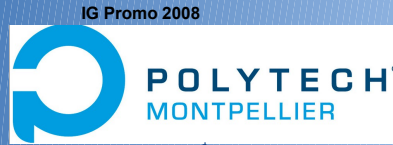


# INFORMATIQUE DÉCISIONNELLE OPÉRATIONNELLE



# Présentation



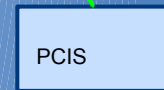
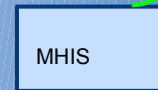
Développement & Conception  
Chargé de 2 projets BI



Expert Technique BI  
Chef de Projet BI

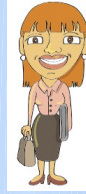


Chef de Projet BI  
Réfèrent décisionnel



## 1 - EXEMPLE CONCRET





***La directrice commerciale en Chef de Audi France aimerait connaître le nombre de ventes du modèle “Audi TT 180cv” par mois pour l’année 2013.***

***Comment doit-elle s’y prendre ?***

Niveaux  
Intermédiaires

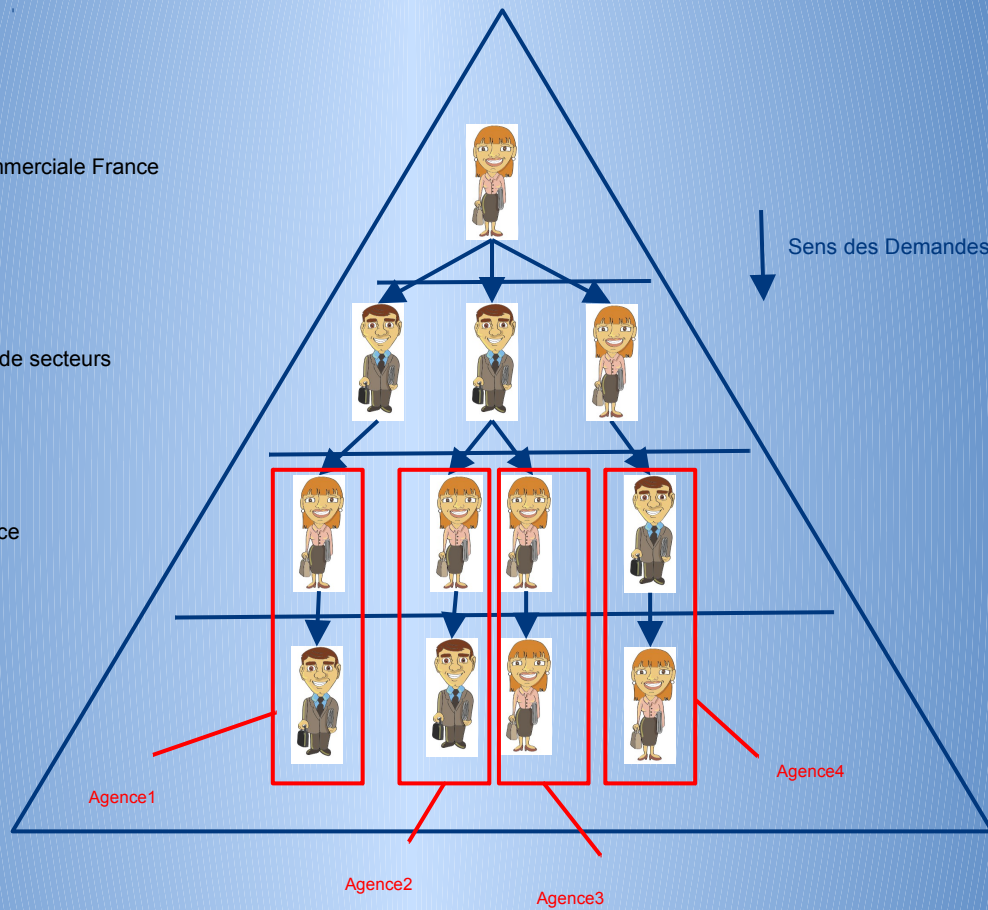
Directrice commerciale France

Directeurs de secteurs

Directeurs d'agence

Comptables

Sens des Demandes d'informations pour Analyses



**AGENCE 1** utilise un logiciel de Gestion maintenu par un prestataire externe



Requête SQL



	Mois2013	NombreDeVentes
1	AVRIL	3668
2	FEVRIER	1520
3	JANVIER	1286
4	MARS	1337

**AGENCE 2** utilise un ERP qui permet une extraction au format suivant



*Fichier txt*

Modele;Annee;Mois;NbVentes

50VBG26;2013;01;2150

50ERD96;2013;01;2362

50BGG45;2013;01;1522

...

**AGENCE 3** gère manuellement sa comptabilité via Excel

Modele	Libelle	Date Achat	Toute option
50ERD96	Audi TT 180cv	03/01/2013	Oui
50VBG26	Audio TT 190cv	05/01/2013	Non
50VBG26	Audio TT 190cv	08/01/2013	Oui
50ERD96	Audi TT 180cv	08/01/2013	Oui
50ERD96	Audi TT 180cv	10/01/2013	Non
50BGG45	Audio TT 225cv	11/01/2013	Oui
50ERD96	Audi TT 180cv	13/01/2013	Oui
50ERD96	Audi TT 180cv	13/01/2013	Non
50ERD96	Audi TT 180cv	13/01/2013	Oui
50BGG45	Audio TT 225cv	14/01/2013	Non
50ERD96	Audi TT 180cv	15/01/2013	Oui
50VBG26	Audio TT 190cv	20/01/2013	Non

sp Semaine 1 Semaine 2 Semaine 3 Semaine 4

**AGENCE 4** partagent un Document Excel sur Internet avec des agrégations pré-calculées



Date	Sommes des Ventes
janv-13	1500
févr-13	1236
mars-13	2300
avr-13	3650
mai-13	2512
juin-13	6321

	Mois2013	NombreDeVentes
1	AVRIL	3668
2	FEVRIER	1520
3	JANVIER	1286
4	MARS	1337



Vérifier que le nombre de Ventes correspond bien au bon modèle attendu

Modele;Annee;Mois;NbVentes

50VBG26;2013;01;2150

50ERD96;2013;01;2362

50BGG45;2013;01;1522

...



Transformer '01' en 'Janvier', '02' en Février...

Ne conserver que le modèle 50ERD96

Modele	Libelle	Date Achat	Toute option
50ERD96	Audi TT 180cv	03/01/2013	Oui
50VBG26	Audio TT 190cv	05/01/2013	Non
50VBG26	Audio TT 190cv	08/01/2013	Oui
50ERD96	Audi TT 180cv	08/01/2013	Oui
50ERD96	Audi TT 180cv	10/01/2013	Non
50BGG45	Audio TT 225cv	11/01/2013	Oui
50ERD96	Audi TT 180cv	13/01/2013	Oui
50ERD96	Audi TT 180cv	13/01/2013	Non
50ERD96	Audi TT 180cv	13/01/2013	Oui
50BGG45	Audio TT 225cv	14/01/2013	Non
50ERD96	Audi TT 180cv	15/01/2013	Oui
50VBG26	Audio TT 190cv	20/01/2013	Non

SP Semaine 1 Semaine 2 Semaine 3 Semaine 4



Compter le nombre de lignes du modèle 50ERD96 en fonction du mois de la date d'achat

Faire une somme du nombre de ventes par mois

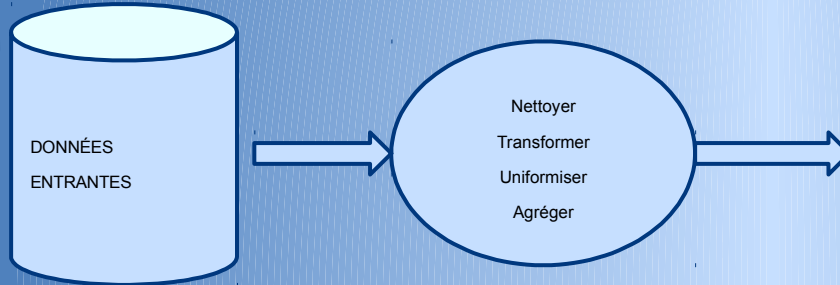
Date	Sommes des Ventes
janv-13	1500
févr-13	1236
mars-13	2300
avr-13	3650
mai-13	2512
juin-13	6321



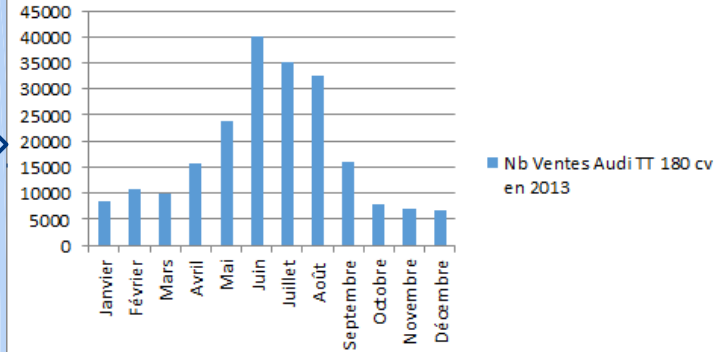
Vérifier que le nombre de Ventes correspond bien au bon modèle attendu



Niveaux Intermédiaires



## Nb Ventes Audi TT 180 cv en 2013



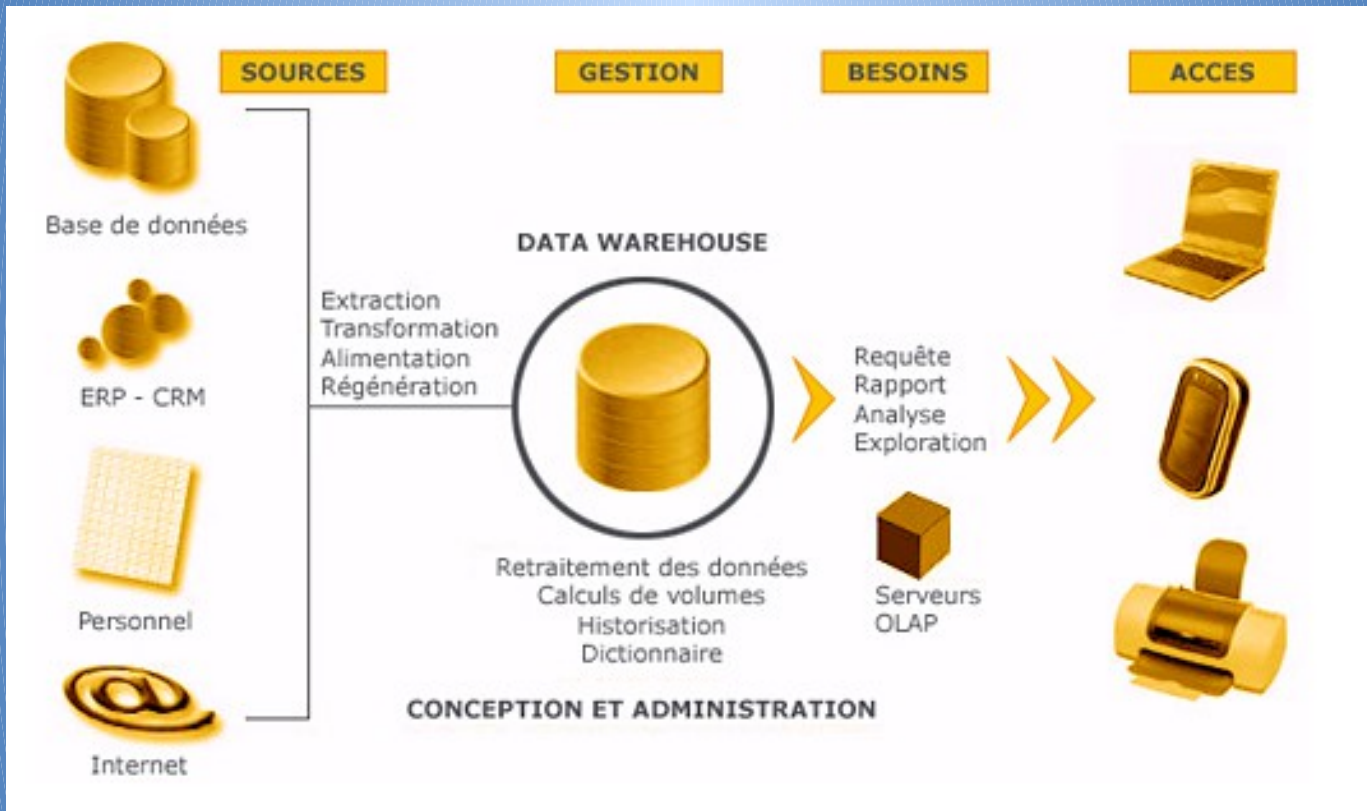
Directrice commerciale France

Exemples d'analyses :

- Les mois "été" Juin / Juillet / Août sont les mois clés des ventes du modèle
- La campagne marketing de 2013 a été lancée en avril avec de bons résultats M+2 M+3 et M+4 (ROI confirmé)

Exemples de décisions pour ce modèle :

- Lancer une pré campagne marketing de 3M€ en janvier 2014 en plus de la campagne officielle de mars 2014 pour booster les ventes de début d'année
- Lancer une campagne d'achat simplifiée pour booster les ventes de fin d'année (au milieu des impôts)



## 2 - INFORMATIQUE DÉCISIONNELLE

(ou BI = Business Intelligence)



C'est l'ensemble des concepts, des méthodes et des technologies permettant de traiter et de présenter les données d'une entreprise à des fins d'analyse et de décision.

Aujourd'hui, les entreprises produisent et stockent d'énormes quantités de données. Ces données, qui constituent un extraordinaire gisement d'informations sur ses activités et son positionnement dans le contexte économique, sont souvent insuffisamment exploitées.

L'informatique décisionnelle permet la collecte, l'organisation et le stockage des données ainsi que l'extraction et la restitution des informations pertinentes.

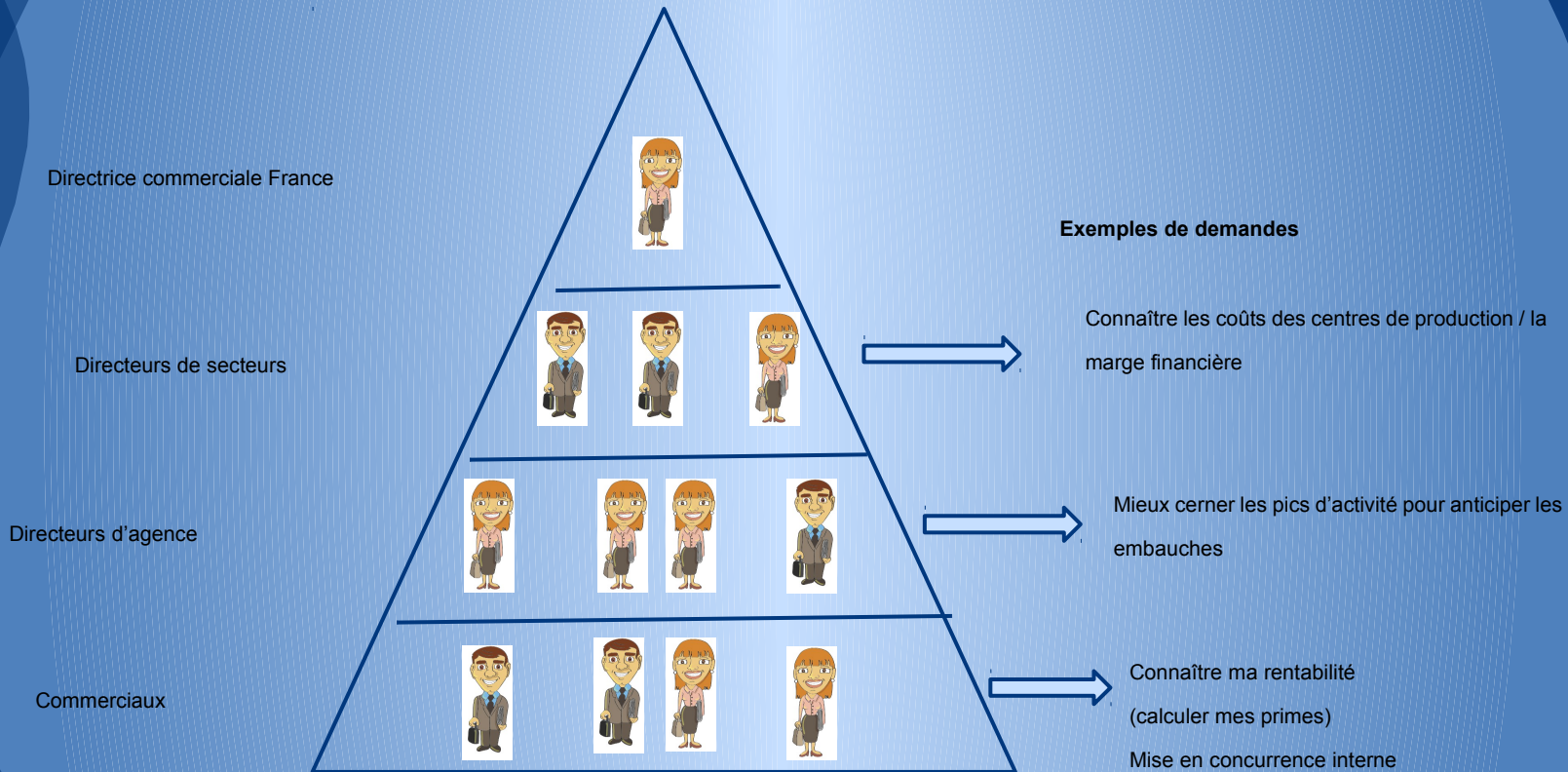
**Elle permet aux responsables de la stratégie d'une entreprise d'avoir une vision synthétique de leurs activités, les aidant ainsi à orienter leurs décisions.**

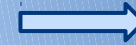
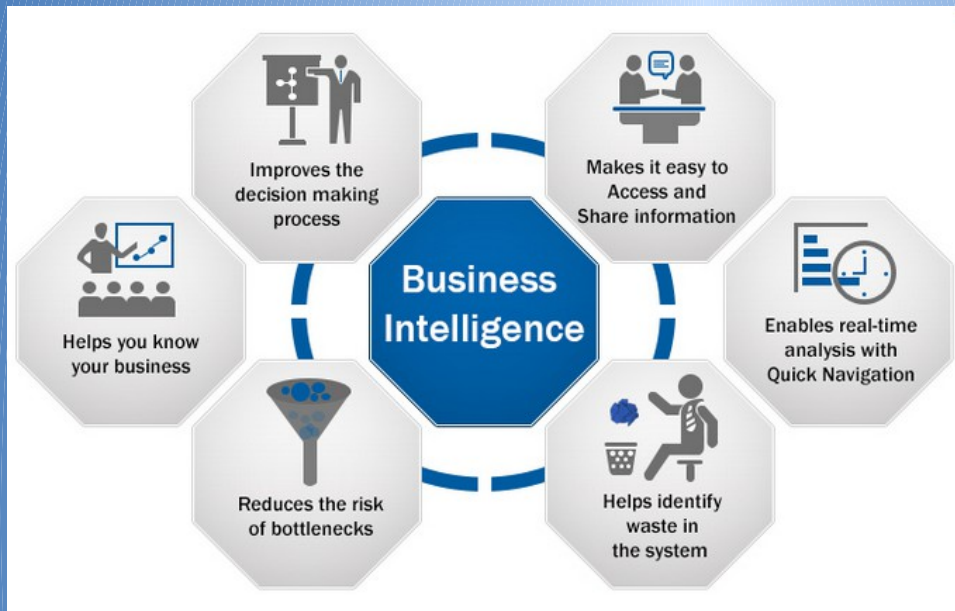
**Regrouper un maximum d'informations pertinentes sous une forme analysable en vue d'en donner du SENS**



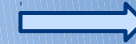
**Entreprises en forte concurrence continue**

- **Besoin de suivre l'évolution du marché**
- **Besoin de réactivité dans les décisions**
- **Besoin de gérer les activités et les coûts liés**
- **Besoin de prévoir l'avenir en analysant le passé**
- ...





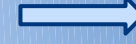
Politiques



Commerciaux



Financiers



Opérationnels



Décideurs au sens large

**Améliorer la prise de décision:** Le maire aimerait connaître les subventions accordées aux associations l'année dernière pour valider / rejeter les nouvelles demandes de subventions des associations cette année

**Evaluer les risques:** Les contrôleurs de gestion de PCIS aimeraient évaluer le risque sur non paiement pour le mois de décembre 2014

**Trouver les déchets:** Le directeur opérationnel de l'usine à Aimargues aimerait analyser les temps des différentes tâches sur sa chaîne de production afin de les comparer avec l'usine Dmitrov (Russie)

...

### 3 - DÉFINITION DU DATAWAREHOUSE





**Définition : Un entrepôt de données est une collection de données thématiques, intégrées, non volatiles et historisées pour la prise de décisions (Bill Inmon)**

Le concept d'entrepôt de données a été formalisé pour la première fois en 1990 par Bill Inmon. Il s'agissait de constituer une base de données **orientée sujet, intégrée et contenant des informations historisées, non volatiles et exclusivement destinées aux processus d'aide à la décision.**

En effet, la simple logique de production (produire pour répondre à une demande) ne suffit plus pour pérenniser l'activité d'une entreprise. Elle est un système ouvert sur son environnement au coeur des systèmes d'informations confrontée à des phénomènes économiques et sociaux lourds de conséquences.

Pour faire face aux nouveaux enjeux, l'entreprise doit collecter, traiter, analyser les informations de son environnement pour anticiper. Mais cette information produite par l'entreprise est **surabondante, non organisée et éparpillée** dans de multiples systèmes opérationnels hétérogènes et peut provenir de toutes les places de marchés (mondialisation des échanges).

Il devient fondamental de rassembler et d'homogénéiser les données afin de permettre l'analyse des **indicateurs** pertinents pour faciliter la **prise de décisions**. L'objet de l'entrepôt de données est de définir et d'intégrer une architecture qui serve de fondation aux applications décisionnelles.

### **Données orientées sujet**

L'entrepôt de données est organisé autour des sujets majeurs et des métiers de l'entreprise. Les données sont organisées par thème, contrairement aux données des systèmes de production, organisées par processus fonctionnels.

### **Données intégrées**

Un Entrepôt de données est un projet d'entreprise et concerne les différents services et métiers de l'entreprise. L'intégration de données, au sein d'un entrepôt de données, est donc un processus déterminant sur la qualité et la quantité d'informations disponibles aux utilisateurs pour le processus de décision.

### **Données historisées**

L'historisation est nécessaire pour suivre dans le temps l'évolution des différentes valeurs des indicateurs à analyser. Ainsi, un référentiel temps doit être associé aux données afin de permettre l'identification dans la durée de valeurs précises.

### **Données non volatiles**

Afin de conserver la traçabilité des informations et des décisions prises, les informations stockées au sein de l'entrepôt de données ne peuvent être supprimées.

Un entrepôt de données peut se structurer en quatre classes de données organisées selon un axe historique et un axe de synthèse.

#### Les données agrégées

Les données agrégées correspondent à des éléments d'analyse représentant les besoins des utilisateurs. Elles constituent déjà un résultat d'analyse et une synthèse de l'information contenue dans le système décisionnel, et doivent être facilement accessibles et compréhensibles.

#### Les données détaillées

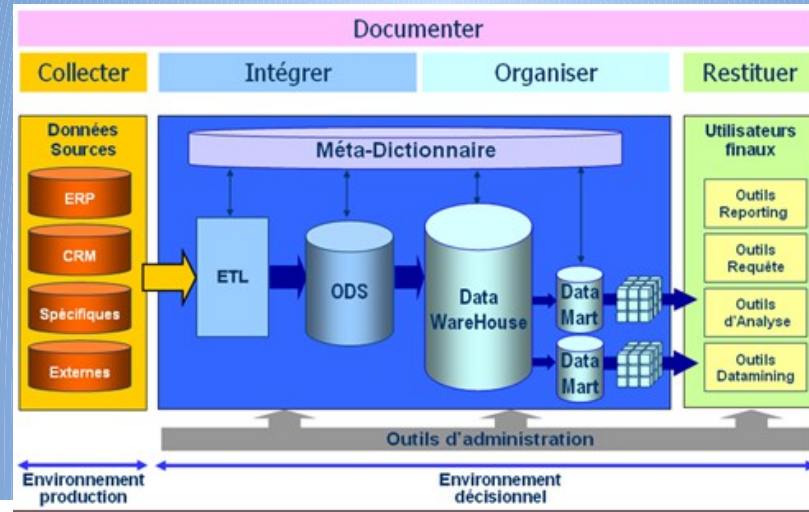
Les données détaillées reflètent les événements les plus récents. Les intégrations régulières des données issues des systèmes de production vont habituellement être réalisées à ce niveau.

#### Les métadonnées

Les métadonnées constituent l'ensemble des données qui décrivent des règles ou processus attachés à d'autres données. Ces dernières constituent la finalité du système d'information.

#### Les données historisées

Chaque nouvelle insertion de données provenant du système de production ne détruit pas les anciennes valeurs, mais crée une nouvelle occurrence de la donnée.



L'infrastructure technique mise en œuvre est capable d'intégrer, d'organiser, de stocker et de coordonner de manière intelligente des données produites au sein du Système d'Information (issues des applications de production) ou importées depuis l'extérieur du SI (louées ou achetées) dans lesquelles les utilisateurs finaux puisent des informations pertinentes à l'aide d'outils de restitution et d'analyse (OLAP<sup>2</sup>, Datamining<sup>3</sup>).

Les points clefs garantissant le succès d'un entrepôt de données sont les suivants :

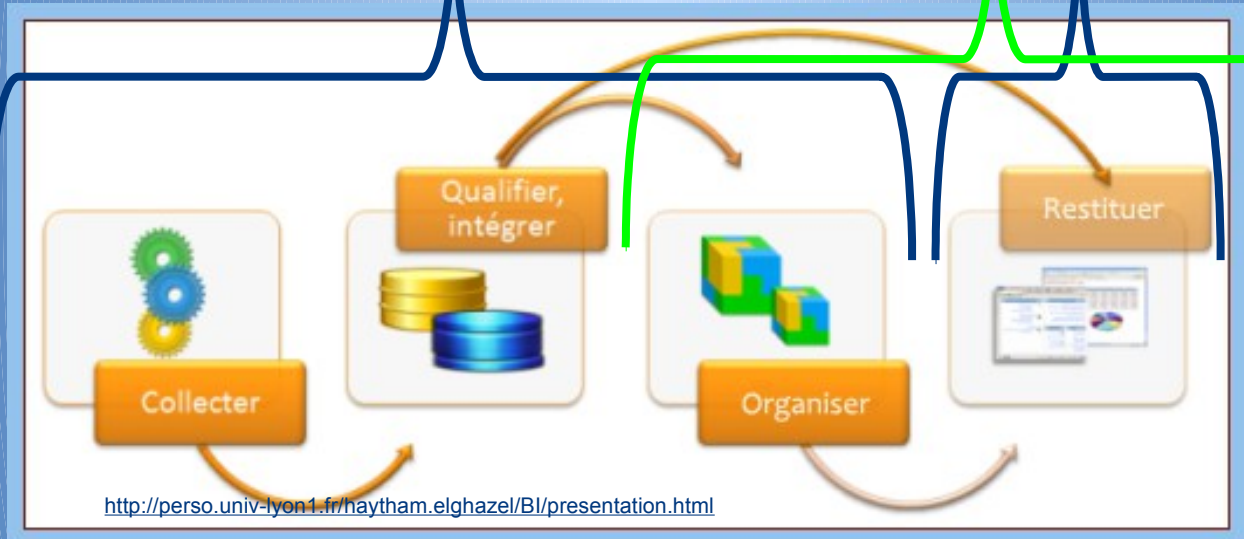
- Les informations d'un entrepôt de données doivent être accessibles et fiables (de qualité).
- La conception d'un entrepôt de données doit répondre à un besoin de ROI<sup>4</sup> élevé.
- La réponse aux demandes très diverses des utilisateurs.
- L'entrepôt de données doit évoluer avec les besoins des utilisateurs et du système d'information.

## 4 - LES OUTILS DU DÉCISIONNEL

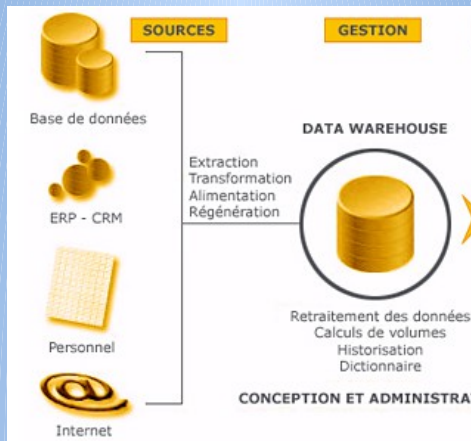


ETL

OUTIL DE RESTITUTION



## ETL



**Extract :** Accéder à la majorité des systèmes de stockage de données (SGBD, ERP, fichiers à plat...) afin de récupérer des données identifiées et sélectionnées. Prendre en compte les questions de synchronisation et de périodicité des rafraichissements.

**Transform :** Toutes les données ne sont pas utilisables telles quelles. Elle méritent d'être vérifiées, reformatées, nettoyées afin d'éliminer les valeurs aberrantes et les doublons et consolidées.

**Load :** Insérer les données dans le DataWarehouse. Elles sont ensuite disponibles pour les différents outils d'analyse et de présentation

**\*\* Problématique de multiples sources entrantes avec des données hétérogènes**

**\*\* Problématique de synchronisation à des bases entrantes "en production" potentiellement en cours de chargement**

**\*\* Problématique des modes de chargement des tables du DW**

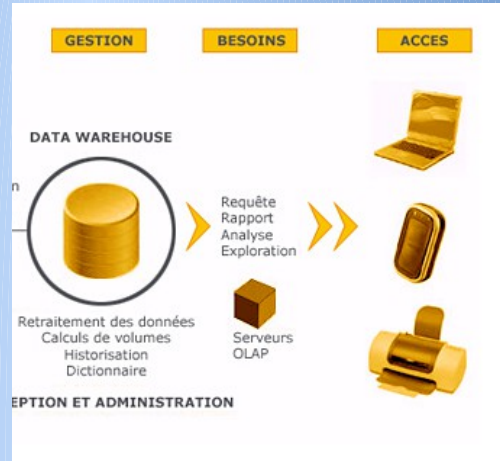
	Informatica Powercenter	IBM DataStage	Microsoft SSIS	Pervasive Data Integrator	Talend Integration Suite	Pentaho Kettle Enterprise	CloverETL
Easy to Use					●	●	●
Web-based UI					●		
Multi-role Team Collaboration	●	●			●		
Process Centric Approach							
Enables SOA							
Reusable Services Repository							
Single-install, Light Footprint				●	●		●
Large Volume Performance	●	●	●	●	●	●	●
Clustering and Job Distribution	●	●			●	●	●
Data Partitioning	●	●		●	●		●
Automatic Recovery of Flows	●	●					
XA-Transaction Rollbacks	●	●					
Meta-Driven Approach vs.Code	●					●	●
Built-in Scheduler	●	●			●		
Real-time Triggers	●	●					
Non-RDBMS Connections	●	●		●	●	●	●
Web Services Client	●	●		●	●	●	
Publish Flows as Web Services	●						
Join Multiple Sources	●	●	●	●	●	●	●
Split Data Streams	●	●	●	●	●	●	●
Graphical Data Mapper	●	●	●	●	●		●
Complex Transformations	●	●	●	●	●	●	●
Data Validations	●	●	●	●	●	●	●
Preview Source Data In-Design	●	●		●	●	●	●
Run Mapping Rules In-Design	●	●			●	●	●
Library of New Mapping Functions	●	●					
Complex Lookups	●	●	●	●	●	●	●
Automatic Documentation	●	●					
Human Workflow for Error Handling							
Plug-in External Programs	●	●		●			
Alerts and Notifications	●	●		●	●	●	●
Version Control	●	●			●		
Deploy in Cloud Option				●		●	
Commerical or Open Source	C	C	C	C	OS	OS	OS

## Principaux ETL :

**Datastage / Informatica / Talend**



## OUTILS DE RESTITUTION



**Les outils de Restitution ne stockent pas les données.** Ils créent une couche "vue logique" sur les données de l'entrepôt

Principaux outils : Business Object / IBM Cognos

Différents types d'utilisateurs :

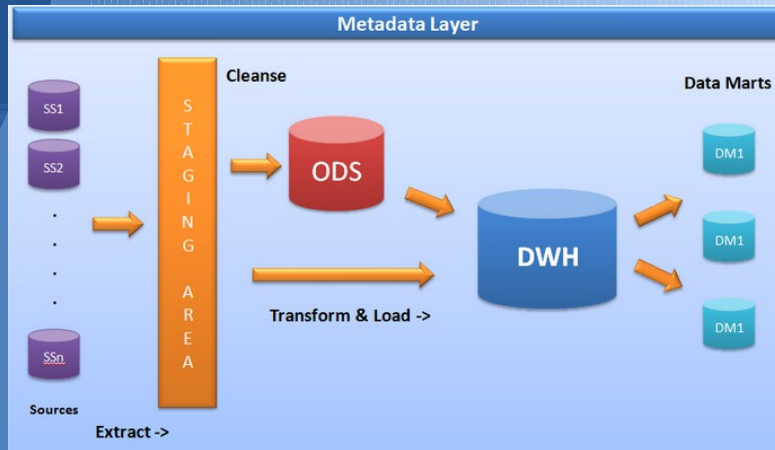
- Utilisateurs "Read Only" qui analysent les rapports
- Utilisateurs créateurs de rapports
- Utilisateurs analystes

\*\* Problématique de sécurité d'accès aux données

\*\*Problématique d'accès aux données lors du chargement de l'entrepôt

## 5 - ALIMENTATION DU DTW





### Chargement Tampon (ODS)

- > Tables ODS à l'identique du format des sources entrantes
- > Prendre le format Varchar par défaut

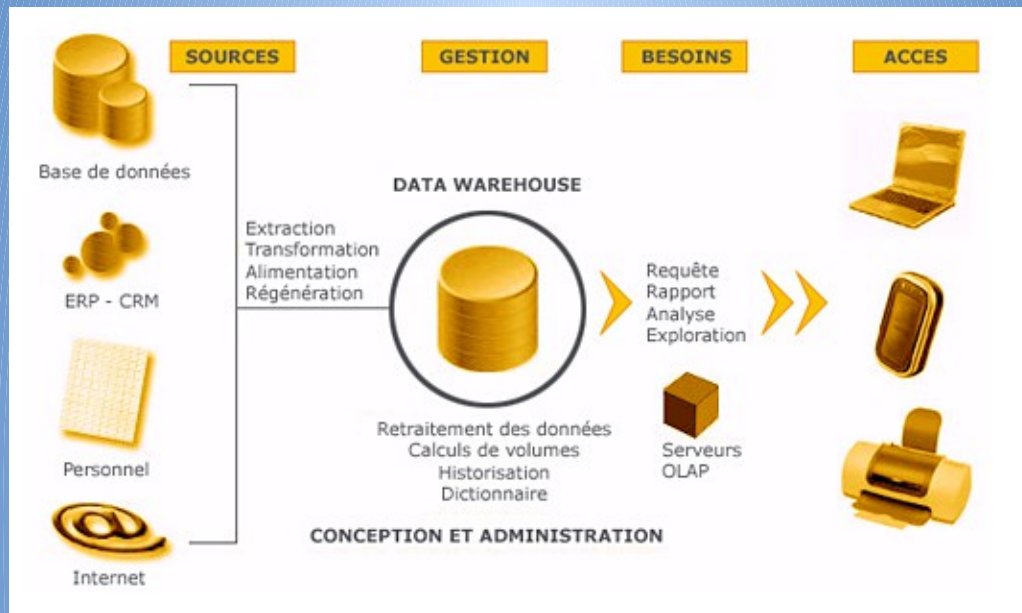
### Chargement Tables Élémentaires (DWH)

Modèle de données généralement en 3NF (normalisation forte)

### Chargement Datamarts / Tables agrégées (DTM)

Modèle en étoile / flocon

Tables agrégées

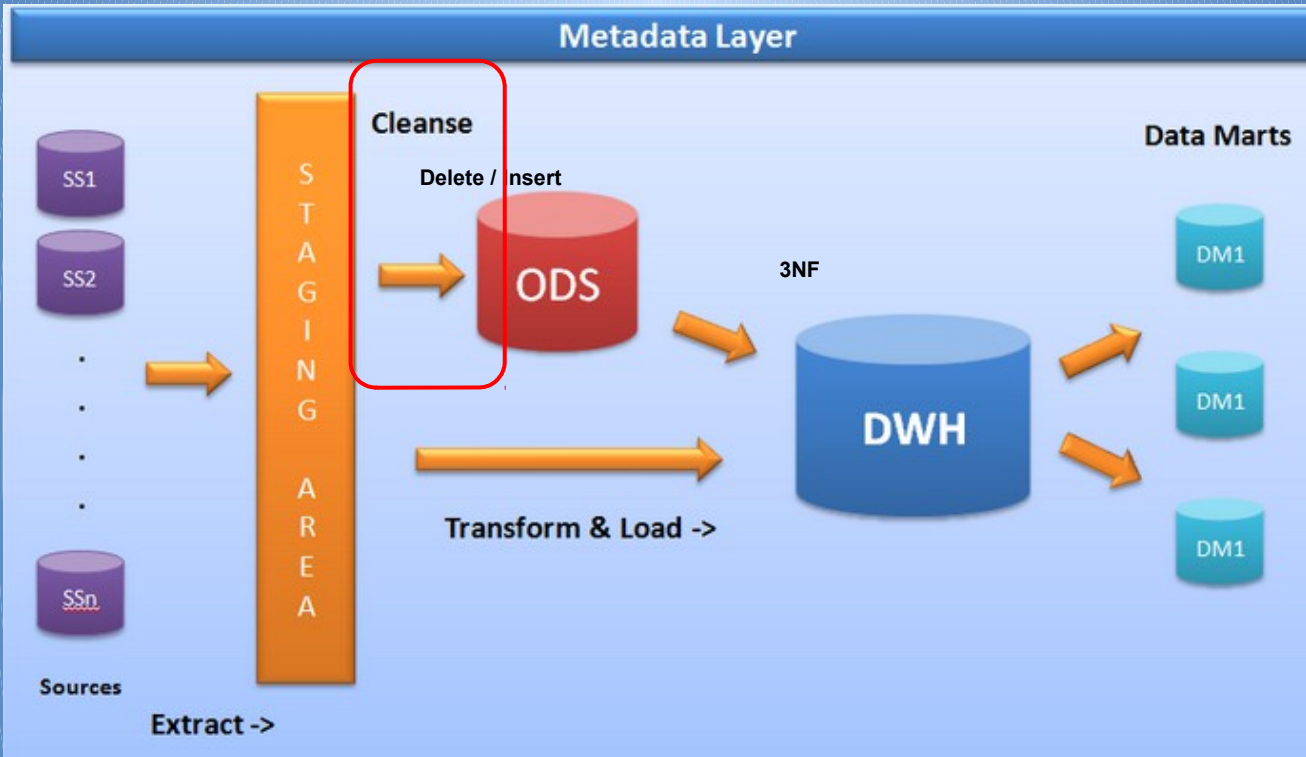


**\*\*Les sources sont souvent des applications opérationnelles(en Production)**

**\*\* Le système décisionnel est un système“Automatique”, les traitements y sont “Ordonnés”**

**\*\* Le système décisionnel doit permettre deschargement en continutout en garantissant laqualité des données**

**\*\* Le systèmene doit pas se substituer à une application opérationnellemais rester une application “intelligente” restititrice**



# Chargement Sources -> ODS

## Fichier Doublons.csv Source ASSO

A
Maitre,"Doublon"
610,"4762"
1001,"44"
1001,"1256"
1001,"1342"
1001,"4987"
1001,"3451"
2205,"4299"
2157,"7852"

- Peu / Aucune Transformation
- Ajout d'une date de chargement

**Mode de chargement (généralement) :**  
**Suppression / Ajout**

ASSO\_ODS

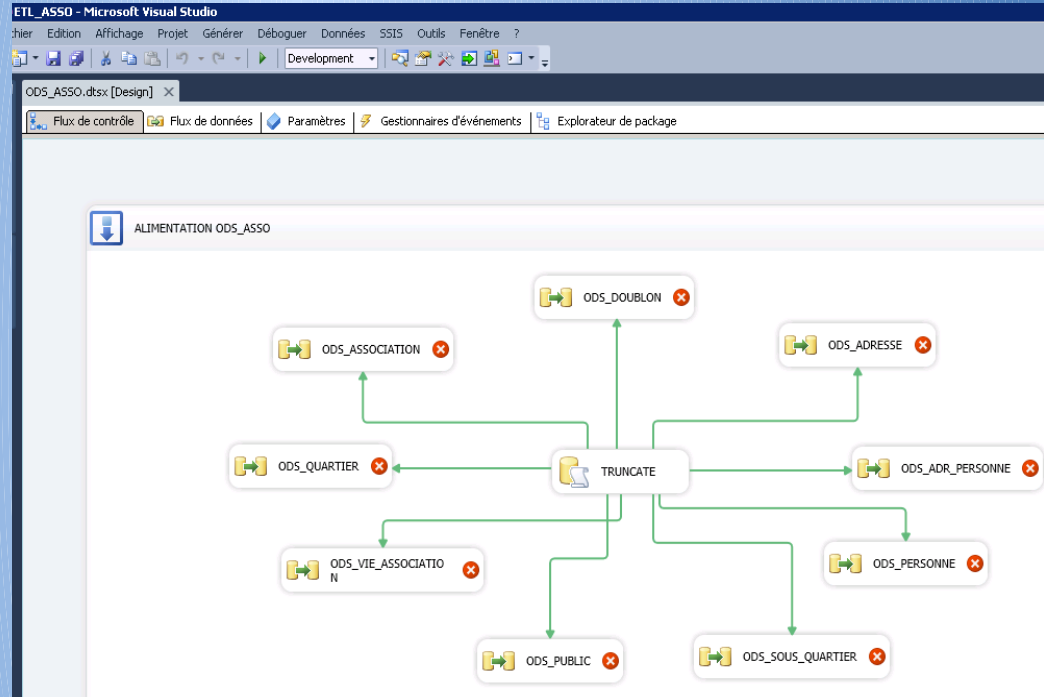
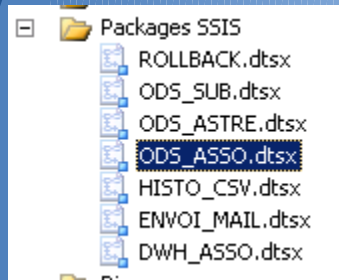
- Schémas de base de données
- Tables
  - Tables système
  - FileTables
  - dbo.ASSO\_ADR\_PERSONNE
  - dbo.ASSO\_ADRESSE
  - dbo.ASSO\_ASSOCIATION
  - dbo.ASSO\_DOUBLONS
    - Colonnes
      - CODE\_ASSO\_MAITRE (int, NULL)
      - CODE\_ASSO\_DOUBLON (int, NULL)
      - DATE\_CHG (datetime, non NULL)
  - Clés
  - Contraintes
  - Déclencheurs
  - Index
  - Statistiques
  - dbo.ASSO\_DOUBLONS\_TCH
  - dbo.ASSO\_PERSONNE
    - Colonnes
      - ID\_AIDE\_DISPOSITIF\_DEF (int, non NULL)
      - VERSION (int, NULL)
      - ARCHIVAGE (int, NULL)
      - LIBELLE (varchar(765), NULL)
      - MONTANT\_PLAFOND (real, NULL)
      - NATURE (varchar(765), NULL)
      - TAUX\_MAXIMUM\_PRISE\_EN\_CHARGE (real, NULL)
      - ID\_DISPOSITIF (int, NULL)
      - ACTIVE (int, NULL)
      - REFERENCE\_TIERS\_FINANCIER (varchar(765), NULL)
      - DATE\_CHG (datetime, non NULL)

## Base Mysql Source SUB

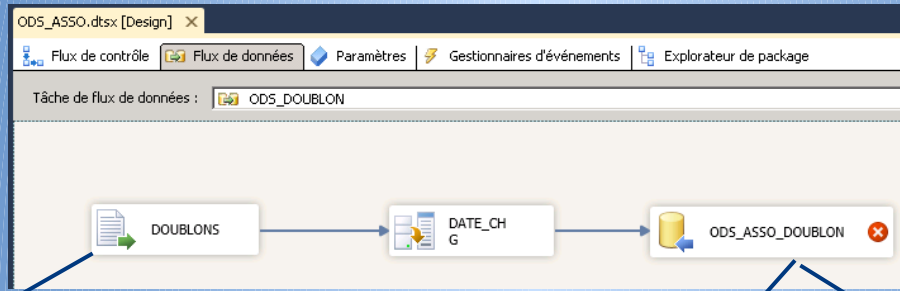
Tables

- T\_ADRESSE
- T\_AIDE\_AXE
- T\_AIDE\_DISPOSITIF
- T\_AIDE\_DISPOSITIF\_DEF
  - Columns
    - ID\_AIDE\_DISPOSITIF\_DEF
    - VERSION
    - ARCHIVAGE
    - LIBELLE
    - MONTANT\_PLAFOND
    - NATURE
    - TAUX\_MAXIMUM\_PRISE\_EN\_CHARGE
    - ID\_DISPOSITIF
    - ACTIVE
    - REFERENCE\_TIERS\_FINANCIER

# Chargement Sources -> ODS



# Chargement Sources -> ODS



**Éditeur du gestionnaire de connexions de fichiers plats**

Nom du gestionnaire de connexions :

Description :

Sélectionnez un fichier et spécifiez les propriétés et le format de fichier.

Nom de fichier :

Paramètres régionaux :   Unicode

Page de codes :

**Gestionnaire de connexions**

Mappages

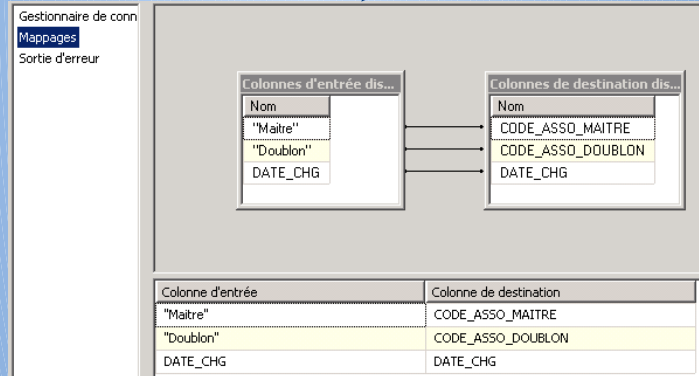
Sortie d'erreur

Spécifiez un gestionnaire de connexions OLE DB, une source de données ou une table, puis sélectionnez le mode d'accès aux données. Si vous utilisez le mode commande SQL, spécifiez la commande SQL soit en tapant, soit en utilisant les requêtes. Pour un accès aux données par chargement rapide, définissez les options de table.

Gestionnaire de connexions OLE DB :

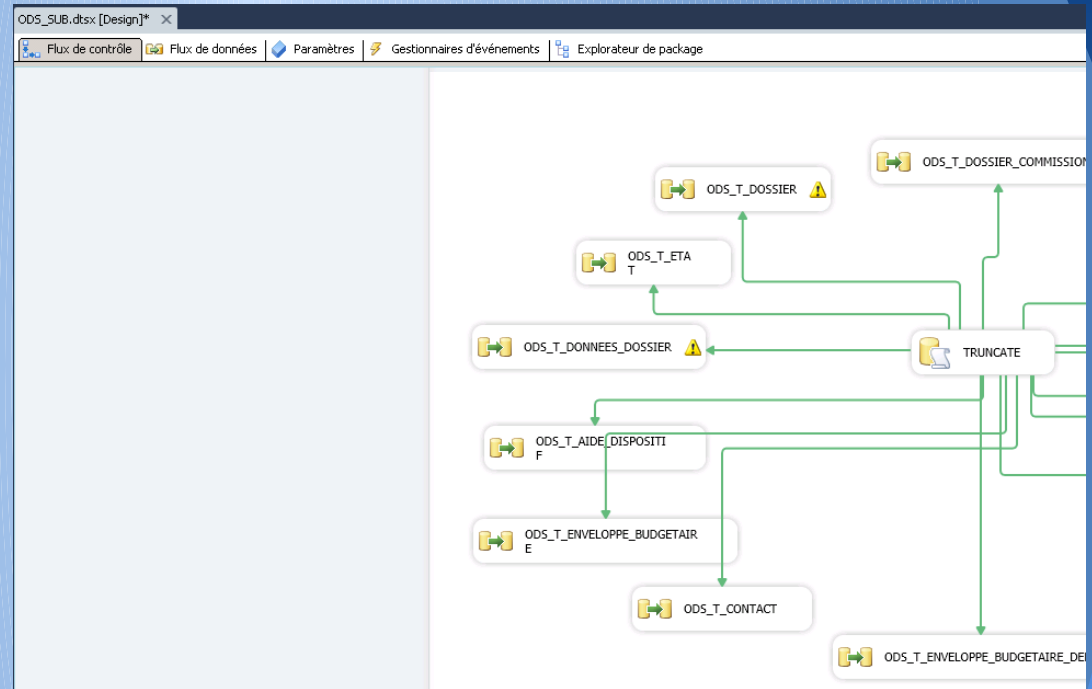
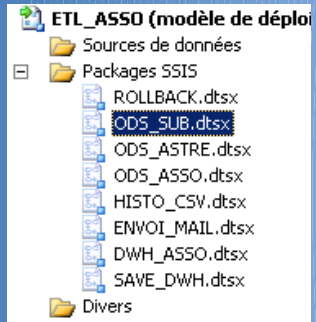
Mode d'accès aux données :

Nom de la table ou de la vue :

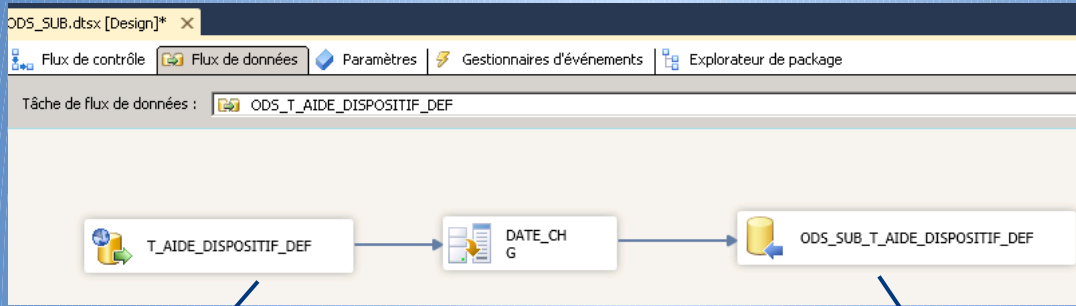




# Chargement Sources -> ODS



# Chargement Sources -> ODS



Gestionnaire de conn  
Colonnes  
Sortie d'erreur

Colonnes externes disponibles

<input checked="" type="checkbox"/>	Nom
<input checked="" type="checkbox"/>	ID_AIDE_DISPOSITIF_DEF
<input checked="" type="checkbox"/>	VERSION
<input checked="" type="checkbox"/>	ARCHIVAGE
<input checked="" type="checkbox"/>	LIBELLE
<input checked="" type="checkbox"/>	MONTANT_PLAFOND
<input checked="" type="checkbox"/>	NATURE
<input checked="" type="checkbox"/>	TAUX_MAXIMUM_PRISE_EN_CHARGE
<input checked="" type="checkbox"/>	ID_DISPOSITIF
<input checked="" type="checkbox"/>	ACTIF

Colonne externe	Colonne de sortie
ID_AIDE_DISPOSITIF_DEF	ID_AIDE_DISPOSITIF_DEF
VERSION	VERSION
ARCHIVAGE	ARCHIVAGE
LIBELLE	LIBELLE

Gestionnaire de conn  
Mappages  
Sortie d'erreur

Colonnes d'entrée disponibles

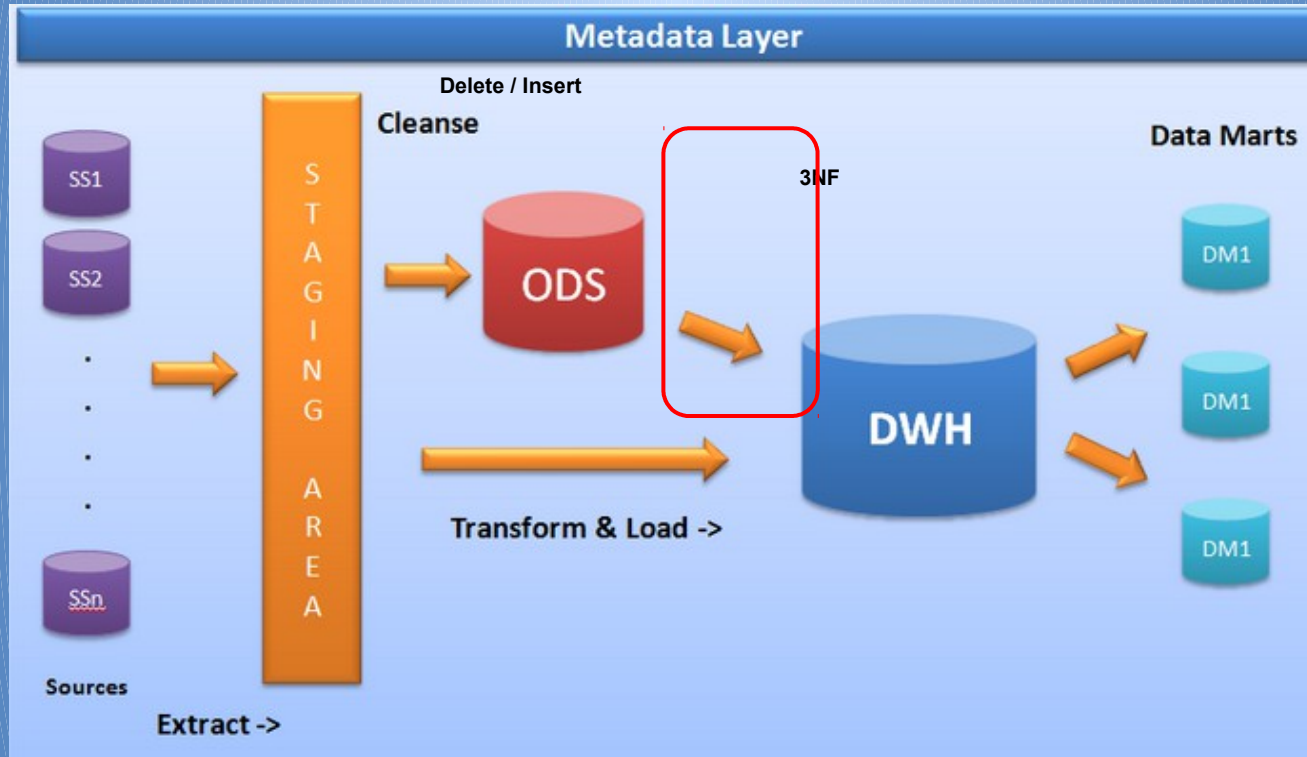
Nom
ID_AIDE_DISPOSITIF_DEF
VERSION
ARCHIVAGE
LIBELLE
MONTANT_PLAFOND
NATURE
TAUX_MAXIMUM_PRISE_EN_CHARGE
ID_DISPOSITIF

Colonnes de destination disp...

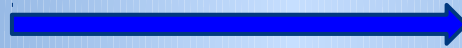
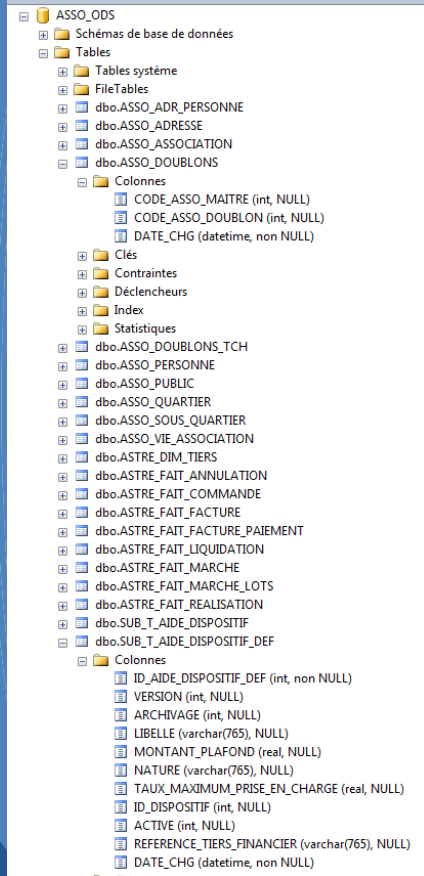
Nom
VERSION
ARCHIVAGE
DATE_CHG
ID_AIDE_DISPOSITIF_DEF
LIBELLE
MONTANT_PLAFOND
NATURE
TAUX_MAXIMUM_PRISE...
ID_DISPOSITIF

Colonne d'entrée	Colonne de destination
VERSION	VERSION
ARCHIVAGE	ARCHIVAGE
DATE_CHG	DATE_CHG

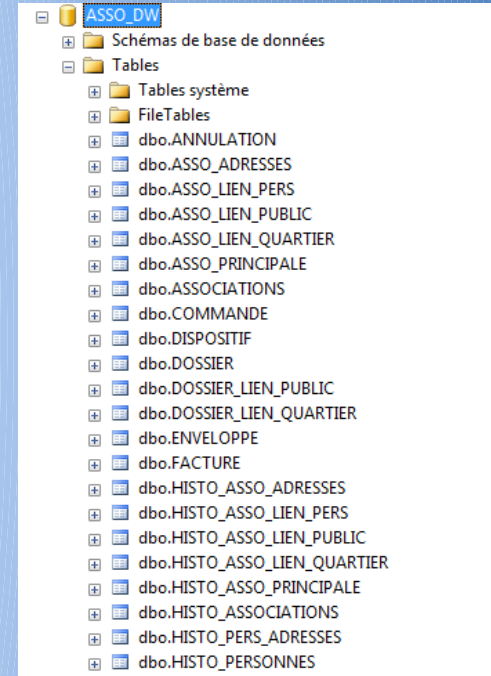
# Chargement ODS -> DWH



# Chargement ODS -> DWH



Mode de chargement :  
Slow Data Change diffèrent selon les tables



# Chargement ODS -> DWH

- Slow Data Change SDCO (rarement utilisé)

On conserve uniquement l'enregistrement initial (toutes les variations suivantes sont ignorées)

Exemple avec une donnée avec le champ Supplier\_Key comme clé primaire

**ODS**

**DWH**

1er lancement

Supplier_Key	Supplier_Code	Supplier_Name	Supplier_State
123	ABC	Acme Supply Co	CA



Supplier_Key	Supplier_Code	Supplier_Name	Supplier_State
123	ABC	Acme Supply Co	CA

2nd lancement

Supplier_Key	Supplier_Code	Supplier_Name	Supplier_State
123	ABC	Acme Supply Co	IL



Supplier_Key	Supplier_Code	Supplier_Name	Supplier_State
123	ABC	Acme Supply Co	CA

# Chargement ODS -> DWH

- Slow Data Change SDC1

On modifie l'enregistrement sans conservation de ces états successifs (pas de gestion d'historique)

**ODS**

**DWH**

1er lancement

Supplier_Key	Supplier_Code	Supplier_Name	Supplier_State
123	ABC	Acme Supply Co	CA



Supplier_Key	Supplier_Code	Supplier_Name	Supplier_State
123	ABC	Acme Supply Co	CA

2nd lancement

Supplier_Key	Supplier_Code	Supplier_Name	Supplier_State
123	ABC	Acme Supply Co	IL



Supplier_Key	Supplier_Code	Supplier_Name	Supplier_State
123	ABC	Acme Supply Co	IL

+ Colonne  
Date de modification

# Chargement ODS -> DWH

- Slow Data Change SDC2 (partie 1)

Des dates de début / fin de validité sont créées afin de conserver tout l'historique des changements d'un enregistrement (et de connaître celui qui est valide à un instant T)

Inconvénient : Gestion de dates au niveau de la restitution

**ODS**

**DWH**

1er lancement le 10/10/2014 à 17h20

Supplier_Key	Supplier_Code	Supplier_Name	Supplier_State
123	ABC	Acme Supply Co	CA



Supplier_Key	Supplier_Code	Supplier_Name	Supplier_State	Date_Debut	Date_Fin
123	ABC	Acme Supply Co	CA	10/10/2014 17:20	-

2nd lancement le 12/10/2014 à 17h20

Supplier_Key	Supplier_Code	Supplier_Name	Supplier_State
123	ABC	Acme Supply Co	IL



Supplier_Key	Supplier_Code	Supplier_Name	Supplier_State	Date_Debut	Date_Fin
123	ABC	Acme Supply Co	CA	10/10/2014 17:20	12/10/2014 17:20
123	ABC	Acme Supply Co	IL	12/10/2014 17:20	-

3e lancement le 12/10/2014 à 18h35

Supplier_Key	Supplier_Code	Supplier_Name	Supplier_State
123	ABC	Acme Supply Co	BO



Supplier_Key	Supplier_Code	Supplier_Name	Supplier_State	Date_Debut	Date_Fin
123	ABC	Acme Supply Co	CA	10/10/2014 17:20	12/10/2014 17:20
123	ABC	Acme Supply Co	IL	12/10/2014 17:20	12/10/2014 18h35
123	ABC	Acme Supply Co	BO	12/10/2014 18h35	-

# Chargement ODS -> DWH

- Slow Data Change SDC2 (partie 2)

Attention : La restitution doit gérer le système de date

Parfois on ajoute un booléen (nommé flag) pour aider les requêtes de restitution

**ODS**

**DWH**

1er lancement le 10/10/2014 à 17h20

Supplier_Key	Supplier_Code	Supplier_Name	Supplier_State
123	ABC	Acme Supply Co	CA



Supplier_Key	Supplier_Code	Supplier_Name	Supplier_State	Date_Debut	Date_Fin	Valide
123	ABC	Acme Supply Co	CA	10/10/2014 17:20	-	1

2nd lancement le 12/10/2014 à 17h20

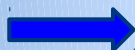
Supplier_Key	Supplier_Code	Supplier_Name	Supplier_State
123	ABC	Acme Supply Co	IL



Supplier_Key	Supplier_Code	Supplier_Name	Supplier_State	Date_Debut	Date_Fin	Valide
123	ABC	Acme Supply Co	CA	10/10/2014 17:20	12/10/2014 17:20	0
123	ABC	Acme Supply Co	IL	12/10/2014 17:20	-	1

3e lancement le 12/10/2014 à 18h35

Supplier_Key	Supplier_Code	Supplier_Name	Supplier_State
123	ABC	Acme Supply Co	BO



Supplier_Key	Supplier_Code	Supplier_Name	Supplier_State	Date_Debut	Date_Fin	Valide
123	ABC	Acme Supply Co	CA	10/10/2014 17:20	12/10/2014 17:20	0
123	ABC	Acme Supply Co	IL	12/10/2014 17:20	12/10/2014 18h35	0
123	ABC	Acme Supply Co	BO	12/10/2014 18h35	-	1



# Chargement ODS -> DWH

- Slow Data Change SDC3  
Seule une partie de l'historique est conservée (dans des colonnes séparées)  
Avantage : Évite les impacts au niveau de la restitution  
Inconvénients : Conservation d'une partie de l'historique seulement

ODS

DWH

1er lancement le 10/10/2014 à 17h20

Supplier_Key	Supplier_Code	Supplier_Name	Supplier_State
123	ABC	Acme Supply Co	CA



Supplier_Key	Supplier_Code	Supplier_Name	Supplier_State_Old	Supplier_State	Date_MAJ
123	ABC	Acme Supply Co		CA	10/10/2014 17:20

2nd lancement le 12/10/2014 à 17h20

Supplier_Key	Supplier_Code	Supplier_Name	Supplier_State
123	ABC	Acme Supply Co	IL



Supplier_Key	Supplier_Code	Supplier_Name	Supplier_State_Old	Supplier_State	Date_MAJ
123	ABC	Acme Supply Co	CA	IL	12/10/2014 17:20

3e lancement le 12/10/2014 à 18h35

Supplier_Key	Supplier_Code	Supplier_Name	Supplier_State
123	ABC	Acme Supply Co	BO



Supplier_Key	Supplier_Code	Supplier_Name	Supplier_State_Old	Supplier_State	Date_MAJ
123	ABC	Acme Supply Co	IL	BO	12/10/2014 18:35

# Chargement ODS -> DWH

- Slow Data Change SDC4

Création d'une table historique qui conserve l'ensemble des modifications des enregistrements

**ODS**

**DWH**

1er lancement le 10/10/2014 à 17h20

Supplier_Key	Supplier_Code	Supplier_Name	Supplier_State
123	ABC	Acme Supply Co	CA



Supplier_Key	Supplier_Code	Supplier_Name	Supplier_State
123	ABC	Acme Supply Co	CA

2nd lancement le 12/10/2014 à 17h20

Supplier_Key	Supplier_Code	Supplier_Name	Supplier_State
123	ABC	Acme Supply Co	IL



*TABLE*

Supplier_Key	Supplier_Code	Supplier_Name	Supplier_State
123	ABC	Acme Supply Co	IL

*TABLE HISTORISEE*

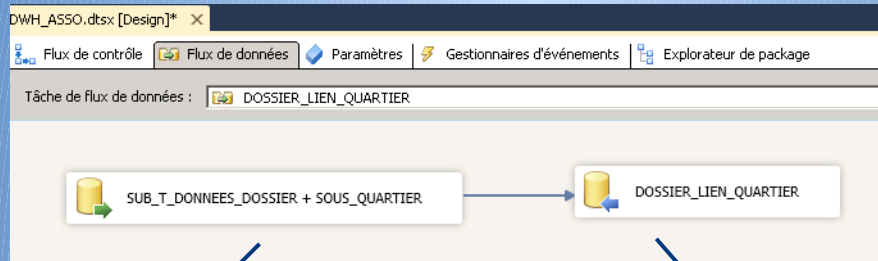
Supplier_Key	Supplier_Code	Supplier_Name	Supplier_State	Date_MAJ
123	ABC	Acme Supply Co	CA	12/10/2014 17:20

### Exemple de Règles d'alimentation de tables ODS vers DWH

Table	Sources	Commentaires	Chargement	Fréquence
DOSSIER_LIEN_QUARTIER	SUB_T_DONNEES_DOSSIER + SOUS_QUARTIER	INNER JOIN ON SUB_T_DONNEES_DOSSIER.VALEUR_REFERENTIEL LIKE SOUS_QUARTIER.LIBELLE WHERE SUB_T_DONNEES_DOSSIER.DONNEES_CHAMPS_NOM In ('LocalisationProjet1', 'LocalisationProjet2', 'LocalisationProjet3', 'LocalisationProjet4', 'LocalisationProjet5', 'LocalisationProjet6', 'LocalisationProjet7', 'LocalisationProjet8', 'LocalisationProjet9', 'LocalisationProjet10')	Annule et remplace	Quotidien
ASSO_PRINCIPALE	ASSO_DOUBLONS	Intégralité de ASSO_DOUBLONS	Historisation	Quotidien

### Exemple de Mapping de champs DWH

Table	Champs	Type	CP	FK	NULL	AJ	Description	Commentaire
DOSSIER_LIEN_QUARTIER	ID_DOSSIER	INTEGER	Y	Y	N		Identifiant du dossier	SUB_T_DONNEES_DOSSIER.ID_DOSSIER
	ID_SOUS_QUARTIER	INTEGER	Y	Y	N		Identifiant du quartier	SOUS_QUARTIER.ID_SOUS_QUARTIER
	DATE_CHG	DATETIME			N		Date de chargement	Date du jour
VIE_ASSOCIATION	ID_VIE_ASSO	INTEGER	Y		N	Y	Numéro incrémentiel de ligne	
	CODE_ASSO	INTEGER		Y	N		Code de l'association	ASSO_VIE_ASSOCIATION.CODE_ASSO
	AUTEUR	VARCHAR(50)			Y		Auteur de la modification sur l'association	ASSO_VIE_ASSOCIATION.AUTEUR
	DATE_SAISIE	DATETIME			N		Date de l'évènement	ASSO_VIE_ASSOCIATION.DATE_SAISIE
	TYPE_EVENT	VARCHAR(38)			N		Type de l'évènement	ASSO_VIE_ASSOCIATION.TYPE_EVNT
	CHAMPS_MODIF	VARCHAR(100)			Y		Champs modifiés par l'évènement	Créer autant de lignes que de champ séparées par ";" ASSO_VIE_ASSOCIATION.CHAMP_MODIF
	OLD_VALEUR	VARCHAR(100)			Y		Ancienne valeur du champ	Idem récupérer les valeurs séparées par des ";" dans l'ordre correspondant au champs ASSO_VIE_ASSOCIATION.OLD_VALEUR
	NEW_VALEUR	VARCHAR(100)			Y		Nouvelle valeur du champ	Idem récupérer les valeurs séparées par des ";" dans l'ordre correspondant au champs ASSO_VIE_ASSOCIATION.NEW_VALEUR
	DATE_CREATION	DATETIME			Y		Date de création	ASSO_VIE_ASSOCIATION.DATE_CREATION
	DATE_AG	DATETIME			Y		Date AG	ASSO_VIE_ASSOCIATION.DATE_AG
DATE_CHG	DATETIME			N		Date du chargement	Date du jour	



**Gestionnaire de conn**  
Colonnes  
Sortie d'erreur

Spécifiez un gestionnaire de connexions OLE DB, une source de données, puis sélectionnez le mode d'accès aux données. Si vous utilisez une commande SQL, spécifiez la commande SQL soit en tapant, soit en utilisant un fichier de requêtes.

Gestionnaire de connexions OLE DB :  
OD5\_ASSO

Mode d'accès aux données :  
Commande SQL

Texte de la commande SQL :

```
SELECT TD.ID_DOSSIER
      ,S.ID_SOUS_QUARTIER
      ,S.LIBELLE
      ,GETDATE() AS DATE_CHG
FROM ASSO_ODS.DBO.SUB_T_DONNEES_DOSSIER TD
INNER JOIN ASSO_DW.dbo.SOUS_QUARTIER S ON (S.LIBELLE =
UPPER(RD.VALEUR_REFERENTIEL) OR TD.VALEUR_REFERENTIEL
LIKE '%sous%')
INNER JOIN ASSO_DW.DBO.DOSSIER D ON D.ID_DOSSIER =
TD.ID_DOSSIER
WHERE TD.DONNEES_CHAMPS_NOM IN ('LocalisationProjet1',
'LocalisationProjet2', 'LocalisationProjet3', 'LocalisationProjet4',
'LocalisationProjet5', 'LocalisationProjet6', 'LocalisationProjet7',
'LocalisationProjet8', 'LocalisationProjet9', 'LocalisationProjet10')
GROUP BY TD.ID_DOSSIER, S.ID_SOUS_QUARTIER, S.LIBELLE
```

SQL volumineux

**Gestionnaire de conn**  
Mappages  
Sortie d'erreur

Colonne d'entrée

Nom
ID_DOSSIER
ID_SOUS_QUARTIER
LIBELLE
DATE_CHG

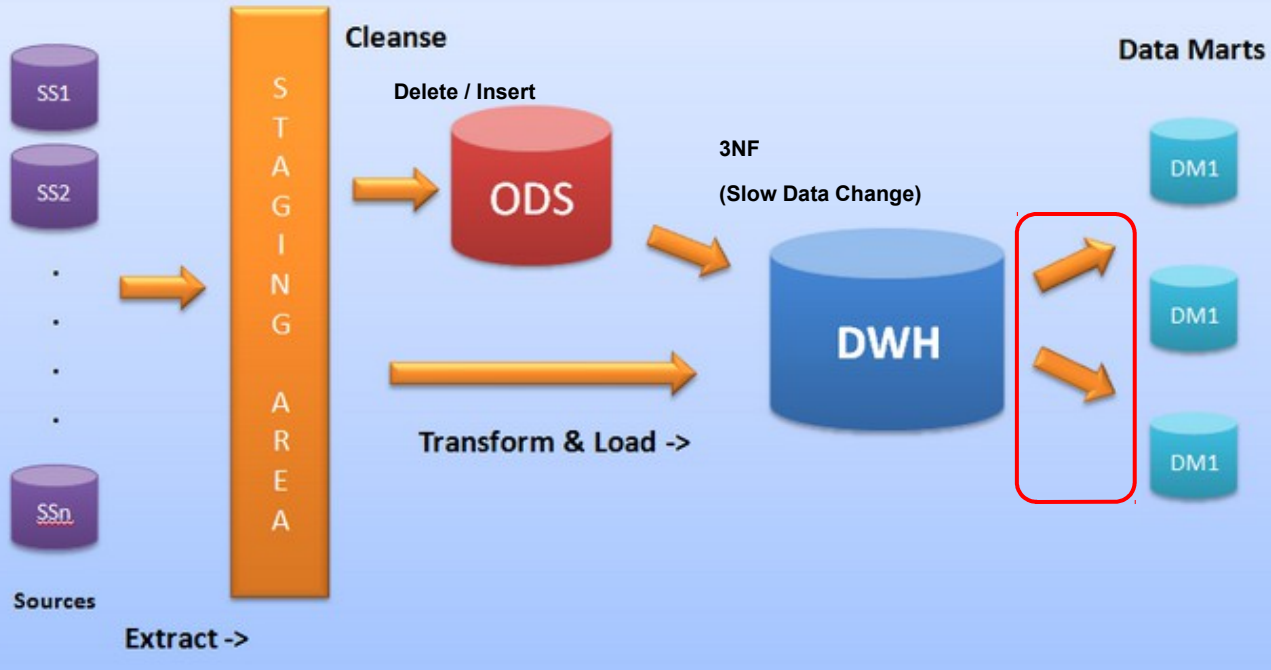
Colonne de destination

Nom
ID_DOSSIER
DATE_CHG
ID_SOUS_QUARTIER

Colonne d'entrée

Colonne d'entrée	Colonne de destination
ID_DOSSIER	ID_DOSSIER
DATE_CHG	DATE_CHG
ID_SOUS_QUARTIER	ID_SOUS_QUARTIER

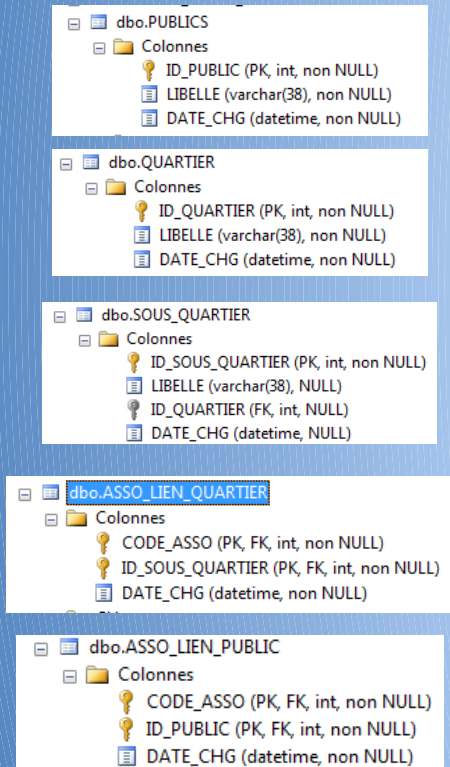
## Metadata Layer



# Chargement DWH -> DTM / AGG

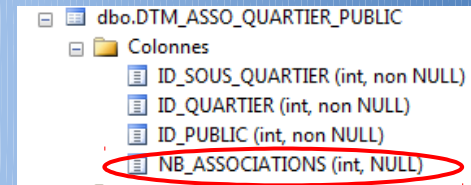
(avec stockage)

## DWH



Somme sur le nombre d'associations

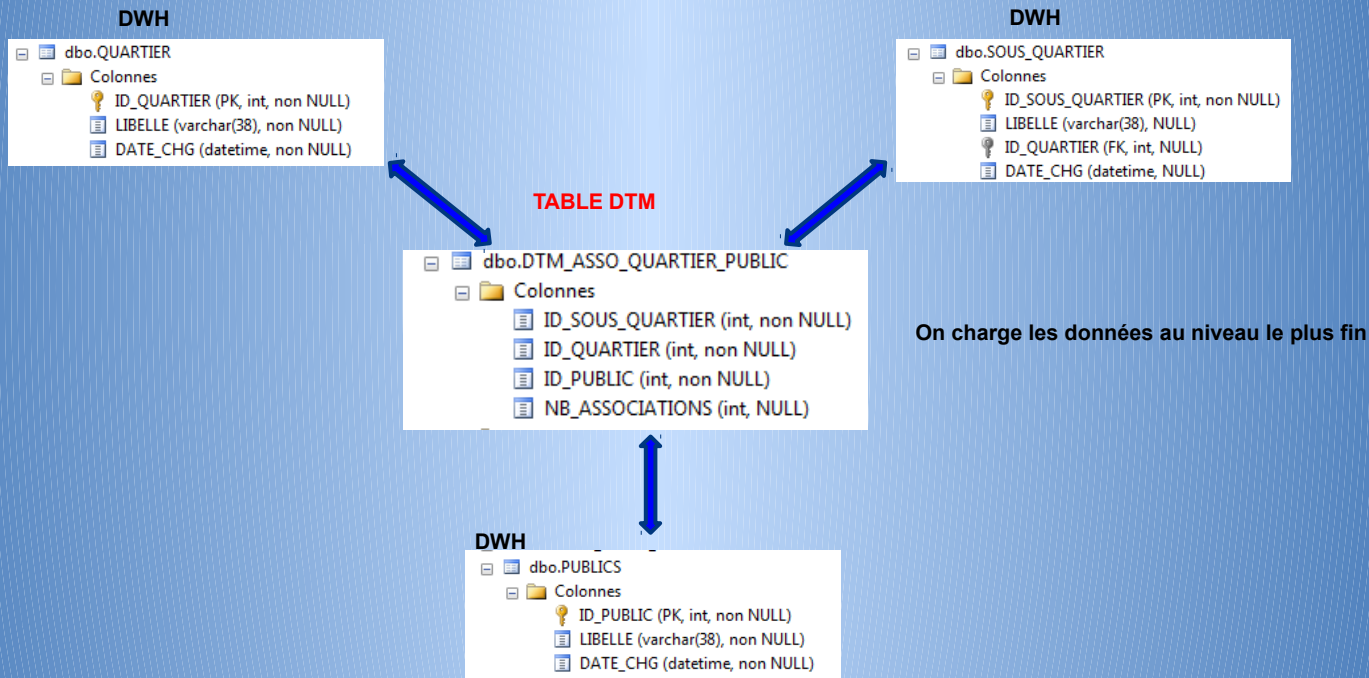
## DTM



ID_SOUS_QUARTIER	ID_QUARTIER	ID_PUBLIC	NB_ASSOCIATIONS
1	1	1	35
1	1	2	48
1	1	3	53
2	1	1	13
2	1	2	26
2	1	3	25
3	1	1	34
3	1	2	58
3	1	3	96
4	2	1	47
4	2	2	54
4	2	3	12

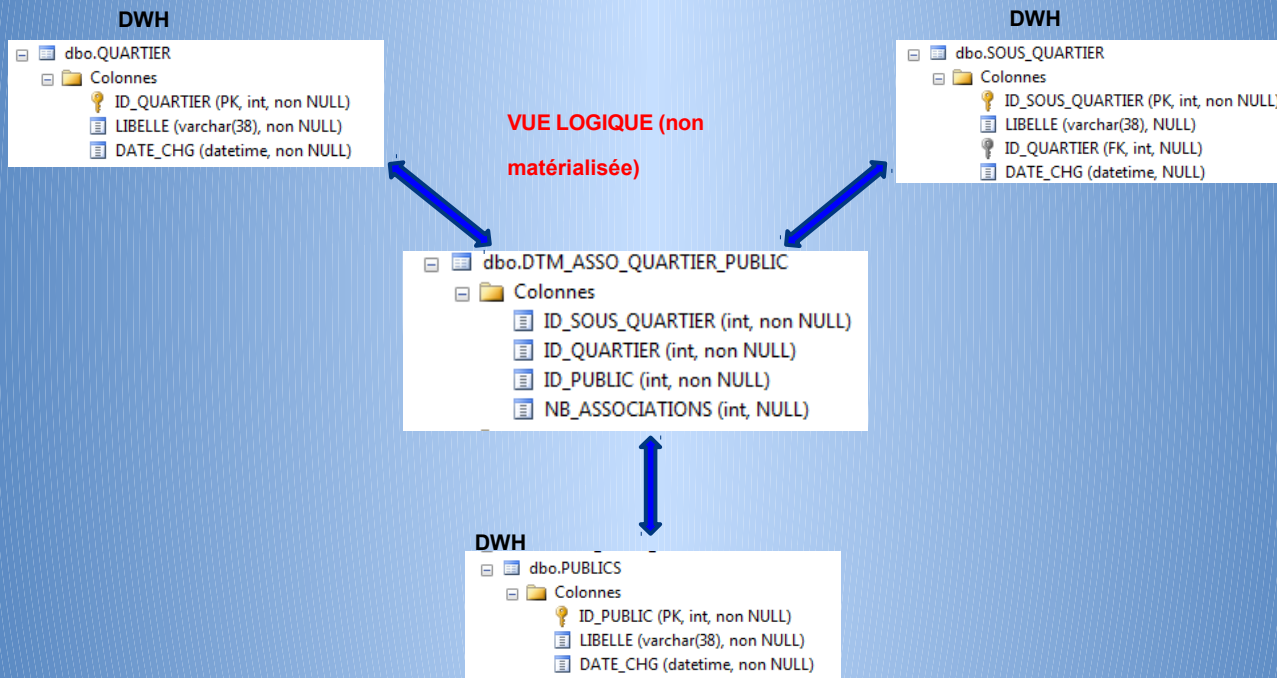
# Chargement DWH -> DTM / AGG

(avec stockage)



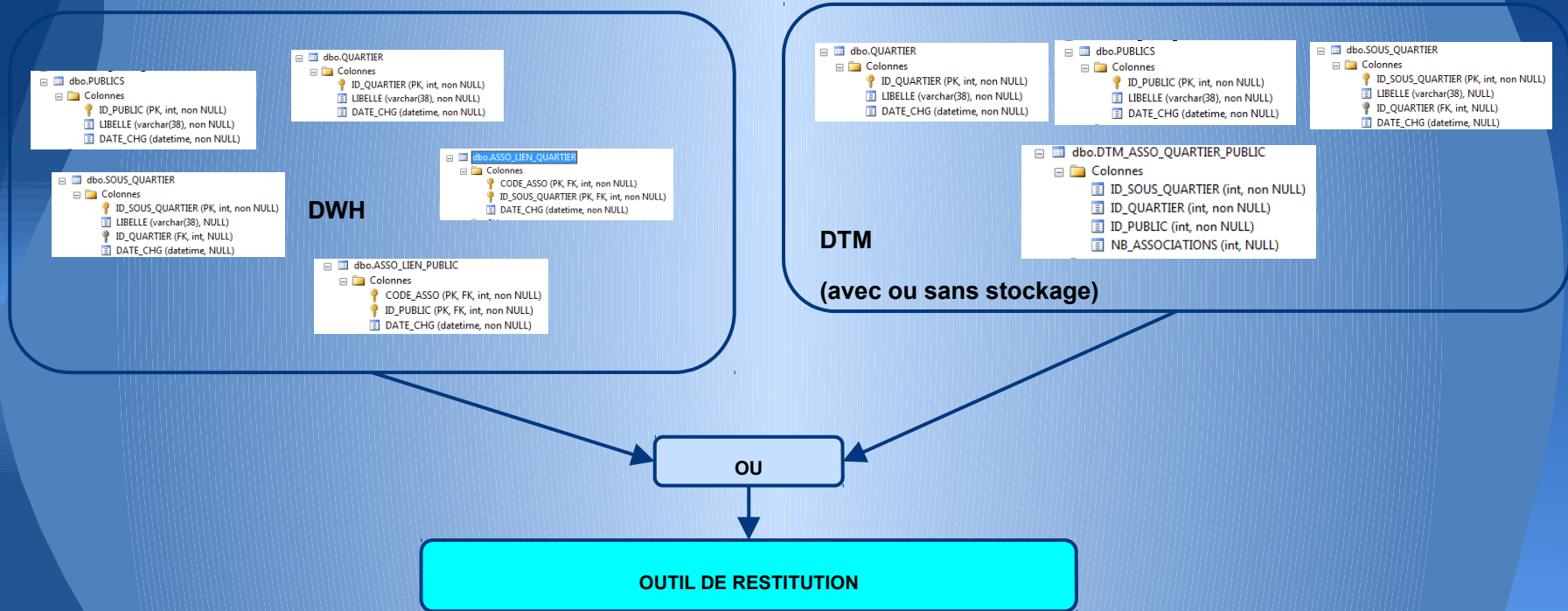
# Chargement DWH -> DTM / AGG

(sans stockage)





# Chargement DWH -> DTM / AGG



## 6 - CONCEPTION ETL



## Fournir une analyse du nombre de ventes par Secteur et par Type de Produits par mois

Application 1 (REFPDT) ouverte tous les jours de 6h à 22h: Table Produits

ProduitCode	MarqueduProduit	TypeduProduit	SousTypeduProduit	LibelleduProduit
198383	DIOR	Maquillage	Mascara	Diorshow Extase
175012	DIOR	Maquillage	Mascara	Diorshow Iconic
151802	DIOR	Maquillage	Mascara	Diorshow Black Out
196826	GIVENCHY	Maquillage	Mascara	Extravagancia Noir
308018	GIVENCHY	Maquillage	Rouge à lèvres	Extravagancia Rouge
263739	GIVENCHY	Maquillage	Rouge à lèvres	Extravagancia Brun Vintage
308020	GIVENCHY	Maquillage	Rouge à lèvres	Extravagancia Rose
166255	DIOR	Soins	Yeux	Duo Yeux Démaquillant Express
166855	DIOR	Soins	Yeux	Eau Micellaire Démaquillante Express

integer    varchar    varchar    varchar    varchar

**Application 2 (VCOM) : Fichiers CSV extraits tous les jours à Mididans un répertoire C:\Extraction\**

Le fichier Ventes contient toutes les ventes du J-1 / peut contenir les ventes d'un jour antérieur

(les ventes qui ne sont plus envoyées pour un jour doivent être supprimées)

FichierVentes.csv

DateVente	NumeroProduit	NumeroClient	SecteurVente
01/07/2014 12:12	196826	AAA	NORD
01/07/2014 12:12	196826	AAA	NORD
01/07/2014 12:12	196826	AAA	NORD
01/07/2014 12:12	263739	AAA	NORD
03/07/2014 06:55	196826	BBB	SUD
03/07/2014 06:55	308018	BBB	SUD
03/07/2014 06:55	528444	BBB	SUD
05/07/2014 12:12	308020	AAA	NORD

Fichier Clients.csv

NumeroClient	Nom	Prenom
AAA	Dupont	Jacques
AAA	Dupont	Jacques
BBB	Martin	Julien

## Fournir une analyse des ventes par Secteur, Type de Produits et par mois

### Conception du modèle décisionnel / traitements ETL

- > Conception du schéma ODS
- > Conception du schéma DWH en 3NF
- > Conception du modèle DTM
- > Conception du mode de chargement ODS
- > Conception du mode de chargement DWH
- > Conception du mode de chargement DTM

\*\*\* Spécifier tous les traitements ETL à effectuer pour chaque alimentation

\*\*\* Gérer les cas d'erreur

\*\*\* Ordonnancer tous les traitements

**ATTENTION: LA CONCEPTION DOIT PERMETTRE PLUSIEURS CHARGEMENTS SUCCESSIFS**

## Fournir une analyse des ventes par Secteur, Type de Produits et par mois

### Conception ODS

TABLE REFPDT PRODUITS	
ProduitCode	Integer
MarqueduProduit	Varchar
TypeduProduit	Varchar
SousTypeduProduit	Varchar
LibelleduProduit	Varchar
Date_Upd	Datetime

3 tables au format identique de l'entrant,

Sans contrainte de clé

TABLE VCOM VENTES	
DateVente	Varchar
NumeroProduit	Varchar
NumeroClient	Varchar
SecteurVente	Varchar
Date_Upd	Datetime

TABLE VCOM CLIENTS	
NumeroClient	Varchar
Nom	Varchar
Prenom	Varchar
Date_Upd	Datetime

## Fournir une analyse des ventes par Secteur, Type de Produits et par mois

Conception DWH 3NF sans réflexion sur le mode de chargement

DWH TYPEPDT	
<b>Id_TypePdt</b>	<b>Integer</b>
Libelle	Varchar
Date_Upd	Datetime

DWH SSTYPEPDT	
<b>Id_SsTypePdt</b>	<b>Integer</b>
<i>Id_TypePdt</i>	<i>Integer</i>
Libelle	Varchar
Date_Upd	Datetime

DWH VENTES	
<b>DateVente</b>	<b>Datetime</b>
<i>Id_Client</i>	<i>Varchar</i>
<i>Id_Pdt</i>	<i>Integer</i>
Quantité	Integer
<i>Id_Secteur</i>	<i>Integer</i>
Date_Upd	Datetime

DWH CLIENTS	
<b>Id_Client</b>	<b>Varchar</b>
Nom	Varchar
Prenom	Varchar
Date_Upd	Datetime

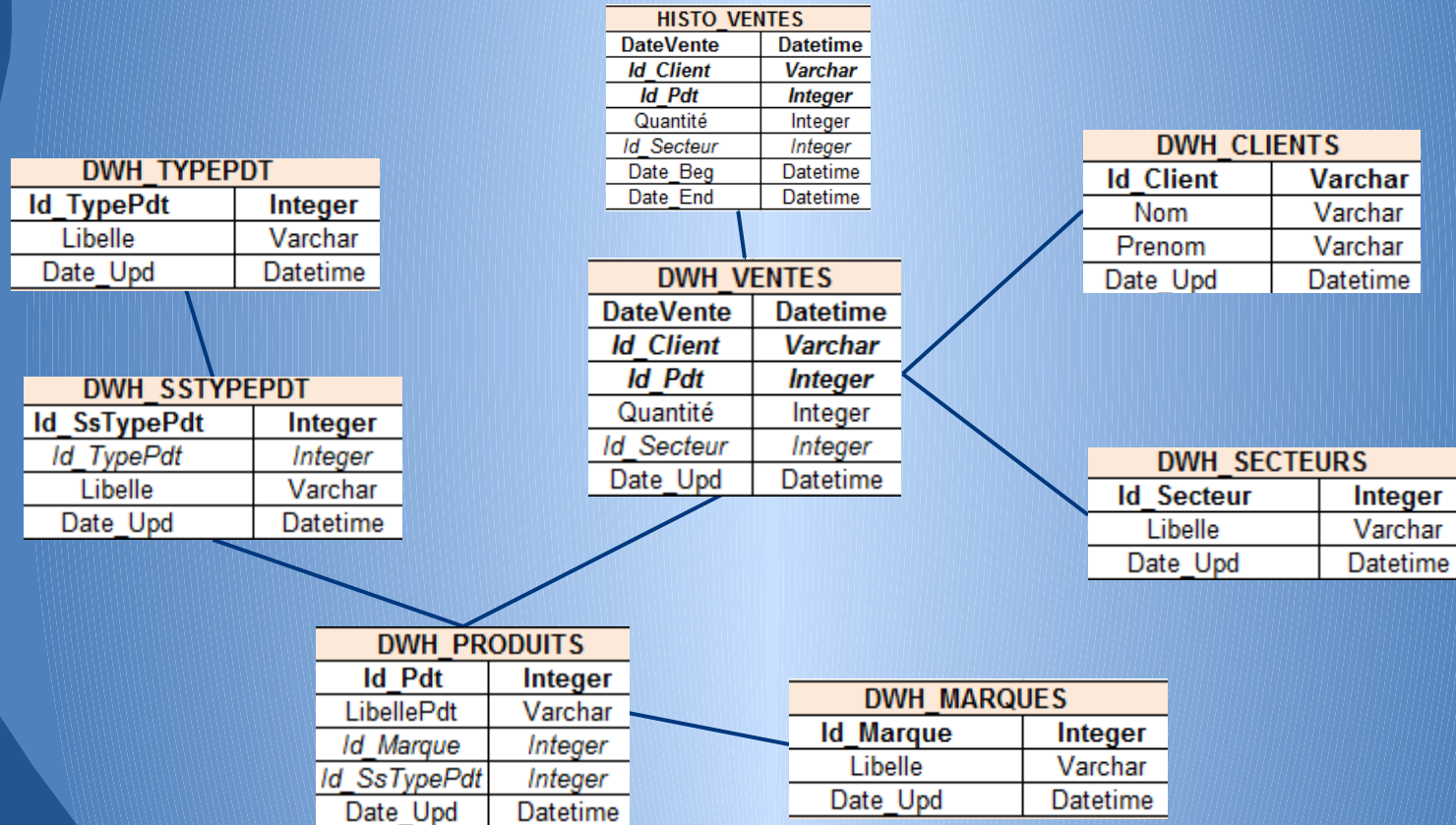
DWH SECTEURS	
<b>Id_Secteur</b>	<b>Integer</b>
Libelle	Varchar
Date_Upd	Datetime

DWH PRODUITS	
<b>Id_Pdt</b>	<b>Integer</b>
LibellePdt	Varchar
<i>Id_Marque</i>	<i>Integer</i>
<i>Id_SsTypePdt</i>	<i>Integer</i>
Date_Upd	Datetime

DWH MARQUES	
<b>Id_Marque</b>	<b>Integer</b>
Libelle	Varchar
Date_Upd	Datetime

## Fournir une analyse des ventes par Secteur, Type de Produits et par mois

Finalisation ConceptionDWH=> finalement tout en SDC1 sauf pour VENTES en SDC4



## Fournir une analyse des ventes par Secteur, Type de Produits et par mois

### Conception du DTM

DWH_TYPEPDT	
<b>Id_TypePdt</b>	<b>Integer</b>
Libelle	Varchar
Date_Upd	Datetime

DWH_SECTEURS	
<b>Id_Secteur</b>	<b>Integer</b>
Libelle	Varchar
Date_Upd	Datetime

DTM_VENTES_TYPEPDT_SECTEUR	
DateDTM	Datetime
<b>Id_TypePdt</b>	<b>Integer</b>
<b>Id_Secteur</b>	<b>Integer</b>
Nb_Ventes	Integer



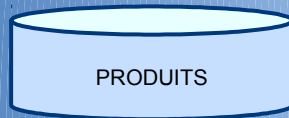
# Fournir une analyse des ventes par Secteur, Type de Produits et par mois

TOUS LES JOURS APRES 22H

## Conception chargement ODS

1

Suppression des données de toutes les tables ODS



Chargement ODS Produit

2

ODS

TABLE REFPDT PRODUITS	
ProduitCode	Integer
MarqueduProduit	Varchar
TypeduProduit	Varchar
SousTypeduProduit	Varchar
LibelleduProduit	Varchar
Date_Upd	Datetime



Chargement ODS Ventes

2

TABLE VCOM VENTES	
DateVente	Varchar
NumeroProduit	Varchar
NumeroClient	Varchar
SecteurVente	Varchar
Date_Upd	Datetime



Chargement ODS Clients

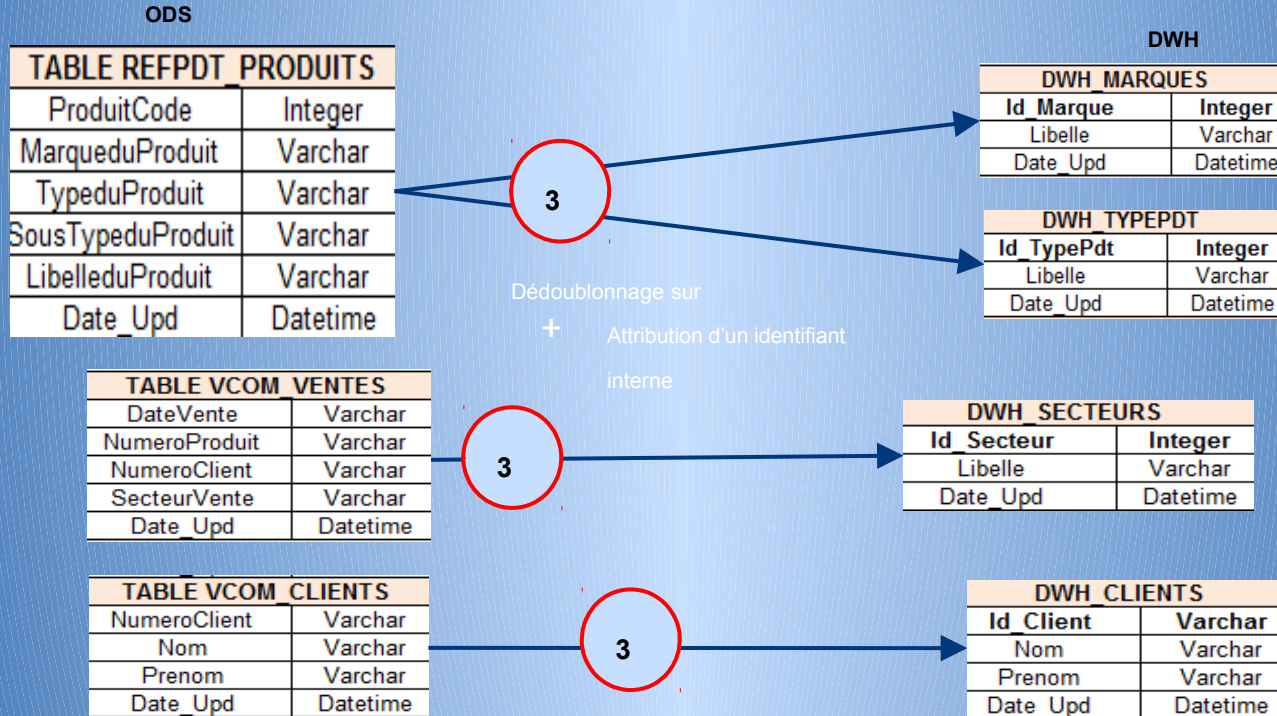
2

TABLE VCOM CLIENTS	
NumeroClient	Varchar
Nom	Varchar
Prenom	Varchar
Date_Upd	Datetime

# Fournir une analyse des ventes par Secteur, Type de Produits et par mois

## Conception Chargement DWH

Alimentation des tables en mode UPSERT  
(MAJ si enregistrement Existe, sinon INSERT)  
Mettre à jour la date Date\_Upd



## Fournir une analyse des ventes par Secteur, Type de Produits et par mois

### Conception Chargement DWH

TOUS LES JOURS UNE FOIS LE  
CHARGEMENT ODS TERMINE



Alimentation des tables en mode UPSERT  
(MAJ si enregistrement Existe, sinon INSERT)  
Mettre à jour la date Date\_Upd

**ODS**

TABLE REFPDT_PRODUICTS	
ProduitCode	Integer
MarqueduProduit	Varchar
TypeduProduit	Varchar
SousTypeduProduit	Varchar
LibelleduProduit	Varchar
Date_Upd	Datetime

DWH TYPEPDT	
Id_TypePdt	Integer
Libelle	Varchar
Date_Upd	Datetime

Dédoublonnage + Attribution Identifiant Interne  
Sélection de l'identifiant Id\_TypePdt  
avec Libelle = TypeduProduit

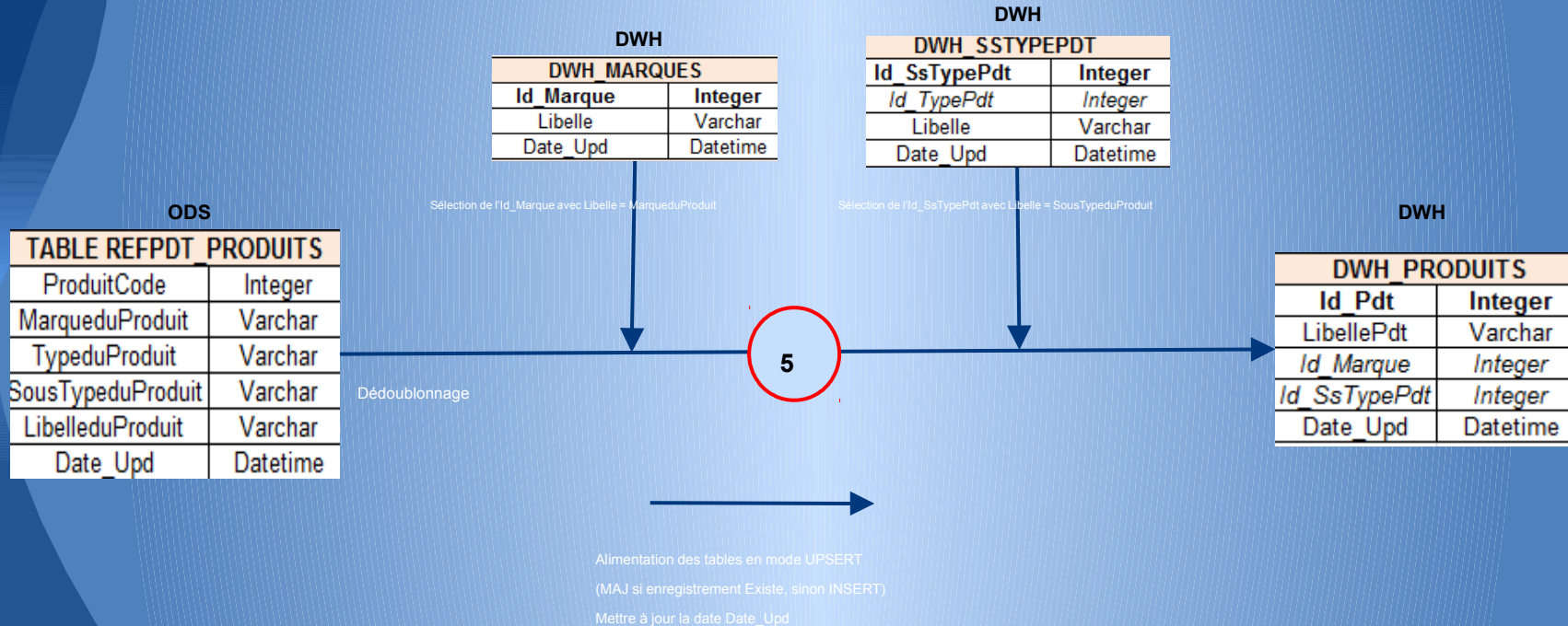
4

**DWH**

DWH SSTYPEPDT	
Id_SsTypePdt	Integer
Id_TypePdt	Integer
Libelle	Varchar
Date_Upd	Datetime

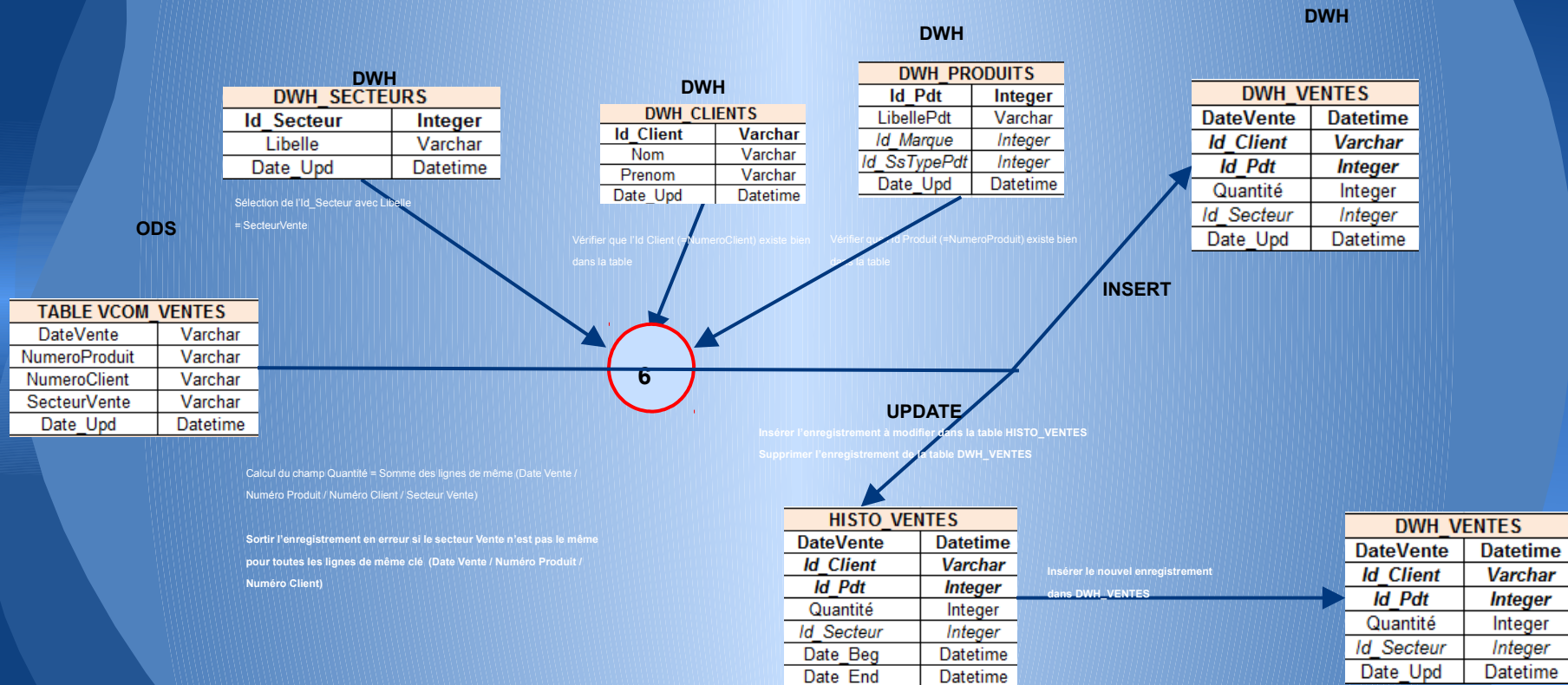
# Fournir une analyse des ventes par Secteur, Type de Produits et par mois

## Conception Chargement DWH



# Fournir une analyse des ventes par Secteur, Type de Produits et par mois

## Conception Chargement DWH



# Fournir une analyse des ventes par Secteur, Type de Produits et par mois

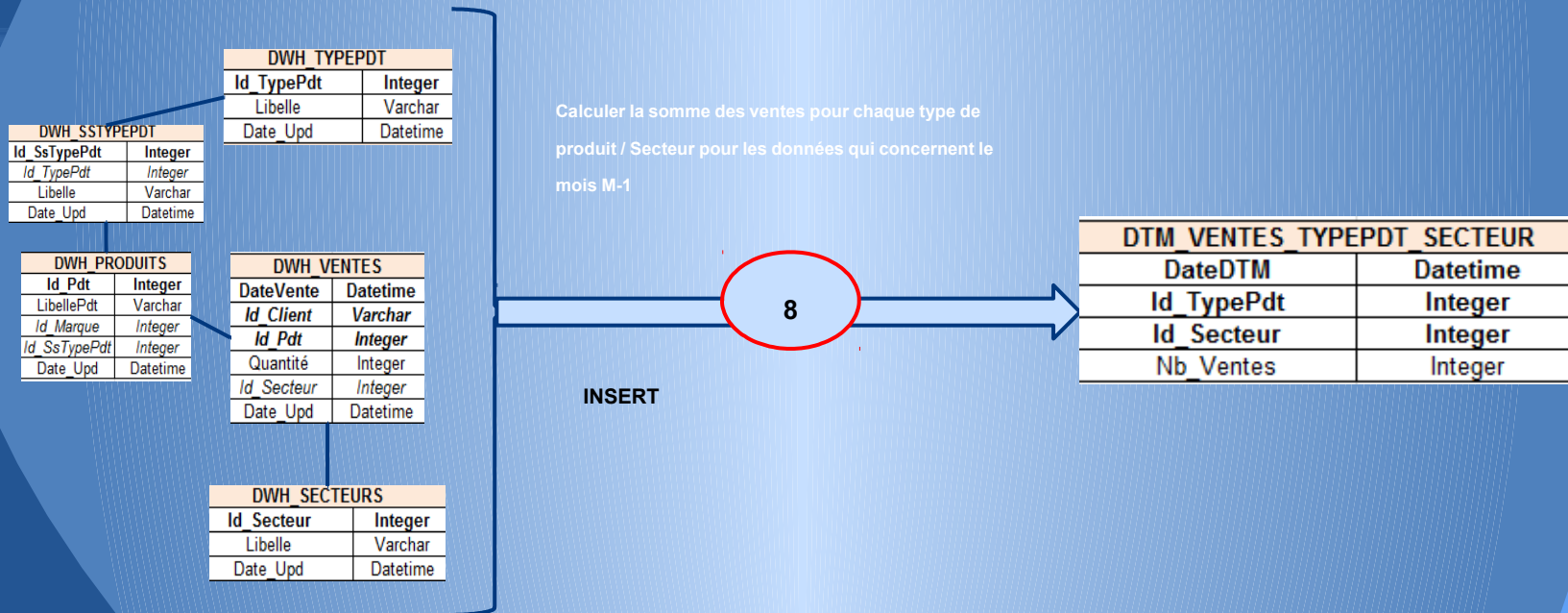
## Conception Chargement DTM

TOUS LES 1er du mois, CHARGER le mois M-

1

7

Suppression des données du mois M-1 dans la table DTM



REGLE DE LANCEMENT	N°ETAPE	ACTIONS	LISTE ACTIONS SPECIFIQUES
Tous les jours à minuit	1	Suppression des données dans l'ODS	
	2		
	2.1	Charger la table ODS_PRODUIITS	
	2.1	Charger la table ODS_VENTES	
	2.1	Charger la table ODS_CLIENTS	
	3		
	3.1	Charger la table DWH_MARQUES	Dédoublonner l'entrant Attribuer identifiant interne Alimenter la table en mode UPSERT
	3.1	Charger la table DWH_TYPEPDT	
	3.1	Charger la table DWH_SECTEURS	
	3.1	Charger la table DWH_CLIENTS	
	4	Charger la table DWH_SSTYPEPDT	Dédoublonner l'entrant Récupérer les autres identifiants internes Attribuer identifiant interne Alimenter la table en mode UPSERT
	5	Charger la table DWH_PRODUIITS	Dédoublonner l'entrant Récupérer les autres identifiants internes Alimenter la table en mode UPSERT
	6	Charger la table DWH_VENTES	Calcul du champs Quantité Récupérer les autres identifiants interne Si nouvel enregistrement => Charger la table en mode INSERT Si enregistrement existant => Insérer enregistrement à modifier dans la table HISTO_VENTES Supprimer cet enregistrement de la table DWH_VENTES Insérer le nouvel enregistrement dans la table DWH_VENTES
	7	Supprimer les données du mois M-1 de la table DTM_VENTES_TYPEPDT_SECTEUR	
8	Charger la table DTM_VENTES_TYPEPDT_SECTEUR	Calculer la somme des ventes par type de produit et secteur pour les données du mois M-1 Alimenter la table en mode INSERT	
Tous les 1er du mois après le chargement DWH			

## Fournir une analyse des ventes par Secteur, Type de Produits et par mois

### Conceptions parallèles au bon fonctionnement du système

-> Récupérer les fichiers distants dans répertoire machine local

( Comment être sûr de charger les bons fichiers au lancement du calcul DTW ? Cas de fichiers vide ? )

-> Gestion des connexions utilisateurs

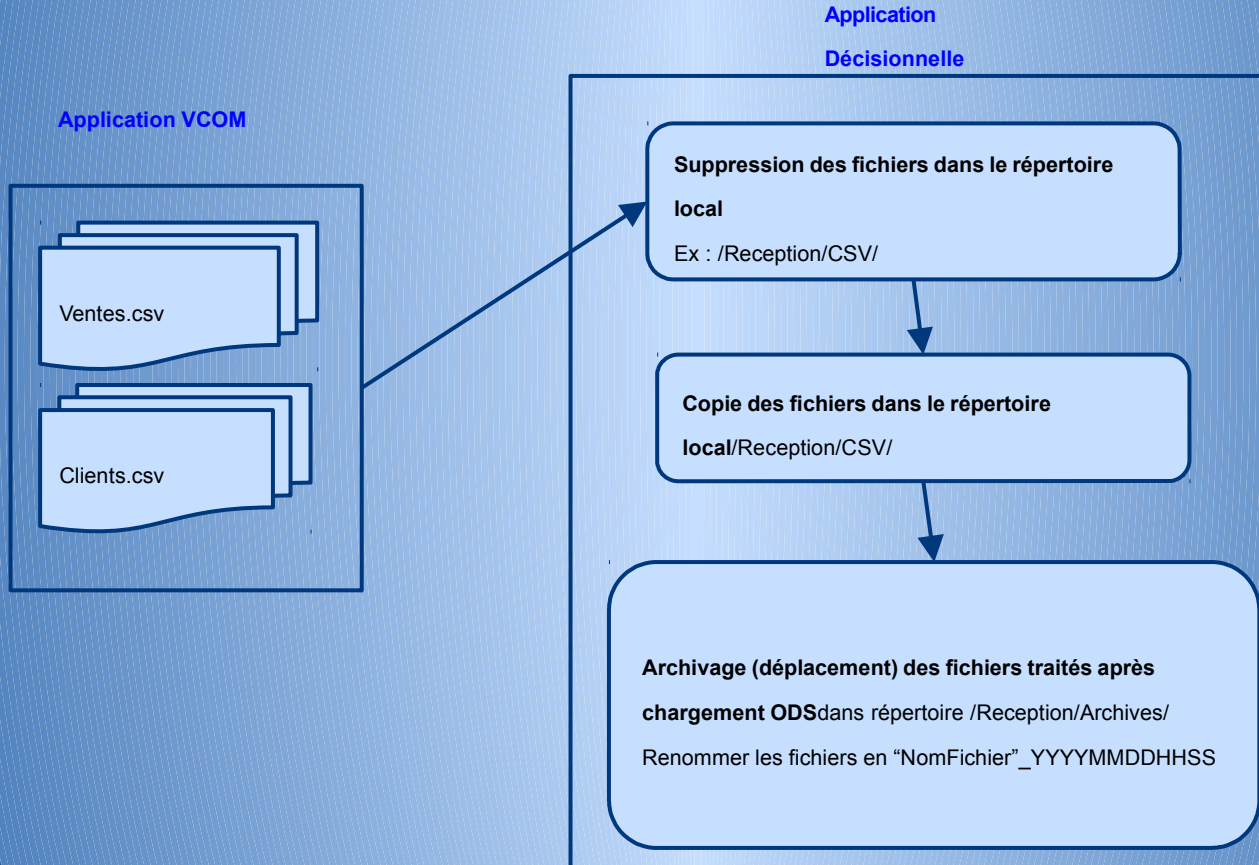
-> Gestion de la sauvegarde des données avant traitement

-> Gestion des logs



## Fournir une analyse des ventes par Secteur, Type de Produits et par mois

### Exemple de Gestion des fichiers reçus



REGLE DE LANCEMENT	N°ETAPE	ACTIONS	LISTE ACTIONS SPECIFIQUES
Tous les jours à minuit	1	Action Système : Couper le service Restitution ET / OU Changer la page d'accès aux restitutions par une page de maintenance	En cas d'erreur => Bloquer la suite de l'exécution / Tracer l'erreur / Réactiver les connexions utilisateurs
	2	Action Système : Sauvegarder les bases DWH et DTM	En cas d'erreur => Bloquer la suite de l'exécution / Tracer l'erreur / Réactiver les connexions utilisateurs
	3	Supprimer les fichiers du répertoire local /Reception/CSV/	En cas d'erreur => Bloquer la suite de l'exécution / Tracer l'erreur / Réactiver les connexions utilisateurs
	4	Copier les Fichiers CSV de l'application VCOM	En cas d'erreur => Bloquer la suite de l'exécution / Tracer l'erreur / Réactiver les connexions utilisateurs
	5	Suppression des données dans l'ODS	En cas d'erreur => Bloquer la suite de l'exécution / Tracer l'erreur / Réactiver les connexions utilisateurs
	6	<b>CHARGEMENT ODS</b>	
	6.1	Charger la table ODS_PRODUITS	Récupérer LOGS
	6.1	Charger la table ODS_VENTES	Récupérer LOGS
	6.1	Charger la table ODS_CLIENTS	Récupérer LOGS
	6.2	Archiver les fichiers CSV reçus	Récupérer LOGS
	6.3	Contrôler que toutes les tables ODS sont bien chargées	En cas d'erreur => Bloquer la suite de l'exécution / Tracer l'erreur / Réactiver les connexions utilisateurs
	7	<b>CHARGEMENT DWH</b>	
	7.1	Charger la table DWH_MARQUES	Dédoublonner l'entrant Attribuer identifiant interne
	7.1	Charger la table DWH_TYPEPDT	Alimenter la table en mode UPSERT Récupérer LOGS
	7.1	Charger la table DWH_SECTEURS	
	7.1	Charger la table DWH_CLIENTS	
	7.2	Charger la table DWH_SSTYPEPDT	Dédoublonner l'entrant Récupérer les autres identifiants internes Attribuer identifiant interne Alimenter la table en mode UPSERT Récupérer LOGS
	7.3	Charger la table DWH_PRODUITS	Dédoublonner l'entrant Récupérer les autres identifiants internes Alimenter la table en mode UPSERT Récupérer LOGS
	7.4	Charger la table DWH_VENTES	Calcul du champs Quantité Récupérer les autres identifiants interne Si nouvel enregistrement => Charger la table en mode INSERT Si enregistrement existant => Insérer enregistrement à modifier dans la table HISTO_VENTES Supprimer cet enregistrement de la table DWH_VENTES Insérer le nouvel enregistrement dans la table DWH_VENTES Récupérer LOGS
7.5	Contrôler que toutes les tables DWH sont bien chargées	En cas d'erreur => Bloquer la suite de l'exécution / Tracer l'erreur	
Tous les 1er du mois	8	Supprimer les données du mois M-1 de la table	En cas d'erreur => Bloquer la suite de l'exécution / Tracer l'erreur
	9	<b>CHARGEMENT DTM</b>	
	9.1	Charger la table DTM_VENTES_TYPEPDT_SECTEUR	Calculer la somme des ventes par type de produit et secteur pour les données du mois M-1 Alimenter la table en mode INSERT Récupérer LOGS
	9.2	Contrôler que toutes les tables DTM sont bien chargées	En cas d'erreur => Bloquer la suite de l'exécution / Tracer l'erreur
	10	Réactiver les connexions utilisateurs	En cas d'erreur => Tracer l'erreur
Tous les jours sauf les 1er du mois	8	Réactiver les connexions utilisateurs	En cas d'erreur => Tracer l'erreur

## Questions utilisateurs

-> Comment se fait il que l'ensemble des ventes de la journée J-1 ne figure pas dans votre application à J+1 ?

-> Je n'ai pas le bon nombre de ventes attendus, votre application est elle bien conforme ?

En effet la vente 308020 n'est pas remontée.

-> Il est 14h30, étant donné qu'une vente n'est pas remontée (dans le DTW) suite à une erreur de saisie, pouvez vous corriger la ligne et rejouer vos traitements ?

-> L'application REFPDT pourrait être fermée de 12h30 à 13h, les fichiers CSV renvoyés le soir à 18h. Quel stratégie peut on mettre en place pour la mise à jour des analyses ?

**Application 1 (REFPDT) ouverte tous les jours de 6h à 22h : Table Produits**

-> Vous est il possible d'intégrer l'ensemble des ventes de 2013 dans le DTW ?

**Application 2 : Fichiers CSV extraits tous les jours à Midi dans un répertoire C:\Extraction\**

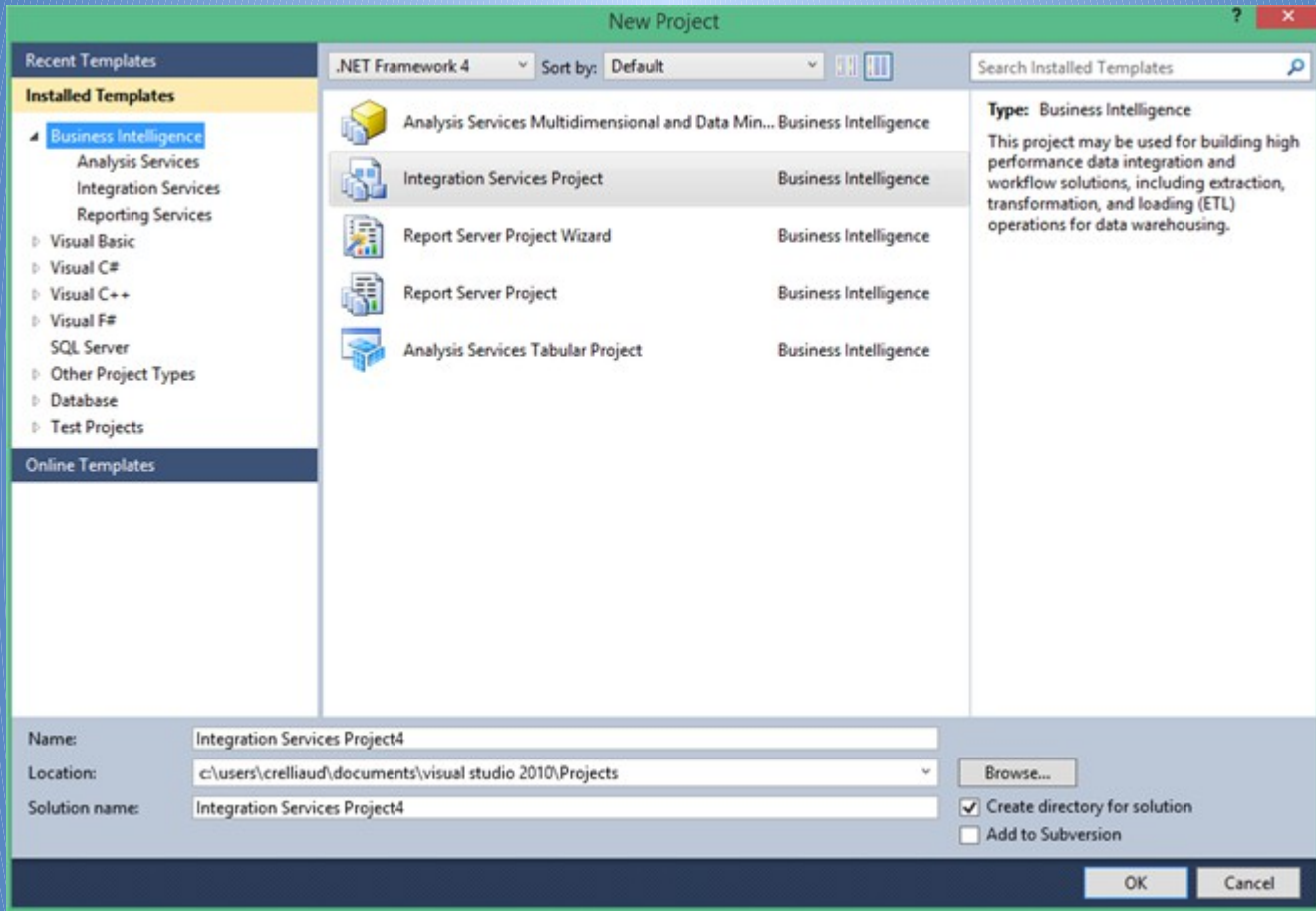
## Autres Questions

-> Il est 17h30, vous apprenez que la source REFPDT ne fonctionne plus. Comment réagissez vous ?

-> L'application REFPDT va évoluer, que faites vous ?

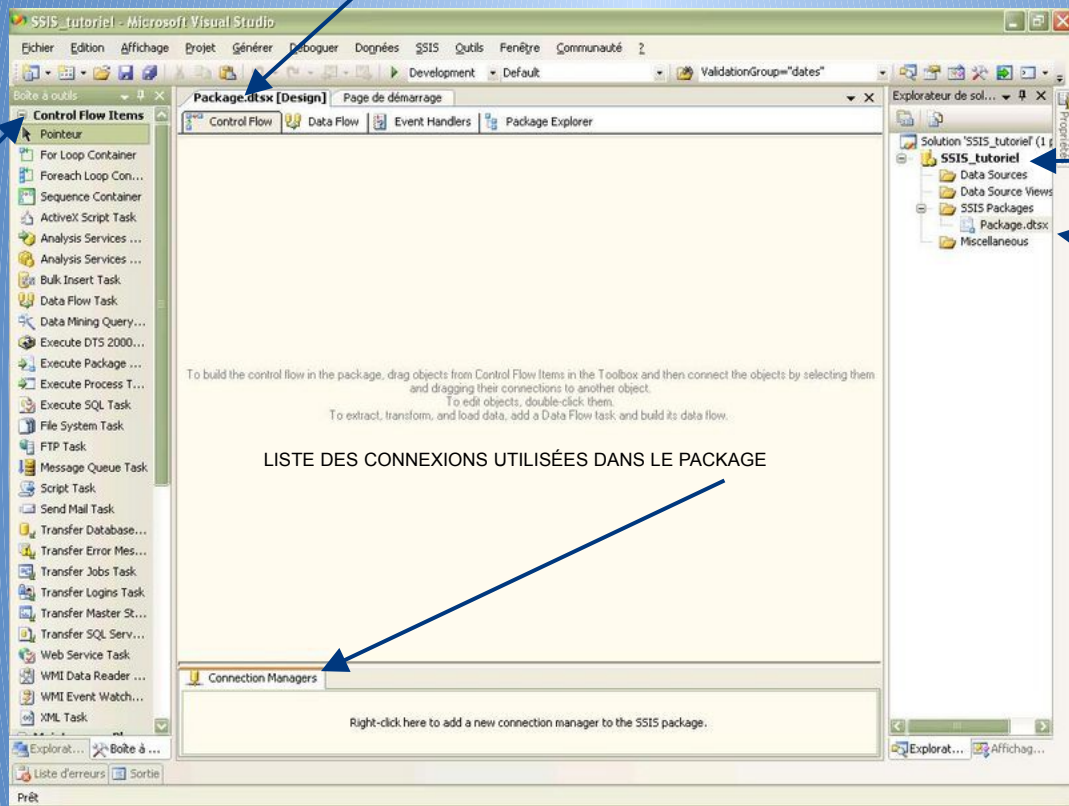
## 7 - ETL SSIS





## LISTE DES TACHES DU PACKAGE ORDONNANCES

LISTE DES TACHES  
DISPONIBLES



PROJET

PACKAGE

LISTE DES CONNEXIONS UTILISÉES DANS LE PACKAGE

### Explorateur de Solution (Déploiement projet)

- Accès aux packages
- Accès aux paramètres et connexions partagées

### Flux de Contrôle

- Unique pour un package
- Point de départ du développement et de l'exécution
- Tâches et conteneurs reliés par des contraintes de précédence

### Flux de Données

- Multiples pour un package
- Tâche du flux de contrôle permettant d'éditer le flux de données dans l'onglet dédié à sa conception
- Contient des sources, des transformations et destinations

### Paramètres

- Gestion des paramètres du package
- Création de références entre les paramètres et les configurations

### Gestionnaire d'événements

- Création d'un flux de contrôle au niveau package ou tâche, déclenché par un événement (OnError, OnWarning, OnProgress...)

### Explorateur de package

- Permet de modifier des propriétés et de supprimer des éléments

### Gestionnaire de connexions

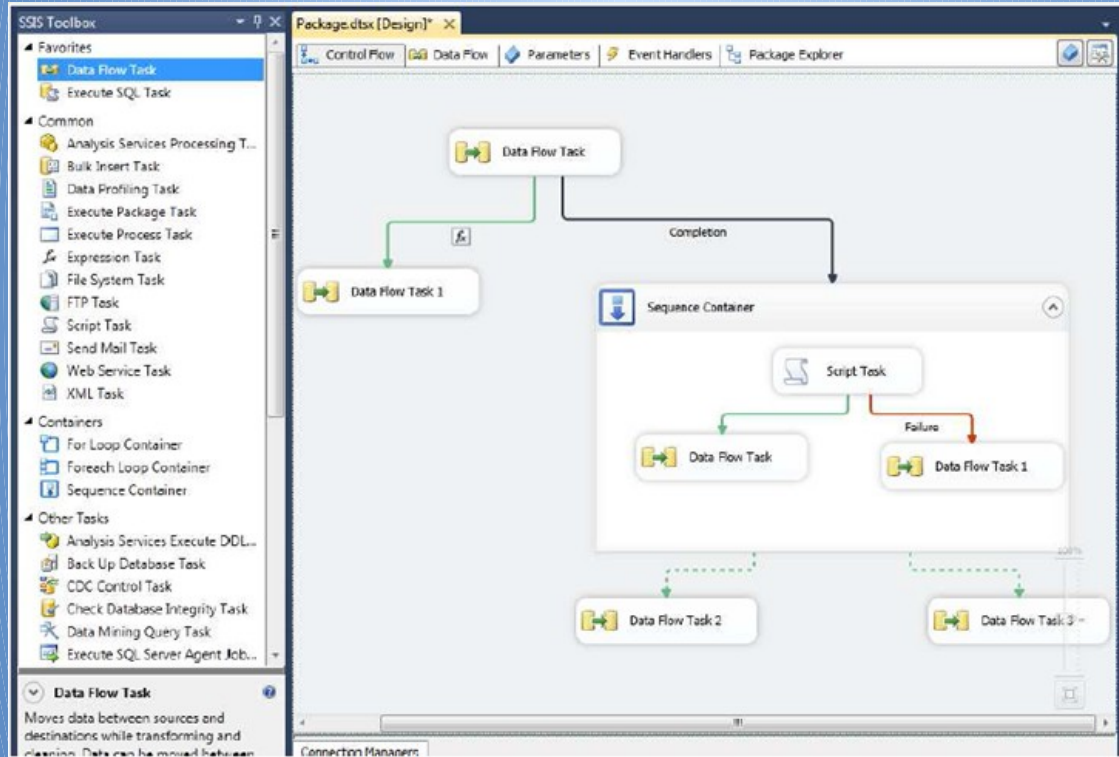
- Concerne le package
- connexions aux sources et destinations et aux autres ressources (FTP, SMTP...)

### Boîte à outils SSIS

- Flux de contrôle: tâches et conteneurs
- Flux de données: sources, transformations et destinations

### Le Menu SSIS

- Travaille hors connexion





## DIFFÉRENTS COMPOSANTS

### CONTENEURS

- ▶ Flux de travail répétitif (boucles For, ForEach )
- ▶ Sous ensemble de flux de contrôle (regroupement des tâches, gestion de transaction, d'erreurs...)

### TÂCHES

- ▶ Flux de données
- ▶ Autres tâches (SQL, fichiers, traitement OLAP, FTP...)

### CONNECTEURS

- ▶ Ordonnancement de tâches
- ▶ Contraintes de précedence

## CONTENEURS

### BOUCLE FOR

- ▶ expression d'initialisation : @Counter = 0
- ▶ expression d'évaluation : @Counter < 4
- ▶ expression d'itération : @Counter = @Counter + 1

### BOUCLE FOREACH (ÉNUMÉRATEUR)



- ▶ Foreach ADO Enumerator (lignes d'une table)
- ▶ Foreach ADO.NET (liste de table d'un schéma)
- ▶ Foreach File (liste des fichier d'un dossier)
- ▶ Foreach à partir d'une variable objet

### SÉQUENCE













- ▶ Sous ensemble du flux de contrôle

## TACHES LES PLUS UTILISÉES

### ▲ Favoris

-  Tâche de flux de données
-  Tâche d'exécution de requêtes SQL

### ▲ Courant

-  Tâche de profilage des données
-  Tâche de script
-  Tâche de service Web
-  Tâche de système de fichiers
-  Tâche de traitement SQL Server Analysis Serv...
-  Tâche d'exécution de package
-  Tâche d'exécution de processus
-  Tâche d'expression
-  Tâche d'insertion en bloc
-  Tâche Envoyer un message
-  Tâche FTP
-  Tâche XML

## TÂCHES

### FLUX DE DONNÉES

- Extraction, transformation, chargement

### PRÉPARATION DES DONNÉES

- Système de fichier, FTP, XML...

### FLUX DE TRAVAIL

- Exécution de packages, de processus, envoi de messages...

### SQL SERVER

- Exécution de requêtes SQL,  
insertion en bloc, Transfert d'objets ou de base de données...

### SCRIPT

- Environnement Microsoft Visual Studio Tools for Applications (VSTA), Microsoft Visual Basic 2010 et Microsoft Visual C# 2010

## TÂCHES

### ANALYSIS SERVICES

- ▶ Exécution d'instruction DDL, traitement d'objets...

### SQL SERVER ADMINISTRATION

- ▶ Sauvegarde, reconstruction ou réorganisation d'index, exécution T-SQL

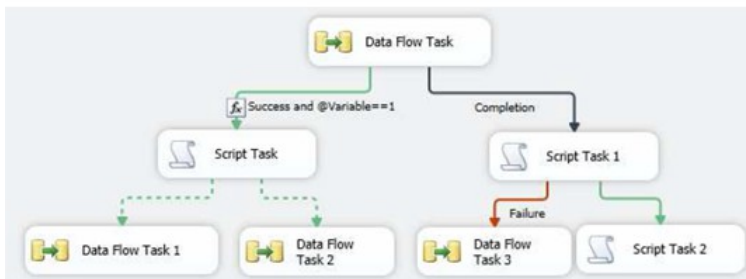
### TACHE D'EXPRESSION COMPLEXE

- ▶ Affectation une valeur à une variable en cours d'exécution (vs Script)

# LES CONTRAINTES DE PRÉCÉDENCE

## CRÉATION

- ▶ Lier les éléments du flux de contrôle par leur connecteur
- ▶ Spécification de l'ordre et des conditions d'exécution des tâches



Constraint options

Evaluation operation: 

Constraint
Constraint
Expression
Expression and Constraint
Expression or Constraint

Value:

Expression:

Multiple constraints

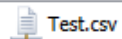
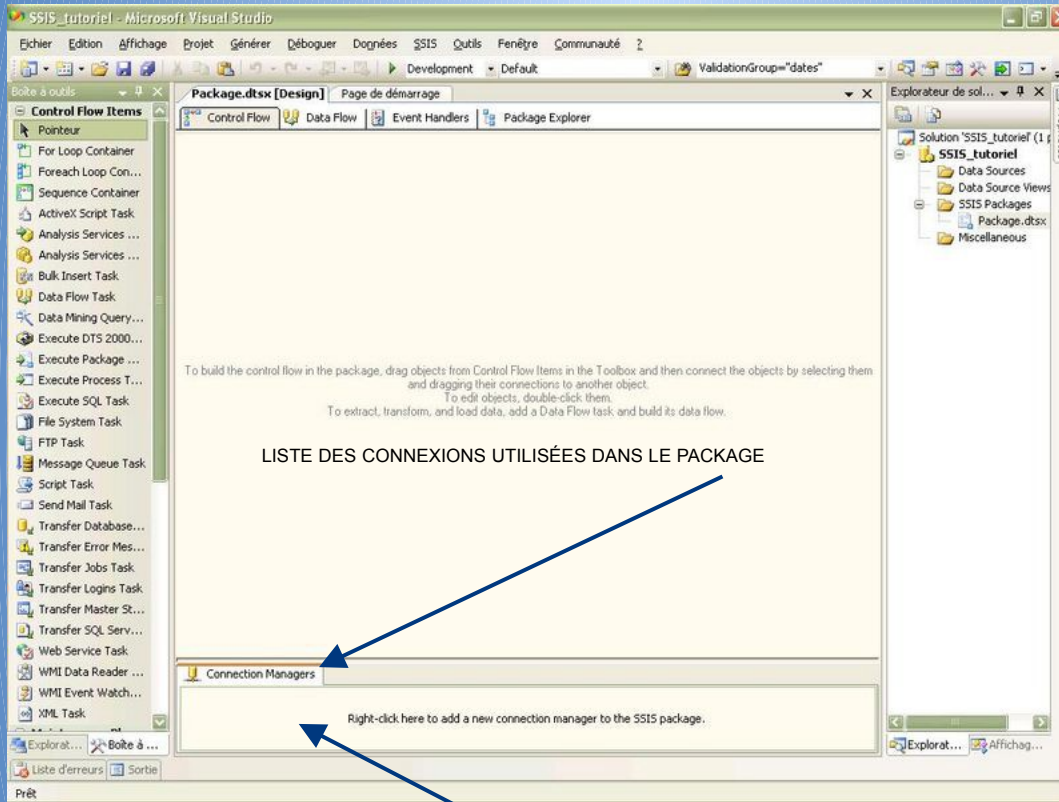
If the constrained task has multiple constraints, you can choose to control the execution of the constrained task.

Logical AND. All constraints must evaluate to True

Logical OR. One constraint must evaluate to True

## CONFIGURATION

- ▶ Evaluation du résultat d'exécution du traitement précédent => Succès ou échec ou terminé
- ▶ Conditionnement de l'exécution du traitement suivant => Expression et/ou contraintes (utilisation de variables)

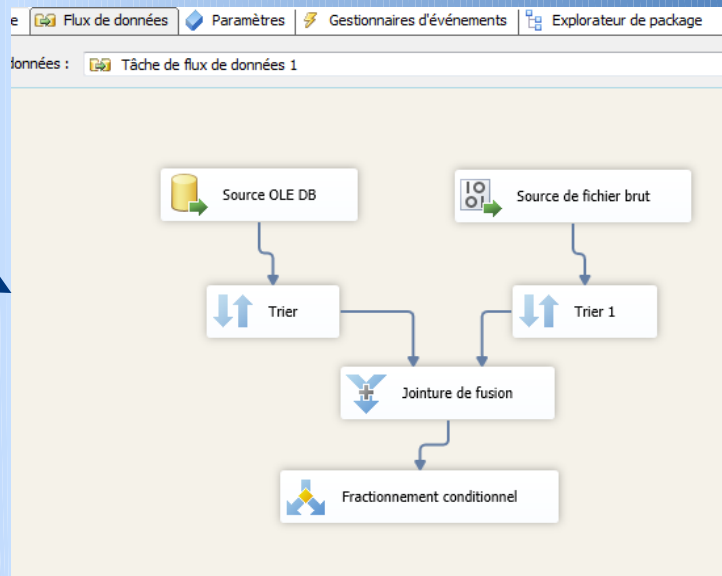
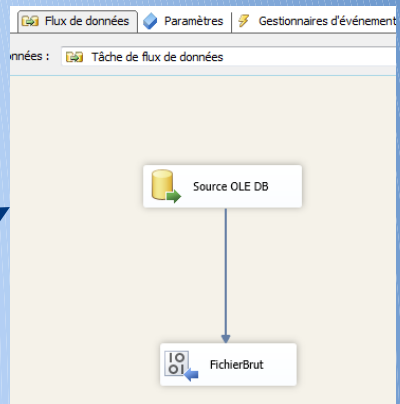
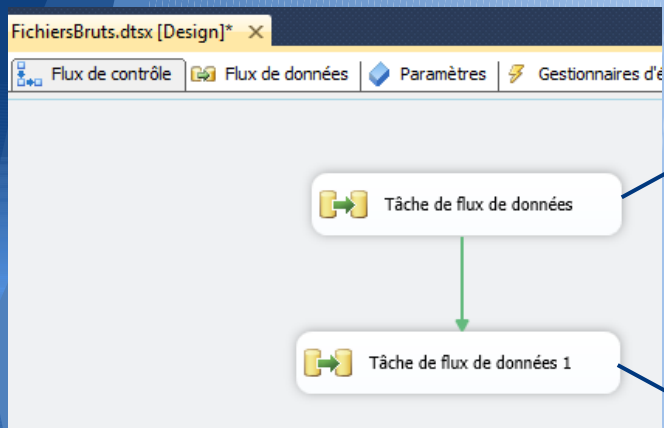


Test.csv



(projet) SI-SANJAY.AdventureWorksLT2008R2

## FLUX DE DONNEES





## FLUX DE DONNÉES

### SOURCES

- ▶ Extraction des données de fichiers ou de bases de données

### TRANSFORMATIONS

- ▶ Implémentation des règles métier









### DESTINATIONS

- ▶ Chargement des données dans des fichiers ou bases de données














### FLOW DE DONNÉES

- ▶ Enchaînement des composants
- ▶ Gestion des sorties d'erreurs

# LES SOURCES ET DESTINATIONS

-  ADO NET Source
-  CDC Source
-  Excel Source
-  Flat File Source
-  ODBC Source
-  OLE DB Source
-  Raw File Source
-  XML Source



-  ADO NET Destination
-  Data Mining Model Training
-  DataReader Destination
-  Dimension Processing
-  Excel Destination
-  Flat File Destination
-  ODBC Destination
-  OLE DB Destination
-  Partition Processing
-  Raw File Destination
-  Recordset Destination
-  SQL Server Compact Destination
-  SQL Server Destination

## TRANSFORMATIONS

### TRANSFORMATIONS DE LIGNES

- ▶ Appliquée à chaque ligne de l'entrée de transformation
- ▶ Mise à jour des valeurs de colonnes, création de nouvelles colonne

### TRANSFORMATIONS D'ENSEMBLE DE LIGNES

- ▶ Création d'ensembles de lignes triées ou agrégées















### FRACTIONNEMENT OU FUSION DE LIGNES

- ▶ Distribution de lignes vers plusieurs sorties (conditionnelles ou non)
- ▶ Copies des lignes en entrée de transformation (mise en cache)
- ▶ Jointures diverses de plusieurs entrées en une
- ▶ Opération de recherche

### AUTRES TRANSFORMATIONS

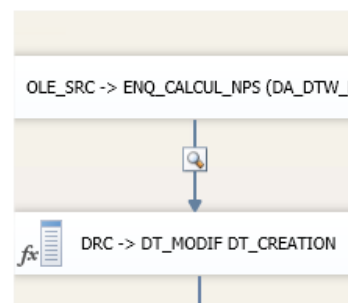
- ▶ Comptage des lignes
- ▶ Dimension à variation lente

#### ◀ Courant

-  Agrégation
-  Colonne dérivée
-  Commande OLE DB
-  Composant Script
-  Conversion de données
-  Dimension à variation lente
-  Fractionnement conditionnel
-  Fusionner
-  Jointure de fusion
-  Multidiffusion
-  Nombre de lignes
-  Recherche
-  Trier
-  Unir tout

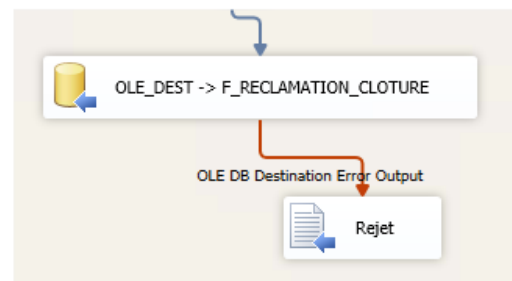
## CONNEXION DES COMPOSANTS

- ▶ Entre sortie d'un composant et entrée du suivant
- ▶ Editeur de chemin d'accès
  - = > Métadonnées, Visionneuse

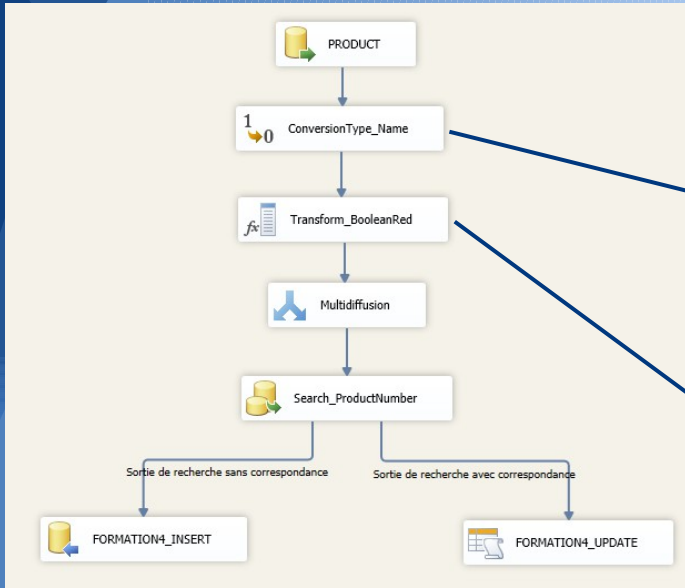


## GESTION DES ERREURS

- ▶ Sortie d'erreur d'un composant
- ▶ Types d'erreurs:
  - => Erreurs de conversion (truncation, cast...)
  - => Erreurs d'évaluation d'expression
  - => Erreurs de recherche (valeur non trouvée)
- ▶ Gestion
  - => Défaillant
  - => Ignorer l'échec
  - => Réacheminer la ligne



## EXEMPLE D'UNE TACHE



Colonne d'entrée disponibles

- Nom
- ProductID
- Name
- ProductNumber
- Color
- StandardCost

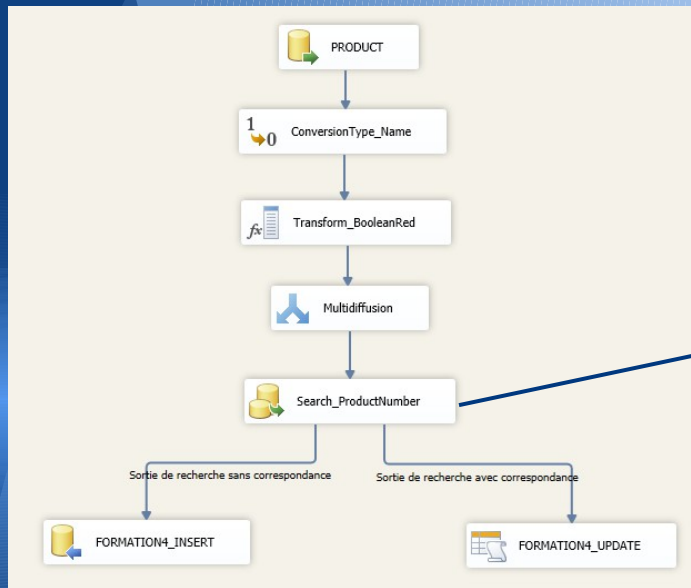
Colonne d'entrée	Alias de sortie	Type de données	Longue
ProductNumber	Copie de ProductNumber	entier non signé (4 bits) [D...	

- ⊕ Variables et paramètres
- ⊕ Colonne

- ⊕ Fonctions mathématiques
- ⊕ Fonctions de chaîne
- ⊕ Fonctions Date/Heure
- ⊕ Fonctions NULL
- ⊕ Casts de type
- ⊕ Opérateurs

Description :

Nom de la colonne dérivée	Colonne dérivée	Expression	Type de données	Longueur
BooleanRed	<ajouter comme nouvelle...	(DT_WSTR,50)Color == "Red" ? "Rouge" : Color == "Black" ? "Noir" : "Autre couleur"	chaîne Unicode [DT_WSTR]	13



Général  
Connexion  
**Colonnes**  
Avancé  
Sortie d'erreur

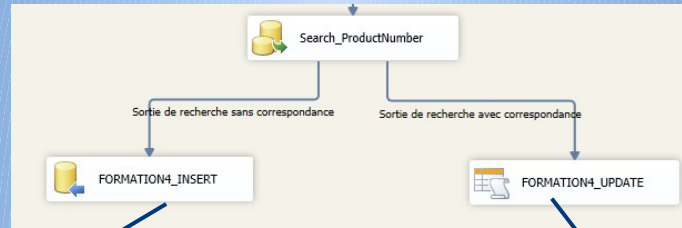
Colonnes d'entrée d...

Nom
ProductID
Name
ProductNumber
Color
StandardCost
Name_Varchar
BooleanRed

Colonnes de recherche disponibles

<input checked="" type="checkbox"/>	Nom	Index
<input checked="" type="checkbox"/>	ProductID	
<input type="checkbox"/>	Name	
<input type="checkbox"/>	ProductNumber	
<input type="checkbox"/>	Color	
<input type="checkbox"/>	StandardCost	
<input type="checkbox"/>	DateDebut	
<input type="checkbox"/>	DateFin	

Colonne de recherche	Opération de recherche	Alias de sortie
ProductID	<ajouter comme nouvelle colonne>	ProductID



Gestionnaire de connexion

**Mappages**

Sortie d'erreur

Colonnes d'entrée d...

- Nom
- ProductID
- Name
- ProductNumber
- Color
- StandardCost
- Name\_Varchar
- BooleanRed

Colonnes de destina...

- Nom
- ProductID
- Name
- ProductNumber
- Color
- StandardCost
- DateDebut
- DateFin

Colonne d'entrée	Colonne de destination
<ignorer>	ProductID
Name	Name
ProductNumber	ProductNumber
Color	Color
StandardCost	StandardCost
<ignorer>	DateDebut
<ignorer>	DateFin

Gestionnaires de connexion

Propriétés du composant

Mappage de colonnes

Propriétés d'entrée et de sortie

Spécifiez les propriétés avancées pour le composant de flux de données.

Propriétés :

**Propriétés communes**

- ComponentClassID
- ContactInfo
- Description
- ID
- IdentificationString
- IsDefaultLocale
- LocaleID
- Name
- PipelineVersion
- UsesDispositions
- ValidateExternalMetadata
- Version

**Propriétés personnalisées**

- CommandTimeout
- DefaultCodePage
- SqlCommand

Éditeur de valeur de chaîne

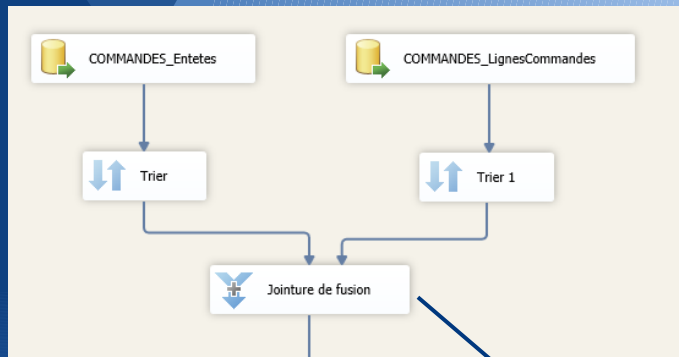
Valeur de chaîne :

```

UPDATE [SalesLT].[ProductCIBLE_FORMATION4]
SET [Name] = ?
    [Color] = ?
    [StandardCost] = ?
WHERE ProductID = ?
          
```

OK    Annuler

UPDATE [SalesLT].[ProductCIBLE\_FORMATION4] SET [Name] = ?



Type de jointure : Jointure externe gauche Échanger les entrées

**Trier**

<input checked="" type="checkbox"/>	Nom	Ordre	Clé de j...
<input checked="" type="checkbox"/>	SalesOrderID	1	<input checked="" type="checkbox"/>
<input checked="" type="checkbox"/>	OrderDate	0	<input type="checkbox"/>
<input checked="" type="checkbox"/>	AccountNumber	0	<input type="checkbox"/>
<input checked="" type="checkbox"/>	Status	0	<input type="checkbox"/>

**Trier 1**

<input type="checkbox"/>	Nom	Ordre	Clé de j...
<input type="checkbox"/>	SalesOrderID	1	<input checked="" type="checkbox"/>
<input checked="" type="checkbox"/>	SalesOrderDetailID	0	<input type="checkbox"/>
<input checked="" type="checkbox"/>	OrderQty	0	<input type="checkbox"/>
<input checked="" type="checkbox"/>	LineTotal	0	<input type="checkbox"/>

Entrée	Colonne d'entrée	Alias de sortie
Trier	SalesOrderID	SalesOrderID
Trier	OrderDate	OrderDate
Trier	AccountNumber	AccountNumber
Trier	Status	Status
Trier 1	SalesOrderDetailID	SalesOrderDetailID
Trier 1	OrderQty	OrderQty
Trier 1	LineTotal	LineTotal





Général  
**Collection**  
 Mappage de variables  
 Expressions

Éditeur de boucle Foreach

Enumerator	Énumérateur Foreach File
Expressions	

**Enumerator**  
 Spécifie le type d'énumérateur.

Configuration de l'énumérateur

Dossier :

Fichiers :

Récupérer le nom de fichier

Conteneur de boucles Foreach

Tâche de système de fichiers

Général  
 Expressions

**Connexion de destination**

IsDestinationPathVariable	True
DestinationVariable	User::vCheminDestination
OverwriteDestination	True

**Connexion source**

IsSourcePathVariable	True
SourceVariable	User::vCheminSource

**Général**

Name	Tâche de système de fichiers
Description	Tâche de système de fichiers

**Opération**

Operation	Copier le fichier
-----------	-------------------

## 8 - QUELQUES BONNES PRATIQUES & STRATÉGIE DE RECETTE



-> Utiliser une charte de nommage (Packages / Variables / Taches / Destinations ...)

-> Un libellé d'un traitement doit permettre d'en comprendre son but

-> Un coup d'oeil doit suffire à comprendre une tâche

-> Seules les données "utiles" doivent être transférées dans le flux de données

(gain de performances)

-> Limiter la volumétrie des données transférées en privilégiant les jointures internes directement via le SGBD (requête entrante)

-> Un ETL est optimisé lorsqu'il travaille sur des fichiers avec son format interne (fichiers bruts SSIS). Dans l'idéal, créer des fichiers bruts à partir des sources SQL avant traitement

-> Des fichiers bruts doivent être créés pour toutes les sources de référence (afin de limiter les accès en base avec des requêtes SQL lourdes).

-> Plusieurs traitements simples ordonnancés valent mieux qu'un unique traitement lourd et difficilement maintenable

Mode opératoire

1. Compter le nombre d'enregistrements en source
2. Lancer un chargement ODS
3. Compter le nombre d'enregistrements dans les tables cibles
4. Vérifier les données (vue macro) => tous les champs sont ils bien chargés ? Erreur mapping ?

Source	Requête	Nb Enregistrements	Cible	Requête	Nb Enregistrements	Ecart Nb Enregistrements	Contrôle Données (select *)	ETAT (OK / KO)
REF ASSO	ASSO_DOUBLONS.csv	13	ODS_ASSO	select count(*) from ASSO_DOUBLONS	13	0	OK	OK
REF ASSO	ASSO_ASSOCIATION.csv	5962	ODS_ASSO	select count(*) from ASSO_ASSOCIATION	5962	0	Attention : - Peu de quartiers présents en entrant - Champs Site & Mail : Seulement 50 caracteres prévus	EC
REF ASSO	ASSO_ADRESSE.csv	6192	ODS_ASSO	select count(*) from ASSO_ADRESSE	6192	0	OK	OK
REF ASSO	ASSO_PERSONNE.csv	10034	ODS_ASSO	select count(*) from ASSO_PERSONNE	10034	0	OK par rapport au CSV Mais des personnes attendues non présentes dans le CSV : (16092,16094,16095,16096,16097,16098,16099,16100,16101)	OK
REF ASSO	ASSO_ADR_PERSONNE.csv	9946	ODS_ASSO	select count(*) from ASSO_ADR_PERSONNE	9946	0	OK	OK
REF ASSO	ASSO_VIE_ASSOCIATION.csv	4592	ODS_ASSO	select count(*) from ASSO_VIE_ASSOCIATION	4592	0	Attention : Dates chargées avec le caractère "	OK
REF ASSO	ASSO_QUARTIER.csv	7	ODS_ASSO	select count(*) from ASSO_QUARTIER	7	0	A voir pour la suite du chargement	OK
REF ASSO	ASSO_SOUS_QUARTIER.csv	31	ODS_ASSO	select count(*) from ASSO_SOUS_QUARTIER	31	0	OK	OK
REF ASSO	ASSO_PUBLIC.csv	8	ODS_ASSO	select count(*) from ASSO_PUBLIC	8	0	OK	OK
SUB	select count(*) from T_DOSSIER	7892	ODS_ASSO	select count(*) from SUB_T_DOSSIER	7892	0	OK	OK
SUB	select count(*) from T_DONNEES_DOSSIER	388905	ODS_ASSO	select count(*) from SUB_T_DONNEES_DOSSIER	388905	0	OK	OK

OK : ID\_TIERS 4374 et 1986 absents car selection du numero de tiers de version max

SUB	select count(*) from T_DOSSIER_COMMISSION	5076	ODS_ASSO	select count(*) from SUB_T_DOSSIER_COMMISSION	5076
SUB	select count(*) from T_CONTACT	7316	ODS_ASSO	select count(*) from SUB_T_CONTACT	7316
ASTRE	select count(*) from FAIT_MARCHE	1540	ODS_ASSO	select count(*) from ASTRE_FAIT_MARCHE	1540

Chargement ODS

Chargement DWH

Sauvegarde-RetourArriere

Mode Alimentation Tables

Fusion des Associations

Purge des Associations

Gestion des erreurs

Mode opératoire

1. Compter le nombre d'enregistrements dans les tables
2. Vérifier les données (vue macro) -> tous les champs sont ils bien chargés ? Erreur mapping ?

Cible	Table	Nb Enregistrement ODS	Nb Enregistrements	Contrôle Données (select *)	ETAT (OK / KO)
DWH_ASSO	ASSOCIATIONS	5962	5942	KO : Champ NB_EMPLOI_AIDES mal alimenté CHAMPS DATE_JO / NAF / NB_ADHERENTS / NB_BENEVOLES / NB_SALARIES / PROPRIETAIRE / LOCATAIRE / COMMISSAIRE jamais renseignés => OK	OK
DWH_ASSO	ASSO_PRINCIPALE	13	13	OK	OK
DWH_ASSO	VIE_ASSOCIATION	4592	4589	OK KO 3 enregistrements non attendus sont présents dans la table PERSONNES	OK
DWH_ASSO	PERSONNES	10034	9991	OK	OK
DWH_ASSO	ASSO_LIEN_PERS	10034	10064	OK	EXCEPT
DWH_ASSO	ASSO_ADRESSES	6192	6170	OK	OK
DWH_ASSO	PERS_ADRESSES	9946	9937	OK	EXCEPT
DWH_ASSO	PUBLICS	8	8	OK	OK
DWH_ASSO	QUARTIER	7	7	OK	OK
DWH_ASSO	ASSO_LIEN_PUBLIC	N/A	6529	OK KO	OK KO
DWH_ASSO	ASSO_LIEN_QUARTIER	N/A	7	(Notion de Tous Quartiers non prise en compte)	OK
DWH_ASSO	SOUS_QUARTIER	31	31	OK	OK
DWH_ASSO	DOSSIER	7892	7780	Erreur alimentation : CHAMP DATE_PROJET mal alimenté Erreur sur les montants (demande / elu / accorde) : Les montants sont au format entier et pas au format numeric, les decimales sont perdues Erreur nommage : Champ OBJET_CULTUEL nommé OBJET_CULTUREL	OK
DWH_ASSO	DISPOSITIF	45	44	OK	OK
DWH_ASSO	ENVELOPPE	67	42	OK	OK
DWH_ASSO	DOSSIER_LIEN_PUBLIC	N/A	1629	OK	OK
DWH_ASSO	DOSSIER_LIEN_QUARTIER	N/A	632	OK	OK
DWH_ASSO	MARCHE_LOTS	22889	80	OK	OK
DWH_ASSO	MARCHE	1540	1540	OK	OK

Mode opératoire

1. Sauvegarder le nombre de lignes des tables DWH
  2. Lancer le package DWH
  3. Vérifier que l'ensemble des tables sont sauvegardées avec le bon nombre de lignes
  4. Vérifier les données dans les tables sauvegardées (select \*)
  5. Vérifier que le nombre de lignes à varier dans les tables DWH
- 6.Lancer le retour arrière et vérifier que les tables DWH contiennent bien leur état initial

Cible	Requetes	Nb Enregistrements (etape 1)	RETOUR ARRIERE(OK / KO) (etape 6)	Cible	Requetes	Nb Enregistrements (etape 1)
DWH_ASSO	select count(*) from ASSOCIATIONS	5942		DWH_ASSO_SAVE	select count(*) from ASSOCIATIONS	5942
DWH_ASSO	select count(*) from ASSO_PRINCIPALE	13		DWH_ASSO_SAVE	select count(*) from ASSO_PRINCIPALE	13
DWH_ASSO	select count(*) from VIE_ASSOCIATION	4591		DWH_ASSO_SAVE	select count(*) from VIE_ASSOCIATION	4591
DWH_ASSO	select count(*) from PERSONNES	9991		DWH_ASSO_SAVE	select count(*) from PERSONNES	9991
DWH_ASSO	select count(*) from ASSO_LIEN_PERS	10064		DWH_ASSO_SAVE	select count(*) from ASSO_LIEN_PERS	10064
DWH_ASSO	select count(*) from ASSO_ADRESSES	6170		DWH_ASSO_SAVE	select count(*) from ASSO_ADRESSES	6170

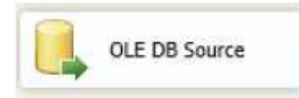
Scénario	Résultat attendu	Résultat Obtenu
Exécuter un lancement ODS avec un fichier CSV manquant	Erreur fatale à l'exécution du chargement Tampon => Blocage des traitements => Pas de lancement de retour arrière	
Exécuter un lancement ODS avec un fichier CSV vide	Erreur fatale à l'exécution du chargement Tampon => Blocage des traitements => Pas de lancement de retour arrière	
Renommer une table ODS SUB en TABLE_wait	Erreur fatale à l'exécution du chargement Tampon => Blocage des traitements => Pas de lancement de retour arrière	
Faire échouer la sauvegarde DWH	Erreur fatale à l'exécution de la sauvegarde DWH => Blocage des traitements => Pas de lancement de retour arrière	
Renommer une table DWH en TABLE_wait	Erreur fatale à l'exécution du chargement DWH => Blocage des traitements => Lancement automatique du retour arrière	

## 9 - EXEMPLES DE COMPOSANTS





## SOURCE OLE DB



- ▶ Extraction d'une base de données relationnelle compatible OLE DB
- ▶ Modes d'accès:
  - => Table ou vue
  - => Commande SQL (paramétrée)
  - => Commande SQL stockée dans une variable

Data access mode:

Table or view

Table or view

Table name or view name variable

SQL command

SQL command from variable

# COLONNE DÉRIVÉE



- ▶ Création de nouvelles colonnes ou modification de colonnes existantes
- ▶ Utilisation d'expressions applicables aux colonnes d'entrée:

A screenshot of the 'Derived Column' dialog box in a software application. The dialog is split into two panes. The left pane shows a tree view with 'Variables and Parameters' and 'Columns'. Under 'Columns', there are five items: 'CalendarQuarter', 'CalendarYear', 'TimeKey', and 'EnglishMonthName'. The right pane is titled 'String Functions' and contains a list of functions with their syntax: CODEPOINT(«character\_expression»), FINDSTRING(«character\_expression», «string», «occurrence»), HEX(«integer\_expression»), LEFT(«character\_expression», «number»), LEN(«character\_expression»), LOWER(«character\_expression»), LTRIM(«character\_expression»), REPLACE(«character\_expression», «search\_expression», «replace\_expression»), and REPLICATE(«character\_expression», «times»). Below the list is a 'Description:' field.

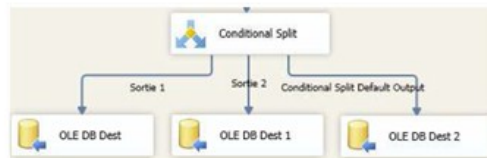
Derived Column Name	Derived Column	Expression	Data Type	
CalendarQtrID	<add as new column>	CalendarYear * 100 + CalendarQuarter	four-byte signed integer [...]	
CalendarQtrName	<add as new column>	(DT_WSTR,4)CalendarYear + "-" + (DT_WSTR,2)Calendar...	Unicode string [DT_WSTR]	7

## FRACTIONNEMENT DU FLUX



- ▶ Chaque ligne entrée est systématiquement dirigée vers chaque sortie
- ▶ Copie logique des données

## FRACTIONNEMENT CONDITIONNEL



- ▶ Chaque ligne entrée est dirigée de façon conditionnelle vers une sortie
- ▶ Si aucune condition n'est vérifiée => Sortie par défaut

Order	Output Name	Condition
1	Sortie 1	ErrorNumber == 1
2	Sortie 2	ErrorNumber == 2
3	Sortie 3	ErrorNumber == 3

## FUSION DES FLUX

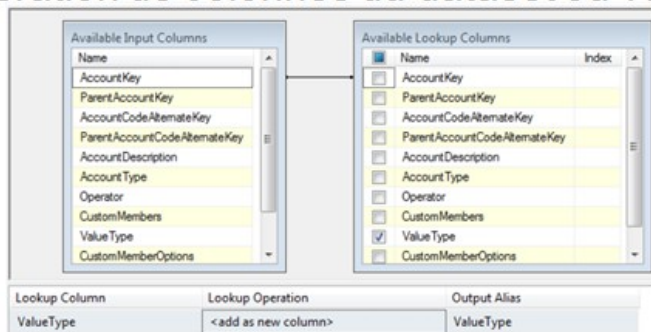


- ▶ Union all : Combinaison de plusieurs entrées en une sortie unique, les métadonnées doivent être identiques
- ▶ Merge : Même chose que Union All mais sur deux entrées triées, option de suppression des doublons (valeurs de clefs de tri identiques)
- ▶ Merge Join : Union de deux entrées triées par une jointure (left / right / inner join)
  
- ▶ Attention transformation semi-bloquante, performance réduite, privilégier le composant de recherche si possible

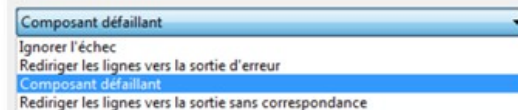
## RECHERCHE



- ▶ Recherche dans un dataset de référence
- ▶ Equijointure entre les données des colonnes d'entrée et celles du dataset
- ▶ Récupération de colonnes du dataset ou vérification de correspondance



Spécifier comment gérer les lignes sans entrées correspondantes



### ▶ Sortie:

Lignes avec 1 correspondance

Lignes sans correspondance (par défaut = erreur)

Lignes avec plusieurs correspondances (première correspondance restituée par la requête de recherche)

## RECHERCHE (CACHE)



- ▶ Le dataset est stocké en mémoire pendant l'exécution du package
  - Cache complet (par défaut):** Dataset chargé dans le cache avant l'exécution du flux de données – le cache est figé (recherche rapide – table de référence peu volumineuse)
  - Cache partiel:** Dataset chargé dans le cache avant l'exécution du flux de données - le cache est mis à jour (options de paramétrage du cache – table de référence volumineuse)
  - Aucun cache:** accès à la base pour chaque ligne traitée (table de référence évolutive – peu de mémoire dispo)

- ▶ Gestionnaire de connexions OLE DB + Cache complet:  
Dataset généré à partir d'une table ou requête SQL et stocké dans le cache
- ▶ Gestionnaire de connexions du cache + Cache complet:  
Dataset généré à partir d'une table ou requête SQL et stocké dans un fichier cache (transformation du cache)  
Cache persistant dans un fichier (.caw):

Cache mode

Full cache

Partial cache

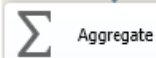
No cache

Connection type

Cache connection manager

OLE DB connection manager

## AGGREGATION



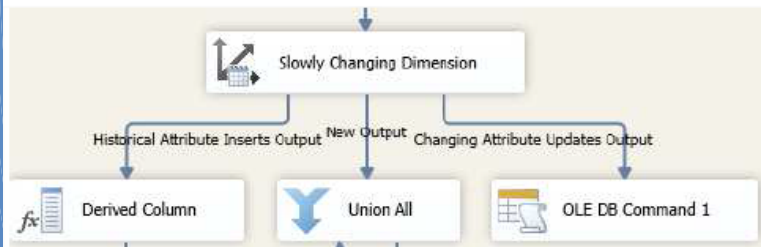
- ▶ Création de groupes de données (GROUP BY)
- ▶ Fonctions d'agrégation des données (COUNT, SUM, MAX....)
- ▶ Peut générer plusieurs sorties avec plusieurs agrégats
- ▶ Transformation bloquante !!



## DIMENSION A VARIATION LENTE



- ▶ **Attribut fixe (Type 0)**  
La valeur de la colonne ne doit pas changer  
Seules les nouvelles lignes sont insérées => « Nouvelle sortie »
- ▶ **Modification d'attribut (Type 1)**  
Les enregistrements existants sont remplacés et les nouveaux insérés  
Deux sortie générées: « Sortie de mises à jour d'attribut de validation » et « Nouvelle sortie »
- ▶ **Attribut d'historique (Type 2)**  
Création de nouveaux enregistrements et mise à jour des enregistrements existants (expiré) Deux sortie générées: « Sortie d'insertions d'attribut d'historique » et « Nouvelle sortie »
- ▶ **Uniquement SQL Server, Performance limitée (faible volumétrie)**



Select a change type for slowly changing dimension columns:

Dimension Columns	Change Type
AccountDescription	Fixed attribute
AccountType	Changing attribute
CustomMemberOptions	Historical attribute
CustomMembers	Fixed attribute
Operator	Changing attribute

## COMMANDE OLE DB



- ▶ Exécution d'une commande SQL pour chaque ligne du flux de données entrant

Seule possibilité d'update de la base dans le flux de données

Problèmes de performance sur volume important

Utilisé pour des requêtes nécessitant des paramètres

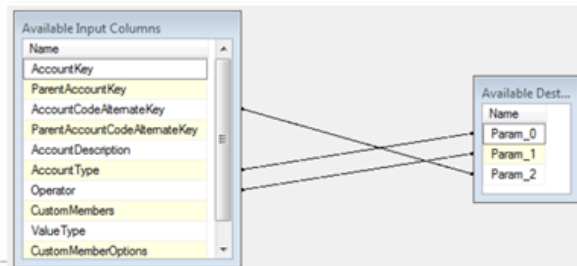
- ▶ Exemples:

```
UPDATE [dbo].[DimProduit]
```

```
SET [Famille] = ?,[Produit] = ?,[SousFamille] = ?,[SousFamilleCode] = ?
```

```
WHERE [ProduitCode] = ? AND [Valide] ='1'
```

- ▶ Mappage des colonnes et paramètres:



## DESTINATION OLE DB



- ▶ Chargement de données dans des bases compatibles OLE DB

- ▶ Mode d'accès:

Data access mode:

- Table or view - fast load
- Table or view
- Table or view - fast load
- Table name or view name variable
- Table name or view name variable - fast load
- SQL command

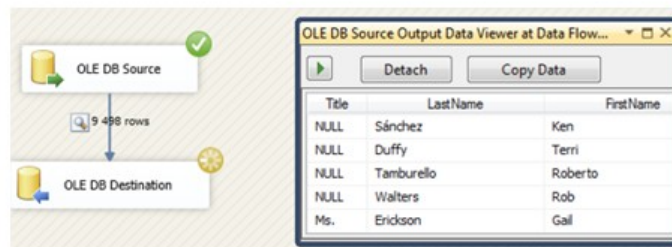
- ▶ Option de chargement:

<input type="checkbox"/> Keep identity	<input checked="" type="checkbox"/> Table lock
<input type="checkbox"/> Keep nulls	<input checked="" type="checkbox"/> Check constraints
Rows per batch:	<input type="text"/>
Maximum insert commit size:	<input type="text" value="2147483647"/>

## EXÉCUTION MANUELLE PACKAGE/COMPOSANTS

- ▶ Activer/Désactiver Conteneurs/Taches

- ▶ Visualiser les données



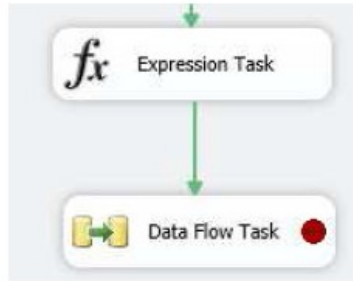
- ▶ Visualiser l'exécution

The screenshot shows the 'Execution Results' window in SQL Server Data Tools. The window has a toolbar with icons for Control Flow, Data Flow, Parameters, Event Handlers, Package Explorer, and Execution Results. The main area displays a tree view of the package execution:

- Package
  - Task Data Flow Task
    - Start, 17:35:29
    - Validation has started
      - [SSIS.Pipeline] Information: Validation phase is beginning.
      - Progress: Validating - 0 percent complete
      - Progress: Validating - 50 percent complete

# EXÉCUTION MANUELLE PACKAGE/COMPOSANTS

- ▶ Positionner des points d'arrêts



Set Breakpoints - DFT -> D\_CENTRE

Select the breakpoints in the task, For Loop, Foreach Loop, or Sequence to enable. Optionally, select the number of breakpoint is ignored before execution is suspended on the breakpoint.

Enabl...	Break Condition	Hit Count Type	Hit Count
<input checked="" type="checkbox"/>	Break when the container receives the OnPreExecute event	Always	0
<input type="checkbox"/>	Break when the container receives the OnPostExecute event	Always	0
<input type="checkbox"/>	Break when the container receives the OnError event	Always	0
<input type="checkbox"/>	Break when the container receives the OnWarning event	Always	0
<input type="checkbox"/>	Break when the container receives the OnInformation event	Always	0
<input type="checkbox"/>	Break when the container receives the OnTaskFailed event	Always	0
<input type="checkbox"/>	Break when the container receives the OnProgress event	Always	0
<input type="checkbox"/>	Break when the container receives the OnQueryCancel event	Always	0
<input type="checkbox"/>	Break when the container receives the OnVariableValueChanged...	Always	0
<input type="checkbox"/>	Break when the container receives the OnCustomEvent event	Always	0

Pour visualiser les variables

Locals		
Name	Value	Type
⊕ User::Date	{06/03/2013 10:39:43}	DateTime
⊕ System::ExecutionInstanceGUID	{{D3A1ECFB-E4A9-4316-8B23-9FB5023E6BD5}}	String
⊕ System::FailedConfigurations	{}	String