user": { "id": "8a4f500d"
}, "message": "Login successful" }
{"create":{ }}
{ "@timestamp": "2023-03-30T11:07:08.000Z", "user": { "id": "l7gk7f82"
}, "message": "Logout successful" }
'
```

Search for the required CAM policy as needed, and click to complete policy association.

Note

- Bulk operations only support `create` .
- Please ensure that the data being written includes the **time field** set when creating the index.

Data Query

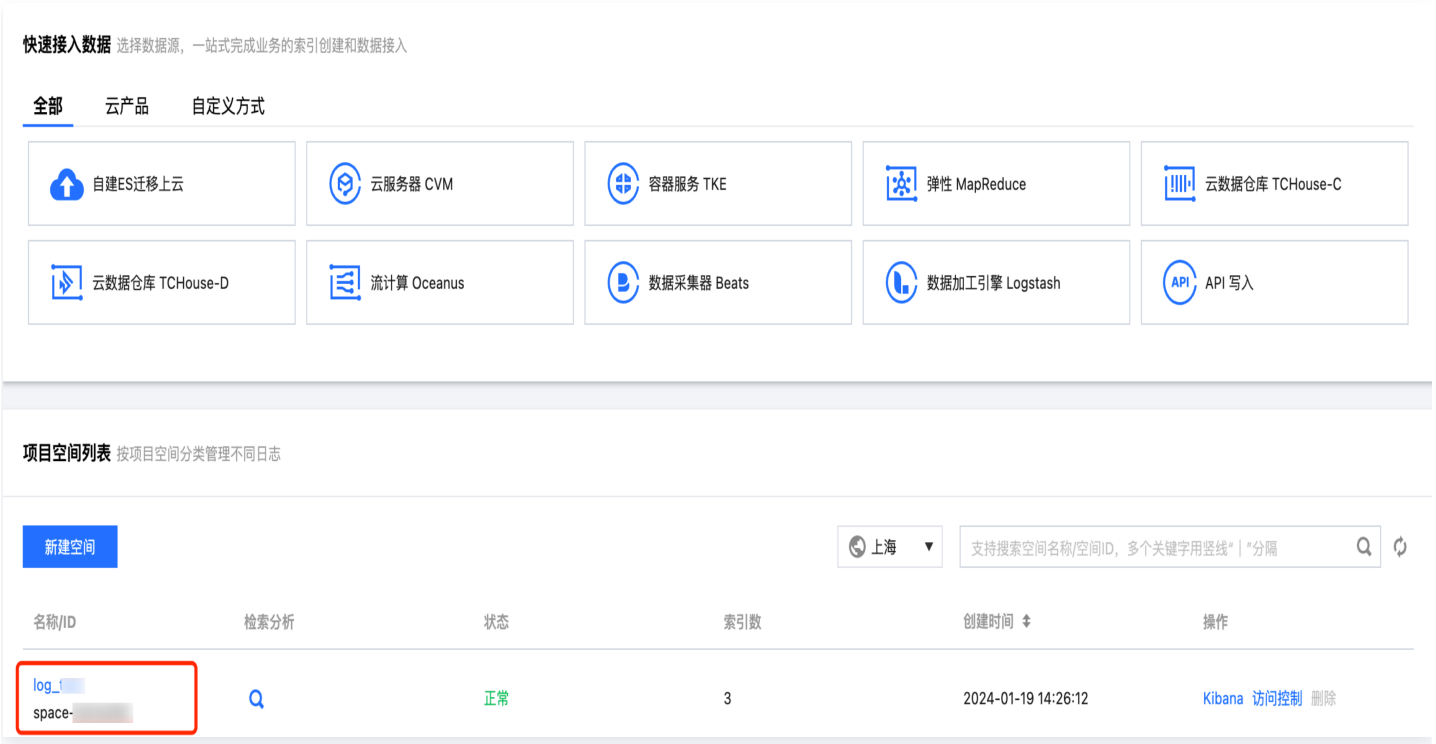
Last updated：2024-10-25 09:14:36

Overview

This document introduces data query operations through **Kibana** and **Curl Command Line**, among other methods.

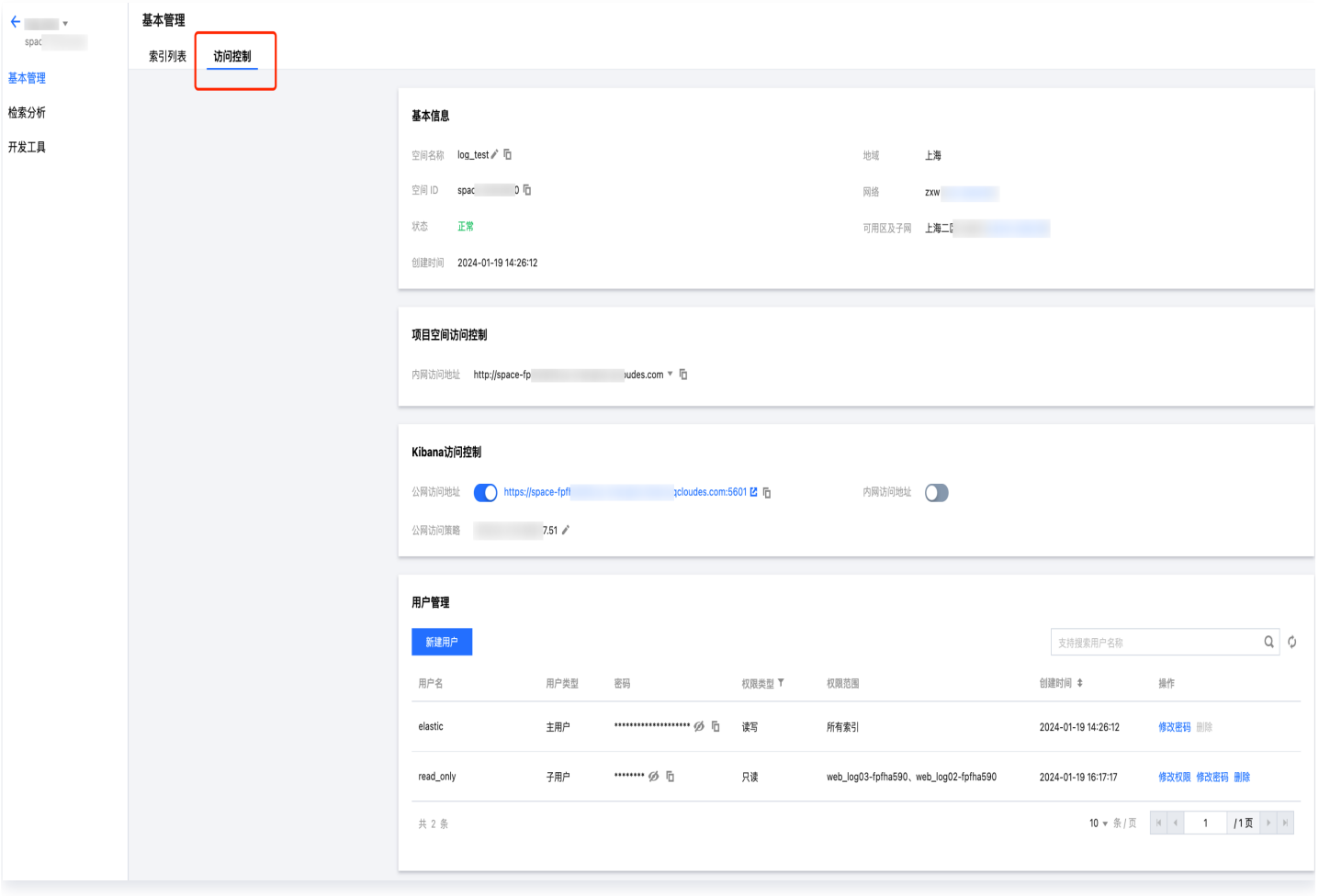
Access Control

1. In the space list, click the corresponding **Space Name** to enter the basic information page. Search for the required CAM policy as needed, and click to complete policy association.



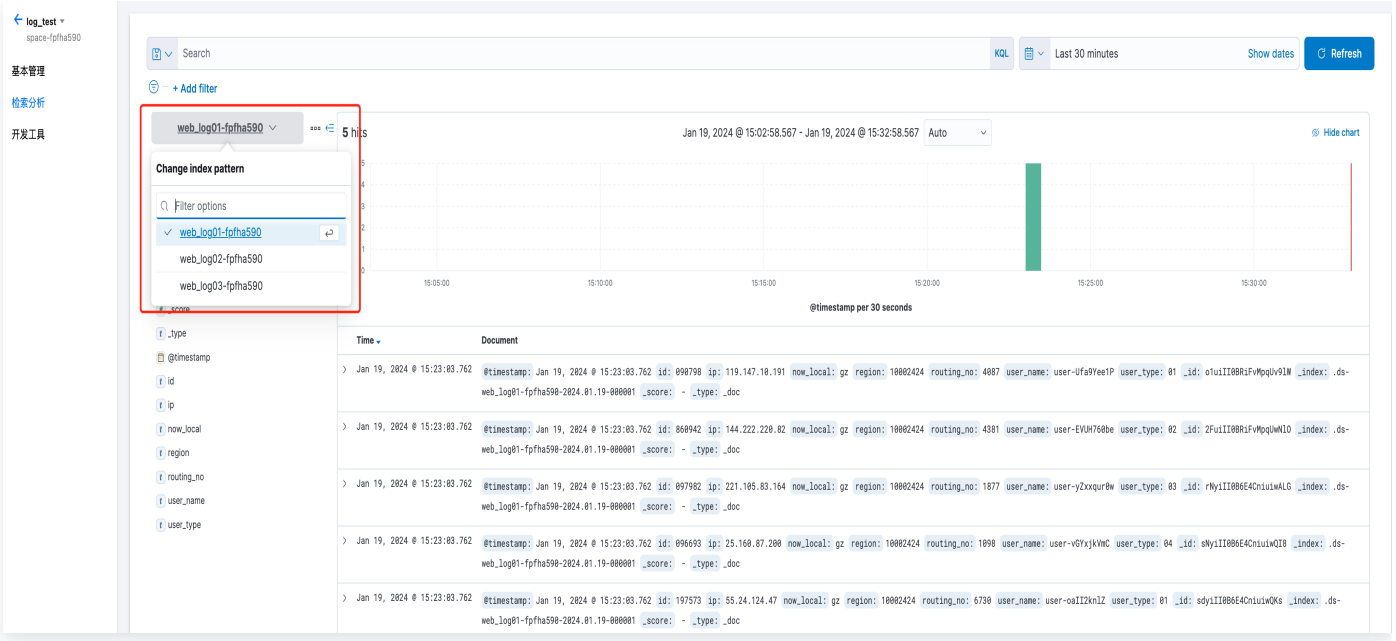
2. In the **Access Control** module, we can obtain the space's Sub-user Information (username, password, Permissions), Intranet Access Address, Kibana Intranet Access Address, and Kibana Public Network Access Address, and set the Kibana Public Network Access Policy.

Search for the required CAM policy as needed, and click to complete policy association.



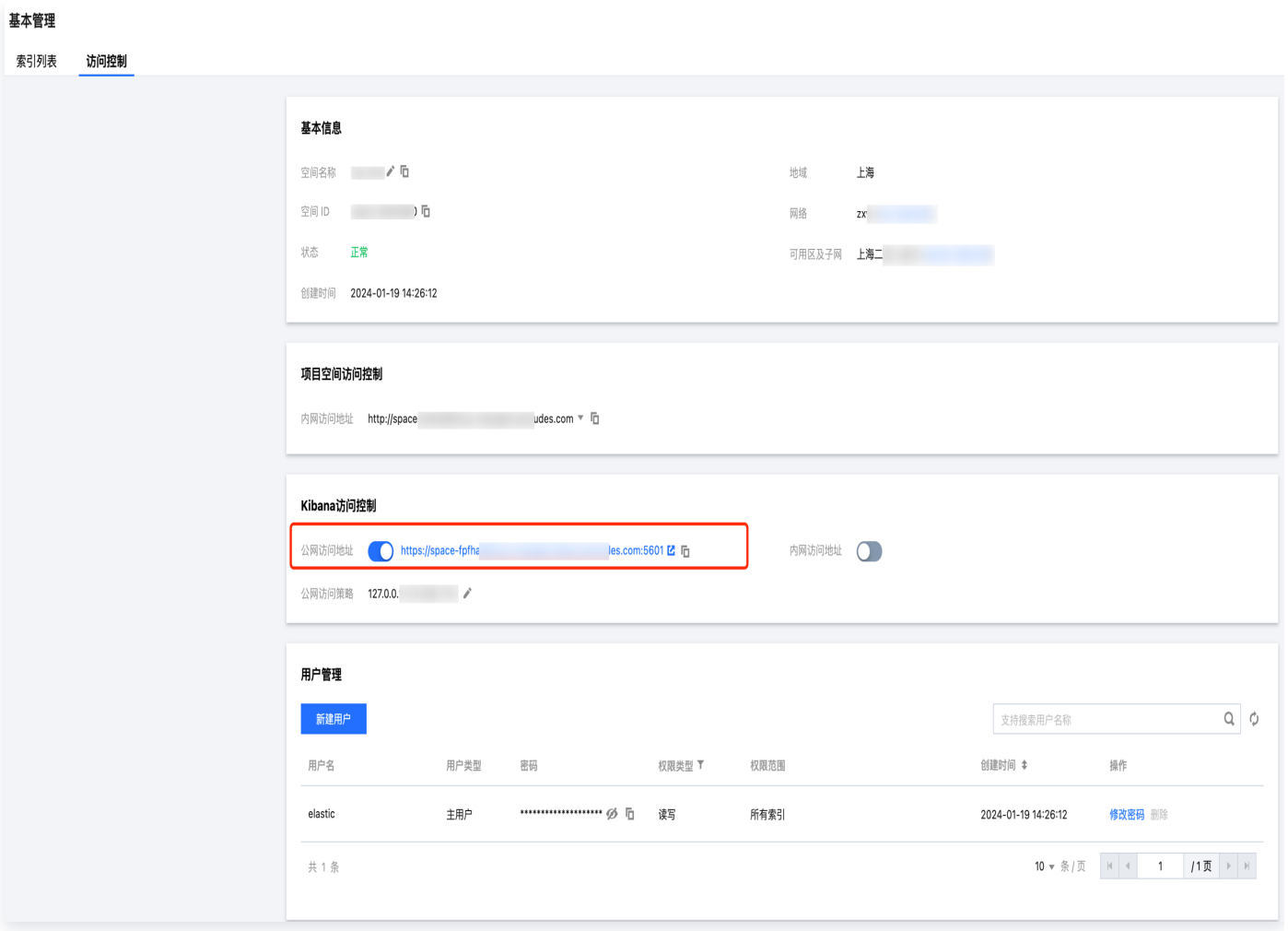
3. Access Kibana: Kibana's **Discover** and **Dev tools** features are embedded in the Tencent Cloud console, allowing us to use search and analysis capabilities directly in the console or access Kibana via external links.

- **Through the console**: Click the **Search and Analysis** on the sidebar to enter the related page directly. Click the index pattern dropdown on the left to switch between different index views. **Log search** corresponds to **Discover**, and **Development tools** corresponds to **Dev tools**.



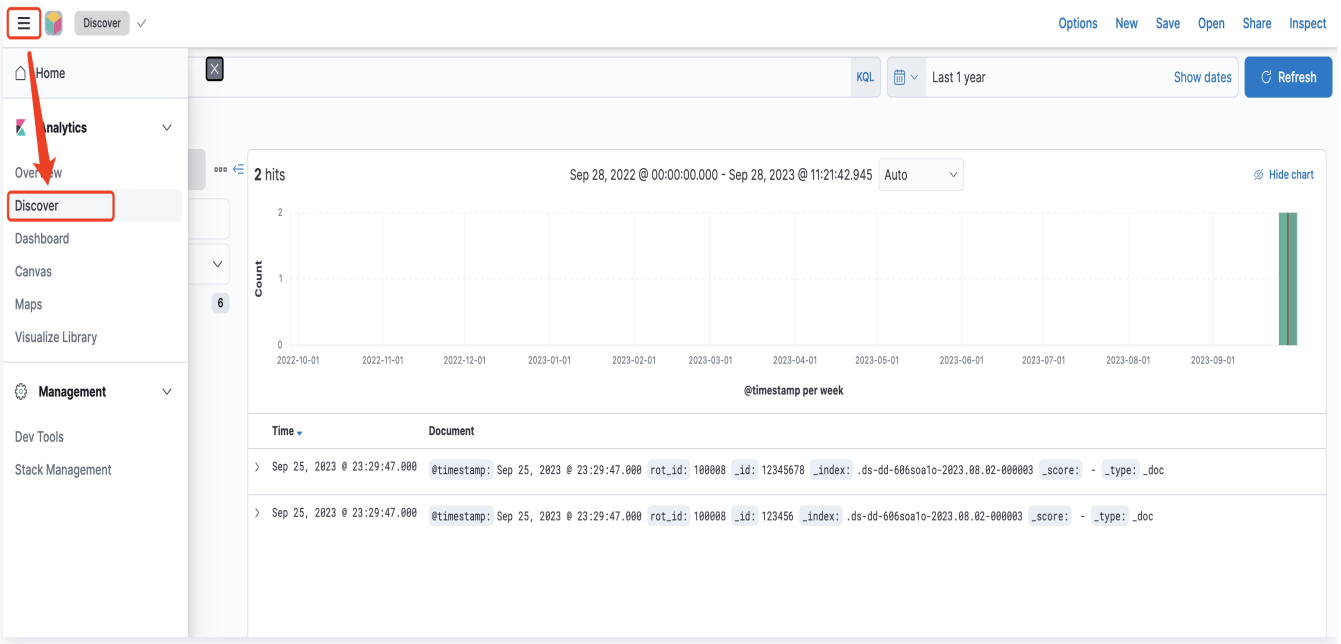
> **Note:**
> Embedded capability requires browser support for third-party cookies. If you encounter issues, please try enabling third-party cookies in your browser settings.

- **Through Kibana public network access address:** click **Kibana public network access address** to enter the Kibana page.



- On the Kibana log in page, enter username and password. The username and password can be copied directly from the user management page.

- After entering the Kibana page, click the three bars at the top right, then click **Discover** to go to the search and analysis page.



Note:

Kibana public network access has an allowlist access mechanism, meaning IPs not in the access policy cannot access Kibana to enhance access security. If the page prompts "Sorry, you do not have permission to access", click "**Kibana Public Network Access Policy**" in the picture above, and in the pop-up window click "**Click to automatically obtain the current IP address**" to fill in the current IP address.

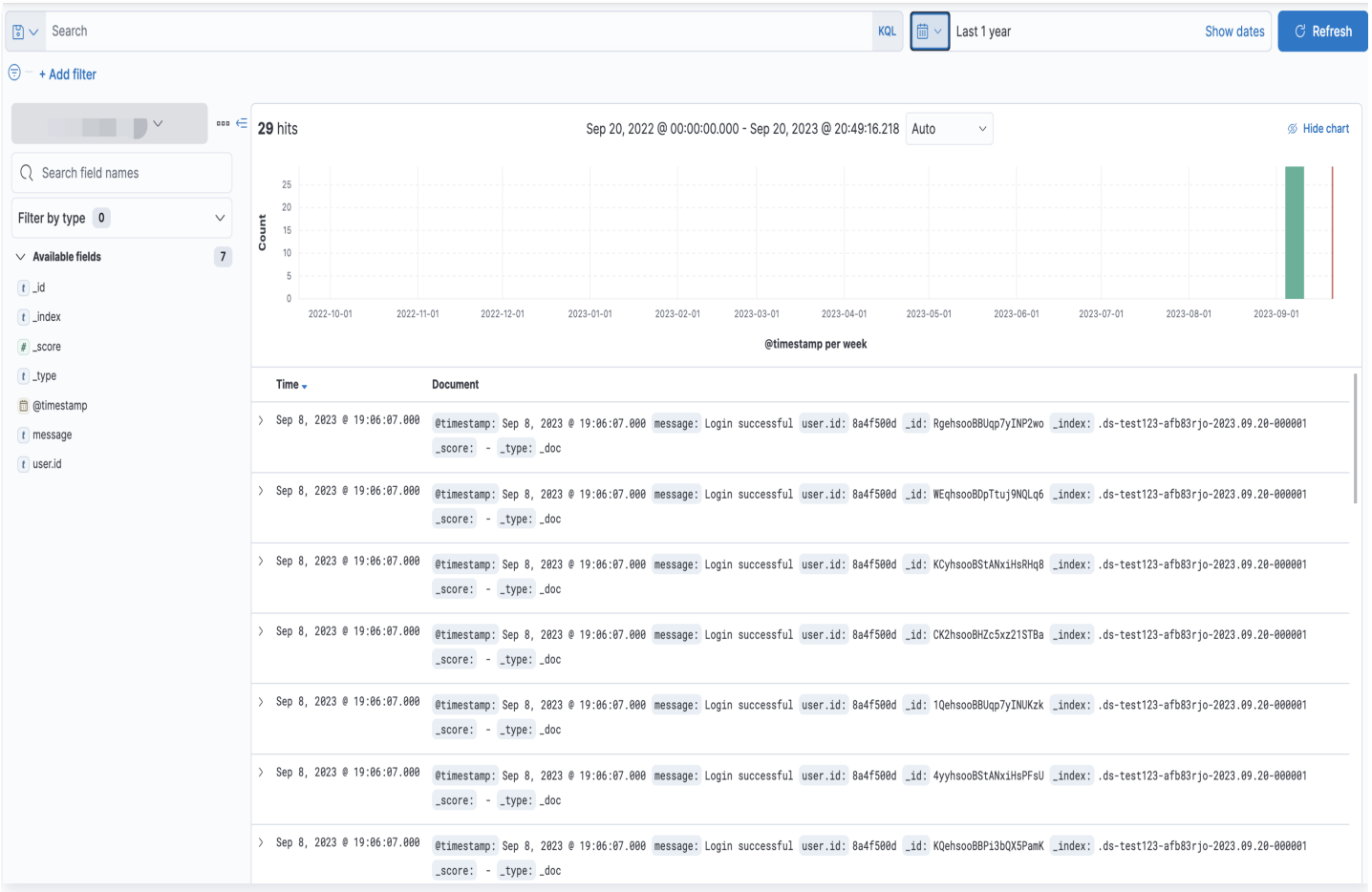
Search and Analysis

Via command line

```
curl -X GET "Index access address/index name/_search?pretty" -H
'Content-Type: application/json' -d'
{
  "query": {
    "term": {
      "user.id": "kimchy"
    }
  }
}
'
```

Via Discover

On the Discover page, you can perform time filtering, keyword searches, etc.:



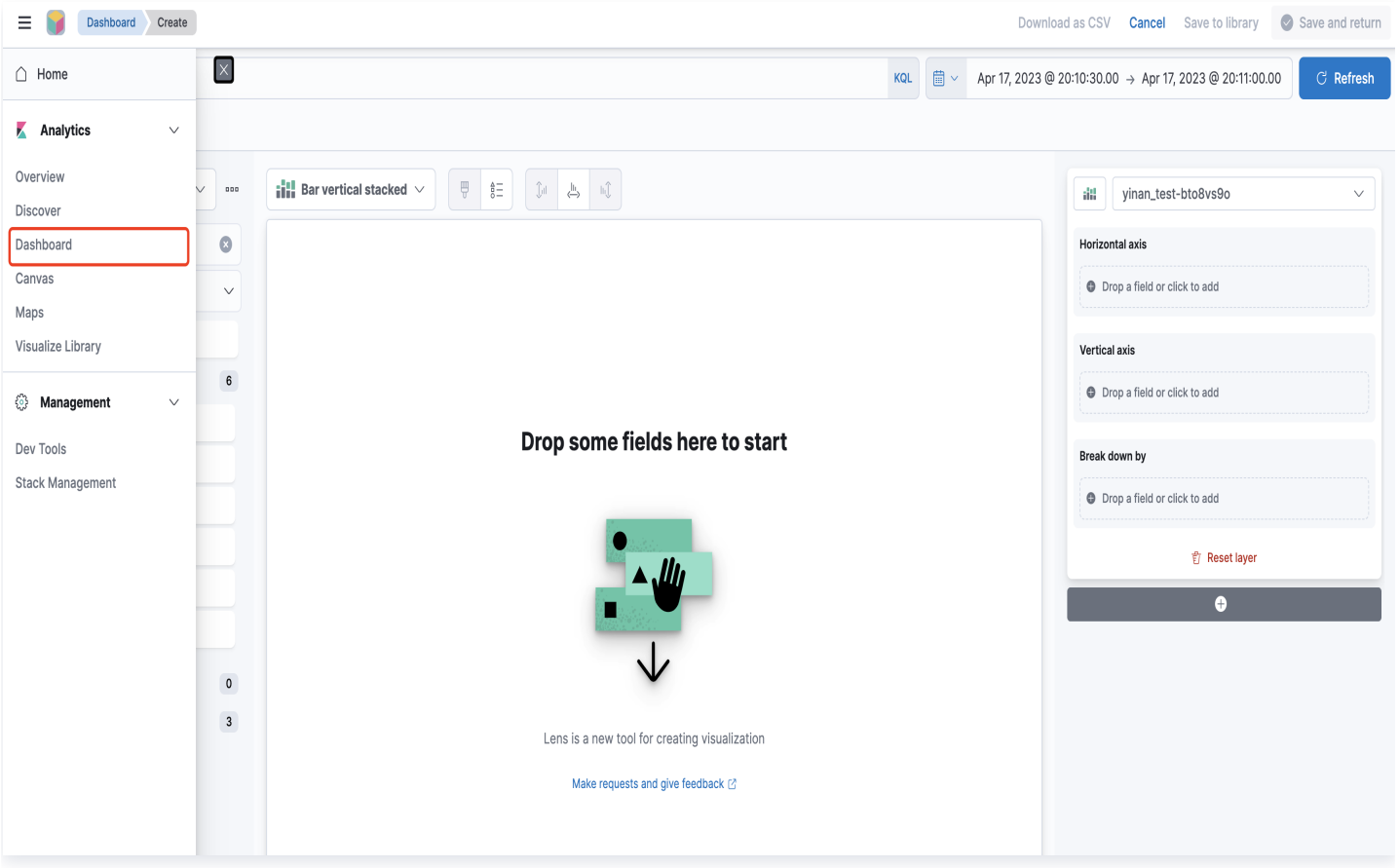
Via Dev Tools

Data queries can be done using DSL. Examples are as follows:

```
GET /Index Name/_search
{
  "query": {
    "term": {
      "user.id": "kimchy"
    }
  }
}
```

Through Kibana Dashboard

After entering Kibana, select **Dashboard** in the left navigation bar to perform data visualization. You can quickly create charts by dragging and dropping.



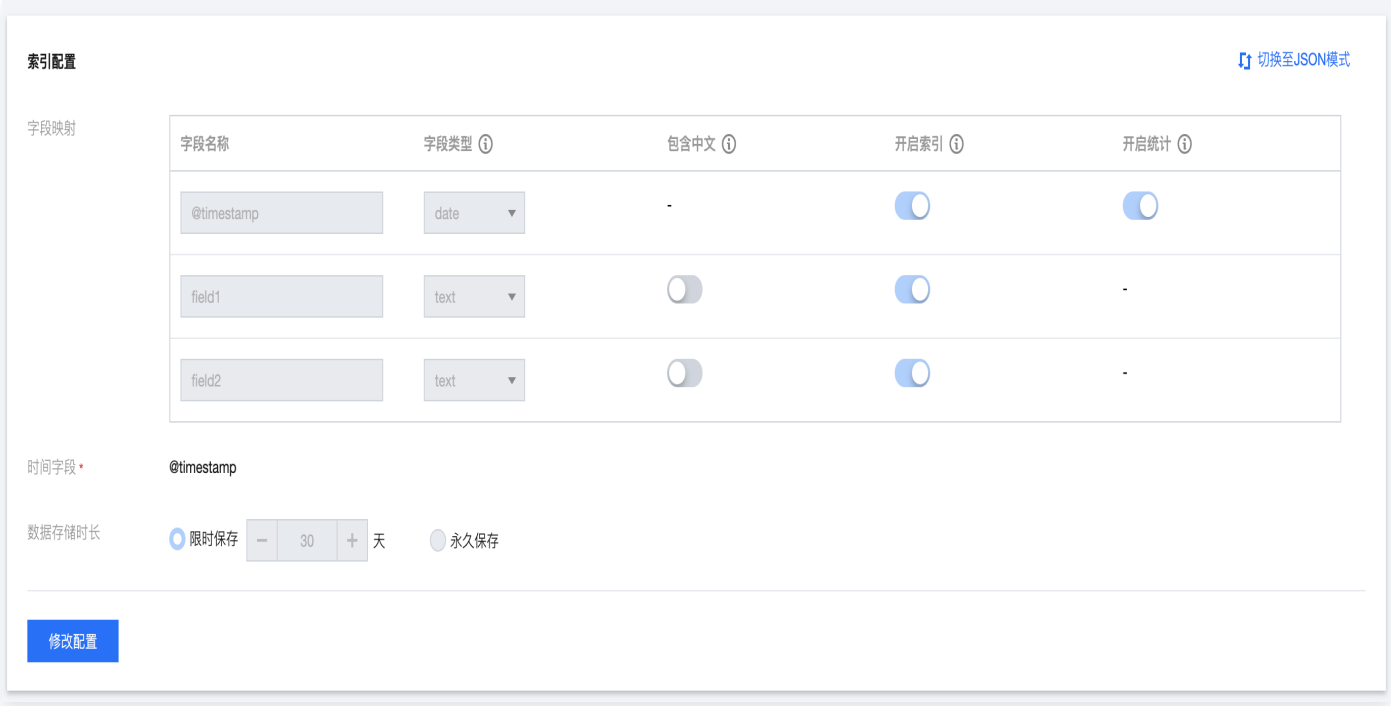
Index management
Configuration Management

Last updated：2024-11-12 21:32:11

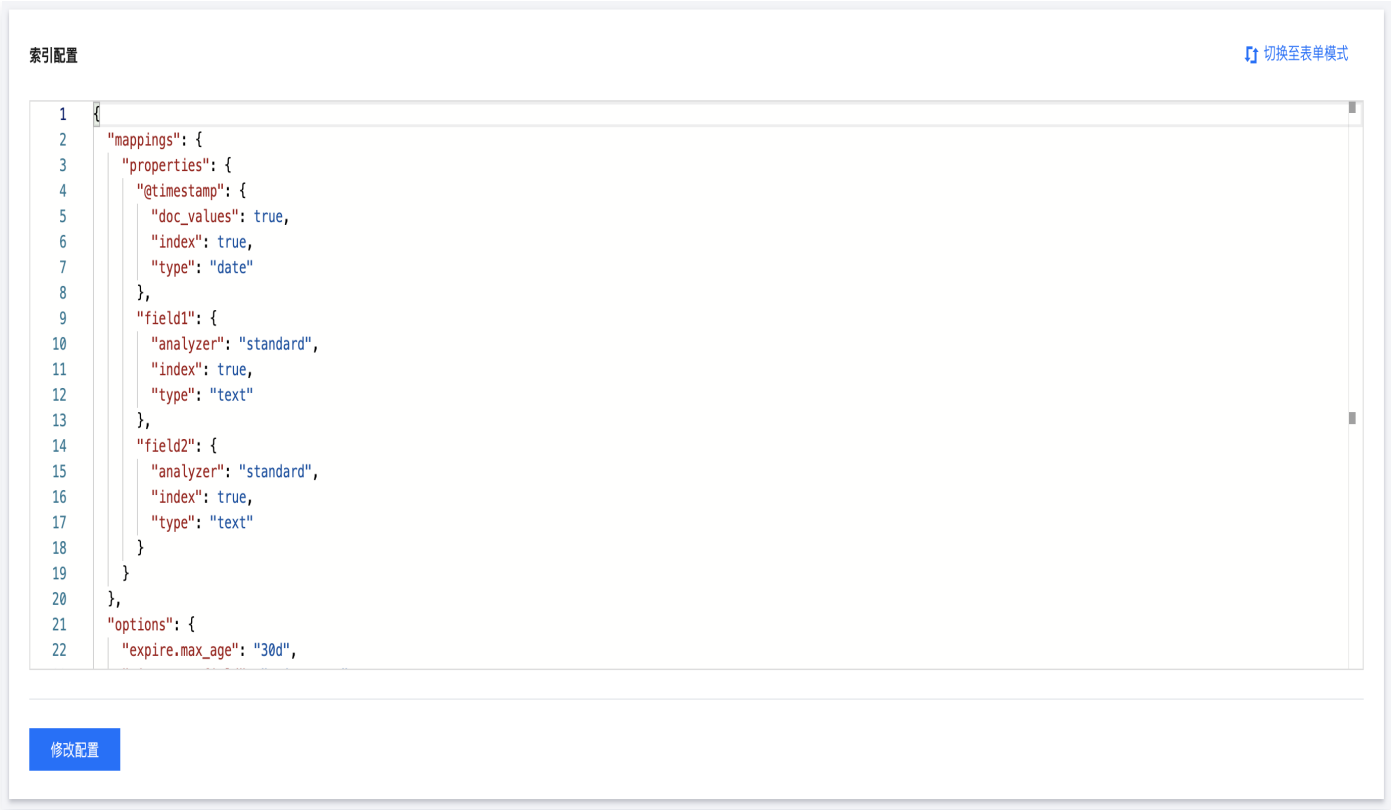
Elasticsearch Serverless service provides the feature to manage index configuration. You can quickly view the index configuration through the configuration management page. At the same time, you can modify the index configuration to quickly adapt to business development.

Checking Index Configuration

1. After entering this page, it defaults to View Mode, including Field Mapping and Data Storage Duration information.

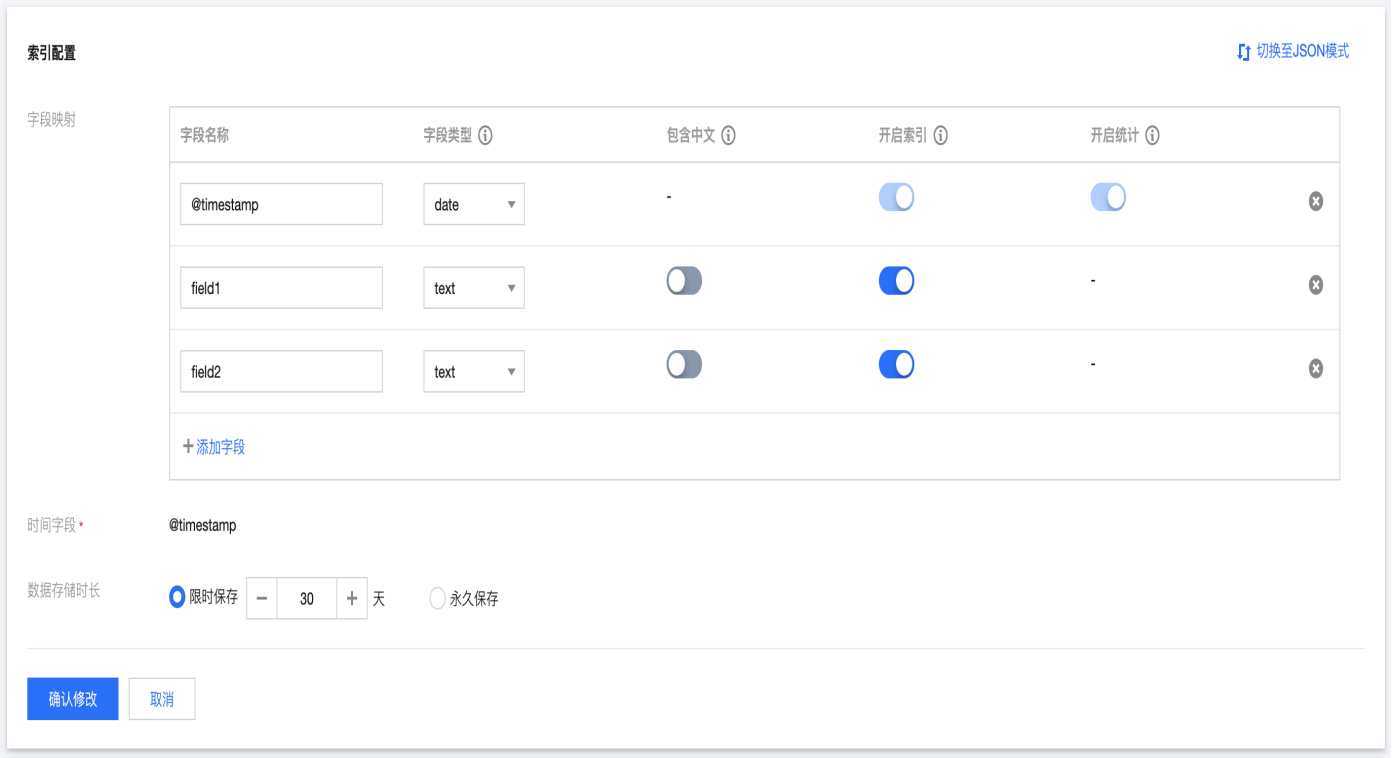


2. At the same time, you can click the top right **to switch to JSON mode** to view the current index configuration in JSON format.

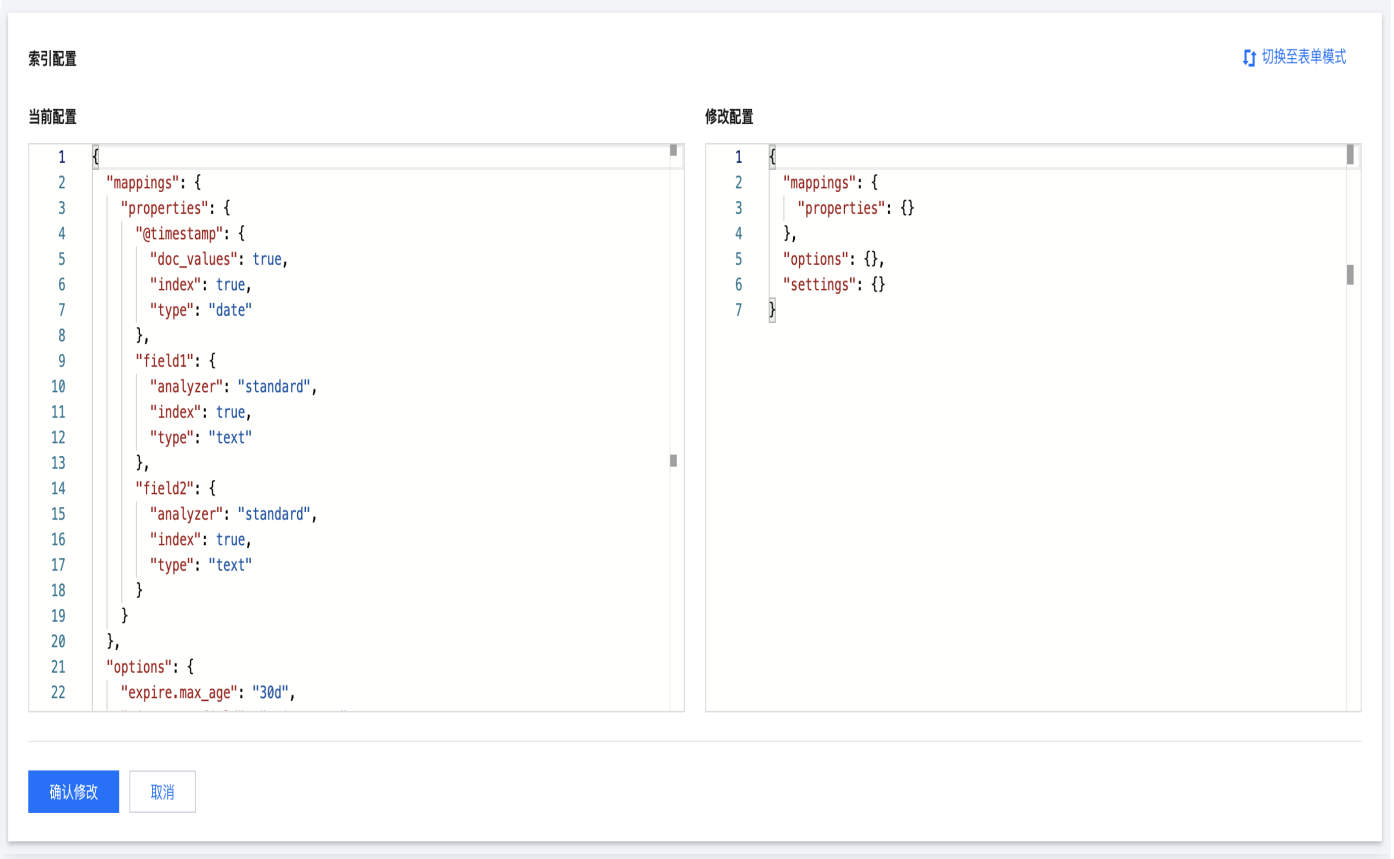


Modifying index configuration

1. Click the bottom left **Modify Configuration** to enter the editing mode, where you can modify the index configuration information, such as editing field mapping or data storage duration.



2. After switching to JSON mode, the left side displays the current configuration information, allowing you to view the config information of the index currently running online. The right side is the configuration input box. Enter the **configuration information to be modified** in the input box. After successful modification, the corresponding index configuration item will be updated.



Alarm management

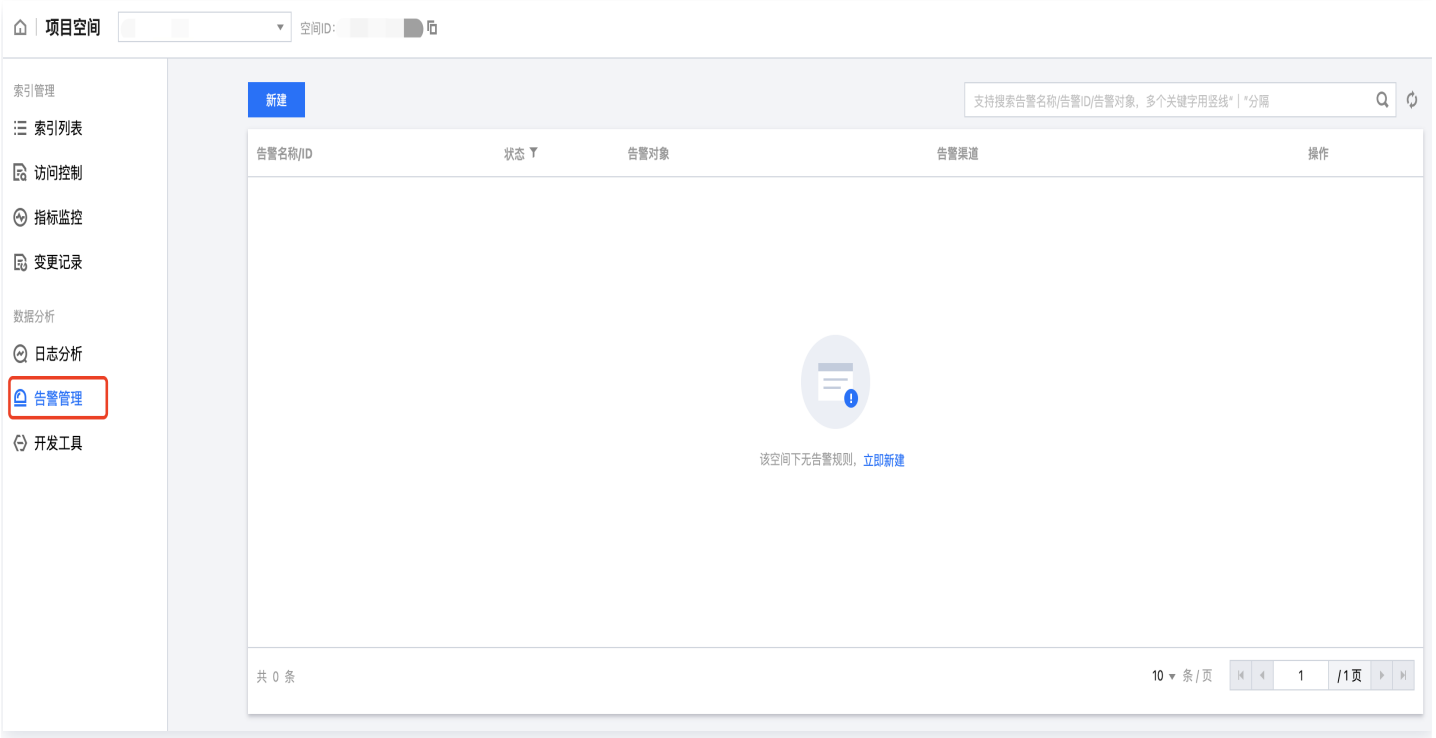
Last updated：2024-11-12 21:32:56

ES Serverless supports alarm management. You can set up alert policies for specific objects in the console. The alert policy will periodically perform search and analysis on the index within the monitored object. When the query results meet the trigger conditions, alarm notifications are sent (currently supporting Email and WeCom), allowing users to quickly identify abnormal issues. It can achieve keyword alarms (for example, the number of occurrences of "error" in logs within a certain time frame), metric monitoring (for example, whether the maximum value of a numeric type field exceeds a threshold within a certain time frame), enhancing full observability in log analysis scenarios, and rapidly discovering and addressing related issues.

Steps to Use

Preconditions

1. log in to ES Serverless Console.
2. In the space list, click the corresponding space name.



Creating Alarm

Basic information

1. In the left navigation bar, click **Alert Management**, and then click the **New** button.
2. Enter the alert name, which must be between 1 and 50 characters, including numbers, letters, Chinese characters, underscores, and the separator "–".
3. Select the alert object; you can choose indexes within the current space.

> **Note:**
> Indexes that are in the process of being created are not selectable.

Alarm rule

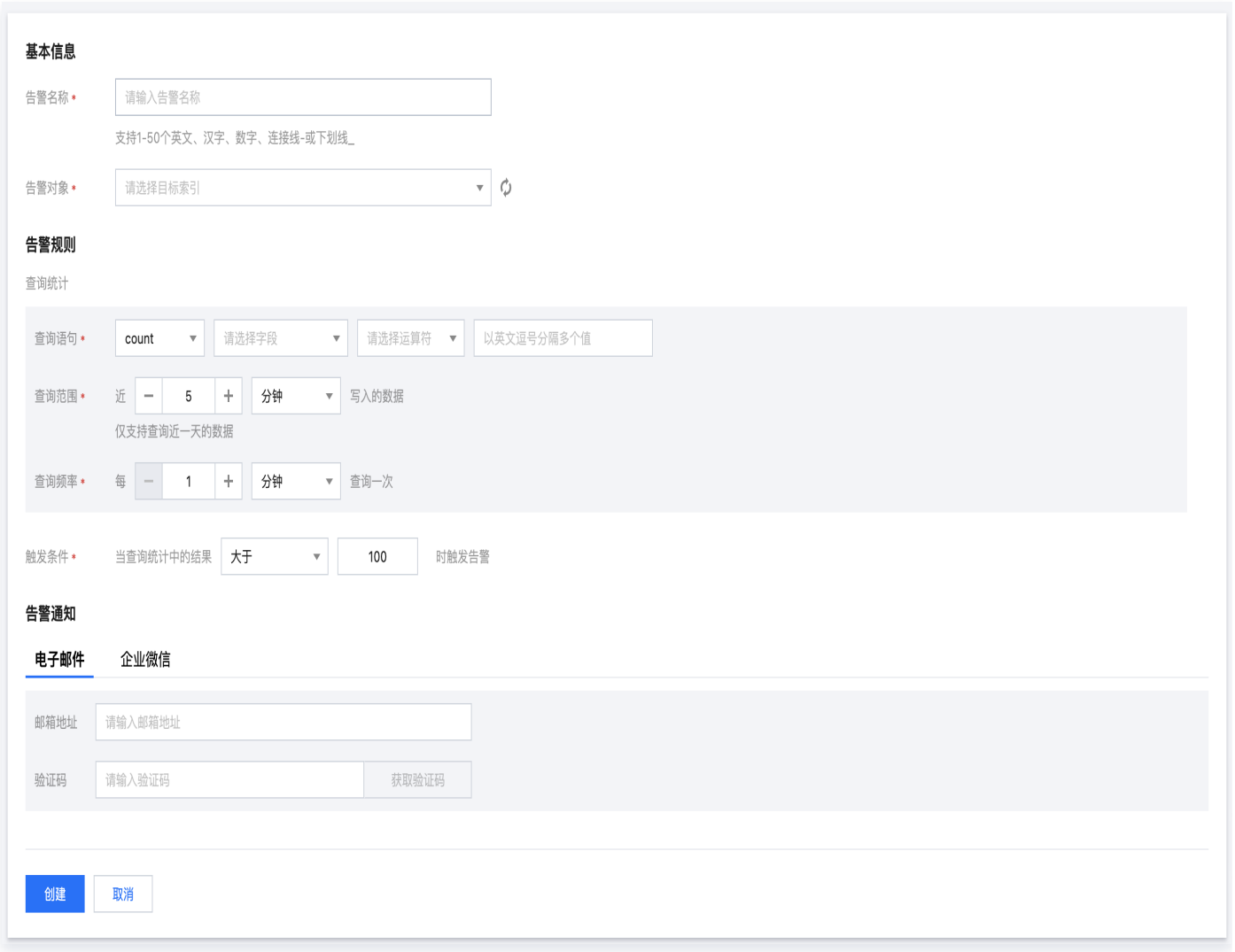
1. Query statement:
- Supported operators include **count, average, sum, max, min**, with the default being count.
- When the operator is **count**, all fields can be selected, and expressions support **equal to, not equal to, belong to, not belong to, name already exists, no existing**.
  - If the expression is **equal to, not equal to**, you need to enter the corresponding value; only a single string is supported.
  - If the expression is **belong to, not belong to**, you need to enter an array of corresponding values, with at least one entry, separated by English commas.
- When the operator is **average, sum, max, min**, only numeric data type fields (e.g., long, integer, short, double, float) are supported.
2. Query Range: The default is data written in the last 5 minutes. Units can be in minutes or hours.
3. Query frequency: The default is querying every 1 minute. Units can be in minutes or hours.
4. Trigger Conditions: Expressions support **greater than, greater than or equal to, equal to, equal to less than, less than, between**, with the default being greater than, and the default value is 100.

Alarm notification

1. Email:
   - To ensure the alarm address is correct, after entering the email address, CAPTCHA verification is required.
   - If the email address is changed, you need to obtain a new CAPTCHA.
2. WeCom: enter the WeCom robot webhook address.

> **Note:**

> The WeCom robot Webhook address must start with `https://qyapi.weixin.qq.com`.



3. After confirming it is correct, click **Create** to complete the alarm creation.

Alarm content

When an alarm is triggered, you will receive the following content:

- Title: Tencent Cloud Elasticsearch Serverless Service Alarm Triggered.
- Content:

[Alarm] Dear Tencent Cloud User, Hello! The Elasticsearch Serverless service used by your Tencent Cloud account (Account ID: xxx) triggered an alarm at Beijing time {Time}.
Alarm Name: {Corresponding Alarm Name}
Alarm Object: {Corresponding Index Name}

Alarm management

1. In the **Alarm Management** list page, you can view the relevant information and operating status of the set alarm policies.
2. If you need to close or delete an alarm, click **More** in the action column.
3. If you need to edit the alarm policy, click **Edit**.

Data Migration
Migration Plan Description

Last updated: 2024-10-25 09:15:36

For self-built ES clusters on Tencent Cloud or ES clusters purchased from other cloud providers, if you need to migrate to Tencent Cloud ES Serverless Service (suitable for most common index migrations), you can use the following methods for offline data transfer.

Migration Methods	Description	Use cases	Limits
COS Snapshot	Use the Offline Migration Tool provided in the console to migrate data from the COS Snapshot to the ES Serverless Service.	• Scenarios with large data volumes (GB, TB, PB levels) • Scenarios requiring high migration speed	• Only supports ES version clusters 6.0.0 – 7.14.2
Logstash	By configuring a Logstash pipeline, migrate data from an existing ES Cluster to the ES Serverless service in full or incremental mode.	• Scenarios for migrating full or incremental data with low real-time requirements • Scenarios requiring simple filtering of migrated data using ES Query • Scenarios requiring complex filtering or processing of migrated data • Data migration scenarios with large version spans, such as version 5.x or 8.x	• Migration speed is related to Logstash instance specifications
Dual write	By using dual write, incremental data is written to the ES Serverless service. After dual writing for a	• Scenarios with short data retention time (e.g., less than 7 days)	• Stock data migration not supported • Business write code needs to be

	period, historical data is eliminated, the original output port is removed, and the write switch is completed.		transformed to add new output ports

Fully migrate existing ES cluster data using the migration tool

Last updated: 2024-10-25 09:16:01

Overview

The ES Serverless service already provides an offline migration tool in the console. This tool uses Elasticsearch's snapshot feature, allowing us to quickly migrate existing ES cluster data in full to the ES Serverless service.

Notes

1. Data migration is supported only for clusters with ES versions 6.0.0 – 7.14.2. For other versions, migration demands can be addressed through Logstash for full or incremental data migration.
2. ES Serverless service will charge for the storage of migrated data. The storage size is calculated based on the size of the primary shard of the migrated index. The calculation method for fees can be consulted in the documentation.
3. Every index in the ES Serverless service must specify a time field of the date type. Please ensure that the migrated data contains a field **with the same name and type**. If the name matches but the type is not date, you can convert the field type through the Reindex method.

Creating Snapshots

The migration tool offers offline snapshot migration capabilities. Therefore, before setting up in the console, it is necessary to perform snapshot backup operations on the existing ES cluster data. If the cluster to be migrated originates from a **self-built ES cluster**, the snapshot operation must be executed by yourself to back up the data under your account; if it comes from a **Tencent Cloud ES cluster**, you can use the automatic snapshot capabilities provided by Tencent Cloud ES (currently free of charge) to back up the data.

Search for the required CAM policy as needed, and click to complete policy association.

Backup self-built ES cluster data

Creating Repository

```
PUT _snapshot/web_log
{
    "type": "cos",
    "settings": {
        "app_id": "xxxxxxx",
        "access_key_id": "xxxxxx",
        "access_key_secret": "xxxxxxx",
        "bucket": "xxxxxx",
        "region": "ap-guangzhou",
        "compress": true,
        "chunk_size": "500mb",
        "max_snapshot_bytes_per_sec" : "50mb",
        "base_path": "/"
    }
}
```

- app_id: APPID of your Tencent Cloud account.
- access_key_id: SecretId of your Tencent Cloud API key.
- access_key_secret: SecretKey of your Tencent Cloud API ley.
- bucket: COS bucket name, which cannot contain the -{appId} prefix.
- region: COS Bucket region, for example, ap-guangzhou. This region must be the same as the ES cluster. You can refer to the document for region codes by clicking here.
- compress: The default is true, compressing the storage of index metadata.

- base_path: The backup directory.
- max_snapshot_bytes_per_sec: The snapshot rate for this repository. It is set at the repository level, and customers can adjust it based on the current cluster load.

Creating Snapshots

```
PUT _snapshot/web_log/snapshot_test
{
   "ignore_unavailable": true,
   "include_global_state": false,
    "indices": "test3*,nginx_log,test2" // where test2 is an exact match
for a regular index, test3* is a fuzzy match for a regular index, and
nginx_log is a datastream
}
```

- ignore_unavailable: Ignore unavailable or non-existent indices.
- include_global_state: Whether to snapshot the cluster state, **it is strongly recommended** to set this to `false`.
- indices: Specify the indices to be snapshot, supporting fuzzy matching with '*', multiple indices separated by commas without spaces. To snapshot all rolled-over indices under a datastream, write the datastream name directly.

> ⚠ **Note:**
> Each index in the ES Serverless service needs to specify a Time Field, and the field type must be date. Ensure that the migrated data contains fields with consistent **names and types** as the Time Field. If names are consistent but the type is not date, you can use the Reindex method to convert the field type.

Viewing Snapshot

```
GET _snapshot/web_log/snapshot_test
```

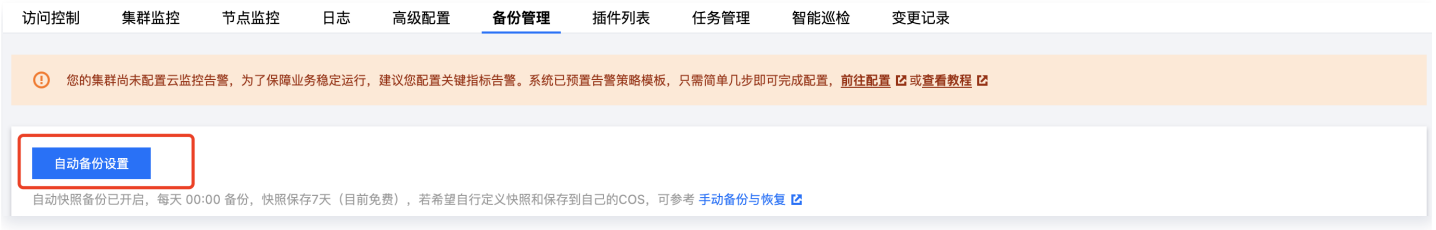
This command returns information about the snapshot. When the `state` field in the information is `SUCCESS`, it indicates that the snapshot backup is complete.

Backing up Tencent Cloud ES Cluster data

Creating Snapshots

1. Enter the ES Cluster Management Interface, select the corresponding region and cluster, click on **cluster name** to enter the cluster details page.
2. Enter the **Backup Management** page, click on **Automated Backup Settings**.

Search for the required CAM policy as needed, and click to complete policy association.



3. Enable automated snapshot backup and set the backup time.

Search for the required CAM policy as needed, and click to complete policy association.

Note:
Automated snapshot backup is currently free. Once set, the system will perform daily backups, with a retention period of 7 days.

Viewing Snapshot

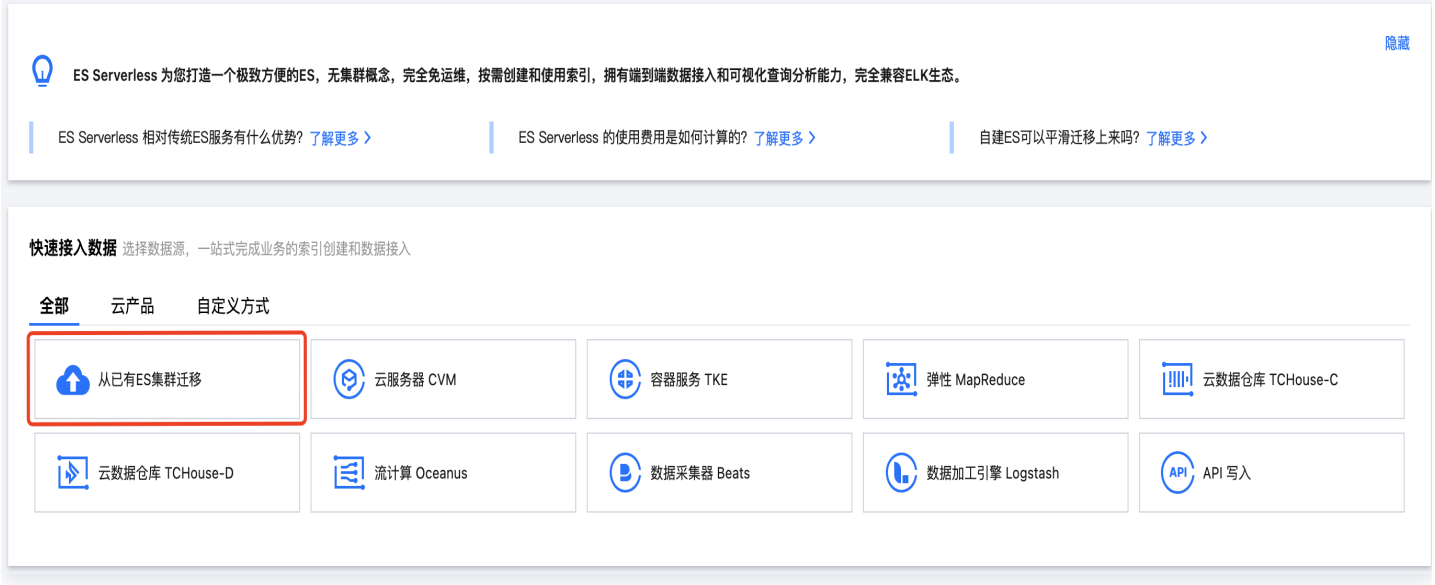
Once the snapshot backup is completed, we can view the corresponding snapshot in the snapshot list.

Search for the required CAM policy as needed, and click to complete policy association.

Operating Procedures

1. Log in to the ES Serverless Service Console.
2. In the **Rapid Data Integration** module, click **Migrate from an Existing ES Cluster**.

Search for the required CAM policy as needed, and click to complete policy association.



3. Enter the settings interface and select the source of the cluster to be migrated.
4. **Set Snapshot Source**

   4.1 The cluster to be migrated is from a self-built ES cluster

   - Region Location: Supported regions are consistent with the ES Serverless service, currently Beijing, Shanghai, Guangzhou, Nanjing, Hong Kong (China).
   - Bucket Name: Select the bucket where the snapshot is located.

- Access Path: Select the directory where the snapshot file is located.

> **Note:**
> The access path currently supports only the root directory. Do not select files in other directories.

- Snapshot Name: Choose the specific snapshot name.

4.2 The cluster to be migrated is from a Tencent Cloud ES cluster

- Region Location: Supported regions are consistent with the ES Serverless service, currently Beijing, Shanghai, Guangzhou, Nanjing, Hong Kong (China).
- Cluster Name: Select the ES cluster that needs to be migrated to the ES Serverless service.

- Snapshot Name: Select the specific snapshot name.

5. **Set the indexes to be migrated.**

After setting the snapshot source, you can choose the index to be migrated from the drop-down box of indexes to be migrated.

Search for the required CAM policy as needed, and click to complete policy association.

Note:

- Please ensure that all fields in the above indices have the same field name and type as the target index. If the name matches but the type is not date, you can convert the field type through reindex.
- Indexes containing failed backup shards cannot be migrated.
- Indexes with write mode set to "partitioned by time" cannot be migrated.

6. Set Target Index

Select the project space and index of the ES Serverless service where the data needs to be migrated.

7. **Migration Verification**

To ensure successful data migration, the platform will verify the overall information filled in above. If the verification is successful, click **OK** to start the data migration.

> **Note:**
> The ES Serverless service will charge storage fees for the migrated data. The storage size is calculated based on the size of the primary shards of the migrated indexes. For the method of fee calculation, you may refer to the documentation.

8. **View Migration Progress**

- After initiating the migration, we can view the data migration progress in the corresponding project space and index.
- After the migration is completed, if you need to review the migration details later, you can check the change records of the target index.

Migrate existing ES cluster data via Logstash in full or incrementally

Last updated: 2024-10-25 09:16:20

Overview

If you need to migrate full or incremental data from your self-hosted ES cluster to the ES Serverless service, you can configure the feature through Logstash's pipeline. This document demonstrates purchasing a Tencent Cloud Logstash instance and using Logstash to migrate the full or incremental data from the self-hosted ES cluster to the ES Serverless service.

Notes

- Logstash instances are recommended to use versions 7.14.2 or 7.10.2 and must be under the same Virtual Private Cloud (VPC) as the ES Serverless service's project space.
- Data migration can be done either in full or incrementally. During the first migration, a full migration is required, followed by incremental migration through the time field.

Directions

Deploying Logstash pipeline

1. Please first log in to Tencent Cloud.
2. Enter the Logstash Management Interface, click **Create New**, go to the Logstash purchase page, fill in the following configuration information and other details, then click **Buy now**.
   - Network: Select the same VPC as the project space.
   - Logstash version: Choose 7.10.2 or 7.14.2.
   - Advanced features: Choose the "X-Pack version".
3. In the console Logstash management interface, click on the corresponding Logstash instance name, enter the sub-page, and then click **Pipeline Management** > **Create New Pipeline**.
4. In the Config configuration, enter the relevant information for both the input terminal and the output terminal. Refer to the subsequent content for details.

Acquire project space access information

To configure data synchronization to the ES Serverless service, acquire access information of the target project space, such as access address, username, and password. The method is as follows:

1. In the project space list, click on the target space name to enter the basic management page, and then click **Access Control**.
2. Retrieve the project space access address:

Search for the required CAM policy as needed, and click to complete policy association.

3. Retrieve the username and password, ensuring that the user's permission type is "read-write" and that the permission scope includes the target index:

Search for the required CAM policy as needed, and click to complete policy association.

Pipeline Configuration: Migrating Existing Data

```
# Configuration for migrating index data from the existing cluster to ES
Serverless
input {
    elasticsearch {
        hosts =>  ["ES cluster access address"]
        user => "elastic"
        password => "Your_password"
        index => "index1,index2"
    }
}

output {
    elasticsearch {
        hosts => ["project space intranet access address"]
        index => "target index name"
        user => "elastic"
        password => "Your_password"
    }
}
```

Pipeline Configuration: Migrating Incremental Data

```
# Configuration for migrating incremental cluster index data to ES
Serverless
input {
    elasticsearch {
        hosts =>  ["ES cluster access address"]
        user => "elastic"
        password => "Your_password"
        index => "index1,index2"
        query => '{"query":{"range":{"@timestamp":{"gte":"now-
5m","lte":"now/m"}}}}'
        schedule => "* * * * *"
        scroll => "5m"
    }
}

output {
    elasticsearch {
        hosts => ["project space intranet access address"]
        index => "target index name"
        user => "elastic"
```

```
        password => "Your_password"
    }
}
```

Attachment: Meanings of each configuration item

Configuration Item	Meaning	Parameter Description
input.elasticsearch.hosts	On-premises ES cluster access address	On-premises ES cluster IP and port number, e.g., http://xxx.xx.x.xx:9200
input.elasticsearch.user	On-premises ES cluster access username	On-premises ES cluster username, e.g., elastic
input.elasticsearch.password	On-premises ES cluster access password	On-premises ES cluster password, e.g., elastic@123
input.elasticsearch.index	Self-built ES Cluster Index Name	Self-built ES Cluster Index Name, multiple indexes separated by commas (,)
input.elasticsearch.query	Self-built ES Cluster Data Query Statement	Query incremental data by time range. The following configuration queries the data from the last 5 minutes (recommended to keep the default settings): `{"query":{"range":{"@timestamp":{"gte":"now-5m","lte":"now/m"}}}}` **Note:** @timestamp is the time field of the target index.
input.elasticsearch.schedule	Self-built ES Cluster Scheduled Data Query	Scheduled tasks, the following configuration executes once per minute (recommended to keep the default settings): * * * *

input.elasticsearch.scroll	Self-built ES Cluster Data Query Scroll Cache Time	For example, the following sets the scroll_id cache to 5 minutes during queries (keeping the default settings is recommended): 5m
output.elasticsearch.hosts	Project Space Access Address	Project Space Intranet Access Address, for example: `http://space-abcdefg.ap-guangzhou.qcloudes.com:80`
output.elasticsearch.index	Target Index Name	Target Endpoint Index Name
output.elasticsearch.user	Project Space Username	Project Space Access Username, for example: elastic
output.elasticsearch.password	Project Space Access Password	Project Space Access Password, for example: elastic@123

ES API Reference

Last updated: 2024-10-25 09:16:39

in use via Command Line or clients like Filebeat

API uri	Applicable Method	Description
/_bulk	"PUT", "POST"	For details, refer to Bulk API
/{index}/_bulk	"PUT", "POST"	For details, refer to Bulk API
/{index}/_doc	"POST"	For details, refer to Index API
/{index}/_create/{id}	"PUT", "POST"	For details, refer to Index API
/{index}/_mapping	"GET"	For more details, refer to the Get mapping API
/{index}/_msearch	"POST", "GET"	For more details, refer to the Multi search API
/_msearch	"POST", "GET"	For more details, refer to the Multi search API
/{index}/_count	"POST", "GET"	For more details, refer to the Count API
/{index}/_search	"POST", "GET"	For more details, refer to the Search API

Use through Kibana

API uri	Applicable Method	Description
/{index}/_bulk	"PUT", "POST"	For details, refer to Bulk API
/{index}/_doc	"POST"	For details, refer to Index API
/{index}/_create/{id}	"PUT", "POST"	For details, refer to Index API
/_security/user/_has_privileges	"POST", "GET"	For details, refer to Has privileges API
/{index}/_field_caps	"POST", "GET"	For more details, refer to Field capabilities API

/{index}/_flush	"POST", "GET"	For more details, refer to Flush API
/{index}/_mapping	"GET"	For more details, refer to the Get mapping API
/{index}/_mappings	"GET"	For more details, refer to the Get mapping API
/_resolve/index/{name}	"GET"	For more details, refer to Resolve API
/{index}/_count	"POST", "GET"	For more details, refer to the Count API
/{index}/_msearch	"POST", "GET"	For more details, refer to the Multi search API
/{index}/_search	"POST", "GET"	For more details, refer to the Search API
/_async_search/{id}	"GET"	–
/{index}/_async_search	"POST"	–
/_security/_authenticate	"GET"	For more details, refer to Authenticate API

FAQs
Upgrade project space

Last updated：2024-11-12 21:34:09

Background

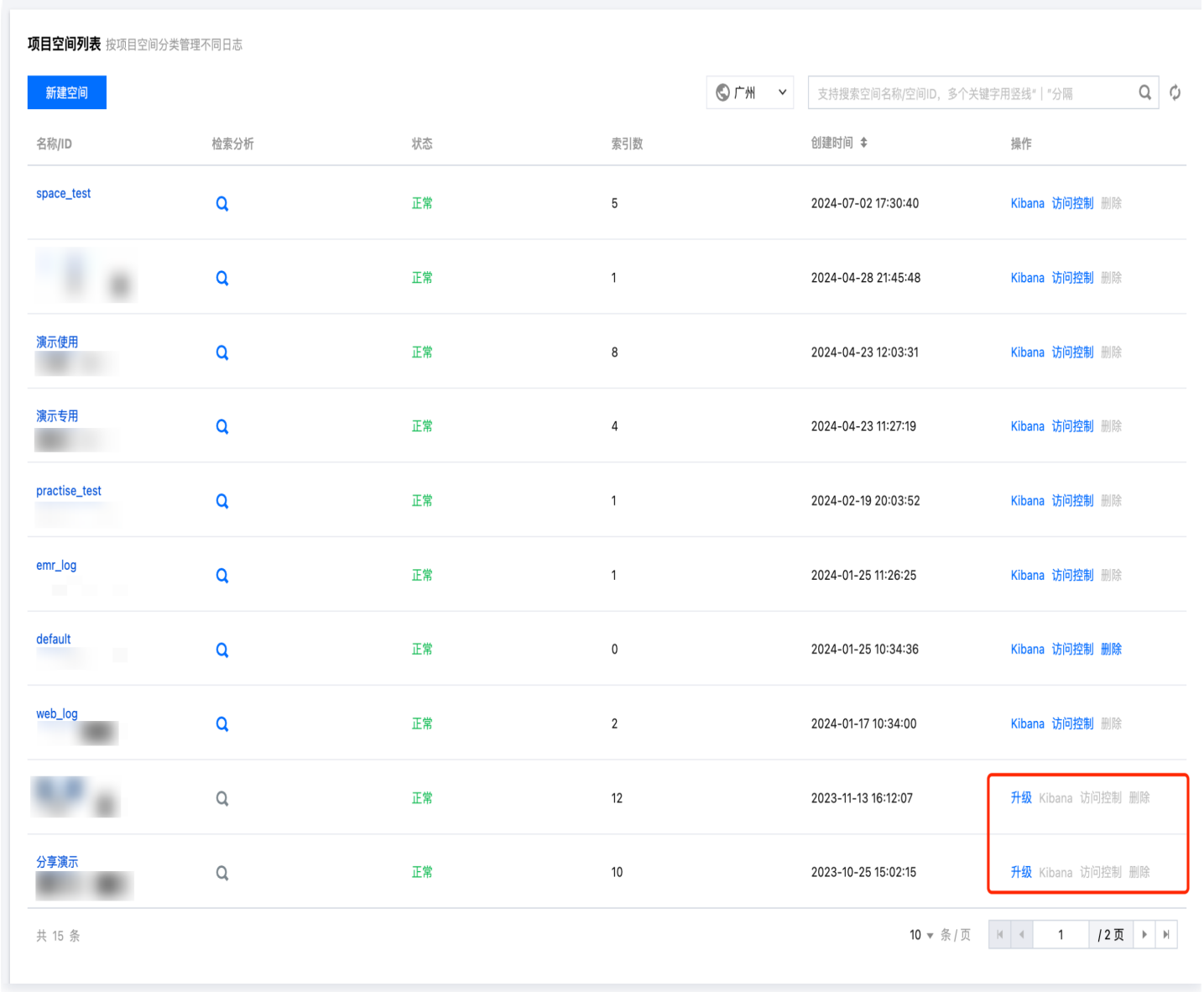
ES Serverless service offers a brand-new experience upgrade, supporting unified access address and Kibana management, accessing multiple indexes, and better fitting original usage habits. The differences before and after the upgrade are as follows:

- For project spaces created before January 23, 2024, there is no Independent Access Control feature. Simultaneous access to multiple indices in Kibana is not supported. Writing and querying are done through the index's access address.
- For project spaces created after January 23, 2024, unified management and access for all indexes within the space via the project space's access address and Kibana is supported. Additionally, you can set permission types and scopes through the visual user management feature, closely aligning with the original ES Cluster usage habits and accommodating various scenarios. The upgrade does not require changes to business code; you only need to migrate the legacy indexes to the new space. We strongly recommend migrating indexes to the new space. After successful migration, the current index access method will be decommissioned. We will provide advance notification.

Upgrade operation

For project spaces that support the **upgrade** operation, we can click the **Upgrade** button in the **Operations** column of the project space list and proceed:

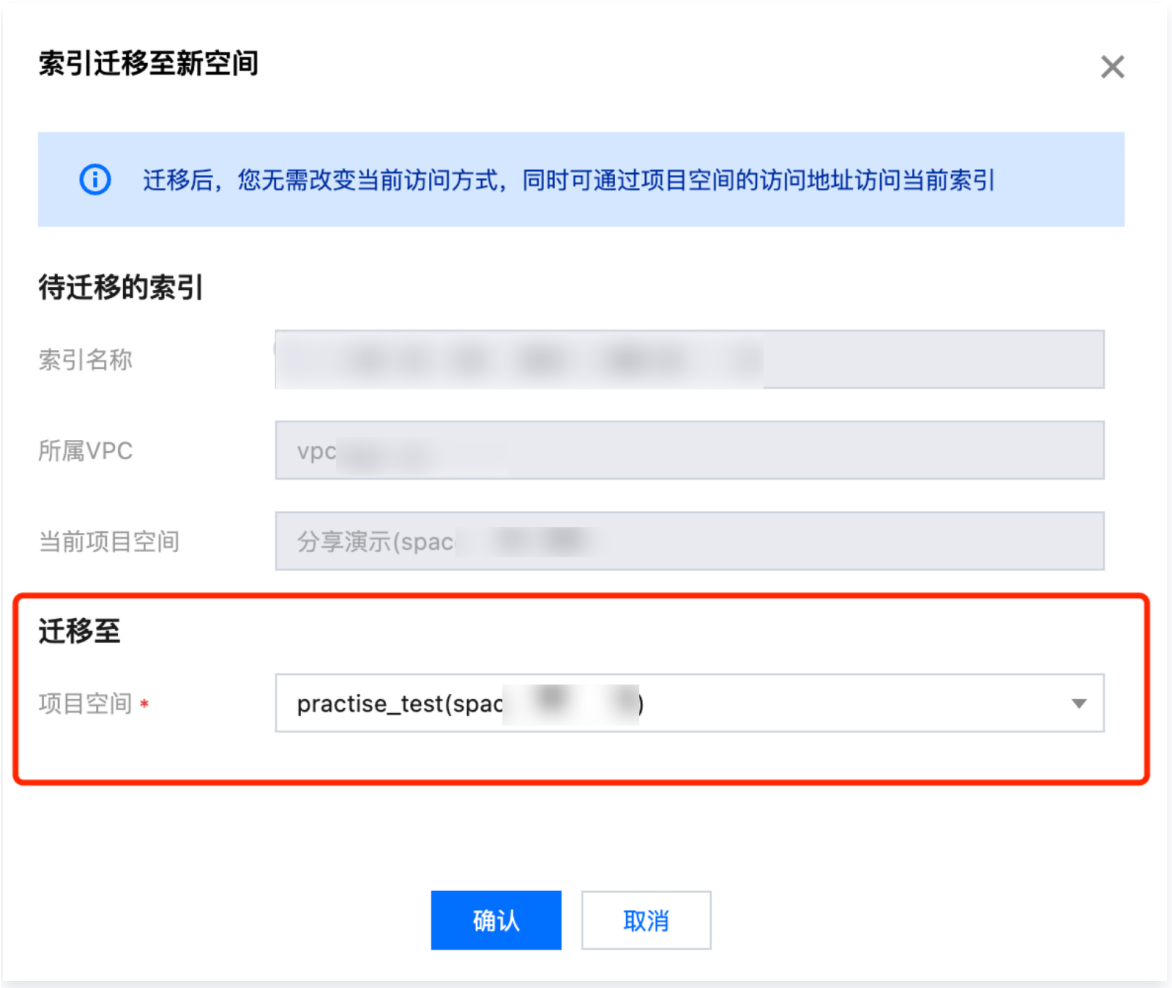
Search for the required CAM policy as needed, and click to complete policy association.



Upgrade operations support migrating indices under the project space to other spaces within the **same VPC**. We click **Operations** column for the corresponding index and click **Upgrade**: Search for the required CAM policy as needed, and click to complete policy association.

Select the target project space, click **Confirm** to complete the index migration operation:

Search for the required CAM policy as needed, and click to complete policy association.



The indices migrated to the new space can continue to use the original access method. From the perspectives of management and convenience, we strongly recommend switching the access method to the project space address.

Index usage issues

Last updated: 2024-10-25 09:17:19

I need to use the ES Serverless service. How do I perform data migration?

Currently, data can be migrated to the ES Serverless service index via COS Snapshot or Logstash. COS Snapshot is suitable for scenarios with large data volumes or high migration speed requirements, while Logstash is suitable for migrating full or incremental data with low real-time requirements. For data migration using Logstash, refer to the documentation. If you need to migrate using a snapshot, you can submit a ticket for consultation.

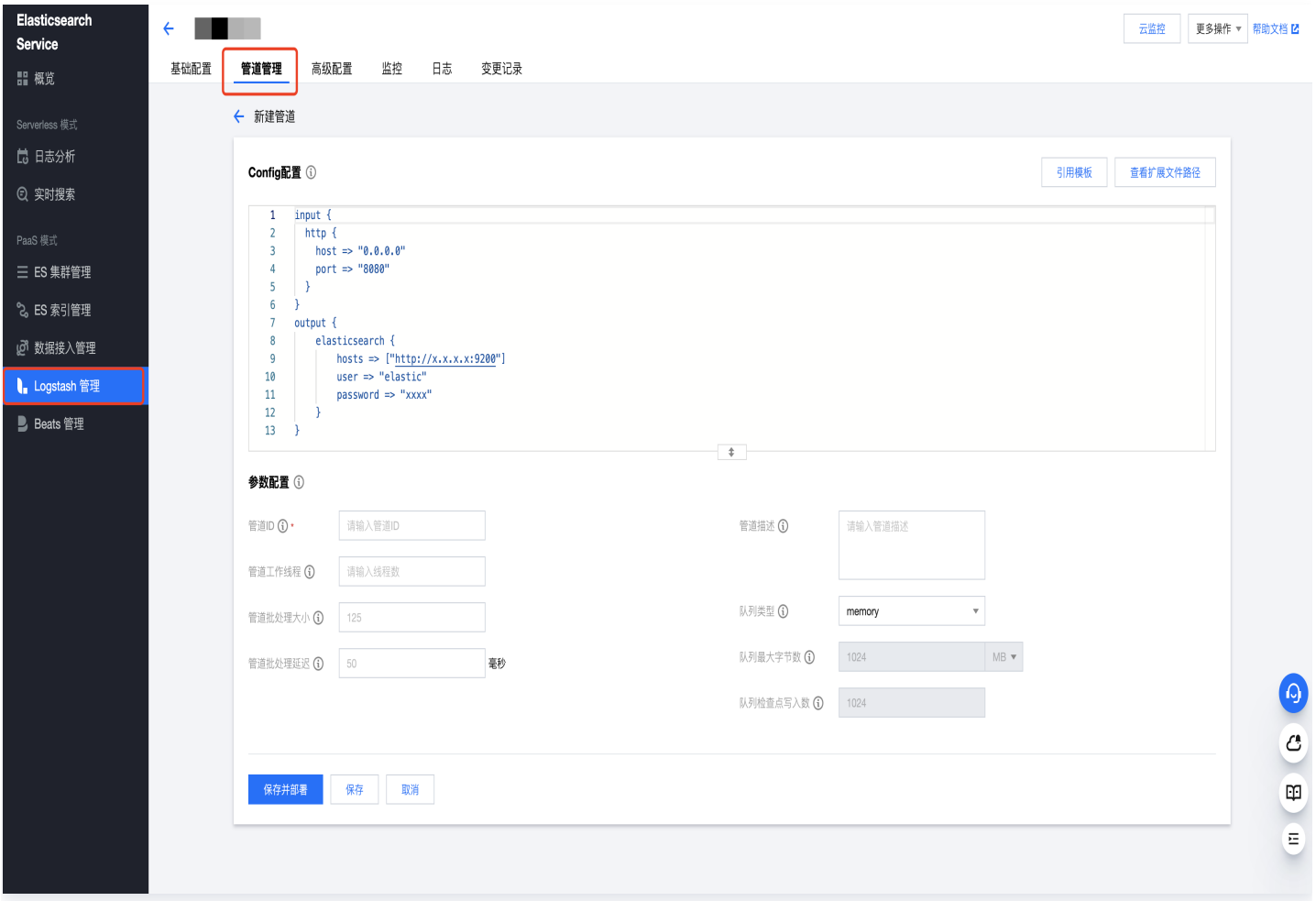
Is it supported to collect data to the ES Serverless service index using Filebeat?

Supported. You can use the Beats Management feature in the cloud console to collect data to the index using Filebeat from CVM, TKE (including EKS). You can also use a self-built Filebeat with the internal network access address of the ES Serverless service to write data into the index.
Search for the required CAM policy as needed, and click to complete policy association.

Does it support writing data to the ES Serverless service index through Logstash?

Supported. You can set the output end to the index of ES Serverless service in the pipeline configuration of Cloud-based Logstash or your own Logstash cluster. For details, please refer to Logstash Data Delivery.

Search for the required CAM policy as needed, and click to complete policy association.



What is the recommended size for batch writing?

The ES Serverless service has a built-in proprietary directed routing capability. It is recommended that the number of documents for a single batch write be set to 2000 – 5000, which helps improve write performance and reduce interface call costs.

What should I do if the required field type is not available in the field settings form?

The form page dropdown supports selecting common field types. If more types are needed, you can **switch to JSON mode** and set the field to the desired type.

Search for the required CAM policy as needed, and click to complete policy association.

Issues with using Kibana

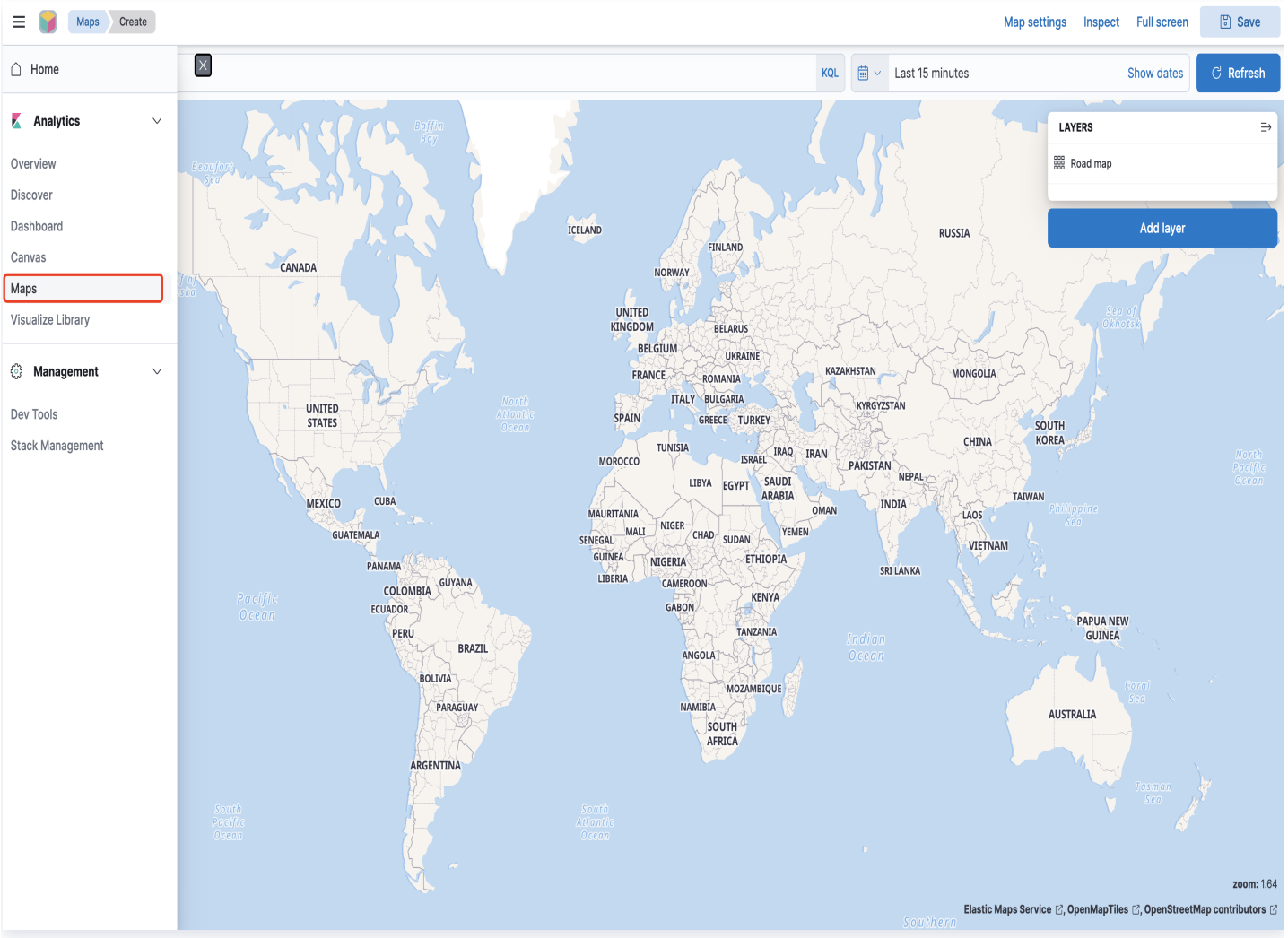
Last updated: 2024-10-25 09:17:39

How to set a field to geo_point type and draw a map?

Before writing data, set the type of the specified field to geo_point in the mapping. After the data is written, you can select **Maps** in the Kibana navigation bar to enter the map drawing interface.

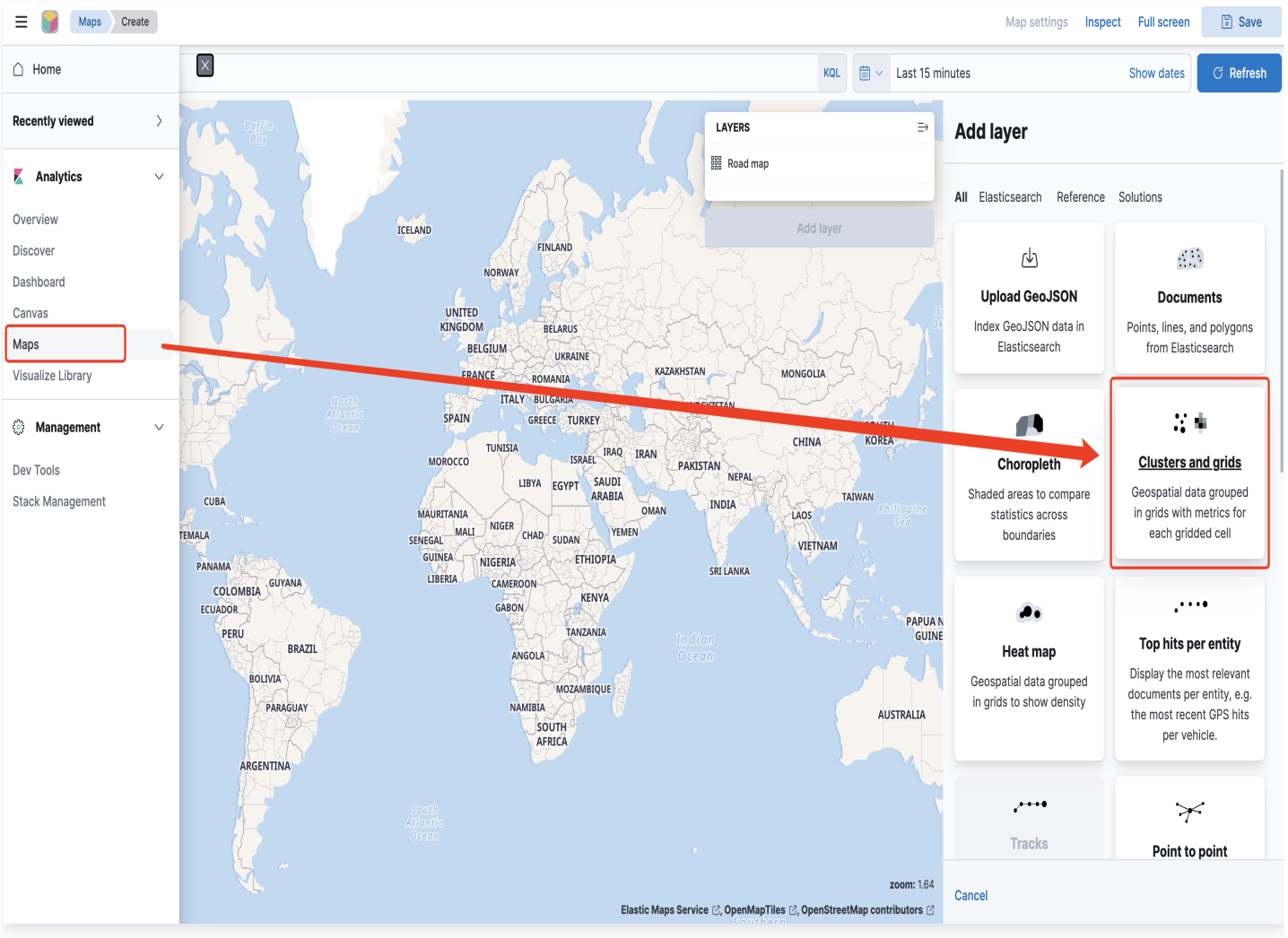
> **Note:**
> For the geo_point type, please set it manually; otherwise, the field might be automatically mapped to an incorrect type, resulting in an inability to render the map.

Search for the required CAM policy as needed, and click to complete policy association.



Where can the Coordinate Map feature of Kibana be found?

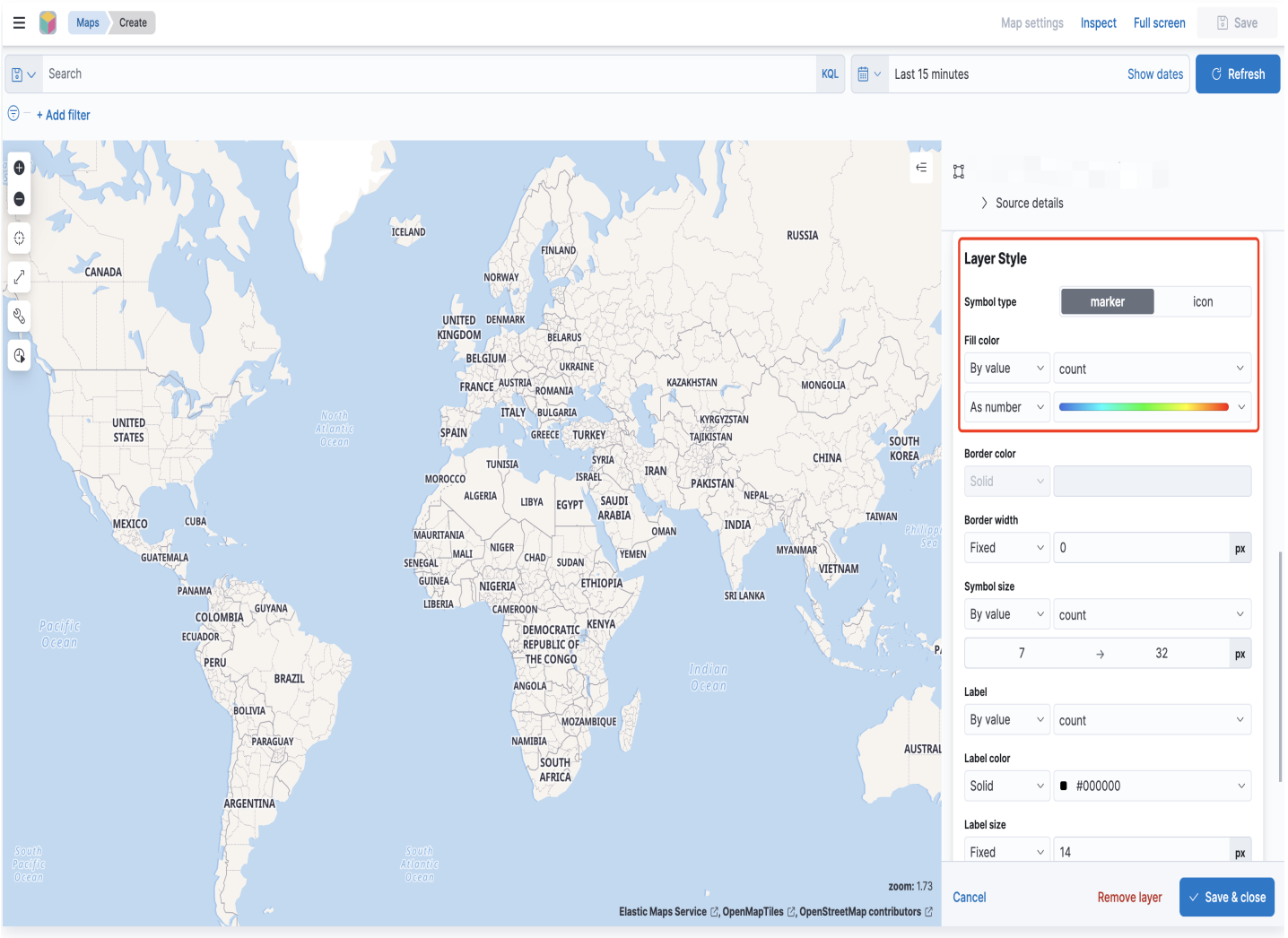
You can use the **Maps** option's **Clusters and grids** type to perform aggregation on specific fields.
Search for the required CAM policy as needed, and click to complete policy association.



How to distinguish and display values in different intervals based on the values of fields aggregated by metrics?

You can adjust the **Fill color**: In the **Layer settings**, scroll down to find the **Layer Style** module, select **Fill color** by value, choose the key you want to differentiate for display, and in the as number field, select the appropriate Gradient Color.

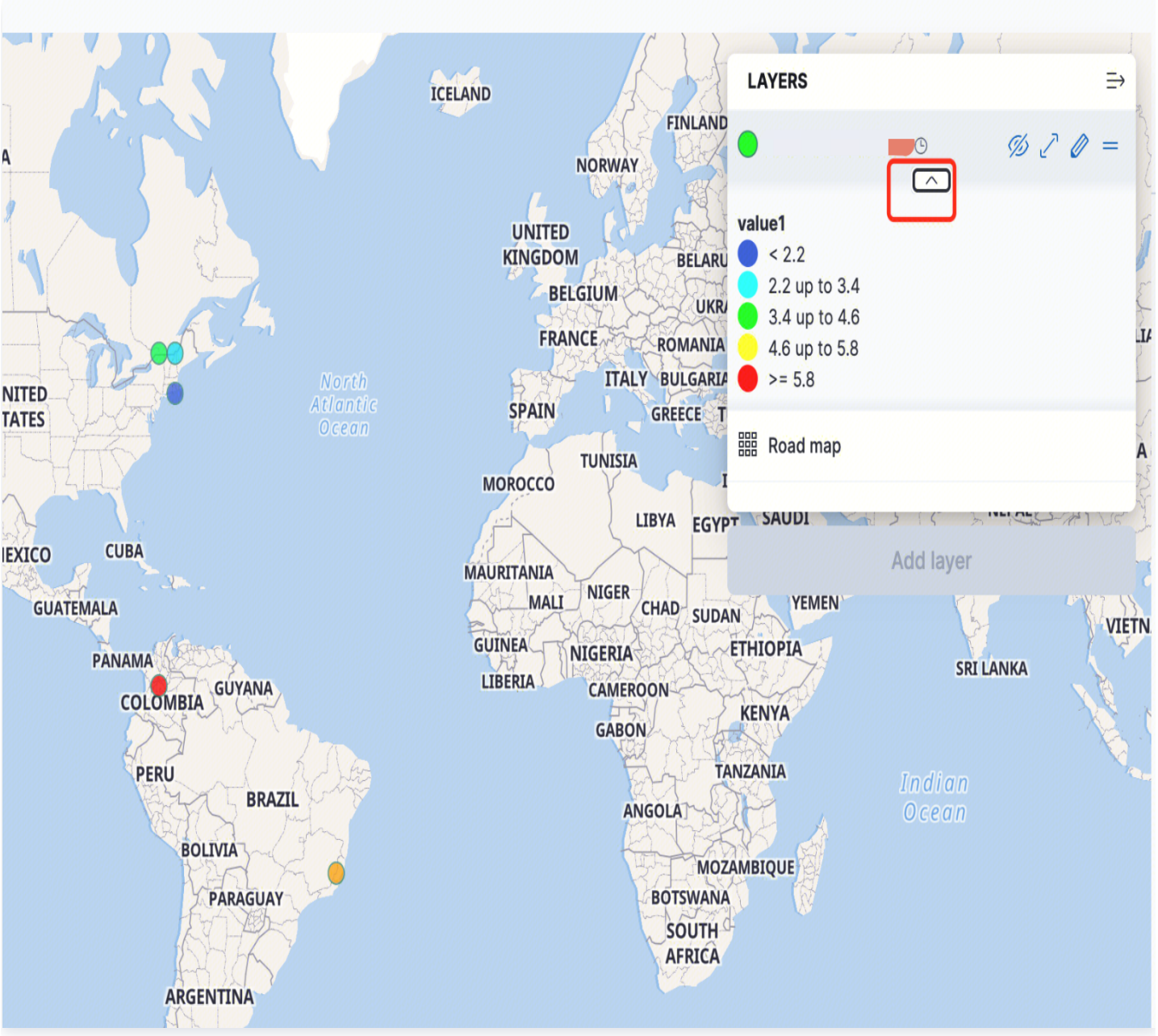
Search for the required CAM policy as needed, and click to complete policy association.



After displaying metrics, how can we know the corresponding value range for the various colored dots on the map?

Just click the arrow in the middle of the corresponding layer under LAYERS to display it (the arrow is not displayed by default and will only be shown when the mouse hovers over it).

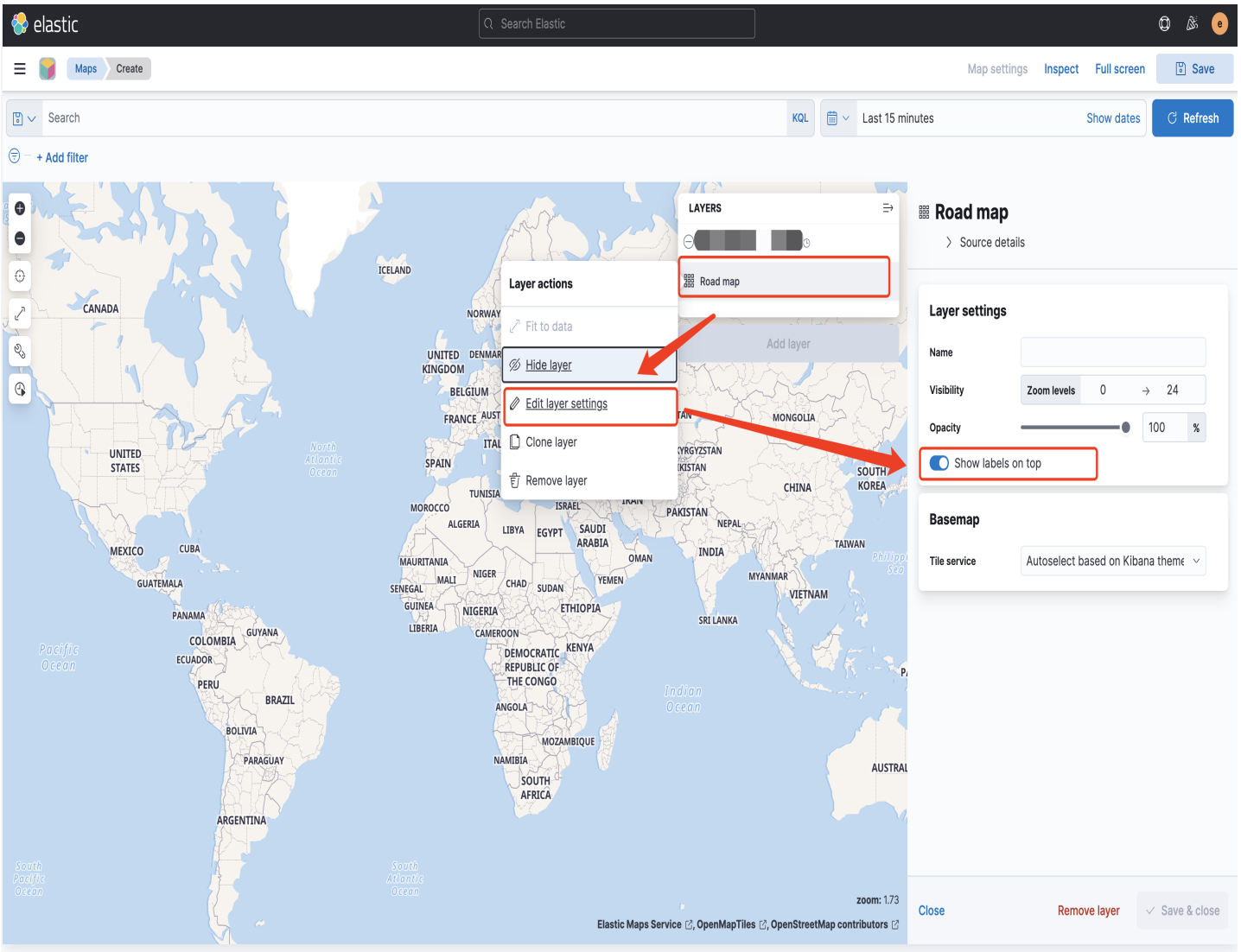
Search for the required CAM policy as needed, and click to complete policy association.



After displaying metrics, each dot is very large, occupying the base map's place names, causing them not to be displayed. How can this be set?

You can click **Layer** under the **Road map**, select **Edit layer settings**, and set the base map display priority to top in the **Layer settings**.

Search for the required CAM policy as needed, and click to complete policy association.



Third party Cookie settings

Last updated：2024-11-12 21:34:54

Using console retrieval and analysis capabilities requires the browser to support third party Cookies. Common browser configurations are as follows:

Chrome

1. Open the Chrome browser.
2. Click the "More Options" icon in the upper right corner .
3. click **Settings** > select**Privacy and Security** > find**Third-party cookies** > choose**Allow third-party cookies**.

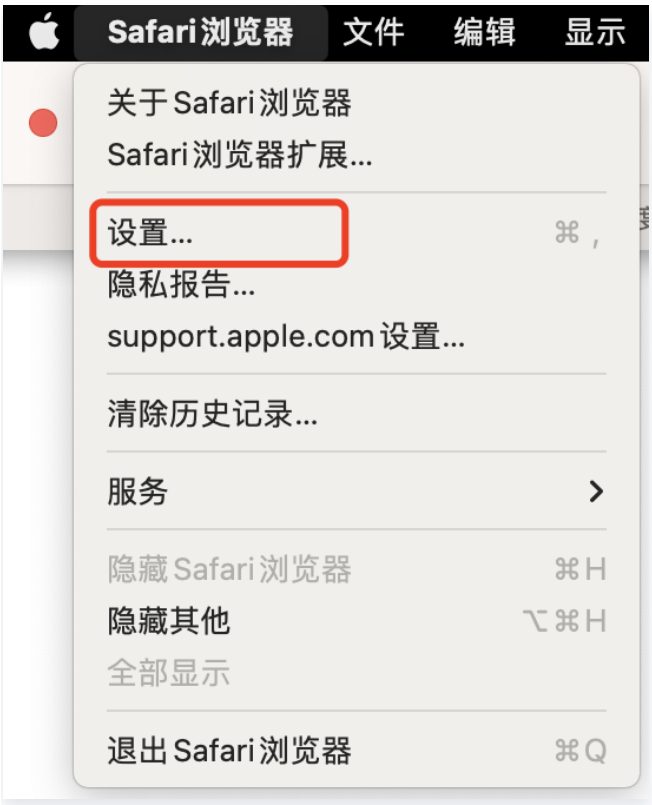
Search for the required CAM policy as needed, and click to complete policy association.

Search for the required CAM policy as needed, and click to complete policy association.

Safari

1. Open Safari on your Mac.
2. Select **Safari** > **Settings** to enter the settings page.

Search for the required CAM policy as needed, and click to complete policy association.



3. In the Privacy Settings module, uncheck **Prevent Cross-site Tracking** and **Block All Cookies**.

Converting field type via Reindex

Last updated: 2024-11-12 21:41:45

Overview

When creating an index in ES Serverless service, you need to specify a time field, and its type must be `date`. When synchronizing data from an existing ES cluster to an index in ES Serverless service, if the data field name matches but the type is different from the time field, the write will fail. In such cases, we can convert the field type using reindex.

Process Description

1. Create the target index for reindex and set the type of the field with the same name as the time field in the ES Serverless service index to `date`.
2. Use the reindex interface to synchronize existing data to the target index.

Case demonstration

1. Suppose we need to synchronize data from the index `source_index` to the index in the ES Serverless service (the time field of this index is `@timestamp`). At this point, check the field configuration of source_index and find that the field type of `source_index` in `@timestamp` is `keyword` type. At this point, synchronizing data will result in a write error.

Search for the required CAM policy as needed, and click to complete policy association.

```
GET source_index/_mapping
```

```
{
  "source_index" : {
    "mappings" : {
      "dynamic_templates" : [
        {
          "message_full" : {
            "match" : "message_full",
            "mapping" : {
              "fields" : {
                "keyword" : {
                  "ignore_above" : 2048,
                  "type" : "keyword"
                }
              },
              "type" : "text"
            }
          }
        },
        {
          "message" : {
            "match" : "message",
            "mapping" : {
              "type" : "text"
            }
          }
        },
        {
          "strings" : {
            "match_mapping_type" : "string",
            "mapping" : {
              "type" : "keyword"
            }
          }
        }
      ],
      "properties" : {
        "@timestamp" : {
          "type" : "keyword"
        },
        "field1" : {
          "type" : "text"
        }
      }
```

2. Check the number of documents in source_index.

Search for the required CAM policy as needed, and click to complete policy association.

```
GET source_index/_search
```


```
{
  "took" : 0,
  "timed_out" : false,
  "_shards" : {
    "total" : 1,
    "successful" : 1,
    "skipped" : 0,
    "failed" : 0
  },
  "hits" : {
    "total" : {
      "value" : 2,
      "relation" : "eq"
    },
    "max_score" : 1.0,
    "hits" : [
      {
        "_index" : "source_index",
        "_type" : "_doc",
        "_id" : "3_4vF40BCM6qeXZO07Kv",
        "_score" : 1.0,
        "_source" : {
          "@timestamp" : "2022-03-13T03:07:34.348+08:00",
          "field1" : "a"
        }
      },
      {
        "_index" : "source_index",
        "_type" : "_doc",
        "_id" : "4P4vF40BCM6qeXZO07Kv",
        "_score" : 1.0,
        "_source" : {
          "@timestamp" : "2022-03-24T10:51:34.348+08:00",
          "field1" : "b"
        }
      }
    ]
  }
}
```


3. Create the target index for reindex `dest_index` and specify the type of the field `@timestamp` in the mapping as `date`.

Search for the required CAM policy as needed, and click to complete policy association.

```
GET dest_index/_mapping
```

```
{
  "dest_index" : {
    "mappings" : {
      "dynamic_templates" : [
        {
          "message_full" : {
            "match" : "message_full",
            "mapping" : {
              "fields" : {
                "keyword" : {
                  "ignore_above" : 2048,
                  "type" : "keyword"
                }
              },
              "type" : "text"
            }
          }
        },
        {
          "message" : {
            "match" : "message",
            "mapping" : {
              "type" : "text"
            }
          }
        },
        {
          "strings" : {
            "match_mapping_type" : "string",
            "mapping" : {
              "type" : "keyword"
            }
          }
        }
      ],
      "properties" : {
        "@timestamp" : {
          "type" : "date"
        }
      }
    }
  }
```

4. Use the reindex API to synchronize data from `source_index` to `dest_index`. Compare the number of documents in the original `source_index`. The count matches exactly.

```
POST _reindex
{
  "source": {
    "index": "source_index"
  },
  "dest": {
    "index": "dest_index"
  }
}
```

Search for the required CAM policy as needed, and click to complete policy association.

5. At this time, searching for `dest_index` data will retrieve the data synchronized from `source_index`.

Search for the required CAM policy as needed, and click to complete policy association.

```
GET dest_index/_search
```


```
{
  "took" : 0,
  "timed_out" : false,
  "_shards" : {
    "total" : 1,
    "successful" : 1,
    "skipped" : 0,
    "failed" : 0
  },
  "hits" : {
    "total" : {
      "value" : 2,
      "relation" : "eq"
    },
    "max_score" : 1.0,
    "hits" : [
      {
        "_index" : "dest_index",
        "_type" : "_doc",
        "_id" : "3_4vF40BCM6qeXZO07Kv",
        "_score" : 1.0,
        "_source" : {
          "@timestamp" : "2022-03-13T03:07:34.348+08:00",
          "field1" : "a"
        }
      },
      {
        "_index" : "dest_index",
        "_type" : "_doc",
        "_id" : "4P4vF40BCM6qeXZO07Kv",
        "_score" : 1.0,
        "_source" : {
          "@timestamp" : "2022-03-24T10:51:34.348+08:00",
          "field1" : "b"
        }
      }
    ]
  }
}
```