

## РАБОТА С HDFS

1. Перешел в директорию, где лежит контейнер  
`cd /Downloads/docker-hadoop-master`
2. Запустил контейнер: (ПОДНИМАЕМ КОНТЕЙНЕРЫ)  
`sudo docker-compose up -d`

```
max@max:~/Downloads/docker-hadoop-master$ sudo docker-compose up -d
historyserver is up-to-date
resourcemanager is up-to-date
datanode is up-to-date
namenode is up-to-date
nodemanager is up-to-date
max@max:~/Downloads/docker-hadoop-master$
```

3. Подключаемся к Name Node:

`sudo docker exec -it namenode bash`

```
max@max:~/Downloads/docker-hadoop-master$ sudo docker exec -it namenode bash
root@08e6c8323525:/#
```

4. Смотрю в корень внутри файловой системы HDFS (нашего контейнера)  
`hdfs dfs -ls /`

```
root@08e6c8323525:/# hdfs dfs -ls /
Found 1 items
drwxr-xr-x - root supergroup 0 2021-12-02 17:36 /rmstate
root@08e6c8323525:/#
```

5. Создам папку на HDFS:  
`hdfs dfs -mkdir user_home`  
просмотрю что она создалась

```
root@08e6c8323525:/test_folder# hdfs dfs -ls
Found 1 items
drwxr-xr-x - root supergroup 0 2021-12-07 11:45 user_home
root@08e6c8323525:/test_folder#
```

6. Создал 3 файла

```
root@08e6c8323525:/test_folder# echo Hello, > test_file_1
root@08e6c8323525:/test_folder# cat test_file_1
Hello,
root@08e6c8323525:/test_folder# echo hdfs > test_file_2
root@08e6c8323525:/test_folder# echo ! > test_file_3
root@08e6c8323525:/test_folder# ls
test_file_1 test_file_2 test_file_3
```

7. Записал их в контейнер в папку user\_home  
`hdfs dfs -put test_file_* user_home`  
и просмотрю, что они появились в контейнере  
`hdfs dfs -ls user_home`

```
root@08e6c8323525:/test_folder# hdfs dfs -ls user_home
root@08e6c8323525:/test_folder# hdfs dfs -put test_file_* user_home
2021-12-07 11:49:27,682 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHo
2021-12-07 11:49:27,801 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHo
2021-12-07 11:49:28,229 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHo
root@08e6c8323525:/test_folder# hdfs dfs -ls user_home
Found 3 items
-rw-r--r-- 3 root supergroup 7 2021-12-07 11:49 user_home/test_file_1
-rw-r--r-- 3 root supergroup 5 2021-12-07 11:49 user_home/test_file_2
-rw-r--r-- 3 root supergroup 2 2021-12-07 11:49 user_home/test_file_3
root@08e6c8323525:/test_folder#
```

8. Делаю 2 репликации с этими файлами

hdfs dfs -setrep 2 user\_home/test\_file\_\*

```
root@08e6c8323525:/test_folder# hdfs dfs -setrep 2 user_home/test_file_1
Replication 2 set: user_home/test_file_1
root@08e6c8323525:/test_folder# hdfs dfs -setrep 2 user_home/test_file_{2,3}
Replication 2 set: user_home/test_file_2
Replication 2 set: user_home/test_file_3
root@08e6c8323525:/test_folder#
```

9. Создаю task\_2

hdfs dfs -mkdir task\_2

10. Далее на hdfs перемещаем их в папку task\_2

hdfs dfs -mv user\_home/test\_file\_\* task\_2

```
root@08e6c8323525:/# hdfs dfs -mv user_home/test_file_* task_2
root@08e6c8323525:/# hdfs dfs -ls task_2
Found 3 items
-rw-r--r--  2 root supergroup 7 2021-12-07 11:49 task_2/test_file_1
-rw-r--r--  2 root supergroup 5 2021-12-07 11:49 task_2/test_file_2
-rw-r--r--  2 root supergroup 2 2021-12-07 11:49 task_2/test_file_3
root@08e6c8323525:/#
```

11. Печатаем размер каждого файла с учётом репликации

hdfs dfs -du task\_2 | awk '{print \$2}'

```
root@08e6c8323525:/# hdfs dfs -du task_2 | awk '{print $2}'
14
10
4
```

12. Печатаем размер каждого файла с учётом репликации в log.txt

echo test\_file\_1..3 `hdfs dfs -du task\_2 | awk '{print \$2}'` > log.txt

```
root@08e6c8323525:/# echo test_file_1..3 `hdfs dfs -du task_2 | awk '{print $2}'` > log.txt
root@08e6c8323525:/# cat log.txt
test_file_1..3 14 10 4
root@08e6c8323525:/#
```

13. Удалил папку task\_2 с HDFS

hdfs dfs -rm -r -skipTrash task\_2

```
test_file_1 test_file_2 test_file_3
root@08e6c8323525:/test_folder# hdfs dfs -rm -r -skipTrash task_2
Deleted task_2
root@08e6c8323525:/test_folder#
```

14. Закинул файл со скриптом (через другой терминал, просто в системе) в папку test\_folder

docker cp ~/Downloads/script.sh namenode:/test\_folder

15. Надо поменять права на исполнение скрипту:

chmod +x script.sh

```
root@08e6c8323525:/test_folder# chmod +x script.sh
root@08e6c8323525:/test_folder# ls -h
script.sh
root@08e6c8323525:/test_folder# ls -lh
total 4.0K
-rwxrwxr-x 1 1000 1000 1.2K Dec  7 18:18 script.sh
root@08e6c8323525:/test_folder#
```

16. Запустим скрипт на выполнение:

```
max@max: ~/Downloads/docker-hadoop-master
root@08e6c8323525:/test_folder# ls
script.sh
root@08e6c8323525:/test_folder# ./script.sh
2021-12-08 10:07:32,487 INFO sasl.SaslDataTransferClient: SASL encryption
rusted = false, remoteHostTrusted = false
2021-12-08 10:07:32,577 INFO sasl.SaslDataTransferClient: SASL encryption
rusted = false, remoteHostTrusted = false
2021-12-08 10:07:32,603 INFO sasl.SaslDataTransferClient: SASL encryption
rusted = false, remoteHostTrusted = false
Replication 2 set: user_home/file_1
Replication 2 set: user_home/file_2
Replication 2 set: user_home/file_3
2021-12-08 10:07:39,469 INFO sasl.SaslDataTransferClient: SASL encryption
rusted = false, remoteHostTrusted = false
2021-12-08 10:07:40,985 INFO sasl.SaslDataTransferClient: SASL encryption
rusted = false, remoteHostTrusted = false
2021-12-08 10:07:42,529 INFO sasl.SaslDataTransferClient: SASL encryption
rusted = false, remoteHostTrusted = false
Deleted task_2
Deleted user_home
2021-12-08 10:07:37 14 10 4
2021-12-08 10:07:38 Hello,
2021-12-08 10:07:40 hdfs
2021-12-08 10:07:41 !
root@08e6c8323525:/test_folder# ls
log.txt  script.sh
root@08e6c8323525:/test_folder#
```

### Текст скрипта:

```
#!/bin/bash
#Создам папку на HDFS:
hdfs dfs -mkdir user_home

#Создал 3 файла
echo Hello, > file_1
echo hdfs > file_2
echo ! > file_3

#Записал их в контейнер в папку user_home
hdfs dfs -put file_* user_home

# Делаю 2 репликации с этими файлами
hdfs dfs -setrep 2 user_home/file_*

#Создаю task_2
hdfs dfs -mkdir task_2

#Далее на hdfs перемещаем файлы в папку task_2
hdfs dfs -mv user_home/file_* task_2

# Печатаем размер каждого файла с учётом репликации в log.txt
echo $(date +"%Y-%m-%d %H:%M:%S") `hdfs dfs -du task_2 | awk '{print $2}` > log.txt
echo $(date +"%Y-%m-%d %H:%M:%S") `hdfs dfs -cat task_2/file_1` >> log.txt
echo $(date +"%Y-%m-%d %H:%M:%S") `hdfs dfs -cat task_2/file_2` >> log.txt
echo $(date +"%Y-%m-%d %H:%M:%S") `hdfs dfs -cat task_2/file_3` >> log.txt

#удаляем все созданные файлы как на hdfs (не забываем -skipTrash), так и на локальной
машине
hdfs dfs -rm -r -skipTrash task_2
hdfs dfs -rm -r -skipTrash user_home
rm file_*
cat log.txt
```