

# Rapport du projet

## Data Warehouse



Réalisé par :

Walid Taoutaou

Brahim Oulhaj

Encadré par :

Pr. Guénaël Cabanes

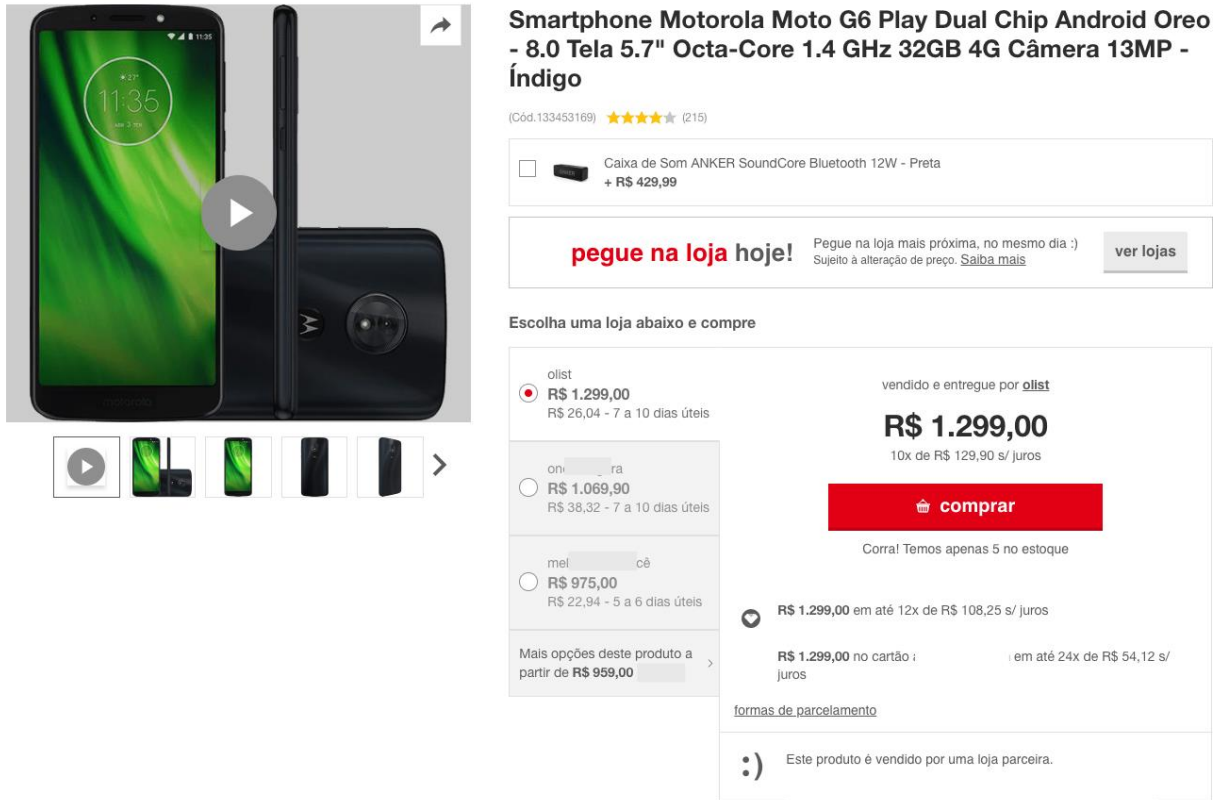
# Sommaire

- I. Introduction et problématique
- II. Modèle E/A
- III. Modèles dimensionnels, Data Marts
- IV. E.T.L et Data Warehouse
- V. Reporting
- VI. Conclusion
- VII. Annexe : la structure du fichier ZIP

# Introduction et problématique

Le processus décisionnel est un projet qui se construit, il est né d'un besoin exprimé par les entreprises à cause du volume important des données à manipuler et à analyser car le système traditionnel ne suffit plus pour pérenniser l'activité de cette dernière, car celui-ci convient bien aux applications gérant l'activité quotidienne de l'entreprise, mais s'avère inadapté au Décisionnel. En entreposant les données, le processus décisionnel apporte la solution au problème de la croissance continue des données provenant de différentes sources et de différents formats en les homogénéisant et en les organisant dans des data mart où elles sont historiées, résumées et consolidées. Le volume de données des entrepôts est important et va de centaines de giga-octets à des téraoctets, voir même encore davantage de nos jours.

Notre projet a été réalisé pour faire des statistiques pour un magasin brésilien en ligne, [www.olist.com](http://www.olist.com), dont le but d'améliorer le nombre de ventes et d'augmenter le chiffre d'affaire de ce magasin.



**Smartphone Motorola Moto G6 Play Dual Chip Android Oreo - 8.0 Tela 5.7" Octa-Core 1.4 GHz 32GB 4G Câmera 13MP - Índigo**

(Cód.133453169) ★★★★★ (215)

☐ Caixa de Som ANKER SoundCore Bluetooth 12W - Preta + R\$ 429,99

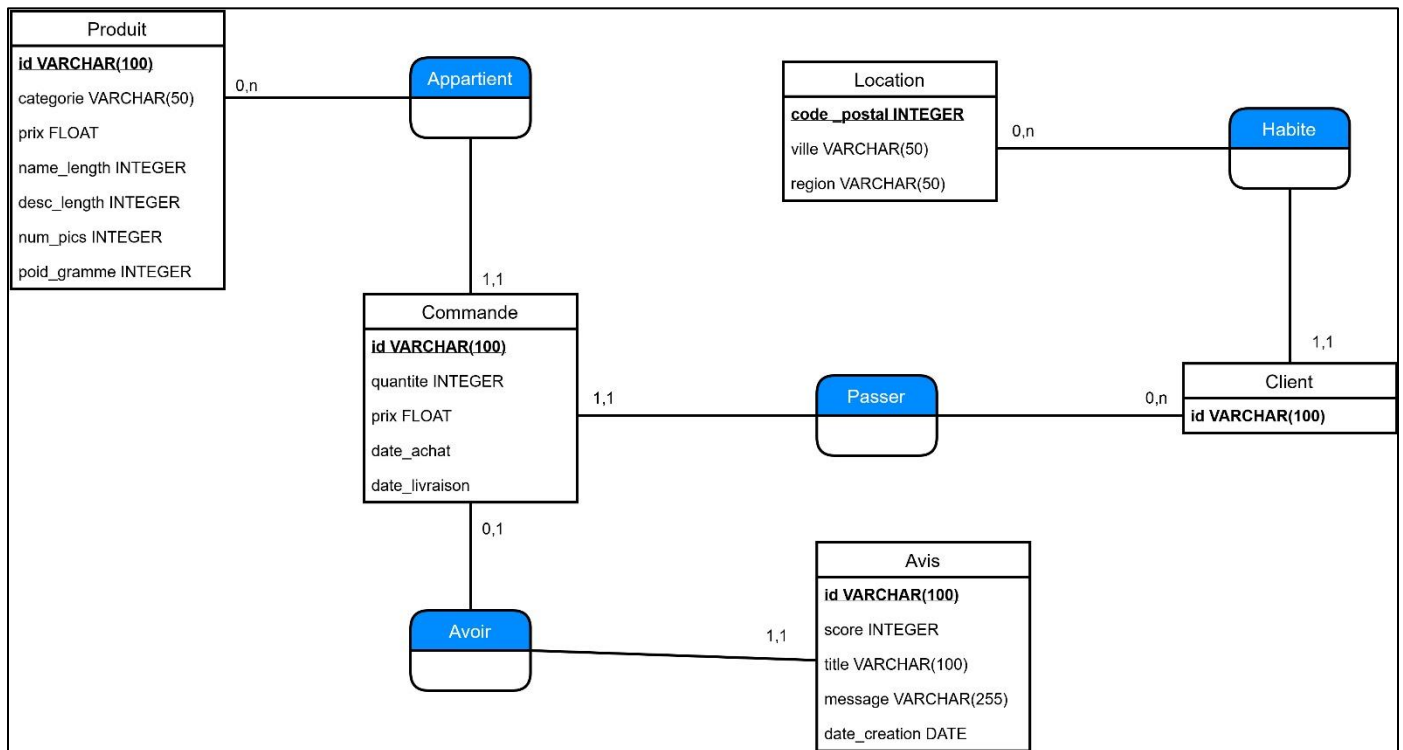
**pegue na loja hoje!** Pegue na loja mais próxima, no mesmo dia :) Sujeito à alteração de preço. [Saiba mais](#) [ver lojas](#)

Escolha uma loja abaixo e compre

olista	vendido e entregue por <b>olista</b>
<input checked="" type="radio"/> <b>R\$ 1.299,00</b> R\$ 26,04 - 7 a 10 dias úteis	<b>R\$ 1.299,00</b> 10x de R\$ 129,90 s/ juros
<input type="radio"/> on-line <b>R\$ 1.069,90</b> R\$ 38,32 - 7 a 10 dias úteis	<b>comprar</b>
<input type="radio"/> mais <b>R\$ 975,00</b> R\$ 22,94 - 5 a 6 dias úteis	Corra! Temos apenas 5 no estoque
Mais opções deste produto a partir de <b>R\$ 959,00</b>	<b>R\$ 1.299,00 em até 12x de R\$ 108,25 s/ juros</b>
	<b>R\$ 1.299,00 no cartão</b>   em até 24x de R\$ 54,12 s/ juros
	<a href="#">formas de parcelamento</a>
	: ) Este produto é vendido por uma loja parceira.

# Modèle Entité-Association

Olist est un magasin qui permet les clients, dans tous les régions du brésil, d'acheter des produits en ligne. Chaque client a un identifiant et une location, la location se caractérise par un code postal, nom de la ville et la région. Le client peut commander des produits en choisi la quantité qui veut dans une date précise et le produit va être livrer dans une autre date. Les produits sont caractérisés par un identifiant, une catégorie, un nom, une description du produit, nombre des photos de ce produit, le poids en gramme et la longueur et la largeur. Une commande peut avoir un avis (Review) qui se caractérise par un ID, le Score (de 0 à 5), le titre, le message ou le commentaire et la date d'écriture du commentaire.



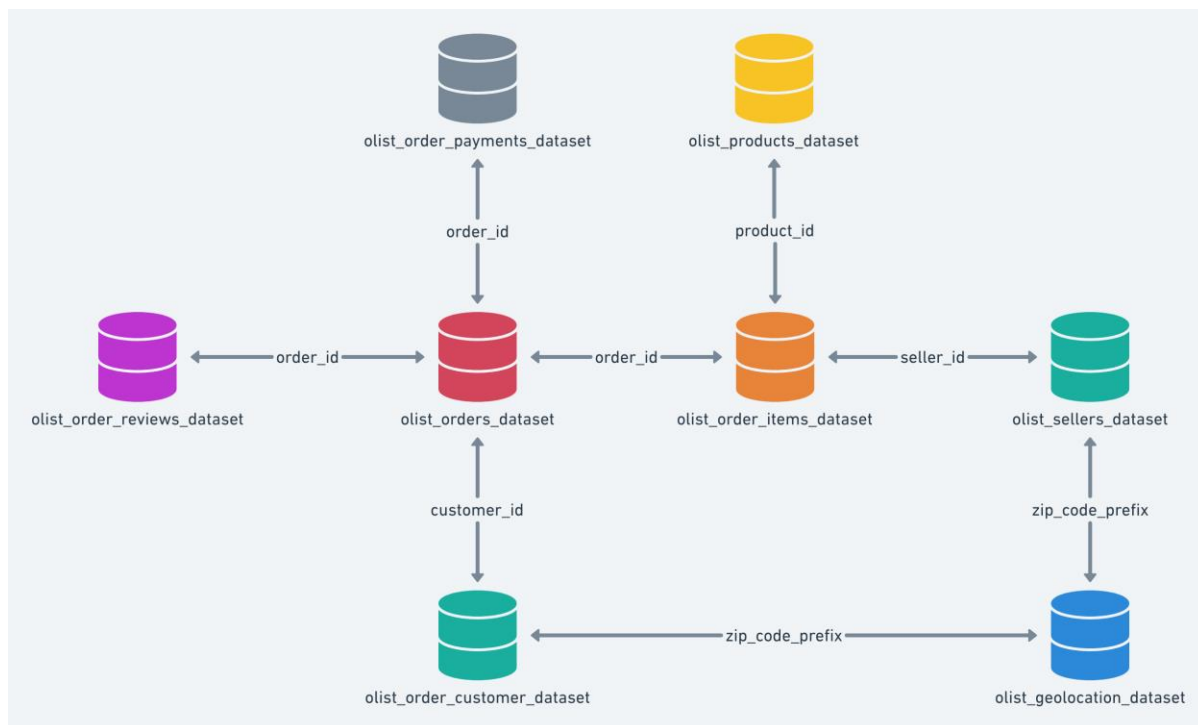
Modèle Entité/Association

# Modèles dimensionnels

## 1. Analyse des données :

Cet ensemble de données a été fourni par Olist, le plus grand magasin du marché brésilien qui relie les petites entreprises de tout le Brésil. Les commerçants peuvent vendre leurs produits par l'intermédiaire du magasin Olist et les expédier directement aux clients en utilisant les partenaires logistiques Olist.

Il s'agit d'un ensemble de données publiques brésiliennes sur le e-commerce des commandes effectuées à Olist Store. L'ensemble de données contient des informations, en format CSV, sur les 100 k commandes de 2016 à 2018 effectuées sur de multiples marchés au Brésil. Ses caractéristiques permettent de visualiser un ordre de multiples dimensions : de l'état de l'ordre, le prix, les performances de paiement et de fret à l'emplacement du client, les attributs du produit et enfin les commentaires écrits par les clients. Ainsi, un ensemble de données de géolocalisation qui relie les codes postaux brésiliens aux coordonnées lat/long.



## 2. Analyse des besoins :

Dans ce projet on considère qu'on a deux départements, le département informatique et le département logistique.

Le département logistique cherche à savoir si le chiffre d'affaire ( $CA = \text{Quantité} \times \text{Prix}$ ) s'influence en fonction de certaines régions et certaines villes, et en fonction des jours de la semaine (jds) et des jours du mois (jdm) et aussi de certaines périodes de l'année. Il voudrait aussi savoir si le nombre des jours de livraison (jdl) a une influence sur le CA.

### Requêtes brutes :

Chiffre d'affaire  
/région  
/ville

Chiffre d'affaire  
/jds  
/jdm  
/mois

Chiffre d'affaire  
/jdl

### Requêtes organisées par dimensions :

Chiffre d'affaire  
/lieux (région, ville)  
/dates (jds, jdm, mois, jdl)

Le département informatique voudrait savoir la nature des produits qui ont un CA élevé et si la description et le nombre des photos du produit dans le site web a une influence sur le CA. Il aimerait également savoir si les commentaires des clients s'influencent sur le CA. Ainsi qu'il voudrait aussi savoir les jours de la semaine et les mois de l'année dont les clients sont très actifs.

### Requêtes brutes :

Chiffre d'affaire  
/catégorie  
/description  
/nbr\_photos

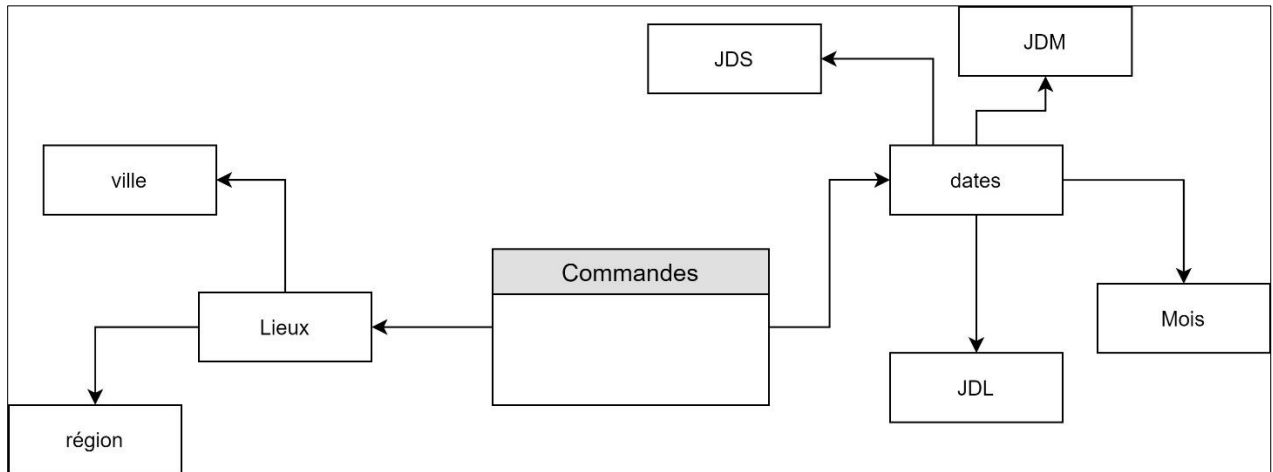
Chiffre d'affaire  
/commentaires

Chiffre d'affaire  
/jds  
/mois

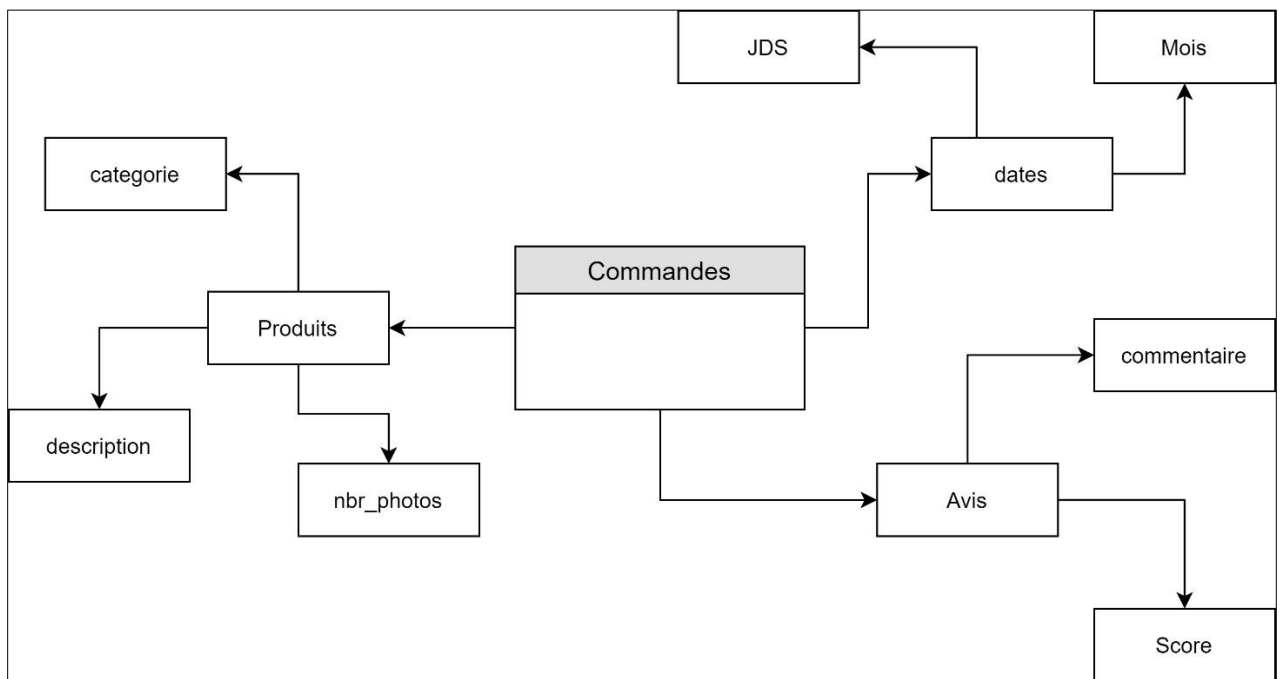
### Requêtes organisées par dimensions :

Chiffre d'affaire  
/produit (catégorie, description, nbr\_photos)  
/avis (commentaire, score)  
/dates (jds, mois)

### 3. Modèles dimensionnelles/Data Marts:

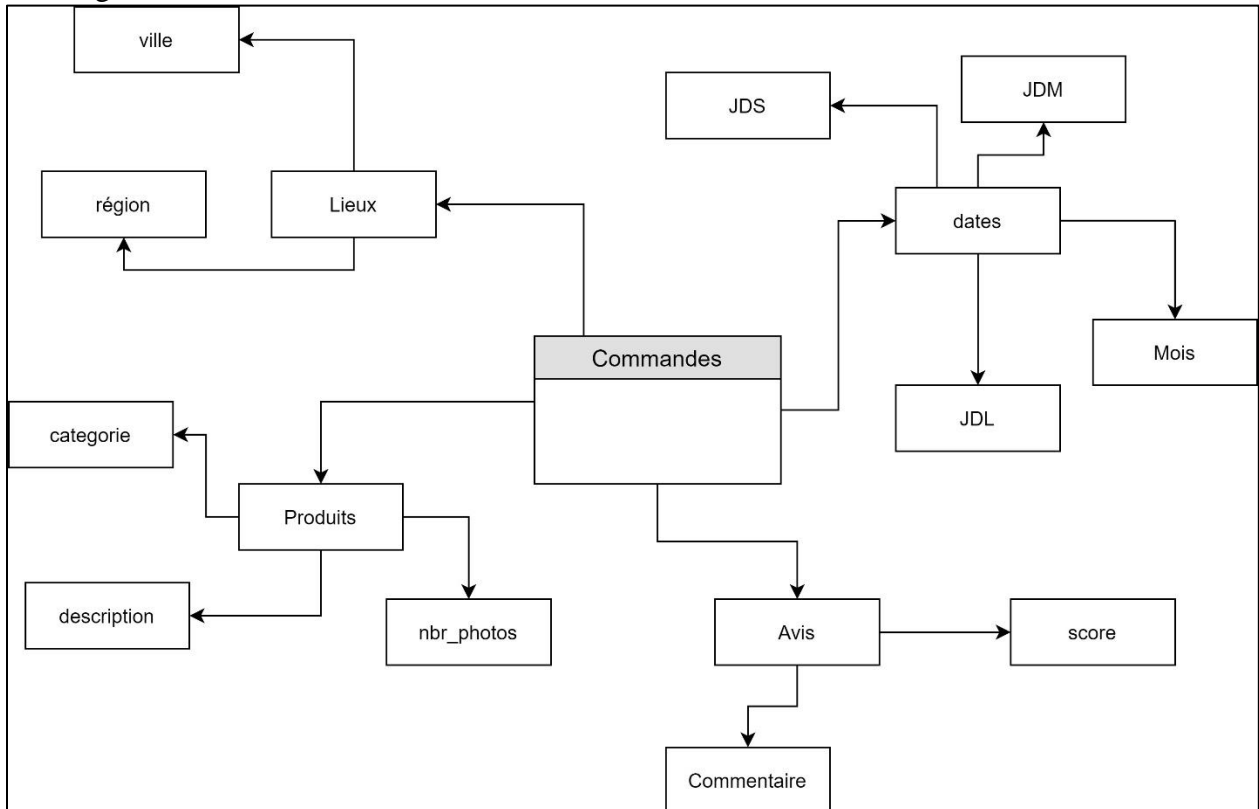


*Vue Département Logistique*



*Vue Département Informatique*

En intégrant les deux modèles on va voir un modèle de Data Warehouse



#### 4. Description des tables de dimensions:

Attribut	Type	Commentaire
#id	NUMBER	Clé primaire et auto incrémente
Date_id	DATE	La date de la commande
Year	NUMBER	L'année de la commande
Month	VARCHAR(20)	Le mois (jan, ..., déc.)
Day_of_week	VARCHAR(20)	Le jour de la semaine
Day_of_month	NUMBER	Le jour du mois (de 1 à 31)
Livraison_days	NUMBER	Nombre des jours entre la date de la commande et la date de livraison
<i>Description de la table de dimension Dates</i>		

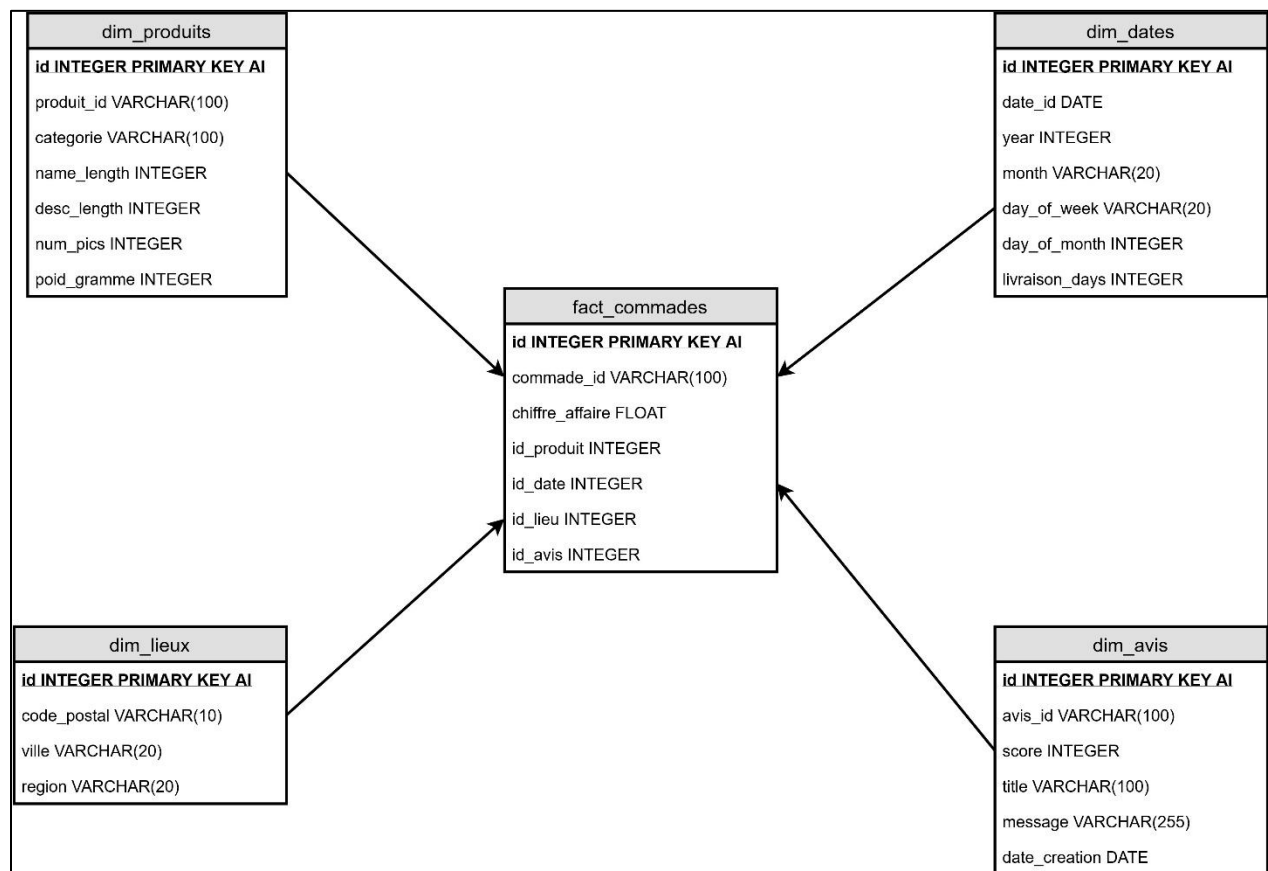
Attribut	Type	Commentaire
#id	NUMBER	Clé primaire et auto incrémente
Code_postal	VARCHAR(10)	Le code postal de chaque lieu (zip code)
Ville	VARCHAR(20)	
Region	VARCHAR(20)	
<i>Description de la table de dimension Lieux</i>		



Attribut	Type	Commentaire
#id	NUMBER	Clé primaire et auto incrémente
Produit_id	VARCHAR(100)	La référence du produit
categorie	VARCHAR(100)	La catégorie dont qu'il appartient
Name_length	NUMBER	La taille du nom du produit
Desc_length	NUMBER	La taille de la description du produit
Num_pics	NUMBER	Le nombre des photos du produit publier en ligne
Poids_gramme	NUMBER	Le poids de produit en Gramme
<i>Description de la table de dimension Produits</i>		

Attribut	Type	Commentaire
#id	NUMBER	Clé primaire et auto incrémente
Avis_id	VARCHAR(100)	La référence de l'avis
Score	NUMBER	Les points donnée par le client à cette commande (0..5)
Title	VARCHAR(100)	Le titre du commentaire
Message	VARCHAR(255)	Le contenu du commentaire
Date_creation	DATE	La date d'écriture de commentaire
<i>Description de la table de dimension Avis</i>		

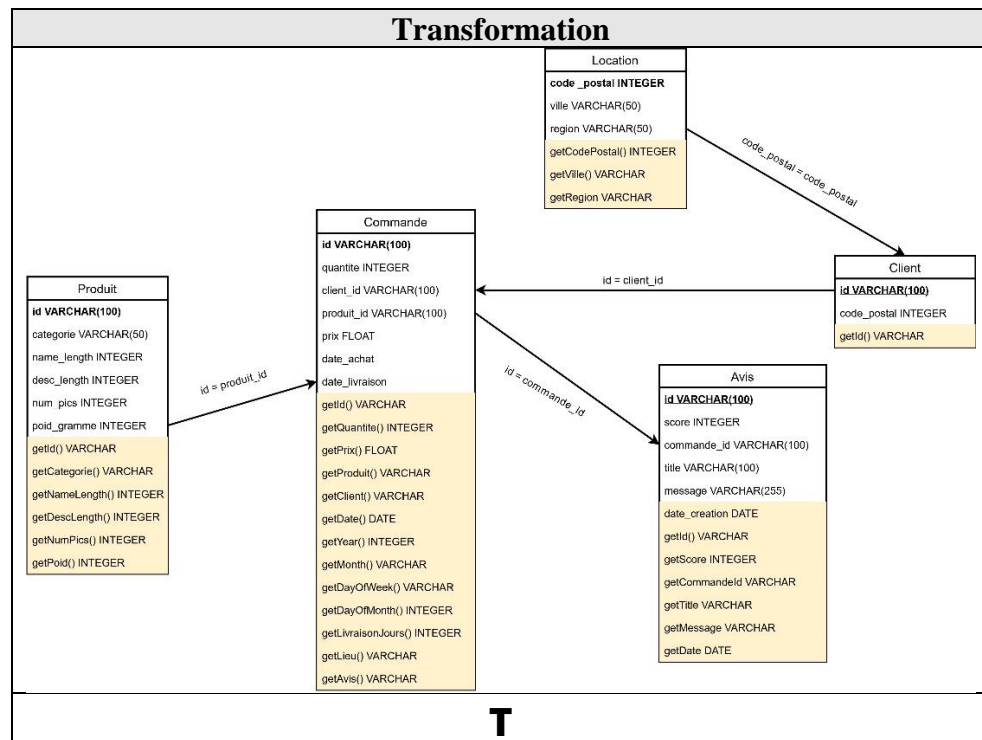
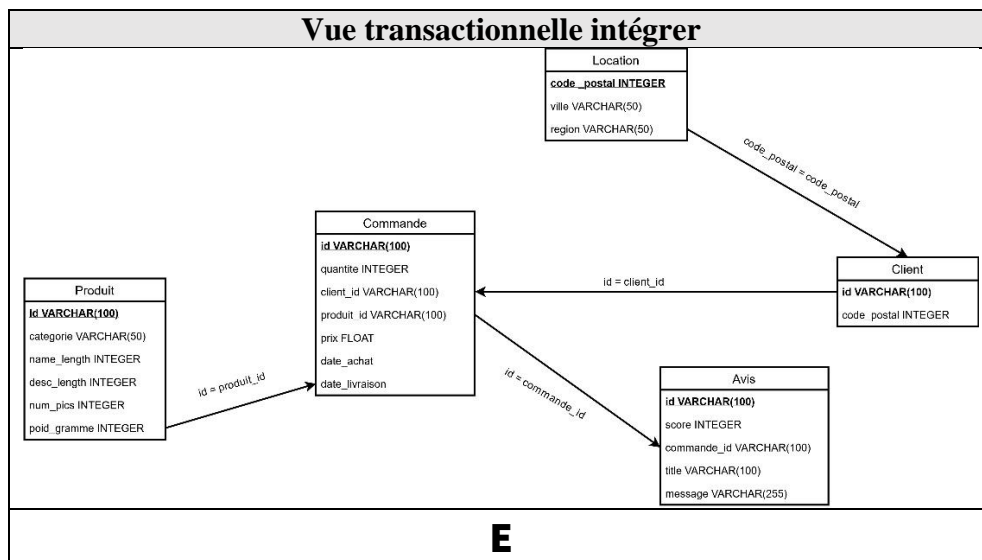
Nous pouvons aussi créer le modèle de Data Warehouse comme ceci

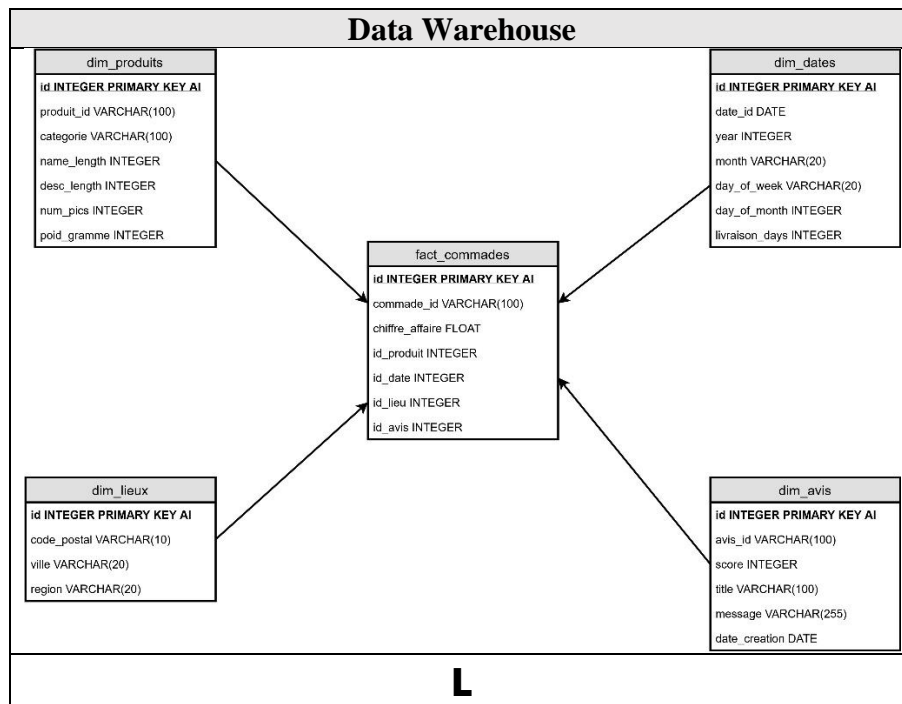


# ET.L

## Extraction, Transformation, Loading

Dans cette partie on va charger les données de Olist Store de puis les fichiers CSV, téléchargé depuis Kaggle ([https://www.kaggle.com/olistbr/brazilian-ecommerce#olist\\_order\\_items\\_dataset.csv](https://www.kaggle.com/olistbr/brazilian-ecommerce#olist_order_items_dataset.csv)), et les mettre dans des tables relationnelle de base de données Oracle. Puis on va créer des tables objet-relationnelle (SQL3) qui ont des fonctions qui permet de transformer les données pour les mettre dans notre Data Warehouse.





Ci-dessous l'exemple de code source de la table « Dates » depuis la partie extraction jusqu'à le chargement des données dans le Data Warehouse :

Tous d'abord on a créé deux dossiers, un dossier source « src » qui contient les fichiers CSV et un autre « log » qui va contenir les fichiers historique Log.

```
CREATE OR REPLACE DIRECTORY dataSrc AS 'E:/dwh/projet_dwh/dataset/brazilian-ecommerce/';
CREATE OR REPLACE DIRECTORY dataLog AS 'E:/dwh/projet_dwh/log/';
```

Puis, on a chargé les données du fichiers CSV, qui contient les dates (date d'achat et date de livraison) qui nous concerne, et on les mettre dans une table dans la base de données Oracle :

```
CREATE TABLE orders (
  order_id VARCHAR2(100),
  customer_id VARCHAR2(100),
  order_status VARCHAR2(50),
  order_purchase_timestamp VARCHAR2(25),
  order_approved_at VARCHAR2(25),
  order_delivered_carrier_date VARCHAR2(25),
  order_delivered_customer_date VARCHAR2(25),
  order_estimated_delivery_date VARCHAR2(25)
)
ORGANIZATION EXTERNAL(
  TYPE ORACLE_LOADER
  DEFAULT DIRECTORY dataSrc
  ACCESS PARAMETERS(
    RECORDS DELIMITED BY '\n'
    SKIP 1
    CHARACTERSET UTF8
    BADFILE dataLog:'orders.bad'
    LOGFILE dataLog:'orders.log'
    FIELDS TERMINATED BY ','
    OPTIONALLY ENCLOSED BY '"'
  )
  LOCATION('olist_orders_dataset.csv')
)
REJECT LIMIT UNLIMITED;
```

C:\Windows\system32\cmd.exe - sqlplus	99434	cfa78b997e3	a2f7428f0caf	delivered	20-12-17 09:52
SQL> SELECT COUNT(*) FROM orders;	99435	9115830be80	da2124f134f	delivered	04-10-17 19:57
COUNT(*)	99436	aa04ef52145	f01a6bfcc730	delivered	27-01-17 00:30
-----	99437	880675dff21	47cd45a6ac7	delivered	23-02-17 09:05
99441	99438	9c5dedf39a9	39bd1228ee8	delivered	09-03-17 09:54
SQL>	99439	63943bddc26	1fca14ff2861	delivered	06-02-18 12:58
	99440	83c1379a015	1aa71eb0421	delivered	27-08-17 14:46
	99441	11c177c8e97	b331b74b18c	delivered	08-01-18 21:28
	99442	66dea50a8b1	edb027a75a1	delivered	08-03-18 20:57
	99443				

Jusqu'à maintenant tout se passe bien et les données sont bien chargées depuis le fichier CSV. Maintenant, on va aborder la partie de transformation

```
CREATE OR REPLACE TYPE type_commande AS OBJECT (
  cid VARCHAR2(100),
  quantite VARCHAR2(4),
  prix VARCHAR2(10),
  produit_id VARCHAR2(100),
  client_id VARCHAR2(100),
  date_achat VARCHAR2(25),
  date_livraison VARCHAR2(25),
  MEMBER FUNCTION getId RETURN VARCHAR2,
  MEMBER FUNCTION getQuantite RETURN NUMBER,
  MEMBER FUNCTION getPrix RETURN NUMBER,
  MEMBER FUNCTION getProduit RETURN VARCHAR2,
  MEMBER FUNCTION getClient RETURN VARCHAR2,
  MEMBER FUNCTION getDate RETURN DATE,
  MEMBER FUNCTION getYear RETURN NUMBER,
  MEMBER FUNCTION getMonth RETURN VARCHAR2,
  MEMBER FUNCTION getDayOfWeek RETURN VARCHAR2,
  MEMBER FUNCTION getDayOfMonth RETURN NUMBER,
  MEMBER FUNCTION getLivraisonJours RETURN NUMBER,
  MEMBER FUNCTION getLieu RETURN VARCHAR2,
  MEMBER FUNCTION getAvis RETURN VARCHAR2
);
/
```

Chaque fonction permet de transformer les données de la table et retourner des nouvelles données dont on a besoin.

```
CREATE OR REPLACE TYPE BODY type_commande AS
.
.
MEMBER FUNCTION getDayOfWeek RETURN VARCHAR2 IS
  ldate DATE;
BEGIN
  ldate := TO_DATE(SUBSTR(date_achat, 1, 10), 'yyyy-mm-dd');
  RETURN TO_CHAR(ldate, 'Day');
END;
.
MEMBER FUNCTION getLivraisonJours RETURN NUMBER IS
  dateachat DATE := TO_DATE(SUBSTR(date_achat, 1, 10), 'yyyy-mm-dd');
  datelivraison DATE := TO_DATE(SUBSTR(date_livraison, 1, 10), 'yyyy-mm-dd');
BEGIN
  RETURN datelivraison-dateachat;
END;
END;
/
```

Après la création de type et la définition des fonctions membres et la création de la table de ce type « commandes » on va charger les données dans cette table en créant une vue « view\_orders » qui prendra les données depuis la jointure des tables « orders », puis on va les enregistrer dans la table « commandes » :

```
CREATE TABLE commandes OF type_commande;

DROP VIEW view_orders;
CREATE VIEW view_orders AS
SELECT oi.ORDER_ID, oi.ORDER_ITEM_ID, oi.PRICE, oi.PRODUCT_ID,
o.CUSTOMER_ID, o.ORDER_PURCHASE_TIMESTAMP, o.ORDER_DELIVERED_CUSTOMER_DATE
FROM orders o JOIN order_items oi
ON (o.order_id=oi.order_id)
WHERE o.order_status = 'delivered';

INSERT INTO commandes SELECT * FROM view_orders;
```

```
SQL> select count(*) from commandes;

COUNT(*)
-----
110197

SQL>
```

On a terminé la partie transformation, on va passer maintenant à la partie « Load » pour mettre les données dans la table de dimension « dates » en utilisant le code suivant :

```
CREATE TABLE dim_dates (
    id NUMBER GENERATED BY DEFAULT ON NULL AS IDENTITY START WITH 1,
    date_id DATE,
    year NUMBER(4),
    month VARCHAR(20),
    day_of_week VARCHAR(20),
    day_of_month NUMBER(2),
    livraison_days NUMBER
);

INSERT INTO dim_dates(date_id, year, month, day_of_week, day_of_month, livraison_days)
SELECT DISTINCT c.getDate(), c.getYear(), c.getMonth(),
c.getDayOfWeek(), c.getDayOfMonth(), c.getLivraisonJours()
FROM commandes c;

ALTER TABLE dim_dates ADD CONSTRAINT dim_dates_pk PRIMARY KEY (id);
```

```
SQL> select count(*) from dim_dates;

COUNT(*)
-----
15405
```

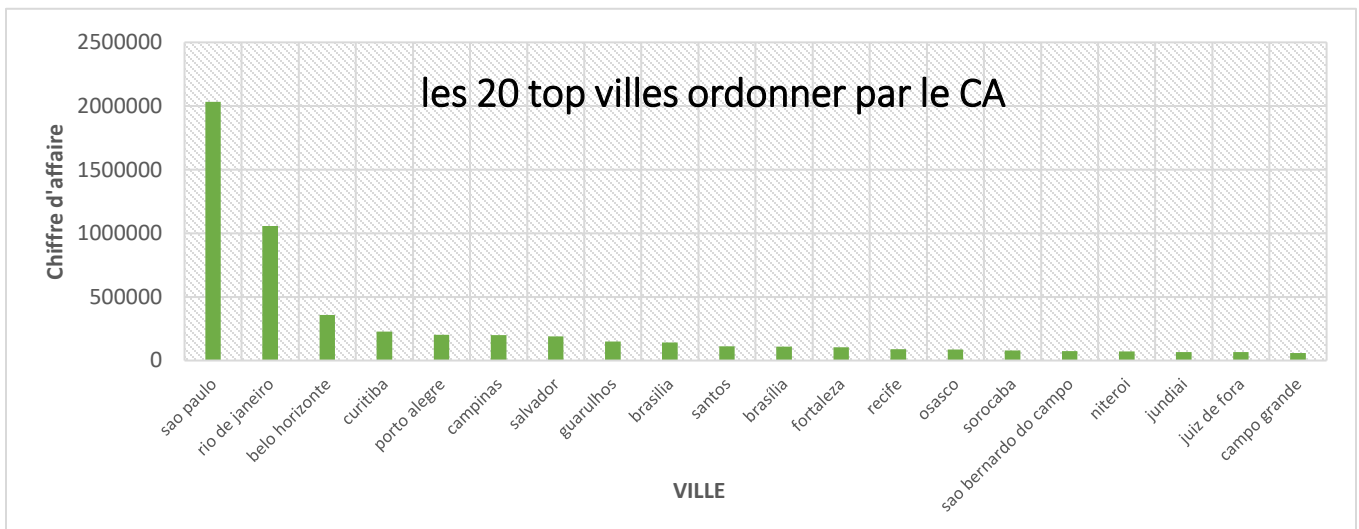
On va faire le même processus pour tous les autres tableaux et on termine la partie ETL.

# Reporting

Dans ce chapitre on va présenter quelques statistiques répondent aux questions des départements pour les permettre à prendre des décisions.

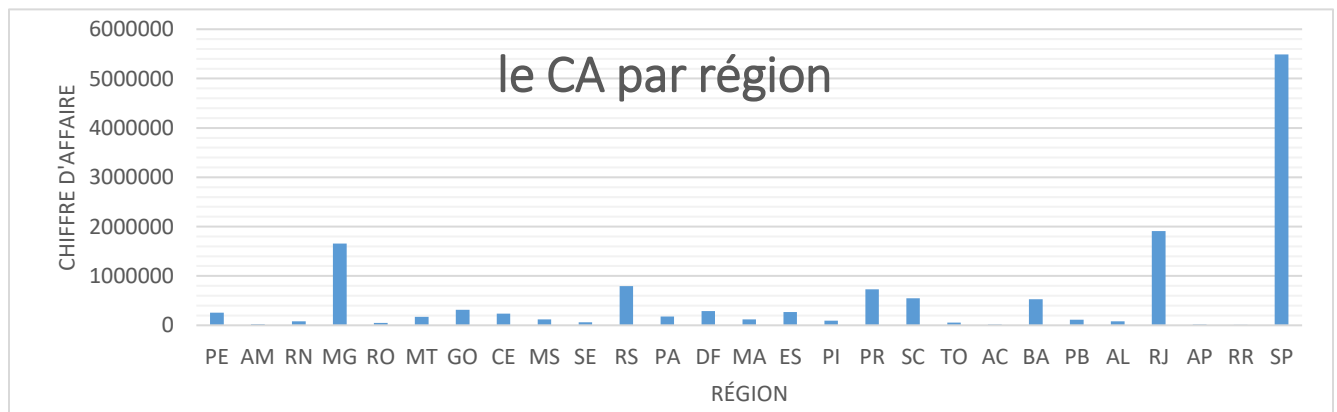
- **Quelle sont les 20 premières villes dans les régions du brésil ayant un chiffre d'affaire élevé ?**

```
SELECT 'VILLE', 'CA' FROM dual;
SELECT * FROM (
  SELECT dl.ville, SUM(f.chiffre_affaire) ca
  FROM fact_commandes f JOIN dim_lieux dl ON dl.id = f.id_lieu
  GROUP BY dl.ville ORDER BY SUM(f.chiffre_affaire) DESC
) WHERE ROWNUM <= 20;
```



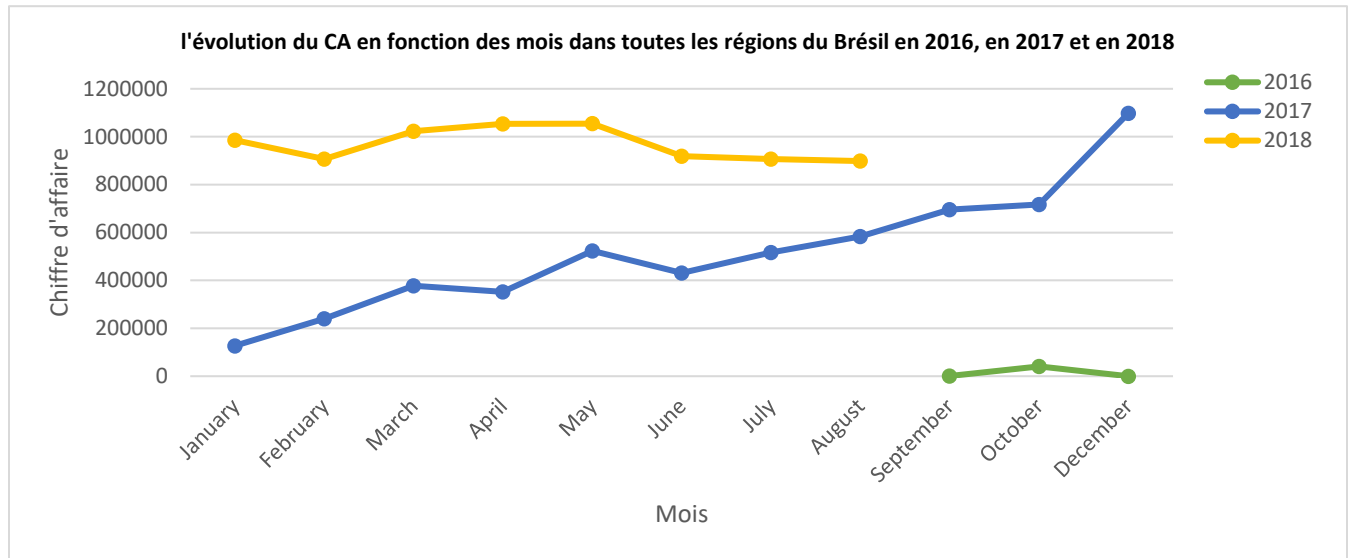
- **Quelle est le chiffre d'affaire de chaque région de Brésil ?**

```
SELECT 'REGION', 'CA' FROM dual;
SELECT dl.region, SUM(f.chiffre_affaire) ca
FROM fact_commandes f JOIN dim_lieux dl ON dl.id = f.id_lieu
GROUP BY dl.region;
```



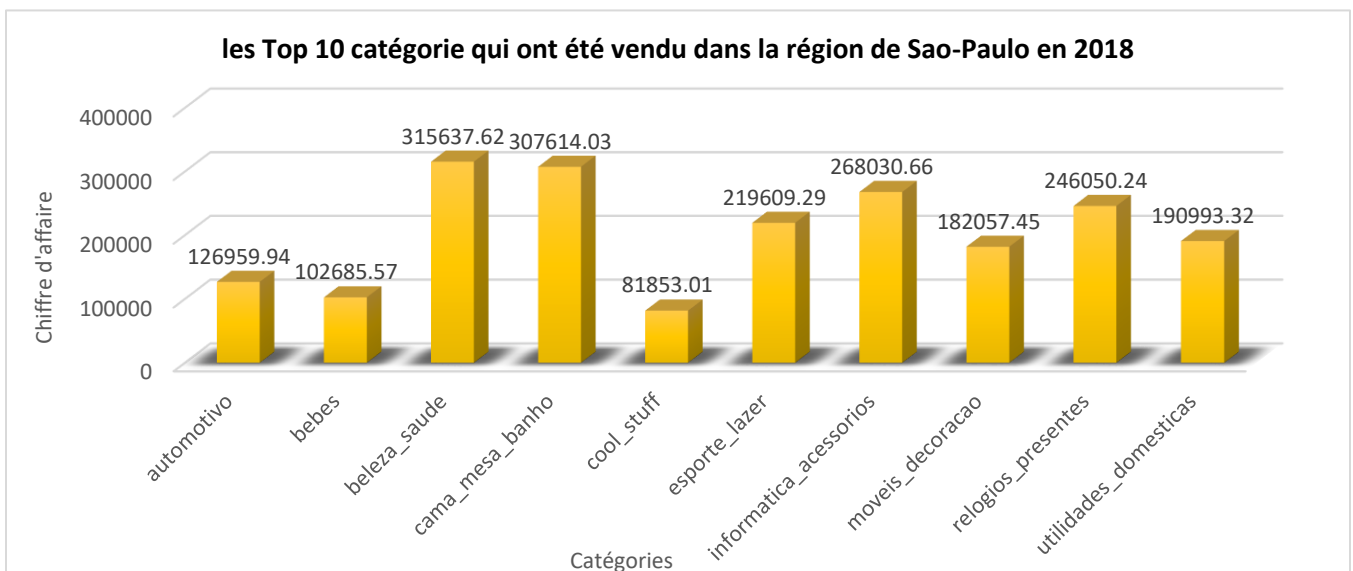
- Comment le chiffre d'affaire s'évalue en fonction des mois dans tous les régions en 2016, 2017 et 2018 ?

```
SELECT 'ANNEE', 'MOIS', 'CA' FROM dual;
SELECT dd.year, dd.month, SUM(fc.chiffre_affaire) ca
FROM fact_commandes fc JOIN dim_dates dd ON fc.id_date = dd.id
GROUP BY dd.year, dd.month, EXTRACT(MONTH FROM dd.date_id)
ORDER BY dd.year, EXTRACT(MONTH FROM dd.date_id);
```



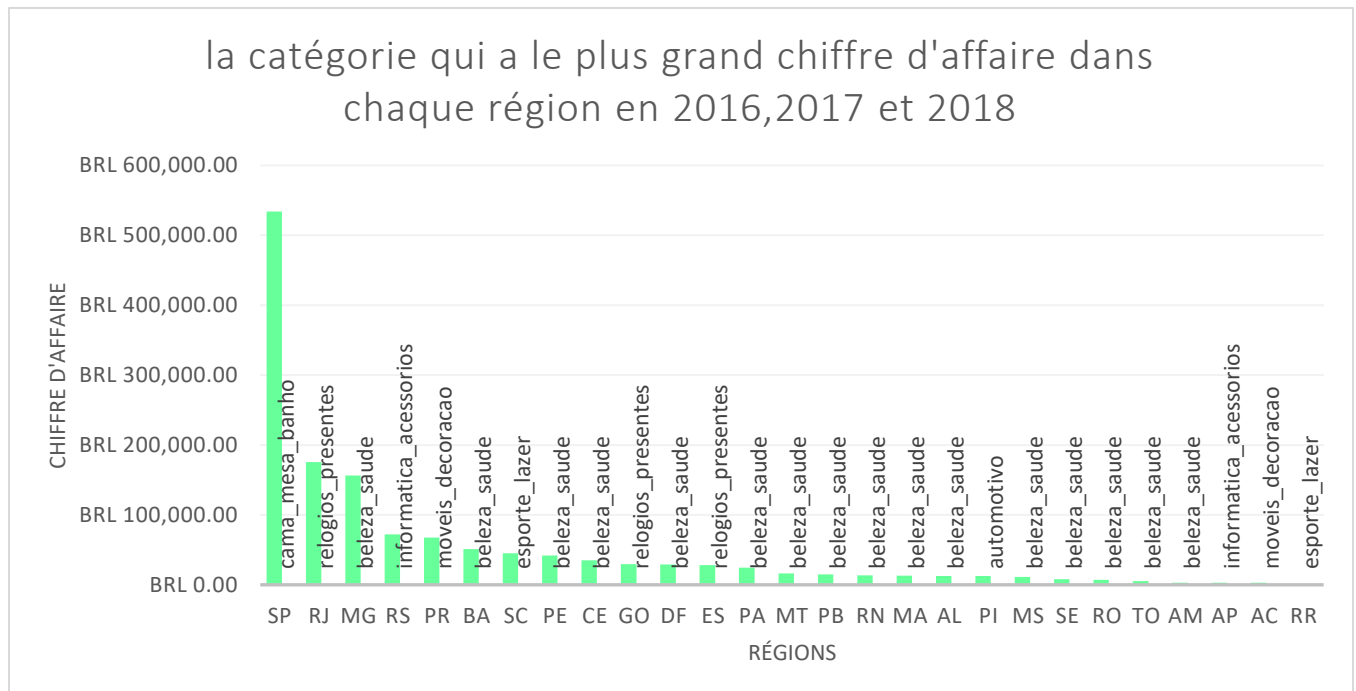
- Quelle sont les catégories des produits qui ont été vendu dans la région de Sao Paulo en 2018 ?

```
SELECT * FROM(
SELECT dp.categorie cat, SUM(fc.chiffre_affaire) ca FROM fact_commandes fc
JOIN dim_produits dp ON fc.id_produit = dp.id JOIN dim_dates dd ON fc.id_date = dd.id
JOIN dim_lieux dl ON fc.id_lieu = dl.id
WHERE dd.year = 2018 AND dl.region = 'SP' AND categorie IS NOT NULL
GROUP BY dp.categorie ORDER BY ca DESC
) WHERE ROWNUM <= 10;
```



- En 2016 - 2018, quelle est la catégorie qui a le plus grand chiffre d'affaire dans chaque région ?

```
SELECT reg AS REGION, cat AS CATEGORIE, ca AS CHIFFRE_AFFAIRE
FROM(
  SELECT l.region reg, p.categorie cat, sum(chiffre_affaire) ca,
  RANK() OVER(PARTITION BY l.region ORDER BY sum(chiffre_affaire) DESC) ranking
  FROM fact_commandes
  JOIN dim_produits p ON (id_produit = p.id)
  JOIN dim_lieux l ON (id_lieu = l.id)
  GROUP BY l.region, p.categorie
)
WHERE ranking = 1
ORDER BY ca DESC;
```



- Quelle est le nombre des jours de livraison des 10 produits qui ont les plus grands chiffre d'affaire (CA) et qui ont les plus petits CA?

Dans cette question nous voulons savoir l'influence des jours de livraison sur le chiffre d'affaire.

```
SELECT produit_id, categorie, livraison_days, ca
FROM(
  SELECT f.commande_id, SUM(f.chiffre_affaire) ca, p.produit_id, p.categorie, d.livraison_days
  FROM fact_commandes f JOIN dim_produits p ON p.id = f.id_produit
  JOIN dim_dates d ON d.id = f.id_date
  GROUP BY f.commande_id, p.produit_id, p.categorie, d.livraison_days
  ORDER BY 2 DESC --ORDER BY 2 dans Le cas des 10 derniers
) WHERE ROWNUM <= 10;
```

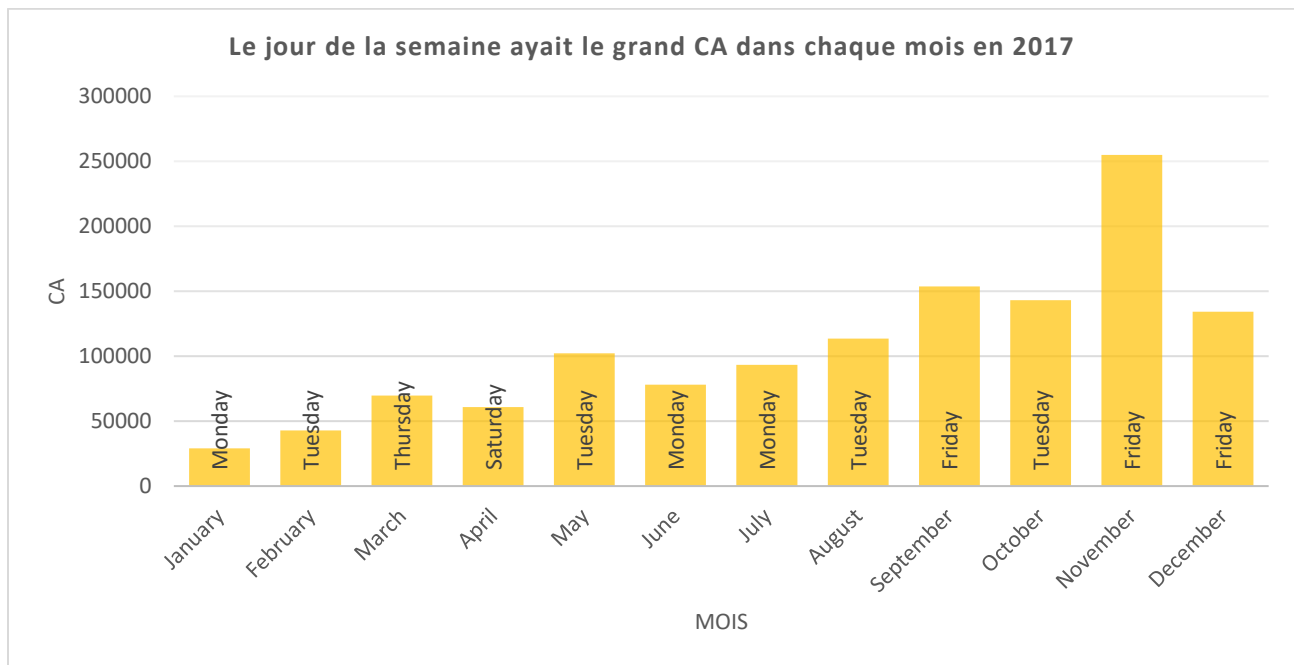


Référence du produit	Jours de livraison	Chiffre d'affaire
5769ef0a239114ac3a854af00df129e4	18	60480
ee3d532c8a438679776d222e997606b3	11	21000
c183fd5d2abf05873fa6e1014ed9e06c	19	20771.1
0c9439abe84dc40a64fc164365c4d064	14	15674.45
a62e25e09e05e6faf31d90c6ec1aa3d1	11	8424
89b190a046022486c635022524a974a8	20	7858.8
d5991653e037ccb7af6ed7d94246b249	14	7584.5
1f411e9a31196b71ed8438c8254b858e	10	7300
e272c7b6c21c38e8ee510666a28460d4	7	7194
6dde44b4172999f35f08654d06bad633	19	7019.64
<i>le nombre des jours de livraison des 10 produits qui ont les plus grands chiffre d'affaire</i>		

Référence du produit	Jours de livraison	Chiffre d'affaire
8a3254bee785a526d548a81a9bc3c9be	13	0.85
8a3254bee785a526d548a81a9bc3c9be	10	0.85
8a3254bee785a526d548a81a9bc3c9be	8	1.7
46fce52cef5caa7cc225a5531c946c8b	4	2.2
310dc32058903b6416c71faff132df9e	5	2.29
680cc8535be7cc69544238c1d6a83fe8	11	2.9
2e8316b31db34314f393806fd7b6e185	14	2.99
44d53f1240d6332232e4393c06500475	12	3
44d53f1240d6332232e4393c06500475	7	3
1716ea399ed8ee62ba811e6f55180f45	9	3.49
<i>le nombre des jours de livraison des 10 produits qui ont les plus petits chiffre d'affaire</i>		

- Quelle est le jour de la semaine de chaque mois dont les clients sont très actifs en 2017 ?

```
select trim(mois)||','||trim(jds)||','||ca from(
  select mois, jds, sum(ca) ca,rank() over(PARTITION BY mois ORDER BY sum(ca) desc) ra
  nking
  from(
    select sum(f.chiffre_affaire) ca, d.month mois, d.day_of_week jds
    from fact_commandes f
    join dim_dates d on d.id = f.id_date
    where d.year = 2017
    group by d.month, d.day_of_week
  )
  group by mois, jds
) where ranking = 1 order by to_date(mois,'MONTH');
```



- Est-ce que le nombre des photos du produit publiées dans le site web s'influence sur le CA ?

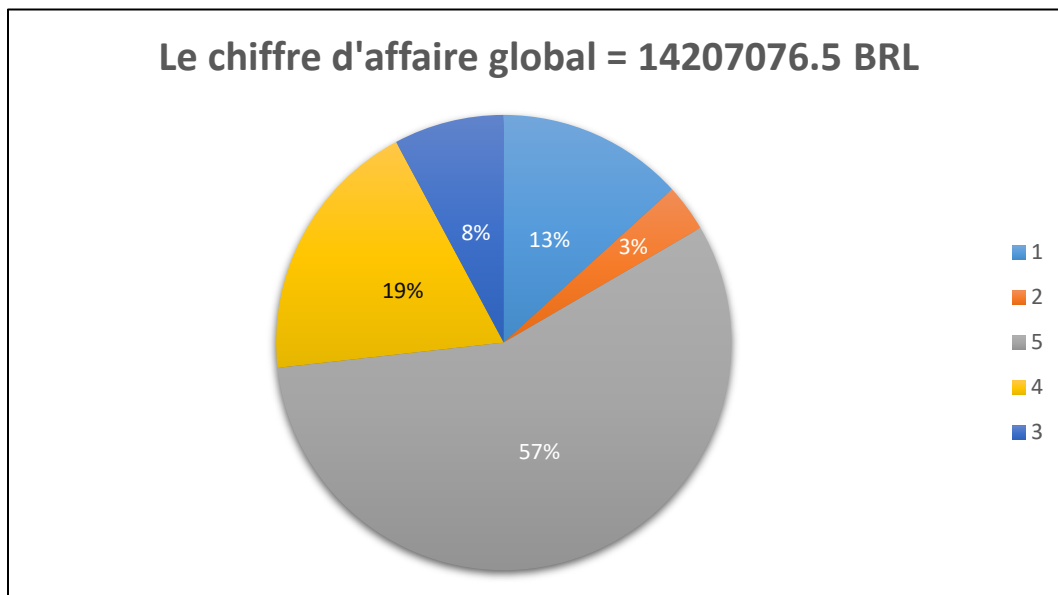
```
SELECT f.commande_id, p.num_pics, SUM(f.chiffre_affaire) ca
FROM fact_commandes f JOIN dim_produits p ON p.id = f.id_produit
WHERE p.num_pics IS NOT NULL
GROUP BY f.commande_id, p.num_pics
ORDER BY ca DESC;
```

COMMANDE_ID	NUM_PICS	CA	COMMANDE_ID	NUM_PICS	CA
1 03caa2c082116e1d31e67e9ae3700499	3	60480	15e5d4cea5f386d7f1d2d2600b830aae	20	110.27
2 1b15974a0141d54e36626dca3fdc731a	1	21000	38ab7ca6747531f5173881e11be23b74	19	138.94
3 2cc9089445046817a7539d90805e6e5a	1	20771.1	0ef8647af5bc37a0a511a194ae927e09	18	88.6
4 e8fa22c3673b1dd17ea315021b1f0f61	1	15674.45	15d975f1c3d089ac26a9b8b40df63d7b	18	88.6
5 912343626f370ead5ef2e85a9d757e22	1	10087.9	284c504a1801fa4911d5808290731274	18	88.6
6 3a213fcdfe7d98be74ea0dc05a8b31ae	2	8424	93b82c635931ca55e5d20e0d095507ce	18	88.6
7 428a2f660dc84138d969ccd69a0ab6d5	2	7858.8	22e4add79c3efc7828d1119a222b76da	17	71.6
8 71dab1155600756af6de79de92e712e3	1	7759.34	2d58df29239f4ee3887103ef2576e857	17	71.6
9 f80549a97eb203e1566e026ab66f045b	1	7584.5	67c41f8c7f43495ef9156232d80a9e0b	17	89.15
10 8171523911786efd1d91c66d69051fcd	2	7300	775747dc325485aedc9970fb0eda8b86	17	71.6
11 b15c7e972c74684414fb2e659fce916a	2	7194	8973a53fcad22a79e0ff315e7f9a0fd9	17	214.8

A partir de ces deux résultats on distingue que le CA ne dépend pas du nombre des photos publiées dans le site web, car on voit que les commandes qui ont un CA grand ont un peu des photos d'autre part les commandes qui ont un CA petit ont beaucoup des photos publiées.

- Quelle est le pourcentage de CA global pour chaque score (1-5) pour tous les produits ?

```
SELECT score, SUM(ca) ca
FROM(
  SELECT f.commande_id, a.score, SUM(f.chiffre_affaire) ca
  FROM fact_commandes f JOIN dim_avis a ON a.id = f.id_avis
  GROUP BY f.commande_id, a.score
) GROUP BY score;
```



- Quelle est le score de chaque produit des 10 produits qui ont été très vendu dans la région de Sao Paulo en 2018 ?

```
SELECT p.produit_id, a.score, SUM(f.chiffre_affaire) ca
FROM fact_commandes f
JOIN dim_produits p ON p.id = f.id_produit
JOIN dim_avis a ON a.id = f.id_avis
JOIN dim_lieux l ON l.id = f.id_lieu
JOIN dim_dates d ON d.id = f.id_date
WHERE d.year = 2018 AND l.region = 'SP'
GROUP BY p.produit_id, a.score
ORDER BY ca DESC;
```

	PRODUIT_ID	SCORE	CA
1	ee3d532c8a438679776d222e997606b3	1	21000
2	bb50f2e236e5eea0100680137654686c	5	18075
3	aca2eb7d00eala7b8ebd4e68314663af	5	8742.6
4	8a12c0e223d9dbd49ccfcd2ad25d1b30	5	8100
5	36f60d45225e60c7da4558b070ce4b60	1	7810.35
6	3dd2a17168ec895c781a9191c1e95ad7	5	7045.3
7	d0825229639d1afa9bd533314e27d7c8	1	5985
8	24daadd871657fb36dacde66b92cf0ff	5	5685
9	16c4e87b98a9370a9cbc3a4658a3f45b	4	5679
10	1f411e9a31196b71ed8438c8254b858e	5	5579.91

# Conclusion

L'informatique décisionnelle est un domaine en pleine évolution pour l'aide à la prise de décisions aux seins des entreprises. Elle est devenue un besoin capital, car la simple logique de produire pour répondre à une demande, ne suffit plus pour pérenniser l'activité de celle-ci, de ce fait sont apparus les entrepôts de données.

A travers ce projet, nous avons exposé les étapes de mise en place et de conception d'un entrepôt de données, qui facilitera la prise de décisions et permettre aux décideurs tout comme aux utilisateurs de sélectionner et filtrer des données utiles à la prise de décision.

Nous avons réparti ce projet en quatre parties principales.

Dans la première partie, nous avons essayé de chercher des données et les comprendre en suite on a créé le modèle Entité-Association pour bien présenter les données pour faciliter la tâche de la deuxième partie.

Dans la 2ème partie, nous avons exposé notre travail, la modélisation dimensionnelle en réalisant notre Data Marts, l'entrepôt de données, ainsi que le schéma en étoile.

Dans la 3ème, on a attaqué la partie ETL (Extraction, Transformation et Loading) pour charger les données de puis les fichiers CSV qu'on a déjà téléchargé, puis on a créé des autres tableaux en utilisant SQL3 contenant des fonctions permettant de transformer les données actuelles pour avoir des nouvelles données pour les mettre dans notre entrepôt de données.

Enfin, on a exécuté quelques requêtes SQL, convertir les résultats en CSV puis les représenter dans un graphe au un tableau.

## ANNEXE : Structure du fichier compresser

Après l'extraction du fichier « [BrahimOULHAJ\\_WalidTAOUTAOU.zip](#) », on va voir des fichiers et des dossiers comme elle représente la capture d'écran ci-dessous :

Name	Date modified	Type	Size
dataset	01-Mar-20 11:53	File folder	
log	01-Mar-20 12:46	File folder	
statistics	14-Mar-20 00:48	File folder	
1.Extraction.sql	01-Mar-20 12:56	SQL File	7 KB
1.Extraction_metadata.sql	01-Mar-20 12:58	SQL File	7 KB
2.Transformation.sql	04-Mar-20 20:26	SQL File	8 KB
3.Dimention_tables.sql	04-Mar-20 18:29	SQL File	4 KB
4.Fact_table.sql	04-Mar-20 23:04	SQL File	2 KB

**Dataset** : contient les sets des données télécharger depuis Kaggle.

**Log** : contient les fichiers log de la partie extractions des données.

**Statistics** : contient les requêtes SQL, permettant de sélectionner quelques données de l'entrepôt de données, et les fichiers CSV qui représente ces données pour les transformer en courbes, histogrammes, tableaux ...

**\*.sql** : des fichiers sql contenant les requêtes SQL des phases Extraction des données, Transformation des données et le remplissage du Data Warehouse.