

БАЗЫ ДАННЫХ

4. Бизнес-анализ в стиле Microsoft Analysis Services



Основные определения (1)

Business analysis (*бизнес-анализ*) – выявление деловых потребностей и нахождение решений деловых проблем. Решения часто заключаются в разработке систем управления, но могут также состоять из усовершенствования процессов (методик работы), организационных изменений, стратегического планирования и разработки управленческих политик.

Business Intelligence (*система бизнес-анализа*) – совокупность технологий, программного обеспечения и способов извлечения из данных значимой информации, которые направлены на достижение целей бизнеса путём наилучшего использования имеющихся ресурсов.

Data Mining (*добыча данных, глубинный анализ данных*) — собирательное название, используемое для обозначения совокупности методов обнаружения в данных ранее неизвестных, нетривиальных, практически полезных и доступных для интерпретации знаний, необходимых для принятия решений в различных сферах человеческой деятельности.

Data Mining концепция



Построение деревьев решений, алгоритмы искусственных нейронных сетей, генетические алгоритмы, алгоритмы эволюционного программирования, ассоциативной памяти, нечёткой логики.

Основные определения (2)

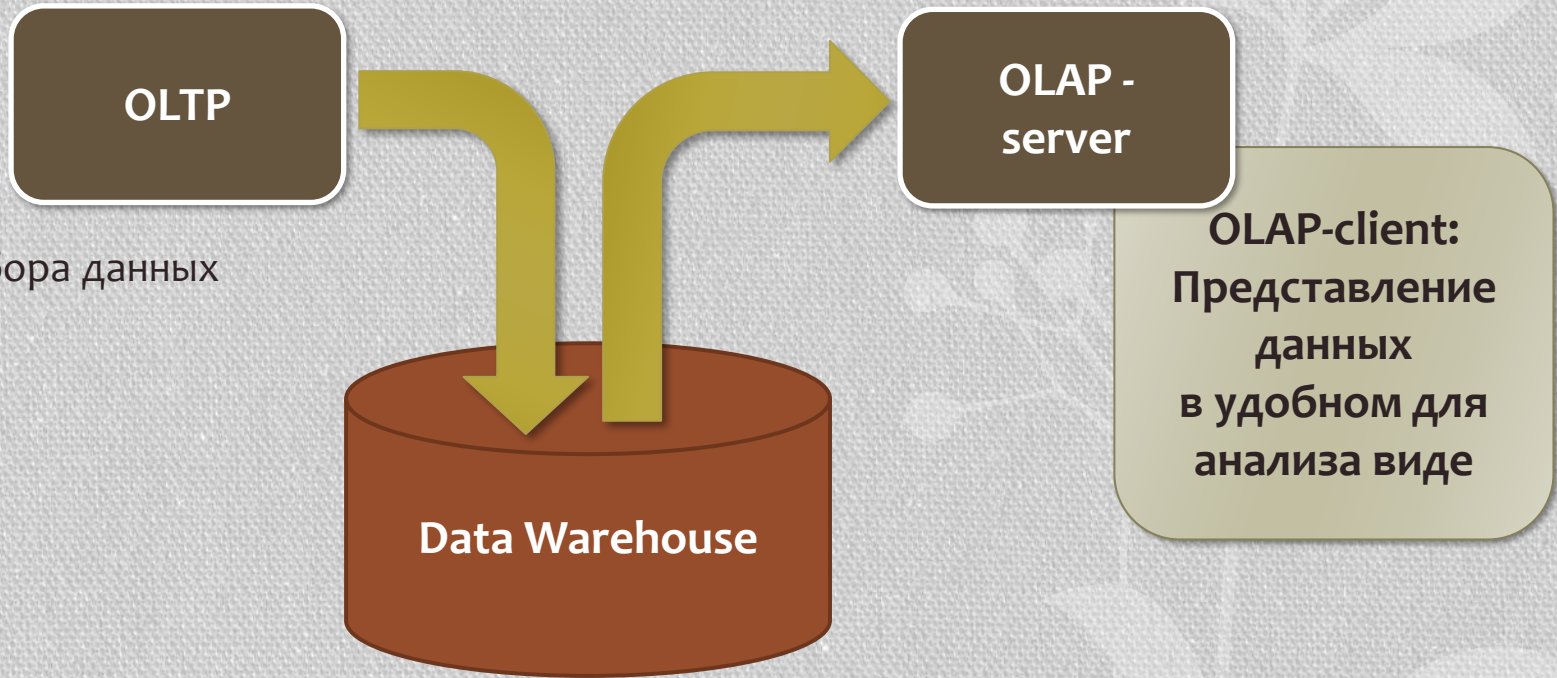
OLAP (*online analytical processing*, аналитическая обработка в реальном времени) – технология обработки данных, заключающаяся в подготовке суммарной (агрегированной) информации на основе больших массивов данных, структурированных по многомерному принципу.

OLTP (*Online Transaction Processing*) — обработка транзакций в реальном времени. Способ организации БД, при котором система работает с небольшими по размерам транзакциями, но идущими большим потоком, и при этом клиенту требуется от системы минимальное время отклика. Объёмы данных, обрабатываемые за один день OLTP-системой, могут достигать нескольких гигабайт в день.

Data Warehouse (хранилище данных) – предметно-ориентированная информационная база данных, специально разработанная и предназначенная для подготовки отчётов и бизнес-анализа с целью поддержки принятия решений в бизнесе. Строится на базе систем управления базами данных и систем поддержки принятия решений.

Data Mining схема

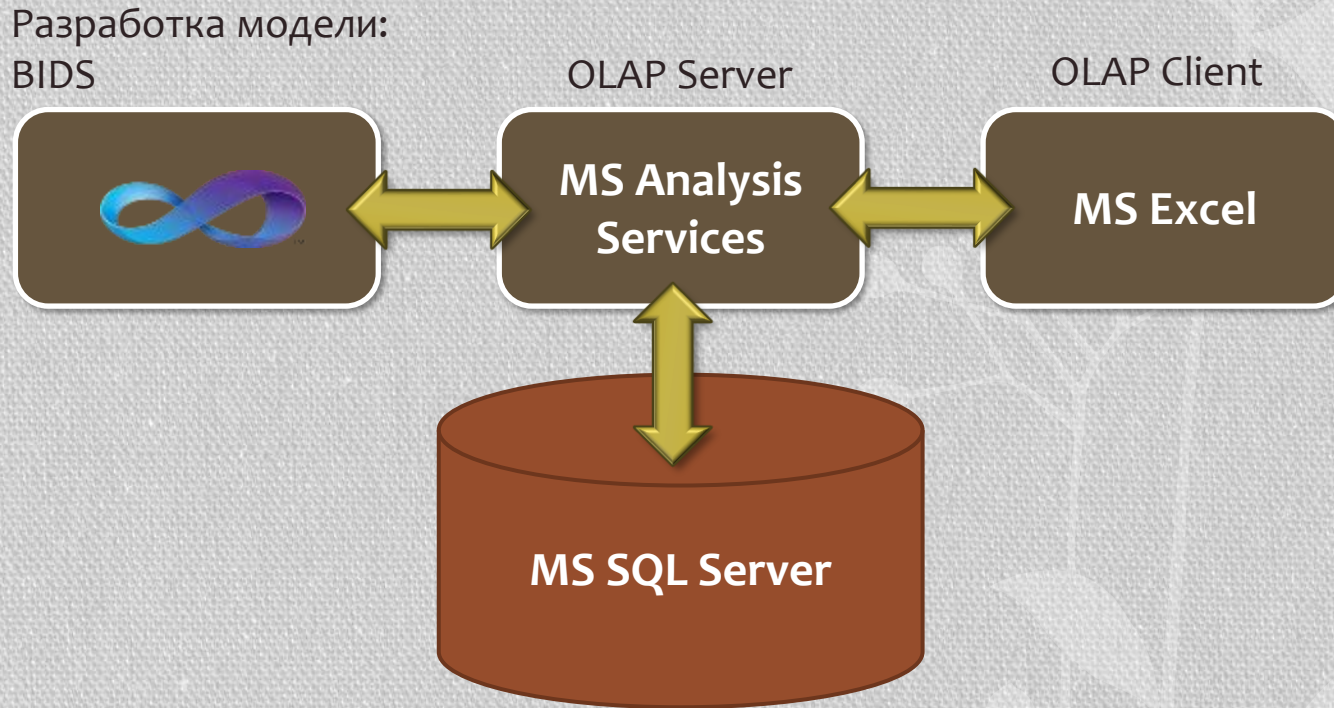
многомерная OLAP (Multidimensional — MOLAP);
реляционная OLAP (Relational — ROLAP);
гибридная OLAP (Hybrid — HOLAP).



Системы сбора данных

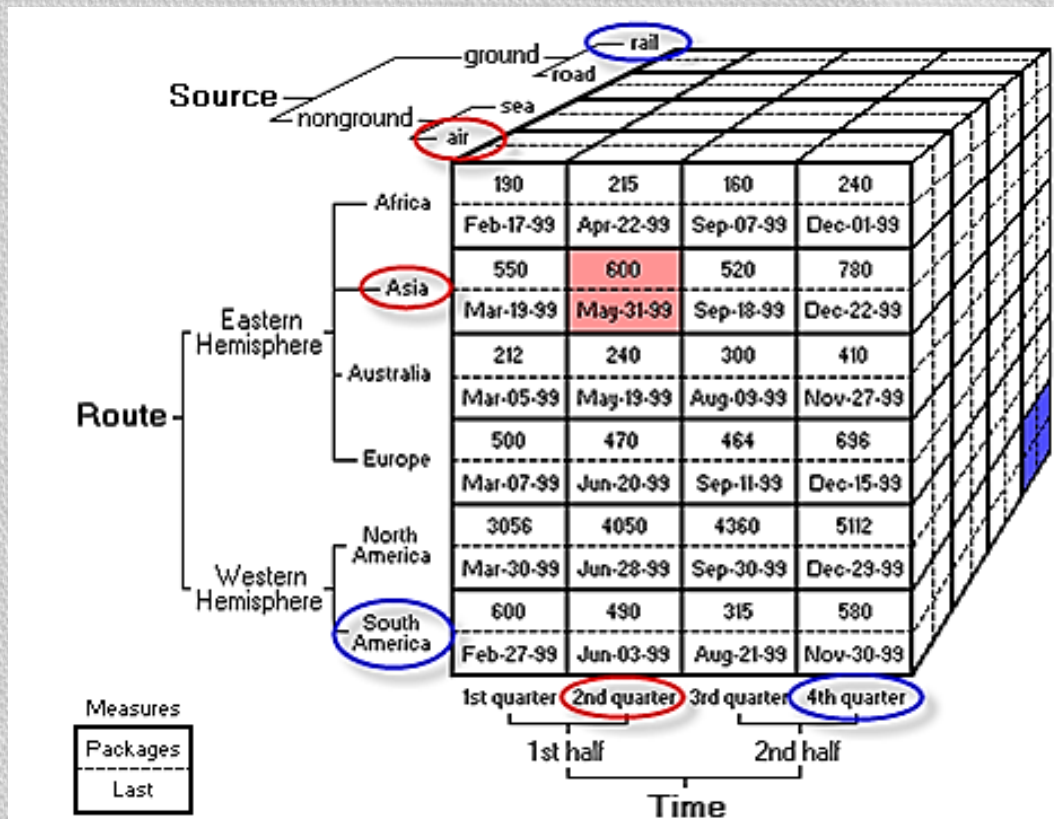
Главное свойство **хранилищ данных**: данные в хранилище *не создаются*: т.е. данные поступают из внешних источников, **не корректируются и не удаляются**. Данные постфактум привязываются к меткам времени, для которых они были актуальными.

Analysis Services – общие сведения (1)



Главная идея – многомерное представление данных для анализа.

Analysis Services – общие сведения (2)



В Analysis Services главным объектом, используемым в многомерном анализе, является **куб**, хотя геометрический аналог здесь мало подходит, поскольку количество единиц информации вдоль каждого из измерений в общем случае различно, а число измерений обычно гораздо больше трёх.

Терминология гиперкуба (1)

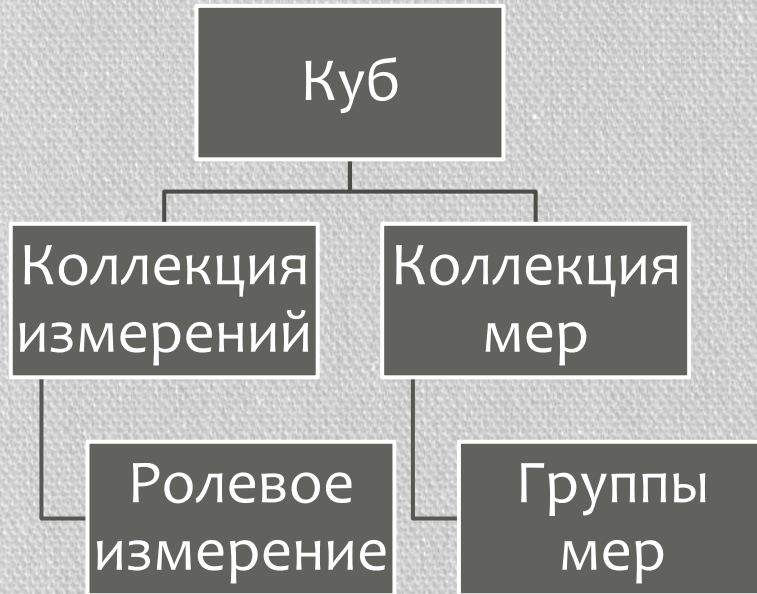
- ❑ *Измерение* (Dimension) – описывает категорию данных, по которой производится анализ. Например, «время».
- ❑ *Иерархия измерений* (Dimension Hierarchy) включает добавление зависимых атрибутов в измерение. Например, для измерения «время» – годы, кварталы и месяцы (при наличии ключевого атрибута *день*).
- ❑ *Элемент* (Member) – именованная точка на измерении. Например, «понедельник».
- ❑ *Значение элемента* (Member Value) – уникальная характеристика элемента. Например, конкретная дата.
- ❑ *Атрибут* (Attribute) – полная коллекция элементов одного типа. Например, перечисление всех дней недели.
- ❑ *Размер измерения* (Size) – количество элементов, которое оно содержит. Например, время, состоящее из дней недели, имеет размерность 7.
- ❑ *Кортеж* (Tuple) – координата в многомерном пространстве.
- ❑ *Срез* (Slice) – секция многомерного пространства, которая может быть определена кортежем.

Терминология гиперкуба (2)



На рисунке показано «пространство продаж» с точкой одной продажи. Срез (*,*,[Март]) определяет подпространство (плоскость) всех продаж в указанном месяце. Значением точки может быть фактическая сумма продажи.

Терминология гиперкуба (3)



Ролевое измерение используется для создания нескольких измерений куба, которые не зависят друг от друга.

Меры определяют, какая информация доступна для анализа, в какой форме, по каким правилам данные будут обработаны и в каком формате представлены пользователю.

Меры могут быть объединены в группы по их отношению к различным измерениям в кубе

Описание модели данных

Для определения и изменения многомерной модели используется объектно-ориентированный язык описания данных (DDL – Data Definition Language). В основе синтаксиса языка лежит XML (eXtended Markup Language).

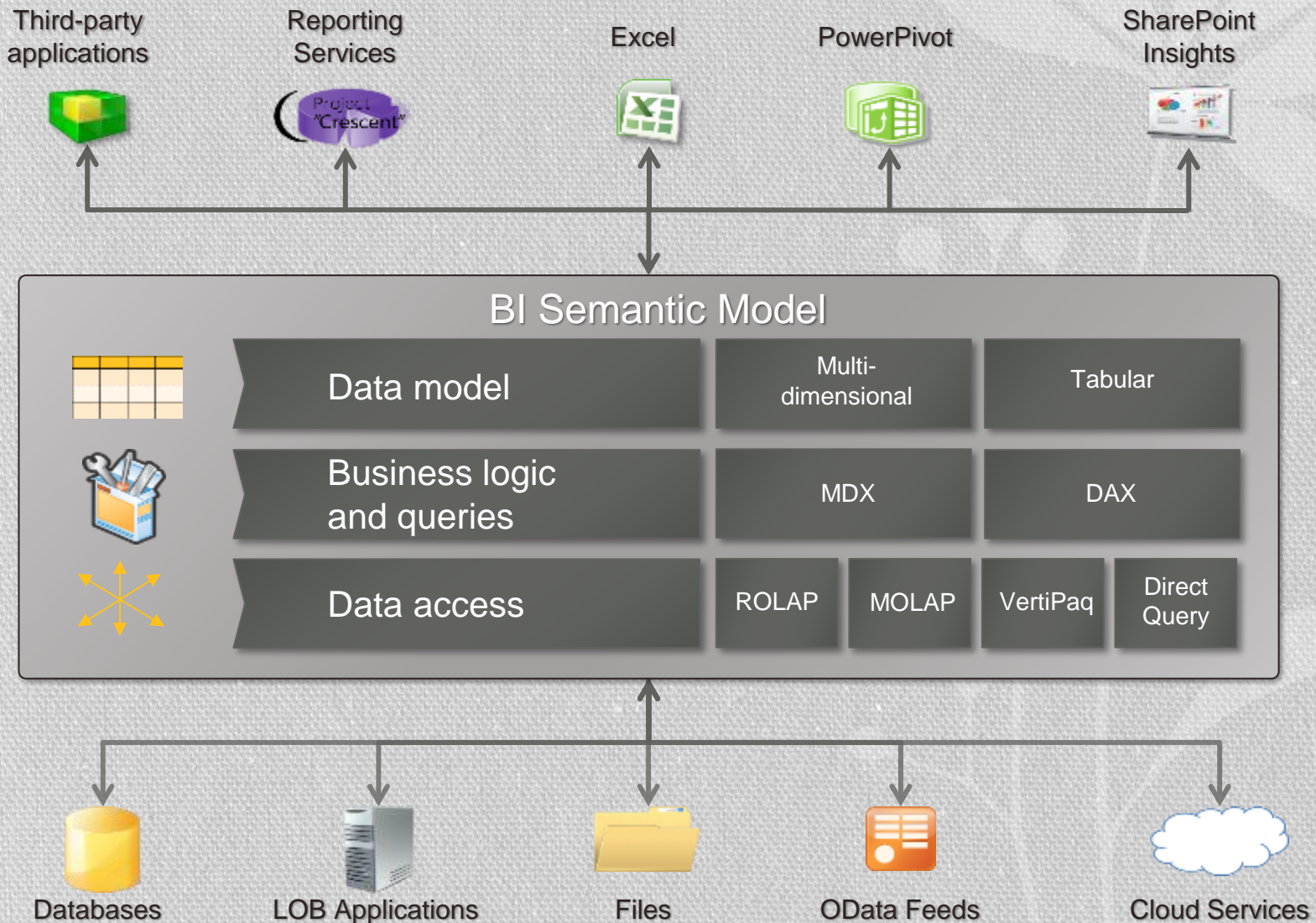
Основными объектами являются объекты, которыми пользователь может манипулировать независимо от их родительских объектов при создании и изменении модели.

Вспомогательными объектами являются дочерние объекты основных объектов, которые можно определять и редактировать только в контексте их основных объектов.

Источник данных — это объект, определяющий способ подключения Analysis Services к реляционной базе данных или к другим источникам данных.

Создание измерения начинается с того, что определяются таблицы из реляционной базы данных, на которых будет построено измерение. Однако основным критерием для построения модели должно быть представление многомерной модели для конечного пользователя, а не структура данных в реляционном источнике данных.

Модель использования MS Analysis Services



Характеристики MS Analysis Services

Items	Limits
Databases per server	Unlimited
Cubes per database	Unlimited
Cubes per virtual cube	64
Dimensions per cube	128
Measures per cube	1,024
Calculated members per cube	65,535
Levels per dimension	64
Partitions per cube	Unlimited

Пример подготовки данных для BI (MS AS)

Проект "Анализ работы компании Adventure Works Cycles с использованием развёрток куба данных".

Тестовая база данных AdventureWorksDW2012

Адреса
Клиенты
Продажи
Даты
Товары

Свойства	Adventure Works DW2012 DataSourceView
Data Source	Adventure Works DW2012
Description	
Name	Adventure Works DW2012
NameMatchingCriteria	(net)
Документально	
RetrieveRelationships	true
SchemaRestriction	

Работа с MS AS (подготовка данных)

Создание измерений

Мастер измерений

Выбор атрибутов измерения

Щелкните атрибуты измерения, а затем выберите "Разрешить обзор", чтобы разместить их в виде иерархий.

Доступные атрибуты:

<input checked="" type="checkbox"/>	Имя атрибута	<input checked="" type="checkbox"/>	Разрешить об...	Тип атрибута
<input checked="" type="checkbox"/>	Date Key	<input checked="" type="checkbox"/>		обычное
<input checked="" type="checkbox"/>	Full Date Alternate Key	<input checked="" type="checkbox"/>		Дата
<input type="checkbox"/>	Day Number Of Week	<input type="checkbox"/>		обычное
<input type="checkbox"/>	English Day Name Of Week	<input type="checkbox"/>		обычное
<input type="checkbox"/>	Spanish Day Name Of Week	<input type="checkbox"/>		обычное
<input type="checkbox"/>	French Day Name Of Week	<input type="checkbox"/>		обычное
<input type="checkbox"/>	Day Number Of Month	<input type="checkbox"/>		обычное
<input type="checkbox"/>	Day Number Of Year	<input type="checkbox"/>		обычное
<input type="checkbox"/>	Week Number Of Year	<input type="checkbox"/>		обычное
<input checked="" type="checkbox"/>	English Month Name	<input checked="" type="checkbox"/>		Месяц
<input type="checkbox"/>	Spanish Month Name	<input type="checkbox"/>		обычное
<input type="checkbox"/>	French Month Name	<input type="checkbox"/>		обычное
<input type="checkbox"/>	Month Number Of Year	<input type="checkbox"/>		обычное
<input checked="" type="checkbox"/>	Calendar Quarter	<input checked="" type="checkbox"/>		Квартал
<input checked="" type="checkbox"/>	Calendar Year	<input checked="" type="checkbox"/>		Год
<input checked="" type="checkbox"/>	Calendar Semester	<input checked="" type="checkbox"/>		Полугодие
<input type="checkbox"/>	Fiscal Quarter	<input type="checkbox"/>		
<input type="checkbox"/>	Fiscal Year	<input type="checkbox"/>		
<input type="checkbox"/>	Fiscal Semester	<input type="checkbox"/>		

Календарь

- Дата
- День полугодия
- День месяца
- День квартала
- День декады
- День триместра
- День недели
- День года
- Дата
- Полугодие года
- Полугодие**
- Месяц полугодия

Обозреватель решений

Измерения

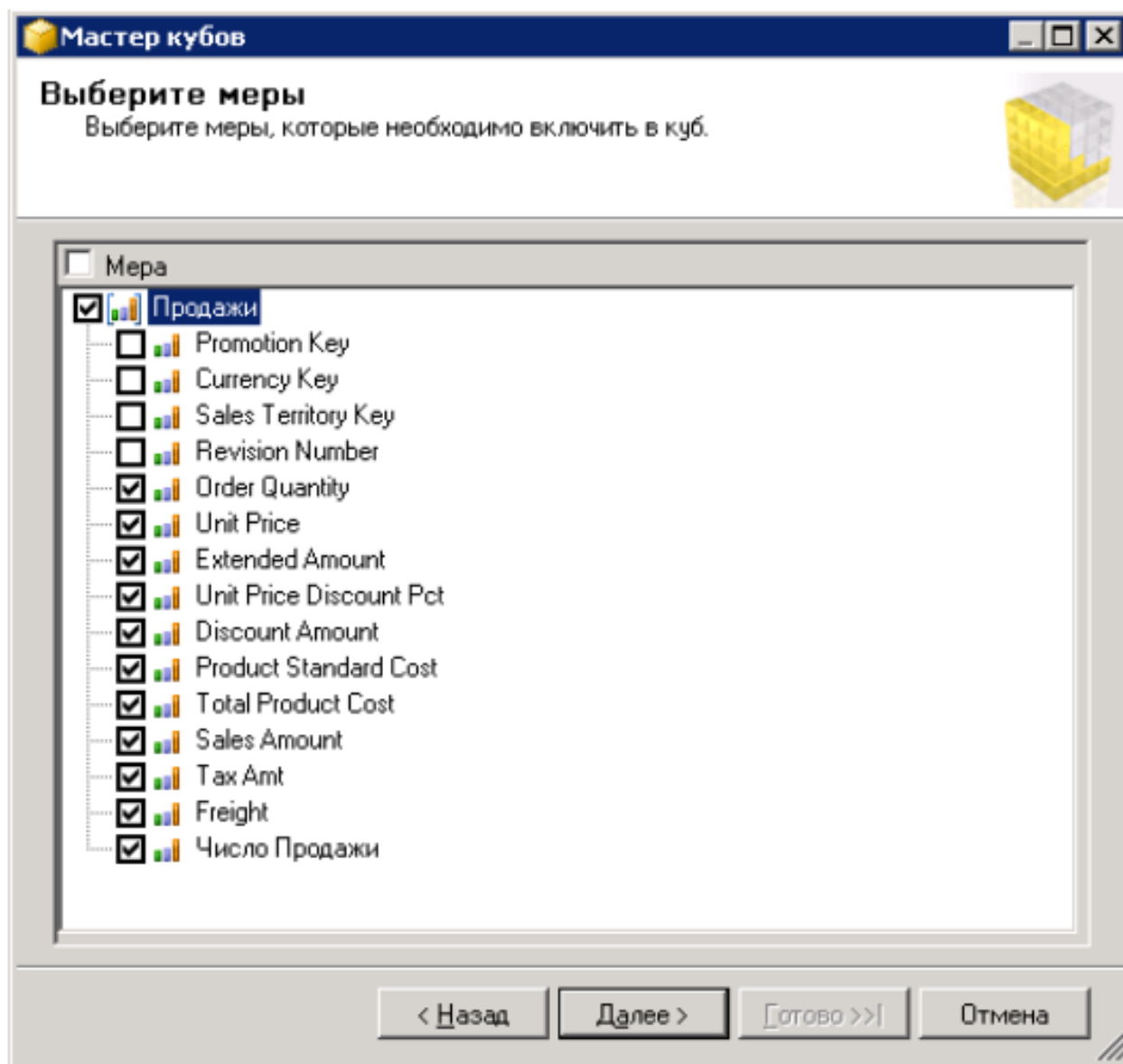
Расположение

Измерения

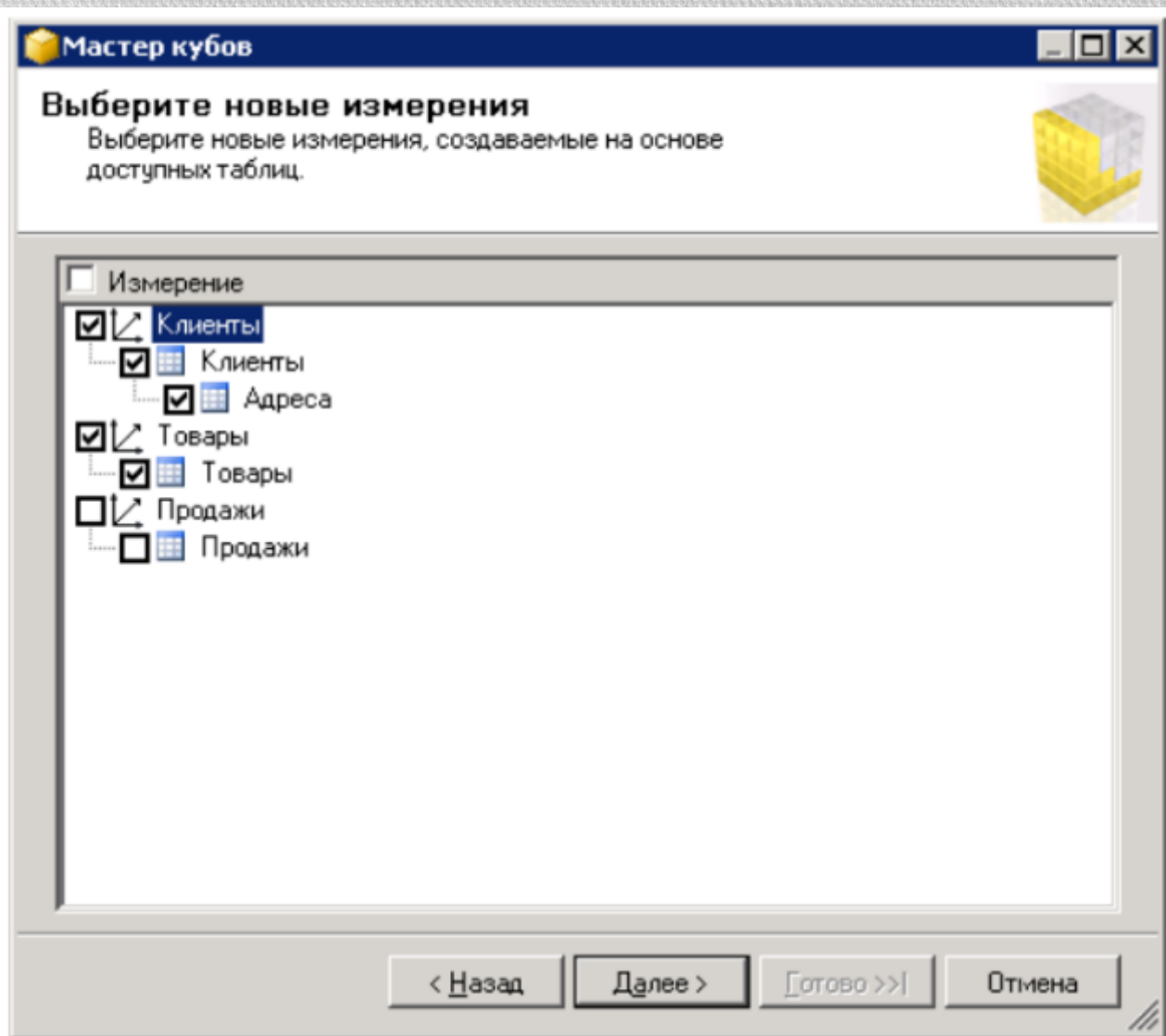
- Атрибуты
- Товары
 - Вес
 - Диапазон размеров
 - Класс
 - Конечная дата
 - Наименование линейки товаров
 - Наименование модели
 - Наименование товара
 - Начальная дата
 - Нормативная стоимость
 - Прейскурантная цена
 - Размер
 - Статус
 - Стиль
 - Точка заказа
 - Уровень резервного запаса
 - Цвет
 - Цена дилера
 - Число дней на изготовление

Работа с MS AS (подготовка данных)

Создание мер



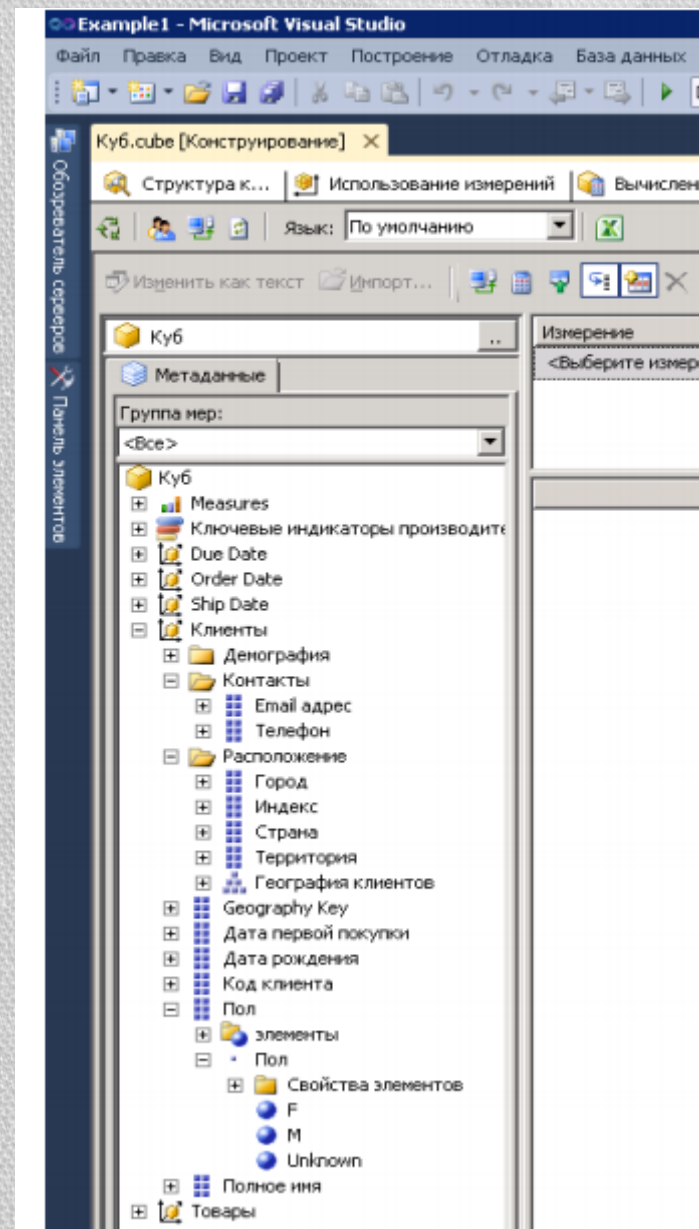
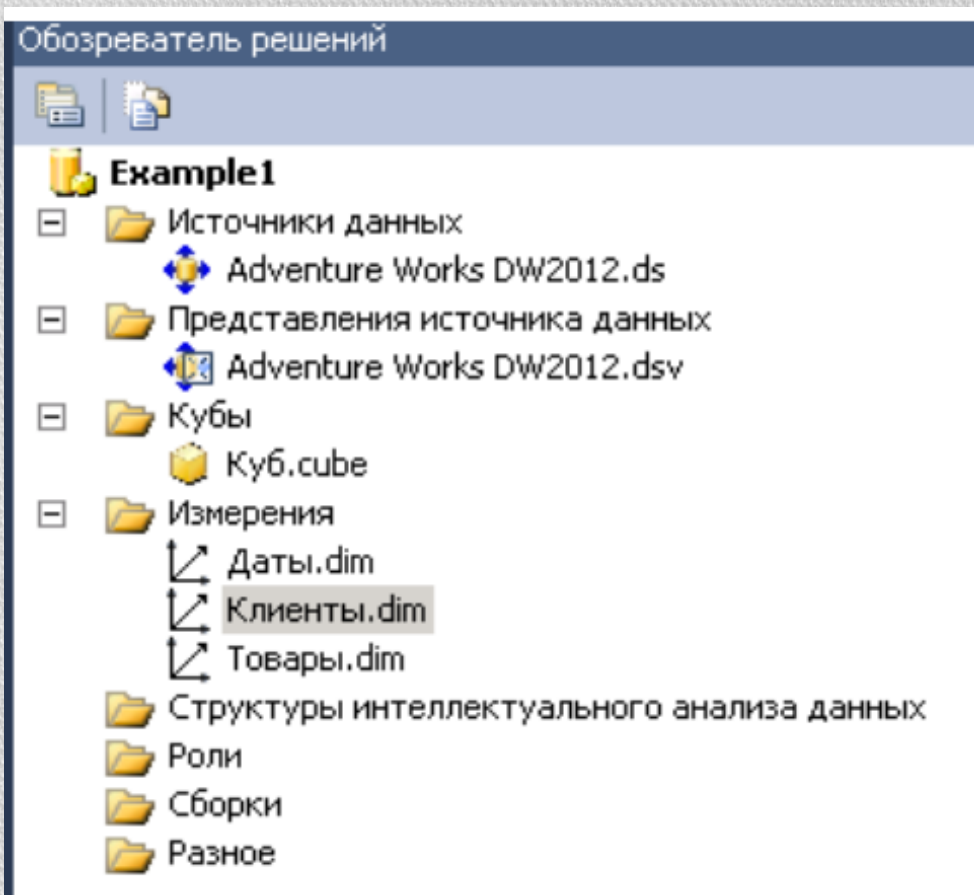
Работа с MS AS (подготовка куба)



Предыдущие шаги являлись подготовительным этапом для формирования представления данных для бизнес-анализа.

После создания куба следует выполнить команду "развёртывание".

Работа с MS AS (редактирование куба)



Работа с MS AS (сводная таблица)

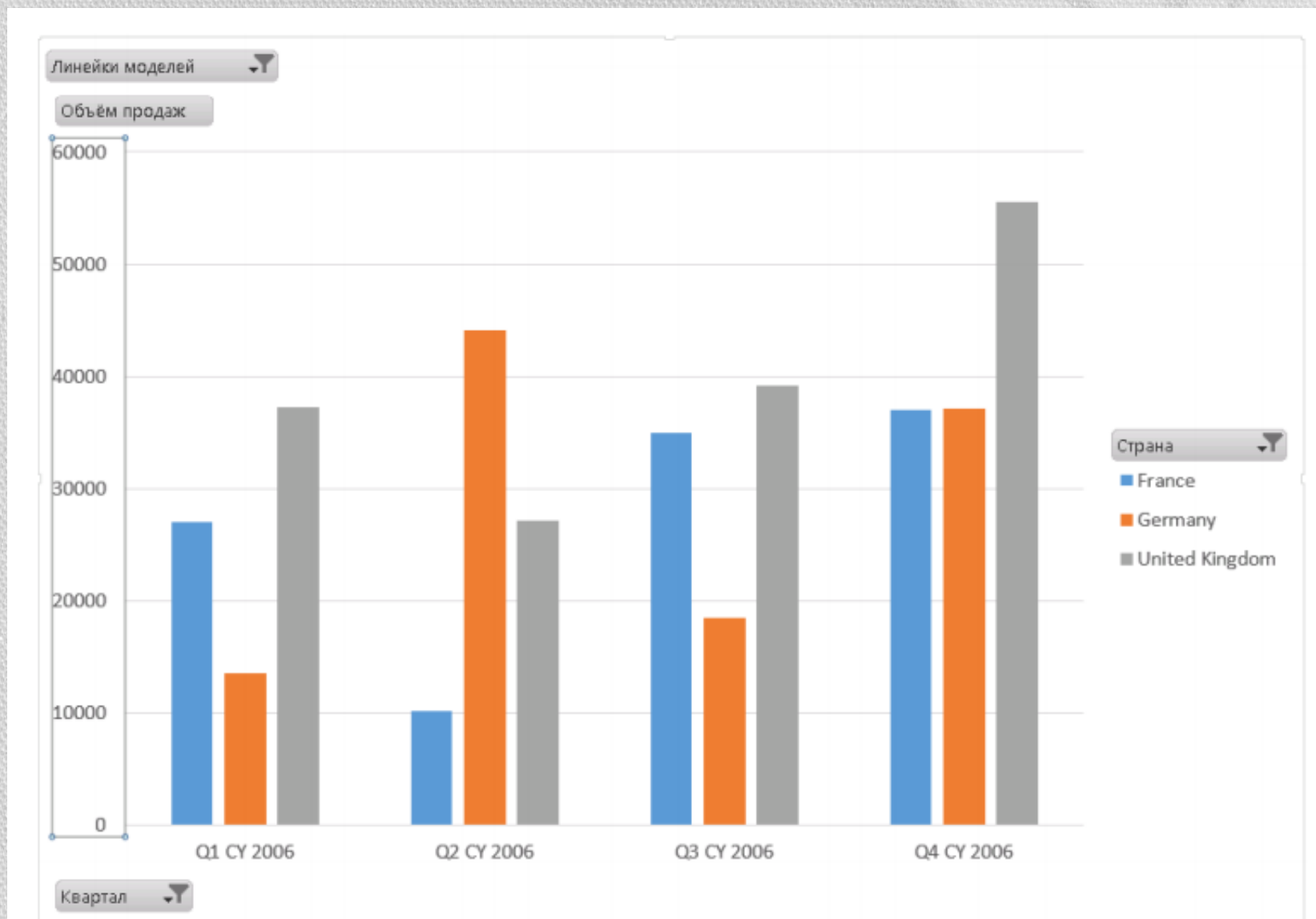
The screenshot displays the Microsoft Excel interface with a PivotTable. The PivotTable is structured as follows:

Названия строк	Road-150	Road-250	Road-550-W	Road-650	Road Итого	Общий итог	
Australia	956728,918	1849965,59	524360,3625	30013,125	102264,0858	2506603,163	3463332,081
Canada	99986,257	504536,07	134122,4625	8003,5	21783,9002	668445,9327	768432,1897
France	139680,7366	325622,57	174786,7875	18007,875	37415,736	555832,9685	695513,7051
Charente-Maritime	2071,4196	3578,27	2181,5625			5759,8325	7831,2521
Essonne	16344,586	39360,97	20157,6375	2000,875	2880,2846	64399,7671	80744,3531
Garonne (Haute)		10734,81	4363,125			15097,935	15097,935
Hauts de Seine	18343,6842	25047,89	11169,6	3001,3125	4362,3728	43581,1753	61924,8595
Loir et Cher		3578,27	2443,35			6021,62	6021,62
Loiret	2049,0982	14313,08	7068,2625	1000,4375	699,0982	23080,8782	25129,9764
Moselle	5471,4096	7156,54	4886,7		2964,1764	15007,4164	20478,826
Nord	23066,8874	46517,51	29145,675	2000,875	4530,1564	82194,2164	105261,1038
Pas de Calais			4624,9125			4624,9125	4624,9125
Seine (Paris)	30705,2376	53674,05	33770,5875	2000,875	11073,7156	100519,2281	131224,4657
Seine et Marne	9566,9274	14313,08	6806,475		1398,1964	22517,7514	32084,6788
Seine Saint Denis	14996,3728	42939,24	29669,25	3001,3125	3747,1664	79356,9689	94353,3417
Somme	2049,0982	3578,27	4624,9125		1482,0882	9685,2707	11734,3689
Val de Marne	8846,3996		2181,5625			2181,5625	11027,9621
Val d'Oise		7156,54			699,0982	7855,6382	7855,6382
Yveline	6169,616	53674,05	11693,175	5002,1875	3579,3828	73948,7953	80118,4113
Germany	157287,851	390031,43	150527,8125	16007	45161,7442	601727,9867	759015,8377

The right sidebar shows the 'Поля сводной табл...' task pane with the following settings:

- Выберите поля для добавления в отчет: Order Date.Календарь, Другие поля
- Фильтры: Order Date.К...
- Колонны: Линейки мод...
- Строки: География.К...
- Значения: Объем продаж

Работа с MS AS (сводная диаграмма)



Доступ к данным

Для доступа к данным используется язык **MDX** (Multi-Dimensional eXpressions).

Как и в языке SQL, для выборки данных в MDX используется конструкция **SELECT** с разделами **FROM** и **WHERE**. Предложение **SELECT** определяет многомерное пространство, которое будет результатом запроса. Предложение **FROM** определяет источник данных, которым может быть либо название куба, содержащего данные, либо другой запрос. Предложение **WHERE** определяет правила, ограничивающие результаты запроса подпространством данных.

Процесс ограничения результатов называется **сечением** (slicing). В Analysis Services сечение может выполняться не только по одной плоскости, но и по более сложным фигурам.

Результат MDX-запроса представляет собой многомерное подпространство — **подкуб** (subcube), который может содержать множество измерений — **осей** (axes). Для указания осей в операторе **SELECT** используется предложение **ON**. Каждая ось задает координаты в многомерном пространстве, которые являются **элементами** (members), проецируемыми на ось.

Пример использования MDX (условие)

Пусть куб имеет три измерения:

- Магазины (указаны страны, где есть данные магазины),
- Товары (продукты, напитки, промтовары и др.)
- Время (по годам)

Пусть также сформировано измерение мер с тремя мерами:

- Продажи (Store Sales),
- Расходы (Store Cost) и
- Продано (Units Sold).

Провести анализ товаров в группах *Продукты* и *Напитки*, проданные за период времени с начала 1997 по конец 1998 г. во всех странах, в которых имеются магазины.

Для решения этой задачи можно спроецировать измерение *Время* (Time) и измерение *Товары* (Product) на столбцы итоговой таблицы, а измерение *Меры* - на строки таблицы.

Пример использования MDX (решение)

SELECT

**{([Drink],[1997]),([Drink],[1998]),([Food],[1997]),([Food],[1998])} ON COLUMNS,
{[Measures].[Store Sales],[Measures].[Store Cost],[Measures].[Unit Sales]} ON ROWS
FROM [Warehouse and Sales]**

	Drink	Drink	Food	Food
	1997	1998	1997	1998
Store Sales	48836.2100000001	93742.1600000003	409035.59	778135.800000001
Store Cost	19477.2346	37498.6690000001	163270.7235	311993.641900003
Unit Sales	24597	46954	191940	365923

aspire invent achieve

