**Результат работы.**

1. Копируем информацию из репозитория с образом Hadoop в заранее созданную папку с помощью Git Bash:

shiva@PC-Tatiana MINGW64 /c/Projects/4\_3

**$ git clone git@github.com:tech4242/docker-hadoop-hive-parquet.git**

2. Для развертывания контейнера необходимо в командной строке войти в скачанную папку docker-hadoop-hive-parquet и выполнить команду

**docker-compose up**

3. После поднятия контейнера попадаем в HUE по адресу http://localhost:8888/hue

Для авторизации использовался логин *shiva*

Личная папка в файловой системе Hadoop(hdfs)  */ user/ shiva*

4. C ресурса <https://all-the-books.ru/authors/tolstoy-lev-nikolaevich/> скачаны все доступные тома произведения «Война и мир» Л.Н. Толстого:

*"C:\Users\shiva\Downloads\voyna-i-mir-tom-1.txt"*

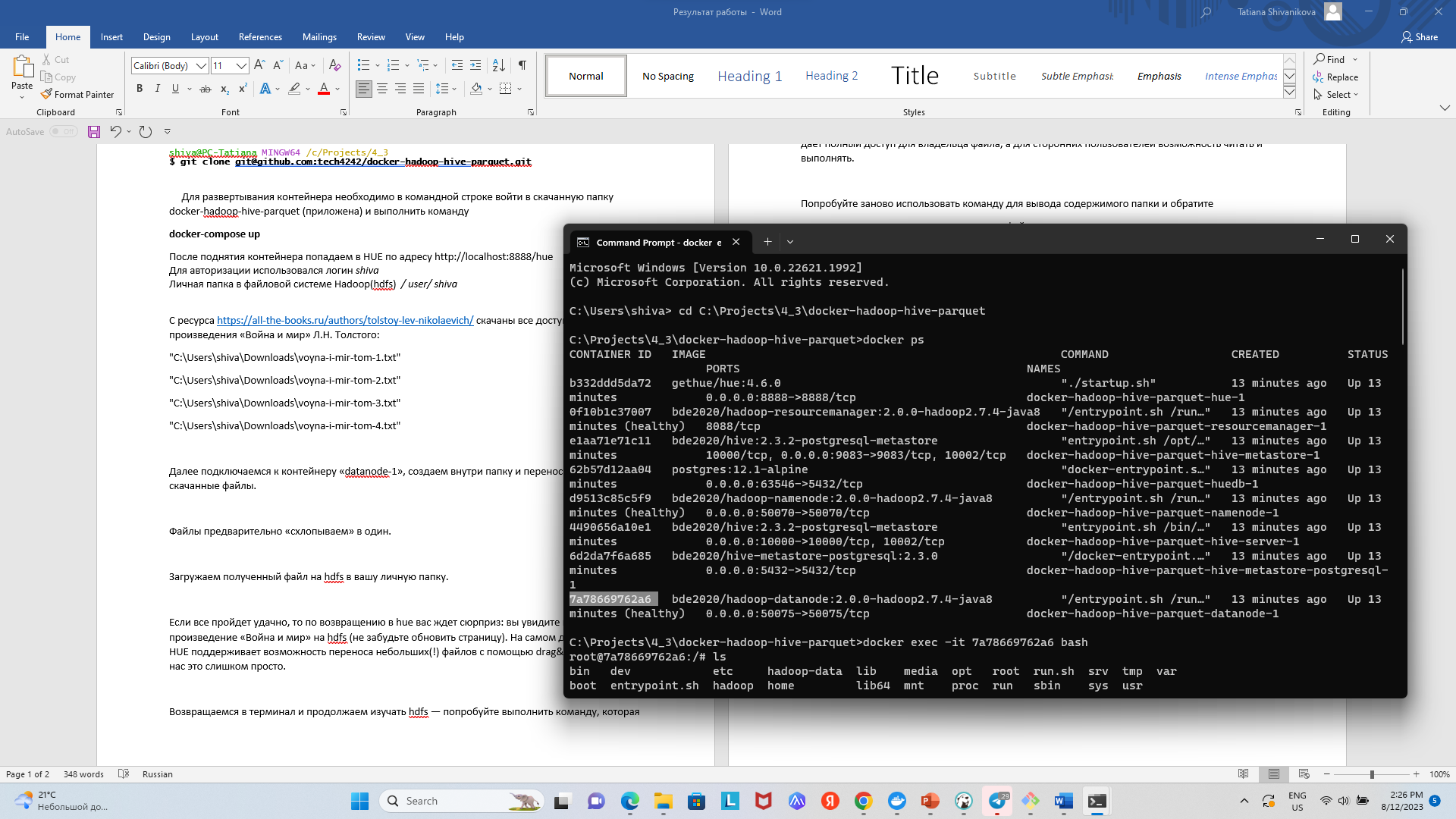
*"C:\Users\shiva\Downloads\voyna-i-mir-tom-2.txt"*

*"C:\Users\shiva\Downloads\voyna-i-mir-tom-3.txt"*

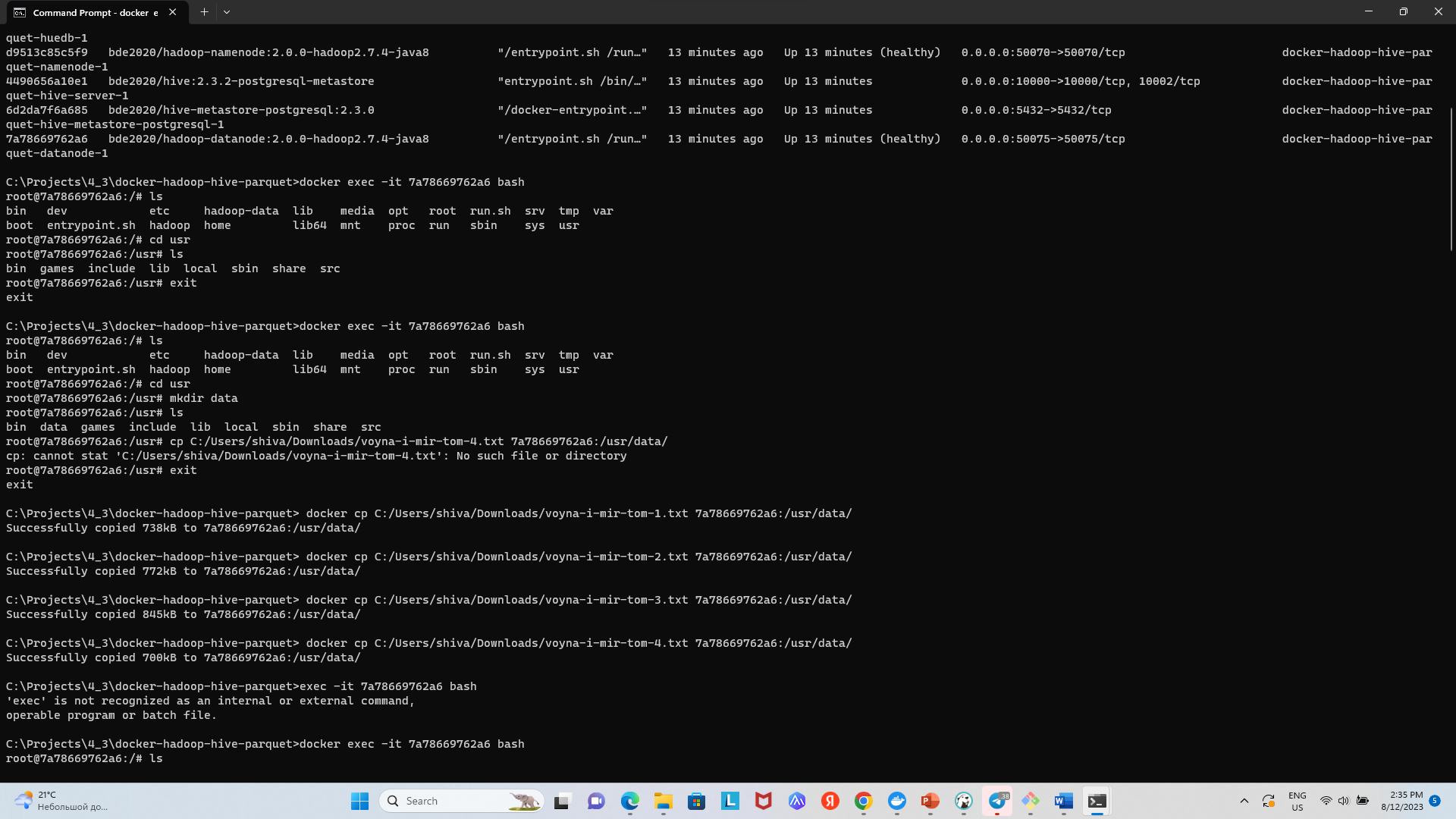
*"C:\Users\shiva\Downloads\voyna-i-mir-tom-4.txt"*

5. Далее подключаемся к контейнеру «datanode-1», создаем внутри папку и переносим в нее скачанные файлы.

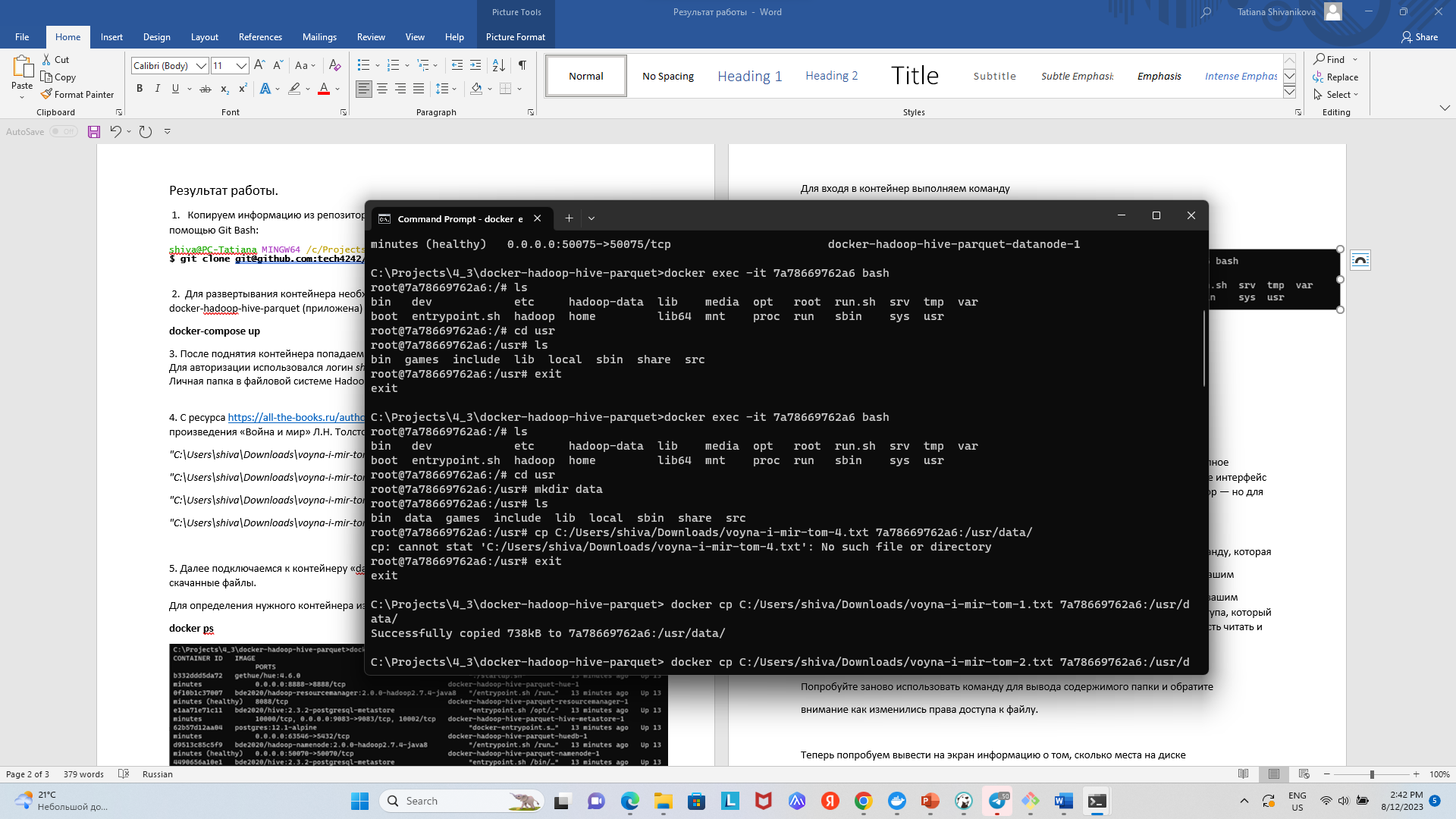
Для определения нужного контейнера из папки, где развернут Hadoop выполняем команду



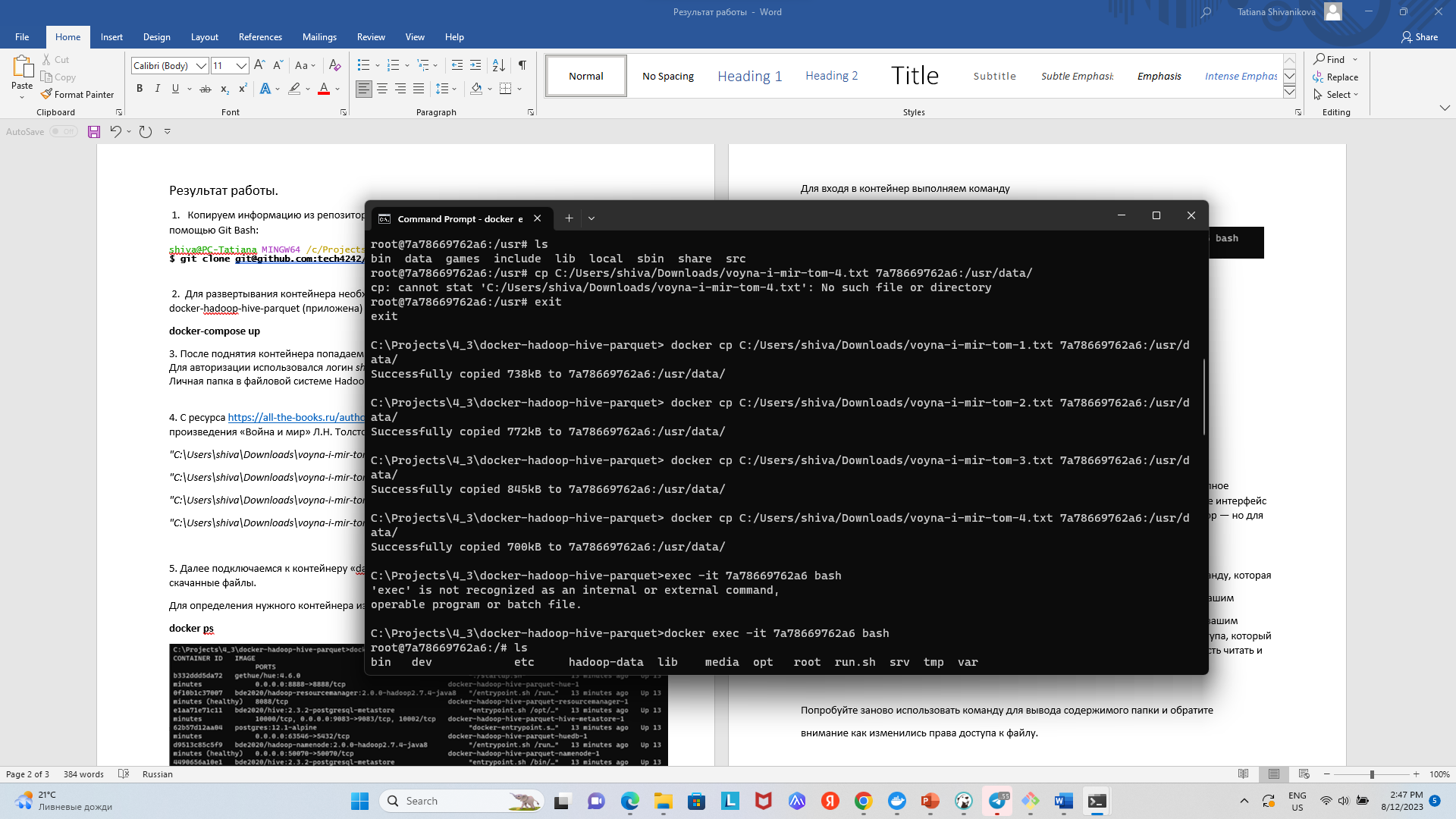
Для входя в контейнер выполняем команду



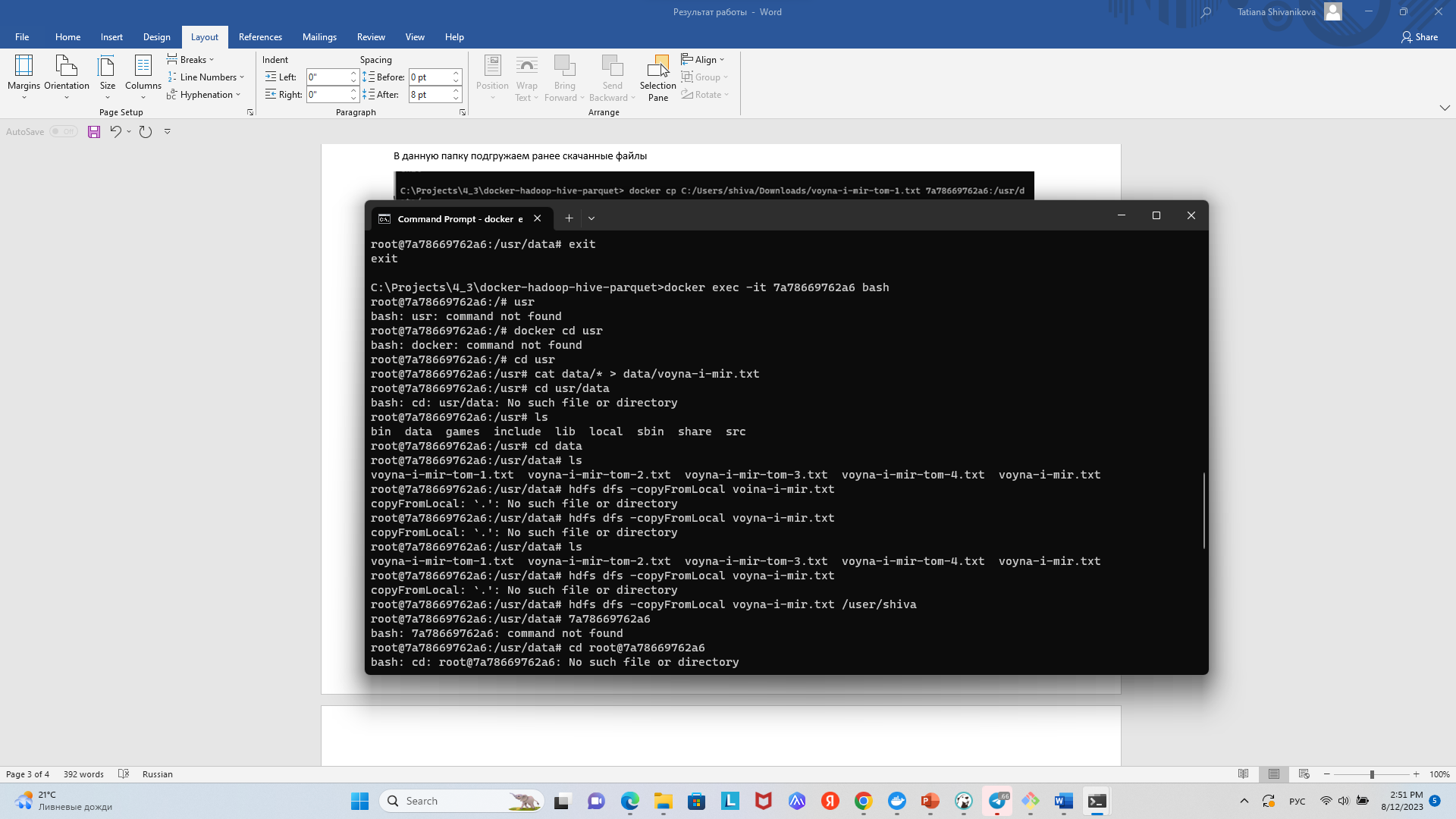
В существующей папке usr cоздаем папку data



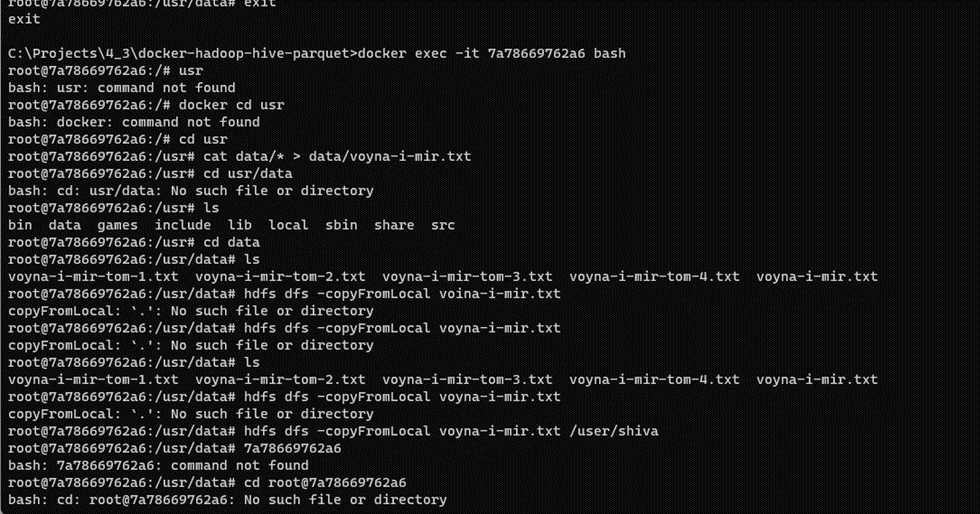
В данную папку подгружаем ранее скачанные файлы



Проверяем наличиые файлов командой



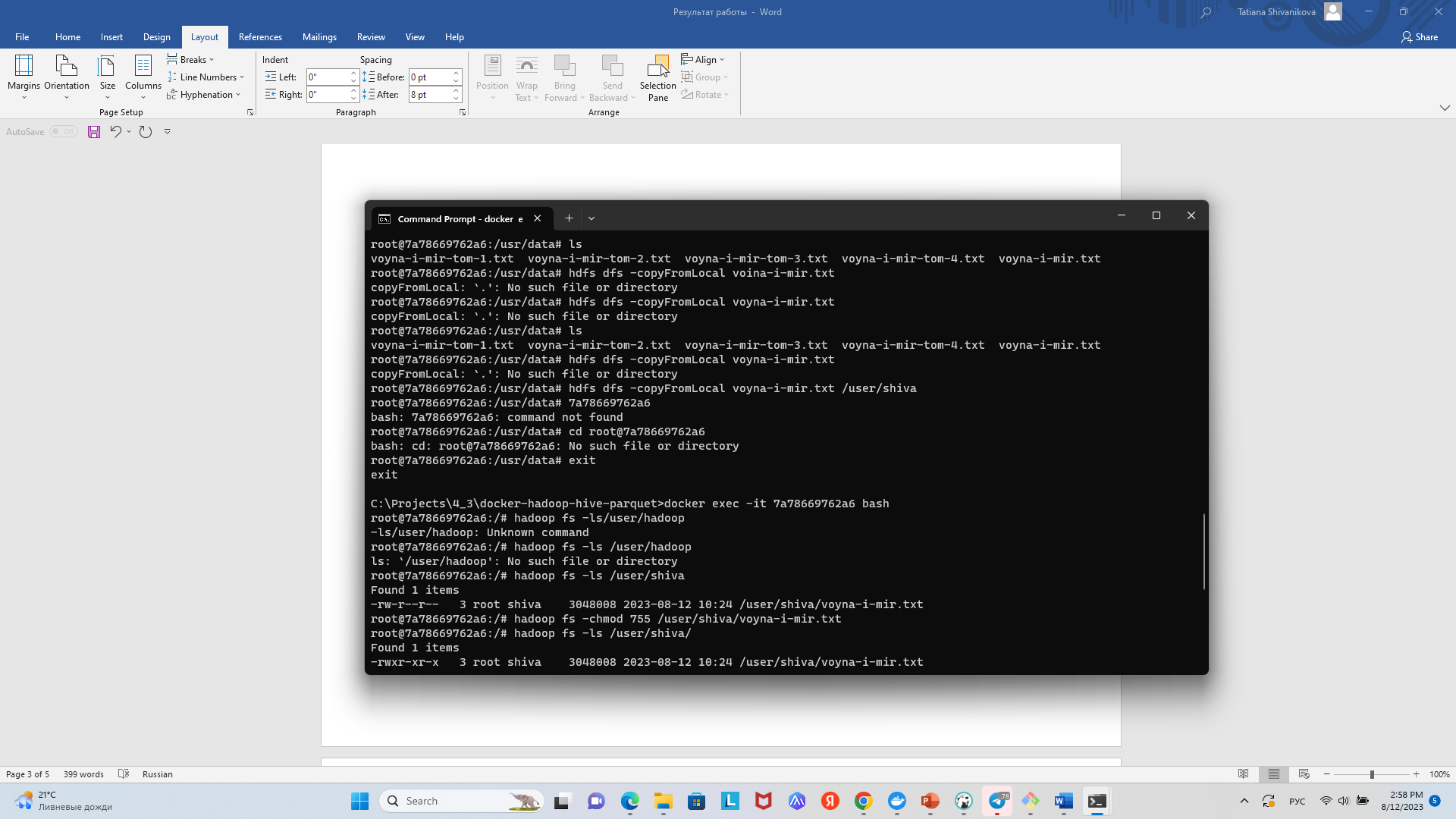
6. Файлы предварительно «схлопываем» в один



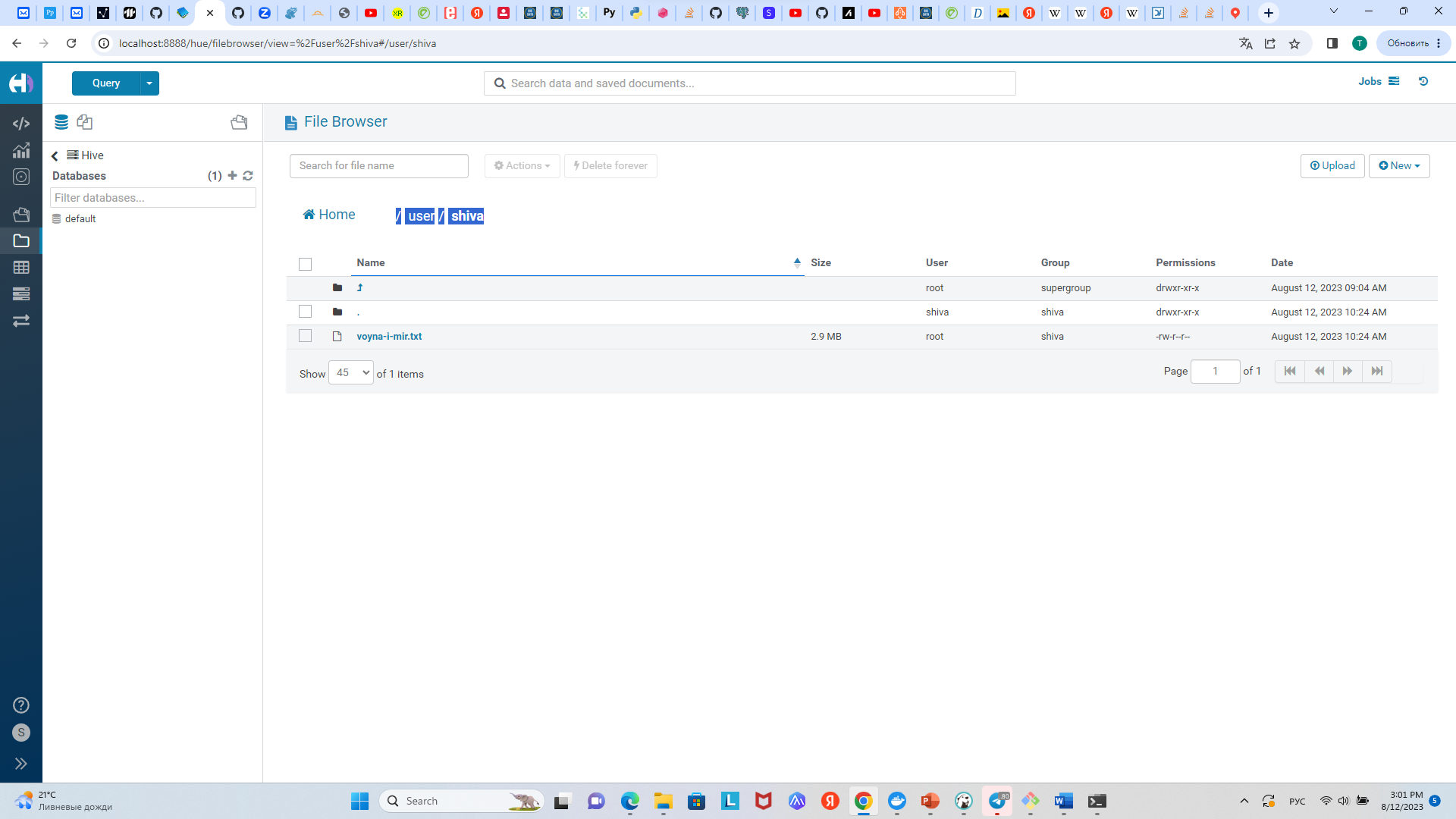
Проверяем наличие файлов



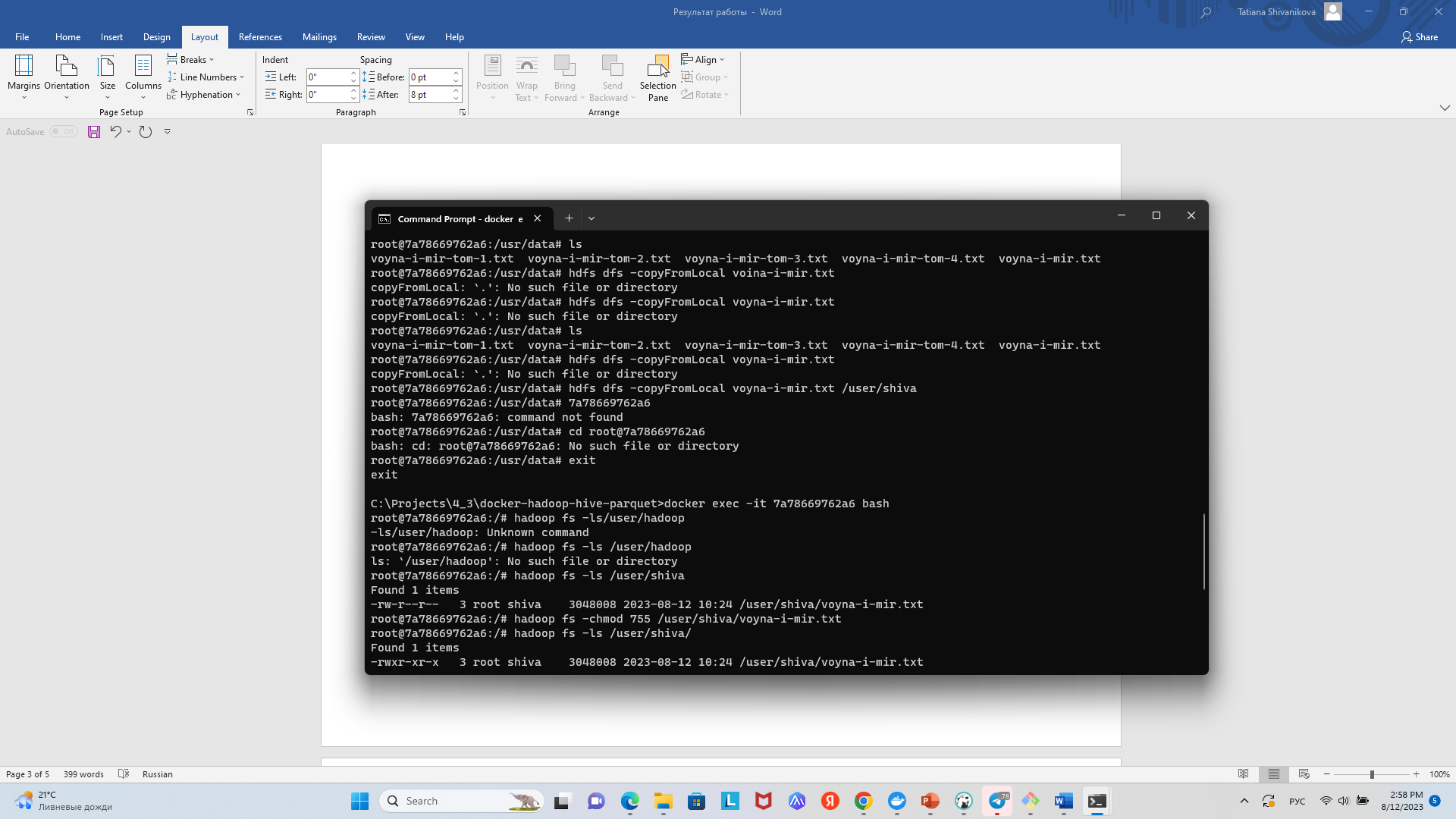
7. Загружаем полученный файл на hdfs в вашу личную папку.



Отображение файла в HUE

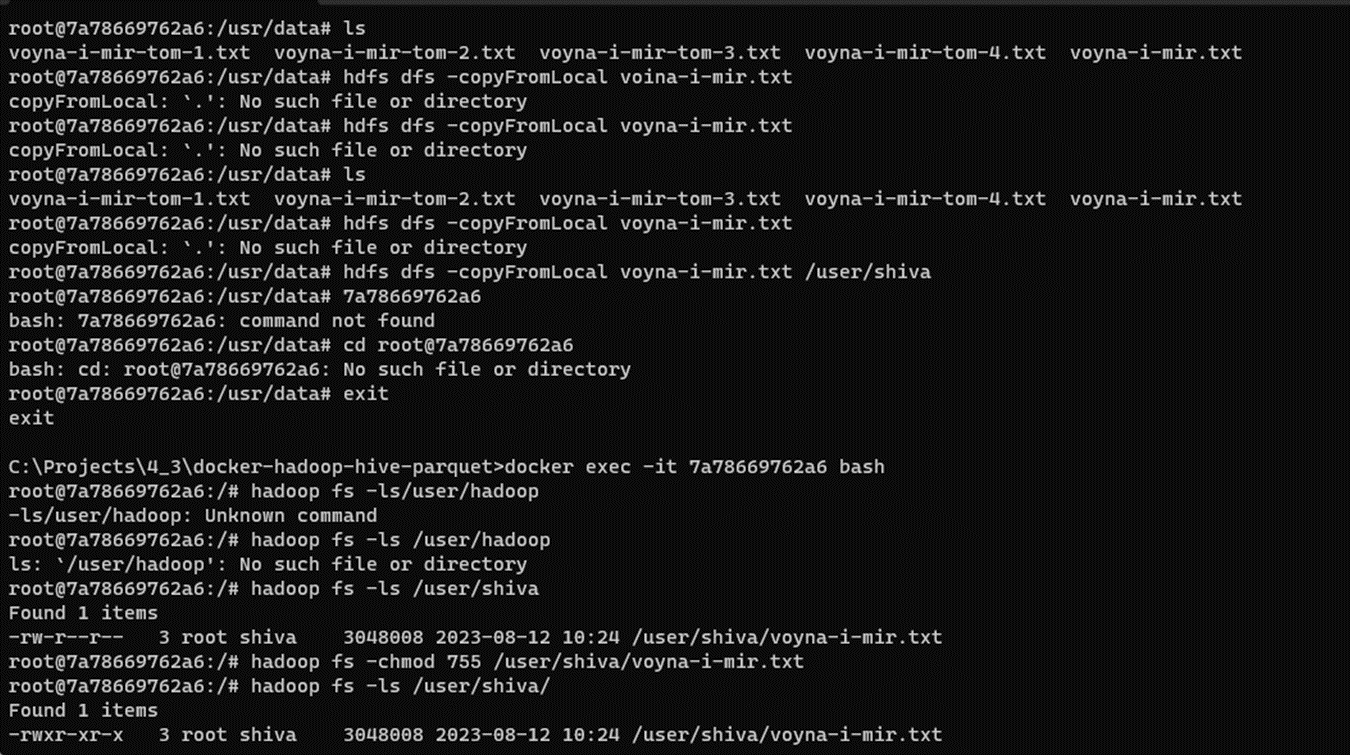


8. Попробуйте выполнить команду, которая выводит содержимое вашей личной папки.



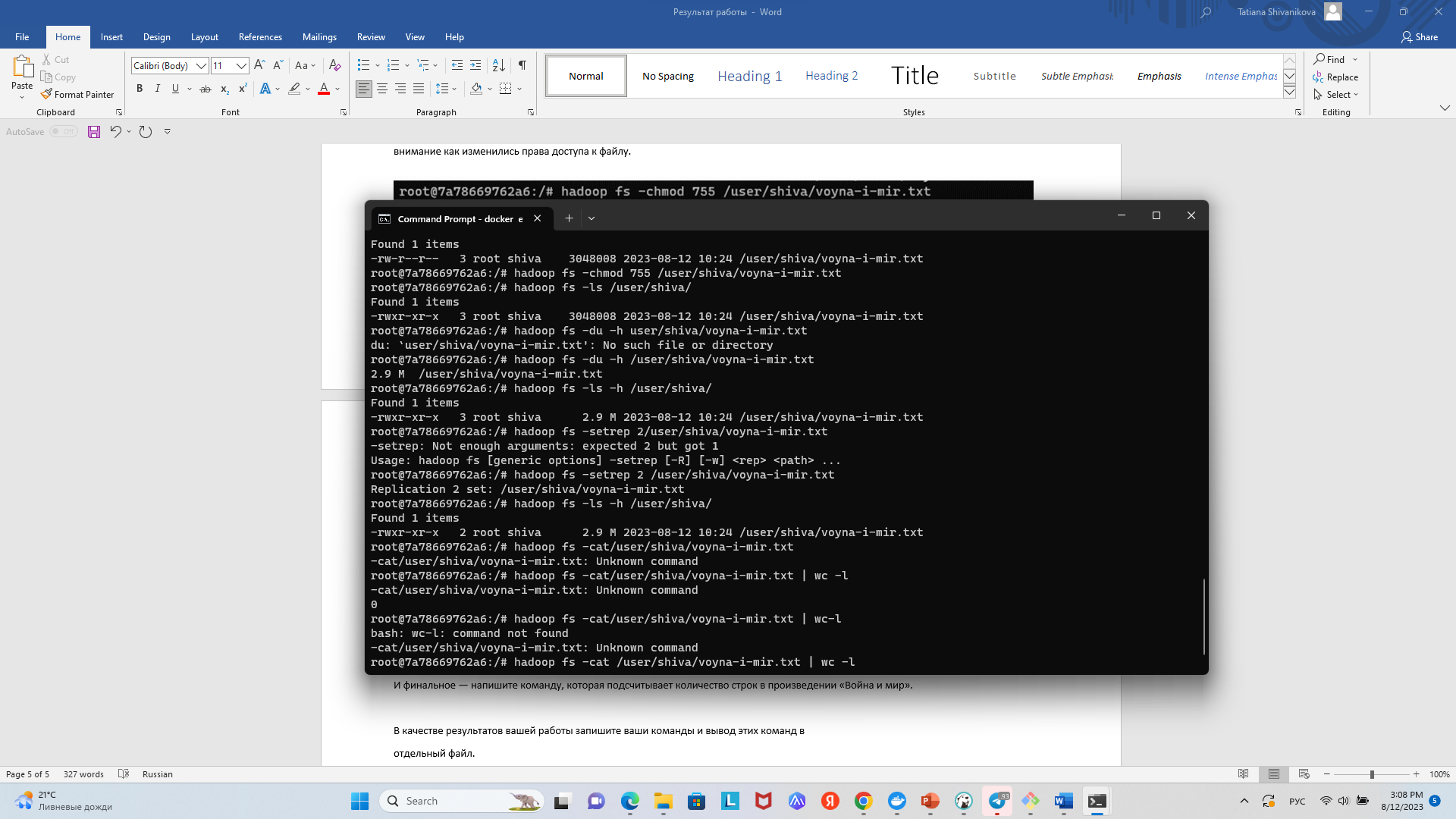
9. Установите режим доступа, который дает полный доступ для владельца файла, а для сторонних пользователей возможность читать и выполнять. Попробуйте заново использовать команду для вывода содержимого папки и обратите

внимание как изменились права доступа к файлу.



10. Теперь попробуем вывести на экран информацию о том, сколько места на диске

занимает наш файл. Желательно, чтобы размер файла был удобочитаемым.

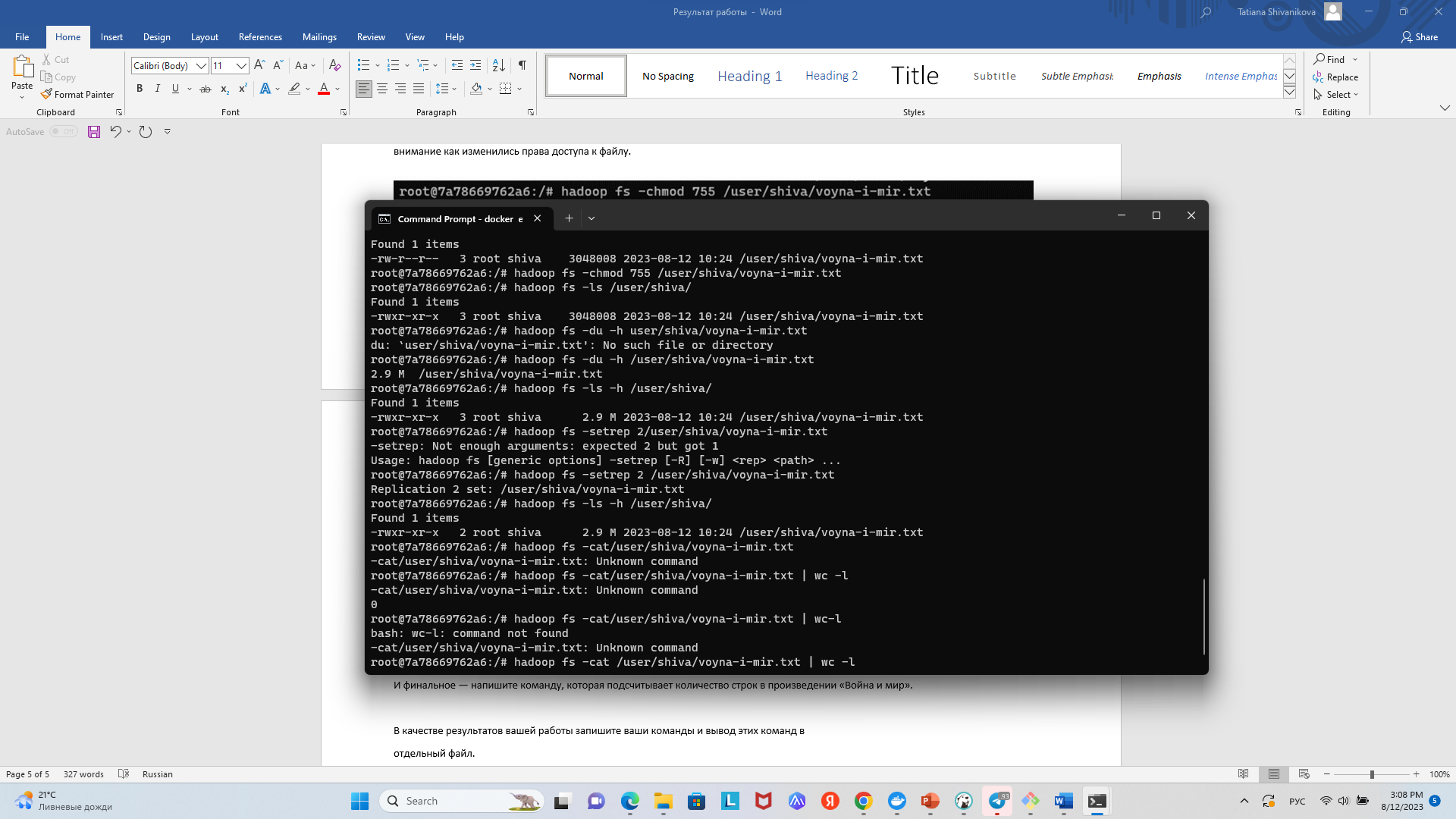


11. На экране вы можете заметить 2 числа. Первое число — это фактический размер файла,

а второе — это занимаемое файлом место на диске с учетом репликации — измените фактор репликации на 2.

Повторите команду, которая выводит информацию о том, какое место на диске

занимает файл и убедитесь, что изменения произошли.



12. И финальное — напишите команду, которая подсчитывает количество строк в произведении «Война и мир».

