Intégration des données

Exercice : ShopPlus – Une plateforme e-commerce en croissance

Une entreprise de commerce électronique, ShopPlus, a mis en place un Entrepôt de Données (ED) Oracle afin de mieux analyser ses activités commerciales. Cet ED a pour objectif de centraliser les informations relatives aux commandes effectuées par les clients. L’entreprise vend des milliers de produits répartis dans plusieurs catégories telles que l’informatique, la maison ou la mode. Chaque produit appartient à une seule catégorie, et chaque catégorie est définie par un nom et une description. Les commandes sont passées par les clients via le site web de ShopPlus. Chaque commande est liée à un seul client et peut contenir plusieurs produits. Pour chaque produit contenu dans une commande, la quantité commandée est précisée. Chaque commande est datée afin de permettre une analyse temporelle des ventes.

Le schéma conceptuel de cet ED est le suivant :

Une image contenant texte, capture d’écran, Police, conception

Le contenu généré par l’IA peut être incorrect.

Figure 1 Schéma Entité-Association

Le schéma relationnel de l'entrepôt de données est le suivant :

Client = (idClient, nomClient, prenomClient, emailClient, villeClient);

Category = (idCategory, nomCategory, description);

Commande = (idCommande, dateCommande, idClient\*);

Produit = (idProduit, nomProduit, prixUnitaire, idCategory\*);

contenir = (idCommande\*, idProduit\*, quantité);

Les informations relatives aux produits et aux commandes sont stockées dans un SGBD Oracle. Les catégories des produits proviennent d’un logiciel logistique et sont disponibles sous forme de fichier CSV. Les profils des clients, quant à eux, sont gérés dans une application web et sont exportés sous forme de fichier JSON.

Les sources de données sont les suivants :

Source 1 : Base de données Oracle – Information des commandes, des produits et des ventes



Figure 2 Table Commande

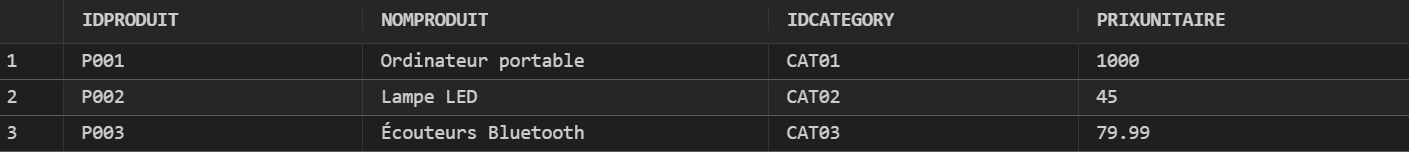


Figure 3 Table Produit

Une image contenant capture d’écran, texte, noir

Le contenu généré par l’IA peut être incorrect.

Figure 4 Table contenir

Source 2 : Ficher CSV – Informations des catégories

Une image contenant texte, capture d’écran, Police

Le contenu généré par l’IA peut être incorrect.

Figure 5 Fichier category.csv

Source 3 : Fichier JSON : Profil clients

Une image contenant texte, capture d’écran, menu

Le contenu généré par l’IA peut être incorrect.

Figure 6 Fichier client.json

1. **Processus ETL pour l’entrepôt de données**
   1. Quelles sont les caractéristiques des 3 sources de données ? (structurées / semi-structurées)
   2. Qu’est-ce qu’il faut faire pour transformer les données dans les sources ?
   3. Expliquez les différences entre le chargement incrémental et le chargement fusion.
   4. Modéliser le processus de ETL.
   5. Il y a un fichier CSV [« 20230823-communes-departement-region »](https://www.data.gouv.fr/fr/datasets/r/dbe8a621-a9c4-4bc3-9cae-be1699c5ff25) qui contient les codes postaux et les noms des villes. Extrayez, transformez et chargez les données en utilisant Talend.
2. **Modélisation et ETL d’un Magasin de Données (MD) multidimensionnel**

Cet ED sert de base à la construction d’un Magasin de Données Multidimensionnelles destiné à l’équipe marketing de l’entreprise ShopPlus. Cette équipe souhaite connaître le chiffre d’affaires par client, par produit et par période (jour, mois, année).

* 1. Donnez le dictionnaire des mesures.
  2. Donnez une représentation graphique du système décisionnel complet.
  3. Extrayez, transformez et chargez les données en utilisant Talend.