Elasticsearch

Duration: 4 Days (8 hrs/day)
Version: Elasticsearch 8.x
Lab Requirement: 3-node Cluster (CentOS 8)
Prerequisites: Basic Linux Knowledge

Course Objective

This course builds core competency in deploying, managing, and optimizing Elasticsearch clusters. Participants will gain hands-on experience performing data ingestion, search, aggregation, lifecycle management, and cross-cluster operations, following best practices aligned with the Elastic Certified Engineer 8.x curriculum.

Module 1 – Getting Started

Topics:

- Overview of Elasticsearch and the Elastic Stack
- Understanding Elastic Stack components: Elasticsearch, Kibana, Beats, Logstash
- Use cases and architecture overview
- Installation and deployment (self-managed and Elastic Cloud)
- Elasticsearch data model: Index, Document, Node, Cluster
- Basic REST operations and CRUD workflow
- Using Kibana for data exploration

Labs:

- Deploy Elasticsearch and Kibana
- Perform CRUD operations via REST API
- Load and view sample data in Kibana
- Create a single-node cluster and verify connectivity

Module 2 – Data Modeling

Topics:

- Understanding fields, mappings, and data types
- Keyword vs Text fields: purpose and behavior
- Creating and updating index mappings
- Custom analyzers and tokenizer configuration
- Dynamic mappings and templates
- Specialized data types (geo_point, ip, range, etc.)

Labs:

- Define custom mappings for structured and unstructured data
- Create and manage dynamic index templates
- Test analyzers using _analyze API
- Create and reindex data with updated mappings

Module 3 – Search

Topics:

- Introduction to Query DSL (Domain Specific Language)
- Structure of search requests and responses
- Full-text vs term-level queries
- Range queries for numeric, date, and IP fields
- Boolean queries and compound logging
- Document scoring and relevance (_score, BM25)
- Pagination, sorting, and highlighting results

Labs:

- Create and execute match, term, and range queries
- Combine queries using Boolean logic
- Analyze document scoring using _explain API
- Implement sorting and pagination
- Highlight search results in Kibana Dev Tools

Module 4 – Aggregations

Topics:

- Understanding aggregation framework
  Metric aggregations: avg, sum, min, max, stats
- Bucket aggregations: terms, range, histogram, date_histogram
- Sub-aggregations and nested aggregations
- Pipeline aggregations: avg_bucket, derivative, moving_avg
- Using transforms to create summary indices

Labs:

- Create metric and bucket aggregations
- Group and analyze data by category and date
- Build nested aggregations for multi-level insights
- Create pipeline aggregations for trend analysis
- Build a transform for summarized analytics

KOENIG
step forward

Module 5 – Data Processing

Topics:

- Ingest pipelines and processors
- Transforming and enriching documents
- Using set, rename, grok, and geoip processors
- Creating and executing enrich policies
- Runtime fields and computed data
- Writing Painless scripts for dynamic calculations

Labs:

- Create and simulate ingest pipelines
- Enrich data with category metadata
- Use Grok to parse unstructured data
- Add runtime fields using Painless scripts
- Create and test enrich pipelines

Module 6 – Distributed Datastore

Topics:

- Shard and replica architecture
- Write and search paths (internal flow)
- Cluster coordination and shard communication
- Scaling clusters up/down
- Shard balancing and reallocation
- Performance optimization for search and indexing

Labs:

- Create multi-shard, multi-replica index
- View shard allocation using _cat/shards
- Simulate node failure and observe recovery
- Scale replica count and monitor balancing
- Manually reroute shards

Module 7 – Data Management

Topics:

- Index templates and patterns
- Data streams for time-series data

- Configuring data tiers: hot, warm, cold, frozen
- Index Lifecycle Management (ILM) policies
- Snapshot and restore operations
- Searchable snapshots for long-term retention

Labs:

- Create index templates and data streams
- Configure tier-based index allocation
- Define and attach ILM policies
- Perform index rollover
- Create, verify, and restore snapshots

Module 8 – Cluster Management

Topics:

- Cluster roles, state, and configuration
- Monitoring cluster health and performance
- Diagnosing cluster issues (allocation, balancing)
- Cross-cluster search (CCS) setup
- Cross-cluster replication (CCR) configuration
- Security and role-based access (brief overview)

Labs:

- Monitor cluster health and node status
- Configure and test cross-cluster search
- Set up cross-cluster replication (leader–follower model)
- Analyze cluster stats and node metrics
- Troubleshoot unassigned shards