## 1. Transfert des résultats de la tâche Map-Reduce vers la datalake Hive :

• Vous configurez une table dans Hive dans la datalake pour accueillir les résultats du traitement Map-Reduce nommé "resultat\_Catalogue\_CO2.csv" effectué sur les fichiers Excel "Catalogue.csv" et "CO2.csv", préalablement chargés dans HDFS.

```
0: jdbc:hive2://localhost:10000> CREATE TABLE IF NOT EXISTS
resultat_catalogue_co2 (
  marque STRING,
  nom STRING,
  puissance INT,
  longueur STRING,
  nbPlaces INT,
  nbPortes INT.
  couleur STRING,
  occasion BOOLEAN,
  prix INT,
  moyenne_bonus_malus DOUBLE,
  moyenne_rejets_co2 DOUBLE,
  cout energie moyen DOUBLE
ROW FORMAT DELIMITED
FIELDS TERMINATED BY ','
STORED AS TEXTFILE
TBLPROPERTIES ("skip.header.line.count"="1");
```

• Vous importez les données du fichier résultat "resultat\_Catalogue\_CO2.csv", qui ont été préalablement chargées dans HDFS, dans cette table :

```
0: jdbc:hive2://localhost:10000> LOAD DATA INPATH '/MBDS_Projet/resultat_Catalogue_CO2.csv' OVERWRITE INTO TABLE resultat catalogue co2;
```

• Vous effectuez une vérification pour vous assurer que les données ont été correctement chargées et vous comptez le nombre de lignes dans la table :

```
0: jdbc:hive2://localhost:10000> Select * from resultat_catalogue_co2;
```

								resultat_catalogue_co2.nbportes catalogue co2.cout energie moyen	resultat_catalogue_co2.couleu 
				<del>-</del>	·	+		+	
/olvo		<del>+</del>   272		très longue		5		5	blanc
false	50500	. I	0.0		42.455		72.727		ľ
/olvo		272		très longue					noir
false	50500		0.0		42.455		72.727		
		272		très longue				5	rouge
false	50500		0.0		42.455		72.727		
	S80 T6	272		très longue		5		5	gris
true	35350		0.0		42.455		72.727		
/olvo	S80 T6	272		très longue				5	bleu
true	35350		0.0		42.455		72.727		
/olvo	S80 T6	272		très longue		5		5	gris
false	50500		0.0		42.455		72.727		l <sub>e</sub>
/olvo	S80 T6	272		très longue				5	bleu
false	50500		0.0		42.455		72.727		l,
/olvo	S80 T6	272		très longue				5	rouge
true	35350		0.0		42.455		72.727		
		272		très longue		5		5	blanc
true	35350		0.0		42.455		72.727		ļ., .
/olvo	S80 T6	272		très longue		5		5	noir
true	35350		0.0		42.455		72.727		',
	Touran 2.0 FSI	150		longue				5	rouge
false	27340		-6000.0		32.8		96.0		'
/olkswagen		150		longue					gris
true	19138		-6000.0		32.8		96.0		
/olkswagen		150		longue			1 96.0		bleu
true	19138		-6000.0		32.8		96.0	1.6	', .
/olkswagen	Touran 2.0 FSI   27340	150		longue	32.8	7	96.0	5	gris
false		1 450	-6000.0				96.0	1.5	Labor
/olkswagen	Touran 2.0 FSI   27340	150		longue	1 32.8		96.0	5	bleu
false		ا 150 ا	-6000.0	longue		1 7	96.0	1.5	   blanc
/olkswagen	19138	126	-6000.0		32.8		96.0	13	l pranc
true /olkswagen		ا 150 ا		longue		1 7	90.0	1.5	Looin
	19138	1 126	-6000.0		32.8		96.0	5	noir
true /olkswagen		ا 150 ا				1 7	90.0	5	Laguage
VOIKSWagen	19138	130	-6000.0	longue	1 32.8		96.0		rouge

## 0: jdbc:hive2://localhost:10000> Select COUNT(\*) from resultat\_catalogue\_co2;

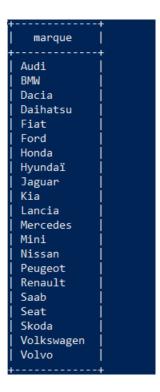
## 2. Exploration des données de la table « resultat\_catalogue\_co2 » :

## 0: jdbc:hive2://localhost:10000>DESC resultat\_catalogue\_co2;

col_name	data_type	comment
narque	string	
nom	string	
puissance	int	
longueur	string	
nbplaces	int	1
nbportes	int	j i
couleur	string	
occasion	boolean	
prix	int	
moyenne_bonus_malus	double	1
moyenne_rejets_co2	double	
cout_energie_moyen	double	j i

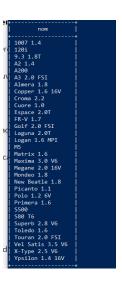
-----Exploration des valeurs de la colonne « marque » ------

0: jdbc:hive2://localhost:10000>Select DISTINCT marque from resultat\_catalogue\_co2;



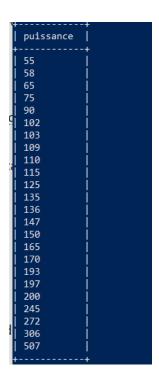
-----Exploration des valeurs de la colonne « nom » ------

0: jdbc:hive2://localhost:10000>Select DISTINCT nom from resultat\_catalogue\_co2;



-----Exploration des valeurs de la colonne « puissance » -----

0: jdbc:hive2://localhost:10000>Select DISTINCT puissance from resultat\_catalogue\_co2;



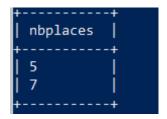
-----Exploration des valeurs de la colonne « longueur » ------

0: jdbc:hive2://localhost:10000>Select DISTINCT longueur from resultat\_catalogue\_co2;



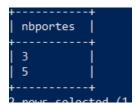
-----Exploration des valeurs de la colonne « nbplaces » -----

0: jdbc:hive2://localhost:10000>Select DISTINCT nbplaces from resultat\_catalogue\_co2;



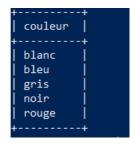
-----Exploration des valeurs de la colonne « nbportes » ------

0: jdbc:hive2://localhost:10000>Select DISTINCT nbportes from resultat\_catalogue\_co2;



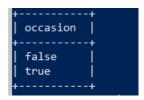
------Exploration des valeurs de la colonne « couleur » -------

0: jdbc:hive2://localhost:10000>Select DISTINCT couleur from resultat\_catalogue\_co2;



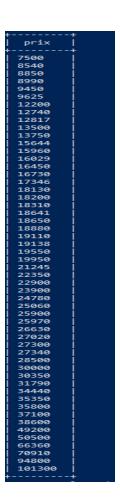
-----Exploration des valeurs de la colonne « occasion » ------

0: jdbc:hive2://localhost:10000>Select DISTINCT occasion from resultat\_catalogue\_co2;



-----Exploration des valeurs de la colonne « prix » ------

0: jdbc:hive2://localhost:10000>Select DISTINCT prix from resultat\_catalogue\_co2;



Nous pouvons conclure que toutes les colonnes associées au catalogue respectent le dictionnaire des données.