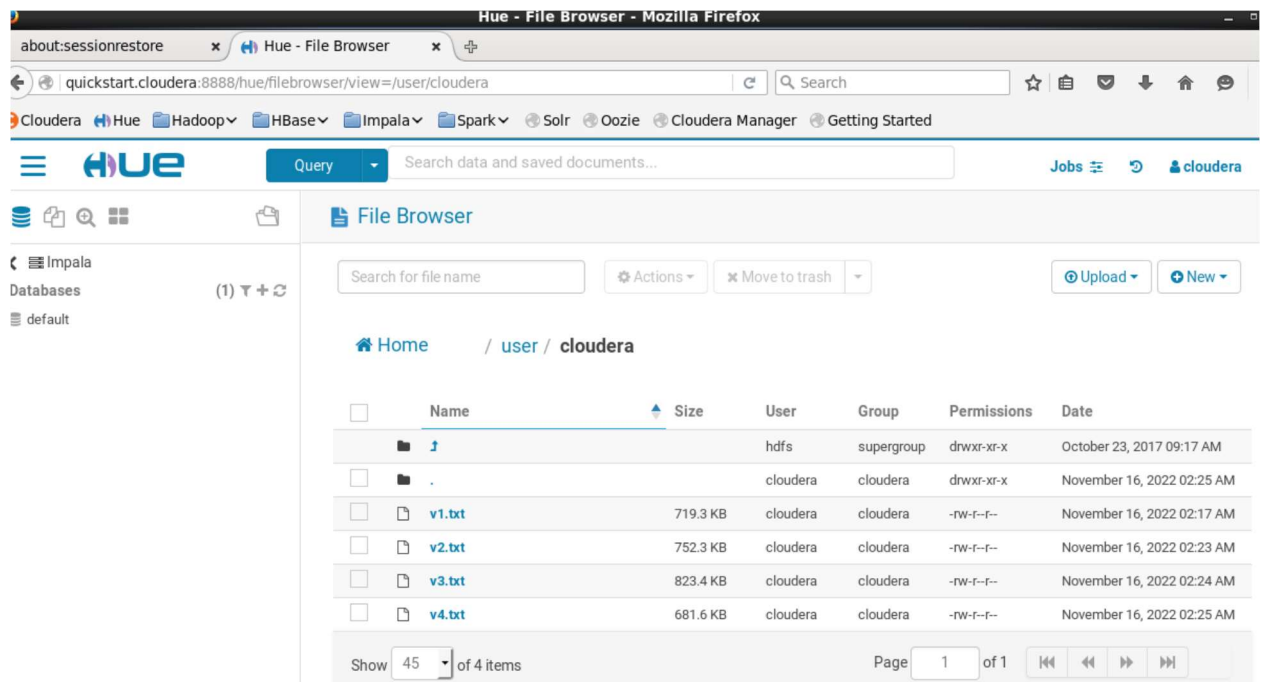


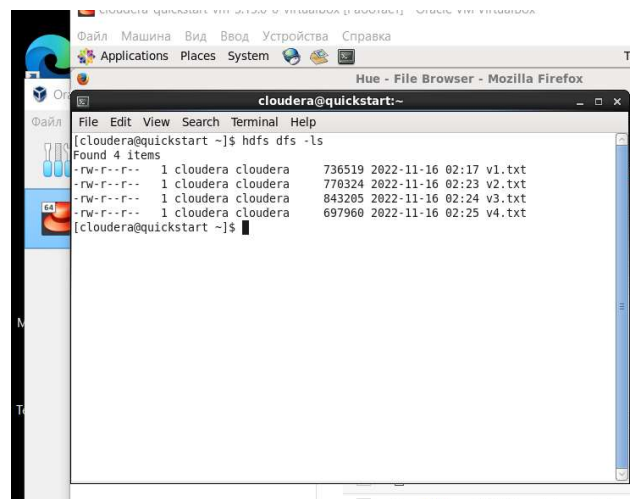
- Когда мы перетащили файлы с произведением Льва Толстого – мы перетащили их в файловую систему виртуальной машины, но не в HDFS, соответственно, в первую очередь нам нужно перенести их в папку нашего пользователя именно на HDFS.

```
hdfs dfs -copyFromLocal //home/cloudera/Desktop/v1.txt
hdfs dfs -copyFromLocal //home/cloudera/Desktop/v2.txt
hdfs dfs -copyFromLocal //home/cloudera/Desktop/v3.txt
hdfs dfs -copyFromLocal //home/cloudera/Desktop/v4.txt
```



- После того, как файлы окажутся на HDFS попробуйте выполнить команду, которая выводит содержимое папки. Особенно обратите внимание на права доступа к вашим файлам.

```
hdfs dfs -ls
```



- Далее сожмите все 4 тома в 1 файл.

```
hdfs dfs -getmerge /user/cloudera /home/cloudera/Desktop/1234.txt
```

и перенесем полученный файл в папку с остальными файлами

```
[cloudera@quickstart ~]$ hdfs dfs -ls
Found 5 items
-rw-r--r-- 1 cloudera cloudera 3048008 2022-11-17 05:51 1234.txt
-rw-r--r-- 1 cloudera cloudera 736519 2022-11-16 02:17 v1.txt
-rw-r--r-- 1 cloudera cloudera 770324 2022-11-16 02:23 v2.txt
-rw-r--r-- 1 cloudera cloudera 843205 2022-11-16 02:24 v3.txt
-rw-r--r-- 1 cloudera cloudera 697960 2022-11-16 02:25 v4.txt
[cloudera@quickstart ~]$
```

- Теперь давайте изменим права доступа к нашему файлу. Чтобы с нашим файлом могли взаимодействовать коллеги, установите режим доступа, который дает полный доступ для владельца файла, а для сторонних пользователей возможность читать и выполнять.

```
hdfs dfs -chmod 755 1234.txt
```

- Попробуйте заново использовать команду для вывода содержимого папки и обратите внимание как изменились права доступа к файлу.

```
[cloudera@quickstart ~]$ hdfs dfs -chmod 777 1234.txt
[cloudera@quickstart ~]$ hdfs dfs -ls
Found 5 items
-rwxrwxrwx 1 cloudera cloudera 3048008 2022-11-17 05:51 1234.txt
-rw-r--r-- 1 cloudera cloudera 736519 2022-11-16 02:17 v1.txt
-rw-r--r-- 1 cloudera cloudera 770324 2022-11-16 02:23 v2.txt
-rw-r--r-- 1 cloudera cloudera 843205 2022-11-16 02:24 v3.txt
-rw-r--r-- 1 cloudera cloudera 697960 2022-11-16 02:25 v4.txt
[cloudera@quickstart ~]$ hdfs dfs -chmod 764 1234.txt
[cloudera@quickstart ~]$ hdfs dfs -ls
Found 5 items
-rwxrw-r-- 1 cloudera cloudera 3048008 2022-11-17 05:51 1234.txt
-rw-r--r-- 1 cloudera cloudera 736519 2022-11-16 02:17 v1.txt
-rw-r--r-- 1 cloudera cloudera 770324 2022-11-16 02:23 v2.txt
-rw-r--r-- 1 cloudera cloudera 843205 2022-11-16 02:24 v3.txt
-rw-r--r-- 1 cloudera cloudera 697960 2022-11-16 02:25 v4.txt
[cloudera@quickstart ~]$ hdfs dfs -chmod 755 1234.txt
[cloudera@quickstart ~]$ hdfs dfs -ls
Found 5 items
-rwxr-xr-x 1 cloudera cloudera 3048008 2022-11-17 05:51 1234.txt
-rw-r--r-- 1 cloudera cloudera 736519 2022-11-16 02:17 v1.txt
-rw-r--r-- 1 cloudera cloudera 770324 2022-11-16 02:23 v2.txt
-rw-r--r-- 1 cloudera cloudera 843205 2022-11-16 02:24 v3.txt
-rw-r--r-- 1 cloudera cloudera 697960 2022-11-16 02:25 v4.txt
[cloudera@quickstart ~]$
```

- **744** - разрешить все для владельца, а остальным только чтение;
- **755** - все для владельца, остальным только чтение и выполнение;
- **764** - все для владельца, чтение и запись для группы, и только чтение для остальных;
- **777** - всем разрешено все.

- Теперь попробуем вывести на экран информацию о том, сколько места на диске занимает наш файл. Желательно, чтобы размер файла был удобочитаемым.

```
hdfs dfs -du /user/cloudera/1234.txt --размер в байтах
```

```
File Edit View Search Terminal Help
[cloudera@quickstart ~]$ hdfs dfs -du /user/cloudera/1234.txt
3048008 3048008 /user/cloudera/1234.txt
[cloudera@quickstart ~]$
```

`hdfs dfs -du -h /user/cloudera/1234.txt` - удобочитаемый

```
Applications Places System
cloudera@quickstart:~
File Edit View Search Terminal Help
[cloudera@quickstart ~]$ hdfs dfs -du /user/cloudera/1234.txt
3048008 3048008 /user/cloudera/1234.txt
[cloudera@quickstart ~]$ ^C
[cloudera@quickstart ~]$ hdfs dfs -du -h /user/cloudera/1234.txt
2.9 M 2.9 M /user/cloudera/1234.txt
[cloudera@quickstart ~]$
```

- На экране вы можете заметить 2 числа. Первое число – это фактический размер файла, а второе – это занимаемое файлом место на диске с учетом репликации. По умолчанию в данной версии HDFS эти числа будут одинаковы – это означает, что никакой репликации нет – нас это не очень устраивает, мы хотели бы, чтобы у наших файлов существовали резервные копии, поэтому напишите команду, которая изменит фактор репликации на 2.

`hdfs dfs -setrep -w 2 /user/cloudera/1234.txt`

- Повторите команду, которая выводит информацию о том, какое место на диске занимает файл и убедитесь, что изменения произошли.

`hdfs dfs -du -h /user/cloudera/1234.txt`

```
cloudera@quickstart:~
File Edit View Search Terminal Help
[cloudera@quickstart ~]$ hdfs dfs -du -h /user/cloudera/1234.txt
2.9 M 5.8 M /user/cloudera/1234.txt
[cloudera@quickstart ~]$
```

- Напишите команду, которая подсчитывает количество строк в вашем файле

`hdfs dfs -cat /user/cloudera/1234.txt | wc -l`

```
[cloudera@quickstart ~]$ hdfs dfs -cat /user/cloudera/1234.txt |wc -l
10272
[cloudera@quickstart ~]$
```