



AMAZON CLOUDSEARCH

CS 410 : TECHNOLOGY REVIEW

Aravind Pillai
Pillai5@illinois.edu

Table of Contents

1. Introduction.....	2
2. Features.....	2
3. Processing Steps.....	3
3.1 Create a search domain.....	3
3.2 Configure indexing options for the data	3
3.3 Upload data for indexing.....	3
3.4 Submit search requests from website or application.....	4
4. Amazon CloudSearch Pricing.....	4
5. Pros and Cons	4
6. Conclusion.....	5
7. References	5

1. Introduction

Amazon CloudSearch is a fully managed service in the cloud that makes it easy to set up, manage, and scale a search solution. Amazon CloudSearch can search large collections of data such as web pages, document files, forum posts, or product information. CloudSearch makes it possible to search large collections of mostly textual data items called documents to quickly find the best matching results. Search requests are usually a few words of unstructured text. The returned results are ranked with the best matching, or most relevant, items listed first.

As a managed search service, Amazon CloudSearch determines the size and number of search instances required to deliver low latency, high throughput search performance. Amazon CloudSearch also automatically scales to handle increases in the amount of search traffic. When a search instance nears its maximum query load, CloudSearch deploys a replica of the search instance. Conversely, when search traffic drops, Amazon CloudSearch removes unneeded replicas to minimize costs.

2.Features

Amazon CloudSearch provides features to index and search both structured data and plain text, including faceted search, free text search, Boolean search expressions, customizable relevance ranking, query time rank expressions, field weighting, searching, and sorting of results using any field, and text processing options including tokenization, stopwords, stemming and synonyms. It also provides near real-time indexing for document updates.

3.Processing Steps

To use Amazon CloudSearch, we need to follow these steps:

3.1 Create a search domain

We can create Amazon CloudSearch search domain for each collection of data that we want to make searchable. A search domain encapsulates data and the hardware and software resources required to operate a search engine. Each search domain has one or more search instances. A search instance is a server instance that has a finite amount of RAM and CPU resources for indexing data and processing requests. The number of search instances in a domain depends on the documents in the collection and the volume and complexity of search requests.

3.2 Configure indexing options for the data

Each document that are added to the search domain has a collection of fields that contain the data that can be searched or returned. Every document must have a unique document ID and at least one field. We need to define an index field for each of the fields that occur in the documents.

3.3 Upload data for indexing

To make the data searchable, we need to format it in JSON or XML and upload it to search domain for indexing. In most cases, Amazon CloudSearch automatically indexes the data, and the changes are visible in search results in just a few minutes. However, certain changes to your domain configuration put the domain in the “*needs Indexing*” state. For those changes to take effect, we must explicitly run indexing to rebuild the index.

3.4 Submit search requests from website or application

We can submit search requests to the domain's search endpoint as HTTP/HTTPS GET requests. Also, can specify a variety of options to constrain the search, request facet information, control ranking, and specify what you want to be returned in the results. Amazon CloudSearch looks up the search terms in the index and identifies all the documents that match the request. To generate a response, Amazon CloudSearch processes this list of search hits to filter and sort the matching documents and compute facets. Amazon CloudSearch then returns the response in JSON or XML.

4.Amazon CloudSearch Pricing

We need to pay only for what we use. There are no set-up fees or upfront commitments to begin using Amazon CloudSearch. The major portion of a typical domain's costs come from search instance usage. All source documents and updates to the domain are stored behind-the-scenes on Amazon S3 for data durability and recovery, but customers get this for free, which is a significant cost saving over self-managed search infrastructure. Customers are billed according to their monthly usage across Search instances, Document batch uploads, Index Documents requests and Data transfer.

5.Pros and Cons

Amazon CloudSearch provides several benefits including easy configuration, auto scaling for data and traffic, self-healing clusters, and high availability with Multi-AZ. Amazon CloudSearch supports many SDKs along with RESTful API calls. The most popular SDKs are in Java, Ruby, Python, .Net, PHP, and Node.js. Amazon CloudSearch indexes and searches both structured data and plain text. It includes most search features that developers have come to expect from a search engine, such as faceted search, free text search, Boolean search,

customizable relevance ranking, query time rank expressions, field weighting, and sorting of results using any field.

One of the Cons with Amazon CloudSearch is the lack of control on spending. It's very hard to pinpoint how much we will spend here. Since it goes by active searches, the small price quote we get in the beginning will skyrocket if we have more search data for a certain month or have bandwidth issues. There is no way should set a maximum price. Also, the ability to customize the features are minimal and require thorough knowledge on AWS services.

6. Conclusion

Amazon CloudSearch is a complete search solution which will allow you to scale and upload new data and make available to search. With Amazon CloudSearch, one should be able to create their search domain, set search attributes, upload the data, and start testing them out in no time.

7. References

- Amazon CloudSearch Service - <https://aws.amazon.com/cloudsearch/>
- Amazon CloudSearch Developer Guide:
<https://docs.aws.amazon.com/cloudsearch/latest/developerguide/cloudsearch-dg.pdf>
- A step-by-step guide to setting up Amazon Cloud Search:
<https://www.cuelogic.com/blog/a-step-by-step-guide-to-setting-up-amazon-cloud-search-with-examples>
- AWS Cloud Search Choices:
<https://cloudacademy.com/blog/elasticsearch-vs-cloudsearch/>