МИНОБРНАУКИ РОССИИ

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ «САРАТОВСКИЙ НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ ИМЕНИ Н.Г.ЧЕРНЫШЕВСКОГО»

Кафедра теории функций и стохастического анализа

ПРАКТИЧЕСКАЯ РАБОТА ПО ДИСЦИПЛИНЕ «ТЕХНОЛОГИЯ МНОГОМЕРНЫХ БАЗ ДАННЫХ»

ОТЧЕТ

студента (ки)	1	_ курса	148	_ группы	
направления	09.04.03 – Прикладная информатика				
	код и наименование направления				
механико-математического факультета					
наименование факультета					
Вайцуля Александра Николаевича					
фамилия, имя, отчество					
Преподаватель:					
Доцент, к.ф м. нау	/К				Е.Ю. Крылова
должность, уч. степен	,		под	пись, дата	инициалы, фамилия
уч. звание					

1 Постановка задачи

Перед созданием OLAP-куба и анализом, необходимо установить необходимое ПО. Для этого были установлены MS SQL Server 2022, SQL Management Studio 2019, и дополнительное расширение Analysis Services для многмерного анализа. Все необходимые компоненты представлены на рисунке 1.

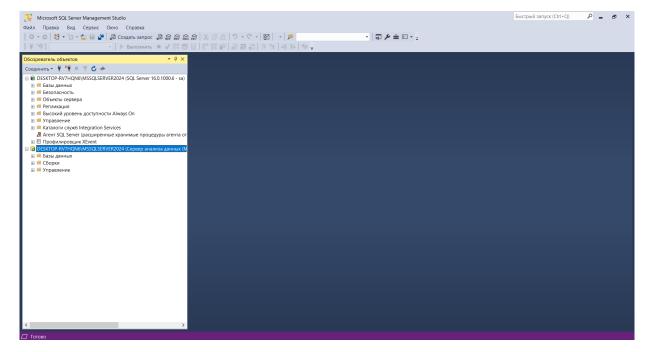


Рисунок 1 - Установленные компоненты.

2 Выбор предметной области

В качестве предметной области выбрана крупная многонациональная компания, которая производит велосипеды из металла и композитных материалов, а продукция экспортируется на рынки Северной Америки, Европы и Азии.

БД построена по схеме "снежинки" (Snowflake scheme). В схеме содержится 2 таблицы фактов FactInternetSales (продажи онлайн), FactResellerSales (продажи офлайн). Обе таблицы фактов соединены со списками либо напрямую, либо через промежуточные таблицы.

Диаграмма исходной БД представлена на рисунке 2.

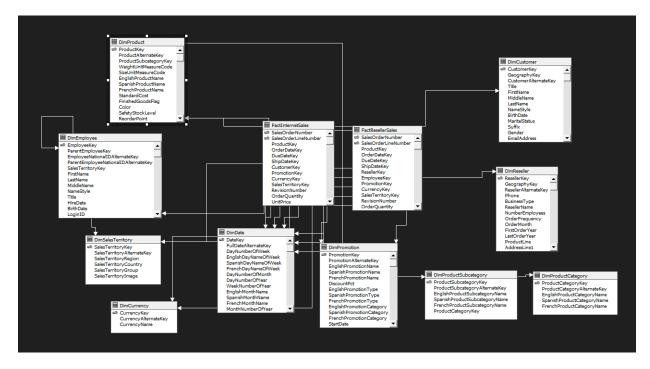


Рисунок 2 - Диаграмма исходной БД для анализа

3 Исходные данные для БД

Таблицами измерений являются следующие таблицы:

DimDate — таблица, содержащая всю информацию по датам. В ней указаны назавания дней недели, месяцев и хранится информация по годам. Пример данных приведен на рисунке 3.

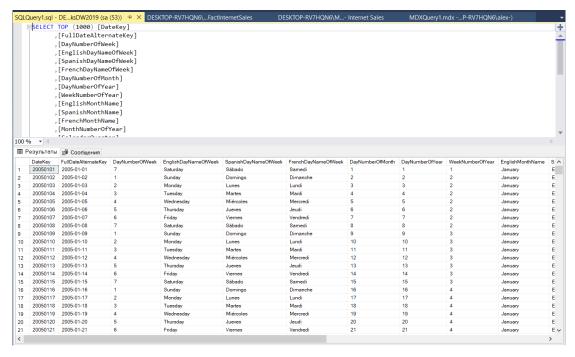


Рисунок 3 - Структура и данные таблицы DimDate

DimCustomer – таблица, содержащая информацию по клиентам, которые зарегистрированы в системе и могут совершать операции.

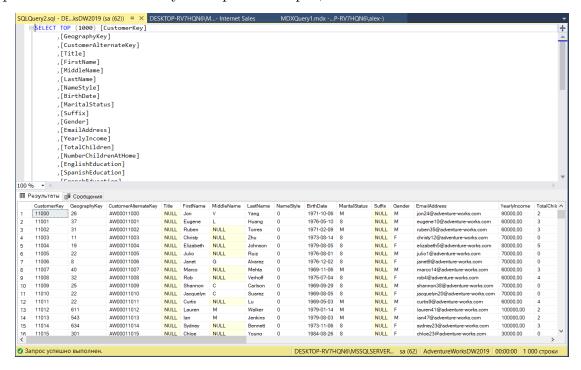


Рисунок 4 - Структура и данные таблицы DimCustomer

DimProduct – таблица, содержащая информацию по продуктам компании.

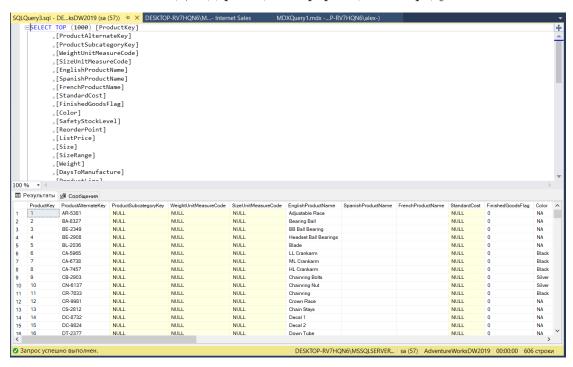


Рисунок 5 - Структура и данные таблицы DimProduct

DimPromotion – таблица, содержащая информацию по акциям и продвижению.

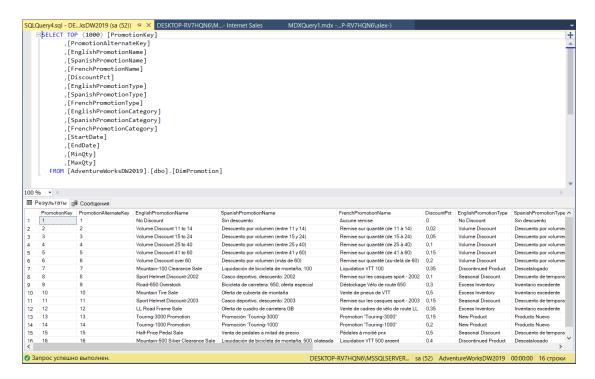


Рисунок 6 - Структура и данные таблицы DimPromotion

DimProductSubcategory — таблица, содержащая информацию по подкатегории товаров.

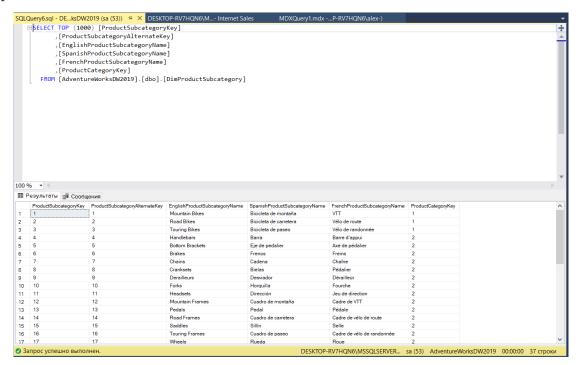


Рисунок 7 - Структура и данные таблицы DimProductSubcategory

DimProductCategory – таблица, содержащая информацию по категории товаров.

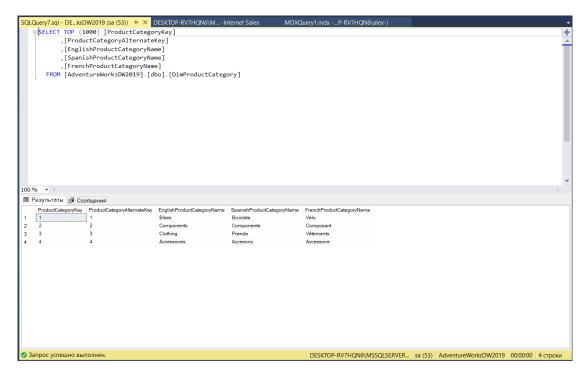


Рисунок 8 - Структура и данные таблицы DimProductCategory

DimSalesTerritory — таблица, содержащая информацию по странам, регионам, штатам.

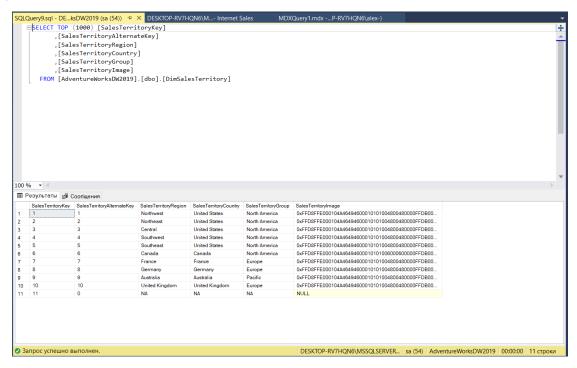


Рисунок 9 - Структура и данные таблицы DimSalesTerritory

DimEmployee – таблица с информацией о сотрудниках.

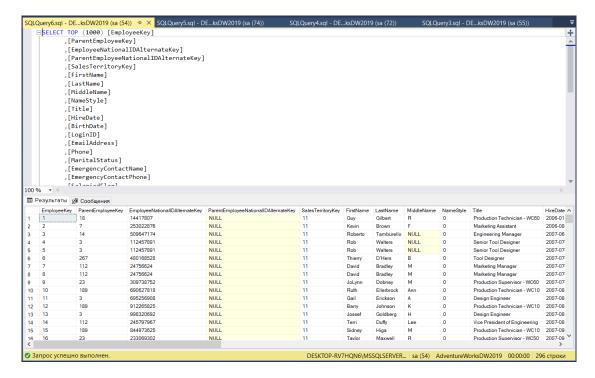


Рисунок 10 - Структура и данные таблицы DimEmployee

DimReseller – таблица с информацией о посреднике.

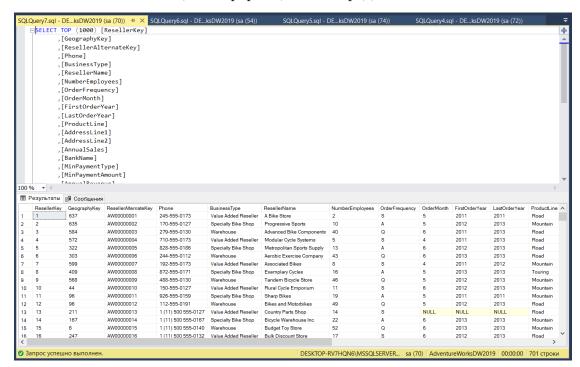


Рисунок 11 - Структура и данные таблицы DimReseller

FactInternetSales — таблица фактов, в которой хранятся меры покупок в интернете и внешние ключи.

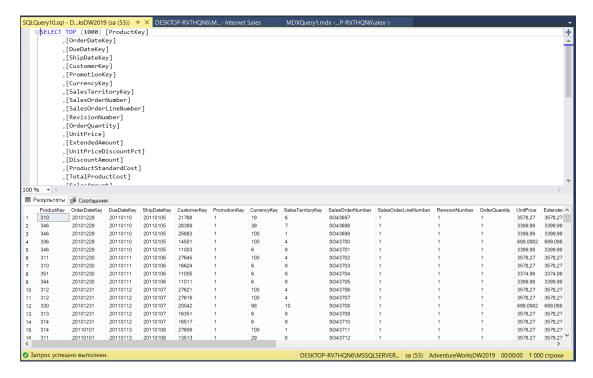


Рисунок 12 - Структура и данные таблицы FactInternetSales

FactResselerSales – таблица фактов, характеризующая покпку товаров в розницу.

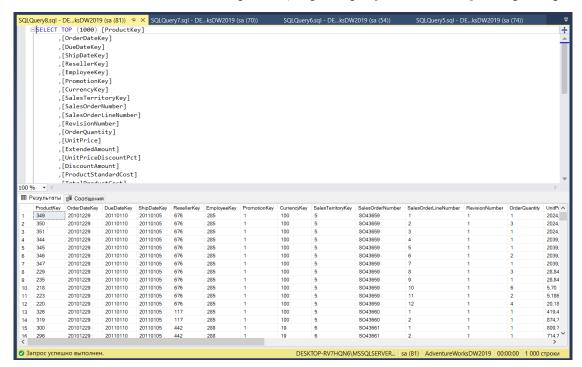


Рисунок 13 - Структура и данные таблицы FactResselerSales

4 OLAP-куб

Перйдем к построению OLAP-куба. Необходимо объявить источник данных. В данном случае это хранилище AdventureWorksDW2019.

Инструмент, позволяющий построить куб является Visual Studio 2022.

Создание источника данных включает в себя несколько этапов:

- 1. Создание нового проекта;
- 2. Выбрать некоторое хранилище в качестве источника данных.

Каждый шаг описан на рисунках ниже.

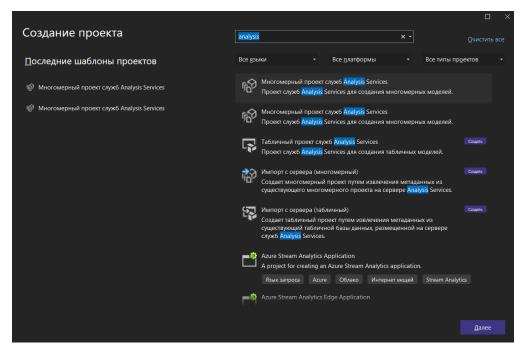


Рисунок 14 - Создание нового проекта многомерных служб

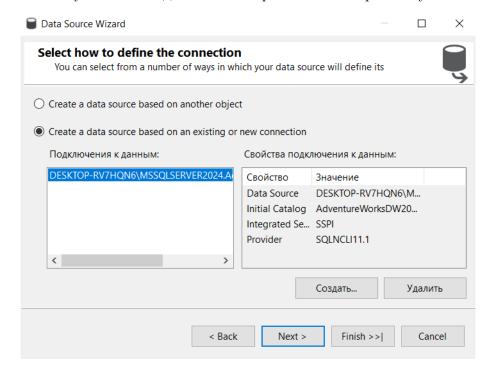


Рисунок 15 - Задание определенного хранилища в качестве источника данных

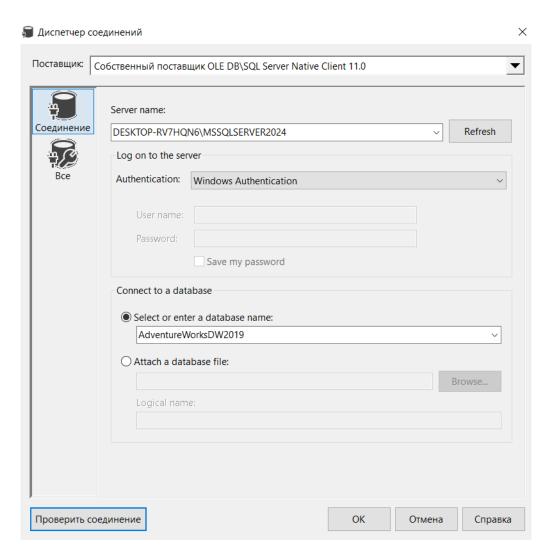


Рисунок 16 - Определение источника данных

После определения источника данных, необходимо определить таблицы и представления из данного хранилища. На этом этапе мы выделяем те факты, с которыми в дальнейшем будем производить аналитику, опираясь на размерности. Таким образом, необходимо задать таблицы фактов и измерений. Выберем в качестве таких таблиц следующие: FactInternetSales, FactResselerSales - таблицы фактов. Таблицы измерений: DimCustomer, DimDate, DimProduct, DimPromotion, DimSalesTerritory, DimEmployee, DimProductCategory, DimProductSubcategory, DimReseller. Этот шаг представлен на следующем рисунке 17.

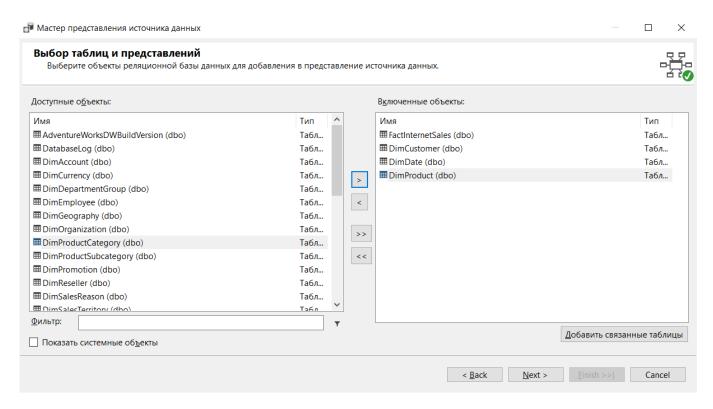


Рисунок 17 - Определение таблиц и представлений из хранилища

В результате выполнения описанных шагов, была получена следующая схема заданного источника данных.

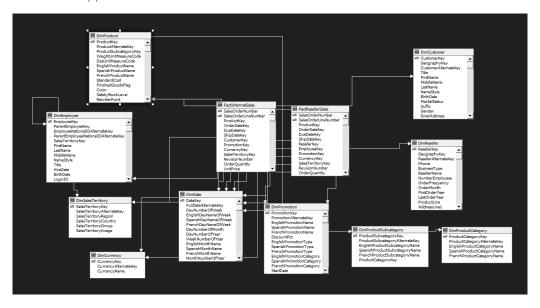


Рисунок 18 - Схема источника данных для OLAP-куба

После того, как подготовительный этап к созданию OLAP-куба был завершен, можно переходить к его построению и обработке. На этом же этапе будут заданы измерения для OLAP-куба. Шаги по результатам такого построения представлены на рисунках (19-21).

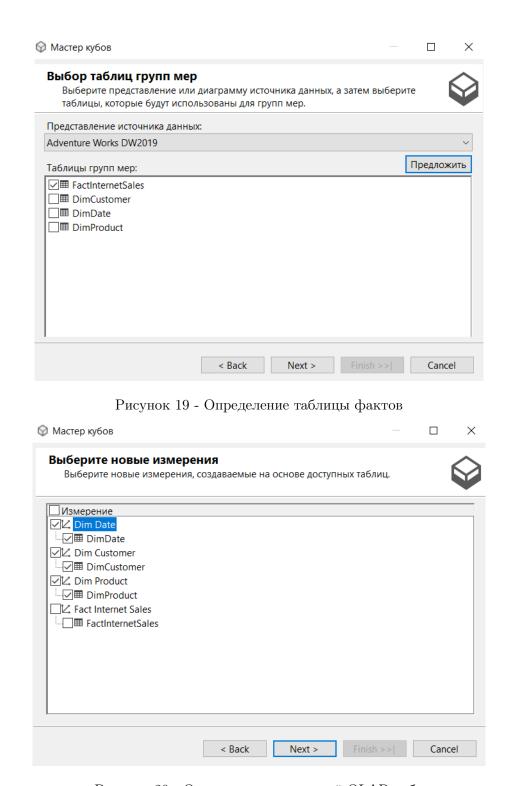


Рисунок 20 - Определение измерений OLAP-куба

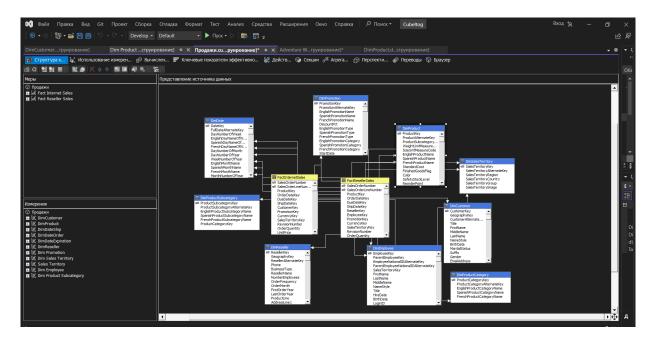


Рисунок 21 - Результат построения OLAP-куба

После того, как создание куба завершено, необходимо его развернуть в MS SQL Server.

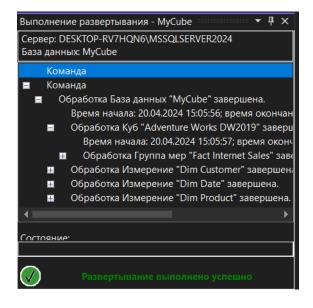


Рисунок 22 - Успешное развертывание OLAP-куба

В результате полученный куб доступен для работы из SQL Management Studio.

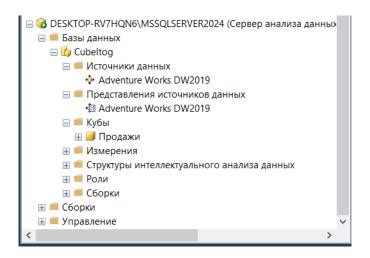


Рисунок 23 - Развернутый OLAP-куб в SQL Management

K развернотому OLAP-кубу можно писать MDX-запросы, для этого будет использована программа SQL Management Studio 19.

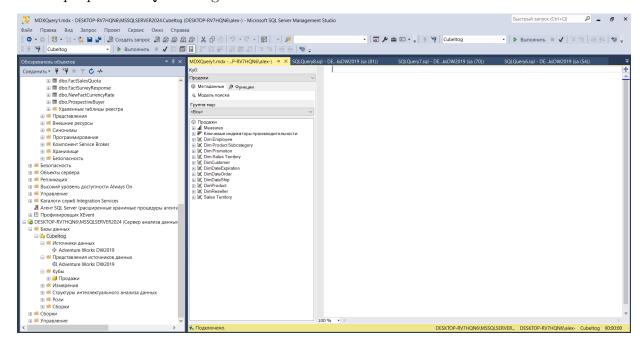


Рисунок 24 - Окно для написания МDX-запросов

5 МДХ-запросы

Перейдем к написанию MDX-запросов к построенному в предыдущих разделах OLAP-кубу. Исходя из предназначения данных куба, сформулируем несколько основных запросов, которые могут быть применимы с практической точки зрения для данных, со-держащихся в кубе.

1 Запрос. Количество товара черного цвета по виду в соответствии с 2010 - 2012 годы, купленные онлайн. Код данного MDX-запроса представлен ниже, а его результат

приведен на рисунке — 25.

SELECT {[DimDateOrder].[Год].&[2010], [DimDateOrder].[Год].&[2011],

[DimDateOrder].[Год].&[2012]} ON COLUMNS,

([DimProduct].[Цвет].&[Black], {[DimProduct].[Вид].&[М],

[DimProduct].[Вид].&[U], [DimProduct].[Вид].&[W]}) ON ROWS

FROM [Продажи]

WHERE [Measures].[Order Quantity - Fact Internet Sales]

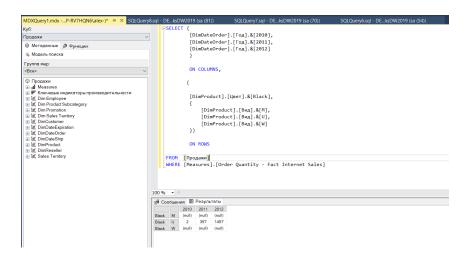


Рисунок 25 - Результат MDX-запроса №1

2 Запрос. Разность в количестве проданного товара через интернет и офлайн в соответствии с годами, полом, типом товаров черного цвета. Код данного MDX-запроса представлен ниже, а его результат приведен на рисунке – 26.

```
WITH MEMBER [Dif] AS ([Measures].[Order Quantity]
-[Measures].[Order Quantity-Fact Internet Sales])

SELECT ({[DimDateOrder].[Год].&[2011], [DimDateOrder].[Год].&[2012]},

{[DimProduct].[Вид].&[M], [DimProduct].[Вид].&[W],

[DimProduct].[Вид].&[U]}) ON COLUMNS,

({[DimProduct].[Цвет].&[Black]},

{[DimCustomer].[Пол].&[F], [DimCustomer].[Пол].&[M]}) ON ROWS
```

FROM [Продажи]
WHERE [Dif]

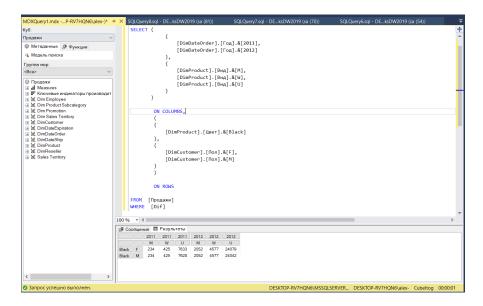


Рисунок 26 - Результат МDХ-запроса №2

3 Запрос. Количество товара, проданного через интернет за 2011 год по кварталам с учетом географии. Код данного MDX-запроса представлен ниже, а его результат приведен на рисунке – 27.

```
SELECT {[Dim Sales Territory].[Sales Territory Group].&[Pacific],

[Dim Sales Territory].[Sales Territory Group].&[Europe],

[Dim Sales Territory].[Sales Territory Group].&[North America]} ON 0,

([DimDateOrder].[Γομ].&[2011],

{[DimDateOrder].[Kвартал].&[1], [DimDateOrder].[Kвартал].&[2],

[DimDateOrder].[Kвартал].&[3], [DimDateOrder].[Kвартал].&[4]}) ON 1
```

FROM [Продажи]

WHERE [Measures].[Order Quantity - Fact Internet Sales]

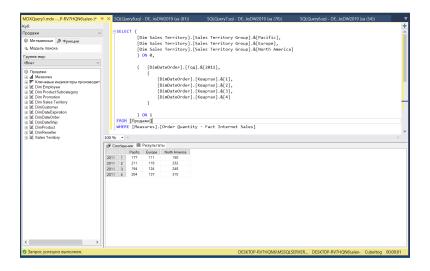


Рисунок 27 - Результат MDX-запроса
 №3

4 Запрос. Разность в объеме продаж через интернет и офлайн в семьях с детьми и без в соответствии с уровнем дохода, семейным положением во Франции в 2011 году. Семьи с высоким достатком не закупались, при этом объем продаж онлайн оказалось больше, чем офлайн. Код данного MDX-запроса представлен ниже, а его результат приведен на рисунке – 28.

```
SELECT ({[DimCustomer].[Годовой доход].&[1.E4],[DimCustomer].[Годовой доход].&[4.E4],

[DimCustomer].[Годовой доход].&[8.E4], [DimCustomer].[Годовой доход].&[1.7E5]},

{[DimCustomer].[Семейное положение].&[M],[DimCustomer].[Семейное положение].&[S]
}) ON O,

{[DimCustomer].[Количество детей].&[O], [DimCustomer].[Количество детей].&[1],
```

[DimCustomer].[Количество детей].&[2], [DimCustomer].[Количество детей].&[3]} ON 1

FROM [Продажи]

WHERE ([Dif], [Dim Sales Territory].[Sales Territory Country].&[France],
[DimDateShip].[Год].&[2011])

WITH MEMBER [Dif] AS ([Measures].[Sales Amount]

- [Measures].[Sales Amount - Fact Internet Sales])

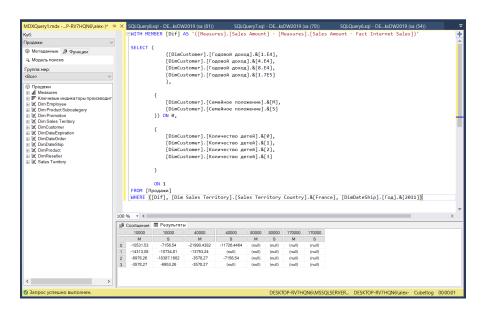


Рисунок 28 - Результат МDХ-запроса №4

5 Запрос. Сумма, потраченная на товар вида U (черного цвета) в Германии в период с 2011 - 2014 год. Код данного MDX-запроса представлен ниже, а его результат приведен на рисунке – 29.

SELECT non empty ([DimDateOrder].[Год].[Год], [DimDateOrder].[Квартал].[Квартал]) ON O,

```
[DimCustomer].[Владелец дома].[Владелец дома] *
[Measures].[Unit Price - Fact Internet Sales] on 1
```

FROM [Продажи]

WHERE ([Dim Sales Territory].[Sales Territory Country].&[Germany],

[DimProduct].[Вид].&[U], [DimProduct].[Цвет].&[Black])

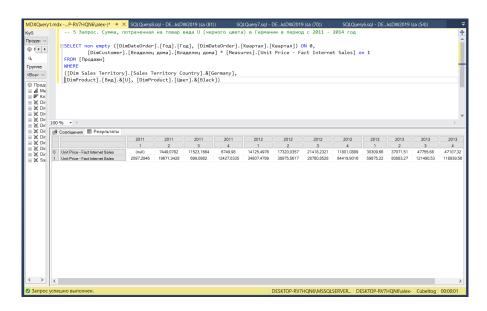


Рисунок 29 - Результат MDX-запроса
 №5

6 Запрос. Сравнение сумм продаж между Францией и Австралией по цвету и размеру товара. Код данного MDX-запроса представлен ниже, а его результат приведен на рисунке – 30.

```
WITH MEMBER [dif] AS ([Dim Sales Territory].[Sales Territory Country].&[France],
[Measures].[Sales Amount]) - ([Dim Sales Territory].[Sales Territory Country].&

[Australia], [Measures].[Sales Amount])

SELECT non empty [DimProduct].[Цвет].[Цвет] on 0,

{[DimProduct].[Размер товара].&[L], [DimProduct].[Размер товара].&[M],

[DimProduct].[Размер товара].&[S], [DimProduct].[Размер товара].&[XL]} on 1
```

FROM [Продажи] WHERE [dif]

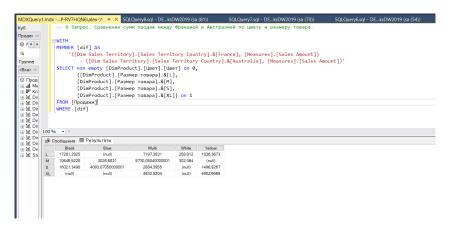


Рисунок 30 - Результат МDХ-запроса №6

7 Запрос. Объем продаж в США в период с 2010-2011 год, с группировкой по уровню образования и продуктовой линейке. Код данного MDX-запроса представлен ниже, а его результат приведен на рисунке – 31.

```
SELECT ([Measures].[Sales Amount - Fact Internet Sales],
```

[Dim Sales Territory].[Sales Territory Country].&[United States]) ON COLUMNS, NonEmptyCrossJoin({[DimProduct].[Продуктовая линейка].[Продуктовая линейка]}, {[DimCustomer].[Уровень образования].[Уровень образования]}) ON ROWS

FROM [Продажи]

WHERE {[DimDateOrder].[Год].&[2010],

[DimDateOrder].[Год].&[2011]}

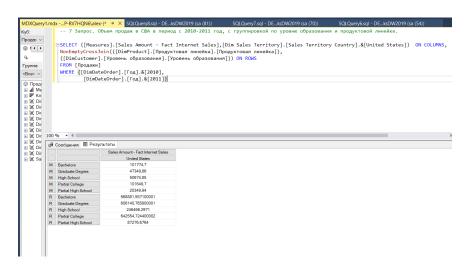


Рисунок 31 - Результат МDХ-запроса №7

8 Запрос. Инофрмация о товарах, для которых сумма продаж по каждой из стран превышает 2000000 долларов. Код данного MDX-запроса представлен ниже, а его результат приведен на рисунке – 32.

```
SELECT [Measures]. [Sales Amount - Fact Internet Sales] ON COLUMNS,
Filter(CrossJoin({[DimProduct].[Продуктовая линейка].[Продуктовая линейка]},
{[Dim Sales Territory].[Sales Territory Country].[Sales Territory Country]}),
[Sales Amount - Fact Internet Sales] > 2000000) ON ROWS
```

FROM [Продажи]

Рисунок 32 - Результат МDХ-запроса №8

9 Запрос. Количество денег, которые потратили француженки на красные велосипеды в каждом месяце 2012 года. Код данного MDX-запроса представлен ниже, а его результат приведен на рисунке -33.

```
SELECT ([DimCustomer].[Пол].&[F],

[Dim Sales Territory].[Sales Territory Country].&[France],

[DimDateOrder].[Год].&[2012], [DimProduct].[Цвет].&[Red],

[Dim Product Subcategory].[English Product Category Name].&[Bikes]

) ON COLUMNS,

{[DimDateOrder].[Mecяц года].&[January],[DimDateOrder].[Mecяц года].&[February],

[DimDateOrder].[Mecяц года].&[March], [DimDateOrder].[Mecяц года].&[April],

[DimDateOrder].[Mecяц года].&[May], [DimDateOrder].[Mecяц года].&[June],

[DimDateOrder].[Mecяц года].&[July], [DimDateOrder].[Mecяц года].&[August],

[DimDateOrder].[Mecяц года].&[September], [DimDateOrder].[Mecяц года].&[October],

[DimDateOrder].[Mecяц года].&[November], [DimDateOrder].[Mecяц года].&[December]

} ON ROWS
```

FROM [Продажи]

```
MDXQuey1.mdx — P-RV7HQN6jalex)* © X SQQuey8xq| -DE_xsDW2019 (as (61))

KyG

-9 Sanpoc. Количество денег, которые потратили француженки на красные велосиледы в каком меслце 2012 года

9 SELECT (
[DimCustomer].[floa], &[F],
[Dim Sales Territory], [Sales Territory Country]. &[France],
[DimDateOder], [Ica], &[2012],
[DimProduct]. [Iliner], &[Red],
[Dim Broduct]. [Iliner], &[Red],
[Dim Broduct]. [Iliner], &[Red],
[DimDateOder], [Mecnu rona], &[February],
[DimDateOder], [Mecnu rona], &[March],
[DimDateOder], [Mecnu rona], &[
```

Рисунок 33 - Результат MDX-запроса №9

10 Запрос. Общее количество товаров онлайн и офлайн, а так же средняя выручка с товара. Код данного MDX-запроса представлен ниже, а его результат приведен на рисунке – 34.

```
WITH MEMBER [Все количество]

AS [Measures].[Order Quantity - Fact Internet Sales] + [Measures].[Order Quantity]

MEMBER [Средняя выручкастовара]

AS ([Measures].[Sales Amount - Fact Internet Sales] +

[Measures].[Sales Amount]) / [Все количество]
```

SELECT

```
non empty [DimDateOrder].[Год].[Год] ON COLUMNS,
{[Measures].[Order Quantity - Fact Internet Sales], [Measures].[Order Quantity],
[Все количество], [Средняя выручкастовара]} ON ROWS
```

FROM [Продажи]

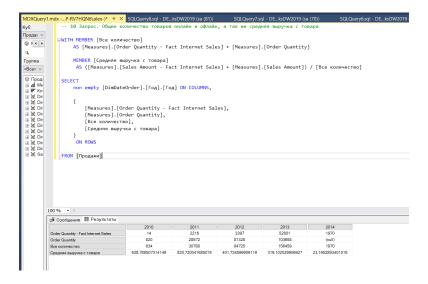


Рисунок 34 - Результат МDХ-запроса №10

6 Анализ результатов MDX-запросов

В процессе работы с развернутым OLAP-кубом, были отмечены следующие основные особенности. На каждой из осей (columns, rows и др.) можно располагать несколько измерений. Кортежи (множества), указанные на осях и после WHERE, образуют декартово произведение, которое и является результатом выполнения запроса. МDX-запросы позволяют осуществлять навигацию по многомерному пространству и определенным над ними иерархиями, что значительно упрощает написание запросов. Функция Members применяется к иерархиям или к уровням. Функция AllMembers работает аналогично функции Members, но Members возвращает все элементы иерархии, кроме вычисляемых, а AllMembers возвращает также и вычисляемые элементы.

MDX-запросы дают возможность получения данных, находящихся в многомерной структуре. Наличие многомерности в представленной OLAP-кубом информации дает возможность формировать запросы на основе реальных бизнес задач. Для построения запросов используются основные инструменты:

- 1. Мера, или факт, или показатель (Measure) параметры предметной области, служащие предметом анализа;
- Измерение (Dimension) понятия предметной области, на основании которых анализируются меры. Каждое измерение может иметь иерархическую структуру уровней.
 В частности, измерение Время может включать уровни Год, Квартал, Месяц, Число, а измерение География может включать уровни Весь мир, Страна, Город;

3. Член (Member) – единица описания, на каждом уровне иерархии.

Все рассчеты поверх срезов осуществляются в онструкции WITH, которая может выступать как в качестве промежуточного расчета значений, так и в качестве фильтрации или агрегирования промежуточных результатов.

В целом, MDX-запросы, которые используются для извлечения информации из баз данных Analysis Services, некоторым образом схожи с SQL, но это сходство выглядит уже не таким явным при более детальном изучении синтаксиса. При этом язык запросов MDX расширяет границы анализа и дает возможности для анализа многомерных данных.

Список литературы

- [1] Мамедли Р.Э. Базы данных: лабораторный практикум. Нижневартовск: Нижневартовский государственный университет, 2021. 160 с.
- [2] Сивакумар Харинатх, Стивен Куинн SQL Server Analysis Services 2005 и MDX для профессионалов, 2005. 822 с.
- [3] Bryan C.Smith Microsoft SQL Server 2008 MDX Step by Step, 2009, 400 c.
- [4] Алексей Федоров, Наталия Елманова Введение в ОLAP, 98 с.
- [5] Барсегян А.А., Куприянов М.С., Степаненко В.В., Холод И.И. Методы и модели анализа данных: OLAP и Data Mining. СПб.: БХВ-Петербург, 2004. 336 с.