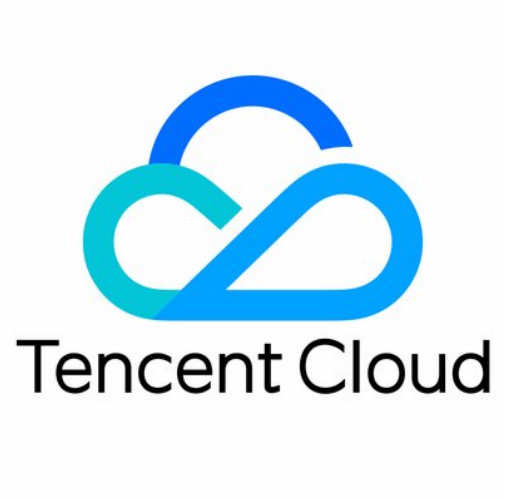
Elasticsearch Service

Product Intro

Product Introduction



Contents

Product Intro
Overview

Last updated：2018-09-14 17:08:26

Tencent Cloud Elasticsearch Service (ES) is a highly available and scalable cloud-hosted service built on the open source search engine Elasticsearch. ES not only has technical and resource advantages in computing, storage, security and other fields of Tencent Cloud Computing, but also maintains the compatibility and openness of Elasticsearch. It provides stable, elastic and scalable cloud-hosted services and various cluster management features, helping you avoid the deployment and debugging of software and hardware and focus on your business. ES is suitable for the storage and search of massive data, real-time log analysis and other scenarios, such as website search navigation, enterprise-grade search, service log exception monitoring, clickstream analysis and so on.

ES is built on the Elasticsearch 5.6.4 and comes with two components (Elasticsearch and Kibana) deployed in a VPC, allowing you to easily integrate it with the existing services in the VPC. It also provides the Kibana monitoring page accessed via the public network for you to easily search and statistically analyze cluster data in a browser.

Benefits

Last updated：2018-09-14 17:08:43

ES provides a highly available, scalable and easily maintained cloud-hosted Elasticsearch cluster service.

Elastic deployment

A variety of node instance specifications are provided to meet your needs in all kinds of scenarios. You can select an appropriate model based on the volume of data and queries, and customize the storage capacity of nodes to minimize wasteful costs caused by resource idleness. As your data volumes increase with business growth, clusters can be dynamically scaled to meet ever-changing business needs.

Ease of use

An ES cluster can be created in just minutes through a simple configuration process, eliminating the need for deploying and debugging hardware and software. ES also provides convenient cluster OPS tools (via Kibana pages) and a comprehensive cluster monitoring and alarming system to meet your daily cluster OPS needs.

Openness

ES supports the complete ELK ecosystem, covering the entire lifecycle from data collection to storage and consumption. It supports various open source plug-ins and RESTful APIs, allowing you to select appropriate plug-ins and clients and build your own applications on ES clusters.

High availability

ES adopts a multi-node and multi-replica deployment scheme which intelligently detects and replaces faulty nodes, reducing the risk of data loss caused by failures and ensuring data security and service stability.

Security

ES is deployed on logically isolated VPCs, enabling you to configure your VPC environment and customize network access control lists and security groups. Further, access from browser to Kibana can be controlled via the blacklist and whitelist of IPs, systematically protecting the security of your cloud resources.

Service integration

ES can be easily integrated with other products of Tencent Cloud such as COS, Flow Logs, Message Queue and CDB to provide you with various data capabilities like data transfer and backup for different usage scenarios.

Scenarios

Last updated：2018-09-14 17:09:10

Log analysis

Logs generated by devices such as web servers, mobile devices and IoT sensors are inherently scattered across nodes, in large scale and vary in type, posing a major challenge for troubleshooting and business analysis through log search. Featuring an elastically scalable and near real-time centralized storage solution and full-text search capability, ES simplifies the unified management and query of logs, helping users quickly locate issues and improve troubleshooting efficiency.

Full-text Search

Ecommerce goods search, mobile application search, enterprise internal information search and other site search services are necessary ways to obtain information efficiently. ES features full-text search and supports both structured and unstructured data. It also provides simple and easy-to-use RESTful API and clients in various languages for you to quickly build stable search services that can be integrated into your existing business frameworks.

Business intelligence (BI)

In the context of data-driven operations, fields such as ecommerce, mobile applications, advertising and media require in-depth statistical analysis and mining of data to assist business decision-making. However, the massive amounts of business data have posed a major challenge to companies in these fields. ES is capable of structural queries and supports complex filtering and aggregated statistics, helping customers perform statistical analysis of large volumes of data in an efficient and customized manner, identify problems and opportunities, make business decisions and fully unleash the value of data.

Glossary

Last updated：2018-09-14 17:12:25

Cluster and nodes

A running Elasticsearch instance is a node. In a network, one or more nodes that have the same cluster name and can communicate with each other form an Elasticsearch cluster. Nodes in the cluster work together to store data and process query requests. When new nodes are added into the cluster or some nodes are removed from the cluster, the cluster will re-distribute the data evenly. Each node knows where a document is located. Any node that receives a query request forwards it directly to the nodes where the document is stored, then collects data from these nodes, and finally returns the result to the client. All the above processes are transparent.

Document

Elasticsearch is document-oriented, which stores the entire object or document, and indexes the content of each document so that it can be searched. Elasticsearch serializes documents to JSON, making them simple, concise and easy to read. JSON serialization is supported by most programming languages and has become a standard format in NoSQL area. In Elasticsearch, you can index, search, sort and filter the entire document instead of searching data from rows and columns. This is a completely different way of thinking about data, and also the reason why Elasticsearch can support complex full-text search.

Index

Common concept

Just like a traditional relational database, an index is a place where relational documents are stored. The plural of index is indices or indexes.

Indexing a document is to store a document in an index, so that is can be searched and queried. An existing document will be replaced with the new version, which is similar to the keyword INSERT in SQL statements.

Inverted index

The relational database accelerates data search speed by adding an index, such as a B-tree index, to the specified column. Elasticsearch and Lucene use a structure called inverted index to achieve the same purpose. By default, each attribute in a document is indexed (with an inverted index) and searchable. An attribute without an inverted index cannot be searched.

Shard

A shard is a data container where documents are stored. A shard is an underlying work unit, which only stores part of the data. Shards are assigned to each node of a cluster. When you scale your cluster up or down, Elasticsearch will automatically migrate shards between nodes to distribute data evenly in the cluster.

A shard can be a primary shard or a replica. Any document in an index belongs to a primary shard, so the number of primary shards determines the maximum volume of data that can be stored in the index. Technically, a primary shard can hold up to Integer.MAX_VALUE - 128 documents.

A replica is a copy of a primary shard. Replicas are used as redundant backups to avoid data loss in case of hardware failure, and provide services for read operations such as searching and returning documents. When an index is created, the number of primary shards is set, but the number of replicas can be modified at any time.