



BUSINESS ANALYTICS

DAFFE Tristan – 3IG – Groupe B2
ROUSSEAU Benjamin – 2IG – Groupe C

Henallux – Informatique de Gestion – 2022-2023

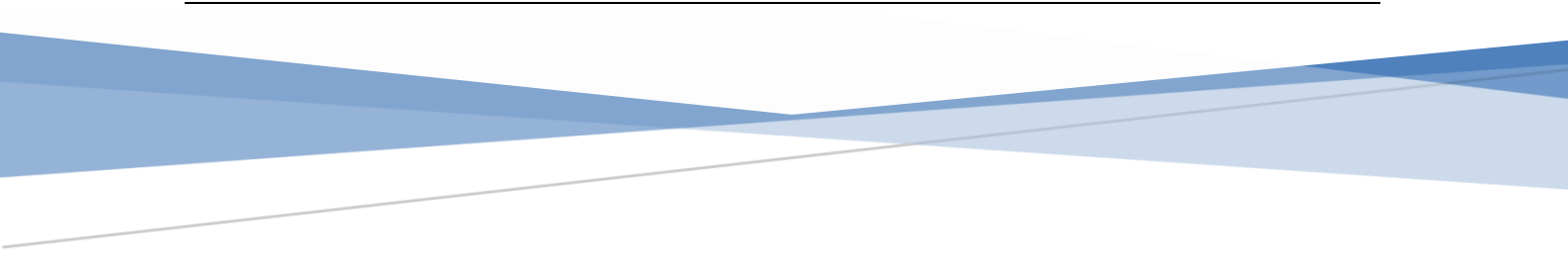


Table des matières

OSolutions Proposées	2
Modèles	2
Data warehouse: schéma star.....	2
Base de données opérationnelle fournie par le client :	3
Table de faits.....	4
Justifications	4
Clés primaires Data Warehouse	4
Modélisation des dimensions et faits.....	4
Limites rencontrées.....	5
Stratégies d'exécution ETL	5

Solutions Proposées

Notre solution se compose d'un projet ETL pour créer et alimenter des tables de dimensions et une table de faits afin de passer d'une base de données opérationnelle à un data warehouse. Cet ETL sera ensuite utilisé pour alimenter un cube OLAP pour visualiser les données et répondre aux besoins métier.

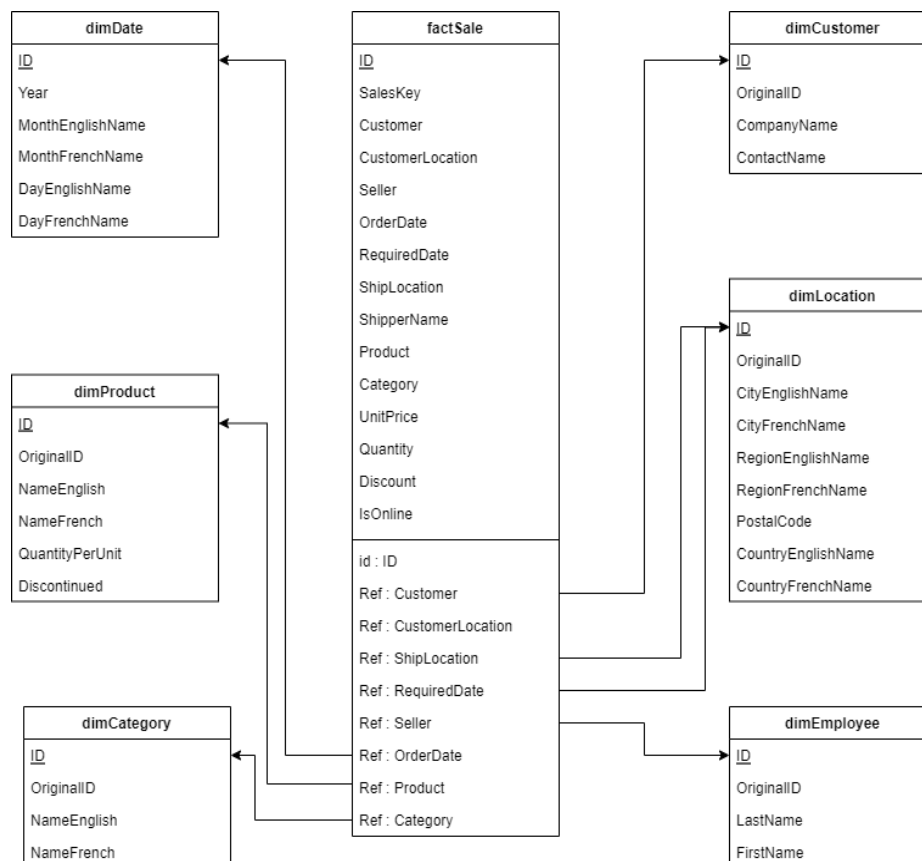
Pour ce faire, nous sommes partis d'une base de données fournit par le client et nous avons réalisé l'ETL grâce aux outils Visual Studio et SSIS toolbox.

Dans Visual Studio, les principaux composants utilisés ont été :

- Colonnes dérivées
- Source et destination OLE DB
- Composants script C# et SQL
- Agrégations pour supprimer les doublons
- Unir tout pour rassembler les données de différentes sources
- Recherche

Modèles

Data warehouse: schéma star



Base de données opérationnelle fournie par le client :

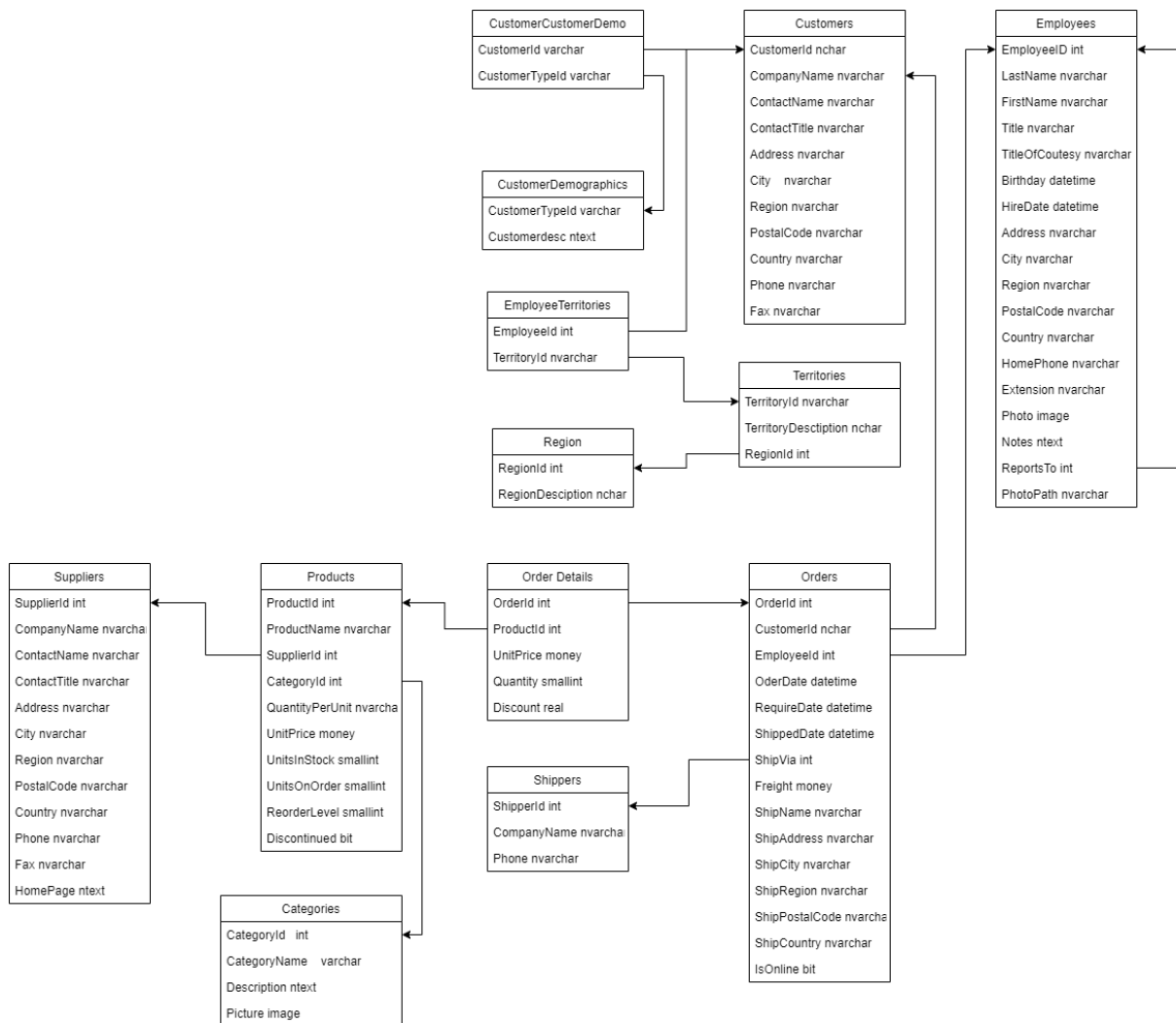


Table de faits

Dans notre table de faits, nous avons gardé toutes les informations concernant une ligne de commande. Nous avons ajouté à cela les informations sur le lieu de livraison, la localisation du client, les informations sur le client, sur l'employé qui a vendu le produit si la vente n'est pas faite en ligne, les informations sur le produit et la catégorie associée et la date de la commande via des clés étrangères vers nos dimensions.

Justifications

Clés primaires Data Warehouse

Les clés primaires dans notre data Warehouse pour notre table de faits sont générées de manière automatique (surrogatekey).

Les clés primaires de nos dimensions sont créées à partir des clés primaires de la base de données opérationnelle avec un ajout. La seule exception est pour la table dimDate. Nous avons créé son identifiant en concaténant le mois, le jour et l'année.

La clé primaire originelle de chaque élément est sauvegardée dans une colonne nommée OriginalID.

Modélisation des dimensions et faits

Nous avons choisi de faire une table des faits sur les ventes car une vente est une donnée quantitative qui permet de suivre les performances de l'entreprise et mesurer les résultats des activités commerciales (caractéristique d'une table de faits). Cette table va donc contenir les informations sur une vente et les clés étrangères vers nos dimensions.

Nous avons mis comme dimension la date pour pouvoir accéder aux informations concernant les ventes sur une période de temps donnée. La dimension de localisation permet d'analyser les données de ventes en les séparant par régions géographiques avec un degré de précision variable (pays, régions, ville).

Une analyse sur les produits et catégories des objets vendus est possible via nos dimensions dimProduct et dimCategory avec les noms des catégories également traduits en français et anglais.

Enfin, les performances de chaque employé et les achats de chacun de nos clients peut être analysé au travers de nos dimensions dimCustomer et dimEmployee.

Limites rencontrées

La limite majeure rencontrée pour la création du data warehouse fut le multilinguisme. Certains champs comme le nom des catégories et le jour de la semaine et mois de la date ont été traduits facilement.

En revanche, le nom des produits n'est pas traduit. Cela est dû à un manque de temps et à la complexité de cette tâche. En effet, les noms de produits étant nombreux et enregistrés dans différentes langues (français, anglais, espagnol, ...) mais dans une seule colonne, notre moyen de traduction s'est révélé inadapté. Il en va de même pour la traduction de la table dimLocation.

Une autre limite vis-à-vis des demandes des clients concerne la superficie de vente. Nous n'avons pas trouvé cette donnée dans les tables et sans explications supplémentaires sur la signification de chaque colonne, nous ne savons pas répondre à cette question.

Stratégies d'exécution ETL

Lors de chaque exécution de notre ETL, les tables et données déjà présentes sont supprimées ce qui permet de relancer l'ETL en cas d'arrêt soudain lors de son exécution.

Les erreurs venant de valeurs NULL dans la base de données opérationnelles sont remplacées par des « dummy values », des valeurs décidées à l'avance pour marquer le champ comme anormal.

En conclusion, les enregistrements erronés sont marqués comme anormaux mais le processus n'est pas arrêté.