

Скачаем Hadoop:

```
user@user: ~  
  
System information as of Wed Dec 16 06:17:58 UTC 2020  
  
System load: 0.21          Processes:          93  
Usage of /: 19.5% of 8.79GB Users logged in:      1  
Memory usage: 4%          IP address for enp0s3: 192.168.1.101  
Swap usage: 0%  
  
33 packages can be updated.  
0 updates are security updates.  
  
New release '20.04.1 LTS' available.  
Run 'do-release-upgrade' to upgrade to it.  
  
Last login: Wed Dec 16 06:16:17 2020  
user@user:~$ wget https://archive.apache.org/dist/hadoop/core/hadoop-2.8.5/hadoop-2.8.5.tar.gz  
--2020-12-16 06:18:24-- https://archive.apache.org/dist/hadoop/core/hadoop-2.8.5/hadoop-2.8.5.tar.gz  
Resolving archive.apache.org (archive.apache.org)... 138.201.131.134, 2a01:4f8:172:2ec5::2  
Connecting to archive.apache.org (archive.apache.org)|138.201.131.134|:443... connected.  
HTTP request sent, awaiting response... 200 OK  
Length: 246543928 (235M) [application/x-gzip]  
Saving to: 'hadoop-2.8.5.tar.gz'  
  
hadoop-2.8.5.tar.gz      100%[=====] 235.12M  3.84MB/s   in 87s  
  
2020-12-16 06:19:52 (2.69 MB/s) - 'hadoop-2.8.5.tar.gz' saved [246543928/246543928]  
  
user@user:~$
```

Установим:

`tar -xvf hadoop-2.8.5.tar.gz`

Установим пути в файл ~/.bashrc:

```
export JAVA_HOME=/usr/lib/jvm/java-1.11.0-openjdk-amd64  
export HADOOP_HOME=$HOME/hadoop-2.8.5  
export HADOOP_HDFS_HOME=$HADOOP_HOME  
export PATH=$PATH:$HADOOP_HOME/sbin:$HADOOP_HOME/bin
```

```
user@user: ~  
GNU nano 2.9.3 /home/user/.bashrc  
  
. ~/.bash_aliases  
fi  
  
# enable programmable completion features (you don't need to enable  
# this, if it's already enabled in /etc/bash.bashrc and /etc/profile  
# sources /etc/bash.bashrc).  
if ! shopt -oq posix; then  
  if [ -f /usr/share/bash-completion/bash_completion ]; then  
    . /usr/share/bash-completion/bash_completion  
  elif [ -f /etc/bash_completion ]; then  
    . /etc/bash_completion  
  fi  
fi  
  
export JAVA_HOME=/usr/lib/jvm/java-1.11.0-openjdk-amd64  
export HADOOP_HOME=$HOME/hadoop-2.8.5  
export HADOOP_HDFS_HOME=$HADOOP_HOME  
export PATH=$PATH:$HADOOP_HOME/sbin:$HADOOP_HOME/bin  
  
^G Get Help  ^O Write Out  ^W Where Is   ^K Cut Text   ^J Justify    ^C Cur Pos    M-U Undo      M-A Mark Text  
^X Exit      ^R Read File  ^_ Replace    ^U Uncut Text ^T To Spell   ^_ Go To Line  M-E Redo      M-6 Copy Text
```

Посмотрим версию:

```
administrator@clickhousestudent: ~  
  
System load: 0.1          Processes:          128  
Usage of /: 59.3% of 19.56GB Users logged in:    1  
Memory usage: 3%         IP address for ens18: 192.168.121.176  
Swap usage: 0%  
  
* Introducing self-healing high availability clusters in MicroK8s.  
Simple, hardened, Kubernetes for production, from RaspberryPi to DC.  
  
https://microk8s.io/high-availability  
  
* Canonical Livepatch is available for installation.  
- Reduce system reboots and improve kernel security. Activate at:  
https://ubuntu.com/livepatch  
  
39 packages can be updated.  
0 updates are security updates.  
  
Last login: Sun Dec 13 12:52:47 2020 from 192.168.121.112  
administrator@clickhousestudent:~$ ls  
50916kek.w.txt btext.txt hadoop-2.8.5 hadoop-2.8.5.tar.gz index.html iris.data temp.txt temp5091614.txt test.txt  
administrator@clickhousestudent:~$ hdfs version  
Hadoop 2.8.5  
Subversion https://git-wip-us.apache.org/repos/asf/hadoop.git -r 0b8464d75227fcee2c6e7f2410377b3d53d3d5f8  
Compiled by jdu on 2018-09-10T03:32Z  
Compiled with protoc 2.5.0  
From source with checksum 9942ca5c745417c14e318835f420733  
This command was run using /home/administrator/hadoop-2.8.5/share/hadoop/common/hadoop-common-2.8.5.jar  
administrator@clickhousestudent:~$
```

Получим список файлов на удаленном сервере:

```
administrator@clickhousestudent: ~  
  
* Canonical Livepatch is available for installation.  
- Reduce system reboots and improve kernel security. Activate at:  
https://ubuntu.com/livepatch  
  
39 packages can be updated.  
0 updates are security updates.  
  
Last login: Sun Dec 13 12:52:47 2020 from 192.168.121.112  
administrator@clickhousestudent:~$ ls  
50916kek.w.txt btext.txt hadoop-2.8.5 hadoop-2.8.5.tar.gz index.html iris.data temp.txt temp5091614.txt test.txt  
administrator@clickhousestudent:~$ hdfs version  
Hadoop 2.8.5  
Subversion https://git-wip-us.apache.org/repos/asf/hadoop.git -r 0b8464d75227fcee2c6e7f2410377b3d53d3d5f8  
Compiled by jdu on 2018-09-10T03:32Z  
Compiled with protoc 2.5.0  
From source with checksum 9942ca5c745417c14e318835f420733  
This command was run using /home/administrator/hadoop-2.8.5/share/hadoop/common/hadoop-common-2.8.5.jar  
administrator@clickhousestudent:~$ hdfs dfs -ls hdfs://192.168.121.147:9000/lab4/  
WARNING: An illegal reflective access operation has occurred  
WARNING: Illegal reflective access by org.apache.hadoop.security.authentication.util.KerberosUtil (file:/home/administrator/hadoop-2.8.5/share/hadoop/common/lib/hadoop-auth-2.8.5.jar) to method sun.security.krb5.Config.getInstance()  
WARNING: Please consider reporting this to the maintainers of org.apache.hadoop.security.authentication.util.KerberosUtil  
WARNING: Use --illegal-access=warn to enable warnings of further illegal reflective access operations  
WARNING: All illegal access operations will be denied in a future release  
Found 2 items  
-rw-r--r-- 1 hadoop supergroup 174484 2020-12-09 09:09 hdfs://192.168.121.147:9000/lab4/Alice.txt  
drwxr-xr-x - hadoop supergroup 0 2020-12-09 09:28 hdfs://192.168.121.147:9000/lab4/test1out  
administrator@clickhousestudent:~$
```

Найдем файл part-r-00000:

```
administrator@clickhousestudent: ~  
  
The general command line syntax is  
command [genericOptions] [commandOptions]  
  
Usage: hadoop fs [generic options] -ls [-C] [-d] [-h] [-q] [-R] [-t] [-S] [-r] [-u] [<path> ...]  
administrator@clickhousestudent:~$ hdfs dfs -lsr hdfs://192.168.121.147:9000/lab4/  
WARNING: An illegal reflective access operation has occurred  
WARNING: Illegal reflective access by org.apache.hadoop.security.authentication.util.KerberosUtil (file:/home/administrator/hadoop-2.8.5/share/hadoop/common/lib/hadoop-auth-2.8.5.jar) to method sun.security.krb5.Config.getInstance()  
WARNING: Please consider reporting this to the maintainers of org.apache.hadoop.security.authentication.util.KerberosUtil  
WARNING: Use --illegal-access=warn to enable warnings of further illegal reflective access operations  
WARNING: All illegal access operations will be denied in a future release  
lsr: DEPRECATED: Please use 'ls -R' instead.  
-rw-r--r-- 1 hadoop supergroup 174484 2020-12-09 09:09 hdfs://192.168.121.147:9000/lab4/Alice.txt  
drwxr-xr-x - hadoop supergroup 0 2020-12-09 09:28 hdfs://192.168.121.147:9000/lab4/test1out  
-rw-r--r-- 1 hadoop supergroup 0 2020-12-09 09:28 hdfs://192.168.121.147:9000/lab4/test1out/_SUCCESS  
-rw-r--r-- 1 hadoop supergroup 61765 2020-12-09 09:28 hdfs://192.168.121.147:9000/lab4/test1out/part-r-00000  
administrator@clickhousestudent:~$ hdfs dfs -ls hdfs://192.168.121.147:9000/lab4/test1out  
WARNING: An illegal reflective access operation has occurred  
WARNING: Illegal reflective access by org.apache.hadoop.security.authentication.util.KerberosUtil (file:/home/administrator/hadoop-2.8.5/share/hadoop/common/lib/hadoop-auth-2.8.5.jar) to method sun.security.krb5.Config.getInstance()  
WARNING: Please consider reporting this to the maintainers of org.apache.hadoop.security.authentication.util.KerberosUtil  
WARNING: Use --illegal-access=warn to enable warnings of further illegal reflective access operations  
WARNING: All illegal access operations will be denied in a future release  
Found 2 items  
-rw-r--r-- 1 hadoop supergroup 0 2020-12-09 09:28 hdfs://192.168.121.147:9000/lab4/test1out/_SUCCESS  
-rw-r--r-- 1 hadoop supergroup 61765 2020-12-09 09:28 hdfs://192.168.121.147:9000/lab4/test1out/part-r-00000  
administrator@clickhousestudent:~$
```

Выполним команду lsr:

```
administrator@clickhousestudent: ~  
  
lsr: DEPRECATED: Please use 'ls -R' instead.  
-rw-r--r-- 1 hadoop supergroup 174484 2020-12-09 09:09 hdfs://192.168.121.147:9000/lab4/Alice.txt  
drwxr-xr-x - hadoop supergroup 0 2020-12-09 09:28 hdfs://192.168.121.147:9000/lab4/test1out  
-rw-r--r-- 1 hadoop supergroup 0 2020-12-09 09:28 hdfs://192.168.121.147:9000/lab4/test1out/_SUCCESS  
-rw-r--r-- 1 hadoop supergroup 61765 2020-12-09 09:28 hdfs://192.168.121.147:9000/lab4/test1out/part-r-00000  
administrator@clickhousestudent:~$ hdfs dfs -ls hdfs://192.168.121.147:9000/lab4/test1out  
WARNING: An illegal reflective access operation has occurred  
WARNING: Illegal reflective access by org.apache.hadoop.security.authentication.util.KerberosUtil (file:/home/administrator/hadoop-2.8.5/share/hadoop/common/lib/hadoop-auth-2.8.5.jar) to method sun.security.krb5.Config.getInstance()  
WARNING: Please consider reporting this to the maintainers of org.apache.hadoop.security.authentication.util.KerberosUtil  
WARNING: Use --illegal-access=warn to enable warnings of further illegal reflective access operations  
WARNING: All illegal access operations will be denied in a future release  
Found 2 items  
-rw-r--r-- 1 hadoop supergroup 0 2020-12-09 09:28 hdfs://192.168.121.147:9000/lab4/test1out/_SUCCESS  
-rw-r--r-- 1 hadoop supergroup 61765 2020-12-09 09:28 hdfs://192.168.121.147:9000/lab4/test1out/part-r-00000  
administrator@clickhousestudent:~$ hdfs dfs -lsr hdfs://192.168.121.147:9000/lab4/  
WARNING: An illegal reflective access operation has occurred  
WARNING: Illegal reflective access by org.apache.hadoop.security.authentication.util.KerberosUtil (file:/home/administrator/hadoop-2.8.5/share/hadoop/common/lib/hadoop-auth-2.8.5.jar) to method sun.security.krb5.Config.getInstance()  
WARNING: Please consider reporting this to the maintainers of org.apache.hadoop.security.authentication.util.KerberosUtil  
WARNING: Use --illegal-access=warn to enable warnings of further illegal reflective access operations  
WARNING: All illegal access operations will be denied in a future release  
lsr: DEPRECATED: Please use 'ls -R' instead.  
-rw-r--r-- 1 hadoop supergroup 174484 2020-12-09 09:09 hdfs://192.168.121.147:9000/lab4/Alice.txt  
drwxr-xr-x - hadoop supergroup 0 2020-12-09 09:28 hdfs://192.168.121.147:9000/lab4/test1out  
-rw-r--r-- 1 hadoop supergroup 0 2020-12-09 09:28 hdfs://192.168.121.147:9000/lab4/test1out/_SUCCESS  
-rw-r--r-- 1 hadoop supergroup 61765 2020-12-09 09:28 hdfs://192.168.121.147:9000/lab4/test1out/part-r-00000  
administrator@clickhousestudent:~$
```

cat:

`hdfs dfs -cat hdfs://192.168.121.147:9000/lab4/test1out/part-r-00000`

```
administrator@clickhousestudent: ~/50916kek  
Let 6  
Lewis 4  
License 8  
License. 2  
Like 1  
Literary 13  
Little 2  
Lizard 2  
Lizard) 1  
Lizard, 2  
Lizard's 1  
Lobster 5  
Lobster; 1  
London 1  
Long 2  
Longitude 2  
Lory 3  
Lory, 3  
Lory. 1  
Luckily 1  
M, 1  
M?" 1  
MARMALADE", 1  
ME," 2  
MERCHANTABILITY 1  
ME" 1  
MILLENNIUM 1  
Ma!" 1  
Mabel 1  
Mabel! 1
```

Самое часто встречающееся слово нельзя быстро найти, так как различных слов в тексте очень много

awk:

```
hdfs dfs -cat hdfs://192.168.121.147:9000/lab4/test1out/part-r-00000 | awk '{print $1":"$2}'
```

```
administrator@clickhousestudent: ~/50916kek  
fact.:1  
fact.:1  
fading:1  
failure.:1  
faint:1  
faintly:1  
fair:1  
fairly,:1  
fairy-:1  
fall:6  
fall,:1  
fallen:4  
falling:2  
familarly:1  
family:1  
fan:8  
fan!:1  
fan,:1  
fancied:2  
fancy:3  
fancy,:2  
fancying:1  
fancy-who's:1  
fanned:1  
fanning:1  
far:9  
far!:1  
far,:1  
far,:2  
farm-yard-while:1
```

Найдем все the в тексте:

```
hdfs dfs -cat hdfs://192.168.121.147:9000/lab4/test1out/part-r-00000 | awk '{print $1":"$2}' | grep the
```

```

administrator@clickhousestudent: ~/50916kekw
others:4
others!":1
others.:3
otherwise:2
otherwise,:1
otherwise,"":1
other-Bill!:1
out-of-the-way:3
rather:25
reality-the:1
reeds-the:1
sisters-they:1
the:1677
their:50
theirs,:1
them:49
them!":5
them,:21
them,"":3
them.:2
them."":2
them::1
themselves:2
themselves,"":1
them-all:1
them-and:2
them-"I:1
them-":1
then:50
then!-Bill's:1

```

Численная сортировка:

```

hdfs dfs -cat hdfs://192.168.121.147:9000/lab4/test1out/part-r-00000 | awk '{print $1":"$2}' | grep the | sort -t: -nk2

```

```

administrator@clickhousestudent: ~/50916kekw
others.:3
out-of-the-way:3
them,"":3
they'd:3
Father:4
others:4
they'll:4
other,:5
them!":5
"the:5
either:6
then,:6
then,"":6
they're:9
whether:11
these:14
there's:15
another:21
them,:21
rather:25
other:40
there:44
them:49
their:50
then:50
they:110
the:1677
administrator@clickhousestudent:~/50916kekw$

```

Найдем слово Alice:

```

hdfs dfs -cat hdfs://192.168.121.147:9000/lab4/test1out/part-r-00000 | awk '{print $1":"$2}' | grep Alice | sort -t: -nk2

```

```
administrator@clickhousestudent: ~/50916kek$ hdfs dfs -cat hdfs://192.168.121.147:9000/lab4/test1out/part-r-00000 | awk '{print $1":"$2}' | grep Alice | sort -t: -nk2
WARNING: An illegal reflective access operation has occurred
WARNING: Illegal reflective access by org.apache.hadoop.security.authentication.util.KerberosUtil (file:/home/administrator/hadoop-2.8.5/share/hadoop/common/lib/hadoop-auth-2.8.5.jar) to method sun.security.krb5.Config.getInstance()
WARNING: Please consider reporting this to the maintainers of org.apache.hadoop.security.authentication.util.KerberosUtil
WARNING: Use --illegal-access=warn to enable warnings of further illegal reflective access operations
WARNING: All illegal access operations will be denied in a future release
Alice:7
Alice!:1
Alice)-and:1
Alice's,:1
Alice's:1
"Alice!":1
Alice,):2
Alice!:3
(Alice:4
Alice's:14
Alice,:16
Alice.:54
Alice.:76
Alice:221
administrator@clickhousestudent:~/50916kek$
```

Получить корректный ответ мешают знаки препинания.

Уберем знаки препинания:

```
hdfs dfs -cat hdfs://192.168.121.147:9000/lab4/test1out/part-r-00000 | awk '{gsub(/[[[:punct:]]/, ""); print $1":"$2}' | grep Alice | sort -t: -nk2
```

```
administrator@clickhousestudent: ~
administrator@clickhousestudent:~$ {gsub(/[[[:punct:]]/, ""); print $1":"$2}' | grep Alice | sort -t: -nk2
-bash: syntax error near unexpected token `/{[:punct:]]/, ""'
dfsadministrator@clickhousestudent:~$ hdfs dfs -cat hdfs://192.168.121.147:9000/lab4/test1out/part-r-00000 | awk '{gsub(/[[[:punct:]]/, ""); print $1":"$2}' | grep Alice | sort -t: -nk2
gs> {gsub(/[[[:punct:]]/, ""); print $1":"$2}' | grep Alice | sort -t: -nk2
WARNING: An illegal reflective access operation has occurred
WARNING: Illegal reflective access by org.apache.hadoop.security.authentication.util.KerberosUtil (file:/home/administrator/hadoop-2.8.5/share/hadoop/common/lib/hadoop-auth-2.8.5.jar) to method sun.security.krb5.Config.getInstance()
WARNING: Please consider reporting this to the maintainers of org.apache.hadoop.security.authentication.util.KerberosUtil
WARNING: Use --illegal-access=warn to enable warnings of further illegal reflective access operations
WARNING: All illegal access operations will be denied in a future release
Alice:1
Alice:1
Aliceand:1
Alices:1
Alices:1
Alice:2
Alice:3
Alice:4
Alice:7
Alices:14
Alice:16
Alice:54
Alice:76
Alice:221
administrator@clickhousestudent:~$
administrator@clickhousestudent:~$
```

Создадим на диске файл и запишем туда состав группы:

```
administrator@clickhousestude x + -
GNU nano 2.9.3 50916kekwait

ИТ - 50916
Старовойтов Иван Николаевич
https://github.com/Emmeron45/BigData
Смирнов Константин Игоревич
https://github.com/konnectr/BigData
Заводиленко Екатерина Игоревна
https://github.com/EkaterinaZavodilenko

[ Read 7 lines ]
^G Get Help  ^O Write Out  ^W Where Is  ^K Cut Text   ^J Justify   ^C Cur Pos   M-U Undo     M-A Mark Text
^X Exit      ^R Read File  ^\ Replace   ^U Uncut Text ^T To Spell  ^_ Go To Line M-E Redo     M-G Copy Text
```

Переместим файл в папку exam:

```
hdfs dfs -put 50916kekwait hdfs://192.168.121.147:9000/exam/
```

```
hdfs dfs -ls 50916kekwait hdfs://192.168.121.147:9000/exam/
```

```
l
WARNING: Use --illegal-access=warn to enable warnings of further illegal reflective access operations
WARNING: All illegal access operations will be denied in a future release
put: '50916kekwait': File exists
administrator@clickhousestudent:~/50916kekwait$ ls
1file._COPYING_ 50916kekwait
administrator@clickhousestudent:~/50916kekwait$ hdfs dfs -put 50916kekwait hdfs://192.168.121.147:9000/exam/
WARNING: An illegal reflective access operation has occurred
WARNING: Illegal reflective access by org.apache.hadoop.security.authentication.util.KerberosUtil (file:/home/administrator/hadoop-2.8.5/share/hadoop/common/lib/hadoop-auth-2.8.5.jar) to method sun.security.krb5.Config.getInstance()
WARNING: Please consider reporting this to the maintainers of org.apache.hadoop.security.authentication.util.KerberosUtil
l
WARNING: Use --illegal-access=warn to enable warnings of further illegal reflective access operations
WARNING: All illegal access operations will be denied in a future release
administrator@clickhousestudent:~/50916kekwait$ hdfs dfs -ls 50916kekwait hdfs://192.168.121.147:9000/exam/
WARNING: An illegal reflective access operation has occurred
WARNING: Illegal reflective access by org.apache.hadoop.security.authentication.util.KerberosUtil (file:/home/administrator/hadoop-2.8.5/share/hadoop/common/lib/hadoop-auth-2.8.5.jar) to method sun.security.krb5.Config.getInstance()
WARNING: Please consider reporting this to the maintainers of org.apache.hadoop.security.authentication.util.KerberosUtil
l
WARNING: Use --illegal-access=warn to enable warnings of further illegal reflective access operations
WARNING: All illegal access operations will be denied in a future release
-rw-rw-r-- 1 administrator administrator      291 2020-12-16 07:52 50916kekwait
Found 1 items
-rw-r--r-- 3 administrator supergroup        291 2020-12-16 07:53 hdfs://192.168.121.147:9000/exam/50916kekwait
administrator@clickhousestudent:~/50916kekwait$ nano 50916kekwait
administrator@clickhousestudent:~/50916kekwait$ hdfs dfs -ls 50916kekwait hdfs://192.168.121.147:9000/exam/
```

Переместим файл в папку lab4:


```
administrator@clickhousestude x + v
tor/hadoop-2.8.5/share/hadoop/common/lib/hadoop-auth-2.8.5.jar) to method sun.security.krb5.Config.getInstance()
WARNING: Please consider reporting this to the maintainers of org.apache.hadoop.security.authentication.util.KerberosUtil
WARNING: Use --illegal-access=warn to enable warnings of further illegal reflective access operations
WARNING: All illegal access operations will be denied in a future release
administrator@clickhousestudent:~/50916kekwa$ hdfs dfs -ls 50916kekwait hdfs://192.168.121.147:9000/exam/
WARNING: An illegal reflective access operation has occurred
WARNING: Illegal reflective access by org.apache.hadoop.security.authentication.util.KerberosUtil (file:/home/administra
tor/hadoop-2.8.5/share/hadoop/common/lib/hadoop-auth-2.8.5.jar) to method sun.security.krb5.Config.getInstance()
WARNING: Please consider reporting this to the maintainers of org.apache.hadoop.security.authentication.util.KerberosUtil
WARNING: Use --illegal-access=warn to enable warnings of further illegal reflective access operations
WARNING: All illegal access operations will be denied in a future release
-rw-rw-r-- 1 administrator administrator 291 2020-12-16 07:52 50916kekwait
Found 1 items
-rw-r--r-- 3 administrator supergroup 291 2020-12-16 07:53 hdfs://192.168.121.147:9000/exam/50916kekwait
administrator@clickhousestudent:~/50916kekwa$ nano 50916kekwait
administrator@clickhousestudent:~/50916kekwa$ hdfs dfs -put 50916kekwait hdfs://192.168.121.147:9000/lab4
WARNING: An illegal reflective access operation has occurred
WARNING: Illegal reflective access by org.apache.hadoop.security.authentication.util.KerberosUtil (file:/home/administra
tor/hadoop-2.8.5/share/hadoop/common/lib/hadoop-auth-2.8.5.jar) to method sun.security.krb5.Config.getInstance()
WARNING: Please consider reporting this to the maintainers of org.apache.hadoop.security.authentication.util.KerberosUtil
WARNING: Use --illegal-access=warn to enable warnings of further illegal reflective access operations
WARNING: All illegal access operations will be denied in a future release
put: Permission denied: user=administrator, access=WRITE, inode="/lab4":hadoop:supergroup:drwxr-xr-x
administrator@clickhousestudent:~/50916kekwa$ hdfs dfs -put 50916kekwait hdfs://192.168.121.147:9000/lab4
```

put: Permission denied: user=administrator, access=WRITE,
inode="/lab4":hadoop:supergroup:drwxr-xr-x