
Adaptation du fichier CO2.csv et son intégration dans la table catalogue

- Pour commencer, nous avons écrit un programme Hadoop pour l'adaptation, accessible à cette URL : <https://github.com/ManoaLoic/TPA/tree/main/MAP-REDUCE/MAP-REDUCE-TPA>, puis nous avons compilé ce programme pour obtenir le fichier JAR nommé MAP-REDUCE-TPA-1.0-SNAPSHOT.jar. Ce JAR sera placé dans le dossier TPA sur Vagrant.

Nom	Modifié le	Type	Taille
data	17/05/2024 22:29	Dossier de fichiers	
MAP-REDUCE-TPA-1.0-SNAPSHOT.jar	17/05/2024 22:22	Fichier JAR	

- Après cela, sur la machine virtuelle (VM), nous avons démarré Hadoop.

```
[vagrant@oracle-21c-vagrant ~]$ start-dfs.sh
Starting namenodes on [localhost]
Starting datanodes
Starting secondary namenodes [oracle-21c-vagrant]
[vagrant@oracle-21c-vagrant ~]$ start-yarn.sh
Starting resourcemanager
Starting nodemanagers
```

- Il est nécessaire d'insérer les fichiers co2 et catalogue dans Le dossier dans le dossier TPA/data du vagrant et de les placer dans le HDFS (étape déjà réalisée lors de la préparation du datalake). Nous allons également créer le répertoire pour les nouvelles données de catalogue.

Nom	Modifié le	Type	Taille
Catalogue.csv	18/05/2024 10:20	Fichier source Comm...	14 Ko
Clients_0.csv	14/12/2006 22:32	Fichier source Comm...	3 748 Ko
Clients_19.csv	14/12/2006 22:36	Fichier source Comm...	3 748 Ko
CO2.csv	09/04/2024 16:22	Fichier source Comm...	39 Ko

```
[vagrant@oracle-21c-vagrant ~]$ MYTPHOME=/vagrant/TPA/data
[vagrant@oracle-21c-vagrant ~]$ hdfs dfs -mkdir /M2_DMA_Catalogue
[vagrant@oracle-21c-vagrant ~]$ hdfs dfs -mkdir /CO2
[vagrant@oracle-21c-vagrant ~]$ hdfs dfs -mkdir /M2_DMA_New_Catalogue
[vagrant@oracle-21c-vagrant ~]$ hdfs dfs -put $MYTPHOME/Catalogue.csv /M2_DMA_Catalogue
[vagrant@oracle-21c-vagrant ~]$ hdfs dfs -put $MYTPHOME/CO2.csv /CO2
```

- Après cela, nous allons exécuter le JAR en utilisant la commande suivante :

```
[vagrant@oracle-21c-vagrant ~]$ hadoop jar /vagrant/TPA/MAP-REDUCE-TPA-1.0-SNAPSHOT.jar  
com.mycompany.map.reduce.tpa.MAPREDUCETPA /M2_DMA_Catalogue/Catalogue.csv /CO2/CO2.csv  
/M2_DMA_New_Catalogue/newCatalogue
```

- Nous allons importer le fichier contenant le résultat de l'adaptation dans le dossier data du TPA avec la commande suivante :

```
[vagrant@oracle-21c-vagrant ~]$ hadoop fs -cat /M2_DMA_New_Catalogue/newCatalogue/* |  
sed 's/\\t/,/g' > /vagrant/TPA/data/newCatalogue.csv
```

- Création de la table externe pour le fichier dans Hive :
 - Pour commencer , nous allons démarrer hive

```
nohup hive --service metastore > hive.log &  
nohup hiveserver2 > hiveServer.log &
```

- Nous allons créer de nouveau la table externe catalogue en prenant en compte les nouvelles données obtenues lors de l'adaptation du fichier co2 dans catalogue pendant le MapReduce.

```
0: jdbc:hive2://localhost:10000>CREATE EXTERNAL TABLE M2_DMA_Catalogue_EXT (  
  MARQUE string,  
  NOM string,  
  PUISSANCE int,  
  LONGUEUR string,  
  NBPLACES int,  
  NBPORTES int,  
  COULEUR string,  
  OCCASION boolean,  
  PRIX int,  
  bonusmalus double,  
  co2 double,  
  coutEnergie double  
) ROW FORMAT DELIMITED FIELDS TERMINATED BY ',' STORED AS TEXTFILE LOCATION  
'hdfs://M2_DMA_New_Catalogue';
```