

Урок 2. HDFS

ssh student3_10@37.139.32.56 -D localhost:8080

Создаем директорию **data** в HDFS: `hdfs dfs -mkdir data`

```
Терминал - student3_10@bigdataanalytics2-head-shdpt-v31-1-0:~
Файл  Правка  Вид  Терминал  Вкладки  Справка
[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$ hdfs dfs -ls /
Found 13 items
drwxrwxrwt - yarn  hadoop      0 2021-04-08 15:19 /app-logs
drwxr-xr-x - hdfs  hdfs       0 2021-03-25 08:38 /apps
drwxr-xr-x - yarn  hadoop      0 2021-03-20 14:24 /ats
drwxr-xr-x - hdfs  hdfs       0 2021-03-20 14:25 /atsv2
drwxr-xr-x - hdfs  hdfs       0 2021-03-20 14:24 /hdp
drwx----- - livy   hdfs       0 2021-03-20 14:31 /livy2-recovery
drwxr-xr-x - mapred hdfs       0 2021-03-20 14:24 /mapred
drwxrwxrwx - mapred hadoop     0 2021-03-31 21:30 /mr-history
drwxrwxrwx - spark  hadoop     0 2021-04-08 21:07 /spark2-history
drwxr-xr-x - hdfs  hdfs       0 2021-04-06 09:11 /system
drwxrwxrwx - hdfs  hdfs       0 2021-04-07 13:46 /tmp
drwxr-xr-x - hdfs  hdfs       0 2021-04-02 16:10 /user
drwxr-xr-x - hdfs  hdfs       0 2021-04-02 08:21 /warehouse
[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$ ^C
[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$ pwd
/home/student3_10
[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$ hdfs dfs -mkdir data
[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$ ls
[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$ hdfs dfs -ls
Found 1 items
drwxr-xr-x - student3_10 student3_10      0 2021-04-08 21:10 data
[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$
```

Удаляем директорию **data**: `hdfs dfs -rm -r data`

```
Терминал - student3_10@bigdataanalytics2-head-shdpt-v31-1-0:~
Файл  Правка  Вид  Терминал  Вкладки  Справка
drwxr-xr-x - hdfs hdfs      0 2021-03-20 14:25 /atsv2
drwxr-xr-x - hdfs hdfs      0 2021-03-20 14:24 /hdp
drwx----- - livy hdfs      0 2021-03-20 14:31 /livy2-recovery
drwxr-xr-x - mapred hdfs     0 2021-03-20 14:24 /mapred
drwxrwxrwx - mapred hadoop    0 2021-03-31 21:30 /mr-history
drwxrwxrwx - spark hadoop    0 2021-04-08 21:07 /spark2-history
drwxr-xr-x - hdfs hdfs      0 2021-04-06 09:11 /system
drwxrwxrwx - hdfs hdfs      0 2021-04-07 13:46 /tmp
drwxr-xr-x - hdfs hdfs      0 2021-04-02 16:10 /user
drwxr-xr-x - hdfs hdfs      0 2021-04-02 08:21 /warehouse
[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$ ^C
[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$ pwd
/home/student3_10
[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$ hdfs dfs -mkdir data
[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$ ls
[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$ hdfs dfs -ls
Found 1 items
drwxr-xr-x - student3_10 student3_10      0 2021-04-08 21:10 data
[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$ hdfs dfs -rm -r data
21/04/08 21:14:03 INFO fs.TrashPolicyDefault: Moved: 'hdfs://bigdataanalytics2-head-shdpt-v31-1-0.novalocal:8020/user/student3_10/data' to trash at: hdfs://bigdataanalytics2-head-shdpt-v31-1-0.novalocal:8020/user/student3_10/.Trash/Current/user/student3_10/data
[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$ █
```

Просмотрим директорию: hdfs dfs -ls

```
Терминал - student3_10@bigdataanalytics2-head-shdpt-v31-1-0:~
Файл  Правка  Вид  Терминал  Вкладки  Справка
drwxrwxrwx - mapred hadoop    0 2021-03-31 21:30 /mr-history
drwxrwxrwx - spark hadoop    0 2021-04-08 21:07 /spark2-history
drwxr-xr-x - hdfs hdfs      0 2021-04-06 09:11 /system
drwxrwxrwx - hdfs hdfs      0 2021-04-07 13:46 /tmp
drwxr-xr-x - hdfs hdfs      0 2021-04-02 16:10 /user
drwxr-xr-x - hdfs hdfs      0 2021-04-02 08:21 /warehouse
[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$ ^C
[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$ pwd
/home/student3_10
[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$ hdfs dfs -mkdir data
[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$ ls
[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$ hdfs dfs -ls
Found 1 items
drwxr-xr-x - student3_10 student3_10      0 2021-04-08 21:10 data
[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$ hdfs dfs -rm -r data
21/04/08 21:14:03 INFO fs.TrashPolicyDefault: Moved: 'hdfs://bigdataanalytics2-head-shdpt-v31-1-0.novalocal:8020/user/student3_10/data' to trash at: hdfs://bigdataanalytics2-head-shdpt-v31-1-0.novalocal:8020/user/student3_10/.Trash/Current/user/student3_10/data
[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$ ls
[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$ hdfs dfs -ls
Found 1 items
drwx----- - student3_10 student3_10      0 2021-04-08 21:14 .Trash
[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$ █
```

Удаляем директорию **.Trash** со всем содержимым: hdfs dfs -rm -r .Trash

```
Терминал - student3_10@bigdataanalytics2-head-shdpt-v31-1-0:~
Файл  Правка  Вид  Терминал  Вкладки  Справка
igor@igor-MS-7808:~$ ssh student3_10@37.139.32.56
Last login: Sat Apr 10 19:20:45 2021 from 109.252.26.235
[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$ ls
urbandict.csv
[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$ hdfs dfs -ls
Found 1 items
drwx-----  - student3_10 student3_10          0 2021-04-10 18:11 .Trash
[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$ hdfs dfs -rm -r .Trash
Deleted .Trash
[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$ hdfs dfs -ls
[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$
```

Создаем директорию **data** в HDFS и в ней файл **1.txt**: `hdfs dfs -touchz data/1.txt`

```
Терминал - student3_10@bigdataanalytics2-head-shdpt-v31-1-0:~
Файл  Правка  Вид  Терминал  Вкладки  Справка
drwx-----  - student3_10 student3_10          0 2021-04-08 21:14 .Trash
[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$ hdfs dfs -rm -r .Trash
Deleted .Trash
[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$ hdfs dfs -rm -r data
rm: `data': No such file or directory
[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$ hdfs dfs -mkdir data
[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$ hdfs dfs -ls
[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$ hdfs dfs -touchz /data/1.txt
touchz: `/data/1.txt': No such file or directory: `hdfs://bigdataanalytics2-head-shdpt-v31-1-0.novalocal:8020/data/1.txt'
[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$ hdfs dfs -touchz data/1.txt
[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$ pwd
/home/student3_10
[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$ hdfs dfs -ls
Found 1 items
drwxr-xr-x  - student3_10 student3_10          0 2021-04-08 21:48 data
[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$ data
-bash: data: команда не найдена
[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$ hdfs dfs -ls /data
ls: `/data': No such file or directory
[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$ hdfs dfs -ls data
Found 1 items
-rw-r--r--  2 student3_10 student3_10          0 2021-04-08 21:48 data/1.txt
[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$
```

Удаляем директорию **data** со всем содержимым: `hdfs dfs -rm -r data`

```
Терминал - student3_10@bigdataanalytics2-head-shdpt-v31-1-0:~
Файл  Правка  Вид  Терминал  Вкладки  Справка
[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$ hdfs dfs -touchz data/1.txt
[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$ pwd
/home/student3_10
[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$ hdfs dfs -ls
Found 1 items
drwxr-xr-x  - student3_10 student3_10          0 2021-04-08 21:48 data
[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$ data
-bash: data: команда не найдена
[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$ hdfs dfs -ls /data
ls: '/data': No such file or directory
[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$ hdfs dfs -ls data
Found 1 items
-rw-r--r--  2 student3_10 student3_10          0 2021-04-08 21:48 data/1.txt
[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$ hdfs dfs -rm -r data
21/04/08 21:54:35 INFO fs.TrashPolicyDefault: Moved: 'hdfs://bigdataanalytics2-head-shdpt-v31-1-0.novalocal:8020/user/student3_10/data' to trash at: hdfs://bigdataanalytics2-head-shdpt-v31-1-0.novalocal:8020/user/student3_10/.Trash/Current/user/student3_10/data
[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$ hdfs dfs -ls
Found 1 items
drwx-----  - student3_10 student3_10          0 2021-04-08 21:54 .Trash
[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$ hdfs dfs -rm -r .Trash
Deleted .Trash
[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$
```

Создаем директорию **data** в HDFS: `hdfs dfs -mkdir data`

```
Терминал - student3_10@bigdataanalytics2-head-shdpt-v31-1-0:~
Файл  Правка  Вид  Терминал  Вкладки  Справка
Found 1 items
drwxr-xr-x  - student3_10 student3_10          0 2021-04-08 21:48 data
[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$ data
-bash: data: команда не найдена
[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$ hdfs dfs -ls /data
ls: '/data': No such file or directory
[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$ hdfs dfs -ls data
Found 1 items
-rw-r--r--  2 student3_10 student3_10          0 2021-04-08 21:48 data/1.txt
[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$ hdfs dfs -rm -r data
21/04/08 21:54:35 INFO fs.TrashPolicyDefault: Moved: 'hdfs://bigdataanalytics2-head-shdpt-v31-1-0.novalocal:8020/user/student3_10/data' to trash at: hdfs://bigdataanalytics2-head-shdpt-v31-1-0.novalocal:8020/user/student3_10/.Trash/Current/user/student3_10/data
[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$ hdfs dfs -ls
Found 1 items
drwx-----  - student3_10 student3_10          0 2021-04-08 21:54 .Trash
[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$ hdfs dfs -rm -r .Trash
Deleted .Trash
[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$ hdfs dfs -mkdir data
[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$ hdfs dfs -ls
Found 1 items
drwxr-xr-x  - student3_10 student3_10          0 2021-04-08 21:56 data
[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$
```

Удаляем директорию **data** со всем содержимым: `hdfs dfs -rm -r data`


```
Терминал - student3_10@bigdataanalytics2-head-shdpt-v31-1-0:~
Файл  Правка  Вид  Терминал  Вкладки  Справка
igor@igor-MS-7808:~$ ssh student3_10@37.139.32.56
Last login: Sat Apr 10 17:47:44 2021 from 109.252.26.235
[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$ hdfs dfs -ls
Found 1 items
drwxr-xr-x  - student3_10 student3_10          0 2021-04-08 21:56 data
[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$ hdfs dfs -rm -r data
21/04/10 18:11:07 INFO fs.TrashPolicyDefault: Moved: 'hdfs://bigdataanalytics2-head-shdpt-v31-1-0.novalocal:8020/user/student3_10/data' to trash at: hdfs://bigdataanalytics2-head-shdpt-v31-1-0.novalocal:8020/user/student3_10/.Trash/Current/user/student3_10/data
[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$
```

```
Терминал - student3_10@bigdataanalytics2-head-shdpt-v31-1-0:~
Файл  Правка  Вид  Терминал  Вкладки  Справка
igor@igor-MS-7808:~$ ssh student3_10@37.139.32.56
Last login: Sat Apr 10 17:47:44 2021 from 109.252.26.235
[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$ hdfs dfs -ls
Found 1 items
drwxr-xr-x  - student3_10 student3_10          0 2021-04-08 21:56 data
[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$ hdfs dfs -rm -r data
21/04/10 18:11:07 INFO fs.TrashPolicyDefault: Moved: 'hdfs://bigdataanalytics2-head-shdpt-v31-1-0.novalocal:8020/user/student3_10/data' to trash at: hdfs://bigdataanalytics2-head-shdpt-v31-1-0.novalocal:8020/user/student3_10/.Trash/Current/user/student3_10/data
[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$ hdfs dfs -ls
Found 1 items
drwx-----  - student3_10 student3_10          0 2021-04-10 18:11 .Trash
[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$
```

Положить файл **urbandict.csv** в HDFS: `scp urbandict.csv student3_10@37.139.32.56:.`

```
Терминал - igor@igor-MS-7808: ~/Загрузки
Файл  Правка  Вид  Терминал  Вкладки  Справка
igor@igor-MS-7808:~/Загрузки$ scp urbandict.csv student3_10@37.139.32.56:
urbandict.csv          62% 321MB 36.5MB/s 00:05 ETA
urbandict.csv          100% 513MB 36.5MB/s 00:14
igor@igor-MS-7808:~/Загрузки$
```

Копируем файл **urbandict.csv** в HDFS: `hdfs dfs -put urbandict.csv`

```
Терминал - student3_10@bigdataanalytics2-head-shdpt-v31-1-0:~
Файл  Правка  Вид  Терминал  Вкладки  Справка
drwxrwxrwt - yarn  hadoop    0 2021-04-10 16:05 /app-logs
drwxr-xr-x - hdfs  hdfs      0 2021-03-25 08:38 /apps
drwxr-xr-x - yarn  hadoop    0 2021-03-20 14:24 /ats
drwxr-xr-x - hdfs  hdfs      0 2021-03-20 14:25 /atsv2
drwxrwxrwx - hdfs  hdfs      0 2021-04-09 17:50 /datasets
drwxr-xr-x - hdfs  hdfs      0 2021-03-20 14:24 /hdp
drwx----- - livy   hdfs      0 2021-03-20 14:31 /livy2-recovery
drwxr-xr-x - mapred hdfs      0 2021-03-20 14:24 /mapred
drwxrwxrwx - mapred hadoop    0 2021-03-31 21:30 /mr-history
drwxrwxrwx - spark  hadoop    0 2021-04-10 18:18 /spark2-history
drwxr-xr-x - hdfs  hdfs      0 2021-04-06 09:11 /system
drwxrwxrwx - hdfs  hdfs      0 2021-04-07 13:46 /tmp
drwxr-xr-x - hdfs  hdfs      0 2021-04-02 16:10 /user
drwxr-xr-x - hdfs  hdfs      0 2021-04-02 08:21 /warehouse
[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$ cd /datasets
-bash: cd: /datasets: Нет такого файла или каталога
[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$ cd urbandict.csv
-bash: cd: urbandict.csv: Это не каталог
[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$ pwd
/home/student3_10
[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$ hdfs dfs -put urbandict.csv
[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$ ls
urbandict.csv
[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$
```

Просмотреть размер: ls -al

```
Терминал - student3_10@bigdataanalytics2-head-shdpt-v31-1-0:~
Файл  Правка  Вид  Терминал  Вкладки  Справка
igor@igor-MS-7808:~$ ssh student3_10@37.139.32.56
Last login: Sat Apr 10 19:20:45 2021 from 109.252.26.235
[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$ ls
urbandict.csv
[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$ hdfs dfs -ls
Found 1 items
drwx----- - student3_10 student3_10      0 2021-04-10 18:11 .Trash
[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$ hdfs dfs -rm -r .Trash
Deleted .Trash
[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$ hdfs dfs -ls
[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$ ls
urbandict.csv
[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$ ls -al
итого 525576
drwx-----  4 student3_10 student3_10      4096 anp  10 17:39 .
drwxr-xr-x. 213 root          root          4096 anp   1 16:31 ..
-rw-----  1 student3_10 student3_10     2648 anp  10 19:15 .bash_history
-rw-r--r--  1 student3_10 student3_10       18 abr  8 2019 .bash_logout
-rw-r--r--  1 student3_10 student3_10      193 abr  8 2019 .bash_profile
-rw-r--r--  1 student3_10 student3_10      231 abr  8 2019 .bashrc
drwxrwxr-x  2 student3_10 student3_10     4096 anp   3 18:35 .oracle_jre_usage
drwx-----  2 student3_10 student3_10     4096 anp   3 18:40 .ssh
-rw-rw-r--  1 student3_10 student3_10 538150516 anp  10 17:39 urbandict.csv
[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$
```

Копируем файл в HDFS: hdfs dfs -put urbandict.csv

```
Терминал - student3_10@bigdataanalytics2-head-shdpt-v31-1-0:~
Файл  Правка  Вид  Терминал  Вкладки  Справка
igor@igor-MS-7808:~$ ssh student3_10@37.139.32.56
Last login: Sat Apr 10 20:37:08 2021 from 109.252.26.235
[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$ ls
urbandict.csv
[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$ hdfs dfs -put urbandict.csv
[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$ ls
urbandict.csv
[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$ hdfs dfs -ls
Found 1 items
-rw-r--r--  2 student3_10 student3_10  538150516 2021-04-11 06:14 urbandict.csv
[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$
```

Смотрим размер файла: hdfs dfs -ls -h

```
Терминал - student3_10@bigdataanalytics2-head-shdpt-v31-1-0:~
Файл  Правка  Вид  Терминал  Вкладки  Справка
igor@igor-MS-7808:~$ ssh student3_10@37.139.32.56
Last login: Sat Apr 10 20:37:08 2021 from 109.252.26.235
[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$ ls
urbandict.csv
[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$ hdfs dfs -put urbandict.csv
[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$ ls
urbandict.csv
[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$ hdfs dfs -ls
Found 1 items
-rw-r--r--  2 student3_10 student3_10  538150516 2021-04-11 06:14 urbandict.csv
[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$ hdfs dfs -ls -h
Found 1 items
-rw-r--r--  2 student3_10 student3_10    513.2 M 2021-04-11 06:14 urbandict.csv
[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$
```

фактор репликации 2: hdfs dfs -du -h urbandict.csv

```
Терминал - student3_10@bigdataanalytics2-head-shdpt-v31-1-0:~
Файл  Правка  Вид  Терминал  Вкладки  Справка
igor@igor-MS-7808:~$ ssh student3_10@37.139.32.56
Last login: Sat Apr 10 20:37:08 2021 from 109.252.26.235
[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$ ls
urbandict.csv
[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$ hdfs dfs -put urbandict.csv
[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$ ls
urbandict.csv
[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$ hdfs dfs -ls
Found 1 items
-rw-r--r--  2 student3_10 student3_10  538150516 2021-04-11 06:14 urbandict.csv
[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$ hdfs dfs -ls -h
Found 1 items
-rw-r--r--  2 student3_10 student3_10    513.2 M 2021-04-11 06:14 urbandict.csv
[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$ hdfs dfs -du -h urbandict.csv
513.2 M  1.0 G urbandict.csv
[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$
```

Изменяем файл репликации 3: `hdfs dfs -setrep 3 urbandict.csv`

```
Терминал - student3_10@bigdataanalytics2-head-shdpt-v31-1-0:~
Файл  Правка  Вид  Терминал  Вкладки  Справка
igor@igor-MS-7808:~$ ssh student3_10@37.139.32.56
Last login: Sun Apr 11 06:09:09 2021 from 109.252.26.235
[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$ hdfs dfs -setrep 3 urbandict.csv
Replication 3 set: urbandict.csv
[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$
```

```
Терминал - student3_10@bigdataanalytics2-head-shdpt-v31-1-0:~
Файл  Правка  Вид  Терминал  Вкладки  Справка
igor@igor-MS-7808:~$ ssh student3_10@37.139.32.56
Last login: Sun Apr 11 06:09:09 2021 from 109.252.26.235
[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$ hdfs dfs -setrep 3 urbandict.csv
Replication 3 set: urbandict.csv
[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$ hdfs dfs -ls -h
Found 1 items
-rw-r--r--   3 student3_10 student3_10   513.2 M 2021-04-11 06:14 urbandict.csv
[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$ hdfs dfs -ls
Found 1 items
-rw-r--r--   3 student3_10 student3_10 538150516 2021-04-11 06:14 urbandict.csv
[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$ hdfs dfs -du -h urbandict.csv
513.2 M 1.5 G urbandict.csv
[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$
```

Посмотрим все папки в корне HDFS: `hdfs dfs -ls -h /`


```
Терминал - student3_10@bigdataanalytics2-head-shdpt-v31-1-0:~
Файл  Правка  Вид  Терминал  Вкладки  Справка
Found 1 items
-rw-r--r--  3 student3_10 student3_10    513.2 M 2021-04-11 06:14 urbandict.csv
[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$ hdfs dfs -ls
Found 1 items
-rw-r--r--  3 student3_10 student3_10  538150516 2021-04-11 06:14 urbandict.csv
[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$ hdfs dfs -du -h urbandict.csv
513.2 M  1.5 G urbandict.csv
[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$ hdfs dfs -ls -h /
Found 14 items
drwxrwxrwt - yarn  hadoop      0 2021-04-10 16:05 /app-logs
drwxr-xr-x - hdfs  hdfs      0 2021-03-25 08:38 /apps
drwxr-xr-x - yarn  hadoop      0 2021-03-20 14:24 /ats
drwxr-xr-x - hdfs  hdfs      0 2021-03-20 14:25 /atsv2
drwxrwxrwx - hdfs  hdfs      0 2021-04-09 17:50 /datasets
drwxr-xr-x - hdfs  hdfs      0 2021-03-20 14:24 /hdp
drwx----- - livy  hdfs      0 2021-03-20 14:31 /livy2-recovery
drwxr-xr-x - mapred hdfs      0 2021-03-20 14:24 /mapred
drwxrwxrwx - mapred hadoop      0 2021-03-31 21:30 /mr-history
drwxrwxrwx - spark  hadoop      0 2021-04-11 06:54 /spark2-history
drwxr-xr-x - hdfs  hdfs      0 2021-04-06 09:11 /system
drwxrwxrwx - hdfs  hdfs      0 2021-04-07 13:46 /tmp
drwxr-xr-x - hdfs  hdfs      0 2021-04-02 16:10 /user
drwxr-xr-x - hdfs  hdfs      0 2021-04-02 08:21 /warehouse
[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$
```

Посмотрим файл: `hdfs dfs -ls -h /user/student3_10`

```
Терминал - student3_10@bigdataanalytics2-head-shdpt-v31-1-0:~
Файл  Правка  Вид  Терминал  Вкладки  Справка
Found 1 items
-rw-r--r--  3 student3_10 student3_10  538150516 2021-04-11 06:14 urbandict.csv
[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$ hdfs dfs -du -h urbandict.csv
513.2 M  1.5 G urbandict.csv
[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$ hdfs dfs -ls -h /
Found 14 items
drwxrwxrwt - yarn  hadoop      0 2021-04-10 16:05 /app-logs
drwxr-xr-x - hdfs  hdfs      0 2021-03-25 08:38 /apps
drwxr-xr-x - yarn  hadoop      0 2021-03-20 14:24 /ats
drwxr-xr-x - hdfs  hdfs      0 2021-03-20 14:25 /atsv2
drwxrwxrwx - hdfs  hdfs      0 2021-04-09 17:50 /datasets
drwxr-xr-x - hdfs  hdfs      0 2021-03-20 14:24 /hdp
drwx----- - livy  hdfs      0 2021-03-20 14:31 /livy2-recovery
drwxr-xr-x - mapred hdfs      0 2021-03-20 14:24 /mapred
drwxrwxrwx - mapred hadoop      0 2021-03-31 21:30 /mr-history
drwxrwxrwx - spark  hadoop      0 2021-04-11 06:54 /spark2-history
drwxr-xr-x - hdfs  hdfs      0 2021-04-06 09:11 /system
drwxrwxrwx - hdfs  hdfs      0 2021-04-07 13:46 /tmp
drwxr-xr-x - hdfs  hdfs      0 2021-04-02 16:10 /user
drwxr-xr-x - hdfs  hdfs      0 2021-04-02 08:21 /warehouse
[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$ hdfs dfs -ls -h /user/student3_10
Found 1 items
-rw-r--r--  3 student3_10 student3_10    513.2 M 2021-04-11 06:14 /user/student3_10/urbandict.csv
[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$
```

Находимся в директории: `pwd`

```
Терминал - student3_10@bigdataanalytics2-head-shdpt-v31-1-0:~
Файл  Правка  Вид  Терминал  Вкладки  Справка
[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$ hdfs dfs -du -h urbandict.csv
513.2 M  1.5 G  urbandict.csv
[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$ hdfs dfs -ls -h /
Found 14 items
drwxrwxrwt - yarn  hadoop      0 2021-04-10 16:05 /app-logs
drwxr-xr-x - hdfs  hdfs      0 2021-03-25 08:38 /apps
drwxr-xr-x - yarn  hadoop      0 2021-03-20 14:24 /ats
drwxr-xr-x - hdfs  hdfs      0 2021-03-20 14:25 /atsv2
drwxrwxrwx - hdfs  hdfs      0 2021-04-09 17:50 /datasets
drwxr-xr-x - hdfs  hdfs      0 2021-03-20 14:24 /hdp
drwx----- - livy  hdfs      0 2021-03-20 14:31 /livy2-recovery
drwxr-xr-x - mapred hdfs      0 2021-03-20 14:24 /mapred
drwxrwxrwx - mapred hadoop      0 2021-03-31 21:30 /mr-history
drwxrwxrwx - spark  hadoop      0 2021-04-11 06:54 /spark2-history
drwxr-xr-x - hdfs  hdfs      0 2021-04-06 09:11 /system
drwxrwxrwx - hdfs  hdfs      0 2021-04-07 13:46 /tmp
drwxr-xr-x - hdfs  hdfs      0 2021-04-02 16:10 /user
drwxr-xr-x - hdfs  hdfs      0 2021-04-02 08:21 /warehouse
[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$ hdfs dfs -ls -h /user/student3_10
Found 1 items
-rw-r--r--  3 student3_10 student3_10  513.2 M 2021-04-11 06:14 /user/student3_10/urbandict.csv
[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$ pwd
/home/student3_10
[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$
```

Смотрим что файл urbandict.csv хранится в home: ls -al

```
Терминал - student3_10@bigdataanalytics2-head-shdpt-v31-1-0:~
Файл  Правка  Вид  Терминал  Вкладки  Справка
drwxr-xr-x - mapred hdfs      0 2021-03-20 14:24 /mapred
drwxrwxrwx - mapred hadoop      0 2021-03-31 21:30 /mr-history
drwxrwxrwx - spark  hadoop      0 2021-04-11 06:54 /spark2-history
drwxr-xr-x - hdfs  hdfs      0 2021-04-06 09:11 /system
drwxrwxrwx - hdfs  hdfs      0 2021-04-07 13:46 /tmp
drwxr-xr-x - hdfs  hdfs      0 2021-04-02 16:10 /user
drwxr-xr-x - hdfs  hdfs      0 2021-04-02 08:21 /warehouse
[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$ hdfs dfs -ls -h /user/student3_10
Found 1 items
-rw-r--r--  3 student3_10 student3_10  513.2 M 2021-04-11 06:14 /user/student3_10/urbandict.csv
[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$ pwd
/home/student3_10
[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$ ls -al
итого 525576
drwx-----  4 student3_10 student3_10  4096 anp 10 17:39 .
drwxr-xr-x. 213 root      root      4096 anp  1 16:31 ..
-rw-----  1 student3_10 student3_10  3041 anp 11 06:32 .bash_history
-rw-r--r--  1 student3_10 student3_10    18 abr  8 2019 .bash_logout
-rw-r--r--  1 student3_10 student3_10   193 abr  8 2019 .bash_profile
-rw-r--r--  1 student3_10 student3_10   231 abr  8 2019 .bashrc
drwxrwxr-x  2 student3_10 student3_10  4096 anp  3 18:35 .oracle_jre_usage
drwx-----  2 student3_10 student3_10  4096 anp  3 18:40 .ssh
-rw-rw-r--  1 student3_10 student3_10 538150516 anp 10 17:39 urbandict.csv
[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$
```

Скопировать файл urbandict.csv: hdfs dfs -get urbandict.csv urbandict2.csv

```
Терминал - student3_10@bigdataanalytics2-head-shdpt-v31-1-0:~
Файл  Правка  Вид  Терминал  Вкладки  Справка
drwxrwxrwx - hdfs hdfs      0 2021-04-07 13:46 /tmp
drwxr-xr-x - hdfs hdfs      0 2021-04-02 16:10 /user
drwxr-xr-x - hdfs hdfs      0 2021-04-02 08:21 /warehouse
[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$ hdfs dfs -ls -h /user/student3_10
Found 1 items
-rw-r--r--  3 student3_10 student3_10    513.2 M 2021-04-11 06:14 /user/student3_10/urbandict.csv
[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$ pwd
/home/student3_10
[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$ ls -al
итого 525576
drwx-----  4 student3_10 student3_10    4096 anp 10 17:39 .
drwxr-xr-x. 213 root          root          4096 anp  1 16:31 ..
-rw-----  1 student3_10 student3_10    3041 anp 11 06:32 .bash_history
-rw-r--r--  1 student3_10 student3_10      18 abr  8 2019 .bash_logout
-rw-r--r--  1 student3_10 student3_10    193 abr  8 2019 .bash_profile
-rw-r--r--  1 student3_10 student3_10    231 abr  8 2019 .bashrc
drwxrwxr-x  2 student3_10 student3_10    4096 anp  3 18:35 .oracle_jre_usage
drwx-----  2 student3_10 student3_10    4096 anp  3 18:40 .ssh
-rw-rw-r--  1 student3_10 student3_10 538150516 anp 10 17:39 urbandict.csv
[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$ hdfs dfs -ls -h
Found 1 items
-rw-r--r--  3 student3_10 student3_10    513.2 M 2021-04-11 06:14 urbandict.csv
[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$ hdfs dfs -get urbandict.csv urbandict2.csv
[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$
```

Смотрим файл: ls -alh

```
Терминал - student3_10@bigdataanalytics2-head-shdpt-v31-1-0:~
Файл  Правка  Вид  Терминал  Вкладки  Справка
-rw-----  1 student3_10 student3_10    3041 anp 11 06:32 .bash_history
-rw-r--r--  1 student3_10 student3_10      18 abr  8 2019 .bash_logout
-rw-r--r--  1 student3_10 student3_10    193 abr  8 2019 .bash_profile
-rw-r--r--  1 student3_10 student3_10    231 abr  8 2019 .bashrc
drwxrwxr-x  2 student3_10 student3_10    4096 anp  3 18:35 .oracle_jre_usage
drwx-----  2 student3_10 student3_10    4096 anp  3 18:40 .ssh
-rw-rw-r--  1 student3_10 student3_10 538150516 anp 10 17:39 urbandict.csv
[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$ hdfs dfs -ls -h
Found 1 items
-rw-r--r--  3 student3_10 student3_10    513.2 M 2021-04-11 06:14 urbandict.csv
[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$ hdfs dfs -get urbandict.csv urbandict2.csv
[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$ ls -alh
итого 1,1G
drwx-----  4 student3_10 student3_10  4,0K anp 11 07:11 .
drwxr-xr-x. 213 root          root          4,0K anp  1 16:31 ..
-rw-----  1 student3_10 student3_10  3,0K anp 11 06:32 .bash_history
-rw-r--r--  1 student3_10 student3_10   18 abr  8 2019 .bash_logout
-rw-r--r--  1 student3_10 student3_10  193 abr  8 2019 .bash_profile
-rw-r--r--  1 student3_10 student3_10   231 abr  8 2019 .bashrc
drwxrwxr-x  2 student3_10 student3_10  4,0K anp  3 18:35 .oracle_jre_usage
drwx-----  2 student3_10 student3_10  4,0K anp  3 18:40 .ssh
-rw-r--r--  1 student3_10 student3_10 514M anp 11 07:11 urbandict2.csv
-rw-rw-r--  1 student3_10 student3_10 514M anp 10 17:39 urbandict.csv
[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$
```

Удаляем файл urbandict2.csv: rm -rf urbandict2.csv

```
Терминал - student3_10@bigdataanalytics2-head-shdpt-v31-1-0:~
Файл  Правка  Вид  Терминал  Вкладки  Справка
drwx-----  4 student3_10 student3_10 4,0K anp 11 07:11 .
drwxr-xr-x. 213 root      root      4,0K anp  1 16:31 ..
-rw-----  1 student3_10 student3_10 3,0K anp 11 06:32 .bash_history
-rw-r--r--  1 student3_10 student3_10  18 abr  8 2019 .bash_logout
-rw-r--r--  1 student3_10 student3_10 193 abr  8 2019 .bash_profile
-rw-r--r--  1 student3_10 student3_10 231 abr  8 2019 .bashrc
drwxrwxr-x  2 student3_10 student3_10 4,0K anp  3 18:35 .oracle_jre_usage
drwx-----  2 student3_10 student3_10 4,0K anp  3 18:40 .ssh
-rw-r--r--  1 student3_10 student3_10 514M anp 11 07:11 urbandict2.csv
-rw-rw-r--  1 student3_10 student3_10 514M anp 10 17:39 urbandict.csv
[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$ rm -rf urbandict2.csv
[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$ ls -alh
итого 514M
drwx-----  4 student3_10 student3_10 4,0K anp 11 07:17 .
drwxr-xr-x. 213 root      root      4,0K anp  1 16:31 ..
-rw-----  1 student3_10 student3_10 3,0K anp 11 06:32 .bash_history
-rw-r--r--  1 student3_10 student3_10  18 abr  8 2019 .bash_logout
-rw-r--r--  1 student3_10 student3_10 193 abr  8 2019 .bash_profile
-rw-r--r--  1 student3_10 student3_10 231 abr  8 2019 .bashrc
drwxrwxr-x  2 student3_10 student3_10 4,0K anp  3 18:35 .oracle_jre_usage
drwx-----  2 student3_10 student3_10 4,0K anp  3 18:40 .ssh
-rw-rw-r--  1 student3_10 student3_10 514M anp 10 17:39 urbandict.csv
[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$
```

Посмотреть как файл хранится на файловой системе: `hdfs fsck urbandict.csv -files -bloks -locations`


```
Терминал - student3_10@bigdataanalytics2-head-shdpt-v31-1-0:~
Файл  Правка  Вид  Терминал  Вкладки  Справка
drwxrwxr-x  2 student3_10 student3_10 4,0K апр  3 18:35 .oracle_jre_usage
[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$ hdfs fsck urbandict.csv -files -bloks -locations
fsck: Illegal option '-bloks'
Usage: hdfs fsck <path> [-list-corruptfileblocks | [-move | -delete | -openforwrite] [-files [-blocks [-locations | -racks | -replicaDetails | -upgradeddomains]]] [-includeSnapshots] [-showprogress] [-storagepolicies] [-maintenance] [-blockId <blk_id>]
    <path> start checking from this path
    -move move corrupted files to /lost+found
    -delete delete corrupted files
    -files print out files being checked
    -openforwrite print out files opened for write
    -includeSnapshots include snapshot data if the given path indicates a snapshottable directory or there are snapshottable directories under it
    -list-corruptfileblocks print out list of missing blocks and files they belong to
    -files -blocks print out block report
    -files -blocks -locations print out locations for every block
    -files -blocks -racks print out network topology for data-node locations
    -files -blocks -replicaDetails print out each replica details
    -files -blocks -upgradeddomains print out upgrade domains for every block
    -storagepolicies print out storage policy summary for the blocks
    -maintenance print out maintenance state node details
    -showprogress show progress in output. Default is OFF (no progress)
    -blockId print out which file this blockId belongs to, locations (nodes, racks) of this block, and other diagnostics info (under replicated, corrupted or not, etc)

Please Note:
  1. By default fsck ignores files opened for write, use -openforwrite to report such files. They are usually tagged CORRUPT or HEALTHY depending on their block allocation status
  2. Option -includeSnapshots should not be used for comparing stats, should be used only for HEALTH check, as this may contain duplicates if the same file present in both original fs tree and inside snapshots.

Generic options supported are:
  -conf <configuration file> specify an application configuration file
  -D <property=value> define a value for a given property
  -fs <file:///|hdfs://namenode:port> specify default filesystem URL to use, overrides 'fs.defaultFS' property from configurations.
  -jt <local|resourcemanager:port> specify a ResourceManager
  -files <file1,...> specify a comma-separated list of files to be copied to the map reduce cluster
  -libjars <jar1,...> specify a comma-separated list of jar files to be included in the classpath
  -archives <archive1,...> specify a comma-separated list of archives to be unarchived on the compute machines

The general command line syntax is:
command [genericOptions] [commandOptions]

[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$
```

Вывести только начало файла: `hdfs dfs -cat urbandict.csv | head`

```
Терминал - student3_10@bigdataanalytics2-head-shdpt-v31-1-0:~
Файл  Правка  Вид  Терминал  Вкладки  Справка
-archives <archive1,...> specify a comma-separated list of archives to be unarchived on the compute machines

The general command line syntax is:
command [genericOptions] [commandOptions]

[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$ hdfs dfs -cat urbandict.csv | head
word_id,word,up_votes,down_votes,author,definition
0000007,Janky,296,255,dc397b2f,"Undesirable; less-than optimum."
0000008,sumpin',16,37,dc397b2f,"low down and funky, but [knee deep] enough to ride to."
0000009,yayeeeyay,19,27,dc397b2f,"affirmation; suggestion of encouragement, approval, or interest."
0000012,hard-core,162,96,d1610749,"anything out of our league that can be good or bad."
0000013,brutal,12,45,40ecelef,"anything that makes you sweat"
0000014,skanky,9,48,485e4db7,"Anything of or pertaining to a $10,000 hooker."
0000015,ho-bag,26,27,b37fba05,"a term of endearment, used affectionately for your roommate. First used in the schools' parking lot after an incident with the hall monitors."
0000016,massive,36,45,b9dcf126,"really really good. excellently good."
0000017,wtf,183,99,a6c97ba3,"what the fuck? ; use it in place of expletives. a more polite alternative."
cat: Unable to write to output stream.
[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$
```

Выведения конца файла: `hdfs dfs -tail urbandict.csv`

```
Терминал - student3_10@bigdataanalytics2-head-shdpt-v31-1-0:~
Файл  Правка  Вид  Терминал  Вкладки  Справка
0000014,skanky,9,48,485e4db7,"Anything of or pertaining to a $10,000 hooker."
0000015,ho-bag,26,27,b37fba05,"a term of endearment, used affectionately for your roommate. First used in the schools' parking lot after an incident with the hall monitors."
0000016,masive,36,45,b9dcf126,"really really good. excellently good."
0000017,wtf,183,99,a6c97ba3,"what the fuck? ;; use it in place of expletives. a more polite alternative."
cat: Unable to write to output stream.
[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$ hdfs dfs -tail urbandict.csv
43437,donald trump,14,14,377ad22f,"The reason hair care products were invented."
9043452,Nungry,0,0,428972c7,"The feeling of being both hungry and nauseated."
9043479,hemo,0,0,b47039c4,"Mix of hipster and emo"
9043502,Canadian bowtie,1,0,95405e7,"Condom, or similar prophylactic, made entirely of denim."
9043515,Slug,0,0,99f7906f,"A mix between a slag and a slut ;; 'That girl is such a (slug)' ;; 'go away you (slug)'"
9043521,Wikipedia,0,0,44335ab6,"An online encyclopedia. This site is often not encouraged by schools due to the fact it can be edited by anyone with an account. Usually used when you have an argument about something with your friends and need to quickly get an answer to prove them all wrong."
9043525,Giuseppe,1,1,dc81c268,"An annoying person but a good friend."
9043541,pacewhore,0,0,b5e3ec68,"Very fit and loving cat with a huge tail"
9043565,FML,1,0,bd352275,"Short for: Forgive me Lord"
9043586,Poptart,0,0,d6084af8,"Someone who is sweet and nice on the outside and slimy and dirty on the inside."
[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$
```

Изменяем файл репликации 2: `hdfs dfs -setrep 2 urbandict.csv`

```
Терминал - student3_10@bigdataanalytics2-head-shdpt-v31-1-0:~
Файл  Правка  Вид  Терминал  Вкладки  Справка
Please Note:
  1. By default fsck ignores files opened for write, use -openforwrite to report such files. They are usually tagged CORRUPT or HEALTHY depending on their block allocation status
  2. Option -includeSnapshots should not be used for comparing stats, should be used only for HEALTH check, as this may contain duplicates if the same file present in both original fs tree and inside snapshots.

Generic options supported are:
-conf <configuration file>      specify an application configuration file
-D <property=value>             define a value for a given property
-fs <file:///|hdfs://namenode:port> specify default filesystem URL to use, overrides 'fs.defaultFS' property from configurations.
-jt <local|resourcemanager:port> specify a ResourceManager
-files <file1,...>              specify a comma-separated list of files to be copied to the map reduce cluster
-libjars <jar1,...>            specify a comma-separated list of jar files to be included in the classpath
-archives <archive1,...>       specify a comma-separated list of archives to be unarchived on the compute machines

The general command line syntax is:
command [genericOptions] [commandOptions]

[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$ hdfs dfs -setrep 2 urbandict.csv
Replication 2 set: urbandict.csv
[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$
```

Удаляем файл `urbandict.csv` из HDFS без помещения в корзину: `hdfs dfs -rm -skipTrash urbandict.csv`

```
Терминал - student3_10@bigdataanalytics2-head-shdpt-v31-1-0:~
Файл  Правка  Вид  Терминал  Вкладки  Справка

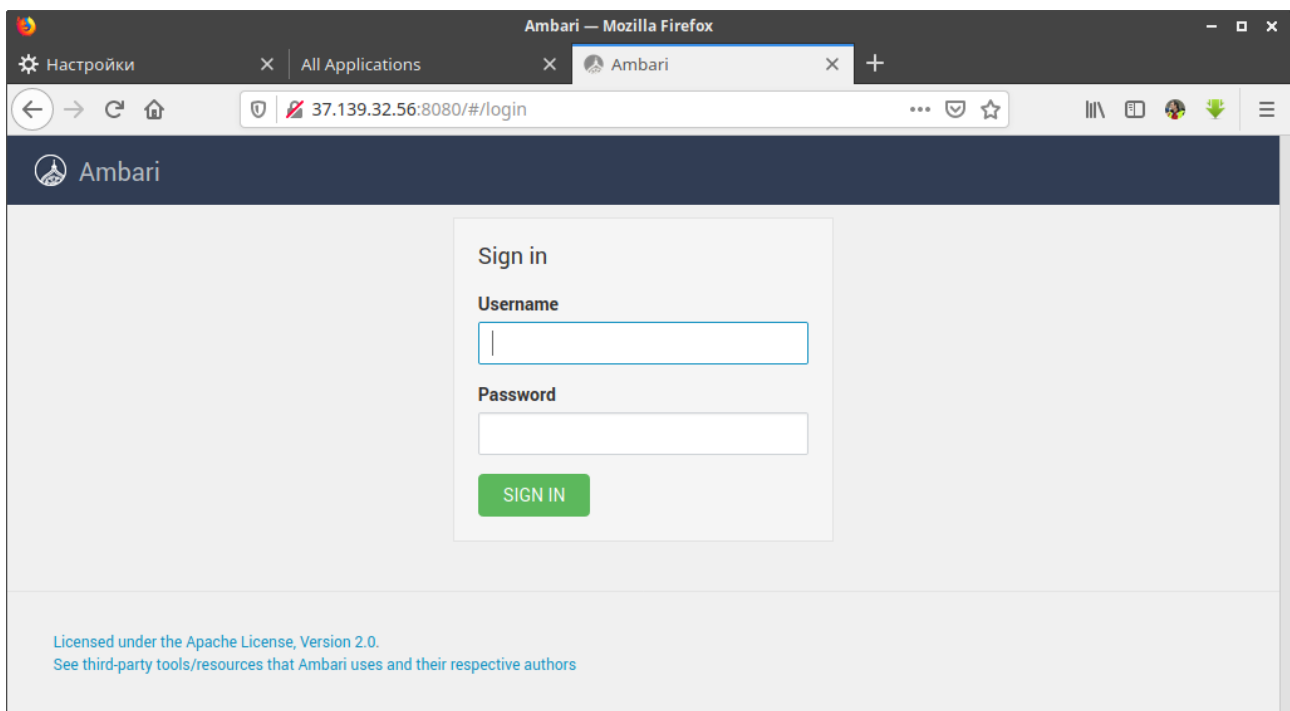
1. By default fsck ignores files opened for write, use -openforwrite to report such files. They are usually tagged
CORRUPT or HEALTHY depending on their block allocation status
2. Option -includeSnapshots should not be used for comparing stats, should be used only for HEALTH check, as this m
ay contain duplicates if the same file present in both original fs tree and inside snapshots.

Generic options supported are:
-conf <configuration file>          specify an application configuration file
-D <property=value>                 define a value for a given property
-fs <file:///|hdfs://namenode:port> specify default filesystem URL to use, overrides 'fs.defaultFS' property from configura
tions.
-jt <local|resourceManager:port>    specify a ResourceManager
-files <file1,...>                  specify a comma-separated list of files to be copied to the map reduce cluster
-libjars <jar1,...>                 specify a comma-separated list of jar files to be included in the classpath
-archives <archive1,...>            specify a comma-separated list of archives to be unarchived on the compute machines

The general command line syntax is:
command [genericOptions] [commandOptions]

[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$ hdfs dfs -setrep 2 urbandict.csv
Replication 2 set: urbandict.csv
[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$ hdfs dfs -rm -skipTrash urbandict.csv
Deleted urbandict.csv
[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$
```

Захожу на Ambari



Захожу на [YARN](#)

Cluster Metrics

Apps Submitted	Apps Pending	Apps Running	Apps Completed
104	0	0	104

Cluster Nodes Metrics

Active Nodes	Decommissioning Nodes
5	0

Scheduler Metrics

Scheduler Type	Scheduling Resource Type
Capacity Scheduler	[memory-mb (unit=Mi), vcores]

Show 20 entries

ID	User	Name	Application Type	Queue	Application Priority
application_1617704751423_0103	antong	HIVE-9696a235-74cd-4d76-a211-78404cdc727d	TEZ	default	0
application_1617704751423_0102	mihaile	HIVE-d9be1869-082a-4030-9b4e-96bbd24094ef	TEZ	default	0
application_1617704751423_0101	mihaile	HIVE-d9be1869-082a-4030-9b4e-96bbd24094ef	TEZ	default	0
application_1617704751423_0100	mihaile	HIVE-d9be1869-082a-4030-9b4e-96bbd24094ef	TEZ	default	0

Подключаюсь к Namenode

Overview 'bigdataanalytics2-head-shdpt-v31-1-0.novalocal:8020' (active)

Started:	Tue Apr 06 11:52:43 +0300 2021
Version:	3.1.1.3.1.4.0-315, r58d0fd3d8ce58b10149da3c717c45e57a60d14
Compiled:	Fri Aug 23 08:15:00 +0300 2019 by jenkins from (detached from 58d0fd3)
Cluster ID:	CID-8d3689bf-7c39-4e3c-a0b5-7fb6efe5fe78
Block Pool ID:	BP-1159087252-10.0.0.7-1616250271725

Summary

Security is off.

Safemode is off.

56 868 files and directories, 46 428 blocks (46 428 replicated blocks, 0 erasure coded block groups) = 103 296 total filesystem object(s).

Heap Memory used 135.97 MB of 1004 MB Heap Memory. Max Heap Memory is 1004 MB.

Non Heap Memory used 110.51 MB of 113.04 MB Committed Non Heap Memory. Max Non Heap Memory is <unbounded>.

просмотрим корневую директорию: `hdfs dfs -ls /`


```
Терминал - student3_10@bigdataanalytics2-head-shdpt-v31-1-0:~
Файл  Правка  Вид  Терминал  Вкладки  Справка
igor@igor-MS-7808:~$ ssh student3_10@37.139.32.56
Last login: Wed Apr  7 19:47:17 2021 from 109.252.26.235
[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$ ls
[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$ hdfs dfs -ls /
Found 13 items
drwxrwxrwt   - yarn   hadoop           0 2021-04-07 21:17 /app-logs
drwxr-xr-x   - hdfs   hdfs             0 2021-03-25 08:38 /apps
drwxr-xr-x   - yarn   hadoop           0 2021-03-20 14:24 /ats
drwxr-xr-x   - hdfs   hdfs             0 2021-03-20 14:25 /atsv2
drwxr-xr-x   - hdfs   hdfs             0 2021-03-20 14:24 /hdp
drwx-----  - livy    hdfs             0 2021-03-20 14:31 /livy2-recovery
drwxr-xr-x   - mapred  hdfs             0 2021-03-20 14:24 /mapred
drwxrwxrwx   - mapred  hadoop           0 2021-03-31 21:30 /mr-history
drwxrwxrwx   - spark   hadoop           0 2021-04-07 22:01 /spark2-history
drwxr-xr-x   - hdfs   hdfs             0 2021-04-06 09:11 /system
drwxrwxrwx   - hdfs   hdfs             0 2021-04-07 13:46 /tmp
drwxr-xr-x   - hdfs   hdfs             0 2021-04-02 16:10 /user
drwxr-xr-x   - hdfs   hdfs             0 2021-04-02 08:21 /warehouse
[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$ ^C
[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$ █
```

Создаем директорию в HDFS: `hdfs dfs -mkdir /data`

```
Терминал - student3_10@bigdataanalytics2-head-shdpt-v31-1-0:~
Файл  Правка  Вид  Терминал  Вкладки  Справка
igor@igor-MS-7808:~$ ssh student3_10@37.139.32.56
Last login: Wed Apr  7 22:00:20 2021 from 109.252.26.235
[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$ hdfs dfs -mkdir dir_name
[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$ hdfs dfs -ls
Found 1 items
drwxr-xr-x   - student3_10 student3_10           0 2021-04-07 22:21 dir_name
[student3_10@bigdataanalytics2-head-shdpt-v31-1-0 ~]$ █
```