



Document de conception – Projet ODE

|  |  |  |
| --- | --- | --- |
| **Suivi des versions** | | |
| **Date** | **Version** | **Objet de la version** |
| 08/07/2015 | 01 | [Olivier] Création du document |
|  |  |  |
|  |  |  |

Sommaire

[Introduction 3](#_Toc424573269)

[Aspects métier 3](#_Toc424573270)

[Quelques chiffres 3](#_Toc424573271)

[Structure organisationnelle 3](#_Toc424573272)

[Architecture de la chaine décisionnel 4](#_Toc424573273)

[Vue d’ensemble 4](#_Toc424573274)

[Partie ETL 5](#_Toc424573275)

[Partie DWH / Cubes 5](#_Toc424573276)

[Partie Reporting 5](#_Toc424573277)

[Base opérationnelle 5](#_Toc424573278)

[Vue d’ensemble 5](#_Toc424573279)

[Table UNIVERS\_PRODUITS 6](#_Toc424573280)

[Table RAYONS\_PRODUITS 7](#_Toc424573281)

[Table FAMILLES\_PRODUITS 7](#_Toc424573282)

[Table SOUS\_FAMILLES\_PRODUITS 8](#_Toc424573283)

[Table PRODUITS 9](#_Toc424573284)

[Table CLIENTS 10](#_Toc424573285)

[Table LIEUX 11](#_Toc424573286)

[Table VILLES 11](#_Toc424573287)

[Table STOCKS 12](#_Toc424573288)

[Table VENTES 13](#_Toc424573289)

[Table TICKETS 14](#_Toc424573290)

[Table PRIXPRODUITS 14](#_Toc424573291)

[Entrepôt de données 15](#_Toc424573292)

[Vue d’ensemble 15](#_Toc424573293)

[Table de dimension CATEGORIES 16](#_Toc424573294)

[Table de dimension PRODUITS 17](#_Toc424573295)

[Table de dimension TEMPS 17](#_Toc424573296)

[Table de dimension LIEUX 18](#_Toc424573297)

[Table de dimension CLIENTS 19](#_Toc424573298)

[Table de dimension VILLES 19](#_Toc424573299)

[Table de faits VENTES 20](#_Toc424573300)

[Table de faits STOCKS 21](#_Toc424573301)

# Introduction

Le projet « Optimisation des Données de l’Entrepôt (ODE) » consiste à utiliser les techniques mathématiques vues dans le Master 2 MIAGE afin de construire l'entrepôt de données (Data-Warehouse - DWH) de manière optimal, en termes de temps de réponse à l’interrogation des cubes et d’occupation disque.

# Aspects métier

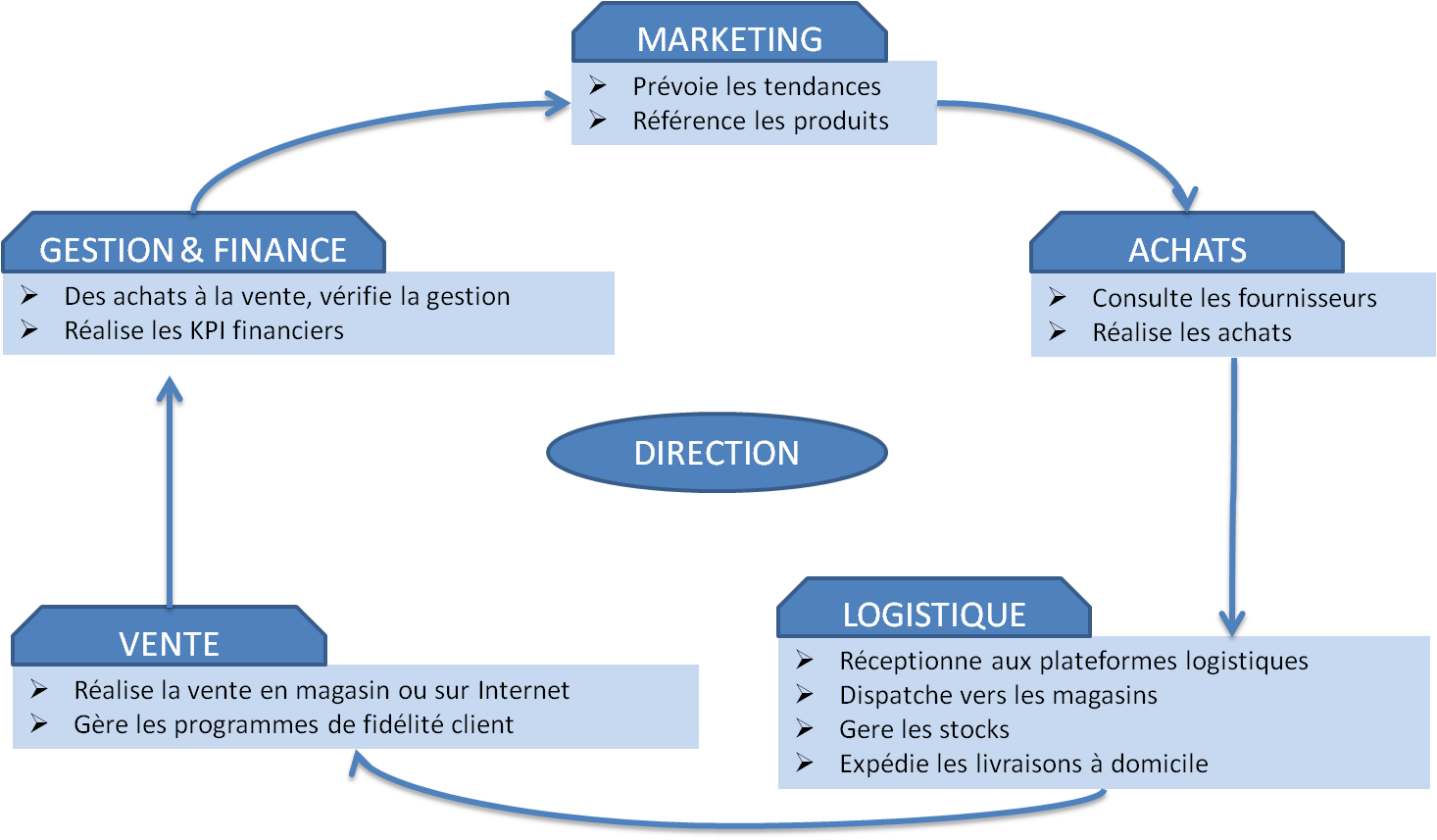
CASTO-MERLIN : Grande distribution de bricolage

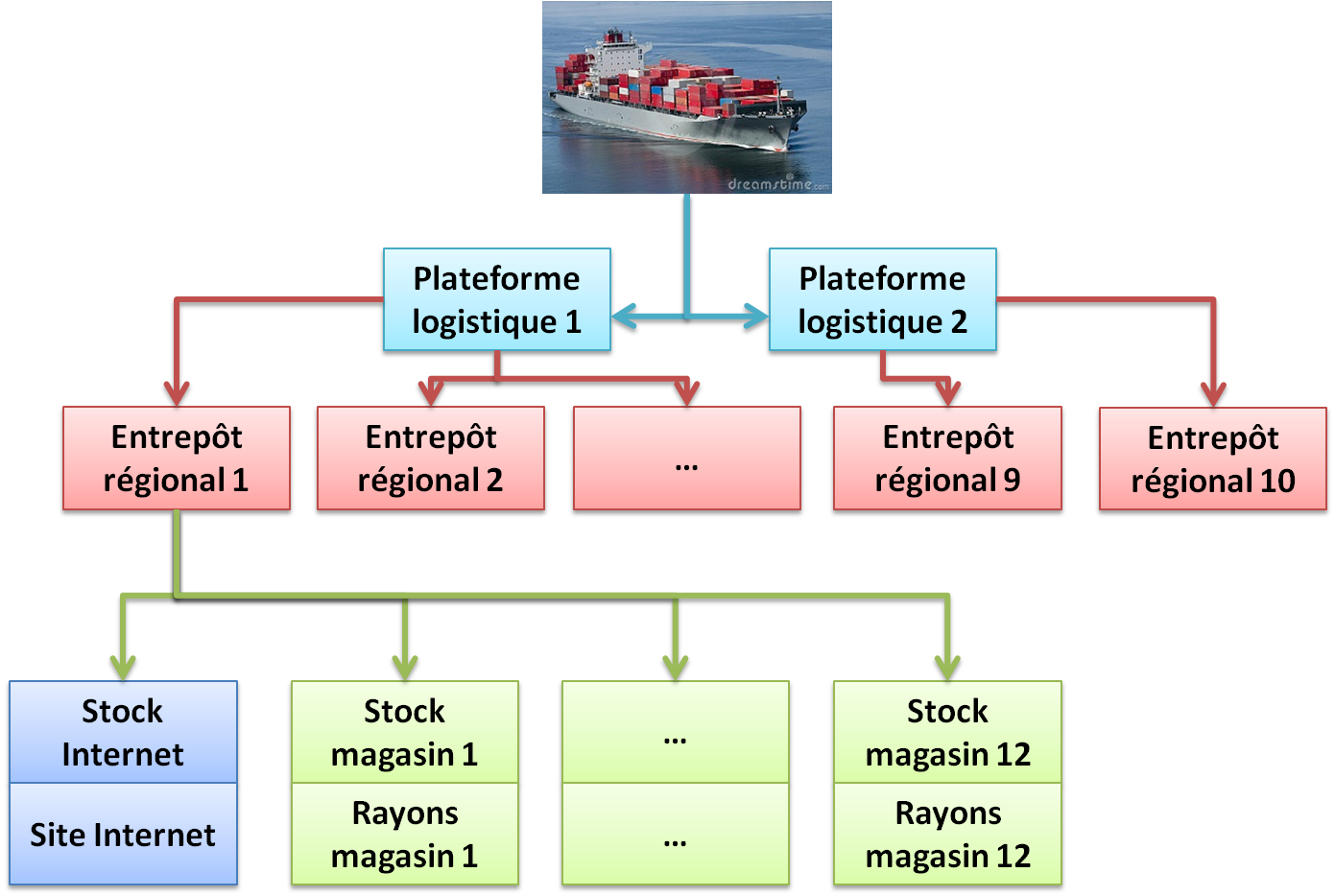


## Quelques chiffres

* 31 300 collaborateurs
* 2 plateformes logistiques nationales
* 10 centrales de distribution régionales
* 121 magasins + 1 site Internet
* 60 000 références magasins plus 20 000 « sur commande »
* 3 856 fournisseurs

## Structure organisationnelle





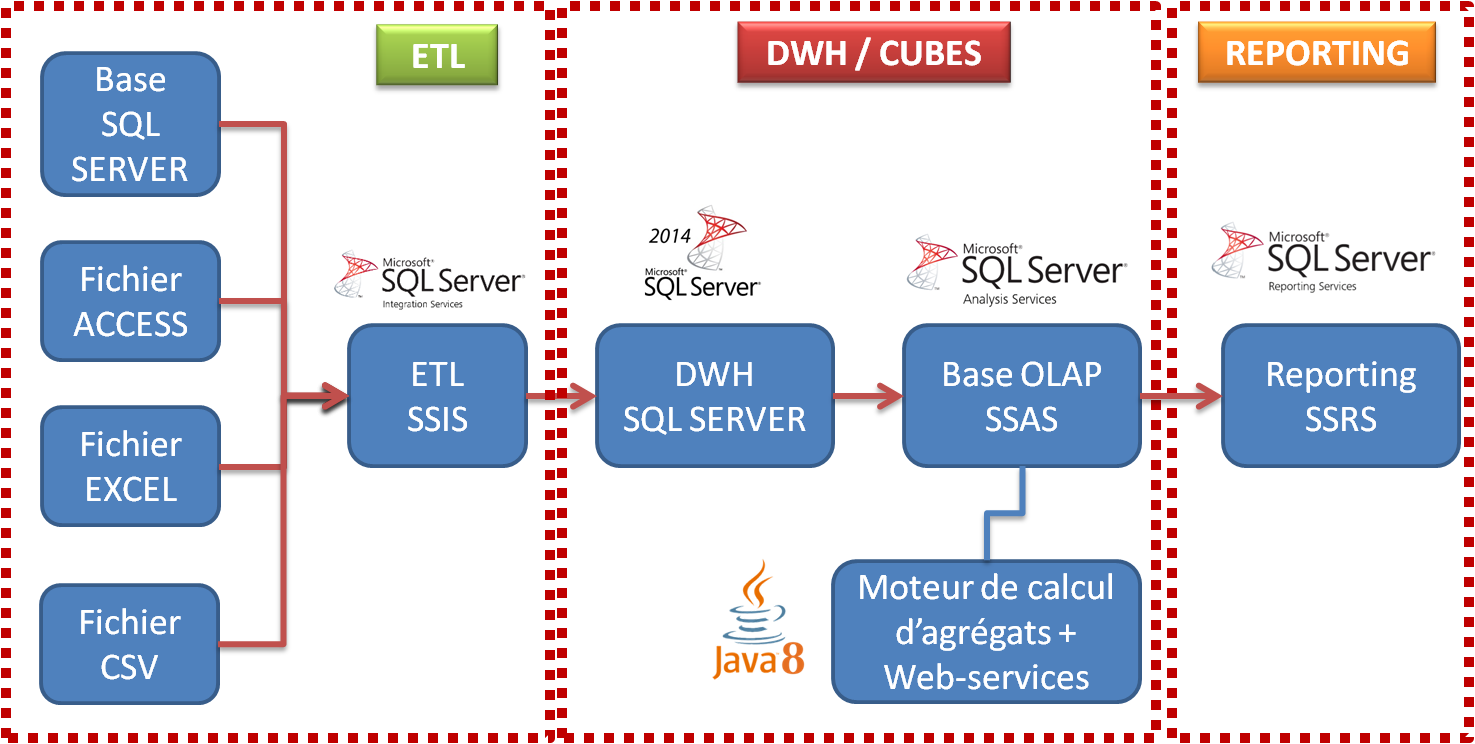
# Architecture de la chaine décisionnel

## Vue d’ensemble

Les données opérationnelles de l’entreprise, c'est-à-dire celle générées et utilisés pour la gestion de l’activité courante de l’entreprise, sont stockées principalement dans une base OLTP SQL Server 2014. Nous la désignerons par « base opérationnelle » dans la suite de ce document.

D’autres sources de données sont possibles :

* Fichiers « plats » type CSV
* Fichiers EXCEL
* Fichiers ACCESS



## Partie ETL

L’ETL sera en charge de collecter les informations à partir des différentes sources de données opérationnelles de l’entreprise :

* Base opérationnelle
* Fichiers CSV
* Fichiers EXCEL
* Fichiers ACCESS

Cette partie sera réalisée avec **SQL Server 2014 – Integration Services (SSIS)**

## Partie DWH / Cubes

C’est le « cœur » de la chaine décisionnelle que nous allons réaliser.

Le Datawarehouse (DWH – Entrepot de données) est modélisé en flocon, et hébergé sur une base OLTP **SQL Server 2014.**

Le cube sera hébergé sur une base OLAP **SQL Server 2014 Analysis Service (SSAS)**

Le moteur de calcul d’agrégats est un programme Java servant à « optimiser » la structure et le calcul des cubes (Agrégats, cf. cours du D111)

## Partie Reporting

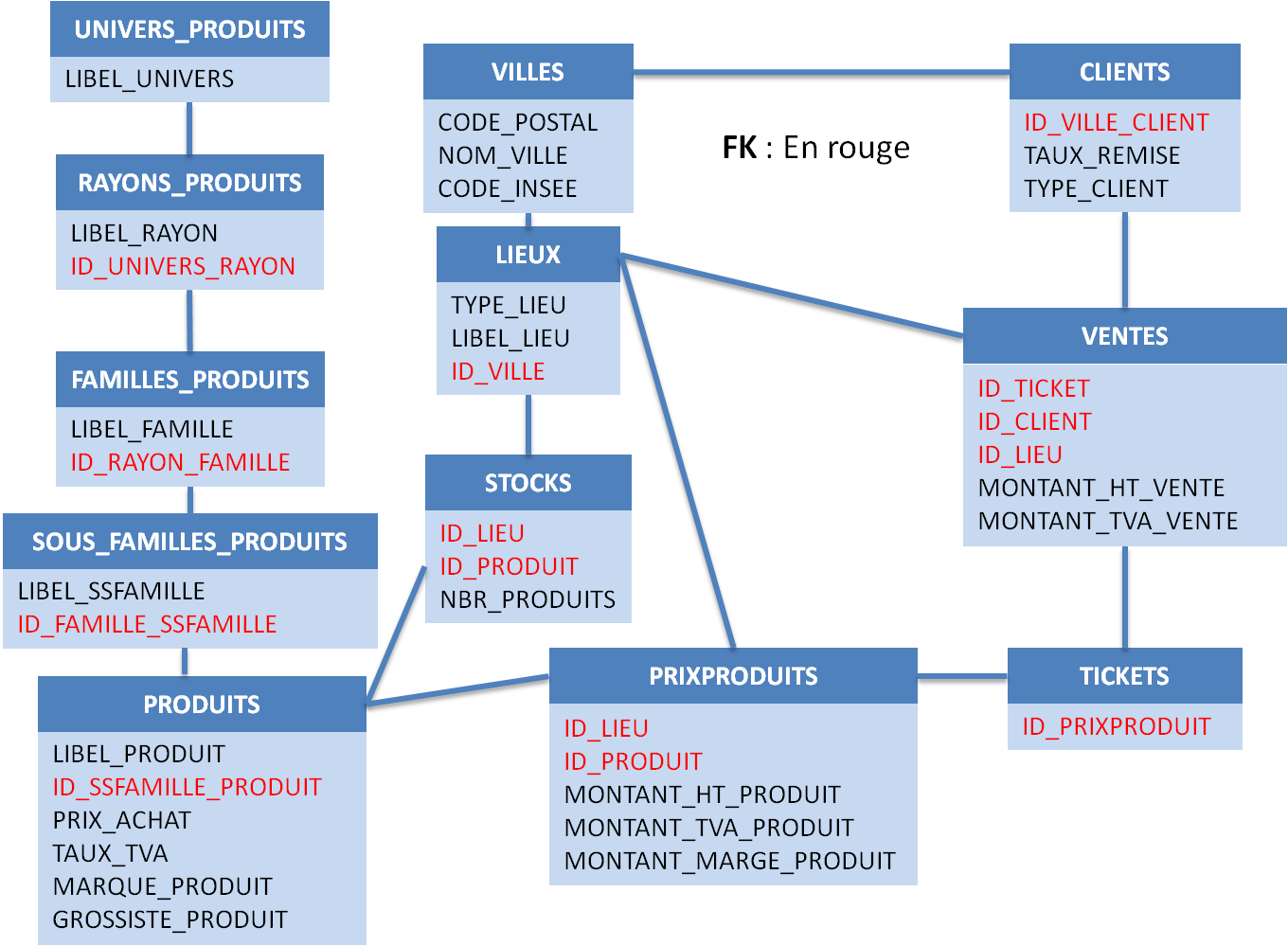
La chaine décisionnelle génère des reports sur un sujet « métier ». Par exemple : Calcul du CA, de la marge, évolution des ventes…

Cette partie sera réalisée avec **SQL Server 2014 – Reporting Services (SSRS)**

# Base opérationnelle

## Vue d’ensemble

Dans la base dédiée « **BaseOperationelleODE** », nous créons un schéma dédié « **ODE\_VENTES** », qui contient les tables suivantes :



## Table UNIVERS\_PRODUITS

But

Les Univers de produits sont les premiers niveaux de classification des produits vendus par CASTO-MERLIN.

Il comprend actuellement 3 valeurs :

* « Interieur Et Decoration »
* « Atelier Et Materiaux »
* « Jardin Et Exterieur »

Chaque libellé est associé à un ID numérique, qui sera utilisé en tant que clé primaire : Les libellés peuvent évoluer, mais pas leurs ID associés.

Comme la quasi-totalité des tables du SI opérationnel de ventes de CASTO-MERLIN, il comprend le nom de l’utilisateur ayant fait l’INSERT de cette ligne, et la date (Resp. les champs « OPER\_CREAT » et « DATE\_CREAT »), ainsi que le nom de l’utilisateur ayant fait le dernier UPDATE de cette ligne, et la date (Resp. les champs « OPER\_MODIF » et « DATE\_MODIF »)

L’opérateur peut être le nom d’une personne (« Jean DUPOND ») ou d’un SI externe (« IBM Point Of Sales System »)

Structure

|  |  |  |
| --- | --- | --- |
| Nom colonne | Type | Nullable ? |
| ID\_UNIVERS | int | Non |
| LIBEL\_UNIVERS | nvarchar(256) | Non |
| DATE\_CREAT | datetime | Non |
| OPER\_CREAT | nvarchar(64) | Non |
| DATE\_MODIF | datetime | Oui |
| OPER\_MODIF | nvarchar(64) | Oui |

Remplissage

/\* A COMPLETER LORS DU LOT 3 : ETL \*/

## Table RAYONS\_PRODUITS

But

Les rayons de produits sont les seconds niveaux de classification des produits vendus par CASTO-MERLIN.

Il comprend actuellement 39 valeurs :

* « Chauffage, Climatisation Et Ventilation »
* « Cuisine »
* « Décoration »
* …

Chaque libellé est associé à un ID numérique, qui sera utilisé en tant que clé primaire : Les libellés peuvent évoluer, mais pas leurs ID associés. Chaque rayon est associé à exactement un Univers, au travers du champ ID\_UNIVERS\_RAYON, qui est une clé étrangère vers la table des Univers de produits.

Comme la quasi-totalité des tables du SI opérationnel de ventes de CASTO-MERLIN, il comprend le nom de l’utilisateur ayant fait l’INSERT de cette ligne, et la date (Resp. les champs « OPER\_CREAT » et « DATE\_CREAT »), ainsi que le nom de l’utilisateur ayant fait le dernier UPDATE de cette ligne, et la date (Resp. les champs « OPER\_MODIF » et « DATE\_MODIF »)

L’opérateur peut être le nom d’une personne (« Jean DUPOND ») ou d’un SI externe (« IBM Point Of Sales System »)

Structure

|  |  |  |
| --- | --- | --- |
| Nom colonne | Type | Nullable ? |
| ID\_RAYON | int | Non |
| LIBEL\_RAYON | nvarchar(256) | Non |
| ID\_UNIVERS\_RAYON | int | Non |
| DATE\_CREAT | datetime | Non |
| OPER\_CREAT | nvarchar(64) | Non |
| DATE\_MODIF | datetime | Oui |
| OPER\_MODIF | nvarchar(64) | Oui |

Remplissage

/\* A COMPLETER LORS DU LOT 3 : ETL \*/

## Table FAMILLES\_PRODUITS

But

Les familles de produits sont les troisièmes niveaux de classification des produits vendus par CASTO-MERLIN.

Il comprend actuellement 317 valeurs :

* « Chauffage Électrique »
* « Chauffage Central »
* « Chauffage Bois Et Bois De Chauffage »
* …

Chaque libellé est associé à un ID numérique, qui sera utilisé en tant que clé primaire : Les libellés peuvent évoluer, mais pas leurs ID associés. Chaque famille est associée à exactement un rayon, au travers du champ ID\_RAYON\_FAMILLE, qui est une clé étrangère vers la table des rayons de produits.

Comme la quasi-totalité des tables du SI opérationnel de ventes de CASTO-MERLIN, il comprend le nom de l’utilisateur ayant fait l’INSERT de cette ligne, et la date (Resp. les champs « OPER\_CREAT » et « DATE\_CREAT »), ainsi que le nom de l’utilisateur ayant fait le dernier UPDATE de cette ligne, et la date (Resp. les champs « OPER\_MODIF » et « DATE\_MODIF »)

L’opérateur peut être le nom d’une personne (« Jean DUPOND ») ou d’un SI externe (« IBM Point Of Sales System »)

Structure

|  |  |  |
| --- | --- | --- |
| Nom colonne | Type | Nullable ? |
| ID\_FAMILLE | int | Non |
| LIBEL\_FAMILLE | nvarchar(256) | Non |
| ID\_RAYON\_FAMILLE | int | Non |
| DATE\_CREAT | datetime | Non |
| OPER\_CREAT | nvarchar(64) | Non |
| DATE\_MODIF | datetime | Oui |
| OPER\_MODIF | nvarchar(64) | Oui |

Remplissage

/\* A COMPLETER LORS DU LOT 3 : ETL \*/

## Table SOUS\_FAMILLES\_PRODUITS

But

Les sous-familles de produits sont les quatrièmes niveaux de classification des produits vendus par CASTO-MERLIN.

Il comprend actuellement 1 212 valeurs :

* « Poêle À Pétrole »
* « Poêle À Gaz »
* « Radiateur Soufflant »
* …

Chaque libellé est associé à un ID numérique, qui sera utilisé en tant que clé primaire : Les libellés peuvent évoluer, mais pas leurs ID associés. Chaque sous-famille est associée à exactement une famille, au travers du champ ID\_FAMILLE\_SSFAMILLE, qui est une clé étrangère vers la table des familles de produits.

Comme la quasi-totalité des tables du SI opérationnel de ventes de CASTO-MERLIN, il comprend le nom de l’utilisateur ayant fait l’INSERT de cette ligne, et la date (Resp. les champs « OPER\_CREAT » et « DATE\_CREAT »), ainsi que le nom de l’utilisateur ayant fait le dernier UPDATE de cette ligne, et la date (Resp. les champs « OPER\_MODIF » et « DATE\_MODIF »)

L’opérateur peut être le nom d’une personne (« Jean DUPOND ») ou d’un SI externe (« IBM Point Of Sales System »)

Structure

|  |  |  |
| --- | --- | --- |
| Nom colonne | Type | Nullable ? |
| ID\_SSFAMILLE | int | Non |
| LIBEL\_FAMILLE | nvarchar(256) | Non |
| ID\_RAYON\_SSFAMILLE | int | Non |
| DATE\_CREAT | datetime | Non |
| OPER\_CREAT | nvarchar(64) | Non |
| DATE\_MODIF | datetime | Oui |
| OPER\_MODIF | nvarchar(64) | Oui |

Remplissage

/\* A COMPLETER LORS DU LOT 3 : ETL \*/

## Table PRODUITS

But

Cette table contient les produits vendus par CASTO-MERLIN.

Chaque libellé de produit est associé à un ID numérique, qui sera utilisé en tant que clé primaire : Les libellés peuvent évoluer, mais pas leurs ID associés. Chaque produit est associé à exactement une sous-famille, au travers du champ ID\_ SSFAMILLE\_PRODUIT, qui est une clé étrangère vers la table des sous-familles de produits.

La table contient toutes les informations relatives au produit, indépendamment du circuit de vente :

* Description courte (LIBEL\_PRODUIT) et détaillée (DESC\_PRODUIT) du produit
* Prix d’achat HT du produit auprès du grossiste, ou directement du fabriquant
* TVA applicable. 20% par défaut
* Fabriquant du produit
* S’il existe : Grossiste ayant assuré l’intermédiaire de l’achat
* Code-barres

Comme la quasi-totalité des tables du SI opérationnel de ventes de CASTO-MERLIN, il comprend le nom de l’utilisateur ayant fait l’INSERT de cette ligne, et la date (Resp. les champs « OPER\_CREAT » et « DATE\_CREAT »), ainsi que le nom de l’utilisateur ayant fait le dernier UPDATE de cette ligne, et la date (Resp. les champs « OPER\_MODIF » et « DATE\_MODIF »)

L’opérateur peut être le nom d’une personne (« Jean DUPOND ») ou d’un SI externe (« IBM Point Of Sales System »)

Structure

|  |  |  |
| --- | --- | --- |
| Nom colonne | Type | Nullable ? |
| ID\_PRODUIT | int | Non |
| LIBEL\_PRODUIT | nvarchar(256) | Non |
| DESC\_PRODUIT | nvarchar(1024) | Non |
| ID\_SSFAMILLE\_PRODUIT | int | Non |
| CODE\_BARRE\_PRODUIT | int | Non |
| PRIX\_ACHAT | money | Non |
| TAUX\_TVA | decimal(4, 1) | Non |
| MARQUE\_PRODUIT | nvarchar(256) | Non |
| GROSSISTE\_PRODUIT | nvarchar(256) | Oui |
| DATE\_CREAT | datetime | Non |
| OPER\_CREAT | nvarchar(64) | Non |
| DATE\_MODIF | datetime | Oui |
| OPER\_MODIF | nvarchar(64) | Oui |

Remplissage

/\* A COMPLETER LORS DU LOT 3 : ETL \*/

## Table CLIENTS

But

Cette table contient tous les clients ayant acheté des produits chez CASTO-MERLIN au cours des 2 derniers mois (Historique flottant). Pour des raisons de volumétrie et de performance, la table est historisée vers une base de données secondaire, configurée pour le stockage en masse en lecture seule (au détriment des performances) pendant 2 ans. Après quoi, les données sont supprimées.

On distingue 5 types de clients (cf. champ « TYPE\_CLIENT ») :

* **A – Anonyme**. Ce sont la majorité des clients de magasins. Des particuliers sans programme de fidélité. L’hôtesse de caisse recueille seulement leur code-postal d’habitation à fins de statistiques.
* **I – Internet**. Ce sont les clients de site Internet sans programme de fidélité. Le site recueille leurs nom et adresse pour livraison et sécurité de la commande (Lutte contre la fraude)
* **N – Nominatif**. Ce sont des clients de type particulier, en magasin ou sur Internet, avec un programme de fidélité. Leur identifiant de carte fidélité est alors renseigné dans le champ « CODE\_CARTE\_FIDEL »
* **P – Professionnel de type artisan**.
* **S – Professionnel de type société**.

Comme la quasi-totalité des tables du SI opérationnel de ventes de CASTO-MERLIN, il comprend le nom de l’utilisateur ayant fait l’INSERT de cette ligne, et la date (Resp. les champs « OPER\_CREAT » et « DATE\_CREAT »), ainsi que le nom de l’utilisateur ayant fait le dernier UPDATE de cette ligne, et la date (Resp. les champs « OPER\_MODIF » et « DATE\_MODIF »)

L’opérateur peut être le nom d’une personne (« Jean DUPOND ») ou d’un SI externe (« IBM Point Of Sales System »)

Structure

|  |  |  |
| --- | --- | --- |
| Nom colonne | Type | Nullable ? |
| ID\_CLIENT | int | Non |
| NOM\_CLIENT | nvarchar(256) | Non |
| TYPE\_CLIENT | char(1) | Non |
| DATE\_NAISSANCE | date | Non |
| DATE\_SOUSCRIPTION | date | Non |
| ID\_VILLE\_CLIENT | int | Non |
| LIBEL\_ADRESSE | nvarchar(256) | Oui |
| CODE\_CARTE\_FIDEL | nvarchar(32) | Oui |
| TAUX\_REMISE | decimal(4, 1) | Non |
| DATE\_CREAT | datetime | Non |
| OPER\_CREAT | nvarchar(64) | Non |
| DATE\_MODIF | datetime | Oui |
| OPER\_MODIF | nvarchar(64) | Oui |

Remplissage

/\* A COMPLETER LORS DU LOT 3 : ETL \*/

## Table LIEUX

But

Cette table contient tous les lieux de ventes ou de stockage de CASTO-MERLIN depuis la création de l’entreprise. Ce peut être des lieux de ventes (Rayonnages des magasins, site Internet) de stockage (Entrepôts régionaux, partie stocks d’un magasin) ou des plateformes logistiques où les marchandises ne font que transiter.

On distingue 5 types de lieux (cf. champ « TYPE\_LIEU ») :

* **R – Rayons de vente du magasin.**
* **M – Partie stocks du magasin.**
* **I – Site Internet.**
* **S – Partie stocks pour Internet.**
* **E – Entrepôt régional.**
* **P – Plateforme logistique.**

Se reporter à la section « **Base opérationnelle – Vue d’ensemble** » pour la signification de ces termes.

Chaque lieu est dans une ville donnée et une seule. En cas de déménagement, la référence vers la ville sera modifiée. En cas de fermeture, le lieu sera supprimé de cette table dans la base opérationnelle.

Comme la quasi-totalité des tables du SI opérationnel de ventes de CASTO-MERLIN, il comprend le nom de l’utilisateur ayant fait l’INSERT de cette ligne, et la date (Resp. les champs « OPER\_CREAT » et « DATE\_CREAT »), ainsi que le nom de l’utilisateur ayant fait le dernier UPDATE de cette ligne, et la date (Resp. les champs « OPER\_MODIF » et « DATE\_MODIF »)

L’opérateur peut être le nom d’une personne (« Jean DUPOND ») ou d’un SI externe (« IBM Point Of Sales System »)

Structure

|  |  |  |
| --- | --- | --- |
| Nom colonne | Type | Nullable ? |
| ID\_LIEU | int | Non |
| TYPE\_LIEU | char(1) | Non |
| LIBEL\_LIEU | nvarchar(256) | Non |
| ID\_VILLE | int | Non |
| DATE\_CREAT | datetime | Non |
| OPER\_CREAT | nvarchar(64) | Non |
| DATE\_MODIF | datetime | Oui |
| OPER\_MODIF | nvarchar(64) | Oui |

Remplissage

/\* A COMPLETER LORS DU LOT 3 : ETL \*/

## Table VILLES

But

La liste et la codification des villes actuelles en France est issue du nommage officiel de l’INSEE :

<http://www.insee.fr/fr/methodes/nomenclatures/cog/telechargement.asp>

Dans la liste de l’INSEE de 2015, il y a 39 806 villes en France, DOM compris. C’est pourquoi la clé primaire a été typée INT. En revanche, les codifications de l’INSEE se font sur 1, 2 ou 3 chiffres, c’est pourquoi ces colonnes sont typées TINYINT.

Comme la quasi-totalité des tables du SI opérationnel de ventes de CASTO-MERLIN, il comprend le nom de l’utilisateur ayant fait l’INSERT de cette ligne, et la date (Resp. les champs « OPER\_CREAT » et « DATE\_CREAT »), ainsi que le nom de l’utilisateur ayant fait le dernier UPDATE de cette ligne, et la date (Resp. les champs « OPER\_MODIF » et « DATE\_MODIF »)

L’opérateur peut être le nom d’une personne (« Jean DUPOND ») ou d’un SI externe (« IBM Point Of Sales System »)

Structure

|  |  |  |
| --- | --- | --- |
| Nom colonne | Type | Nullable ? |
| ID\_VILLE | int | Non |
| CODE\_POSTAL | nvarchar(6) | Non |
| CODE\_COMMUNE | tinyint | Non |
| CODE\_REGION | tinyint | Non |
| CODE\_DEPARTEMENT | tinyint | Non |
| CODE\_ARRONDISEMENT | tinyint | Non |
| CODE\_CANTON | tinyint | Non |
| NOM\_VILLE\_MAJ | varchar(64) | Oui |
| NOM\_VILLE\_MIN | varchar(64) | Oui |
| DATE\_CREAT | datetime | Non |
| OPER\_CREAT | nvarchar(64) | Non |
| DATE\_MODIF | datetime | Oui |
| OPER\_MODIF | nvarchar(64) | Oui |

Remplissage

/\* A COMPLETER LORS DU LOT 3 : ETL \*/

## Table STOCKS

But

/\* A COMPLETER SI VALORISATION ULTERIEURE \*/

Comme la quasi-totalité des tables du SI opérationnel de ventes de CASTO-MERLIN, il comprend le nom de l’utilisateur ayant fait l’INSERT de cette ligne, et la date (Resp. les champs « OPER\_CREAT » et « DATE\_CREAT »), ainsi que le nom de l’utilisateur ayant fait le dernier UPDATE de cette ligne, et la date (Resp. les champs « OPER\_MODIF » et « DATE\_MODIF »)

L’opérateur peut être le nom d’une personne (« Jean DUPOND ») ou d’un SI externe (« IBM Point Of Sales System »)

Structure

|  |  |  |
| --- | --- | --- |
| Nom colonne | Type | Nullable ? |
| ID\_STOCK | int | Non |
| DATE\_RECENSEMENT | datetime | Non |
| OPER\_RECENSEMENT | nvarchar(64) | Non |
| ID\_LIEU | int | Non |
| ID\_PRODUIT | int | Non |
| NBR\_DISPO | int | Non |
| NBR\_DEFECTUEUX | int | Non |
| NBR\_RETOUR\_SAV | int | Non |
| DATE\_CREAT | datetime | Non |
| OPER\_CREAT | nvarchar(64) | Non |
| DATE\_MODIF | datetime | Oui |
| OPER\_MODIF | nvarchar(64) | Oui |

Remplissage

/\* A COMPLETER LORS DU LOT 3 : ETL \*/

/\* A COMPLETER SI VALORISATION ULTERIEURE \*/

## Table VENTES

But

La table des ventes est une table d’association entre :

* **Un ticket**, c’est-à-dire de 1 à N articles vendus (Clé étrangère vers la table TICKETS)
* **Un client** (Clé étrangère vers la table CLIENTS)
* **Un lieu**, de type Internet ou magasin (Clé étrangère vers la table LIEUX)

La vente a un montant global, et une somme de TVA, tous deux typés MONEY, plus adapté que DECIMAL au stockage de valeurs financières.

Le champ « Operateur Ventes » stocke, pour information, le matricule de l’employée ayant réalisé la vente (0 si site Internet)

Exception du SI opérationnel de ventes de CASTO-MERLIN : Il ne comprend que le nom de l’utilisateur ayant fait l’INSERT de cette ligne, et la date (Resp. les champs « OPER\_CREAT » et « DATE\_CREAT »)

Il n’y a pas de champs pour le nom d’utilisateur ayant fait le dernier UPDATE de cette ligne, et la date car cette table n’est pas mise à jour une fois la vente effectuée. En cas d’annulation de la vente, la ligne n’est jamais insérée, et en cas de remboursement, il est créé une vente de montant opposé à la vente à rembourser.

L’opérateur peut être le nom d’une personne (« Jean DUPOND ») ou d’un SI externe (« IBM Point Of Sales System »)

Structure

|  |  |  |
| --- | --- | --- |
| Nom colonne | Type | Nullable ? |
| ID\_VENTE | int | Non |
| OPER\_VENTE | nvarchar(64) | Non |
| ID\_TICKET | int | Non |
| ID\_CLIENT | int | Non |
| ID\_LIEU | int | Non |
| MONTANT\_HT\_VENTE | money | Non |
| MONTANT\_TVA\_VENTE | money | Non |
| DATE\_CREAT | datetime | Non |
| OPER\_CREAT | nvarchar(64) | Non |

Remplissage

/\* A COMPLETER LORS DU LOT 3 : ETL \*/

## Table TICKETS

But

Comme en magasin, le ticket est un ensemble de 1 à N produits vendus. On ne précise ni le client ni le lieu (Ces informations sont reprises dans la table VENTES)

On en relie pas directement les produits, mais les couples « Prix – Produits », car un même produit peut avoir un prix diffèrent en fonction des magasins et de leur politique propre de prix.

Si dans un même ticket, le client achète plusieurs fois le même produits, alors le champ QUANTITE sera diffèrent de 1.

Exception du SI opérationnel de ventes de CASTO-MERLIN : Il ne comprend que le nom de l’utilisateur ayant fait l’INSERT de cette ligne, et la date (Resp. les champs « OPER\_CREAT » et « DATE\_CREAT »)

Il n’y a pas de champs pour le nom d’utilisateur ayant fait le dernier UPDATE de cette ligne, et la date car cette table n’est pas mise à jour une fois la vente effectuée. En cas d’annulation de la vente, la ligne n’est jamais insérée, et en cas de remboursement, il est créé une vente de montant opposé à la vente à rembourser.

L’opérateur peut être le nom d’une personne (« Jean DUPOND ») ou d’un SI externe (« IBM Point Of Sales System »)

Structure

|  |  |  |
| --- | --- | --- |
| Nom colonne | Type | Nullable ? |
| ID\_TICKET | int | Non |
| ID\_PRIXPRODUIT | int | Non |
| QUANTITE | int | Non |
| DATE\_CREAT | datetime | Non |
| OPER\_CREAT | nvarchar(64) | Non |

Remplissage

/\* A COMPLETER LORS DU LOT 3 : ETL \*/

## Table PRIXPRODUITS

But

La tale PRIXPRODUITS fixe le prix d’un produit dans un lieu donné. En effet, un même produit peut avoir un prix diffèrent en fonction des magasins et de leur politique propre de prix.

Le champ « Operateur Prix Produit » stocke, pour information, le matricule de l’employée ayant fixé le prix du produit sur le lieu de vente donné.

On peut noter une légère entorse à la normalisation du modèle relationnel :

* La table VENTES relie (Entre autre) un lieu à un ticket
* La table TICKETS relie un ticket à un Prix-produit
* La table PRIXPRODUIT relie un produit à un prix dans un lieu donné

On voit donc une redondance de l’information « lieu » entre la table des VENTES et celle des PRIXPRODUITS. Cette information a été dupliquée pour améliorer la lisibilité et l’exploitation de la table des VENTES.

Comme la quasi-totalité des tables du SI opérationnel de ventes de CASTO-MERLIN, il comprend le nom de l’utilisateur ayant fait l’INSERT de cette ligne, et la date (Resp. les champs « OPER\_CREAT » et « DATE\_CREAT »), ainsi que le nom de l’utilisateur ayant fait le dernier UPDATE de cette ligne, et la date (Resp. les champs « OPER\_MODIF » et « DATE\_MODIF »)

L’opérateur peut être le nom d’une personne (« Jean DUPOND ») ou d’un SI externe (« IBM Point Of Sales System »)

Structure

|  |  |  |
| --- | --- | --- |
| Nom colonne | Type | Nullable ? |
| ID\_PRIXPRODUIT | int | Non |
| OPER\_PRIXPRODUIT | nvarchar(64) | Non |
| ID\_LIEU | int | Non |
| ID\_PRODUIT | int | Non |
| MONTANT\_HT\_PRODUIT | money | Non |
| MONTANT\_TVA\_PRODUIT | money | Non |
| MONTANT\_MARGE\_PRODUIT | money | Non |
| DATE\_CREAT | datetime | Non |
| OPER\_CREAT | nvarchar(64) | Non |
| DATE\_MODIF | datetime | Oui |
| OPER\_MODIF | nvarchar(64) | Oui |

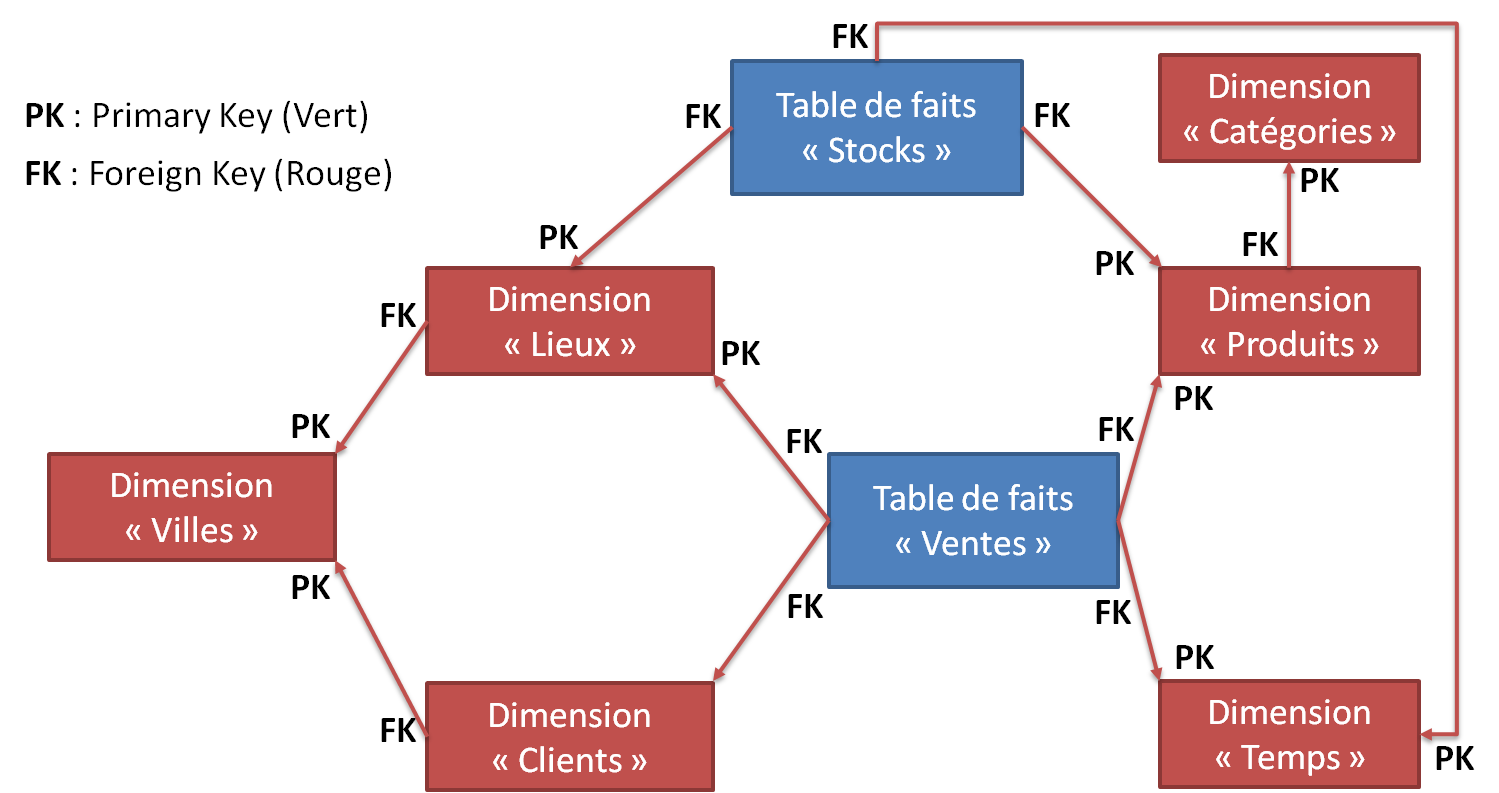
Remplissage

/\* A COMPLETER LORS DU LOT 3 : ETL \*/

# Entrepôt de données

## Vue d’ensemble

Dans la base dédiée « **DataWarehouseODE** », nous créons un schéma dédié « **ODE\_DATAWAREHOUSE** », qui contient les tables suivantes :



## Table de dimension CATEGORIES

But

/\***BRICE** : A COMPLETER \*/

Structure

|  |  |  |
| --- | --- | --- |
| Nom colonne | Type | Nullable ? |
| CATEGORIE\_PK | int | Non |
| LIBEL\_UNIVERS | nvarchar(256) | Non |
| ID\_UNIVERS | int | Non |
| LIBEL\_RAYON | nvarchar(256) | Non |
| ID\_RAYON | int | Non |
| LIBEL\_FAMILLE | nvarchar(256) | Non |
| ID\_FAMILLE | int | Non |
| LIBEL\_SSFAMILLE | nvarchar(256) | Non |
| ID\_SSFAMILLE | int | Non |

**Clé primaire** : CATEGORIE\_PK (Clé technique)

**Clé étrangère** : ∅ (C’est une table de dimension en fin de branche du flocon)

Remplissage via l’ETL

/\* A COMPLETER LORS DU LOT 3 : ETL \*/

Remplissage pour les tests

/\***BRICE** : A COMPLETER \*/

## Table de dimension PRODUITS

But

/\***THOMAS** : A COMPLETER \*/

Structure

|  |  |  |
| --- | --- | --- |
| Nom colonne | Type | Nullable ? |
| PRODUIT\_PK | int | Non |
| CATEGORIE\_FK | int | Non |
| LIBEL\_PRODUIT | nvarchar(256) | Non |
| PRIX\_ACHAT | money | Non |
| TAUX\_TVA | decimal(4, 1) | Non |
| MARQUE\_PRODUIT | nvarchar(256) | Non |
| GROSSISTE\_PRODUIT | nvarchar(256) | Oui |

**Clé primaire** : PRODUIT\_PK (Clé technique)

**Clé étrangère** : CATEGORIE\_FK → CATEGORIES.CATEGORIE\_PK

Remplissage via l’ETL

/\* A COMPLETER LORS DU LOT 3 \*/

Remplissage pour les tests

/\***THOMAS** : A COMPLETER \*/

## Table de dimension TEMPS

But

La table TEMPS est une table de dimension utilisée pour quantifier la dimension d’analyse « Temps » des faits de ventes et de stocks.

Cette dimension est une quasi-obligatoire dans les SI décisionnels d’entreprise.

Nous la imiterons à la granularité « Jour », et non heure. En effet, les métiers du Marketing et du Commercial ne requièrent pas une telle finesse.

On définit 4 « groupes » de temps :

* Année
* Trimestre
* Mois
* Semaine

Chaque groupe est défini simultanément par :

* Un code numérique, pour les recherches
* Un libellé, pour l’affichage à l’utilisateur
* Une date, qui correspond au Jour-Mois-Année du début de la période.

Par exemple, le 1er juillet année 2014 sera identifiée par :

* Le code année « 2014 »
* Le libellé année « Calendrier 2014 »
* La date année « 01-01-2014 »
* Le code mois « 052014 »
* Le libellé mois « Juillet 2014 »
* La date mois « 01-07-2014 »

Structure

|  |  |  |
| --- | --- | --- |
| Nom colonne | Type | Nullable ? |
| TEMPS\_PK | int | Non |
| DATE | smalldatetime | Non |
| JOUR | nvarchar (64) | Oui |
| ANNEE\_CODE | int | Oui |
| ANNEE\_DATE | smalldatetime | Non |
| ANNEE\_NOM | nvarchar (64) | Oui |
| TRIMESTRE\_CODE | int | Oui |
| TRIMESTRE\_DATE | smalldatetime | Non |
| TRIMESTRE\_NOM | nvarchar (64) | Oui |
| MOIS\_CODE | int | Oui |
| MOIS\_DATE | smalldatetime | Non |
| MOIS\_NOM | nvarchar (64) | Oui |
| SEMAINE\_CODE | int | Oui |
| SEMAINE\_DATE | smalldatetime | Non |
| SEMAINE\_NOM | nvarchar(64) | Oui |

**Clé primaire** : TEMPS\_PK (Clé technique)

Contrairement aux autres tables de dimensions, la clé primaire technique n’est pas auto-incrémentée (Mot-clé IDENTITY dans les scripts de création de table) En effet, pour des raisons d’optimisation de performances et de clarté, les clés primaires sont au format Année – Mois – Jour « AAAAMMDD ». Par exemple : « 20151231 » pour 31 décembre 2015.

**Clé étrangère** : ∅

Il n’y a pas de clé étrangère car c’est une table de dimension en fin de branche du flocon. Elle ne prend donc plus de référence dans d’autres tables de dimensions.

Remplissage via l’ETL

/\* A COMPLETER LORS DU LOT 3 \*/

Remplissage pour les tests

Afin de prendre large, elle a été remplie du 01 janvier 2009 au 31 décembre 2018 (Soit 3 652 jours)

Les outils graphiques de SQL Server 2014 proposent des assistants de création de cette dimension « standard », avec une granularité journalière. Toutefois, il reste un travail d’optimisation sur cette dimension auto-générée par l’outil. Il a donc été choisi de charger un fichier CSV en table de dimension DIM\_TEMPS, issue du livre « Business Intelligence avec SQL Server 2014 » de Sébastien FANTINI et Franck GAVAND aux Editions ENI.

## Table de dimension LIEUX

But

/\***THOMAS** : A COMPLETER \*/

Structure

|  |  |  |
| --- | --- | --- |
| Nom colonne | Type | Nullable ? |
| LIEU\_PK | int | Non |
| VILLE\_FK | int | Non |
| TYPE\_LIEU | char(1) | Non |
| LIBEL\_LIEU | nvarchar(256) | Non |
| DATE\_OUVERTURE | date | Non |
| DATE\_FERMETURE | date | Non |
| SURFACE\_M2 | numeric(6, 1) | Oui |

**Clé primaire** : LIEU\_PK (Clé technique)

**Clé étrangère** : VILLE\_FK → VILLES.VILLE\_PK

Remplissage via l’ETL

/\* A COMPLETER LORS DU LOT 3 \*/

Remplissage pour les tests

/\***THOMAS** : A COMPLETER \*/

## Table de dimension CLIENTS

But

/\***CEDRIC** : A COMPLETER \*/

Structure

|  |  |  |
| --- | --- | --- |
| Nom colonne | Type | Nullable ? |
| CLIENT\_PK | int | Non |
| VILLE\_FK | int | Non |
| TAUX\_REMISE | decimal(6, 2) | Non |
| TYPE\_CLIENT | char(1) | Non |
| NOM\_CLIENT | nvarchar(256) | Non |
| DATE\_NAISSANCE | date | Non |
| DATE\_SOUSCRIPTION | date | Non |
| CODE\_FIDELITE | nvarchar(32) | Oui |

**Clé primaire** : CLIENT\_PK (Clé technique)

**Clé étrangère** : VILLE\_FK → VILLES.VILLE\_PK

Remplissage via l’ETL

/\* A COMPLETER LORS DU LOT 3 \*/

Remplissage pour les tests

/\***CEDRIC** : A COMPLETER \*/

## Table de dimension VILLES

But

/\***BERNARD** : A COMPLETER \*/

Structure

|  |  |  |
| --- | --- | --- |
| Nom colonne | Type | Nullable ? |
| VILLE\_PK | int | Non |
| CODE\_POSTAL | nvarchar(6) | Non |
| CODE\_COMMUNE | int | Non |
| CODE\_REGION | int | Non |
| CODE\_DEPARTEMENT | int | Non |
| CODE\_ARRONDISEMENT | int | Non |
| CODE\_CANTON | int | Non |
| NOM\_VILLE\_MAJ | nvarchar(256) | Non |
| NOM\_VILLE\_MIN | nvarchar(256) | Non |
| POPULATION | int | Oui |

**Clé primaire** : VILLE\_PK (Clé technique)

**Clé étrangère** : ∅ (C’est une table de dimension en fin de branche du flocon)

Remplissage via l’ETL

/\* A COMPLETER LORS DU LOT 3 \*/

Remplissage pour les tests

/\***BERNARD** : A COMPLETER \*/

## Table de faits VENTES

But

Cette table de faits contient toutes les ventes depuis la mise en service du SI décisionnel, le 01 janvier 2010.

Chaque ligne enregistre la vente d’un produit (Ou de plusieurs identiques, cf. champs quantité) à un client, sur un magasin ou sur Internet, et à une date précise.

Structure

|  |  |  |
| --- | --- | --- |
| Nom colonne | Type | Nullable ? |
| DATE\_VENTE\_FK | int | Non |
| PRODUIT\_FK | int | Non |
| CLIENT\_FK | int | Non |
| LIEU\_FK | int | Non |
| MONTANT\_HT\_VENTE | money | Non |
| MONTANT\_TVA\_VENTE | money | Non |
| MARGE\_BRUTE | money | Non |
| UNITES\_VENDUES | int | Non |
| NUM\_TICKET | nvarchar(256) | Non |

**Clé primaire** : ∅

La table de Ventes étant une table de faits, elle ne contient aucune clé primaire, qui n’a aucune utilité puisque les faits ne sont pas exploités individuellement, ais collectivement par le biais de regroupements et de statistiques.

**Clés étrangères** :

* DATE\_VENTE\_FK → TEMPS.TEMPS\_PK
* PRODUIT\_FK → PRODUITS.PRODUIT\_PK
* CLIENT\_FK → CLIENTS.CLIENT\_PK
* LIEU\_FK → LIEUX.LIEU\_PK

Fonctionnellement, chaque Vente lie un client, un lieu (Le magasin ou le site Internet) et un produit, à une date donnée. Techniquement, la table de faits Ventes utilise des clés étrangère vers chacune de ces dimensions.

Remplissage via l’ETL

/\* A COMPLETER LORS DU LOT 3 \*/

Remplissage pour les tests

Pour chaque année entre 2010 et 2015, on considère un volume de ventes stable :

* 2.2 Millions de ventes annuelles sur Internet
* 9.6 Millions de ventes annuelles pour l’ensemble des magasins physiques

Concernant le champ « Marge brute » :

* Entre 24.3 et 64.3 % de marge sur Internet - Moyenne à 44.3 %
* Entre 14.8 et 54.8 % de marge en Magasin - Moyenne à 34.8 %

Qui correspond à de « bons » chiffres dans la grande distribution de Bricolage.

Concernant le nombre moyens d’articles par panier :

* Entre 1 et 9 articles sur Internet - Moyenne de 5
* Entre 1 et 35 articles en Magasin - Moyenne de 18

Dans ces deux cas : On commence par choisir un client et un ticket de caisse, puis on ajoute un nombre aléatoire d’articles pour rentrer dans cette fourchette de nombre d’articles. Le nombre d’articles identiques ajoutés dans le panier d’un client est stocké dans le champ « Unités vendues ». En effet, il est fréquent qu’au cours d’un acte d’achat, le client prenne plusieurs fois exactement le même produit (Exemple : 5 planches, 20 paquets de clous, 2 tournevis…) Afin d’économiser de la place, nous regroupons ces ventes d’articles identiques à un même client dans une seule ligne de table de faits Ventes.

Le champ « Montant HT » va contenir le montant total HT de la vente, c’est-à-dire le prix d’achat du produit (Cf. table de dimension « Produits ») plus la marge brute, multiplié par le nombre d’articles identiques vendus (Cf. champ « Unités vendues »)

Le champ « Montant TVA » représente la TVA de cette même vente. Le taux appliqué sur ce produit est précisé dans la table de dimension « Produits ».

Enfin, le champ « Numero de Ticket » est dit « dégénéré » : Il s’agit d’une clé primaire du SI opérationnel des ventes. Ce champs n’est utile que pour remonter à la source de la vente (« Pour information ») et ne sera jamais valorisé dans le SI décisionnel.

## Table de faits STOCKS

But

/\* A COMPLETER SI VALORISATION ULTERIEURE \*/

Structure

|  |  |  |
| --- | --- | --- |
| Nom colonne | Type | Nullable ? |
| DATE\_INVENTAIRE\_FK | int | Non |
| PRODUIT\_FK | int | Non |
| LIEU\_FK | int | Non |
| NBR\_DISPO | int | Non |
| NBR\_DEFECTUEUX | int | Non |
| ID\_INVENTAIRE | int | Non |

**Clé primaire** : ∅ (C’est une table de faits)

**Clés étrangères** :

* DATE\_INVENTAIRE\_FK → TEMPS.TEMPS\_PK
* PRODUIT\_FK → PRODUITS.PRODUIT\_PK
* LIEU\_FK → LIEUX.LIEU\_PK

Remplissage via l’ETL

/\* A COMPLETER SI VALORISATION ULTERIEURE \*/

Remplissage pour les tests

/\* A COMPLETER SI VALORISATION ULTERIEURE \*/