

## - Système d'Aide à la Décision -

### TP 2 : Pentaho Data Integration (PDI)

« Création et Alimentation du DW »

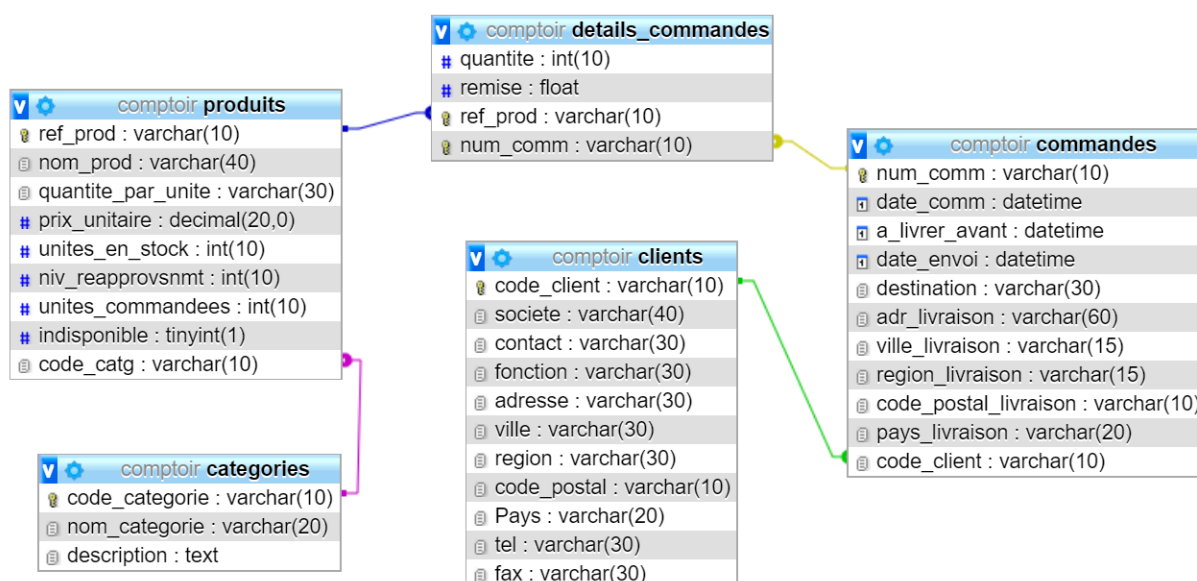
#### Objectifs du TP :

1. Créer la base de données source de production.
2. Concevoir le data warehouse cible sous la forme d'un schéma en étoile.
3. Configurer le processus ETL pour alimenter notre entrepôt de données.

#### Phase 1 : Création de la Base de Données Source

Supposons que vous collectez des informations sur les ventes d'une société dans des tables liées dans une base de données relationnelle. A la fin de chaque mois, vous devez connaître la **quantité de produits vendus** ainsi que le **montant total de ventes** pour chaque client, pour chaque produit.

1. Pour commencer, créer une nouvelle base de données « *comptoir* » via MySQL en respectant le schéma suivant :

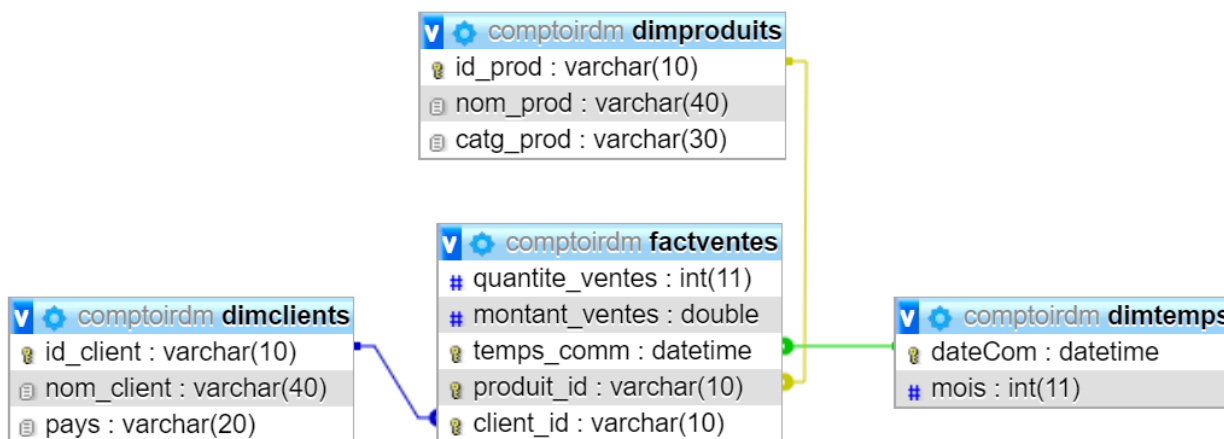


2. Insérez dans vos tables les données suivantes :

- 3 catégories différentes.
- 10 produits.
- 10 commandes avec des quantités différentes et des commandes sur plusieurs produits.
- 10 clients.

## Phase 2 : Conception de la Base de Données Cible (Datamart)

A partir de cette base de données opérationnelle « *comptoir* », on vise à produire un entrepôt de données (plus précisément, un datamart) pour étudier les performances de *Ventes* selon plusieurs dimensions : *Temps*, *Clients* et *Produits*. Pour ce faire, créer une base de données « **comptoirDM** » contenant les 3 tables *DimTime*, *DimRegion* et *DimProduit* qui serviront de dimensions, et *FactVentes* comme table de fait (le diagramme ci-dessous).



## Phase 3 : Intégration des données via la couche ETL (Data Flow)

A cette étape, utiliser *Pentaho Data Integration* pour alimenter l'entrepôt en utilisant les données provenant de la base de données en production.

1. Créer une nouvelle transformation.
2. Ajouter une fonction d'extraction (*Extraction depuis Table*).
3. Renommer la fonction d'extraction : *Extraction depuis Produits*.

4. Créer une nouvelle connexion à la base de données « **comptoir** » : *cn\_comptoirDB*. Tester la connexion.
5. Intégrer la requête SQL permettant d'extraire les granularités de la dimension produit.
6. Ajouter une fonction d'alimentation : *Insertion dans table*.
7. Renommer la fonction d'alimentation : *Alimentation dimension produit*.
8. Créer une nouvelle connexion à l'entrepôt « **comptoirDM** » : *cn\_datamart*.
9. Configurer correctement le mapping des champs.
10. Refaire le même processus précédent pour les deux autres dimensions et la table de faits ventes.
11. Démarrer le workflow ETL.
12. Vérifier que votre datamart est bien rempli.

