

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.marque	resultat_catalogue_co2.nom	resultat_catalogue_co2.puissance	resultat_catalogue_co2.longueur	resultat_catalogue_co2.nbplaces	resultat_catalogue_co2.nbportes	resultat_catalogue_co2.couleur
resultat_catalogue_co2.occasion	resultat_catalogue_co2.prix	resultat_catalogue_co2.moyenne_bonus_malus	resultat_catalogue_co2.moyenne_rejets_co2	resultat_catalogue_co2.cout_energie_moyen		
Volvo	S80 T6	272	très longue	5	5	blanc
false	50500	0.0		42.455	72.727	
Volvo	S80 T6	272	très longue	5	5	noir
false	50500	0.0		42.455	72.727	
Volvo	S80 T6	272	très longue	5	5	rouge
false	50500	0.0		42.455	72.727	
Volvo	S80 T6	272	très longue	5	5	gris
true	35350	0.0		42.455	72.727	
Volvo	S80 T6	272	très longue	5	5	bleu
true	35350	0.0		42.455	72.727	
Volvo	S80 T6	272	très longue	5	5	gris
false	50500	0.0		42.455	72.727	
Volvo	S80 T6	272	très longue	5	5	bleu
false	50500	0.0		42.455	72.727	
Volvo	S80 T6	272	très longue	5	5	rouge
true	35350	0.0		42.455	72.727	
Volvo	S80 T6	272	très longue	5	5	blanc
true	35350	0.0		42.455	72.727	
Volvo	S80 T6	272	très longue	5	5	noir
true	35350	0.0		42.455	72.727	
Volkswagen	Touran 2.0 FSI	150	longue	7	5	rouge
false	27340	-6000.0		32.8	96.0	
Volkswagen	Touran 2.0 FSI	150	longue	7	5	gris
true	19138	-6000.0		32.8	96.0	
Volkswagen	Touran 2.0 FSI	150	longue	7	5	bleu
true	19138	-6000.0		32.8	96.0	
Volkswagen	Touran 2.0 FSI	150	longue	7	5	gris
false	27340	-6000.0		32.8	96.0	
Volkswagen	Touran 2.0 FSI	150	longue	7	5	bleu
false	27340	-6000.0		32.8	96.0	
Volkswagen	Touran 2.0 FSI	150	longue	7	5	blanc
true	19138	-6000.0		32.8	96.0	
Volkswagen	Touran 2.0 FSI	150	longue	7	5	noir
true	19138	-6000.0		32.8	96.0	
Volkswagen	Touran 2.0 FSI	150	longue	7	5	rouge
true	19138	-6000.0		32.8	96.0	

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

```
24/05/02 18:18:32 INFO lockmgr.DbTxnH
24/05/02 18:18:32 INFO txn.TxnHandler
24/05/02 18:18:32 INFO txn.TxnHandler
+-----+
| _c0 |
+-----+
| 270 |
+-----+
1 row selected (1.92 seconds)
0: jdbc:hive2://localhost:10000>
```

2. Exploration des données de la table « resultat_catalogue_co2 » :

```
0: jdbc:hive2://localhost:10000>DESC resultat_catalogue_co2 ;
```

col_name	data_type	comment
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	

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

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

marque
Audi
BMW
Dacia
Daihatsu
Fiat
Ford
Honda
Hyundai
Jaguar
Kia
Lancia
Mercedes
Mini
Nissan
Peugeot
Renault
Saab
Seat
Skoda
Volkswagen
Volvo

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

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

nom
1007 1.4
120i
9.3 1.8T
A2 1.4
A200
A3 2.0 FSI
Almera 1.8
Copper 1.6 16V
Croma 2.2
Cuore 1.0
Espace 2.0T
FR-V 1.7
Golf 2.0 FSI
Laguna 2.0T
Logan 1.6 MPI
M5
Matrix 1.6
Maxima 3.0 V6
Megane 2.0 16V
Mondeo 1.8
New Beetle 1.8
Picanto 1.1
Polo 1.2 16V
Prius 1.6
S500
S80 T6
Superb 2.8 V6
Toledo 1.6
Touran 2.0 FSI
V6i Satis 3.5 V6
X-Type 2.5 V6
Ypsilon 1.4 16V

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

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

puissance
55
58
65
75
90
102
103
109
110
115
125
135
136
147
150
165
170
193
197
200
245
272
306
507

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

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

longueur
courte
longue
moyenne
très longue

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

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

nbplaces
5
7

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

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

nbportes
3
5

2 rows selected (1

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

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

couleur
blanc
bleu
gris
noir
rouge

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

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

occasion
false
true

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

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

prix
7500
8540
8850
8990
9450
9625
12200
12740
12817
13500
13750
15644
15960
16029
16450
16730
17346
18130
18200
18310
18641
18650
18880
19110
19138
19550
19950
21245
22350
22900
23900
24780
25060
25900
25970
26630
27020
27300
27340
28500
30000
30350
31790
34440
35350
35800
37100
38600
49200
50500
66360
70910
94800
101300

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