

Data Management

Lab 01 - Création et exploration d'un cube OLAP

Robin Chappatte

Frédéric Montet

Brian Nydegger

Rendu le 5 novembre 2016

à Lausanne

Professeurs :

Dr. Laura Elena Raileanu

Fabien Dutoit

Table des matières

Introduction	1
1 Installation	2
1.1 MySQL	2
1.2 icCube	2
2 Configuration	3
2.1 Importation des données dans MySQL	3
2.1.1 Création d'une vue	4
2.2 Configuration d'icCube	4
2.2.1 Dimension temporelle	4
2.2.2 Dimension des produits	7
3 Exploration	8
3.1 Slice	8
3.2 Dice	9
3.3 Roll-up	10
3.4 Drill-down	12
4 Conclusion	14

Introduction

Ce laboratoire va nous permettre de mettre en pratique le contenu du cours de data warehousing et OLAP. Pour cela, nous allons utiliser un serveur de base de donnée MySQL et un serveur OLAP icCube.

La première partie du laboratoire consistera à l'installation et la configuration des serveurs cités ci-dessus. Dans un deuxième temps, nous allons explorer les possibilités fournies par un tel système à l'aide d'une série de requêtes. Ces dernières seront composées des opérations typiques qu'il est possible de faire sur un cube OLAP (slice, dice, etc.).

1 | Installation

L'infrastructure nécessaire pour l'utilisation d'un cube OLAP consiste en l'installation de 2 serveurs :

1. MySQL
2. icCube

1.1 MySQL

L'installation du serveur MySQL a été simplifiée en utilisant la plateforme de développement web Wamp, respectivement Mamp pour OSX ou encore Xamp pour Linux. Ainsi, la configuration de ces outils est déjà faite et, de plus, des outils d'administration de base de donnée comme PhpMyAdmin y sont disponibles.

1.2 icCube

L'installation du serveur OLAP icCube se fait de la même manière qu'une application commerciale grand public, il suffit de suivre l'assistant d'installation. Sa seule dépendance est Java 1.8.

Une fois l'installation faite, le serveur OLAP est disponible à l'adresse `http://localhost:8282/icCube/icCube_en.html`.

2 | Configuration

Les étapes de configuration de MySQL et icCube converge vers l'objectif commun d'avoir un cube OLAP sur lequel il est possible d'effectuer des requêtes multi-dimensionnelles.

2.1 Importation des données dans MySQL

Après avoir créé un nouvel utilisateur MySQL, nous avons importé les données du dump `aventure2014.sql.gz`.

Les données importées proviennent d'une base de données fournie par Microsoft. Il s'agit de 120'000 ventes d'articles de vélo, par internet et en magasin dans plusieurs pays. Le diagramme ER¹ est visible sur la figure Figure 2.1.

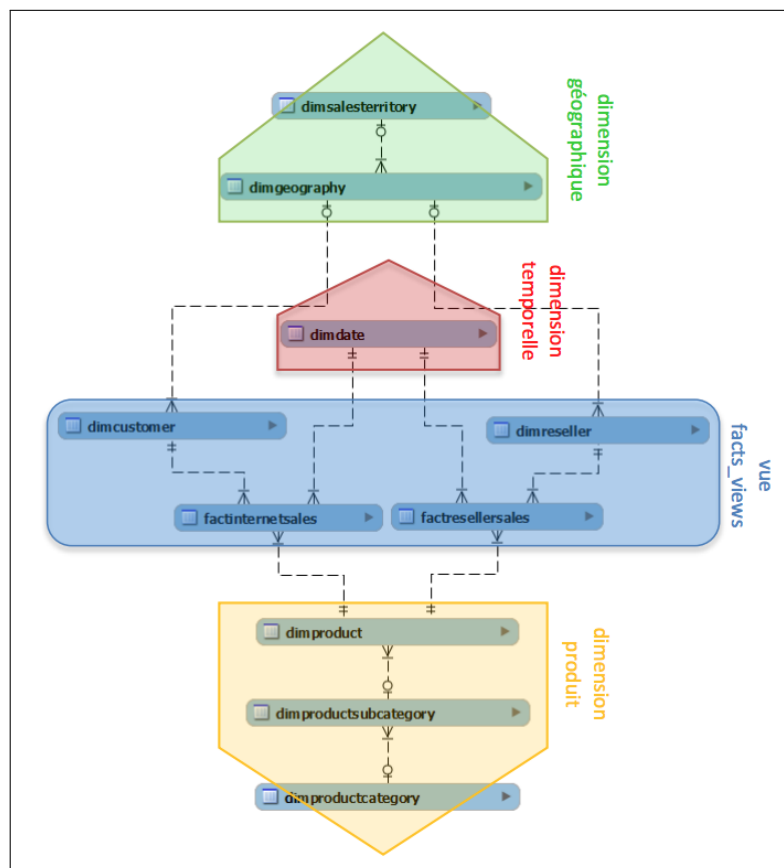


FIGURE 2.1 – Entity Relationship diagramm

1. Entity-Relationship

2.1.1 Création d'une vue

Pour préparer les différentes dimensions que nous souhaitons avoir avec notre cube OLAP, une vue supplémentaire est nécessaire. Il s'agit de la `product_view` qui a été créée avec la requête Figure 2.2

```
1 CREATE VIEW product_views AS
2 SELECT dp.ProductKey, dp.EnglishProductName, dp.ProductSubcategoryKey, dps.
   ↳ EnglishProductSubcategoryName, dps.ProductCategoryKey, dpc.
   ↳ EnglishProductCategoryName
3 FROM dimproduct AS dp
4 LEFT JOIN dimproductsubcategory AS dps
5 ON dps.ProductSubcategoryKey = dp.ProductSubcategoryKey
6 LEFT JOIN dimproductcategory AS dpc
7 ON dpc.ProductCategoryKey = dps.ProductCategoryKey
```

FIGURE 2.2 – Requête SQL pour la création de la vue des produits

2.2 Configuration d'icCube

2.2.1 Dimension temporelle

Hierarchie Année - Mois - Jour

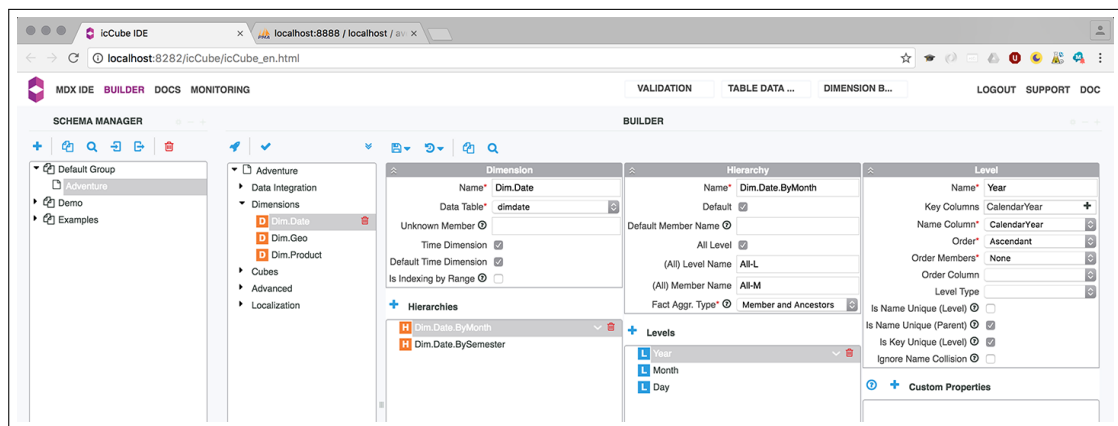


FIGURE 2.3 – Configuration Dim.Date.ByMonth - Year

2.2. Configuration d'icCube

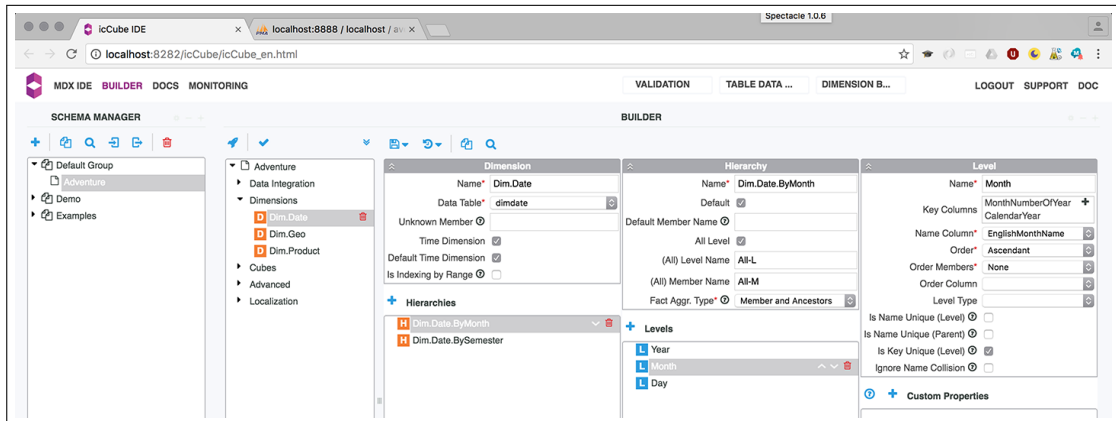


FIGURE 2.4 – Configuration Dim.Date.ByMonth - Month

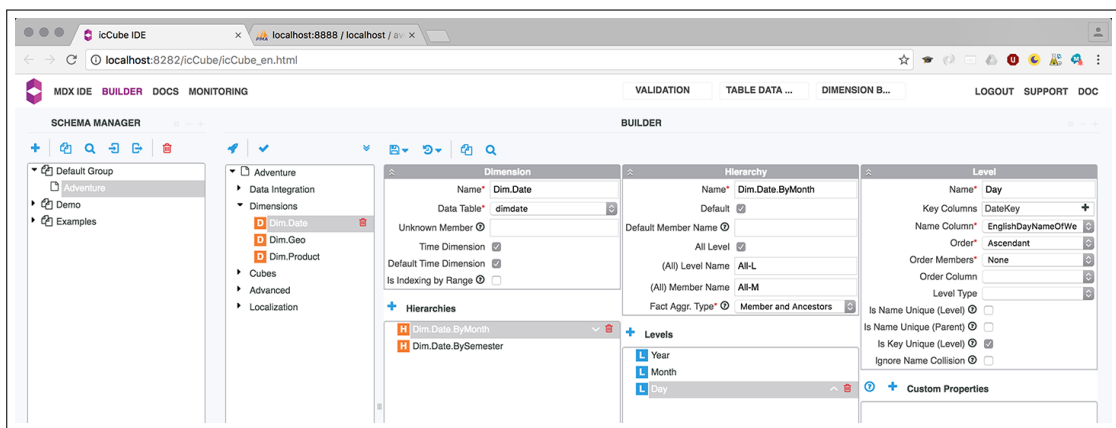


FIGURE 2.5 – Configuration Dim.Date.ByMonth - Day

Hiérarchie Année - Semestre - Trimestre - Jour

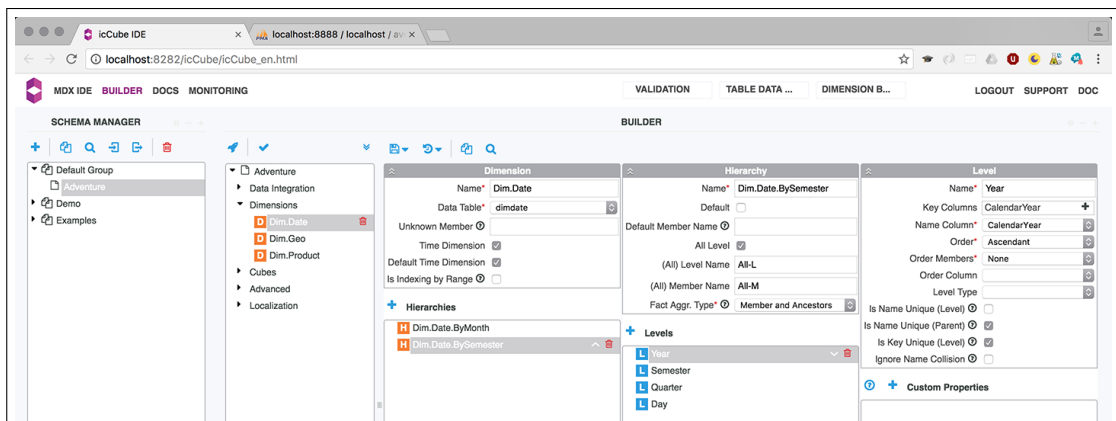


FIGURE 2.6 – Configuration Dim.Date.BySemester - Year

2.2. Configuration d'icCube

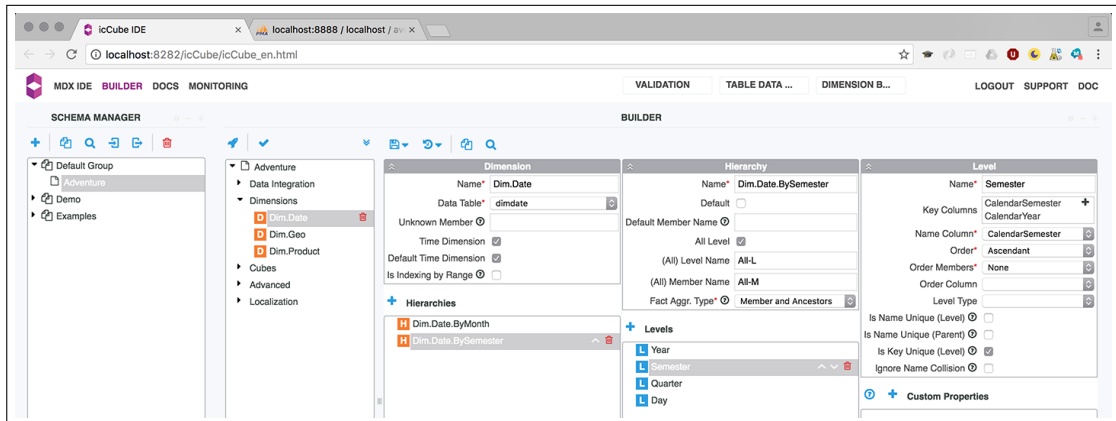


FIGURE 2.7 – Configuration Dim.Date.BySemester - Semester

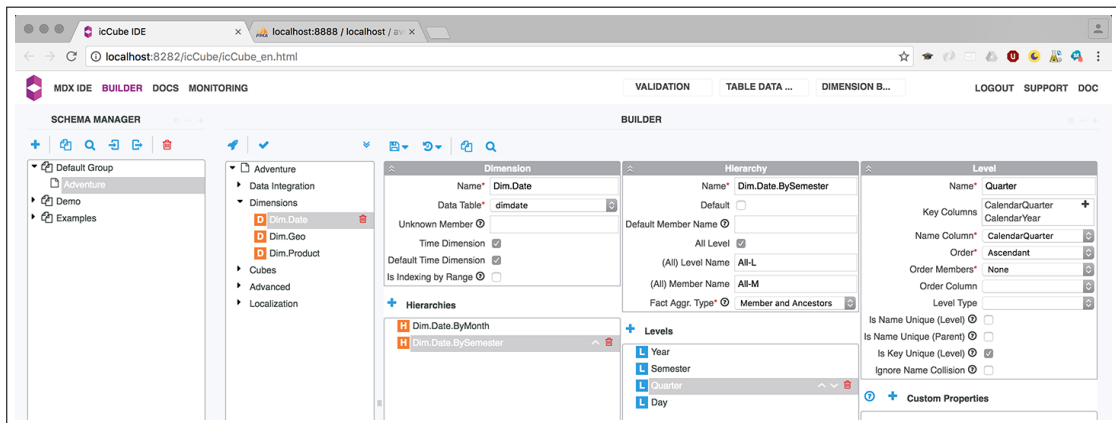


FIGURE 2.8 – Configuration Dim.Date.BySemester - Quarter

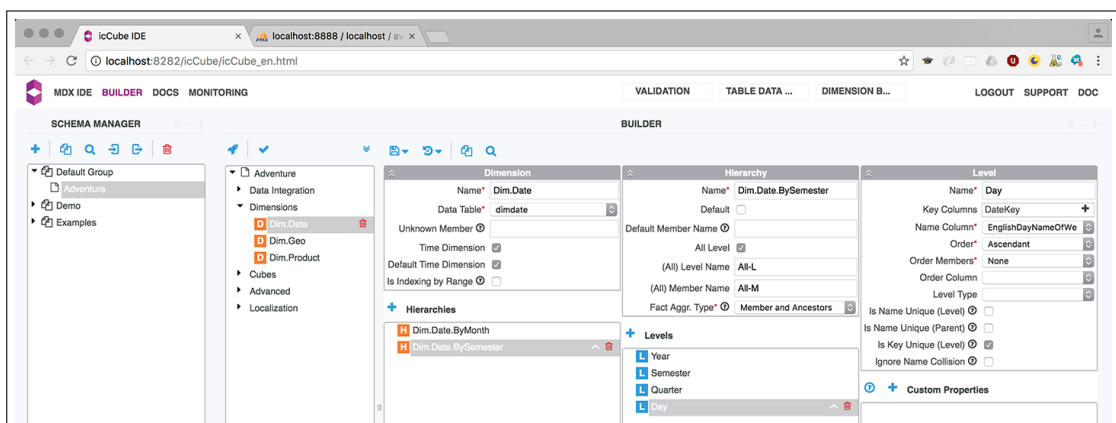


FIGURE 2.9 – Configuration Dim.Date.BySemester - Day

2.2. Configuration d'icCube

2.2.2 Dimension des produits

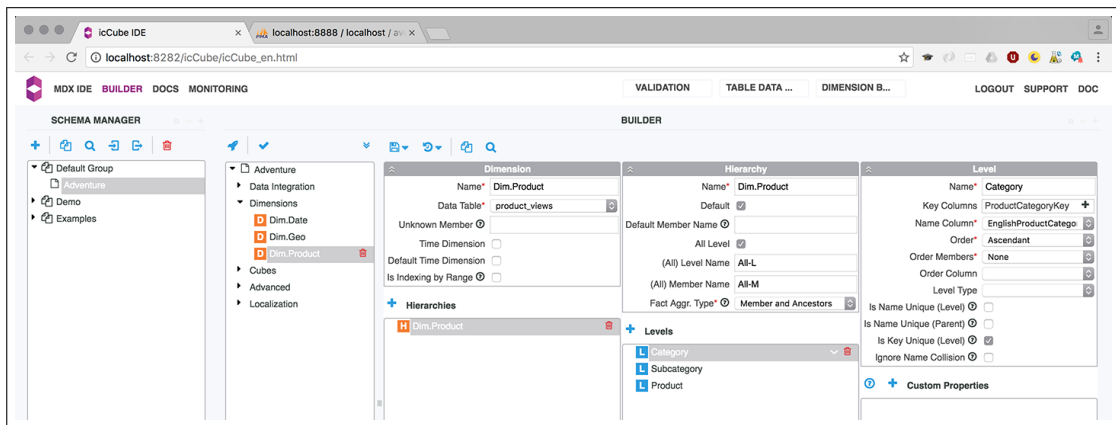


FIGURE 2.10 – Configuration Dim.Product - Category

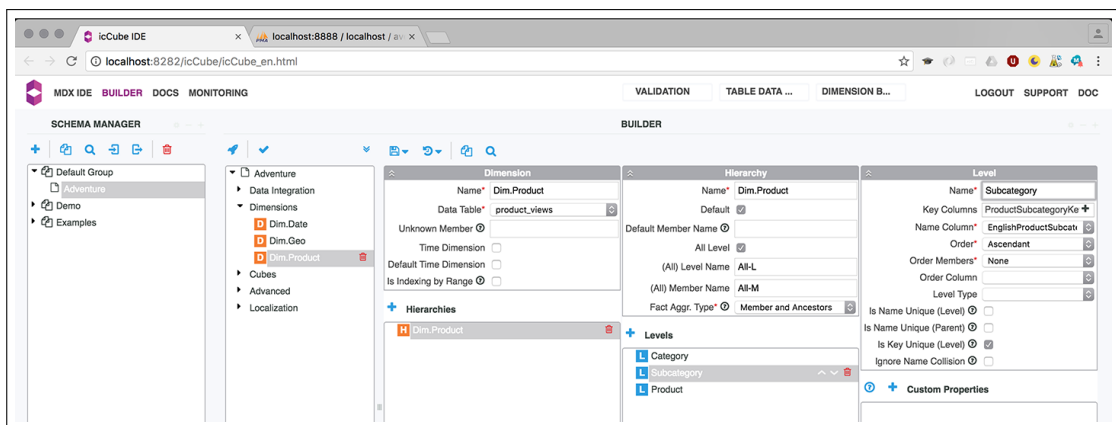


FIGURE 2.11 – Configuration Dim.Product - Subcategory

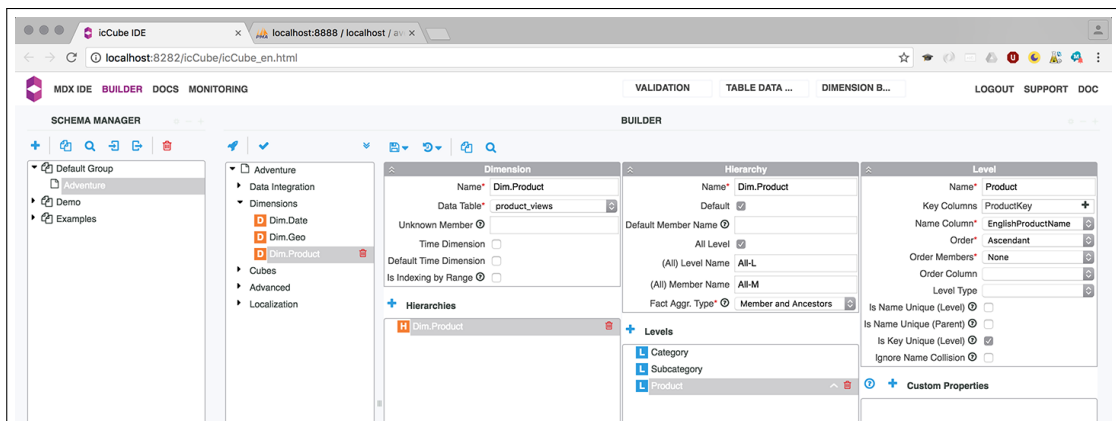


FIGURE 2.12 – Configuration Dim.Product - Product

3 | Exploration

3.1 Slice

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

```
1 select [Dim.Date].[Dim.Date.ByMonth].[Year].[2008] on rows,  
2 [Measures].[SalesAmount] on columns from [AdventureCube]
```

FIGURE 3.1 – Requête Slice : montant total des ventes pour l'année 2008



The screenshot shows the MDX IDE interface. On the left, a tree view shows the 'ADVENTURE' schema with dimensions: AdventureCube, Dim.Geo, Dim.Date, and Dim.Product. The main editor displays the query: `select [Dim.Date].[Dim.Date.ByMonth].[Year].[2008] on rows, [Measures].[SalesAmount] on columns from [AdventureCube]`. Below the editor, the 'RESULT' section shows a table with the following data:

	SalesAmount
2008	26461317.1

FIGURE 3.2 – Résultat requête Slice : montant total des ventes pour l'année 2008

Le résultat obtenu est : 26461317.1

3.2 Dice

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

```

1 select [Dim.Date].[Dim.Date.ByMonth].[Year].[2008] on rows,
2        [Measures].[SalesAmount] on columns
3 from [AdventureCube]
4 where (
5        [Dim.Geo].[Dim.Geo].[Country].[United States].[California],
6        [Dim.Product].[Dim.Product].[Product].[&[485]]
7 )

```

FIGURE 3.3 – Requête Dice : montant des ventes de l'article « Fender Set – Mountain » pour l'année 2008

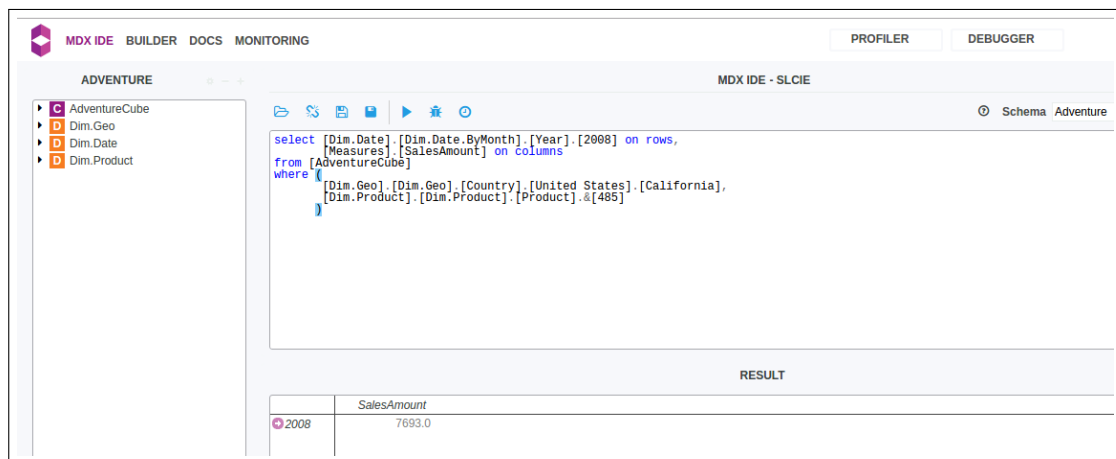


FIGURE 3.4 – Résultat requête Dice : montant des ventes de l'article « Fender Set – Mountain » pour l'année 2008

Le résultat obtenu est : 7693

3.3 Roll-up

Énoncé : 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.

Étas-Unis

```

1 select ([Dim.Geo].[Country].[United States], [Dim.Product].[Product].&[485])
   ↪ on columns,
2 [Measures].[SalesAmount] on rows
3 from [AdventureCube]
4 where [Dim.Date].[Dim.Date.ByMonth].[Year].[2008]

```

FIGURE 3.5 – Requête Roll-up : montant des ventes de l'article « Fender Set – Mountain » États-Unis en 2008



FIGURE 3.6 – Résultat requête Roll-up : montant des ventes de l'article « Fender Set – Mountain » États-Unis en 2008

Le résultat obtenu pour les États-Unis est : 13473.7

Tous les pays

```

1 select ([Dim.Geo].[Dim.Geo],[Dim.Product].[Dim.Product].[Product].&[485]) on
   ↪ columns,
2 [Measures].[SalesAmount] on rows
3 from [AdventureCube]
4 where [Dim.Date].[Dim.Date.ByMonth].[Year].[2008]

```

FIGURE 3.7 – Requête Roll-up : montant des ventes de l'article « Fender Set – Mountain » États-Unis en 2008

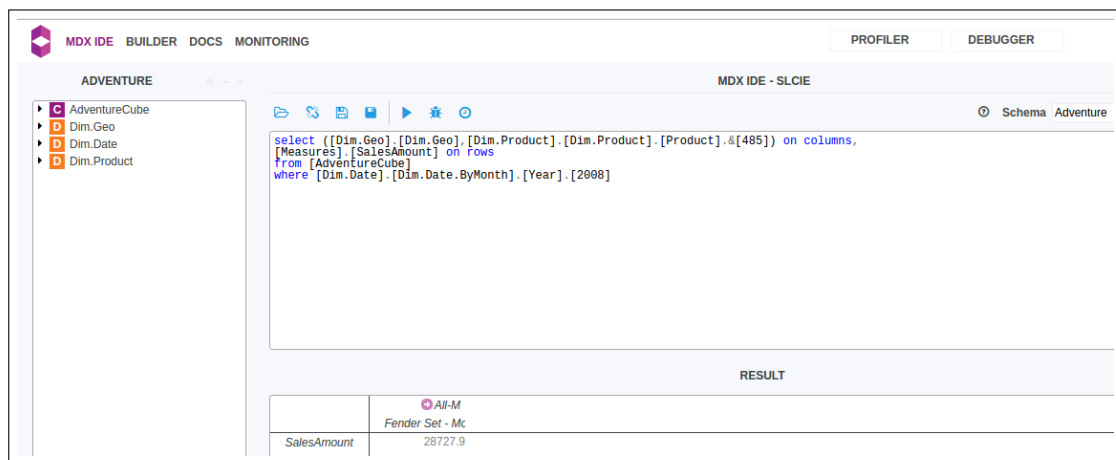


FIGURE 3.8 – Résultat requête Roll-up : montant des ventes de l'article « Fender Set – Mountain » Tous les pays en 2008

Le résultat obtenu pour tous les pays est : 28727.9

3.4 Drill-down

Énoncé : 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 ?

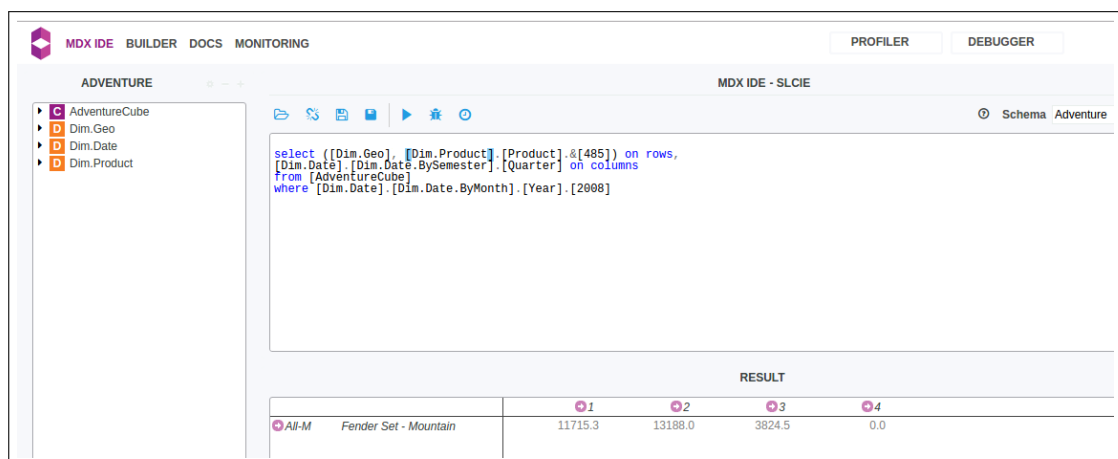
Ordre chronologique

```

1 select ([Dim.Geo], [Dim.Product].[Product].&[485]) on rows,
2 [Dim.Date].[Dim.Date.BySemester].[Quarter] on columns
3 from [AdventureCube]
4 where [Dim.Date].[Dim.Date.ByMonth].[Year].[2008]

```

FIGURE 3.9 – Requête Drill-Down - Chronologique



The screenshot shows the MDX IDE interface. On the left, a tree view shows the 'ADVENTURE' schema with dimensions: Dim.Geo, Dim.Date, and Dim.Product. The main area displays the MDX query from Figure 3.9. Below the query, the 'RESULT' table is shown with columns for the dimensions and a measure. The results are for the year 2008, showing sales amounts for each quarter.

		Q1	Q2	Q3	Q4
All-M	Fender Set - Mountain	11715.3	13188.0	3824.5	0.0

FIGURE 3.10 – Résultat requête Drill-Down - Chronologique

Ordre par trimestre le plus fructueux

```

1 select ([Dim.Geo], [Dim.Product].[Product].&[485]) on rows,
2 order([Dim.Date].[Dim.Date.BySemester].[Quarter], [Measures].[SalesAmount],
3     ↪ DESC) on columns
4 from [AdventureCube]
5 where [Dim.Date].[Dim.Date.ByMonth].[Year].[2008]

```

FIGURE 3.11 – Requête Drill-Down - Trimestre le plus fructueux

3.4. Drill-down

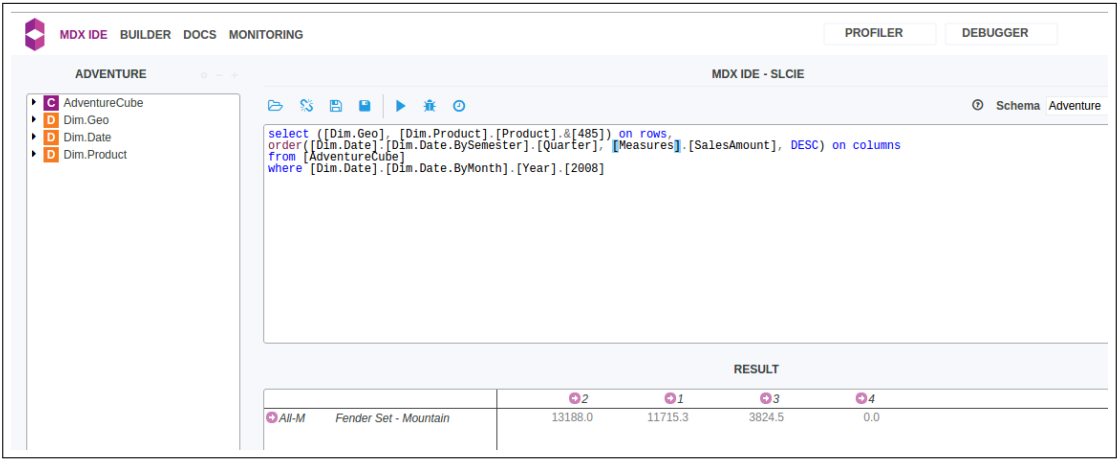


FIGURE 3.12 – Résultat requête Drill-Down - Trimestre le plus fructueux

4 | Conclusion

Comme prévu, ce laboratoire de data warehousing et cube OLAP, nous a permis de mettre en pratique les notions vues en cours.

La création et l'interrogation d'un cube s'est révélée aisée à l'aide du serveur OLAP icCube. Nous constatons donc l'utilité d'un tel outils pour faire de la Business Intelligence en entreprise.

Concernant les différentes opérations OLAP, nous avons trouvé que les noms attribués à ces dernières : Drill down, Roll up, etc. conviennent bien à des cubes à 3 dimensions mais, lorsque le nombre de dimensions du cube n dimensionnel augmentent, il devient plus difficile de se représenter les rotations du cube.