# Методы сбора, хранения, обработки и анализа данных

Лекция 13

Бизнес-аналитика

### ВВЕДЕНИЕ В БИЗНЕС-АНАЛИТИКУ

#### Бизнес-аналитика

- Бизнес-аналитика business intelligence BI
   область технологии баз данных
- Процесс интегрирования всех необходимых данных в один источник данных, к которому конечные пользователи могут осуществлять нерегламентированные запросы для анализа этих данных
- Цель: принятие лучших бизнес решений

### Работа с данными

- OLTP online transaction processing оперативная обработка транзакций
- OLAP системы бизнес-аналитики

#### **OLTP**

- реляционность обычно
- обработка большого числа коротких транзакций
- большое количество пользователей
- постоянное выполнение операций чтения и записи над небольшим объемом строк

#### Системы бизнес-аналитики

- данные собираются из различных источников
- данные приводятся в согласованное состояние
- данные регулярно загружаются в базу данных
  - хранилище данных data warehouse
  - витрина данных data mart
- основной режим только чтение
- количество одновременных пользователей небольшое
- создание отчетов по различным аспектам
- выполнение сложных запросов для сравнения данных

## Системы бизнес-аналитики и OLTP-системы

- OLTP-система доступна непрерывно
- Система бизнес-аналитики доступна только после очистки данных и их загрузки в базу данных

- OLTP-системы хранит текущие данные
- Система бизнес-аналитики сохраняет историю данных

#### Хранилище данных

- Хранилище данных база данных всех данных организации, к которым пользователи могут иметь единообразный доступ
- Проблемы:
  - большой объем данных
  - получены в разное время
  - хранятся в разных базах данных
- Решение: Консолидация данных

#### Консолидация

- Консолидация данных все одинаковые запросы, выполняемые к хранилищу данных в разное время, должны возвращать одинаковые результаты
- Процесс консолидации ETL (Extract, Transform, Load)
- сбор данных из различных источников
- очистка данных
- обеспечение качества данных

#### Организация данных

- Витрина (киоск) данных база данных всех данных на уровне подразделения организации, к которым пользователи могут иметь единообразный доступ
- Хранилище данных база данных всех данных на уровне организации, к которым пользователи могут иметь единообразный доступ

#### Проектирование

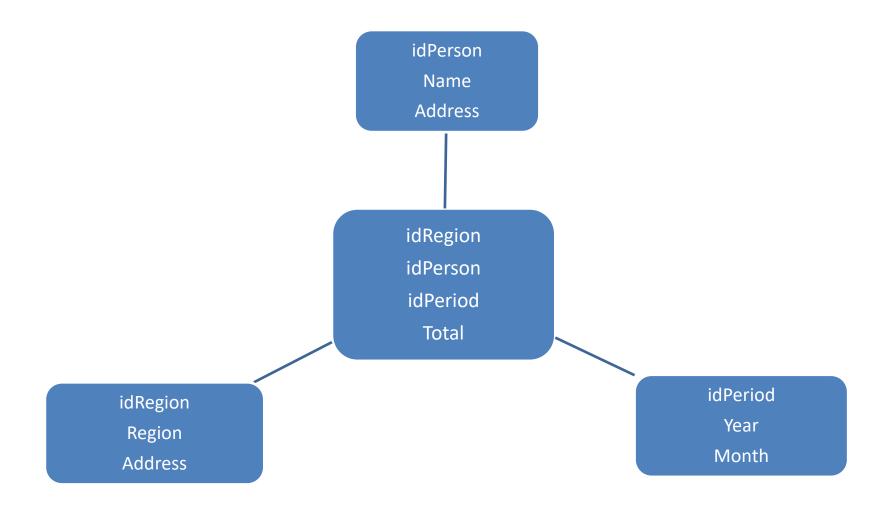
- Для проектирования хранилищ и витрин данных применяется размерная или пространственная модель
- Центральная таблица таблица фактов fact table
- Таблицы измерений dimension table
- Числовые столбцы таблицы фактов, кроме первичного ключа, называются мерами measure
- Имена столбцов таблиц измерений используются в качестве заголовков в отчетах

#### Проектирование хранилищ данных

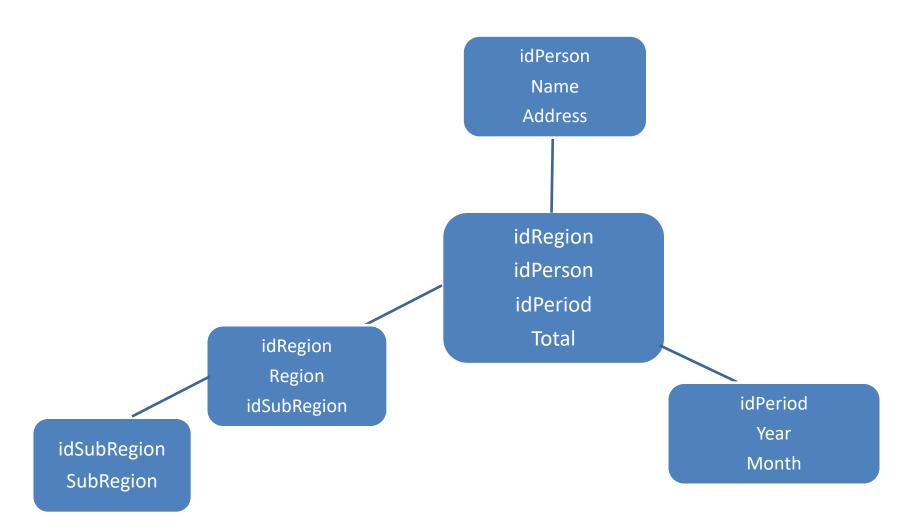
- Таблица фактов
  - Одна таблица обычно
  - Содержит большинство строк хранилища данных
  - Имеет составной первичный ключ
  - Неключевые столбцы имеют числовой тип данных
  - Основная задача дать сумму, максимальное, минимальное, среднее и пр. значение

- Таблица измерений
  - Много таблиц
  - Содержит сравнительно небольшой набор строк
  - Первичный ключ состоит из одного столбца
  - Столбцы содержат описания измерений
  - Разреженность измерений

### Звезда



#### Снежинка



#### Агрегирование

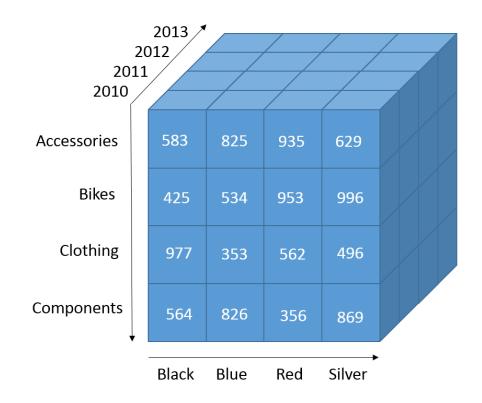
- Агрегатные таблицы aggregate tables промежуточные таблицы для хранения заранее подсчитанных низкоуровневых данных
- Процесс создания агрегирование данных

#### Уровень агрегирования

- Не выполнять агрегирование
- Выполнять исчерпывающее агрегирование для каждой возможной комбинации запросов
- Выполнить необходимое агрегирование

### Кубы

- Куб это подмножество данных из хранилища данных
- Может иметь любое количество измерений



#### Виды архитектуры

- реляционная OLAP ROLAP relational online analytical processing – предварительно вычисленные данные не сохраняются
- многомерная OLAP MOLAP multidimensional online analytical processing – данные на уровне листьев и их агрегированные данные сохраняются в кубе
- гибридная OLAP HOLAP hybrid online analytical processing агрегированные данные сохраняются, данные на уровне листьев не сохраняются

#### Доступ к данным

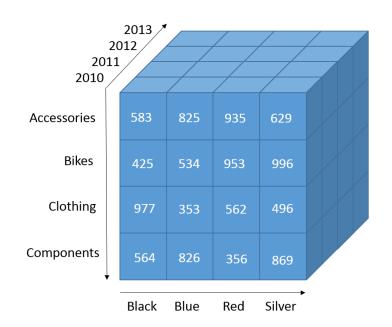
- Составление отчетов
- Оперативная аналитика
- Извлечение информации из данных data mining

### **SQL Server Analysis Services**

- Службы SQL Server Analysis Services группа сервисов, которые используются для создания многомерных баз данных и управления этими данными
- 1994 -1995 Панорама
- 1996 Microsoft
- 1998 Microsoft OLAP Services и MDX

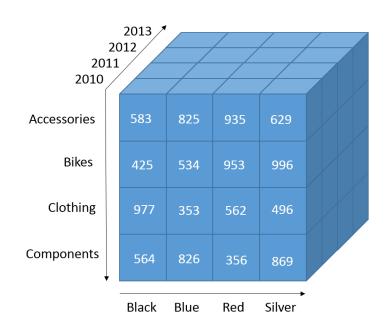
#### Термины

- куб cube
- измерение dimension
- элемент member
- кортеж tuple
- cpe3 slice
- иерархия hierarchy
- ячейка cell
- мера measure
- группа мер measure group
- подкуб subcube



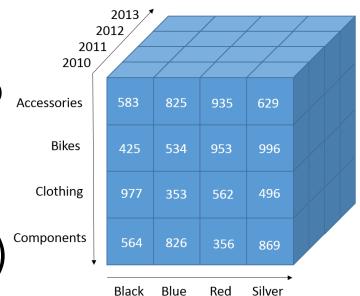
### Куб

- Куб основа многомерной БД
- Куб подмножество данных из хранилища данных, которое можно организовать в многомерную структуру



#### Измерение

 Измерение – описание для элемента данных, по которому проводится анализ



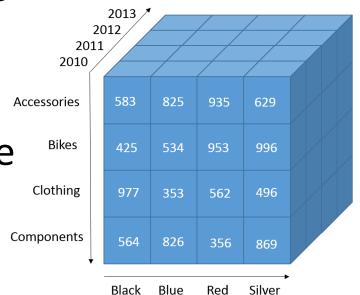
Размер (кардинальность)
 измерения – количество
 элементов измерения

#### Элемент

• Элемент – одно значение измерения

• Значение элемента — Acc значение, ассоциируемое с элементом

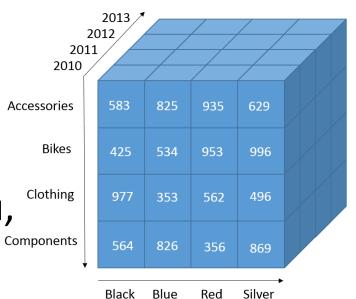
 Атрибут элемента – полная коллекция элементов измерения



#### Кортеж и срез

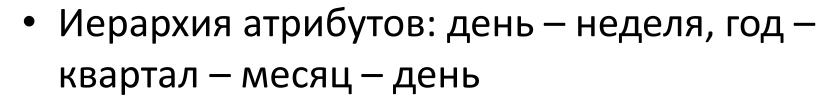
 Кортеж – одно значение данных с координатами по всем измерениям

 Срез – одна или более координат произвольная, остальные фиксированные

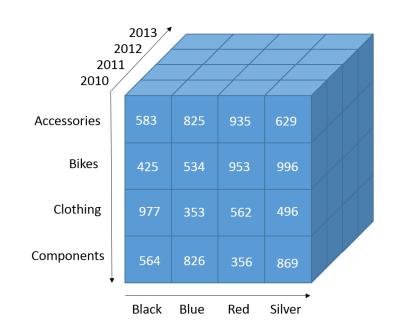


#### Иерархия

- День
- Неделя
- Месяц
- Квартал
- Год, 5 лет, 10 лет

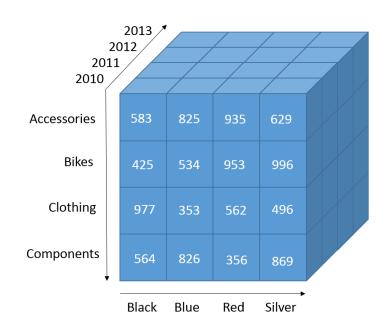


• Уровень иерархии



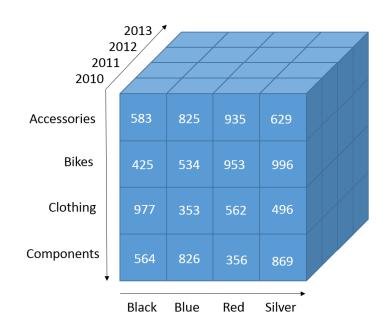
#### Ячейки

 Ячейка – набор точек пространства многомерного куба



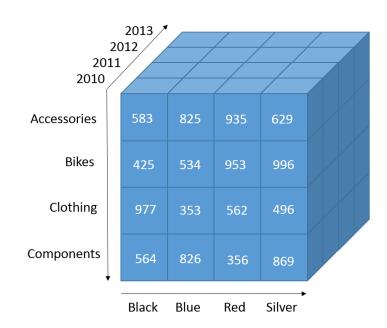
#### Меры

- Мера значения в ячейке
- Может быть несколько– сумма, количество
- Все вместе измерение мер
- Частичное объединение – группа мер



#### Подкуб

• Подкуб – часть пространства куба



# Создание и обработка многомерного куба

- Создание проекта бизнес-аналитики
- Определение источников данных
- Создание представлений источников данных
- Создание куба
- Агрегирование
- Обработка куба
- Просмотр куба

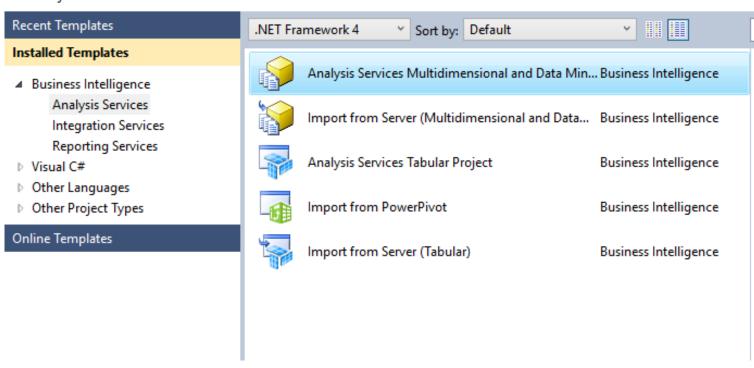
#### **SQL Server Data Tools**

- SQL Server Data Tools
- AdventureWorksDW

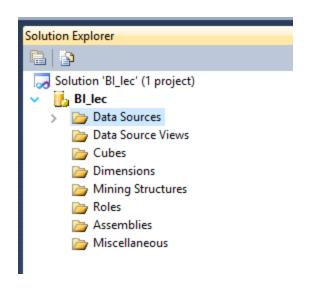


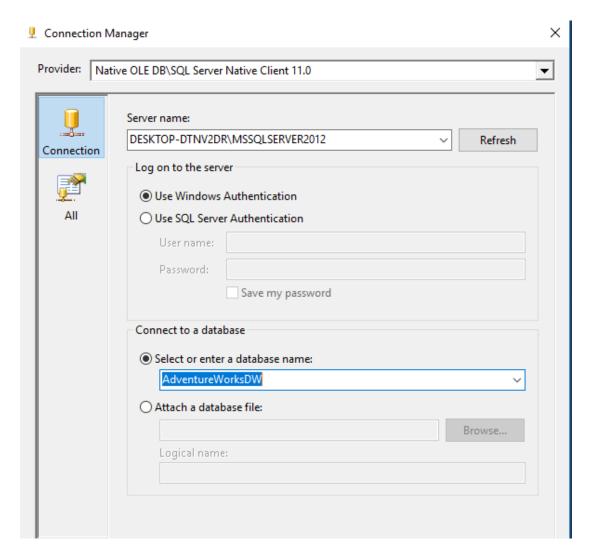
# Создание проекта бизнес-

#### New Project

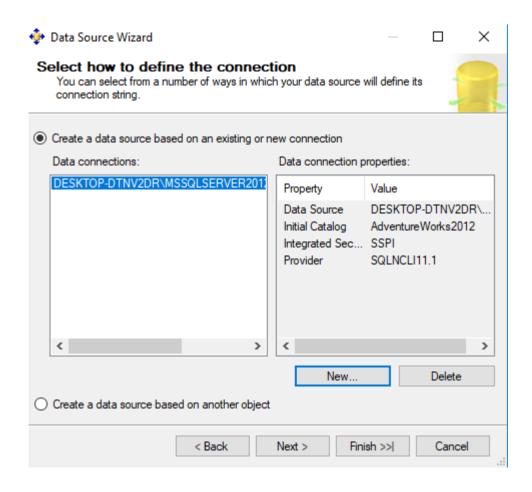


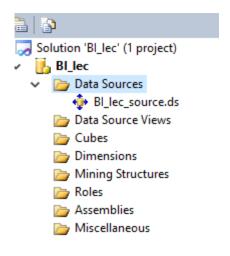
#### Определение источника данных



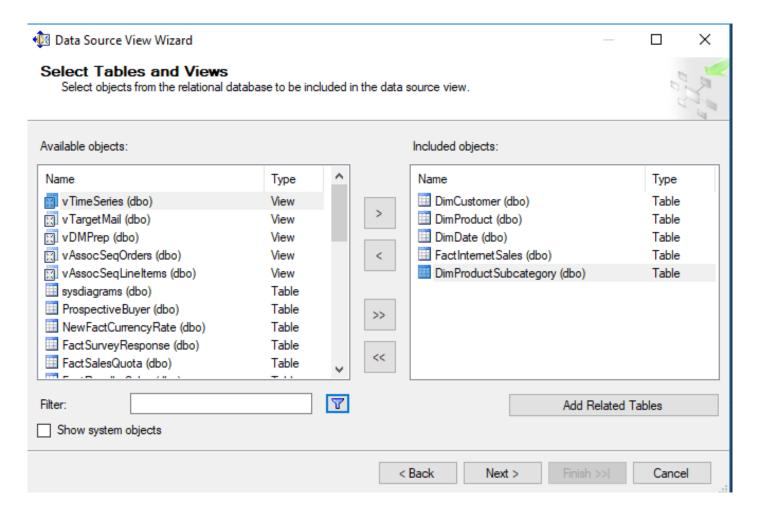


#### Определение источника данных

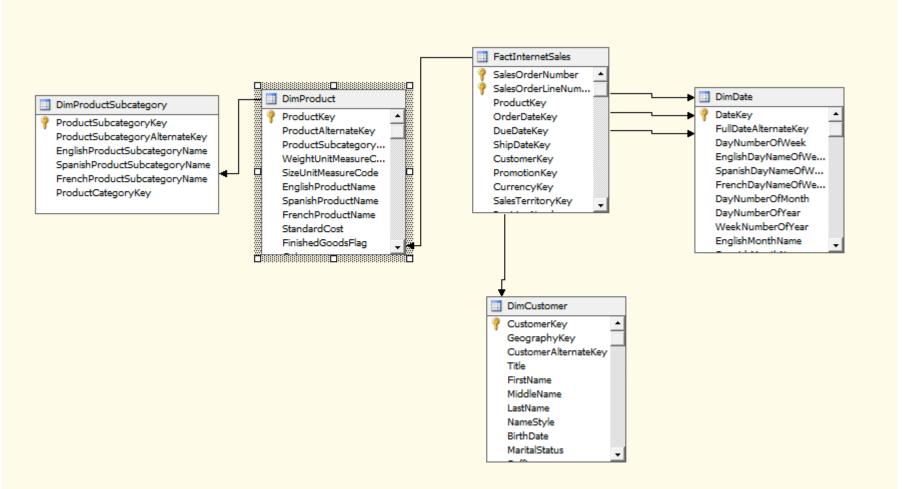




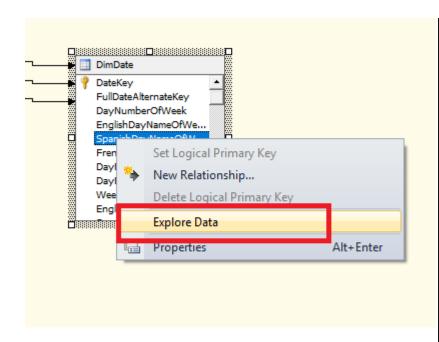
# Создание представлений источников данных



# Создание представлений источников данных

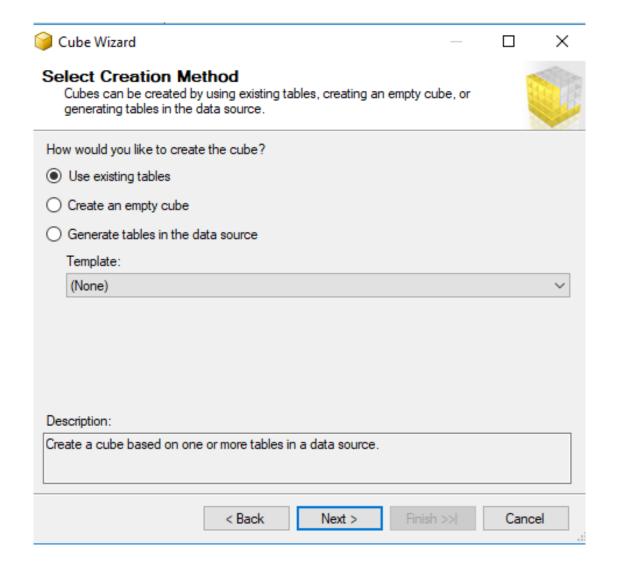


# Создание представлений источников данных

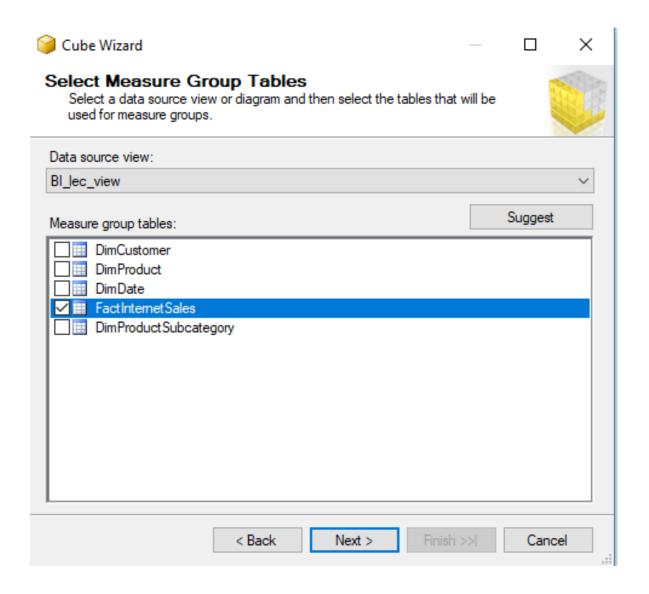


DateKey	FullDateAlternateKey	DayNumberOfWeek	EnglishDayNa
20050101	2005-01-01 00:00:00Z	7	Saturday
20050102	2005-01-02 00:00:00Z	1	Sunday
20050103	2005-01-03 00:00:00Z	2	Monday
20050104	2005-01-04 00:00:00Z	3	Tuesday
20050105	2005-01-05 00:00:00Z	4	Wednesday
20050106	2005-01-06 00:00:00Z	5	Thursday
20050107	2005-01-07 00:00:00Z	6	Friday
20050108	2005-01-08 00:00:00Z	7	Saturday
20050109	2005-01-09 00:00:00Z	1	Sunday
20050110	2005-01-10 00:00:00Z	2	Monday
20050111	2005-01-11 00:00:00Z	3	Tuesday
20050112	2005-01-12 00:00:00Z	4	Wednesday
20050113	2005-01-13 00:00:00Z	5	Thursday
20050114	2005-01-14 00:00:00Z	6	Friday
20050115	2005-01-15 00:00:00Z	7	Saturday
20050116	2005-01-16 00:00:00Z	1	Sunday
20050117	2005-01-17 00:00:00Z	2	Monday
20050118	2005-01-18 00:00:00Z	3	Tuesdav

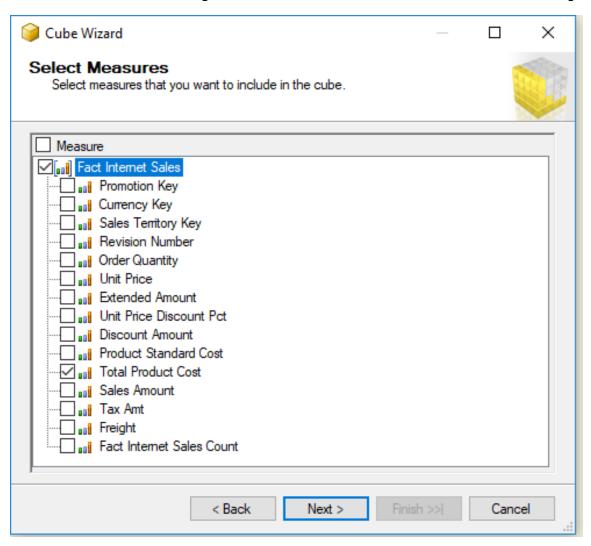
## Создание куба



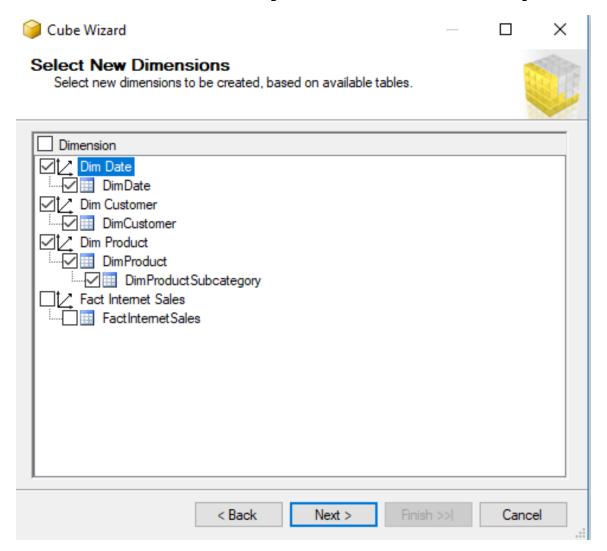
## Создание куба – таблица фактов



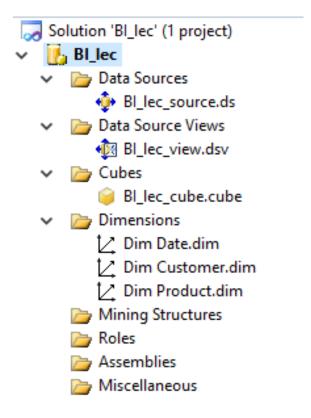
## Создание куба – таблица фактов

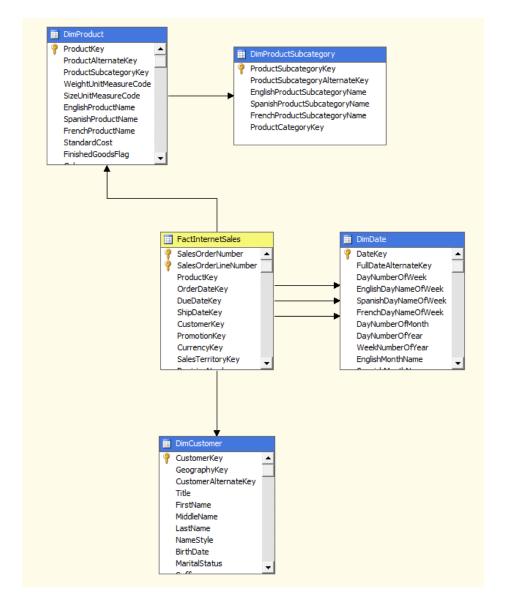


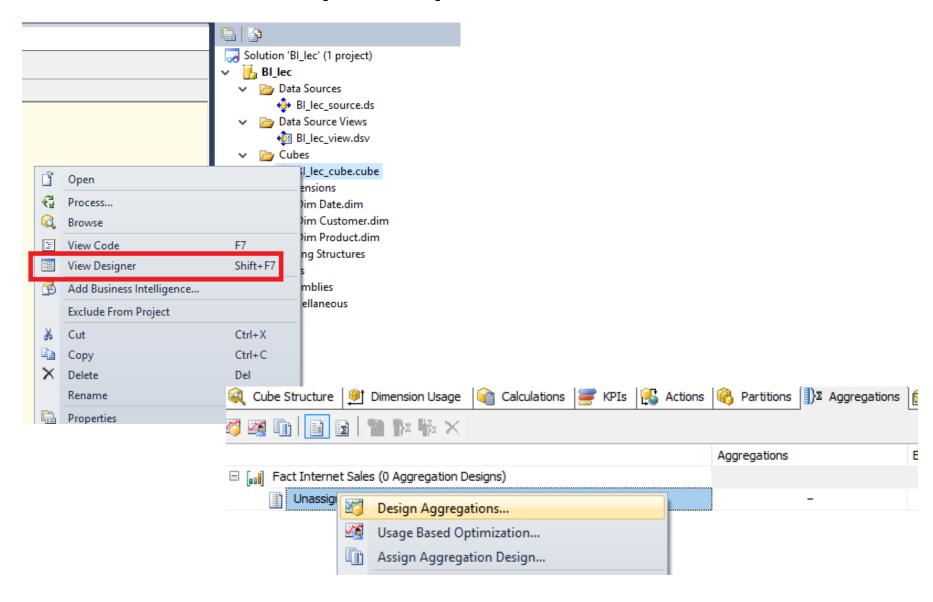
## Создание куба –измерения



## Создание куба







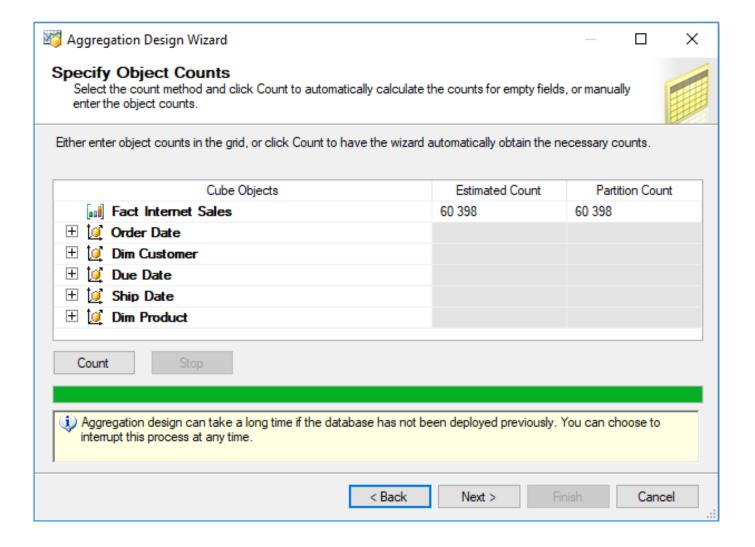


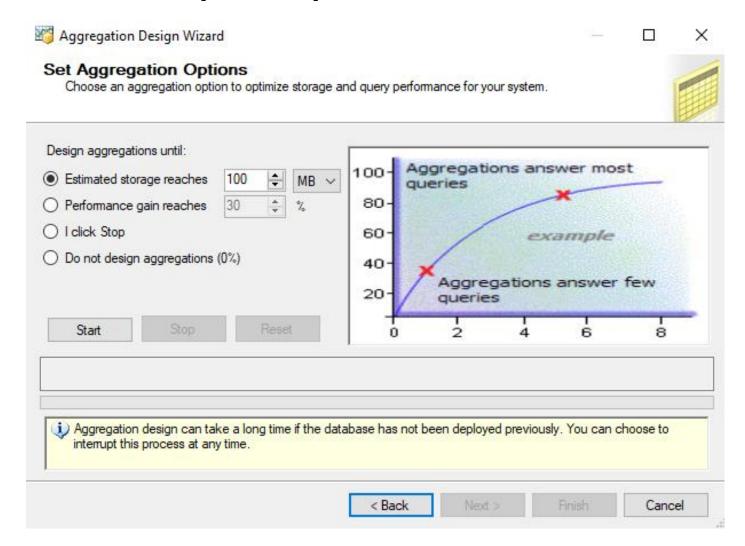
#### **Review Aggregation Usage**

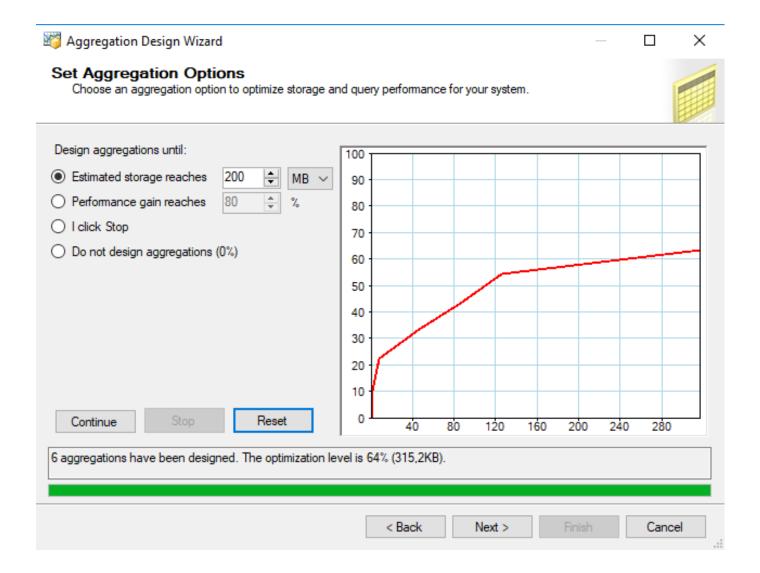
Aggregation usage settings can be modified to include or exclude attributes from consideration.



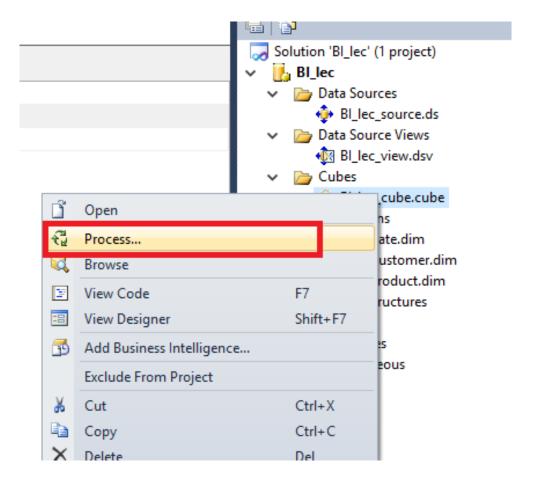
Customer Key  Due Date Date Key	1 ① 1 ② 1 ①	0 0 0	0 0 0	0 0 0
Date Key  Dim Customer  Customer Key  Due Date  Date Key  Ship Date	1 ① 1 ① 1	0 0 0	0 0 0	0 0
Customer Key  Due Date Date Key  Ship Date	1 0	0 0	0 0	0 0
Due Date Date Key Ship Date	1 0	0	0	0
Date Key  Ship Date	1	0	0	0
Ship Date	1	0	0	0
-	1	0	0	
Date Key	- James		U	0
	$\odot$	0	0	0
∃ 💆 Dim Product	2	0	0	0
Product Key	•	0	0	0
Product Subcategory Key	•	0	0	0
Set All to Default				



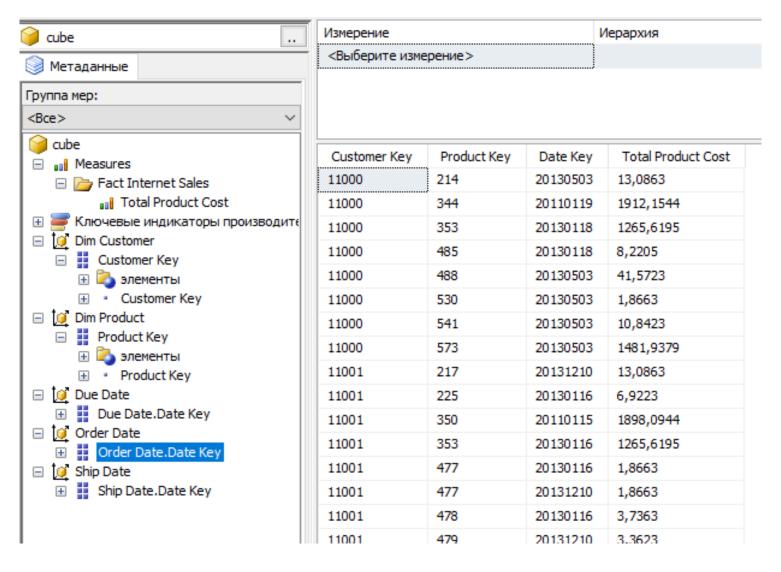




## Обработка куба

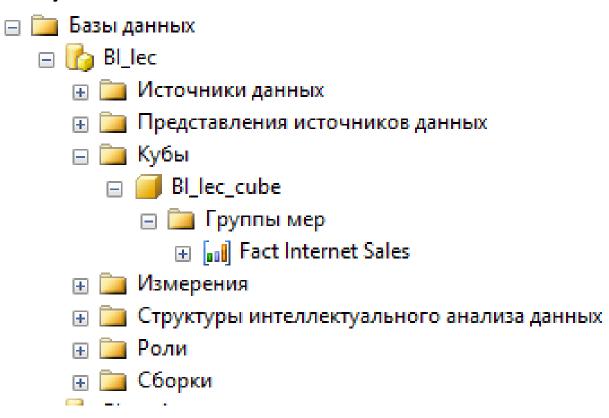


## Просмотр куба

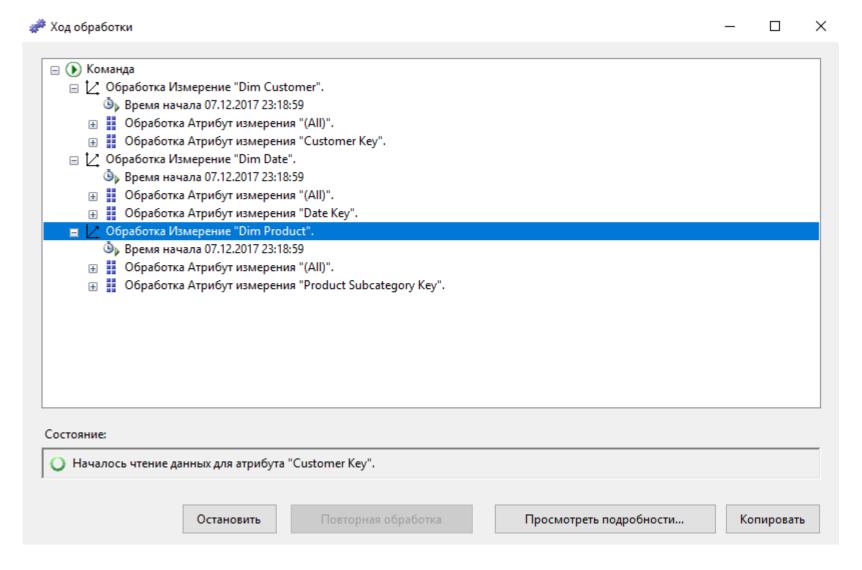


## Просмотр куба

• При подключении к службе SQL Server Analysis Services



## Просмотр куба



## Извлечение данных

- Язык MDX Multi-Dimensional eXpressions
- Power Pivot for Excel

#### **MDX**

- MDX язык запросов для извлечения данных из многомерных БД
- Разработан в Microsoft 1998
- Является частью спецификации OLEDB для OLAP

## Извлечение данных в MDX

- SELECT задает результирующий набор (подкуб), который содержит подмножество многомерных данных, выбранных из куба
- SELECT ... FROM ... WHERE ...

#### Оси

- Необходимо указать оси (подмножества измерений) ON axis(...):
- ON COLUMNS
- ON ROWS
- ON PAGES

```
SELECT [Measures].[Order Quantity] ON COLUMNS
FROM [BI_lec_cube];

100 % 
Сообщения Результаты

Order Quantity
60398
```

## Извлечение данных MDX

```
□SELECT [Measures].[Total Product Cost] ON Columns,
     ({[Dim Customer].[Customer Key].[11000]:[Dim Customer].[Customer Key].[36450]}) ON ROWS
     FROM [BI lec cube]
100 %
 Сообщения
               Результаты
                Total Product Cost
 11000
                   4735.2995
 11001
                   3587.9961
 11002
                   4659.1599
 11003
                   4672.1636
 11004
                   4694.1049
 11005
                   4661 4646
 11006
                   4665.3144
 11007
                    4695.423
 11008
                   4651.559
 11009
                   4645.9564
 11010
                   4652.9959
 11011
                   4687,6459
 11012
                    30.3914
                    42 6212
 11013
```

## Наборы

 Набор – совокупность кортежей, определенных с помощью одних и тех же измерений

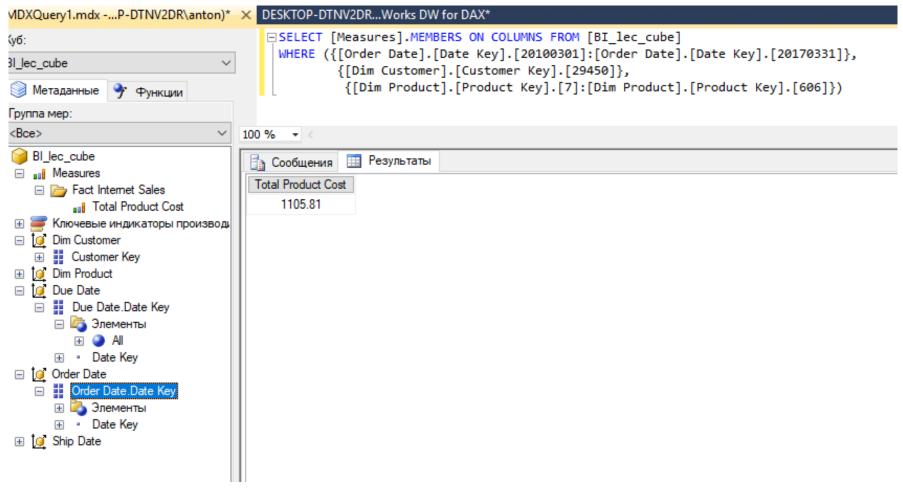
```
☐SELECT [Measures].[Order Quantity] ON COLUMNS

   FROM [BI lec cube]
   WHERE {[Dim Product].[Product Key].&[600]:[Dim Product].[Product Key].&[606]};
             Результаты
Сообщения
Order Quantity
   1150
   □SELECT [Measures].[Order Quantity] ON COLUMNS
    FROM [BI lec cube]
     WHERE {[Dim Product].[Product Key].&[600], [Dim Product].[Product Key].&[606]};
100 % ▼ <
              Результаты
 Сообщения
 Order Quantity
     427
```

## Извлечение данных MDX

```
□SELECT {[Due Date].[Date Key].[20110110]: [Due Date].[Date Key].[20110121]} ON rows,
       ({[Dim Customer].[Customer Key].[11000]:[Dim Customer].[Customer Key].[11012]}) ON columns
       FROM [BI lec cube]
100 % ▼ <
                     Результаты
 Сообщения
                11000
                           11001
                                       11002
                                                      11003
                                                                   11004
                                                                               11005
                                                                                           11006
                                                                                                        11007
                                                                                                                     11008
                                                                                                                               11009
                                                                                                                                         11010
                                                                                                                                                      11011
                                                                                                                                                                  11012
 20110110
                                                    1912.1544
                 (null)
                            (null)
                                        (null)
                                                                   (null)
                                                                                (null)
                                                                                             (null)
                                                                                                         (null)
                                                                                                                      (null)
                                                                                                                                (null)
                                                                                                                                          (null)
                                                                                                                                                       (null)
                                                                                                                                                                   (null)
 20110111
                                                                             1898.0944
                                                                                                                                                    1912.1544
                 (null)
                            (null)
                                        (null)
                                                       (null)
                                                                   (null)
                                                                                                         (null)
                                                                                                                      (null)
                                                                                                                                (null)
                                                                                                                                          (null)
                                                                                                                                                                   (null)
                                                                                             (null)
 20110112
                 (null)
                            (null)
                                                       (null)
                                                                                                         (null)
                                                                                                                      (null)
                                                                                                                                (null)
                                                                                                                                          (null)
                                                                                                                                                                   (null)
                                        (null)
                                                                    (null)
                                                                                (null)
                                                                                             (null)
                                                                                                                                                       (null)
                                                                                                                                                                   (null)
 20110113
                 (null)
                            (null)
                                        (null)
                                                       (null)
                                                                    (null)
                                                                                (null)
                                                                                             (null)
                                                                                                         (null)
                                                                                                                      (null)
                                                                                                                                (null)
                                                                                                                                          (null)
                                                                                                                                                       (null)
 20110114
                 (null)
                            (null)
                                                                                (null)
                                                                                                                      (null)
                                                                                                                                (null)
                                        (null)
                                                       (null)
                                                                    (null)
                                                                                             (null)
                                                                                                         (null)
                                                                                                                                          (null)
                                                                                                                                                       (null)
                                                                                                                                                                   (null)
 20110115
                 (null)
                            (null)
                                        (null)
                                                       (null)
                                                                    (null)
                                                                                (null)
                                                                                             (null)
                                                                                                         (null)
                                                                                                                      (null)
                                                                                                                                (null)
                                                                                                                                          (null)
                                                                                                                                                       (null)
                                                                                                                                                                   (null)
 20110116
                                                                                                                      (null)
                 (null)
                            (null)
                                        (null)
                                                       (null)
                                                                   (null)
                                                                                (null)
                                                                                             (null)
                                                                                                         (null)
                                                                                                                                (null)
                                                                                                                                          (null)
                                                                                                                                                       (null)
                                                                                                                                                                   (null)
 20110117
                 (null)
                            (null)
                                        (null)
                                                       (null)
                                                                   (null)
                                                                                (null)
                                                                                             (null)
                                                                                                         (null)
                                                                                                                      (null)
                                                                                                                                (null)
                                                                                                                                          (null)
                                                                                                                                                                   (null)
                                                                                                                                                       (null)
 20110118
                 (null)
                            (null)
                                        (null)
                                                       (null)
                                                                   (null)
                                                                                (null)
                                                                                                         (null)
                                                                                                                      (null)
                                                                                                                                (null)
                                                                                                                                          (null)
                                                                                                                                                                   (null)
                                                                                             (null)
                                                                                                                                                       (null)
 20110119
                 (null)
                            (null)
                                     1912.1544
                                                       (null)
                                                                   (null)
                                                                                (null)
                                                                                             (null)
                                                                                                         (null)
                                                                                                                      (null)
                                                                                                                                (null)
                                                                                                                                          (null)
                                                                                                                                                       (null)
                                                                                                                                                                   (null)
 20110120
                 (null)
                            (null)
                                                                                                                      (null)
                                        (null)
                                                       (null)
                                                                    (null)
                                                                                (null)
                                                                                             (null)
                                                                                                         (null)
                                                                                                                                (null)
                                                                                                                                          (null)
                                                                                                                                                       (null)
                                                                                                                                                                   (null)
 20110121
                 (null)
                            (null)
                                        (null)
                                                       (null)
                                                                    (null)
                                                                                (null)
                                                                                             (null)
                                                                                                      1912.1544
                                                                                                                      (null)
                                                                                                                                (null)
                                                                                                                                          (null)
                                                                                                                                                       (null)
                                                                                                                                                                   (null)
```

## Извлечение данных MDX



#### Вычисляемые значения

• WITH MEMBER ... AS ... - позволяет получить все значения измерения

```
WITH MEMBER [Measures].[Profit] AS '([Measures].[Total Product Cost]*0.25)'

SELECT [Measures].[Profit] ON COLUMNS,

{[Dim Customer].[Customer Key].[11000],[Dim Customer].[Customer Key].[12000]} ON ROWS

FROM [BI_lec_cube];

100 % ▼

Profit

11000 1183,824875

12000 1171,237775
```

## Операторы

- Комментарии -- // /\* ...\*/
- Арифметические операторы + \*/
- Операторы для наборов + \*
- Логические операторы > < <= >= = <> AND OR XOR NOT
- Специальные операторы {},:

## Функции

- Числовые
- Обработки строковых значений
- Для наборов
- Для элемента измерения
- Для иерархий
- Прочие

## Функции - FILTER

```
□ SELECT {[Measures].[Order Quantity], [Measures].[Total Product Cost]} ON COLUMNS,

{FILTER([Dim Customer].[Customer Key].MEMBERS, [Measures].[Total Product Cost] > 7000)} ON ROWS

FROM [BI_lec_cube];
```

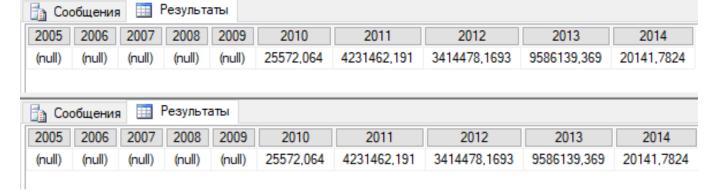
100 % 🔻 <								
🛅 Cool	бщения	Pe	езультаты					
	Order Quantity		Total Product Cost					
All	603	98	17277793,5757001					
11433	12		7713,07					
11439	14		7413,2821					
12124	12		7987,1323					
12131	11		8033,9674					
12132	14		8020,46					
12296	12		7975,5383					
12300	10		8005,4609					
12301	13		8044,9593					
12307	11		7975,176					
12308	14		8014,67					
12321	15		7994,6162					

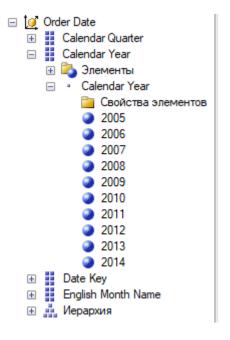
## Функции

```
☐ SELECT [Order Date].[Calendar Year].[2005].LEVEL ON COLUMNS | FROM [BI_lec_cube];

☐ SELECT [Order Date].[Calendar Year].[2011].LEVEL ON COLUMNS | FROM [BI_lec_cube];

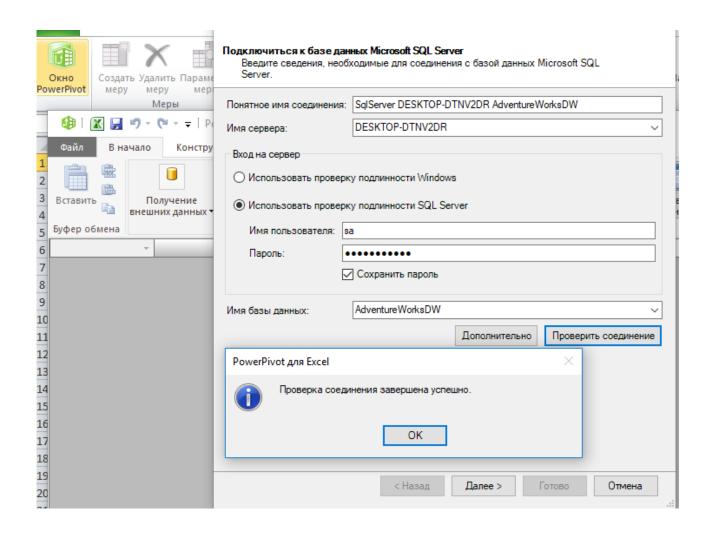
☐ SELECT [Order Date].[Calendar Year].[2015].LEVEL ON COLUMNS | FROM [BI_lec_cube];
```





Выполнение запроса...

Query (1, 8) Не удалось применить функцию многомерных выражений, поскольку при вычислении выражения аргумента получен неопределенный элемент. Выполнение завершено



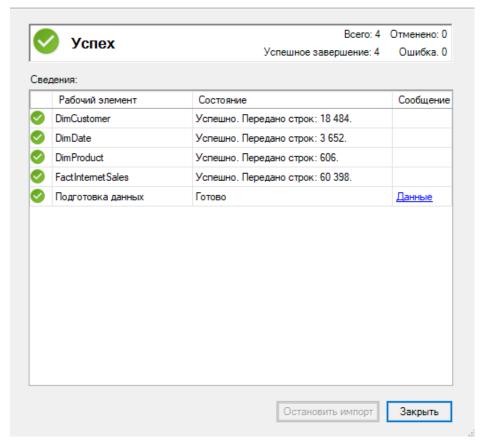
Мастер импорта таблиц

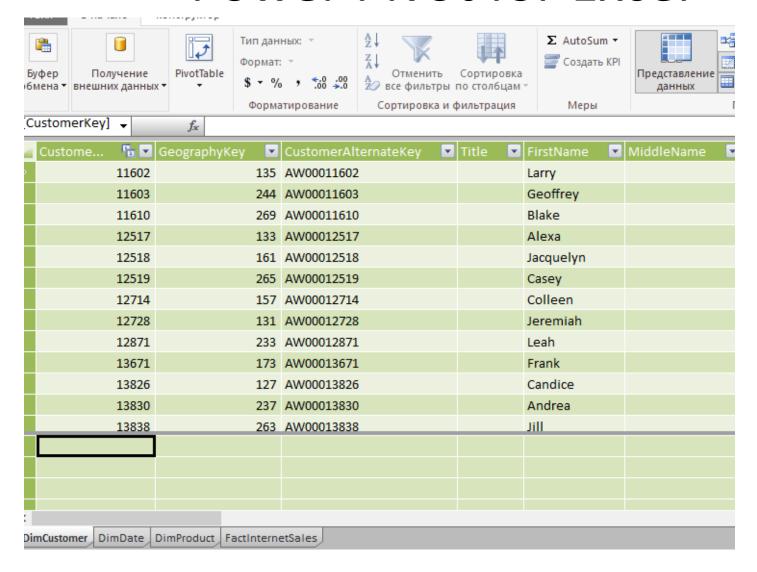
?

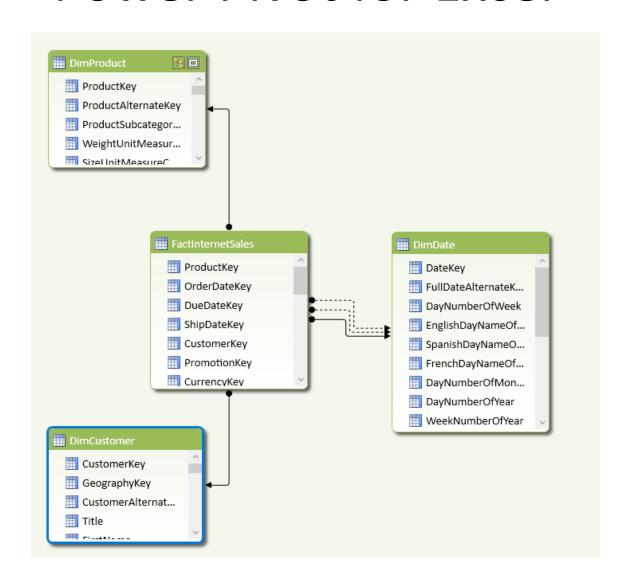
 $\times$ 

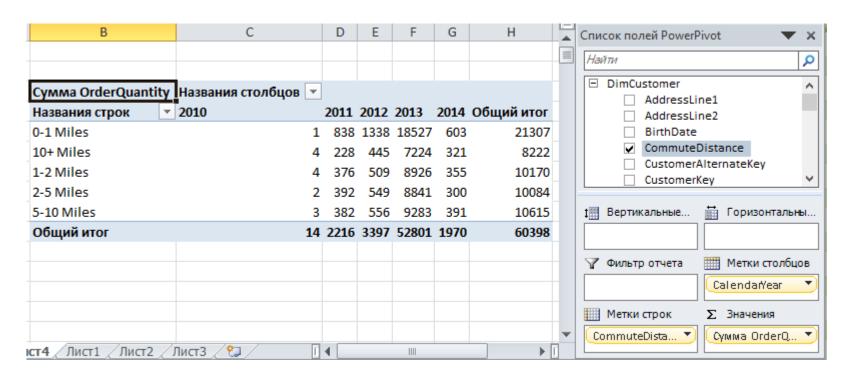
#### Идет импорт

Операция импорта может занять несколько минут. Чтобы остановить операцию импорта, нажмите кнопку "Остановить импорт".









# Вопросы?