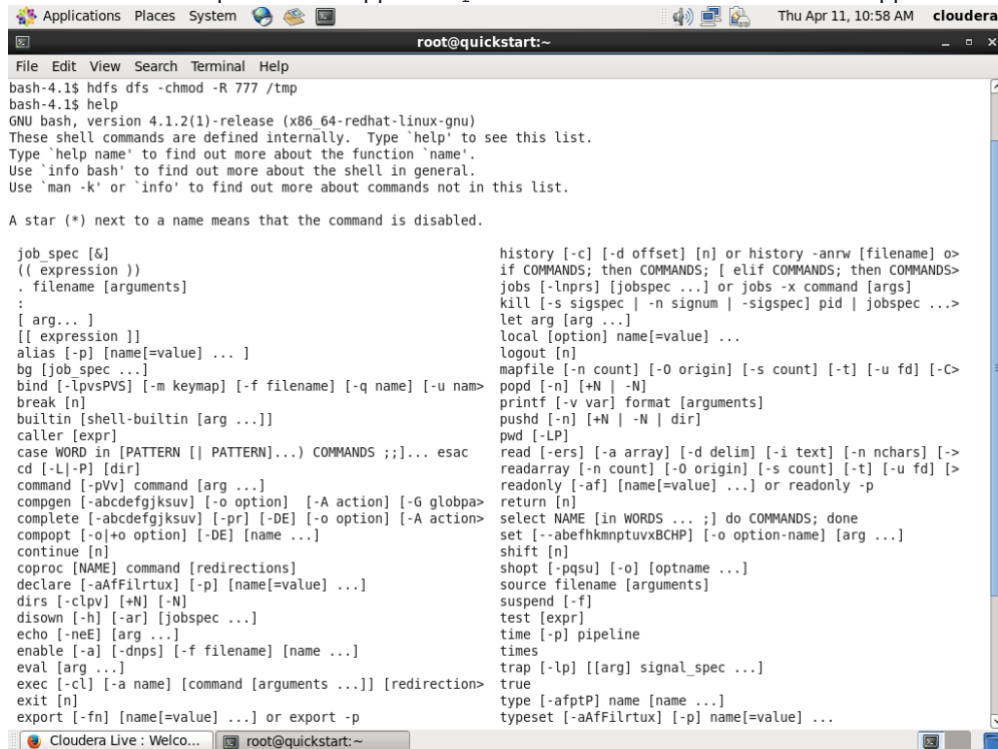


Отчёт по
ПРАКТИЧЕСКОЙ РАБОТЕ № 3-1. Корпоративные платформы. cloudera
студента группы БД-231м
Осокина Артёма Ибрагимовича

3.1.1. Развернуть виртуальное окружение.

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



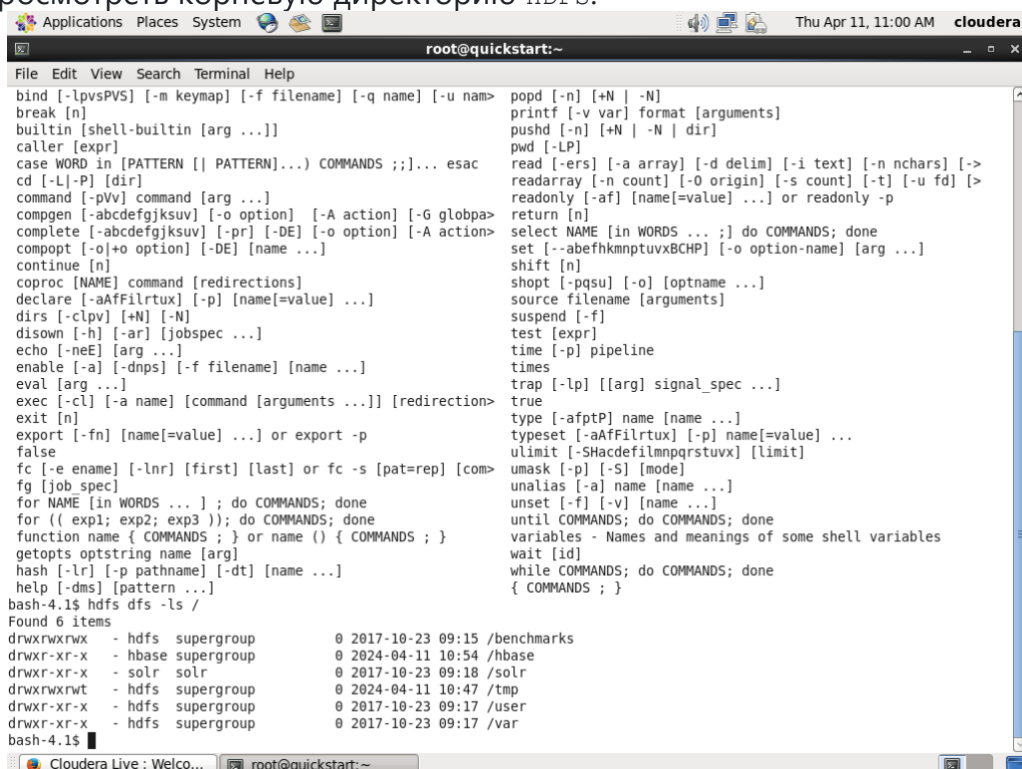
```
Applications Places System Thu Apr 11, 10:58 AM cloudera
root@quickstart:~
File Edit View Search Terminal Help
bash-4.1$ hdfs dfs -chmod -R 777 /tmp
bash-4.1$ help
GNU bash, version 4.1.2(1)-release (x86_64-redhat-linux-gnu)
These shell commands are defined internally. Type 'help' to see this list.
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 [-lpsPVS] [-m keymap] [-f filename] [-q name] [-u name]
break [n]
builtin [shell-builtin [arg ...]]
caller [expr]
case WORD in [PATTERN [| PATTERN]...] COMMANDS ;;... esac
cd [-L|-P] [dir]
command [-pVv] command [arg ...]
compgen [-abcdfgjkuv] [-o option] [-A action] [-G globpa>
complete [-abcdfgjkuv] [-pr] [-DE] [-o option] [-A action>
comptop [-o|+o option] [-DE] [name ...]
continue [n]
coproc [NAME] command [redirections]
declare [-aAfFIlrtux] [-p] [name[=value] ...]
dirs [-clpv] [+N] [-N]
disown [-h] [-ar] [jobspec ...]
echo [-neE] [arg ...]
enable [-a] [-dnps] [-f filename] [name ...]
eval [arg ...]
exec [-cl] [-a name] [command [arguments ...]] [redirection>
exit [n]
export [-fn] [name[=value] ...] or export -p

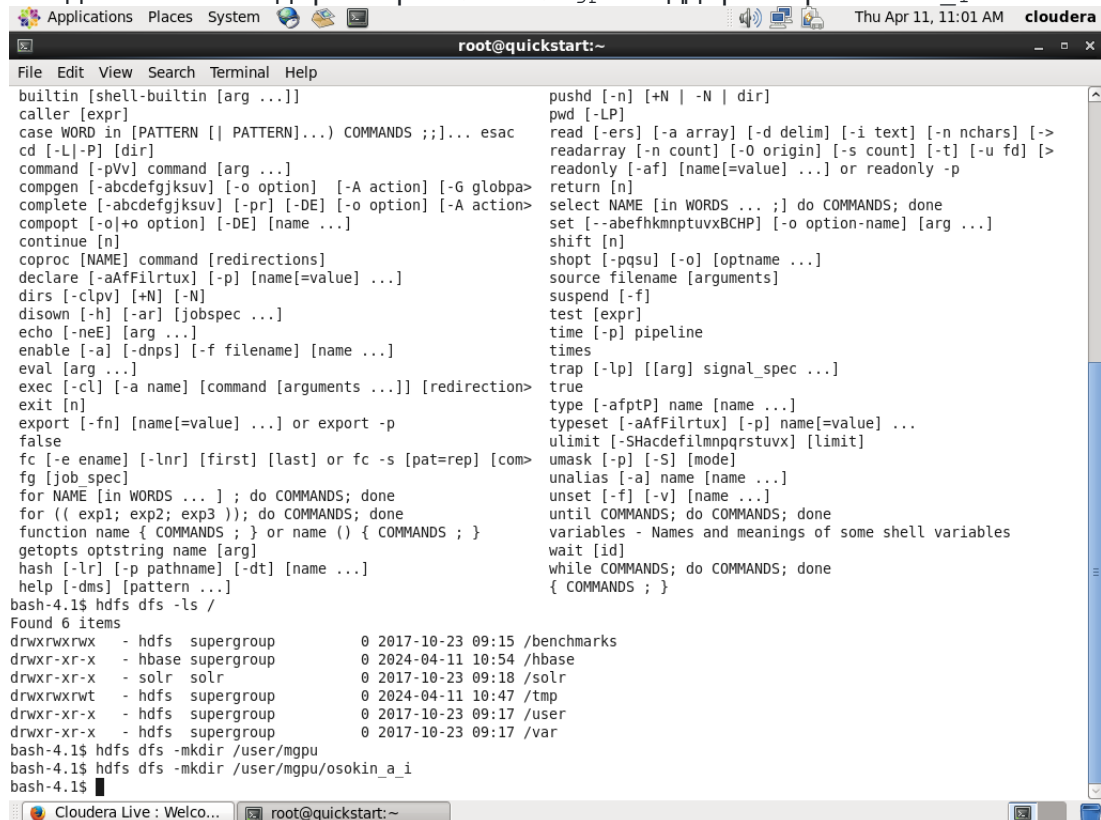
history [-c] [-d offset] [n] or history -anrw [filename] o>
if COMMANDS; then COMMANDS; [ elif COMMANDS; then COMMANDS>
jobs [-lnprs] [jobspec ...] or jobs -x command [args]
kill [-s sigspec | -n signum | -sigspec] pid | jobspec ...>
let arg [arg ...]
local [option] name[=value] ...
logout [n]
mapfile [-n count] [-O origin] [-s count] [-t] [-u fd] [-C>
popd [-n] [+N | -N]
printf [-v var] format [arguments]
pushd [-n] [+N | -N | dir]
pwd [-LP]
read [-ers] [-a array] [-d delim] [-i text] [-n nchars] [->
readarray [-n count] [-O origin] [-s count] [-t] [-u fd] [>
readonly [-af] [name[=value] ...] or readonly -p
return [n]
select NAME [in WORDS ... ;] do COMMANDS; done
set [--abefhkmnptuvxBCHP] [-o option-name] [arg ...]
shift [n]
shopt [-pqsu] [-o] [optname ...]
source filename [arguments]
suspend [-f]
test [expr]
time [-p] pipeline
times
trap [-lp] [[arg] signal_spec ...]
true
type [-afptP] name [name ...]
typeset [-aAfFIlrtux] [-p] name[=value] ...
```

3.1.3. Просмотреть корневую директорию HDFS.



```
Applications Places System Thu Apr 11, 11:00 AM cloudera
root@quickstart:~
File Edit View Search Terminal Help
bind [-lpsPVS] [-m keymap] [-f filename] [-q name] [-u nam>
break [n]
builtin [shell-builtin [arg ...]]
caller [expr]
case WORD in [PATTERN [| PATTERN]...] COMMANDS ;;... esac
cd [-L|-P] [dir]
command [-pVv] command [arg ...]
compgen [-abcdfgjkuv] [-o option] [-A action] [-G globpa>
complete [-abcdfgjkuv] [-pr] [-DE] [-o option] [-A action>
comptop [-o|+o option] [-DE] [name ...]
continue [n]
coproc [NAME] command [redirections]
declare [-aAfFIlrtux] [-p] [name[=value] ...]
dirs [-clpv] [+N] [-N]
disown [-h] [-ar] [jobspec ...]
echo [-neE] [arg ...]
enable [-a] [-dnps] [-f filename] [name ...]
eval [arg ...]
exec [-cl] [-a name] [command [arguments ...]] [redirection>
exit [n]
export [-fn] [name[=value] ...] or export -p
false
fc [-e name] [-lnr] [first] [last] or fc -s [pat=rep] [com>
fg [job_spec]
for NAME [in WORDS ... ] ; do COMMANDS; done
for (( exp1; exp2; exp3 )); do COMMANDS; done
function name { COMMANDS ; } or name () { COMMANDS ; }
getopts optstring name [arg]
hash [-lr] [-p pathname] [-dt] [name ...]
help [-dms] [pattern ...]
bash-4.1$ hdfs dfs -ls /
Found 6 items
drwxrwxrwx - hdfs supergroup 0 2017-10-23 09:15 /benchmarks
drwxr-xr-x - hbase supergroup 0 2024-04-11 10:54 /hbase
drwxr-xr-x - solr solr 0 2017-10-23 09:18 /solr
drwxrwxrwt - hdfs supergroup 0 2024-04-11 10:47 /tmp
drwxr-xr-x - hdfs supergroup 0 2017-10-23 09:17 /user
drwxr-xr-x - hdfs supergroup 0 2017-10-23 09:17 /var
```

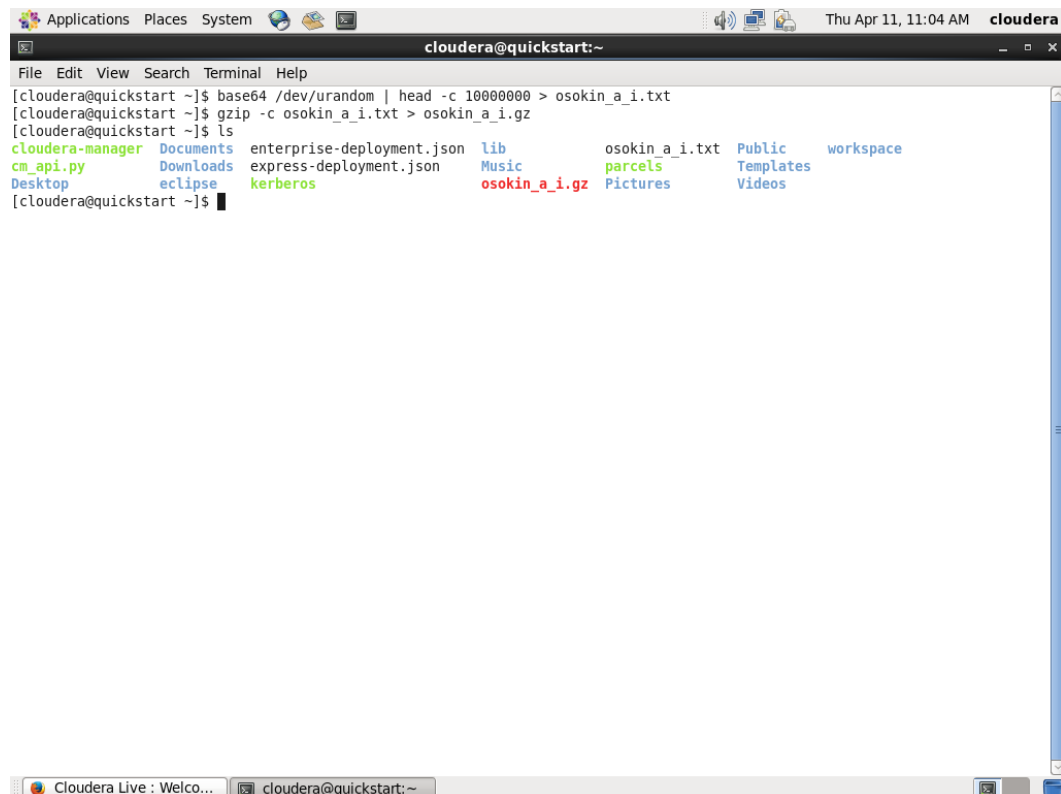
3.1.4. Создать в HDFS в директории /user/mgpu поддиректорию ваше_фio.



```
Applications Places System root@quickstart:~ Thu Apr 11, 11:01 AM cloudera
File Edit View Search Terminal Help
builtin [shell-builtin [arg ...]]
caller [expr]
case WORD in [PATTERN] [...] COMMANDS ;;)... esac
cd [-L|-P] [dir]
command [-pVv] command [arg ...]
compgen [-abcdefgksuv] [-o option] [-A action] [-G globpa>
complete [-abcdefgksuv] [-pr] [-DE] [-o option] [-A action>
comptopt [-o]+o option] [-DE] [name ...]
continue [n]
coproc [NAME] command [redirections]
declare [-aAfFiltux] [-p] [name=value] ...
dirs [-clpv] [+N] [-N]
disown [-h] [-ar] [jobspec ...]
echo [-neE] [arg ...]
enable [-a] [-dnps] [-f filename] [name ...]
eval [arg ...]
exec [-cl] [-a name] [command [arguments ...]] [redirection>
exit [n]
export [-fn] [name=value] ... or export -p
false
fc [-e ename] [-lnr] [first] [last] or fc -s [pat=rep] [com>
fg [job spec]
for NAME in WORDS ... ; do COMMANDS; done
for (( exp1; exp2; exp3 )); do COMMANDS; done
function name { COMMANDS ; } or name () { COMMANDS ; }
getopts optstring name [arg]
hash [-lr] [-p pathname] [-dt] [name ...]
help [-dms] [pattern ...]
bash-4.1$ hdfs dfs -ls /
Found 6 items
drwxrwxrwx - hdfs supergroup 0 2017-10-23 09:15 /benchmarks
drwxr-xr-x - hbase supergroup 0 2024-04-11 10:54 /hbase
drwxr-xr-x - solr solr 0 2017-10-23 09:18 /solr
drwxrwxrwt - hdfs supergroup 0 2024-04-11 10:47 /tmp
drwxr-xr-x - hdfs supergroup 0 2017-10-23 09:17 /user
drwxr-xr-x - hdfs supergroup 0 2017-10-23 09:17 /var
bash-4.1$ hdfs dfs -mkdir /user/mgpu
bash-4.1$ hdfs dfs -mkdir /user/mgpu/osokin_a_i
bash-4.1$
```

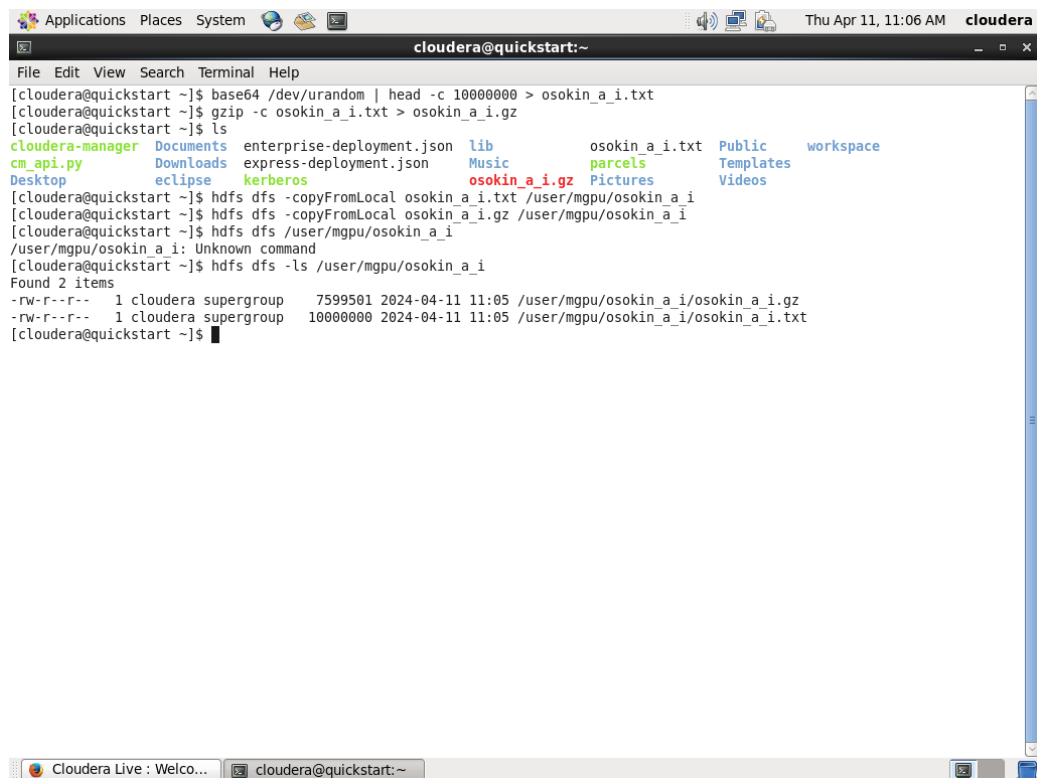
3.1.5. Создать в локальной файловой системе случайный текстовый файл размером 10 Mb с именем, образованным вашими инициалами base64 /dev/urandom | head -с 10000000 > file.txt .

3.1.6. Заархивировать созданный текстовый файл gzip -c file.txt > file.gz .



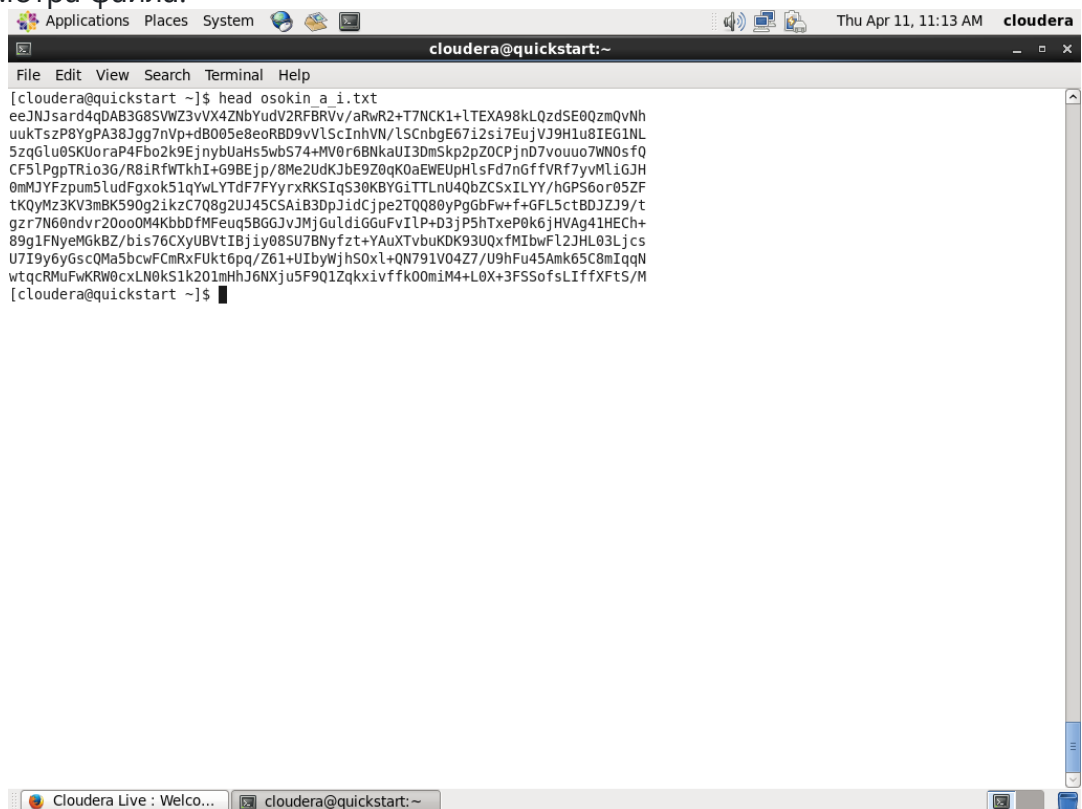
```
Applications Places System cloudera@quickstart:~ Thu Apr 11, 11:04 AM cloudera
File Edit View Search Terminal Help
[cloudera@quickstart ~]$ base64 /dev/urandom | head -c 10000000 > osokin_a_i.txt
[cloudera@quickstart ~]$ gzip -c osokin_a_i.txt > osokin_a_i.gz
[cloudera@quickstart ~]$ ls
cloudera-manager Documents enterprise-deployment.json lib osokin_a_i.txt Public workspace
cm_api.py Downloads express-deployment.json Music parcels Templates
Desktop eclipse kerberos osokin_a_i.gz Pictures Videos
[cloudera@quickstart ~]$
```

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



```
cloudera@quickstart:~$ base64 /dev/urandom | head -c 1000000 > osokin_a_i.txt
cloudera@quickstart:~$ gzip -c osokin_a_i.txt > osokin_a_i.gz
cloudera@quickstart:~$ ls
cloudera-manager  Documents  enterprise-deployment.json  lib  osokin_a_i.txt  Public  workspace
cm_api.py         Downloads  express-deployment.json    Music  parcels        Templates
Desktop          eclipse    kerberos                  osokin_a_i.gz  Pictures       Videos
cloudera@quickstart:~$ hdfs dfs -copyFromLocal osokin_a_i.txt /user/mgpu/osokin_a_i
cloudera@quickstart:~$ hdfs dfs -copyFromLocal osokin_a_i.gz /user/mgpu/osokin_a_i
cloudera@quickstart:~$ hdfs dfs /user/mgpu/osokin_a_i
/user/mgpu/osokin_a_i: Unknown command
cloudera@quickstart:~$ hdfs dfs -ls /user/mgpu/osokin_a_i
Found 2 items
-rw-r--r-- 1 cloudera supergroup 7599501 2024-04-11 11:05 /user/mgpu/osokin_a_i/osokin_a_i.gz
-rw-r--r-- 1 cloudera supergroup 1000000 2024-04-11 11:05 /user/mgpu/osokin_a_i/osokin_a_i.txt
cloudera@quickstart:~$
```

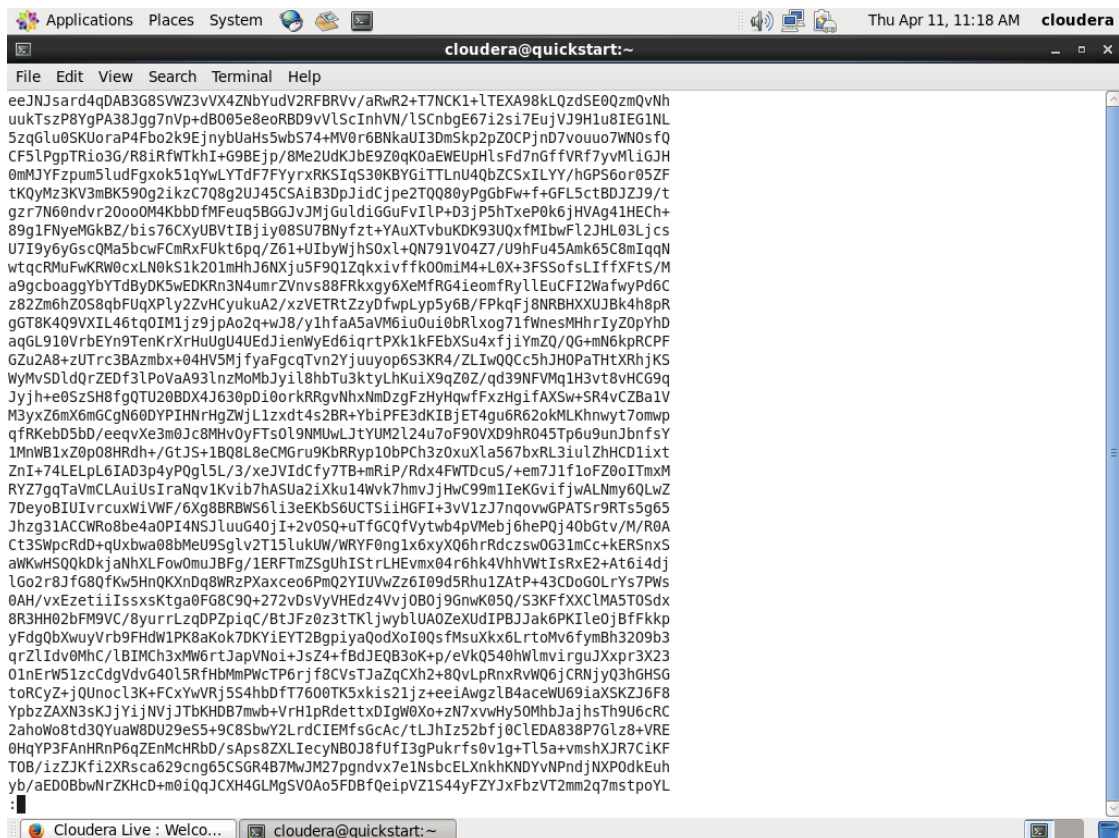
3.1.8. Просмотреть файл и архив с помощью утилит cat, text в комбинации с каналами и утилитами head, tail -- привести не менее 3 вариантов команд и просмотра файла.



```
cloudera@quickstart:~$ head osokin_a_i.txt
eeJNJsard4qDAB3G8SVWZ3vVX4ZNbYudV2RFBRVv/aRwR2+T7NCK1+lTEXA98KLQzdSE0QzmQvNh
uukTszP8YgPA3Jgg7nVp+dB005e8eoRBD9vVLScInhVN/LSnbgE67i2si7EujVJ9H1u8IEG1NL
5zqGlu0SKUoraP4Fbo2k9EjnybUaH55wbS74+MV0r6BNkaUI3DmSkp2pZ0CPjnd7vouuo7WN0sfQ
CF5lPgpTRio3G/R8iRfWtkhI+G9BEjp/8Me2UdKJbE9Z0qK0aEWEUpHLSfd7ngffVrf7yvMLiGJH
0mMJYFzpum5ludF0xok51qYwLYTdF7FYyrxRKSIqS30KBYGiTTLnU40bZCSxILYY/hGpS6or05ZF
tkQyMz3Kv3mBK590g2ikzC7Q8g2UJ45CSAiB3DpJidCjpe2TQ080yPgGbFw+f+GFL5ctBDJZJ9/t
gzr7N60ndvr20oo0M4KbbdFMFeuq5BG6JvJMjGuldiGGuFvILP+D3jP5hTxeP0K6jHVAg41HECH+
89g1FNyemGkBZ/bis76CXyUBVtIBjiy08SU7BNyftz+YAuXTvbuKDK93UqxFMiBwFLZJHL03Ljcs
U7I9y6YgscQMa5bcwFcmRxFUKt6pq/Z61+UIbyWjh50xL+QN791V04Z7/U9hFu45Amk65C8mIqqN
wtqCRMuFwKRW0xLNOk51k201mHhJ6NXju5F9Q1Zqkxivffk00miM4+L0X+3FSSofsLIffXfTs/M
cloudera@quickstart:~$
```

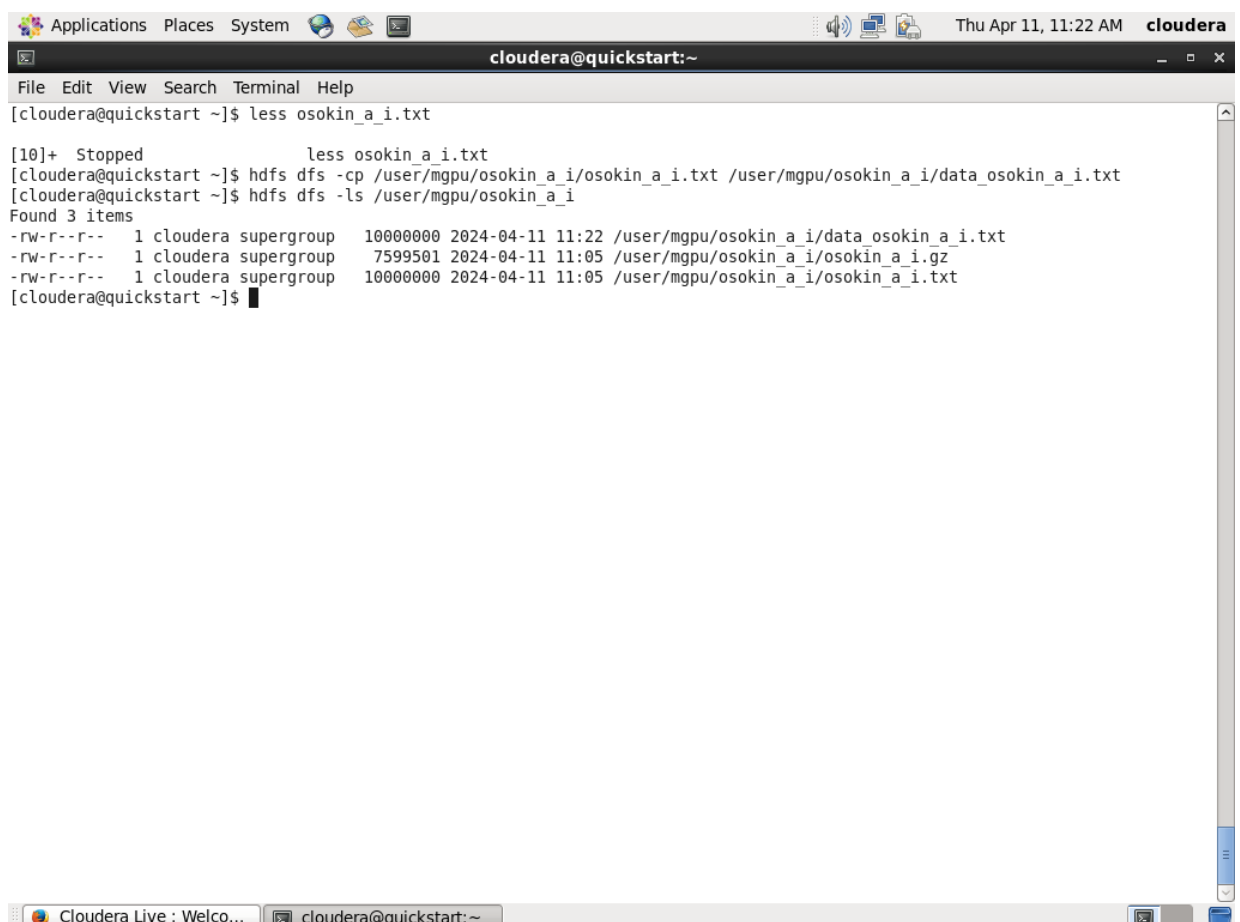
```
Applications Places System Thu Apr 11, 11:14 AM cloudera
cloudera@quickstart:~
File Edit View Search Terminal Help
[cloudera@quickstart ~]$ tail osokin_a.i.gz -n 5
600X00U10003000000
00,000j0y2_50:000n!d#0a0000<000
07Z0p 0000S0z0_00,00:00K0000C000
p00:000000~000000Hv000s0v0000ys
00S0{00000Fke0ifG0000,000T0B0700'00
00)00t*S000 0000w000x~6.0000o0P#0000-00000n00~0000end0V0000p0007
[am0000
0G0i!000!000000]000000300p0z00s0y00s0#00$~:0y000T000k*00v000I00)0AA7!000s0y0
00M
0Y00Y{x00000G00r~5000A0;C000090+<000c`0000{0000A0B00050z00Y
000020,0;'0a00"0000000
T-F00000`
0000000000
q000[cloudera@quickstart ~]$
[cloudera@quickstart ~]$
```

```
Applications Places System Thu Apr 11, 11:18 AM cloudera
cloudera@quickstart:~
File Edit View Search Terminal Help
[cloudera@quickstart ~]$ less osokin_a.i.txt
```



```
Applications Places System Thu Apr 11, 11:18 AM cloudera
cloudera@quickstart:~
File Edit View Search Terminal Help
eeJN3sard4QdAB3G8SVWZ3vVX4Z2NbYudV2RFBVv/aRwR2+T7NCK1+lTEXA98kLQzSE0QmQvNh
uukTsZp8YgPA38Jgg7Nvp+dB005e8eoRBD9vVLScInhVN/LScNbge6712s17EuJVJ9H1u8IEG1NL
5zqGLu0SKUoraP4Fbo2k9EjnybUaHs5wb574+MV0r6BNkaUI3DmSkp2pZOCpJnD7vuuu7WN0sFQ
CF5LPgpTRio3G/R8iRfWtkhI+G9BEjp/8Me2UdKJbE9Z0Q0aEWEUpHLSFd7NgffVrf7yvMLiGJH
0mMJYFzpm5LudFgXok51qYwLYTdfFYyrxRKS1qS30KBYGiTTLnU4QbZCSxILYY/hGP56or05ZF
tKQymZ3KV3mBK590g2ikZC7Q8g2UJ45CSAIB3DpJidCjpe2TQ080yPgGbFw+f+GFL5ctBDJZJ9/t
gzr7N60ndvr200o0M4KbbdfmFeuq5B6GjvJMjGuLdiGGuFvILP+D3jP5htXeP0K6jHvAg41HECH+
89g1FNyeMGkKBZ/bis76CXyUBVtIBjiy085U7BNyftz+YAuXTvbuKDK93UQxfMIbwFL2JHL03Ljcs
U719y6GscQMa5bcwFcmRxFukT6pq/Z61+UIbyWjh50xl+QN791V04Z7/U9hFu45Amk65C8mIqqN
wtqCRMuFwKRW0cXLN0k51k201mHhJ6NXju5F9Q1Zqkxvffk00miM4+L0X+3F5SofsLIffXfTs/M
a9gcb0aggYbYtDkBYD5wEDKRn3N4umrZVnvs88FRKxgy6XeMfRG4ieomFRyLEuCFI2WafwyPd6C
z82Zm6hZ0S8qbFuQXPLy2ZvHCyukuA2/xzVETrtZzyDfwlpyp5y6B/FpkqFj8NRBHXUJBk4h8pR
ggT8K4Q9VXL46tq0IM1jz9jpaO2q+wJ8/y1hfa5aVM6iu0ui0bRLxog71fWnesMHhrIyZ0pYhD
aq6L910VrbEYn9TenKrXrHuUg4U4EDJienWyEd6iqrTPXk1kFEBXSux4fjiYmZQ/QG+mN6kPRCPF
GZu2A8+zUTrc3BAZmbx+04HV5MjfyafGcqTvn2Yjuuyop653KR4/ZLIwQCC5hJH0PaThtXrhjKS
WymVSDldqrZEDf3LPoVaA93lnzMoMbJyil8hbTu3ktyLhKuiX9qZ0Z/qd39NFVMq1H3vt8vHC6G9q
Jyj+h+e0S5SH8fgUT20BDX4J630pD10orkRRgvNhxNmDzgFzHyHqwfFzHgiLFAxSw+SR4vCZBa1V
M3yx26mX6mGcGn60DYPHNRHgZwJL1zxdT4s2BR+YbIPFE3dKIBJET4gu6R62okMLKhnwyt70mwp
qfKebD5bd/eeqVXe3m0Jc8MHvOyFTs0L9NMUwLJtYUM2L24u7oF90VXD9hR045Tp69unJbnfsY
1MnWb1xZ0p08Hrdh+GtJ5+1B08L8eCMGr9KbRRyp10bPch3z0xuXla567bxRL3iulZHCd1ixt
Zn1+74LElPL6IAD3pyPqgl5L/3/xEJVIdcFy7TB+mRiP/Rdx4FWTDcuS/+em7J1f1oF20oITmxM
RYZ7gBIUvrcuxWVWF/6Xg8BRBWS6li3eEkB56UCTSiiHGF1+3vV1zJ7nqovwGPA5r9RTS5g65
Jhzg3IACCWR08be4a0PI4NSJluuG40jI+2v0SQ+uTfGcQfVytwb4pVMebj6hePQj40bGtv/M/R0A
Ct35WpCRd0+qUxwba88MeU9Sglv2T15luKUW/WRYF0ng1x6xyX06hrRdczsw0G3lmCc+kERSnxs
aWkWhSQQdKjAhnXLFow0muJBFG/1ERFTmZSgUHIstrLHEvmx04r6hk4VhhVwtIsRxE2+At6i4dj
lGo2r8JfG8QfKw5HnQKXndQ8WRzPXaxceo6PmQ2YIUvWZz6I09d5Rhu1ZATP+43CD0G0LrYs7PWs
0AH/vxEzetiiIsxxktga0FG8C9Q+272vDsVyVHedz4Vvj0B0j9GnwK05Q/53KFXXCLMA5T0Sdx
8R3HH02bFM9VC/8yurrLzqDPZpiqC/BtJFz0z3TKLjwybLUA0ZeXUdIPBJak6PKIe0jBfFKkp
yFdgQbXwuyVrb9FHDW1PK8aKok7DKYiEYt2BgpiyaQodXoI0QsfMsuXk6LrtoMv6fymB8h3209b3
qrZLIdv0MhC/LBIMCh3xMW6rtJapVNoi+JsZ4+fBdJEQ83oK+p/evKQ540hWlvmvrguJXxpr3X23
0lnErW51zCcdgvdvG40L5RfHbMmPwCTP6rjf8CVsTJaZqCkh2+8QvLpRnxRvW06jCRNjyQ3hGSG
toRCYz+YqUnocL3K+FCXyWVRj554hbdFT7600TK5xk1s2lj+ee1AwgzL84aceWU69iaXSKZJ6F8
YpbzZAXN3sKjYijjVjJtBKHDB7mwb+VrH1pRdetxDIgW0Xo+zN7xvWHy50MhbJajhsTh9U6CRC
2ahow08td3QYuaW8DU29e55+9C8SbwY2LrdCIEMfSgCac/tLJhIz52bfj0CLED838P7glz8+VRE
0HQYP3FAnHRnP6qZEnMCHRBd/sAps8ZXLTEcyNB0J8fufI3gPukrfS0v1g+TL5a+vmshXJR7CiKF
TOB/izZJKfi2XRscab29cng65CSGR4B7MwJM27pgndvx7e1NsbcELXnkKhNDYvNPndjNXP0dkEuh
yb/aED0BbwNRZKHcd+m0iQqJCXh4GLMgSV0Ao5FDBfQeipVZ1544yFZYJxvFzVT2mm2q7mstpyL
:
```

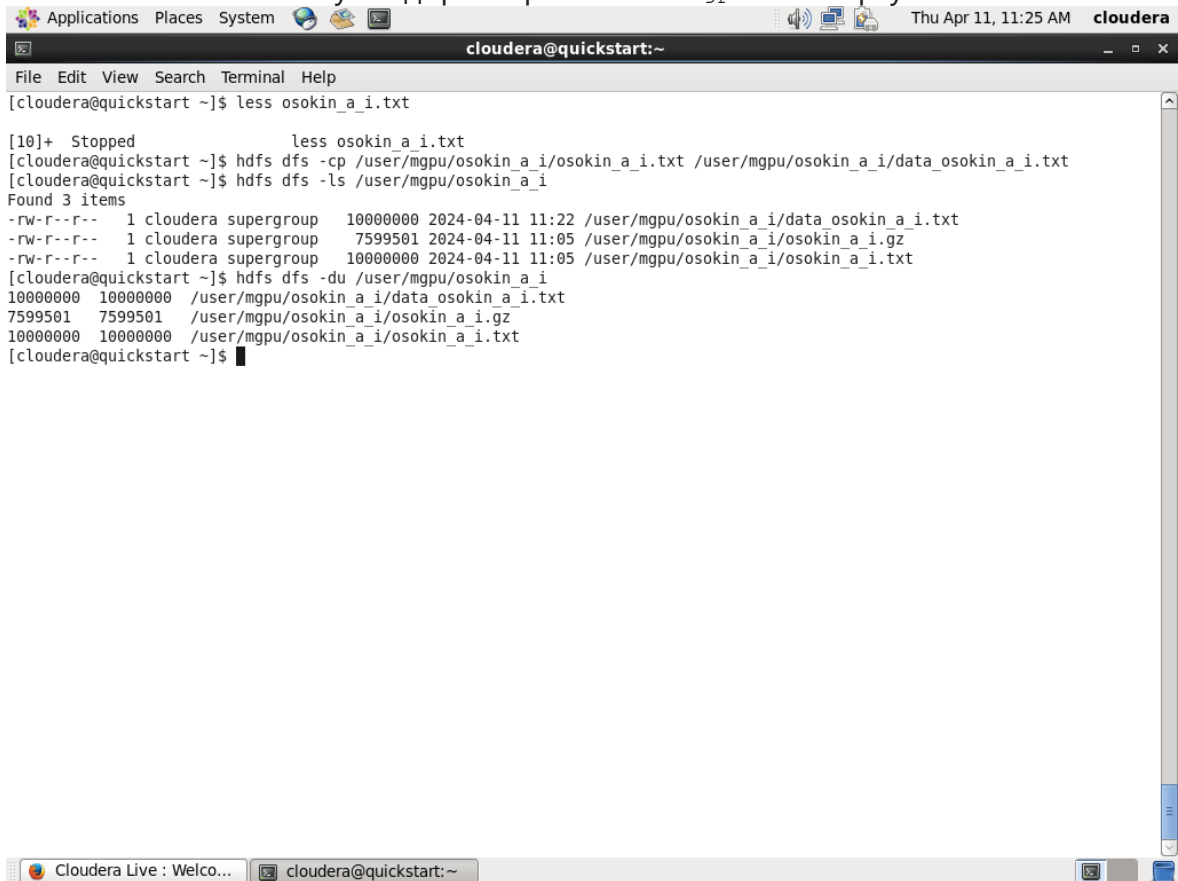
3.1.9. Создать копию файла file.txt вида date_file.txt, где в начале имени файла-копии указана текущая дата. Вывести листинг.



```
Applications Places System Thu Apr 11, 11:22 AM cloudera
cloudera@quickstart:~
File Edit View Search Terminal Help
[cloudera@quickstart ~]$ less osokin_a_i.txt

[10]+  Stopped                  less osokin_a_i.txt
[cloudera@quickstart ~]$ hdfs dfs -cp /user/mgpu/osokin_a_i/osokin_a_i.txt /user/mgpu/osokin_a_i/data_osokin_a_i.txt
[cloudera@quickstart ~]$ hdfs dfs -ls /user/mgpu/osokin_a_i
Found 3 items
-rw-r--r--  1 cloudera supergroup    10000000 2024-04-11 11:22 /user/mgpu/osokin_a_i/data osokin_a_i.txt
-rw-r--r--  1 cloudera supergroup    7599501 2024-04-11 11:05 /user/mgpu/osokin_a_i/osokin_a_i.gz
-rw-r--r--  1 cloudera supergroup    10000000 2024-04-11 11:05 /user/mgpu/osokin_a_i/osokin_a_i.txt
[cloudera@quickstart ~]$
```

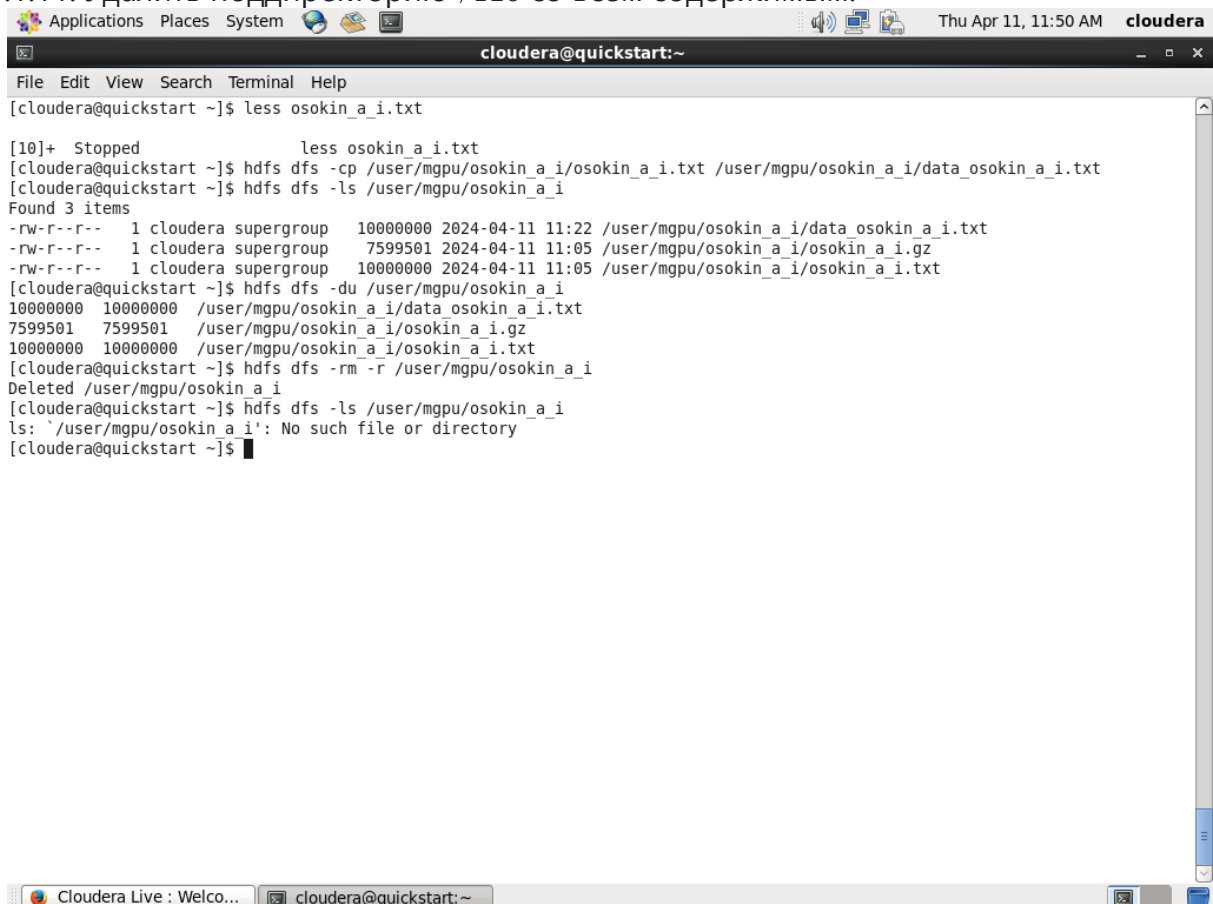
3.1.10. Вывести статистику по директории /user/mgpu/fio виртуальной машины.



```
Applications Places System Thu Apr 11, 11:25 AM cloudera
cloudera@quickstart:~
File Edit View Search Terminal Help
[cloudera@quickstart ~]$ less osokin_a_i.txt

[10]+  Stopped                  less osokin_a_i.txt
[cloudera@quickstart ~]$ hdfs dfs -cp /user/mgpu/osokin_a_i/osokin_a_i.txt /user/mgpu/osokin_a_i/data_osokin_a_i.txt
[cloudera@quickstart ~]$ hdfs dfs -ls /user/mgpu/osokin_a_i
Found 3 items
-rw-r--r--  1 cloudera supergroup  10000000 2024-04-11 11:22 /user/mgpu/osokin_a_i/data_osokin_a_i.txt
-rw-r--r--  1 cloudera supergroup   7599501 2024-04-11 11:05 /user/mgpu/osokin_a_i/osokin_a_i.gz
-rw-r--r--  1 cloudera supergroup  10000000 2024-04-11 11:05 /user/mgpu/osokin_a_i/osokin_a_i.txt
[cloudera@quickstart ~]$ hdfs dfs -du /user/mgpu/osokin_a_i
10000000 10000000 /user/mgpu/osokin_a_i/data_osokin_a_i.txt
7599501  7599501 /user/mgpu/osokin_a_i/osokin_a_i.gz
10000000 10000000 /user/mgpu/osokin_a_i/osokin_a_i.txt
[cloudera@quickstart ~]$
```

3.1.11. Удалить поддиректорию /fio со всем содержимым.



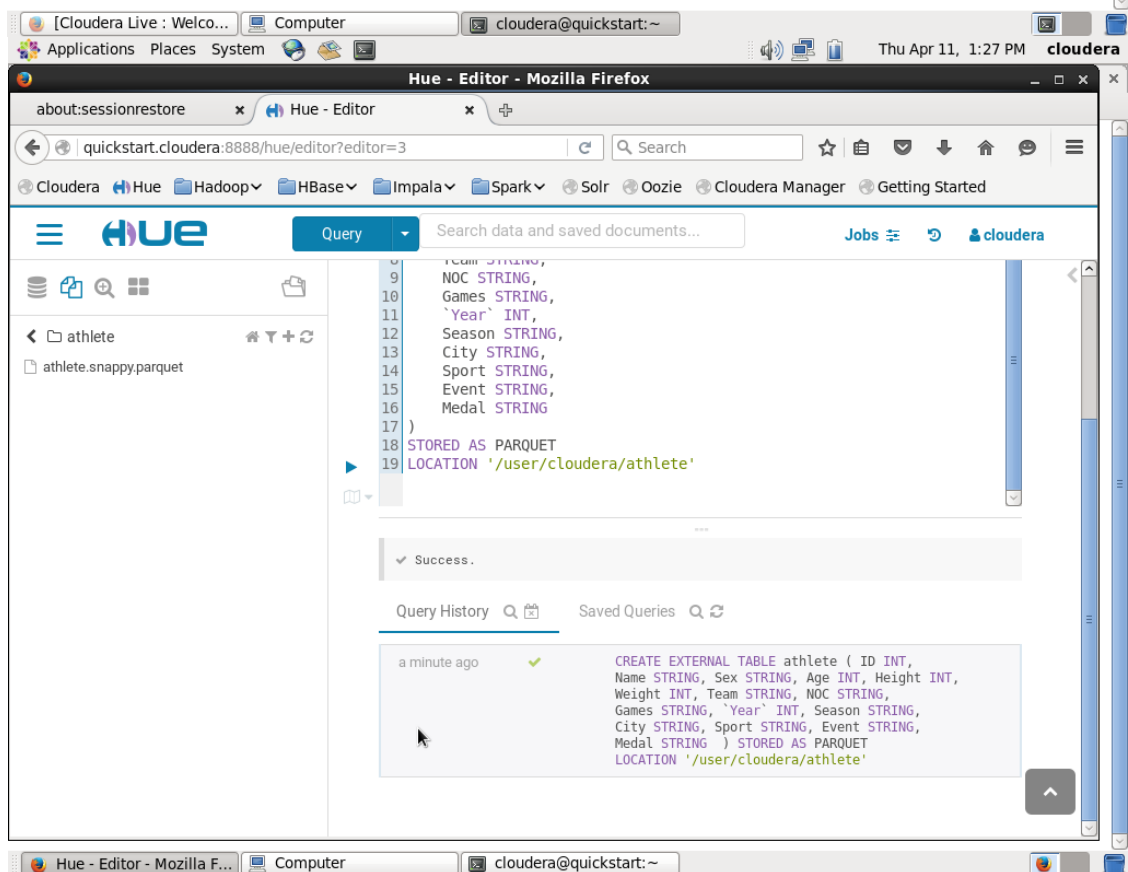
```
Applications Places System Thu Apr 11, 11:50 AM cloudera
cloudera@quickstart:~
File Edit View Search Terminal Help
[cloudera@quickstart ~]$ less osokin_a_i.txt

[10]+  Stopped                  less osokin_a_i.txt
[cloudera@quickstart ~]$ hdfs dfs -cp /user/mgpu/osokin_a_i/osokin_a_i.txt /user/mgpu/osokin_a_i/data_osokin_a_i.txt
[cloudera@quickstart ~]$ hdfs dfs -ls /user/mgpu/osokin_a_i
Found 3 items
-rw-r--r--  1 cloudera supergroup  10000000 2024-04-11 11:22 /user/mgpu/osokin_a_i/data_osokin_a_i.txt
-rw-r--r--  1 cloudera supergroup   7599501 2024-04-11 11:05 /user/mgpu/osokin_a_i/osokin_a_i.gz
-rw-r--r--  1 cloudera supergroup  10000000 2024-04-11 11:05 /user/mgpu/osokin_a_i/osokin_a_i.txt
[cloudera@quickstart ~]$ hdfs dfs -du /user/mgpu/osokin_a_i
10000000 10000000 /user/mgpu/osokin_a_i/data_osokin_a_i.txt
7599501  7599501 /user/mgpu/osokin_a_i/osokin_a_i.gz
10000000 10000000 /user/mgpu/osokin_a_i/osokin_a_i.txt
[cloudera@quickstart ~]$ hdfs dfs -rm -r /user/mgpu/osokin_a_i
Deleted /user/mgpu/osokin_a_i
[cloudera@quickstart ~]$ hdfs dfs -ls /user/mgpu/osokin_a_i
ls: '/user/mgpu/osokin_a_i': No such file or directory
[cloudera@quickstart ~]$
```

3.1.12. Подсчитать количество слов в файле внутри HDFS с помощью методологии Map Reduce (размер файла не менее 128 Мб).

3.1.13 Создание таблицы в Hive

```
Applications Places System cloudera@quickstart:~ Thu Apr 11, 1:22 PM cloudera
cloudera@quickstart:~$ ls
athlete.snappy.parquet Downloads lib Pictures
cloudera-manager eclipse Music Public
cm_api.py enterprise-deployment.json osokin_a_i.gz Templates
Desktop express-deployment.json osokin_a_i.txt Videos
Documents kerberos parcels workspace
[cloudera@quickstart ~]$ hdfs dfs -mkdir /user/cloudera/athlete
[cloudera@quickstart ~]$ hdfs dfs -copyFromLocal athlete.snappy.parquet /user/cloudera/athlete
copyFromLocal: /home/cloudera/athlete.snappy.parquet (Permission denied)
[cloudera@quickstart ~]$ sudo chmod 777 athlete.snappy.parquet
sudo: chmod: command not found
[cloudera@quickstart ~]$ sudo chmod 777 athlete.snappy.parquet
[cloudera@quickstart ~]$ hdfs dfs -copyFromLocal athlete.snappy.parquet /user/cloudera/athlete
[cloudera@quickstart ~]$ hdfs dfs -ls /user/cloudera/athlete
Found 1 items
-rw-r--r-- 1 cloudera cloudera 5572921 2024-04-11 13:22 /user/cloudera/athlete/athlete.snappy.parquet
[cloudera@quickstart ~]$
```



Applications Places System Thu Apr 11, 1:35 PM cloudera

Hue - Editor - Mozilla Firefox

about:sessionrestore x Hue - Editor x

quickstart.cloudera:8888/hue/editor?editor=5

Cloudera Hue Hadoop HBase Impala Spark Solr Oozie Cloudera Manager Getting Started

HUE Query Search data and saved documents... Jobs cloudera

Hive Add a name... Add a description...

default (1) Tables athlete

```
1 SELECT * from default.athlete limit 10
```

Query History Saved Queries Results (10)

	athlete.id	athlete.name	athlete.sex	athlete.age	athlet
1	1	A Dijiang	M	24	180
2	2	A Lamusi	M	23	170
3	3	Gunnar Nielsen Aaby	M	24	NULL
4	4	Edgar Lindenau Aabye	M	34	NULL
5	5	Christine Jacoba Aaftink	F	21	185
6	5	Christine Jacoba Aaftink	F	21	185

Hue - Editor - Mozilla F... Computer cloudera@quickstart:~

Applications Places System Thu Apr 11, 2:00 PM cloudera

Hue - Editor - Mozilla Firefox

about:sessionrestore x Hue - Editor x

quickstart.cloudera:8888/hue/editor?type=hive

Cloudera Hue Hadoop HBase Impala Spark Solr Oozie Cloudera Manager Getting Started

HUE Query Search data and saved documents... Jobs cloudera

5m, 30s default text

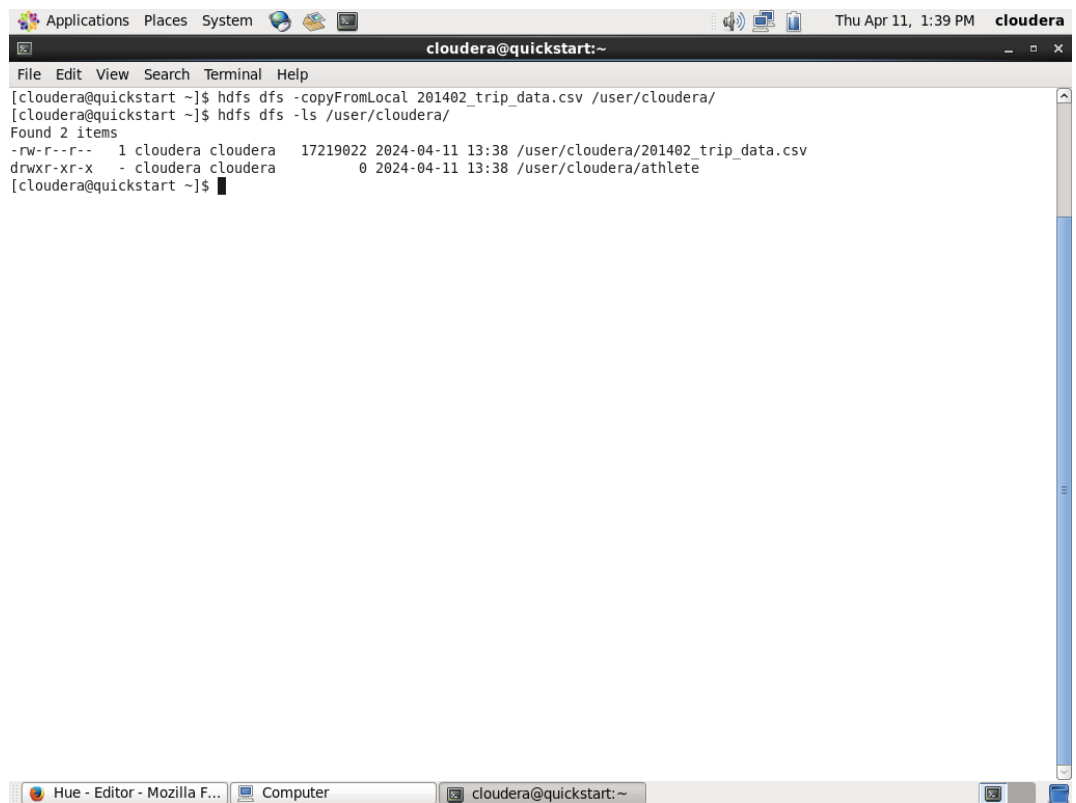
```
1 select `startstation`, `endstation`, count(*) as trips
2 from `default`.`201402_trip_data`
3 group by `startstation`, `endstation`
4 order by trips desc;
```

Query History Saved Queries Results (1,600)

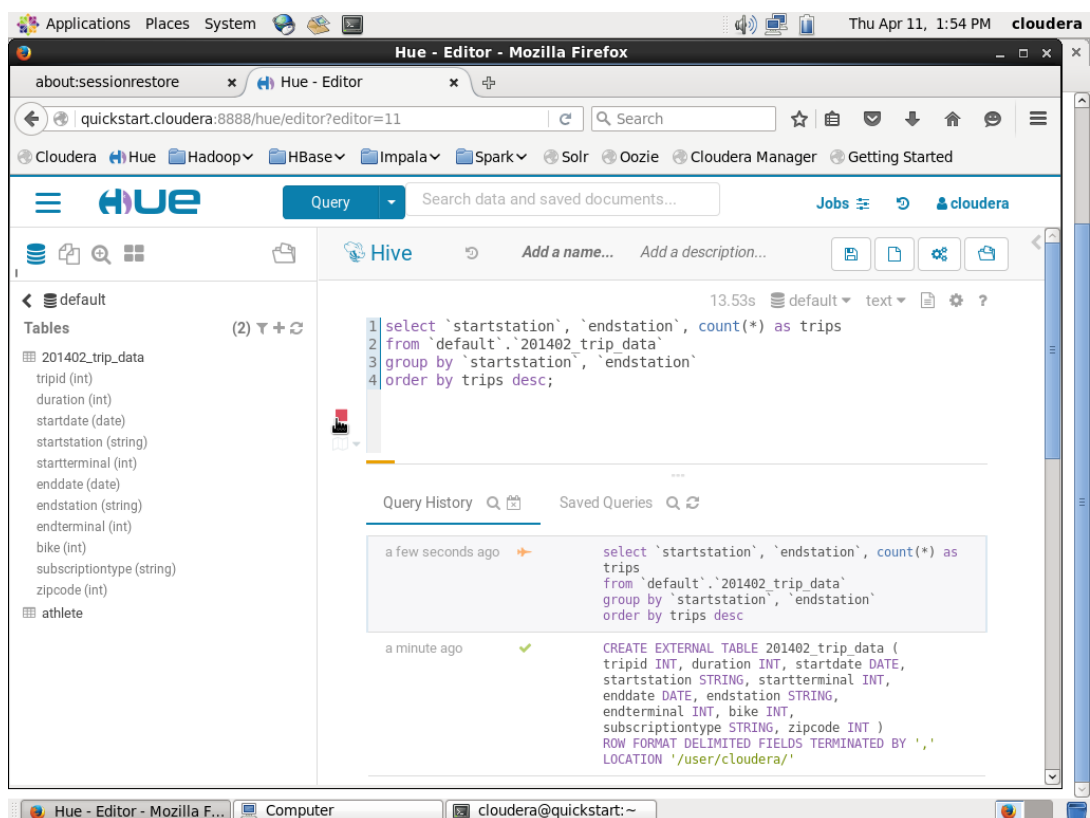
	startstation	endstation	trips
1	Harry Bridges Plaza (Ferry Building)	Embarcadero at Sansome	1330
2	Townsend at 7th	San Francisco Caltrain (Townsend at 4th)	1322
3	San Francisco Caltrain 2 (330 Townsend)	Townsend at 7th	1116
4	Market at Sansome	2nd at South Park	866
5	Embarcadero at Sansome	Steuart at Market	811
6	2nd at South Park	Market at Sansome	798
7	San Francisco Caltrain (Townsend at 4th)	Harry Bridges Plaza (Ferry Building)	782
8	2nd at Townsend	Harry Bridges Plaza (Ferry Building)	757

Hue - Editor - Mozilla F... [Computer] cloudera@quickstart:~

3.1.14 Проанализировать и визуализировать данные с помощью Impala(высокоскоростной механизм запросов SQL) или Hive.



```
Applications  Places  System  Thu Apr 11, 1:39 PM  cloudera
cloudera@quickstart:~
File Edit View Search Terminal Help
[cloudera@quickstart ~]$ hdfs dfs -copyFromLocal 201402_trip_data.csv /user/cloudera/
[cloudera@quickstart ~]$ hdfs dfs -ls /user/cloudera/
Found 2 items
-rw-r--r-- 1 cloudera cloudera 17219022 2024-04-11 13:38 /user/cloudera/201402_trip_data.csv
drwxr-xr-x - cloudera cloudera 0 2024-04-11 13:38 /user/cloudera/athlete
[cloudera@quickstart ~]$
```



The screenshot shows the Hue web interface in a Mozilla Firefox browser. The left sidebar displays the 'default' database with two tables: '201402_trip_data' and 'athlete'. The '201402_trip_data' table schema is listed as follows:

Field	Type
tripid	(int)
duration	(int)
startdate	(date)
startstation	(string)
startterminal	(int)
enddate	(date)
endstation	(string)
endterminal	(int)
bike	(int)
subscriptiontype	(string)
zipcode	(int)

The main panel shows a Hive query being executed:

```
1 select `startstation`, `endstation`, count(*) as trips
2 from `default`.`201402_trip_data`
3 group by `startstation`, `endstation`
4 order by trips desc;
```

The query history shows two recent queries:

- a few seconds ago: `select `startstation`, `endstation`, count(*) as trips from `default`.`201402_trip_data` group by `startstation`, `endstation` order by trips desc`
- a minute ago: `CREATE EXTERNAL TABLE 201402_trip_data (tripid INT, duration INT, startdate DATE, startstation STRING, startterminal INT, enddate DATE, endstation STRING, endterminal INT, bike INT, subscriptiontype STRING, zipcode INT) ROW FORMAT DELIMITED FIELDS TERMINATED BY ',' LOCATION '/user/cloudera/'`

Applications Places System Thu Apr 11, 1:56 PM cloudera

Hue - Job Browser - Mozilla Firefox

about:sessionrestore x Hue - Job Browser x

quickstart.cloudera:8888/hue/jobbrowser#!jobs

Cloudera Hue Hadoop HBase Impala Spark Solr Oozie Cloudera Manager Getting Started

HUE Query Search data and saved documents... Jobs 1 cloudera

Job Browser Jobs Workflows Schedules Bundles SLAs

default (2) Tables

201402_trip_data

tripid (int)
duration (int)
startdate (date)
startstation (string)
startterminal (int)
enddate (date)
endstation (string)
endterminal (int)
bike (int)
subscriptiontype (string)
zipcode (int)

athlete

user:cloudera

Succeeded Running Failed in the last 7 days

Kill

Id	Name	User	Type	Status	Progress
application_1712865687770_0001	select `startstation`, `endstation`, ...desc(Stage-1)	cloudera	MAPREDUCE	RUNNING	5%

1 jobs

Hue - Job Browser - M... [Computer] cloudera@quickstart:~

Applications Places System Thu Apr 11, 2:04 PM cloudera

Hue - Editor - Mozilla Firefox

about:sessionrestore x Hue - Editor x

quickstart.cloudera:8888/hue/editor?editor=12

Cloudera Hue Hadoop HBase Impala Spark Solr Oozie Cloudera Manager Getting Started

HUE Query Search data and saved documents... Jobs cloudera

Hive test Add a description...

5m, 30s default text

```
1 select `startstation`, `endstation`, count(*) as trips
2 from `default`.`201402_trip_data`
3 group by `startstation`, `endstation`
4 order by trips desc;
```

Query History Saved Queries Results (1,600)

	startstation	endstation
1	Harry Bridges Plaza (Ferry Building)	Embarcadero at Sansome
2	Townsend at 7th	San Francisco Caltrain (Townsend at 4th)
3	San Francisco Caltrain 2 (330 Townsend)	Townsend at 7th
4	Market at Sansome	2nd at South Park
5	Embarcadero at Sansome	Steuart at Market
	2nd at South Park	Market at Sansome

COLUMNS (3)

startstation STRING

endstation STRING

trips BIGINT

CSV

Excel

Download first 100000 rows as XLS

javascript:void(0)

Hue - Editor - Mozilla F... [Computer] cloudera@quickstart:~

