Datawarehouse: Cubes OLAP

Marlyse Dieungang – Khaoula Ghilani
Table des matières

1 Data Warehouse 3
  1.1 Introduction ............................................. 3
     1.1.1 Définition ........................................ 3
     1.1.2 Architecture ....................................... 3
  1.2 Data Warehouse vs Opérationnel ....................... 5
  1.3 Concepts relatifs ....................................... 5
     1.3.1 Business Intelligence ............................... 5

2 OLAP 7
  2.1 Vocabulaire ............................................. 7
  2.2 Architecture OLAP ....................................... 9
     2.2.1 Configurations de la base de données ............ 10
     2.2.2 Structures des métadonnées ....................... 10
  2.3 Opérateurs OLAP ........................................ 11
     2.3.1 Opérateurs d’agrégation .......................... 11
     2.3.2 Opérateurs de présentation pour la navigation ... 14
  2.4 Langage de requêtes .................................... 15
  2.5 Logiciels ................................................. 15
     2.5.1 Serveurs ........................................... 15
     2.5.2 Clients ............................................ 16

3 Démonstration 17
  3.1 Structure de la base de donnée : Oracle 10g .......... 17
  3.2 OLAPCube Writer ........................................ 18
  3.3 OLAPCube Reader ........................................ 19
# Table des figures

1.1 Architecture du Data Warehouse .............................................. 4  
2.1 Exemple d’utilisation d’un cube ............................................. 8  
2.2 Modèle en étoile ................................................................. 11  
2.3 Modèle en flocon ................................................................. 12  
2.4 Modèle mixte ................................................................. 12  
2.5 Modèle en constellation ...................................................... 13  
3.1 Base de données exemple - modèle en étoile ......................... 17  
3.2 Interface de OLAPCubeWriter .............................................. 18  
3.3 Interface de OLAPCubeReader .............................................. 19
Chapitre 1

Data Warehouse

1.1 Introduction

1.1.1 Définition

Selon Bill Inmon, connu comme étant le père du Data Warehouse, “un datawarehouse est une collection de données thématiques, intégrées, non volatiles et historisées pour la prise de décisions.”
De manière plus concrète, nous pouvons le définir comme une structure pour l’organisation des systèmes d’information. Il s’agit un process d’aide à la prise de décision et la gestion de la connaissance tant pour l’usage quotidien que pour l’élaboration de stratégies à long terme.

1.1.2 Architecture

La figure 1.1 offre une vision générale de l’architecture du Data Warehouse. Parmi les composantes de cette architecture, on distingue :

1. Sources de données ...
2. Back-end tier ...
3. Data Warehouse tier ...
4. OLAP tier ...
5. Front-end tier ...
Fig. 1.1 – Architecture du Data Warehouse

Datawarehouse : Cubes OLAP
1.2 Data Warehouse vs Opérationnel

Le Data Warehouse diffèrent sur de nombreux points avec le domaine dit Opérationnel (ou transactionnel). Ci-après un aperçu de ces différences.

<table>
<thead>
<tr>
<th>Data Warehouse</th>
<th>Opérationnel/transactionnel</th>
</tr>
</thead>
<tbody>
<tr>
<td>orienté sujet</td>
<td>orienté application</td>
</tr>
<tr>
<td>Résumé, affiné, détaillé</td>
<td>Détaillé</td>
</tr>
<tr>
<td>évolue avec le temps</td>
<td>Statique</td>
</tr>
<tr>
<td>Besoins décisionnels</td>
<td>Besoins quotidiens</td>
</tr>
<tr>
<td>Lecture seule</td>
<td>Mise à jour possible</td>
</tr>
<tr>
<td>Traitement par lots</td>
<td>Transactions temps réel</td>
</tr>
<tr>
<td>Analyse</td>
<td>Transaction</td>
</tr>
<tr>
<td>Model dimensionnel</td>
<td>Diagramme Entité-Relation</td>
</tr>
<tr>
<td>Large amount of data</td>
<td>Petite quantité de données</td>
</tr>
<tr>
<td>Diponibilité relative</td>
<td>Grande disponibilité</td>
</tr>
<tr>
<td>Structure flexible</td>
<td>Structure statique</td>
</tr>
</tbody>
</table>

1.3 Concepts relatifs

L’objet de ce travail concerne un point particulier du Data Warehouse. Mais avant de le détailler, il serait intéressant de s’arrêter sur les concepts qui sont relatifs et de les définir.

Ainsi, le Data Mart est un sous-ensemble du Data Warehouse qui concerne une branche particulière de l’entreprise (dépendant ou indépendant). Ex : ... On y fait appel au Business Intelligence (BI). Il s’agit d’un ensemble de procédés pour la collecte, la comparaison et l’analyse d’informations décisionnelles.

1.3.1 Business Intelligence

Parmi les procédés existants du Business Intelligence, il y a OLAP qui fait l’objet de ce travail. OLAP ou On-Line Analytical Processing est une “catégorie de logiciels axés sur l’exploration et l’analyse rapide des données selon une approche multidimensionnelle à plusieurs niveaux d’agrégation.”
Les objectifs attendus par l’utilisation d’OLAP sont les suivants :
1. Assistance pour une analyse optimale des données sans se focaliser sur les moyens utilisés (abstraction),
2. Rapidité et facilité,
3. Visualisation multidimensionnelle des données (lacune des R-DBs).

Datawarehouse : Cubes OLAP
Chapitre 2

OLAP

2.1 Vocabulaire


La figure 2.1 est un exemple de cube d’analyse où “la marge de profit en pourcent pour les vélos au mois de février” est un fait qui exprime la valeur de la mesure “marge de profit” pour le membre “février” du niveau “mois” de la dimension “temps” et le membre “vélos” de la dimension “produits”.
FIG. 2.1 – Exemple d’utilisation d’un cube

Datawarehouse : Cubes OLAP
2.2 Architecture OLAP

Le vocabulaire étant établi, nous pouvons dès lors aborder le vif du sujet. Nous commencerons par détailler l’architecture d’un OLAP. Celle-ci est constituée de trois parties qui s’emboîtent :

La base de données

– constitue un support de données agrégées ou résumées (notion de niveaux hiérarchiques).
– Les données qu’elle contient peuvent provenir d’un entrepôt de données.
– Elle possède une structure multidimensionnelle c’est-à-dire basée sur un SGDB multidimensionnel ou relationnel.

La serveur OLAP permet

– la gestion de la structure multidimensionnelle dans le SGDB.
– la gestion de l’accès aux données de la part des utilisateurs.

Le module client permet

– à l’utilisateur de manipuler et d’explorer les données.
– l’affichage des données sous formes de graphiques ou de tableaux.
En ce qui concerne la base de données, il existe plusieurs configurations possibles.
2.2.1 Configurations de la base de données

<table>
<thead>
<tr>
<th></th>
<th>ROLAP</th>
<th>MOLAP</th>
<th>HOLAP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stockage des données de base</td>
<td>BD relationnelle</td>
<td>BD multidimensionnelle (hypercube)</td>
<td>BD relationnelle</td>
</tr>
<tr>
<td>Stockage des agrégations</td>
<td>BD relationnelle</td>
<td>BD multidimensionnelle (hypercube)</td>
<td>BD multidimensionnelle (hypercube)</td>
</tr>
<tr>
<td>Structure de la BD</td>
<td>Modèle particulier (étoile, flocon, etc)</td>
<td>Structure propriétaire au logiciel utilisé</td>
<td>Croisement des architectures ROLAP et MOLAP</td>
</tr>
<tr>
<td>Fonctionnement</td>
<td>Le serveur extrait les données par des requêtes SQL et interprète les données selon une vue multidimensionnelle avant de les présenter au module client.</td>
<td>Le serveur MOLAP extrait les données de l’hypercube et les présente directement au module client.</td>
<td>Accède aux deux BD et les présente au module client selon leur méthode respective.</td>
</tr>
<tr>
<td>Performance des requêtes</td>
<td>Le moins performant</td>
<td>Le plus performant</td>
<td>Performance moyenne</td>
</tr>
</tbody>
</table>

2.2.2 Structures des métadonnées

L’usage des configurations ROLAP et HOLAP nécessite de simuler une structure multidimensionnelle dans un SGDB relationnel.

Pour cela, il existe des modèles prédéfinis :
- en étoile (Star Schema) caractérisé par une simplicité d’utilisation.
- en flocon (Snowflake Schema) représente la vision des données du point de vue de l’utilisateur : respect de l’héritage.
- mixte (Mixed Schema). Fusion des modèles en étoile et en flocon et Datawarehouse : Cubes OLAP
consiste en une normalisation des grandes tables lorsqu’il y a trop de redondance.
– en constellation (Fact Constellation Schema) consiste à relier plusieurs modèles en étoile ayant une dimension commune. Les figures 2.2, 2.3, 2.4 et 2.5 illustrent ces modèles.

2.3 Opérateurs OLAP

Les opérateurs OLAP pour la manipulation des cube de données sont de deux types : les opérateurs d’agrégation et les opérateurs de présentation pour la navigation.

2.3.1 Opérateurs d’agrégation

Etant donné le principe de granularité, la navigation dans le cube de données permet à l’utilisateur de passer de données détaillées à des données moins détaillées. Ce genre de manipulation nécessite de résumer les données.

Datawarehouse : Cubes OLAP
Fig. 2.3 – Modèle en flocon

Fig. 2.4 – Modèle mixte

Datawarehouse : Cubes OLAP
Pour cela, il faut mettre en place des opérateurs adéquats. Ex : Passer du nombre de vente par ville au nombre de vente par région consiste à agréger ces données par une addition. Il s’agit, en fait, d’opérations élémentaires telles la somme, la moyenne, le comptage, le min, le max ou toute autre fonction statistique. Afin de déterminer l’opération à utiliser, il faut restrecter des conditions d’agrégation des mesures et fournir une classification de celles-ci.

**Conditions d’agrégation**

- Disjonction des instances : le regroupement d’instances d’un niveau relatif à leurs parents produit des ensembles disjoints. Par exemple, une ville ne peut appartenir à plusieurs régions ou provinces.
- Architecture complète : chaque instance est rattachée à un niveau qui est relatif à un parent de niveau supérieur. Ex : les magasins d’une multinationale se trouvent dans une ville qui se trouve à son tour dans un pays.
- Usage convenable des fonctions d’agrégation : les différents types de mesures déterminent quelle fonction d’agrégation utiliser.

**Classification des mesures**

Il existe deux classifications différentes des mesures.

Datawarehouse : Cubes OLAP
La première tient compte du caractère additif ou non de la mesure. Ainsi, on distingue :

- Les mesures additives : elles peuvent être additionnées le long de toutes les dimensions (temporelle, spatiale, catégories, etc.)
- Les mesures semi-additives : elles peuvent être sommées le long de certaines dimensions seulement. Ex : il n’est pas pertinent de sommer les volumes de stocks le long de la dimension temporelle.
- Les mesures non additives ou value-per-unit measures ne peuvent être sommées le long d’aucune dimension. Ex : Prix d’un produit.

La seconde classification permet d’éviter des calculs redondants en utilisant des résultats déjà obtenus afin d’optimiser les calculs. On distingue :


### 2.3.2 Opérateurs de présentation pour la navigation

Les outils OLAP utilisent des opérateurs particuliers pour la navigation dans les hypercubes.

**Roll-up** Passage de mesures détaillées à résumées en remontant dans la hiérarchie de la dimension.

**Drill-down** Descendre dans la hiérarchie de la dimension.

**Rotate** Rotation des axes du cube pour fournir une vue alternative des données.

**Slicing** Extraction d’une tranche d’informations : Sélection d’une dimension pour passer à un sous-cube.

**Dice** Extraction d’un bloc de données : Sélection de deux ou plusieurs dimensions.

**Drill-across** Exécution de requêtes impliquant plus d’un cube ayant une dimension commune.

Datawarehouse : Cubes OLAP
Drill-through Passage d’une mesure à l’autre ou d’un membre d’une dimension à un autre.

2.4 Langage de requêtes

Comme SQL pour les bases de données relationnels, il existe des langages de requêtes pour l’utilisation des OLAP. Il s’agit de langage de calcul avec une syntaxe similaire à celle des tableurs. L’inconvénient est qu’il n’y a pas de langage universel. Néanmoins, on peut distinguer deux tendances :

– **OLAP DML** (Data Manipulation Language) intégré à Oracle 10g.

2.5 Logiciels


2.5.1 Serveurs

– SAS Business Analytics
– SAP
– PALO OLAP Server (Open Source)
– OlapCubes de AderSoft
– Oracle
– SQL Server Services analysis
– ALG Software
– Applix
– Microstrategy
– Descisys
– INEA/Cartesis

Datawarehouse : Cubes OLAP
2.5.2 Clients

- Panorama Software
- Proclarity
- AppSource
- Cognos
- Business Objects
- Brio Technology
- Crystal Reports
- Microsoft Excel
- Microsoft Reporting Services.

Datawarehouse : Cubes OLAP
Chapitre 3

Démonstration

3.1 Structure de la base de donnée : Oracle 10g

La base de données qui a été utilisée pour les tests porte sur la vente d’articles (chaussures) d’une entreprise qui possède des magasins dans plusieurs villes différentes à travers le monde. Elle est constituée de 5 tables : Les ventes qui enregistre toutes les ventes effectuées dans différents magasins et concernant les chaussures de différents modèles pour différentes pointures et couleurs.

**Fig. 3.1 – Base de données exemple - modèle en étoile**
3.2 OLAPCube Writer

OLAPCube Writer est un outil développé par AderSoft, il permet de se connecter à une base de données, de récupérer les tables dont on a besoin pour créer les cubes OLAP, de créer les dimensions ainsi que les mesures et champs calculés du cube. On peut dès lors visualiser les données relatives au cube grâce à OLAPCube Reader.

![Interface de OLAPCubeWriter](image)

**FIG. 3.2 – Interface de OLAPCubeWriter**

Datawarehouse : Cubes OLAP
3.3 OLAPCube Reader

OLAPCube Reader est un outil de visualisation (reporting) des cubes OLAP. Il permet entre autre d’appliquer des filtres sur les données, de générer des diagrammes de différentes formes (camembert, histogrammes, tableaux, etc).

![Diagramme OLAPCube Reader](image)

**Fig. 3.3 – Interface de OLAPCubeReader**

Datawarehouse : Cubes OLAP
Bibliographie


Datawarehouse : Cubes OLAP