Université Mohammed V de Rabat
Ecole Nationale Supérieure d'Arts et Métiers
Cycle Ingénieur Filière INDIA (3ème année) A.U. : 2023/2024 Pr. A. El Qadi

TP n7 en Mise en œuvre des Framework d'IA et Big data

Exercice 1. Hive : Utilisation du référentiel docker-Hadoop **https://github.com/Marcel-Jan/docker-hadoop-spark.git**

1. Lancer Docker desktop
2. Accéder à la ligne de commande du serveur Hive et démarrer hiveserver2
   C:\docker-hadoop-spark**>docker exec -it hive-server bash**
   root@baf977faf87f:/opt# **hiveserver2**

Vérification que quelque chose est à l'écoute sur le port 10000 maintenant
root@baf977faf87f:/opt# netstat -anp | grep 10000

3. Connecter à hiveserver2 (Beeline est l'interface en ligne de commande de Hive)

root@baf977faf87f:/opt# beeline -u jdbc:hive2://localhost:10000 -n root

0: jdbc:hive2://localhost:10000> !connect jdbc:hive2://127.0.0.1:10000 scott tiger

4. Créer la BD dbHive
5. Afficher le contenu du dossier HDFS : /user/hive/warehouse
6. Afficher toutes les bases de données et choisir la BD dbHive
7. Créer la table interne livre_interne (codelivre, isbn , titre ,auteur, editeur, dateedition, prixunite)
8. Afficher la liste des tables de la BD courante
9. Consulter le Metastore pour avoir le schéma de la table livre : DESCRIBE livre_interne **;**
10. Copier le fichier datalIvre.csv dans le dossier du référentiel cloné namenode :/tmp
11. Copier le fichier de dataLivre.csv /tmp vers le fichier d'entrée /user/root/input
12. Charger le fichier local dataLivre.csv dans la table livre
13. Effectuer des requête HQL sur la table livre_interne
    a. Afficher le premier livre : select codelivre, titre from livre limit 2;
    b. Lister la liste des éditeurs présents dans la base
    c. Compter le nombre des livres d'Editeur 'Dunod'
    d. Chercher tous les livres de prixunite > 300).
14. Créer la table externe livre_externe (codelivre, isbn , titre ,auteur, editeur, dateedition, prixunite) en indiquant son dossier HDFS de données qu'il faut créer, par exemple: /user//hive/data/db
15. Afficher la liste des tables de la BD courante.
16. Copier le fichier local dataLivre.csv dans le dossier HDFS: /user/hive/data/db
17. Afficher le contenu du dossier HDFS: /user/hive/data/db
18. Afficher les 5 premiers livres dans la table livre_externe.
19. Exécuter séparément les deux requêtes :

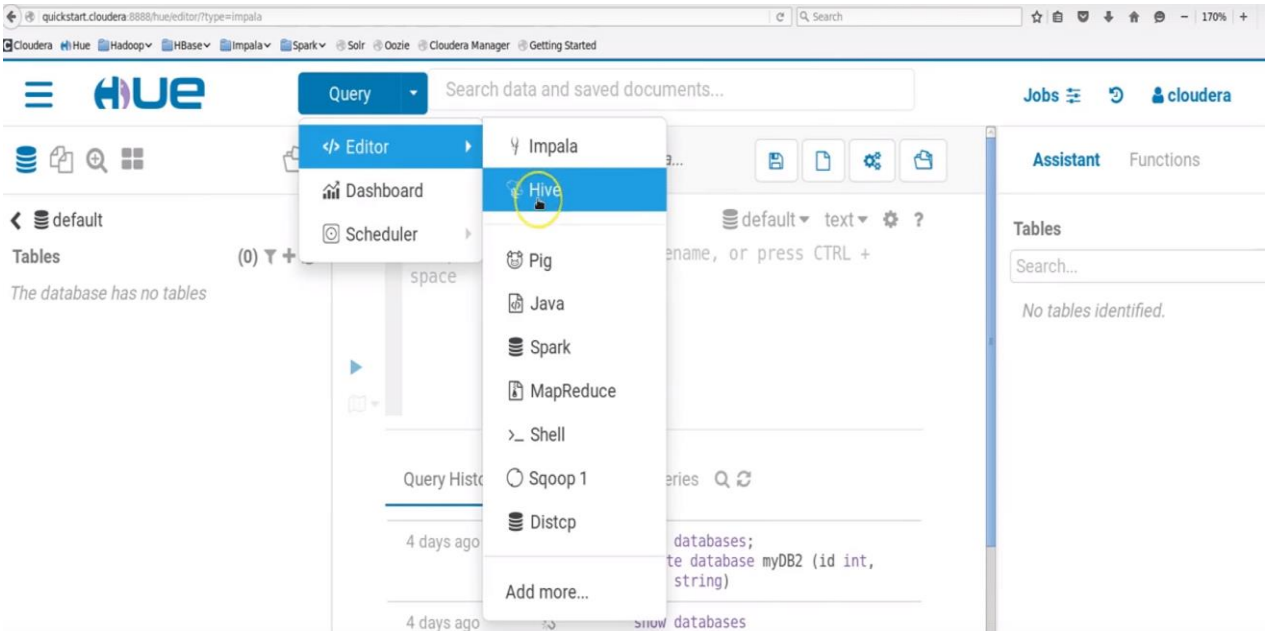
SELECT editeur, COUNT(*) FROM livre_externe GROUP BY editeur;	SELECT editeur, COUNT(*) FROM livre_interne GROUP BY editeur;

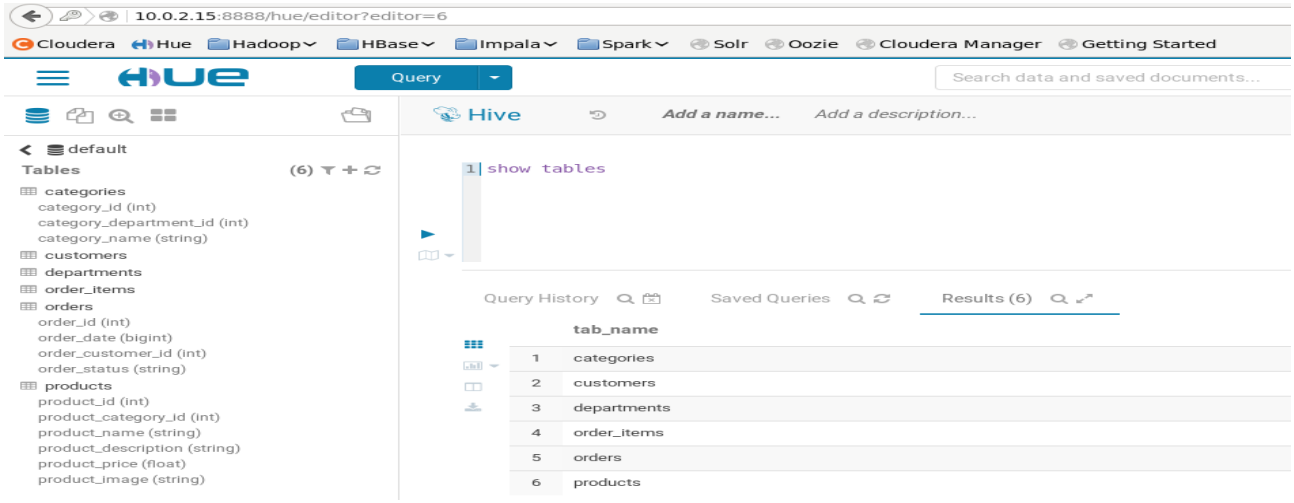
Quelle est la différence entre les deux lors de leur exécution ?

Remarqurer les jobs Map-Reduce crées.

Exercice 2. Gérer les données avec Apache **Hue** (**H**adoop **U**ser **E**xperience)

Hue est une application Web open source qui propose des interfaces utilisateurs nettement plus élaborées que celles fournies de base par Apache Hadoop.
Nous utiliserons le QuickStart de Cloudera, qui dispose de la configuration Hadoop sur un seul nœud.
Le mot de passe par défaut du site Web Hue dans la machine virtuelle Hadoop est le suivant :
Nom d'utilisateur : cloudera
Mot de passe : cloudera





1. Créer la table interne livre_interne (codelivre, isbn , titre ,auteur, editeur, dateedition, prixunite)
2. Afficher la liste des tables de la BD courante
3. Consulter le Metastore pour avoir le schéma de la table livre : DESCRIBE livre_interne **;**
4. Charger le fichier local dataLivre.csv dans la table livre
5. Effectuer des requête HQL sur la table livre_interne
   a. Afficher le premier livre : select codelivre, titre from livre limit 2;
   b. Lister la liste des éditeurs présents dans la base
   c. Compter le nombre des livres d'Editeur 'Dunod'
   d. Chercher tous les livres de prixunite > 300).