Elles ressemblent à quoi mes données ?

Date : 8 avril 2015
Format : Tools in action
Speakers : David Pilato et Colin Surprenant, développeurs Logstash et Elasticsearch chez Elastic

David se met dans la peau d'un développeur a qui son boss donne 3 semaines pour exploiter des données.

Besoin métier :

1. Qui sont les clients ?
2. Ce qu'ils pensent de Twitter
3. Comment ils utilisent nos services ?

Les données viennent de :

- Notre application
- Base de données CRM
- Twitter
- Logs de production

Tools in action en 3 démos.

Démo 1

Extraire de l'information depuis la base Postgres du CRM.
Requêtes SQL via Hibernate puis injection dans Elasticsearch.
Création d'un dashboard Kibana.

Etapes :

- Cluster Elasticsearch démarré
- Configuration d'un TransportClient d'Elasticsearch
- Utilisation d'un BulkProcessor et implémentation du Listener avec des logs de debug ou d'erreur
- Lecture des personnes avec Hibernate
- Utilisation d'un mapper Jackson pour indexer en JSON les personnes dans ES
- Utilisation de Marvel pour requêter l'index ES
- Depuis Kibana, ajout d'un Camembert : agrégation homme/femme puis sous-répartition par pays, puis par ville

Démo 2

Extraction des données depuis les logs avec Logstash.
Sauvegarder les données dans Elasticsearch ?
Les logs viennent de Apache, Nginx et Postgres
Consultation dans un dashboard Kibana

Logstash transporte des données d'une source vers une destination. Tourne en continue (streaming) ou utilisable au besoin.
Démo réalisée avec Logstash 1.5.0-rc2 et avec les logs de download d'Elasticsearch.

Etapes :

- Utilisation de logs Nginx formatter en json
- 3 étapes de logstash : input, filter et output
- Codec json_lines
- Hétérogénéité des dates uniformisée avec le filtre date
- Enrichissement de l'événement par géolocalisation en fonction de l'adresse IP
- Utilisation du filtre useragent pour connaître l'OS et le navigateur des utilisateurs
- Les 5 000 000 lignes de log sont indexés dans ES
- Kibana : carte du monde ou de France

Démo 3

Utilisation de Logstash pour indexer des tweets.
Démo réalisée par Colin en 4mn.

Etapes :

- Configuration de Logstash pour se connecter à Twitter
- Utilisation du plugin Logstash dot pour afficher des points ?
- Suppression dans Kibana des tweets qui n'ont aucun rapport avec Elasticsearch