

Профессиональная программа “Анализ данных на языке SQL”

Блок 1. Реляционные базы данных СУБД SQL Server

Ваш лектор профессиональной программы

- Смирнов Михаил Вячеславович
- Кандидат экономических наук
- Текущая предметная область: технологии хранения и защиты данных, администрирование баз данных
- Контакты:



Электронная почта: smirnovmgupi@gmail.com



Образовательный сайт: <http://msuniversity.ru>

Содержание блока

- Язык запросов SQL
- ПО для работы с СУБД MS SQL Server (SSMS)
- Базы данных и таблицы
- SQL запросы: написание и исполнение
- Выгрузка данных из Excel

В конце блока предусмотрено решение ряда практических задач.

1-ый учебный вопрос

Язык запросов SQL

Общие сведения: язык запросов SQL



Формальный непроцедурный язык программирования, применяемый для создания, модификации и управления данными в произвольной реляционной базе данных, управляемой соответствующей системой управления базами данных (СУБД). SQL основывается на исчислении кортежей.

Изначально SQL был основным способом работы пользователя с базой данных и позволял выполнять следующий набор операций: создание в базе данных новой таблицы, добавление в таблицу новых записей, изменение записей, удаление записей, выборка записей из одной или нескольких таблиц (в соответствии с заданным условием), изменение структур таблиц.

Со временем SQL усложнился — обогатился новыми конструкциями, обеспечил возможность описания и управления новыми хранимыми объектами (например, индексы, представления, триггеры и хранимые процедуры)...

Группы команд языка SQL (операторы)

Операторы определения данных
(*Data Definition Language, DDL*)

Операторы манипуляции данными
(*Data Manipulation Language, DML*)

Операторы определения доступа к
данным (*Data Control Language, DCL*)

Операторы управления транзакциями
(*Transaction Control Language, TCL*)

Операторы определения данных (DDL)

Используются в СУБД для создания и изменения структуры базы данных и ее составных частей - таблиц, индексов, представлений (виртуальных таблиц), а также триггеров и сохраненных процедур.

Группа CREATE

- CREATE database
- CREATE table
- CREATE view
- CREATE index
- CREATE trigger
- CREATE procedure

Группа ALTER

- ALTER database
- ALTER table
- ALTER view
- ALTER index
- ALTER trigger
- ALTER procedure

Группа DROP

- DROP database
- DROP table
- DROP view
- DROP index
- DROP trigger
- DROP procedure

Типовой скрипт создания БД.

```
CREATE DATABASE test_createdb
ON
( NAME = crdb_dat,
  FILENAME = 'C:\Program Files\Microsoft SQL
Server\MSSQL13.FORLAB\MSSQL\DATA\crdb_dat.mdf',
  SIZE = 30MB, MAXSIZE = 50MB,
  FILEGROWTH = 20%
)
LOG ON
( NAME = crdb_log,
  FILENAME = 'C:\Program Files\Microsoft SQL
Server\MSSQL13.FORLAB\MSSQL\DATA\crdb_log.ldf',
  SIZE = 5MB,
  MAXSIZE = UNLIMITED,
  FILEGROWTH = 2MB
);
```

Операторы манипулирования данными (DML)

Используются, для манипулирования данными в таблицах баз данных.

SELECT считывает данные, удовлетворяющие заданным условиям.

INSERT добавляет новые данные.

UPDATE изменяет существующие данные.

DELETE удаляет данные.

Операторы определения доступа к данным (DCL)

Используются, для манипулирования данными в таблицах баз данных.

SELECT считывает данные, удовлетворяющие заданным условиям.

INSERT добавляет новые данные.

UPDATE изменяет существующие данные.

DELETE удаляет данные.

Типовые скрипты группы GRANT/REVOKE.

```
GRANT INSERT ON dbo.paintings TO test;
```

```
GRANT SELECT ON dbo.paintings TO test;
```

```
REVOKE INSERT ON dbo.paintings TO test AS dbo;
```

```
REVOKE SELECT ON dbo.paintings TO test AS dbo;
```

Операторы управления транзакциями (TCL)

Часть SQL, используемая для обработки транзакций.

COMMIT применяет транзакцию.

ROLLBACK откатывает все изменения, сделанные в контексте текущей транзакции.

SAVEPOINT делит транзакцию на более мелкие участки.

Типовой скрипт транзакции.

```
BEGIN TRANSACTION
INSERT INTO work VALUES
(001, 'Севастополь', 'На картине изображена бухта города в утренней дымке',
001),
(002, 'Шхуна под парусами', 'На картине изображена шхуна с флагом', 001),
(003, 'Ялта. Горы ночью', 'Группа людей прогуливаются вдоль берега Крыма',
001);
IF (@@error <> 0)
ROLLBACK
ELSE
COMMIT;
```

2-ой учебный вопрос

Программное обеспечение для
взаимодействия с СУБД. MS
SQL Server Management Studio.

Общие сведения: принцип передачи команд в СУБД

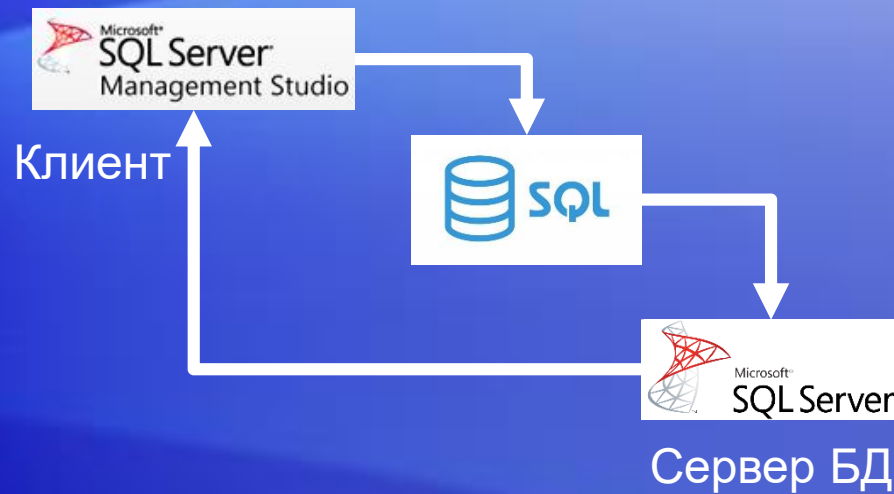


База данных – емкость мультиварки. В ней готовятся данные (или хранятся до готовки).

СУБД (система управления базами данных) – это кнопки вокруг дисплея, которые позволяют настраивать режим приготовления данных.

Воздействовать на СУБД можно командами языка SQL. Написать эти команды и передать СУБД можно в специальных программных средствах – универсальных (как, например, ErWin или DBeaver) или же специализированных (как Microsoft SQL Server Management Studio)

Упрощенный принцип работы ПО БД Microsoft

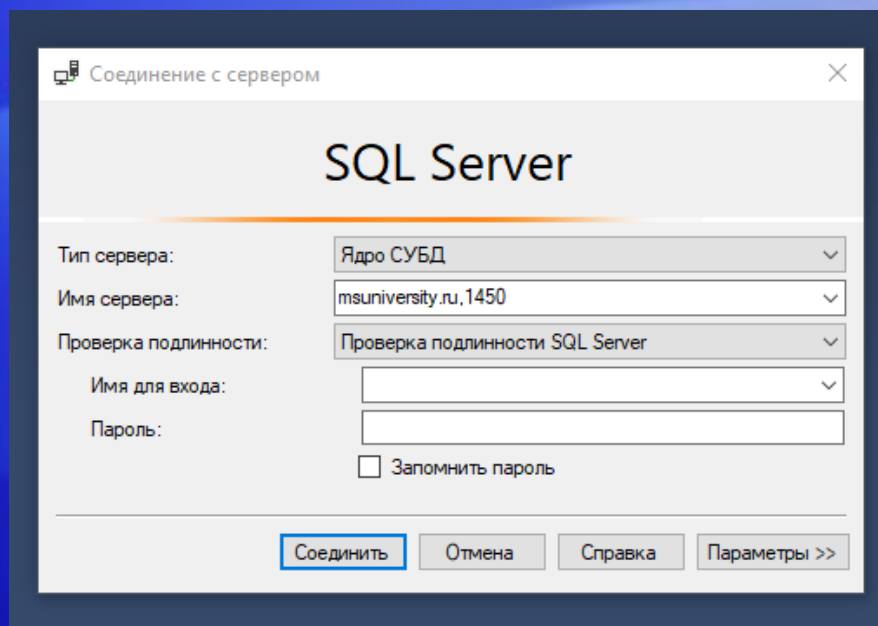


Клиент набирает в Management Studio код в языке SQL.

Management Studio передает код на локальный или выделенный сервер БД

Сервер БД исполняет или не исполняет SQL код и результат исполнения или ошибки возвращается в Management Studio.

Соединение с сервером SQL Server

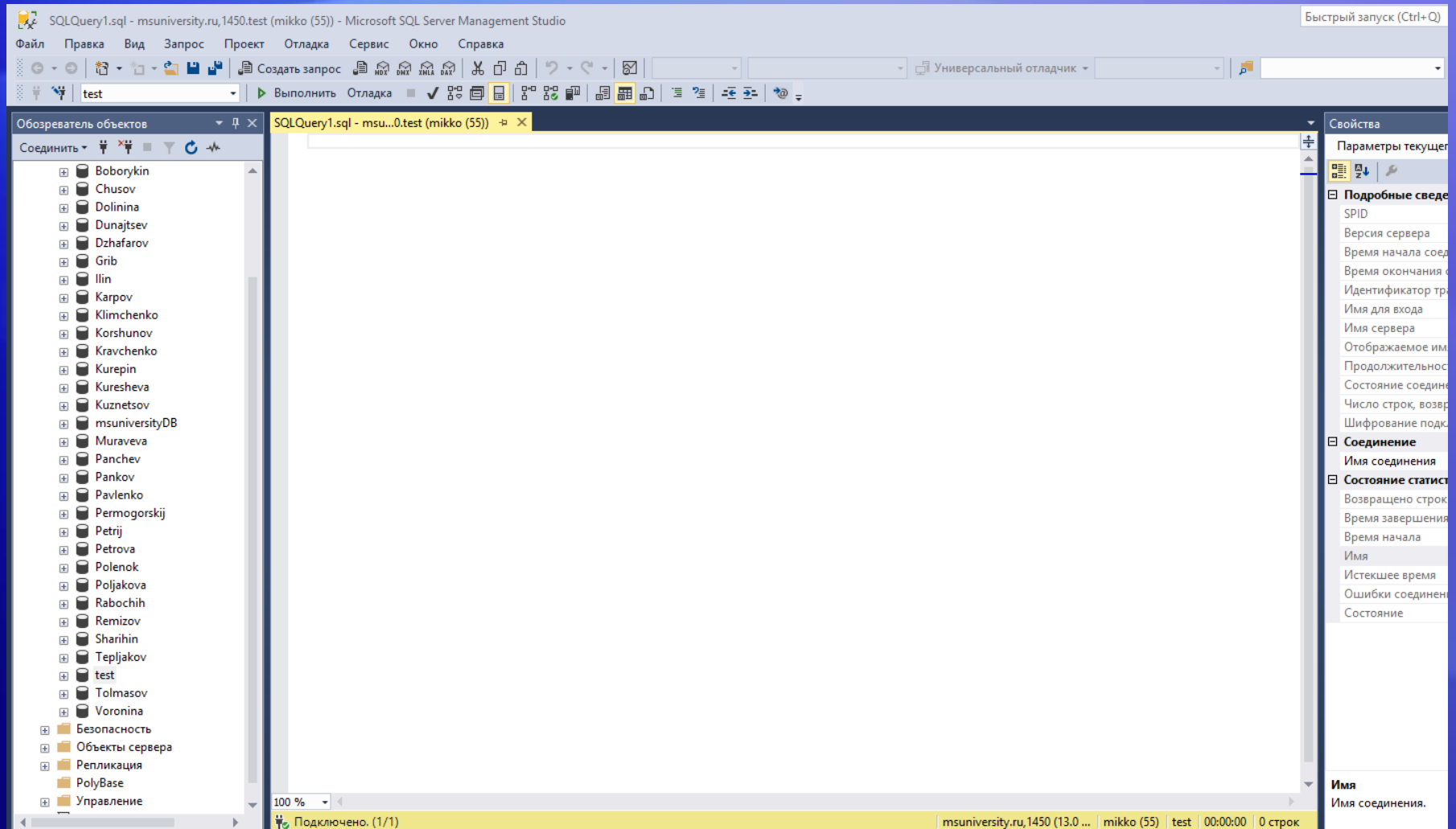


Ядро СУБД –
инструментарий
исполнения SQL запросов
всех групп (есть еще AS,
RS, IS, у этих
компонентов свои
функции).

Имя сервера может быть
прописано в любом
доступном виде.

Проверок подлинности несколько – внутренний
пользователь SQL Server, пользователь OS Windows,
группы Active Directory.

Рабочее окно MS SQL Server Management Studio



3-ий учебный вопрос

Базы данных и таблицы

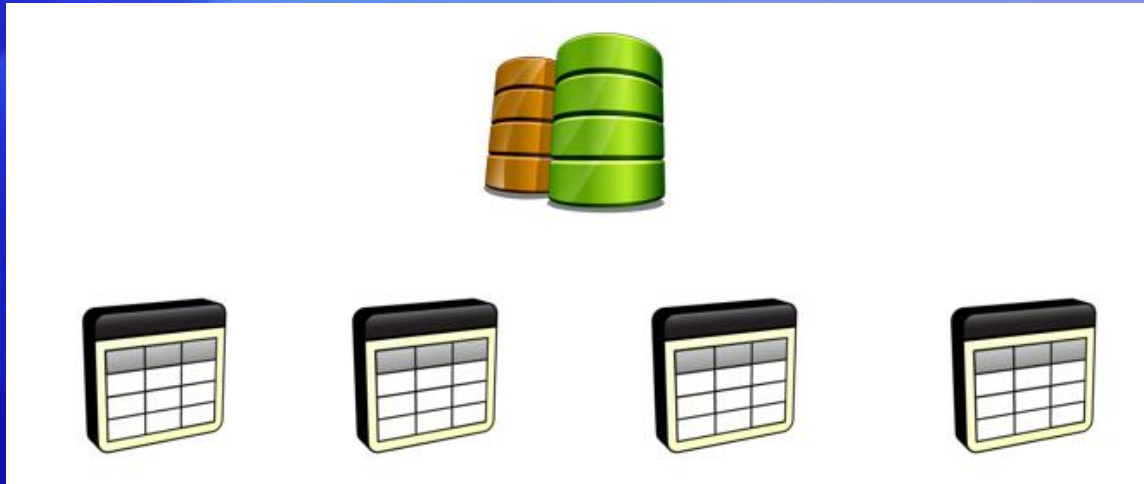
Принципы взаимодействия с базами данных



На рисунке показано несколько типовых способов реализации функционирования пользовательского приложения баз данных

Конечным пунктом движения информации (если она добавляется или изменяется), или запроса (если нужно получить выборку данных или аналитику) является **база данных** - совокупность данных, хранимых в соответствии со схемой данных, манипулирование которыми выполняют в соответствии с правилами средств моделирования данных.

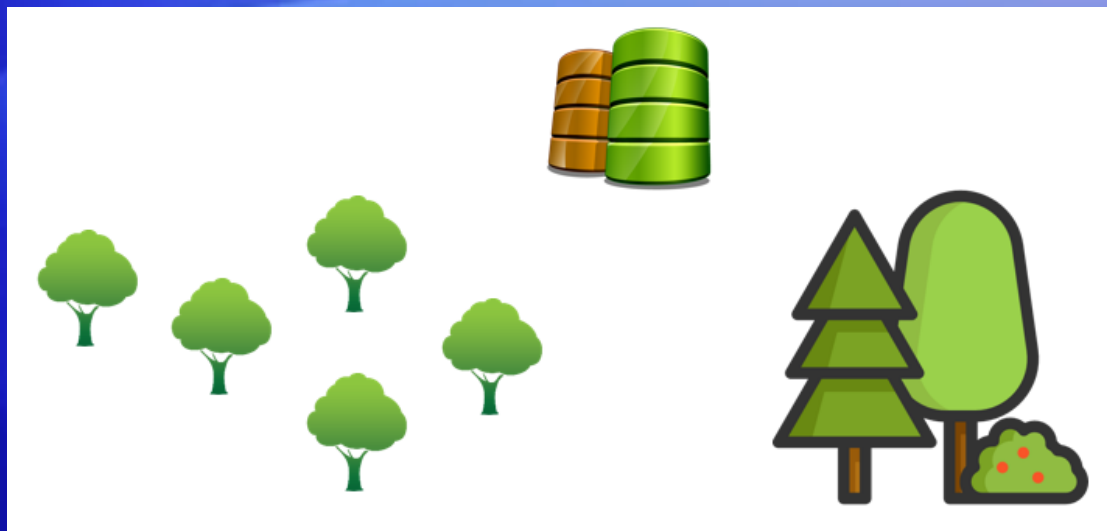
Структура реляционной базы данных



Если база данных построена в рамках реляционной модели данных (на настоящий момент времени – наиболее популярная модель), то ее ключевым элементом являются таблицы.

Реляционная база данных – это коллекция таблиц, организованная согласно правилам реляционной модели. Каждая ячейка этих таблиц имеет соответствующее формальное описание.

Структура базы данных noSQL



В данной модели данных место таблиц занимают документы.

Документы могут быть деревьями, а могут быть лесом.

Если объект, аккумулирующий на ветках данные один (экземпляр), то принято говорить о дереве.

Если объектов несколько (группа экземпляров), то говорят о лесе.

Таблица реляционной базы данных

отношение

первичный ключ

атрибуты

СОТР_НОМЕР	СОТР_ИМЯ	СОТР_ЗАРПЛ	СОТР_ОТД_НОМЕР
2934	Иванов	112,000	310
2935	Петров	144,000	310
2936	Сидоров	92,000	313
2937	Федоров	110,000	310

кортежи

Выше приведена логическая модель типовой таблицы реляционной модели данных. В физической модели используется иная терминология: таблица, индекс первичного ключа, столбцы, строки, поля.

4-ий учебный вопрос

Написание и исполнение
запросов

Общие сведения: запрос SQL



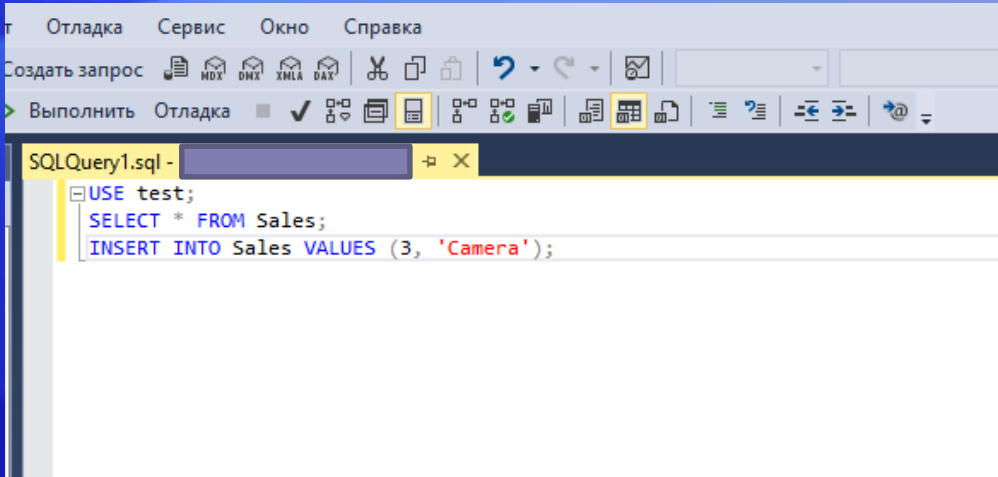
Команда, написанная на языке запросов SQL. Представляет собой совокупность операторов, инструкций, вычисляемых функций.

Структура каждого запроса в целом статична и в типовых случаях изменению при написании подлежат пользовательские конструкции (названия таблиц, индексов, столбцов и т.д.).

Отдельно стоит выделить группу запросов, связанных с управлением хранимыми процедурами и триггерами. Их можно назвать своего рода “микропрограммами”, так как они в теле могут содержать элементы задания условий, переменных и циклов.

Типовые запросы SQL

На рисунке
последовательно
показаны три
запроса языка SQL

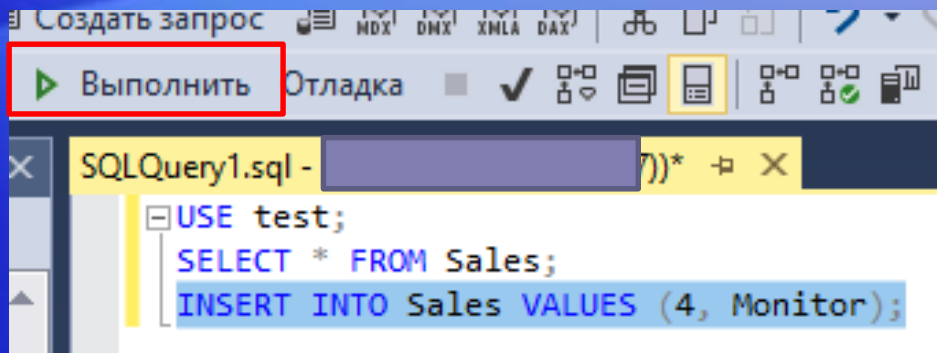


```
SQLQuery1.sql - [X]  
USE test;  
SELECT * FROM Sales;  
INSERT INTO Sales VALUES (3, 'Camera');
```

Конструкция USE указывает на существующую БД, в которой мы собираемся оперировать данными. Конструкция SELECT позволяет по заданным условиям сформировать выборку данных из таблицы или таблиц. Конструкция INSERT INTO позволяет добавить строку или строки с данными в таблицу.

Действия при запуске запроса на исполнение

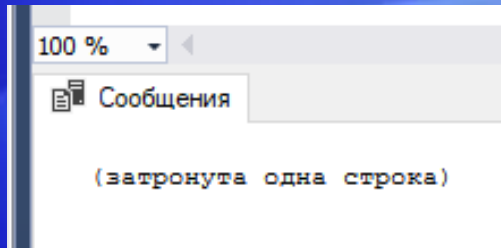
На рисунке показаны действия при выполнении запроса в среде MS SQL Server Management Studio



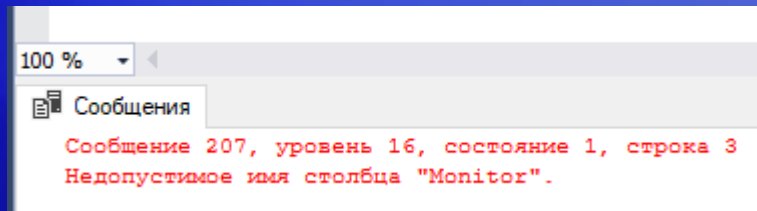
Выделяется законченный фрагмент листинга запросов от начального оператора, до символа “точка с запятой”.

Нажатие кнопки выполнить. Результаты запроса будут варьироваться от того оператора, который был в нем использован.

Результат исполнения запроса



На рисунке показан результат выполнения запроса без ошибки и с ошибкой.



Так как на исполнение был запущен запрос на добавление строки в таблицу, результатом его выполнения стало системное сообщение об успешно выполненной операции.

В случае ошибки при запуске запроса, в результатах будет выведена подробная информация об ошибке. Запись, соответственно добавлена не будет.

5-ый учебный вопрос

Доступ к данным из Excel

Общие сведения: импорт данных из Excel



Процедура переноса данных из таблицы табличного редактора Excel в базу данных MS SQL Server



Перенос данных из таблицы Excel в целом отражает способ переноса данных и из других источников. Мы рассмотрим два способа – один будет реализован через специальный оператор языка Transact-SQL, а второй – с помощью одного из мастеров MS SQL Server.

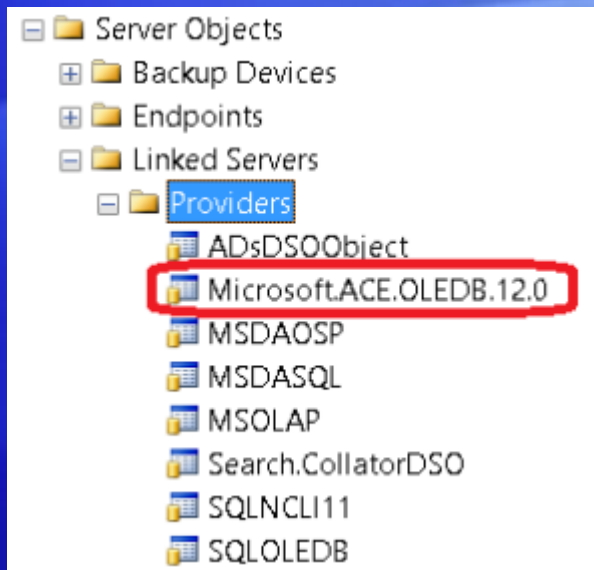
Конфигурация SQL Server для распределенного запроса

На рисунке показаны этапы изменения конфигурации SQL Server.

```
sp_configure 'show advanced options', 1;  
RECONFIGURE;  
GO  
  
sp_configure 'Ad Hoc Distributed Queries', 1;  
RECONFIGURE;  
GO
```

Будьте внимательны, прибегая к такой вариации импорта данных. Убедитесь, что политика безопасности организации допускает такой способ реконфигурации!

Проверка наличия необходимого сервиса



Проверяем на сервере наличие поставщика услуг (протокола), необходимого для выбранного вида импорта.

По умолчанию этот поставщик услуг не установлен. Если вы намереваетесь осуществить этот вид импорта, данную службу необходимо заранее установить на сервер БД.

Импорт данных из файла Excel

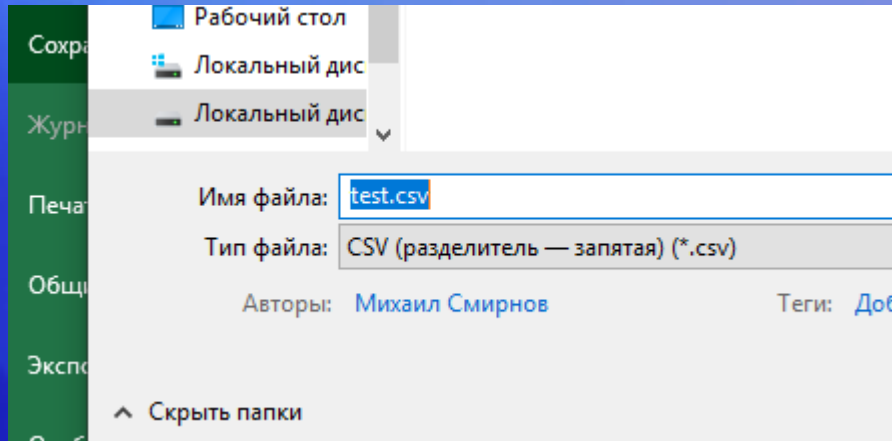
Набираем команду
для импорта данных
из файла Excel

```
USE ImportFromExcel;  
GO  
SELECT * INTO Data_dq  
FROM OPENROWSET('Microsoft.ACE.OLEDB.12.0',  
    'Excel 12.0; Database=C:\Temp\Data.xlsx', [Sheet1$]);  
GO
```

Необходимо обязательно закреть Excel таблицу на которую
ссылается запрос, перед его выполнением!

Импорт неструктурированных файлов

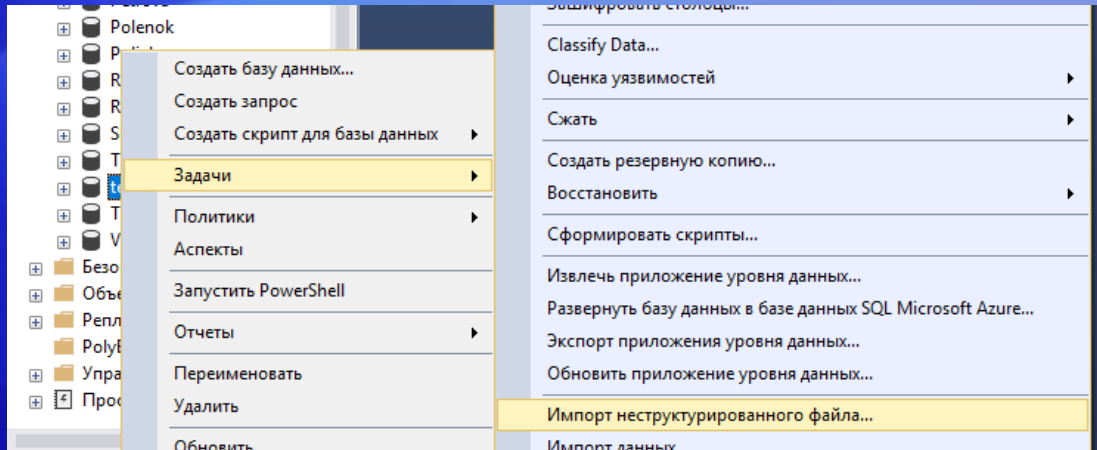
На рисунке показано, как из таблицы Excel сделать неструктурированный файл.



Таблицу Excel необходимо сохранить в файл формата *.csv или *.txt с разделителями.

Мастер импорта неструктурированных файлов

На рисунке показано, как вызвать мастер импорта неструктурированного файла.



Дальнейшие действия и работа с мастером будет продемонстрирована на практической части этого блока.