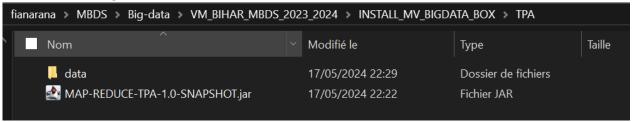
Adaptation du fichier CO2.csv et son intégration dans la table catalogue

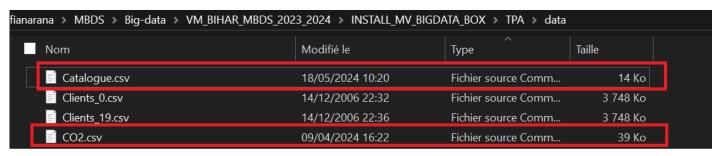
Pour commencer, nous avons écrit un programme Hadoop pour l'adaptation, accessible à cette URL : https://github.com/ManoaLoic/TPA/tree/main/MAP-REDUCE/MAP-REDUCE-TPA, puis nous avons compilé ce programme pour obtenir le fichier JAR nommé MAP-REDUCE-TPA-1.0-SNAPSHOT.jar. Ce JAR sera placé dans le dossier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



• 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.



```
[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 exporter le fichier contenant le résultat de l'adaptation dans le dossier data du TPA dans vagrant avec la commande suivante pour vérifier si les données (bonusmalus, CO2, coût) ont bien été ajoutées dans le catalogue :

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

- Modification de la table externe M2 DMA Catalogue EXT dans Hive :
 - o Pour commencer, nous allons démarrer hive

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

 Nous allons modifier la table externe catalogue en prenant en compte les nouvelles données obtenues lors de l'adaptation du fichier co2 dans catalogue pendant le MapReduce et en changeant la propriété location correspondant au chemin du résultat sur HDFS (/M2_DMA_New_Catalogue/NewCatalogue).

```
0: jdbc:hive2://localhost:10000>drop table M2_DMA_Catalogue_EXT ;
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/newCatalogue';
```