Insper

PROJETO 2 -
Estudo de Rotas de Táxi em NY

Megadados
- 04 de Novembro de 2019 -

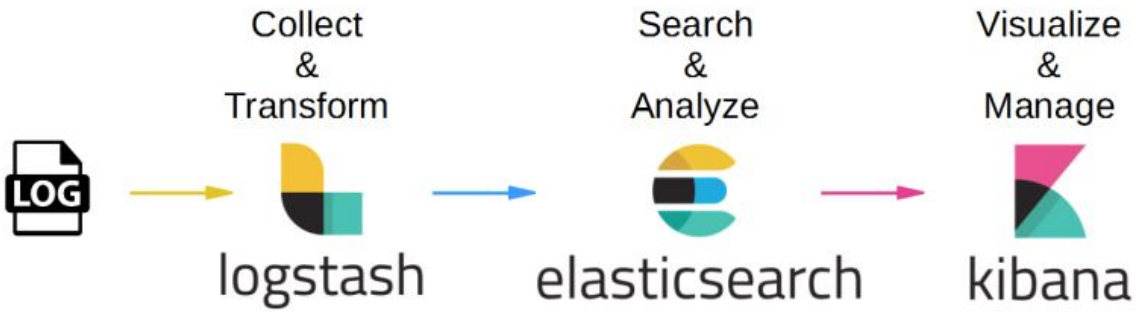
João Pedro Castro
Rachel Bottino

ÍNDICE

ELASTIC STACK

ELK, da Elastic, é a junção das ferramentas: Elasticsearch, Logstash e Kibana. Com ela é possível filtrar logs de maneira muito eficiente, eliminando o que não interessa, utilizando a engine para decodificar os dados (XML, Json, Multilines, Netflow, entre outros) e acessá-los em um único lugar.

Seu funcionamento é basicamente o seguinte: O **Logstash** coleta e recebe os logs de distintas fontes, realiza as transformações, normaliza, agrupa e indexa no **Elasticsearch** que faz buscas de maneira eficiente. Por fim, os dados podem ser visualizados na interface montada pelo **Kibana**.



Essa tecnologia pode ser usada para **análise de negócios**, **análise de logs**, **análise de segurança**, **análise de métricas**, **buscas de sites**, **buscas de empresas**, entre outras. Algumas das empresas que utilizam essas ferramentas são: Netshoes, Telefonica, Adobe, Ebay, BBC, Netflix e Blackboard.

1. Instalação ElasticSearch

a. Faça o download do Elasticsearch;
b. Descompacte o arquivo;
c. Rode bin/elasticsearch.bat na linha de comando;

2. Instalação Kibana

a. Faça o download do Kibana;
b. Descompacte o arquivo;
c. Abra config/kibana.yml em um editor de texto;
d. Rode bin/kibana.bat;
e. Acesse o dashboard do kibana

3. Instalação Logstash

a. Faça o download do Logstash;
b. Crie um arquivo **logstash-csv.conf** com as seguintes configurações:

```
input {
  file {
    path => "C:\Users\rache\Documents\INSPER\20192\Megadados\Projeto2\2017_Yellow_Taxi_Trip_Data.csv"
    start_position => "beginning"
    sincedb_path => "nul"
  }
}
filter {
  csv {
    separator => ","
    columns => ["VendorID","tpep_pickup_datetime","tpep_dropoff_datetime","passenger_count","trip_distance","RatecodeID","s
    tore_and_fwd_flag","PULocationID","DOLocationID","payment_type","fare_amount","extra","mta_tax","tip_amount","tolls_amo
    unt","improvement_surcharge","total_amount"]
  }
}
output {
   elasticsearch {
     hosts => "http://localhost:9200"
     index => "ny-taxi"
  }
stdout {}
}
```

4. Injetando dados via Logstash

a. Abra uma nova janela de comando;
b. Rode bin/logstash -f logstash-csv.conf
c. Aguarde

Dependendo do tamanho do banco de dados, esse processo pode demorar um pouco. Porém, a vantagem do ELK Stack, é que isso é feito apenas uma única vez, e depois disso é muito mais rápido e fácil ter acesso e trabalhar com seus dados.

5. Visualizando seus dados no Kibana

Quando a etapa anterior foi concluída, você finalmente pode encontrar seu índice ny-taxi no dashboard do Kibana