

Recette technique de la réalisation d'un Data Warehouse (entrepôt de données) de bout en bout avec SSIS.

La société Orion

Cette société, présente au niveau mondial, est spécialisée dans la commercialisation d'articles de sport et vêtements. Les données disponibles regroupent des informations :

- les employés
- les produits
- les clients
- les commandes
- les fournisseurs

Le siège social aux États-Unis, gère des filiales en Belgique (depuis 1999), Pays Bas, Allemagne, Royaume-Uni, Danemark, France, Italie, Espagne et Australie. Les produits sont vendus en magasin, par catalogue et par internet. Une carte de fidélité : 'Orion Star Club', propose beaucoup d'avantages. L'historique d'information va du 1er janvier 1998 au 31 décembre 2002.

Structure de l'organisation

Le siège social héberge la majeure partie des fonctions administratives, soit un nombre important d'employés, entre 600 et 800. Le siège social centralise aussi la gestion des stocks, la vente par catalogue, la vente par internet et l'import - export. Néanmoins, certains employés gèrent aussi ces fonctions depuis les différentes filiales.

Les employés sont enregistrés dans la base de données selon cinq niveaux :

- Pays
- Compagnie
- Département
- Section
- Groupe

Les informations complémentaires sur les employés sont notamment :

- Date d'entrée et de départ de l'employé
- Date de début et de fin de contrat (pour certain contrat)
- Adresse
- Sexe
- Salaire
- Responsable hiérarchique

L'offre

La société propose environ 5 500 références. Certaines ne sont pas commercialisées dans tous les pays, tandis que d'autres, en raison des volumes écoulés, reflètent des particularités régionales ou des sports typiquement nationaux. L'ensemble des produits est organisé selon quatre niveaux.

- Ligne de produit
- Catégorie de produit
- Groupe de produit

- Produit

Chaque produit a un coût et un prix de vente. Le système informatique gère tous les prix en dollars. En utilisant les dates de début et de fin, ces prix varient en fonction du temps. Cet historique est sauvegardé. Le système gère aussi les remises pour certains produits, à certaines périodes. Les prix sont généralement uniques de dans le monde.

Les clients

Les clients sont repartis à travers le monde, notamment dans les pays où se trouvent des filiales, mais pas uniquement. La base de données enregistre environ 89 000 clients, pas tous actifs. L'adresse des clients comprend tout ou partie des informations suivantes :

- Rue
- Code postal
- Ville
- Région / département / conté
- Etat
- Pays
- Continent

Les clients sont classés dans des groupes en fonction de leur activité d'achat.

Les commandes

Chaque commande pointe vers le commercial qui a enregistré la vente. Environ 951 000 commandes sont enregistrées, commandes qui reflètent notamment les saisonnalités. Chaque commande comprend une ou plusieurs lignes, une ligne par produit.

Les fournisseurs

Chaque produit provient d'un fournisseur qui est basé dans un pays, mais toutes les commandes sont passées par le siège social. Il y a environ 64 fournisseurs, mais un seul fournisseur par produit.

Mise en place d'un système décisionnel

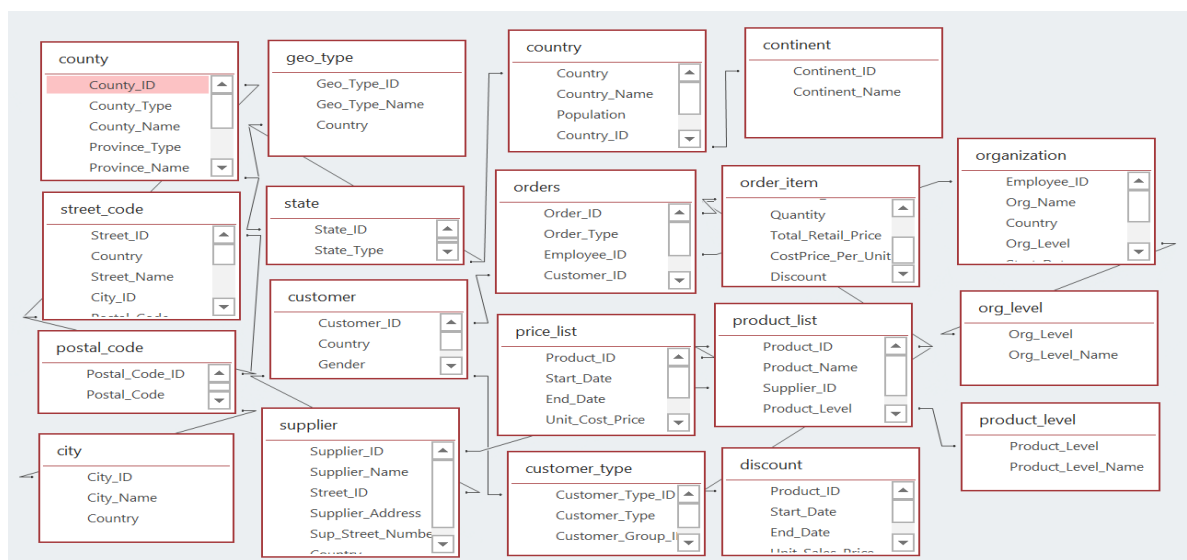
La société Orion souhaite améliorer sa performance à l'aide d'un système décisionnel. Voici quelques questions qui ont été recensées et auxquelles devrait répondre le système mis en place :

- ✓ Quels sont les 5 produits les plus vendus ?
- ✓ Quel est le chiffre d'affaires généré par produit ?
- ✓ Quels sont les 5 meilleurs fournisseurs ?
- ✓ Quel est le chiffre d'affaires généré par continent ?
- ✓ Quelle est la marge générée par année et par mois ?
- ✓ Quel est le chiffre d'affaires généré par année et par mois ?
- ✓ Quel est le coût total par année et par mois ?
- ✓ Quels sont les commerciaux réalisant le plus de ventes ?
- ✓ Quelles sont les caractéristiques des commerciaux (pays, sexe, âge, salaire, chiffre d'affaires) ?
- ✓ Y a-t-il une différence significative entre la moyenne du chiffre d'affaires généré par les commerciaux de sexe féminin et ceux de sexe masculin ?
- ✓ Quels sont les 5 meilleurs clients ?
- ✓ Déterminer l'âge des clients ainsi que leur tranche d'âge afin de leur proposer une offre adaptée :
 - Moins de 30 ans : "Tranche "moins de 30 ans"
 - De 30 à 45 ans : "Tranche "30-45 ans"
 - De 46 à 60 ans : "Tranche "46-60 ans"
 - De 61 à 75 ans : "Tranche "61-75 ans"
 - Plus de 75 ans : "Tranche "plus de 75 ans"
- ✓ Quel est le chiffre d'affaires généré par groupe de clients (par tranche d'âge) ?

Il faut donc construire un entrepôt de données capable de répondre aux besoins de requêtage, de reporting et d'analyses avancées, et le mettre en œuvre efficacement.

Sources de données

Voici le schéma relationnel de la base de données opérationnelle de l'entreprise, à partir de laquelle seront extraites les données de l'entrepôt :



Employee_ID	Start_Date	End_Date	Job_Title	Salary	Gender	Birth_Date	Emp_Hire_Date	Emp_Term_Date	Manager_ID
120101	01/07/1999	31/12/9999	Director	163040	M	18/08/1972	01/07/1999		120261
120102	01/06/1985	31/12/9999	Sales Manager	108255	M	11/08/1965	01/06/1985		120101
120103	01/01/1970	31/12/9999	Sales Manager	87975	M	22/01/1945	01/01/1970		120101
120104	01/01/1977	31/12/9999	Administration IV	46230	F	11/05/1950	01/01/1977		120101
120105	01/05/1995	31/12/9999	Secretary I	27110	F	21/12/1970	01/05/1995		120101
120106	01/01/1970	31/12/9999	Office Assistant	26960	M	23/12/1940	01/01/1970		120104
120107	01/02/1970	31/12/9999	Office Assistant	30475	F	21/01/1945	01/02/1970		120104
120108	01/08/2002	31/12/9999	Warehouse Assi	27660	F	23/02/1980	01/08/2002		120104
120109	01/10/2002	31/12/9999	Warehouse Assi	26495	F	15/12/1982	01/10/2002		120104
120110	01/11/1975	31/12/9999	Warehouse Assi	28615	M	20/11/1945	01/11/1975		120104
120111	01/11/1970	31/12/9999	Security Guard I	26895	M	23/07/1945	01/11/1970		120104
120112	01/07/1986	31/12/9999	Security Guard I	26550	F	17/02/1965	01/07/1986		120104
120113	01/01/1970	31/12/9999	Security Guard I	26870	F	10/05/1940	01/01/1970		120104

Ces tables sont stockées dans la base de données Microsoft Access nommée orion.mdb, hormis la table Staff stockée dans le fichier Microsoft Excel nommé staff.xls.

Création du datamart

Pour la suite, un datamart sera construit avec uniquement les clients membres du club Orion Gold et ayant acheté des vêtements ou des chaussures (pour accélérer les requêtes).

Planification des Packages

Les données sources évoluent en permanence. Les packages doivent donc être planifiés régulièrement. Le lancement des packages pourra se faire par exemple toutes les nuits pour prendre en compte les données modifiées pendant la journée.

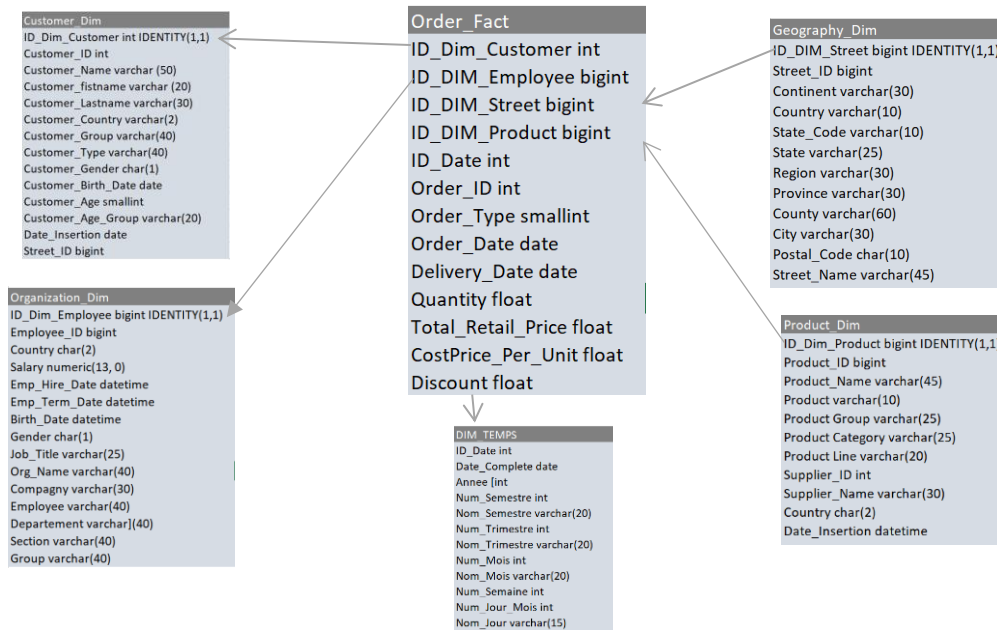
Solutions

Modélisation multidimensionnelle : Il existe trois formes de modèles multidimensionnels :

- ✓ Le modèle en étoile
- ✓ Le modèle en flocon de neige
- ✓ Le modèle en constellation

Schéma de l'entrepôt

Voici le schéma en étoile de l'entrepôt de données :



Une fois le schéma en étoile validé, il convient de créer l'entrepôt de données sous SQL Server.

Création des tables des dimensions :

Table de dimension : CUSTOMER_DIM

```

CREATE TABLE [dbo].[Customer_Dimension](
    [ID_Dim_Customer] [int] IDENTITY(1,1) NOT NULL,
    [Customer_ID] [int] NULL,
    [Customer_Name] [varchar](50) NULL,
    [Customer_firstname] [varchar](20) NULL,
    [Customer_lastname] [varchar](30) NULL,
    [Customer_Country] [char](2) NULL,
    [Customer_Group] [char](40) NULL,
    [Customer_Type] [char](40) NULL,
    [Customer_Gender] [char](1) NULL,
    [Customer_Birth_Date] [date] NULL,
    [Customer_Age] [smallint] NULL,
    [Customer_Age_Group] [char](20) NULL,
    [Date_Insertion] [date] NULL,
    [Street_ID] [bigint] NULL,
    CONSTRAINT [ID_Dim_Customerkey] PRIMARY KEY ([ID_Dim_Customer])
);
  
```

Table de dimension : GEOGRAPHY_DIM

```

CREATE TABLE [dbo].[Geography_Dimension](
    [ID_DIM_Street] [bigint] IDENTITY(1,1) NOT NULL,
    [Street_ID] [bigint] NULL,
    [Continent] [varchar](30) NULL,
    [Country] [varchar](10) NULL,
    [State_Code] [varchar](10) NULL,
    [State] [varchar](25) NULL,
    [Region] [varchar](30) NULL,
    [Province] [varchar](30) NULL,
    [County] [varchar](60) NULL,
    [City] [varchar](30) NULL,
    [Postal_code] [char](10) NULL,
    [Street_Name] [varchar](45) NULL,
    CONSTRAINT [ID_DIM_StreetPK] PRIMARY KEY ([ID_DIM_Street])
);
  
```

Table de dimension : ORGANIZATION_DIM

```

CREATE TABLE [dbo].[Organization_Dimension](
    [ID_DIM_Employee] [int] IDENTITY(1,1) NOT NULL,
    [Employee_ID] [bigint] NULL,
    [Employee_Country] [char](2) NULL,
    [Company] [char](30) NULL,
    [Department] [varchar](40) NULL,
    [Section] [varchar](40) NULL,
    [Org_Group] [varchar](40) NULL,
    [Job_Title] [varchar](25) NULL,
    [Employee_Name] [varchar](40) NULL,
    [Employee_Gender] [char](1) NULL,
    [Salary] [decimal](13, 0) NULL,
    [Employee_Birth_Date] [date] NULL,
    [Employee_Hire_Date] [date] NULL,
    [Employee_Term_Date] [date] NULL,
    CONSTRAINT [ID_DIM_EmployeePK] PRIMARY KEY ([ID_DIM_Employee])
);
  
```

Table de dimension : **PRODUCT_DIM**

```
CREATE TABLE [Product_Dimension] (
    [ID_Dim_Product] bigint,
    [Product_ID] bigint,
    [Product_Name] varchar(45),
    [Supplier_ID] int,
    [Product] varchar(10),
    [Product Group] varchar(25),
    [Product Category] varchar(25),
    [Product Line] varchar(20),
    [Date Insertion] datetime,
    [Supplier_Name] varchar(30),
    [Country] varchar(2),
    CONSTRAINT [ID_Dim_ProductPK] PRIMARY KEY ([ID_Dim_Product])
);
```

Table de dimension : **TIME_DIM**

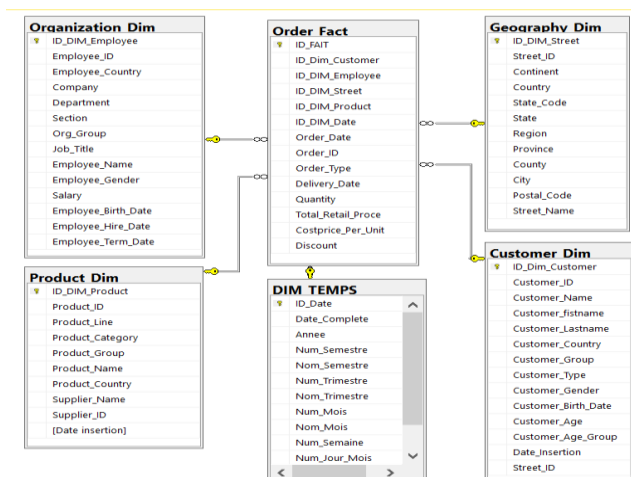
```
CREATE TABLE [dbo].[DIMENSION_TEMPS](
    [ID_Date] [int] NOT NULL,
    [Date_Complete] [date] NULL,
    [Annee] [int] NULL,
    [Num_Semestre] [int] NULL,
    [Nom_Semestre] [varchar](20) NULL,
    [Num_Trimestre] [int] NULL,
    [Nom_Trimestre] [varchar](20) NULL,
    [Num_Mois] [int] NULL,
    [Nom_Mois] [varchar](20) NULL,
    [Num_Semaine] [int] NULL,
    [Num_Jour_Mois] [int] NULL,
    [Nom_Jour] [varchar](15) NULL,
    CONSTRAINT [ID_DatePK] PRIMARY KEY ([ID_Date])
);
```

Table de faits : **FACT_ORDER**

```
CREATE TABLE [dbo].[Fact_Order](
    [ID_FAIT] [bigint] IDENTITY(1,1) PRIMARY KEY ,
    [ID_Dim_Customer] [int] NULL,
    [ID_DIM_Employee] [int] NULL,
    [ID_DIM_Street] [bigint] NULL,
    [ID_DIM_Product] [bigint] NULL,
    [ID_DIM_Date] [int] NULL,
    [Order_Date] [date] NULL,
    [Order_ID] [int] NULL,
    [Order_Type] [smallint] NULL,
    [Delivery_Date] [date] NULL,
    [Quantity] [smallint] NULL,
    [Total_Retail_Proce] [float] NULL,
    [Costprice_Per_Unit] [float] NULL,
    [Discount] [float] NULL,
)
```

Modèle de données (étoile) :

Voici le modèle en étoile final de l'entrepôt de données sous SQL Server.



Maintenant que les tables de l'entrepôt sont créées, il faut réaliser les processus qui vont remplir ces tables à partir des données sources.

Mise en place de solution SSIS :

Création de projet :

- ✓ Ouvrir Visual studio.
- ✓ Créer un nouveau projet nommé orion_project

La fenêtre de SSIS est composée des onglets suivants :

- ✓ Barres d'outils et menus (en haut)
- ✓ Onglet Flux de contrôle : Dédié aux contrôles de données
- ✓ Onglet Flux de données : Dédié aux données
- ✓ Onglet paramètres : Les paramètres d'Intégrations Services (SSIS) permettent d'affecter des valeurs aux propriétés des packages au moment de l'exécution du package.
- ✓ Onglet Gestionnaires d'événements : Gère des événements du composant de package
- ✓ Onglet Explorateur de package : décrit tous les éléments qui composent les packages.
- ✓ Onglet Progression : L'onglet **Progression** indique la progression de l'exécution du package
- ✓ Boîte à outils SSIS : Tous les composants installés sur l'ordinateur local apparaissent automatiquement dans la **Boîte à outils SSIS**
- ✓ Zone Gestionnaires de connexion : Les packages utilisent des gestionnaires de connexion pour se connecter aux sources de données telles que les fichiers, les bases de données relationnelles et les serveurs.
- ✓ Explorateur de solutions : Qui contient le nom de package

Spécification des données sources :

Placer les différents fichiers Excel et Access dans un répertoire C:/orion.

- ✓ Établir une connexion orion_BD à la base Access orion.mdb :
- ✓ Établir une connexion fichier Excel(staff.xls)

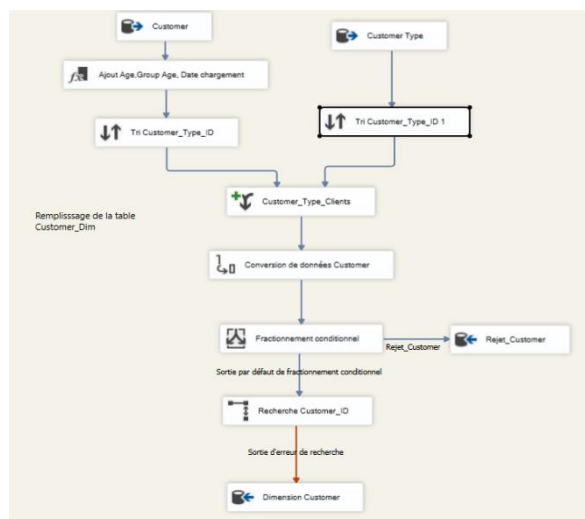
Spécification des données cibles :

Établir une connexion orion_DW à l'entrepôt de données :

Les connexions sources et destination sont maintenant disponibles dans la zone Gestionnaires de connexion. Il faut alors construire les différents flux pour remplir les tables de l'entrepôt.

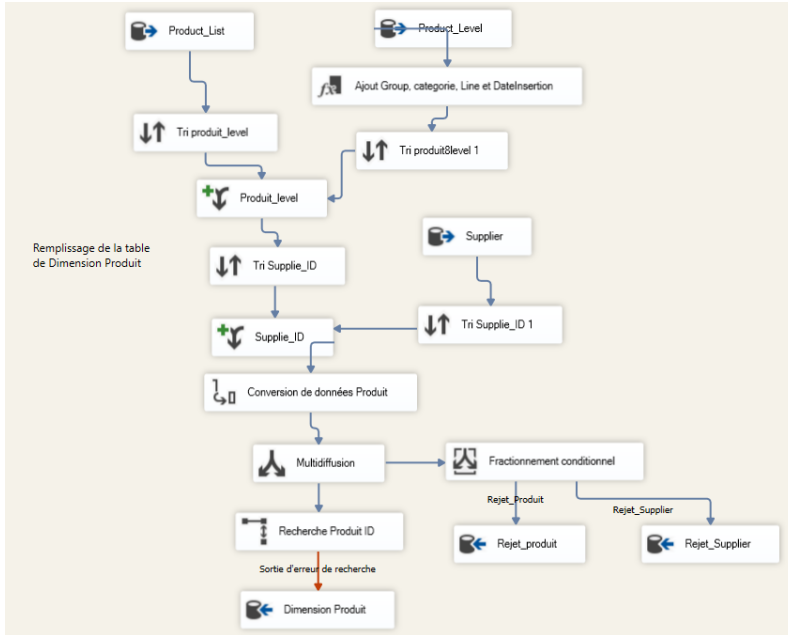
Remplissage de la table **Customer Dim** :

- ✓ Pour chaque colonne de la table Dimension Customer, on spécifie de quelle(s) donnée(s) source elle dépend.
- ✓ On crée la tâche Dimension Customer et l'ajoute tous les composants nécessaires.
- ✓ On lance la tâche et vérifie le résultat sous SQL Server.



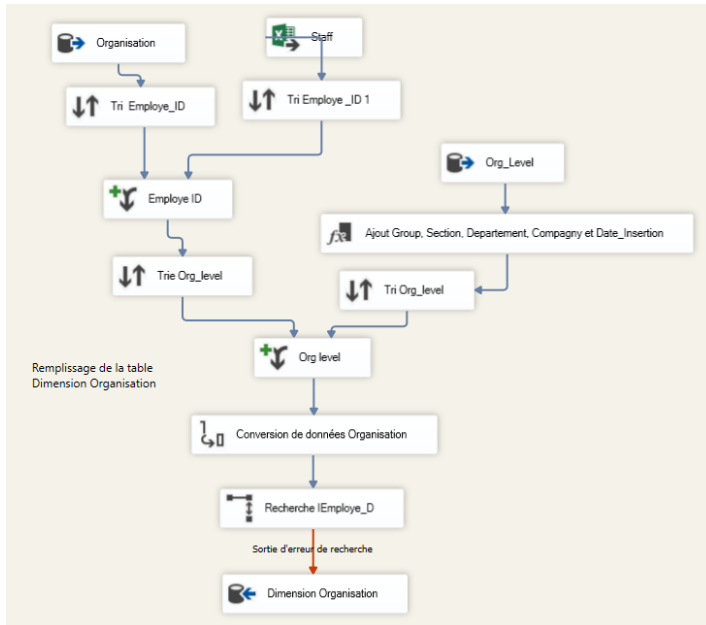
Remplissage de la table **Dim_Product** :

- Pour chaque colonne de la table Dim_Product, on spécifie de quelle(s) donnée(s) source elle dépend.
- On crée la tâche Dimension Product et ajoute tous les composants nécessaires.
- On lance la tâche et vérifie le résultat sous SQL Server.



Remplissage de la table **Dim Organization** :

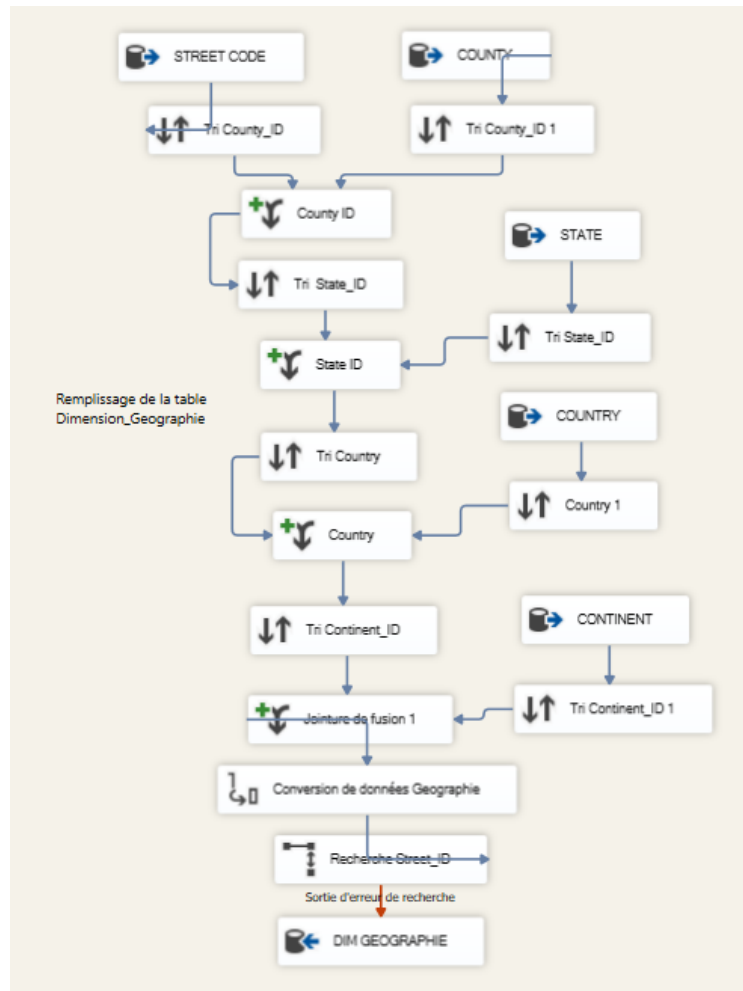
- Pour chaque colonne de la table Dim_Organization, on spécifie de quelle(s) donnée(s) source elle dépend.
- On crée la tâche Dimension Organization et l'on ajoute tous les composants nécessaires.
- On lance la tâche et vérifie le résultat sous SQL Server.



Remplissage de la table **Dim_Geography** :

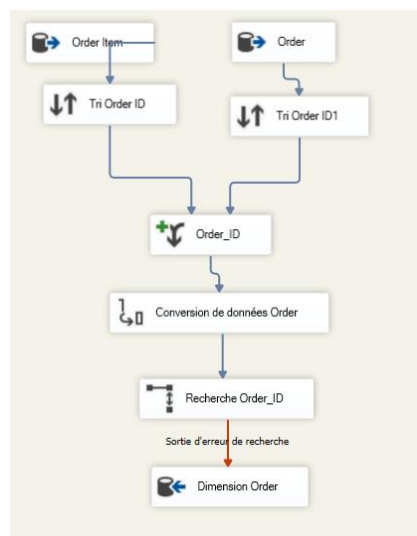
- Pour chaque colonne de la table Dim_Geography, on spécifie de quelle(s) donnée(s) source elle dépend.

- On crée la tâche Dimension Geography et l'on ajoute tous les composants nécessaires.
- On lance la tâche et vérifie le résultat sous SQL Server.

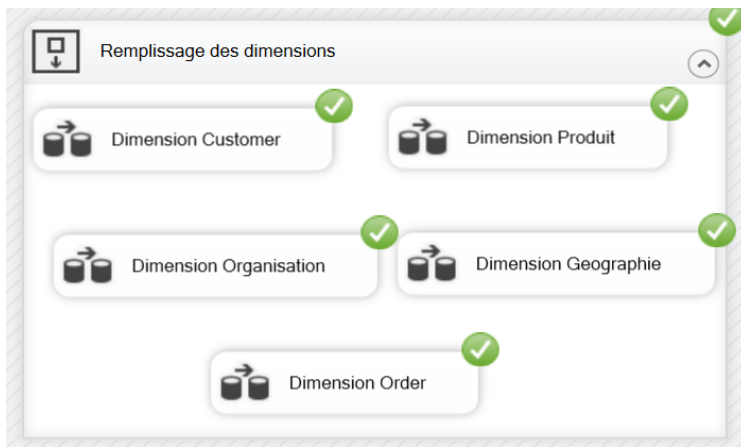


Remplissage de la table Dim Order :

- Pour chaque colonne de la table Dim_Order, on spécifie de quelle(s) donnée(s) source elle dépend.
- On crée la tâche Dimension Order et l'on ajoute tous les composants nécessaires.
- On lance la tâche et vérifie le résultat sous SQL Server.



Regroupement des dimensions dans un conteneur de séquence :

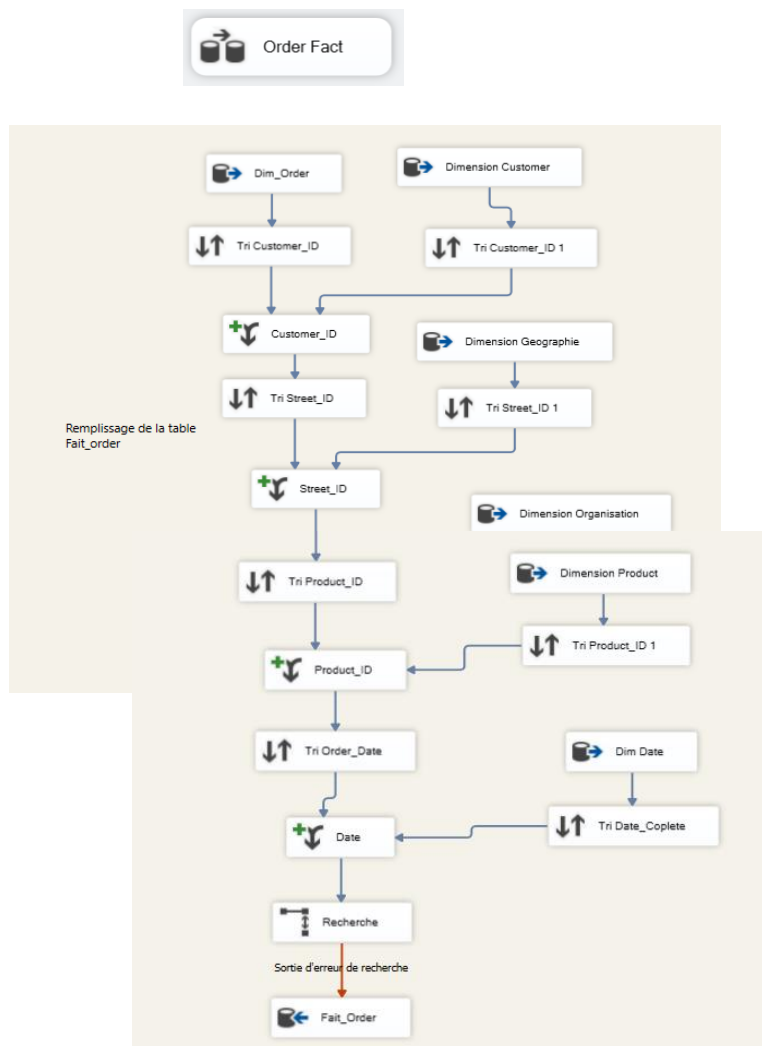


Chargement de la table de **faits** :

Un package nommé `Chargement de la table de faits.dtsx` qui fera le chargement de la table de faits en rassemblant les mesures existantes et en ajoutant les dimensions une à une pour avoir les clés des substitutions ou clés primaires des tables des dimensions.

Ce package contiendra un flux de données comme ci-dessous :

- Pour chaque colonne de la table `Order_fact`, on spécifie de quelle(s) donnée(s) source elle dépend.
- On crée la tâche « Order Fact » et l'on ajoute tous les composants nécessaires.
- On lance la tâche et vérifie le résultat sous SQL Server.



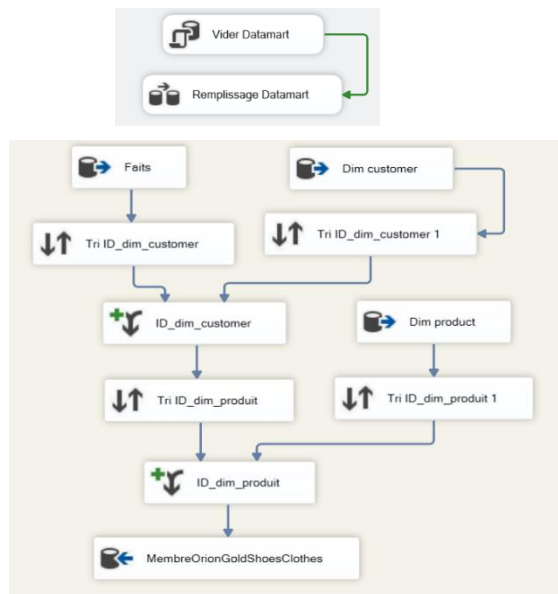
Création du datamart :

Pour la suite, un datamart sera construit avec uniquement les clients membres du club Orion Gold et ayant acheté des vêtements ou des chaussures (pour accélérer les requêtes).

Travail à réaliser :

Base de données indépendante + Package de remplissage

- ✓ On crée la BD puis le package permettant de remplir le datamart en même temps le script de création de la table.
- ✓ Conditions à spécifier:
 - ✓ WHERE Customer_Group = 'Orion Club Gold members';
 - ✓ WHERE Product_Name LIKE '%Clothes%' OR Product_Name LIKE '%Shoes%';
- ✓ On sélectionne l'option Inner join dans les jointures.
- ✓ Vérifier que tout a bien fonctionné.



Ordonnancement des packages :

Un dernier package est nécessaire pour appeler l'ensemble des packages créés précédemment. Il sera nommé **Chargement Initial.dtsx** en conséquence. Lors de son exécution, ce package devra créer et charger les dimensions, la table de faits ainsi que le Datamart. Ci-dessous, un aperçu de l'ordonnancement des packages



Les tâches ont été exécuté avec succès sous SSIS

Planification du package Chargement Initial :

Travail à réaliser :

Il existe deux solutions pour planifier l'exécution des packages : soit utiliser le Planificateur de tâches de Windows, soit utiliser SQL Server Agent."

Planificateur de tâches de Windows :

- ✓ Ouvrir le planificateur de tâches de Windows

- ✓ Accessoires / Outils système / Planificateur de tâches
- ✓ On clique sur Créer une tâche... et l'on donne un nom : orion_jobsETL
- ✓ On crée un déclencheur (dans 15 min par exemple)
- ✓ On crée une action :
 - ✓ Action: Démarrer un programme
- ✓ Programme:
 - C:\BI\Orion DW\Orion DW\Chargement Initial.dtsx Run_Package.bat
- ✓ On vérifie que tout a bien fonctionné.

Conclusion :

Ce mini-projet a permis de réaliser un projet SSIS de bout en bout. Le Data Warehouse construit, contenant les dimensions et la table de faits, peut être utilisé pour le reporting, l'analyse de données et la création de tableaux de bord.

Mise en œuvre d'un système BI : réponses aux questions stratégiques d'Orion :

- ✓ Quels sont les 5 produits les plus vendus ?

```
SELECT TOP 5
    P.Product_Name,
    SUM(F.Quantity) AS Total_Quantity_Sold
FROM
    Order_Fact F
JOIN
    Product_Dim P ON F.ID_DIM_Product = P.ID_DIM_Product
GROUP BY
    P.Product_Name
ORDER BY
    Total_Quantity_Sold DESC;
```

	Product_Name	Total_Quantity_Sold
1	Petanque Balls Chromium 8-pack	8436
2	Hurricane 4	6638
3	Lucky Tech Intergal Wp/B Rain Pants	6514
4	Comfort Shelter	6273
5	Family Holiday 4	5951

- ✓ Quel est le chiffre d'affaires généré par produit ?

```
SELECT
    P.Product_Name,
    SUM(F.Quantity * F.Total_Retail_Price) AS Chiffre_Affaires
FROM
    Order_Fact F
JOIN
    Product_Dim P ON F.ID_DIM_Product = P.ID_DIM_Product
GROUP BY
    P.Product_Name
ORDER BY
    Chiffre_Affaires DESC;
```

	Product_Name	Chiffre_Affaires
1	Family Holiday 4	6746166,5
2	Hurricane 4	3996665,9
3	Family Holiday 6	3771664,8
4	Expedition Dome 3	3165120
5	Comfort Shelter	2956140
6	Lucky Tech Intergal Wp/B Rain Pants	1800708
7	Big Guy Men's Air Tuned Sirocco Shoes	1777301,4
8	Basic 10, Left, Yellow/Black	1437275,4
9	Weston F4 Men's Hockey Skates	1409516
10	Jl Legacy Curig I.A.Jacket	1193008
11	Perfect Fit Men's Stunt Skates	1180874,66
12	G.A.M. Bongo	1173705,8

✓ Quels sont les 5 meilleurs fournisseurs ?

```
SELECT TOP 5
    P.Supplier_Name,
    SUM(F.Quantity * F.Total_Retail_Price) AS Chiffre_Affaires
FROM
    Order_Fact F
JOIN
    Product_Dim P ON F.ID_DIM_Product = P.ID_DIM_Product
GROUP BY
    P.Supplier_Name
ORDER BY
    Chiffre_Affaires DESC;
```

	Supplier_Name	Chiffre_Affaires
1	Eclipse Inc	39841740,22
2	Magnifico Sports	20418297,02
3	3Top Sports	18767523,65
4	Twain Inc	17844094,085
5	Petterson AB	17137695,7

✓ Quel est le chiffre d'affaires généré par continent

```
SELECT
    G.Continent,
    SUM(F.Quantity * F.Total_Retail_Price) AS Chiffre_Affaires
FROM
    Order_Fact F
JOIN
    Geography_Dim G ON F.ID_DIM_Street = G.ID_DIM_Street
GROUP BY
    G.Continent
ORDER BY
    Chiffre_Affaires DESC;
```

	Continent	Chiffre_Affaires
1	Europe	199612466,015
2	North America	74495401,9650001
3	Australia/Pacific	16494810,595
4	Asia	520843,48
5	Africa	329841,09

✓ Quelle est la marge générée par année et mois ?

Remarque : Comme je n'ai pas la colonne Num_Année dans mon modèle, je l'ai remplacée par la colonne Année.

```
SELECT
    T.Année,
    T.Num_Mois,
    SUM((F.Total_Retail_Price - F.CostPrice_Per_Unit) * F.Quantity) AS Marge_Totale
FROM
    Order_Fact F
JOIN
    DIM_TEMPS T ON F.ID_Date = T.ID_Date
GROUP BY
    T.Année,
    T.Num_Mois
ORDER BY
    T.Année,
    T.Num_Mois;
```

	Année	Num_Mois	Marge_Totale
1	1998	1	2434766,875
2	1998	2	2125332,625
3	1998	3	1027835,825
4	1998	4	2693043,63
5	1998	5	3411832,55
6	1998	6	5423078,935
7	1998	7	5427793,92
8	1998	8	5066598,59
9	1998	9	1136562,895

✓ Quel est le chiffre d'affaires généré par année et par mois ?

```
SELECT
    T.Annee,
    T.Num_Mois,
    SUM(F.Quantity * F.Total_Retail_Price) AS Chiffre_Affaires
FROM
    Order_Fact F
JOIN
    DIM_TEMPS T ON F.ID_Date = T.ID_Date
GROUP BY
    T.Annee,
    T.Num_Mois
ORDER BY
    T.Annee,
    T.Num_Mois;
```

	Annee	Num_Mois	Chiffre_Affaires
1	1998	1	3167822,565
2	1998	2	2713256,375
3	1998	3	1561949,815
4	1998	4	3503420,95
5	1998	5	4359079,29
6	1998	6	6627249,225
7	1998	7	6646459,76
8	1998	8	6271423,85
9	1998	9	1639499,455

✓ Quel est le coût total par année et par mois ?

```
SELECT
    T.Annee,
    T.Num_Mois,
    SUM(F.Quantity * F.CostPrice_Per_Unit) AS Cout_Total
FROM
    Order_Fact F
JOIN
    DIM_TEMPS T ON F.ID_Date = T.ID_Date
GROUP BY
    T.Annee,
    T.Num_Mois
ORDER BY
    T.Annee,
    T.Num_Mois;
```

	Annee	Num_Mois	Cout_Total
1	1998	1	733055,69
2	1998	2	587923,75
3	1998	3	534113,99
4	1998	4	810377,32
5	1998	5	947246,74
6	1998	6	1204170,29
7	1998	7	1218665,84
8	1998	8	1204825,26
9	1998	9	502936,56

✓ Quels sont les commerciaux réalisant le plus de ventes ?

```
SELECT
    E.Org_Name AS Commercial,
    SUM(F.Quantity * F.Total_Retail_Price) AS Chiffre_Affaires
FROM
    Order_Fact F
JOIN
    Organization_Dim E ON F.ID_DIM_Employee = E.ID_DIM_Employee
GROUP BY
    E.Org_Name
ORDER BY
    Chiffre_Affaires DESC;
```

	Commercial	Chiffre_Affaires
1	Agnes de Fournou	1584135,48
2	Christelle Bourrier	1566861,545
3	Jaime Wetherington	1543943,605
4	Joseph Robbin-Coker	1510638,175
5	Donald Washington	1505163,335
6	Signid Kagarise	1478357,67
7	Jan-Reijer Rutten	1474864,92
8	Ray Abbott	1473737,725
9	Brienne Darrohn	1462788,5

✓ Quelles sont les caractéristiques des commerciaux (pays, sexe, âge, salaire, chiffre d'affaires) ?

```
SELECT
    E.Org_Name AS Commercial,
    G.Country AS Pays,
    E.Gender AS Sexe,
    DATEDIFF(YEAR, E.Birth_Date, GETDATE()) AS Age,
    E.Salary AS Salaire,
    SUM(F.Quantity * F.Total_Retail_Price) AS Chiffre_Affaires
FROM
    Organization_Dim E
JOIN
    Geography_Dim G ON E.ID_Geography = G.ID_Geography
LEFT JOIN
    Order_Fact F ON E.ID_DIM_Employee = F.ID_DIM_Employee
GROUP BY
    E.Org_Name, G.Country, E.Gender, E.Birth_Date, E.Salary
ORDER BY
    Chiffre_Affaires DESC;
```

✓ Y a-t-il une différence significative entre la moyenne du chiffre d'affaires généré par les commerciaux de sexe féminin et ceux de sexe masculin ?

```
WITH CA_Par_Commercial AS (
    SELECT
        E.ID_DIM_Employee,
        E.Gender,
        SUM(F.Quantity * F.Total_Retail_Price) AS Chiffre_Affaires
    FROM
        Order_Fact F
    JOIN
        Organization_Dim E ON F.ID_DIM_Employee = E.ID_DIM_Employee
    GROUP BY
        E.ID_DIM_Employee, E.Gender
)

SELECT
    Gender,
    COUNT(*) AS Nombre_Commerciaux,
    AVG(Chiffre_Affaires) AS Moyenne_CA_Par_Commercial
FROM
    CA_Par_Commercial
GROUP BY
    Gender;
```

	Gender	Nombre_Commerciaux	Moyenne_CA_Par_Commercial
1	F	291	333779,587113402
2	M	356	316176,542064607

✓ Quels sont les 5 meilleurs clients ?

```
SELECT TOP 5
    C.[Customer_Name] AS Client,
    SUM(F.Quantity * F.Total_Retail_Price) AS Chiffre_Affaires
FROM
    Order_Fact F
JOIN
    Customer_Dim C ON F.ID_DIM_Customer = C.ID_DIM_Customer
GROUP BY
    C.[Customer_Name]
ORDER BY
    Chiffre_Affaires DESC;
```

	Client	Chiffre_Affaires
1	Carlos Martínez Sanmamed	65652,2
2	Kala Leitch	63458,7
3	Joaquin Torrico	48199,6
4	Ralf Bondy	47848,1
5	Francesca Romana Sartori	44223,24

- ✓ Déterminer l'âge des clients ainsi que leur tranche d'âge afin de leur proposer une offre adaptée :
- Moins de 30 ans : "Tranche "moins de 30 ans"
 - De 30 à 45 ans : "Tranche "30-45 ans"
 - De 46 à 60 ans : "Tranche "46-60 ans"
 - De 61 à 75 ans : "Tranche "61-75 ans"
 - Plus de 75 ans : "Tranche "plus de 75 ans"

```
SELECT
  C.Customer_Name AS Client,
  DATEDIFF(YEAR, C.Customer_Birth_Date, GETDATE()) AS Age,
  CASE
    WHEN DATEDIFF(YEAR, C.Customer_Birth_Date, GETDATE()) < 30 THEN 'Tranche moins de 30 ans'
    WHEN DATEDIFF(YEAR, C.Customer_Birth_Date, GETDATE()) BETWEEN 30 AND 45 THEN 'Tranche 30-45 ans'
    WHEN DATEDIFF(YEAR, C.Customer_Birth_Date, GETDATE()) BETWEEN 46 AND 60 THEN 'Tranche 46-60 ans'
    WHEN DATEDIFF(YEAR, C.Customer_Birth_Date, GETDATE()) BETWEEN 61 AND 75 THEN 'Tranche 61-75 ans'
    ELSE 'Tranche plus de 75 ans'
  END AS Tranche_Age
FROM
  Customer_Dim C;
```

	Client	Age	Tranche_Age
1	Tapani Pukkinen	70	Tranche 61-75 ans
2	Pierson Delagula	65	Tranche 61-75 ans
3	Veronique Van de Meent	95	Tranche plus de 75 ans
4	Margreet Bastiampillia	60	Tranche 46-60 ans
5	Philip Rierson	75	Tranche 61-75 ans
6	Mercedes Iñiesta Micó	41	Tranche 30-45 ans
7	T.I. Heuvelman	70	Tranche 61-75 ans
8	Doretha Haakinson	65	Tranche 61-75 ans
9	Ditte Strandgård Jensen	50	Tranche 46-60 ans
10	Phong Rehr	41	Tranche 30-45 ans
11	Laurent Darchy	43	Tranche 30-45 ans
12	Anupam Umbarger	95	Tranche plus de 75 ans

- ✓ Quel est le chiffre d'affaires généré par groupe de clients (par tranche d'âge) ?

```
SELECT
  CASE
    WHEN DATEDIFF(YEAR, C.Customer_Birth_Date, GETDATE()) < 30 THEN 'Tranche moins de 30 ans'
    WHEN DATEDIFF(YEAR, C.Customer_Birth_Date, GETDATE()) BETWEEN 30 AND 45 THEN 'Tranche 30-45 ans'
    WHEN DATEDIFF(YEAR, C.Customer_Birth_Date, GETDATE()) BETWEEN 46 AND 60 THEN 'Tranche 46-60 ans'
    WHEN DATEDIFF(YEAR, C.Customer_Birth_Date, GETDATE()) BETWEEN 61 AND 75 THEN 'Tranche 61-75 ans'
    ELSE 'Tranche plus de 75 ans'
  END AS Tranche_Age,
  SUM(F.Quantity * F.Total_Retail_Price) AS Chiffre_Affaires
FROM
  Order_Fact F
JOIN
  Customer_Dim C ON F.ID_DIM_Customer = C.ID_DIM_Customer
GROUP BY
  CASE
    WHEN DATEDIFF(YEAR, C.Customer_Birth_Date, GETDATE()) < 30 THEN 'Tranche moins de 30 ans'
    WHEN DATEDIFF(YEAR, C.Customer_Birth_Date, GETDATE()) BETWEEN 30 AND 45 THEN 'Tranche 30-45 ans'
    WHEN DATEDIFF(YEAR, C.Customer_Birth_Date, GETDATE()) BETWEEN 46 AND 60 THEN 'Tranche 46-60 ans'
    WHEN DATEDIFF(YEAR, C.Customer_Birth_Date, GETDATE()) BETWEEN 61 AND 75 THEN 'Tranche 61-75 ans'
    ELSE 'Tranche plus de 75 ans'
  END
ORDER BY
  Chiffre_Affaires DESC;
```

	Tranche_Age	Chiffre_Affaires
1	Tranche 46-60 ans	77571113,5699999
2	Tranche 30-45 ans	73945811,09
3	Tranche 61-75 ans	72452672,67
4	Tranche plus de 75 ans	67483765,815