

# Projet ODE

Optimisation des Données de l'Entrepôt

## ➤ **Projet Master 2 MIAGE – Bordeaux 1**

Thomas CHOURREAU

Brice ELISHA

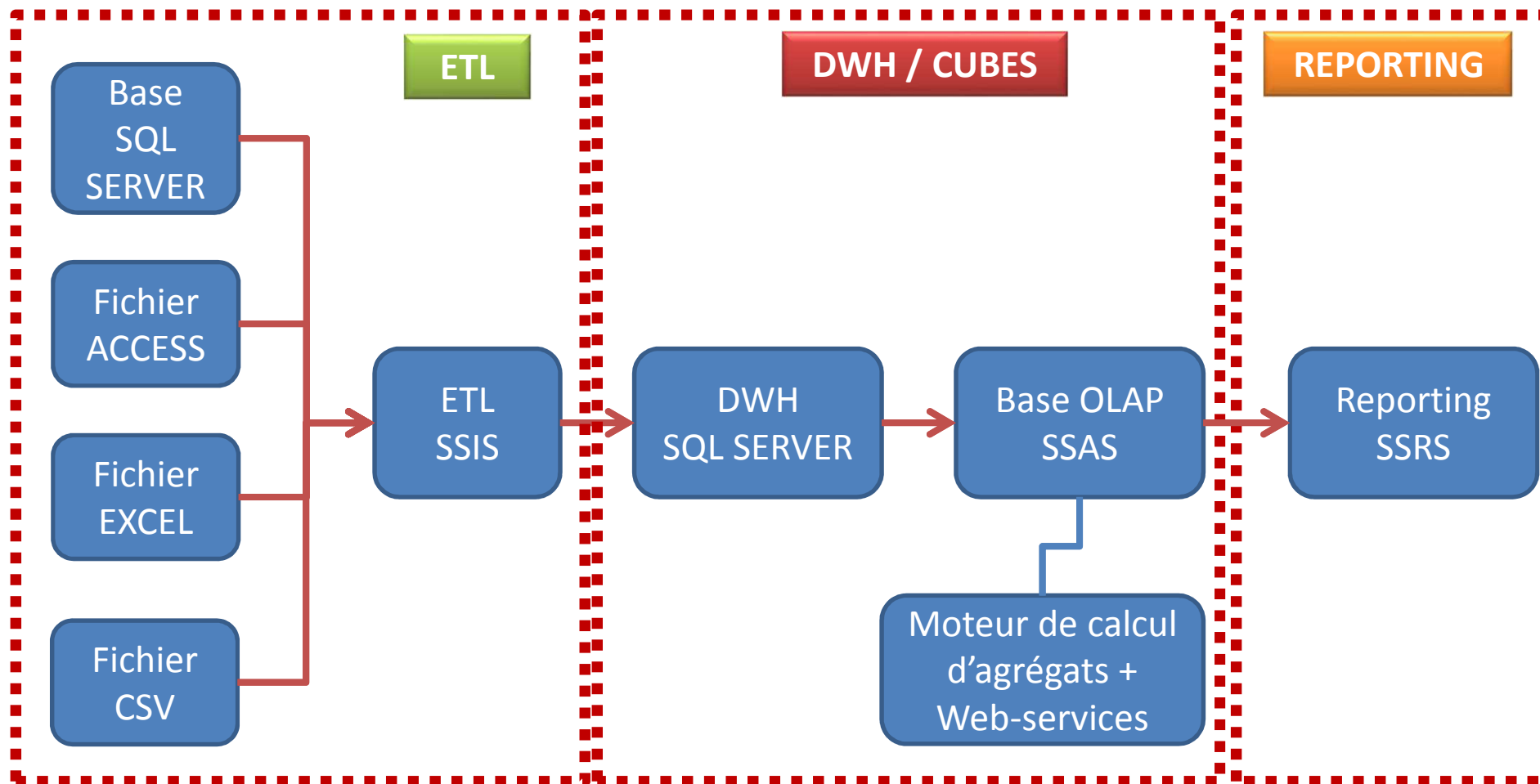
Olivier ESSNER <[olivier.essner \(at\) free.fr](mailto:olivier.essner@free.fr)>

Bernard MOUMY NJANGA

Cédric VANDEVORDE

# CŒUR DU PROJET





### ETL

« Cerise sur le gâteau » : L'ETL et ses fichiers-sources de différents formats. Cette partie du projet est optionnelle, et sera donc traitée à la fin.

### DWH / CUBES

C'est le « cœur » de la chaine BI que nous allons réaliser.

On s'appuie sur les livres & ressources Internet pour sa création et sa configuration

Le DWH est modélisé en étoile / flocon / galaxie (à voir)

Le moteur de calcul d'agrégats est un programme Java servant à « optimiser » la structure et le calcul des cubes (Agrégats, cf. cours du D111)

### REPORTING

Reports de sortie, sur un sujet « métier ». Permet de benchmarker entre elles les différentes optimisations faites par le moteur de calcul d'agrégats sur un vrai cas métier : Calcul du CA, de la marge, évolution des ventes...

Si nous réalisons ces 3 parties, nous démontrerons la mise en œuvre d'une chaîne complète de BI en stack Microsoft.

## Détails de la partie ETL

Base  
SQL  
SERVER



C'est une base relationnelle, suivant un schéma normalisé (3FN)

### **Pour créer sa structure**

Utiliser le script de la structure de la base opérationnelle

### **Pour la peupler**

Générer des données de tests via un script SQL / T-SQL

ETL  
SSIS



C'est l'ETL de la suite BI de Microsoft. Se référer aux documents & livres pour créer le traitement.

Fichier ACCESS



Fichier EXCEL



Fichier CSV

Ce sont des fichiers d'exemples, sur un périmètre restreint :

- **Access** : Fichier de suivi client par exemple
- **Excel** : Fichier du SAV par exemple

La prise en charge par l'ETL des formats Excel, CSV et ACCESS est vraiment optionnelle dans ce projet.

### DWH SQL SERVER



C'est une base multidimensionnelle, suivant un schéma en étoile.

#### **Pour créer sa structure**

Se baser sur la structure de la base opérationnelle (SGBD-R)

#### **Pour la peupler**

Générer des données de tests via un script SQL / T-SQL en attendant la mise en œuvre de l'ETL qui l'alimentera

### Base OLAP SSAS



C'est la base OLAP de la suite BI de Microsoft. Elle contient les cubes, un moteur d'interrogation et un moteur de calcul d'agrégats, simplifié et datant de 2005. Se référer aux documents pour sa configuration.

### Moteur de calcul d'agrégats + Web-services



#### **Point technique**

Il faut « sortir » de SSAS le plan d'exécution des requêtes sur les cubes OLAP, puis générer le meilleur arrangement possible, et enfin retourner dans SSAS pour le faire appliquer.

## Détails de la partie REPORTING

Reporting  
SSRS



C'est le créateur / générateur de reports fixes de la suite BI de Microsoft.  
Se référer aux documents & livres pour les créer.

## Les objectifs du projet

**Rappel :** La formation fait 60 ECTS

- Un module est coefficient 3
- **Le projet est coefficient 9**

Si vous êtes exempté de stage, ce projet représente **22% de votre note finale**

A minima

**Note visée : 14**

### Périmètre initial

Nous réalisons le **DWH avec SQL Server**, les cubes avec **SSAS** et un **moteur de calcul d'agrégat « simple »**

Pas d'ETL avec SSIS ni de report avec SSRS

En plus

**Note visée : 16**

### Aller au-delà...

Nous réalisons une chaine partielle de BI : **DWH, base OLAP et report SSRS simple**  
Notre **moteur de calcul d'agrégat est globalement équivalent** de celui de SSAS

Idéalement

**Note visée : 18**

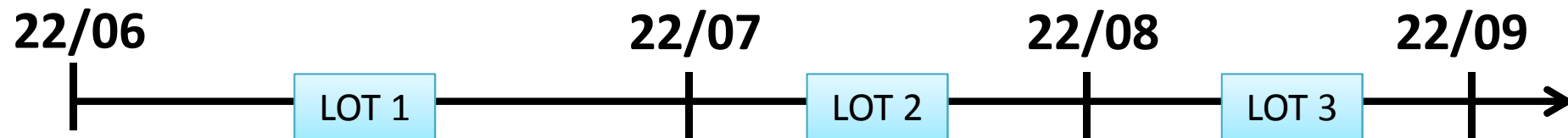
### Effet « Waouh » garanti !

Nous réalisons la **chaine complète de BI** : ETL, DWH, OLAP et plusieurs reports SSRS  
Notre **moteur de calcul d'agrégat est plus performant** que celui de SSAS



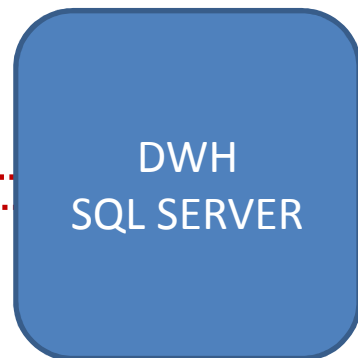


## Macro-planning prévisionnel

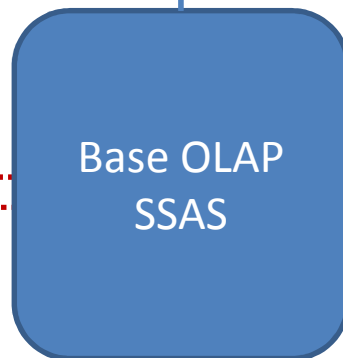


### Equipe 1

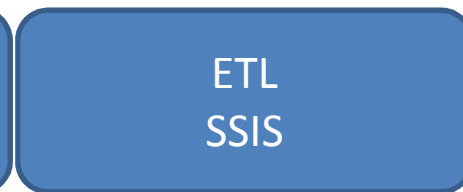
Commun 1 & 2



Commun 1 & 2

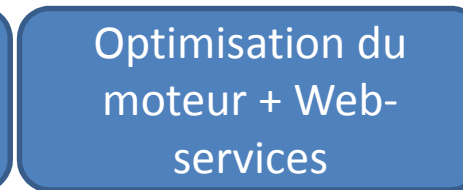
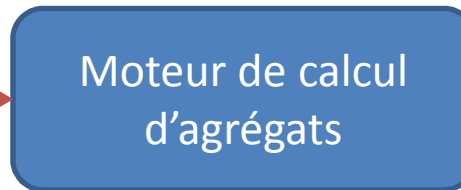


Membres de  
l'équipe 1



### Equipe 2

Membres de  
l'équipe 2



1 mois par lot // Chaque lot inclut sa documentation, ses développements et ses tests

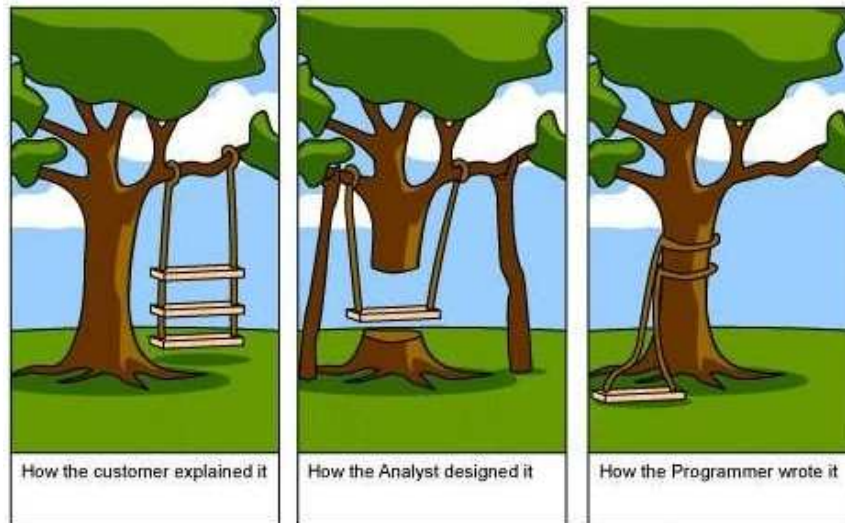


LE MOT DE LA FIN...



## Communication d'équipe

Nous allons privilégier SKYPE pour les visio-conférences hebdomadaires  
Entre-temps, vous pouvez m'envoyer un mail à [olivier.essner \(at\) free.fr](mailto:olivier.essner@free.fr)  
Penser à mettre tout le monde en copie des échanges de mails !



### Keep it simple !

En 3 mois de travail, les enseignants du Master ne nous demandent pas un projet « techniquement parfait » mais « un projet qui marche »...

### En cas de blocage ou de bug récalcitrant

N'hésitez pas à demander de l'aide aux autres membres de notre équipe. Nos compétences & expériences pro sont variées, il est donc très probable que l'un d'entre nous ai déjà votre solution !



Début du projet le lundi  
22/06/2015



**D'ici là, bonne chance pour les examens de Juin !**



# ANNEXES



E-Doc “**Analysis Services Performance Guide for SQL Server 2012 and SQL Server 2014**”  
sur le site de Microsoft (Anglais)

Double-cliquez sur l’icone Word  
pour ouvrir le document ci-joint

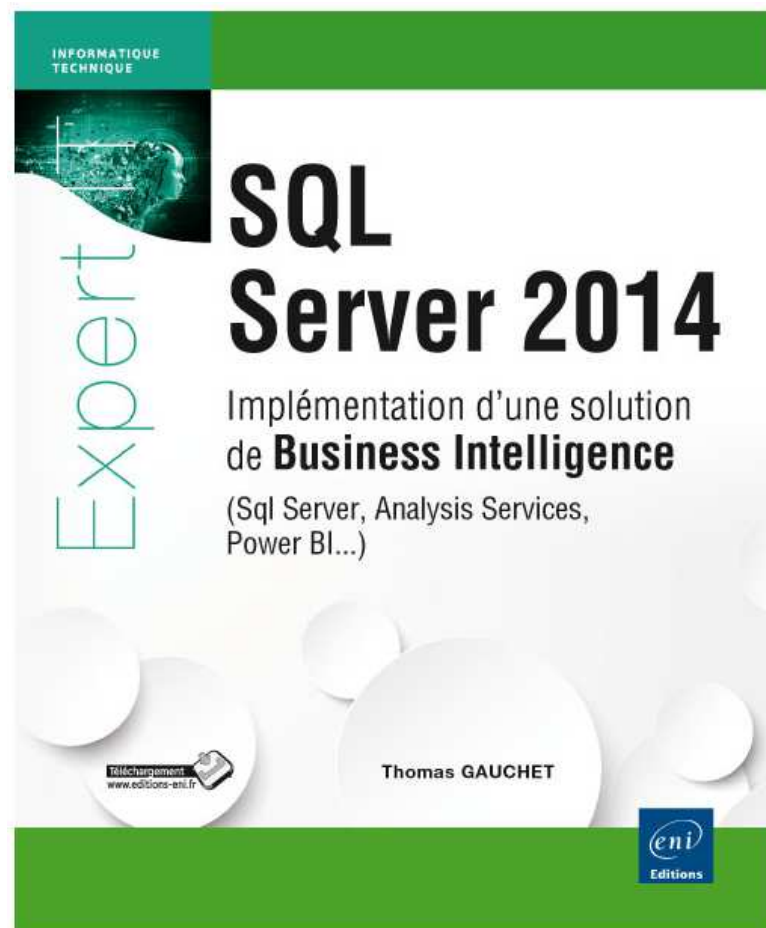


Document  
Microsoft Word

E-Docs du MSDN sur SQL Server (Anglais)

<https://msdn.microsoft.com/en-us/library/ms130214.aspx>

Source : **SQL Server 2014 : Implémentation d'une solution de Business Intelligence**  
Thomas GAUCHET - Edition ENI



Le DM est un ensemble de tables de données organisées dans une structure qui favorise la lecture pour du *reporting* analytique sur un historique plus important que celui conservé en production

...

Le cube est très proche du DM, mais il contient en plus des données pré-agrégées sur les divers niveaux des axes d'analyse. Ces agrégats constitués à l'avance permettent de réduire considérablement les temps de réponse aux demandes des utilisateurs. Il existe plusieurs types d'agrégat, le plus courant étant la somme. Toute donnée qui peut être sommée sur n'importe quel axe d'analyse tirera un grand avantage du cube, à partir du moment où vous en avez plusieurs millions de lignes.

...

Un cube est un schéma en étoile dans lequel un certain nombre d'agrégats ont été précalculés. Le cube offre aussi une couche métier au-dessus des données stockées dans le schéma en étoile.



Le cube peut être construit à partir du DM ou directement à partir d'une autre source.

...

La réunion dans une même base de plusieurs DM prend souvent le nom de DW décisionnel (*Data Warehouse* décisionnel).

Enfin, certains appellent la réunion de plusieurs cubes un hypercube.

...

### **Data Mart (DM)**

...

Un DM est un ensemble de données isolé des systèmes opérationnels, dédié à l'aide à la prise de décision, et son périmètre fonctionnel est généralement focalisé sur un point précis de l'activité de l'entreprise. Les données du DM sont entre autres exprimées sur un axe temporel, avec une profondeur définie.

...

Comme le DM est créé pour être lu par des outils de décision, les données y sont structurées d'une manière adaptée à la lecture. Les créateurs du système OLTP normalisent les tables. Ceux du système décisionnel effectuent l'opération inverse : la dénormalisation.

Le DM peut consolider plusieurs sources de données OLTP. Pour ce faire, les données sont préalablement nettoyées et rapprochées.

Source : **SQL Server 2014 : Implémentation d'une solution de Business Intelligence** // Thomas GAUCHET - Edition ENI

## Data Warehouse

Le DW opérationnel est normalisé logiquement comme les applications source dont il conserve l'historique des données. Il est précieux comme source de construction des DM.

...

Le DW décisionnel est dénormalisé. C'est un ensemble cohérent de DM.

...

Pour Ralph Kimball, le DW est l'ensemble des DM ; chaque nouveau DM vient enrichir le DW : « Le DW n'est rien d'autre que l'union de tous les DM »

...

## Modélisation OLAP

Le modèle de référence pour les DM est le modèle dit *en étoile*.

...

Dans le cas d'un DM périodiquement mis à jour, la *dénormalisation* n'est pas sans poser de problèmes. Un schéma *semi-dénormalisé* est souvent préféré. Il porte le nom de schéma en flocon.

Source : SQL Server 2014 : Implémentation d'une solution de Business Intelligence // Thomas GAUCHET - Edition ENI

Le cube est particulièrement performant dans certaines conditions. Tout d'abord, il doit être utilisé lorsque le besoin est d'extraire, de lire des résultats agrégés et non des faits détaillés atomiques, comme les lignes de factures.

Ensuite l'apport du cube dépend de la nature des mesures métier à analyser. Pour chaque mesure un opérateur d'agrégat est à définir par le métier : par exemple pour le CA, l'agrégat est obtenu en réalisant une somme ; pour la date de commercialisation, l'agrégat peut être le minimum ; pour un niveau de stock, la somme est adaptée sauf sur la dimension calendrier. Cela n'a pas de sens d'additionner le niveau de stock de mars avec celui d'avril.

## SSRS

SQL Server Reporting Services, le serveur de rapports. Conçu pour travailler sur des bases relationnelles, il permet également de lire la base SSAS et de produire des rapports intégrant les principales demandes du *reporting statique* de Business Intelligence : les tableaux, les graphiques, les jauges, les indicateurs visuels et la cartographie. Le serveur de rapports est accompagné d'un concepteur de rapports orienté utilisateur, Report Builder, mais dont l'emploi reste du développement.

...

## Serveur SSAS multidimensionnel

Les modèles OLAP sont basés sur des dimensions et des tables de faits permettant de construire des agrégats précalculés. Le tout est stocké sur disque aux formats MOLAP, ROLAP ou HOLAP, au choix.

Le Multidimensional OLAP (MOLAP) enregistre les données fines ainsi que les agrégats dans un format multidimensionnel.

Le Relational OLAP (ROLAP) enregistre les données fines ainsi que les agrégats dans la base relationnelle source.

Il existe également un mode hybride (HOLAP) qui conserve les données fines au format relationnel dans la source et stocke les agrégats au format multidimensionnel.

Le langage MDX (*Multidimensional Dynamic eXpression*) est utilisé avec le moteur multidimensionnel pour les expressions et requêtes

...

## Excel

Une source à partir d'une requête ou d'une vue sur le DW suffit à créer un *tableau*, un *tableau croisé dynamique* ou un *graphique croisé dynamique*.

Une source à partir d'une base Analysis Services peut alimenter un *tableau croisé dynamique*, un *graphique croisé dynamique* ou une *formule de cube*.

...

les deux principaux outils : SSDT (*SQL Server Data Tools*) et SSMS (*SQL Server Management Studio*),

- SSDT est l'interface dédiée au développement. C'est une édition de Microsoft Visual Studio consacrée à la BI.
- SSMS est dédié à la gestion des serveurs, ou des instances de serveur.

## Index

Le DM est fortement indexé : il n'est pas rare que la place occupée par les index sur le disque soit équivalente à celle occupée par les données.

...

## Présentation de SSAS multidimensionnel

MDX, le langage de requête, est plus adapté que le SQL pour écrire des requêtes décisionnelles et des formules de calcul. Obtenir le CA du mois précédent nécessite une requête SQL complexe et peu performante. En MDX, une simple expression permet de retourner instantanément la valeur. La puissance de calcul s'en trouve démultipliée

...

Le moteur OLAP construit des agrégats qui permettent une exécution très rapide des requêtes. En SQL, le calcul du CA du mois précédent implique une lecture de toutes les lignes des ventes du mois, alors que dans le cube le moteur n'a qu'à lire les bons agrégats précalculés.

Source : **SQL Server 2014 : Implémentation d'une solution de Business Intelligence** // Thomas GAUCHET - Edition ENI

La base de données SSAS est l'équivalent de la base de données SQL

...

## **Agrégations**

Une agrégation est un ensemble de données qui matérialise les agrégats du groupe de mesures dans un fichier, pour éviter au moteur OLAP de les calculer à la volée lors de la requête. L'agrégation est la clé de la performance du cube. Elle repose sur la conception juste et efficace des dimensions.

Le travail de conception des agrégations commence en développement dans SSDT une fois le cube stabilisé, puis continue tout au long de la vie du cube en production dans SSMS

La dernière étape est de Définir les options d'agrégation. Il s'agit de demander au moteur SSAS de créer la *conception d'agrégation* en fonction des paramètres saisis précédemment. L'objectif est de trouver un juste milieu entre créer toutes les agrégations possibles en réalisant un produit cartésien des membres des dimensions et ne créer aucune agrégation. SSAS vous propose de trouver un compromis entre *gains de performance*, mesuré en %, et *l'espace de stockage* utilisé par les agrégats.

100 % de *gains de performance* correspond au maximum que peut apporter SSAS avec les agrégations. 0 % correspond à aucune agrégation créée.

Concevoir les agrégations est une opération qui peut prendre du temps à SSAS étant donné qu'il doit tenir compte d'un nombre élevé de combinaisons possibles. Un bouton Démarrer (ou Continuer) permet de démarrer la conception. Lors de la conception, le graphique se met à jour, présentant les *gains de performance* en ordonnée pour *l'espace de stockage* en abscisse.



Trois options permettent de conditionner l'arrêt de ce traitement :

- L'espace de stockage estimé atteint : la conception s'arrête automatiquement au-delà d'une limite d'espace de stockage calculé.
- Les gains de performance atteignent : la conception s'arrête automatiquement au-delà d'une limite de gains de performance calculés.

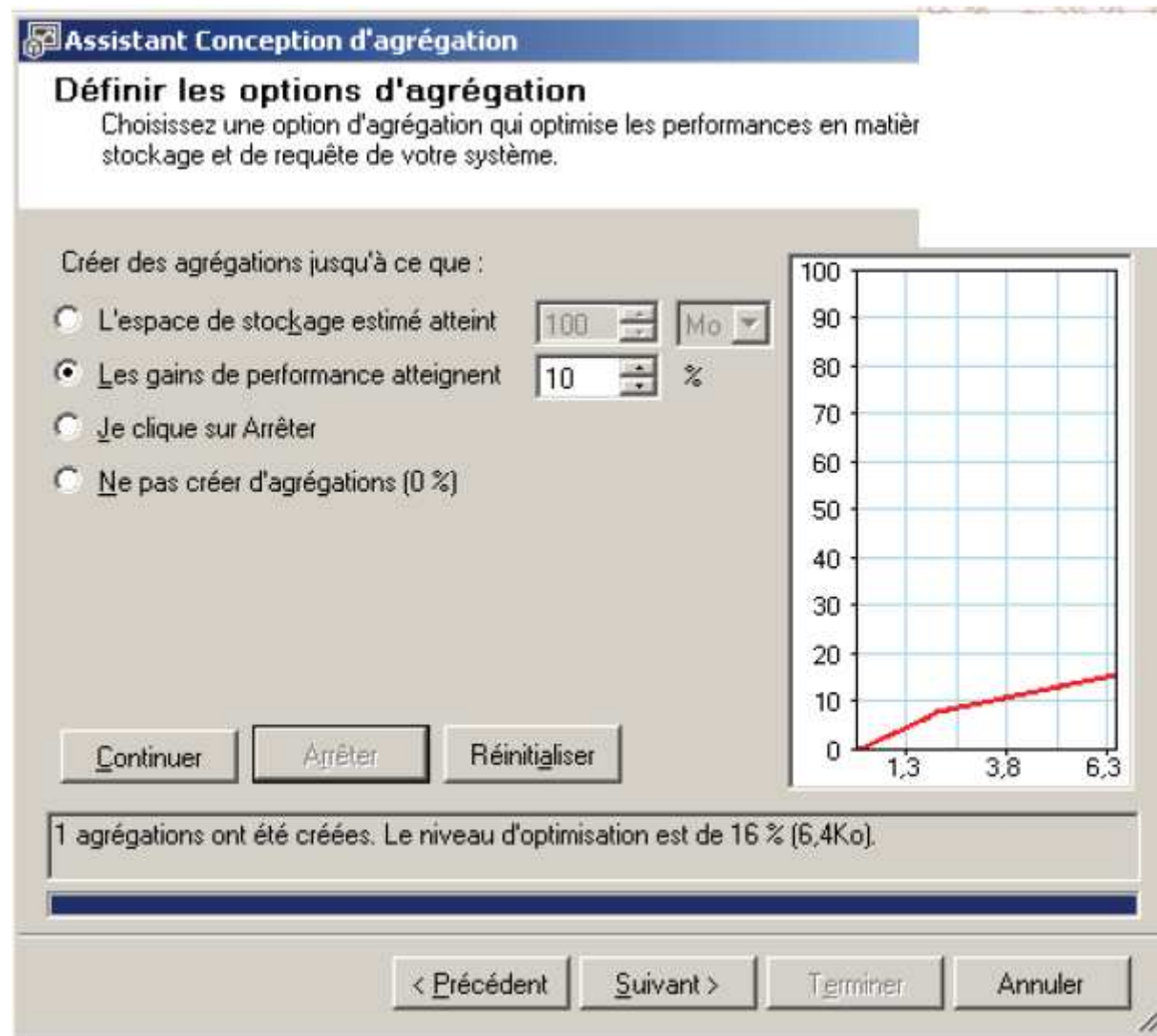
La bonne stratégie à adopter varie selon les développeurs de cube. Partir sur un gain de performance initial de 30 % est sage.

...

### **Optimisation basée sur l'utilisation**

Pour affiner les *conceptions d'agrégation* créées et déployées en production, SSAS fournit un outil qui prend en compte l'activité réelle des utilisateurs du cube. Le principe est simple. Un échantillonnage de l'activité est enregistré dans un journal (*Query Log*) au niveau du serveur. Ce journal est lu par le moteur de *conception d'agrégation*. Il en déduit des poids pour chaque agrégation qui, ajoutés aux paramètres déjà exposés dans la section précédente, permettent de calculer au plus juste une nouvelle conception d'agrégation

Source : **SQL Server 2014 : Implémentation d'une solution de Business Intelligence** // Thomas GAUCHET - Edition ENI



Source : **SQL Server 2014 : Implémentation d'une solution de Business Intelligence** // Thomas GAUCHET - Edition ENI