



Énoncé de TP : Analyse des tendances musicales avec MapReduce, HappyBase et Matplotlib

Objectif :

Ce TP a pour objectif d'analyser la relation entre les caractéristiques musicales (danceability et énergie) et la popularité (streams).

- **Mapper :**
 - Extraire le pourcentage de danceability (**danceability_%**), le pourcentage d'énergie (**energy_%**) et le nombre de streams (**streams**) pour chaque chanson.
- **Reducer :**
 - Agréger les données par combinaison de **danceability_%** et **energy_%**.
 - Calculer la somme totale des streams pour chaque combinaison, ainsi que le nombre d'occurrences de chaque combinaison (pour ensuite calculer la moyenne des streams pour chaque combinaison de danceability et d'énergie).

Vous utiliserez **Hadoop MapReduce** pour traiter les données, **HappyBase** pour stocker les résultats dans **HBase**, et **Matplotlib** pour visualiser les résultats sous forme de graphique.

Livrables attendus :

1. **Le code source** de votre job MapReduce.
2. **Le graphique généré** avec Matplotlib (au format PDF).
3. **Un rapport détaillé** expliquant chaque étape du processus, les résultats obtenus, et une analyse des relations observées entre la danceability et les streams.

Étapes du TP :

1. Connexion à la machine virtuelle

- Indiquez comment vous vous connectez à la machine virtuelle avec l'IP, le port, le nom d'utilisateur et le mot de passe.

- Décrivez les étapes pour établir la connexion SSH.

2. Lancer les conteneurs et services Hadoop

- Expliquez les services nécessaires pour Hadoop (HDFS, YARN, Zookeeper, HBase).
- Indiquez les commandes nécessaires pour démarrer ces services sur la machine virtuelle.

3. Importer les données dans HDFS

- Décrivez la manière dont vous allez importer les données pour qu'elles puissent être traitées avec MapReduce.
- Précisez où stocker les fichiers.

4. Créer et exécuter un job MapReduce

- Expliquez la structure d'un job MapReduce (mapper, reducer).
- Donnez les commandes à exécuter pour soumettre le job sur Hadoop.

5. Visualiser les résultats

- Décrivez comment récupérer les résultats de MapReduce.
- Indiquez comment visualiser les données avec **Matplotlib** (scatter plot, bar chart).

6. Récupérer les résultats

- Expliquez comment récupérer les fichiers de sortie du job MapReduce depuis HDFS sur votre machine locale.
- Donnez la procédure pour récupérer le graphique généré par Matplotlib.