

TP3: Modélisation, publication des cubes OLAP et exploitation des données

Le but principal de ce TP est d'apprendre à modéliser et publier des cubes. Ainsi, vous devez d'abord installer les outils nécessaires à la mise en place de l'architecture Mondrian avec Pentaho. Ensuite vous êtes amenés à modéliser le schéma du Cube Mondrian à partir de la base de données du DW du TP2 (comptoirDM) en créant les dimensions, la table de fait... Et enfin, après avoir publié ledit schéma sur le serveur Pentaho, vous pouvez exploiter les données du cube publié sur le serveur en créant des vues métier et ce en utilisant un plugin spécifique.

Technologies utilisées :

Pentaho Schema Workbench & Pentaho Server & le plugin « *Pivot4j* » ou « *Saiko* ».

1ère Partie:

1. La solution Mondrian

Les cubes OLAP permettent à un utilisateur de « naviguer » au travers de ses données. Pour ce, une mesure va être définie, puis analysée suite à un croisement avec des données organisées sous forme de dimensions. Les cubes sont classés en deux grandes familles :

- ✓ les cubes ROLAP (solution Mondrian)
- ✓ les cubes MOLAP (solution Palo)

Dans ce TP, on s'intéresse aux cubes ROLAP. Ces derniers sont des cubes OLAP qui sont basés sur une structure relationnelle. Ils permettent d'accéder directement aux données physiques qui sont contenues dans les tables relationnelles. Les outils présentent les données regroupées selon des vues métier, organisées autour des mesures et des dimensions. **MONDRIAN** présente la solution OpenSource de l'implémentation des cubes ROLAP. Il s'agit d'un moteur OLAP (*Online Analytical Processing*) open source qui se présente sous la forme d'un ensemble de bibliothèques Java. Aussi, permet-il la représentation des données relationnelles dans un cube ROLAP, grâce à un fichier XML décrivant les axes de mesures et les dimensions. En général, il permet la conception, la publication et le requêtage (*en langage MDX*) de cubes multidimensionnels.

S'appuyant sur une modélisation OLAP standard, Mondrian peut se connecter à n'importe quel entrepôt de données permettant ainsi d'accéder aux résultats dans un format multidimensionnel qui est compréhensible par une API de présentation côté client, le plus souvent en mode WEB, avec par exemple JPivot, Pentaho Analyzer, Pentaho Analysis Tool..

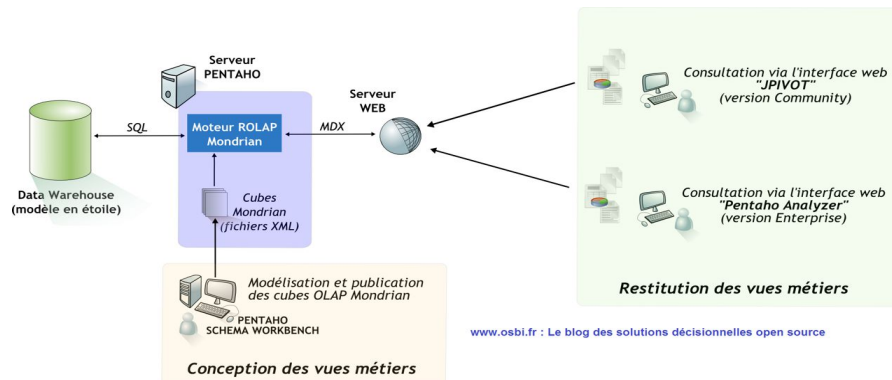
Les caractéristiques de Mondrian sont les suivantes :

- ✓ Stockage des données dans un SGBDR: les données sont entreposées dans des tables de faits et de dimension, selon la modélisation habituelle en étoile et/ou flocons.
- ✓ Utilisation des dimensions partagées (« *shared dimension* », ou « *dimensions conformes* ») pour la mutualisation de celles-ci entre plusieurs cubes.
- ✓ Gestion des hiérarchies multiples, dimensions dégénérées.
- ✓ Gestion de la sécurité d'accès aux cubes via des rôles : sécurisation au niveau du cube, des dimensions, des hiérarchies et des membres à afficher ou non.

- ✓ Internationalisation possible (multilinguisme).

Les éléments caractéristiques de son architecture sont les suivants : (figure ci-dessous)

- ✓ Datawarehouse avec stockage des données au format OLAP, en étoile et en flocons.
- ✓ Serveur web J2EE Pentaho. Ce serveur embarque le moteur ROLAP Mondrian qui permet d'effectuer des requêtes multi-dimensionnelles (langage MDX) sur des données stockées dans un SGBD relationnel.
- ✓ Pour la partie conception, les cubes sont modélisés avec « Pentaho Schema Workbench », dédié à la modélisation et la publication de schémas Mondrian sur un serveur Pentaho.
- ✓ Pour la partie restitution, les vues métiers sont accessibles via un navigateur web (Internet, Explorer, Firefox, Safari, Opera, ...):
 - La mise en place d'un serveur Pentaho Enterprise permettra d'effectuer des analyses avancées avec le composant Pentaho Analyzer.
 - Si on veut rester en open source on utilisera JPivot, qui reste encore malgré son « grand âge » un outil de requête puissant et stable.



Pour la définition des cubes, **Mondrian** s'appuie sur des schémas XML. Ainsi, un **schéma Mondrian** permet donc de définir le modèle logique ainsi que le mapping sur le modèle physique :

- ✓ Le modèle logique décrit les cubes, les dimensions, les hiérarchies, les niveaux et les membres (et plus encore...) sur lesquels vont s'appuyer les requêtes MDX.
- ✓ Le modèle physique correspond à la source de données sur laquelle s'appuie le modèle logique (le modèle en étoile et/ou flocon).

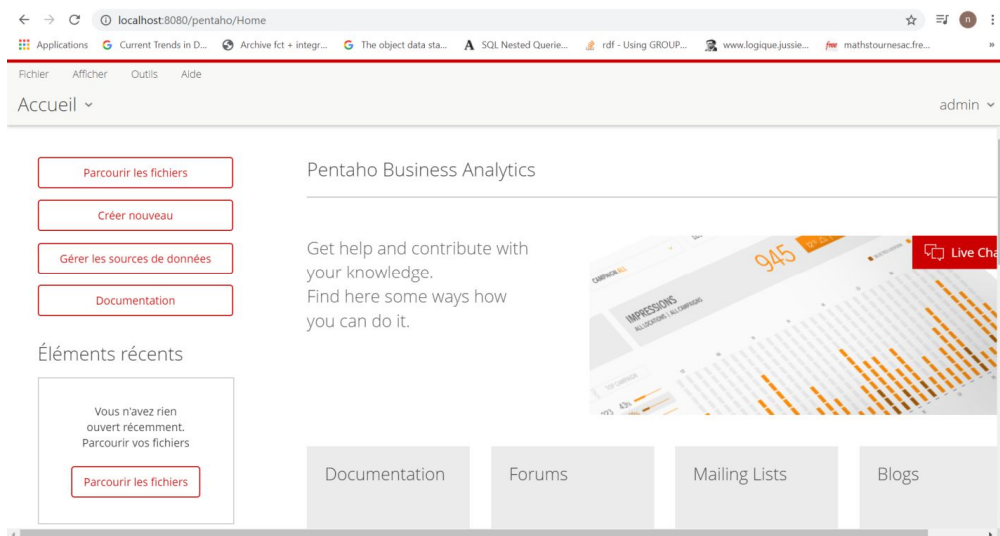
Ces schémas sont créés par les concepteurs et analystes avec l'outil « **Pentaho Schema Workbench** » qui est un client open source (écrit en Java) et qui permet de créer des schémas de cubes sans avoir à connaître la syntaxe XML de Mondrian. L'outil Pentaho Schema Workbench présente une interface avancée qui permet d'effectuer les actions suivantes :

- ✓ Connexion au datawarehouse (via JDBC)
- ✓ Création de schémas et cubes Mondrian
- ✓ Création des mesures, dimensions et hiérarchies d'un cube
- ✓ Jointures pour les tables floconnées
- ✓ Tables d'agrégation à utiliser (le cas échéant)
- ✓ Création de dimensions partagées, de cubes virtuels.
- ✓ Définition des rôles (sécurité d'accès à l'intérieur d'un cube)
- ✓ Publication des cubes sur le serveur Pentaho

2. Installation des modules Pentaho nécessaires

Afin de réaliser les objectifs précédents, vous devez tout d'abord mettre en place les deux modules suivants :

1. « **Pentaho Server** » qui embarque le moteur ROLAP Mondrian pour effectuer une analyse multidimensionnelle.
 - a) Télécharger le module via le lien suivant :
<https://sourceforge.net/projects/pentaho/files/Pentaho%209.0/server/>
 - b) Décompresser le, et placer le dans le répertoire en C:\ qui contient tous les modules de Pentaho utilisés.
 - c) Créer la variable d'environnement `PENTAHO_JAVA_HOME` ayant comme valeur le chemin du JDK.
 - d) Copier le fichier « `mysqlconnector-java-5.1.17-bin.jar` » dans les répertoires suivants :
`C:\Pentaho\pentaho-server\tomcat\lib`
`C:\Pentaho\pentaho-server\data\lib`
 - e) Lancer « `start-pentaho.bat` »
 - f) Ouvrir le navigateur web pour accéder au serveur Pentaho via l'URL suivant :
<http://localhost:8080/pentaho/Login>
 - g) Utiliser les informations ci-dessous pour l'authentification :
 - **login** : admin
 - **password** : password



2. « **Pentaho Schema Workbench** » pour définir le schéma xml du cube de notre data warehouse « *comptoirDM* ».
 - a) Télécharger le module via le lien ci-dessous et décompresser le dans le même répertoire précédent :
<https://sourceforge.net/projects/pentaho/files/Pentaho%209.0/client-tools/psw-9.0.0-423.zip/download>
 - b) Copier le fichier « `mysqlconnector-java-5.1.17-bin.jar` » dans le répertoire :
`C:\Pentaho\schema-workbench\lib`
 - c) Lancer « `workbench.bat` »

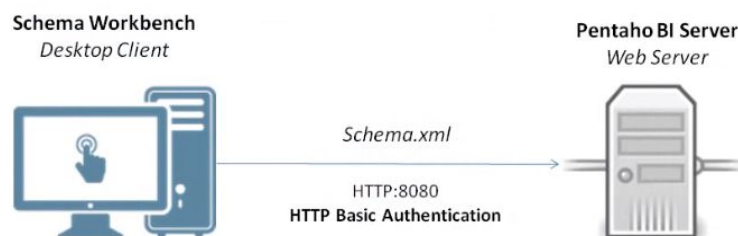
3. Modélisation du schéma du cube avec PSW

Après avoir créé et alimenté notre data warehouse « *comptoirDM* » à partir de la base de données relationnelle « *comptoir* » (TP 2), on doit concevoir le schéma de notre cube dans *Pentaho Schema Workbench* au premier lieu et le publier par la suite sur le serveur pour continuer l'analyse. Dans l'environnement du Pentaho Schema Workbench :

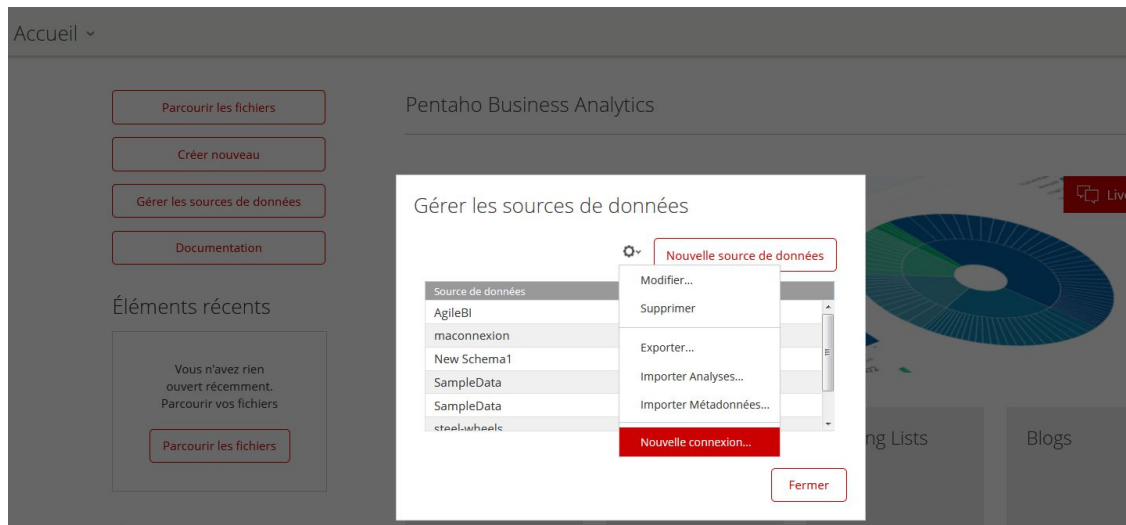
1. Créer une connexion avec le data warehouse « *comptoirDM* » : *Options>connexion*. Tester la réussite de cette configuration.
2. Créer un nouveau schéma : *File > New > Schema*.
3. Créer dedans les trois dimensions « *dimTemps* », « *dimProduits* » et « *dimClients* » en définissant ses types.
4. Pour chaque dimension :
 - a) ajouter une hiérarchie (*Add hierarchy*) sous le nom de « *default* » (une et une seule puisqu'on travaille avec un DW ayant un schéma en étoile).
 - b) Sous l'hiérarchie *default*, ajouter la table de la dimension convenable avec l'option (*Add Table*).
 - c) Ajouter les attributs de la dimension avec (*Add Level*) en spécifiant au minimum le nom de l'attribut et la colonne d'origine.
5. Ajouter un cube à votre schéma nommé *cubeVente*.
6. Ajouter à ce cube :
 - a) La table de fait « *factVents* ».
 - b) Les mesures (*Add Measure*) en précisant son nom, sa colonne d'origine et sa fonction d'agrégation.
 - c) les dimensions créées précédemment (*Add Dimension Usage*). Identifier pour chaque dimension son nom, la clé de référence et la clé étrangère.

4. Publication du schéma Mondrian du cube sur le serveur Pentaho

1. Lancer le serveur Pentaho.
2. Créer une source de données liée au DW « *comptoirDM* » avec : *Gérer les sources de données > Nouvelle source de données*, et configurer votre connexion.
3. Dans l'environnement PSW, publier le schéma du cube créé dans le serveur avec : *File > Publish*. Configurer le avec le login et mot de passe d'authentification utilisés pour se connecter au serveur ; préciser aussi la source de donnée qui sera liée à celle créée dans l'étape précédente.



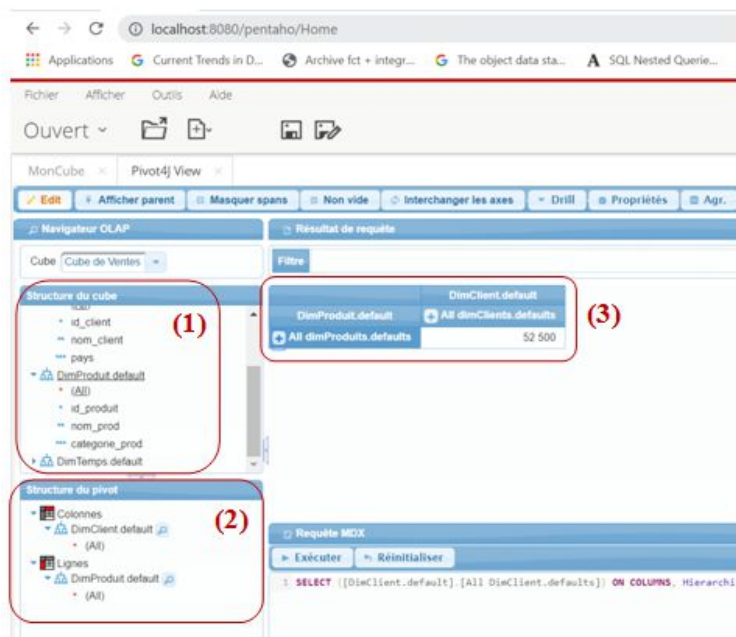
4. Après la réussite de la publication du schéma par PSW, aller au serveur et vider le cache du schéma Mondrian : *Tools > Refresh > Schema Mondrian Cache*.
5. Lier le schéma déployé à la source de données pour commencer une nouvelle analyse.



2ème Partie:

1. Plugin de restitution des vues métiers

On vise dans cette partie, la mise en pratique de l'architecture Mondrian décrite ci-dessus dans la page 2 de ce TP. Pour ce, vous pouvez utiliser l'un des deux fameux plugins : **Pivot4j** ou **Saiku**. Le plugin **Pivot4j** qui est successeur de JPivot représente le cube sous forme d'un tableau croisé (figure ci-dessous). Il permet d'insérer en ligne ou en colonne les différentes dimensions afin d'étudier facilement les mesures du cube. Il permet aussi de visualiser les vues générées selon des diagrammes de divers types et exporter ces analyses en format PDF ou Excel.

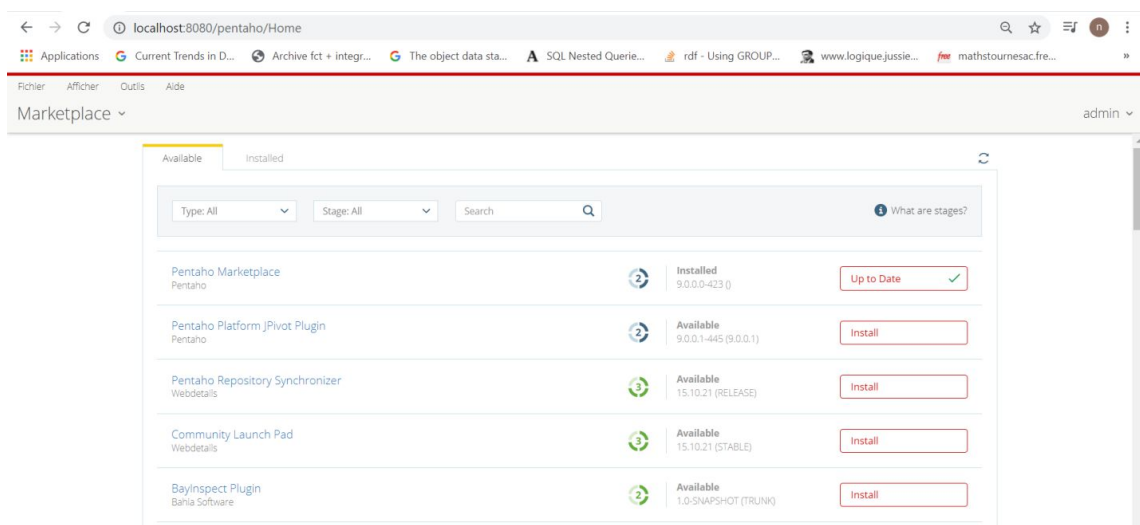


La *zone (1)* affiche la structure de votre cube ; vous pouvez cliquer et glisser ces éléments (selon votre besoin) dans la *zone (2)* pour concevoir la structure de votre vue qui sera affichée dans la *zone (3)*.

Pour la suite de ce TP, il est fortement conseillé d'ajouter plus de données que possible sur la BD initiale « comptoir » et relancer le processus dès d'alimentation du DW via *Pentaho Data Integration* afin de travailler sur une BD « comptoirDM » riche qui générera par la suite des rapports plus expressifs.

2. Installation du plugin et création des vues

Afin d'installer facilement des modules additionnels pour notre serveur Pentaho, on utilise le Marketplace (figure ci-dessous).



1. Lancer le serveur et accéder au Marketplace : *Accueil>marketplace*.
2. Si vous choisissez de travailler avec Pivot4j par exemple, tapez « *Pivot4j Analytics* » dans la barre de recherche et lancer l'installation.
3. Redémarrer votre serveur pour prendre ces modifications en considération.
4. Créer une nouvelle vue Pivo4j : *créer nouveau>Pivot4j view*
5. Pointer sur le schéma et le cube de la première partie de ce TP pour commencer la création de votre rapport d'analyse.
6. Concevoir des rapports avec les vues métiers suivantes :
 - a) Extraire le montant et la quantité des ventes totales.
 - b) Modifier la vue précédente (a) pour extraire le montant et la quantité des ventes par pays.
 - c) Extraire la quantité des ventes par produit et par client.
 - d) En utilisant une agrégation, ajouter à la vue (c) :
 - le total des produits vendus,
 - le total des ventes par client.
 - le nombre de ventes par client.
7. Exporter le rapport de la vue (d) en format pdf.
8. Créer les diagrammes suivants :
 - a) Graphique en barre de la vue (d),
 - b) Graphique en Secteur de la vue (b).