### Департамент образования и науки города Москвы

# Государственное автономное образовательное учреждение высшего образования города Москвы «Московский городской педагогический университет»

Институт цифрового образования Департамент информатики, управления и технологий

Башкатова Анна Денисовна

#### ОТЧЕТ

по дисциплине «Инструменты для хранения и обработки больших данных» Тема: «5.2. Практическая работа Hadoop»

Направление подготовки (специальность) 38.03.05 – бизнес-информатика Направленность (профиль) образовательной программы «Аналитика данных и эффективное управление»

Курс обучения: 3

Форма обучения: очная

Руководитель: Босенко Т. М.

## Вывести с помощью команды help описание основных команды shellклиента:

```
hadoop@bashkatova-VirtualBox:~$ help
GNU bash, версия 4.4.20(1)-release (x86 64-pc-linux-qnu)
Показанные ниже команды определены внутри командного процессора. Чтобы вывести
полный список, введите «help».
Чтобы вывести справку по функции «name», введите «help name».
Чтобы вывести справку по командному процессору, введите «info bash».
Чтобы вывести справку по командам, которые отсутствуют в этом списке, введите «m
an -k» или «info».
Звёздочка (*) рядом с названием команды означает, что эта команда отключена.
 задание [&]
                                           history [-c] [-d смещение] [n] или h>
                                           if КОМАНДЫ; then КОМАНДЫ; [ elif KOM>
jobs [-lnprs] [задание ...] или jobs>
 (( выражение ))
 . файл [аргументы]
                                           kill [-s назв_сигнала | -п номер_сиг>
                                           let аргумент [аргумент ...]
 [ аргумент... ]
 [[ выражение ]]
                                           local [параметр] имя[=значение] ...
 alias [-p] [имя[=значение] ...]
                                           logout [n]
 bg [задание ...]
                                           mapfile [-d delim] [-n count] [-0 or>
 bind [-lpvsPSVX] [-m раскладка] [-f ф>
                                           popd [-n] [+N | -N]
                                           printf [-v переменная] формат [аргум>
 break [n]
 builtin [встр_команда [аргумент ...]>
                                           pushd [-n] [+N | -N | каталог]
 caller [выражение]
                                           pwd [-LP]
 case СЛОВО in [ШАБЛОН [| ШАБЛОН]...) >
                                           read [-ers] [-а массив] [-d разделит>
 cd [-L|[-P [-e]] [-@]] [каталог]
                                           readarray [-n число] [-О начало] [-s>
 command [-pVv] команда [аргумент ...>
                                           readonly [-aAf] [имя[=значение] ...]>
 compgen [-abcdefgjksuv] [-o option] [>
                                           return [n]
 complete [-abcdefgjksuv] [-pr] [-DE] > compopt [-o|+o параметр] [-DE] [имя .>
                                           select ИМЯ [in СЛОВА ... ;] do KOMAH>
                                           set [-abefhkmnptuvxBCHP] [-о парамет>
                                           shift [n]
 continue [n]
 соргос [ИМЯ] команда [перенаправления>
                                           shopt [-pqsu] [-o] [параметр ...]
 declare [-aAfFgilnrtux] [-р] [имя[=зн>
                                           source файл [аргументы]
 dirs [-clpv] [+N] [-N]
                                           suspend [-f]
 disown [-h] [-ar] [jobspec ... | pid >
                                           test [выражение]
 есho [-neE] [аргумент ...]
                                           time [-p] конвейер
 enable [-a] [-dnps] [-f файл] [имя ..>
                                           times
 eval [apryment ...]
                                           trap [-lp] [[аргумент] сигнал ...]
 exec [-cl] [-а имя] [команда] [аргуме>
                                           true
 exit [n]
                                           type [-afptP] имя [имя ...]
 export [-fn] [имя[=значение ...] или > typeset [-aAfFgilnrtux] [-p] имя[=зн>
```

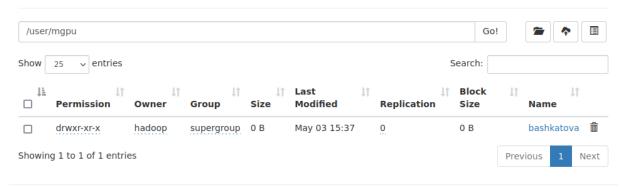
### Просмотреть корневую директорию HDFS:

```
hadoop@bashkatova-VirtualBox:~$ hadoop fs -ls /
2023-05-03 15:34:44,978 WARN util.NativeCodeLoader: Unable to load native-hadoop
library for your platform... using builtin-java classes where applicable
Found 2 items
drwx----- - hadoop supergroup 0 2023-05-03 14:54 /tmp
drwxr-xr-x - hadoop supergroup 0 2023-05-03 14:52 /user
hadoop@bashkatova-VirtualBox:~$
```

## Создать в HDFS в директории /user/mgpu поддиректорию ваше\_фио:

```
hadoop@bashkatova-VirtualBox:~$ hadoop fs -mkdir -p /user/mgpu/bashkatova
2023-05-03 15:37:54,319 WARN util.NativeCodeLoader: Unable to load native-hadoop
library for your platform... using builtin-java classes where applicable
hadoop@bashkatova-VirtualBox:~$
```

## **Browse Directory**



Создать в локальной файловой системе случайный текстовый файл размером 10 Mb с именем, образованным вашими инициалами:

```
hadoop@bashkatova-VirtualBox:~$ base64 /dev/urandom | head -c 10000000 > bad.txt
```

Заархивировать созданный текстовый файл:

```
hadoop@bashkatova-VirtualBox:~$ gzip -c bad.txt > bad.gz
```

Скопировать текстовый файл и архив в директорию /user/mgpu/fio HDFS виртуальной машины:

```
hadoop@bashkatova-VirtualBox:~$ hadoop fs -put bad.txt /user/mgpu/bashkatova 2023-05-03 15:49:16,876 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable hadoop@bashkatova-VirtualBox:~$ hadoop fs -put bad.gz /user/mgpu/bashkatova 2023-05-03 15:49:49,964 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable hadoop@bashkatova-VirtualBox:~$
```

Просмотреть файл и архив с помощью утилит cat, text в комбинации с каналами и утилитами head, tail -- привести не менее 3 вариантов команд и просмотра файла:

```
hadoop@bashkatova-VirtualBox:~$ hadoop fs -cat /user/mgpu/bashkatova/bad.txt | h ead -3 2023-05-03 15:55:30,761 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable ZRGfENkjYwi7xBopTjcHV/9dRZDa5CEeid2FrNAbmqLKzpZMdXKjrdn/8J1AaogIYEQtwN7twvjI /ZH2XgNRathY1GhxPYkA0WIwOwHZBBz41z5fI7DqYyFcY5KTGNlOlBE1zt1X0gtRrW2EqvMuuQWz VQoKw+EicNY4UbQmL78NKP4SHFhdVK7H9g4N5mg77XjXkEkTdFDyEAn1rc4r2lGvfq5oSKEQxt9D cat: Unable to write to output stream. hadoop@bashkatova-VirtualBox:~$
```

```
hadoop@bashkatova-VirtualBox:~$ hadoop fs -text /user/mgpu/bashkatova/bad.txt |
head -3
2023-05-03 16:06:30,104 WARN util.NativeCodeLoader: Unable to load native-hadoop
library for your platform... using builtin-java classes where applicable
ZRGfENkjYwi7xBopTjcHV/9dRZDa5CEeid2FrNAbmqLKzpZMdXKjrdn/8J1AaogIYEQtwN7twvjI
/ZH2XgNRathY1GhxPYkA0WIw0wHZBBz41z5fI7DqYyFcY5KTGNl0lBE1zt1X0gtRrW2EqvMuuQWz
VQoKw+EicNY4UbQmL78NKP4SHFhdVK7H9q4N5mq77XjXkEkTdFDyEAn1rc4r2lGvfq5oSKEQxt9D
text: Unable to write to output stream.
hadoop@bashkatova-VirtualBox:~$ hadoop fs -cat /user/mgpu/bashkatova/bad.gz | he
ad -3
2023-05-03 16:18:48,566 WARN util.NativeCodeLoader: Unable to load native-hadoop
 library for your platform... using builtin-java classes where applicable
/◆◆췸[點t◆◆|◆SB◆[點◆◆◆gXt◆◆S◆◆◆V-F◆Ra&◆j%[點◆2◆◆:[◆Le$◆M7î ,}o;◆[ឍR.◆◆S◆}1◆◆+◆◆◆◆[卧
關|◆癸閏₩◆C◆◆◆5B◆◆4◆ÉⅢ◆
                        00
18 0 L000
                              :I000y00-B000 1200& 00:00
                                                     •@$••@$•••:w@$•••:w@$•••
ožBounoowooDoAooood\ooo1űao盟oooDq?oF ooL[弱o閉園oio$oyooDooroo`O
◆_||弱◆◆V◆◆◆b||1◆◆||3H2N◆◆◆◆%◆H|?◆◆'||間對U]S◆F◆:◆N◆◆V:$ ||Z◆◆0◆||8→◆◆◆||i◆!◆S|3\W◆J◆ut◆V◆◆
◆聞◆聞++◆問◆◆女[��]別Y(◆{◆◆{廳◆◆SC◆◆H}Q◆?賢;}◆◆問◆賢◆~=◆`◆◆!◆5◆-"◆◆Lh◆◆◆U◆問◆U◆◆
000'Ito BPP 00Z80100 BB0000L: yF000000/ Bomo; Po Bb000ffK00m pX00@po\),00,00[top
Ke | Se, e~Teeee
◆,◆◆◆聞動◆?◆◆◆問題◆J
                                       KIROO (III | gooovPgoHMo osino iii) jo/o\gE
                   000000d#K0`000{
                                                                           ⊕3
◆~◆I-開發4a閉y◆◆BoN間 | ◆◆器が◆◆◆◆◆tG◆◆◆◆◆◆◆Q門3q)◆◆◆◆◆!◆B(◆G門 DZ肥Y◆<◆,_門]配3門
cat: Unable to write to output stream.
hadoop@bashkatova-VirtualBox:~$
cal: Unable to write to outbut stream.
hadoop@bashkatova-VirtualBox:~$ hadoop fs -text /user/mgpu/bashkatova/bad.gz | h
ead -3
2023-05-03 16:19:36,124 WARN util.NativeCodeLoader: Unable to load native-hadoop
library for your platform... using builtin-java classes where applicable
ZRGfENkjYwi7xBopTjcHV/9dRZDa5CEeid2FrNAbmqLKzpZMdXKjrdn/8J1AaoqIYE0twN7twvjI
```

! Команда чтения файлов text — это аналог команды cat, но разархивирует архивы.

/ZH2XgNRathY1GhxPYkA0WIwOwHZBBz41z5fI7DqYyFcY5KTGNl0lBE1zt1X0gtRrW2EqvMuuQWz VQoKw+EicNY4UbQmL78NKP4SHFhdVK7H9g4N5mg77XjXkEkTdFDyEAn1rc4r2lGvfq5oSKEQxt9D

text: Unable to write to output stream.

Создать копию файла file.txt вида date\_file.txt, где в начале имени файлакопии указана текущая дата. Вывести листинг:

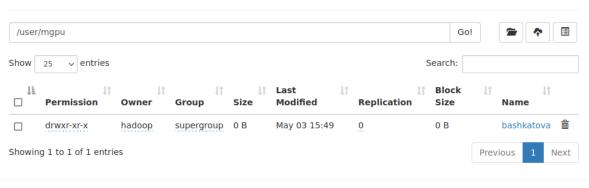
```
hadoop@bashkatova-VirtualBox:~$ hadoop fs -get /user/mgpu/bashkatova/bad.txt 03-
05-2023-bad.txt
2023-05-03 16:25:01,646 WARN util.NativeCodeLoader: Unable to load native-hadoop
library for your platform... using builtin-java classes where applicable
2023-05-03 16:25:03,457 WARN hdfs.DFSClient: Found Checksum error for BP-1242662
641-127.0.1.1-1683114589891:blk_1073741868_1044 from DatanodeInfoWithStorage[127
.0.0.1:9866,DS-942c327e-8b21-4d3d-b7dd-291f295a0f9a,DISK] at 4579328
2023-05-03 16:25:03,467 WARN hdfs.DFSClient: No live nodes contain block BP-1242
662641-127.0.1.1-1683114589891:blk_1073741868_1044 after checking nodes = [Datan
odeInfoWithStorage[127.0.0.1:9866,DS-942c327e-8b21-4d3d-b7dd-291f295a0f9a,DISK]]
, ignoredNodes = null
2023-05-03 16:25:03,473 INFO hdfs.DFSClient: Could not obtain BP-1242662641-127.
0.1.1-1683114589891:blk_1073741868_1044 from any node: No live nodes contain cu
rrent block Block locations: DatanodeInfoWithStorage[127.0.0.1:9866,DS-942c327e-
8b21-4d3d-b7dd-291f295a0f9a,DISK] Dead nodes: DatanodeInfoWithStorage[127.0.0.1
:9866,DS-942c327e-8b21-4d3d-b7dd-291f295a0f9a,DISK]. Will get new block location
s from namenode and retry...
2023-05-03 16:25:03,474 WARN hdfs.DFSClient: DFS chooseDataNode: got # 1 IOExcep
tion, will wait for 2427.4810781171923 msec.
get: Checksum CRC32C not matched for file /user/mgpu/bashkatova/bad.txt at posit
ion 4579328: expected=A5AFF22C but computed=FF53E4EB, algorithm=PureJavaCrc32C
hadoop@bashkatova-VirtualBox:~$ ls
                  Faust_1_Count_Output.csv hs_err_pid15635.log
bad.txt
                  Faust_1_Output.csv
                                            hs_err_pid15746.log
Bosenko
                  hadoop
                                            hs_err_pid15850.log
examples.desktop
                                            hs_err_pid16835.log
Faust_1_Count_0
                  hadoopdata
                                            replay_pid15635.log
hadoop@bashkatova-VirtualBox:~$
```

### Вывести статистику по директории /user/mgpu/fio виртуальной машины:

```
hadoop@bashkatova-VirtualBox:~$ hadoop fs -du -h /user/mgpu/bashkatova
2023-05-03 16:27:51,418 WARN util.NativeCodeLoader: Unable to load native-hadoop
library for your platform... using builtin-java classes where applicable
7.2 M 7.2 M /user/mgpu/bashkatova/bad.gz
9.5 M 9.5 M /user/mgpu/bashkatova/bad.txt
```

## Удалить поддиректорию /fio со всем содержимым:

## **Browse Directory**

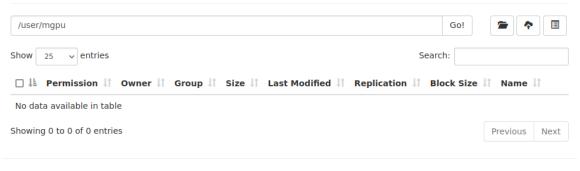


Hadoop, 2018.

hadoop@bashkatova-VirtualBox:~\$ hadoop fs -rm -r /user/mgpu/bashkatova
2023-05-03 16:30:54,670 WARN util.NativeCodeLoader: Unable to load native-hadoop
library for your platform... using builtin-java classes where applicable
Deleted /user/mgpu/bashkatova
hadoop@bashkatova-VirtualBox:~\$



## **Browse Directory**



Hadoop, 2018.