

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

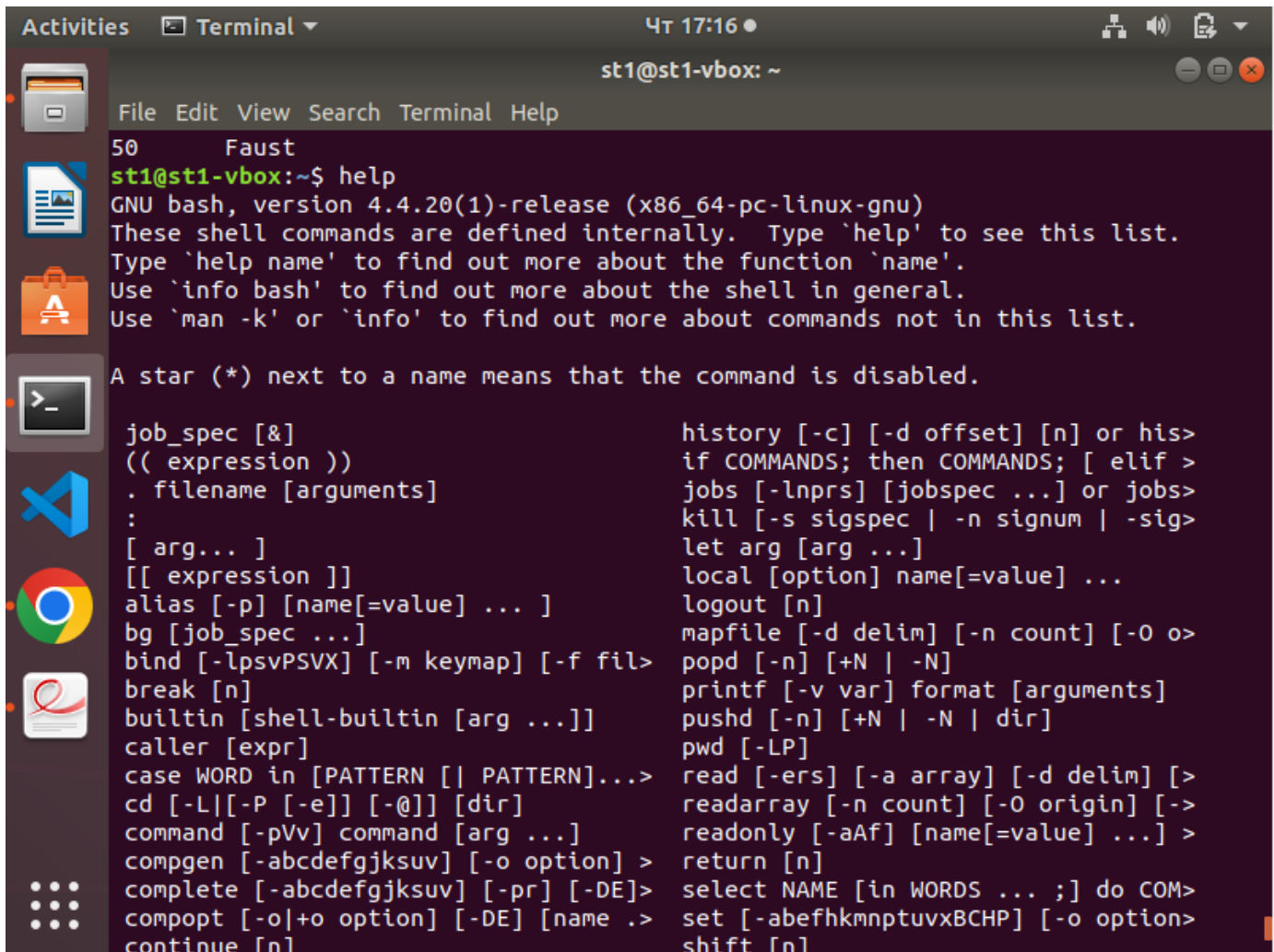
ДИСЦИПЛИНА:
«Инструменты для хранения и обработки больших данных»
пр_05-02
Тема:
«Introduction to HDFS»

Выполнила: Николаева С.Г., АДЭУ-201
Преподаватель: Босенко Тимур Муртазович

Москва
2023

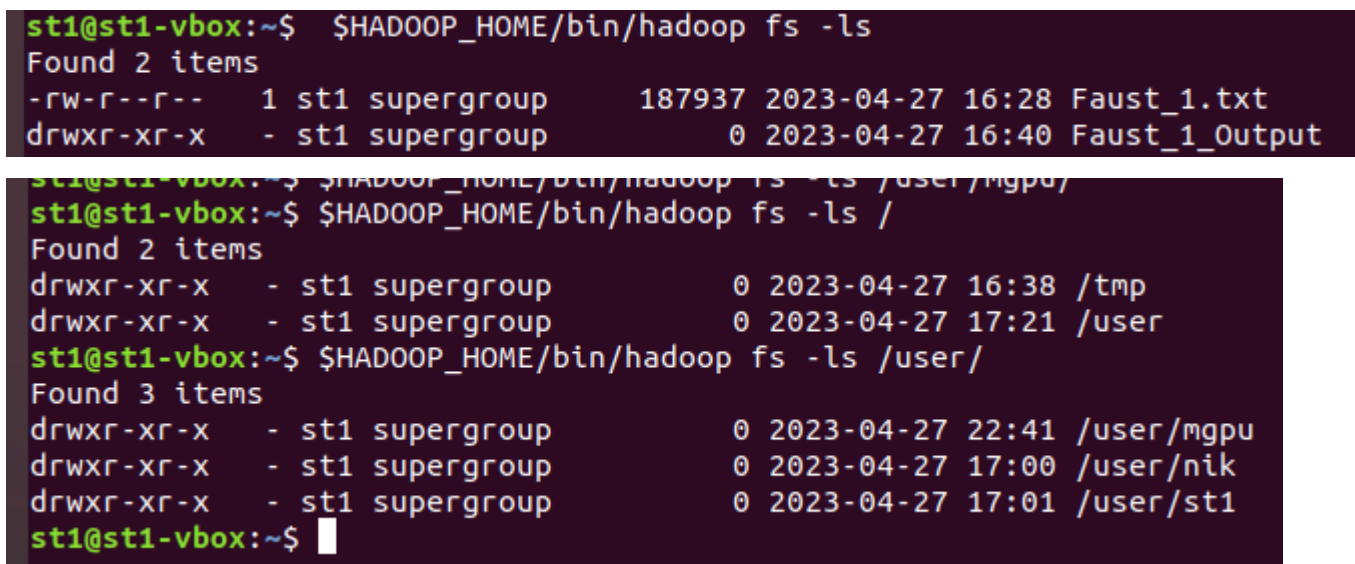
1. Развернуть виртуальное окружение

2. Вывести с помощью команды `help` описание основных команды shell-клиента



```
st1@st1-vbox: ~  
File Edit View Search Terminal Help  
50 Faust  
st1@st1-vbox:~$ help  
GNU bash, version 4.4.20(1)-release (x86_64-pc-linux-gnu)  
These shell commands are defined internally. Type 'help' to see this list.  
Type 'help name' to find out more about the function 'name'.  
Use 'info bash' to find out more about the shell in general.  
Use 'man -k' or 'info' to find out more about commands not in this list.  
  
A star (*) next to a name means that the command is disabled.  
  
job_spec [&  
(( expression ))  
. filename [arguments]  
:  
[ arg... ]  
[[ expression ]]  
alias [-p] [name[=value] ... ]  
bg [job_spec ...]  
bind [-lpsvPSVX] [-m keymap] [-f file]  
break [n]  
builtin [shell-builtin [arg ...]]  
caller [expr]  
case WORD in [PATTERN [| PATTERN]...>  
cd [-L|-P [-e]] [-@]] [dir]  
command [-pVv] command [arg ...]  
compgen [-abcdefgksuv] [-o option] >  
complete [-abcdefgksuv] [-pr] [-DE]>  
comptop [-o|+o option] [-DE] [name .>  
continue [n]  
  
history [-c] [-d offset] [n] or his>  
if COMMANDS; then COMMANDS; [ elif >  
jobs [-lnprs] [jobspec ...] or jobs>  
kill [-s sigspec | -n signum | -sig>  
let arg [arg ...]  
local [option] name[=value] ...  
logout [n]  
mapfile [-d delim] [-n count] [-O o>  
popd [-n] [+N | -N]  
printf [-v var] format [arguments]  
pushd [-n] [+N | -N | dir]  
pwd [-LP]  
read [-ers] [-a array] [-d delim] [>  
readarray [-n count] [-O origin] [->  
readonly [-aAf] [name[=value] ...] >  
return [n]  
select NAME [in WORDS ... ;] do COM>  
set [-abefhkmnptuvxBCHP] [-o option>  
shift [n]
```

3. Просмотреть корневую директорию HDFS



```
st1@st1-vbox:~$ $HADOOP_HOME/bin/hadoop fs -ls  
Found 2 items  
-rw-r--r-- 1 st1 supergroup 187937 2023-04-27 16:28 Faust_1.txt  
drwxr-xr-x - st1 supergroup 0 2023-04-27 16:40 Faust_1_Output  
  
st1@st1-vbox:~$ $HADOOP_HOME/bin/hadoop fs -ls /user/mgpu/  
st1@st1-vbox:~$ $HADOOP_HOME/bin/hadoop fs -ls /  
Found 2 items  
drwxr-xr-x - st1 supergroup 0 2023-04-27 16:38 /tmp  
drwxr-xr-x - st1 supergroup 0 2023-04-27 17:21 /user  
st1@st1-vbox:~$ $HADOOP_HOME/bin/hadoop fs -ls /user/  
Found 3 items  
drwxr-xr-x - st1 supergroup 0 2023-04-27 22:41 /user/mgpu  
drwxr-xr-x - st1 supergroup 0 2023-04-27 17:00 /user/nik  
drwxr-xr-x - st1 supergroup 0 2023-04-27 17:01 /user/st1  
st1@st1-vbox:~$
```

Local FS

```
st1@st1-vbox:~/hadoop-3.3.5$ ls
bin      lib      licenses-binary  NOTICE-binary  sbin
etc      libexec  LICENSE.txt      NOTICE.txt      share
include  LICENSE-binary  logs            README.txt
st1@st1-vbox:~/hadoop-3.3.5$
```

```
st1@st1-vbox:~/hadoop-3.3.5$ cd /bin/
st1@st1-vbox:/bin$ ls
bash          fuser          networkctl     static-sh
brltty        fusermount     nisdomainname  stty
bunzip2       getfacl        ntfs-3g        su
busybox       grep           ntfs-3g.probe  sync
bzip2         gunzip         ntfsctl        systemctl
bzip          gzexe          ntfscluster    systemd
bzdiff        gzip           ntfsdcomp      systemd-ask-password
bzegrep       hciconfig      ntfsfallocate  systemd-escape
bzexe         hostname       ntfsfix         systemd-hwdb
bzfgrep       ip             ntfsinfo        systemd-inhibit
bzgrep        journalctl     ntfsls          systemd-machine-id-setup
bzip2         kbd_mode       ntfsmove        systemd-notify
bzip2recover  kill           ntfsrecover     systemd-sysusers
bzless        kmod           ntfssecaudit    systemd-tmpfiles
bzip          less           ntfstruncate    systemd-tty-ask-password-agent
cat           lessecho       ntfsusermap     tar
chacl         lessfile       ntfswipe        tempfile
chgrp         lesskey        open            touch
```

4. Создать в HDFS в директории /user/mgpu поддиректорию ваше_фio

```
st1@st1-vbox:/bin$ $HADOOP_HOME/bin/hadoop fs -mkdir -p /user/mgpu/nikolaevaSG
```

5. Создать в локальной файловой системе случайный текстовый файл размером 10 Мб с именем, образованным вашими инициалами

```
st1@st1-vbox:~$ base64 /dev/urandom | head -c 10000000 > nsg.txt
```

6. Заархивировать созданный текстовый файл

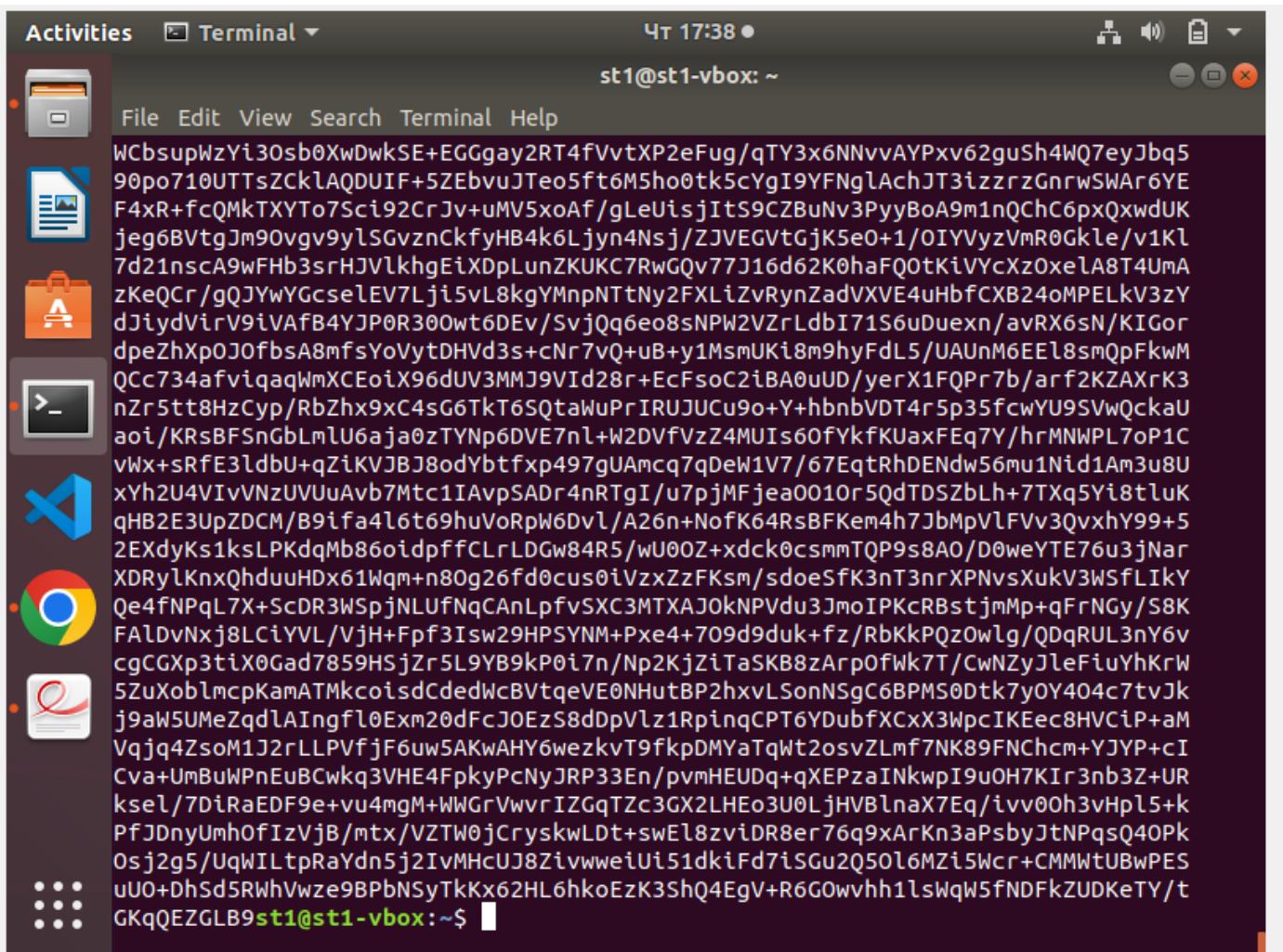
```
st1@st1-vbox:~$ gzip -c nsg.txt > nsg.gz
```

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

```
st1@st1-vbox:~$ $HADOOP_HOME/bin/hadoop fs -copyFromLocal nsg.txt /user/mgpu/nikolaevaSG
st1@st1-vbox:~$ $HADOOP_HOME/bin/hadoop fs -copyFromLocal nsg.gz /user/mgpu/nikolaevaSG
```

8. 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.

```
st1@st1-vbox:~$ $HADOOP_HOME/bin/hadoop fs -cat /user/mgpu/nikolaevaSG/nsg.txt  
part-m-*
```



```
Activities Terminal 4T 17:38 ●  
st1@st1-vbox: ~  
File Edit View Search Terminal Help  
WCbsupWzYi30sb0XwDwkSE+EGGgay2RT4fVvtXP2eFug/qTY3x6NNvvAYPxxv62guSh4WQ7eyJbq5  
90po710UTTsZCKlAQDUIF+5ZEbvUJTTeo5ft6M5ho0tk5cYgI9YFNglAchJT3izzrzGnrwSWAr6YE  
F4xR+fcQMkTXYTo7Sci92CrJv+uMV5xoAf/gLeUisjItS9CZBuNv3PyyBoA9m1nQChC6pxQxwdUK  
jeg6BvtgJm90vgv9yLSGvznCkfyHB4k6Ljyn4Nsj/ZJVEGVtGjK5e0+1/OIYVyzVmR0Gkle/v1Kl  
7d21nscA9wFhb3srHJVLkhgEiXDPunZKUKC7RwGQv77J16d62K0haFQ0tKiVYcXz0xelaA8T4UmA  
zKeQCr/gQJYwYGcseLEV7Lji5vL8kgYMnpNTNy2FXLiZvRynZadvXVE4uHbfCXB24oMPELkV3zy  
djiydVirV9iVAFB4YJP0R300wt6DEv/SvjQq6eo8sNPW2VZrLdbI71S6uDuexn/avRX6sN/KIGor  
dpeZhXp0J0fbsA8mfsYoVytDhVd3s+cNr7vQ+uB+y1MsmUKi8m9hyFdL5/UAUnM6EEL8smQpFkwM  
QCc734afviqaqWmXCEoiX96dUV3MMJ9VId28r+EcFsoC2iBA0uUD/yerX1FQPr7b/arf2KZAXrK3  
nZr5tt8HzCyp/RbZhx9xC4sG6TkT6SQtawuPrIRUJUcu9o+Y+hbnbVDT4r5p35fcwYU9SVwQckaU  
aoi/KRsBFSnGbLmLU6aja0zTYNp6DVE7nl+W2DVfVzZ4MUIs60fYkFKUaxFEq7Y/hrMNWPL7oP1C  
vWx+sRfE3ldbU+qZiKVJBj8odYbtfxp497gUAmcq7qDeW1V7/67EqRhDENdw56mu1Nid1Am3u8U  
xYh2U4VivVNzUVUuAvb7Mtc1IAvpSADr4nRTGI/u7pjMFjea0010r5QdTDSZbLh+7TXq5Yi8tluk  
qHB2E3UpZDCM/B9ifa4l6t69huVoRpw6DvL/A26n+NofK64RsBFKem4h7JbMpVLFVv3QvvhY99+5  
2EXdyKs1ksLPKdqMb86oidpffCLrLDGw84R5/wU00Z+xdck0csmmTQP9s8A0/D0weYTE76u3jNar  
XDRylKnXQhduuHDx61Wqm+n80g26fd0cus0iVzXzZFKsm/sdoeSfK3nT3nrXPNvsXukV3WSfLiKY  
Qe4fNPqL7X+ScDR3WSpjNLUFnQCanLpfvSXC3MTXAJ0kNPVdu3JmoIPKcRBstjmMp+qFrNGy/S8K  
FALDvNxj8LCiYVL/VjH+Fpf3IsW29HPSYNN+Pxe4+709d9duk+fz/RbKkPQz0wlg/QDqRUL3nY6v  
cgCGXp3tiX0Gad7859HSjZr5L9YB9kP0i7n/Np2KjZiTasKB8zArp0fWk7T/CwNZyJleFiuYhKrW  
5ZuXoblmcPkamATMkcoisdCdedWcBVTqeVE0NHutBP2hxxVLSonNSGc6BPMS0Dtk7y0Y404c7tvJk  
j9aW5UMEzqdLAIingfl0Exm20dFcJOEzS8dDpVlZ1RpinqCPT6YDufXCcxX3WpcIKEec8HVCiP+aM  
Vqjq4ZsOm1J2rLLPVfjF6uw5AKwAHY6wezkvT9fkdMYaTqWt2osvZLmf7NK89FNChcm+YJYP+cI  
Cva+UmBuWpNeuBCwkq3VHE4FpkyPcNyJRP33En/pvmHEUDq+qXEPzaInkwpI9u0H7KIr3nb3Z+UR  
kseL/7DiRaEDF9e+vu4mgM+WWGrVwvriZGqTZc3GX2LHEo3U0LjHVBlnaX7Eq/ivv00h3vHpl5+k  
PfJDnyUmhOfIzVjB/mtx/VZTW0jCryskwLdt+swEL8zviDR8er76q9xArKn3aPsbyJtNPqsQ40Pk  
OsJ2g5/UqWILtpRaYdn5j2IvMHCUJ8ZivwweiUi51dkiFd7iSGu2Q50L6MZi5Wcr+CMMWtUBWPES  
uUO+DhSd5RwHvWze9BPbNSyTkKx62HL6hkoEzK3ShQ4EgV+R6G0wvhh1lsWqW5fNDFkZUDKeTY/t  
GKqQEZGLB9st1@st1-vbox:~$
```

```
GKqQEZGLB9st1@st1-vbox:~$ $HADOOP_HOME/bin/hadoop fs -cat /user/mgpu/nikolaevaS  
| head  
pke70skyX2z2l/3FEBUSvVZH4S3i7LEY+VZWge5dVI1qyQeHadrFfXvnEW40u3Sjr2JyCpVSdivo  
7fkVJK0GA+wwHfnry27uNzui7Je/8GWeTW9R6kgfClrl7o687l8NuHYbqT0c1oCwvSD9IgVtVHfQ  
jlk+VT0f/x21V3LXAAM183k9IWNuc7eJzyvcfceXCXrXCeB7vEPFen51JTKqI24pNhWVGisx5w0G  
IaYiVnZLnS6QzfgmtxV1/vCJSPbdVqUSF9Ran0B3ivLF7oQfzKlFYITS+ugW5Eyr9WBe6urBX85d  
PFo0Zq5fSEGma8ygax7DelilJxvQ6mZqyzd1X4eMNPnwo6+QZPgmz3b8GkShw3v0ILJ+BM3QBmdQ  
/oS4t0cIB9wDA0KPSRA00Zl6oU0oZVL236ZBR5hQ0hVcg/CQZ6UgpySbeBxbhi3rvx6MmZ1jRUzn  
Bv2e2AtfVJA3mjomqvDbfPdzSyci7mLE6KgkBAHjfs8JHs/aHAHPTU0sm9h1hqPfZsj6FTjWCU0D  
3kh65n0ZXVA1V7VCcHY8wbAvfJPbLgmtwE3GndZgt/+PEYnoK96ZahBjpkLRhZs7+PKsjuxQKNst  
ukKzEoskkuw5cK8eDzxyJaSxmPBbrnbKylXmips1JKbK2qmCefemI5vNE4AF4HzYT0JSh9Sy0BZnX  
J0tg2hATagtXDUX78IyfQRPfPvTiMviabE684U+eutdE+o02fKwej+ZGNBjQo6KLSwH/nNbqFVGz  
cat: Unable to write to output stream.  
st1@st1-vbox:~$
```



```
st1@st1-vbox:~$ $HADOOP_HOME/bin/hadoop fs -cat /user/mgpu/nikolaevaSG/nsg.txt
| tail
cgCGXp3tiX0Gad7859HSjZr5L9YB9kP0i7n/Np2KjZiTasKB8zArp0fWk7T/CwNZyJleFiuYhKrw
5ZuXoblmcPkamATMkcoisdCdedWcBvtqeVE0NHutBP2hXvLSonNSgC6BPMS0Dtk7y0Y404c7tvJk
j9aW5UMeZqdlAIngfl0Exm20dFcJOEzS8dDpVlZ1RpinqCPT6YDubfXCxX3WpcIKEec8HVCiP+aM
Vqqj4ZsoM1J2rLLPVfjF6uw5AKwAHY6wezkvT9fkdMYaTqWt2osvZLmf7NK89FNChcm+YJYP+cI
Cva+UmBuWPnEuBCwkq3VHE4FpkyPcNyJRP33En/pvmHEUDq+qXEPzaINkwpI9u0H7KIr3nb3Z+UR
ksel/7DiRaEDF9e+vu4mgM+WWGrVwvrIZGqTZc3GX2LHEo3U0LjHVBlnaX7Eq/ivv00h3vHpl5+k
PfJDnyUmhOfIzVjB/mtx/VZTW0jCryskwLdt+swEl8zviDR8er76q9xArKn3aPsbyJtNPqsQ40Pk
Osj2g5/UqWILtpRaYdn5j2IvMHcUJ8ZivwweiUi51dkifd7iSGu2Q50l6MZi5Wcr+CMMWtUBwPES
uU0+DhSd5RwhVwze9BPbNSyTkKx62HL6hkoEzK3ShQ4EgV+R6G0wvhh1lsWqW5fNDFkZUDKeTY/t
GKqQEZGLB9st1@st1-vbox:~$
```

```
st1@st1-vbox:~$ $HADOOP_HOME/bin/hadoop fs -text /user/mgpu/nikolaevaSG/nsg.txt
| head
pke70skyX2z2l/3FEBUSvVZH4S3i7lEY+VZWge5dVI1qyQeHadrFfXvnEW40u3Sjr2JyCpVSdivo
7fkVJK0GA+wWHfnry27uNzui7Je/8GWeTW9R6kgfClrL7o687l8NuHYbqT0c1oCwvSD9IgVtVHFQ
jlk+VT0f/x21V3LXAAM183k9IWNuc7eJzyvcfceXCXrXCeB7vEPFen51JTKqI24pNhWVGIsx5w0G
IaYiVnZLnS6QzfgmtxV1/vCJSPbdVqUSF9Ran0B3ivLF7oQfzKlFYITS+ugW5Eyr9WBe6urBX85d
PFo0Zq5fSEGma8ygax7DelilJxVQ6mZqyzd1X4eMNPnwo6+QZPgmz3b8GkShw3v0ILJ+BM3QbmdQ
/oS4t0cIB9wDA0KPSRA00ZL6oU0oZVL236ZBR5hQ0hVcg/CQZ6UgpYSbeBxbHi3rvx6MmZ1jRUZn
Bv2e2AtfVJA3mjomqvDbfPdZSyci7mLE6KgbAHjfs8JHs/aHAHPTU0sm9h1hqPfZsj6FTjWCU0D
3kh65n0ZXVA1V7VCcHY8wbAvfJPbLgmtwE3GndZgt/+PEYnoK96ZahBjpkLRhZs7+PKsjuXQKNst
ukKzEoskkuw5cK8eDzxyJaSxmPBnrbKylXmips1JKbK2qmCefemI5vNE4AF4HzYT0JSh9Sy0BZnX
JOtg2hATagTxDUX78IyfQRPfPvTiMviabE684U+eutdE+o02fKwej+ZGNBjQo6KLSwH/nNbqFVgz
text: Unable to write to output stream.
st1@st1-vbox:~$
```

```
st1@st1-vbox:~$ $HADOOP_HOME/bin/hadoop fs -text /user/mgpu/nikolaevaSG/nsg.txt
| tail
cgCGXp3tiX0Gad7859HSjZr5L9YB9kP0i7n/Np2KjZiTasKB8zArp0fWk7T/CwNZyJleFiuYhKrw
5ZuXoblmcPkamATMkcoisdCdedWcBvtqeVE0NHutBP2hXvLSonNSgC6BPMS0Dtk7y0Y404c7tvJk
j9aW5UMeZqdlAIngfl0Exm20dFcJOEzS8dDpVlZ1RpinqCPT6YDubfXCxX3WpcIKEec8HVCiP+aM
Vqqj4ZsoM1J2rLLPVfjF6uw5AKwAHY6wezkvT9fkdMYaTqWt2osvZLmf7NK89FNChcm+YJYP+cI
Cva+UmBuWPnEuBCwkq3VHE4FpkyPcNyJRP33En/pvmHEUDq+qXEPzaINkwpI9u0H7KIr3nb3Z+UR
ksel/7DiRaEDF9e+vu4mgM+WWGrVwvrIZGqTZc3GX2LHEo3U0LjHVBlnaX7Eq/ivv00h3vHpl5+k
PfJDnyUmhOfIzVjB/mtx/VZTW0jCryskwLdt+swEl8zviDR8er76q9xArKn3aPsbyJtNPqsQ40Pk
Osj2g5/UqWILtpRaYdn5j2IvMHcUJ8ZivwweiUi51dkifd7iSGu2Q50l6MZi5Wcr+CMMWtUBwPES
uU0+DhSd5RwhVwze9BPbNSyTkKx62HL6hkoEzK3ShQ4EgV+R6G0wvhh1lsWqW5fNDFkZUDKeTY/t
GKqQEZGLB9st1@st1-vbox:~$
```

9. Создать копию файла file.txt вида date_file.txt, где в начале имени файла-копии указана текущая дата. Вывести листинг

```
st1@st1-vbox:~$ $HADOOP_HOME/bin/hadoop fs -copyToLocal /user/mgpu/nikolaevaSG
/home/st1/Documents/27-04-2023-nsg.txt
```

```
st1@st1-vbox:~$ cd /home/st1/Documents/
st1@st1-vbox:~/Documents$ ls
27-04-2023-nsg.txt
st1@st1-vbox:~/Documents$
```

10. Вывести статистику по директории /user/mgpu/fio виртуальной машины

```
st1@st1-vbox:~$ $HADOOP_HOME/bin/hadoop fs -dus /user/mgpu/nikolaevaSG
dus: DEPRECATED: Please use 'du -s' instead.
17599480 17599480 /user/mgpu/nikolaevaSG
st1@st1-vbox:~$ $HADOOP_HOME/bin/hadoop fs -du -s /user/mgpu/nikolaevaSG
17599480 17599480 /user/mgpu/nikolaevaSG
```

```
st1@st1-vbox:~$ $HADOOP_HOME/bin/hadoop fs -du /user/mgpu/nikolaevaSG
7599480 7599480 /user/mgpu/nikolaevaSG/nsg.gz
10000000 10000000 /user/mgpu/nikolaevaSG/nsg.txt
st1@st1-vbox:~$
```

11. Удалить поддиректорию /fio со всем содержимым

Hadoop

Overview Datanodes Datanode Volume Failures Snapshot Startup Progress Utilities

Browse Directory

/user/mgpu/nikolaevaSG Go!

Show 25 entries Search:

Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name
-rw-r--r--	st1	supergroup	7.25 MB	Apr 27 17:36	1	128 MB	nsg.gz
-rw-r--r--	st1	supergroup	9.54 MB	Apr 27 17:35	1	128 MB	nsg.txt

Showing 1 to 2 of 2 entries

Previous 1 Next

```
st1@st1-vbox:~$ $HADOOP_HOME/bin/hadoop dfs -rm -r /user/mgpu/nikolaevaSG
WARNING: Use of this script to execute dfs is deprecated.
WARNING: Attempting to execute replacement "hdfs dfs" instead.

Deleted /user/mgpu/nikolaevaSG
st1@st1-vbox:~$
```

Activities

Google Chrome

22:42

GitHub - Bosenko x

Browsing HDFS x

All Applications x

+

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

Update

Hadoop

OverviewDatanodesDatanode Volume FailuresSnapshotStartup ProgressUtilities

Browse Directory

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

Hadoop, 2023.