

### TP TALEN ETL : Proposition de solution

MariaDB est la base de données utilisée sous Debian durant la réalisation de ce TP.

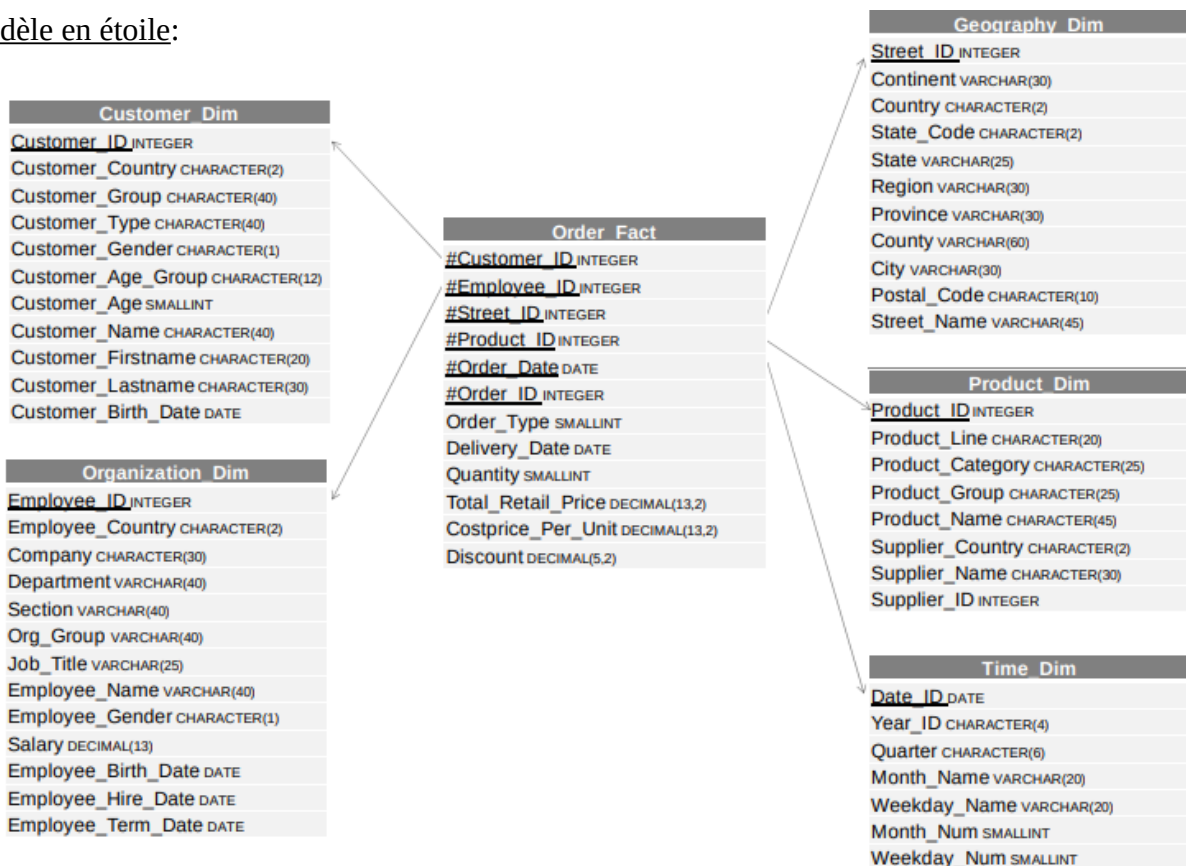
Le type des colonne Street\_ID et Product\_ID des tables respectives Geography\_Dim et Product\_Dim ont été modifié en BIGINT car en remplissant ces tables avec INTEGER la valeur maximale fut atteint.

Le fichier SQL/tables\_assoc.pdf contient le tableau des tables sources et cibles avec des remarques .

Le fichier SQL/tables.sql est le script de création des tables de l'entrepot.

Le fichier timedim.sql contient le script de remplissage du table Time\_Dim.

#### Modèle en étoile:



#### 1. Création de l'utilisateur orion\_DW\_user et les tables de l'entrepot

```
CREATE USER 'orion_DW_user' IDENTIFIED BY 'orion_DW_USER';  
CREATE DATABASE db_entrepot;  
GRANT ALL ON db_entrepot.* TO 'orion_DW_user'@'%' IDENTIFIED BY  
'orion_DW_USER' WITH GRANT OPTION;  
FLUSH PRIVILEGES;
```

SHOW GRANTS FOR orion\_DW\_user pour s'assurer de la création de l'utilisateur et des droits:

```
MariaDB [(none)]> SHOW GRANTS FOR orion_DW_user;
+-----+
| Grants for orion_DW_user@% |
+-----+
| GRANT USAGE ON *.* TO 'orion_DW_user'@'%' IDENTIFIED BY PASSWORD '*76D64CEB9C7A90 |
| GRANT ALL PRIVILEGES ON `db_entrepot`.* TO 'orion_DW_user'@'%' WITH GRANT OPTION |
+-----+
2 rows in set (0.000 sec)

MariaDB [(none)]> |
```

Le script de création des tables de l'entrepot est stocké dans un fichier entrepot.sql remis avec ce rapport. Toutefois nous visualisation les tables sous la session orion\_DW\_user:

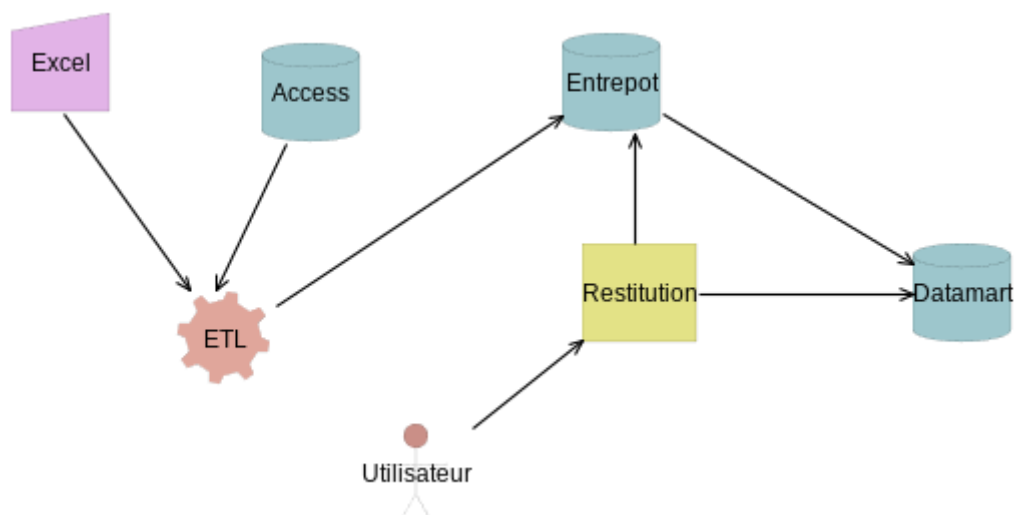
```
MariaDB [db_entrepot]> show tables;
+-----+
| Tables_in_db_entrepot |
+-----+
| Customer_Dim           |
| Geography_Dim          |
| Order_Fact             |
| Organization_Dim       |
| Product_Dim            |
| Time_Dim               |
+-----+
6 rows in set (0.446 sec)

MariaDB [db_entrepot]> |
```

## 2. Création du projet nommé orion\_project sous talend

Projet attaché avec ce cempte rendu.

## 3. Business model



#### 4. Spécification des données sources :

Etablissement de la connexion (nommée orion sous talend) à la base de données acces et importation du source orion.mdb.

## 5. Spécification des données cibles :

Etablissement de la connexion (nommée orion\_DW\_user sous talend) à la bae de donnée mariaDB (db\_entrepot) puis recupération des schémas des tables créer au préalable.

#### 6. Remplissage de la table Customer Dim :

Un fichier excel nommé tables\_assoc.xls, attaché avec ce rapport, contient les données sources et cibles ainsi que des remarques pour chaque table de l'entrepot.

## 7. Création de Job01 Customer Dim :

Le Job a été exécuté avec succès sous talend studio.

Vérification du résultat de l'exécution du job sous MariaDB en lançant les requêtes :

```
MariaDB [db_entrepot]> select count(*) from Customer_Dim;
+-----+
| count(*) |
+-----+
|      89954 |
+-----+
1 row in set (0.035 sec)

MariaDB [db_entrepot]> |
```

Les dix premiers lignes du table Customer\_Dim :

```
MariaDB [db_entrepot]> SELECT * FROM Customer_Dim LIMIT 10;
```

Customer_ID	Customer_Country	Customer_Group	Customer_Type	Customer_Gender	Customer_Age_Group	Customer_Age	Customer_Name	Customer_Firstname	Customer_Lastname	Customer_BirthDate
1	FR	20.0	Orion Club	Gold	members	high	activity	M	>75 years	79   Albert Collet
								Albert	Collet	1940-11-24

#### 8. Remplissage de la table Product\_Dim :

Le Job a été exécuté avec succès sous talend studio.

Vérification du résultat de l'exécution du job sous MariaDB :

```
MariaDB [db_entrepot]> SELECT COUNT(*) FROM Product_Dim;
+-----+
| COUNT(*) |
+-----+
|      5504 |
+-----+
1 row in set (0.005 sec)

MariaDB [db_entrepot]> |
```

Trois premières lignes de la table Product\_Dim

```
MariaDB [db_entrepot]> SELECT * FROM Product_Dim LIMIT 3;
+-----+-----+-----+-----+-----+
| Product_ID | Product_Line | Product_Category | Product_Group | Product_Name | Supplier_ID | Supplier_Country | Supplier_Name |
+-----+-----+-----+-----+-----+
| 210100100001 | Children | Children Outdoors | Outdoor things, Kids | Boy's and Girl's Ski Pants with Braces | 50 | NO | Scandinavian Clothing A/S |
+-----+-----+-----+-----+-----+
```

#### 9. Remplissage de la table Organization\_Dim :

Le Job a été exécuté avec succès sous talend studio.

Vérification des résultats sous MariaDB :

```
MariaDB [db_entrepot]> SELECT COUNT(*) FROM Organization_Dim;
+-----+
| COUNT(*) |
+-----+
|      1049 |
+-----+
1 row in set (0.002 sec)

MariaDB [db_entrepot]> |
```

Trois premières lignes de table Organization\_Dim

```
MariaDB [db_entrepot]> SELECT * FROM Organization_Dim LIMIT 3;
+-----+-----+-----+-----+-----+
| Employee_ID | Employee_Country | Company | Department | Section | Org_Group | Job_Title | Employee_Name | Employee_Gender | Salary | Employee_Birth_Date | Employee_Hire_Date | Employee_Term_Date |
+-----+-----+-----+-----+-----+
| 120101 | AU | Orion Australia | Sales Management | Sales Management | Sales Management | Director | Patrick Lu | M | 163040 | 1972-08-18 | 1999-07-01 | NULL |
| 120102 | AU | Orion Australia | Sales Management | Sales Management | Sales Management | Director | Patrick Lu | M | 163040 | 1972-08-18 | 1999-07-01 | NULL |
+-----+-----+-----+-----+-----+
```

#### 10. Remplissage de la table Time\_Dim :

La table Time\_Dim est remplie grâce au procédure nommé fillTimeDim() attaché avec le rapport sous le nom de timeDim.sql

L'exécution du procédure sous MariaDB donne:

```
MariaDB [db_entrepot]> call fillTimeDim();
Query OK, 1096 rows affected (56.032 sec)

MariaDB [db_entrepot]> |
```

Trois premières lignes de la table Time\_Dim:

```
MariaDB [db_entrepot]> select * from Time_Dim limit 3;
+-----+-----+-----+-----+-----+-----+-----+
| Date_ID | Year_ID | Quarter | Month_Name | Weekday_Name | Month_Num | Weekday_NUM |
+-----+-----+-----+-----+-----+-----+-----+
| 1998-01-01 | 1998 | 1 | January | Thursday | 1 | 5 |
| 1998-01-02 | 1998 | 1 | January | Friday | 1 | 6 |
| 1998-01-03 | 1998 | 1 | January | Saturday | 1 | 7 |
+-----+-----+-----+-----+-----+-----+-----+
3 rows in set (0.000 sec)

MariaDB [db_entrepot]> |
```

#### 11. Remplissage du table Geograpy\_Dim :

Le job a été exécuté avec succès et les résultats sont bien visibles sous MariaDB.

```
MariaDB [db_entrepot]> select count(*) from Geography_Dim;
+-----+
| count(*) |
+-----+
|      89807 |
+-----+
1 row in set (0.669 sec)

MariaDB [db_entrepot]> |
```

Les trois premières lignes de la table Geography\_Dim:

```
MariaDB [db_entrepot]> SELECT * FROM Geography_Dim LIMIT 3;
+-----+-----+-----+-----+-----+-----+-----+
| Street_ID | Continent | Country | State_Code | State | Region | Province |
| unty | City | Postal_Code | Street_Name |
+-----+-----+-----+-----+-----+-----+-----+
| 1400100001 | NULL | NULL | NULL | NULL | NULL | NULL |
| 1400100002 | NULL | NULL | NULL | NULL | NULL | NULL |
| 1400100003 | NULL | NULL | NULL | NULL | NULL | NULL |
+-----+-----+-----+-----+-----+-----+-----+
```

## 12. Remplissage du table Order\_Fact :

Le job a été bien exécuté et a insérer les données.

Visualisation des données ayant été enregistrer dans la table.

```
MariaDB [db_entrepot]> select count(*) from Order_Fact;
+-----+
| count(*) |
+-----+
|    552700 |
+-----+
1 row in set (6.268 sec)

MariaDB [db_entrepot]> |
```

Les 3 premières lignes :

```
MariaDB [db_entrepot]> select * from Order_Fact limit 3;
+-----+-----+-----+-----+-----+-----+-----+
| Customer_ID | Employee_ID | Street_ID | Product_ID | Order_Date | Order_ID |
| Order_Type | Delivery_Date | Quantity | Total_Retail_Proce | Costprice_Per_ |
| Discount |
+-----+-----+-----+-----+-----+-----+-----+
| 1 | 120349 | 3500101176 | 240100100281 | 2000-04-10 | 123544155 |
| 1 | 2000-04-10 | 1 | 41.30 | 10 |
| NULL |
| 1 | 120349 | 3500101176 | 240100300006 | 2000-04-10 | 123544155 |
| 1 | 2000-04-10 | 2 | 106.00 | 23 |
| NULL |
| 1 | 120353 | 3500101176 | 210201000167 | 2000-02-13 | 123486454 |
| 1 | 2000-02-13 | 1 | 33.80 | 15 |
| NULL |
```