UNIVERSITÉ MOHAMMED V

Faculté des Sciences Rabat

Master Ingénierie de Données et Développement Logiciel

Rapport fin de module Business Intelligence

Professeur : Ziyati Houssaine

Réalisé par:

Fatima Bouya

Année universitaire 2020-2021

Introduction : **3**

Chapitre1 : Présentation de la Business Intelligence **4**

1.1 Context: 4

1.2 Qu'est-ce qu'un système décisionnel ? 5

1. 3 La Business Intelligence : 5

1. 4 Les étapes de processus de la BI : 6

2.5 Business Intelligence, Architecture du SID **11**

2.5.1 Architecture du Système d'Information Décisionnelle (SID) 11

2.5.2 Les 4 fonctions de la chaîne décisionnelle 11

1.6 Les outils utiliser au Business Intelligence **12**

1.6.1 Collecter : Les outils d'ETL (Extract Transform and Load) 12

1.6.2 Distribuer les informations : Portail décisionnel EIP Enterprise Information Portal 14

1.6.3 Exploiter : Tableau de bord, analyse OLAP , data mining,... 14

1.7 PROBLÉMATIQUE: **15**

1.8 CONCEPTION DE L'ENTREPÔT DE DONNÉES: **17**

Chapitre 2 : PRÉPARATION DE L'ENVIRONNEMENT **22**

2.1.Pentaho Data Integration PDI: 22

2.1.1 Étapes d'installation : 22

2.2 Oracle SQL Developer 24

2.3 Pentaho workbench 26

2.3.1 Étapes d'installation : 27

Chapitre 3 : MANIPULATION DES DONNÉES AVEC L'OUTIL PDI **27**

3.1 CRÉER UNE CONNEXION ORACLE DANS PDI **27**

3.2 EXTRAIRE LES DONNÉES EXCEL: **28**

3.3 SÉLECTIONNER LE FLUX DE DONNÉES **29**

3.4 TRIER LES CLÉS, SUPPRIMER LES DOUBLONS: **31**

3.5 création du champs month, year: **34**

3.6 APPLIQUER UNE FONCTION D'AGRÉGATION AUX MESURES: **36**

3.7 ORDONNANCER LE FLUX: **39**

3.8 AJOUTER LES DIMENSIONS: **40**

3.9 LANCER LA TRANSFORMATION: **44**

Chapitre 4 :REPRÉSENTATION MULTIDIMENSIONNELLE AVEC SCHEMA WORKBENCH **47**

4-1 CRÉER UNE CONNEXION JDBC DANS PDI **47**

4.2 OLAP CUBE **48**

4.2.1 DIMENSIONS 50

4.2.2 MESURES 51

Chapitre 5: VISUALISATION Power BI **52**
5.1 Définition sur le POWER BI : **52**
5.2 Les Points Forts de Power BI : **53**
5.3.Les étapes de visualisation **53**
5.3.1 Création de la connexion : 53
5.3.2 importer le schéma dans la base de données d'utilisateur target 54
Chapitre 6: Conclusion **57**

Introduction :

La business intelligence (BI) est un sujet en pleine évolution, s'adressant à la direction générale tout comme aux métiers. Outil d'aide à la décision, la BI permet d'avoir une vue d'ensemble des différentes activités de l'entreprise, et son environnement. Cette vue transversale nécessite de connaître les différents métiers de l'entreprise et implique certaines spécificités organisationnelles et managériales. L'organisation de la BI dans l'entreprise est fortement dépendante de l'organisation de l'entreprise elle-même. Cependant, la BI peut avoir un impact structurant pour l'entreprise, notamment par la formalisation de référentiels de données et par la mise en place de centres de compétences .la BI a aussi un rôle fédérateur entre DSI. Métiers, DG, et renforce le rôle du DSI par rapport à la performance globale de l'entreprise. La mise en place de projets BI ne peut se faire sans avoir défini préalablement une stratégie décisionnelle globale. Plus que pour des projets SI classiques, les projets BI nécessitent une grande maturité dans les relations DSI-Métiers, et s'inscrivent dans une démarche d'amélioration continue. De ce fait, les méthodologies utilisées se veulent agiles et itératives, pour coller au plus près de la demande client. Aujourd'hui ,le marché BI propose des solutions assez complètes concernant les aspects de reporting et de consolidation de données ,tant du domaine propriétaire que de l'open source .les évolutions possibles ,à court-moyen terme, porteraient sur les outils d'analyse proactive et de simulation, ainsi que sur l'interactivité et la convivialité des accès aux données ,et sur la combinaison de données structurées et non structurées issues des sources de données internes et externes .Mais le manque de visibilité des clients sur le marché BI (fusions-acquisitions , absence de roadmap ,méconnaissance des stratégies des éditeurs ,…)crée une grande incertitude quant à la pérennité des offres actuelles et nuit grandement à la définition d'une stratégies décisionnelle à moyen-long terme.

Chapitre1 : Présentation de la Business Intelligence

1.1 Context:

Aujourd'hui ,dans un contexte où les sources d'information sont éclatées ,volumineuses et complexes ,il y a un réel besoin de consolider et d'analyser ces dernières pour pouvoir avoir une vision globale et optimiser la patrimoine informationnel de l'entreprise .Or « trop d'information tue l'information » … l'objectif de la BI est de créer ,à partir des données de l'entreprise mais aussi externe à celle-ci ,l'information et le savoir aidant les membres de l'entreprise ,des cadres dirigeants aux opérationnels ,dans leur pilotage. De récentes études montrent que la BI est l'une des préoccupations principales au sein des DSI de grandes entreprises. En effets, dans le contexte actuel de crise et d'hyper-concurrence, la BI représente une opportunité pour les entreprises d'optimiser le pilotage de leurs activités, et d'anticiper sur les évolutions du marché, des comportements des clients/ consommateurs, … les domaines d'utilisation de la BI touchent la plupart du Métiers de l'entreprise :

- Finance, avec les reportings financiers et budgétaires par exemples ;
- Vente et Commercial, avec l'analyse des points de ventes, l'analyse de la profitabilité et de l'impact des promotions par exemples ;
- Marketing, avec la segmentation clients, les analyses comportementales par exemples ;
- Logistique, avec l'optimisation de la gestion de stock, le suivi des livraisons par exemples ;

- Ressources humaines, avec l'optimisation de l'allocation des ressources par exemples ;
- Etc...

1.2 Qu'est-ce qu'un système décisionnel ?

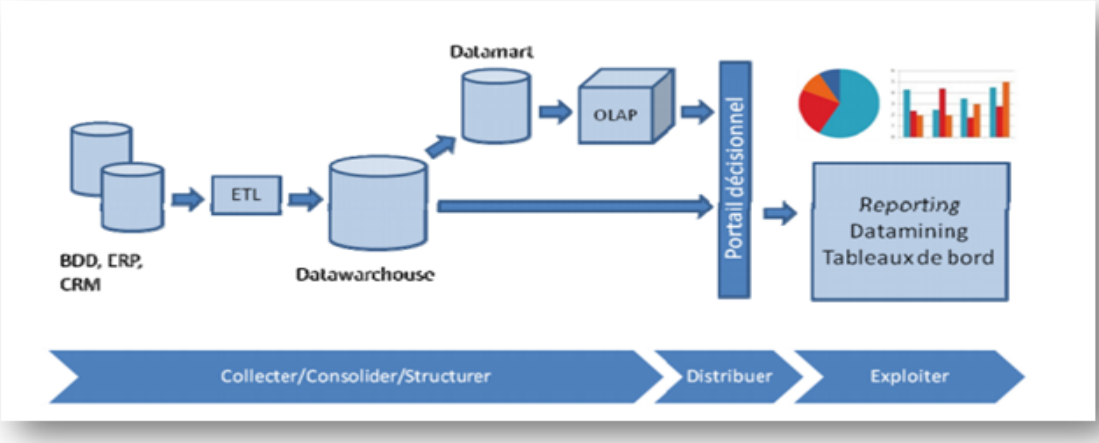
Ou bien plus généralement L'informatique décisionnelle est un ensemble de solutions informatiques qui permet une analyse claire des données de cette entreprise de construction.

1. 3 La Business Intelligence :

Le but principal du Business Intelligence est de dégager des informations qualitatives nouvelles, qui serviront de base aux décisions tactiques ou stratégiques d'une organisation. Conduire un projet de Business Intelligence, c'est adopter une nouvelle logique de l'informatisation de l'entreprise axée sur l'anticipation des besoins et des attentes des utilisateurs. L'architecture décisionnelle mise en place a ainsi pour mission de rendre accessible, de mettre en forme et de présenter les informations clés afin de faciliter la prise de décision.

Le processus BI peut se schématiser de la manière suivante :

1. 4 Les étapes de processus de la BI :



Un système d'information décisionnel assure quatre fonctions que sont la collecte, l'intégration, la diffusion et la restitution des données.

Le principe de fonctionnement d'un SID peut être résumé par les schéma suivants :

Diverses sources de données d'entreprise sont utilisées en entrée pour être mises en commun dans un datawarehouse. Ces données peuvent ensuite être réparties en divers datamarts, chacun physiquement représenté par une base de données multidimensionnelle. Ces données sont ensuite mises à disposition des utilisateurs et décideurs par divers moyens. Généralement elles sont mises en forme dans des tableaux de bords et rapports avant de leur être présentées via un serveur web.

- **La collecte :**

La première étape de collecte des données va permettre à terme de produire les indicateurs nécessaires au périmètre du SID. Pour cela il convient d'aller chercher les données où elles se trouvent. Les données applicatives métier sont stockées dans une ou plusieurs bases de données correspondant à chaque application utilisée.

La collecte est donc l'ensemble des tâches consistant à détecter, à sélectionner, à extraire et à filtrer les données brutes issues des environnements pertinents pour obtenir des indicateurs utiles dans le cadre de l'aide à la décision. Les sources de données internes et/ou externes étant souvent hétérogènes tant sur le plan technique que sur le plan sémantique, cette fonction est la plus délicate à mettre en place dans un système décisionnel complexe.

Ces données applicatives sont donc extraites, transformées et chargées dans un entrepôt de données par un outil de type ETL ou en français ETC (Extraction-Transformation-Chargement).

Ainsi un outil d'ETL permet la synchronisation de données de tous types dans notre datawarehouse et nos datamarts

Un ETL repose sur des connecteurs permettant l'extraction ou l'importation des données de types divers (bases de données de tout type, fichiers xml ou autres formats, …) et sur des transformateurs qui manipulent les données : agrégations, filtres, conversions, mises en correspondance. Ainsi un tel outil va aller chercher les données d'une entreprise, les transformer pour les mettre en commun et les rendre utilisables dans le cadre de l'aide à la décision pour enfin les injecter dans un entrepôt de données.

Ces actions peuvent être effectuées de manière périodique via l'utilisation de batchs. Par exemple l'action d'extraction, de transformation et de chargement peut être lancée tous les soirs à minuit lorsque plus aucune application n'est utilisée.

- **L'intégration :**

Cette deuxième étape est l'intégration des données. Une fois les données centralisées par un outil d'ETL, celles-ci doivent être structurées au sein de l'entrepôt de données. Cette étape est toujours faite par un ETL grâce à un connecteur permettant l'écriture dans le datawarehouse. L'intégration est en fait un pré-traitement ayant pour but de faciliter l'accès aux données centralisées aux outils d'analyse.

Ainsi l'intégration consiste à concentrer les données collectées dans un espace unifié, dont le socle informatique essentiel est l'entrepôt de données. Ce dernier est l'élément central du dispositif dans le sens où il permet aux applications d'aide à la décision de bénéficier d'une source d'information homogène, commune, normalisée et fiable. Cette centralisation permet surtout de s'abstraire de la diversité des sources de données.

Lors de cette étape les données sont transformées et filtrées en vue du maintien de la cohérence d'ensemble (les valeurs acceptées par les filtres de l'outil d'ETL de la fonction de collecte mais qui peuvent introduire des incohérences dans les données centralisées sont soit rejetées, soit intégrées après une phase d'adaptation)

Enfin, c'est aussi durant cette étape que sont effectués les éventuels calculs et agrégations communs à l'ensemble du SID.

- **La diffusion :**

Cette étape de diffusion met les données à la disposition des utilisateurs. Elle permet la gestion de droits d'accès et respecte donc des schémas correspondant au profil ou au métier de chacun. Ainsi l'accès direct à l'entrepôt de données n'est pas autorisé. En effet ce genre de pratique ne correspond généralement pas aux besoins des décideurs ou analystes. L'objectif principal de l'étape de diffusion est de segmenter les données collectées en contextes qui soient cohérents, simples à utiliser et qui correspondent à une activité décisionnelle particulière (par exemple aux besoins d'un service particulier). En comparaison de l'entrepôt de données peut héberger de nombreuses variables ou indicateurs, un contexte de diffusion n'en présente que quelques dizaines pour rester simple d'exploitation. Chaque contexte peut correspondre à un datamart, bien que le stockage physique ne soit pas sujet à des règles particulières. Généralement un contexte de diffusion est multidimensionnel : il est modélisable sous la forme d'un hypercube et peut donc être mis à disposition via un outil OLAP.

Enfin les différents contextes d'un même SID n'ont pas forcément tous besoin du même niveau de détail selon la cible visée. En effet de nombreux agrégats n'intéressent que certaines applications et ne sont donc pas considérés comme des agrégats communs. Ces cumuls ne sont donc pas gérés par la fonction d'intégration mais par la diffusion. Ils peuvent être soit calculés dynamiquement soit stockés de manière persistante.

· La restitution :

Cette dernière étape, également appelée reporting, se charge de présenter les informations à valeur ajoutée de telle sorte qu'elles apparaissent de la façon la plus lisible possible dans le cadre de l'aide à la décision. Les données sont principalement modélisées par des représentations à base de requêtes afin de constituer des tableaux de bord ou des rapports via des outils d'analyse décisionnelle.

Cette quatrième fonction, la plus visible pour l'utilisateur assure le fonctionnement du poste de travail, le contrôle d'accès aux rapports, la prise en charge des requêtes et la visualisation des résultats sous quelque forme que ce soit.

Le reporting est l'application la plus utilisée dans l'informatique décisionnelle, il permet aux décideurs :

- De sélectionner des données par période, production, secteur de clientèle, etc.,
- De trier, regrouper ou répartir ces données selon des critères de choix,
- De réaliser des calculs (totaux, moyennes, sommes, pourcentages, écarts, comparatif, ...),
- De présenter les résultats de manière synthétique ou détaillée, généralement sous forme de graphiques.

Les programmes utilisés pour le reporting permettent de faire varier certains critères pour affiner l'analyse. Des instruments de type tableau

de bord équipés de fonctions d'analyses multidimensionnelles de type Olap sont aussi utilisés pour cette dernière étape du SID.

Enfin, à ce stade, il est essentiel de définir certains termes, qui seront utilisés dans la suite de ce rapport :

a. **Datawarehouse** (ou entrepôt de données) : base de données utilisée pour collecter et stocker des informations volatiles provenant d'autres bases de données.

b. **Datamart** : sous-ensemble logique d'un datawarehouse. Il est généralement exploité en entreprise pour restituer des informations ciblées sur un métier spécifique.

c. **Datamining** : les outils de datamining permettent d'extraire des hypothèses à partir de grandes quantités de données, par des procédés typiquement statistiques.

d. **ETL** (Extract, Transform, Load) : il s'agit d'une technologie informatique intergicielle (comprendre middleware) permettant d'effectuer des synchronisations massives d'informations d'une base de données vers une autre.

e. **Cube OLAP** : représentation abstraite d'informations multidimensionnelles. Les données sont rangées selon un principe de dimensions correspondant étroitement aux axes de recherche des utilisateurs (par exemple les ventes de produits dans le temps parkzone géographique).

2.5 Business Intelligence, Architecture du SID

2.5.1 Architecture du Système d'Information Décisionnelle (SID)

La Business Intelligence (informatique décisionnelle) propose d'utiliser les données transitant par le Système d'information, données de production le plus souvent, en informations susceptibles d'être

exploitées à des fins décisionnelles. Sur le plan pratique et technique, la Business Intelligence se compose d'une famille d'outils informatiques et de progiciels assurant le fonctionnement de la chaîne de traitement de l'information.

2.5.2 Les 4 fonctions de la chaîne décisionnelle

Il est coutumier de présenter les éléments et outils composant la chaîne décisionnelle en quatre catégories correspondant chacune à une fonction spécifique, à une phase du processus.

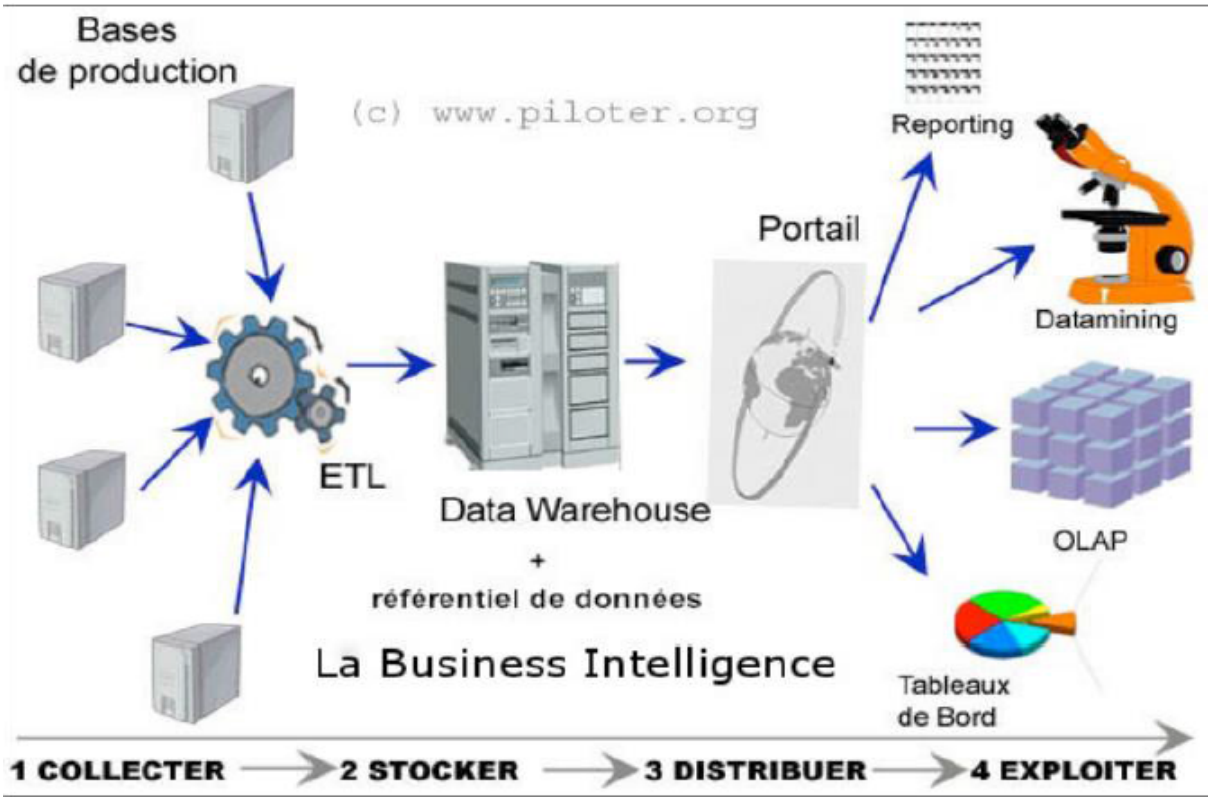


Figure 3: Les 4 phases du processus de Business Intelligence, de la donnée à l'information

•-1- Collecter, nettoyer et consolider les données Extraire les données des systèmes de production et les adapter à un usage décisionnel.
• -2- Stocker Centraliser les données structurées et traitées afin qu'elles soient disponibles pour un usage décisionnel.

- -3- Distribuer ou plutôt faciliter l'accessibilité des informations selon les fonctions et les types d'utilisation.
- -4- Exploiter ou comment assister du mieux possible l'utilisateur afin qu'il puisse extraire la substance de l'information des données stockées à cet usage.

1.6 Les outils utiliser au Business Intelligence

Le système d'information de l'entreprise ne s'est pas bâti en un temps unique. La majorité des systèmes d'information d'entreprise sont de nature hétérogène pour la plupart. Bien que la standardisation des échanges entre les divers outils informatiques avance à grands pas, la disparité des formats des données en circulation est toujours une réalité. C'est le principal obstacle technologique aux échanges étendus d'informations.

1.6.1 Collecter : Les outils d'ETL (Extract Transform and Load)

sas
Data Integration
CloverETL
PowerCenter 8
INFORMATICA
tableau
SOFTWARE
IBM
COGNOS
pentaho
A Hitachi Group Company
rapidminer
talend
open data solutions
SAP
Business Intelligence
scriptella
Microsoft
SQL Server
Integration Services
SISENSE
VanillaETL
ORACLE
FUSION MIDDLEWARE
DATA INTEGRATOR
11g
Google BigQuery
Amazon Redshift
ab INITIO
CloverDX
TERADATA
MicroStrategy

1.6.2 Distribuer les informations : Portail décisionnel EIP Enterprise Information Portal

1.6.3 Exploiter : Tableau de bord, analyse OLAP , data mining,...

1.7 PROBLÉMATIQUE:

Une société de vente voulait déployer un entrepôt de données pour analyser le profit en fonction du Temps ,les Clients, les Produits et les Régions. sachant que toutes les informations sont stockées sous forme d’un fichier Excel. Pour vulgariser les informations pour en faciliter leur interprétation au travers de rapports et tableaux de bord de suivi des activités de l'entreprise, qu'ils soient analytiques ou prospectifs. Alors pour bien gérer nos données on a utilisé Oracle 11g, pentaho et power BI.
Une société de vente voulait déployer un entrepôt de données pour analyser les montant de ventes par rapport aux temps, sachant que toutes les informations sont stockées sous forme d’un fichier CSV dont le schéma est indiqué dans la gure suivante :(Train.CSV)

Order Date	Ship Date	Ship Mode	Customer ID	Customer Name	Segment	Country	City	State
08/11/2017	11/11/2017	Second Class	CG-12520	Claire Gute	Consumer	United States	Henderson	Kentucky
08/11/2017	11/11/2017	Second Class	CG-12520	Claire Gute	Consumer	United States	Henderson	Kentucky
12/06/2017	16/06/2017	Second Class	DV-13045	Darrin Van Huff	Corporate	United States	Los Angeles	California
11/10/2016	18/10/2016	Standard Class	SO-20335	Sean O'Donnell	Consumer	United States	Fort Lauderdale	Florida
11/10/2016	18/10/2016	Standard Class	SO-20335	Sean O'Donnell	Consumer	United States	Fort Lauderdale	Florida
09/06/2015	14/06/2015	Standard Class	BH-11710	Brosina Hoffman	Consumer	United States	Los Angeles	California
09/06/2015	14/06/2015	Standard Class	BH-11710	Brosina Hoffman	Consumer	United States	Los Angeles	California
09/06/2015	14/06/2015	Standard Class	BH-11710	Brosina Hoffman	Consumer	United States	Los Angeles	California
09/06/2015	14/06/2015	Standard Class	BH-11710	Brosina Hoffman	Consumer	United States	Los Angeles	California
09/06/2015	14/06/2015	Standard Class	BH-11710	Brosina Hoffman	Consumer	United States	Los Angeles	California
09/06/2015	14/06/2015	Standard Class	BH-11710	Brosina Hoffman	Consumer	United States	Los Angeles	California
09/06/2015	14/06/2015	Standard Class	BH-11710	Brosina Hoffman	Consumer	United States	Los Angeles	California
15/04/2018	20/04/2018	Standard Class	AA-10480	Andrew Allen	Consumer	United States	Concord	North Carolina
05/12/2017	10/12/2017	Standard Class	IM-15070	Irene Maddox	Consumer	United States	Seattle	Washington
22/11/2016	26/11/2016	Standard Class	HP-14815	Harold Pawlan	Home Office	United States	Fort Worth	Texas
22/11/2016	26/11/2016	Standard Class	HP-14815	Harold Pawlan	Home Office	United States	Fort Worth	Texas
11/11/2015	18/11/2015	Standard Class	PK-19075	Pete Kriz	Consumer	United States	Madison	Wisconsin
13/05/2015	15/05/2015	Second Class	AG-10270	Alejandro Grove	Consumer	United States	West Jordan	Utah
27/08/2015	01/09/2015	Second Class	ZD-21925	Zuschuss Donatelli	Consumer	United States	San Francisco	California
27/08/2015	01/09/2015	Second Class	ZD-21925	Zuschuss Donatelli	Consumer	United States	San Francisco	California

Region	Product ID	Category	Sub-Category	Product Name	Sales
South	FUR-BO-10001798	Furniture	Bookcases	Bush Somerset Collection Bookcase	261.96
South	FUR-CH-10000454	Furniture	Chairs	Hon Deluxe Fabric Upholstered Stacking Chairs, Rounde...	731.94
West	OFF-LA-10000240	Office Supplies	Labels	Self-Adhesive Address Labels for Typewriters by Univer...	14.62
South	FUR-TA-10000577	Furniture	Tables	Bretford CR4500 Series Slim Rectangular Table	957.5775
South	OFF-ST-10000760	Office Supplies	Storage	Eldon Fold 'N Roll Cart System	22.368
West	FUR-FU-10001487	Furniture	Furnishings	Eldon Expressions Wood and Plastic Desk Accessories,...	48.86
West	OFF-AR-10002833	Office Supplies	Art	Newell 322	7.28
West	TEC-PH-10002275	Technology	Phones	Mitel 5320 IP Phone VoIP phone	907.152
West	OFF-BI-10003910	Office Supplies	Binders	DXL Angle-View Binders with Locking Rings by Samsill	18.504
West	OFF-AP-10002892	Office Supplies	Appliances	Belkin F5C206VTEL 6 Outlet Surge	114.9
West	FUR-TA-10001539	Furniture	Tables	Chromcraft Rectangular Conference Tables	1706.184
West	TEC-PH-10002033	Technology	Phones	Konftel 250 ConferenceÂ phoneÂ - Charcoal black	911.424
South	OFF-PA-10002365	Office Supplies	Paper	Xerox 1967	15.552
West	OFF-BI-10003656	Office Supplies	Binders	Fellowes PB200 Plastic Comb Binding Machine	407.976
Central	OFF-AP-10002311	Office Supplies	Appliances	Holmes Replacement Filter for HEPA Air Cleaner, Very L...	68.81
Central	OFF-BI-10000756	Office Supplies	Binders	Storex DuraTech Recycled Plastic Frosted Binders	2.544
Central	OFF-ST-10004186	Office Supplies	Storage	Stur-D-Stor Shelving, Vertical 5-Shelf: 72"H x 36"W x 18...	665.88
West	OFF-ST-10000107	Office Supplies	Storage	Fellowes Super Stor/Drawer	55.5
West	OFF-AR-10003056	Office Supplies	Art	Newell 341	8.56
West	TEC-PH-10001949	Technology	Phones	Cisco SPA 501G IP Phone	213.48

Extrait du fichier Excel contenant les données en question

La direction a besoin d’un système décisionnel pour analyser quelques mesures pertinentes, par exemple :

- montants des ventes par client, par produit et par région en fonction du temps·

1.8 CONCEPTION DE L'ENTREPÔT DE DONNÉES:

Sujet : Analyse des ventes de l'entreprise.

• Dimensions :

Une dimension est un ensemble de données structurées correspondant à un axe d’analyse sur un événement mesurable. Dans ce contexte, les événements sont appelés «faits».
La plupart du temps une table de dimension est dénormalisée, cependant il faut toujours veiller à respecter la première forme normale. Il est préférable d’utiliser une clé physique générique sans aucun sens de type auto-incrémentée. Une autre approche, avec une clé logique, permet de rendre plus simple la lecture et la maintenance due à l’évolution des dimensions. Elle simplifie aussi les jointures avec les tables de faits et réduit son volume de données.

• Table de faits :

Une table de fait est une table qui contient les données observables (les faits) que l’on possède sur un sujet et que l’on veut étudier, selon divers axes d’analyse (les dimensions). Les « faits », dans un entrepôt de données, sont normalement numériques, puisque d’ordre quantitatif. Il peut s’agir du montant en argent des ventes, du nombre d’unités vendues d’un produit. ainsi les clés associées aux dimensions. "Il s’agit des clés étrangères dans la table de faits"

• Modélisation du data warehouse :

Dans un Data Warehouse (et au niveau de chaque Data mart), les données et leurs relations sont organisées suivant un modèle de données spécifique. Le choix du modèle de données structure et définit le design du Data Warehouse. Nous avons vu qu'il existait trois modélisations possibles :

-La modélisation en étoile: est le paradigme de modélisation simple et commun dans lequel l'entrepôt de données comprend une table de faits avec une seule table pour chaque dimension. Le schéma imite une étoile, avec une table de dimension présentée dans un motif étalé entourant la table de faits centrale. La table des dimensions en fait est connectée à la table de dimension via la clé primaire et la clé étrangère.

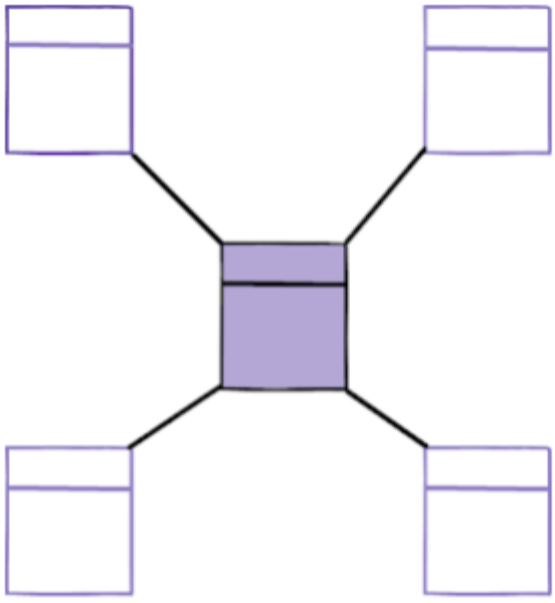
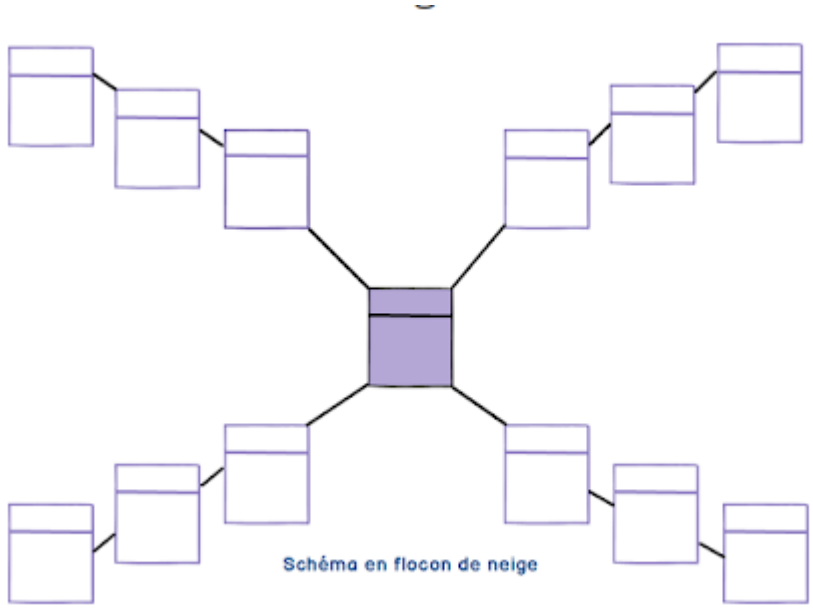


Schéma en étoile

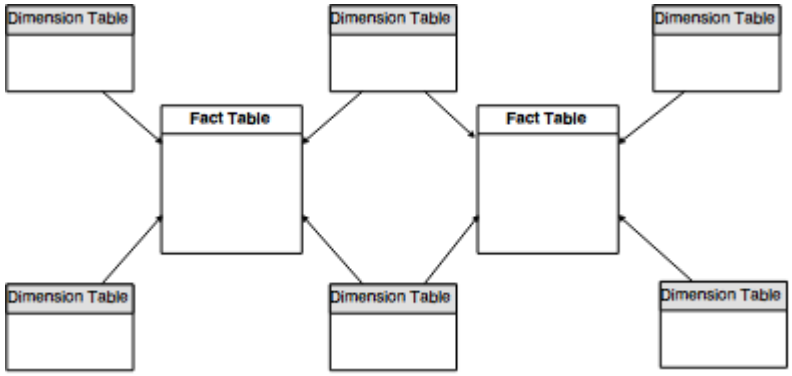
-La modélisation en flocons de neige: est un type de **schéma en étoile** qui inclut la forme hiérarchique des tables dimensionnelles. Dans ce schéma, il existe une table de faits composée de différentes tables de dimension et de sous-dimension reliées par des

clés primaires et étrangères à la table de faits. Il porte le nom de **flocon de neige** car sa structure ressemble à un flocon de neige.

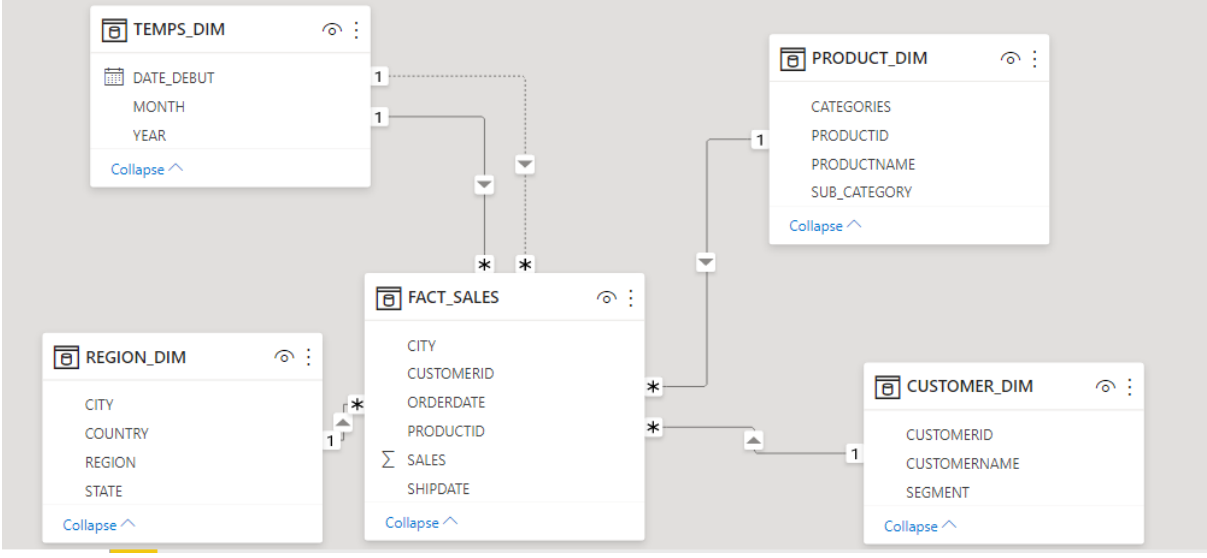
⇒ Il utilise la normalisation qui divise les données en tables supplémentaires. Le fractionnement permet de réduire la redondance et de prévenir les pertes de mémoire. Un **schéma de flocon de neige** est plus facile à gérer mais complexe à concevoir et à comprendre. Cela peut également réduire l'efficacité de la navigation car davantage de jointures seront nécessaires pour exécuter une requête

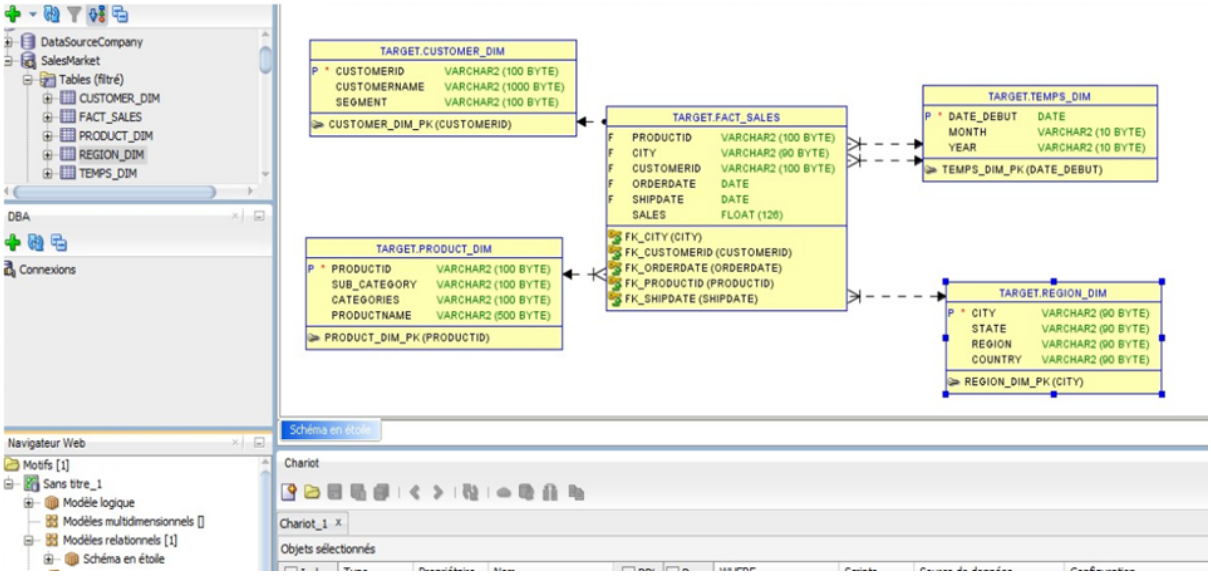


-La modélisation en constellation se compose de plusieurs tables des faits avec leurs tables de dimensions respectives. Les tables de dimensions communes aux différentes tables des faits ne font pas l'objet de redondances : c'est l'un des principaux avantages de cette modélisation. Cela permet de réduire l'espace de stockage nécessaire. Idéalement, il faut que les tables de dimensions partagées soient identiques et contiennent les mêmes valeurs, les mêmes attributs. Dans le cas contraire, des ajustements sont nécessaires pour que les tables de dimensions partagées conviennent aux deux besoins métiers.



Dans notre projet, nous allons utiliser le schéma en étoile.




DataSourceCompany
SalesMarket
Tables (filtré)
CUSTOMER_DIM
FACT_SALES
PRODUCT_DIM
REGION_DIM
TEMPS_DIM
DBA
Connexions
Navigateur Web
Motifs [1]
Sans titre_1
Modèle logique
Modèles multidimensionnels []
Modèles relationnels [1]
Schéma en étoile
TARGET.CUSTOMER_DIM
P * CUSTOMERID VARCHAR2 (100 BYTE)
CUSTOMERNAME VARCHAR2 (1000 BYTE)
SEGMENT VARCHAR2 (100 BYTE)
CUSTOMER_DIM_PK (CUSTOMERID)
TARGET.FACT_SALES
F PRODUCTID VARCHAR2 (100 BYTE)
F CITY VARCHAR2 (90 BYTE)
F CUSTOMERID VARCHAR2 (100 BYTE)
F ORDERDATE DATE
F SHIPDATE DATE
SALES FLOAT (126)
FK_CITY (CITY)
FK_CUSTOMERID (CUSTOMERID)
FK_ORDERDATE (ORDERDATE)
FK_PRODUCTID (PRODUCTID)
FK_SHIPDATE (SHIPDATE)
TARGET.TEMPS_DIM
P * DATE_DEBUT DATE
MONTH VARCHAR2 (10 BYTE)
YEAR VARCHAR2 (10 BYTE)
TEMPS_DIM_PK (DATE_DEBUT)
TARGET.PRODUCT_DIM
P * PRODUCTID VARCHAR2 (100 BYTE)
SUB_CATEGORY VARCHAR2 (100 BYTE)
CATEGORIES VARCHAR2 (100 BYTE)
PRODUCTNAME VARCHAR2 (500 BYTE)
PRODUCT_DIM_PK (PRODUCTID)
TARGET.REGION_DIM
P * CITY VARCHAR2 (90 BYTE)
STATE VARCHAR2 (90 BYTE)
REGION VARCHAR2 (90 BYTE)
COUNTRY VARCHAR2 (90 BYTE)
REGION_DIM_PK (CITY)
Schéma en étoile
Chariot
Chariot_1
Objets sélectionnés

Chapitre 2 : PRÉPARATION DE L'ENVIRONNEMENT

2.1.Pentaho Data Integration PDI:

PDI (Pentaho Data Integration), qui était auparavant connu sous le nom de Kettle, est un logiciel d'ETL (Extract, Transform, Load) Open Source qui permet la conception ainsi que l'exécution des opérations de manipulation et de transformation de données très complexes.

2.1.1 Étapes d'installation :

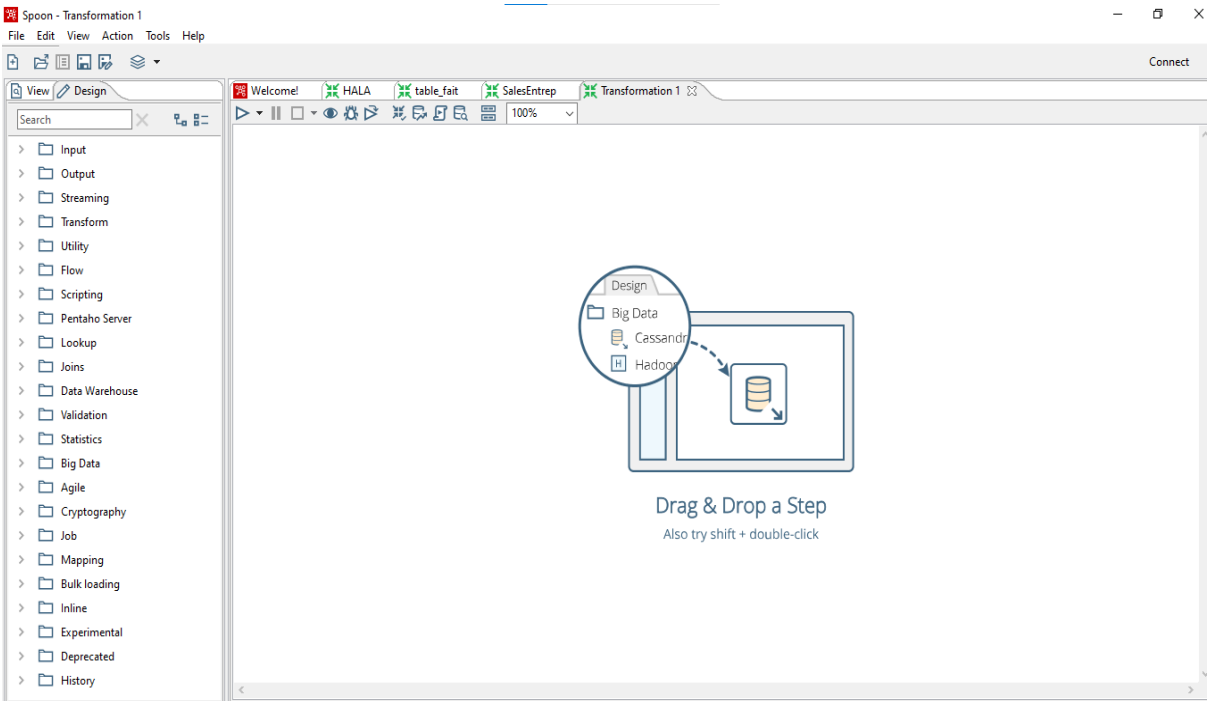
- PDI 9.0 nécessite l'installation d'un Java Development Kit (JDK)

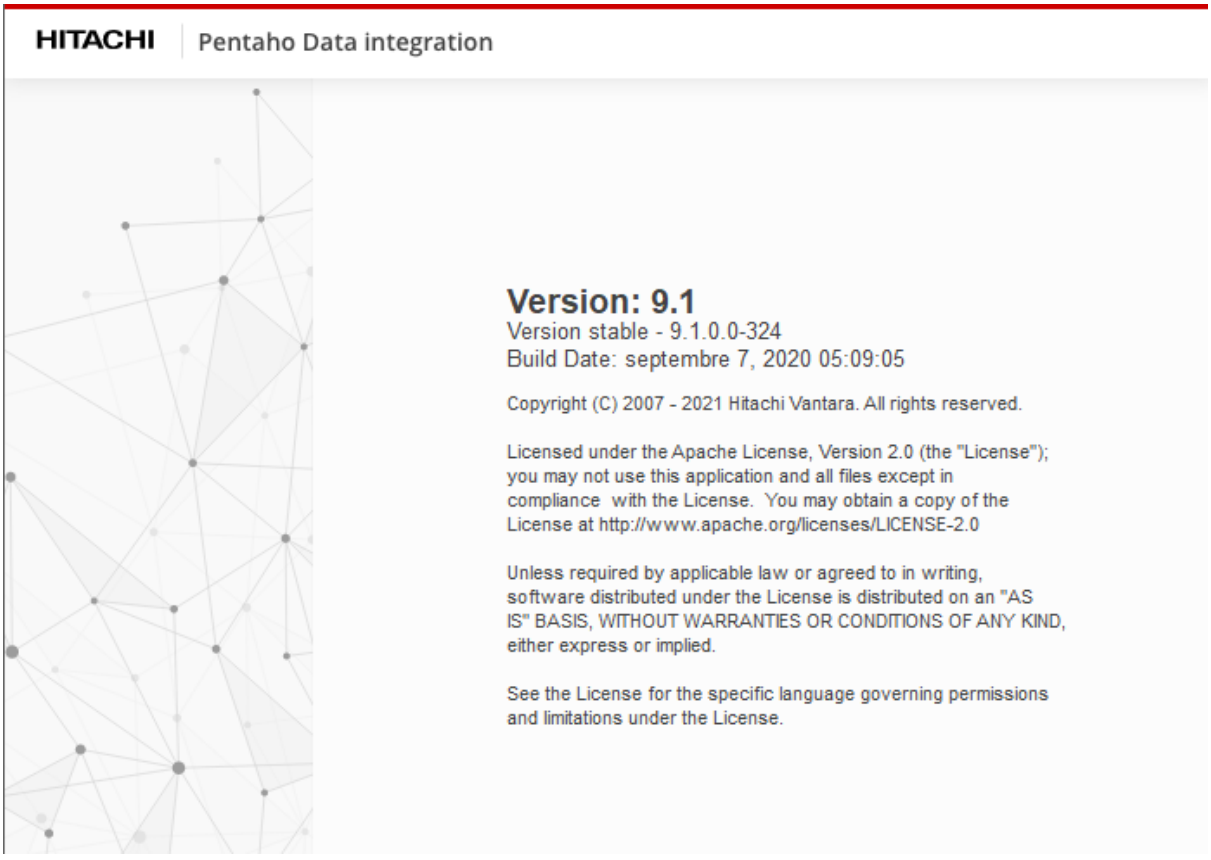
version 8.0 ou supérieure (la version choisie ici est 8.0u241)

- Téléchargement de data-integration (connu également sous le nom Kettle), c'est le lien de téléchargement https://sourceforge.net/projects/pentaho/files/Data%20Integration/
- Dézipper l'archive dans un répertoire quelconque

Il est maintenant possible d'utiliser data-integration en exécutant le programme spoon.bat sous Windows ou spoon.sh sous Linux et Mac OSX.

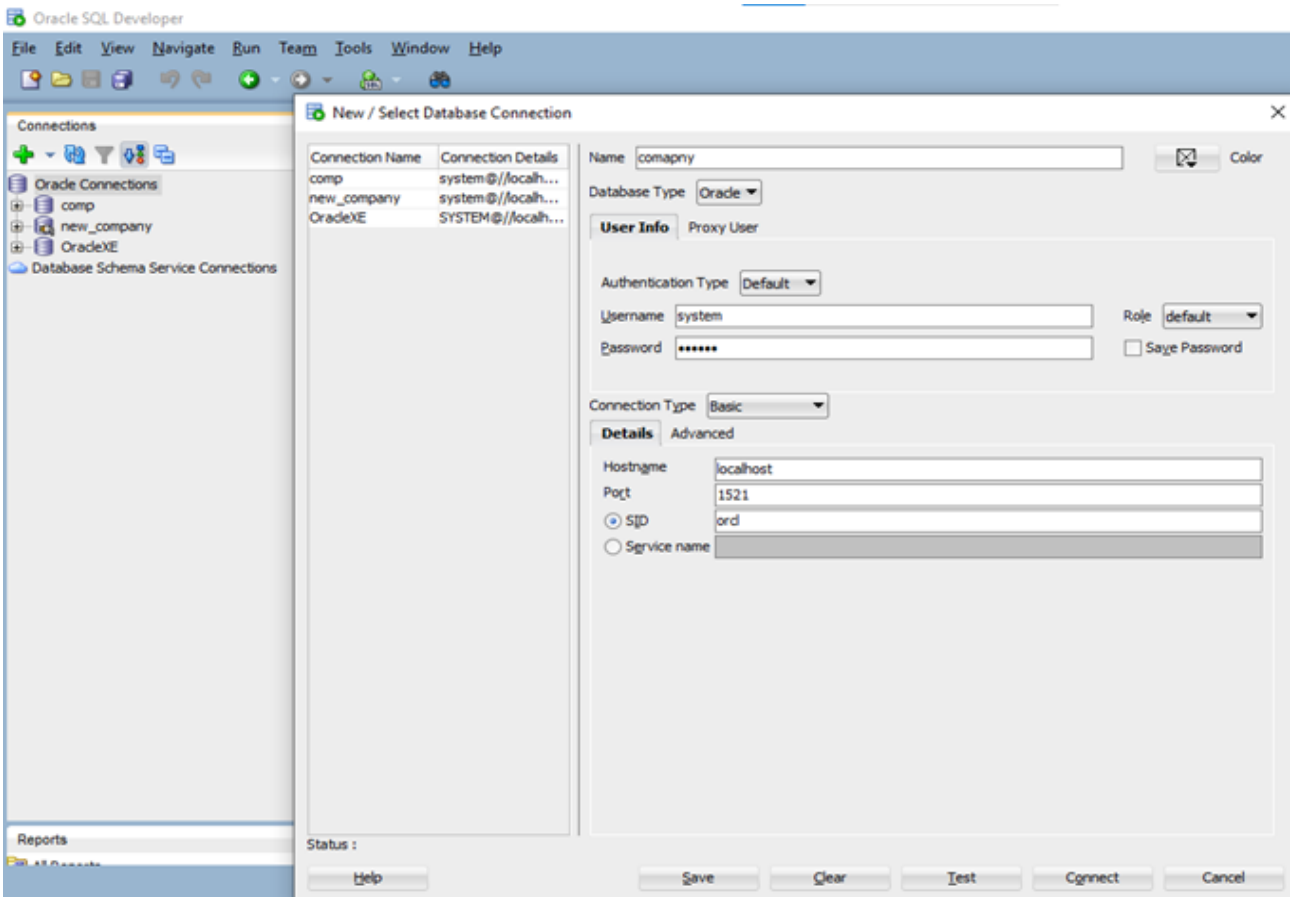
Espace de travail :


Spoon - Transformation 1
File Edit View Action Tools Help
Connect
View
Design
Welcome!
HALA
table_fait
SalesEntrep
Transformation 1
100%
Search
Input
Output
Streaming
Transform
Utility
Flow
Scripting
Pentaho Server
Lookup
Joins
Data Warehouse
Validation
Statistics
Big Data
Agile
Cryptography
Job
Mapping
Bulk loading
Inline
Experimental
Deprecated
History
Design
Big Data
Drag & Drop a Step
Also try shift + double-click


HITACHI
Pentaho Data integration
Version: 9.1
Version stable - 9.1.0.0-324
Build Date: septembre 7, 2020 05:09:05
Copyright (C) 2007 - 2021 Hitachi Vantara. All rights reserved.
Licensed under the Apache License, Version 2.0 (the "License"); you may not use this application and all files except in compliance with the License. You may obtain a copy of the License at http://www.apache.org/licenses/LICENSE-2.0
Unless required by applicable law or agreed to in writing, software distributed under the License is distributed on an "AS IS" BASIS, WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied.
See the License for the specific language governing permissions and limitations under the License.

2.2 Oracle SQL Developer

Oracle SQL Developer est un environnement de développement intégré (EDI) multi-plateforme, fourni gratuitement par Oracle Corporation et utilisant la technologie Java (Java Development Kit). C'est un outil graphique permettant d'interroger des bases de données Oracle à l'aide du langage SQL.



En premier temps, nous avons créé un utilisateur “ Target” avec le mot de passe “Test” pour faire la connexion à tous les outils qu’on va utiliser pour réaliser notre projet.

Sélection SQL Plus

```
SQL*Plus: Release 11.2.0.1.0 Production on Mar. Juin 29 22:22:27 2021

Copyright (c) 1982, 2010, Oracle.  All rights reserved.

Entrez le nom utilisateur : system
Entrez le mot de passe :

ConnectÚ Ó :
Oracle Database 11g Enterprise Edition Release 11.2.0.1.0 - 64bit Production
With the Partitioning, OLAP, Data Mining and Real Application Testing options

SQL> create user Target IDENTIFIED BY test;

Utilisateur crÚÚ.

SQL> grant all PRIVILEGES to Target;

Autorisation de privilÞges (GRANT) acceptÚe.

SQL>
```

Sur [pentaho data integration]

En premier lieu, vous devrez créer une connexion à une base de données. La plupart des bases de données disponibles sur le marché sont fonctionnelles sous **Pentaho**, donc en cliquant sur le bouton New, il vous sera possible de sélectionner votre base de données et d'en indiquer les paramètres de connexion.

2.3 Pentaho workbench

Le Schema workbench, du projet Pentaho d'analyse de données OLAP «Mondrian», est une interface graphique permettant de produire des schémas Mondrian au format XML et de les publier dans une solution Pentaho

2.3.1 Étapes d’installation :

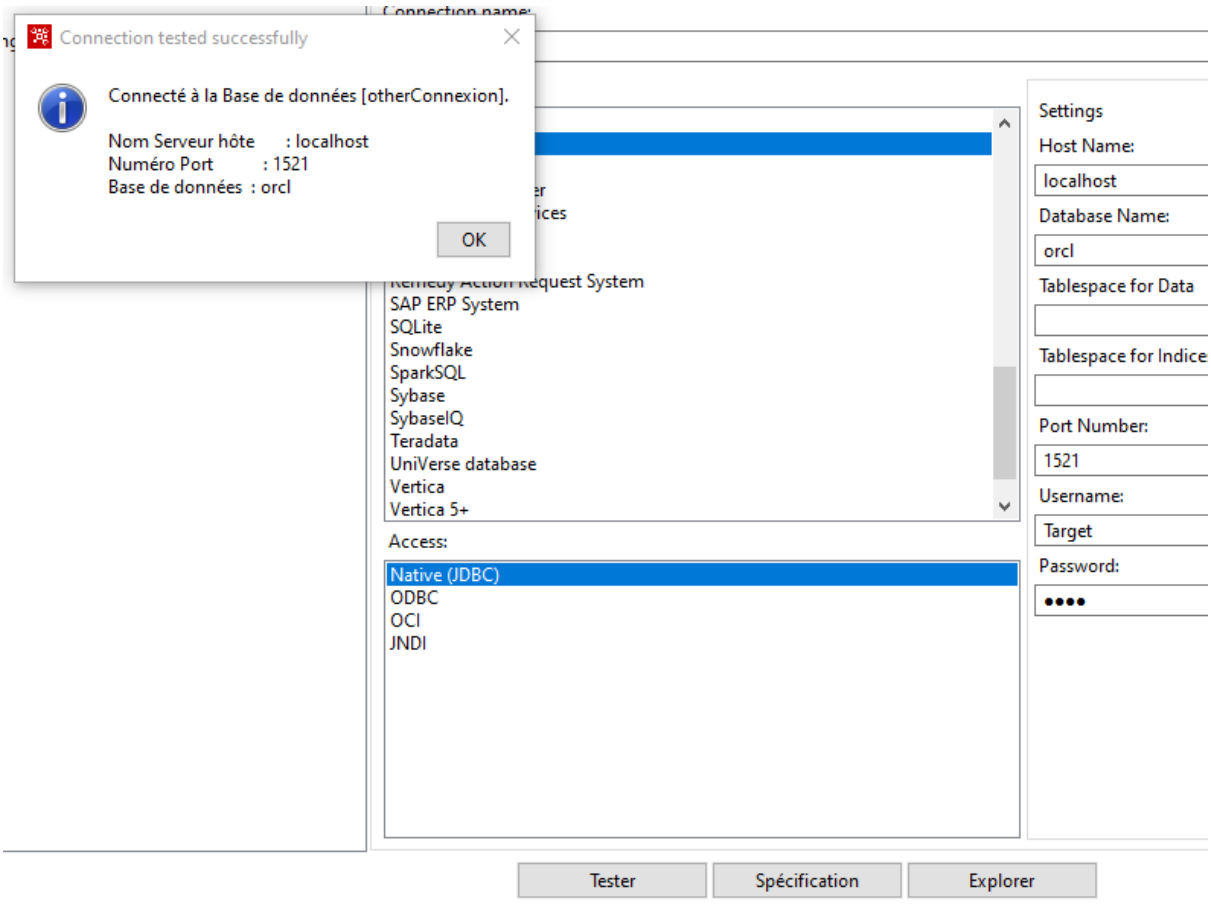
1. Le logiciel Pentaho Schema Workbench (PSW) est disponible pour le téléchargement depuis Sourceforge.net.
2. Conditions préalables : psw 3.14 nécessite l’installation d’un Java Development Kit (JDK) version 1.5 ou supérieur (la version choisie ici est 8.0u241)
3. PSW ne nécessite pas d’installation. Décompressez simplement le chier zip dans un dossier de votre choix.
4. Exécution : Lancer directement le chier workbench.bat pour l’ouvrir.

Chapitre 3 : MANIPULATION DES DONNÉES AVEC L’OUTIL PDI

3.1 CRÉER UNE CONNEXION ORACLE DANS PDI

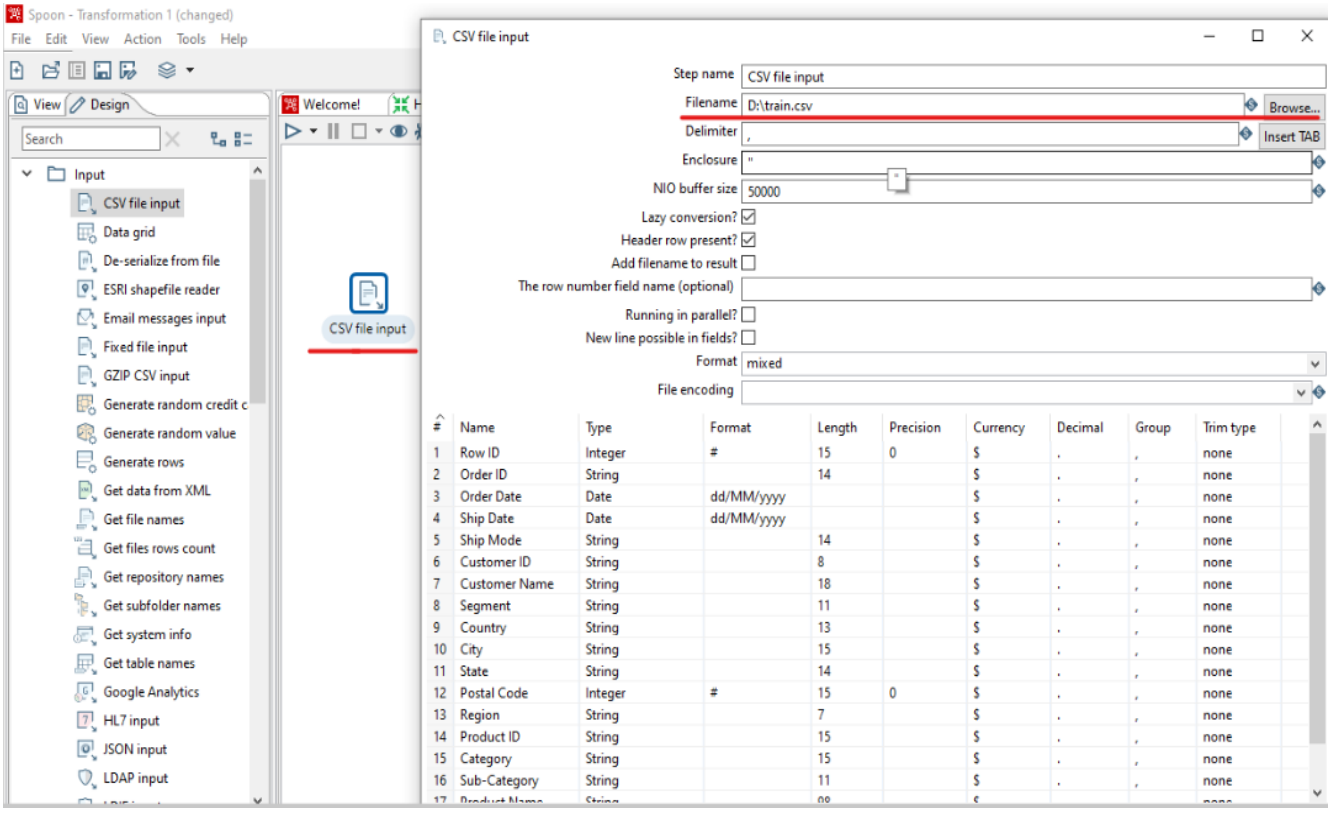
Avant d’établir la connexion avec la base de données nouvellement créée, nous devons installer : ojdbc6.jar (2,739,670 bytes)

Puis copier-coller dans pentaho\data-integration\lib



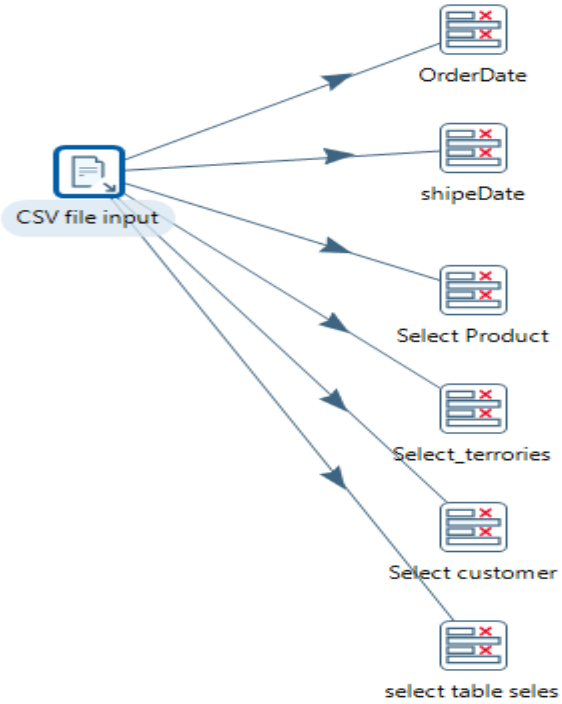
3.2 EXTRAIRE LES DONNÉES EXCEL:

Vu qu’il n’y a qu’une seule source et que le volume de données traitées est assez faible, nous n'utilisons pas de staging tables comme moyen de traitement temporaire, et nous satisferons par le processus ETL classique pour la création et l’alimentation des tables de dimensions ainsi que la table de faits.

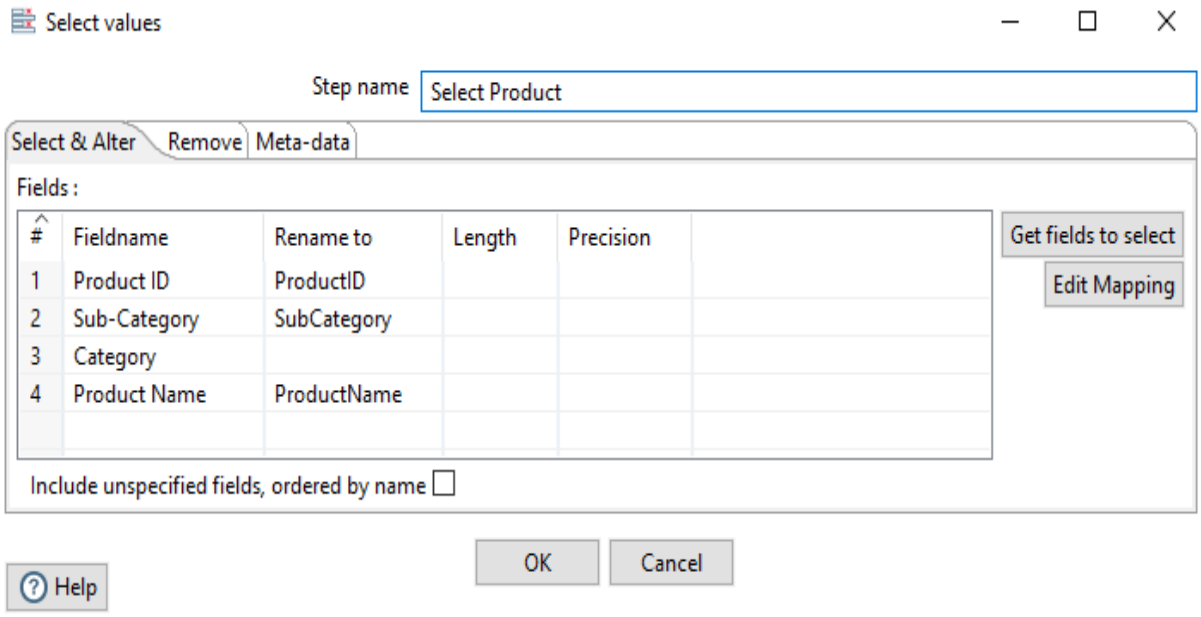


3.3 SÉLECTIONNER LE FLUX DE DONNÉES

Cliquez sur l'onglet puis Design puis transformez l'étape Select values dans l'espace de travail faites cette opération 6 fois .



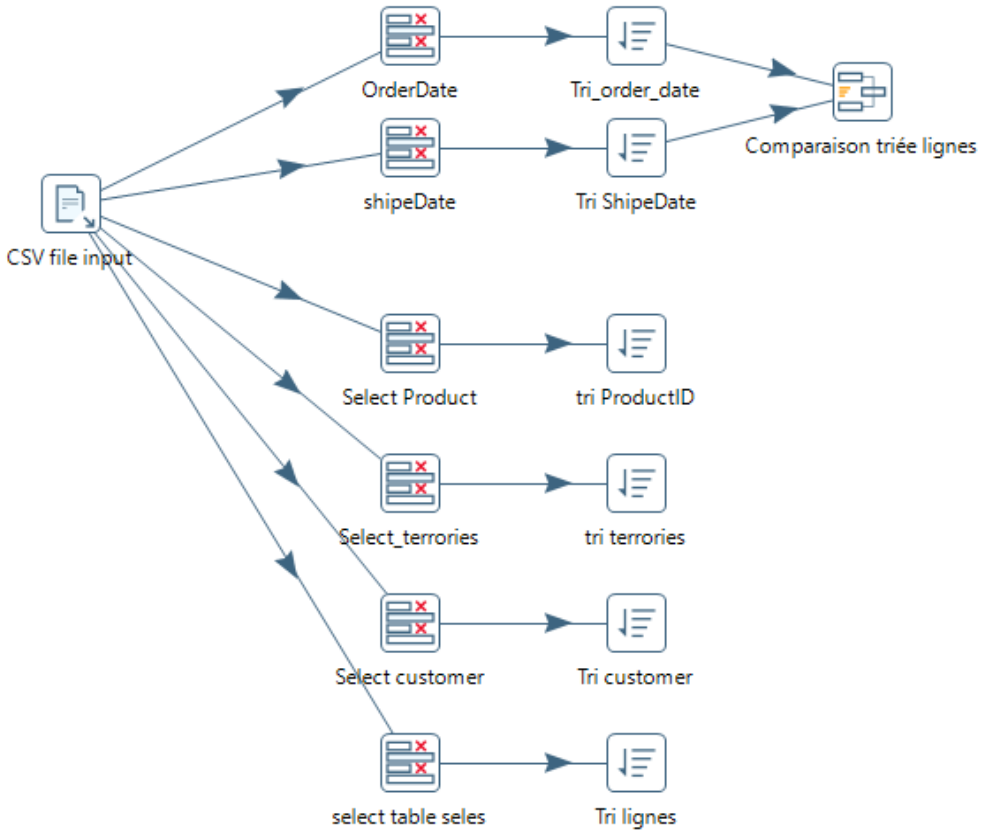
Pour chacune de ces 6 étapes, sélectionnez uniquement le flux de données qui vous intéressent.



3.4 TRIER LES CLÉS, SUPPRIMER LES DOUBLONS:

Cliquez sur l'onglet design → Transform , puis glissez l'étape Sort rows dans l'espace de travail. Dupliquer cette étape 6 fois (pour les 6 flux provenant de OrderDate, shipDate, Select Product, Select_terrories, Select customer, select table seles).

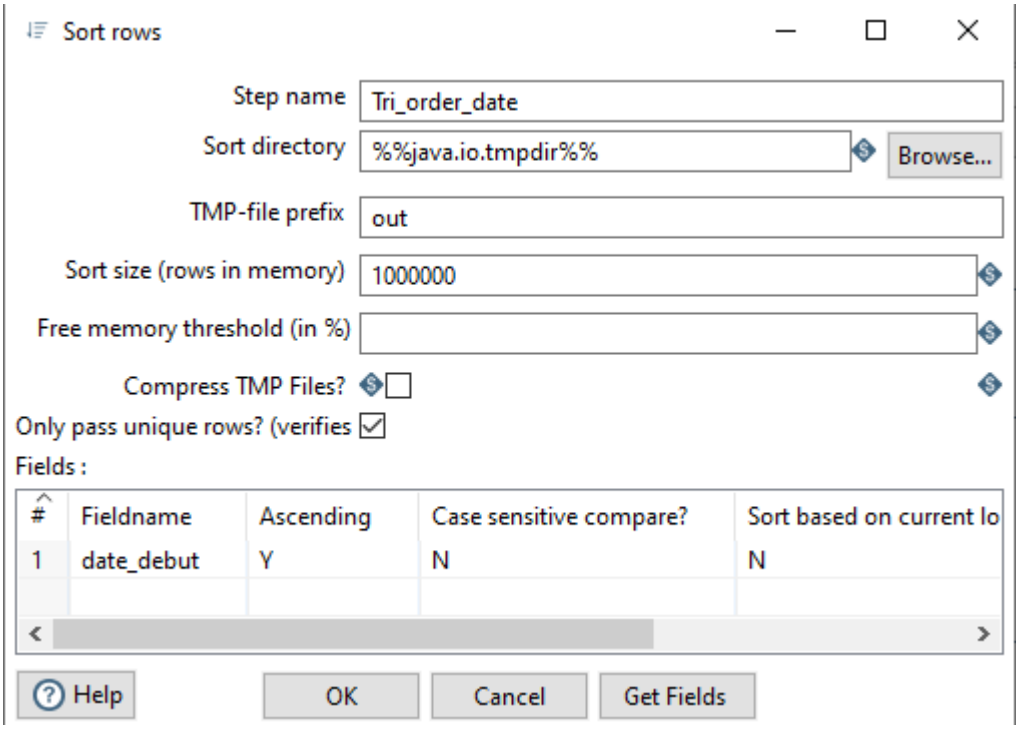
Ajoutez l'étape **Sorted merge** cette étape trier et fusionner les linge provenant de OrderDate et shipDate.



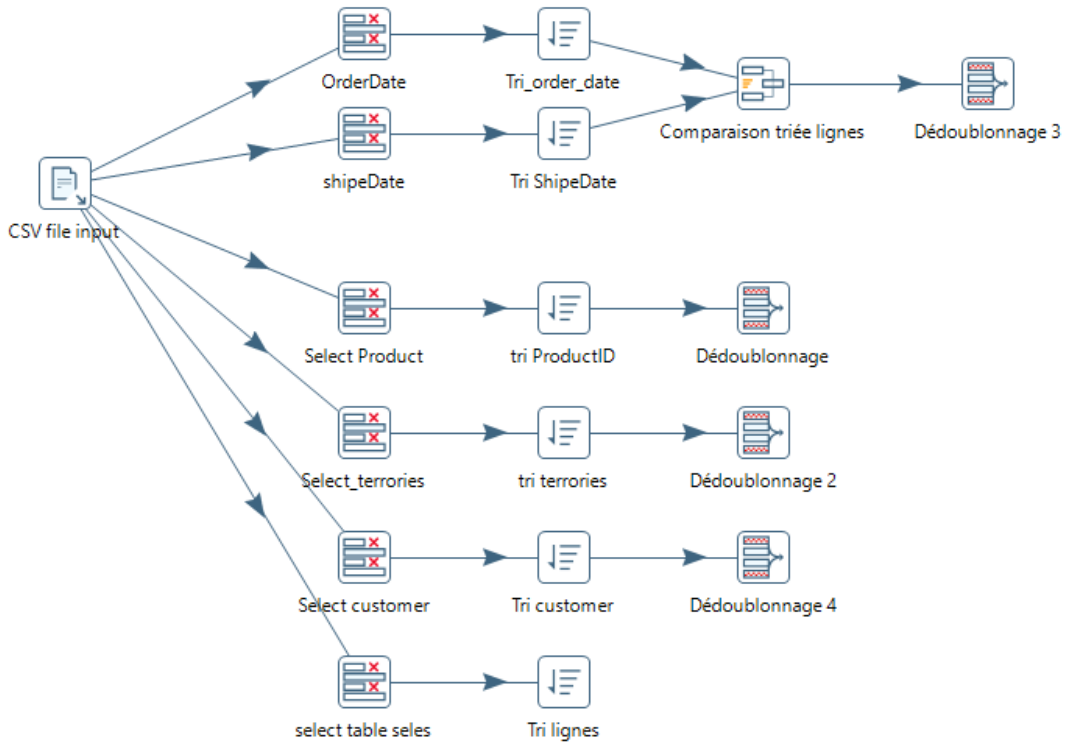
Pour chacune de ces 6 étapes (tri_order_date, Tri shipDate, tri ProductID, tri terrories, Tri customer, Tri lignes) :

- cliquez sur Get Fields puis enlevez ces champs et ne laissez que les champs déterminant la clé primaire.

- Cochez la case Only pass unique rows ? pour supprimer les groupes de clés doublons.

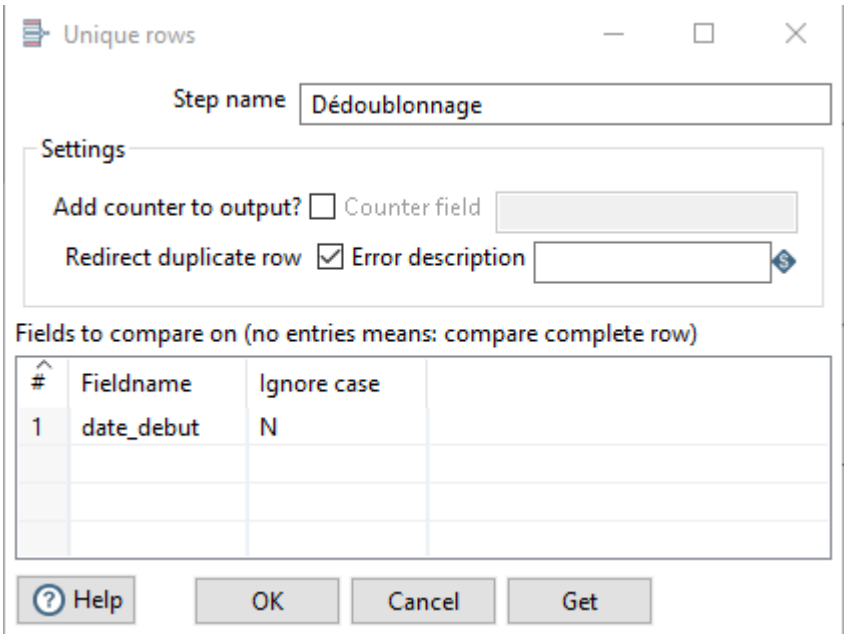


Cliquez sur l'onglet design → Transform , puis glissez l'étape Unique rows cette étape permet d'éliminer les doublons du ID, Dupliquer cette étape 4 fois.



Pour chacune de ces 4 étapes :

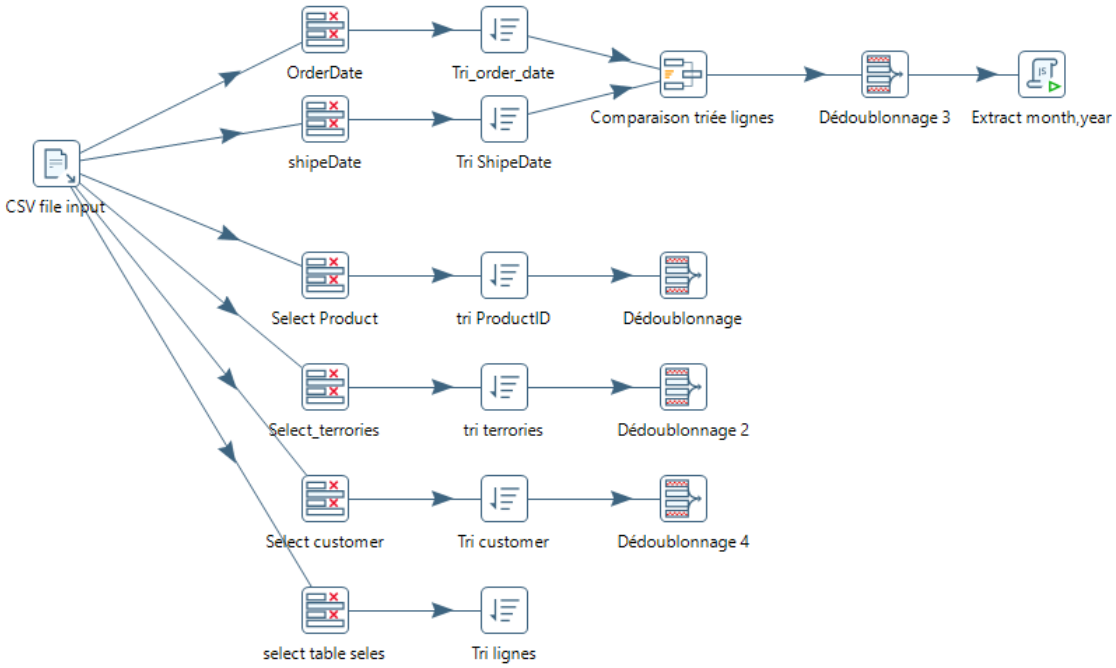
- cliquez sur Get puis enlevez ces champs et ne laissez que les champs déterminant la clé primaire.
- Cochez la case Redirect duplicate row.



3.5 création du champs month, year:

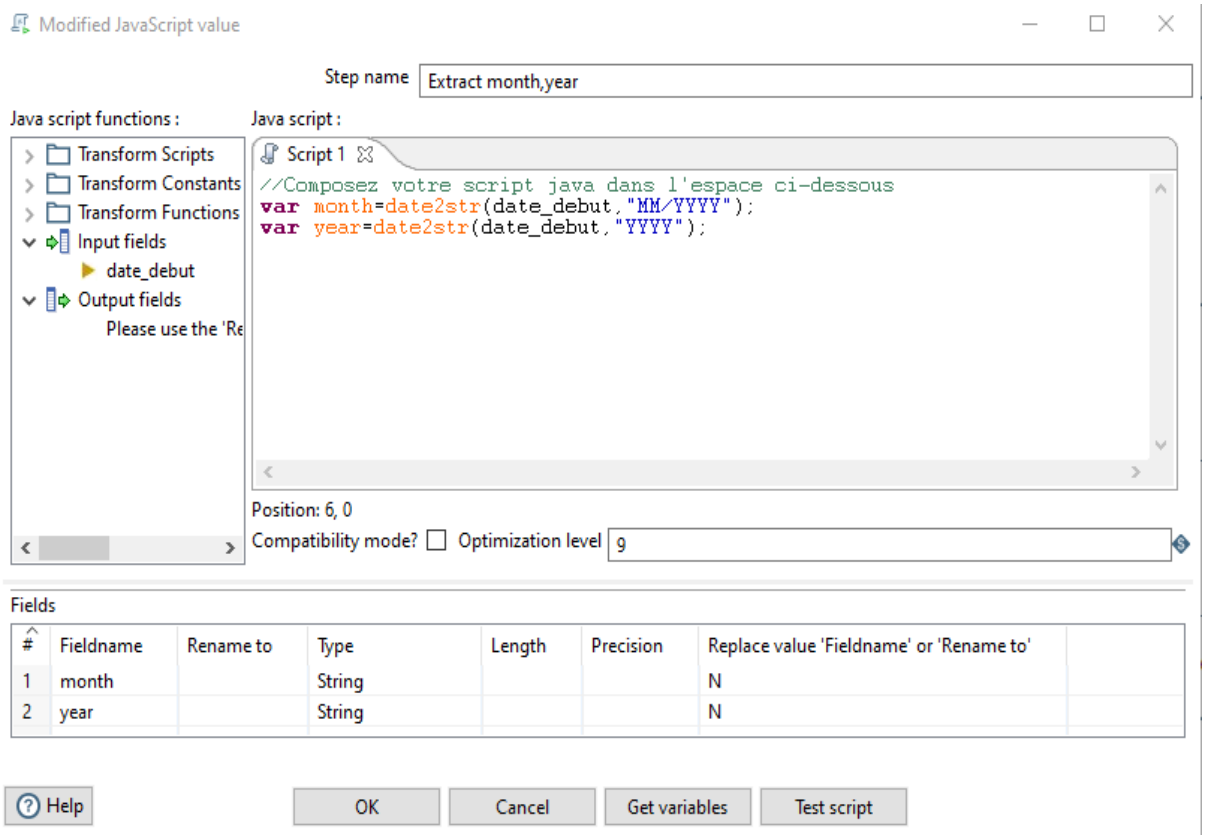
Cliquez sur l'onglet design → Scripting , puis glissez l'étape Modified JavaScript value dans l'espace de travail.

L'étape Modified JavaScript value fournit une interface utilisateur pour créer des expressions JavaScript que vous pouvez l'utiliser pour générer un flux de données additionnel.



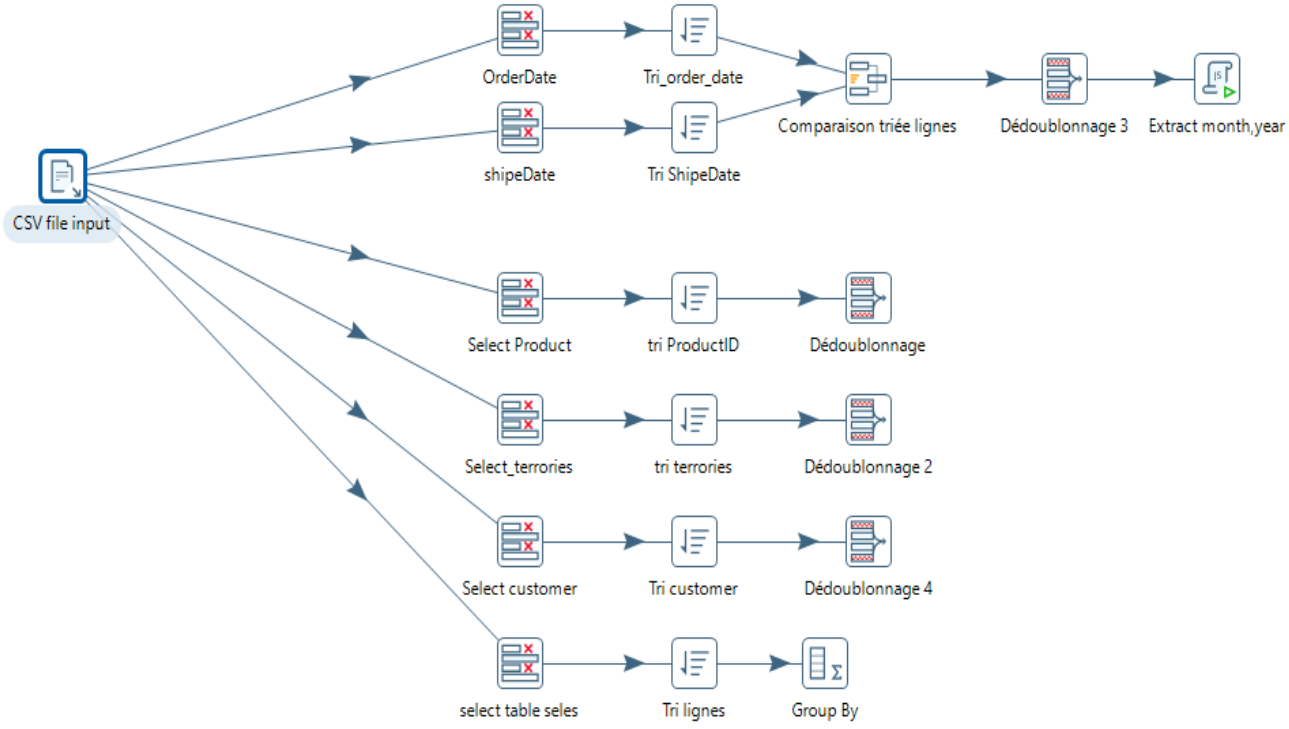
Utilisez le script ci-dessous pour créer ces deux champs de mois, et annee :

```
var month=date2str(date_debut,"MM/YYYY");
var year=date2str(date_debut,"YYYY");
```



3.6 APPLIQUER UNE FONCTION D'AGRÉGATION AUX MESURES:

Cliquez sur l'onglet design → Statistics , puis glissez l'étape Group by dans l'espace de travail.



Double-cliquez sur l'étape Group by et remplissez les champs qui composent le groupe, ainsi que les agrégats. Cette étape regroupe les clés en appliquant une fonction d'agrégation (COUNT, MAX, MIN, SUM, AVG,...) sur les mesures de nos choix.

Group by

Step name: Group By

Include all rows? ☐

Temporary files directory: %%java.io.tmpdir%% Browse...

TMP-file prefix: grp

Add line number, restart in each group ☐

Line number field name:

Always give back a result row ☐

The fields that make up the group:

#	Group field
1	OrderDate
2	ShipDate
3	CustomerID
4	City
5	ProductID

Get Fields

Aggregates :

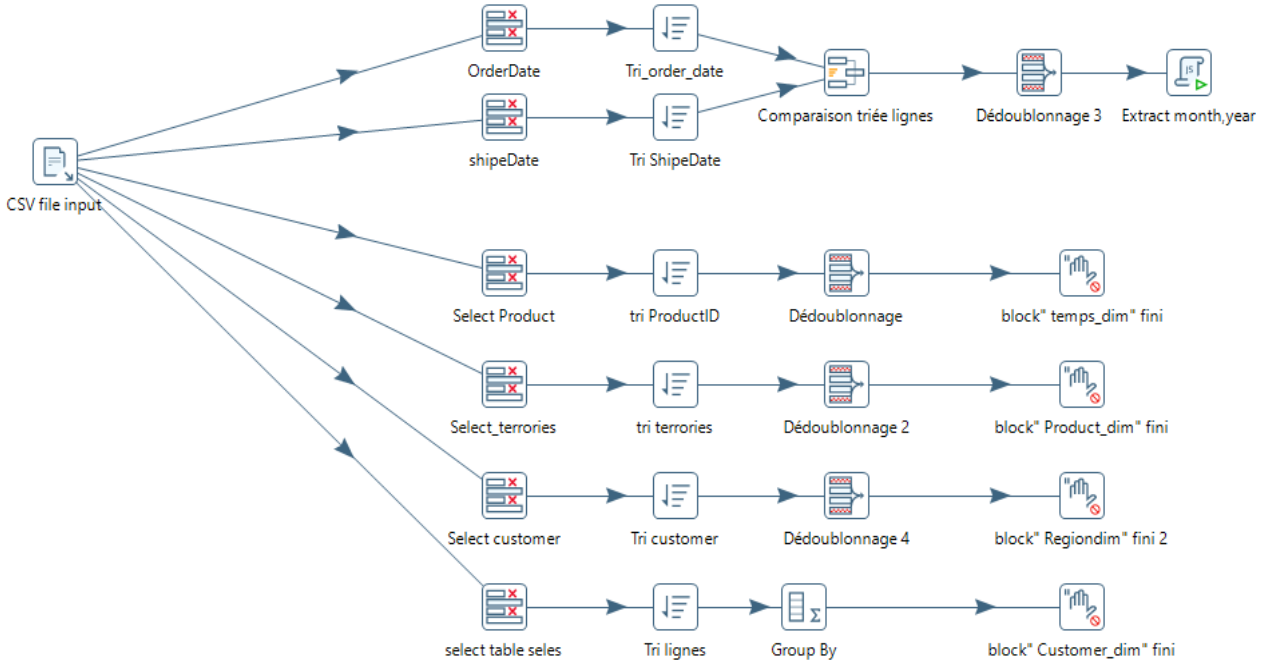
#	Name	Subject	Type	Value
1	Sales	Sales_Price	Sum	

Get lookup fields

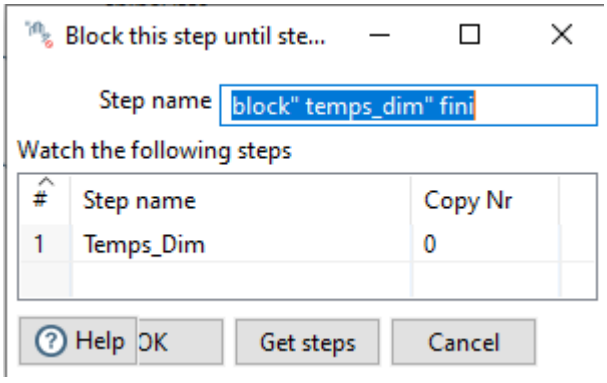
Help OK Cancel

3.7 ORDONNANCER LE FLUX:

Cliquez sur l'onglet design → Flow , puis glissez l'étape Block this step until steps finish dans l'espace de travail.Puis dupliquez cette étape 4 fois.

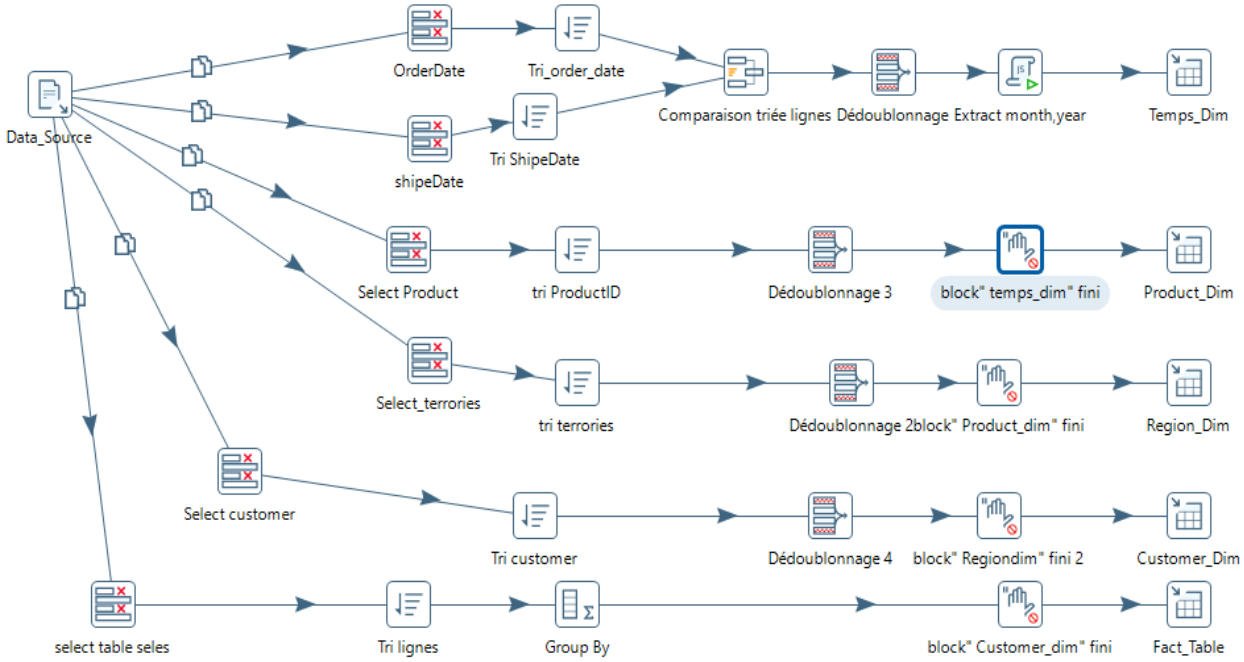


Pour chacune de ces 4 étapes, ajoutez les étapes à surveiller:



3.8 AJOUTER LES DIMENSIONS:

Cliquez sur l'onglet design → Output , puis glissez l'étape Table output dans l'espace de travail. Puis Dupliquer cette étape 5 fois .



avant de fair la configuration du étapes Dim en crée les table dim sur oracle:

- création du table Customer_Dim:

```
CREATE  TABLE Customer_Dim
(
    CustomerID VARCHAR(100) PRIMARY KEY ,
    CustomerName VARCHAR(1000),
    Segment VARCHAR(100)
);
```

- création du table Product_Dim:

```sql
CREATE TABLE Product_Dim
(
    ProductID VARCHAR(100) PRIMARY KEY,
    Sub_Category VARCHAR(100),
    Categories VARCHAR(100),
    ProductName VARCHAR(500)
);
```

- création du table Region_Dim:

```sql
CREATE TABLE Region_Dim
(
    City VARCHAR(90) PRIMARY KEY ,
    State VARCHAR(90),
    Region VARCHAR(90),
    Country VARCHAR(90)
);
```

- création du table Temps_Dim:

```sql
CREATE  TABLE Temps_Dim
(
    date_debut DATE PRIMARY KEY,
    month VARCHAR(10),
    year VARCHAR(10)
);
```

- création du table Fact_Sales:

```sql
CREATE TABLE Fact_Sales
(
    Sales_ID integer PRIMARY KEY,
    ProductID VARCHAR(100),
    City VARCHAR(90),
    CustomerID VARCHAR(100),
    OrderDate   DATE,
    ShipDate    DATE,
    Sales FLOAT,
    CONSTRAINT FK_ProductID FOREIGN KEY (ProductID) REFERENCES Product_Dim (ProductID),
    CONSTRAINT FK_City FOREIGN KEY (City) REFERENCES Region_Dim (City),
    CONSTRAINT FK_CustomerID FOREIGN KEY (CustomerID) REFERENCES Customer_Dim (CustomerID),
    CONSTRAINT FK_OrderDate FOREIGN KEY (OrderDate) REFERENCES Temps_Dim (date_debut),
    CONSTRAINT FK_ShipDate FOREIGN KEY (ShipDate) REFERENCES Temps_Dim (date_debut)
);
```

Pour chacune de ces 4 Table output :

- Entrez le nom de la connexion et celui de la table cible.
- Choisir la table dim depui Target table.
- Cliquez sur Get fields.

Table output

Step name: Temps_Dim

Connection: Company [Edit...] [New...] [Wizard...]

Target schema: ORADB1 [Browse...]

Target table: TEMPS_DIM [Browse...]

Commit size: 1000

Truncate table ☐

Ignore insert errors ☐

Specify database fields ☑

Main options | Database fields

Fields to insert:

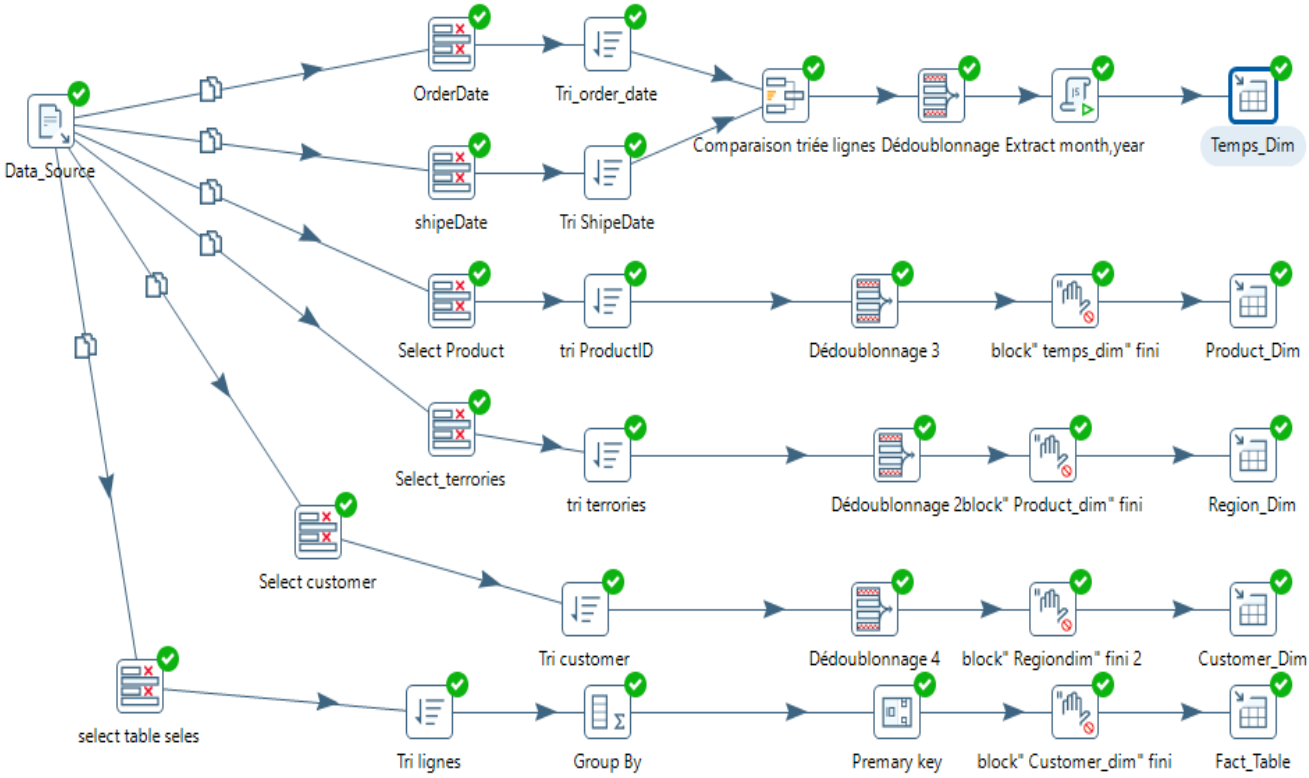
#	Table field	Stream field
1	date_debut	date_debut
2	month	month
3	year	year

[Get fields]

[Enter field mapping]

[Help] [OK] [Cancel] [SQL]

3.9 LANCER LA TRANSFORMATION:



Les tables de dimension et la table de faits sont maintenant mises à jour avec les données.

- Table de la dimension temps_dim:

```
SELECT * FROM temps_dim;
```

Query Result

SQL | Fetched 50 rows in 0.024 seconds

	DATE_DEBUT	MONTH	YEAR
1	18-OCT-15	10/2015	2015
2	19-OCT-15	10/2015	2015
3	20-OCT-15	10/2015	2015
4	21-OCT-15	10/2015	2015
5	22-OCT-15	10/2015	2015
6	23-OCT-15	10/2015	2015
7	24-OCT-15	10/2015	2015
8	25-OCT-15	10/2015	2015

- Table de la dimension customer_dim:

```
SELECT * FROM customer_dim;
```

Query Result

SQL | Fetched 50 rows in 0.029 seconds

	CUSTOMERID	CUSTOMERNAME	SEGMENT
1	TS-21505	Tony Sayre	Consumer
2	TS-21610	Troy Staebel	Consumer
3	TS-21655	Trudy Schmidt	Consumer
4	TT-21070	Ted Trevino	Consumer
5	TT-21220	Thomas Thornton	Consumer
6	TT-21265	Tim Taslimi	Corporate
7	TT-21460	Tonja Turnell	Home Office
8	TW-21025	Tamara Willingham	Home Office

- Table de la dimension product_dim:

Worksheet | Query Builder

```
SELECT * FROM product_dim;
```

Query Result

SQL | Fetched 50 rows in 0.022 seconds

	PRODUCTID	SUB_CATEGORY	CATEGORIES	PRODUCTNAME
1	FUR-FU-10003274	Furnishings	Furniture	Regeneration Desk Collection
2	FUR-FU-10003347	Furnishings	Furniture	Coloredge Poster Frame
3	FUR-FU-10003374	Furnishings	Furniture	Electrix Fluorescent Magnifier Lamps & Weighted Base
4	FUR-FU-10003394	Furnishings	Furniture	Tenex "The Solids" Textured Chair Mats
5	FUR-FU-10003424	Furnishings	Furniture	Nu-Dell Oak Frame
6	FUR-FU-10003464	Furnishings	Furniture	Seth Thomas 8 1/2" Cubicle Clock
7	FUR-FU-10003489	Furnishings	Furniture	Contemporary Borderless Frame
8	FUR-FU-10003535	Furnishings	Furniture	Howard Miller Distant Time Traveler Alarm Clock

- Table de la dimension region_dim:

```
SELECT * FROM region_dim;
```

Query Result

SQL | Fetched 50 rows in 0.018 seconds

	CITY	STATE	REGION	COUNTRY
1	Newark	Ohio	East	United States
2	Newport News	Virginia	South	United States
3	Niagara Falls	New York	East	United States
4	Noblesville	Indiana	Central	United States
5	Norfolk	Nebraska	Central	United States
6	Normal	Illinois	Central	United States
7	Norman	Oklahoma	Central	United States
8	North Charleston	South Carolina	South	United States
9	North Las Vegas	Nevada	West	United States

- Table de fait:

Worksheet | Query Builder

```
SELECT * FROM fact_sales;
```

Query Result

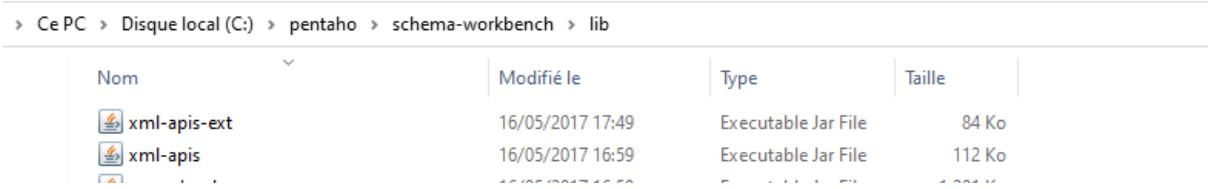
SQL | Fetched 50 rows in 0.136 seconds

	PRODUCTID	CITY	CUSTOMERID	ORDERDATE	SHIPDATE	SALES
1	FUR-BO-10001798	Henderson	CG-12520	08-NOV-17	11-NOV-17	261.96
2	FUR-CH-10000454	Henderson	CG-12520	08-NOV-17	11-NOV-17	731.94
3	OFF-LA-10000240	Los Angeles	DV-13045	12-JUN-17	16-JUN-17	14.62
4	FUR-TA-10000577	Fort Lauderdale	SO-20335	11-OCT-16	18-OCT-16	957.5775
5	OFF-ST-10000760	Fort Lauderdale	SO-20335	11-OCT-16	18-OCT-16	22.368
6	FUR-FU-10001487	Los Angeles	BH-11710	09-JUN-15	14-JUN-15	48.86
7	OFF-AR-10002833	Los Angeles	BH-11710	09-JUN-15	14-JUN-15	7.28
8	TEC-PH-10002275	Los Angeles	BH-11710	09-JUN-15	14-JUN-15	907.152
9	OFF-BI-10003910	Los Angeles	BH-11710	09-JUN-15	14-JUN-15	18.504
10	OFF-AP-10002892	Los Angeles	BH-11710	09-JUN-15	14-JUN-15	114.9
11	FUR-TA-10001539	Los Angeles	BH-11710	09-JUN-15	14-JUN-15	1706.184
12	TEC-PH-10002033	Los Angeles	BH-11710	09-JUN-15	14-JUN-15	911.424
13	OFF-PA-10002365	Concord	AA-10480	15-APR-18	20-APR-18	15.552
14	OFF-BI-10003656	Seattle	IM-15070	05-DEC-17	10-DEC-17	407.976
15	OFF-AP-10002311	Fort Worth	HP-14815	22-NOV-16	26-NOV-16	68.81
16	OFF-BI-10000756	Fort Worth	HP-14815	22-NOV-16	26-NOV-16	2.544
17	OFF-ST-10004186	Madison	PK-19075	11-NOV-15	18-NOV-15	665.88
18	OFF-ST-10000107	West Jordan	AG-10270	13-MAY-15	15-MAY-15	55.5
19	OFF-AR-10003056	San Francisco	ZD-21925	27-AUG-15	01-SEP-15	8.56
20	TEC-PH-10001949	San Francisco	ZD-21925	27-AUG-15	01-SEP-15	213.48
21	OFF-BI-10002215	San Francisco	ZD-21925	27-AUG-15	01-SEP-15	22.72

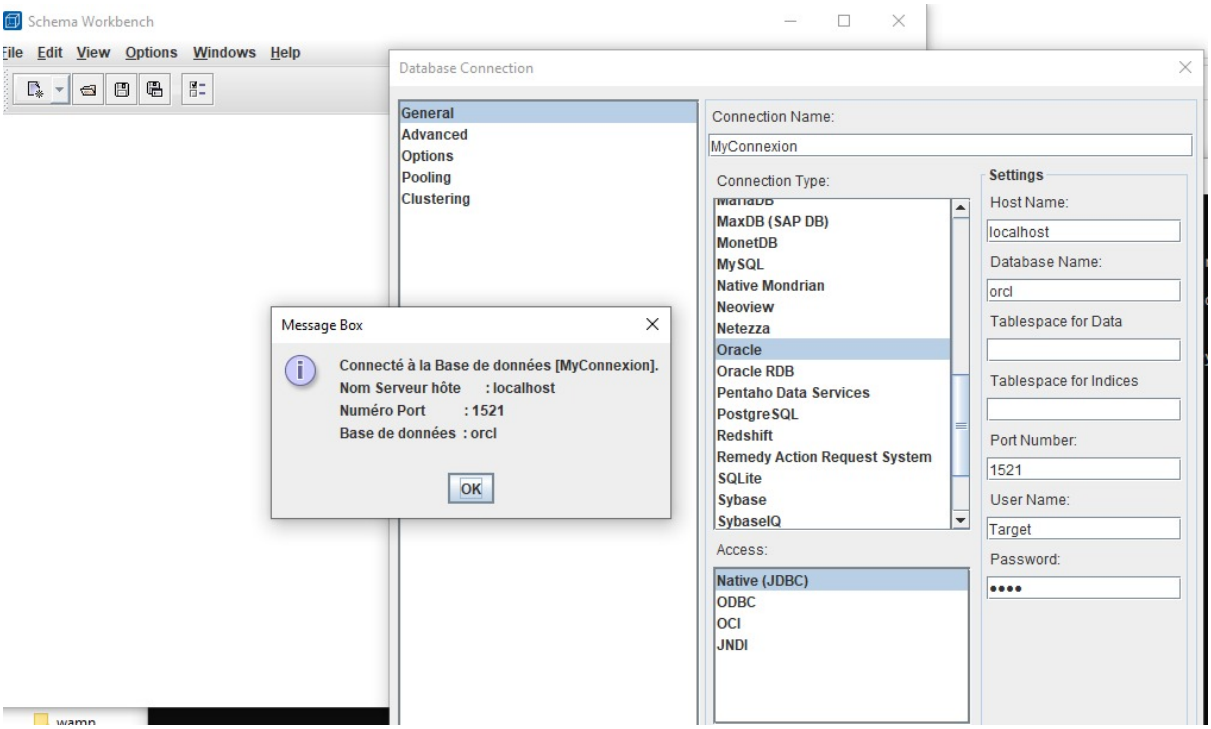
Chapitre 4 :REPRÉSENTATION MULTIDIMENSIONNELLE AVEC SCHEMA WORKBENCH

4-1 CRÉER UNE CONNEXION JDBC DANS PDI

a- Avant d'établir la connexion avec la base de données nouvellement créée, nous devons installer le pilote Oracle pour JDBC on le mettre dans le chemin suivant:



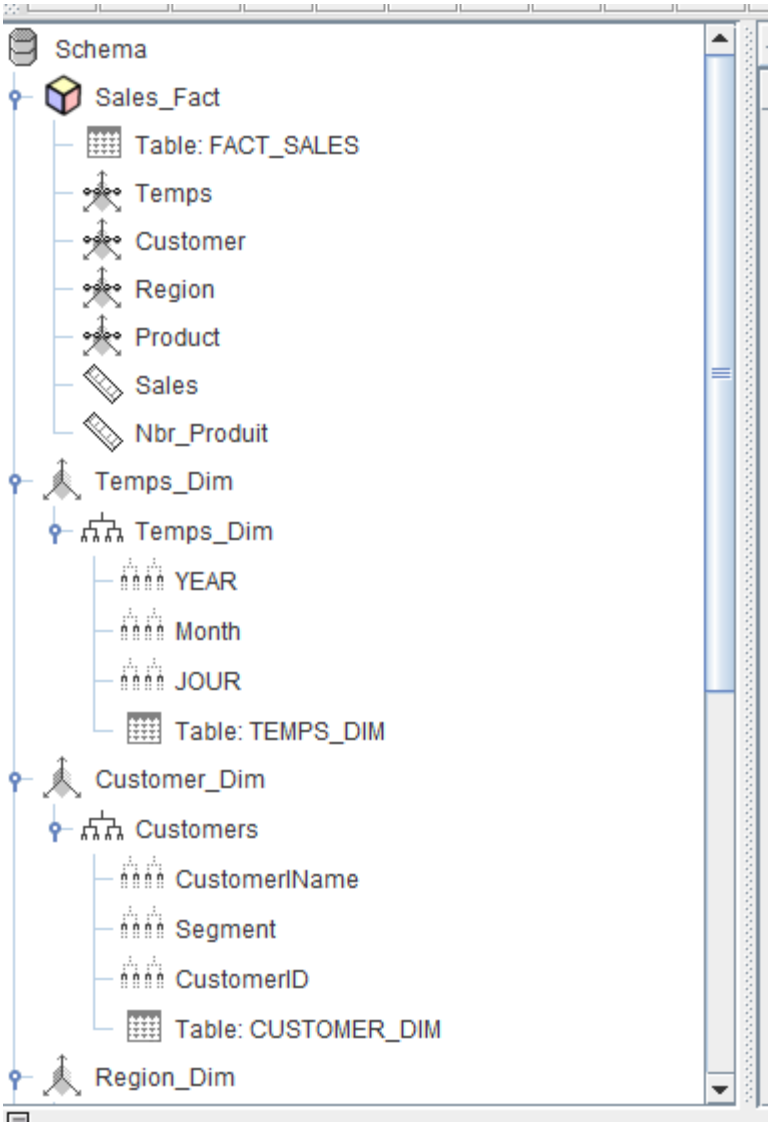
b. Allez dans le logiciel Schema Workbench, cliquez sur Options ! Connection... et remplissez la configuration comme indiqué dans la gure ci-dessous :

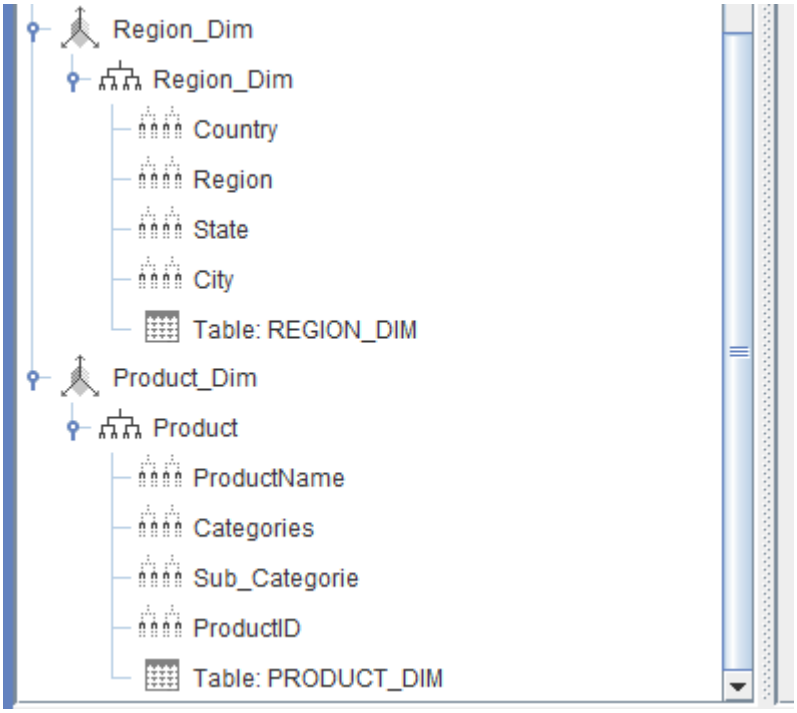


c. Testez la connexion; Cliquez sur OK en cas de succès an de sauvegarder la connexion.

4.2 OLAP CUBE

Un cube OLAP est une base de données multidimensionnelle optimisée pour les Data Warehouses et les applications OLAP. Il s'agit d'une méthode permettant de stocker les données sous forme multidimensionnelle, notamment pour le reporting. En général, ces cubes sont présumés pour accélérer le temps de requête par rapport aux bases de données relationnelles.


Schema
Sales_Fact
Table: FACT_SALES
Temps
Customer
Region
Product
Sales
Nbr_Produit
Temps_Dim
Temps_Dim
YEAR
Month
JOUR
Table: TEMPS_DIM
Customer_Dim
Customers
CustomerIName
Segment
CustomerID
Table: CUSTOMER_DIM
Region_Dim



4.2.1 DIMENSIONS

Code XML: Dimension Temps

Shared Dimension

```
<Dimension type="StandardDimension" visible="true" highCardinality="false" name="Temps_Dim">
  <Hierarchy name="Temps_Dim" visible="true" hasAll="true" primaryKey="DATE_DEBUT">
    <Table name="TEMPS_DIM" schema="TARGET">
    </Table>
    <Level name="YEAR" visible="true" column="YEAR" nameColumn="YEAR" type="String" uniqueMembers="false" levelType="Regular" hideMemberIf="Never">
    </Level>
    <Level name="Month" visible="true" column="MONTH" nameColumn="MONTH" type="String" uniqueMembers="false" levelType="Regular" hideMemberIf="Never">
    </Level>
    <Level name="JOUR" visible="true" column="DATE_DEBUT" nameColumn="DATE_DEBUT" type="Date" uniqueMembers="false" levelType="Regular" hideMemberIf="Never">
    </Level>
  </Hierarchy>
</Dimension>
```

Code XML: Dimension Customer

Shared Dimension

```
<Dimension type="StandardDimension" visible="true" highCardinality="false" name="Customer_Dim">
  <Hierarchy name="Customers" visible="true" hasAll="true" primaryKey="CUSTOMERID">
    <Table name="CUSTOMER_DIM" schema="TARGET">
    </Table>
    <Level name="CustomerIName" visible="true" column="CUSTOMERNAME" nameColumn="CUSTOMERNAME" type="String" uniqueMembers="false" levelType="Regular" hideMemberIf="N
ever">
    </Level>
    <Level name="Segment" visible="true" column="SEGMENT" nameColumn="SEGMENT" type="String" uniqueMembers="false" levelType="Regular" hideMemberIf="Never">
    </Level>
    <Level name="CustomerID" visible="true" column="CUSTOMERID" nameColumn="CUSTOMERID" type="String" uniqueMembers="false" levelType="Regular" hideMemberIf="Never">
    </Level>
  </Hierarchy>
</Dimension>
```

Code XML: Dimension Product

Shared Dimension

```
<Dimension type="StandardDimension" visible="true" highCardinality="false" name="Product_Dim">
  <Hierarchy name="Product" visible="true" hasAll="true" primaryKey="PRODUCTID">
    <Table name="PRODUCT_DIM" schema="TARGET">
    </Table>
    <Level name="ProductName" visible="true" column="PRODUCTNAME" nameColumn="PRODUCTNAME" type="String" uniqueMembers="false" levelType="Regular" hideMemberIf="Never">
    </Level>
    <Level name="Categories" visible="true" column="CATEGORIES" nameColumn="CATEGORIES" type="String" uniqueMembers="false" levelType="Regular" hideMemberIf="Never">
    </Level>
    <Level name="Sub_Categorie" visible="true" column="SUB_CATEGORY" nameColumn="SUB_CATEGORY" type="String" uniqueMembers="false" levelType="Regular" hideMemberIf="Never
">
    </Level>
    <Level name="ProductID" visible="true" column="PRODUCTID" nameColumn="PRODUCTID" type="String" uniqueMembers="false" levelType="Regular" hideMemberIf="Never">
    </Level>
  </Hierarchy>
</Dimension>
```

Code XML: Dimension de fait

Cube

```
<Cube name="Sales_Fact" visible="true" cache="true" enabled="true">
  <Table name="FACT_SALES" schema="TARGET">
  </Table>
  <DimensionUsage source="Temps_Dim" name="Temps" visible="true" foreignKey="ORDERDATE" highCardinality="false">
  </DimensionUsage>
  <DimensionUsage source="Customer_Dim" name="Customer" visible="true" foreignKey="CUSTOMERID" highCardinality="false">
  </DimensionUsage>
  <DimensionUsage source="Region_Dim" name="Region" visible="true" foreignKey="CITY" highCardinality="false">
  </DimensionUsage>
  <DimensionUsage source="Product_Dim" name="Product" visible="true" foreignKey="PRODUCTID" highCardinality="false">
  </DimensionUsage>
  <Measure name="Sales" column="SALES" datatype="Numeric" aggregator="sum" visible="true">
  </Measure>
  <Measure name="Nbr_Produit" column="PRODUCTID" datatype="Integer" aggregator="count" visible="true">
  </Measure>
</Cube>
```

4.2.2 MESURES

```
<Measure name="Sales" column="SALES" datatype="Numeric" aggregator="sum" visible="true">
</Measure>
```

Chapitre 5: VISUALISATION Power BI

5.1 Définition sur le POWER BI :

Microsoft Power BI est une suite de solutions de Business Intelligence (BI), développée par Microsoft et intégrée à Office 365. Elle permet la création de **rapports et d'outils de visualisation interactifs et dynamiques** tout en offrant **des fonctionnalités de BI puissantes et avancées** permettant de répondre à toutes les problématiques, quelle que soit la taille d'entreprise et le secteur d'activité.

L'écosystème Power BI s'articule autour des composants suivants :

• Power BI Desktop : C'est l'application Desktop de Power BI. Elle est utilisée pour concevoir le modèle de données, ainsi que les rapports et les publier sur Power BI Service. Anciennement appelé Power BI Designer.

• Power BI Service : C'est la partie cloud de Power BI (en mode "SaaS") utilisée principalement pour la visualisation des rapports ainsi que pour la gestion du partage et la diffusion des rapports au sein et à l'extérieur de l'organisation.

• Power BI Mobile Apps : Ce sont les applications mobiles et tablettes pour Android et IOS de Power BI. Elles sont utilisées pour visualiser les rapports depuis un périphérique mobile.
• Power BI Gateway : Les passerelles Power BI permettent une actualisation des données stockées en locales depuis Power BI Service, sans procéder par une actualisation dite "manuelle"

• Power VI Embedded : Il s'agit de l'API REST mise à disposition des développeurs pour intégrer Power BI dans leurs propres solutions.

• Power BI Repot Server : C’est le pendant “On-Premises” de Power BI Service. Cette solution donne la possibilité aux organisations d’héberger leur propre instance Power BI sur leurs serveurs.

• Power BI Visuals Marketplace : Il s’agit d’un magasin d’applications où les utilisateurs peuvent télécharger de nouveaux visuels basés sur le langage R pour leurs rapports Power BI.

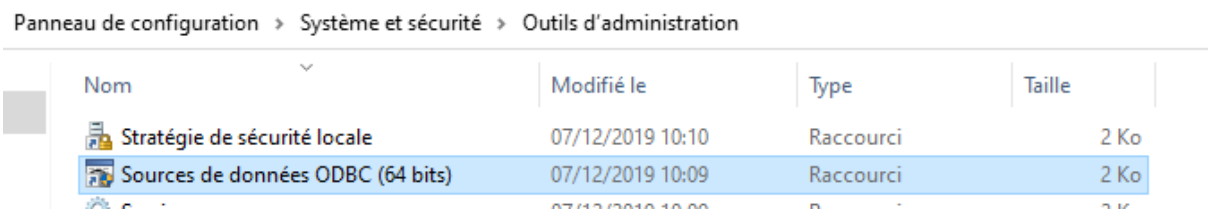
5.2 Les Points Forts de Power BI :

SIMPLE	COMPLÈTE	INTÉGRÉE
• Interface familière • Simplicité d’utilisation • Rapidité d’implémentation • Plus grande autonomie des équipes	• Toutes les fonctionnalités attendues d’une solution de BI moderne • Des tableaux de bords visuels et percutants	• Intégration avec les autres produits Microsoft • Facilité du partage et de diffusion de l’information
Efforts et coûts minimisés	**Exploitation simplifiée de toutes les données d’entreprise**	**Aide à la décision facilitée**

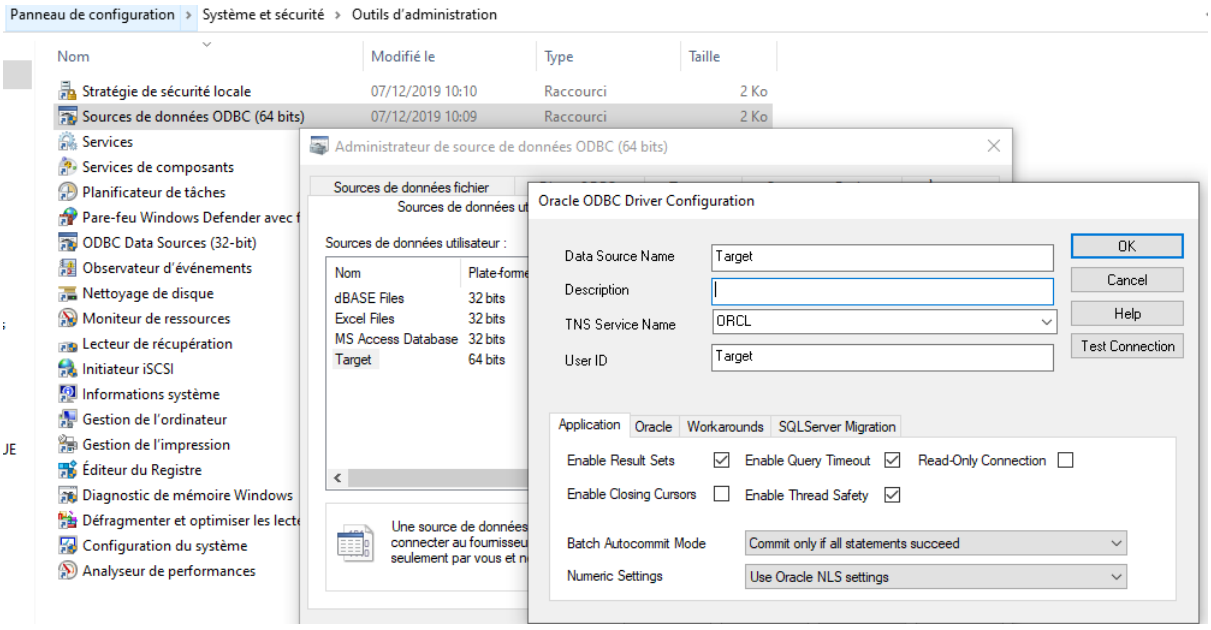
5.3.Les étapes de visualisation

5.3.1 Création de la connexion :

La configuration du pilote ODBC Oracle .Nous avons choisi 64bit car notre machine est 64bit (power BI 64bit)

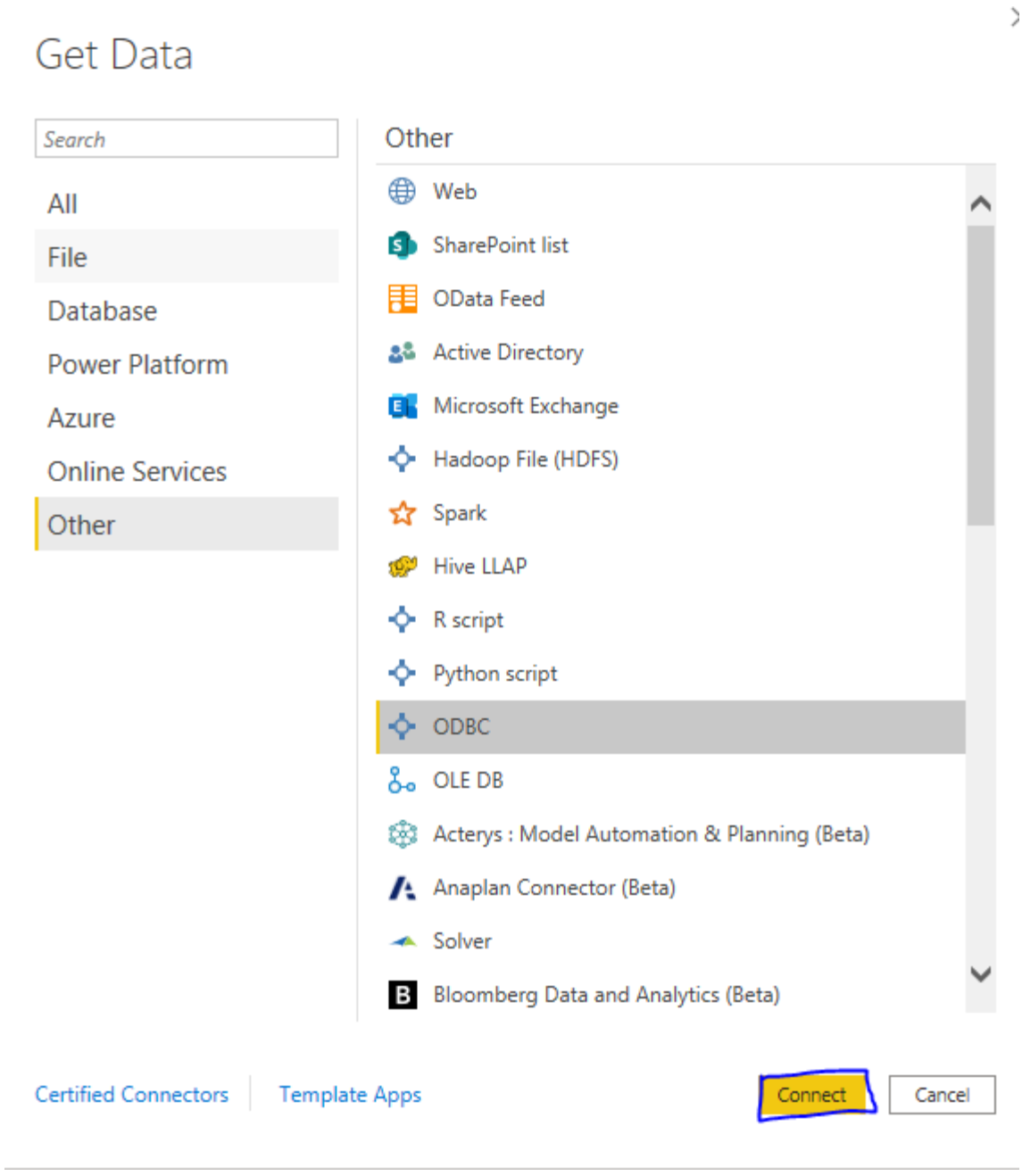


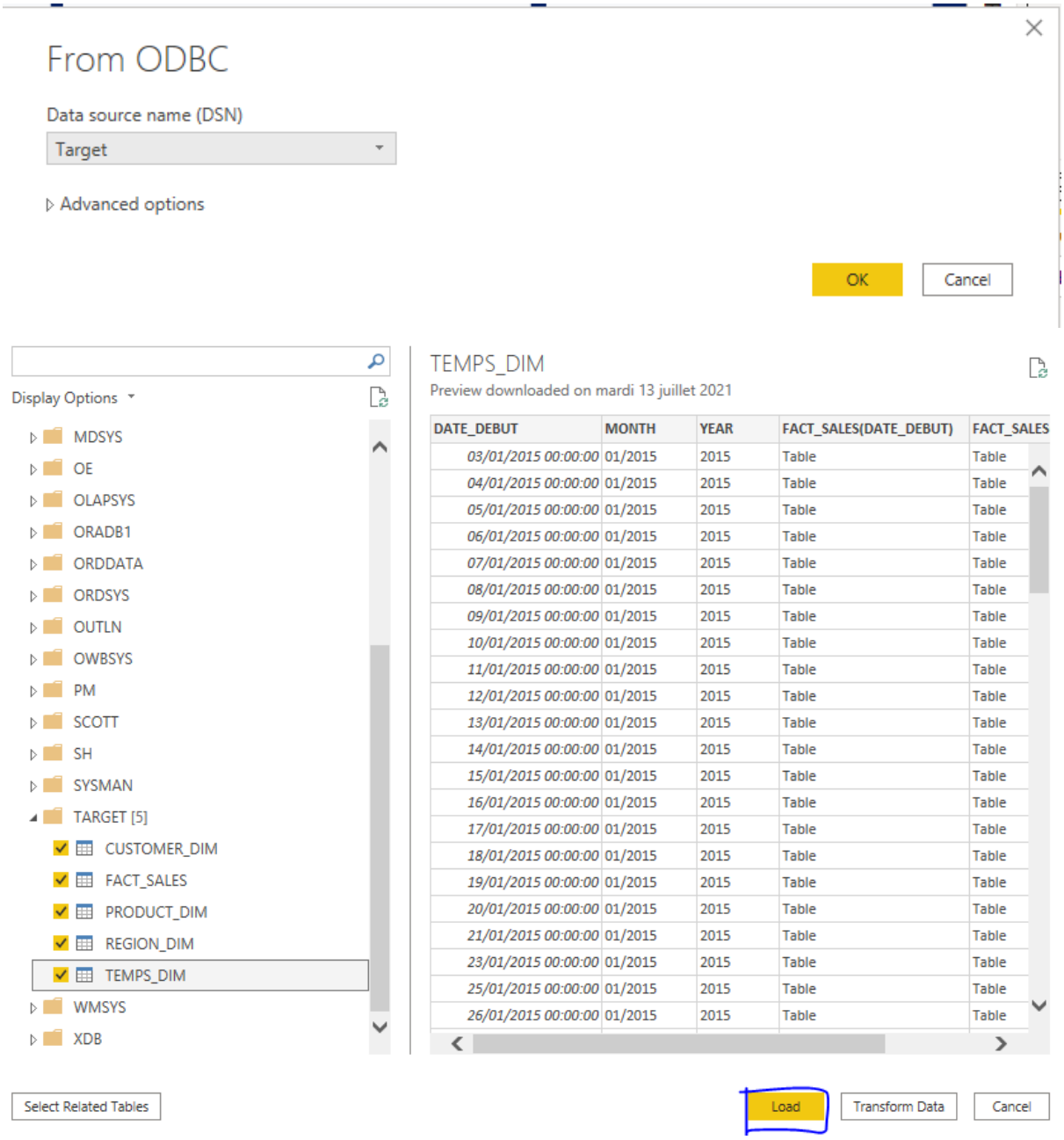
Saisissez les informations de connexion et cliquez sur Tester la connexion :



5.3.2 importer le schéma dans la base de données d'utilisateur target

Tester la connexion et vérifier que la source de données JDBC se trouve dans la fenêtre. en cliquant sur get data ⇒ other⇒ ODBC et nous sélectionnons toutes les tables: temps_dim, customer_Dim ,Product_Dim, region_dim et la table de fait. En suite on clique sur load.


Get Data
Search
All
File
Database
Power Platform
Azure
Online Services
Other
Other
Web
SharePoint list
OData Feed
Active Directory
Microsoft Exchange
Hadoop File (HDFS)
Spark
Hive LLAP
R script
Python script
ODBC
OLE DB
Acterys : Model Automation & Planning (Beta)
Anaplan Connector (Beta)
Solver
Bloomberg Data and Analytics (Beta)
Certified Connectors
Template Apps
Connect
Cancel



Après l’importation des tables, nous choisissons les diagrammes dont nous avons besoin , après nous sélectionnons les mesures, dans notre cas Sales par région par temps par client et par produit.

Cette interface permet à l'entreprise de voir un tableau détaillée du sales (qui pourra l'exporter dans un fichier Excel) ainsi que d'autre graphe variant par la date , la région et le pays et par client.

Chapitre 6: Conclusion

La business intelligence est devenue un sujet incontournable et au cœur des préoccupations des DSI dans leur impact positif sur la performance de l’entreprise. Elle s’apparente plus à un processus continu au sein des entreprises, cadré par une stratégie, qu’à un projet ponctuel.

S’adressant aussi bien aux Directions Générales qu’aux Métiers, la BI, pour être efficace, nécessite de la maturité dans les relations DSI-Métiers, en demandant notamment que DSI et Métiers se considèrent réciproquement comme des partenaires et construisent ensemble sur leurs différences.

Aujourd’hui, les principaux enjeux pour la réussite des projets BI résident dans la capacité de ces acteurs à consolider et structurer l’information en ligne avec la stratégie de l’entreprise et sans problème de performance du fait de volumes en augmentation constante, et à réussir l’accompagnement du changement :

• Assurer l’appropriation des outils par les utilisateurs et une utilisation partagée des standards à partir de reports ad hoc ;
• Offrir un espace de liberté aux utilisateurs, à côté des solutions existantes, pour permettre la montée en puissance de l'utilisation des reports standards ;
• Instaurer un climat de confiance entre la DSI et les utilisateurs ;
• Nommer un process owner sur chaque projet ;
• Définir et suivre les indicateurs clés ;
• Etc.

Cependant, les solutions actuelles, à l’intégration pas toujours finalisée et à la roadmap souvent floue, doivent évoluer pour aider l’entreprise à mieux répondre aux besoins des utilisateurs et des décideurs. La démocratisation de l’information doit passer par des interfaces web plus accessibles, interactives et collaboratives, ainsi que par des outils de simulation et d’analyse prospective. Ces perspectives d’évolution constituent le fer de lance de la nouvelle génération d’outils BI. Cette tendance BI 2.0 traduit l’idée « d’intelligence » et d’interactivité qu’il faudra apporter aux futurs utilisateurs.