



[Анализ данных на языке SQL]

День 3

Вечер 6

www.specialist.ru

Программа курса

- День 1. Работа с одной таблицей
- День 2. Работа с несколькими таблицами
- **День 3. Другие вопросы (более сложные)**

День 3: Другие вопросы (продолжение)

- Ранжирование
- Постраничный вывод
- Смещение (LAG, LEAD)
- Оконная агрегация
- Хранилища и витрины данных (обзор)
- Доступ к данным из Экселя
- Итоговая работа

Ранжирование

- Оконные функции ранжирования

```
SELECT ProductName, UnitPrice,  
       Row_Number () OVER (ORDER BY UnitPrice DESC),  
       Rank () OVER (ORDER BY UnitPrice DESC),  
       Dense_Rank () OVER (ORDER BY UnitPrice DESC)  
FROM Products  
ORDER BY UnitPrice DESC
```

100 %

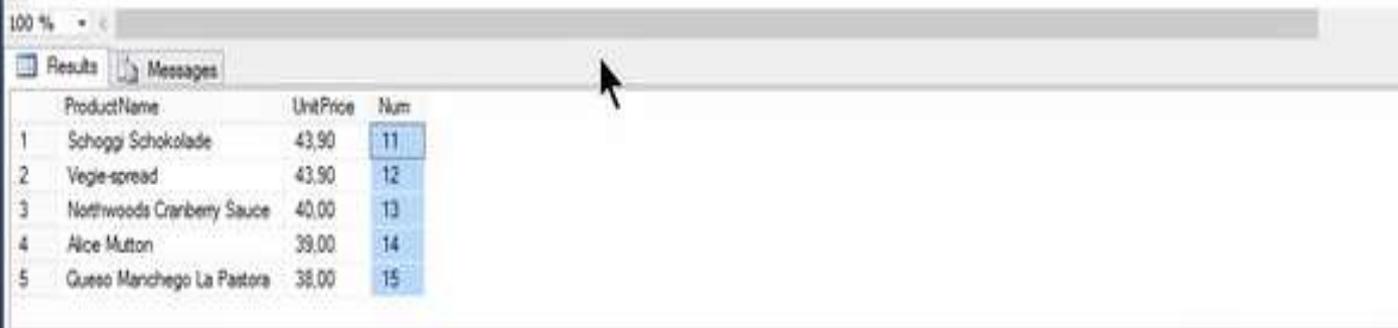
Results Messages

	ProductName	UnitPrice	(No column name)	(No column name)	(No column name)
1	Cote de Blaye	263.50	1	1	1
2	Thuringer Rostbratwurst	123.79	2	2	2
3	Mahi Kobe Niku	97.00	3	3	3
4	Sir Rodney's Marmalade	81.00	4	4	4
5	Camaron Tigers	62.50	5	5	5
6	Raclette Courdavault	55.00	6	6	6
7	Manjimup Dried Apples	53.00	7	7	7
8	Tarte au sucre	49.30	8	8	8
9	Ipoh Coffee	46.00	9	9	9
10	Rossle Sauerkraut	45.60	10	10	10
11	Schoggi Schokolade	43.90	11	11	11
12	Vegie-spread	43.90	12	11	11
13	Northwoods Cranberry Sauce	40.00	13	13	12
14	Alice Mutton	39.00	14	14	13
15	Queso Manchego La Pastora	38.00	15	15	14
16	Gnocchi di nonna Alice	38.00	16	15	14
17	Gudbrandsdalsost	36.00	17	17	15

Постраничный вывод

```
SET @Page = 3
SET @PageSize = 5

SELECT *
FROM (
    SELECT ProductName, UnitPrice,
           Row_Number () OVER (
                                   ORDER BY UnitPrice DESC
                               ) AS Num
    FROM Products
) MyScroller
WHERE Num BETWEEN (@Page - 1) * @PageSize + 1 AND @Page * @PageSize
```



	ProductName	UnitPrice	Num
1	Schoggi Schokolade	43.90	11
2	Vegie-spread	43.90	12
3	Northwoods Cranberry Sauce	40.00	13
4	Alice Mutton	39.00	14
5	Queso Manchego La Pastora	38.00	15

Функции оконной навигации

- Оконные функции LAG, LEAD

```
SELECT OrderID, Freight,  
       Lag (Freight) OVER (ORDER BY OrderID ASC),  
       Lag (Freight, 7) OVER (ORDER BY OrderID ASC),  
       Lag (Freight, 7, 0) OVER (ORDER BY OrderID ASC),  
       Lead (Freight) OVER (ORDER BY OrderID ASC)  
FROM Orders
```

100 % - Lag Lead

OrderID	Freight	(No column name)	(No column name)	(No column name)	(No column name)
1	10248	32.38	NULL	NULL	0.00
2	10249	11.61	32.38	NULL	0.00
3	10250	65.83	11.61	NULL	0.00
4	10251	41.34	65.83	NULL	0.00
5	10252	51.30	41.34	NULL	0.00
6	10253	58.17	51.30	NULL	0.00
7	10254	22.98	58.17	NULL	0.00
8	10255	148.33	22.98	32.38	32.38
9	10256	13.97	148.33	11.61	11.61
10	10257	81.91	13.97	65.83	65.83
11	10258	140.51	81.91	41.34	41.34
12	10259	3.25	140.51	51.30	51.30

Оконная агрегация

- Нарастающие итоги в оконной функции SUM
- Обычно SUM требует GROUP BY



The screenshot shows a SQL Server Enterprise Manager interface with three query windows open. The active window, titled 'SQLQuery34.sql - (...ор Самородов (61))', contains the following SQL query:

```
SELECT OrderID, Freight,  
       Sum (Freight)  
FROM Orders  
ORDER BY OrderID
```

The cursor is positioned at the end of the query. Below the query window, a 'Messages' pane displays the following error message:

```
State 1, line 1  
'r10' is invalid in the select list because it is not contained in either an aggregate function or the GROUP BY clause.
```

Оконная агрегация (продолжение)

- Но можно использовать в окне
- Сперва только окно с сортировкой

```
SELECT OrderID, Freight,
       Sum (Freight) OVER (
           ORDER BY OrderID ASC
       )
FROM Orders
ORDER BY OrderID
```

100 %

	OrderID	Freight	(No column name)
1	10248	32.38	32.38
2	10249	11.61	43.99
3	10250	65.83	109.82
4	10251	41.34	151.16
5	10252	51.30	202.46
6	10253	58.17	260.63
7	10254	22.98	283.61
8	10255	148.33	431.94
9	10256	13.97	445.91
10	10257	81.91	527.82
11	10258	140.51	668.33
12	10259	3.25	671.58

Оконная агрегация (скользящее окно)

- Задается выражением ROWS BETWEEN
- Верхняя граница окна неподвижна UNBOUNDED PRECEDING, а нижняя – скользит вниз CURRENT ROW

```
SELECT OrderID, Freight,
       Sum (Freight) OVER (
         ORDER BY OrderID ASC
         ROWS BETWEEN
           UNBOUNDED PRECEDING
           AND
           CURRENT ROW
       )
FROM Orders
ORDER BY OrderID
```

100 %

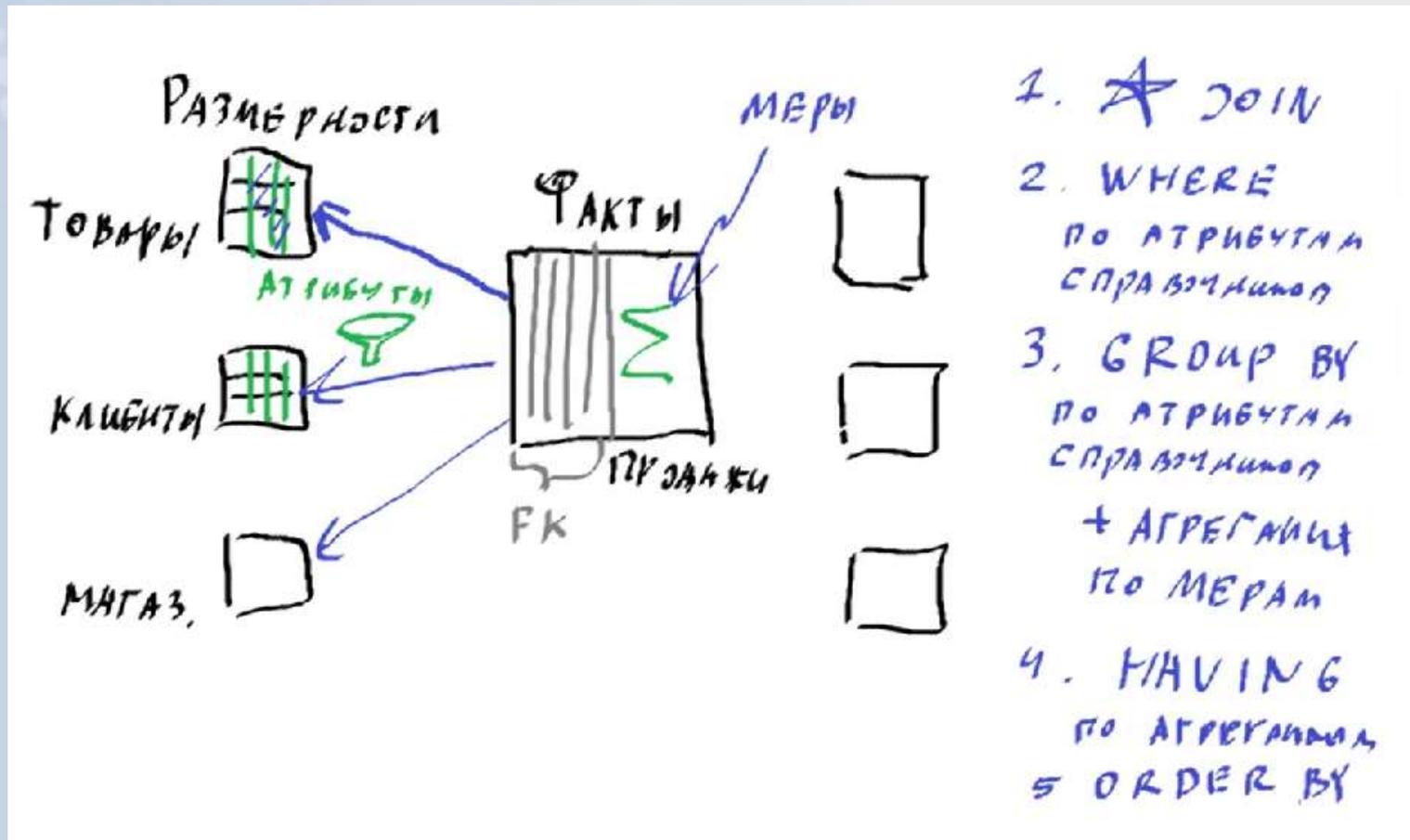
Results Messages

	OrderID	Freight	(No column name)
1	10248	32,38	32,38
2	10249	11,61	43,99
3	10250	65,83	109,82
4	10251	41,34	151,16
5	10252	51,30	202,46
6	10253	58,17	260,63
7	10254	22,98	283,61
8	10255	148,33	431,94
9	10256	13,97	445,91
10	10257	81,91	527,82
11	10258	140,51	668,33
12	10259	3,25	671,58

Хранилища и витрины данных (обзор)

- Обычно анализ выполняется не на операционных данных, а на специально подготовленных более простых БД, которые называются хранилищами или витринами.
- Пример:
AdventureWorks – операционная БД
- **AdventureWorksDW** – БД хранилища с меньшим кол-вом таблиц и более простой структурой (звезда)
- Но потребуются много JOIN –выражений
- Таблица фактов – таблица продаж – часто модифицируется
Другие таблицы – это справочники (таблицы измерений) – они редко изменяются

Схема хранилища - звезда



Пример работы с хранилищем

- JOIN - при работе с хранилищем – предсказуемые
- Типовой запрос из БД хранилища:

Показать продажи по месяцам, покупательницам в зависимости от рода занятий

```
SELECT EnglishMonthName,
EnglishOccupation,
Sum (SalesAmount)
FROM FactInternetSales S
INNER JOIN DimCustomer C
ON S.CustomerKey = C.CustomerKey
INNER JOIN DimDate D
ON S.OrderDateKey = D.DateKey
WHERE D.CalendarYear = 2011
AND C.Gender = 'F'
GROUP BY EnglishMonthName,
EnglishOccupation
```

	EnglishMonthName	EnglishOccupation	(No column name)
1	September	Clerical	46231,451
2	November	Clerical	55128,7256
3	January	Management	32725,2482
4	June	Professional	115075,6946
5	January	Professional	88312,07
6	January	Manual	25047,89
7	May	Management	49689,22
8	May	Clerical	54690,6864

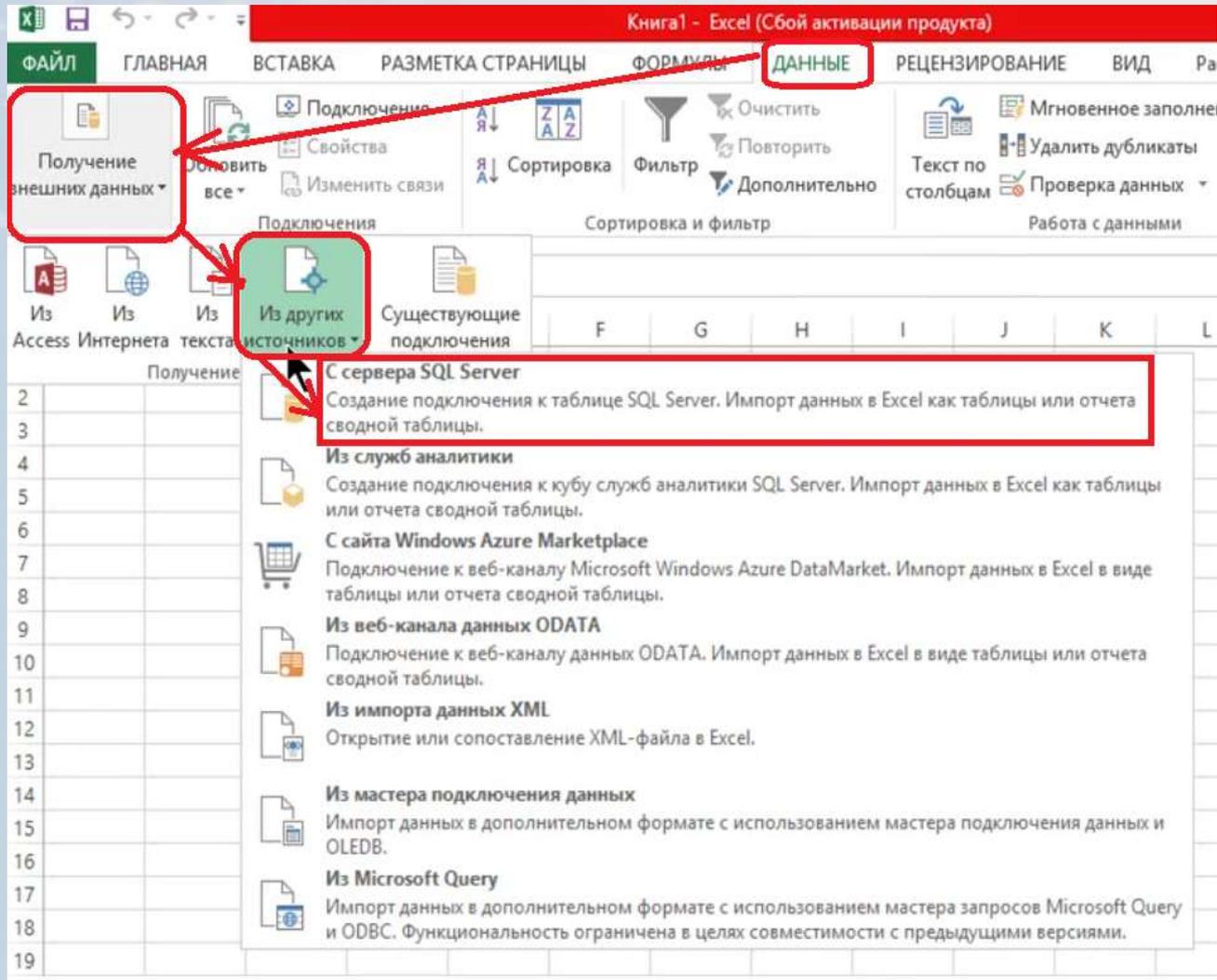
Query executed successfully. (local) (12.0 RTM) CKO-ALM4NO8U701\Фёдор ... AdventureWorksDW2014 00:00:00 60 rows

Доступ к данным из Excel

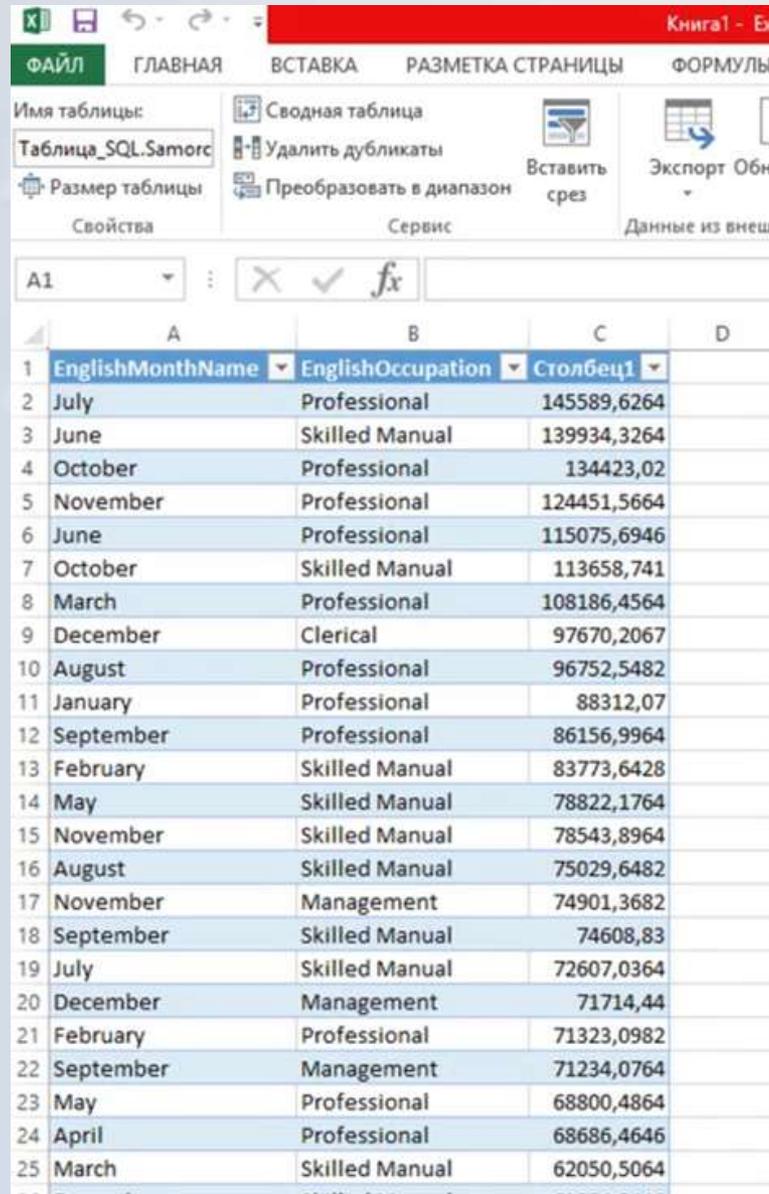
Как обращаться из Excel (4-способа):

- ✓ PowerQuery
- ✓ PowerPivot
- ✓ VBA
- ✓ Древний способ, получения в листе Excel данных из внешнего источника, в качестве которого выступит SQL Server

Как настроить



Итог



The screenshot shows the Microsoft Excel interface with a pivot table. The pivot table has three columns: 'EnglishMonthName', 'EnglishOccupation', and 'Столбец1'. The data is sorted by the 'EnglishMonthName' column. The numerical values in the 'Столбец1' column range from 62050,5064 to 145589,6264.

	A	B	C	D
1	EnglishMonthName	EnglishOccupation	Столбец1	
2	July	Professional	145589,6264	
3	June	Skilled Manual	139934,3264	
4	October	Professional	134423,02	
5	November	Professional	124451,5664	
6	June	Professional	115075,6946	
7	October	Skilled Manual	113658,741	
8	March	Professional	108186,4564	
9	December	Clerical	97670,2067	
10	August	Professional	96752,5482	
11	January	Professional	88312,07	
12	September	Professional	86156,9964	
13	February	Skilled Manual	83773,6428	
14	May	Skilled Manual	78822,1764	
15	November	Skilled Manual	78543,8964	
16	August	Skilled Manual	75029,6482	
17	November	Management	74901,3682	
18	September	Skilled Manual	74608,83	
19	July	Skilled Manual	72607,0364	
20	December	Management	71714,44	
21	February	Professional	71323,0982	
22	September	Management	71234,0764	
23	May	Professional	68800,4864	
24	April	Professional	68686,4646	
25	March	Skilled Manual	62050,5064	

Контрольная работа

-- Как зовут продавцов, которые оформили заказов в Берлин больше, чем в Париж?

Вопросы?