codurance | CRAFT AT HEART

AWS OpenSearch: What and Why

Ignacio Gonzalez

codurance.com

aws PARTNER Advanced Tier Services

Table of Contents

Introduction

Humans are constantly producing data: applications and systems dumping information and recording issues about health checks of our systems and monitoring our applications. We are now living in a time in which we have tools to metabolise humongous amounts of information in real time, enabling us to take extremely well informed decisions and take actions when necessary for our business/lives. This resource is the introduction to one of those tools which became extremely popular recently because of different reasons: Amazon OpenSearch Service. Full-Text search engine with other capabilities that will unleash the way you and your organisation ingest/process information.

Search Engines

Search engines offer services that allow the user to find information by querying it and retrieving relevant information from its indexes based on the user input.
Different types of search engines exist like web search engines, image, video, news search engines, etc.

A full text search engine is an engine that is designed to retrieve information based on text documents, enabling users to search for documents via words or sentences. Usually, search engines rank results based on the relevance of the information given to them.
The relevance of the results depends on different algorithms and other factors like word proximity, frequency or contextual analysis. They make use of inverted indexes (list of words that points to the document they live in) and tokenization strategies as well as filtering in order to narrow down the scope of the queries.

Amazon OpenSearch Service

Amazon OpenSearch, previously known as Amazon Elasticsearch (do not confuse it with ElasticSearch from Elastic.co, this is another service from another company, and Amazon ElasticSearch is a fork of ElasticSearch from Elastic.co) is a service that allows you to store huge amounts of data and run operations on it. Based on Apache Lucene, it is open source and offers the features of a search and analytics engine often used for full-text search, log analysis, monitoring and data exploration. Data inputs like metrics, logs, traces, etc, can be ingested into OpenSearch service and then that info can be analysed to gain real-time insights.
Thanks to the power of Amazon OpenSearch you can deploy, operate and scale OpenSearch clusters in the cloud on a large scale.

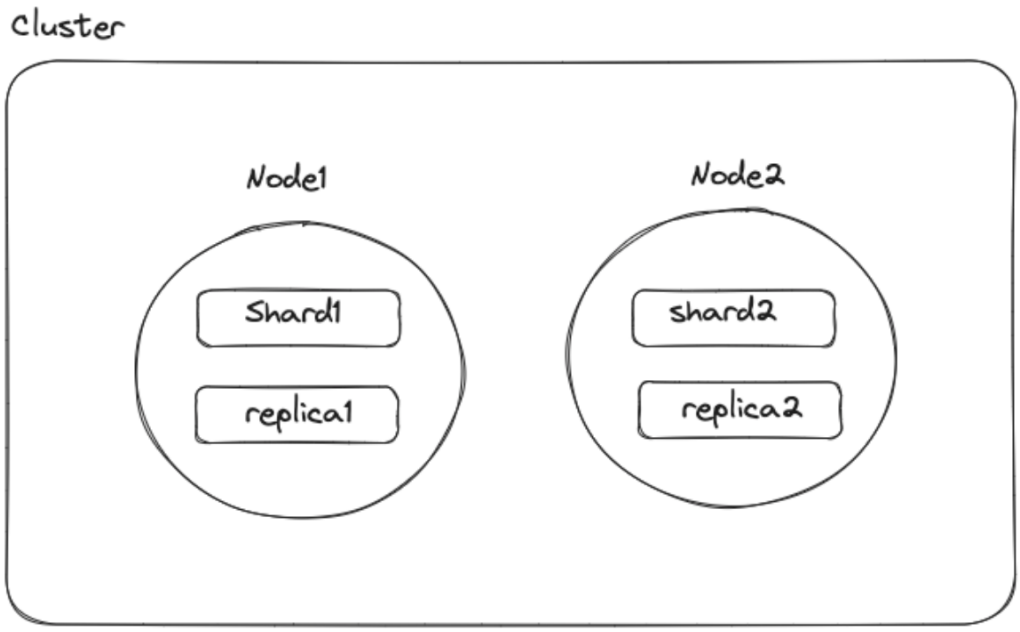
Some of the advantages of this service include:

- **Scalability:** Totally based on your needs. It supports different cluster sizes.
- **Availability**: Fault tolerant via distributing nodes across multiple availability zones.
- **Security:** Provide integration with AWS IAM, for the admin to define roles if necessary.

- Monitoring and Alerting: Integrates with AWS Cloudwatch, allowing the user to monitor different performance metrics, set up alarms and notifications.
- Integration with other AWS services.

Amazon OpenSearch Serverless also offers a service that reduces the time and effort of managing OpenSearch infrastructure allowing you to use a powerful search/analytics tool, but you will lose some degree of control and customisation.

In Amazon OpenSearch we have the following concepts that here are worth mentioning:



- **Cluster:** A cluster is a collection of one or more nodes.
- **Node:** Stores your data and process search requests.
- **Indices:** The way that data is organised is through the use of indices, each index is a collection of JSON documents.
- **Shards:** Indices are split into shards for an even distribution of the data across nodes in a cluster.

"For example, a 400 GB index might be too large for any single node in your cluster to handle, but split into ten shards, each one 40 GB, OpenSearch can distribute the shards across ten nodes and work with each shard individually.
By default, OpenSearch creates a replica shard for each primary shard. If you split your index into ten shards, for example, OpenSearch also creates ten replica shards. These replica shards act as backups in the event of a node failure.

Despite being a piece of an OpenSearch index, each shard is actually a full Lucene index.

This detail is important, though, because each instance of Lucene is a running process that consumes CPU and memory. More shards is not necessarily better. Splitting a 400 GB index into 1,000 shards, for example, would place needless strain on your cluster. A good rule of thumb is to keep shard size between 10–50 GB" Link

Origins of Amazon OpenSearch

Its origins can be traced back to 2010, when Elastic.co created the open source Elasticsearch. Distributed search and analytics engine built on top of Apache Lucene. Designed to provide scalable and efficient full text search engine capabilities and data exploration, among other features.

Recognising how useful it was, Amazon Web Services decided to introduce Amazon Elasticsearch Service in 2015. Providing Elasticsearch service in the cloud, making it easier for users to deploy and operate Elasticsearch clusters.

In 2021, AWS platform announced the launching of Amazon OpenSearch Service, the successor of Amazon Elasticsearch Service. Providing the same functionality while being open source and independent from Elastic.co.

Uses for Amazon OpenSearch Service

Some of the most popular and interesting use cases for this service are features like log analysis, as OpenSearch is well equipped with tools to ingest huge quantities of log records in order to perform searches, visualise and monitor via setting up alarms for anomaly detection.
Linked to this, application monitoring is also available in the form of metric analysis and performance using the ingested data, allowing the users to run health checks and identify issues in their applications.
Search, discovery and visualisation are also well known capabilities in the form of full-text search feature, enabling the users to build search engines for any kind of application allowing clients to search information. These capabilities open the door to other more generic features like business analysis and intelligence, by having access to different view/metrics of your data you can reach informed conclusions and take actions over the domain you are working in.

All of these can be done in real time as OpenSearch is capable of metabolising and analysing in real time.

Within the search and discovery capability, it is worth mentioning a few of the tools that the user can count on, features like the following:

- Full-text search
- Filtering
- Faceted Search
- Autocomplete
- Relevance scoring

Regarding visualisation, note that Amazon OpenSearch does not have Kibana but a similar tool called OpenSearch Dashboards which basically pursue the same purpose.

Scalability and Reliability

Regarding capabilities like scalability, It is worth mentioning elastic scaling, which is the ability to scale your cluster depending on the amount of work that it is holding. This 'elasticity' enables the users to adapt to different quantities of traffic in different contexts.

Amazon OpenSearch follows a strategy in which it stores information in different shards inside a cluster, this improves performance at the same time of automating the process due to the fact it is a managed service: Amazon OpenSearch takes care of infrastructure provisioning, configuration, and maintenance.

On the reliability side, data durability and availability shine as Amazon OpenSearch makes sure automated data backup and replication by storing data across different nodes and replicating information avoiding data loss. Monitoring and alarms are two other options related to reliability that work extremely well with purposes like anomaly detection or notifications.
Amazon OpenSearch also offers a fully managed service which means AWS takes care of the infrastructure such as patching, upgrades, hardware maintenance, etc.

Multi-AZ Deployment is another popular characteristic of multiple AWS services, allowing the users to deploy clusters in different Availability Zones (AZ), improving fault tolerance and avoiding disruptions. This, combined with other features like automatic backups that can be stored in S3 buckets, provide protection for your data.

Last but not least, AWS support, documentation and forums where you can find information about troubleshooting and assistance.

Cost and Usage Considerations

This service follows a pay-as-you-go scheme that depends on the type, storage size and the number of instances. The instances need to be well considered before picking as they need to be refined to get maximum performance/efficiency considering variables like CPU, memory and network workload. For the storage options, EBS and S3 are available. Usually, it is suggested to use EBS for low latency whereas S3 provides better cost-effective capacity for less frequent access data.

AWS also charges for data transfer, so considering the volume of data to be ingested is something important: only pay for the resources consumed by your workload. OpenSearch Ingestion charges for only the computer needed to ingest, transform, and route data in an OpenSearch Ingestion pipeline.

Reserved instances are available for one or three years which could save money during that period of time depending on your strategy compared with On-Demand instances. It is also worth mentioning UltraWarm and cold storage pricing. UltraWarm allows you to keep large amounts of data while still being efficient and cold storage is a lower cost storage option for more infrequently accessed data in AWS S3 and you pay for compute only when needed.

AWS also offers a free tier which is composed of 750 hours per month for a ‘t2.small.search’ or a ‘t3.small.search’ instance typically used for testing as well as 10 GB month of EBS. For more information you can refer to this link.

Security and Compliance

Compliance and security are extremely popular topics in the in the world of business. Nobody can forget about these two without having issues down the line when interacting with other entities, even in the same country or industrial context.

Regarding data protection, IAM is the AWS service that allows you to define accesses in a granular way and this service is deeply related and interacts with Amazon OpenSearch in order to make sure only authorised personnel can access the platform. Another possibility is isolating your cluster in your own private network and controlling its traffic via Virtual Private Cloud (VPC) creating security groups.

Amazon OpenSearch also supports encryption at rest and transit. At rest means that you can enable encryption for your data using AWS Key Management Service. In transit, you can use SSL/TLS encryption to secure communication between clients and clusters.

Something that is also important is data retention can be customised for different periods for your backups.

At the same time, compliance and auditing are available via integration with other services like AWS CloudTrail or built in features like access logs giving away information about usage inside clusters. CloudWatch also can be configured to receive trails/logs providing detailed information about the actions taken during the use of the service.

Amazon OpenSearch is also compliant with regulations like GDPR, HIPAA and ISO 27001, helping it achieve specific industry requirements for data privacy. Demonstrating AWS has reached a high level of security control over its infrastructures.

Hypothetical case of use

Imagine a retail e-commerce company that needs to enhance its product search and recommendation system because the one in place is slow and does not provide the most accurate information.

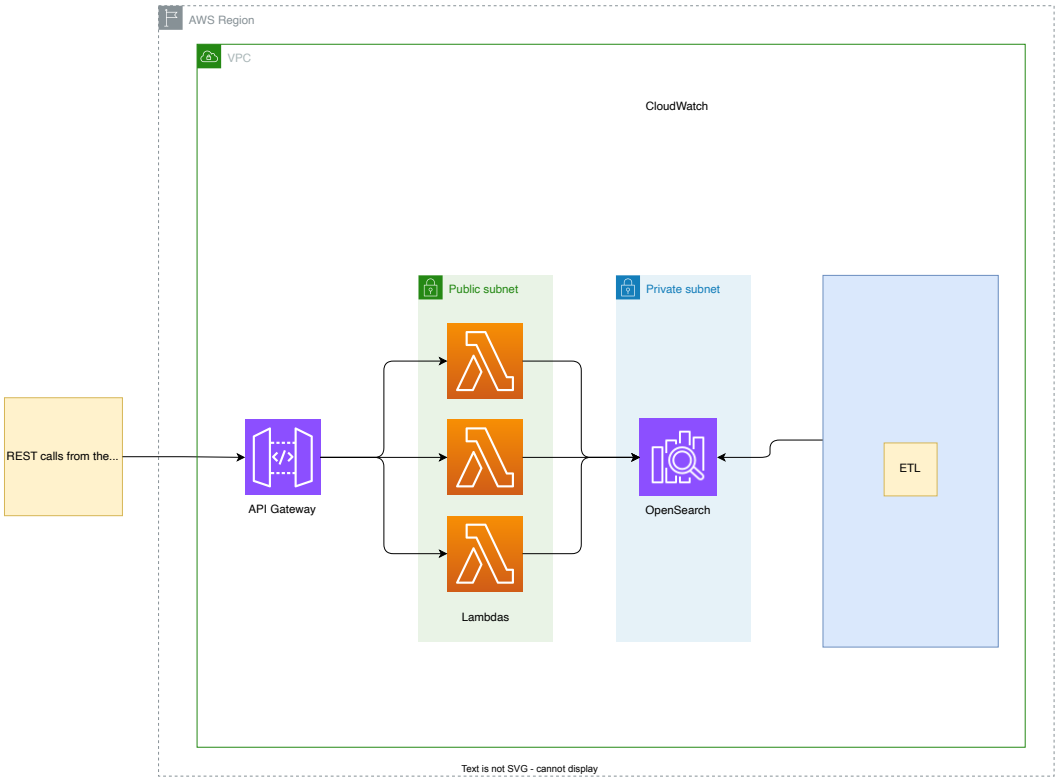
In this particular context, Amazon OpenSearch would be helpful because of different reasons, for starters the company would be able to index their products in their catalogue and customers would be able to perform queries and apply filters as well as sort results. Amazon OpenSearch offers the features of a full-text search engine as well as personal recommendations by analysing the customer browsing and purchasing history along with product attributes.

We talked already about scalability and reliability, allowing the company catalogue to grow at the same time that the OpenSearch instance can adapt to it.

Another good reason is the alerting and monitoring capabilities that are offered, being able to set up alarms and notifications when anomalies/issues happen.

Generic architecture for the hypothetical use-case

The diagram shows the recommended architecture for an AWS OpenSearch implementation for the use case described above. The use of OpenSearch as a logging and monitoring platform (as an alternative to AWS CloudWatch) is a different use case and a different architecture must be considered.



In the use-case under consideration, the one which uses OpenSearch as a search engine exposed to the final users, there are these main components to consider:

- AWS OpenSearch as the search engine
- An ETL process to feed the search engine indexes with the data to be searched
- A front-end architecture to expose the search functionality to the users' browsers
- A logging and monitoring solution.

The OpenSearch engine is in the middle and must be deployed in a network with these characteristics:

- A private subnet: the search engine must not be deployed in a public subnet, as per any DB that can be deployed in AWS
- OpenSearch must be open to receive calls from the front end and the ETL component.

The ETL process designed in this diagram is a very generic placeholder because every company has its own network of back-end components. The ETL component must take care of inserting, updating and deleting the documents inside the OpenSearch indexes by calling the indexing APIs. This component is represented as part of the OpenSearch VPC, but with the right network configurations can be deployed in another VPC.

The searches done by the final users must be processed ideally in milliseconds, and this is the right scenario for the Serverless components: an API Gateway and one or more Lambda functions are the right choices. This stack will call the search function APIs of OpenSearch.

AWS CloudWatch can be used as an out-of-the-box solution for logging and monitoring.

Conclusion

We are constantly on the lookout to improve our systems and Amazon OpenSearch is a really powerful tool that anybody can leverage to gain valuable insights and make sense of your data. It offers different capabilities like real-time analysis, complex aggregations and queries as well as strategies to monitor the health of your system.

Scalability and performance are really important in any system, the distributed nature of OpenSearch allows it to handle large amounts of data flawlessly, able to adapt dynamically to increasing amount of information ingested. Amazon OpenSearch supports different data formats allowing users to collect data from different sources and has an insane amount of compatibility with other AWS Services.

With a thriving community, more and more support can be found apart from the one that is already available, also another positive point to take into consideration when assessing tools like this. Amazon OpenSearch will empower the way you analyse and ingest data in your systems.

codurance | CRAFT AT HEART

Software | People | Process | Community

hello@codurance.com @codurance codurance.com