

Université de Montpellier

# Projet de Base de Données Avancées

## Rapport de projet

JEUDI 10 NOVEMBRE 2016

**Auteurs :**

Thomas LEFEBVRE  
Rudy BARDOUT  
Hani GUENOUNE  
Yasin MOHAMMED

**Encadrants :**

Anne-Muriel CHIFOLLEAU  
Federico ULLIANA



## Table des matières

<b>1</b>	<b>Introduction</b>	<b>2</b>
<b>2</b>	<b>Traitements</b>	<b>3</b>
2.1	Actions / Opérations . . . . .	3
2.2	Requêtes analytiques . . . . .	4
<b>3</b>	<b>Conception</b>	<b>5</b>
3.1	Opérations importantes . . . . .	5
3.2	Réponse aux requêtes . . . . .	5
3.3	Détails . . . . .	6
<b>4</b>	<b>Réalisation</b>	<b>8</b>
4.1	Requêtes analytiques . . . . .	8
4.2	Surrogate keys . . . . .	8
4.3	Bridge tables (Multivaluated attributes) . . . . .	9
4.4	Vues virtuelles (Role-playing) . . . . .	9
4.5	Problème d'insertion Oracle . . . . .	10
<b>5</b>	<b>Conclusion</b>	<b>11</b>

## Table des figures

1	Premier Datamart . . . . .	5
2	Deuxième Datamart . . . . .	6
3	Table de fait F_Commande . . . . .	7
4	Constellation de Datamarts . . . . .	8
5	Surrogate keys . . . . .	9
6	Tables pont . . . . .	9
7	Creation de vues . . . . .	10
8	Erreur lors de l'insertion dans la base . . . . .	10

# 1 Introduction

Le sujet sur lequel nous nous sommes penchés lors du mini-projet objet-relationnel proposait de construire une base de données de commandes de plateaux repas constitués de produits du terroir (pain, charcuterie, vins et fromage de différents types).

Nous adoptons ici une vision plus business. Les activités d'un organisme dans son terrain d'expertise se déclinent en un corpus important de données regroupant l'historique de sa production et de ses événements passés (faits, ventes, clients, produits...). L'évolution du volume de ces données est parallèle à un savoir-faire également grandissant au fil de son expérience. La rapidité de cette évolution dépend de la capacité des organismes à tirer profit de leurs expériences et de la qualité des décisions prises par sa direction.

Il s'agit maintenant de supposer la prospérité des services de livraison du produit du terroir, l'analyse suivante consiste à proposer les potentielles directions stratégiques que pourraient emprunter les entreprises proposant le service en question afin d'améliorer leurs productivités, ainsi que les analyses à effectuer pour mener à bien cette démarche.

## 2 Traitements

Les données que nous aurons besoin de stocker dans l'entrepôt concernent :

- Les produits commandés : Ceux-ci sont stockés dans la table repas de notre base de données.
- Clients ayant passé des commandes de produits : Tuples de la table Client.
- Date : Il s'agit de la dimension temps (Jour, mois, années) permettant de poser un suivi des activités du business à travers le temps.

Afin de permettre des analyses plus détaillées, nous choisissons d'étendre le modèle du premier projet en ajoutant les notions de Fournisseur et du Local. Pour souligner l'hétérogénéité des données des entrepôts nous supposons que ces ajouts proviennent de sources de données différentes ; **Attention : les éléments suivants sont dépréciés depuis les derniers changements dans la conception.**

- Livreur : Informations relatives aux livreurs de produits du terroir.
- Locaux : Sortes de magasins servant à recueillir, stocker, et envoyer les commandes aux domiciles des clients.

### 2.1 Actions / Opérations

Pour parvenir à des stratégies cohérentes et prometteuses d'évolution, nous proposons un petit ensemble de démarches décisionnelles stratégiques (opérations). Pour chacune de ces démarches, une entité dirigeante fictive propose un certain nombre d'idées d'améliorations.

Il s'agira ensuite de déduire les faits et les traitements (analyses) à partir de ces propositions.

- a) Opération de communication et marketing : Ce volet vise à promouvoir les services de l'entreprise, l'intérêt étant de cibler une clientèle de plus en plus grande, ainsi que de fidéliser les clients déjà existants de l'entreprise.
  - 1-Communication : Les dirigeants de l'entreprise souhaiteraient organiser des foires de dégustation des différents produits du terroir, et en profiter pour parler de leur entreprise. Pour une meilleure efficacité de cette opération de communication, ils se demandent quelles sont les régions prioritaires où pourront avoir lieu ces foires (Les régions les moins intéressées par la restauration du terroir), et quels sont les produits les moins connus des consommateurs par région.
  - 2-Marketing : Afin d'améliorer les ventes et les chiffres d'affaire de l'entreprise, les idées suivantes sont proposées :
    - Création de menus (commandes) formés des assortiments et associations des produits les plus populaires (couple, triplet de produits) auprès des clients par région et par saison.
    - Proposer des offres d'abonnement/réductions aux clients les plus fidèles.
    - Création de menus à moindre prix pour une certaine tranche d'âge de clients.
    - Création de menus (commandes) formés des assortiments et associations des produits les plus populaires (couple, triplet de produits) auprès des clients par région et par saison.
    - Proposer des offres d'abonnement/réductions aux clients les plus fidèles.
    - Création de menus à moindre prix pour une certaine tranche d'âge de clients.
- b) Opération ouverture de nouveaux locaux en optimisant les trajets de livraison : Une équipe de recherche développe des solutions d'optimisation des trajets du service de livraison. En complément aux heuristiques développées, ils proposent afin de réduire le coût engendré par la livraison des produits, les dispositions suivantes : Pour chaque région, établir les nouveaux locaux à des emplacements proches des fournisseurs et du périmètre le plus desservi (clients les plus souvent livrés)
- c) Opération optimisation stockage des produits avant leurs livraisons : L'équipe des stocks suggère également d'identifier les produits les plus demandés par région afin de gérer les approvisionnements (décider des quantités à stocker), et ainsi éviter l'épuisement de ces produits.

## 2.2 Requêtes analytiques

**Attention : certains de ces traitements ne sont plus valides dû aux changement dans le modèle effectués suite aux remarque de l'encadrant, voir la partie réalisation pour les nouveaux traitements**

### **Operation optimisation gain**

- Quelles sont les villes ayant le moins d'habitants qui ont déjà commandé ? (Ou les villes dont les habitants n'ont jamais commandé de produits)
- Quels sont les produits les moins commandés par régions ?
- Quels sont les produits les plus souvent associés dans une commande ?
- Quels sont les produits commandés par les clients ayant un âge compris dans une certaine tranche ?
- Opération nouveaux locaux/optimisation trajets et route Quels sont les clients effectuant des commandes régulièrement (Tous les x jours) ?
- Quels sont les fournisseurs dont les produits sont les plus appréciés par région ?

### **Opération optimisation stockage**

- Quels sont les produits disponibles dans un local donné ?
- Quel jour un local de stockage donné a été saturé ?
- Quelle est la capacité restante (aujourd'hui) d'un local donné ?

### **Ordre d'importance des opérations**

- a)-Commandes/Ventes de produits
- c)-Approvisionnement et Stockage
- b)-Ouverture de nouveaux locaux de stockage

## 3 Conception

### 3.1 Opérations importantes

- a) Commandes/Ventes de plateau
- c) Approvisionnement et stockage

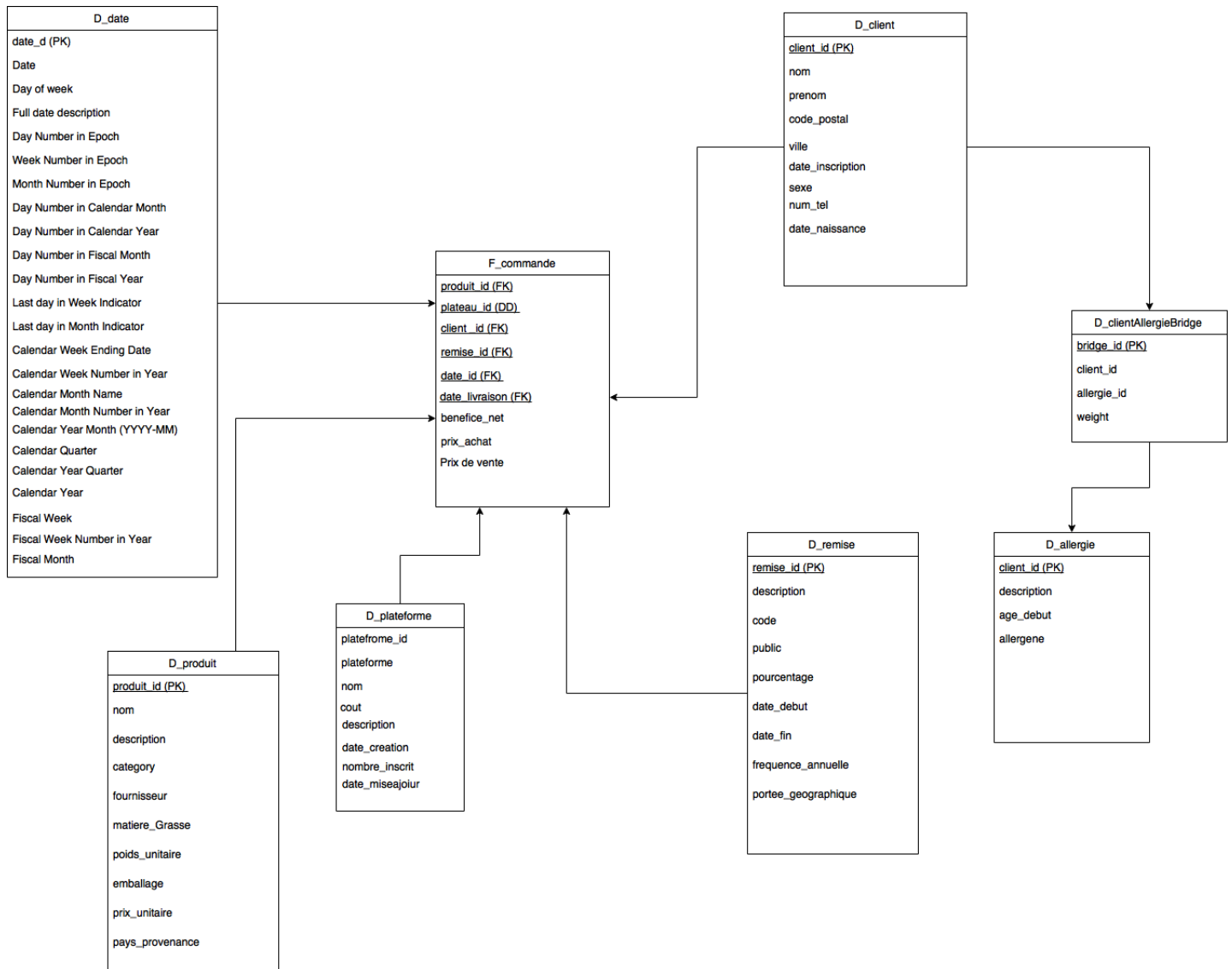


FIGURE 1 – Premier Datamart

### Datamarts

Les mesures : Prix\_Achat et Prix\_Vente et Benefice\_Net sont toutes trois additives.

Datamart opération c : Il s'agit d'un snapshot périodique. La mesure de capacité restante est semi-additive.

### 3.2 Réponse aux requêtes

Le datamart ayant changé, on ne peut pas interroger la DW avec les traitements prévus précédemment, voir nouveaux traitement dans la partie réalisation

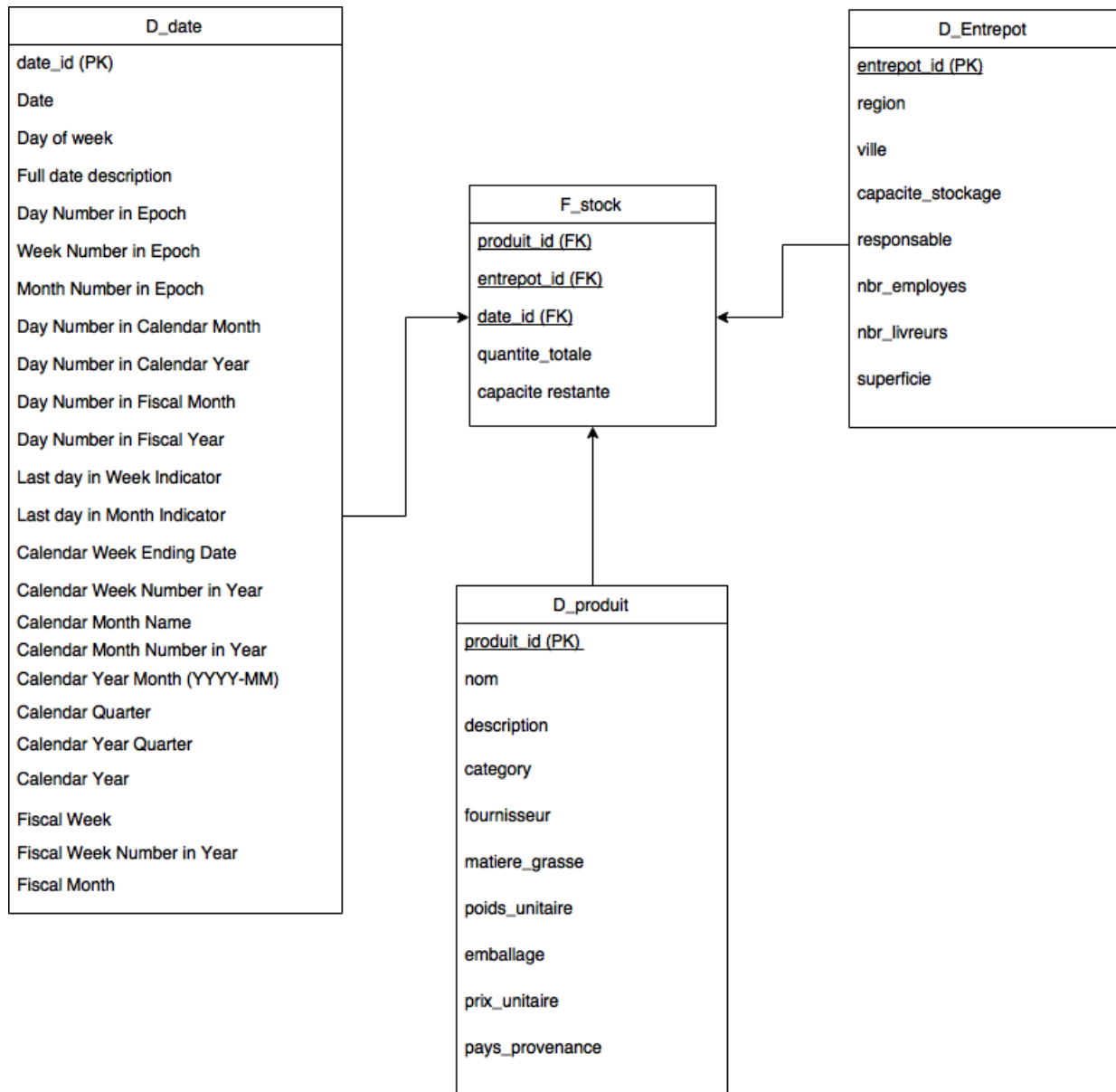


FIGURE 2 – Deuxième Datamart

### 3.3 Détails

La notion de plateau était celle présente dans notre modèle objet-relationnel, et qui consiste en un regroupement de produits constituant un repas commandé.

En raison de la granularité différente de notre entrepôt (plus détaillée/Transaction concernant un seul produit et non un plateau), la majorité des attributs de la table plateau ont migré vers la notion atomique de produit. La dimension associée est ainsi devenue superflue.

Il reste néanmoins utile d'identifier les produits achetés au sein d'une même commande.

Afin de permettre cela, nous décidons de mettre l'attribut "plateau\_id" dans la table de fait "F\_commande". Il devient donc possible de retrouver les associations de produits (produits commandés ensemble), en regroupant les commandes ayant la même valeur de "plateau\_id".

F_commande
<u>produit_id (FK)</u>
<u>plateau_id (DD)</u>
<u>client_id (FK)</u>
<u>remise_id (FK)</u>
<u>date_id (FK)</u>
<u>plateforme_id (FK)</u>
benefice_net
prix_achat
prix_vente

FIGURE 3 – Table de fait F\_Commande



## 4 Réalisation

### 4.1 Requêtes analytiques

Traitements :

- On cherche à savoir quel fromage est le plus vendu dans le sud est de la France.
- On veut savoir quel duo vin-charcuterie est le plus vendu en France, afin de pouvoir faire une remise dessus.
- On veut chercher les clients résidant en région parisienne qui ont déjà une remise actuellement active afin de ne pas leur en offrir une 2ème.
- On cherche à savoir quel est la région ou la clientèle est la plus importante, et celle qui est la moins importante, afin de faire des remises sur ces régions.
- On veut savoir quel plate-forme est la plus utilisée parmi nos clients.
- On cherche à savoir quel produit a été vendu le plus pour chaque région et pour chaque produit, afin de définir des box promotionnels.
- On cherche à savoir quel région a eu le plus de remise, afin de diminuer le nombre de ces dernières dans cette région.
- On cherche à savoir si notre clientèle est plutôt vieille ou jeune afin de faire profiter d'une certaines remises pour ces tranches d'âge.
- On cherche les produits les moins vendus par type, afin de potentiellement les supprimer de notre menu.
- Inversement, on cherche les produits les plus vendus par type, afin d'augmenter notre stock de ces produits.

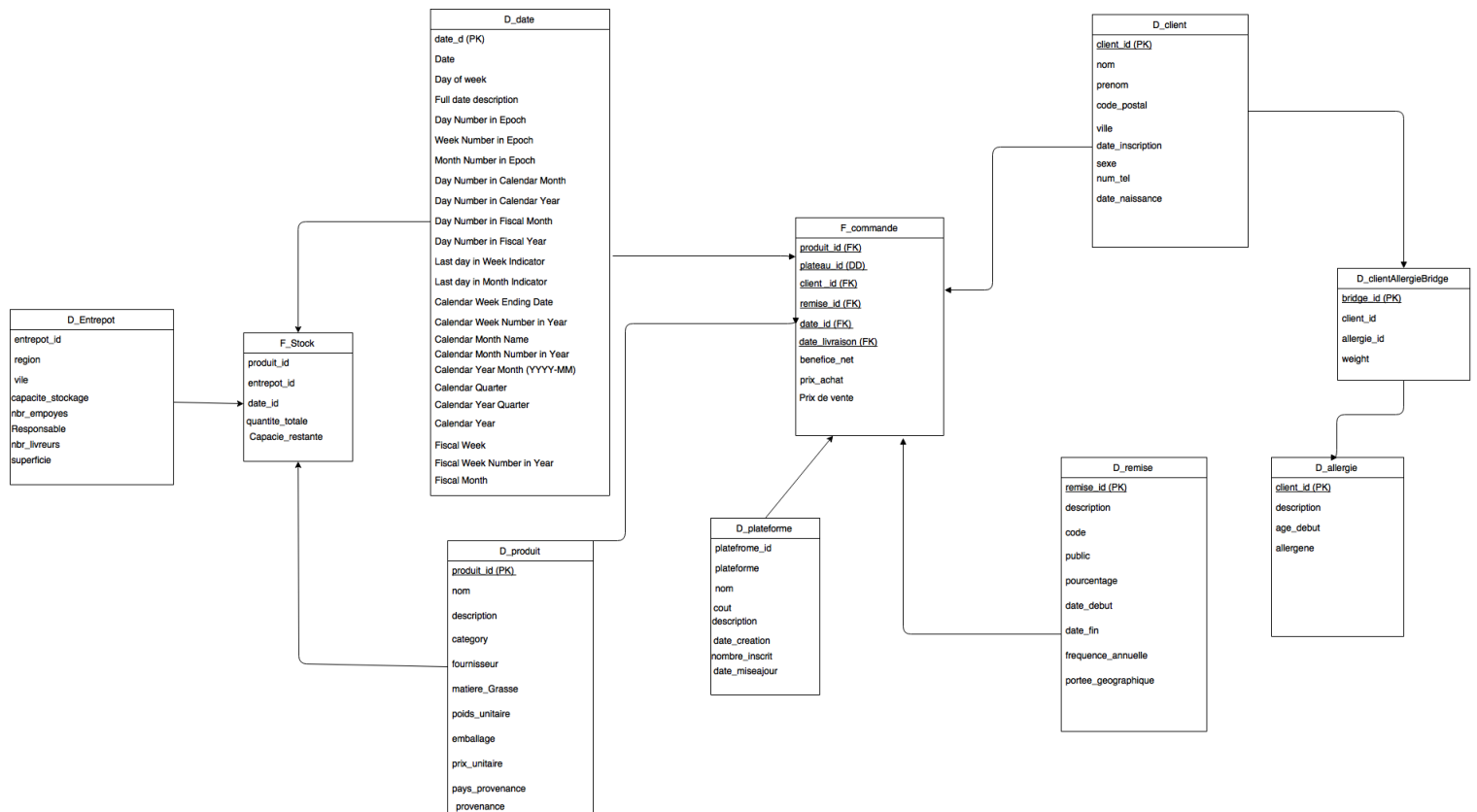


FIGURE 4 – Constellation de Datamarts

### 4.2 Surrogate keys

Etant donné que l'utilisation des clés artificielles (Surrogate keys) est conseillée pour les tables dimensions. Nous avons choisi d'identifier les tuples insérés de façon incrémentale, en créant et utilisant des séquences pour l'insertion dans chaque table.

```

CREATE SEQUENCE clientSeq INCREMENT BY 1 START WITH 1 ;

INSERT INTO D_client VALUES(
  clientSeq.nextval, 'Tonneau', 'Mathieu', 'Montpellier', 34000, 06123456, '16/02/1984', '13/09/2011', '1');

INSERT INTO D_client VALUES(
  clientSeq.nextval, 'Bruges', 'Jon', 'Marseille', 124, 02079460777, '24/05/1990', '15/12/2015', '1');

INSERT INTO D_client VALUES(
  clientSeq.nextval, 'Richard', 'Julie', 'Paris', 75015, 0607815637, '03/11/1972', '11/06/2009', '2');

CREATE SEQUENCE allergieSeq INCREMENT BY 1 START WITH 1 ;

INSERT INTO D_allergie VALUES(
  allergieSeq.nextval, 'Intolérant au gluten', 'souvent dès la naissance', 'Gluten');

INSERT INTO D_allergie VALUES(
  allergieSeq.nextval, 'Intolérant au lactose', 'se développe assez souvent chez les jeunes', 'Lactose');

INSERT INTO D_allergie VALUES(
  allergieSeq.nextval, 'Intolérant au sésame', 'souvent dès la naissance', 'Sésame');

```

FIGURE 5 – Surrogate keys

### 4.3 Bridge tables (Multivaluated attributes)

Un attribut multi-valué est un attribut qui peut avoir plusieurs valeurs pour un seul tuple. Une occurrence de ce type d'attribut apparaît dans notre modèle, il s'agit de l'attribut allergie de la dimension "D\_client".

Afin de gérer la multiplicité des valeurs possibles de l'attribut "allergie", nous avons dû faire intervenir une table pont (Bridge table) liant la dimension "D\_client" et la table allergie nouvellement créée.

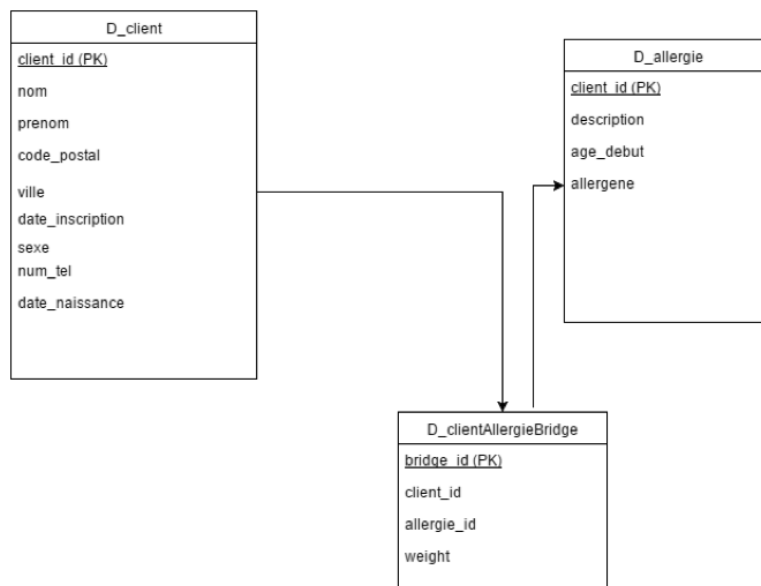


FIGURE 6 – Tables pont

### 4.4 Vues virtuelles (Role-playing)

Étant donné que quelques dimensions sont partagées par les deux datamarts (en l'occurrence D\_produit et D\_date), les deux devront "jouer les rôles" de dimensions pour chacune des tables de faits des deux datamarts.

```

/*
Création des vues pour les dimensions partagées.
<dimension>s partagées : D_produit, D_date.

Une vue pour chaque table de fait référençant la dimension ;
<dimension>ForComm : attributs --> vc_<nom_attribut>.
<dimension>ForStocks : attributs --> vs_<nom_attribut>.

*/

--D_produit

CREATE VIEW D_prodForComm (vc_produit_id PRIMARY KEY DISABLE NOVALIDATE,vc_nom,vc_description,vc_category,vc_type,vc_fournisseur,vc_matiere_grasse,vc_poids_unitaire,vc_emballage,vc_prix_unitaire)
SELECT produit_id,nom,description,category ,type,fournisseur,matiere_grasse,poids_unitaire,emballage,prix_unitaire
FROM D_produit

CREATE VIEW D_prodForStocks (vs_produit_id PRIMARY KEY DISABLE NOVALIDATE,vs_nom,vs_description,vs_category,vs_type,vs_fournisseur,vs_matiere_grasse,vs_poids_unitaire,vs_emballage,vs_prix_unitaire)
SELECT produit_id,nom,description,category,type,fournisseur,matiere_grasse,poids_unitaire,emballage,prix_unitaire
FROM D_produit

--D_date

CREATE VIEW D_dateForComm (vc_date_id PRIMARY KEY DISABLE NOVALIDATE,vc_date_type,vc_full_date,vc_day_num_of_week,vc_day_num_of_month,vc_day_num_of_quarter,vc_day_num_of_year,vc_day_num_of_absolute,vc_week_num_of_month,vc_week_num_of_quarter,vc_week_num_of_year)
SELECT date_id,date_type,full_date,day_num_of_week ,day_num_of_month ,day_num_of_quarter ,day_num_of_year ,day_num_of_absolute,week_num_of_month ,week_num_of_quarter ,week_num_of_year
FROM D_date

CREATE VIEW D_dateForStocks (vs_date_id PRIMARY KEY DISABLE NOVALIDATE,vs_date_type,vs_full_date,vs_day_num_of_week,vs_day_num_of_month,vs_day_num_of_quarter,vs_day_num_of_year,vs_day_num_of_absolute,vs_week_num_of_month,vs_week_num_of_quarter,vs_week_num_of_year)
SELECT date_id,date_type,full_date,day_num_of_week ,day_num_of_month ,day_num_of_quarter ,day_num_of_year ,day_num_of_absolute,week_num_of_month ,week_num_of_quarter ,week_num_of_year
FROM D_date

```

FIGURE 7 – Creation de vues

## 4.5 Problème d'insertion Oracle

Après l'étape de création des tables de faits et des dimensions des deux datamarts, nous avons malheureusement rencontré un problème que nous n'avons pas pu contourner lors des insertions de tuples dans les tables. La capture d'écran suivante montre le message d'erreur affiché à chaque tentative d'insertion.

```

Terminal - hguenoune@q12: ~/Bureau/DW-Mini-Projet
Fichier Éditer Affichage Terminal Onglets Aide

SQL> INSERT INTO D.allergie VALUES(
      1, 'Intolérant au gluten', 'souvent dès la naissance', 'Gluten'); 2
^
      1, 'Intolérant au gluten', 'souvent dès la naissance', 'Gluten')
ERROR at line 2:
ORA-00600: internal error code, arguments: [kqludp2], [0x090919760], [1], [],
[1], [], [], [], [], [], [], [], []

SQL> SQL> ^C
SP2-0042: unknown command "SQL" - rest of line ignored.
SQL> INSERT INTO D.allergie VALUES(1, 'Intolérant au gluten', 'souvent dès la na
issance', 'Gluten');
INSERT INTO D.allergie VALUES(1, 'Intolérant au gluten', 'souvent dès la naissan
ce', 'Gluten')
*
ERROR at line 1:
ORA-00600: internal error code, arguments: [kqludp2], [0x090919760], [1], [],
[1], [], [], [], [], [], [], [], []

SQL>

```

FIGURE 8 – Erreur lors de l'insertion dans la base

## 5 Conclusion

En conclusion, ce projet a été pour nous l'occasion d'aller au delà de la théorie vue en cours et de pouvoir enfin aborder la partie pratique et concrète de la réalisation d'un entrepôt de données.

Ce projet nous a également permis de travailler en groupe à l'apprentissage et l'approfondissement de certaines notions des bases de données avancé.