Elasticsearch Ops

Durée 2 jour(s) (ES-OPS-02)

Administrer et exploiter Elasticsearch

Description

ElasticSearch est un moteur de recherche et d'indexation Open Source nouvelle génération. Basé sur la librairie Apache Lucene, ce moteur de recherche offre des fonctionnalités avancées telles que les recherches par coordonnées géographiques, l'analyse et la catégorisation par facettes, le filtrage de résultats ou encore la recherche sur plusieurs index et types de documents différents. Taillé pour le Cloud, ElasticSearch a été spécialement conçu pour indexer de très gros volumes de données tout en assurant une montée en charge performante et une forte tolérance aux pannes.

Objectifs

- Comprendre le rôle d'Elasticsearch
- Dimensionner et installer un cluster Elasticsearch
- Maintenir un cluster Elasticsearch: backup (snapshot/restore), upgrades...
- Monitorer un cluster Elasticsearch et détecter les problèmes
- Configurer Elasticsearch pour obtenir des performances optimales
- Connaitre les bonnes et mauvaises pratiques

Public

- Ops
- Architecte
- Exploitant

Prérequis

- Avoir des connaissances de ce que sont REST/HTTP
- Avoir des connaissances du format Json, Yaml
- Avoir des connaissances sur les outils utilisés classiquement en exploitation de SI (Graphite, Patrol, Nagios, ...)
- Avoir des connaissances sur les outils utilisés classiquement en tests de performance
- Avoir des connaissances minimales en sécurité: PKI, SSL, ACLs
- **Obligatoire:** Avoir des connaissances de Linux et des lignes de commandes et notamment Curl, wget ...
- **Obligatoire:** Avoir des connaissances d'administration Linux (swap, systemctl, ulimit,réseau)
- **Obligatoire:** Avoir des connaissances d'administration Java (Mémoire, options classiques)

Les connaissances non obligatoires ralentissent fortement les travaux pratiques lorsqu'elles sont toutes à acquérir.

Répartition

60% Théorie, 40% Pratique

Programme

Introduction à Elasticsearch

- De l'analyse à la recherche
- Index, documents, noeuds, shards...: les concepts
- L'analyse et l'indexation
- La recherche et les agrégations
- TP1 Installation 1 noeud, indexations et recherche (CRUD)

Dimensionnement et installation

- Préquis
- Clustering, Zen Discovery, Split brain
- CPU : concurrence, thread pools
- Mémoire: heap Java et le garbage collector, coupe-circuit, swap
- Réseau: multicast/unicast, firewall
- Disque: organisation du système de fichiers, le flush (translog)
- Méthode de dimensionnement
- TP2 Installation cluster ES de 3 noeuds

Monitoring et surveillance

- API Cluster Health, Cat...
- Les logs et le slow query logging
- Les métriques à monitorer
- TP3 Cluster/Node stats, Cat, Marvel

Maintenance au jour le jour

- Alias d'index
- Configuration de l'allocation
- Affectation shard/noeud (API Cluster Route)
- Redémarrage: warmers
- Backups: snapshot et restore
- Upgrades
- Commissionner/décommissionner une machine
- Curator
- TP4 Backup et restore, Retirer/ajouter un noeud...

Sécurité

- Authentification et autorisations
- Scripting
- Shield, reverse proxy

Performances et optimisation

- Indexation:
  - API Bulk
  - Alias
- Recherche:
  - filtering, post filtering et cache
  - wildcard, parent/child
  - Agrégations et tris: doc_values et fielddata