Architecting the OpenSearch service at CERN

Sokratis Papadopoulos
it-opensearch-experts@cern.ch
Overview

• Introduction
• Motivation for change
• The OpenSearch service architecture
• Migration process
• Service usage at CERN
• Roadmap
What is Elasticsearch and OpenSearch?

- **Elasticsearch** is a distributed, search and analytics engine based on Apache Lucene
- **Kibana** is the web user interface that lets you visualise your Elasticsearch data

- **OpenSearch** is a [fork](#) of Elasticsearch 7.10.2 open source codebase
- **OpenSearch Dashboards** is the [fork](#) of Kibana 7.10.2 open source codebase
Legacy Elasticsearch service design at CERN

Download from artifacts.elastic.co

- Elasticsearch 7.1.1 binary package
  - Apache 2.0 licensed
- Kibana 7.1.1 binary package
  - Provided by Elastic

Open-source plugins

- ReadOnlyRest
- OwnHome
- Curator
- elasticsearch-status
- elasticsearch-ingest
- elasticsearch-analysis-icu
- elasticsearch-repository-s3
OpenSearch service design

Apache 2.0 licensed core + plugins
- Security (SearchGuard)
- Anomaly Detection
- Alerting
- SQL
- k-NN
- Index Management
- ...

OpenSearch +

Apache 2.0 licensed tools
- PerfTop
- Perf. Analyzer
- sql-cli
- ...

OpenSearch Dashboards +

Notebooks
- Trace Analytics
- Visualizations
- SQL Workbench
- Reports
- Security UI
- ...

OSS Community + AWS
Motivation for change

- **Licensing**
  - As of v7.10.2 Elastic no longer provides Apache 2.0 releases
  - OpenSearch is licensed under Apache 2.0

- **Maintainability**

- **Streamlined deployment**

- **Customers isolation**

- **Features**
  - Many native plugins (alerting, index-management, etc.)
  - Fine-grained security access control
The OpenSearch service architecture
Elasticsearch to OpenSearch migration

1. Index templates
2. Index retention policies
3. Reindex data
4. Kibana objects (patterns/viz/dashboards)

5. Live data

6. Finalize config on new cluster (e.g., ACLs, landing page)
7. Decommission old endpoint
8. Move alias to new cluster
Number of clusters per version

![Graph showing the number of clusters per version from 2021-07-01 to 2023-04-01. The graph compares ElasticSearch v6, ElasticSearch v7, OpenDistro, OpenSearch v1, and OpenSearch v2. The data shows a notable increase in clusters for OpenDistro and a decrease for OpenSearch v1.]
Lessons learned and challenges emerged

- OpenSearch integration with CERN internal tools was quite easy
- Upstream puppet module does not support multiple instances
- Elastic burning bridges with OpenSearch
  - Some adjustments were needed on user side clients (e.g. logstash, filebeat, python, etc.)
- Users side engagement
  - Maintainers have left the organization
  - Deprioritizing migration
- Maintaining a service on 5 different major versions at a time
- Providing dedicated clusters now, users must respect their quotas
Service usage at CERN

- ALICE, ATLAS, CMS, LHCb, NA62, …
- Beams, INSPIRE, Zenodo, …
- IT Monitoring, IT Security, IT Storage, …

350 TB on SSD and 150 TB of CEPH storage
Roadmap

• Complete OpenDistro **migration** to OpenSearch and jump to **ALMA9**
• **Automate** further cluster bootstrapping and interventions
• Engage more with the OpenSearch **community**
• Explore **data streams** functionality for append-only logs data
• Explore OpenSearch plugins and functionalities (Anomaly Detection, Observability, Snapshots, …)
• Evaluate **logstash** and **beats** alternatives
Summary

• A service with growing interest for the last 7 years
• OpenSearch brought significant changes both internally and on user side
• New service fully operational on physical machines
• OpenSearch migration is 60% completed
• Plethora of opportunities to further enhance the service