

EXAMEN deuxième session : HADOOP MapReduce

Durée 2h, documents non autorisés

Exercice 1 : Traitement de données de vente

Objectif

Développer un job Hadoop MapReduce pour calculer le total des ventes par ville à partir d'un fichier texte (ventes.txt) contenant les ventes d'une entreprise dans différentes villes. La structure du fichier ventes.txt est la suivante :

date ville produit prix

Travail à faire

1. Développer un job MapReduce pour calculer le total des ventes par ville.
2. Développer un deuxième job pour calculer le prix total des ventes des produits par ville pour une année donnée.

Exercice 2 : Analyse des fichiers journaux Web

Objectif

Analyser un grand ensemble de fichiers journaux (logs) Web générés par un serveur. Chaque ligne du fichier journal contient des informations sur les requêtes HTTP, y compris l'adresse IP du client, la date, le chemin d'accès demandé, le code de réponse HTTP, etc.

Utiliser Hadoop MapReduce pour calculer le nombre total de requêtes et le nombre de requêtes réussies (code de réponse HTTP 200) par adresse IP.