

# Automatisation infrastructure de donnée

CILINGIR Elif  
HUANG Chaofu  
MZABI Hakim  
HUCHARD Théo

# Sommaire

**01**

**Présentation du projet**

**02**

**Récupération de donnée**

**03**

**Transformations Spark**

**04**

**Architecture**

**05**

**Data visualisation**

# Présentation du projet





# Présentation du projet

Objectif :

- Permettre à notre client de voir en avant les artistes suivants :
  - Nekfeu
  - JUL
  - Petit Biscuit
- Voir les dernières tendances musicales afin d'aider à la décision des artistes
- Automatiser le workflow pour faciliter la prise de décision

# Récupération de donnée

---





# Récupération de donnée

Pour la récupération des données nous avons créé un compte Spotify développeur.

Puis, nous avons récupérer les données en Python à l'aide de l'API.

Par la suite nous allons garder les données pour tout ce qui est artiste, album....

Ces données sont ensuite envoyées dans le HDFS.

# Transformations

## Spark



# Transformations Spark



## Etape 02

Faire ressortir les meilleurs artistes



## Etape 01

Récupération des données

## Etape 03

Les classer par popularité

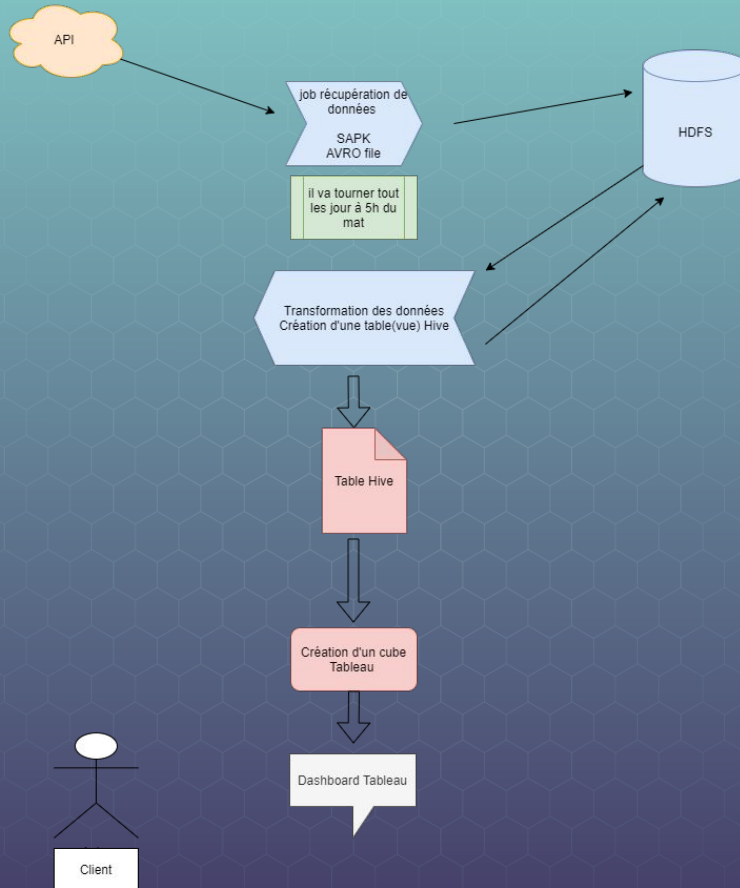


# Architecture

---



# Architecture



# Data visualisation

---

Concernant la data visualisation nous avons  
essayé d'installer la dernière version de Tableau  
Server sur le edge, sans succès :

```
[datagang@79663291-27f5-4fb7-ac28-634c79cec301 /]$ sudo mkdir tableau-server  
[datagang@79663291-27f5-4fb7-ac28-634c79cec301 /]$ sudo rpm -i --prefix /tableau-server ~/tableau-server.rpm  
installing package tableau-server-20203.20.1110.1623-20203-20.1110.1623.x86_64 needs 2375MB on the / filesystem
```



# Data visualisation

On a alors également essayé une solution Cloud, mais la configuration minimale de Tableau Server nécessitait un certain coût :

Estimation mensuelle : 504,37 \$

Soit un coût horaire d'environ 0,691 \$

Vous payez ce que vous consommez : facturation à la seconde, sans frais initiaux à supporter



**Merci à vous**