



PROJET EXAM BIGDATA

Groupe 3 :
Badr MOURAFIQ
Harouna Ba
Lucienne Sayou
Nour Ben Rejeb



estiam





PRÉSENTATION DU PROJET

1

CONTEXTE

Aujourd'hui, nous vivons dans un monde où chaque clic, recherche ou écoute génère des données. Imaginez que chaque chanson que vous écoutez, chaque artiste que vous découvrez ou chaque playlist que vous créez sur Spotify contribue à un océan d'informations. Ce flot constant de données, est au cœur de ce que l'on appelle le Big Data.

2

PRÉSENTATION

Dans notre projet, nous avons exploité l'API Spotify pour collecter des données musicales en temps réel. L'objectif étant d'identifier les artistes populaires, analyser les genres tendance, et observer l'impact de la musique selon le temps et le lieu. En utilisant Kafka, nous avons traité les 5V de notre data, permettant ainsi de transformer ces informations brutes en insights exploitables.

3

LES 5 V

- Volume : Des millions de morceaux et d'utilisateurs.
- Vitesse : Des données constamment mises à jour en temps réel.
- Variété : Différents types de données (textuelles, numériques, temporelles).
- Véracité : L'importance de garantir l'exactitude des données collectées.
- Valeur : Transformer ces données en informations exploitables pour des analyses et des insights

PRÉSENTATION DES TECHNOLOGIES KAFKA

1

POURQUOI KAFKA ?

PRODUCER :

- LES ACTIONS UTILISATEUR (LECTURE DE CHANSON, PAUSE, RECHERCHE) SONT CAPTURÉES VIA L'INTERFACE UTILISATEUR OU L'API SPOTIFY ET TRANSMISES À KAFKA.
- LES LOGS SYSTÈME OU MÉTRIQUES DES SERVEURS PEUVENT ÉGALEMENT ÊTRE DES PRODUCTEURS

PRODUCER API :

- L'API COLLECTE DES DONNÉES SUR LES CHANSONS JOUÉES, LES PLAYLISTS CRÉÉES OU LES RECHERCHES EFFECTUÉES.
- CES INFORMATIONS SONT TRANSMISES AU PRODUCTEUR KAFKA POUR UN TRAITEMENT ULTÉRIEUR.

- **SCALABILITÉ** : GÈRE LE VOLUME MASSIF DE DONNÉES GÉNÉRÉ PAR SES MILLIONS D'UTILISATEURS ACTIFS.
- **TRAITEMENT EN TEMPS RÉEL** : GARANTIT UNE TRANSMISSION ET UN TRAITEMENT RAPIDE DES ÉVÉNEMENTS, CE QUI EST ESSENTIEL POUR DES FONCTIONNALITÉS COMME LES RECOMMANDATIONS PERSONNALISÉES.
- **FIABILITÉ** : GÈRE LES PANNES EN S'ASSURANT QUE LES DONNÉES CRITIQUES NE SONT PAS PERDUES.

PRODUCER PAR API

PRODUCER SCRAPING

- SPOTIFY PEUT SCRAPER DES SITES DE RÉSEAUX SOCIAUX COMME TWITTER, REDDIT, OU DES BLOGS MUSICAUX POUR IDENTIFIER DES ARTISTES OU DES CHANSONS QUI GAGNENT EN POPULARITÉ, ET AINSI ADAPTER SES RECOMMANDATIONS MUSICALES EN FONCTION DE CES TENDANCES.
- SPOTIFY PEUT AUSSI SCRAPER DES PLATEFORMES COMME YOUTUBE OU DES FORUMS SPÉCIALISÉS POUR OBTENIR DES CRITIQUES SUR DES ALBUMS OU DES ARTISTES, CE QUI POURRAIT ENSUITE INFLUENCER LA RECOMMANDATION MUSICALE.

PRÉSENTATION DES TECHNOLOGIES HADOOP

2

POURQUOI HADOOP ?

- HADOOP EST PARTICULIÈREMENT ADAPTÉ POUR GÉRER DES DATASETS DE GRANDE TAILLE, ET C'EST UN SYSTÈME DE STOCKAGE ET DE TRAITEMENT IDÉAL POUR LES GRANDES VOLUMÉTRIES DE DONNÉES QUI PEUVENT DEVENIR COMPLEXES À ANALYSER EN RAISON DE LEUR TAILLE.

APACHE SPARK

APACHE ZEPPELIN

- UTILISÉ POUR EFFECTUER DES TRAITEMENTS BATCH SUR LES DONNÉES STOCKÉES DANS HDFS. SPARK PEUT EFFECTUER DES ANALYSES COMPLEXES SUR DE GRANDES VOLUMÉTRIES DE DONNÉES DE MANIÈRE DISTRIBUÉE

- AGIT COMME UN OUTIL DE NOTEBOOK INTERACTIF PERMETTANT D'ÉCRIRE, EXÉCUTER ET VISUALISER DES SCRIPTS SPARK EN TEMPS RÉEL. AVEC ZEPPELIN, TU PEUX FACILEMENT GÉRER DES ANALYSES SUR DES DONNÉES VOLUMINEUSES ET CRÉER DES VISUALISATIONS INTERACTIVES (PAR EXEMPLE, POUR LES TENDANCES MUSICALES OU LES RECOMMANDATIONS).



DÉMONSTRATION

- MISE EN SITUATION PRATIQUE AVEC LE PROJET FONCTIONNEL
- 



CONCLUSION

Le projet Spotify intègre Kafka, Hadoop, Spark, et Zeppelin pour offrir une solution scalable et performante. Il permet de traiter de grandes quantités de données en temps réel et en batch, tout en fournissant des recommandations musicales personnalisées grâce à l'usage de Machine Learning. Zeppelin facilite l'analyse interactive et la visualisation des données. Cette architecture flexible et robuste assure une gestion efficace des données, tout en offrant une expérience utilisateur optimisée et adaptable aux futurs besoins d'expansion.





**MERCI POUR
VOTRE ATTENTION**

