



Dans cette partie, nous allons expliquer comment nous créons des tables externes pointant vers les tables Mongo DB et Oracle NoSQL en Driver Access ainsi que des tables externes à partir du fichier CO2.txt

En commençant bien évidemment par démarrer la machine virtuelle.

On lance HDFS

start-dfs.sh start-yarn.sh

#### On lance HIVE

nohup hive --service metastore > /dev/null & nohup hiveserver2 > /dev/null &

On se connecte au serveur HIVE

beeline -u jdbc:hive2://localhost:10000 vagrant

0: idbc:hive2://localhost:100002

On crée une base de données

CREATE DATABASE IF NOT EXISTS projetBigData; USE projetBigData;

Afin d'avoir une vision plus claire de l'affichage des données, on va utiliser les lignes ci-dessous :

```
set hive.cli.print.header=true;
set hive.cli.print.row.to.vertical=true;
set hive.cli.print.current.db=false;
set hive.cli.print.row.to.vertical.limit=10000;
set hive.resultset.use.unique.column.names=false;
set hive.cli.print.header.separator="\t";
set hive.cli.print.max_column_width=30;
```

# Immatriculations\_ext

Nous allons créer une table externe Immatriculation\_ext qui pointe vers la table Immatriculation d'Oracle NoSQL

On commence par drop la table dans le cas où elle existe déjà

drop table Immatriculations\_ext;

#### On crée ensuite la table

```
CREATE EXTERNAL TABLE IF NOT EXISTS Immatriculations_ext (
immatriculation string,
marque string,
nom string,
puissance int,
longueur String,
nbplaces int,
nbportes int,
couleur String,
occasion boolean,
prix int
)

STORED BY 'oracle.kv.hadoop.hive.table.TableStorageHandler'
TBLPROPERTIES (
"oracle.kv.kvstore" = "kvstore",
"oracle.kv.hosts" = "localhost:5000",
"oracle.kv.tableName" = "IMMATRICULATIONS_ESTIA2223_PACHOT");
```

On vérifie en comptant les lignes

select count(\*) from Immatriculations\_ext;

```
CO<sub>2</sub> ext
```

On drop la table au cas où elle serait déjà existante

```
DROP TABLE IF EXISTS CO2 EXT;
```

Si ce n'est pas le cas, on la crée

```
CREATE EXTERNAL TABLE IF NOT EXISTS CO2_EXT(
    marque STRING,
    avgBonusMalus INT,
    avgCO2 INT,
    avgCoutEnergie INT)

COMMENT 'CO2 text data'

ROW FORMAT DELIMITED

FIELDS TERMINATED BY ';'

STORED AS TEXTFILE

location 'hdfsFileSortedCO2Final';
```

#### Et on passe à la vérification

```
SELECT * from CO2_EXT;
```

+   marque	+   avgbonusmalus	avgco2	+   avgcoutenergie
AUDI	-2400	26	191
BMW	-631	39	80
HYUNDAI	-4000	8	151
JAGUAR	-6000	0	271
KIA	-4000	10	157
MERCEDES	7790	187	749
MINI	-3000	21	126
NISSAN	5802	160	681
PEUGEOT	-3000	15	144
RENAULT	-6000	0	206
SKODA	-666	27	98
VOLKSWAGEN	-1714	23	96
V0LV0	0	42	72
ZDACIA	-1370	42	232
ZDAIHATSU	-1370	42	232
ZFIAT	-1370	42	232
ZFORD	-1370	42	232
ZHONDA	-1370	42	232
ZLANCIA	-1370	42	232
ZSAAB	-1370	42	232
ZSEAT	-1370	42	232
+			

On crée une table interne dans laquelle on va implémenter la data de la table externe

### CO<sub>2</sub>

Nous allons créer une table interne Catalogue

Le drop habituel

#### DROP TABLE IF EXISTS CO2:

On crée ensuite la table

```
CREATE TABLE IF NOT EXISTS CO2(
marque STRING, avgBonusMalus INT, avgCO2 INT,
avgCoutEnergie INT)
COMMENT 'Data about Bonus Malus, emission of CO2 and ernergy cost for each brand';
```

On met cette table en interne

INSERT OVERWRITE TABLE CO2 SELECT \* FROM CO2\_EXT;

#### On passe à la vérification

### SELECT \* from CO2;

+   marque	+   avgbonusmalus	+   avgco2	+   avgcoutenergie
AUDI	-2400	26	191
BMW	-631	39	80
HYUNDAI	-4000	8	151
JAGUAR	-6000	0	271
KIA	-4000	10	157
MERCEDES	7790	187	749
MINI	-3000	21	126
NISSAN	5802	160	681
PEUGEOT	-3000	15	144
RENAULT	-6000	0	206
SKODA	-666	27	98
VOLKSWAGEN	-1714	23	96
V0LV0	0	42	72
ZDACIA	-1370	42	232
ZDAIHATSU	-1370	42	232
ZFIAT	-1370	42	232
ZFORD	-1370	42	232
ZHONDA	-1370	42	232
ZLANCIA	-1370	42	232
ZSAAB	-1370	42	232
ZSEAT	-1370	42	232

On supprime ensuite la table externe (pas obligatoire)

DROP TABLE IF EXISTS CO2 EXT;

# Catalogue\_ext

Nous allons créer une table externe Catalogue\_ext qui pointe vers la table Catalogue de MongoDB

On drop la table au cas où elle existerait

DROP TABLE IF EXISTS Catalogue\_ext;

#### On crée donc la table

CREATE EXTERNAL TABLE Catalogue\_ext ( id STRING, marque STRING, nom STRING, puissance INT, longueur STRING, nbPlaces INT, nbPortes INT, couleur STRING, occasion STRING, prix INT)

STORED BY 'com.mongodb.hadoop.hive.MongoStorageHandler'

 $WITH\ SERDEPROPERTIES ('mongo.columns.mapping'='\{"id":"\_id", "nbPlaces": "nbPlaces", nbPlaces", nbPlaces", nbPlaces": "nbPlaces": "nbPla$ 

"nbPortes": "nbPortes"}'

TBLPROPERTIES('mongo.uri'='mongodb://localhost:27017/projetBigData.Catalogue');

#### On vérifie son existence

#### SELECT \* FROM Catalogue\_ext;

			tr⊠s longue			
			tr⊡s longue			
			tr⊡s longue		rouge	
			tr⊡s longue			
			tr⊡s longue		rouge	
			tr⊡s longue			
	Volkswagen		longue		rouge	
	Volkswagen					
	Volkswagen		longue			
	Volkswagen		longue			
	Volkswagen		longue			
	Volkswagen		longue			
	Volkswagen		longue			
	Volkswagen		longue		rouge	
	Volkswagen		longue			
	Volkswagen		longue			
	Volkswagen					
	Volkswagen					
	Volkswagen					
	Volkswagen					
	Volkswagen					
641028cef84f684cbe744d48		Polo 1.2 6V	courte			12200

On ne va évidemment pas prendre toutes les lignes en capture d'écran.

On crée une table interne dans laquelle on va implémenter la data de la table externe

### Catalogue

Nous allons créer une table interne Catalogue

On drop la table au cas où elle existerait

#### DROP TABLE IF EXISTS Catalogue;

#### On crée la table

CREATE TABLE IF NOT EXISTS Catalogue(

id STRING, marque STRING, nom STRING, puissance INT, longueur STRING, nbPlaces INT, nbPortes INT, couleur STRING, occasion STRING, prix INT)

COMMENT 'Catalogue table intern table';

On met cette table en interne

INSERT OVERWRITE TABLE Catalogue SELECT \* FROM Catalogue\_ext;

On vérifie

SELECT \* from Catalogue;

id							
+   641028cef84f684cbe744d16		272	tr⊠s longue		blanc	false	+   50500
641028cef84f684cbe744d18	Volvo	272	tres longue				50500
641028cef84f684cbe744d1a	Volvo	272	trEs longue		rouge	false	50500
			tr@s longue				
			tr®s longue				
641028cef84f684cbe744d20			tr®s longue				
			tr®s longue				
			tr⊡s longue		rouge		
			trls longue				
			tr⊡s longue				
	Volkswagen		longue		rouge		
	Volkswagen						
	Volkswagen		longue				
	Volkswagen						
	Volkswagen		longue				
	Volkswagen		longue				
	Volkswagen		longue				
	Volkswagen		longue		rouge		
	Volkswagen		longue				
	Volkswagen		longue				
	Volkswagen						
	Volkswagen						
	Volkswagen						
	Volkswagen						
	Volkswagen						
	Volkswagen		moyenne				

On ne va évidemment pas prendre toutes les lignes en capture d'écran.

# Marketing\_ext

Nous allons créer une table externe Marketing\_ext qui pointe vers la table Marketing de MongoDB

On drop la potentielle table

DROP TABLE IF EXISTS Marketing\_ext;

On crée la table Marketing\_ext dans Hive qui pointera vers la table Marketing de Mongo DB

CREATE EXTERNAL TABLE Marketing\_ext ( id STRING, age INT, sexe STRING, taux INT, situationFamiliale STRING, nbEnfantsAcharge INT, 2emevoiture STRING)

STORED BY 'com.mongodb.hadoop.hive.MongoStorageHandler' WITH SERDEPROPERTIES('mongo.columns.mapping'='{"id":"\_id", "situationFamiliale" : "situationFamiliale", "nbEnfantsAcharge" : "nbEnfantsAcharge", "2emevoiture" : "2emevoiture"}') TBLPROPERTIES('mongo.uri'='mongodb://localhost:27017/projetBigData.Marketing');

#### On vérifie

#### SELECT \* FROM Marketing\_ext;

		taux	situationfamiliale	nbenfantsacharge	2emevoiture
 641028cef84f684cbe744cee	21	1396	Celibataire		false
641028cef84f684cbe744cf0	35	223	Celibataire		false
641028cef84f684cbe744cf2	48	401	Celibataire		false
641028cef84f684cbe744cf4		420	En Couple		true
641028cef84f684cbe744cf6		530	En Couple		false
641028cef84f684cbe744cf8	27	153	En Couple		false
641028cef84f684cbe744cfa		572	En Couple		false
641028cef84f684cbe744cfc	43	431	Celibataire		false
641028cef84f684cbe744cfe	64	559	Celibataire		false
641028cef84f684cbe744d00	22	154	En Couple		false
641028cef84f684cbe744d02	79	981	En Couple		false
641028cef84f684cbe744d04	55	588	Celibataire		false
641028cef84f684cbe744d06	19	212	Celibataire		false
641028cef84f684cbe744d08	34	1112	En Couple		false
641028cef84f684cbe744d0a		524	En Couple		true
641028cef84f684cbe744d0c	22	411	En Couple		true
641028cef84f684cbe744d0e	58	1192	En Couple		false
541028cef84f684cbe744d10	54	452	En Couple		true
541028cef84f684cbe744d12	35	589	Celibataire		false
641028cef84f684cbe744d14			En Couple		true

# Clients 4 ext

Nous allons créer une table externe Clients\_4\_ext qui pointe vers la table Clients\_4 de MongoDB

On drop au cas où la table existerait

#### DROP TABLE IF EXISTS Clients\_4\_ext;

On crée la table Clients\_4\_ext dans Hive qui pointera vers la table Clients\_4 de Mongo DB

CREATE EXTERNAL TABLE Clients\_4\_ext ( id STRING, age INT, sexe STRING, taux INT, situationFamiliale STRING, nbEnfantsAcharge INT, 2emevoiture STRING, immatriculation STRING)

STORED BY 'com.mongodb.hadoop.hive.MongoStorageHandler' WITH SERDEPROPERTIES('mongo.columns.mapping'='{"id":"\_id", "situationFamiliale" : "situationFamiliale", "nbEnfantsAcharge" : "nbEnfantsAcharge", "2emevoiture" : "2emevoiture"}') TBLPROPERTIES('mongo.uri'='mongodb://localhost:27017/projetBigData.Clients\_4');

On vérifie

### SELECT COUNT(\*) FROM Clients\_4\_ext;



### Clients 12 ext

Nous allons créer une table externe Clients\_12\_ext qui pointe vers la table Clients\_12 de MongoDB

On drop au cas où la table existerait

#### DROP TABLE IF EXISTS Clients\_12\_ext;

On crée la table Clients\_12\_ext dans Hive qui pointera vers la table Clients\_12 de Mongo DB

CREATE EXTERNAL TABLE Clients\_12\_ext ( id STRING, age INT, sexe STRING, taux INT, situationFamiliale STRING, nbEnfantsAcharge INT, 2emevoiture STRING, immatriculation STRING)

STORED BY 'com.mongodb.hadoop.hive.MongoStorageHandler'
WITH SERDEPROPERTIES('mongo.columns.mapping'='{"id":"\_id", "situationFamiliale" :
"situationFamiliale", "nbEnfantsAcharge" : "nbEnfantsAcharge", "2emevoiture" : "2emevoiture"}')
TBLPROPERTIES('mongo.uri'='mongodb://localhost:27017/projetBigData.Clients\_12');

On vérifie

#### SELECT COUNT(\*) FROM Clients\_12\_ext;



# CatalogueCO2

On drop la table au cas où elle existerait

DROP TABLE IF EXISTS CatalogueCO2;

On crée la table

#### CREATE TABLE IF NOT EXISTS CatalogueCO2(

id STRING, marque STRING, nom STRING, puissance INT, longueur STRING, nbPlaces INT, nbPortes INT, couleur STRING, occasion STRING, prix INT, avgBonusMalus INT, avgCO2 INT, avgCoutEnergie INT)

COMMENT 'CatalogueCO2 table intern table';

#### On met cette table en interne

INSERT OVERWRITE TABLE CatalogueCO2 SELECT c.id, c.marque, c.nom, c.puissance, c.longueur, c.nbPlaces, c.nbPortes, c.couleur, c.occasion, c.prix, CO2.avgBonusMalus, CO2.avgCO2, CO2.avgCoutEnergie FROM Catalogue c, CO2 where UPPER(c.marque) LIKE REPLACE(UPPER(CO2.marque), 'Z', '');

#### On vérifie

#### SELECT \* from CatalogueCO2 WHERE UPPER(margue) = "SAAB";

		150	longue		blanc		42	
641028cff84f684cbe744d98					gris			
541028cff84f684cbe744d9e								

# Vues

#### On passe à la création des vues.

DROP VIEW IF EXISTS clients\_ext;

CREATE VIEW clients\_ext AS

SELECT \*

FROM clients\_4\_ext

UNION ALL

SELECT \*

FROM clients\_12\_ext;

select \* from clients\_ext;

			1			
64102c05f84f684cbe7a51e0			871			8186LJ30
64102c05f84f684cbe7a51e2			1206			2529FF55
64102c05f84f684cbe7a51e4				Celibataire		9116TC64
64102c05f84f684cbe7a51e6			524	Celibataire		6019CB73
64102c05f84f684cbe7a51e8				Celibataire		607XD90
64102c05f84f684cbe7a51ea			1250	Celibataire		1410XX60
64102c05f84f684cbe7a51ec			1180	Celibataire		8705EI35
64102c05f84f684cbe7a51ee						7072UC48
64102c05f84f684cbe7a51f0			1120	Celibataire		8872HS65
64102c05f84f684cbe7a51f2						8919KD36
64102c05f84f684cbe7a51f4						596ZK95
64102c05f84f684cbe7a51f6			587	Celibataire		9885BQ70
64102c05f84f684cbe7a51f8			218	Celibataire		2655QZ35
64102c05f84f684cbe7a51fa						8577NX41
64102c05f84f684cbe7a51fc			520	Celibataire		1182FP76
64102c05f84f684cbe7a51fe				Celibataire		73733065
64102c05f84f684cbe7a5200			182			8750XC67
64102c05f84f684cbe7a5202				Celibataire		5604ZD76
64102c05f84f684cbe7a5204				Celibataire		3353CA65
64102c05f84f684cbe7a5206			432	Celibataire		
64102c05f84f684cbe7a5208	32			Celibataire		7073RW64
64102c05f84f684cbe7a520a						
64102c05f84f684cbe7a520c			417			657ND65
64102c05f84f684cbe7a520e				Celibataire		4016NZ55
64102c05f84f684cbe7a5210						6321W096
64102c05f84f684cbe7a5212						30990W82
64102c05f84f684cbe7a5214						6256BB86
64102c05f84f684cbe7a5216				Celibataire		
64102c05f84f684cbe7a5218						8537LB31
64102c05f84f684cbe7a521a			1327			9813JM16
64102c05f84f684cbe7a521c	21		1105			241WM57
64102c05f84f684cbe7a521e				Celibataire		669KP41
64102c05f84f684cbe7a5220						3828YT59
+						
196,984 rows selected (10.65	second	s)				

Voici ensuite une vue qui affiche la liaison entre la table immatriculations\_ext et l'immatriculation de la vue créée auparavant.

```
DROP VIEW IF EXISTS immatriculations clients ext;
CREATE VIEW immatriculations_clients_ext AS
SELECT immatriculations ext.immatriculation,
immatriculations ext.marque,
immatriculations_ext.nom,
immatriculations_ext.puissance,
immatriculations ext.longueur,
immatriculations_ext.nbplaces,
immatriculations ext.nbportes,
immatriculations_ext.couleur,
immatriculations_ext.occasion,
immatriculations ext.prix,
clients ext.age,
clients ext.sexe,
clients_ext.taux,
clients ext.situationFamiliale,
clients_ext.nbEnfantsAcharge,
clients ext.2emevoiture
FROM clients ext
JOIN immatriculations_ext ON clients_ext.immatriculation =
immatriculations ext.immatriculation;
```