

Alibaba Cloud Website Search Solution

Enabling organizations to build intuitive and customer-oriented websites

Introduction





Alibaba Cloud Website Search solution offers multiple stages of infrastructure for organizations to develop website search engines. Users can successfully minimize the complexity of search engine development along with reducing IT costs using this solution. Additionally, it enhances the accuracy of site search results for visitors.

Background

To process search requests, website search engines require natural language processing (NLP) and reverse indexing. Furthermore, they also need to ensure the presence of data-intensive computing to provide accurate search results consistently.

This calls for an infrastructure with high CPU computing power and I/O capacity. To support NLP, the system will need to use Hadoop or Spark to build a computing cluster. Additionally, Bigtable storage is needed to store source data.

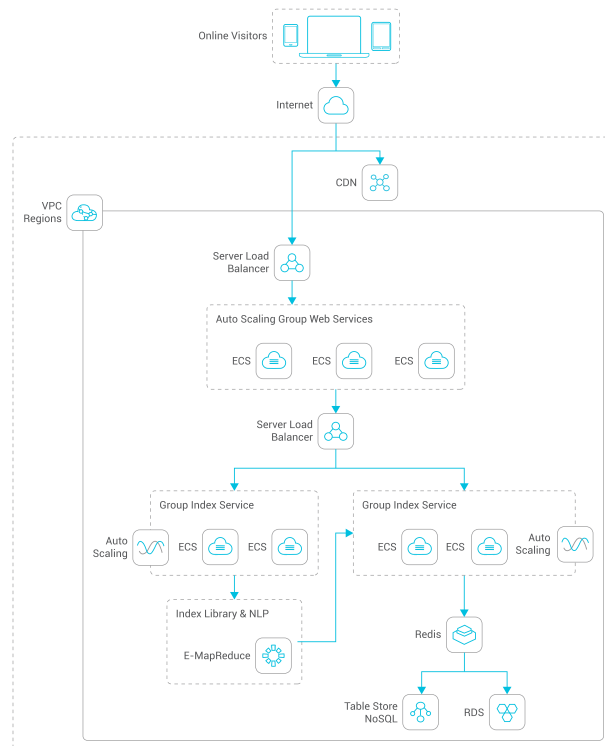
Highlights

-  Higher I/O performance
-  Storage of massive unstructured data
-  Automatic failover of E-MapReduce cluster
-  Multi-layer cache

Benefits

- ✓ Generate reverse index from the indexing framework
- ✓ Ensure high-performance network and disk IO
- ✓ Facilitate storage and real-time access to massive unstructured data
- ✓ Obtain relevant content through the indexing framework and fetch complete content from Table Store

Recommended Solution Architecture



Organizations can deploy the front-end search service using Elastic Compute Service (ECS), Server Load Balancer, and Content Delivery Network (CDN). To build the backend search web-service, one can use Server Load Balancer, ECS, and ApsaraDB for RDS.

The backend search service performs NLP on incremental data and generates a reverse index from the indexing framework. The ECS Input-Output (IO) optimization instance provides high-performance network IO and disk IO, helping the index framework to complete the index operation faster.

Table Store stores the processed data for subsequent searches. The Alibaba Cloud Table Store NoSQL database service is similar to HBase and provides storage and real-time access to massive unstructured data.

When users initiate a search request, the backend search service obtains relevant content through the indexing framework and fetches the complete content from Table Store. The results are then presented to the user through the front web service. The search results use ApsaraDB for Redis to accelerate the turnaround time for the user.

The NLP model constantly analyzes the corpus (linguistic database) on E-MapReduce and provides the latest data model to the search framework.