

Création et exploration d'un cube OLAP

Laboratoire N° 1

A. Introduction

Ce laboratoire a pour but de mettre en pratique les concepts de data warehousing vus en cours.

Pour ce faire, vous aurez besoin d'un serveur de base de données (*MySQL*) et d'un serveur *OLAP* (*icCube*). Vous serez dans un premier temps guidé pour l'installation et la configuration des serveurs, pour l'importation des données, la création du cube *OLAP* et son exploration.

Vous intégrerez les réponses aux différentes questions dans le rapport à rendre pour ce module.

B. Installation

Le laboratoire a été préparé en utilisant un PC avec *Windows 7*, mais tous les logiciels nécessaires (ou leurs équivalents) sont disponibles également sous *Linux* et *Mac OS*.

1. Base de données

Nous aurons tout d'abord besoin d'un serveur de base de données *MySQL*. Bien qu'il soit possible d'installer directement *MySQL*, nous privilégierons une plateforme de développement Web intégrant les outils usuels d'administration de base de données. Par exemple *WampServer*¹, dont l'installation ne devrait pas vous poser de problème. Sur certaines configurations il peut y avoir un conflit avec *Skype* qui s'approprie les ports 80 et 443 (désactivable dans les options de *Skype*).

2. OLAP

Pour le serveur OLAP nous allons privilégier une solution récente et gratuite : *icCube*². Son installation ne devrait pas non plus vous poser de problème particulier, notez toutefois qu'il requiert une version récente de *Java* (1.7) et qu'une fois installé, on y accède par une interface Web (http://localhost:8282/icCube/icCube_en.html) avec les Username et Password vides.

C. Données

1. Importation des données

Une fois *Wamp* installé et lancé (icône W vert), vous pourrez accéder à l'outil de gestion de la base de données par une interface Web (<http://localhost/phpmyadmin>). Une fois que vous aurez pris le soin de créer un utilisateur (vous pouvez directement créer une base de données qui lui est attribuée et qui porte son nom), il vous suffira d'importer le fichier *aventure2014.sql.gz* fourni. Vous serez peut être amené à augmenter la taille maximale autorisée pour l'upload de fichiers, cela se passe dans le fichier *php.ini* de *Wamp* (variables *upload_max_filesize* et *post_max_size*) et il faudra ensuite redémarrer le serveur.

2. Provenance des données

Les données proviennent d'un échantillon fourni par *Microsoft*³ comme exemple pour ses produits *SQL*. Nous avons adapté et simplifié ces données afin que l'on puisse les utiliser dans le cadre de ce laboratoire.

Ces données représentent environs 120'000 ventes d'articles de vélo, par internet et en magasin, dans plusieurs pays.

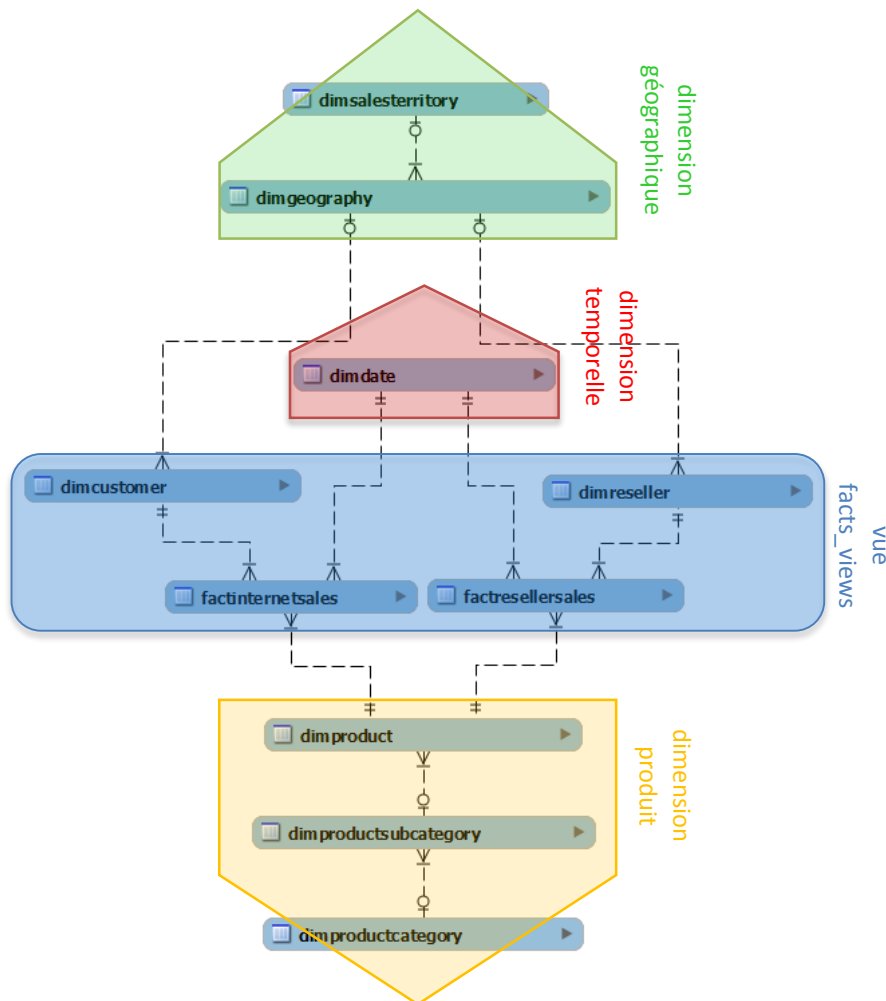
¹ *WampServer* : <http://www.wampserver.com/>

² *icCube* : <http://www.iccube.com/download/>

³ Adventure Works : <http://msftdbprodsamples.codeplex.com/releases/view/55330>

3. Organisation des données

Vous trouverez ci-dessous le diagramme ER représentant ces données :



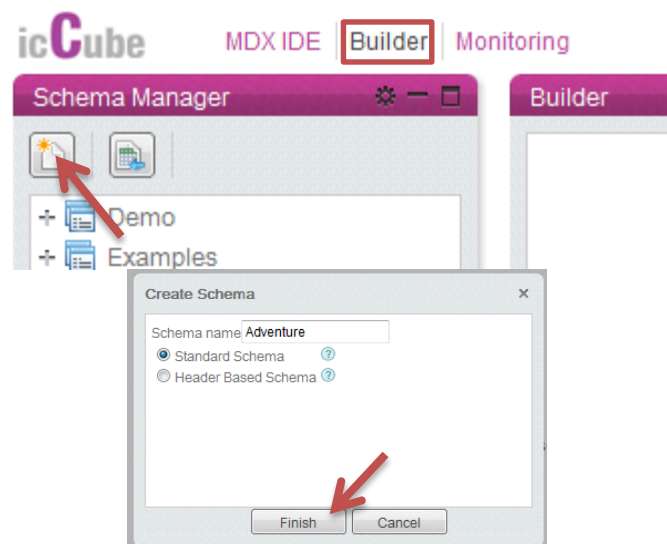
Quelques points à noter :

- Pour simplifier les requêtes et la création du cube, une vue *facts_views* regroupant les ventes en magasin et sur internet vous est fournie ;
- Pour la dimension géographique, nous n'utiliserons pas la table *dimesalesterritory* ;
- Vous devez créer une vue regroupant les champs utiles des tables *dimproduct*, *dimproductsubcategory* et *dimproductcategory* pour la dimension produit. Veuillez indiquer dans le rapport la requête utilisée pour générer cette vue.

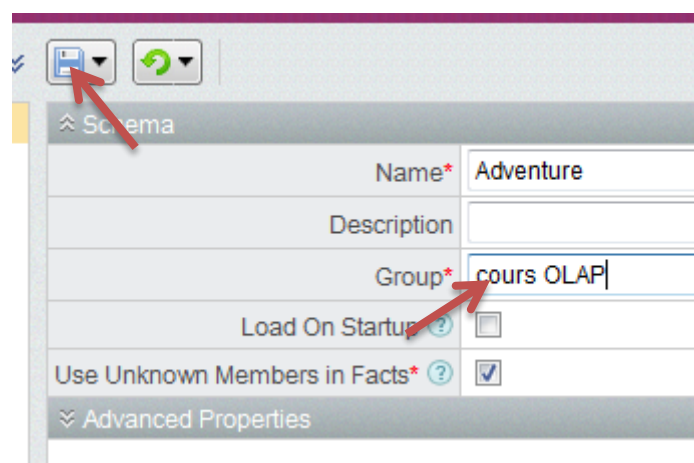
Point 1

D. Configuration d'*icCube*

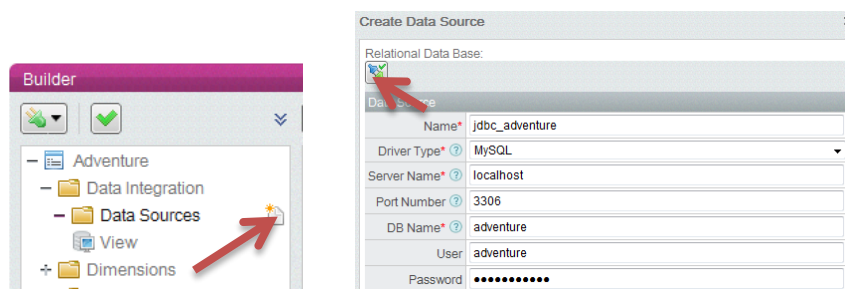
1. Démarrez le serveur *icCube*, puis ouvrez l'interface web. Allez dans la catégorie *Builder* et créez un nouveau *Schema*.



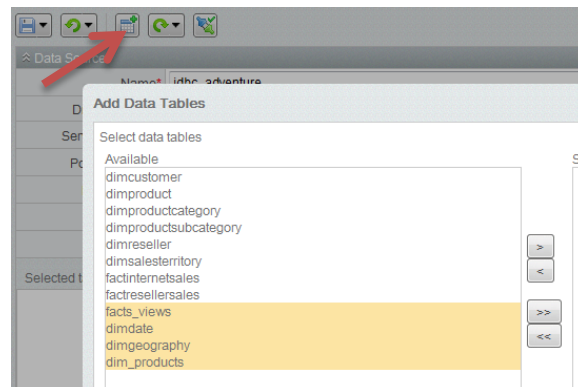
2. Donnez un nom de groupe au *Schema* créé. Et sauvez.



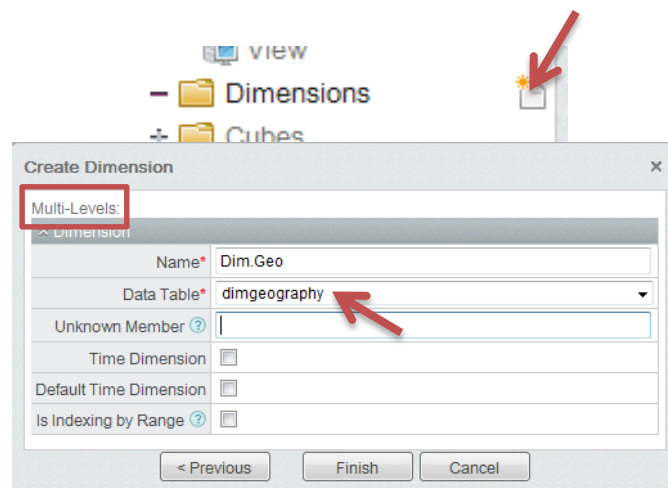
3. On va maintenant configurer notre source de données (*Relational Database*). Vérifiez bien que vos paramètres sont corrects avant de continuer.



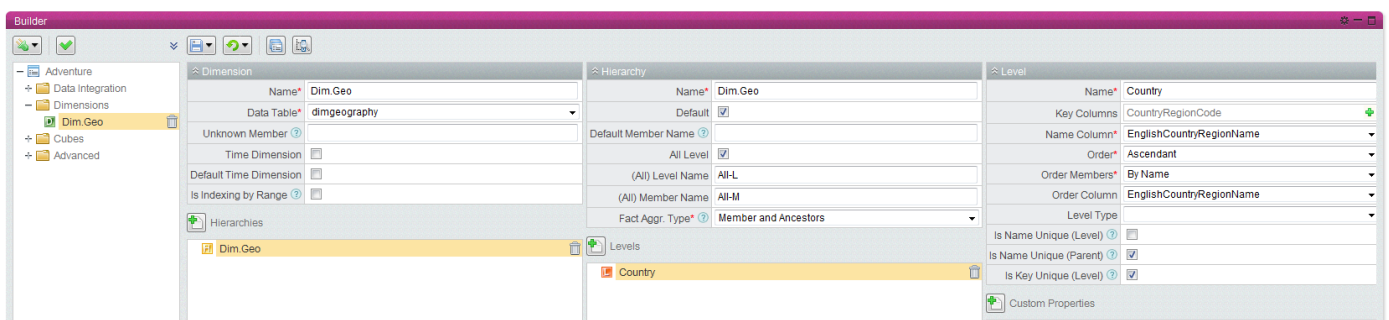
- On sélectionne les tables souhaitées. La version libre d'*icCube* permet uniquement de travailler avec 3 dimensions, nous utiliserons les dimensions géographique, temporelle et par produit. Ainsi que la vue regroupant les ventes effectuées.



- Nous allons maintenant créer les 3 dimensions de notre cube. Nous allons voir l'exemple pour la dimension géographique. On crée une dimension de type *Multi-Levels* :



- On crée le premier niveau pour le pays, on va définir la colonne comportant les clés des pays [*CountryRegionCode*] ainsi que la colonne contenant le nom des pays [*EnglishCountryRegionName*], il est bien sûr aussi possible d'utiliser la même colonne.



- Créez les 2 autres niveaux en utilisant les colonnes adéquates. Pour le dernier niveau (ville) il faut utiliser la clé qui est commune avec la *fact_view*: *GeographyKey*.

MAS-ICT - Data Mining - 2014

8. Vous créez ensuite les 2 autres dimensions. Quelques points importants à noter :

- Pour la dimension *temporelle*, vous créez 2 hiérarchies (Année → Mois → Jour / Année → Semestre → Trimestre → Jour). Vous mettez, dans votre rapport, des captures d'écran des différents paramètres.
- Pour la dimension des *produits*, les informations nécessaires sont réparties dans différentes tables, vous devrez donc utiliser la vue *MySQL* que vous avez créée (point 1). Veuillez aussi mettre des captures d'écran dans votre rapport.

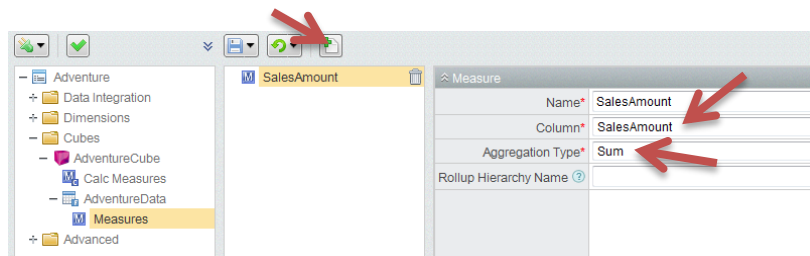
Point 2

Point 3

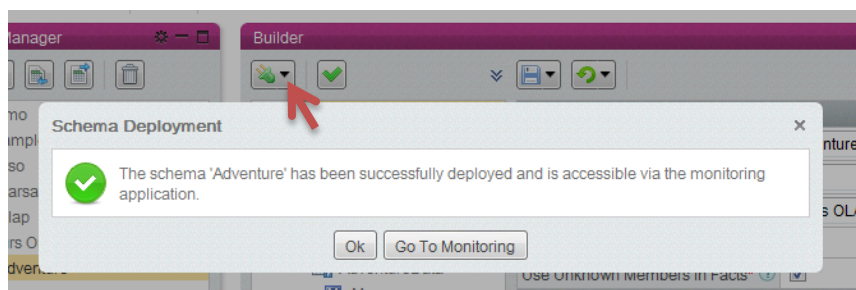
9. Nous allons maintenant pouvoir créer le cube :

10. Et les *Facts* qui permettront de définir les données (vue *facts_view*) et les lier avec les dimensions en utilisant les clés partagées :

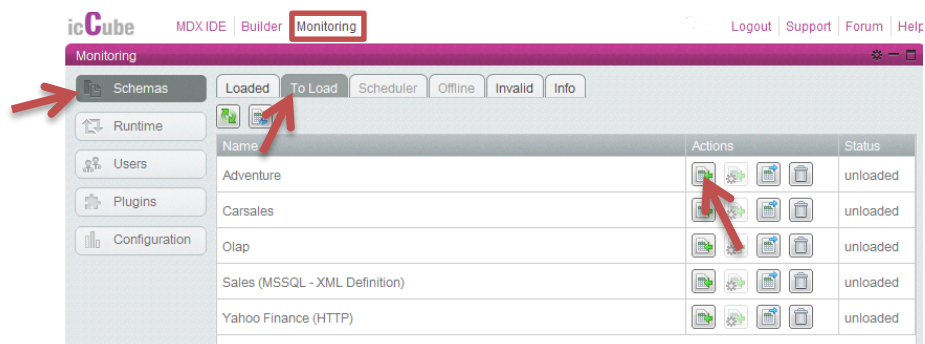
11. Dans la dernière étape, nous allons définir une mesure. Nous allons nous intéresser au montant total des ventes.



12. Il faut maintenant déployer (sur *local*) notre *Schema* afin de pouvoir l'utiliser. Un message devrait confirmer que tout c'est bien déroulé.



13. Il faut ensuite charger notre *Schema* sur le serveur. Cela se passe dans la catégorie *Monitoring*, sous l'onglet *To Load* :



14. Vous devrez corriger les erreurs éventuelles avant de pouvoir effectuer les premières requêtes.

15. Dans la catégorie *MDX IDE* vous pourrez construire et exécuter les requêtes demandées dans le chapitre suivant. N'oubliez pas de sélectionner votre *Schema* dans le menu déroulant en haut à droite.

E. Application des concepts vus en cours

Pour les points suivants, nous vous demandons de bien vouloir indiquer dans votre rapport les requêtes utilisées ainsi que les résultats obtenus sous la forme d'une capture d'écran.

1. Slice

Point 4

Effectuez une requête simple afin de déterminer le montant total des ventes pour l'année 2008.

2. Dice

Point 5

Déterminer le montant des ventes de l'article « Fender Set – Mountain » pour l'année 2008 en Californie. *[Accessoires -> Fenders -> Fender Set – Mountain]*

3. Roll-up

Point 6

En utilisant le principe du Roll-up, déterminez le montant des ventes de l'article « Fender Set – Mountain » pour l'ensemble des États-Unis ainsi que pour l'ensemble des pays, toujours pour l'année 2008.

4. Drill-down

Point 7

Pour l'ensemble des pays, déterminez quel trimestre de 2008 a été le plus fructueux au niveau des ventes de l'article « Fender Set – Mountain ». Quel a été le montant total des ventes ?

5. Pivot (rotate)

Point 8

Faire le même exercice qu'au point précédent, mais en effectuant une rotation entre les produits et les dates. Retrouvez-vous les mêmes valeurs ?