



Pratique de Hadoop n°05

SQOOP (en Python)

Table des matières

I.	Objectif du TP	2
II.	Démarrer avec Sqoop	2
A.	Installation et configuration de SQOOP (Linux)	2
B.	Installation et configuration de MySQL (Linux)	3
C.	Création d'une Base de données sur MySQL	4

I. Objectif du TP

- Initiation à l'outil Sqoop.

II. Démarrer avec Sqoop

A. Installation et configuration de SQOOP (Linux)

1. Télécharger le package Sqoop (la version 1.4.7) :

wget <https://archive.apache.org/dist/sqoop/1.4.2/sqoop-1.4.2.tar.gz>

2. Décompresser le fichier et le déplacer dans le répertoire /usr/local

```
tar -xvf sqoop-1.4.2.tar.gz
```

```
mv /usr/local/sqoop-1.4.2 /usr/local/sqoop
```

3. Ajouter les variables nécessaires sur l'env :

```
vi .bashrc
```

```
export SQOOP_HOME=/usr/local/sqoop
```

```
export PATH=$PATH:$SQOOP_HOME/bin
```

```
source .bashrc
```

4. Configurer Sqoop:

```
cd $SQOOP_HOME/conf
```

```
mv sqoop-env-template.sh sqoop-env.sh
```

```
vi sqoop-env.sh
```

5. sur le fichier ajouter les deux chemins Hadoop : (echo \$HADOOP_HOME) et Hbase (echo \$HBASE_HOME)

```
export HADOOP_COMMON_HOME=/usr/local/Hadoop
```

```
export HADOOP_MAPRED_HOME=/usr/local/Hadoop
```

6. Configurer le sqoop jar dans sqoop home :

```
wget https://repo1.maven.org/maven2/org/apache/sqoop/sqoop/1.4.2/sqoop-1.4.2-hadoop200.jar
```

```
cp sqoop-1.4.2-hadoop200.jar $SQOOP_HOME
```

7. Configurer le connecteur mysql :

```
wget http://ftp.ntu.edu.tw/MySQL/Downloads/Connector-J/mysql-connector-java-8.0.29.tar.gz
```

```
tar -xvf mysql-connector-java-8.0.29.tar.gz
```

```
mv mysql-connector-java-8.0.29/mysql-connector-java-8.0.29.jar  
/$SQOOP_HOME/lib
```

8. Vérifier l'installation de Sqoop :

```
cd $SQOOP_HOME/bin  
sqoop version
```

```
root@hadoop-master:~# sqoop version  
Sqoop 1.4.2  
git commit id 2c4400c3ae432806ea7049cffd6802b783955a0a  
Compiled by jarcec on Wed Apr 24 19:41:20 PDT 2013  
root@hadoop-master:~#
```

B. Installation et configuration de MySQL (Linux)

Afin d'utiliser Sqoop, on va créer une base de données sur Mysql pour pouvoir la migrée sur HDFS. Decoup, on va commencer par installer Mysql :

1. Installation de Mysql :

```
apt update
```

```
apt install mysql-server
```

```
mysql -V
```

2. Lancer Mysql:

```
service mysql start
```

```
mysql -u root -p
```

C. Création d'une Base de données sur MySQL

On va créer une base de données « movies », utilisez le fichier movies.sql. Exécuter les commandes sql sur le mysql shell.

1. Vérifiez que la base de données est bien créée
2. Affichez « *movie_id, title, budget, revenue, popularity, vote_count* » pour chaque film.
3. Afficher la liste des films qui ont apporté des revenus par ordre de vote.

III. Exercice suite au TP

1. Afficher la liste des bases de données disponibles sur notre serveur MySQL.
2. Afficher la liste des tables disponibles sur notre serveur MySQL pour la base de données movies.
3. Afficher le résultats de la requête qui permet d'afficher « *movie_id, title, budget, revenue, popularity, vote_count* » pour chaque film.
4. Importez toute la table movie sur HDFS.
5. Vérifier que les données sont bien importées sur HDFS.
6. Importez sur HDFS le résultat pour la requête qui permet d'afficher la liste des films qui ont apporté des revenus ordonnées par vote en utilisant sqoop.
7. Vérifier que les données sont bien importées sur HDFS.
8. Exporter les données stockées sur HDFS pour la partie salaire (age, max salaire, min salaire, nombre) vers le mysql.