

Лабораторная работа №4

Часть1. Язык MDX

В рамках данной лабораторной работы необходимо рассмотреть основы языка MDX (Multidimensional Expressions) и применить его к созданному хранилищу данных Northwind.

Из теории:

Хранилища данных

Информационные системы серьезного предприятия, как правило, содержат приложения, предназначенные для комплексного анализа данных, их динамики, тенденций и т.п. Соответственно, основными потребителями результатов анализа становится топ-менеджмент. Такой анализ, в конечном итоге, призван содействовать принятию решений. А чтобы принять любое управленческое решение необходимо обладать необходимой для этого информацией, обычно количественной. Для этого необходимо эти данные собрать из всех информационных систем предприятия, привести к общему формату и потом анализировать. Для этого создают хранилища данных

OLAP

Системы поддержки принятия решений обычно обладают средствами предоставления пользователю агрегатных данных для различных выборок из исходного набора в удобном для восприятия и анализа виде (таблицы, диаграммы и т.п.). Традиционный подход сегментирования исходных данных использует выделение из исходных данных одного или нескольких многомерных наборов данных (нередко называемый гиперкубом или метакубом), оси которых содержат атрибуты, а ячейки – агрегируемые количественные данные. Вдоль каждой оси атрибуты могут быть организованы в виде иерархий, представляющих различные уровни их

детализации. Благодаря такой модели данных пользователи могут формулировать сложные запросы, генерировать отчеты, получать подмножества данных.

MDX

Назначение языка MDX (Multidimensional Expressions) — предоставить в распоряжение разработчиков средство для более простого и эффективного доступа к многомерным структурам данных.

Язык SQL изначально был спроектирован не для программистов, а для аналитиков (и поэтому имеет синтаксис, напоминающий естественный язык). Но он со временем все больше усложнялся и теперь мало кто из аналитиков хорошо умеет им пользоваться, если умеет вообще. Он стал инструментом программистов. Язык запросов MDX тоже изначально должен был ориентирован на аналитиков, но его концепции и синтаксис еще сложнее, чем SQL.

Для нас важно, что его open-source реализация используется в виде проекта Mondrian от компании **Pentaho**. Pentaho Analysis Mondrian OLAP Server — OLAP сервер, позволяющий создавать отчеты для онлайн анализа данных, поддерживает язык запросов MDX. (Выбор именно этих средств обоснован в Лабораторной работе №4.)

Работа с аналитическими массивами данных подразумевает только их чтение и не подразумевает запись, т.е. в языке MDX нет предложений для изменения данных, а есть только одно предложение выборки — select.

Синтаксис языка MDX

Запрос на языке MDX представляет собой набор команд, который выглядит следующим образом:

```
SELECT [<axis_specification>
      [, <axis_specification>...]]
FROM [<cube_specification>]
[WHERE
 [< slicer_specification>]]
```

где:

- axis_specification — содержит описание осей куба;
- cube_specification — содержит название куба;
- slicer_specification — содержит описание срезов куба.

В языке MDX выражение SELECT используется для задания набора данных, содержащего подмножество многомерных данных. Простейший SELECT-запрос может выглядеть так:

```
SELECT FROM Sales
```

В этом примере мы получили общее число продаж (Unit Sales) для всего куба. Поскольку в запросе мы не указали имена членов измерений, были выбраны члены по умолчанию из каждого измерения. Наш запрос эквивалентен следующему:

```
SELECT {[Measures].[Unit Sales]} ON COLUMNS  
FROM SALES
```

Более полный SELECT-запрос должен содержать следующую информацию:

- число осей (в одном запросе можно указать до 128 осей);
- список членов измерения, которые должны быть включены для каждой оси;
- имя куба, к которому производится запрос;
- список членов среза.

Рассмотрим более сложный пример, который позволит нам разобраться с различными элементами MDX-запроса:

```
SELECT  
  { [Measures].[Unit Sales], [Measures].[Store Sales] } ON COLUMNS,  
  { [Time].[1997], [Time].[1998] } ON ROWS  
FROM Sales  
WHERE ( [Store].[USA].[CA] )
```

SELECT определяет используемые оси. В нашем примере их две: одна — задает значения для колонок, другая — для рядов:

```
{ [Measures].[Unit Sales], [Measures].[Store Sales] } ON COLUMNS,  
{ [Time].[1997], [Time].[1998] } ON ROWS
```

Выражение FROM определяет источник многомерных данных, к которому обращен наш запрос. В данном примере — это куб Sales.

Выражение WHERE задает размерности или члены, используемые в качестве среза. В нашем примере мы ограничили данные размерностью Store.

Источники:

Более подробную информацию на данную тему можно найти в следующих интернет-источниках.

1. Pentaho. Mondrian Documentation. Официальное описание MDX запросов в Mondrian – Источник: <http://mondrian.pentaho.com/documentation/mdx.php>
2. Введение в многомерный анализ. – Источник: <http://habrahabr.ru/post/126810/>

Выполнение

4-я часть: Формирование MDX запросов

4.1 Подготовительный этап. Запуск Pentaho Community Edition.

На предыдущем этапе мы установили соединение между созданным хранилищем данных и Pentaho Community Edition (см. пункты 5.4 – 5.6 Лабораторной работы №5).

4.2 Настраиваем источник данных для запросов. (Создание хранилища данных)

Выберите нужное соединение. Укажите “Reporting and Analysis (Requires Star Schema)” при создании источника данных” и нажмите “Next”.

Data Source Wizard

> Select Source Type
Select Tables
Define Joins

Data Source Name:
Northwind

Source Type:
Database Table(s)

Select a database connection and click Next to choose from a list of the available database tables.

Connection:

northwind_connection

Create data source for:

☐ Reporting only

☒ Reporting and Analysis (Requires Star Schema)

< Back Next > Finish Cancel

Рисунок 6.2.1 – Выбор источника данных

Далее выбираются таблицы, которые будут участвовать при создании MDX запросов, доступ к данным которых будет разрешён.

Выберем их все и перенесём с помощью стрелок в правый список (рис. 6.2.2). В поле “Fact Table” укажите таблицу фактов (в данном примере sales_fact) (рис. 6.2.3). Она будет определять в дальнейшем меры для формирования запроса.

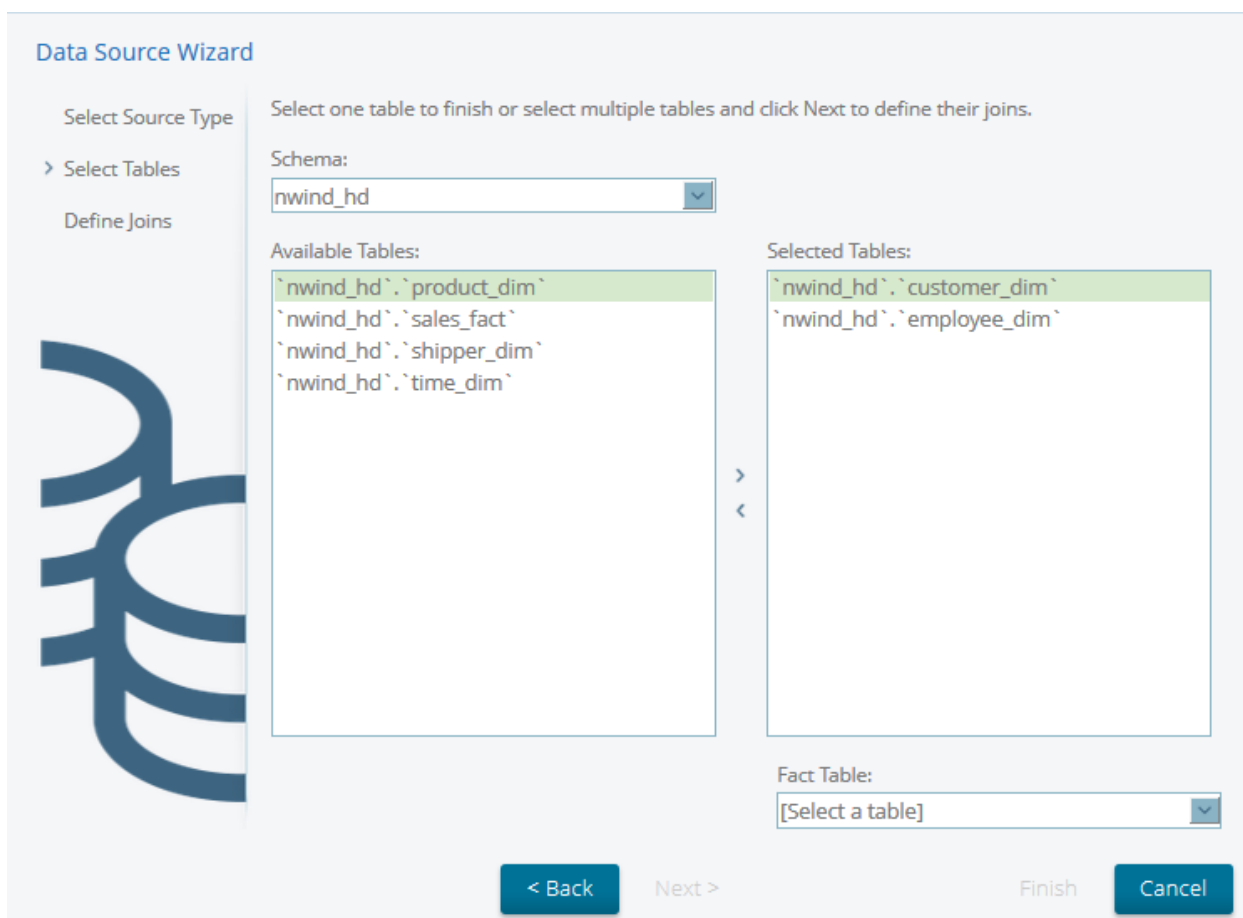


Рисунок 6.2.2 – Выбор таблиц

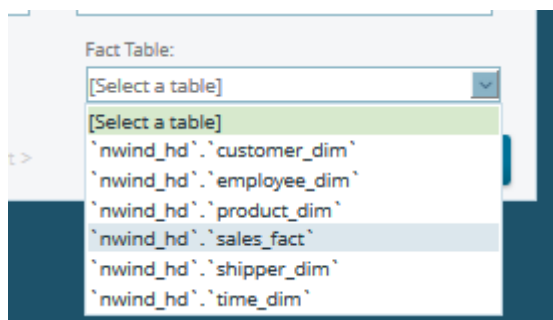


Рисунок 6.2.3 – Указание таблицы фактов

Далее устанавливаем связи между таблицей фактов и всеми остальными. В левой таблице выбирается поле с внешним ключом, а справа соответствующая таблица, в которой хранится тоже ключ в качестве уникального. Нажимаем “create join”, и связь добавляется к списку имеющихся. (См. рис. 6.2.4).

The screenshot shows the 'Data Source Wizard' window, specifically the 'Define Joins' step. The window has a sidebar on the left with three options: 'Select Source Type', 'Select Tables', and 'Define Joins' (which is selected). The main area is titled 'Define how the tables join to each other. All tables must have at least one join defined.' It contains two columns for table selection. The 'Left Table' is set to '`nwind_hd`.`sales_fact`' and the 'Right Table' is set to '`nwind_hd`.`time_dim`'. Below each table name is a list of 'Key Field' options. For the left table, the options are 'TimeKey', 'CustomerKey', 'ShipperKey', 'ProductKey', 'EmployeeKey', 'RequiredDate', and 'LineItemFreight'. For the right table, the options are 'TimeKey', 'TheDate', 'DayOfWeek', 'Month', 'Year', 'Quarter', and 'DayOfYear'. The 'TimeKey' field is selected in both lists. A 'Create Join' button is located to the right of the key field lists. Below these lists is a 'Join(s):' section containing a text box with the text: '`nwind_hd`.`sales_fact`.TimeKey - INNER JOIN - `nwind_hd`.`time_dim`.TimeKey'. A 'Delete Join' button is located to the right of the join list. At the bottom of the window are four buttons: '< Back', 'Next >', 'Finish', and 'Cancel'.

Рисунок 6.2.4 – Установление связей между таблицами

Создание хранилища данных почти завершено. Во всплывшем окне пользователю предлагается создать модель по умолчанию (“Keep default model”) или настроить её вручную (“Customize model new”). Выберите второй пункт для пользовательской настройки модели.

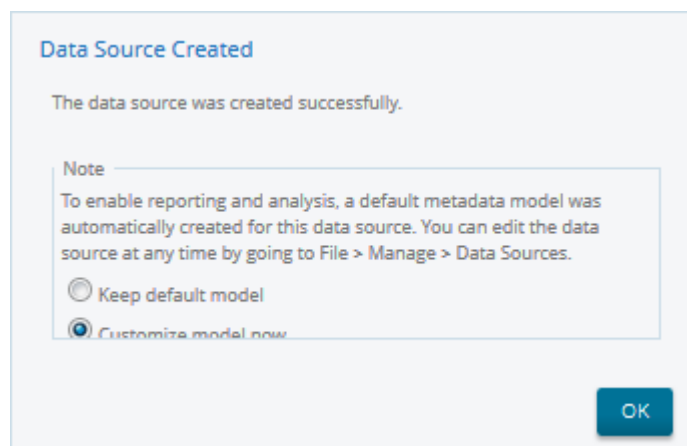


Рисунок 6.2.5 – Выбор модели хранилища данных

При создании мер необходимо создать те меры, которые в дальнейшем будут использоваться, Автоматически создаются меры для всех столбцов в таблице фактов, однако в данном примере нам понадобятся только две:

- LineItemQuantity – количество проданных товаров;
- LineItemTotal – стоимость проданных товаров.

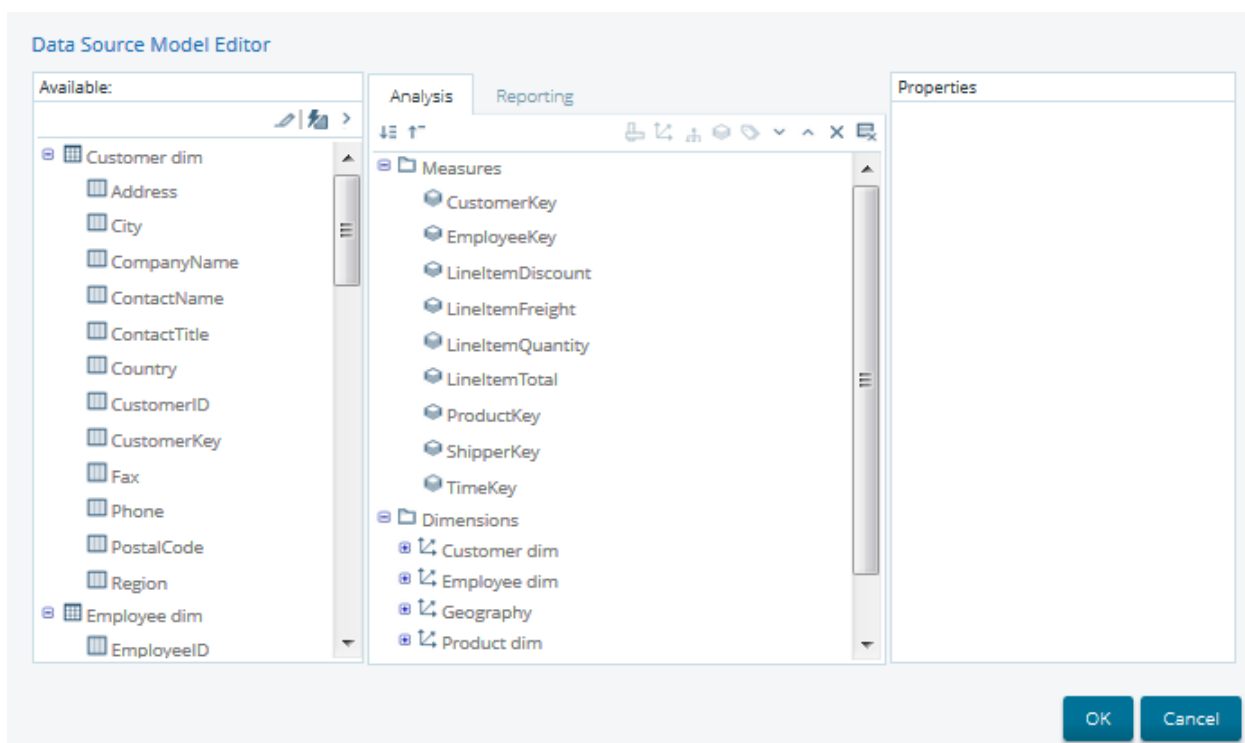


Рисунок 6.2.6 – Автоматически созданные меры

Отредактируем список мер, оставив лишь необходимые и задав им поясняющие имена:

- UnitSales – единица товаров;
- CostSales – стоимость товаров.

И параметр агрегации (default aggregation) обозначим суммирование (SUM). (См. рис. 6.2.7)

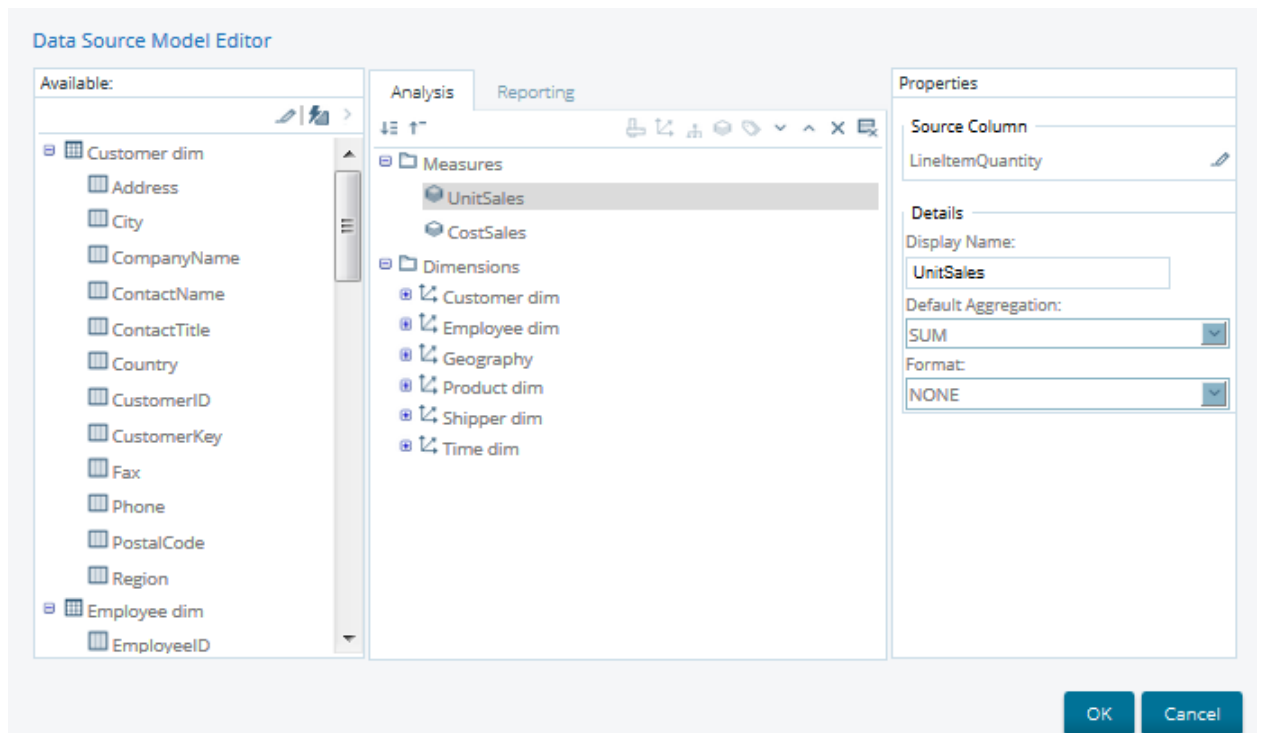


Рисунок 6.2.7 – Редактирование мер

Перейдём к редактированию измерений (Dimensions).

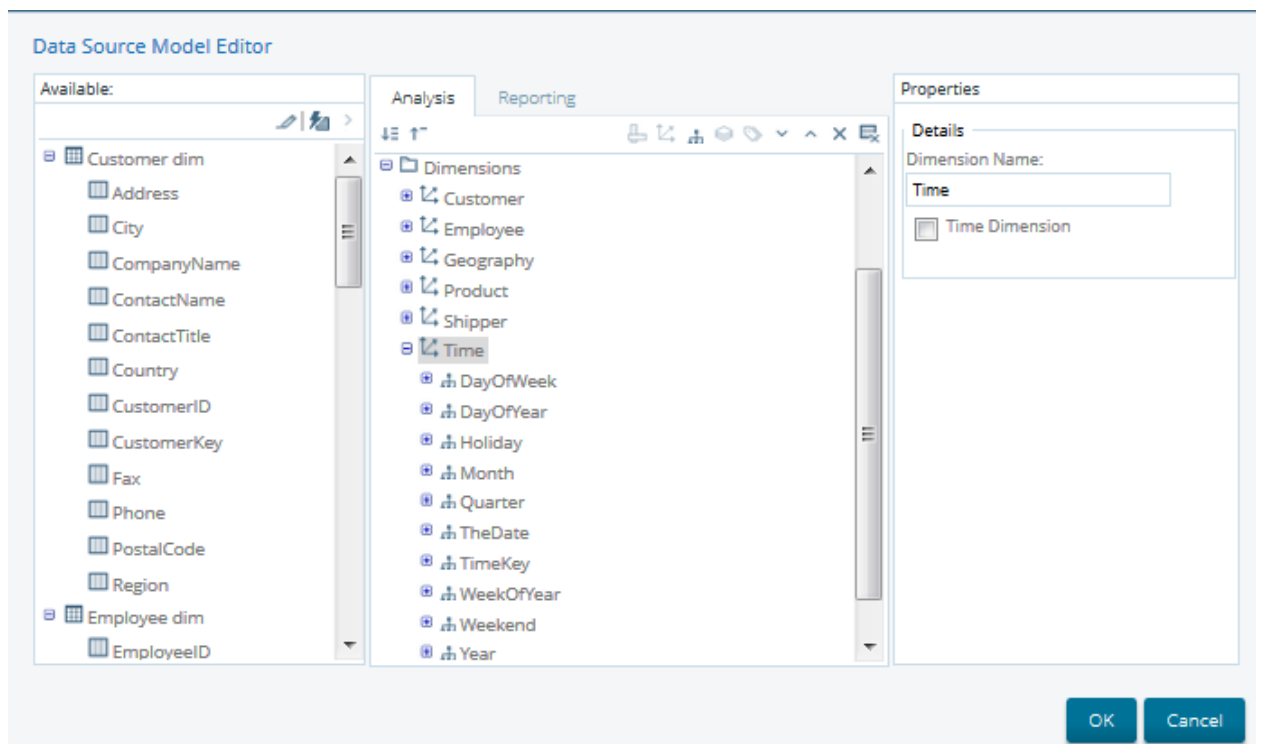


Рисунок 6.2.8 – Список измерений и поля измерения “time”

Из всего длинного списка интервалов времени (см. рис. 6.2.8) нас интересуют лишь год, квартал и месяц. Обязательно укажите, что данное измерение является временем. После того, как поставили галочку, необходимо каждому полю (год, месяц, квартал) указать соответствующий уровень (рис. 6.2.10).

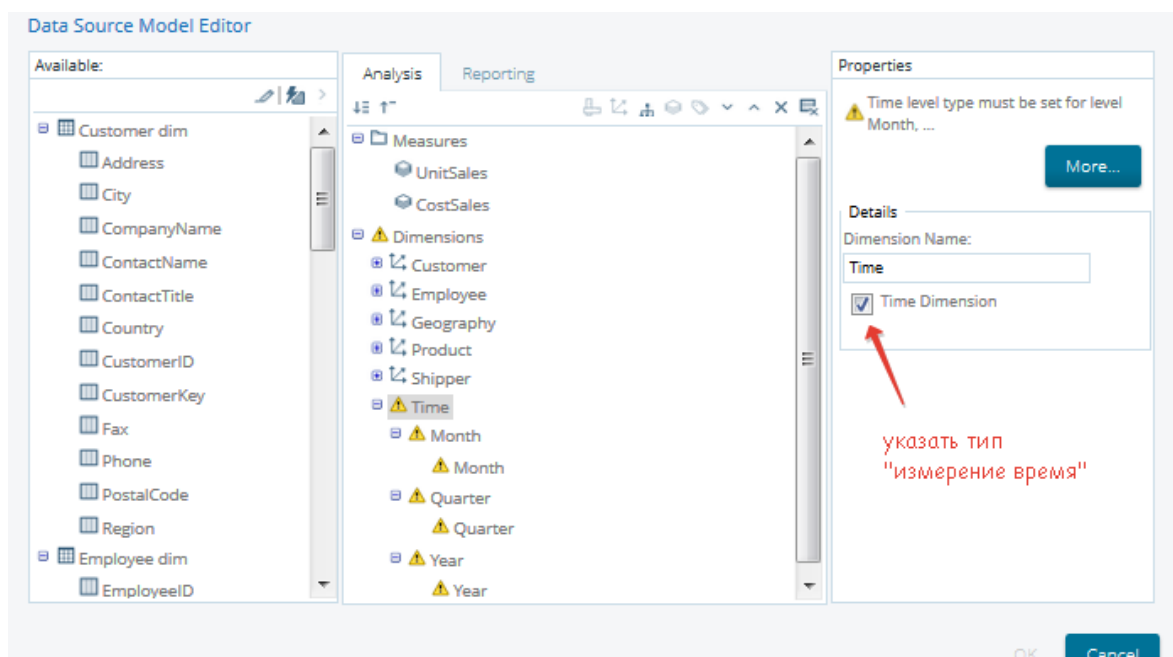


Рисунок 6.2.9 – Редактирование измерения время

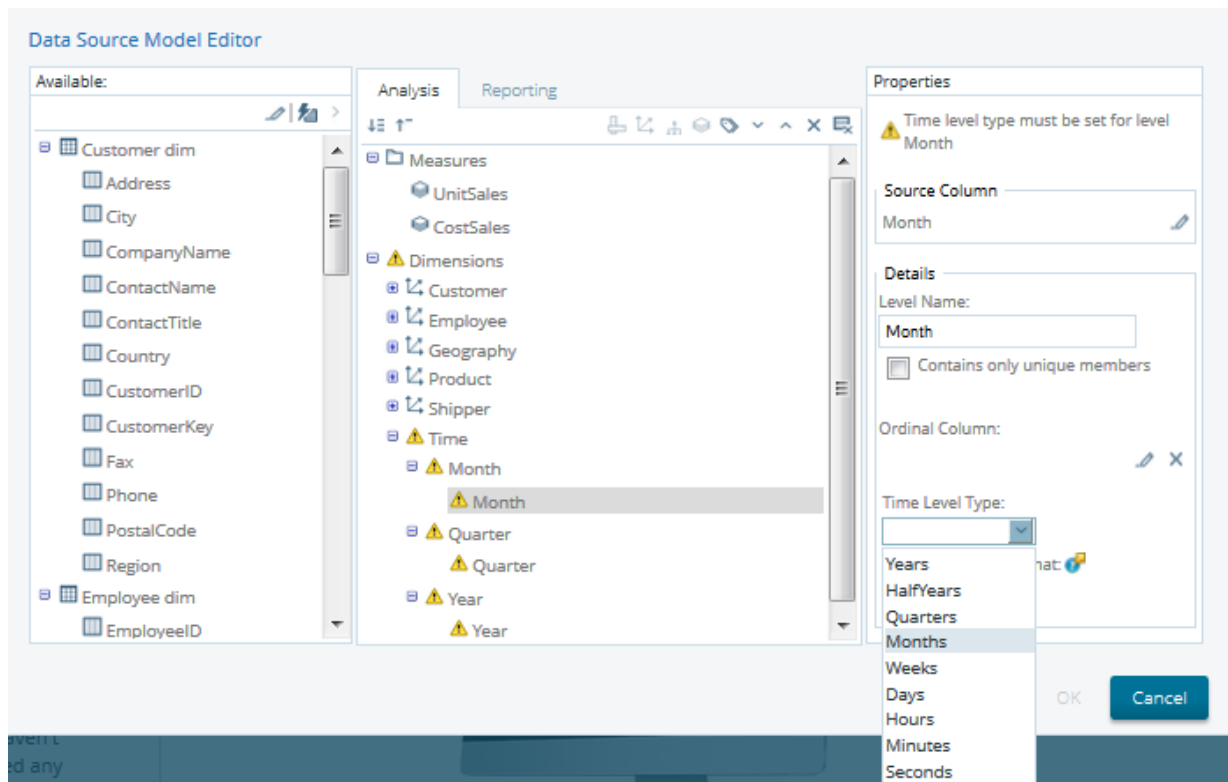


Рисунок 6.2.10 – Указание интервала времени

Заметьте, что в измерении “geography” тип полей указывается автоматически, но стоит проверить (рис. 6.2.11). В остальных измерениях тип полей не определён и по умолчанию стоит NONE.

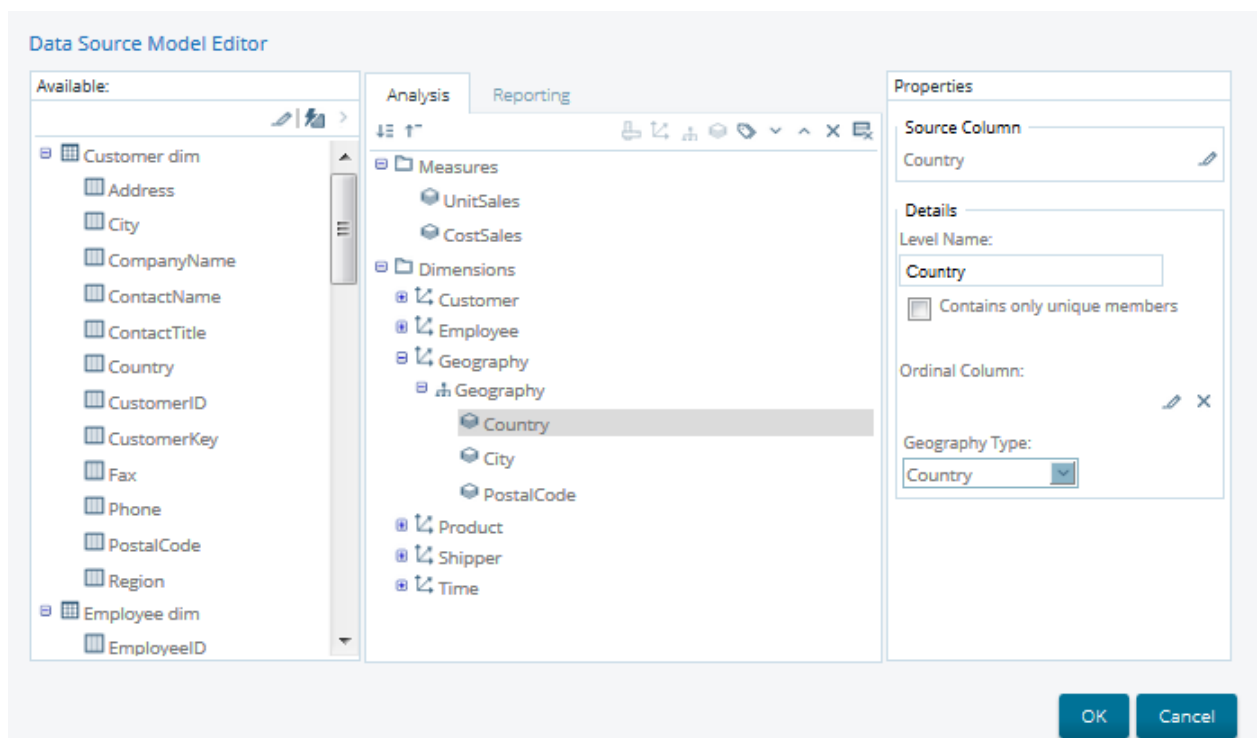


Рисунок 6.2.11 – Тип полей измерения geography

Хранилище данных настроено, и можно переходить к формированию MDX запросов.

4.3 MDX-запросы

На главной странице Pentaho User Console выберите File -> New -> Saiku Analytics. На открывшейся странице, на вкладке Enterprise в разделе Quick Links нажмите “создать новый запрос” (Create a new query).

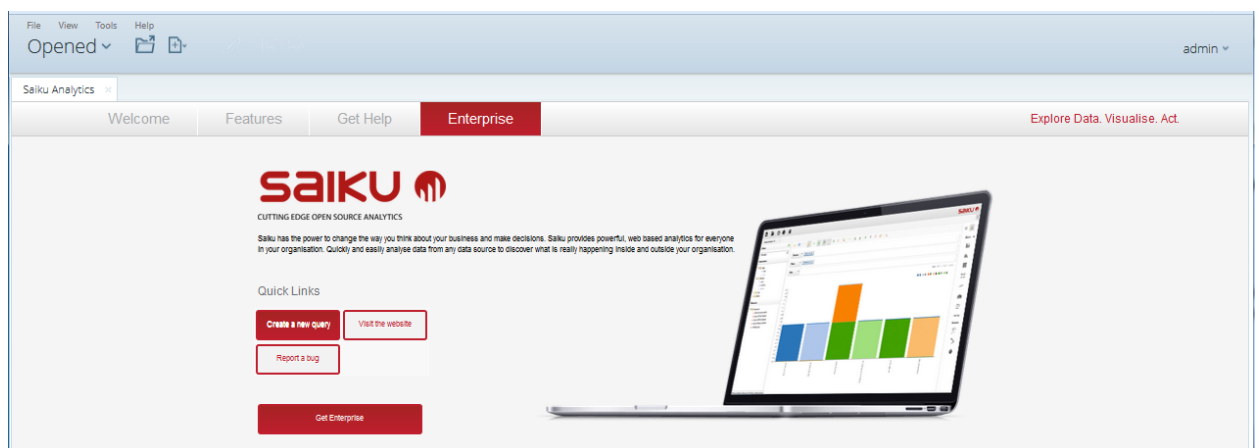


Рисунок 6.3.1 – Создание нового запроса

В первую очередь необходимо указать куб, с которым будем работать в процессе формирования запросов. Saiku Analytics позволяет визуально создавать нужный запрос к хранилищу. Рассмотрим на примере Northwind.

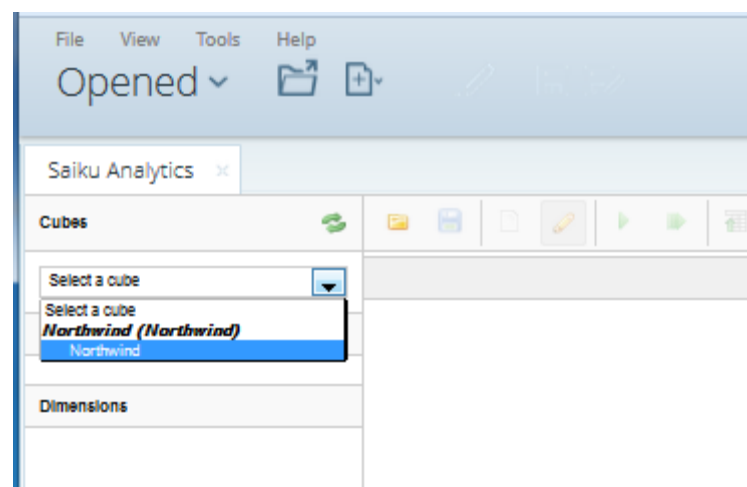


Рисунок 6.3.2 – Выбор куба, к которому формируется запрос

Запрос №1: *Каково число продаж было совершено в каждом месяце в каждом конкретном городе с офисом компании за три года?*

1) Укажите меру. В нашем случае UnitSales – число продаж.

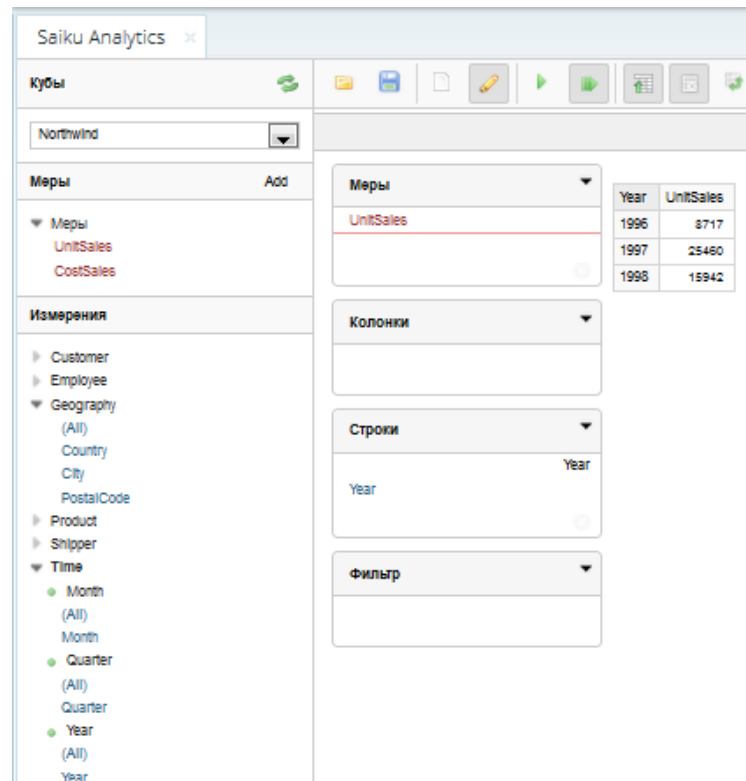


Рисунок 6.3.3 – Выбор меры

2) Укажите измерения времени – год и месяц. Заметьте, что сначала указываются измерения в строке, а затем заполняются столбцы. Измерение можно вручную переместить из столбца в строку. (См. рис. 6.3.5)

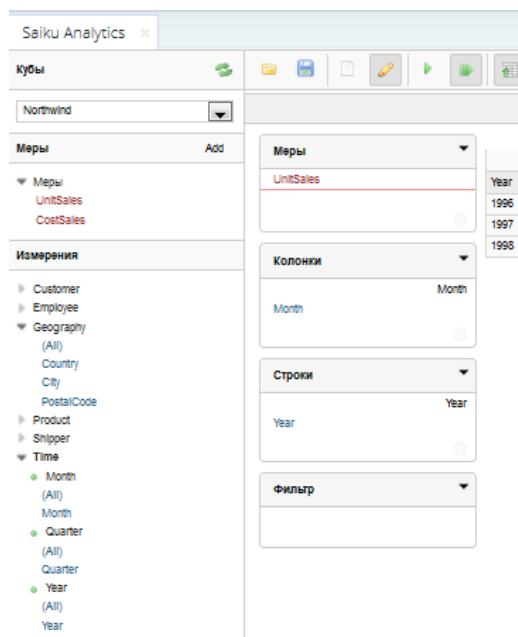


Рисунок 6.3.4 – Задание строк и столбцов

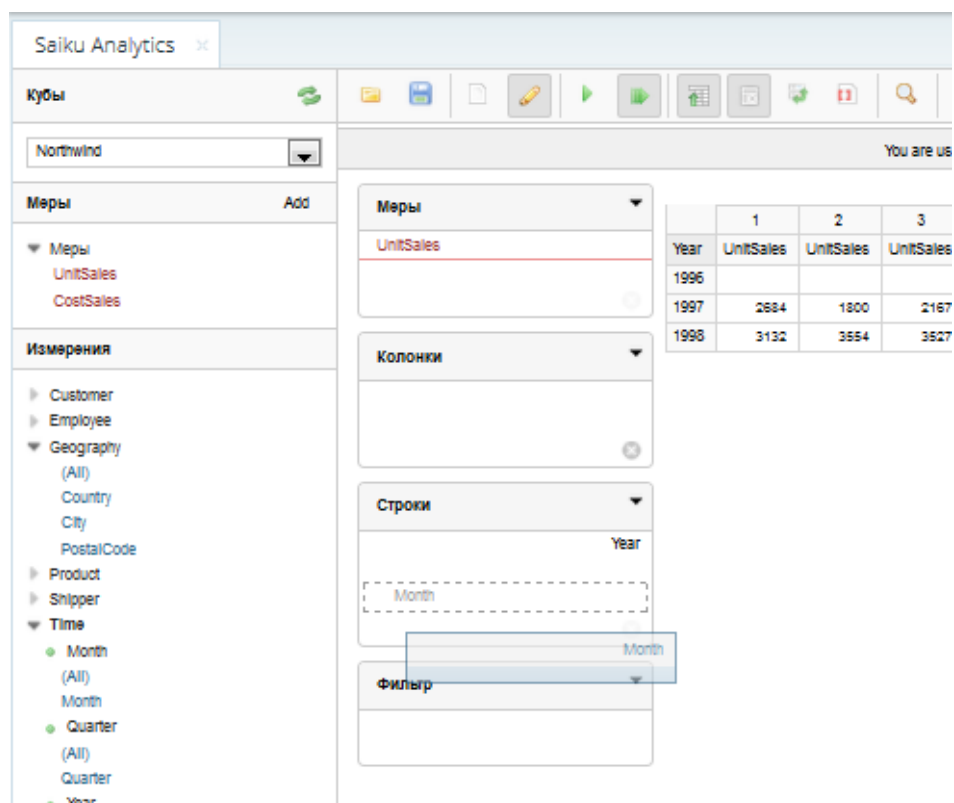
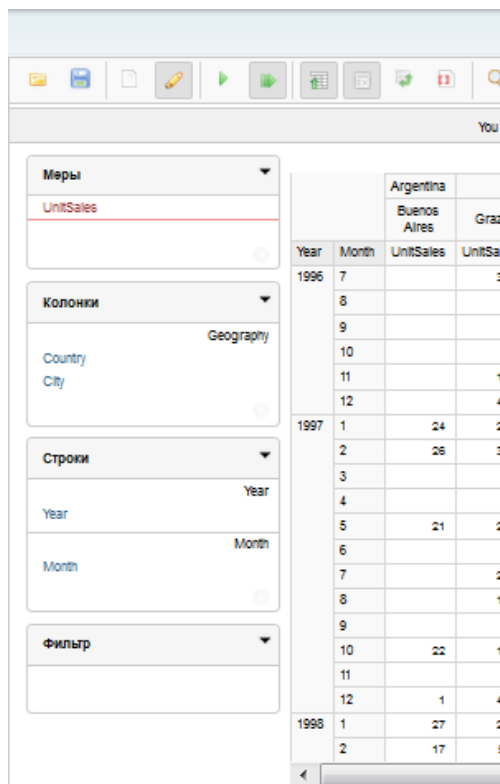


Рисунок 6.3.5 – Настройка строк и столбцов вручную

3) Укажите измерение месторасположения – страна, город.

По мере того, как детализируется запрос, справа меняется таблица вывода. Наглядно можно убедиться в разнице запросов, если в строках сначала указывается месяц, а потом год и наоборот.



		Argentina	
		Buenos Aires	Graz
Year	Month	UnitSales	UnitSal
1996	7		3
	8		
	9		
	10		
	11		1
	12		4
1997	1	24	2
	2	26	3
	3		
	4		
	5	21	2
	6		
	7		2
	8		1
	9		
	10	22	1
	11		
	12	1	4
1998	1	27	2
	2	17	5



		Argentina	
		Buenos Aires	Graz
Month	Year	UnitSales	UnitSal
1	1997	24	
	1998	27	
2	1997	26	
	1998	17	
3	1997		
	1998	96	
4	1997		
	1998	70	
5	1997	21	
	1998		
6	1997		
7	1996		
	1997		
8	1996		
	1997		
9	1996		
	1997		
10	1996		
	1997	22	
11	1996		

Рисунок 6.3.6 а) – Года, разложенные по месяцам

Рисунок 6.3.3 б) – Месяцы по каждому году

Так, мы научились создавать запрос в конструкторе запросов. Панель инструментов Saiku Analytics предоставляет возможность:

- открывать сохранённые запросы;

- сохранять сформированные запросы;
- сбросить настройки запроса;
- экспортировать данные, например в pdf файл;
- а также написать запрос вручную и связанные с этим функции построения запроса.

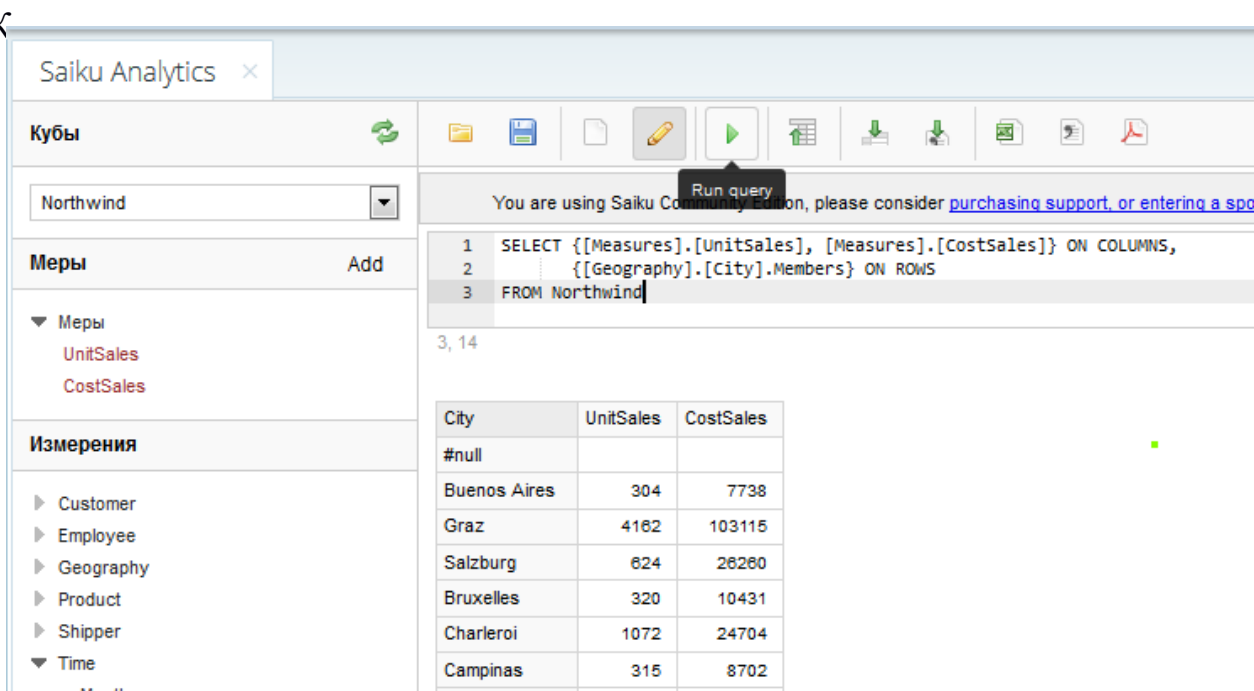
как
выглядит
написание
приведённого
выше
запроса
№1
вручную,

можно посмотреть, кликнув на панели управления “switch to MDX Mode”:

Листинг.1 – Запрос №1

```

1 WITH
2 SET [~COLUMNS] AS
3   Hierarchize({ [{[Geography].[Country].Members}],
4 SET [~ROWS_Time .Month] AS
5   {[Time .Month].[Month].Members}
6 SET [~ROWS_Time .Year] AS
7   {[Time .Year].[Year].Members}
8 SELECT
9   NON EMPTY CrossJoin([~COLUMNS], {[Measures].[UnitSales]}) ON COLUMNS,
10  NON EMPTY NonEmptyCrossJoin([~ROWS_Time .Month], [~ROWS_Time .Year]) ON ROWS
11 FROM [Northwind]
```



Рассмотрим написание простого запроса вручную.

Запрос №2: Сравнительный анализ числа продаж и стоимостей в различных городах.

Листинг.2 – Запрос №2

```

1 SELECT {[Measures].[UnitSales], [Measures].[CostSales]} ON COLUMNS,
2       {[Geography].[City].Members} ON ROWS
3 FROM Northwind
```

Результатом данного запроса является длинная таблица со списком всех городов базы данных.

Рисунок 6.3.7 – Написание запроса вручную

Отметим, что первая строка таблицы пустая. Чтобы этого избежать, требуется ввести NON

1	SELECT {[Measures].[UnitSales], [Measures].[CostSales]} ON COLUMNS,
2	NON EMPTY {[Geography].[City].Members} ON ROWS
3	FROM Northwind
4	

2, 17

City	UnitSales	CostSales
Buenos Aires	304	7738
Graz	4162	103115
Salzburg	624	26260
Bruxelles	320	10431
Charleroi	1072	24704
Campinas	315	8702

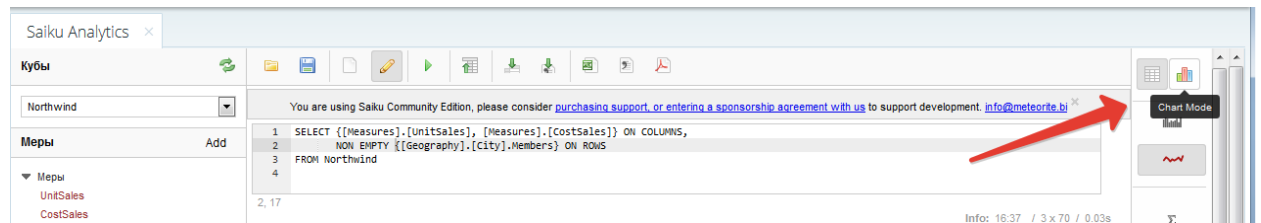
EMPTY перед определением строк, как это сделано в автоматическом запросе.

Рисунок 6.3.8 – Исключение пустой строки

Рисунок 6.3.10 – Графическое отображение запроса №2

Помимо таблиц результат запроса можно выводить в графическом виде. Это могут быть кольцевые или столбчатые диаграммы, широкий выбор

Для более наглядного примера кольцевых диаграмм



которых представлен справа при смене режима Table Mode на Chart Mode.

выполним следующий запрос.

Запрос №3: Сравнительный анализ числа продаж и стоимостей в различных странах.

Рисунок 6.3.9 – Смена режима

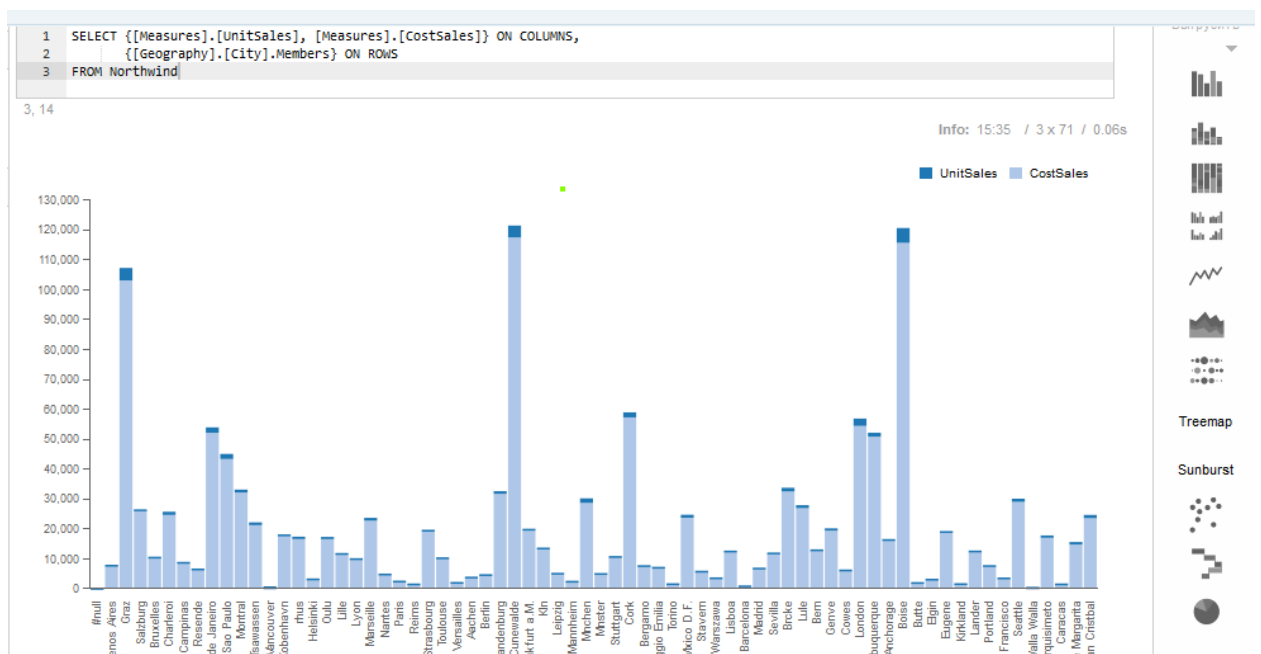
Листинг.3 – Запрос №3

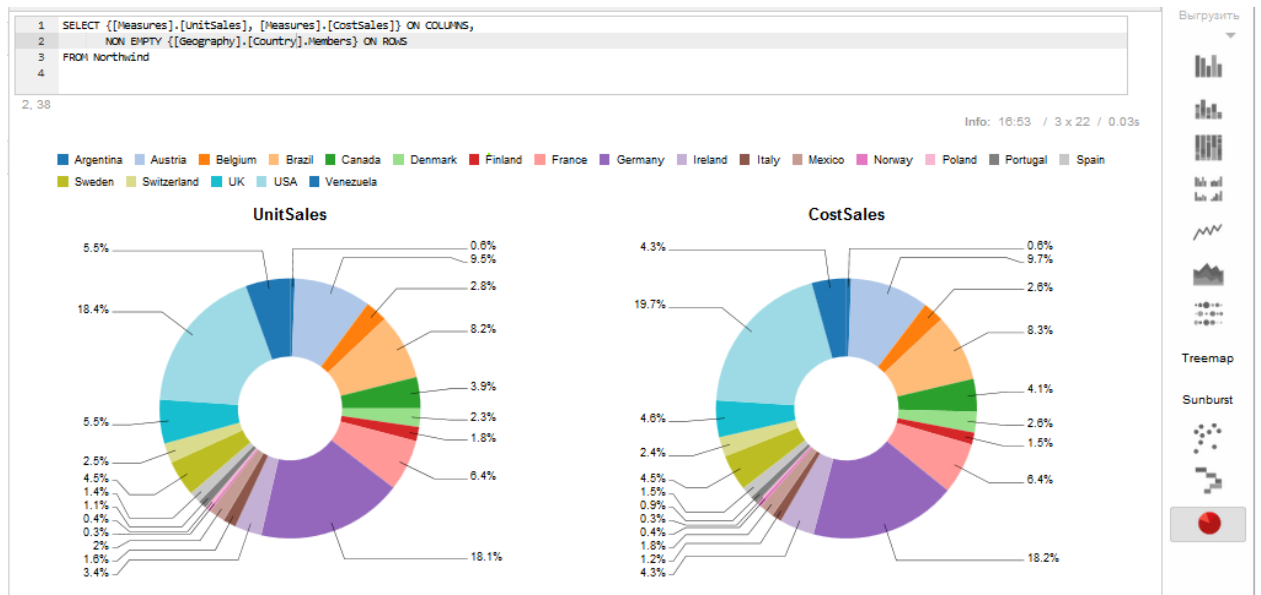
Графическим результатом запроса №2 является следующая столбчатая диаграмма, наглядно отображающая число продаж в каждом городе.

```

1 SELECT {[Measures].[UnitSales], [Measures].[CostSales]} ON COLUMNS,
2     NON EMPTY {[Geography].[Country].Members} ON ROWS
3 FROM Northwind
  
```

Рисунок 6.3.11 – Графическое





отображение запроса №3.
Кольцевая диаграмма

Часть 2. Написать запросы к ХД «Northwind_hd» со следующими функциями:

1. Функция- .Members
2. .PrevMember
3. .NextMember
4. Lag(N),
5. Lead(-N),
6. Функция – .children
7. Функция CrossJoin
8. Функция Filter()
9. Функция ORDER()
10. Функция NONEMPTY

В отчете привести тексты
запросов, табличные и
графические результаты
запросов