Ingest node : (ré)indexer et enrichir ses documents dans Elasticsearch

Speakers : David Pilato (Elastic)
Format : Tool in action
Date : 6 avril 2017

Développeur dans l'équipe Elasticsearch, en particulier sur les plugins.
Suite Open Source + outils commerciaux

Pourquoi Ingest Node ?
Parti du besoin utilisateur « I just want to tail a file ».
Grosse usine à gaz-
Pour simplifier l'architecture, l'agent beats peut envoyer les lignes de log dans Elasticseach.
Ingestnode va permettre de structurer la ligne de log afin de l'indexer.
On passe de Logstash à une partie directement gérée dans ES.
Utilisation des pipelines afin de retravailler les données à indexer.

Plugin ingest-bano en cours de développement
Objectif : transformer des adresses postales en coordonnées géographiques, et inversement.
Utilisation de la base de données d'adresse Bano gérée par Open Street Map.

Démo :
ES et Kibana démarrés en local.
Dans ES, plugin ingest-bano installé.
Utilisation de l'API *simulate en lui passant la liste des processeurs à appliquer et les documents.*
Ajout d'un processeur grok afin de parser le log.
Ajout d'un pipeline pour créer un champs timestamp de type date. Le champs @timestamp a été ajouté au document.
Suppression du champs message et de l'ancienne date.
Conversion des champs response et bytes en entier.
Gestion des exceptions prévues dans le pipeline : on envoie le document en erreur (ex : date invalide) dans un autre index nommé « error-index ».
Possibilité d'avoir une gestion des erreurs pour chaque processeur. Exemple : date par défaut en cas d'erreur de parsing du timestamp.

Des départements français sont indexés à partir de la base Bano.
Le plugin ingest-bano, via un processor bano, enrichit une indexation d'adresse en ajoutant lattitude + longitude et en normalisant l'adresse.
Utilisation du preprocessor geoip pour récupérer les coordonnées géographique d'une adresse IP.
Geoip reconnaît la ville, le pays, longitude et latitude. Bano renvoie l'adresse la plus proche de la position géographique.

Création d'un index contenant des documents de type Personne créés aléatoirement.
Création d'un pipelne bano-person afin d'enrichir l'adresse.

Création de l’API ES re-index pour lire les personnes d’un index, les réécrire dans un autre index tout en passant par le pipeline d’enrichissement d’adresse.