

Master 2 SIID – Projet ODE

Introduction

Le projet « Optimisation des Données de l'Entrepôt (ODE) » consiste à utiliser les techniques mathématiques vues dans le Master 2 MIAGE afin de construire l'entrepôt de données (Data-Warehouse - DWH) de manière optimal, en termes de temps de réponse à l'interrogation des cubes et d'occupation disque.

Extrait du module D111, section « Construction d'un entrepôt de données »

*Vaut-il mieux calculer tout le cube à l'avance ? N'en calculer qu'une partie en (cf. matérialisation partielle) ?
Ou ne rien calculer et tout indexer ?*

Aspects projet

Le projet fera appel à 3 ou 4 participants, moi y compris, et se déroulera entre la fin des examens du S2 2014/2015 et le début des enseignements du S1 2015/2016.

La durée totale du projet dépendra du nombre de participants et des difficultés rencontrées, mais ne dépassera pas 3 mois (Juillet à Septembre 2015, inclus)

Considération techniques

1. Base de données SQL Server configurée pour le DWH. On utilisera « l'Explain Plan », une estimation de temps faite par le moteur de SQL Server pour une requête SQL donnée.
2. Deux web-services :
 - **Calcul de schéma** : Analysera les données d'entrée, estimera (Ou fera estimer par l'Explain Plan) les performances et l'occupation disque, puis générera (En utilisant la méthode exposée dans le module D111, et l'heuristique de Metropolis présenté dans le module D351, cf. "problème du sac à dos") le schéma optimal à appliquer sur la base.
 - **Interrogation du cube** : Seulement pour les tests de performances comparatifs (Benchmarks)

Chaque participant devra télécharger **SQL Server 2015** sur le site officiel de Microsoft. Une notice d'installation et de configuration sous Windows sera fournie aux participants.

Un compte GitHub permet de s'échanger et d'archiver les sources et les documents du projet :

<https://github.com/M2SIID-ODE/ODE/wiki>

Jeu de test

Les données « brutes » d'alimentation du DWH sont celles simulées de la base relationnelle d'une chaîne de magasins de bricolage. Elles sont supposées « propres » et conformes aux référentiels. Nous ne nous attarderons pas sur la partie ETL du DWH.

Un script de création des données de test sera fourni aux participants, afin de le lancer sur leur propre instance de base de données et réaliser les tests unitaires.

Un environnement d'intégration sera également mis en place, pour fournir un environnement de référence pour les tests comparatifs et le maintien de la version « release ».