

| Tables fait-dimension

| Proposition de Modèle de Données (Star Schema)

| 1. Table de Faits : Fact_Sales (Table des Ventes)

C'est le cœur de notre modèle. Chaque ligne représente une ligne de commande unique.

- **Clés Étrangères (Foreign Keys - FK) :**
 - `Order_Date_Key` (FK vers `Dim_Date`)
 - `Ship_Date_Key` (FK vers `Dim_Date`)
 - `Product_Key` (FK vers `Dim_Product`)
 - `Customer_Key` (FK vers `Dim_Customer`)
 - `Geography_Key` (FK vers `Dim_Geography`)
 - `ShipMode_Key` (FK vers `Dim_ShipMode`)
- **Mesures (Measures) / Faits :**
 - `Sales` (déjà présente)
 - `Quantity` (déjà présente)
 - `Profit` (déjà présente)
 - `Order_Line_ID` (équivalent de `Row ID` si chaque Row ID représente une ligne de commande unique)
- **Colonnes à Exclure (ou à ne pas mettre dans la table de faits si elles sont déjà dans une dimension) :**
 - `Order ID` (peut être une colonne dans la table de faits si on veut l'analyser, mais pas une clé étrangère vers une dimension `Dim_Order` sauf si on veut ajouter des attributs à la commande elle-même, ce qui n'est pas le cas ici). Pour nos objectifs, le `Row ID` ou un `Order_Line_ID` est suffisant.

| 2. Tables de Dimensions :

Ces tables contiendront les attributs qui décrivent "qui", "quoi", "où", "quand" et "comment" s'est passée la vente.

I a. **Dim_Date** (Dimension Temps)

Cette dimension est essentielle pour toute analyse temporelle. Puisque tu auras **Order Date** et **Ship Date** en format AAAA-MM-JJ, tu pourras générer cette table à partir de ces dates ou créer une table calendaire complète.

- **Clé Primaire (Primary Key – PK) :**

- **Date_Key** (format entier, par exemple **20160722** pour le 22 juillet 2016)

- **Attributs :**

- **Full_Date** (la date complète : **AAAA-MM-JJ**)
- **Day_Of_Week** (ex: "Vendredi")
- **Day_Of_Month** (ex: 22)
- **Month** (ex: 7)
- **Month_Name** (ex: "Juillet")
- **Quarter** (ex: Q3)
- **Year** (ex: 2016)
- **Week_Number** (Numéro de semaine dans l'année)
- **Is_Weekend** (Booléen, Vrai/Faux)
- *Potentiellement* : **Fiscal_Year** , **Fiscal_Quarter** si l'entreprise a un calendrier fiscal différent.

I b. **Dim_Product** (Dimension Produit)

Décrit les produits vendus.

- **Clé Primaire (PK) :**

- **Product_Key** (basée sur **Product ID**)

- **Attributs :**

- **Product_ID** (l'ID original du produit)
- **Product_Name**

- `Sub_Category`
- `Category`

c. `Dim_Customer` (Dimension Client)

Décrit les clients.

- **Clé Primaire (PK) :**
 - `Customer_Key` (basée sur `Customer ID`)
- **Attributs :**
 - `Customer_ID` (l'ID original du client)
 - `Customer_Name`
 - `Segment`

d. `Dim_Geography` (Dimension Géographie)

Décrit les lieux de vente.

- **Clé Primaire (PK) :**
 - `Geography_Key` (peut être générée ou un agrégat de `State` et `City` si `Country` est toujours "United States")
- **Attributs :**
 - `Country` (sera toujours "United States" ici, mais bonne pratique de l'inclure)
 - `Region`
 - `State`
 - `City`

e. `Dim_ShipMode` (Dimension Mode d'Expédition)

Décrit les modes d'expédition.

- **Clé Primaire (PK) :**
 - `ShipMode_Key` (clé numérique simple)

- **Attributs :**
 - **Ship_Mode_Name** (ex: "First Class", "Same Day", "Second Class", "Standard Class")
-

| Comment Construire ce Modèle dans Power BI (avec Power Query) :

1. **Charger le Dataset original** dans Power BI.
2. **Dupliquer la requête originale** plusieurs fois pour créer les dimensions :
 - Pour **Dim_Product** : Sélectionner **Product ID** , **Product Name** , **Sub-Category** , **Category** . Supprimer les doublons sur **Product ID** . Ajouter une colonne d'index ou une clé générée si **Product ID** n'est pas unique.
 - Pour **Dim_Customer** : Sélectionner **Customer ID** , **Customer Name** , **Segment** . Supprimer les doublons sur **Customer ID** .
 - Pour **Dim_Geography** : Sélectionner **Country** , **Region** , **State** , **City** . Supprimer les doublons sur la combinaison de ces colonnes.
 - Pour **Dim_ShipMode** : Sélectionner **Ship Mode** . Supprimer les doublons.
3. **Créer la Dim_Date** : C'est une table souvent générée soit :
 - À partir des dates uniques de **Order Date** et **Ship Date** de ta table de faits.
 - En créant une table de dates complètes à partir de la date minimale et maximale de ton dataset.
 - Power BI dispose de fonctions DAX pour créer une table de dates ou tu peux le faire via Power Query.
4. **Préparer la Fact_Sales** :
 - Conserver **Row ID** , **Order ID** , **Sales** , **Quantity** , **Profit** .
 - **Ajouter les clés étrangères** : Utiliser des fusions (Merge Queries) dans Power Query pour ramener les **Product_Key** , **Customer_Key** , **Geography_Key** , **ShipMode_Key** à partir de leurs dimensions

respectives en utilisant les attributs correspondants (**Product ID** , **Customer ID** , **Country/State/City** , **Ship Mode**).

- **Convertir les dates en clés de date** : Pour **Order Date** et **Ship Date** , tu devras soit les convertir en format entier (AAAAJJMM) pour qu'elles correspondent à la **Date_Key** de **Dim_Date** , soit les conserver comme des colonnes de type **Date** et laisser Power BI gérer la relation sur le champ **Full_Date** de la dimension Date. La première méthode est souvent plus performante et plus robuste.

5. **Gérer les relations** : Une fois les tables chargées dans le modèle Power BI, tu devras établir les relations entre la table de faits et chaque table de dimension (une relation *un à plusieurs* de la dimension vers le fait, filtrage *simple* dans le sens de la dimension vers le fait).

Avantages de cette approche :

- **Performances** : Power BI est optimisé pour les modèles en étoile.
 - **Simplicité des Mesures DAX** : Les calculs de KPIs seront plus simples et plus lisibles.
 - **Flexibilité** : Facile d'ajouter de nouveaux attributs aux dimensions sans modifier la table de faits.
 - **Compréhension** : Le modèle est intuitif à comprendre pour les utilisateurs.
-