

Tweets

Projet Big Data – Analyse de Tweets avec Hadoop

Partie 1 : Installation et Préparation des Données

1. Installation et Configuration de Hadoop

- Installer Hadoop et configurer un cluster multi-nœuds.
- Configurer HDFS avec un facteur de réplication approprié pour assurer la tolérance aux pannes.

2. Préparation des Données

- Charger les données Twitter dans HDFS.
 - Organiser les données en répertoires selon le mois et l'année de publication des tweets.
-

Partie 2 : Analyse de Données avec MapReduce

1. Analyse des Tendances de Hashtags

- Écrire un programme MapReduce qui identifie les 10 hashtags les plus populaires par mois.
- Étudier l'évolution des tendances de hashtags au fil du temps.

2. Analyse de Sentiments

- Utiliser une bibliothèque d'analyse de sentiments pour évaluer l'émotion générale (positive, négative, neutre) de chaque tweet.
- Écrire un programme MapReduce qui calcule le score moyen de sentiment pour chaque jour.
- Identifier les événements ayant eu un impact significatif sur l'humeur collective des utilisateurs de Twitter.

3. Distribution Géographique des Tweets

- Écrire un programme MapReduce qui regroupe les tweets par pays ou région (à partir des métadonnées de localisation disponibles).
 - Analyser les différences thématiques dans les tweets selon les régions géographiques.
-