

## Домашнее задание по HDFS

В этом домашнем задании Вам предстоит поближе познакомиться с распределенной файловой системой HDFS. Ваша цель - написать скрипты, которые будут делать HTTP запросы к различным демонам HDFS или вызывать стандартные утилиты командной строки (shell) с целью получения запрашиваемой информации.

Можно использовать любой скриптовый язык программирования. Самый простой вариант - использовать bash, а также утилиты curl, grep, wc и т.п.

### Задачи

**01 (0.1 балла)** На вход скрипту подается имя файла, на выходе нужно получить имя сервера или IP-адрес, с которого будет читаться первый блок данных (реплик может быть несколько, засчитываться будет любой из них). Пример:

```
$ ./run.sh /data/access_logs/big_log/access.log.2015-12-10
mipt-node01.atp-fivt.org
```

**02 (0.1 балла)** На вход скрипту подается имя файла, на выходе нужно получить первые 10 байт этого файла (hadoop fs и hdfs dfs использовать нельзя)

```
$ ./run.sh /data/access_logs/big_log/access.log.2015-12-10
41.190.60.
```

**03 (0.1 балла)** На вход скрипту подается полный путь до файла в HDFS, на выходе нужно получить размер файла в блоках (см. hdfs fsck -h).

```
$ ./run.sh /data/access_logs/big_log/access.log.2015-12-10
8
```

**04 (0.3 балла)** На вход скрипту подается идентификатор блока, на выходе нужно получить имя сервера (если их несколько, то выбрать любой), где хранится данный блок и физический путь в локальной файловой системе до этого блока данных.

```
$ ./run.sh blk_1075127191
bds03.vdi.mipt.ru:/dfs/dn/current/BP-76251478-10.55.163.141-1427134131440/current/finalized/
```