Практическое задание к лекции по HDFS
1. Опробовать консольные утилиты для работы с кластером
* Создать/скопировать/удалить папку
* Положить в HDFS любой файл
* Скопировать/удалить этот файл
* Просмотреть размер любой папки
* Посмотреть как файл хранится на файловой системе (см. команду fsck)
* Установить нестандартный фактор репликации (см. команду setrep)
3. Опробовать rest-доступ для работы с кластером
* Используя утилиту CURL
* Используя python3
5. [Для любителей администрирования] Опробовать NFS доступ. Предварительно связаться со мной чтобы я открыл нужные порты.
6. [Для любителей программирования] Достучаться до файловой системы используя python и библиотеку libhdfs3

1.
1) Создал папку   `student2 11@manager ~]$ hdfs dfs -mkdir /t f 1`

2) Скопировал папку и убедился в наличии их обеих

```
[student2_11@manager ~]$ hdfs dfs -cp /t_f_1 /t_f_2
[student2_11@manager ~]$ hdfs dfs -ls /
Found 35 items
drwxr-xr-x   - student2_4    supergroup          0 2020-02-08 15:05 /Kornienko_I
drwxrwxrwx   - dekan         supergroup          0 2019-12-23 20:51 /NetCat
drwxr-xr-x   - student2_4    supergroup          0 2020-02-08 14:13 /Petrov_22
drwxr-xr-x   - student2_9    supergroup          0 2020-02-09 20:05 /SeredinGS
drwxr-xr-x   - student2_10 supergroup            0 2020-02-08 20:46 /TEST
drwxr-xr-x+  - centos        supergroup          0 2019-11-25 15:33 /acldir
-rw-r--r--   3 student7       supergroup     260933 2019-12-16 19:35 /amazon_dataset
drwxrwxrwx   - hdfs          supergroup          0 2019-12-03 11:19 /data
drwxr-xr-x   - student7       supergroup          0 2019-12-16 19:43 /dataset
drwxr-xr-x   - student2_6    supergroup          0 2020-02-08 16:02 /filkin
drwxr-xr-x   - hbase         hbase               0 2020-02-01 11:29 /hbase
drwxr-xr-x   - student2      supergroup          0 2019-12-03 09:08 /ho
drwxr-xr-x   - student2      supergroup          0 2019-12-03 09:10 /home
drwxr-xr-x   - student2_5    supergroup          0 2020-02-10 08:10 /ivanov_1
drwxr-xr-x   - student2_7    supergroup          0 2020-02-10 05:59 /new_1
drwxr-xr-x   - student2_7    supergroup          0 2020-02-10 06:01 /new_2
drwxr-xr-x   - hdfs          supergroup          0 2020-02-04 18:12 /newfolder
drwxr-xr-x   - student2_10 supergroup            0 2020-02-08 19:46 /pugach
drwxr-xr-x   - student2_3    supergroup          0 2020-02-08 13:06 /shlyapka_test
drwxr-xr-x   - student2_3    supergroup          0 2020-02-08 11:19 /shlyapka_test1
drwxr-xr-x   - student2_11 supergroup            0 2020-02-10 12:59 /t f 1
drwxr-xr-x   - student2_11 supergroup            0 2020-02-10 13:03 /t f 2
drwxr-xr-x   - student7       supergroup          0 2019-11-28 16:14 /test
drwxr-xr-x   - student7       supergroup          0 2019-11-28 16:23 /test2
drwxr-xr-x   - hdfs          supergroup          0 2019-12-03 11:11 /test_datasets
-rw-r--r--   3 student7       supergroup      35074 2019-11-26 20:14 /test_dir1
-rw-r--r--   3 student7       supergroup      35074 2019-11-26 20:07 /test_dir2
drwxr-xr-x   - student2_11 supergroup            0 2020-02-10 12:49 /test_f1
drwxr-xr-x   - student2_5    supergroup          0 2020-02-10 08:02 /test_folder1
drwxr-xr-x   - student2_11 supergroup            0 2020-02-09 16:16 /test_larin1
drwxrwxrwx   - hdfs          supergroup          0 2020-01-14 13:08 /tmp
drwxr-xr-x   - student9      supergroup          0 2019-12-01 18:00 /tvoretz_dir
drwxr-xr-x   - hdfs          supergroup          0 2019-12-03 17:55 /ufctable
drwxr-xr-x   - hdfs          supergroup          0 2020-01-21 08:35 /user
drwxr-xr-x   - student2_4    supergroup          0 2020-02-08 13:56 /user_Petrov_Egor
```

3) При попытке удалить получил длинное сообщение об ошибке в правах доступа

```
[student2_ll@manager ~]$ hdfs dfs -rm -r /t_f_1
20/02/10 13:09:46 WARN fs.TrashPolicyDefault: Can't create trash directory: hdfs://manager.novalocal:8020/user/student2_ll/.
Trash/Current
org.apache.hadoop.security.AccessControlException: Permission denied: user=student2_ll, access=WRITE, inode="/user":hdfs:sup
ergroup:drwxr-xr-x
        at org.apache.hadoop.hdfs.server.namenode.DefaultAuthorizationProvider.checkFsPermission(DefaultAuthorizationProvide
r.java:279)
        at org.apache.hadoop.hdfs.server.namenode.DefaultAuthorizationProvider.check(DefaultAuthorizationProvider.java:260)
        at org.apache.hadoop.hdfs.server.namenode.DefaultAuthorizationProvider.check(DefaultAuthorizationProvider.java:240)
        at org.apache.hadoop.hdfs.server.namenode.DefaultAuthorizationProvider.checkPermission(DefaultAuthorizationProvider.
java:162)
        at org.apache.hadoop.hdfs.server.namenode.FSPermissionChecker.checkPermission(FSPermissionChecker.java:152)
        at org.apache.hadoop.hdfs.server.namenode.FSDirectory.checkPermission(FSDirectory.java:3885)
        at org.apache.hadoop.hdfs.server.namenode.FSDirectory.checkPermission(FSDirectory.java:3868)
        at org.apache.hadoop.hdfs.server.namenode.FSDirectory.checkAncestorAccess(FSDirectory.java:3850)
        at org.apache.hadoop.hdfs.server.namenode.FSNamesystem.checkAncestorAccess(FSNamesystem.java:6826)
        at org.apache.hadoop.hdfs.server.namenode.FSNamesystem.mkdirsInternal(FSNamesystem.java:4562)
        at org.apache.hadoop.hdfs.server.namenode.FSNamesystem.mkdirsInt(FSNamesystem.java:4532)
        at org.apache.hadoop.hdfs.server.namenode.FSNamesystem.mkdirs(FSNamesystem.java:4505)
        at org.apache.hadoop.hdfs.server.namenode.NameNodeRpcServer.mkdirs(NameNodeRpcServer.java:884)
        at org.apache.hadoop.hdfs.server.namenode.AuthorizationProviderProxyClientProtocol.mkdirs(AuthorizationProviderProxy
ClientProtocol.java:328)
        at org.apache.hadoop.hdfs.protocolPB.ClientNamenodeProtocolServerSideTranslatorPB.mkdirs(ClientNamenodeProtocolServe
rSideTranslatorPB.java:641)
        at org.apache.hadoop.hdfs.protocol.proto.ClientNamenodeProtocolProtos$ClientNamenodeProtocol$2.callBlockingMethod(Cl
ientNamenodeProtocolProtos.java)
        at org.apache.hadoop.ipc.ProtobufRpcEngine$Server$ProtoBufRpcInvoker.call(ProtobufRpcEngine.java:617)
        at org.apache.hadoop.ipc.RPC$Server.call(RPC.java:1073)
        at org.apache.hadoop.ipc.Server$Handler$1.run(Server.java:2278)
        at org.apache.hadoop.ipc.Server$Handler$1.run(Server.java:2274)
        at java.security.AccessController.doPrivileged(Native Method)
        at javax.security.auth.Subject.doAs(Subject.java:422)
        at org.apache.hadoop.security.UserGroupInformation.doAs(UserGroupInformation.java:1924)
        at org.apache.hadoop.ipc.Server$Handler.run(Server.java:2272)
```

Получилось удалить папку используя ключ -skipTrash

```
[student2_ll@manager ~]$ hdfs dfs -rm -r -skipTrash /t_f_1
Deleted /t_f_1
```

4) Создал файл на сервере

```
[student2_ll@manager ~]$ vi test.txt
[student2_ll@manager ~]$ ll
total 8
-rw-rw-r--. 1 student2_ll student2_ll 4 Feb 10 15:11 123.txt
-rw-rw-r--. 1 student2_ll student2_ll 9 Feb 10 15:27 test.txt
```

Скопировал в директорию

```
[student2_ll@manager ~]$ hdfs dfs -copyFromLocal test.txt /t_f_2
[student2_ll@manager ~]$ hdfs dfs -ls /t_f_2
Found 1 items
-rw-r--r--   3 student2_ll supergroup          9 2020-02-10 15:32 /t_f_2/test.tx
t
```

Удалил файл

```
[student2_ll@manager ~]$ hdfs dfs -rm -skipTrash /t_f_2/test.txt
Deleted /t_f_2/test.txt
```

5) Посмотрел размер папки

```
[student2_ll@manager ~]$ hdfs dfs -du -h -s /t_f_2
4  12  /t_f_2
```

6) Посмотрел как файл хранится в файловой системе

```
[student2_11@manager ~]$ hdfs fsck /t_f_2/123.txt
Connecting to namenode via http://manager.novalocal:50070/fsck?ugi=student2_11&path=%2Ft_f_2%2F123.txt
FSCK started by student2_11 (auth:SIMPLE) from /89.208.221.132 for path /t_f_2/123.txt at Mon Feb 10 15:4
9:12 UTC 2020
.Status: HEALTHY
 Total size:    4 B
 Total dirs:    0
 Total files:   1
 Total symlinks:                0
 Total blocks (validated):      1 (avg. block size 4 B)
 Minimally replicated blocks:   1 (100.0 %)
 Over-replicated blocks:        0 (0.0 %)
 Under-replicated blocks:       0 (0.0 %)
 Mis-replicated blocks:         0 (0.0 %)
 Default replication factor:    3
 Average block replication:     3.0
 Corrupt blocks:                0
 Missing replicas:              0 (0.0 %)
 Number of data-nodes:          3
 Number of racks:               1
FSCK ended at Mon Feb 10 15:49:12 UTC 2020 in 0 milliseconds


The filesystem under path '/t_f_2/123.txt' is HEALTHY
```

7) Сделал количество репликаций файла равное 4

```
[student2_11@manager ~]$ hdfs dfs -setrep 4 /t_f_2/123.txt
Replication 4 set: /t_f_2/123.txt
```

3.
1) Зашел через curl под юзером hdfs

```
[student2_11@manager ~]$ curl -X GET 'http://node2.novalocal:14000/webhdfs/v1/acldir?user.name=hdfs&op=L
ISTSTATUS'
{"FileStatuses":{"FileStatus":[{"pathSuffix":"etc","type":"DIRECTORY","length":0,"owner":"centos","group
":"supergroup","permission":"755","accessTime":0,"modificationTime":1574696487181,"blockSize":0,"replica
tion":0}]}}
```

Зашел через curl под юзером student2_11

```
[student2_11@manager ~]$ curl -X GET 'http://node2.novalocal:14000/webhdfs/v1/t_f_2?user.name=student2_1
1&op=LISTSTATUS'
{"FileStatuses":{"FileStatus":[{"pathSuffix":"123.txt","type":"FILE","length":4,"owner":"student2_11","g
roup":"supergroup","permission":"644","accessTime":1581349149704,"modificationTime":1581349149958,"block
Size":134217728,"replication":4},{"pathSuffix":"boot.log","type":"FILE","length":104,"owner":"student2_1
1","group":"supergroup","permission":"644","accessTime":1581350351561,"modificationTime":1581350352372,"
blockSize":134217728,"replication":3}]}}
[student2_11@manager ~]$ curl -X GET 'http://node2.novalocal:14000/webhdfs/v1/t_f_2/boot.log?user.name=s
tudent2_11&op=LISTSTATUS'
{"FileStatuses":{"FileStatus":[{"pathSuffix":"","type":"FILE","length":104,"owner":"student2_11","group"
:"supergroup","permission":"644","accessTime":1581350351561,"modificationTime":1581350352372,"blockSize"
:134217728,"replication":3}]}}
[student2_11@manager ~]$
```