

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 Nadoor»**

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.

Москва  
2023

## Вывести с помощью команды **help** описание основных команды **shell**-клиента:

```
hadoop@bashkatova-VirtualBox:~$ help
GNU bash, версия 4.4.20(1)-release (x86_64-pc-linux-gnu)
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 «man -k» или «info».

Звёздочка (*) рядом с названием команды означает, что эта команда отключена.

задание [&]
(( выражение ))
. файл [аргументы]
:
[ аргумент... ]
[[ выражение ]]
alias [-p] [имя[=значение] ... ]
bg [задание ...]
bind [-lpvsPSVX] [-m раскладка] [-f ф>
break [n]
builtin [встр_команда [аргумент ...]>
caller [выражение]
case СЛОВО in [ШАБЛОН [| ШАБЛОН]...) >
cd [-L|[-P [-e]] [-@]] [каталог]
command [-pVv] команда [аргумент ...>
compgen [-abcdefgjkusv] [-o option] [>
complete [-abcdefgjkusv] [-pr] [-DE] >
comport [-o|+o параметр] [-DE] [имя .>
continue [n]
coproc [ИМЯ] команда [перенаправления>
declare [-aAfFgIlNrtux] [-p] [имя[=зн>
dirs [-clpv] [+N] [-N]
disown [-h] [-ar] [jobspec ... | pid >
echo [-neE] [аргумент ...]
enable [-a] [-dnps] [-f файл] [имя ..>
eval [аргумент ...]
exec [-cl] [-a имя] [команда] [аргуме>
exit [n]
export [-fn] [имя[=значение ...] или >
history [-c] [-d смещение] [n] или h>
if КОМАНДЫ; then КОМАНДЫ; [ elif КОМ>
jobs [-lnprs] [задание ...] или jobs>
kill [-s назв_сигнала | -n номер_сиг>
let аргумент [аргумент ...]
local [параметр] имя[=значение] ...
logout [n]
mapfile [-d delim] [-n count] [-O or>
popd [-n] [+N | -N]
printf [-v переменная] формат [аргум>
pushd [-n] [+N | -N | каталог]
pwd [-LP]
read [-ers] [-a массив] [-d разделит>
readarray [-n число] [-O начало] [-s>
readonly [-aAf] [имя[=значение] ...]>
return [n]
select ИМЯ [in СЛОВА ... ;] do КОМАН>
set [-abefhkmnptuvxBCHP] [-o парамет>
shift [n]
shopt [-pqsu] [-o] [параметр ...]
source файл [аргументы]
suspend [-f]
test [выражение]
time [-p] конвейер
times
trap [-lp] [[аргумент] сигнал ...]
true
type [-afptP] имя [имя ...]
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

Show  entries

Search:

<input type="checkbox"/>	Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name
<input type="checkbox"/>	drwxr-xr-x	hadoop	supergroup	0 B	May 03 15:37	0	0 B	<a href="#">bashkatova</a>

Showing 1 to 1 of 1 entries

**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
/ZH2XgNRathY1GhxPYKA0WIwOwHZBBz41z5fI7DqYyFcY5KTGNl0lBE1zt1X0gtRrW2EqvMuuQWz
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/9dRZDa5CEid2FrNAbmqLKzpZMdXKjrdn/8J1AaogIYEQtWn7twvjI
/ZH2XgNRathY1GhxPYkA0WIwOwHZBBz41z5fI7DqYyFcY5KTGNl0lBE1zt1X0gtRrW2EqvMuuQWz
VQoKw+EicNY4UbQmL78NKP4SHFhdVK7H9g4N5mg77XjXkEkTdFDyEAn1rc4r2lGvfq5oSKEQxt9D
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
y뵡췡뵡t뵡|sB뵡뵡뵡뵡Xt뵡S뵡뵡뵡V-F뵡Ra&뵡j%뵡뵡2뵡:[뵡Le$뵡M7뵡,}o;뵡뵡R.뵡뵡S뵡}1뵡+뵡뵡뵡뵡
뵡뵡|뵡뵡뵡뵡w뵡c뵡뵡뵡5B뵡뵡4뵡E뵡뵡
뵡뵡뵡L뵡뵡
: I뵡뵡y뵡뵡-B뵡뵡뵡뵡&뵡뵡뵡;뵡뵡
뵡뵡뵡뵡뵡뵡^:뵡뵡<Mp뵡뵡뵡: w뵡뵡뵡
zBoun뵡w뵡뵡D뵡A뵡뵡뵡d\뵡뵡뵡1űa뵡뵡뵡뵡Dq?뵡F 뵡뵡L[뵡뵡뵡뵡뵡ie$뵡y뵡e뵡p뵡e뵡r뵡뵡`o'
뵡뵡뵡뵡V뵡뵡b뵡뵡1뵡뵡뵡H2N뵡뵡뵡%뵡H|?뵡뵡'뵡뵡뵡U]S뵡F뵡: en뵡뵡V뵡뵡Iz뵡0뵡뵡뵡뵡뵡|i!뵡S|3\W뵡Jout뵡V뵡
뵡뵡뵡t뵡뵡뵡g[뵡뵡뵡Xy(뵡{뵡뵡뵡뵡S뵡C뵡뵡H]Q?뵡뵡;뵡뵡뵡뵡뵡~=뵡`뵡!뵡5뵡- "뵡Lh뵡e뵡뵡U뵡뵡뵡u뵡
0뵡뵡'It뵡뵡P뵡뵡뵡Z8i뵡뵡뵡뵡뵡L:űF뵡e뵡뵡뵡/ 뵡뵡뵡뵡;P뵡뵡b뵡뵡뵡ffK뵡e뵡m pX뵡뵡@p뵡\),뵡,뵡[tp
K뵡|s뵡,뵡~T뵡뵡뵡
뵡,뵡뵡뵡뵡載뵡?뵡뵡뵡뵡J
뵡뵡뵡뵡d#K뵡뵡뵡{ K뵡뵡(뵡뵡|g뵡뵡o뵡VPg뵡Hm뵡 뵡5뵡뵡뵡j뵡/뵡\gE
뵡3
뵡뵡
뵡~뵡I-뵡뵡4a뵡뵡y뵡뵡BoN뵡뵡|뵡뵡뵡뵡뵡tG뵡뵡뵡뵡q뵡뵡3q)뵡뵡뵡!뵡뵡(뵡G뵡뵡pZ뵡뵡Y뵡<뵡,_뵡뵡]뵡뵡3뵡뵡
!뵡뵡Q뵡
뵡0y뵡뵡뵡뵡뵡0V뵡뵡}뵡뵡뵡뵡ă}뵡뵡뵡뵡뵡뵡c뵡2[뵡뵡뵡뵡|뵡뵡
cat: Unable to write to output stream.
hadoop@bashkatova-VirtualBox:~$
```

```
cat: Unable to write to output 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/8J1AaogIYEQtwN7twvjI
/ZH2XgNRathY1GhxPYKA0WiOwHZBBz41z5fI7DqYyFcY5KTGNl0lBE1zt1X0gtRrW2EqvMuuQWz
VQoKw+EicNY4UbQmL78NKP4SHFhdVK7H9g4N5mg77XjXkEkTfDyEAn1rc4r2lGvfq5oSKEQxt9D
text: Unable to write to output stream.
```

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

**Создать копию файла file.txt вида date\_file.txt, где в начале имени файла-копии 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-1242662641-127.0.1.1-1683114589891: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-1242662641-127.0.1.1-1683114589891:blk_1073741868_1044 after checking nodes = [DatanodeInfoWithStorage[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 current 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 locations from namenode and retry...
2023-05-03 16:25:03,474 WARN hdfs.DFSClient: DFS chooseDataNode: got # 1 IOException, will wait for 2427.4810781171923 msec.
get: Checksum CRC32C not matched for file /user/mgpu/bashkatova/bad.txt at position 4579328: expected=A5AFF22C but computed=FF53E4EB, algorithm=PureJavaCrc32C
hadoop@bashkatova-VirtualBox:~$ ls
bad.gz          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 hadoop-3.1.2.tar.gz      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

<input type="text" value="/user/mgpu"/>								Go!			
Show	25	entries	Search: <input type="text"/>								
<input type="checkbox"/>	Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name			
<input type="checkbox"/>	drwxr-xr-x	hadoop	supergroup	0 B	May 03 15:49	0	0 B	bashkatova			
Showing 1 to 1 of 1 entries									Previous 1 Next		

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:~$
```

← → ↺

localhost:9870/explorer.html#/user/mgpu

📄 ☆

🔍 ☰

Hadoop Overview Datanodes Datanode Volume Failures Snapshot Startup Progress Utilities ▾

## Browse Directory

/user/mgpu

Go!

📁

📄

📄

Show 

25 ▾

 entries 

Search:

☐

📄

Permission

Owner

Group

Size

Last Modified

Replication

Block Size

Name

No data available in table

Showing 0 to 0 of 0 entries

Previous

Next