Работа 1

Получение статистики файла реализовано в файлах пакета company. Сначала запрашивается весь текст в файле, затем ведется посимвольный анализ, игнорируя специальные символы.

Работа 2

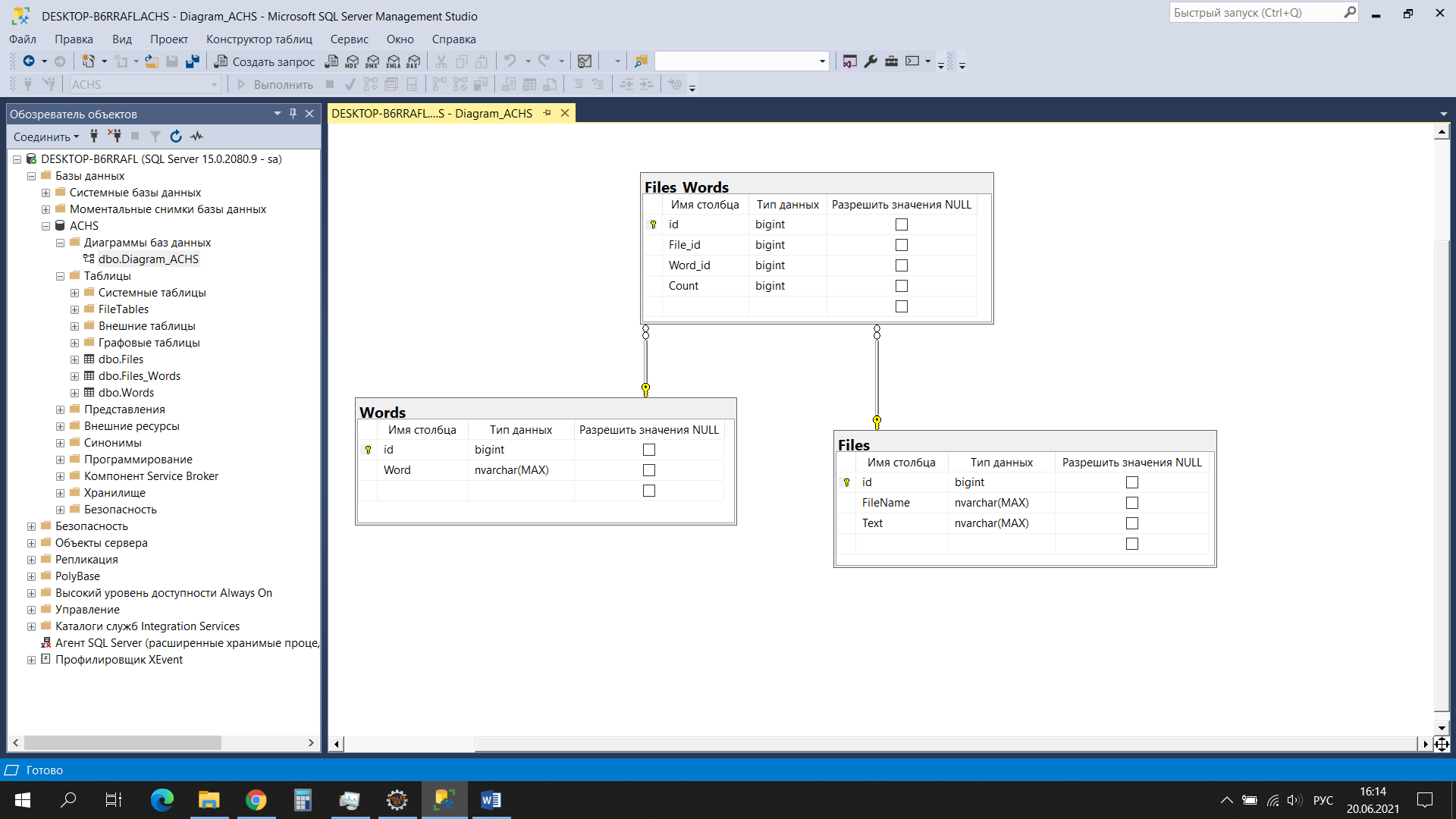
Анализ указанной директории также реализован в файлах пакета company. Работает по принципу получить список полных путей файлов, хранящихся в директории, а затем получить статистику по каждому из них.

Работа 3

Многопоточность реализована в пакете Parallel, которая применяется в сервлете UpdateDB. Работает по принципу поделить для потоков поровну файлы в директории (если есть остаток, то он достается последнему потоку), тем самым каждый поток обрабатывает свою порцию файлов и создает свою статистику. Впоследствии все эти мини-статистики объединяются в 1 большую как конечный результат.

Работа 4

Для работы используется СУБД MSSQL Server. Схема бд выглядит следующим образом:



Запросы к бд реализованы в пакете sqlwork, которые используются в сервлетах.

Работа 5

В процессе работы были созданы сервлеты со следующим функционалом:

1. Главная страница
2. Статистика отдельного файла
3. Текст отдельного файла
4. Получение списка файлов с указанным словом, а также с количеством их повторений
5. Обновление бд