

OpenSearch Beginner

Duration: 2 days (8hrs/day)

Prerequisites:

- Understand the basics of containerization.
- Participants should be comfortable using a terminal/command line.

Course Objective: This comprehensive course will help you understand all core OpenSearch concepts – index, document, sharding, replication, mapping, search relevance, etc. Be able to index data into OpenSearch and retrieve it using search and realtime get APIs. Have a solid grasp of the underlying query parsing, analysis, tokenization, and various types of queries. Learn about a number of different types of OpenSearch aggregations.

Module 1 – Data Flow in OpenSearch

Overview of OpenSearch and its applications

Concepts of indexing, IDs, mappings, and stored fields

Basics of analysis and real-time data retrieval

Distribution of searches across shards

Introduction to aggregations and document values

Versioning and deletions

Lab: CRUD operations

Lab: Query and filter

Lab: Aggregations

Module 2 – Indexing and Data Storage

Using the Bulk API

Understanding mappings and types

Defining subfields and default mappings

Core data types like text, keywords, and integers

Separating stored fields from the source

Lab: Using the bulk API

Lab: Changing mapping

Module 3 – Text Analysis

Built-in and custom analyzers

Char filters, tokenizers, and token filters

Utilizing the Analyze API

Lab: Add stemming support

Lab: Add support for non-ASCII characters

Module 4 – Data Search

Field selection, sorting, and pagination

Basics of term, range, and boolean queries

Main options of match and query string queries

Lab: Configure sorting, pagination and select the right fields

Lab: Using a bool query to combine different match, range and term queries

Module 5 – Aggregations

Metrics aggregations: stats, cardinality, percentiles

Understanding terms, cardinality, and percentiles as approximations

Multi-bucket aggregations and nesting

Lab: Computing the cardinality of a field

Lab: Sorting buckets by results of sub-aggregations

Lab: Nest the sum and histogram aggregations

Module 6 – Clustering Essentials

Nodes, shards, and replicas

Replication and distributed search mechanisms

RAM and heap size considerations

Cluster bootstrapping and Cat APIs

Lab: Create an index

Lab: Verify the distribution of shards

Lab: Add a new node to the cluster