


Name	Last Update
../(agorokhov/mcs-2017/tree/master/workshops/01-hdfs/..)	
 README.md (/agorokhov/mcs-2017/blob/master/workshops/01-hdfs/README.md)	5 months ago

 **README.md** (/agorokhov/mcs-2017/blob/master/workshops/01-hdfs/README.md)

Семинар: HDFS (Курс МВС, Распределенная часть)

Структура кластера и роли серверов

Кластер:

```
Namenode: mipt-master.atp-fivt.org
Datanodes: mipt-nodeXX.atp-fivt.org    (7 штук)
```

Клиентская машина:

```
mipt-client.atp-fivt.org
```

На клиентскую машину заходим по ssh, только с нее есть доступ к кластеру. На сервера кластера по ssh в обычной жизни не ходим, работаем с ним клиентскими утилитами (hdfs, yarn, hadoop).

Но мы в учебных целях будем ходить на ноды, чтобы посмотреть как там устроено хранение. Для этого есть специальный юзер:

```
login: hdfsuser
password: hdfsuser
```

Зайти на любую датаноду, например на mipt-node08, можно командой:

```
$ sudo -u hdfsuser ssh hdfsuser@mipt-node08.atp-fivt.org
```

Основные команды

Зайти на клиентский сервер. Показать и дать попробовать команды:

- Содержимое директорий:

```
$ hdfs dfs -ls /  
$ hdfs dfs -ls .  
$ hdfs dfs -ls
```

последние 2 эквивалентны и относятся к домашней директории; показать, где находится домашняя директория в hdfs

- Создать директорию, загрузить в нее какой-нибудь текстовый файл:

```
$ hdfs dfs -mkdir <dir>  
$ hdfs dfs -put <file>  
$ hdfs dfs -cat <file>
```

В последней можно поменять -cat на -text. Создать файл README:

```
$ echo "read it" > README  
$ echo "read it again" >> README  
$ cat README
```

Загрузить его в HDFS, распечатать оттуда с помощью (-cat)

- Переименование и копирование:

```
$ hdfs dfs -mv <file1> <file2>  
$ hdfs dfs -cp <file1> <file2>
```

Важно: копирование идет через клиента, для избежания этого или копирования вообще между кластерами есть команда `hadoop distcp`

- Исследовать директорию /data

```
$ hdfs dfs -ls /data
```

Вывести на экран часть какого-нибудь файла с помощью `more` (`less`) и `head`.
Например, 2 первые строки из файла с Википедией:

```
$ hdfs dfs -cat /data/wiki/en_articles/articles | head -2
```

- Объяснить вывод команды `ls`, например:

```
$ hdfs dfs -ls /data/wiki/en_articles
Found 1 items
-rw-r--r--    3 hdfs supergroup 12328051927 2017-07-03 23:21 /data/wiki/en_art
```

выводит пермиссии, число реплик, пользователя, группу, размер, время модификации. Посмотреть, что это совпадает со справкой в

```
$ hdfs dfs -help
```

- Сколько места занимают файлы и директории, например:

```
$ hdfs dfs -du -h /data/wiki
11.5 G   34.4 G   /data/wiki/en_articles
```

выводит размер и занимаемое место. Последнее учитывает фактор репликации (3)

- Изменить фактор репликации:

```
$ hdfs dfs -setrep <path>
```

Посмотреть на эффект с помощью ls. (ставим флаг -w если хотим дожидаться завершения команды, а не выполнять в фоновом режиме)

- Удалить файл:

```
$ hdfs dfs -rm <file>
```

или директорию

```
$ hdfs dfs -rm -r <file>
```

Остальные команды: -help

Веб-интерфейс Namenode

Зайти на <http://mipt-master.atp-fivt.org:50070> (<http://mipt-master.atp-fivt.org:50070>)

Если доступа нет, пробросить порт 50070:

```
$ ssh -L 50070:mipt-master.atp-fivt.org:50070 mipt-client.atp-fivt.org
```

Тогда пока сессия ssh открыта, интерфейс будет тут: <http://localhost:50070>
(<http://localhost:50070>)

Показать:

- Heap memory used/total - сколько памяти на NN и насколько занята. Сильно занята - проблемы.
- Capacity - сколько всего места в DFS
- DFS Used - сколько занято
- Live nodes - сколько живых датанод и насколько заполнены
- Utilities -> Browse the Filesystem

Структура хранения в HDFS

HDFS fsck

Посмотреть, на какие блоки разбит файл, какого они размера, их идентификаторы, где они расположены:

```
$ hdfs fsck /data/wiki/en_articles_part -files -blocks -locations
```

Посмотреть информацию по одному блоку, например:

```
$ hdfs fsck -blockId blk_1073971176
```

WebHDFS REST API

Открыть файл на чтение, получив редирект на нужную датаноду:

```
$ curl -i "http://mipt-master.atp-fivt.org:50070/webhdfs/v1/data/wiki/en_articles"
```

Из результата предыдущей команды взять Location, добавить &length=100 - получить 100 первых символов файла.

Подробнее: <http://hadoop.apache.org/docs/r1.2.1/webhdfs.html>
(<http://hadoop.apache.org/docs/r1.2.1/webhdfs.html>)

Оценка характеристик кластера

Задача 1: Для HDFS ёмкостью 2PB с размером блока 64MB и фактором репликации 3 оценить минимальный объем RAM на Namenode. Считаем, что каждый блок (независимо от числа реплик) занимает в памяти 600B.

Решение: 2PB / 64MB - число блоков.

Разных блоков будет: 2PB / 64MB / 3

Требуемая память: $(2PB / 64MB / 3) * 600B = 6.25GB$

Задача 2: На кластере используются HDD с характеристиками: seek time - 12мс, скорость линейного чтения - 210 MB/с. Какой минимальный размер блока в HDFS стоит выбрать, чтобы время поиска блока составляло не больше 0.5% от времени чтения блока?

Решение: 0.5% от времени чтения блока размера X должен быть не меньше 12мс:

$$X / 210 \text{ MB/с} * 0.005 \geq 0.012 \text{ с}$$

$$X \geq 0.012 * 210 / 0.005 = 504 \text{ MB}$$