Elasticsearch Ops

Duration 2 day(s) (ES-OPS-02)

Administer and operate Elasticsearch

Description

ElasticSearch is a new generation Open Source search and indexing engine. Based on the Apache Lucene library, this search engine offers advanced features such as geographic coordinates searches, facet analysis and categorization, result filtering, and searching on several different indexes and document types. Designed for the Cloud, ElasticSearch has been specifically designed to index very large volumes of data while ensuring high performance scalability and high fault tolerance.

Goals

- Understand the role of Elasticsearch
- Dimension and install an Elasticsearch cluster
- Maintain an Elasticsearch cluster: backup (snapshot / restore), upgrades ...
- Monitor an Elasticsearch cluster and detect problems
- Configure Elasticsearch for optimal performance
- Know the good and bad practices

Public

- Ops
- Architect
- Operator

Prerequisites

- Have knowledge of what are REST / HTTP
- Have knowledge of Json, Yaml format
- To have knowledge on the tools used conventionally in exploitation of SI (Graphite, Patrol, Nagios, ...)
- Have knowledge of the tools conventionally used in performance tests
- Have minimum security knowledge: PKI, SSL, ACLs
- **Required:** Have knowledge of Linux and command lines including Curl, wget ...
- **Required:** Have knowledge of Linux administration (swap, systemctl, ulimit, network)
- **Mandatory:** Have Java administrative skills (Memory, classic options)

Non-compulsory knowledge greatly reduces practical work when it is all to be acquired.

Structure

60% Theory, 40% Practice

Program

Introduction to Elasticsearch

- From analysis to research
- Index, documents, nodes, shards ...: the concepts
- Analysis and indexing
- Research and aggregations
- Lab 1: Installation 1 node, indexations and research (CRUD)

Sizing and installation

- Prepared
- Clustering, Zen Discovery, Split brain
- CPU: competition, thread pools
- Memory: Java heap and garbage collector, cutout, swap
- Network: multicast / unicast, firewall
- Disk: organization of the file system, the flush (translog)
- Sizing method
- Lab 2: ES cluster installation of 3 nodes

Monitoring and surveillance

- Cluster Health API, Cat ...
- Logs and slow query logging
- Metrics to monitor
- Lab 3: Cluster / Node stats, Cat, Marvel

Day-to-day maintenance

- Index aliases
- Configuration of the allocation
- Shard / node assignment (Cluster Route API)
- Restart: warmers
- Backups: snapshot and restore
- Upgrades
- Commission / decommission a machine
- Curator
- Lab 4: Backup and restore, Remove / add a node ...

Security

- Authentication and permissions
- Scripting
- Shield, reverse proxy

Performance and optimization

- Indexation:
  - Bulk API
  - Alias
- Research:
  - filtering, post filtering and cache
  - wildcard, parent / child
  - Aggregations and sorts: doc_values and fielddata