



POC Nougatine

Kamel MALKI, Hedi BOUFADEN, Ilès BENKOUSSA, Cécile COUSIN

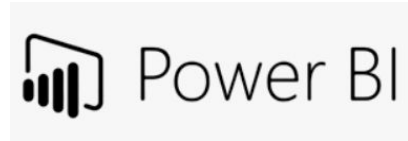


Problématique

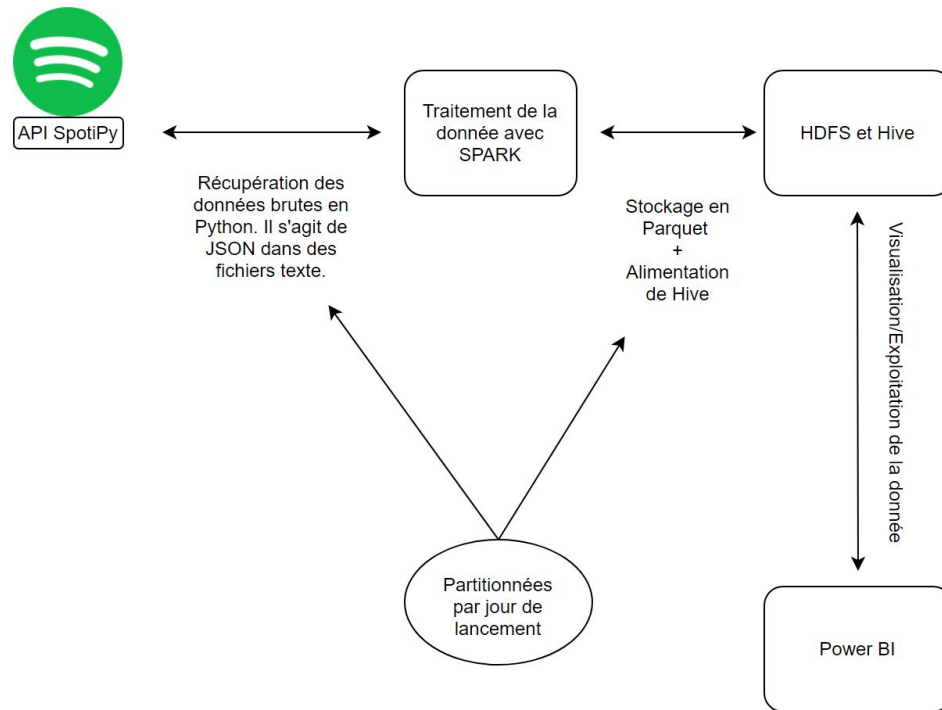
- Anticipation des styles de musiques les plus populaires sur Spotify
- Chaîne de traitement de données complète

Technologies utilisées (et pourquoi)

- Python (bibliothèque SpotiPy)
- Spark/Scala
- Airflow
- HDFS/Hive
- PowerBI



Pipeline



Planification avec
Airflow



Démo



Mises en places futures?

- Tests unitaires à finaliser
- Automatisation, gérer le problème d'accès à Hive (dernière étape du process)
- Ajouter de nouvelles sources de données telles que Deezer et Apple Music

Des questions?
