ALIBABA CLOUD

Alibaba Cloud

Elasticsearch

Product Introduction

Document Version: 20220610

Alibaba Cloud

Legal disclaimer

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

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

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

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

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

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

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

Document conventions

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

Table of Contents

1.What is Alibaba Cloud Elasticsearch?

Elasticsearch is an open source, distributed, real-time search and analytics engine built on Apache Lucene. It is released under the Apache License and is a popular search engine for enterprises. It provides services based on RESTful APIs and allows you to store, query, and analyze large amounts of datasets in near real time. Elasticsearch is typically used to support complex queries and high-performance applications.

Alibaba Cloud Elasticsearch is a fully managed cloud service that is developed based on open source Elasticsearch. This service is fully compatible with the features provided by open source Elasticsearch. It is out-of-the-box and supports the pay-as-you-go billing method. In addition to Elastic Stack components such as Elasticsearch, Logstash, Kibana, and Beats, Alibaba Cloud provides the X-Pack plug-in free of charge together with Elastic. X-Pack is integrated into Kibana to provide features, such as security, alerting, monitoring, and machine learning. It also provides SQL capabilities. Alibaba Cloud Elasticsearch is widely used in scenarios such as real-time log analysis and processing, information retrieval, multidimensional data queries, and statistical data analytics.

Overview

Alibaba Cloud Elasticsearch is committed to providing a low-cost, scenario-based Elasticsearch service on the cloud based on the open source Elastic Stack ecosystem. Alibaba Cloud Elasticsearch originates from but is not limited to this ecosystem. Alibaba Cloud has superior computing and storage capabilities on the cloud and technical expertise in the fields of cluster security and O&M. This enables Alibaba Cloud Elasticsearch to support one-click deployment, auto scaling, intelligent O&M, and various kernel optimization features. Alibaba Cloud Elasticsearch also provides a complete set of solutions such as migration, disaster recovery, backup, and monitoring.

Alibaba Cloud Elasticsearch features high security, performance, and availability and provides powerful search and analytics capabilities. It simplifies cluster deployment and management, reduces resource and O&M costs, ensures data security and reliability, enables upstream and downstream data links, and optimizes read and write performance. Based on these features and optimizations, Alibaba Cloud Elasticsearch allows you to build business applications with ease, such as applications that perform log analysis, exception monitoring, enterprise search, and big data analytics. Alibaba Cloud Elasticsearch enables you to focus on the business applications themselves and add value to your business.

Components

The Alibaba Cloud Elastic Stack ecosystem contains the following components: Elasticsearch, Kibana, Beats, and Logstash. Elasticsearch is a real-time, distributed search and analytics engine. Kibana provides a visual interface for data analytics. Beats collects data from various machines and systems. Logstash collects, converts, processes, and generates data. Integrated with Kibana, Beats, and Logstash, Alibaba Cloud Elasticsearch can be used for real-time log processing, full-text searches, and data analytics.

- X-Pack

  X-Pack is a commercial extension of Elasticsearch. It provides security, alerting, monitoring, graphing, reporting, and machine learning capabilities. When you create an Alibaba Cloud Elasticsearch cluster, the system integrates X-Pack into Kibana to provide free services. The services include authorization and authentication, role-based access control, real-time monitoring, visual reporting, and machine learning. X-Pack facilitates cluster O&M and application development.

- Beats

  Beats is a lightweight data collection tool that integrates a variety of single-purpose data shippers. These data shippers collect data from various machines or systems and send the collected data to Logstash or Elasticsearch.

  Beats allows you to create the following types of data shippers: Filebeat, Metricbeat, Auditbeat, and Heartbeat. You can create and configure a shipper to collect various types of data from Elastic Compute Service (ECS) instances or Container Service for Kubernetes (ACK) clusters. The data include logs, network data, and container metrics. Beats also allows you to manage your shippers in a centralized manner.

- Logstash

  Logstash is a server-side data processing pipeline. It uses input, filter, and output plug-ins to dynamically collect data from a variety of sources, process and convert the data, and then save it to a specific location.

  Alibaba Cloud Logstash is a fully managed service and is fully compatible with open source Logstash. Logstash allows you to quickly deploy pipelines, configure them by using a visual interface, and centrally manage them. It provides multiple types of plug-ins to connect to cloud services, such as Object Storage Service (OSS) and MaxCompute.

- Kibana

  Kibana is a flexible data analytics and visualization tool. Multiple users can log on to the Kibana console at the same time. You can use Kibana to search for, view, and manage data in Elasticsearch indexes. When you create an Alibaba Cloud Elasticsearch cluster, the system automatically deploys an independent Kibana node. This node allows you to present diversified data analytics reports and dashboards by using graphs, tables, or maps based on your business requirements.

Related items

- AliES and its provided plug-ins

  In addition to all the features provided by the open source Elasticsearch kernel, Alibaba Cloud Elasticsearch develops the AliES kernel. This kernel enables Alibaba Cloud Elasticsearch to provide optimizations in multiple aspects, such as thread pools, monitoring metric types, circuit breaking policies, and query and write performance. The kernel also provides a variety of self-developed plug-ins to improve cluster stability, enhance performance, reduce costs, and optimize monitoring and O&M.

- EYou

  EYou is an intelligent O&M system provided by Alibaba Cloud Elasticsearch. This system can detect the health of more than 20 items, such as clusters, nodes, and indexes. EYou simplifies cluster O&M. It comprehensively observes and records the running statuses of clusters and automatically summarizes cluster diagnostic results. It also detects the possible risks of your clusters. If your clusters are abnormal, the system quickly provides key information and reasonable optimization suggestions.

References

- Benefits
  - Comparison between Alibaba Cloud Elasticsearch clusters and self-managed Elasticsearch clusters
  - High availability
  - High security
  - High performance

- Cluster purchase
  - Purchase an Alibaba Cloud Elasticsearch cluster
  - Purchase an Alibaba Cloud Logstash cluster
- Quick start
  - Quick start of Elasticsearch
  - Quick start of Logstash
  - Quick start of Beats

2.Product edition

2.1. Overview

Alibaba Cloud Elasticsearch offers two editions: Standard Edition and Advanced Edition. The two editions support different cluster versions. This topic describes how to view the edition and version of an Elasticsearch cluster, compares the two editions, and compares the various versions.

View the edition and version of a cluster

You can view the edition and version of your cluster on the **Basic Information** page of the cluster in the Elasticsearch console. For more information, see View the basic information of a cluster.

Version comparisons

Open source Elasticsearch version	Alibaba Cloud Elasticsearch version	Feature change
	V7.10	Alibaba Cloud Elasticsearch: • Some features, such as pruning for time series indexes and slow query isolation, are introduced based on the AliES kernel to improve query performance. • The faster-bulk and gig plug-ins are provided in addition to the plug-ins provided by open source Elasticsearch. The two plug-ins are used to improve cluster stability. Open source Elasticsearch: • The compression of storage fields is improved, which reduces storage costs. • Event Query Language (EQL) is used to improve security. • The default value of search.max_buckets is changed from 10000 to 65535. • Queries that are not case-sensitive are supported. To implement such queries, you must set the case_insensitive parameter to true. For more information about feature changes, see Breaking changes in 7.10.

Open source Elasticsearch version	Alibaba Cloud Elasticsearch version	Feature change
7.x	V7.7	Open source Elasticsearch: • The default number of shards in the index template is changed from 5 to 1. • Mapping types are removed. You do not need to specify a mapping type when you define a mapping or an index template. For more information, see Removal of mapping types. • By default, a maximum of 10,000 documents can be returned for each request. If more than 10,000 matching documents exist, Elasticsearch returns only 10,000 matching documents. For more information, see track_total_hits 10000 default. • By default, a single data node can store a maximum of 1,000 shards. You can use the cluster.max_shards_per_node parameter to change this limit. For more information, see Cluster Shard Limit. • By default, a maximum of 500 scrolls can be performed. You can use the search.max_open_scroll_context parameter to change this limit. For more information, see Scroll Search Context. • The parent circuit breaker works based on the current memory usage. This is controlled by the indices.breaker.total.use_real_memory parameter. By default, the parent circuit breaker starts to work when the current memory usage reaches 95% of JVM heap memory usage. This indicates that Elasticsearch uses the maximum memory availability to avoid out of memory (OOM) issues. For more information, see Circuit Breaker. • The _all field is removed to improve search performance. • Intervals queries are supported. Elasticsearch searches for and returns documents based on the order and proximity of matching terms. • After the audit logging feature is enabled, audit events are persisted to *<clustername>_audit.json* in the file system of each node. The audit events cannot be stored in indexes. For more information, see Enabling audit logging. For more information about feature changes, see Breaking changes in 7.0.

Open source Elasticsearch version	Alibaba Cloud Elasticsearch version	Feature change
6.x	V6.3, V6.7, and V6.8	Open source Elasticsearch: • An index can have only one type, and the _doc type is recommended. • The index lifecycle management (ILM) feature is introduced from V6.6.0 to reduce index O&M costs. • The historical data rollup feature is introduced to help summarize historical data. • Elasticsearch SQL, an X-Pack component, is supported in V6.3 and later. It enables SQL statements to be converted to domain-specific language (DSL) statements. This reduces costs for learning DSL. • The Composite, Parent, and Weighted Avg aggregation functions are supported. For more information about feature changes, see Breaking changes in 6.0.
	V6.7	Alibaba Cloud Elasticsearch: • Some features, such as pruning for time series indexes and slow query isolation, are introduced based on the AliES kernel to improve query performance. • The faster-bulk and gig plug-ins are provided in addition to the plug-ins provided by open source Elasticsearch. The two plug-ins are used to improve cluster stability. • The Advanced Monitoring and Alerting service is introduced. This service implements fine-grained monitoring and alerting, such as shard- or segment-level monitoring and alerting. • Upgrades from V6.3.2 to V6.7.0 are supported. For more information, see Upgrade the version of a cluster.

Open source Elasticsearch version	Alibaba Cloud Elasticsearch version	Feature change
5.x	V5.5 and V5.6	Open source Elasticsearch: • An index can have multiple types, and custom types are supported. • The STRING data type is replaced by the TEXT or KEYWORD data type. • The values of fields in indexes are changed from not_analyzed or no to true or false. • The DOUBLE data type is replaced by the FLOAT data type to reduce storage costs. • Java High Level REST Client is launched to replace Transport Client. For more information about feature changes, see Breaking changes in 5.0.

Create an Elasticsearch cluster

- Create an Alibaba Cloud Elasticsearch cluster

2.2. X-Pack advanced features

This topic provides a purchase guideline for X-Pack advanced features, describes the commonly used X-Pack advanced features, and compares the features offered by different editions.

Overview

X-Pack advanced features are the commercial features developed by the open source Elasticsearch team based on the X-Pack commercial plug-in. The features include security, SQL plug-in, machine learning, alerting, and monitoring. These features enhance the service capabilities of open source Elasticsearch in terms of application development and O&M management.

Alibaba Cloud Elasticsearch provides editions that support the advanced features. You can purchase the features when you create a cluster. The following sections describe the detailed information of these features.

Purchase guideline

You can log on to the Alibaba Cloud Elasticsearch console and click Create on the Elasticsearch Clusters page to purchase X-Pack advanced features. Only the Standard Edition of Alibaba Cloud Elasticsearch supports the advanced features. The following table lists the related information of the Standard Edition.

Item	Standard Edition
Whether X-Pack is included	Yes
Whether all X-Pack features are provided	Yes

Note

In addition to X-Pack advanced features, Alibaba Cloud Elasticsearch clusters of the Standard Edition provide features such as O&M management, security, plug-ins, and high availability. For more information, see Alibaba Cloud Elasticsearch clusters of the Standard Edition.

Feature description

This section describes only a few commonly used advanced features. For more information about all the X-Pack advanced features, see Elastic Stack subscriptions and X-Pack APIs.

Notice The X-Pack advanced features provided by Elastic Stack differ in various editions, such as FREE AND OPEN, GOLD, and PLATINUM. Alibaba Cloud subscribes to Elasticsearch of the PLATINUM edition.

Feature	Description
Security	Manages indexes and fields in a decentralized manner and strictly controls access permissions to improve data security.
Machine learning	Monitors data in real time, provides the auto alerting feature, and reports alerts.
Monitoring	Monitors objects, such as clusters, nodes, and indexes, in real time to improve development efficiency and reduce O&M costs.
SQL plug-in	• Implements full-text searches and statistical analysis on Elasticsearch data based on traditional SQL databases. • Supports access methods such as CLI and REST. In the PLATINUM edition, the SQL plug-in also supports the Java Database Connectivity (JDBC) method. • Seamlessly integrates with original business systems, which reduces the costs for learning new techniques. Note In the FREE AND OPEN edition, other SQL plug-ins are integrated. For more information, see elasticsearch-sql.

Feature comparisons between editions

This section compares some of the key X-Pack features between the FREE AND OPEN and PLATINUM editions. This helps you understand the differences between features in the FREE AND OPEN edition and those in the PLATINUM edition. Elasticsearch is rapidly developing. Therefore, features supported in different editions are constantly updated. For more information about the latest comparisons of open source Elasticsearch features among different editions, see Elastic Stack subscriptions.

The following table compares some of the key X-Pack features in the FREE AND OPEN and PLATINUM editions.

Note In the following table, the symbols ✓, ○, and × are used to indicate the integrity of the features.

- ✓: indicates that all the subfeatures of a feature are provided.
- ○: indicates that only some of the subfeatures of a feature are provided.
- ×: indicates that a feature is not provided.

Module	Feature	FREE AND OPEN	PLATINUM
Elasticsearch	Scalability and resiliency	○	✓
	Query and analysis	○	✓
	Stack management	○	✓
	Security	×	✓
	Machine learning	×	✓
	Watcher	×	✓
Kibana	Exploration and visualization	○	✓
	Stack management	○	✓
	Stack detection	×	✓
	Kibana alerting	×	✓
	Security	×	✓
	Machine learning	×	✓
Beats	Data collection	○	✓
	Data transmission	○	✓
	Monitoring and management	×	✓
	Module	○	✓
Logstash	Data collection	✓	✓
	Data enrichment	✓	✓
	Data transmission	✓	✓
	Module	○	✓

Module	Feature	FREE AND OPEN	PLATINUM
	Monitoring and management	×	✓
Elastic APM	APM server	✓	✓
	APM agent	✓	✓
	Kibana APM dashboard	✓	✓
	APM UI	×	✓
	Distributed tracing	×	✓
	Integration of machine learning	×	✓
Elastic logs	Log shipper (Filebeat)	✓	✓
	Common data source dashboard	✓	✓
	Logs UI	×	✓
Elastic infrastructure	Metric shipper	✓	✓
	Common data source dashboard	✓	✓
	Infrastructure UI	×	✓
Elastic status monitoring	Status monitoring (Heartbeat)	✓	✓
	Status dashboard in Kibana	✓	✓
	Status monitoring UI	×	✓

Some features provided by open source Elasticsearch

The following table lists some features provided by open source Elasticsearch.

2.3. Alibaba Cloud Elasticsearch clusters of the Standard Edition

Alibaba Cloud Elasticsearch offers two editions of clusters: Standard Edition and Advanced Edition. This topic provides an overview of Alibaba Cloud Elasticsearch clusters of the Standard Edition. Alibaba Cloud Elasticsearch clusters of the Standard Edition provide fully managed Elasticsearch services. Such clusters support all the features provided by open source Elasticsearch and all the advanced features provided by the X-Pack plug-in. Multiple versions from V5.X to V7.X are available for the clusters. In addition, a variety of capabilities are provided to facilitate cluster management, configuration, and O&M and ensure security and high availability.

Features

Category	Feature	Description
Cluster management and O&M	Cluster creation	Enables you to quickly create and configure a cluster based on your business requirements. Multiple types of nodes and disk types are available for the cluster.
	Cluster upgrade	Enables you to quickly upgrade the version or update the kernel of a cluster.
	Scaling	Allows you to configure auto scaling rules or manually perform scaling for nodes to flexibly cope with your business fluctuations.
	Cluster topology visualization	• Allows you to view the status of a cluster, the statuses of zones where the cluster resides, and the basic information of nodes in the cluster. • Allows you to perform a switchover for faulty nodes, migrate nodes, or restart a cluster or nodes with one click.
	Authorization management	Allows you to grant permissions to users by using the Resource Access Management (RAM) service or role-based access control (RBAC) policies.
	Monitoring and alerting	Allows you to use the Cloud Monitor or Advanced Monitoring and Alerting service with one click to monitor the metrics of clusters and nodes and report alerts.
	Log viewing	Allows you to view multiple types of logs, such as cluster logs, slow logs, and access logs.
	EYou	Diagnoses clusters from multiple dimensions, analyzes possible risks, and provides the optimal solutions.
	Disaster recovery	Enables you to quickly deploy a cluster across zones, which improves the stability of upper-layer services.

Category	Feature	Description
Security and high availability	Network configuration	Allows you to configure public and private IP address whitelists based on your business requirements. Cluster access over virtual private clouds (VPCs) is enabled by default.
	Security configuration	• Allows you to enable HTTPS with one click. • Allows you to use Key Management Service (KMS) to encrypt data at rest before you store the data in Elasticsearch.
	Data backup	• Enables the system to automatically back up data on a regular basis. • Allows you to manually back up and restore data across clusters.
Configuration and plug-in center	Scenario-based configuration	Provides scenario-based configuration templates to optimize default cluster configurations.
	Cluster configuration center	Allows you to customize configurations, such as the YML file, synonym dictionary, and garbage collector (GC).
	Plug-in center	Provides a variety of self-developed plug-ins by default. You can use the plug-ins to perform advanced searches, improve cluster stability, or optimize cluster performance based on your business requirements.
	Plug-in customization	Allows you to upload custom plug-ins and dictionaries based on your business requirements.
Integration with other services in the Elastic Stack ecosystem	Fully managed Elastic Stack ecosystem	Provides the Logstash and Beats services to help you collect and process data.
	Visualization management center	Provides the Kibana, Grafana, and DataV services for you to manage visualization.
	Data migration and synchronization	Allows you to migrate or synchronize data from various self-managed Elasticsearch clusters, databases, and other big data services.

Capabilities and advanced commercial features provided by the open source Elastic Stack ecosystem

Alibaba Cloud Elasticsearch clusters of the Standard Edition support the following open source Elasticsearch features. For more information about the features, see Elastic Stack features. Due to the rapid iteration of open source Elasticsearch versions, the features supported by each version are constantly updated. For more information about the features supported by each version, see Version comparisons.

Category	Subcategory	Feature
	Scalability and resiliency	Clustering and high availability
		Automatic node recovery
		Automatic data rebalancing
		Horizontal scalability
		Rack awareness
		Cross-cluster replication
		Cross-data center replication
	Monitoring	Full stack monitoring
		Multi-stack monitoring
		Configurable retention policy
		Automatic alerts on stack issues
	Management	Index lifecycle management
		Data tiers
		Frozen indexes
		Snapshot creation and data restoration
		Searchable snapshots
		Source-only snapshots
		Snapshot lifecycle management
		Data rollup
		Data streams
		CLI tools
		Upgrade assistant UI

Category	Subcategory	Feature
Management and operations		Upgrade assistant APIs
		User and role management
		Transforms
	Alerting	Highly available, scalable alerting
		Notifications
		Alerting UI
	Stack security	Security settings
		Encrypted communications
		Support for encryption at rest
		RBAC
		Field- and document-level security
		Audit logging
		IP address filtering
		Security realms
		Single sign-on (SSO)
		Third-party security integration
	Clients	RESTful APIs
		Language clients
		Console
		DSL
		SQL
		Event query language (EQL)
		JDBC client

Category	Subcategory	Feature
		ODBC client
Data collection and enrichment	Data sources	Operating systems
		Web servers and proxies
		Data repositories and queues
		Cloud services
		Containers
		Network data
		Security data
		Running status data
		File import
	Data enrichment	Processors
		Analyzers
		Tokenizers
		Filters
		Language analyzers
		Grok
		Field transformation
		External lookups
		Match enrich processor
		Geo-match enrich processor
	Modules and integrations	Clients and APIs
		Beats
		Community shippers
		Logstash
		Elasticsearch-Hadoop

Category	Subcategory	Feature
		Plug-ins and integrations
Data storage	Flexibility	Data types
		Full-text searches
		Document databases
		Time series and analysis
		Geospatial
	Security	Support for encryption at rest
		Field-level security
	Management	Clustered indexes
		Snapshot creation and data restoration
		Index rollup
	Full-text searches	Inverted indexes
		Cross-cluster searches
		Relevance scoring
		Query DSL
		Asynchronous searches
		Highlighters
		Automatic completion
		Spelling checks and corrections
		Suggesters
		Percolators
		Query optimizer
		Permissions-based search results
		Query cancellation

Category	Subcategory	Feature
Search and analysis	Analytics	Aggregations
		Graph searches
		Threshold-based alerting
	Machine learning	Inference
		Forecasting on time series
		Anomaly detection on time series
		Alerting on anomalies
	APM	APM server
		APM agents
		APM applications
		Distributed tracing
		Alerting
		Service maps
	Visualization	Dashboards
		Canvas
		Kibana Lens
		Time Series Visual Builder (TSVB)
		Graph analysis
		Geospatial analysis
		Container monitoring
		Kibana plug-ins
		Data import tutorial
	Maps	Map layers
		Custom area maps

Category	Subcategory	Feature
		GeoJSON upload
	Elastic logs	Log shipper
		Log dashboards
		Detection on log rate anomalies
	Elastic metrics	Metric shipper
		Metric dashboards
		Alerting
	Uptime	Uptime monitoring
		Uptime dashboards
		Alerting
		Certificate monitoring
		Synthetic monitoring
	Security analysis	Common Schema
		Security analysis
		Timeline events
		Case management
		Anomaly detection

Create an Elasticsearch cluster

- Create an Alibaba Cloud Elasticsearch cluster

Get started with the Standard Edition

Quick start

3.Benefits

3.1. Comparison between Alibaba Cloud Elasticsearch clusters and self-managed Elasticsearch clusters

Alibaba Cloud Elasticsearch provides a fully managed Elasticsearch service and is fully compatible with open source Elasticsearch. Alibaba Cloud Elasticsearch optimizes kernel performance and provides the commercial plug-in X-Pack free of charge. Alibaba Cloud Elasticsearch is out-of-the-box. It features high availability and auto scaling, and supports the pay-as-you-go billing method. This topic describes the comparisons between Alibaba Cloud Elasticsearch clusters and self-managed Elasticsearch clusters in terms of costs, cluster management, supported capabilities, security, and availability.

Costs

Item	Alibaba Cloud Elasticsearch cluster	Self-managed Elasticsearch cluster hosted on an ECS instance
Resource costs	• Auto scaling is supported, which allows you to change the specifications, number, disk type, and disk space of nodes.	• Resources may be insufficient during peak hours and may be wasted during off-peak hours.
Network fees	• You can access a cluster over an internal network free of charge by using an Elastic Compute Service (ECS) instance that resides in the same region as the cluster. • You can access a cluster over the Internet free of charge by using an ECS instance that resides in a different region from the cluster. Alibaba Cloud Elasticsearch provides the Public Network Access feature, which is enabled by default. The maximum network bandwidth for access to a cluster over the Internet is 4 GB/s.	• You can access a cluster over an internal network free of charge by using an ECS instance that resides in the same region as the cluster. • You can access a cluster over the Internet by using an ECS instance that resides in a different region from the cluster. However, you are charged for the access. For more information about the billing standards of the access, see Public bandwidth.
Labor and time costs	• Alibaba Cloud Elasticsearch, Logstash, and Kibana (ELK) are fully managed services. They are out-of-the-box and support the pay-as-you-go billing method. • A visual interface is provided for cluster O&M. This reduces O&M costs.	• You must purchase a machine and manually deploy a cluster. This process requires a long period of time, and the iteration of the cluster is slow. • A professional Elasticsearch engineer team is required to manage resources and perform cluster O&M. This results in high labor costs.
Costs for risk reduction	Alibaba Cloud Elasticsearch guarantees 99.9% reliability and features few IT risks, and upper-level business risks are controllable.	Service reliability is not guaranteed. In this case, technical expertise and a large number of investments are required to reduce business risks.

Item	Alibaba Cloud Elasticsearch cluster	Self-managed Elasticsearch cluster hosted on an ECS instance
Feature costs	- All the advanced features of the commercial plug-in X-Pack are provided free of charge. - The OSS-based data backup feature is provided free of charge.	- You must pay USD 5,000 for the X-Pack plug-in. - You must manually back up data, and you are charged for the storage used by backups.

Capabilities

Item	Alibaba Cloud Elasticsearch cluster	Self-managed Elasticsearch cluster hosted on an ECS instance
Usability	- Clusters are out-of-the-box and support auto scaling. You can modify cluster configurations based on your business requirements with one click. - You can upgrade the version of a cluster with one click. - The intelligent O&M system EYou is provided. It can detect the health of more than 20 items, such as clusters, nodes, and indexes. It can also diagnose and analyze exceptions.	- Cluster deployment is complex, and resources must be manually adjusted. - Data must be migrated before you upgrade the version of a cluster. - Cluster O&M is complex. You must run commands to view the health statuses of clusters, nodes, and indexes.
Capabilities for scenario support	- All the advanced features of the X-Pack plug-in are provided free of charge. - Scenario-based configuration templates are used to provide appropriate parameter configurations. - A natural language processing (NLP) plug-in developed by Alibaba DAMO Academy, a vector search plug-in, and a self-developed SQL plug-in are provided. These plug-ins improve the performance of clusters in search scenarios. - Cold and hot data separation is supported, and an index compression plug-in is provided. This improves the performance of clusters in logging scenarios.	You must develop the capabilities or integrate the capabilities of open source Elasticsearch on your own.
Performance	- An enhanced kernel is provided to improve read and write performance.	You must ensure that the cluster performance meets your requirements, which is a complex process.

Item	Alibaba Cloud Elasticsearch cluster	Self-managed Elasticsearch cluster hosted on an ECS instance
Availability	• Data can be automatically backed up. • Data and service reliability reaches 99.9%. • A self-developed throttling plug-in and the slow query isolation feature are provided to ensure cluster stability. • Multi-zone cluster deployment is supported, and an active zone-redundancy architecture is provided.	• You must ensure cluster availability on your own and manually back up data. • Disaster recovery is difficult to implement.
Security	• Clusters are accessed over virtual private clouds (VPCs) by default. • X-Pack security components are provided free of charge. • Field-level access control is supported. • Data can be transmitted after HTTPS-based encryption and can be stored after encryption.	• The security of ECS instances is ensured, but clusters have security risks. • X-Pack security components must be separately purchased.

3.2. High availability

Alibaba Cloud Elasticsearch provides the data backup and restoration, load balancing, and cross-zone deployment features. It also provides various kernel optimization policies to ensure cluster stability. These features and policies ensure comprehensive data reliability and service availability.

Data backup and restoration

Backup and restoration mode	Description
Automatic snapshot creation and data restoration from automatic snapshots	Alibaba Cloud Elasticsearch supports automatic snapshot creation. You can specify the time at which snapshots are automatically created every day. After automatic snapshots are created, you can restore data from an automatic snapshot that is created within three days to the original Elasticsearch cluster. For more information, see Create automatic snapshots and restore data from automatic snapshots.
Manual snapshot creation and data restoration from manual snapshots	Alibaba Cloud Elasticsearch allows you to manually run a command to create a snapshot for a specific index. Then, you can save the snapshot in an Object Storage Service (OSS) bucket in the same region as your Elasticsearch cluster. After the snapshot is created, you can manually run a command to restore the data in the snapshot to the original Elasticsearch cluster or an Elasticsearch cluster that is in the same region as the original Elasticsearch cluster. For more information, see Create manual snapshots and restore data from manual snapshots.

Backup and restoration mode	Description
Shared OSS repository	Alibaba Cloud Elasticsearch allows you to configure shared OSS repositories for your Elasticsearch cluster. This way, you can restore data from the automatic snapshots of an Elasticsearch cluster that are stored in these repositories to your Elasticsearch cluster. For more information, see Configure a shared OSS repository.

Load balancing

Alibaba Cloud Elasticsearch supports load balancing. You can specify the public or internal endpoint of your Elasticsearch cluster on your application. Your requests are evenly distributed to all the data nodes in your Elasticsearch cluster to achieve load balancing.

Notice Load balancing among these data nodes depends on the number and size of index shards. When you create an index, you must set the number and size of index shards to appropriate values. For more information, see Shard evaluation.

Cross-zone deployment

Alibaba Cloud Elasticsearch allows you to deploy an Elasticsearch cluster across zones. In cross-zone deployment, the system automatically selects the zones. If replica shards are configured and nodes in one zone fail, the nodes in the remaining zones can still provide services without interruptions. This significantly enhances the availability of the cluster. In addition, you can perform a switchover in the console to isolate the faulty nodes. During the switchover, the system adds computing resources to the remaining zones to make up for the resources lost in the zone that contains the faulty nodes. After the nodes recover, you can perform a recovery for the zone in the Elasticsearch console. During the recovery, the system adds the nodes that were removed during the switchover to the zone again. It also removes the computing resources that were added to the remaining zones during the switchover. The switchover and recovery are imperceptible to customer services and improve service stability. For more information, see Deploy and use a multi-zone Elasticsearch cluster.

AliES enhancements

The Alibaba Cloud Elasticsearch team continuously develops and optimizes the Elasticsearch kernel to improve cluster stability and availability. The following table describes the features that are provided by the optimized kernel.

Feature	Description
Pruning for time series indexes	When you query data from a time series index, you can specify a time range to filter the data. This feature improves the query performance of time series indexes.
Slow query isolation	This feature allows you to track the overheads for a single query request and logically isolate the request. This reduces the impact of anomalous queries on cluster stability.
gig plug-in	When an exception occurs in a cluster, the gig plug-in can perform a switchover within seconds. This prevents query jitters caused by anomalous nodes.

For information about other features provided by the optimized kernel, see AliES release notes.

3.3. High reliability

Alibaba Cloud Elasticsearch automatically creates snapshots for Elasticsearch clusters. You can directly restore snapshots stored in Alibaba Cloud Elasticsearch, or save snapshots to Object Storage Service (OSS) and then restore data from the snapshots. Alibaba Cloud Elasticsearch also supports load balancing. You can use these features to ensure the high reliability of Alibaba Cloud Elasticsearch clusters.

Automatic snapshot creation and data restoration from the snapshots

- Automatic snapshot creation

  Alibaba Cloud Elasticsearch supports automatic creation of snapshots. You can enable this feature on the **Snapshots** page in the Elasticsearch console. You can specify the time when daily snapshots are automatically created based on your business requirements. This feature facilitates disaster recovery. For more information, see Data backup overview.

- Data restoration from automatically created snapshots

  After snapshots are created, you can restore data from the specified snapshots based on your business requirements. For more information, see Create automatic snapshots and restore data from automatic snapshots.

  > **Note**
  >
  > - Alibaba Cloud Elasticsearch stores only snapshots that were created in the last three days.
  > - You can restore data from automatically created snapshots only to the Alibaba Cloud Elasticsearch cluster where the snapshots are created.

Manual snapshot creation and data restoration from the snapshots

- Manual snapshot creation

  Alibaba Cloud Elasticsearch allows you to save snapshots on your Elasticsearch cluster to OSS. If you want to use OSS, you must activate OSS and create a bucket in the same region as your Alibaba Cloud Elasticsearch cluster. You can also call the snapshot operation to create snapshots for specified indexes. For more information, see Create manual snapshots and restore data from manual snapshots.

  > **Note** You can store the snapshots only in OSS buckets of the Standard storage class. Snapshots cannot be stored in OSS buckets of the Archive storage class.

- Manual data restoration from snapshots

  Alibaba Cloud Elasticsearch allows you to call the restore operation to restore index data from a specified snapshot. This feature facilitates disaster recovery. For more information, see Create manual snapshots and restore data from manual snapshots.

Note

- You can restore data from snapshots in an OSS bucket only to an Alibaba Cloud Elasticsearch cluster that is in the same region as the bucket.
- By default, each Alibaba Cloud Elasticsearch data node can process 40 MiB of data. You can modify the `max_restore_bytes_per_sec` parameter to adjust the data processing capability of the data nodes. For more information, visit Snapshot and restore on the open-source Elasticsearch website.

Load balancing

Alibaba Cloud Elasticsearch supports load balancing. You can specify the public or internal endpoint of your Alibaba Cloud Elasticsearch cluster on your client for access. Your requests are evenly distributed to all data nodes of your Alibaba Cloud Elasticsearch cluster based on load balancing.

Notice Load balancing among these data nodes depends on the number and size of index shards. When you create indexes, you must consider the number and size of index shards. For more information, see Shard evaluation.

3.4. High security

Alibaba Cloud Elasticsearch clusters are deployed in logically isolated virtual private clouds (VPCs). In addition, access control, authentication and authorization, encryption, and the advanced security features provided by X-Pack are used for the clusters. All the preceding features ensure the high security of Alibaba Cloud Elasticsearch clusters. This topic describes the features.

Background information

Open source software is often the first target of attacks. MongoDB ransomware attacks are an example. Elasticsearch has also become the target of attacks. Attackers may attack self-managed Elasticsearch clusters that do not have professional security protection, delete important data, or interfere with business systems.

Alibaba Cloud Security Center released a warning about the security risks associated with Elasticsearch and provided an array of security hardening strategies and solutions. Alibaba Cloud Elasticsearch provides solutions that are more reliable and professional for data and service security than those provided by open source Elasticsearch.

Security features

Alibaba Cloud released the fully managed Elasticsearch service in November 2017. Alibaba Cloud Elasticsearch provides security protection features for you to safeguard your clusters.

The following table compares the security protection of an Alibaba Cloud Elasticsearch cluster with that of a self-managed Elasticsearch cluster.

Category	Built-in security feature of an Alibaba Cloud Elasticsearch cluster	Security protection of a self-managed Elasticsearch cluster

Category	Built-in security feature of an Alibaba Cloud Elasticsearch cluster	Security protection of a self-managed Elasticsearch cluster
Access control	• Clusters are deployed in VPCs. This way, the clusters can be isolated at the data link layer. • Both Elasticsearch and Kibana support whitelists for access control. You can specify IPv4 addresses, IPv6 addresses, and Classless Inter-Domain Routing (CIDR) blocks in whitelists. By default, no IP address is allowed to access the public endpoint of a cluster. If you want to allow access requests, you must configure a whitelist. For more information, see Configure a public or private IP address whitelist for an Elasticsearch cluster. • Users are not allowed to log on to the node servers that are contained in a cluster. • Users can use only ports 9200 and 9300 to access the public and internal endpoints of clusters.	• Purchase cloud security products, such as security groups or Cloud Firewall, to manage and quarantine source IP addresses. • Disable port 9200 unless you plan to use it. • Bind source IP addresses. • Change the default port.
Authentication and authorization	• Cluster-level permission policies in Resource Access Management (RAM), such as the ReadOnlyAccess policy that grants read-only permissions and the FullAccess policy that grants administrator permissions. • RAM-based access control, such as the permissions on clusters, accounts, and GET, POST, and PUT commands. • Role-based access control (RBAC) provided by X-Pack. Access control policies can be specific to data fields. • Single sign-on (SSO) based on X-Pack. Active Directory, LDAP, and Elasticsearch-native Realm are supported for identity verification.	Install third-party security plug-ins, such as Search Guard and Shield.
Encryption	• HTTPS is supported. • Encryption at rest is provided based on Key Management Service (KMS). • X-Pack is integrated to support data transmission encryption by using SSL or TLS.	• Use storage media that support encryption at rest. • Disable HTTP in YML configuration files.

Category	Built-in security feature of an Alibaba Cloud Elasticsearch cluster	Security protection of a self-managed Elasticsearch cluster
Monitoring and auditing	• Operations log auditing based on X-Pack. • Cloud Monitor-based cluster monitoring with multiple metrics, such as cluster workload.	Use third-party tools to audit logs and monitor services.
Disaster recovery	• Snapshots are automatically created at a scheduled time. • A cluster can be deployed across zones in one city to implement disaster recovery.	• Purchase file systems to periodically back up data. • Use multiple clusters to implement disaster recovery.

Access control

Alibaba Cloud Elasticsearch uses the following methods to control access:

- Access over VPCs

  You can use the internal endpoint of an Alibaba Cloud Elasticsearch cluster to access the cluster over a VPC. If you require a secure environment where your applications can access your Alibaba Cloud Elasticsearch cluster, you can purchase an Alibaba Cloud Elastic Compute Service (ECS) instance in the same zone, region, and VPC as the Elasticsearch cluster. Then, deploy the applications on the ECS instance and use the ECS instance to access the internal endpoint of the Elasticsearch cluster.

  > **Note** A VPC is a private network in the cloud and is isolated from the Internet. It provides secure access for your applications.

- Whitelist-based access control

  If you want to use the internal endpoint of an Alibaba Cloud Elasticsearch cluster to access the cluster, configure a whitelist for the cluster to control access. Only clients whose IP addresses are in the whitelist can be used to access the cluster. For more information, see Configure a public or private IP address whitelist for an Elasticsearch cluster.

  If you want to use the public endpoint of an Alibaba Cloud Elasticsearch cluster to access the cluster, configure a whitelist for the cluster to control access. Only clients whose IP addresses are in the whitelist can be used to access the cluster. For more information, see Configure a public or private IP address whitelist for an Elasticsearch cluster.

Authentication and authorization

- RAM-based access control

  The Alibaba Cloud Elasticsearch console supports RAM users. You can use RAM users to isolate resources. A RAM user can view and manage only Alibaba Cloud Elasticsearch clusters on which the user has permissions. For more information, see Policy evaluation process.

- RBAC provided by X-Pack

Alibaba Cloud Elasticsearch provides the X-Pack plug-in, which is a commercial extension of Elasticsearch. The plug-in is an easy-to-install bundle that provides security, alerting, monitoring, graphing, and reporting capabilities. The plug-in is integrated into Kibana to provide more capabilities, such as authentication and authorization, RBAC, real-time monitoring, visual reporting, and machine learning. RBAC can be specific to indexes. For more information, see Use the RBAC mechanism provided by Elasticsearch X-Pack to implement access control and Security APIs in the open source Elasticsearch documentation.

3.5. High performance

Developed based on open source Elasticsearch, Alibaba Cloud Elasticsearch provides various features to optimize write and query performance in different scenarios. The features also help you reduce costs. This topic describes the features that are provided by Alibaba Cloud Elasticsearch to achieve high performance.

High-performance hardware and high-speed access

Alibaba Cloud Elasticsearch supports a variety of servers and storage hardware, and follows the latest hardware iterations to fully ensure cluster performance and stability at the hardware level. In addition, communications over internal networks are used to reduce the response time of applications.

Scenario-based templates

Alibaba Cloud Elasticsearch provides scenario-based templates. All parameters in the templates are developed and optimized based on years of experience. You can select an appropriate template based on your business requirements to optimize the read and write performance of your Elasticsearch cluster in the related scenario. This reduces cluster performance issues caused by inappropriate configurations.

Kernel performance optimization

The AliES kernel of Alibaba Cloud Elasticsearch is developed and optimized to enhance cluster performance. Alibaba Cloud Elasticsearch allows you to update the kernel for higher cluster performance. For more information about the release notes of the AliES kernel, see AliES release notes.

Alibaba Cloud Elasticsearch provides the following high-performance features:

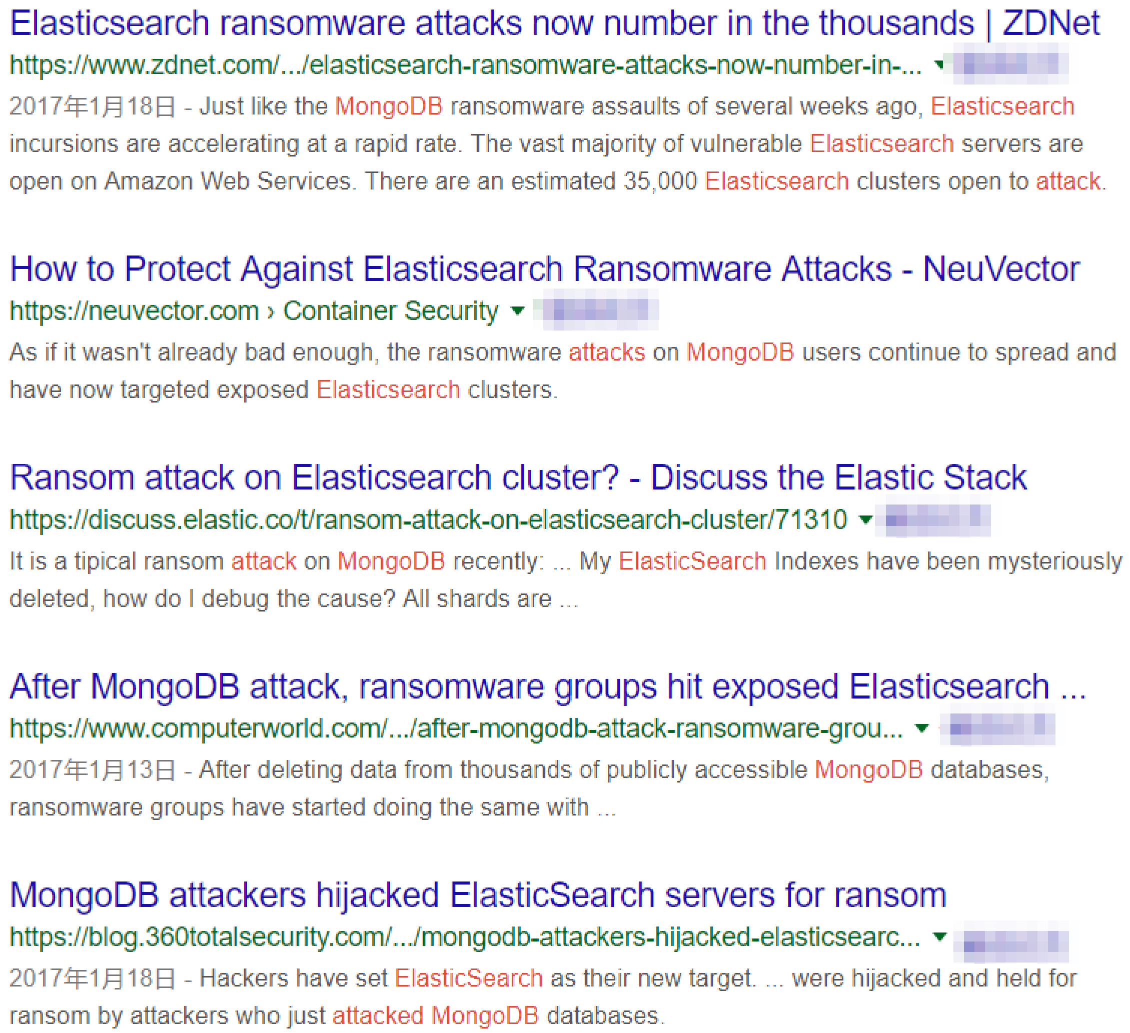
- Physical replication: This feature improves the write performance of indexes that have replica shards. For more information, see Use the physical replication feature of the apack plug-in.
- Pruning for time series indexes: This feature improves the query performance of time series indexes. For more information, see Use the pruning feature for a time series index.
- Primary key-based data deduplication during queries: This feature is optimized. It improves the write performance by 10% for documents that contain primary keys.
- Shard scheduling by dedicated master nodes: This feature is improved. It improves the shard scheduling performance of dedicated master nodes by 10 times. Each dedicated master node is allowed to schedule more shards.
- Translog optimization: This feature reduces the overheads of translog flush and improves write performance by 10%.

3.6. Security features

This topic compares Alibaba Cloud Elasticsearch clusters with user-created Elasticsearch clusters to describe the security protection advantages of Alibaba Cloud Elasticsearch.

Background information

Open-source software is often the first target of attacks. The MongoDB ransomware attacks are an example. Elasticsearch has also become the target of attacks. Attackers may attack user-created Elasticsearch clusters that do not have professional security protection, and then delete important data or interfere with the business system.



Elasticsearch ransomware attacks now number in the thousands | ZDNet
https://www.zdnet.com/.../elasticsearch-ransomware-attacks-now-number-in-... ▾ [redacted]
2017年1月18日 - Just like the MongoDB ransomware assaults of several weeks ago, Elasticsearch incursions are accelerating at a rapid rate. The vast majority of vulnerable Elasticsearch servers are open on Amazon Web Services. There are an estimated 35,000 Elasticsearch clusters open to attack.

How to Protect Against Elasticsearch Ransomware Attacks - NeuVector
https://neuvector.com › Container Security ▾ [redacted]
As if it wasn't already bad enough, the ransomware attacks on MongoDB users continue to spread and have now targeted exposed Elasticsearch clusters.

Ransom attack on Elasticsearch cluster? - Discuss the Elastic Stack
https://discuss.elastic.co/t/ransom-attack-on-elasticsearch-cluster/71310 ▾ [redacted]
It is a tipical ransom attack on MongoDB recently: ... My ElasticSearch Indexes have been mysteriously deleted, how do I debug the cause? All shards are ...

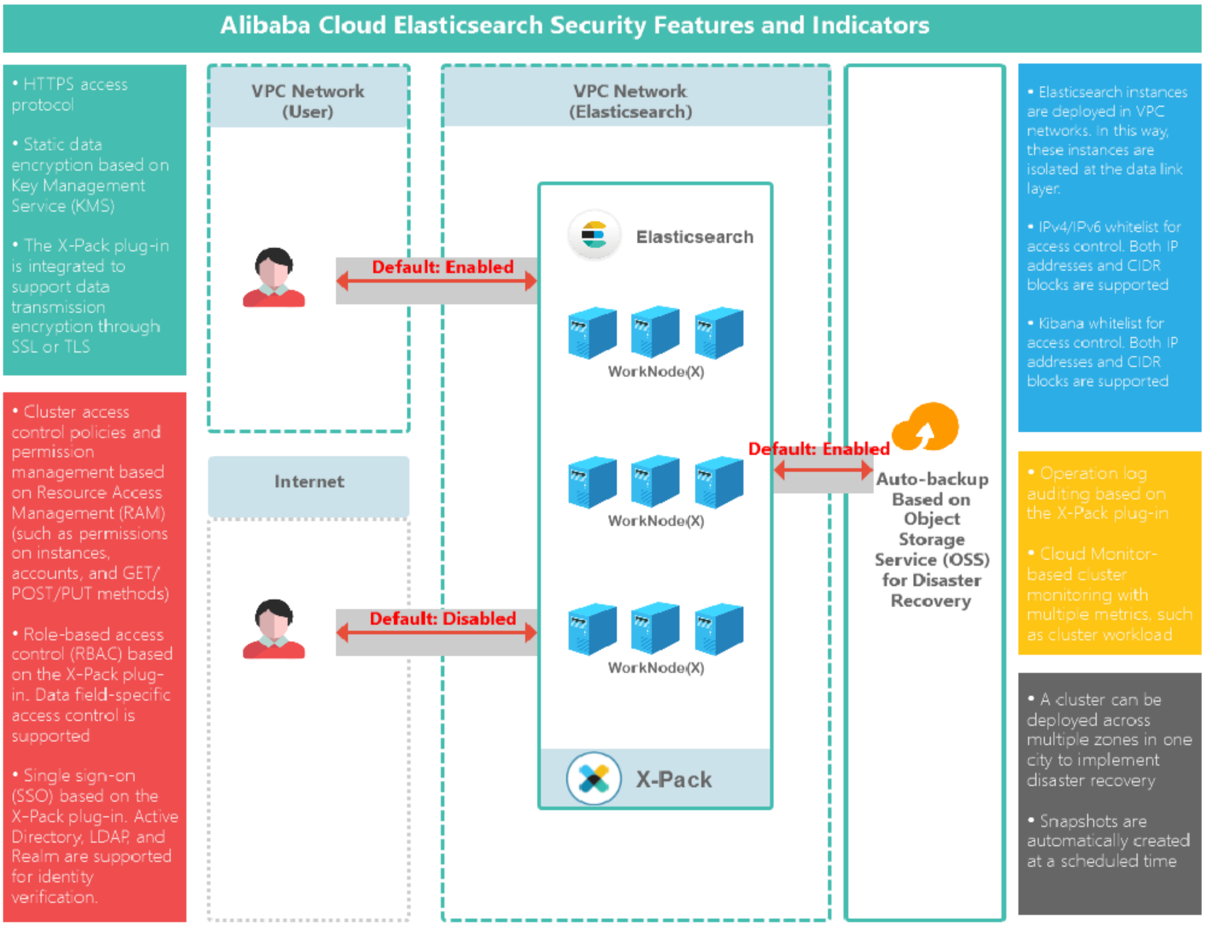
After MongoDB attack, ransomware groups hit exposed Elasticsearch ...
https://www.computerworld.com/.../after-mongodb-attack-ransomware-grou... ▾ [redacted]
2017年1月13日 - After deleting data from thousands of publicly accessible MongoDB databases, ransomware groups have started doing the same with ...

MongoDB attackers hijacked ElasticSearch servers for ransom
https://blog.360totalsecurity.com/.../mongodb-attackers-hijacked-elasticsearc... ▾ [redacted]
2017年1月18日 - Hackers have set ElasticSearch as their new target. ... were hijacked and held for ransom by attackers who just attacked MongoDB databases.

Alibaba Cloud Security Center released a warning about the security risks of Elasticsearch and provided multiple security hardening strategies and solutions. Alibaba Cloud Elasticsearch provides more reliable and professional solutions for data and service security than user-created Elasticsearch.

Security feature descriptions

Alibaba Cloud released the fully hosted Elasticsearch service in November, 2017. Alibaba Cloud Elasticsearch provides security protection features for you to safeguard your clusters.


Alibaba Cloud Elasticsearch Security Features and Indicators
VPC Network (User)
Default: Enabled
Internet
Default: Disabled
VPC Network (Elasticsearch)
Elasticsearch
WorkNode(X)
WorkNode(X)
WorkNode(X)
X-Pack
Default: Enabled
Auto-backup Based on Object Storage Service (OSS) for Disaster Recovery

- HTTPS access protocol
- Static data encryption based on Key Management Service (KMS)
- The X-Pack plug-in is integrated to support data transmission encryption through SSL or TLS

- Cluster access control policies and permission management based on Resource Access Management (RAM) (such as permissions on instances, accounts, and GET/POST/PUT methods)
- Role-based access control (RBAC) based on the X-Pack plug-in. Data field-specific access control is supported
- Single sign-on (SSO) based on the X-Pack plug-in. Active Directory, LDAP, and Realm are supported for identity verification.

- Elasticsearch instances are deployed in VPC networks. In this way, these instances are isolated at the data link layer.
- IPv4/IPv6 whitelist for access control. Both IP addresses and CIDR blocks are supported
- Kibana whitelist for access control. Both IP addresses and CIDR blocks are supported

- Operation log auditing based on the X-Pack plug-in
- Cloud Monitor-based cluster monitoring with multiple metrics, such as cluster workload

- A cluster can be deployed across multiple zones in one city to implement disaster recovery
- Snapshots are automatically created at a scheduled time

The following table compares the security protection of Alibaba Cloud Elasticsearch with that of user-created Elasticsearch.

Security metric	Security protection of user-created Elasticsearch	Integrated security features of Alibaba Cloud Elasticsearch
Access control	• Purchase cloud security products, such as security groups or firewalls, to manage and quarantine source IP addresses. • Disable port 9200 unless it is necessary. • Bind source IP addresses. • Change the default port.	• Alibaba Cloud Elasticsearch clusters that are deployed in VPCs. This way, they can be isolated at the data link layer. • IPv4 and IPv6 whitelists for access control. Both IP addresses and Classless Inter-Domain Routing (CIDR) blocks are supported. • Kibana whitelists for access control. Both IP addresses and CIDR blocks are supported.

Security metric	Security protection of user-created Elasticsearch	Integrated security features of Alibaba Cloud Elasticsearch
Authentication and authorization	Install third-party security plug-ins, such as Search Guard and Shield.	• Cluster-level permission policies in Resource Access Management (RAM), such as the ReadOnlyAccess policy that grants the read-only permissions and the FullAccess policy that grants the administrator permissions. • Access control based on RAM, such as the permissions on clusters, accounts, and GET, POST, and PUT commands. • Role-based access control (RBAC) based on X-Pack. Access control policies can be specific to data fields. • Single sign-on (SSO) based on X-Pack. Active Directory, LDAP, and Elasticsearch-native Realm are supported for identity verification.
Data encryption	• Use storage media that support encryption at rest. • Disable HTTP in YML configuration files.	• HTTPS is supported. • Encryption at rest is provided based on Key Management Service (KMS). • X-Pack is integrated to support data transmission encryption by using SSL or TLS.
Monitoring and auditing	Use third-party tools to audit logs and monitor services.	• Operation log auditing based on X-Pack. • CloudMonitor-based cluster monitoring with multiple metrics, such as cluster workload.
Disaster recovery	• Purchase file systems to back up data periodically. • Use multiple clusters to implement disaster recovery.	• A cluster can be deployed across multiple zones in one city to implement disaster recovery. • Snapshots are automatically created at a scheduled time.

4.Common scenarios

Alibaba Cloud Elasticsearch is suitable for a wide range of scenarios, such as integrated monitoring, intelligent operation and maintenance (O&M), information retrieval, and data intelligence.

Integrated monitoring and intelligent O&M

In complex business scenarios, large amounts of metrics, logs, and application performance monitoring (APM) data of different structures and types exist on various devices. This brings great challenges to end-to-end exception identification, business analysis, and business O&M. The devices include servers, physical machines, Docker containers, mobile devices, and IoT sensors. In these scenarios, it is difficult for users to obtain useful information from the logs, but they are still charged for the high storage costs of the logs.

Alibaba Cloud Elasticsearch allows you to use its services, such as Beats and Logstash, to connect to various data sources. This way, it can provide scalable, centralized collection capabilities and out-of-the-box storage and analysis capabilities for the data sources. The Kibana service provided by Alibaba Cloud Elasticsearch allows you to efficiently create dashboards for data visualization and O&M. You can configure items, such as hostnames, IP addresses, deployment statuses, and display colors, in the dashboards. This helps you quickly identify issues from large amounts of logs, improves troubleshooting efficiency, and makes the logs more useful.

Information searches

Searches for various information occur every day on the mobile Internet, such as searches for credit card bills, electronic invoices, nearby restaurants or hotels, news, shopping orders, transportation, and logistics. To improve search efficiency, enterprises need to implement search services for large amounts of data.

Elasticsearch has stronger full-text search capabilities than traditional relational databases. It also provides easy-to-use RESTful APIs and clients in various languages. Only a few milliseconds are required to find matching information from petabytes of structured and unstructured data. Alibaba Cloud Elasticsearch features high availability and is easy to use. It helps you quickly build search systems, such as e-commerce commodity or order search systems, application search systems, and customer relationship management (CRM) systems. These systems can be integrated into your existing business framework. This way, you can implement complex, condition-based, and fuzzy searches and easily achieve high-performance read and write operations for various types of data. The data include texts, numbers, dates, geographic data (such as IP addresses), images, audio, and videos.

Reference: Use DataWorks to synchronize data from a MySQL database to an Alibaba Cloud Elasticsearch cluster

Data intelligence

Various industries, such as gaming, education, and retail, are rapidly developing. In addition to the logs and metrics of underlying systems, large amounts of business data are generated, such as user behavior, driving trajectory, and transaction records. In the context of data-driven operations, in-depth statistical analysis and mining are required for business data. This helps identify the issues and opportunities of upper-level business, assists business decision-making, and truly makes the data valuable.

Alibaba Cloud Elasticsearch allows you to query structured data and supports complex data filtering and aggregation-based statistics. You can use Alibaba Cloud Elasticsearch to quickly and efficiently analyze various types of data, such as user behavior, attributes, and tags, to accurately identify the desired population. You can use the Kibana service provided by Alibaba Cloud Elasticsearch to collect statistics on, classify, and analyze business data and create dashboards. This way, you can extract more value from the data in various scenarios, such as e-commerce, mobile applications, and advertising media.

5.Terms

This topic describes the terms commonly used in Alibaba Cloud Elasticsearch.

cluster

An Elasticsearch cluster consists of one or more nodes. A cluster provides compound indexes and search capabilities for its nodes. All nodes in a cluster are used to store data. Each cluster has a unique name. The default cluster name is elasticsearch. Before a node joins a cluster, the name of the cluster is required.

You must make sure that clusters in different environments use different names. Otherwise, you may add nodes to the wrong cluster.

node

A node runs on a server in an Elasticsearch cluster. Nodes are used to store data and support indexing and query activities in the cluster. Same to a cluster, each node has a unique name. By default, a random UUID is assigned to a node as its name when the node is started. UUID is short for universally unique identifier. You can also assign a custom name to the node. Node names are required to complete management work. You must determine which node runs on a specific server based on the name of the node.

You can add a node to a cluster with a specified name. By default, nodes are added to the cluster named elasticsearch. Assume that these nodes can discover each other in a network. After you start these nodes, a cluster named elasticsearch is automatically created.

The number of nodes that a cluster can contain is not limited. If no Elasticsearch nodes are running in your network, after you start a node, a single-node cluster named elasticsearch is created.

index

An index is a set of documents that have similar features. It is similar to a relational database. For example, you can create three indexes to store customer data, commodity catalog data, and order data, respectively. In most cases, a name is assigned to an index to identify the index. Index names must be in lowercase. When you index, query, update, or delete a document, you must specify the name of the index to which the document belongs.

Mappings between terms in Elasticsearch and relational databases

Elasticsearch	Relational database
index	database
type	table
document	row
field	column
mapping	schema

type

A type is a logical class or partition of an index. It is similar to a table in a relational database. An index can store different types of documents, such as the user type and blog type. You are not allowed to create multiple types in an index. In later versions of Elasticsearch, this concept will be removed. For more information, see Open-source Elasticsearch documentation.

document

A document is a basic information unit that can be indexed. It is similar to a row in a table of a relational database. For example, you can create a document for a customer or commodity. A document is a JSON object. The number of documents that are stored in an index is not limited and these documents must be indexed.

field

A field is the smallest unit that makes up a document. It is similar to a column in a table of a relational database.

mapping

A mapping is used to define how a document and the fields that the document contains are stored and indexed. For example, you can use mappings to define field names, field types, and the tokenizer that you want to use. A mapping is similar to a schema in a relational database.

shard

An index can be divided into multiple shards. These shards can be distributed among different nodes to support distributed searches. When you create an index, you must specify the number of shards for the index. After the index is created, you cannot change the number.

A shard can be a primary or replica shard. In versions earlier than Elasticsearch 7.0, each index is configured with five primary shards and one replica shard for each primary shard by default. In Elasticsearch 7.0 and later, each index is configured with one primary shard and one replica shard by default. The following table describes the differences between primary and replica shards.

Shard type	Supported request type	Whether the number of shards can be changed	Remarks
Primary shard	Query and indexing requests	The number of primary shards in an index cannot be changed. This number is specified when the index is created.	Each document in an index belongs to a single primary shard. Therefore, the number and sizes of primary shards determine the maximum volume of data that an index can store. **Notice** The more primary shards, the more performance overheads your Elasticsearch cluster incurs.

Shard type	Supported request type	Whether the number of shards can be changed	Remarks
Replica shard	Query requests	The number of replica shards can be changed at any time.	Replica shards are important to search performance and provide the following benefits: • Improved fault tolerance: If a primary shard on a node is damaged or lost, you can restore the shard from replica shards. • Improved search efficiency: Elasticsearch automatically balances the load of queries among replica shards.

recovery

Data recovery (or data redistribution) is the process of redistributing shards for a node. This ensures the integrity of data when the node joins or leaves a cluster, or when the node recovers from a failure.

gateway

A gateway is used to store snapshots of indexes. By default, a node stores all the indexes in its memory. When the node memory is full, the node stores the indexes in local disks. When an Elasticsearch cluster is rebooted, its indexes are restored from the snapshots that are stored on the gateway. Restoring indexes from snapshots is faster than reading indexes from local disks. Elasticsearch supports multiple types of gateways, including the local file system (default), distributed file system, Hadoop Distributed File System (HDFS), and Alibaba Cloud Object Storage Service (OSS).

discovery.zen

discovery.zen is an automatic node discovery mechanism. Elasticsearch is a peer to peer (P2P) system that sends broadcasts to discover nodes. Nodes communicate with each other by using multicast and P2P technologies.

transport

Transport refers to the method that is used by an Elasticsearch cluster or the nodes in the cluster to communicate with clients. By default, TCP is used. You can integrate plug-ins into Elasticsearch to use other protocols, such as HTTP over JSON, Thrift, Servlet, Memcached, and ZeroMQ.

6.Performance

6.1. Overview

This topic describes a stress test performed on Alibaba Cloud Elasticsearch V5.5.3 clusters that have different specifications and reside in the China (Hangzhou) region. The test is performed by using a Rally script that is provided by open source Elasticsearch for benchmarking Elasticsearch clusters. This topic also describes the metrics and the operation parameter used in the stress test.

Overview

Rally is a stress test tool provided by open source Elasticsearch. In this example, Rally is used to perform a stress test on the Alibaba Cloud Elasticsearch clusters that have different specifications. You can view the stress test results in the following topics:

- Performance test of an Elasticsearch cluster with three 4-vCPU 16-GiB data nodes
- Performance test of an Elasticsearch cluster with three 2-vCPU 8-GiB data nodes
- Performance test of an Elasticsearch cluster with three 8-vCPU 32-GiB data nodes

The stress test result of an Elasticsearch cluster with three 4-vCPU 16-GiB data nodes and that of an Elasticsearch cluster with three 2-vCPU 8-GiB data nodes are compared. For more information, see Comparison of stress testing results between an Elasticsearch cluster with three 4-vCPU 16-GiB data nodes and an Elasticsearch cluster with three 2-vCPU 8-GiB data nodes.

You can refer to the Metrics used in the stress test and Description of the operation parameter sections in this topic to have a good command of the metrics and the operation parameter used in the stress test.

Metrics used in the stress test

Before you perform a stress test on an Elasticsearch cluster, you can refer to the following table to understand the related metrics.

> **Note** The following table describes only some important metrics for your reference. You can infer the meanings of other metrics based on the metrics described in the following table. For more information about other metrics, see the documentation for metrics for a stress test by using Rally.

Metric type	Metric name	Description

Metric type	Metric name	Description
	Cumulative indexing time of primary shards	The cumulative time used for indexing of all primary shards. **Notice** The time is not wall-clock time. It is the sum of the CPU time consumed by multiple threads used for indexing. For example, M threads are used for indexing, and each thread runs for N minutes. In this case, the time collected by this metric is calculated by using the following formula: M × N (unit: minutes).
	Min cumulative indexing time across primary shards	The minimum cumulative time used for indexing across primary shards.
	Median cumulative indexing time across primary shards	The average cumulative time used for indexing across primary shards.
	Max cumulative indexing time across primary shards	The maximum cumulative time used for indexing across primary shards.
	Cumulative indexing throttle time of primary shards	The cumulative time that indexing of all primary shards is throttled. **Notice** The time is not wall-clock time. It is the sum of the CPU time consumed by multiple threads used for indexing when indexing is throttled.
	Min cumulative indexing throttle time across primary shards	The minimum cumulative time that indexing across primary shards is throttled.
	Median cumulative indexing throttle time across primary shards	The average cumulative time that indexing across primary shards is throttled.
	Max cumulative indexing throttle time across primary shards	The maximum cumulative time that indexing across primary shards is throttled.
	Cumulative merge time of primary shards	The cumulative runtime used for merge operations for primary shards. The time also indicates the sum of the CPU time consumed by all threads.

Metric type	Metric name	Description
Metrics related to indexing of primary shards	Cumulative merge count of primary shards	The cumulative number of merges of primary shards. Notice Some primary shards may not be merged.
	Min cumulative merge time across primary shards	The minimum cumulative time used for merge operations across primary shards.
	Median cumulative merge time across primary shards	The average cumulative time used for merge operations across primary shards.
	Max cumulative merge time across primary shards	The maximum cumulative time used for merge operations across primary shards.
	Cumulative merge throttle time of primary shards	The cumulative time that merge operations for primary shards are throttled. The time also indicates the sum of the CPU time consumed by all threads.
	Min cumulative merge throttle time across primary shards	The minimum cumulative time that merge operations across primary shards are throttled. The time also indicates the sum of the CPU time consumed by all threads.
	Median cumulative merge throttle time across primary shards	The average cumulative time that merge operations across primary shards are throttled. The time also indicates the sum of the CPU time consumed by all threads.
	Max cumulative merge throttle time across primary shards	The maximum cumulative time that merge operations across primary shards are throttled. The time also indicates the sum of the CPU time consumed by all threads.
	Cumulative refresh time of primary shards	The cumulative time used for index refresh of primary shards. The time also indicates the CPU time consumed by all threads.
	Cumulative refresh count of primary shards	The cumulative number of refreshes of primary shards.
	Min cumulative refresh time across primary shards	The minimum cumulative time used for index refresh across primary shards.
	Median cumulative refresh time across primary shards	The average cumulative time used for index refresh across primary shards.

Metric type	Metric name	Description
	Max cumulative refresh time across primary shards	The maximum cumulative time used for index refresh across primary shards.
	Cumulative flush time of primary shards	The cumulative time used for flushing transactional data of indexing of primary shards from the cache to a disk. The time also indicates the sum of the CPU time consumed by all threads.
	Cumulative flush count of primary shards	The cumulative number of flushes for transactional data of indexing of primary shards from the cache to a disk.
	Min cumulative flush time across primary shards	The minimum cumulative time used for flushing transactional data of indexing across primary shards from the cache to a disk. The time also indicates the sum of the CPU time consumed by all threads.
	Median cumulative flush time across primary shards	The average cumulative time used for flushing transactional data of indexing across primary shards from the cache to a disk. The time also indicates the sum of the CPU time consumed by all threads.
	Max cumulative flush time across primary shards	The maximum cumulative time used for flushing transactional data of indexing across primary shards from the cache to a disk. The time also indicates the sum of the CPU time consumed by all threads.
	Store size	The size of data stored in indexes. The size does not include the size of translogs and that of data stored in replica shards.
	Translog size	The size of translogs.
	Heap used for segments	The size of heap memory occupied by the segments of all primary shards.
	Heap used for doc values	The size of heap memory occupied by the documents in indexes of all primary shards.
	Heap used for terms	The size of heap memory occupied by terms factors of indexes of all primary shards.
	Heap used for norms	The size of heap memory occupied by norms factors of indexes of all primary shards.
	Heap used for points	The size of heap memory occupied by points of indexes of all primary shards.
	Heap used for stored fields	The size of heap memory occupied by fields in indexes of all primary shards.

Metric type	Metric name	Description
	Segment count	The number of segments of indexes of all primary shards.
Metrics related to garbage collectors	Total Young Gen GC	The total runtime of the young-generation garbage collector in the entire cluster.
	Total Old Gen GC	The total runtime of the old-generation garbage collector in the entire cluster.
Metrics related to throughput	Min Throughput	The minimum queries per second (QPS) for each task.
	Median Throughput	The average QPS for each task.
	Max Throughput	The maximum QPS for each task.
Metrics related to latency	50th percentile latency	The maximum latency for the fastest 50% of all requests.
	90th percentile latency	The maximum latency for the fastest 90% of all requests.
	99.9th percentile latency	The maximum latency for the fastest 99.9% of all requests.
	100th percentile latency	The maximum latency for all requests.
Metrics related to service time	50th percentile service time	The service time for the fastest 50% of all requests.
	90th percentile service time	The service time for the fastest 90% of all requests.
	99.9th percentile service time	The service time for the fastest 99.9% of all requests.
	100th percentile service time	The service time for all requests.
Metrics related to error rates	error rate	The rate of responses that contain errors to all responses.

Note

- The latency indicates the period of time from the point in time when a request is submitted to the point in time when a complete response is received. The latency includes the waiting period before Elasticsearch starts to process the request.
- The service time indicates the period of time from the point in time when a request starts to be processed to the point in time when a response is received.
- The error rate indicates the rate of responses that contain errors to all responses.

Description of the operation parameter

You can refer to the values of the operation parameter that are listed in the following table to analyze data collected based on metrics such as throughput, latency, service time, and error rate.

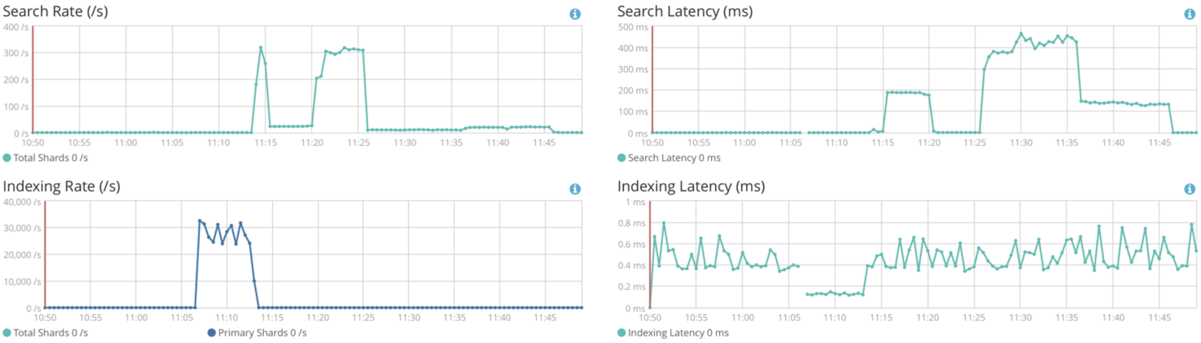
Value	Description
index-append	The index creation operation.
index-stats	The status of an index.
node-stats	The status of a node.
default	The default dimension.
term	The term query.
phrase	The exact queries for phrases.
country_agg_uncached	The aggregate operation that is not cached.
country_agg_cached	The aggregate operation that is cached.
scroll	The scroll operation.
expression	The expression.
painless_static	The static script.
painless_dynamic	The dynamic script.
large_terms	The combination of multiple term queries.
large_filtered_terms	The combination of multiple filtered term queries.
large_prohibited_terms	The combination of multiple prohibited term queries.

6.2. Performance test of an Elasticsearch cluster with three 2-vCPU 8-GiB data nodes

This topic lists the performance metrics of an Elasticsearch cluster that contains three data nodes. Each data node has 2 vCPUs and 8 GiB of memory. The metrics include the Kibana metrics during the performance test and the performance metrics that are used to calculate these Kibana metrics.

Note

Kibana metrics during the performance test



Performance metrics

Metric	Operation	Value	Unit
Indexing time	None	23.9479	min
Merge time	None	14.3001	min
Refresh time	None	5.26405	min
Flush time	None	0.0308333	min
Merge throttle time	None	1.27945	min
Total Young Gen GC	None	183.74	s
Total Old Gen GC	None	1.125	s
Heap used for segments	None	18.8167	MB
Heap used for doc values	None	0.452751	MB
Heap used for terms	None	17.2004	MB
Heap used for norms	None	0.0852051	MB
Heap used for points	None	0.241465	MB
Heap used for stored fields	None	0.836876	MB
Segment count	None	140	items
Min Throughput	index-append	28115.4	docs/s
Median Throughput	index-append	28645.5	docs/s
Max Throughput	index-append	30037.8	docs/s
50th percentile latency	index-append	1447.76	ms

Metric	Operation	Value	Unit
90th percentile latency	index-append	1847.05	ms
99th percentile latency	index-append	2264.68	ms
99.9th percentile latency	index-append	2515.95	ms
100th percentile latency	index-append	2608.68	ms
50th percentile service time	index-append	1447.76	ms
90th percentile service time	index-append	1847.05	ms
99th percentile service time	index-append	2264.68	ms
99.9th percentile service time	index-append	2515.95	ms
100th percentile service time	index-append	2608.68	ms
error rate	index-append	0	%
Min Throughput	force-merge	2.1	ops/s
Median Throughput	force-merge	2.1	ops/s
Max Throughput	force-merge	2.1	ops/s
100th percentile latency	force-merge	475.984	ms
100th percentile service time	force-merge	475.984	ms
error rate	force-merge	0	%
Min Throughput	index-stats	97.75	ops/s
Median Throughput	index-stats	100.05	ops/s
Max Throughput	index-stats	100.07	ops/s
50th percentile latency	index-stats	5.09015	ms
90th percentile latency	index-stats	10.7365	ms
99th percentile latency	index-stats	234.761	ms

Metric	Operation	Value	Unit
99.9th percentile latency	index-stats	277.393	ms
100th percentile latency	index-stats	281.866	ms
50th percentile service time	index-stats	5.01096	ms
90th percentile service time	index-stats	5.30021	ms
99th percentile service time	index-stats	12.0005	ms
99.9th percentile service time	index-stats	141.631	ms
100th percentile service time	index-stats	150.153	ms
error rate	index-stats	0	%
Min Throughput	node-stats	100.01	ops/s
Median Throughput	node-stats	100.08	ops/s
Max Throughput	node-stats	100.49	ops/s
50th percentile latency	node-stats	4.90659	ms
90th percentile latency	node-stats	5.29285	ms
99th percentile latency	node-stats	29.3245	ms
99.9th percentile latency	node-stats	43.3885	ms
100th percentile latency	node-stats	44.6019	ms
50th percentile service time	node-stats	4.83552	ms
90th percentile service time	node-stats	5.12694	ms
99th percentile service time	node-stats	9.08739	ms
99.9th percentile service time	node-stats	39.744	ms

Metric	Operation	Value	Unit
100th percentile service time	node-stats	44.5383	ms
error rate	node-stats	0	%
Min Throughput	default	47.83	ops/s
Median Throughput	default	48.28	ops/s
Max Throughput	default	48.73	ops/s
50th percentile latency	default	617.465	ms
90th percentile latency	default	1033.98	ms
99th percentile latency	default	1083.23	ms
99.9th percentile latency	default	1095.4	ms
100th percentile latency	default	1097.14	ms
50th percentile service time	default	18.646	ms
90th percentile service time	default	24.9381	ms
99th percentile service time	default	35.7667	ms
99.9th percentile service time	default	57.3679	ms
100th percentile service time	default	151.505	ms
error rate	default	0	%
Min Throughput	term	199.43	ops/s
Median Throughput	term	200.07	ops/s
Max Throughput	term	200.13	ops/s
50th percentile latency	term	2.9728	ms
90th percentile latency	term	7.10648	ms
99th percentile latency	term	22.4487	ms

Metric	Operation	Value	Unit
99.9th percentile latency	term	29.0737	ms
100th percentile latency	term	29.6253	ms
50th percentile service time	term	2.87833	ms
90th percentile service time	term	3.08983	ms
99th percentile service time	term	19.9777	ms
99.9th percentile service time	term	29.0082	ms
100th percentile service time	term	29.5597	ms
error rate	term	0	%
Min Throughput	phrase	199.71	ops/s
Median Throughput	phrase	200.04	ops/s
Max Throughput	phrase	200.07	ops/s
50th percentile latency	phrase	3.61484	ms
90th percentile latency	phrase	16.5523	ms
99th percentile latency	phrase	31.394	ms
99.9th percentile latency	phrase	33.902	ms
100th percentile latency	phrase	34.5784	ms
50th percentile service time	phrase	3.47402	ms
90th percentile service time	phrase	3.90958	ms
99th percentile service time	phrase	19.3773	ms
99.9th percentile service time	phrase	22.7947	ms

Metric	Operation	Value	Unit
100th percentile service time	phrase	27.8164	ms
error rate	phrase	0	%
Min Throughput	country_agg_uncached	4.63	ops/s
Median Throughput	country_agg_uncached	4.65	ops/s
Max Throughput	country_agg_uncached	4.67	ops/s
50th percentile latency	country_agg_uncached	14864.3	ms
90th percentile latency	country_agg_uncached	21046	ms
99th percentile latency	country_agg_uncached	22902	ms
99.9th percentile latency	country_agg_uncached	22997.6	ms
100th percentile latency	country_agg_uncached	23018.7	ms
50th percentile service time	country_agg_uncached	204.174	ms
90th percentile service time	country_agg_uncached	242.492	ms
99th percentile service time	country_agg_uncached	345.382	ms
99.9th percentile service time	country_agg_uncached	378.302	ms
100th percentile service time	country_agg_uncached	422.53	ms
error rate	country_agg_uncached	0	%
Min Throughput	country_agg_cached	98.37	ops/s
Median Throughput	country_agg_cached	100.06	ops/s
Max Throughput	country_agg_cached	100.13	ops/s
50th percentile latency	country_agg_cached	3.2638	ms
90th percentile latency	country_agg_cached	4.69259	ms
99th percentile latency	country_agg_cached	189.143	ms

Metric	Operation	Value	Unit
99.9th percentile latency	country_agg_cached	249.851	ms
100th percentile latency	country_agg_cached	256.028	ms
50th percentile service time	country_agg_cached	3.18679	ms
90th percentile service time	country_agg_cached	3.42086	ms
99th percentile service time	country_agg_cached	20.4171	ms
99.9th percentile service time	country_agg_cached	117.273	ms
100th percentile service time	country_agg_cached	255.951	ms
error rate	country_agg_cached	0	%
Min Throughput	scroll	59.16	ops/s
Median Throughput	scroll	60.44	ops/s
Max Throughput	scroll	61.02	ops/s
50th percentile latency	scroll	168347	ms
90th percentile latency	scroll	240658	ms
99th percentile latency	scroll	257048	ms
100th percentile latency	scroll	258853	ms
50th percentile service time	scroll	402.962	ms
90th percentile service time	scroll	431.267	ms
99th percentile service time	scroll	455.632	ms
100th percentile service time	scroll	601.214	ms
error rate	scroll	0	%
Min Throughput	expression	2	ops/s

Metric	Operation	Value	Unit
Median Throughput	expression	2	ops/s
Max Throughput	expression	2	ops/s
50th percentile latency	expression	409.417	ms
90th percentile latency	expression	434.858	ms
99th percentile latency	expression	501.498	ms
100th percentile latency	expression	517.438	ms
50th percentile service time	expression	409.165	ms
90th percentile service time	expression	434.749	ms
99th percentile service time	expression	498.681	ms
100th percentile service time	expression	517.332	ms
error rate	expression	0	%
Min Throughput	painless_static	1.96	ops/s
Median Throughput	painless_static	1.97	ops/s
Max Throughput	painless_static	1.97	ops/s
50th percentile latency	painless_static	3163.94	ms
90th percentile latency	painless_static	3679.27	ms
99th percentile latency	painless_static	3994.52	ms
100th percentile latency	painless_static	4006.5	ms
50th percentile service time	painless_static	503.588	ms
90th percentile service time	painless_static	528.807	ms
99th percentile service time	painless_static	600.103	ms
100th percentile service time	painless_static	623.666	ms

Metric	Operation	Value	Unit
error rate	painless_static	0	%
Min Throughput	painless_dynamic	2	ops/s
Median Throughput	painless_dynamic	2	ops/s
Max Throughput	painless_dynamic	2	ops/s
50th percentile latency	painless_dynamic	611.305	ms
90th percentile latency	painless_dynamic	786.806	ms
99th percentile latency	painless_dynamic	973.432	ms
100th percentile latency	painless_dynamic	982.484	ms
50th percentile service time	painless_dynamic	494.097	ms
90th percentile service time	painless_dynamic	518.082	ms
99th percentile service time	painless_dynamic	606.748	ms
100th percentile service time	painless_dynamic	638.903	ms
error rate	painless_dynamic	0	%
Min Throughput	large_filtered_terms	1.39	ops/s
Median Throughput	large_filtered_terms	1.4	ops/s
Max Throughput	large_filtered_terms	1.4	ops/s
50th percentile latency	large_filtered_terms	65601.1	ms
90th percentile latency	large_filtered_terms	82494.7	ms
99th percentile latency	large_filtered_terms	86452.2	ms
100th percentile latency	large_filtered_terms	86857.3	ms
50th percentile service time	large_filtered_terms	707.17	ms
90th percentile service time	large_filtered_terms	747.949	ms
99th percentile service time	large_filtered_terms	847.069	ms

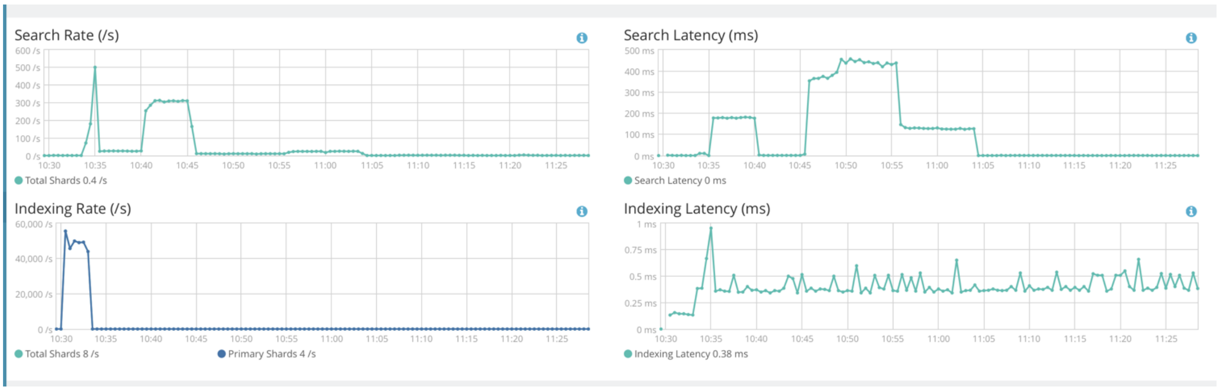
Metric	Operation	Value	Unit
100th percentile service time	large_filtered_terms	927.917	ms
error rate	large_filtered_terms	0	%
Min Throughput	large_prohibited_terms	1.46	ops/s
Median Throughput	large_prohibited_terms	1.46	ops/s
Max Throughput	large_prohibited_terms	1.46	ops/s
50th percentile latency	large_prohibited_terms	55916.3	ms
90th percentile latency	large_prohibited_terms	70529.7	ms
99th percentile latency	large_prohibited_terms	73769.1	ms
100th percentile latency	large_prohibited_terms	74143.9	ms
50th percentile service time	large_prohibited_terms	679.394	ms
90th percentile service time	large_prohibited_terms	717.476	ms
99th percentile service time	large_prohibited_terms	782.085	ms
100th percentile service time	large_prohibited_terms	822.723	ms
error rate	large_prohibited_terms	0	%

6.3. Performance test of an Elasticsearch cluster with three 4-vCPU 16-GiB data nodes

This topic lists the performance metrics of an Elasticsearch cluster that contains three data nodes. Each data node has 4 vCPUs and 16 GiB of memory. The metrics include the Kibana metrics during the performance test and the performance metrics that are used to calculate these Kibana metrics.

> **Note** The official GeoNames database is used for the test. The database contains 11,520,617 DOC files whose total size is 3.3 GB.

Kibana metrics during the performance test



Performance metrics

Metric	Operation	Value	Unit
Indexing time	None	26.3543	min
Merge time	None	11.0297	min
Refresh time	None	3.05238	min
Flush time	None	0.04485	min
Merge throttle time	None	1.39282	min
Total Young Gen GC	None	92.902	s
Total Old Gen GC	None	0.4	s
Heap used for segments	None	18.7955	MB
Heap used for doc values	None	0.360752	MB
Heap used for terms	None	17.2739	MB
Heap used for norms	None	0.0877075	MB
Heap used for points	None	0.241213	MB
Heap used for stored fields	None	0.831932	MB
Segment count	None	133	items
Min Throughput	index-append	51751.7	docs/s
Median Throughput	index-append	52303	docs/s
Max Throughput	index-append	54076.3	docs/s
50th percentile latency	index-append	743.939	ms

Metric	Operation	Value	Unit
90th percentile latency	index-append	1045.7	ms
99th percentile latency	index-append	1325.21	ms
100th percentile latency	index-append	1794.39	ms
50th percentile service time	index-append	743.939	ms
90th percentile service time	index-append	1045.7	ms
99th percentile service time	index-append	1325.21	ms
100th percentile service time	index-append	1794.39	ms
error rate	index-append	0	%
Min Throughput	force-merge	0.95	ops/s
Median Throughput	force-merge	0.95	ops/s
Max Throughput	force-merge	0.95	ops/s
100th percentile latency	force-merge	1052.54	ms
100th percentile service time	force-merge	1052.54	ms
error rate	force-merge	0	%
Min Throughput	index-stats	100.04	ops/s
Median Throughput	index-stats	100.05	ops/s
Max Throughput	index-stats	100.09	ops/s
50th percentile latency	index-stats	4.85232	ms
90th percentile latency	index-stats	5.14185	ms
99th percentile latency	index-stats	77.3127	ms
99.9th percentile latency	index-stats	123.888	ms
100th percentile latency	index-stats	128.01	ms
50th percentile service time	index-stats	4.78006	ms

Metric	Operation	Value	Unit
90th percentile service time	index-stats	4.9831	ms
99th percentile service time	index-stats	9.66475	ms
99.9th percentile service time	index-stats	48.4445	ms
100th percentile service time	index-stats	127.945	ms
error rate	index-stats	0	%
Min Throughput	node-stats	100.05	ops/s
Median Throughput	node-stats	100.1	ops/s
Max Throughput	node-stats	100.55	ops/s
50th percentile latency	node-stats	4.55259	ms
90th percentile latency	node-stats	4.78784	ms
99th percentile latency	node-stats	18.8034	ms
99.9th percentile latency	node-stats	43.7684	ms
100th percentile latency	node-stats	48.1474	ms
50th percentile service time	node-stats	4.48138	ms
90th percentile service time	node-stats	4.69386	ms
99th percentile service time	node-stats	5.64618	ms
99.9th percentile service time	node-stats	27.8155	ms
100th percentile service time	node-stats	43.6905	ms
error rate	node-stats	0	%
Min Throughput	default	49.81	ops/s
Median Throughput	default	50	ops/s

Metric	Operation	Value	Unit
Max Throughput	default	50	ops/s
50th percentile latency	default	19.7245	ms
90th percentile latency	default	94.1457	ms
99th percentile latency	default	133.091	ms
99.9th percentile latency	default	137.285	ms
100th percentile latency	default	138.043	ms
50th percentile service time	default	19.1469	ms
90th percentile service time	default	19.9554	ms
99th percentile service time	default	25.3462	ms
99.9th percentile service time	default	54.7931	ms
100th percentile service time	default	133.771	ms
error rate	default	0	%
Min Throughput	term	200.05	ops/s
Median Throughput	term	200.08	ops/s
Max Throughput	term	200.12	ops/s
50th percentile latency	term	3.07948	ms
90th percentile latency	term	3.37296	ms
99th percentile latency	term	22.3272	ms
99.9th percentile latency	term	26.9648	ms
100th percentile latency	term	28.1562	ms
50th percentile service time	term	3.00599	ms
90th percentile service time	term	3.15279	ms

Metric	Operation	Value	Unit
99th percentile service time	term	4.22302	ms
99.9th percentile service time	term	26.9017	ms
100th percentile service time	term	28.0823	ms
error rate	term	0	%
Min Throughput	phrase	199.84	ops/s
Median Throughput	phrase	200.04	ops/s
Max Throughput	phrase	200.09	ops/s
50th percentile latency	phrase	3.76927	ms
90th percentile latency	phrase	13.6055	ms
99th percentile latency	phrase	28.0245	ms
99.9th percentile latency	phrase	34.7198	ms
100th percentile latency	phrase	35.551	ms
50th percentile service time	phrase	3.67227	ms
90th percentile service time	phrase	4.08037	ms
99th percentile service time	phrase	16.9256	ms
99.9th percentile service time	phrase	24.4886	ms
100th percentile service time	phrase	29.8604	ms
error rate	phrase	0	%
Min Throughput	country_agg_uncached	4.95	ops/s
Median Throughput	country_agg_uncached	4.99	ops/s
Max Throughput	country_agg_uncached	5	ops/s
50th percentile latency	country_agg_uncached	330.923	ms

Metric	Operation	Value	Unit
90th percentile latency	country_agg_uncached	2780.17	ms
99th percentile latency	country_agg_uncached	2866	ms
99.9th percentile latency	country_agg_uncached	2880.39	ms
100th percentile latency	country_agg_uncached	2882.11	ms
50th percentile service time	country_agg_uncached	197.883	ms
90th percentile service time	country_agg_uncached	213.402	ms
99th percentile service time	country_agg_uncached	256.649	ms
99.9th percentile service time	country_agg_uncached	290.496	ms
100th percentile service time	country_agg_uncached	296.875	ms
error rate	country_agg_uncached	0	%
Min Throughput	country_agg_cached	99.92	ops/s
Median Throughput	country_agg_cached	100.06	ops/s
Max Throughput	country_agg_cached	100.11	ops/s
50th percentile latency	country_agg_cached	3.30479	ms
90th percentile latency	country_agg_cached	3.52514	ms
99th percentile latency	country_agg_cached	52.8258	ms
99.9th percentile latency	country_agg_cached	112.895	ms
100th percentile latency	country_agg_cached	119.435	ms
50th percentile service time	country_agg_cached	3.23149	ms
90th percentile service time	country_agg_cached	3.41319	ms
99th percentile service time	country_agg_cached	7.60955	ms

Metric	Operation	Value	Unit
99.9th percentile service time	country_agg_cached	26.2229	ms
100th percentile service time	country_agg_cached	119.365	ms
error rate	country_agg_cached	0	%
Min Throughput	scroll	61.59	ops/s
Median Throughput	scroll	61.67	ops/s
Max Throughput	scroll	61.94	ops/s
50th percentile latency	scroll	164549	ms
90th percentile latency	scroll	237443	ms
99th percentile latency	scroll	253860	ms
100th percentile latency	scroll	255710	ms
50th percentile service time	scroll	399.964	ms
90th percentile service time	scroll	424.303	ms
99th percentile service time	scroll	523.877	ms
100th percentile service time	scroll	639.45	ms
error rate	scroll	0	%
Min Throughput	expression	2	ops/s
Median Throughput	expression	2	ops/s
Max Throughput	expression	2	ops/s
50th percentile latency	expression	409.927	ms
90th percentile latency	expression	434.544	ms
99th percentile latency	expression	532.412	ms
100th percentile latency	expression	537.618	ms
50th percentile service time	expression	409.812	ms

Metric	Operation	Value	Unit
90th percentile service time	expression	428.156	ms
99th percentile service time	expression	532.33	ms
100th percentile service time	expression	537.495	ms
error rate	expression	0	%
Min Throughput	painless_static	2	ops/s
Median Throughput	painless_static	2	ops/s
Max Throughput	painless_static	2	ops/s
50th percentile latency	painless_static	497.626	ms
90th percentile latency	painless_static	643.32	ms
99th percentile latency	painless_static	700.559	ms
100th percentile latency	painless_static	704.679	ms
50th percentile service time	painless_static	490.705	ms
90th percentile service time	painless_static	500.663	ms
99th percentile service time	painless_static	642.124	ms
100th percentile service time	painless_static	683.621	ms
error rate	painless_static	0	%
Min Throughput	painless_dynamic	2	ops/s
Median Throughput	painless_dynamic	2	ops/s
Max Throughput	painless_dynamic	2	ops/s
50th percentile latency	painless_dynamic	473.087	ms
90th percentile latency	painless_dynamic	554.729	ms
99th percentile latency	painless_dynamic	668.363	ms
100th percentile latency	painless_dynamic	706.557	ms

Metric	Operation	Value	Unit
50th percentile service time	painless_dynamic	469.145	ms
90th percentile service time	painless_dynamic	501.774	ms
99th percentile service time	painless_dynamic	606.61	ms
100th percentile service time	painless_dynamic	624.751	ms
error rate	painless_dynamic	0	%
Min Throughput	large_filtered_terms	1.64	ops/s
Median Throughput	large_filtered_terms	1.64	ops/s
Max Throughput	large_filtered_terms	1.65	ops/s
50th percentile latency	large_filtered_terms	33013.5	ms
90th percentile latency	large_filtered_terms	40869	ms
99th percentile latency	large_filtered_terms	42644	ms
100th percentile latency	large_filtered_terms	42936.2	ms
50th percentile service time	large_filtered_terms	598.001	ms
90th percentile service time	large_filtered_terms	626.81	ms
99th percentile service time	large_filtered_terms	771.815	ms
100th percentile service time	large_filtered_terms	796.884	ms
error rate	large_filtered_terms	0	%
Min Throughput	large_prohibited_terms	1.69	ops/s
Median Throughput	large_prohibited_terms	1.69	ops/s
Max Throughput	large_prohibited_terms	1.7	ops/s
50th percentile latency	large_prohibited_terms	27732.3	ms
90th percentile latency	large_prohibited_terms	34305.5	ms

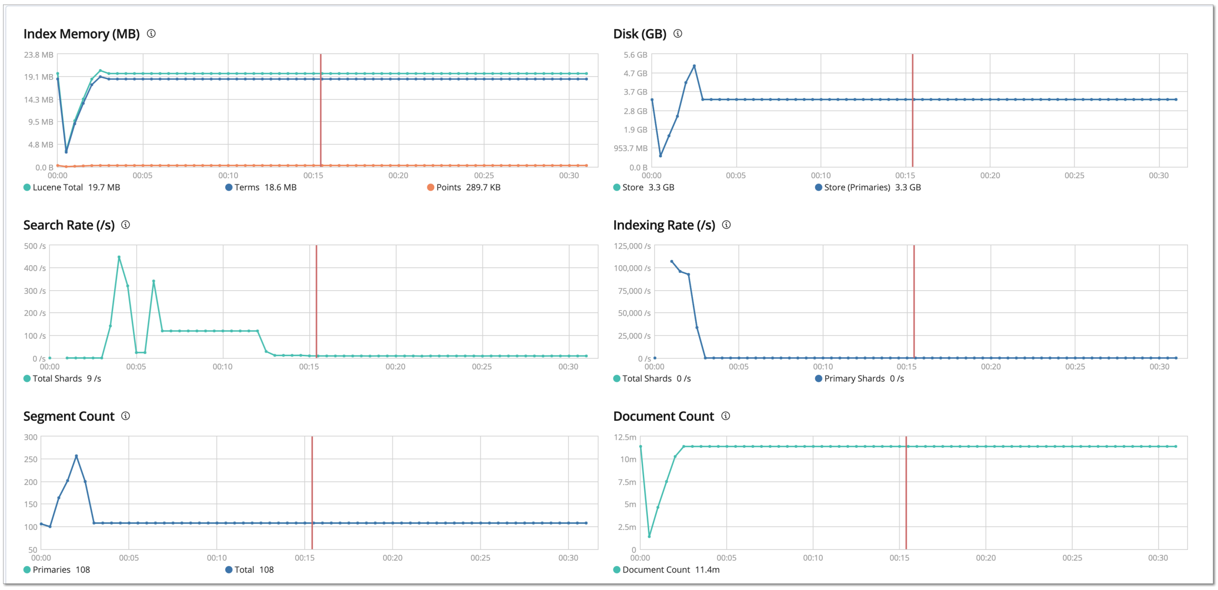
Metric	Operation	Value	Unit
99th percentile latency	large_prohibited_terms	35840.4	ms
100th percentile latency	large_prohibited_terms	35993.5	ms
50th percentile service time	large_prohibited_terms	586.382	ms
90th percentile service time	large_prohibited_terms	618.185	ms
99th percentile service time	large_prohibited_terms	661.378	ms
100th percentile service time	large_prohibited_terms	823.782	ms
error rate	large_prohibited_terms	0	%

6.4. Performance test of an Elasticsearch cluster with three 8-vCPU 32-GiB data nodes

This topic lists the performance metrics of an Elasticsearch cluster that contains three data nodes. Each data node has 8 vCPUs and 32 GiB of memory. The metrics include the Kibana metrics during the performance test and the performance metrics that are used to calculate these Kibana metrics.

Note

Kibana metrics during the performance test



Performance metrics

Metric	Operation	Value	Unit
Cumulative indexing time of primary shards	None	42.3229	min
Min cumulative indexing time across primary shards	None	0.000133333	min
Median cumulative indexing time across primary shards	None	6.85448	min
Max cumulative indexing time across primary shards	None	7.15663	min
Cumulative indexing throttle time of primary shards	None	0	min
Min cumulative indexing throttle time across primary shards	None	0	min
Median cumulative indexing throttle time across primary shards	None	0	min
Max cumulative indexing throttle time across primary shards	None	0	min
Cumulative merge time of primary shards	None	14.212	min
Cumulative merge count of primary shards	None	897	items
Min cumulative merge time across primary shards	None	0	min
Median cumulative merge time across primary shards	None	2.0934	min
Max cumulative merge time across primary shards	None	2.1898	min

Metric	Operation	Value	Unit
Cumulative merge throttle time of primary shards	None	1.16502	min
Min cumulative merge throttle time across primary shards	None	0	min
Median cumulative merge throttle time across primary shards	None	0.157883	min
Max cumulative merge throttle time across primary shards	None	0.238017	min
Cumulative refresh time of primary shards	None	3.59033	min
Cumulative refresh count of primary shards	None	6963	items
Min cumulative refresh time across primary shards	None	0.000366667	min
Median cumulative refresh time across primary shards	None	0.401783	min
Max cumulative refresh time across primary shards	None	0.8976	min
Cumulative flush time of primary shards	None	0.514533	min
Cumulative flush count of primary shards	None	17	items
Min cumulative flush time across primary shards	None	0	min
Median cumulative flush time across primary shards	None	0.0558833	min
Max cumulative flush time across primary shards	None	0.102767	min
Total Young Gen GC	None	2.867	s

Metric	Operation	Value	Unit
Total Old Gen GC	None	0	s
Store size	None	3.36733	GB
Translog size	None	3.40745	GB
Heap used for segments	None	20.2287	MB
Heap used for doc values	None	0.102116	MB
Heap used for terms	None	18.976	MB
Heap used for norms	None	0.0888672	MB
Heap used for points	None	0.286792	MB
Heap used for stored fields	None	0.774918	MB
Segment count	None	135	items
Min Throughput	index-append	92809.2	docs/s
Median Throughput	index-append	92809.2	docs/s
Max Throughput	index-append	92809.2	docs/s
50th percentile latency	index-append	1092.82	ms
90th percentile latency	index-append	1482.98	ms
100th percentile latency	index-append	1655.12	ms
50th percentile service time	index-append	1092.82	ms
90th percentile service time	index-append	1482.98	ms
100th percentile service time	index-append	1655.12	ms
error rate	index-append	0	%
Min Throughput	index-stats	90.03	ops/s
Median Throughput	index-stats	90.04	ops/s
Max Throughput	index-stats	90.07	ops/s

Metric	Operation	Value	Unit
50th percentile latency	index-stats	6.66079	ms
90th percentile latency	index-stats	7.18313	ms
99th percentile latency	index-stats	8.61447	ms
99.9th percentile latency	index-stats	14.5834	ms
100th percentile latency	index-stats	14.7291	ms
50th percentile service time	index-stats	6.58463	ms
90th percentile service time	index-stats	7.08993	ms
99th percentile service time	index-stats	8.03104	ms
99.9th percentile service time	index-stats	14.5087	ms
100th percentile service time	index-stats	14.6548	ms
error rate	index-stats	0	%
Min Throughput	node-stats	90.03	ops/s
Median Throughput	node-stats	90.05	ops/s
Max Throughput	node-stats	90.18	ops/s
50th percentile latency	node-stats	6.75975	ms
90th percentile latency	node-stats	7.45763	ms
99th percentile latency	node-stats	9.49112	ms
99.9th percentile latency	node-stats	21.5716	ms
100th percentile latency	node-stats	23.0975	ms
50th percentile service time	node-stats	6.68247	ms
90th percentile service time	node-stats	7.34734	ms

Metric	Operation	Value	Unit
99th percentile service time	node-stats	8.80335	ms
99.9th percentile service time	node-stats	19.4196	ms
100th percentile service time	node-stats	23.0214	ms
error rate	node-stats	0	%
Min Throughput	default	50	ops/s
Median Throughput	default	50.01	ops/s
Max Throughput	default	50.03	ops/s
50th percentile latency	default	13.4787	ms
90th percentile latency	default	16.4997	ms
99th percentile latency	default	21.6313	ms
99.9th percentile latency	default	27.411	ms
100th percentile latency	default	27.7961	ms
50th percentile service time	default	13.3791	ms
90th percentile service time	default	15.4398	ms
99th percentile service time	default	21.5527	ms
99.9th percentile service time	default	27.3316	ms
100th percentile service time	default	27.7195	ms
error rate	default	0	%
Min Throughput	term	200.02	ops/s
Median Throughput	term	200.03	ops/s
Max Throughput	term	200.03	ops/s
50th percentile latency	term	4.39089	ms

Metric	Operation	Value	Unit
90th percentile latency	term	4.51443	ms
99th percentile latency	term	5.61225	ms
99.9th percentile latency	term	10.4551	ms
100th percentile latency	term	10.6584	ms
50th percentile service time	term	4.32217	ms
90th percentile service time	term	4.43736	ms
99th percentile service time	term	4.61988	ms
99.9th percentile service time	term	5.42257	ms
100th percentile service time	term	5.83687	ms
error rate	term	0	%
Min Throughput	phrase	176.79	ops/s
Median Throughput	phrase	179.33	ops/s
Max Throughput	phrase	180.82	ops/s
50th percentile latency	phrase	581.347	ms
90th percentile latency	phrase	754.845	ms
99th percentile latency	phrase	791.243	ms
99.9th percentile latency	phrase	794.407	ms
100th percentile latency	phrase	794.55	ms
50th percentile service time	phrase	5.38911	ms
90th percentile service time	phrase	5.61755	ms
99th percentile service time	phrase	6.07193	ms

Metric	Operation	Value	Unit
99.9th percentile service time	phrase	10.8131	ms
100th percentile service time	phrase	11.2903	ms
error rate	phrase	0	%
Min Throughput	country_agg_uncached	4	ops/s
Median Throughput	country_agg_uncached	4.01	ops/s
Max Throughput	country_agg_uncached	4.01	ops/s
50th percentile latency	country_agg_uncached	145.729	ms
90th percentile latency	country_agg_uncached	157.922	ms
99th percentile latency	country_agg_uncached	169.1	ms
100th percentile latency	country_agg_uncached	171.331	ms
50th percentile service time	country_agg_uncached	145.573	ms
90th percentile service time	country_agg_uncached	157.744	ms
99th percentile service time	country_agg_uncached	168.924	ms
100th percentile service time	country_agg_uncached	171.149	ms
error rate	country_agg_uncached	0	%
Min Throughput	country_agg_cached	100.02	ops/s
Median Throughput	country_agg_cached	100.05	ops/s
Max Throughput	country_agg_cached	100.09	ops/s
50th percentile latency	country_agg_cached	4.72138	ms
90th percentile latency	country_agg_cached	4.97254	ms
99th percentile latency	country_agg_cached	5.36374	ms
99.9th percentile latency	country_agg_cached	18.7445	ms
100th percentile latency	country_agg_cached	21.2341	ms

Metric	Operation	Value	Unit
50th percentile service time	country_agg_cached	4.64525	ms
90th percentile service time	country_agg_cached	4.89226	ms
99th percentile service time	country_agg_cached	5.18021	ms
99.9th percentile service time	country_agg_cached	14.9384	ms
100th percentile service time	country_agg_cached	21.1512	ms
error rate	country_agg_cached	0	%
Min Throughput	scroll	20.04	pages/s
Median Throughput	scroll	20.05	pages/s
Max Throughput	scroll	20.06	pages/s
50th percentile latency	scroll	514.808	ms
90th percentile latency	scroll	550.392	ms
99th percentile latency	scroll	576.239	ms
100th percentile latency	scroll	584.526	ms
50th percentile service time	scroll	514.028	ms
90th percentile service time	scroll	549.609	ms
99th percentile service time	scroll	575.47	ms
100th percentile service time	scroll	583.767	ms
error rate	scroll	0	%
Min Throughput	expression	2	ops/s
Median Throughput	expression	2	ops/s
Max Throughput	expression	2	ops/s
50th percentile latency	expression	460.795	ms

Metric	Operation	Value	Unit
90th percentile latency	expression	482.177	ms
99th percentile latency	expression	491.312	ms
100th percentile latency	expression	493.126	ms
50th percentile service time	expression	460.65	ms
90th percentile service time	expression	481.979	ms
99th percentile service time	expression	491.119	ms
100th percentile service time	expression	492.936	ms
error rate	expression	0	%
Min Throughput	painless_static	1.5	ops/s
Median Throughput	painless_static	1.5	ops/s
Max Throughput	painless_static	1.5	ops/s
50th percentile latency	painless_static	493.747	ms
90th percentile latency	painless_static	513.06	ms
99th percentile latency	painless_static	601.632	ms
100th percentile latency	painless_static	678.591	ms
50th percentile service time	painless_static	493.504	ms
90th percentile service time	painless_static	512.17	ms
99th percentile service time	painless_static	601.373	ms
100th percentile service time	painless_static	678.352	ms
error rate	painless_static	0	%
Min Throughput	painless_dynamic	1.5	ops/s
Median Throughput	painless_dynamic	1.5	ops/s

Metric	Operation	Value	Unit
Max Throughput	painless_dynamic	1.5	ops/s
50th percentile latency	painless_dynamic	492.847	ms
90th percentile latency	painless_dynamic	507.818	ms
99th percentile latency	painless_dynamic	528.414	ms
100th percentile latency	painless_dynamic	529.858	ms
50th percentile service time	painless_dynamic	492.553	ms
90th percentile service time	painless_dynamic	507.568	ms
99th percentile service time	painless_dynamic	528.154	ms
100th percentile service time	painless_dynamic	529.592	ms
error rate	painless_dynamic	0	%
Min Throughput	large_terms	1.5	ops/s
Median Throughput	large_terms	1.5	ops/s
Max Throughput	large_terms	1.5	ops/s
50th percentile latency	large_terms	567.113	ms
90th percentile latency	large_terms	586.85	ms
99th percentile latency	large_terms	704.151	ms
100th percentile latency	large_terms	757.566	ms
50th percentile service time	large_terms	566.891	ms
90th percentile service time	large_terms	584.571	ms
99th percentile service time	large_terms	678.04	ms
100th percentile service time	large_terms	757.381	ms
error rate	large_terms	0	%

Metric	Operation	Value	Unit
Min Throughput	large_filtered_terms	1.5	ops/s
Median Throughput	large_filtered_terms	1.5	ops/s
Max Throughput	large_filtered_terms	1.5	ops/s
50th percentile latency	large_filtered_terms	570.444	ms
90th percentile latency	large_filtered_terms	588.125	ms
99th percentile latency	large_filtered_terms	734.918	ms
100th percentile latency	large_filtered_terms	775.135	ms
50th percentile service time	large_filtered_terms	569.793	ms
90th percentile service time	large_filtered_terms	585.437	ms
99th percentile service time	large_filtered_terms	682.943	ms
100th percentile service time	large_filtered_terms	774.924	ms
error rate	large_filtered_terms	0	%
Min Throughput	large_prohibited_terms	1.5	ops/s
Median Throughput	large_prohibited_terms	1.5	ops/s
Max Throughput	large_prohibited_terms	1.5	ops/s
50th percentile latency	large_prohibited_terms	567.806	ms
90th percentile latency	large_prohibited_terms	586.277	ms
99th percentile latency	large_prohibited_terms	623.088	ms
100th percentile latency	large_prohibited_terms	636.81	ms
50th percentile service time	large_prohibited_terms	567.621	ms
90th percentile service time	large_prohibited_terms	586.127	ms
99th percentile service time	large_prohibited_terms	622.914	ms

Metric	Operation	Value	Unit
100th percentile service time	large_prohibited_terms	636.642	ms
error rate	large_prohibited_terms	0	%

6.5. Comparison of stress testing results between an Elasticsearch cluster with three 4-vCPU 16-GiB data nodes and an Elasticsearch cluster with three 2-vCPU 8-GiB data nodes

This topic compares the stress testing results of two Elasticsearch clusters: a cluster with three 4-vCPU 16-GiB data nodes and a cluster with three 2-vCPU 8-GiB data nodes.

? Note

Metric	Operation	Cluster with 4-vCPU 16-GiB data nodes	Cluster with 2-vCPU 8-GiB data nodes	Difference	Unit
Merge time	None	11.0297	14.3001	3.27042	min
Refresh time	None	3.05238	5.26405	2.21167	min
Flush time	None	0.04485	0.0308333	-0.01402	min
Merge throttle time	None	1.39282	1.27945	-0.11337	min
Total Young Gen GC	None	92.902	183.74	90.838	s
Total Old Gen GC	None	0.4	1.125	0.725	s
Heap used for segments	None	18.7955	18.8167	0.02126	MB
Heap used for doc values	None	0.360752	0.452751	0.092	MB
Heap used for terms	None	17.2739	17.2004	-0.07343	MB

Metric	Operation	Cluster with 4-vCPU 16-GiB data nodes	Cluster with 2-vCPU 8-GiB data nodes	Difference	Unit
Heap used for norms	None	0.0877075	0.0852051	-0.0025	MB
Heap used for points	None	0.241213	0.241465	0.00025	MB
Heap used for stored fields	None	0.831932	0.836876	0.00494	MB
Segment count	None	133	140	7	items
Min Throughput	index-append	51751.7	28115.4	-23636.2	docs/s
Median Throughput	index-append	52303	28645.6	-23657.5	docs/s
Max Throughput	index-append	54076.3	30037.8	-24038.5	docs/s
50th percentile latency	index-append	743.939	1447.76	703.818	ms
90th percentile latency	index-append	1045.7	1847.05	801.342	ms
99th percentile latency	index-append	1325.21	2264.68	939.47	ms
100th percentile latency	index-append	1794.39	2608.68	814.293	ms
50th percentile service time	index-append	743.939	1447.76	703.818	ms
90th percentile service time	index-append	1045.7	1847.05	801.342	ms
99th percentile service time	index-append	1325.21	2264.68	939.47	ms

Metric	Operation	Cluster with 4-vCPU 16-GiB data nodes	Cluster with 2-vCPU 8-GiB data nodes	Difference	Unit
100th percentile service time	index-append	1794.39	2608.68	814.293	ms
error rate	index-append	0	0	0	%
Min Throughput	force-merge	0.950072	2.10087	1.1508	ops/s
Median Throughput	force-merge	0.950072	2.10087	1.1508	ops/s
Max Throughput	force-merge	0.950072	2.10087	1.1508	ops/s
100th percentile latency	force-merge	1052.54	475.984	-576.556	ms
100th percentile service time	force-merge	1052.54	475.984	-576.556	ms
error rate	force-merge	0	0	0	%
Min Throughput	index-stats	100.037	97.7524	-2.28456	ops/s
Median Throughput	index-stats	100.049	100.048	-0.00112	ops/s
Max Throughput	index-stats	100.085	100.068	-0.01745	ops/s
50th percentile latency	index-stats	4.85232	5.09015	0.23784	ms
90th percentile latency	index-stats	5.14185	10.7365	5.59466	ms
99th percentile latency	index-stats	77.3127	234.761	157.448	ms
99.9th percentile latency	index-stats	123.888	277.393	153.505	ms

Metric	Operation	Cluster with 4-vCPU 16-GiB data nodes	Cluster with 2-vCPU 8-GiB data nodes	Difference	Unit
100th percentile latency	index-stats	128.01	281.866	153.856	ms
50th percentile service time	index-stats	4.78006	5.01096	0.23091	ms
90th percentile service time	index-stats	4.9831	5.30021	0.31711	ms
99th percentile service time	index-stats	9.66475	12.0005	2.33576	ms
99.9th percentile service time	index-stats	48.4445	141.631	93.186	ms
100th percentile service time	index-stats	127.945	150.153	22.2078	ms
error rate	index-stats	0	0	0	%
Min Throughput	node-stats	100.054	100.007	-0.04689	ops/s
Median Throughput	node-stats	100.098	100.085	-0.01341	ops/s
Max Throughput	node-stats	100.551	100.494	-0.0566	ops/s
50th percentile latency	node-stats	4.55259	4.90659	0.354	ms
90th percentile latency	node-stats	4.78784	5.29285	0.50501	ms
99th percentile latency	node-stats	18.8034	29.3245	10.5211	ms
99.9th percentile latency	node-stats	43.7684	43.3885	-0.3799	ms

Metric	Operation	Cluster with 4-vCPU 16-GiB data nodes	Cluster with 2-vCPU 8-GiB data nodes	Difference	Unit
100th percentile latency	node-stats	48.1474	44.6019	-3.54548	ms
50th percentile service time	node-stats	4.48138	4.83552	0.35414	ms
90th percentile service time	node-stats	4.69386	5.12694	0.43308	ms
99th percentile service time	node-stats	5.64618	9.08739	3.44121	ms
99.9th percentile service time	node-stats	27.8155	39.744	11.9285	ms
100th percentile service time	node-stats	43.6905	44.5383	0.84783	ms
error rate	node-stats	0	0	0	%
Min Throughput	default	49.8129	47.8334	-1.97948	ops/s
Median Throughput	default	50.0009	48.281	-1.71982	ops/s
Max Throughput	default	50.0045	48.7269	-1.2776	ops/s
50th percentile latency	default	19.7245	617.465	597.74	ms
90th percentile latency	default	94.1457	1033.98	939.834	ms
99th percentile latency	default	133.091	1083.23	950.137	ms
99.9th percentile latency	default	137.285	1095.4	958.114	ms

Metric	Operation	Cluster with 4-vCPU 16-GiB data nodes	Cluster with 2-vCPU 8-GiB data nodes	Difference	Unit
100th percentile latency	default	138.043	1097.14	959.1	ms
50th percentile service time	default	19.1469	18.646	-0.50082	ms
90th percentile service time	default	19.9554	24.9381	4.98271	ms
99th percentile service time	default	25.3462	35.7667	10.4206	ms
99.9th percentile service time	default	54.7931	57.3679	2.57481	ms
100th percentile service time	default	133.771	151.505	17.7337	ms
error rate	default	0	0	0	%
Min Throughput	term	200.055	199.431	-0.62401	ops/s
Median Throughput	term	200.076	200.072	-0.00349	ops/s
Max Throughput	term	200.119	200.13	0.01076	ops/s
50th percentile latency	term	3.07948	2.9728	-0.10668	ms
90th percentile latency	term	3.37296	7.10648	3.73353	ms
99th percentile latency	term	22.3272	22.4487	0.12153	ms
99.9th percentile latency	term	26.9648	29.0737	2.10889	ms

Metric	Operation	Cluster with 4-vCPU 16-GiB data nodes	Cluster with 2-vCPU 8-GiB data nodes	Difference	Unit
100th percentile latency	term	28.1562	29.6253	1.46915	ms
50th percentile service time	term	3.00599	2.87833	-0.12766	ms
90th percentile service time	term	3.15279	3.08983	-0.06296	ms
99th percentile service time	term	4.22302	19.9777	15.7546	ms
99.9th percentile service time	term	26.9017	29.0082	2.10648	ms
100th percentile service time	term	28.0823	29.5597	1.4774	ms
error rate	term	0	0	0	%
Min Throughput	phrase	199.842	199.711	-0.13145	ops/s
Median Throughput	phrase	200.04	200.038	-0.00174	ops/s
Max Throughput	phrase	200.087	200.074	-0.0125	ops/s
50th percentile latency	phrase	3.76927	3.61484	-0.15442	ms
90th percentile latency	phrase	13.6055	16.5523	2.94681	ms
99th percentile latency	phrase	28.0245	31.394	3.36944	ms
99.9th percentile latency	phrase	34.7198	33.902	-0.81778	ms

Metric	Operation	Cluster with 4-vCPU 16-GiB data nodes	Cluster with 2-vCPU 8-GiB data nodes	Difference	Unit
100th percentile latency	phrase	35.551	34.5784	-0.97253	ms
50th percentile service time	phrase	3.67227	3.47402	-0.19825	ms
90th percentile service time	phrase	4.08037	3.90958	-0.17079	ms
99th percentile service time	phrase	16.9256	19.3773	2.45168	ms
99.9th percentile service time	phrase	24.4886	22.7947	-1.69386	ms
100th percentile service time	phrase	29.8604	27.8164	-2.04399	ms
error rate	phrase	0	0	0	%
Min Throughput	country_agg_uncached	4.95005	4.6328	-0.31724	ops/s
Median Throughput	country_agg_uncached	4.99422	4.65258	-0.34163	ops/s
Max Throughput	country_agg_uncached	5.00022	4.67361	-0.32661	ops/s
50th percentile latency	country_agg_uncached	330.923	14864.3	14533.3	ms
90th percentile latency	country_agg_uncached	2780.17	21046	18265.8	ms
99th percentile latency	country_agg_uncached	2866	22902	20036	ms
99.9th percentile latency	country_agg_uncached	2880.39	22997.6	20117.2	ms

Metric	Operation	Cluster with 4-vCPU 16-GiB data nodes	Cluster with 2-vCPU 8-GiB data nodes	Difference	Unit
100th percentile latency	country_agg_uncached	2882.11	23018.7	20136.6	ms
50th percentile service time	country_agg_uncached	197.883	204.174	6.29064	ms
90th percentile service time	country_agg_uncached	213.402	242.492	29.0907	ms
99th percentile service time	country_agg_uncached	256.649	345.382	88.7335	ms
99.9th percentile service time	country_agg_uncached	290.496	378.302	87.8056	ms
100th percentile service time	country_agg_uncached	296.875	422.53	125.655	ms
error rate	country_agg_uncached	0	0	0	%
Min Throughput	country_agg_cached	99.9249	98.3659	-1.55896	ops/s
Median Throughput	country_agg_cached	100.064	100.056	-0.00795	ops/s
Max Throughput	country_agg_cached	100.112	100.135	0.02245	ops/s
50th percentile latency	country_agg_cached	3.30479	3.2638	-0.04099	ms
90th percentile latency	country_agg_cached	3.52514	4.69259	1.16745	ms
99th percentile latency	country_agg_cached	52.8258	189.143	136.317	ms
99.9th percentile latency	country_agg_cached	112.895	249.851	136.956	ms

Metric	Operation	Cluster with 4-vCPU 16-GiB data nodes	Cluster with 2-vCPU 8-GiB data nodes	Difference	Unit
100th percentile latency	country_agg_cached	119.435	256.028	136.593	ms
50th percentile service time	country_agg_cached	3.23149	3.18679	-0.0447	ms
90th percentile service time	country_agg_cached	3.41319	3.42086	0.00767	ms
99th percentile service time	country_agg_cached	7.60955	20.4171	12.8075	ms
99.9th percentile service time	country_agg_cached	26.2229	117.273	91.0502	ms
100th percentile service time	country_agg_cached	119.365	255.951	136.586	ms
error rate	country_agg_cached	0	0	0	%
Min Throughput	scroll	61.5897	59.1628	-2.42689	ops/s
Median Throughput	scroll	61.6735	60.4406	-1.23292	ops/s
Max Throughput	scroll	61.9387	61.019	-0.91967	ops/s
50th percentile latency	scroll	164549	168347	3798.13	ms
90th percentile latency	scroll	237443	240658	3214.79	ms
99th percentile latency	scroll	253860	257048	3187.91	ms
100th percentile latency	scroll	255710	258853	3143.03	ms

Metric	Operation	Cluster with 4-vCPU 16-GiB data nodes	Cluster with 2-vCPU 8-GiB data nodes	Difference	Unit
50th percentile service time	scroll	399.964	402.962	2.99858	ms
90th percentile service time	scroll	424.303	431.267	6.96397	ms
99th percentile service time	scroll	523.877	455.632	-68.2449	ms
100th percentile service time	scroll	639.45	601.214	-38.236	ms
error rate	scroll	0	0	0	%
Min Throughput	expression	1.9994	1.9998	0.0004	ops/s
Median Throughput	expression	2.00113	2.00113	0	ops/s
Max Throughput	expression	2.00186	2.00189	2e-05	ops/s
50th percentile latency	expression	409.927	409.417	-0.5091	ms
90th percentile latency	expression	434.544	434.858	0.31406	ms
99th percentile latency	expression	532.412	501.498	-30.914	ms
100th percentile latency	expression	537.618	517.438	-20.1798	ms
50th percentile service time	expression	409.812	409.165	-0.64674	ms
90th percentile service time	expression	428.156	434.749	6.59297	ms

Metric	Operation	Cluster with 4-vCPU 16-GiB data nodes	Cluster with 2-vCPU 8-GiB data nodes	Difference	Unit
99th percentile service time	expression	532.33	498.681	-33.6493	ms
100th percentile service time	expression	537.495	517.332	-20.1637	ms
error rate	expression	0	0	0	%
Min Throughput	painless_static	1.99752	1.96306	-0.03446	ops/s
Median Throughput	painless_static	1.99998	1.96607	-0.03391	ops/s
Max Throughput	painless_static	2.00041	1.96914	-0.03127	ops/s
50th percentile latency	painless_static	497.626	3163.94	2666.31	ms
90th percentile latency	painless_static	643.32	3679.27	3035.95	ms
99th percentile latency	painless_static	700.559	3994.52	3293.97	ms
100th percentile latency	painless_static	704.679	4006.5	3301.82	ms
50th percentile service time	painless_static	490.705	503.588	12.8834	ms
90th percentile service time	painless_static	500.663	528.807	28.1439	ms
99th percentile service time	painless_static	642.124	600.103	-42.021	ms
100th percentile service time	painless_static	683.621	623.666	-59.9546	ms

Metric	Operation	Cluster with 4-vCPU 16-GiB data nodes	Cluster with 2-vCPU 8-GiB data nodes	Difference	Unit
error rate	painless_static	0	0	0	%
Min Throughput	painless_dynamic	1.99721	1.99513	-0.00209	ops/s
Median Throughput	painless_dynamic	2.00032	1.99838	-0.00194	ops/s
Max Throughput	painless_dynamic	2.00089	2.00053	-0.00036	ops/s
50th percentile latency	painless_dynamic	473.087	611.305	138.218	ms
90th percentile latency	painless_dynamic	554.729	786.806	232.077	ms
99th percentile latency	painless_dynamic	668.363	973.432	305.069	ms
100th percentile latency	painless_dynamic	706.557	982.484	275.926	ms
50th percentile service time	painless_dynamic	469.145	494.097	24.9528	ms
90th percentile service time	painless_dynamic	501.774	518.082	16.3086	ms
99th percentile service time	painless_dynamic	606.61	606.748	0.13817	ms
100th percentile service time	painless_dynamic	624.751	638.903	14.1524	ms
error rate	painless_dynamic	0	0	0	%
Min Throughput	large_filtered_terms	1.64076	1.38866	-0.2521	ops/s
Median Throughput	large_filtered_terms	1.6443	1.39554	-0.24876	ops/s

Metric	Operation	Cluster with 4-vCPU 16-GiB data nodes	Cluster with 2-vCPU 8-GiB data nodes	Difference	Unit
Max Throughput	large_filtered_terms	1.65048	1.39764	-0.25283	ops/s
50th percentile latency	large_filtered_terms	33013.5	65601.1	32587.5	ms
90th percentile latency	large_filtered_terms	40869	82494.7	41625.7	ms
99th percentile latency	large_filtered_terms	42644	86452.2	43808.2	ms
100th percentile latency	large_filtered_terms	42936.2	86857.3	43921.1	ms
50th percentile service time	large_filtered_terms	598.001	707.17	109.169	ms
90th percentile service time	large_filtered_terms	626.81	747.949	121.139	ms
99th percentile service time	large_filtered_terms	771.815	847.069	75.2534	ms
100th percentile service time	large_filtered_terms	796.884	927.917	131.032	ms
error rate	large_filtered_terms	0	0	0	%
Min Throughput	large_prohibited_terms	1.6893	1.45607	-0.23323	ops/s
Median Throughput	large_prohibited_terms	1.69452	1.46074	-0.23379	ops/s
Max Throughput	large_prohibited_terms	1.69856	1.46248	-0.23608	ops/s
50th percentile latency	large_prohibited_terms	27732.3	55916.3	28184	ms

Metric	Operation	Cluster with 4-vCPU 16-GiB data nodes	Cluster with 2-vCPU 8-GiB data nodes	Difference	Unit
90th percentile latency	large_prohibited_terms	34305.5	70529.7	36224.2	ms
99th percentile latency	large_prohibited_terms	35840.4	73769.1	37928.7	ms
100th percentile latency	large_prohibited_terms	35993.5	74143.9	38150.4	ms
50th percentile service time	large_prohibited_terms	586.382	679.394	93.0121	ms
90th percentile service time	large_prohibited_terms	618.185	717.476	99.2908	ms
99th percentile service time	large_prohibited_terms	661.378	782.085	120.707	ms
100th percentile service time	large_prohibited_terms	823.782	822.723	-1.05804	ms
error rate	large_prohibited_terms	0	0	0	%

7.Specifications

Alibaba Cloud Elasticsearch supports a variety of specifications. When you purchase an Elasticsearch cluster, you can select specifications for the cluster based on your business requirements. This topic describes the supported specifications and cloud disk types.

Specifications

Cluster type	vCPU and memory	Family
elasticsearch.n4.small	1 vCPU and 2 GiB of memory	1:2
elasticsearch.sn1ne.large	2 vCPUs and 4 GiB of memory	
elasticsearch.sn2ne.large	2 vCPUs and 8 GiB of memory	1:4
elasticsearch.sn2ne.xlarge	4 vCPUs and 16 GiB of memory	
elasticsearch.sn2ne.2xlarge	8 vCPUs and 32 GiB of memory	
elasticsearch.sn2ne.4xlarge	16 vCPUs and 64 GiB of memory	

Note

- Alibaba Cloud Elasticsearch no longer provides data nodes with 1 vCPU and 2 GiB of memory or data nodes with 2 vCPUs and 2 GiB of memory due to inventory issues and the impact on performance stability. Existing data nodes with these specifications are not affected.
- Data nodes with 1 vCPU and 2 GiB of memory are designed only for testing purposes. Do not use clusters that contain such data nodes for production purposes. The service-level agreement (SLA) does not apply to these clusters. Therefore, we recommend that you upgrade your data nodes with 1 vCPU and 2 GiB of memory.
- Data node specifications supported in different regions may vary. The specifications on the buy page or the specifications provided in Pricing of Alibaba Cloud Elasticsearch prevail.

Cloud disk types

Cloud disk type	Description
cloud_efficiency	Ultra disk
cloud_ssd	SSD

Note For more information about cloud disks, see Disks.

8.Compatibility matrixes

This topic describes the compatibility among the versions of open source Elasticsearch, Logstash, and Beats.

Elasticsearch compatibility (5.x and later)

Note

- ^: The compatibility between Elasticsearch and other services when Elasticsearch is used as an output (index data is synchronized to Elasticsearch by using Beats or Logstash).
- *: We recommend that you use the latest versions of Beats, Logstash, and ES-Hadoop. Earlier versions may not support all the desired features.
- **: In Elasticsearch 6.3 and later, all the features of X-Pack are delivered with the default distributions of Elastic Stack. For more information, see X-Pack.

Elasticsearch	Kibana	X-Pack	Beats^*	Logstash ^*	ES-Hadoop (jar)*	APM Server	App Search
5.0.x	5.0.x	5.0.x	1.3.x to 5.6.x	2.4.x to 5.6.x	5.0.x to 5.6.x	N/A	N/A
5.1.x	5.1.x	5.1.x	1.3.x to 5.6.x	2.4.x to 5.6.x	5.0.x to 5.6.x	N/A	N/A
5.2.x	5.2.x	5.2.x	1.3.x to 5.6.x	2.4.x to 5.6.x	5.0.x to 5.6.x	N/A	N/A
5.3.x	5.3.x	5.3.x	1.3.x to 5.6.x	2.4.x to 5.6.x	5.0.x to 5.6.x	N/A	N/A
5.4.x	5.4.x	5.4.x	1.3.x to 5.6.x	2.4.x to 5.6.x	5.0.x to 5.6.x	N/A	N/A
5.5.x	5.5.x	5.5.x	1.3.x to 5.6.x	2.4.x to 5.6.x	5.0.x to 5.6.x	N/A	N/A
5.6.x	5.6.x	5.6.x	1.3.x to 6.0.x	2.4.x to 6.0.x	5.0.x to 6.0.x	N/A	N/A
6.0.x	6.0.x	6.0.x	5.6.x to 6.8.x	5.6.x to 6.8.x	6.0.x to 6.8.x	N/A	N/A
6.1.x	6.1.x	6.1.x	5.6.x to 6.8.x	5.6.x to 6.8.x	6.0.x to 6.8.x	N/A	N/A
6.2.x	6.2.x	6.2.x	5.6.x to 6.8.x	5.6.x to 6.8.x	6.0.x to 6.8.x	6.2.x to 6.8.x	N/A
6.3.x	6.3.x	N/A**	5.6.x to 6.8.x	5.6.x to 6.8.x	6.0.x to 6.8.x	6.2.x to 6.8.x	N/A

Elasticsearch	Kibana	X-Pack	Beats^*	Logstash ^*	ES-Hadoop (jar)*	APM Server	App Search

6.4.x	6.4.x	N/A**	5.6.x to 6.8.x	5.6.x to 6.8.x	6.0.x to 6.8.x	6.2.x to 6.8.x	N/A
6.5.x	6.5.x	N/A**	5.6.x to 6.8.x	5.6.x to 6.8.x	6.0.x to 6.8.x	6.2.x to 6.8.x	N/A
6.6.x	6.6.x	N/A**	5.6.x to 6.8.x	5.6.x to 6.8.x	6.0.x to 6.8.x	6.2.x to 6.8.x	N/A
6.7.x	6.7.x	N/A**	5.6.x to 6.8.x	5.6.x to 6.8.x	6.0.x to 6.8.x	6.2.x to 6.8.x	N/A
6.8.x	6.8.x	N/A**	5.6.x to 6.8.x	5.6.x to 6.8.x	6.0.x to 6.8.x	6.2.x to 6.8.x	N/A
7.0.x	7.0.x	N/A**	6.8.x to 7.13.x	6.8.x to 7.13.x	7.0.x to 7.13.x	7.0.x to 7.13.x***	N/A
7.1.x	7.1.x	N/A**	6.8.x to 7.13.x	6.8.x to 7.13.x	7.0.x to 7.13.x	7.0.x to 7.13.x***	N/A
7.2.x	7.2.x	N/A**	6.8.x to 7.13.x	6.8.x to 7.13.x	7.0.x to 7.13.x	7.0.x to 7.13.x***	7.2.x
7.3.x	7.3.x	N/A**	6.8.x to 7.13.x	6.8.x to 7.13.x	7.0.x to 7.13.x	7.0.x to 7.13.x***	7.3.x
7.4.x	7.4.x	N/A**	6.8.x to 7.13.x	6.8.x to 7.13.x	7.0.x to 7.13.x	7.0.x to 7.13.x***	7.4.x
7.5.x	7.5.x	N/A**	6.8.x to 7.13.x	6.8.x to 7.13.x	7.0.x to 7.13.x	7.0.x to 7.13.x***	7.5.x
7.6.x	7.6.x	N/A**	6.8.x to 7.13.x	6.8.x to 7.13.x	7.0.x to 7.13.x	7.0.x to 7.13.x***	7.6.x
7.7.x	7.7.x	N/A**	6.8.x to 7.13.x	6.8.x to 7.13.x	7.0.x to 7.13.x	7.0.x to 7.13.x***	N/A****
7.8.x	7.8.x	N/A**	6.8.x to 7.13.x	6.8.x to 7.13.x	7.0.x to 7.13.x	7.0.x to 7.13.x***	N/A****
7.9.x	7.9.x	N/A**	6.8.x to 7.13.x	6.8.x to 7.13.x	7.0.x to 7.13.x	7.0.x to 7.13.x***	N/A****
7.10.x	7.10.x	N/A**	6.8.x to 7.13.x	6.8.x to 7.13.x	7.0.x to 7.13.x	7.0.x to 7.13.x***	N/A****

We recommend that all Elasticsearch, Kibana, Filebeat, and Logstash clusters use the same minor version.

Logstash compatibility

Note

- *: The compatibility of monitoring and managing Elasticsearch clusters, including clusters specified by `xpack.monitoring.elasticsearch.url` and `xpack.management.elasticsearch.url`. We recommend that all Elasticsearch, Kibana, and Logstash clusters use the same minor version. This helps achieve the best performance of cluster monitoring and management. If you want to monitor and manage clusters of 6.2 or earlier, you must install X-Pack on all these services.
- **: Only Elasticsearch can be used as an output for Functionbeat in versions earlier than 7.4. Logstash and other services are not supported. Both Logstash and Elasticsearch can be used as outputs for Functionbeat in 7.4 and later.

Logstash	Beats**	Monitoring and management of Elasticsearch clusters*
2.4.x	1.0.x to 5.6.x	N/A
5.0.x	1.3.x to 5.6.x	N/A
5.1.x	5.0.x to 5.6.x	N/A
5.2.x	5.0.x to 5.6.x	5.2.x to 5.6.x
5.3.x	5.0.x to 5.6.x	5.3.x to 5.6.x
5.4.x	5.0.x to 5.6.x	5.4.x to 5.6.x
5.5.x	5.0.x to 5.6.x	5.5.x to 5.6.x
5.6.x	5.6.x to 6.8.x	5.6.x to 6.0.x
6.0.x	5.6.x to 6.8.x	6.0.x to 6.8.x
6.1.x	5.6.x to 6.8.x	6.1.x to 6.8.x
6.2.x	5.6.x to 6.8.x	6.2.x to 6.8.x
6.3.x	5.6.x to 6.8.x	6.3.x to 6.8.x
6.4.x	5.6.x to 6.8.x	6.4.x to 6.8.x
6.5.x	5.6.x to 6.8.x	6.5.x to 6.8.x
6.6.x	5.6.x to 6.8.x	6.6.x to 6.8.x
6.7.x	5.6.x to 6.8.x	6.7.x to 6.8.x

Logstash	Beats**	Monitoring and management of Elasticsearch clusters*
6.8.x	5.6.x to 6.8.x	6.8.x
7.0.x	6.8.x to 7.13.x	7.0.x to 7.13.x
7.1.x	6.8.x to 7.13.x	7.1.x to 7.13.x
7.2.x	6.8.x to 7.13.x	7.2.x to 7.13.x
7.3.x	6.8.x to 7.13.x	7.3.x to 7.13.x
7.4.x	6.8.x to 7.13.x	7.4.x to 7.13.x
7.5.x	6.8.x to 7.13.x	7.5.x to 7.13.x
7.6.x	6.8.x to 7.13.x	7.6.x to 7.13.x
7.7.x	6.8.x to 7.13.x	7.7.x to 7.13.x
7.8.x	6.8.x to 7.13.x	7.8.x to 7.13.x
7.9.x	6.8.x to 7.13.x	7.9.x to 7.13.x
7.10.x	6.8.x to 7.13.x	7.10.x to 7.13.x

Beats compatibility

Note

- *: The compatibility of monitoring Elasticsearch clusters, including clusters specified by `xpack.monitoring.elasticsearch`. We recommend that all Elasticsearch, Kibana, and Logstash clusters use the same minor version. This helps achieve the best performance of cluster monitoring. If you want to monitor clusters of 6.2 or earlier, you must install X-Pack on all these services.
- **: Only Elasticsearch can be used as an output for Functionbeat. Logstash and other services are not supported.

Beats**	Logstash	Monitoring of Elasticsearch clusters*
1.3.x	2.0.x to 5.0.x	N/A
5.0.x	2.0.x to 5.6.x	N/A
5.1.x	2.0.x to 5.6.x	N/A
5.2.x	2.0.x to 5.6.x	N/A
5.3.x	2.0.x to 5.6.x	N/A

Beats**	Logstash	Monitoring of Elasticsearch clusters*
5.4.x	2.0.x to 5.6.x	N/A
5.5.x	2.0.x to 5.6.x	N/A
5.6.x	5.6.x to 6.8.x	N/A
6.0.x	5.6.x to 6.8.x	N/A
6.1.x	5.6.x to 6.8.x	N/A
6.2.x	5.6.x to 6.8.x	6.2.x
6.3.x	5.6.x to 6.8.x	6.3.x to 6.8.x
6.4.x	5.6.x to 6.8.x	6.4.x to 6.8.x
6.5.x	5.6.x to 6.8.x	6.5.x to 6.8.x
6.6.x	5.6.x to 6.8.x	6.6.x to 6.8.x
6.7.x	5.6.x to 6.8.x	6.7.x to 6.8.x
6.8.x	5.6.x to 6.8.x	6.8.x to 7.13.x
7.0.x	6.8.x to 7.13.x	7.0.x to 7.13.x
7.1.x	6.8.x to 7.13.x	7.1.x to 7.13.x
7.2.x	6.8.x to 7.13.x	7.2.x to 7.13.x
7.3.x	6.8.x to 7.13.x	7.3.x to 7.13.x
7.4.x	6.8.x to 7.13.x	7.4.x to 7.13.x
7.5.x	6.8.x to 7.13.x	7.5.x to 7.13.x
7.6.x	6.8.x to 7.13.x	7.6.x to 7.13.x
7.7.x	6.8.x to 7.13.x	7.7.x to 7.13.x
7.8.x	6.8.x to 7.13.x	7.8.x to 7.13.x
7.9.x	6.8.x to 7.13.x	7.9.x to 7.13.x
7.10.x	6.8.x to 7.13.x	7.10.x to 7.13.x

For more information about the compatibility between these services, see Product Compatibility.