Создание OLAP куба в MS SQL Server 2012

## Немного теории.

Куб - многомерный массив данных.

Факты – это фактические записи (records) о каком-то процессе, который мы хотим анализировать, например процесс изменения цены товара на бирже. Очень часто факты содержат какие-нибудь числовые данные, например, фактическое значение голоса или цены.

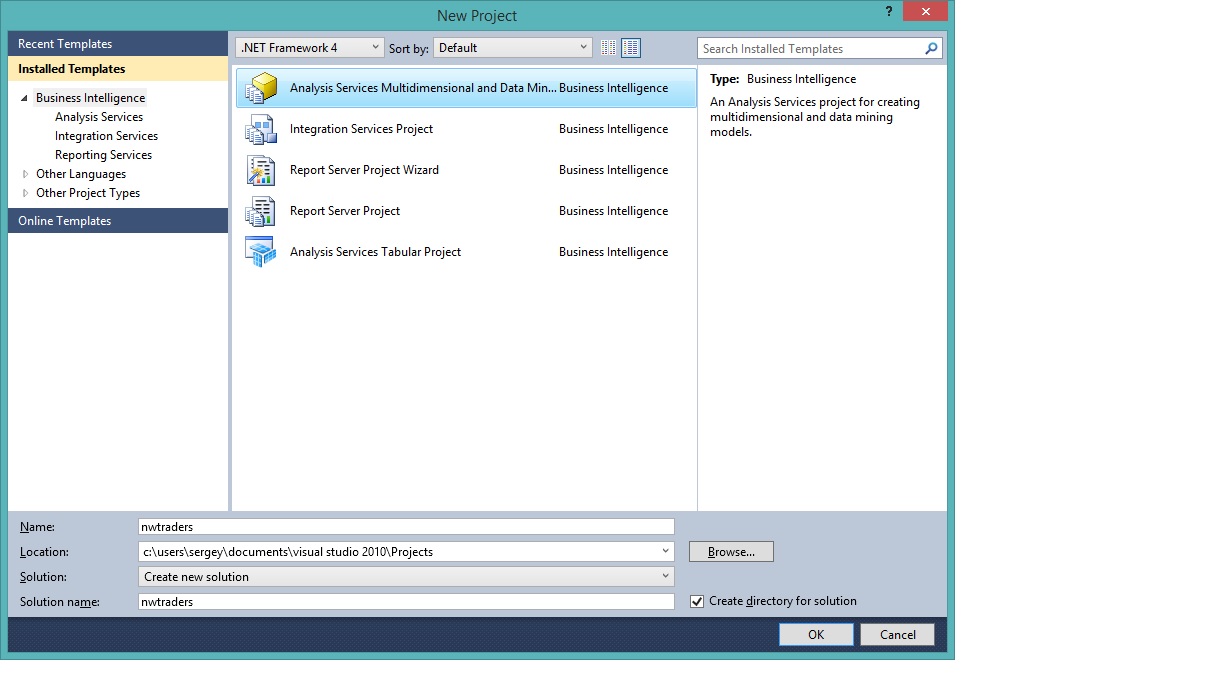
Измерения – это определяющие атрибуты фактов, и обычно отвечают на всякие вопросы: когда произошел факт, над чем или с чем именно, кто был объектом или субъектом и т.п. В основном, измерения имеют более описательный (то есть текстовый) характер, например, имя пользователя или название месяца, так как конечному пользователю будет намного легче воспринимать результаты, описанные текстом (например, название месяца), нежели цифрами (номер месяца в году).

Меры – данные, находящиеся на пересечениях осей - измерений, количественно характеризующие процесс. Это могут быть объемы продаж в штуках или в денежном выражении, остатки на складе, издержки и т. п. Пользователь, анализирующий информацию, может "разрезать" куб по разным направлениям, получать сводные (например, по годам) или, наоборот, детальные (по неделям) сведения и осуществлять прочие манипуляции, которые ему придут в голову в процессе анализа.

Для выполнения лабораторной будем использовать базу данный Northwind traders которая изначально шла в 2000 SQL сервере, но затем было заменена. Эта база данных удобна тем, что она имеет небольшой размер и довольно простую структуру. Скачать её можно с сайта Microsoft (<http://www.microsoft.com/en-us/download/details.aspx?id=23654>). Для развертывания в SQL Server 2012 нужно установить пакет, найти файл скрипта создания БД (instnwnd.sql) и запустить его в SQL Server 2012.

## Создание OLAP куба.

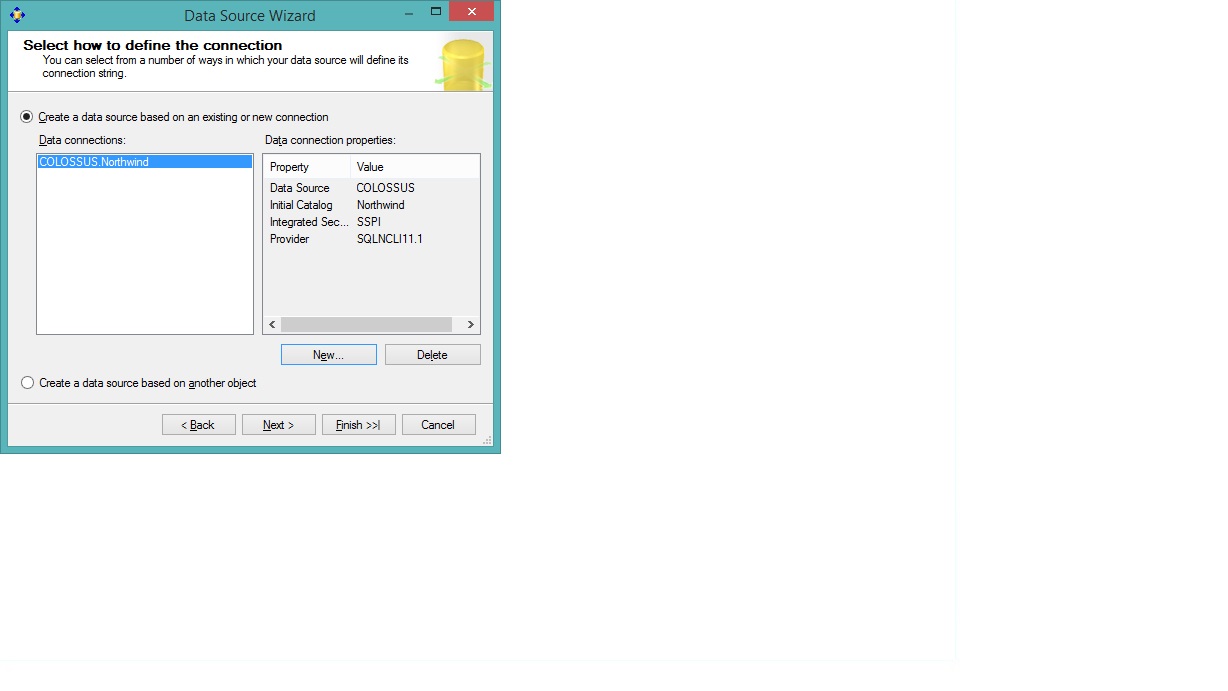
Для создания куба откройте Visual Studio 2010, которая входит в состав 2012 SQL сервера и создайте новый проект File -> New -> Project. В качестве шаблона выберете аналитические системы (Analysis Services Multidimensional and Data Mining Project). Назовите проект «nwtraders» и нажмите кнопку ОК.



**Рисунок 1.** Диалоговое окно создания проекта

Справа вы увидите окно **Solution Explorer**, здесь для создания проекта первым делом необходимо определить источник данных на основе которых будет строиться куб.

1. Правый щелчок мыши на Data Source -> New Data Source. В появившемся окне “Data Source Wizard”, нажмите “Next”
2. Выберете “Create a data source based on an existing or new connection”, нажмите кнопку “New…”
3. На экране появится диалоговое окно “Connection Manager”
   1. Из списка “Server name” выберете имя вашего сервера
   2. Выберете “Use SQL Server Authentication” и введите в поля User name и Password логин и пароль, используемые при аутентификации на SQL сервере. Установите галочку напротив “Save my password”
   3. Нажмите “Select or enter a database name” и из списка выберете базу данных “Northwind” и нажмите ОК
   4. Нажмите кнопку “Next”

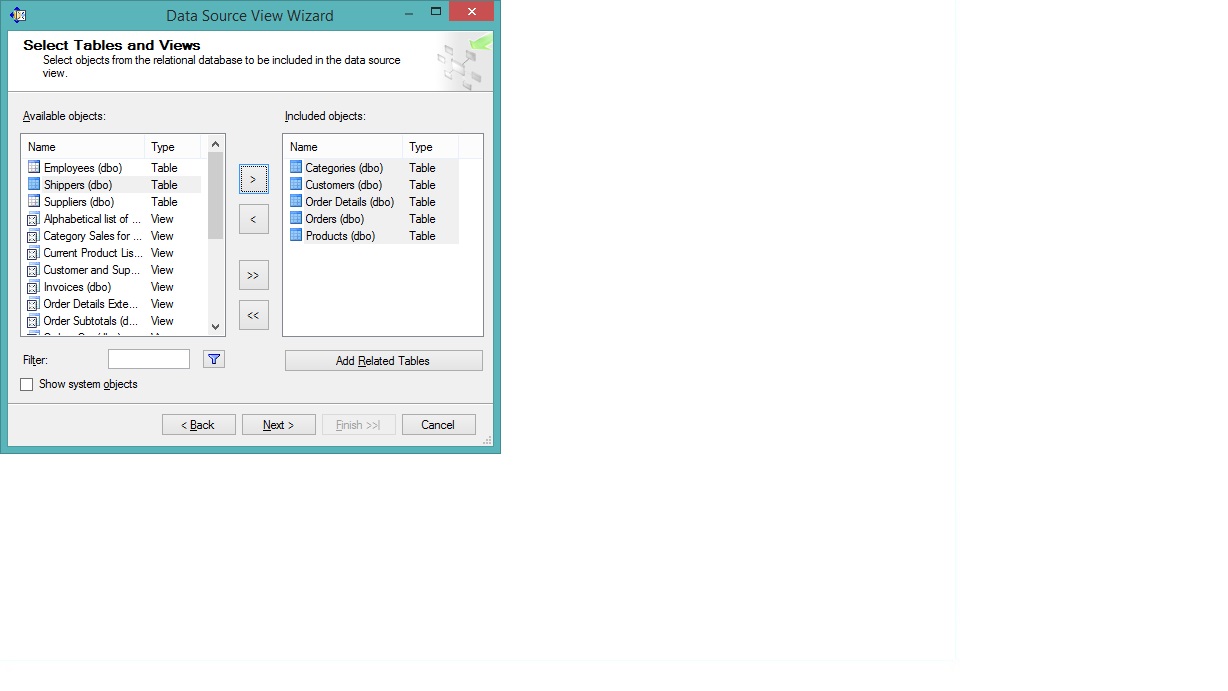


**Рисунок 2.** Определение источника данных

1. Выберете “Inherit” и нажмите кнопку “Next”
2. Источник данных подключен, осталось нажать кнопку “Finish”

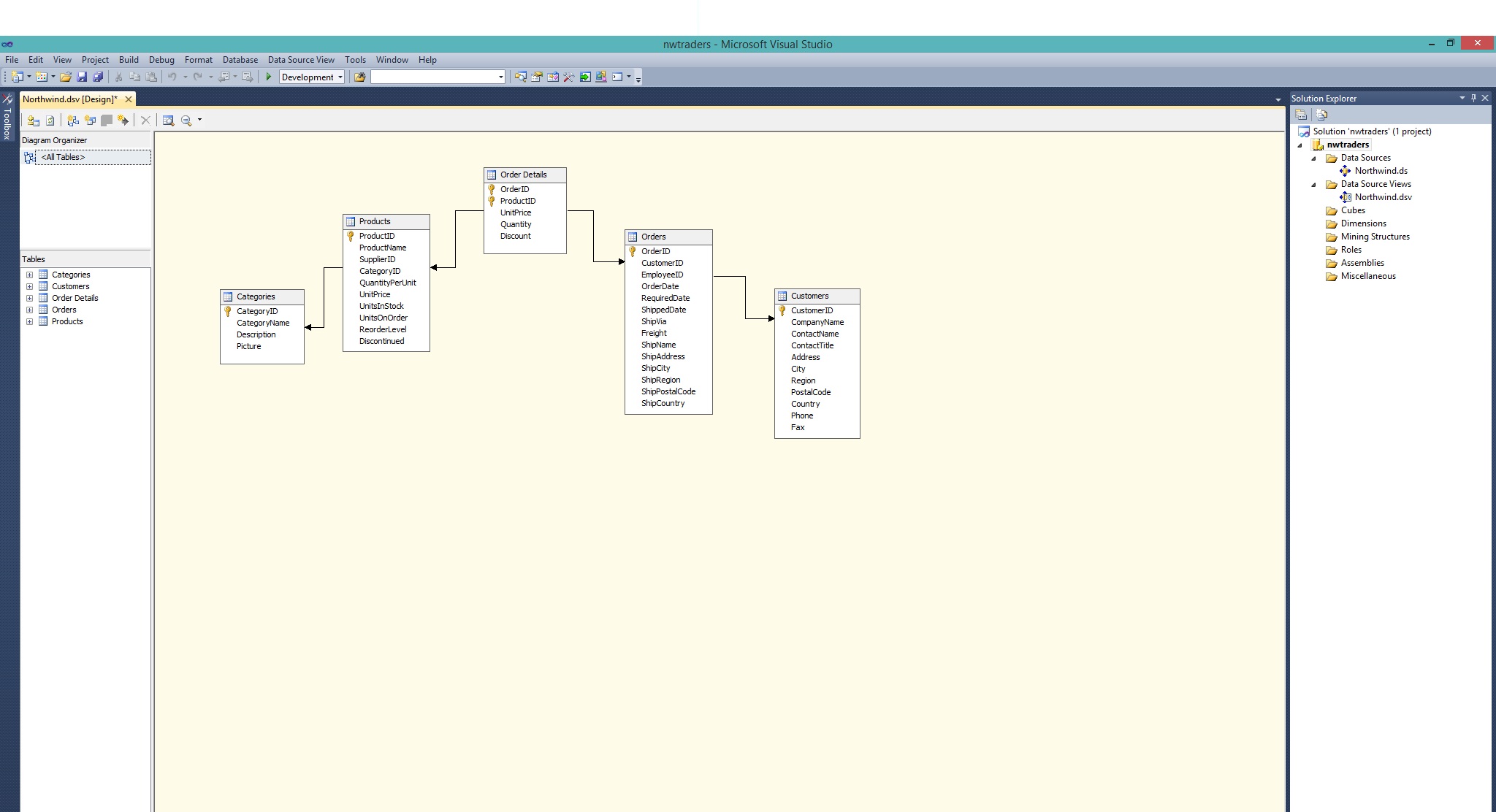
Следующее действие – это создание представления данных.

1. Правый щелчок мыши на Data Source Views -> New Data Source Views. В появившемся окне “Data Source Views Wizard”, нажмите “Next”
2. Укажите источник данных (в вашем случае это Northwind) и нажмите кнопку “Next”
3. Укажите следующие таблицы: **Categories** (таблица категорий товаров), **Customers** (таблица заказчиков), **Order** **Details** (таблица подробностей заказов), **Orders** (таблица заказов) и **Products** (таблица товаров), переместите их в правый список нажав кнопку “>” и нажмите кнопку “Next”



**Рисунок 3.** Выбор используемых таблиц

1. Название представления данных оставьте без изменений (Northwind) и нажмите кнопку “Finish”



**Рисунок 4.** Диаграмма содержимого таблиц

В результате на вашей диаграмме появится 5 таблиц(рисунок 4). Давайте посмотрим, что хранят эти таблицы. У вас есть некоторые Products (Товары), которые сгруппированы по **Categories** (Категориям). Есть **Orders** (Заказы), которые заказывают **Customers** (Заказчики). И для того, чтобы узнать какие товары в каких заказах содержатся есть таблица **Order** **Details** (Подробности заказов). Соответственно из таблиц **Categories** (Категорий) и **Products** (Товаров) вы сделаете **измерение** товаров. Из таблицы **Customers** (Заказчиков) вы сделаете **измерение** заказчиков. А таблицы **Order** **Details** (Подробности заказов) и **Orders** (Заказы) будет использоваться в качестве **фактов**. Фактами у вас будут факт продаж. У каждого факта есть атрибуты – это дата заказа, заказчик и товар, который продается заказчику.

Теперь вам понадобиться измерение. С его помощью вы сможете делать разрезы(срезы), выборку по определённым товарам или по определённым категориям.

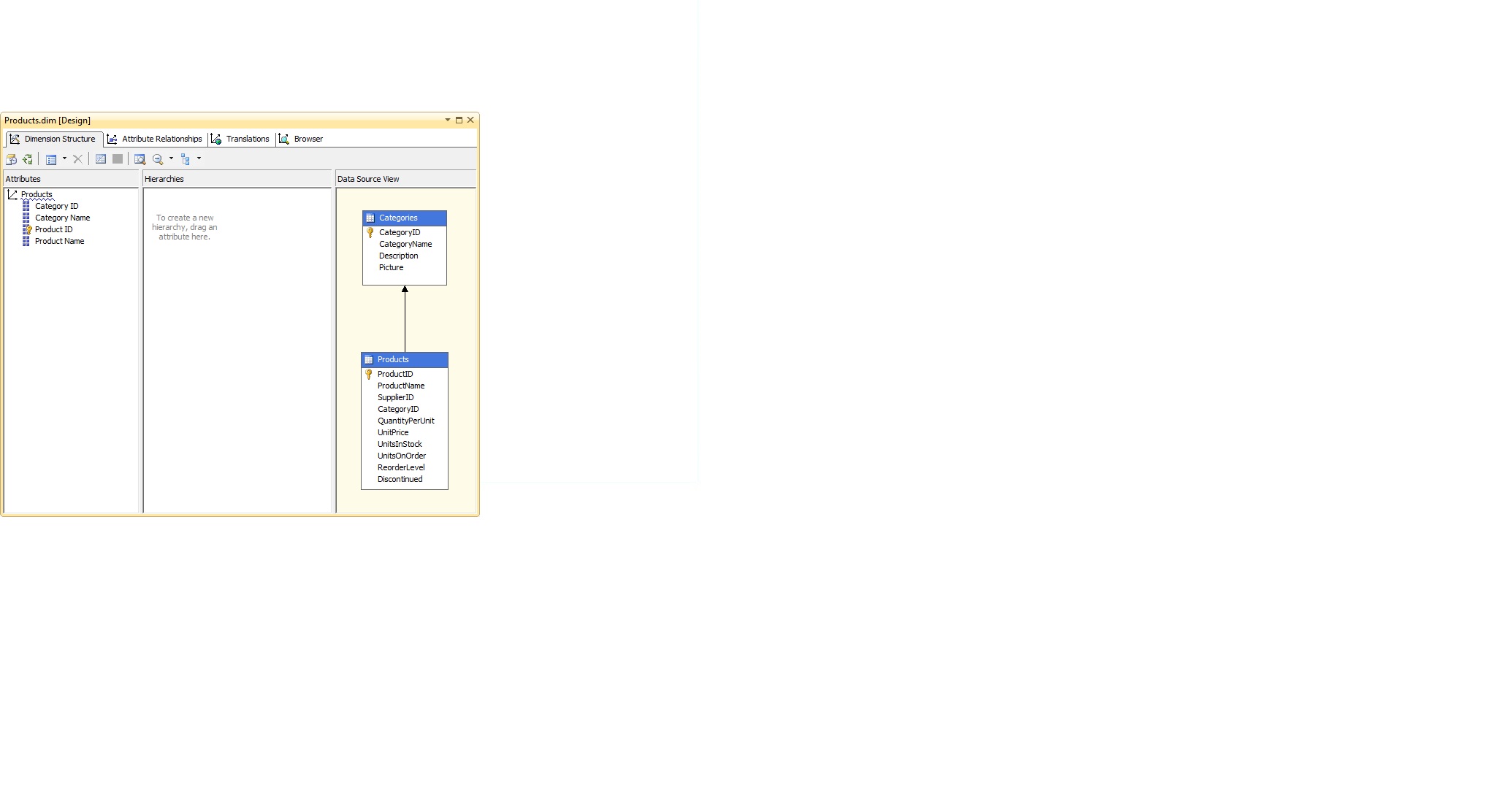
1. Правый щелчок мыши на Dimensions -> New Dimensions. В появившемся окне “Dimensions Wizard”, нажмите “Next”
2. Выберете “Use an existing table” и нажмите кнопку “Next”
3. Укажите Data source views – Northwind и в Main table – Products, нажмите “Next”

**Внимание!** В Main table вы указываете ту таблицу, которая будет иметь **первичный ключ**, т.е. главная таблица. В данном случае первичный ключ – это тот ключ товара, при помощи которого вы будете анализировать свои заказы, соответственно из списка таблиц выбирается таблица Products и автоматически определятся ключевое поле Product ID.

1. На следующей странице установите галочку напротив таблицы Categories и нажмите кнопку “Next”
2. Теперь нужно выбрать столбцы необходимые для использования в измерениях. Установите галочки напротив Product ID, Product Name, Category ID, Category Name. Нажмите кнопку “Next”
3. Название измерения оставьте без изменений (Products) и нажмите кнопку “Finish”

В окне, на рисунке 5, слева находятся те атрибуты которые можно использовать для построения измерений, справа отображается структура таблиц из Data Source Views, а в середине находятся иерархии.

Для продолжения работы вам необходима двухуровневая структура иерархии. На первом уровне будет **Category Name** (название категорий), а на втором – **Product Name** (название товара), т.е. при анализе продаж вы будите иметь возможность сначала выбрать категорию и проанализировать данные продаж по все категории или по нескольким категориям, а также можете спуститься на более низкий уровень и выполнить анализ по отдельным продуктам.



**Рисунок 5.** Результат создания измерения.

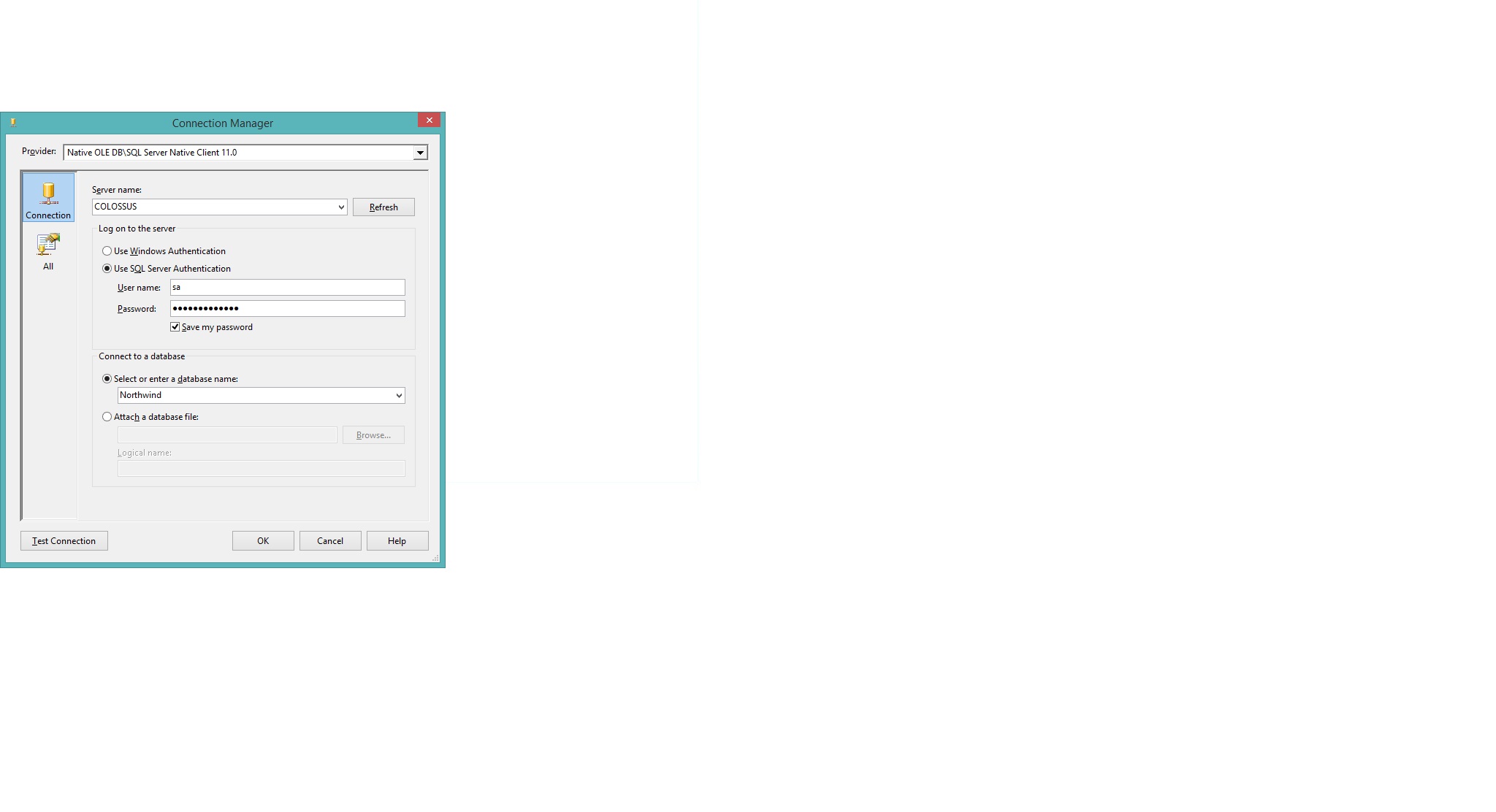
Для создания иерархии наведите курсор на атрибут **Category Name**, находящийся в левой части таблицы, и перетащите его в окно списка иерархий. Тоже самое проделайте с **Product Name**, только добавьте его в уже имеющуюся иерархию. Переименуйте иерархию в “Товары”.



**Рисунок 6.** Иерархия “Товары”

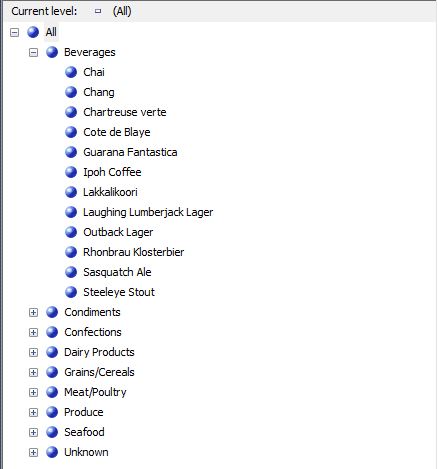
Посмотрим на результат работы через **Browser**. Выберете вкладку, разверните проект нажав кнопку **Process**. В открывшемся окне нажмите кнопку “Yes”, в новом окне “Process Dimensions - Products” нажмите кнопку “Run”. По завершения процесса закройте лишние окна и щёлкните по click here to try loading the browser again

**Внимание!** Есть вероятность что при создании OLAP-куба в пункте 3.b вы выбрали “Use Windows Authentication” вместо “Use SQL Server Authentication”, что приведет к ошибке подключения. Для устранения ошибки необходимо в окне Solution Explorer дважды щёлкнуть по источнику данных Northwind.ds. В появившемся окне нажать кнопку “Edit” после чего откроется окно “Connection Manager”. Здесь вам необходимо выбрать “Use SQL Server Authentication” и ввести логин и пароль, используемые для подключения к SQL серверу, как показано на рисунке 7.



**Рисунок 7.** Окно “Connection Manager”

Можно увидеть, как ваше измерение будет выглядеть для пользователя (рисунок 8). Первый уровень — это категории и второй уровень — это название самих товаров.

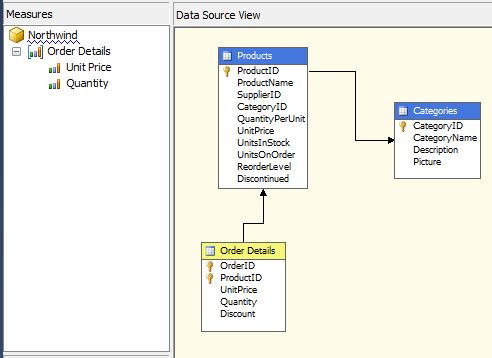


**Рисунок 8.** Вид измерения для пользователя

Теперь, имея одно измерение можно создать куб

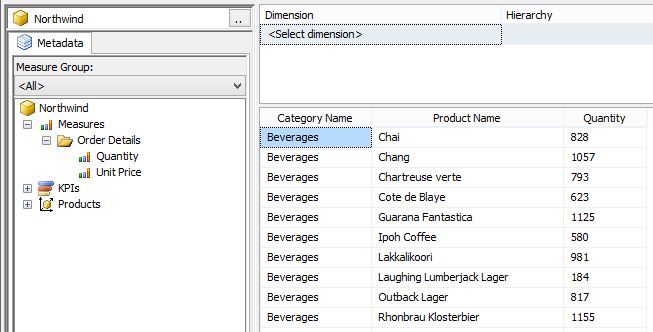
1. Правый щелчок мыши на Cubes -> New Cubes. В появившемся окне “Cubes Wizard”, нажмите “Next”
2. Выберете “Use existing table” и нажмите кнопку “Next”
3. Отметьте в списке таблиц “Order Details” и нажмите кнопку “Next”
4. Далее отметьте только меры “Unit Price” и “Quantity” и нажмите кнопку “Next”
5. Выберете измерение “Products” и нажмите кнопку “Next”
6. Далее мастер предложит создать новое измерение. Откажитесь убрав все галочки и нажмите кнопку “Next”
7. Назовите куб “Northwind” и нажмите кнопку “Finish”

В результате вы создали одномерный куб в котором есть: таблица с фактами – order details и измерение, которое использует две таблицы – Products и Categories. В качестве мер присутствуют Unit Price и Quantity.



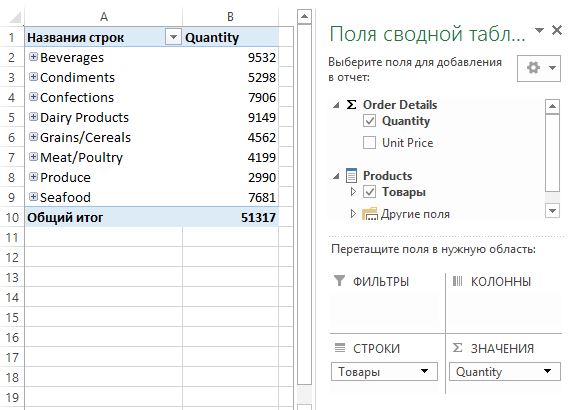
**Рисунок 9.** Одномерный куб Northwind

Перейдите во вкладку **Browser**, нажмите кнопку **Process**. В открывшемся окне нажмите кнопку “Yes”, в новом окне “Process Cube - Northewind” нажмите кнопку “Run”. По завершения процесса закройте лишние окна и щёлкните по click here to try loading the browser again. Из левого окна **Metadata** перетащите измерение **Products** и меру **Quantity**. В результате у вас должна получиться таблица, изображённая на рисунке 10.



**Рисунок 10.** Представление куба в Visual Studio

Этот инструмент для работы с кубами, который изначально присутствует в Visual Studio, не очень удобен. В частности, здесь идут категории измерения, категории товаров и суммарное количество проданных товаров, но нет возможности как-то сортировать или группировать данные. Для конечных пользователей Microsoft предлагает пользоваться MS Excel. Его можно открыть из браузера нажав кнопку **Analyze in Excel**. После открытия Excel перетащите иерархию **Товары** в колонку **СТРОКИ**, а в колонку **ЗНАЧЕНИЯ** добавьте **Quantity**. В результате вы должны увидеть на экране картину, аналогичную рисунку 11.



**Рисунок 11.** Представление куба в MS Excel

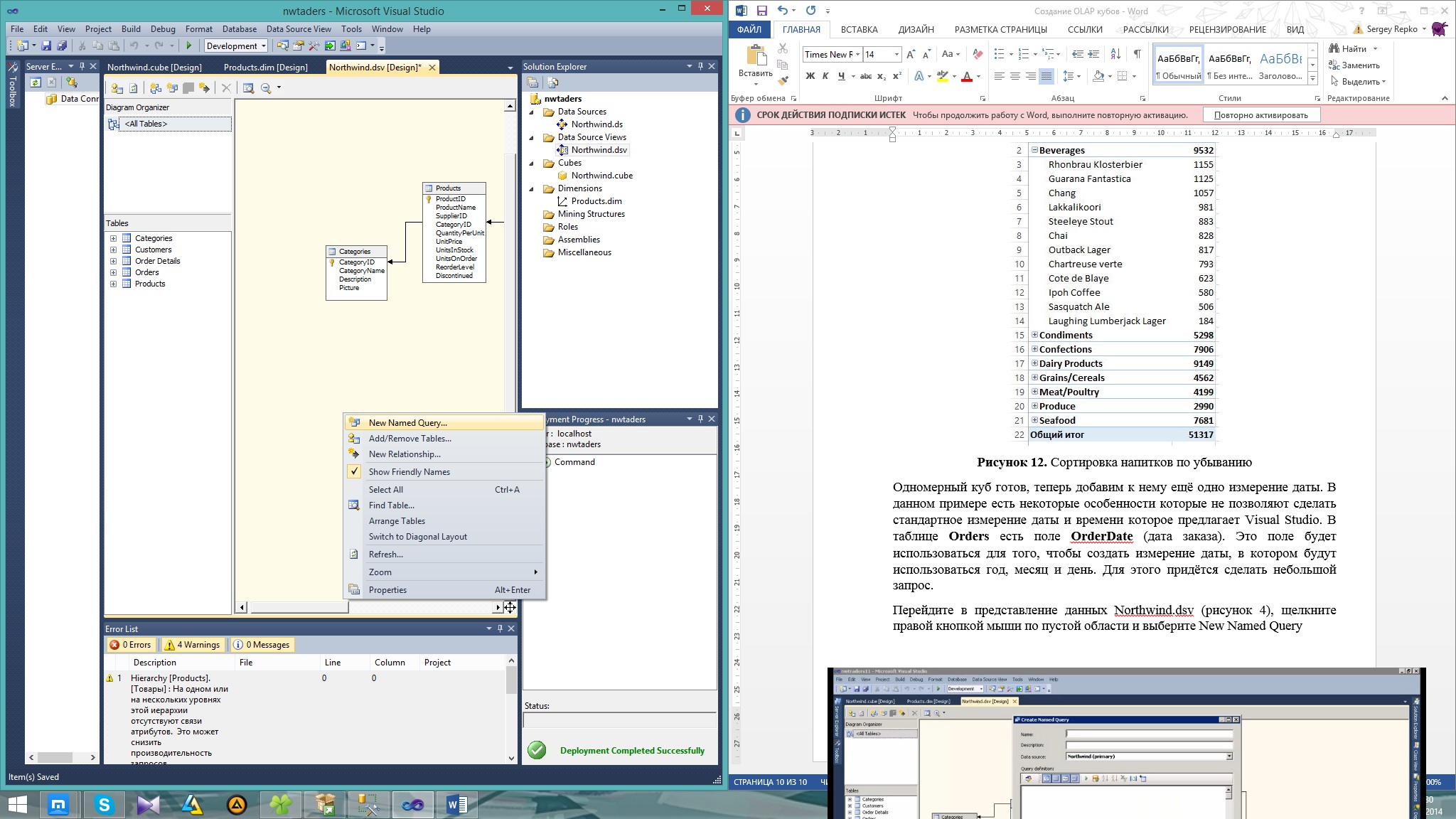
Прелесть OLAP-кубов для аналитики заключается в том, что можно зайти внутрь измерения (например, внутрь напитков/Beverages) и посмотреть более детальный анализ по группам товаров. Так же можно сделать сортировку (Главная -> Сортировка и фильтр) по возрастанию или убыванию и найти наиболее и наименее продаваемые напитки (рисунок 12).



**Рисунок 12.** Сортировка напитков по убыванию

Одномерный куб готов, теперь добавим к нему ещё одно измерение даты. В данном примере есть некоторые особенности которые не позволяют сделать стандартное измерение даты и времени которое предлагает Visual Studio. В таблице **Orders** есть поле **OrderDate** (дата заказа). Это поле будет использоваться для того, чтобы создать измерение даты, в котором будут использоваться год, месяц и день. Для этого придётся сделать небольшой запрос.

Перейдите в представление данных Northwind.dsv (рисунок 4), щелкните правой кнопкой мыши по пустой области и выберите New Named Query

. Назовите запрос “OrderDate” и введите SQL запрос:

SELECT DISTINCT

OrderDate,

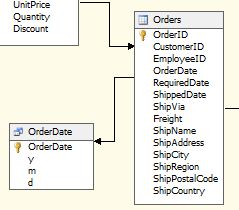
YEAR(OrderDate) AS y,

MONTH(OrderDate) AS m,

DAY(OrderDate) AS d

FROM Orders

И нажмите “OK”. Связь между таблицей Orders и запросом OrderDate будет происходить по ключу OrderDate, поэтому их необходимо связать друг с другом (наводите курсор на поле OrderDate таблицы Orders, зажимаете ЛКМ и тащите до поля OrderDate запроса OrderDate). В итоге должна получиться такая связь (рисунок 13).

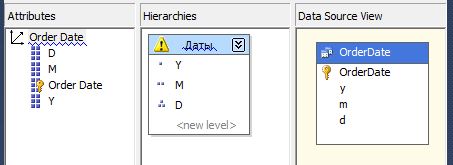


**Рисунок 13.** Связь между таблицей Orders и запросом OrderData

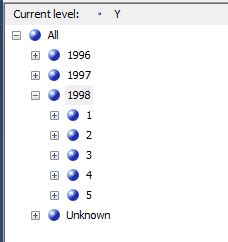
Дальше создайте новое измерение

1. Правый щелчок мыши на Dimensions -> New Dimensions. В появившемся окне “Dimensions Wizard”, нажмите “Next”
2. Выберете “Use an existing table” и нажмите кнопку “Next”
3. Укажите Data source views – Northwind и в Main table – OrderDate, нажмите “Next”
4. На следующей странице установите галочки напротив столбцов Order Date, Y, M, D и нажмите кнопку “Next”
5. Название измерения оставьте без изменений (Order Date) и нажмите кнопку “Finish”

Выстройте иерархию Y -> M -> D (год -> месяц -> день) и переименуйте её в “Даты”, нажмите кнопку **Process**, далее в открывшемся окне нажмите “Yes”, в следующем снова “Yes”. В окне “Process Dimension – Order Date” нажмите кнопку “Run”, подождите пока процесс завершиться и закройте лишние окна. В браузере можно посмотреть на результат. В итоге у вас должно получиться что-то похожее на рисунки 14.a и 14.b.



**Рисунок 14.a.** Структура измерения Order Data



**Рисунок 14.b.** Вид измерения Order Date во вкладке Browser

Следующий шаг – это привязка измерения Order Date к кубу. Для начала нужно добавиться новое измерение к кубу.

1. В окне Solution Explorer дважды щелкните по кубу Northwind.cube
2. Внизу слева в окне Dimensions щелкните правой кнопкой мыши по Northwind -> Add Cube Dimensions
3. В открывшемся окне выберите Order Date и нажмите “OK”

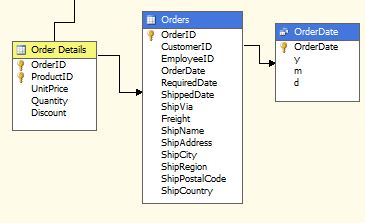
Измерение добавлено, но не связано, потому как факты лежат в таблице **Order Details**, а измерение связано с таблицей **Orders**. Для решения этой проблемы следует добавить новое, вспомогательное измерение, которое будет использовать только таблицу **Orders**.

1. Правый щелчок мыши на Dimensions -> New Dimensions. В появившемся окне “Dimensions Wizard”, нажмите “Next”
2. Выберете “Use an existing table” и нажмите кнопку “Next”
3. Укажите Data source views – Northwind и в Main table – Orders, нажмите “Next”
4. На следующей странице снимаем все галочки таблиц и нажмите кнопку “Next”
5. Теперь нужно выбрать столбцы необходимые для использования в измерениях. Установите галочки напротив Order ID и Order Date. Нажмите кнопку “Next”
6. Название измерения оставьте без изменений (Orders) и нажмите кнопку “Finish”

Ни каких иерархий создавать не нужно, этим измерением пользователи пользоваться не будет. Теперь возвращайтесь в куб и добавьте еще одно измерение Orders.

1. В окне Solution Explorer дважды щелкните по кубу Northwind.cube
2. Внизу слева в окне Dimensions щелкните правой кнопкой мыши по Northwind -> Add Cube Dimensions
3. В открывшемся окне выберите Order Date и нажмите “OK”

В результате настраивается связь (рисунок 15).

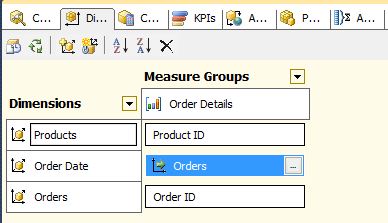


**Рисунок 15.** Настроенное соединение измерения Order Date с кубом

Чтобы проверить корректно ли измерение Order Date привязалось к таблице фактов перейдите на вкладку **Dimension Usage**. Как вы можете заметить, измерение Order Date никак не связано с группой мер, поэтому нужно указать что связь была настроена в ручную.

1. Наведите на серый прямоугольник расположенный за измерением Orders Date и нажмите кнопку
2. В отрывшемся окне “Define Relationship”
   1. В списке “Select relationship” выберете “References”
   2. В списке “Intermediate dimension” выберете “Orders”
   3. В списке “Reference dimension attribute” выберете “Order Date”
   4. В списке “Intermediate dimension attribute” тоже выберете “Order Date”
   5. Установите галочку напротив Materialize
3. Нажмите кнопку “OK”

Должна добавиться связь в списке групп мер (рисунок 16). Нажмите кнопку **Process** и завершите процесс постройки куба.



**Рисунок 16.** Содержимое вкладки Dimension Usage

Перейдите во вкладку **Browser** и нажмите **Analyze in Excel** чтобы посмотреть на построенный куб. Перетащите иерархию **Товары** в колонку **СТРОКИ**, иерархию **Даты** в колонку **КОЛОННЫ** и **Quantity** в колонку **ЗНАЧЕНИЯ**.

В результате вы получили таблицу, позволяющую проанализировать по категориям (напитки, кондитерские изделия, морепродукты) как по годам, месяцам или конкретным числам шли продажи. Вы можете углубиться в любой месяц и посмотреть продажи по каждому дню, отсортировать по популярности и по желанию сформировать статистику.