

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

## **Практическая работа № 6**

Тема: «Hadoop»

Выполнила: Бринькован В. А., ТП-191

Руководитель: Босенко Т.М.

Москва

2022

1. Установим OpenJDK (JDK 8):

```
lemp001@u20-01: ~  
lemp001@u20-01:~$ sudo apt-get update  
[sudo] password for lemp001:  
Hit:1 http://ru.archive.ubuntu.com/ubuntu focal InRelease  
Hit:2 http://ru.archive.ubuntu.com/ubuntu focal-updates InRelease  
Hit:3 http://ru.archive.ubuntu.com/ubuntu focal-backports InRelease  
Hit:4 http://deb.anydesk.com all InRelease  
Hit:5 https://dl.google.com/linux/chrome/deb stable InRelease  
Hit:6 http://security.ubuntu.com/ubuntu focal-security InRelease  
Hit:7 https://download.docker.com/linux/ubuntu focal InRelease  
Reading package lists... Done  
lemp001@u20-01:~$ sudo apt-get install openjdk-8-jdk  
Reading package lists... Done  
Building dependency tree  
Reading state information... Done  
The following packages were automatically installed and are no longer required:  
  chromium-codecs-ffmpeg-extra gstreamer1.0-vaapi libfwupdplugin1  
  libgstreamer-plugins-bad1.0-0 libva-wayland2 libxmlb1  
Use 'sudo apt autoremove' to remove them.  
The following additional packages will be installed:  
  ca-certificates-java fonts-dejavu-extra java-common libatk-wrapper-java  
  libatk-wrapper-java-jni libice-dev libpthread-stubs0-dev libsm-dev  
  libx11-dev libxau-dev libxcb1-dev libxdmcp-dev libxt-dev  
  openjdk-8-jdk-headless openjdk-8-jre openjdk-8-jre-headless  
  x11proto-core-dev x11proto-dev xorg-sgml-doctools xtrans-dev
```

2. Проверим установку:

```
lemp001@u20-01:~$ java -version  
openjdk version "1.8.0_342"  
OpenJDK Runtime Environment (build 1.8.0_342-8u342-b07-0ubuntu1~20.04-b07)  
OpenJDK 64-Bit Server VM (build 25.342-b07, mixed mode)
```

3. Установим JAVA\_HOME и JRE\_HOME:

```
lemp001@u20-01: ~  
GNU nano 4.8 /etc/environment Modified  
PATH="/usr/local/sbin:/usr/local/bin:/usr/sbin:/usr/bin:/sbin:/bin:/usr/games:/>  
JAVA_HOME="/usr/lib/jvm/java-8-openjdk-amd64"  
JRE_HOME="/usr/lib/jvm/java-8-openjdk-amd64/jre"  
  
[ Read 3 lines ]  
^G Get Help ^O Write Out ^W Where Is ^K Cut Text ^J Justify ^C Cur Pos  
^X Exit ^R Read File ^\ Replace ^U Paste Text ^T To Spell ^_ Go To Line
```

4. Создадим нового пользователя

```
lemp001@u20-01:~$ sudo adduser hadoop  
Adding user `hadoop' ...  
Adding new group `hadoop' (1001) ...  
Adding new user `hadoop' (1001) with group `hadoop' ...  
Creating home directory `/home/hadoop' ...  
Copying files from `/etc/skel' ...  
New password:  
Retype new password:  
passwd: password updated successfully  
Changing the user information for hadoop  
Enter the new value, or press ENTER for the default  
Full Name []:  
Room Number []:  
Work Phone []:  
Home Phone []:  
Other []:  
Is the information correct? [Y/n] y
```

5. Переключимся на него:

```
lemp001@u20-01:~$ sudo su hadoop  
hadoop@u20-01:/home/lemp001$
```

6. Переключимся назад к корневому пользователю:

```
hadoop@u20-01:/home/lemp001$ exit  
exit
```

7. Установим SSH и PDSH:

```

lemp001@u20-01:~$ sudo apt-get install ssh pdsh
Reading package lists... Done
Building dependency tree
Reading state information... Done
The following packages were automatically installed and are no longer required:
  chromium-codecs-ffmpeg-extra gstreamer1.0-vaapi libfwupdplugin1
  libgstreamer-plugins-bad1.0-0 libva-wayland2 libxmlb1
Use 'sudo apt autoremove' to remove them.
The following additional packages will be installed:
  genders libgenders0 ncurses-term openssh-server openssh-sftp-server
  ssh-import-id
Suggested packages:
  rdist molly-guard monkeysphere ssh-askpass
The following NEW packages will be installed:
  genders libgenders0 ncurses-term openssh-server openssh-sftp-server pdsh ssh
  ssh-import-id
0 upgraded, 8 newly installed, 0 to remove and 1 not upgraded.
Need to get 860 kB of archives.

```

8. Создадим приватную/публичную ключевую пару для пользователя Hadoop:

```

lemp001@u20-01:~$ sudo su hadoop
hadoop@u20-01:/home/lemp001$ cd
hadoop@u20-01:~$ ssh-keygen -t rsa -N "" -f /home/hadoop/.ssh/id_rsa
Generating public/private rsa key pair.
Created directory '/home/hadoop/.ssh'.
Your identification has been saved in /home/hadoop/.ssh/id_rsa
Your public key has been saved in /home/hadoop/.ssh/id_rsa.pub
The key fingerprint is:
SHA256:5xRD/lWiTg284G78Ye13SzEqNV/WQLGZunqhgvg47uQ hadoop@u20-01
The key's randomart image is:
+---[RSA 3072]---+
|                 . . . +.. |
|                o. . = *  |
|               .+.o.B   |
|               *.o  ..  |
|              Soo ++ o+ |
|               ++ =.+.+ |
|              .. . ..=. o |
|             oo.. . ..+ o o|
|            oEo.   ...  oo|
+-----[SHA256]-----+

```

9. Добавим публичный ключ для авторизованных ключей файла:

```

hadoop@u20-01:~$ cat /home/hadoop/.ssh/id_rsa.pub >> /home/hadoop/.ssh/authorized_keys
hadoop@u20-01:~$ chmod 0600 /home/hadoop/.ssh/authorized_keys

```

10. Проверим, работает ли SSH:



```
hadoop@u20-01: ~  
hadoop@u20-01:~$ ssh localhost  
The authenticity of host 'localhost (127.0.0.1)' can't be established.  
ECDSA key fingerprint is SHA256:lWxB3l0e5Y1qU4UmVsunCezg1K8dCUsrTRlq/H/yJIE.  
Are you sure you want to continue connecting (yes/no/[fingerprint])? y  
Please type 'yes', 'no' or the fingerprint: yes  
Warning: Permanently added 'localhost' (ECDSA) to the list of known hosts.  
Welcome to Ubuntu 20.04.5 LTS (GNU/Linux 5.15.0-52-generic x86_64)  
  
* Documentation:  https://help.ubuntu.com  
* Management:    https://landscape.canonical.com  
* Support:       https://ubuntu.com/advantage  
  
0 updates can be applied immediately.  
  
Your Hardware Enablement Stack (HWE) is supported until April 2025.  
  
The programs included with the Ubuntu system are free software;  
the exact distribution terms for each program are described in the  
individual files in /usr/share/doc/*/copyright.  
  
Ubuntu comes with ABSOLUTELY NO WARRANTY, to the extent permitted by  
applicable law.  
hadoop@u20-01:~$
```

```
hadoop@u20-01:~$ exit  
logout  
Connection to localhost closed.  
hadoop@u20-01:~$ S
```

## 11. Установим Hadoop (v3.1.1):

```
hadoop@u20-01:~$ wget https://archive.apache.org/dist/hadoop/common/hadoop-3.1.2/hadoop-3.1.2.tar.gz  
--2022-11-01 10:15:38-- https://archive.apache.org/dist/hadoop/common/hadoop-3.1.2/hadoop-3.1.2.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: 332433589 (317M) [application/x-gzip]  
Saving to: 'hadoop-3.1.2.tar.gz'  
  
hadoop-3.1.2.tar.gz 100%[=====>] 317,03M 1,70MB/s in 2m 43s  
  
2022-11-01 10:18:21 (1,95 MB/s) - 'hadoop-3.1.2.tar.gz' saved [332433589/332433589]
```

## 12. Извлечём Binaries:

```
hadoop@u20-01: ~  
hadoop-3.1.2/share/doc/hadoop/images/apache-maven-project-2.png  
hadoop-3.1.2/share/doc/hadoop/images/icon_error_sml.gif  
hadoop-3.1.2/share/doc/hadoop/images/logo_apache.jpg  
hadoop-3.1.2/share/doc/hadoop/images/icon_warning_sml.gif  
hadoop-3.1.2/share/doc/hadoop/images/newwindow.png  
hadoop-3.1.2/share/doc/hadoop/images/maven-logo-2.gif  
hadoop-3.1.2/share/doc/hadoop/images/banner.jpg  
hadoop-3.1.2/share/doc/hadoop/images/logo_maven.jpg  
hadoop-3.1.2/share/doc/hadoop/images/collapsed.gif  
hadoop-3.1.2/share/doc/hadoop/images/external.png  
hadoop-3.1.2/share/doc/hadoop/images/h5.jpg  
hadoop-3.1.2/share/doc/hadoop/images/logos/  
hadoop-3.1.2/share/doc/hadoop/images/logos/build-by-maven-white.png  
hadoop-3.1.2/share/doc/hadoop/images/logos/build-by-maven-black.png  
hadoop-3.1.2/share/doc/hadoop/images/logos/maven-feather.png  
hadoop-3.1.2/share/doc/hadoop/index.html  
hadoop-3.1.2/share/doc/hadoop/project-reports.html  
hadoop-3.1.2/include/  
hadoop-3.1.2/include/hdfs.h  
hadoop-3.1.2/include/Pipes.hh  
hadoop-3.1.2/include/SerialUtils.hh  
hadoop-3.1.2/include/StringUtils.hh  
hadoop-3.1.2/include/TemplateFactory.hh  
hadoop@u20-01:~$
```

13.Перенесём Binaries:

```
hadoop@u20-01:~$ mv hadoop-3.1.2 hadoop
```

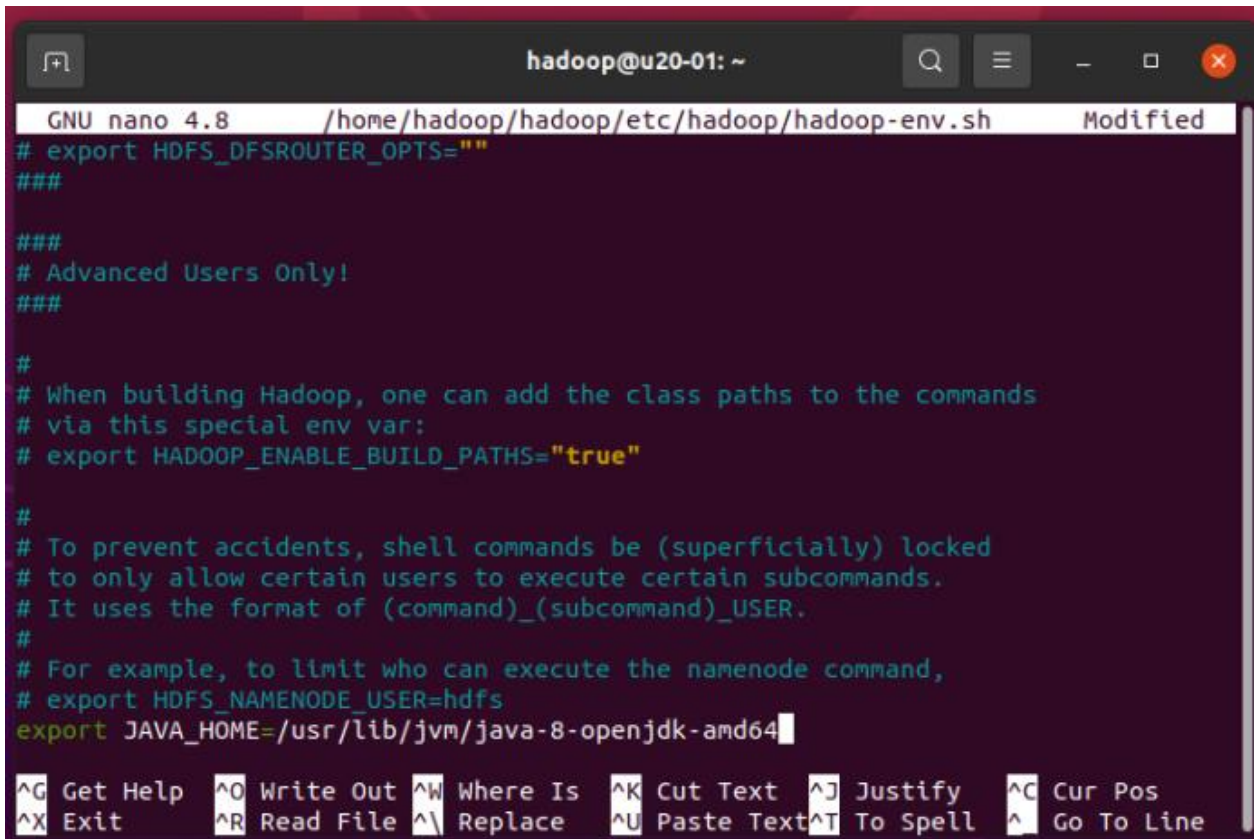
14.Установим переменные для среды UNIX:

```
GNU nano 4.8 .bashrc Modified  
# 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 HADOOP_HOME=/home/hadoop/hadoop  
export HADOOP_INSTALL=$HADOOP_HOME  
export HADOOP_MAPRED_HOME=$HADOOP_HOME  
export HADOOP_COMMON_HOME=$HADOOP_HOME  
export HADOOP_HDFS_HOME=$HADOOP_HOME  
export YARN_HOME=$HADOOP_HOME  
export HADOOP_COMMON_LIB_NATIVE_DIR=$HADOOP_HOME/lib/native  
export PATH=$PATH:$HADOOP_HOME/sbin:$HADOOP_HOME/bin  
export PDSH RCMD TYPE=ssh  
File Name to Write: .bashrc  
^G Get Help      M-D DOS Format  M-A Append      M-B Backup File  
^C Cancel        M-M Mac Format  M-P Prepend     ^T To Files
```



```
hadoop@u20-01:~$ source .bashrc
```

15. Добавим переменные для среды Hadoop:



```
GNU nano 4.8 /home/hadoop/hadoop/etc/hadoop/hadoop-env.sh Modified
# export HDFS_DFSROUTER_OPTS=""
###

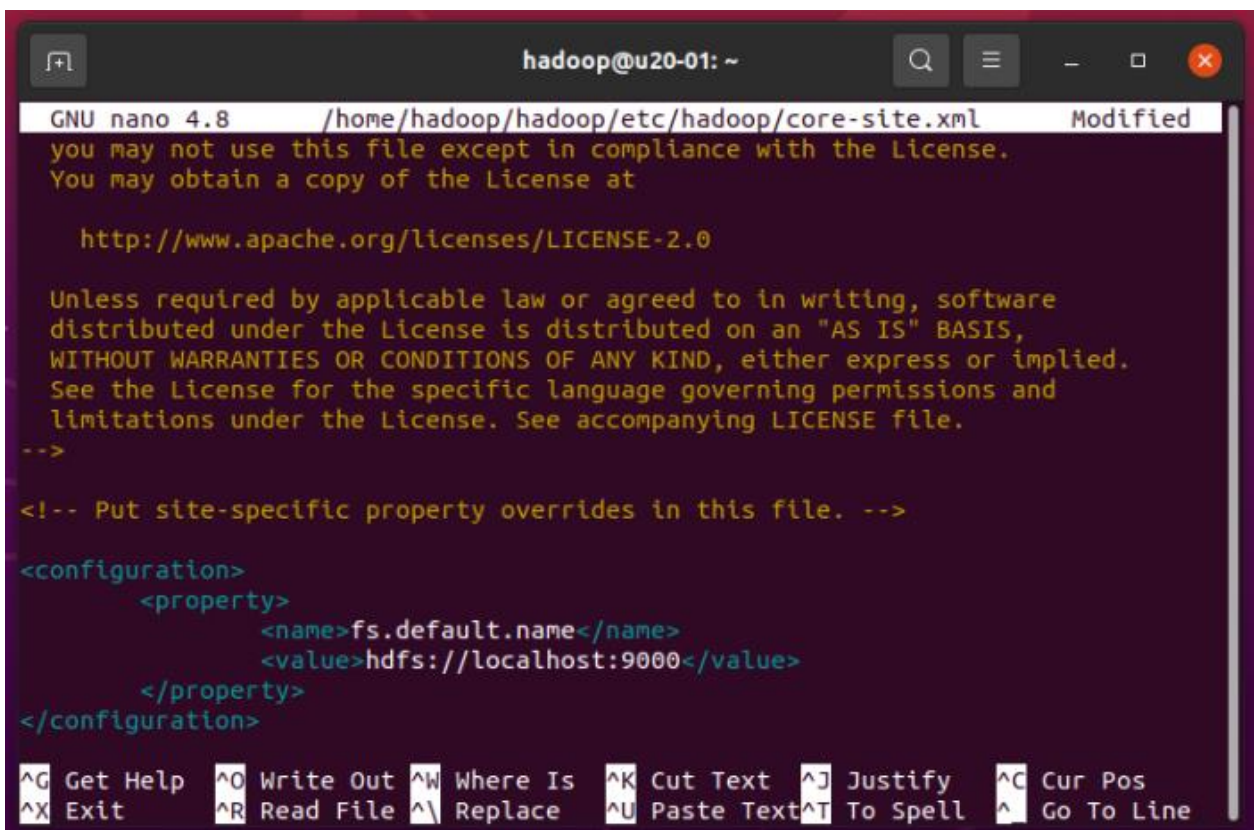
###
# Advanced Users Only!
###

#
# When building Hadoop, one can add the class paths to the commands
# via this special env var:
# export HADOOP_ENABLE_BUILD_PATHS="true"

#
# To prevent accidents, shell commands be (superficially) locked
# to only allow certain users to execute certain subcommands.
# It uses the format of (command)_(subcommand)_USER.
#
# For example, to limit who can execute the namenode command,
# export HDFS_NAMENODE_USER=hdfs
export JAVA_HOME=/usr/lib/jvm/java-8-openjdk-amd64

^G Get Help  ^O Write Out  ^W Where Is  ^K Cut Text  ^J Justify   ^C Cur Pos
^X Exit      ^R Read File  ^\ Replace   ^U Paste Text ^T To Spell  ^_ Go To Line
```

16. Установим переменные CORE:



```
GNU nano 4.8 /home/hadoop/hadoop/etc/hadoop/core-site.xml Modified
you may not use this file except in compliance with the License.
You may obtain a copy of the License at

http://www.apache.org/licenses/LICENSE-2.0

Unless required by applicable law or agreed to in writing, software
distributed under the License is distributed on an "AS IS" BASIS,
WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied.
See the License for the specific language governing permissions and
limitations under the License. See accompanying LICENSE file.
-->

<!-- Put site-specific property overrides in this file. -->

<configuration>
  <property>
    <name>fs.default.name</name>
    <value>hdfs://localhost:9000</value>
  </property>
</configuration>

^G Get Help  ^O Write Out  ^W Where Is  ^K Cut Text  ^J Justify   ^C Cur Pos
^X Exit      ^R Read File  ^\ Replace   ^U Paste Text ^T To Spell  ^_ Go To Line
```

17. Установим переменные HDFS:

```
hadoop@u20-01: ~  
GNU nano 4.8 /home/hadoop/hadoop/etc/hadoop/hdfs-site.xml Modified  
  
<!-- Put site-specific property overrides in this file. -->  
  
<configuration>  
  <property>  
    <name>dfs.replication</name>  
    <value>1</value>  
  </property>  
  
  <property>  
    <name>dfs.name.dir</name>  
    <value>file:///home/hadoop/hadoopdata/hdfs/namenode</value>  
  </property>  
  
  <property>  
    <name>dfs.data.dir</name>  
    <value>file:///home/hadoop/hadoopdata/hdfs/datanode</value>  
  </property>  
</configuration>  
  
^G Get Help ^O Write Out ^W Where Is ^K Cut Text ^J Justify ^C Cur Pos  
^X Exit ^R Read File ^\ Replace ^U Paste Text ^T To Spell ^_ Go To Line
```

18. Установим переменные MapReduce:

```
hadoop@u20-01: ~  
GNU nano 4.8 /home/hadoop/hadoop/etc/hadoop/mapred-site.xml Modified  
  
<!-- Put site-specific property overrides in this file. -->  
  
<configuration>  
  <property>  
    <name>mapreduce.framework.name</name>  
    <value>yarn</value>  
  </property>  
  <property>  
    <name>yarn.app.mapreduce.am.env</name>  
    <value>HADOOP_MAPRED_HOME=${HADOOP_HOME}</value>  
  </property>  
  <property>  
    <name>mapreduce.map.env</name>  
    <value>HADOOP_MAPRED_HOME=${HADOOP_HOME}</value>  
  </property>  
  <property>  
    <name>mapreduce.reduce.env</name>  
    <value>HADOOP_MAPRED_HOME=${HADOOP_HOME}</value>  
  </property>  
</configuration>  
  
^G Get Help ^O Write Out ^W Where Is ^K Cut Text ^J Justify ^C Cur Pos  
^X Exit ^R Read File ^\ Replace ^U Paste Text ^T To Spell ^_ Go To Line
```

19. Установить переменные YARN:



```
hadoop@u20-01: ~  
GNU nano 4.8 /home/hadoop/hadoop/etc/hadoop/yarn-site.xml Modified  
  
http://www.apache.org/licenses/LICENSE-2.0  
  
Unless required by applicable law or agreed to in writing, software  
distributed under the License is distributed on an "AS IS" BASIS,  
WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied.  
See the License for the specific language governing permissions and  
limitations under the License. See accompanying LICENSE file.  
-->  
<configuration>  
  <property>  
    <name>yarn.nodemanager.aux-services</name>  
    <value>mapreduce_shuffle</value>  
  </property>  
  <property>  
    <name>yarn.nodemanager.resource.memory-mb</name>  
    <value>16384</value>  
  </property>  
</configuration>  
  
^G Get Help ^O Write Out ^W Where Is ^K Cut Text ^J Justify ^C Cur Pos  
^X Exit ^R Read File ^\ Replace ^U Paste Text ^T To Spell ^_ Go To Line
```

20.Очистим HDFS:

```
hadoop@u20-01:~$ hdfs namenode -format  
WARNING: /home/hadoop/hadoop/logs does not exist. Creating.  
2022-11-01 15:52:09,760 INFO namenode.NameNode: STARTUP_MSG:  
/*****  
STARTUP_MSG: Starting NameNode  
STARTUP_MSG: host = u20-01/172.26.35.121  
STARTUP_MSG: args = [-format]  
STARTUP_MSG: version = 3.1.2  
STARTUP_MSG: classpath = /home/hadoop/hadoop/etc/hadoop:/home/hadoop/hadoop/sh  
are/hadoop/common/lib/netty-3.10.5.Final.jar:/home/hadoop/hadoop/share/hadoop/co  
mmon/lib/log4j-1.2.17.jar:/home/hadoop/hadoop/share/hadoop/common/lib/kerb-core-  
1.0.1.jar:/home/hadoop/hadoop/share/hadoop/common/lib/commons-logging-1.1.3.jar:  
/home/hadoop/hadoop/share/hadoop/common/lib/jackson-core-asl-1.9.13.jar:/home/ha  
dooop/hadoop/share/hadoop/common/lib/audience-annotations-0.5.0.jar:/home/hadoo  
p/hadoop/share/hadoop/common/lib/jersey-server-1.19.jar:/home/hadoop/hadoop/sha  
re/hadoop/common/lib/woodstox-core-5.0.3.jar:/home/hadoop/hadoop/share/hadoop/commo  
n/lib/kerb-crypto-1.0.1.jar:/home/hadoop/hadoop/share/hadoop/common/lib/kerb-uti  
l-1.0.1.jar:/home/hadoop/hadoop/share/hadoop/common/lib/guava-11.0.2.jar:/home/h  
adoop/hadoop/share/hadoop/common/lib/kerb-identity-1.0.1.jar:/home/hadoop/hadoo  
p/share/hadoop/common/lib/paranamer-2.3.jar:/home/hadoop/hadoop/share/hadoop/comm  
on/lib/jsr305-3.0.0.jar:/home/hadoop/hadoop/share/hadoop/common/lib/jetty-server  
-9.3.24.v20180605.jar:/home/hadoop/hadoop/share/hadoop/common/lib/kerb-server-1.  
0.1.jar:/home/hadoop/hadoop/share/hadoop/common/lib/commons-beanutils-1.9.3.jar:
```

21.Запустим HDFS:

```
hadoop@u20-01:~$ start-dfs.sh
Starting namenodes on [localhost]
Starting datanodes
Starting secondary namenodes [u20-01]
u20-01: Warning: Permanently added 'u20-01,172.26.35.121' (ECDSA) to the list of
known hosts.
2022-11-01 15:55:19,128 WARN util.NativeCodeLoader: Unable to load native-hadoop
library for your platform... using builtin-java classes where applicable
hadoop@u20-01:~$
```

22. Запустим YARN:

```
hadoop@u20-01:~$ start-yarn.sh
Starting resourcemanager
Starting nodemanagers
```

23. Запустим Admin Status Report:

```
hadoop@u20-01:~$ hdfs dfsadmin -report
2022-11-01 16:00:27,719 WARN util.NativeCodeLoader: Unable to load native-hadoop
library for your platform... using builtin-java classes where applicable
Configured Capacity: 51989970944 (48.42 GB)
Present Capacity: 35232309248 (32.81 GB)
DFS Remaining: 35232284672 (32.81 GB)
DFS Used: 24576 (24 KB)
DFS Used%: 0.00%
Replicated Blocks:
    Under replicated blocks: 0
    Blocks with corrupt replicas: 0
    Missing blocks: 0
    Missing blocks (with replication factor 1): 0
    Low redundancy blocks with highest priority to recover: 0
    Pending deletion blocks: 0
Erasure Coded Block Groups:
    Low redundancy block groups: 0
    Block groups with corrupt internal blocks: 0
    Missing block groups: 0
    Low redundancy blocks with highest priority to recover: 0
    Pending deletion blocks: 0
```

```
-----
Live datanodes (1):

Name: 127.0.0.1:9866 (localhost)
Hostname: u20-01
Decommission Status : Normal
Configured Capacity: 51989970944 (48.42 GB)
DFS Used: 24576 (24 KB)
Non DFS Used: 14083530752 (13.12 GB)
DFS Remaining: 35232284672 (32.81 GB)
DFS Used%: 0.00%
DFS Remaining%: 67.77%
Configured Cache Capacity: 0 (0 B)
Cache Used: 0 (0 B)
Cache Remaining: 0 (0 B)
Cache Used%: 100.00%
Cache Remaining%: 0.00%
Xceivers: 1
Last contact: Tue Nov 01 16:00:26 MSK 2022
Last Block Report: Tue Nov 01 15:55:14 MSK 2022
Num of Blocks: 0
```

24. Проверим страницу Ressource Manager:

The screenshot displays the Hadoop All Applications page. The left sidebar contains navigation links: Cluster, About, Nodes, Node Labels, Applications, NEW, NEW SAVING, SUBMITTED, ACCEPTED, RUNNING, FINISHED, FAILED, KILLED, and Scheduler. The main content area shows the following metrics:

Cluster Metrics						
Apps Submitted	Apps Pending	Apps Running	Apps Completed	Containers Running	Memory Used	Memory Total
0	0	0	0	0	0 B	16 GB

Cluster Nodes Metrics				
Active Nodes	Decommissioning Nodes	Decommissioned Nodes	Lost Nodes	
1	0	0	0	0

Scheduler Metrics			
Scheduler Type	Scheduling Resource Type	Minimum Allocation	Maximum Allocation
Capacity Scheduler	[memory-mb (unit=Mi), vcores]	<memory:1024, vCores:1>	<memory:8192, vCores:1>

Show 20 entries

ID	User	Name	Application Type	Queue	Application Priority	StartTime	FinishTime	State	FinalStatus	Running Containers	Allocated CPU VCores	Allocated Memory MB
No data available in table												

Showing 0 to 0 of 0 entries

25. Проверим страницы NameNode и Status Page:



Namenode information x +

localhost:9870/dfshealth.html#tab-overview

Hadoop Overview Datanodes Datanode Volume Failures Snapshot Startup Progress Utilities +

## Overview 'localhost:9000' (active)

Started:	Tue Nov 01 15:55:06 +0300 2022
Version:	3.1.2, r1019dde65bcf12e05ef48ac71e84550d589e5d9a
Compiled:	Tue Jan 29 04:39:00 +0300 2019 by sunilg from branch-3.1.2
Cluster ID:	CID-6b9ab22b-a9ba-41e0-aea9-bb13239ee217
Block Pool ID:	BP-1075887091-172.26.35.121-1667307131247

## Summary

Security is off.  
 Safemode is off.  
 1 files and directories, 0 blocks (0 replicated blocks, 0 erasure coded block groups) = 1 total filesystem object(s).  
 Heap Memory used 63.66 MB of 119.94 MB Heap Memory. Max Heap Memory is 1.86 GB.  
 Non Heap Memory used 50.47 MB of 51.65 MB Committed Non Heap Memory. Max Non Heap Memory is <unbounded>.

Configured Capacity:	48.42 GB
----------------------	----------

Namenode information x +

localhost:9870/dfshealth.html#tab-datanode

Hadoop Overview Datanodes Datanode Volume Failures Snapshot Startup Progress Utilities +

## Datanode Information

✔ In service 
 ❌ Down 
 ⚠ Decommissioned 
 ⚠ Decommissioned & dead 
 🔧 In Maintenance 
 🔧 In Maintenance & dead

### Datanode usage histogram

Disk usage of each DataNode (%)

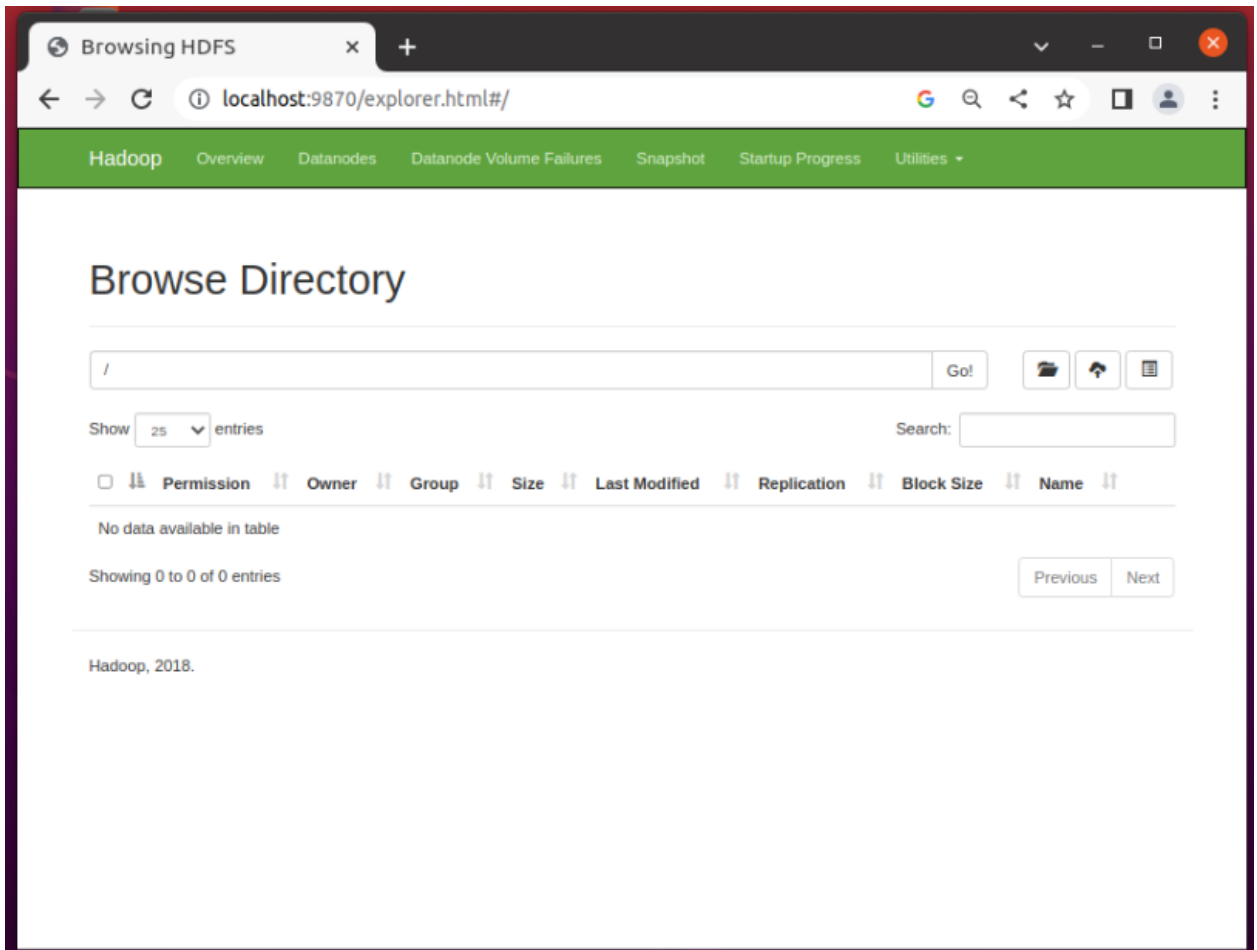
### In operation

Show  entries

Search:

Node	Http Address	Last contact	Last Block Report	Capacity	Blocks	Block pool used	Version
------	--------------	--------------	-------------------	----------	--------	-----------------	---------

26. Проверим HDFS File Browser:



27. Создадим директорию пользователя:

```
hadoop@u20-01:~$ hadoop fs -mkdir /user
2022-11-01 16:26:21,869 WARN util.NativeCodeLoader: Unable to load native-hadoop
library for your platform... using builtin-java classes where applicable
hadoop@u20-01:~$ hadoop fs -mkdir /user/hadoop
2022-11-01 16:26:38,269 WARN util.NativeCodeLoader: Unable to load native-hadoop
library for your platform... using builtin-java classes where applicable
hadoop@u20-01:~$
```

28. Выведем список директорий:

```
hadoop@u20-01:~$ hadoop fs -ls /
2022-11-01 16:28:36,054 WARN util.NativeCodeLoader: Unable to load native-hadoop
library for your platform... using builtin-java classes where applicable
Found 1 items
drwxr-xr-x - hadoop supergroup 0 2022-11-01 16:26 /user
```

29. Скопируем файл лога из локальной директории в HDFS:

```
hadoop@u20-01:~$ hadoop fs -put /var/log/dpkg.log /user/hadoop/dpkg.log
2022-11-01 16:32:56,948 WARN util.NativeCodeLoader: Unable to load native-hadoop
library for your platform... using builtin-java classes where applicable
```

30. Использование MapReduce WordCount Jar, предоставляемого Hadoop, для подсчета слов в файле:

```
hadoop@u20-01: ~  
hadoop@u20-01:~$ hadoop jar hadoop/share/hadoop/mapreduce/hadoop-mapreduce-examples-3.1.2.jar wordcount /user/hadoop/dpkg.log /user/hadoop/test_output  
2022-11-01 16:35:18,510 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable  
  
2022-11-01 16:35:19,478 INFO client.RMProxy: Connecting to ResourceManager at /0.0.0.0:8032  
2022-11-01 16:35:20,044 INFO mapreduce.JobResourceUploader: Disabling Erasure Code for path: /tmp/hadoop-yarn/staging/hadoop/.staging/job_1667307482261_0001  
2022-11-01 16:35:20,401 INFO input.FileInputFormat: Total input files to process : 1  
2022-11-01 16:35:20,542 INFO mapreduce.JobSubmitter: number of splits:1  
2022-11-01 16:35:20,776 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1667307482261_0001  
2022-11-01 16:35:20,777 INFO mapreduce.JobSubmitter: Executing with tokens: []  
2022-11-01 16:35:21,097 INFO conf.Configuration: resource-types.xml not found  
2022-11-01 16:35:21,098 INFO resource.ResourceUtils: Unable to find 'resource-types.xml'.  
2022-11-01 16:35:21,492 INFO impl.YarnClientImpl: Submitted application application_1667307482261_0001  
2022-11-01 16:35:21,570 INFO mapreduce.Job: The url to track the job: http://u20-01:8088/proxy/application_1667307482261_0001/  
2022-11-01 16:35:21,571 INFO mapreduce.Job: Running job: job_1667307482261_0001
```

31.Просмотр Запущенного MapReduce Job:

The screenshot shows the Hadoop web interface in a browser window. The address bar indicates the URL is `localhost:8088/cluster`. The page features the Hadoop logo and a sidebar with navigation links: **Cluster** (About, Nodes, Node Labels, Applications), **NEW**, **NEW SAVING**, **SUBMITTED**, **ACCEPTED**, **RUNNING**, **FINISHED**, **FAILED**, **KILLED**, and **Scheduler**. The main content area displays various metrics:

- Cluster Metrics:** A table showing 1 App Submitted, 0 Apps Pending, 0 Apps Running, 1 Apps Completed, 0 Containers Running, and 0 MB Memory.
- Cluster Nodes Metrics:** A table showing 1 Active Node, 0 Decommissioning Nodes, and 0 Decommissioned Nodes.
- Scheduler Metrics:** A table showing Capacity Scheduler, Scheduling Resource Type [memory-mb (unit=Mi), vcores], and Minimum Allocation <memory:1024, vCores:1>.
- Job List:** A table showing 20 entries. The first entry is highlighted: application\_1667307482261\_0001, user hadoop, name word count, Application Type MAPREDUCE, Queue default, Priority 0, Start Time Tue Nov 1 16:35:21 +0300 2022, Finish Time Tue Nov 1 16:35:42 +0300 2022, and State FINISHED.

At the bottom, it says "Showing 1 to 1 of 1 entries".

32.Проверить результат На Output/Result (via Bash):



```

hadoop@u20-01:~$ hadoop fs -cat /user/hadoop/test_output/part-r-00000
2022-11-01 16:40:09,178 WARN util.NativeCodeLoader: Unable to load native-hadoop
library for your platform... using builtin-java classes where applicable
0.0.2-ubuntu1 7
0.104-0ubuntu2~20.04.1 8
0.104-0ubuntu2~20.04.2 12
Code ubuntu6.7 4
0.17-2 12
0.17.0-ubuntu-focal 4
0.21.0-ubuntu-focal 6
0.24-1ubuntu3 12
0.3.6-2build1~20.04.1 7
0.36-6ubuntu1 4
0.37.1-1 14
0.4-1 7
0.7.9-1ubuntu2 12
0.7.9-1ubuntu3.20.04.2 18
0.72 7
0.9.12-1 7
0.9.4git20200323-0ubuntu6.2 4
02:47:27 9

```

33. Проверить результат На Output/Result (via Web HDFS File Browser):

The screenshot shows a web browser window titled "Browsing HDFS" with the address bar displaying "localhost:9870/explorer.html#/" and a green navigation bar with links: Hadoop, Overview, Datanodes, Datanode Volume Failures, Snapshot, Startup Progress, and Utilities. The main content area is titled "Browse Directory" and shows a directory listing for the root path "/".

Directory listing details:

- Search bar: /
- Show: 25 entries
- Search:
- Table headers: Permission, Owner, Group, Size, Last Modified, Replication, Block Size, Name
- Table entries:
 

Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name
drwx-----	hadoop	supergroup	0 B	Nov 01 16:35	0	0 B	tmp
drwxr-xr-x	hadoop	supergroup	0 B	Nov 01 16:26	0	0 B	user
- Showing 1 to 2 of 2 entries
- Navigation: Previous, 1, Next
- Footer: Hadoop, 2018.

34. Клонировать репозиторий git (чтобы получить пример данных):

```

hadoop@u20-01:~$ git clone https://github.com/BosenkoTM/ds_practice.git
Cloning into 'ds_practice'...
remote: Enumerating objects: 76, done.
remote: Total 76 (delta 0), reused 0 (delta 0), pack-reused 76
Unpacking objects: 100% (76/76), 6.24 MiB | 9.65 MiB/s, done.

```

35.Скопировать образец файла (/ds\_practice/exercises/winter\_semester\_2021-2022/01\_hadoop/sample\_data/Faust\_1.txt) из репозитория Git в HDFS.

```
hadoop@u20-01:~$ hadoop fs -put ds_practice/exercises/winter_semester_2021-2022/05_hadoop/sample_data/Faust_1.txt
2022-11-07 22:16:57,851 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
2022-11-07 22:16:59,105 WARN hdfs.DataStreamer: Caught exception
java.lang.InterruptedException
    at java.lang.Object.wait(Native Method)
    at java.lang.Thread.join(Thread.java:1257)
    at java.lang.Thread.join(Thread.java:1331)
    at org.apache.hadoop.hdfs.DataStreamer.closeResponder(DataStreamer.java:986)
    at org.apache.hadoop.hdfs.DataStreamer.endBlock(DataStreamer.java:640)
    at org.apache.hadoop.hdfs.DataStreamer.run(DataStreamer.java:810)
```

36.Запустить MapReduce Jar по умолчанию (hadoop/share/hadoop/mapreduce/hadoop-mapreduce-examples-3.1.2.jar) для вычисления количества слов в текстовом файле

```
hadoop@u20-01:~$ hadoop jar hadoop/share/hadoop/mapreduce/hadoop-mapreduce-examples-3.1.2.jar wordcount /user/hadoop/Faust_1.txt /user/hadoop/Faust_1_Output
2022-11-07 22:25:41,601 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
2022-11-07 22:25:42,480 INFO client.RMProxy: Connecting to ResourceManager at /0.0.0.0:8032
2022-11-07 22:25:42,952 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/hadoop/.staging/job_1667307482261_0002
2022-11-07 22:25:43,209 WARN hdfs.DataStreamer: Caught exception
java.lang.InterruptedException
    at java.lang.Object.wait(Native Method)
    at java.lang.Thread.join(Thread.java:1257)
    at java.lang.Thread.join(Thread.java:1331)
    at org.apache.hadoop.hdfs.DataStreamer.closeResponder(DataStreamer.java:986)
    at org.apache.hadoop.hdfs.DataStreamer.endBlock(DataStreamer.java:640)
    at org.apache.hadoop.hdfs.DataStreamer.run(DataStreamer.java:810)
2022-11-07 22:25:43,294 INFO input.FileInputFormat: Total input files to process
```

37.Проверка в Диспетчере ресурсов выполнение задания:

The screenshot shows the Hadoop web interface at localhost:8088/cluster. The left sidebar contains navigation links: Cluster (selected), About, Nodes, Node Labels, Applications, NEW, NEW SAVING, SUBMITTED, ACCEPTED, RUNNING, FINISHED, FAILED, KILLED, Scheduler, and Tools. The main content area displays various metrics and a table of applications.

**Cluster Metrics**

Apps Submitted	Apps Pending	Apps Running	Apps Completed	Containers Running	Memory
2	0	0	2	0	0 B

**Cluster Nodes Metrics**

Active Nodes	Decommissioning Nodes	Decommissioned Nodes
1	0	0

**Scheduler Metrics**

Scheduler Type	Scheduling Resource Type	Minimum Allocation
Capacity Scheduler	[memory-mb (unit=Mi), vcores]	<memory:1024, vCores:1>

Show 20 entries

ID	User	Name	Application Type	Queue	Application Priority	StartTime	FinishTime	State
application_1667307482261_0002	hadoop	word count	MAPREDUCE	default	0	Mon Nov 7 22:25:43 +0300 2022	Mon Nov 7 22:26:02 +0300 2022	FINISHED
application_1667307482261_0001	hadoop	word count	MAPREDUCE	default	0	Tue Nov 1 16:35:21 +0300 2022	Tue Nov 1 16:35:42 +0300 2022	FINISHED

Showing 1 to 2 of 2 entries

38. Копирование полученного файла MapReduce обратно в локальную файловую систему ubuntu (using bash):

```
hadoop@u20-01:~$ hadoop fs -get /user/hadoop/Faust_1_Output/part-r-00000 Faust_1_Output.csv
2022-11-07 22:39:44,316 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
```

```
hadoop@u20-01:~$ head -10 Faust_1_Output.csv
"Allein, 1
"Alles 1
"Als 1
"Der 1
"Die 2
"Er 2
"Ich 4
"Im 1
"Mein 1
"Nur 1
```

39. Используйте и запустите по умолчанию MapReduce Jar (hadoop/share/hadoop/mapreduce/hadoop-mapreduce-examples-3.1 .2.jar), чтобы получить количество вхождений точной строки 'Faust' в текстовом файле.



```

hadoop@u20-01:~$ hadoop jar hadoop/share/hadoop/mapreduce/hadoop-mapreduce-examples-3.1.2.jar grep /user/hadoop/Faust_1.txt /user/hadoop/Faust_1_Count_Output 'Faust'
2022-11-07 22:49:35,519 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
2022-11-07 22:49:36,407 INFO client.RMProxy: Connecting to ResourceManager at /0.0.0.0:8032
2022-11-07 22:49:36,953 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/hadoop/.staging/job_1667307482261_0003
2022-11-07 22:49:37,262 INFO input.FileInputFormat: Total input files to process : 1
2022-11-07 22:49:37,392 INFO mapreduce.JobSubmitter: number of splits:1
2022-11-07 22:49:37,606 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1667307482261_0003
2022-11-07 22:49:37,607 INFO mapreduce.JobSubmitter: Executing with tokens: []
2022-11-07 22:49:37,885 INFO conf.Configuration: resource-types.xml not found
2022-11-07 22:49:37,885 INFO resource.ResourceUtils: Unable to find 'resource-types.xml'.
2022-11-07 22:49:38,001 INFO impl.YarnClientImpl: Submitted application application_1667307482261_0003
2022-11-07 22:49:38,077 INFO mapreduce.Job: The url to track the job: http://u20-01:8088/proxy/application_1667307482261_0003/
2022-11-07 22:49:38,077 INFO mapreduce.Job: Running job: job_1667307482261_0003
2022-11-07 22:49:46,271 INFO mapreduce.Job: Job job_1667307482261_0003 running in uber mode : false
2022-11-07 22:49:46,275 INFO mapreduce.Job: map 0% reduce 0%
2022-11-07 22:49:52,367 INFO mapreduce.Job: map 100% reduce 0%
2022-11-07 22:49:57,408 INFO mapreduce.Job: map 100% reduce 100%
2022-11-07 22:49:58,443 INFO mapreduce.Job: Job job_1667307482261_0003 completed successfully
2022-11-07 22:49:58,588 INFO mapreduce.Job: Counters: 53
  File System Counters
    FILE: Number of bytes read=22
    FILE: Number of bytes written=432881
    FILE: Number of read operations=0
    FILE: Number of large read operations=0
    FILE: Number of write operations=0
    HDFS: Number of bytes read=188047
    HDFS: Number of bytes written=108
    HDFS: Number of read operations=8
    HDFS: Number of large read operations=0
    HDFS: Number of write operations=2

```

40. Скопируйте результат MapReduce job обратно в локальную файловую систему ubuntu.

```

hadoop@u20-01:~$ hadoop fs -get /user/hadoop/Faust_1_Count_Output/part-r-00000 Faust_1_Count_Output.csv
2022-11-07 22:55:32,910 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
hadoop@u20-01:~$ cat Faust_1_Count_Output.csv
50      Faust

```

Самостоятельное задание

Текст «War and Peace»

```

hadoop@u20-01:~$ git clone https://github.com/BrinkovanVA/practicds.git
Cloning into 'practicds'...
remote: Enumerating objects: 56, done.
remote: Counting objects: 100% (56/56), done.
remote: Compressing objects: 100% (52/52), done.
remote: Total 56 (delta 16), reused 0 (delta 0), pack-reused 0
Unpacking objects: 100% (56/56), 9.31 MiB | 3.33 MiB/s, done.

```

- 1.
2. Копирование файла из репозитория GIT в HDFS каталог пользователя:

```

hadoop@u20-01:~$ hadoop fs -put practicds/WarAndPeace.txt /user/hadoop/WarAndPeace.txt
2022-11-07 23:18:03,509 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable

```

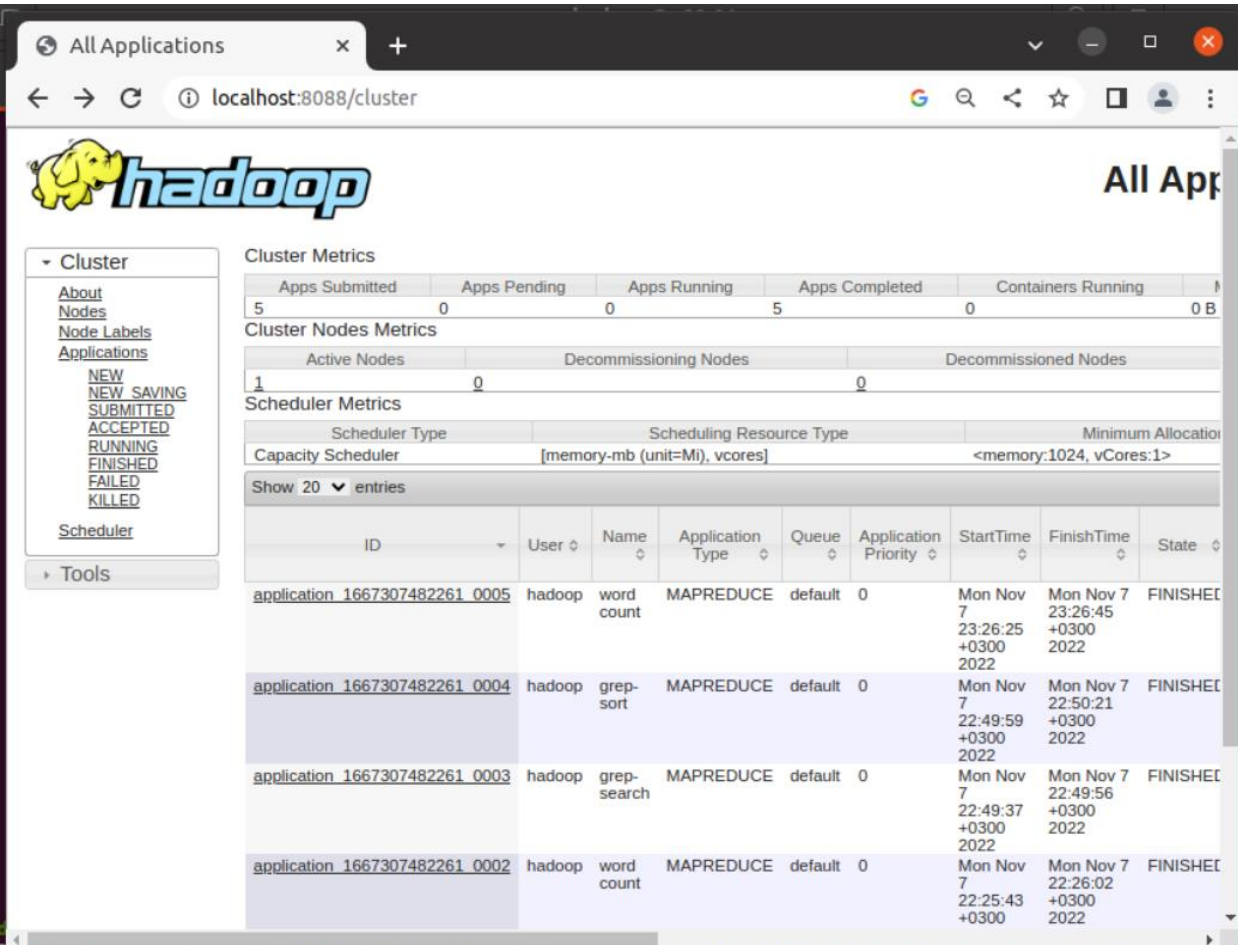
3. Запуск MapReduce Jar по умолчанию (hadoop-mapreduceexamples-3.1.2.jar ) для вычисления количества слов для текстового файла «WarAndPeace.txt»:

```

hadoop@u20-01:~$ hadoop jar hadoop/share/hadoop/mapreduce/hadoop-mapreduce-examples-3.1.2.jar wordcount
/user/hadoop/WarAndPeace.txt /user/hadoop/WarAndPeace_Output
2022-11-07 23:26:23,514 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platfo
rm... using builtin-java classes where applicable
2022-11-07 23:26:24,372 INFO client.RMProxy: Connecting to ResourceManager at /0.0.0.0:8032
2022-11-07 23:26:24,857 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding for path: /tmp/hado
op-yarn/staging/hadoop/.staging/job_1667307482261_0005
2022-11-07 23:26:25,086 WARN hdfs.DataStreamer: Caught exception
java.lang.InterruptedException
    at java.lang.Object.wait(Native Method)
    at java.lang.Thread.join(Thread.java:1257)
    at java.lang.Thread.join(Thread.java:1331)
    at org.apache.hadoop.hdfs.DataStreamer.closeResponder(DataStreamer.java:986)
    at org.apache.hadoop.hdfs.DataStreamer.endBlock(DataStreamer.java:640)
    at org.apache.hadoop.hdfs.DataStreamer.run(DataStreamer.java:810)
2022-11-07 23:26:25,156 INFO input.FileInputFormat: Total input files to process : 1
2022-11-07 23:26:25,261 WARN hdfs.DataStreamer: Caught exception
java.lang.InterruptedException
    at java.lang.Object.wait(Native Method)
    at java.lang.Thread.join(Thread.java:1257)
    at java.lang.Thread.join(Thread.java:1331)
    at org.apache.hadoop.hdfs.DataStreamer.closeResponder(DataStreamer.java:986)
    at org.apache.hadoop.hdfs.DataStreamer.endBlock(DataStreamer.java:640)
    at org.apache.hadoop.hdfs.DataStreamer.run(DataStreamer.java:810)
2022-11-07 23:26:25,262 INFO mapreduce.JobSubmitter: number of splits:1
2022-11-07 23:26:25,540 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1667307482261_0005
2022-11-07 23:26:25,541 INFO mapreduce.JobSubmitter: Executing with tokens: []
2022-11-07 23:26:25,871 INFO conf.Configuration: resource-types.xml not found
2022-11-07 23:26:25,871 INFO resource.ResourceUtils: Unable to find 'resource-types.xml'.
2022-11-07 23:26:26,081 INFO impl.YarnClientImpl: Submitted application application_1667307482261_0005
2022-11-07 23:26:26,271 INFO mapreduce.Job: The url to track the job: http://u20-01:8088/proxy/applicati
on_1667307482261_0005/
2022-11-07 23:26:26,277 INFO mapreduce.Job: Running job: job_1667307482261_0005

```

4. Проверка в Диспетчере ресурсов выполнение задания:



The screenshot shows the Hadoop web interface at `localhost:8088/cluster`. The 'All Applications' page is displayed, showing a list of applications in the 'FINISHED' state. The table below summarizes the data shown in the screenshot.

ID	User	Name	Application Type	Queue	Application Priority	StartTime	FinishTime	State
application_1667307482261_0005	hadoop	word count	MAPREDUCE	default	0	Mon Nov 7 23:26:25 +0300 2022	Mon Nov 7 23:26:45 +0300 2022	FINISHED
application_1667307482261_0004	hadoop	grep-sort	MAPREDUCE	default	0	Mon Nov 7 22:49:59 +0300 2022	Mon Nov 7 22:50:21 +0300 2022	FINISHED
application_1667307482261_0003	hadoop	grep-search	MAPREDUCE	default	0	Mon Nov 7 22:49:37 +0300 2022	Mon Nov 7 22:49:56 +0300 2022	FINISHED
application_1667307482261_0002	hadoop	word count	MAPREDUCE	default	0	Mon Nov 7 22:25:43 +0300 2022	Mon Nov 7 22:26:02 +0300 2022	FINISHED

5. Копирование полученного файла MapReduce обратно в локальную файловую систему ubuntu (using bash):



```
hadoop@u20-01:~$ hadoop fs -get /user/hadoop/WarAndPeace_Output/part-r-000000 WarAndPeace_Output.csv
2022-11-07 23:32:30,292 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platfo
rm... using builtin-java classes where applicable
hadoop@u20-01:~$ head -10 WarAndPeace_Output.csv
192; 5
193;NNA 2
201;cole 2
201;douard 1
201;lie, 1
201;tablissement 1
211;SIP 1
224; 257
224; Vous 1
224; 1
```

6. Запуск MapReduce Jar по умолчанию (hadoop-mapreduce-examples-3.1.2.jar) для поиска в грег строки «Peace» в текстовом файле «m.txt» и подсчета количества повторений данного слова:

```
hadoop@u20-01:~$ hadoop jar hadoop/share/hadoop/mapreduce/hadoop-mapreduce-examples-3.1.2.jar grep /user
/hadoop/WarAndPeace.txt /user/hadoop/WarAndPeace_Count_Output 'war'
2022-11-07 23:40:35,078 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platfo
rm... using builtin-java classes where applicable
2022-11-07 23:40:35,979 INFO client.RMProxy: Connecting to ResourceManager at /0.0.0.0:8032
2022-11-07 23:40:36,508 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding for path: /tmp/hado
op-yarn/staging/hadoop/.staging/job_1667307482261_0006
2022-11-07 23:40:36,816 INFO input.FileInputFormat: Total input files to process : 1
2022-11-07 23:40:36,958 INFO mapreduce.JobSubmitter: number of splits:1
2022-11-07 23:40:37,196 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1667307482261_0006
2022-11-07 23:40:37,197 INFO mapreduce.JobSubmitter: Executing with tokens: []
2022-11-07 23:40:37,477 INFO conf.Configuration: resource-types.xml not found
2022-11-07 23:40:37,478 INFO resource.ResourceUtils: Unable to find 'resource-types.xml'.
2022-11-07 23:40:37,597 INFO impl.YarnClientImpl: Submitted application application_1667307482261_0006
2022-11-07 23:40:37,689 INFO mapreduce.Job: The url to track the job: http://u20-01:8088/proxy/applicati
on_1667307482261_0006/
2022-11-07 23:40:37,689 INFO mapreduce.Job: Running job: job_1667307482261_0006
2022-11-07 23:40:45,893 INFO mapreduce.Job: Job job_1667307482261_0006 running in uber mode : false
2022-11-07 23:40:45,895 INFO mapreduce.Job: map 0% reduce 0%
2022-11-07 23:40:51,997 INFO mapreduce.Job: map 100% reduce 0%
2022-11-07 23:40:57,042 INFO mapreduce.Job: map 100% reduce 100%
2022-11-07 23:40:58,074 INFO mapreduce.Job: Job job_1667307482261_0006 completed successfully
2022-11-07 23:40:58,239 INFO mapreduce.Job: Counters: 53
File System Counters
FILE: Number of bytes read=20
FILE: Number of bytes written=432879
FILE: Number of read operations=0
FILE: Number of large read operations=0
FILE: Number of write operations=0
HDFS: Number of bytes read=3586560
HDFS: Number of bytes written=106
HDFS: Number of read operations=8
HDFS: Number of large read operations=0
HDFS: Number of write operations=2
```

7. Копирование полученного файла MapReduce обратно в локальную файловую систему ubuntu (using bash):

```
hadoop@u20-01:~$ hadoop fs -get /user/hadoop/WarAndPeace_Count_Output/part-r-000000 WarAndPeace_Count_Out
put.csv
2022-11-07 23:43:36,438 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platfo
rm... using builtin-java classes where applicable
hadoop@u20-01:~$ cat WarAndPeace_Count_Output.csv
1335 war
```

## Индивидуальное задание

1. Развернуть виртуальное окружение:



```

lemp001@u20-01:~$ sudo su hadoop
[sudo] password for lemp001:
hadoop@u20-01:/home/lemp001$ cd
hadoop@u20-01:~$ jps
145142 Jps
14005 NodeManager
13672 SecondaryNameNode
13865 ResourceManager
13290 NameNode
13451 DataNode

```

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

```

hadoop@u20-01:~$ hdfs dfs -help

```

```

at <path> with current time as the timestamp of that <path>.
-a Change only the access time
-m Change only the modification time
-t TIMESTAMP Use specified timestamp (in format yyyyMMddHHmmss) instead of
current time
-c Do not create any files

touchz <path> ... :
Creates a file of zero length at <path> with current time as the timestamp of
that <path>. An error is returned if the file exists with non-zero length

truncate [-w] <length> <path> ... :
Truncate all files that match the specified file pattern to the specified
length.

-w Requests that the command wait for block recovery to complete, if
necessary.

usage [cmd ...] :
Displays the usage for given command or all commands if none is specified.

generic options supported are:
conf <configuration file>          specify an application configuration file
D <property=value>                  define a value for a given property

```

3. Просмотреть корневую директорию HDFS:

```

hadoop@u20-01:~$ hdfs dfs -ls / hdfs://localhost:9000/user
2022-11-14 16:56:01,953 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platfo
rm... using builtin-java classes where applicable
Found 2 items
drwx----- - hadoop supergroup          0 2022-11-01 16:35 /tmp
drwxr-xr-x - hadoop supergroup          0 2022-11-01 16:26 /user
Found 1 items
drwxr-xr-x - hadoop supergroup          0 2022-11-14 16:33 hdfs://localhost:9000/user/hadoop

```

4. Создать в HDFS в директории /user/mgpu поддиректорию /Brinkovan:

```

hadoop@u20-01:~$ hadoop fs -mkdir /user/mgpu
2022-11-14 16:56:45,415 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platfo
rm... using builtin-java classes where applicable
hadoop@u20-01:~$ hadoop fs -mkdir /user/mgpu/Brinkovan
2022-11-14 16:56:59,579 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platfo
rm... using builtin-java classes where applicable

```

Show 25 entries
Search:

<input type="checkbox"/>	Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name
<input type="checkbox"/>	<a href="#">drwxr-xr-x</a>	<a href="#">hadoop</a>	<a href="#">supergroup</a>	0 B	Nov 14 16:57	<a href="#">0</a>	0 B	<a href="#">Brinkovan</a>

Showing 1 to 1 of 1 entries

- Создать в локальной файловой системе случайный текстовый файл размером 10 Mb с именем VA.txt:

```
hadoop@u20-01:~$ base64 /dev/urandom | head -c 10000000 > VA.txt
```

- Заархивировать созданный текстовый файл:

```
hadoop@u20-01:~$ gzip -c VA.txt > VA.gz
```

- Скопировать текстовый файл и архив в директорию

/user/mgpu/Brinkovan HDFS виртуальной машины:

```
hadoop@u20-01:~$ hadoop fs -copyFromLocal VA.txt /user/mgpu/Brinkovan
2022-11-14 17:00:14,507 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platfo
rm... using builtin-java classes where applicable
hadoop@u20-01:~$ hadoop fs -copyFromLocal VA.gz /user/mgpu/Brinkovan
2022-11-14 17:00:35,231 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platfo
rm... using builtin-java classes where applicable
```

Show 25 entries
Search:

<input type="checkbox"/>	Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name
<input type="checkbox"/>	<a href="#">-rw-r--r--</a>	<a href="#">hadoop</a>	<a href="#">supergroup</a>	7.25 MB	Nov 14 17:00	<a href="#">1</a>	128 MB	<a href="#">VA.gz</a>
<input type="checkbox"/>	<a href="#">-rw-r--r--</a>	<a href="#">hadoop</a>	<a href="#">supergroup</a>	9.54 MB	Nov 14 17:00	<a href="#">1</a>	128 MB	<a href="#">VA.txt</a>

Showing 1 to 2 of 2 entries



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:

```
hadoop@u20-01:~$ hadoop fs -cat /user/mgpu/Brinkovan/VA.txt
7qe5UySpinXLQpmwUCeSyUvRV3EYzPErt++MRT1u/UUMEX1iEBLn4JSievZ/pzWYeYkDP4QX2Eq0
H+Zv0NVo4A3WZiQ0BAENL2QXcVFZzGsL+jgNc1NkmTsw34+SJ7n0LDarAJG+mNQsF21ip8MNof60
Fi7Cpks6Iaoehn3ARTCiIN9Gq0/FA2/KK1h+ZnWFXE1YfCZIEUj4B42qiAieCHMJRD4h2Guc1sw
y1uapmZErIpnAXDzQWQeQvJeJEme2wSDx7n+oWgaa33pVZURhKPCsdafwf0S30Kxos2VNpls7eKi
ohVGL5qkduU/qvdQ8CxfJD7h2mLFFQgTsLaiizvh6vkNu6fqdkeYtQeeGg1EbkhQIP45WtGvpHL
hKfcmnybRemZ3x1mH5N6JZfGEVmxwPMXeFEssrQbY7ECFC16hw5RCu8ABF8wi8opjwhk4IIEedrm
2T8nU3ePmfQAGvg3w5+Mfgv/+Hc5aKMx/vYbqeRGRoBFMF3xJGXWtjkdOfLfnx7417Zf8JqeOzX
4a3gJVW4bsHDhmqwz6q/Pz5d8CxMWfttgSduIk2Hf508mD3LJngrHEDcrAdX2aqYSH5bvJjT8w7w
1I4sEgdnSPWTHR9/Oez9Zxs4o3nbz3ZBNCLfrgPYIL7TPZL9iUeeYUVCvRwo7Dz/GRXqQ7U69+tF
Ez9EpeRwtBFpob+uRpQSwEDQrmI9Sykjj/ly7EtiznsW3ces695LiELKYfqWW+eMB6/uOKhOXUNA
R1hAvKHqQCEvgxUdGvrxnkBptcKTvSInS3AaP8qP7MERmBCZKPKy9bQmLrM2G1EwYILXJSdxr3eX
HJ0lcXyCZ5nUdrjkyUdf8JoypGARr55RPnDVqmwfQ9GQIjp+W0o//eAv9k/rLexjUZDqdMwKT901
/c8ryiAAxf0Qp8rgTRU5Kbb1tra9H4hgmz2Yr5KESzt90cwLDSiZgk4DmwgiumeTklpczf7YEmjC
DueJFPBwHWQPXs61au0GwkvoQ09UqZYS646TCTRD+YmChLTztJzM+rRlDnL8G6+hv8HR3ixL0vB
9U8RhciBhNLDUYUCJQrBK9Y1WuIT5Hb+g+sLbDQVbRw0Zflto3o0/xrheZcmsfJY6qQzXfk9hUP
u000HUsACL+4znMwvYg0R0nSfTdtScqzoFxC48LrTZ1FCxVMg+3gkiHMPgYcb8C9Ate5tkyaXzI6
wnLhyZk9/HeANYZct62uuRupETzPWdsdleqoLk5DK8wePHq+b/Ul5wWXDhHgcJnidN6866MJwFVd
tw2w0jncLuyYpiKlmsVAF/YJLYV3kz4+k5ZtEwbw8d4Cz+Guc3EyK73qFRQBiUWxQYXOkhUDjo58
omhYrXxExIE8Dg7FbLw3qpFpCqRFjyVzKUCaIE3YyQqgNhoUZUtWvtdvs3/XcsrsONP9x+2IKxqn
05XgS8M4Spa1oB4nBBWznZqvcnod24M3KbX1CzZ+Fm/BuU27eVVJELpC+idFVRND++SFfBpSfZ15
Kg/i7tPolU0SFCgk7PYydCR+KDNQXZeMmroV3JW8IbSW1sy7U1BA9JkvfgRqTAcw98LMmX9gxUt7
+Q7JtW4PBfc5eK2bR56eouIX3wlr7i/rzfk1X4rWCCdCXMiThZ0+g7t0zWiPN0sypQZ8zDpykHH
4VfL2mp1sWRw3TA+qDgptnVN2eOKOr0dbI3+fs0pniDn05yt12zWIFSoGLJ2wIhz5zSjKNDM3fJZ
vd4+afbUjnSMGidRUYbV+b3Cy7cFKDGJ3LcxCPmrQAWQNE0wTV5dZ4APznJWr/MX9Kr2jY3qD3f
ATouyXdehBYr2CVpbSgpmRqNKI1V9k88Qn3Nnlpw2Al+HTZr6HcmC8DYtdwel+fX8ZE3VcQtXAM
8A9JhmpMPcMq7JETHnqqZ0pQhlpK5PbnAu05vXd3x92Ss2eQMzC/19KmTNEiGsIDTrmnhQW1tGsy
HJYB5GejhnEiAIXm0pC5vkSyEs3RQC7OZNwWgtST6ycBG9K9/0qY5tdlPMUWhFcmehefn03ndXSW
jEbtwFrLhrVi0fA2GMWEz102URLyeQMqGyTLxCgytMEAano2Xtd+PYAlF4T7gX9ybOXToxMiNeaF
h8oZ271pT4N68m//AfDPGE7DxApuh9xpGxmvNNDJKPHCK9yFfRrRsFUHnHreZAP+vZfAtRWIwGah
M3nVcCmVzy80EFjt19Cc5vmUZ7aaKnP7dmRCcn580fVFB1viWoaowV4WyPEKrkDjaIy2Bgbj9rGm
Ac0YVG86Tt07Lbhbi0+Hea5M3Wqoz/ANG8S39xMrTo4oUbx6epV9ExXkTO6i8+8UIS7xS5Dr06jX
QGTyOGAsuJEoUe4x/uaF527gmo9TpnaFZb9WguQzYTQjh1Ip89jFHMIn3gnd8SU6vrtqLa4dTJ4d
7sxAuBHCDuj84nng0c8vEwJBLnW1YRKzfGjD4FeRQWxFvXUqyQbj1UDBZGnY0/snKTVFhdCiNVmf
N4UTOav32J6ZqsGwHNNoiMcfXxTFGffxGOILY50FJYPJ2GEC14gz068bpImH45YoQY0QWb2LmWQM
AvEA1g1NUB5lidVznrXagL0wZygKtPqEKTICebGE/D+TYTNLcb0+9U0Khu8k1Gi21Bt/3ckH13Iv
XgSSksxJtxhadoop@u20-01:~$
hadoop@u20-01:~$ hadoop fs -cat /user/mgpu/Brinkovan/VA.gz
```





```
.0E; \GJHBMc+8xδ+k$/$0(♦HF/◊<♦♦♦♦T♦L♦♦♦X^♦◊\♦♦]♦♦♦30♦♦♦♦♦|♦♦♦-♦0!♦"πT♦♦8
♦◊`♦♦♦v!S♦♦♦[bj♦♦=3)♦,♦♦0♦t♦S♦♦♦♦F♦R5y♦♦y.♦4♦d>I♦♦♦♦♦T!♦♦+♦♦.♦N♦♦♦♦♦[j♦Z♦
♦♦♦♦♦"yV`♦Wu♦Q♦[♦♦♦♦♦Q5
.y6w♦P♦♦♦FW♦♦♦♦♦A♦ZQ♦(♦D♦j'♦♦♦[♦♦&♦♦♦♦♦♦♦[♦♦S♦q♦♦
J♦♦♦X♦♦♦♦♦♦"qu7♦k♦♦♦♦♦♦♦b]K;♦♦♦
♦◊(♦)9K<♦♦L♦♦♦♦♦♦♦λ♦_♦2♦m♦w$♦♦♦♦M♦G♦♦>♦♦語m1♦F♦♦cd- ]♦y>♦[♦@r7$Z♦d♦♦W♦o♦s♦X1$|♦f♦♦♦]♦9w2m_♦♦3♦♦♦♦K♦♦E♦♦♦♦a1♦
>♦k♦♦♦♦♦:XH♦<t2♦,♦♦♦♦♦|♦♦{\S♦♦♦♦♦1&6♦♦ 5♦♦♦l♦♦♦E'f♦♦♦*"♦♦(vv♦X♦♦♦♦♦b#♦w♦♦♦\♦9\♦♦X3AXF♦ySw8-♦♦♦♦A9♦♦t♦♦
♦FΔ.i♦♦♦_♦Rj♦♦j♦9
♦.eUi♦eHz`♦♦*♦♦V♦♦7♦!♦,♦♦p♦♦X♦GE♦♦C♦\♦Q♦-♦♦&♦♦♦\♦8♦+♦w?♦◊\♦♦!♦♦,/$/♦♦♦♦i♦w♦B♦♦♦g0♦♦♦♦♦♦♦k♦<-♦(♦V;%YX)~♦♦
♦>D~♦%ζl♦♦]8♦j♦♦'♦S♦♦=♦♦♦♦%m,♦4♦X♦Ysbm♦8♦♦♦♦♦♦♦
♦rF♦t♦♦♦fm#♦.♦-Y♦oQ♣18♦:rC♦♦r[♦♦N♦♦T♦♦LQ♦♦♦♦♦♦♦O[♦
♦z♦♦♦c(T>♦♦576D♦♦GrP♦♦♦c5♦$6 ♦♦♦]♦4♦♦♦T.♦
♦*♦♦♦?♦♦♦JRH>6♦♦♦N♦♦[u2♦♦,♦R♦7rW♦[♦♦]♦♦ ♦♦zN=AiS♦n♦♦♦N♦♦♦♦♦>♦♦♦2♦05 ♦♦@♦♦♦v♦♦♦♦)♦gd♦:<K♦♦♦♦
♦h0♦j ♦♦j
♦♦♦♦lH#♦2♦U♦7♦(▯♦a♦0♦j♦♦♦♦7C+♦^♦♦♦0♦AS♦
♦♦♦♦♠SE0X♦+|♦♦♦♦k4#♦♦♦♦$♦♦0♦♦]♦♦j♦♦ē♦♦{:♦♦♦ ♦♦♦$♦♦♦♦♦♦♦.R"W♦♦"^♦♦♦v♦sB♦♦|
g♦♦♦♦♦♦♦▯♦/
♦ ♦
+♦3♦♦♦♦♦♦&冉{♦♦♦r♦♦}F5♦♦2♦♦y♦p<"♦pΞ{$!~-♦♦♦♦♦a♦♦♦vY^♦♦♦e1♦_♦&}♦♦n(♦♦1♦♦♦♦ZPr♦♦%♦Kq♦♦y
♦;♦♦♦0k<rV_♦♦♦♦;♦>
M<♦U;♦♦♦♦♦%♦ ♦N♦♦47♦♦♦Z♦♦♦♦E机♦+♦GVq♦
♦SH\;j♦_=Y♦W0♦♦b]♦>♦♦♦說♦♦♦.T♦|♦F♦♦♦k♦♦3♦$^N{♦♦j<♦♦♦h♦L+♦Y♦♦|g8♦Z♦c♦♦!BiDt♦♦u♦♦!♦$w2♦V♦♦(♦չջ♦♦♦PX♦♦♦Ks1♦
GeZ♦♦qhX♦l♦♦♦♦♦Eh6%▯8V{♦♦:♦M{♦♦♦♦ {♦♦♦{♦♦:♦♦/♦/♦}&z;♦I♦♦♦♦♦~♦OR♦♦♦♦♦θ♦n♦♦K♦♦♦S♦♦♦♦|♦G♦g♦♦♦@fme n
♦♦~`♦N♦♦♦S♦♦♦♦♦b♦♦@jgc!Jhe♦♦;te_♦♦#♦E♦I♦u♦♦3♦v_♦♦N♦♦♦♦r9♦♦-♦♦♦♦/♦)♦♦OQ♦♦♦>Z♦]♦?r♦♦0a♦♦2♦♦r♦H♦CZ♦'♦)ck♦♦
@♦]♦♦♦♦iIZ♦♦♦t♦(N♦+♦!♦♦♦PP♦g[♦♦♦♦♦r'D8♦!l♦ ♦|S♦♦♦♦~♦♦♦♦3♦♦♦ .?♦]♦♦♦♦:2♦'"q♦:♦'M`v♦♦v♦ky♦♦6♦♦;♦♦,♦
♦♦z{♦♦♦♦hadoop@u20-01:~$ 1;1;120;120;1;0x65;1;9c
1: command not found
1: command not found
120: command not found
120: command not found
1: command not found
0x65: command not found
1: command not found
9c: command not found
hadoop@u20-01:~$ hadoop fs -text /user/mgpu/Brinkovan/VA.txt
2022-11-14 17:03:33,048 WARN util.NativeCodeLoader: Unable to load native-had
rm... using builtin-java classes where applicable
ypQZ8zDpykHH
4VfL2mp1sWRw3TA+qDgptnVN2eOK0r0dbI3+fs0pnidN05yt12zwIFSoGLJ2wIhz5zsJKnDM3fJJ
vd4+afBuJnSMGidRUybV+b3Cy7cFKDGJ3LcxCPmrRQAWQNE0wTV5dZ4APznJWr/MX9Kr2jY3qd3f
AIouyXdehBYr2CVpbSgpmRqNKI1V9k88Qn3Nnlpw2Al+HTZR6HcmC8DYTdweI+fX8ZEvs3VcQtXAM
8A9JhmpMPcMQ7JEThnqqZ0pQhlPk5PbnAu05vXd3x92Ss2eQMzC/19KmTNEiGsIDTrmhQW1tGsy
HJYB5GejhneAiAxm0pC5vkSyEs3RQC70ZNwwGtST6ycBG9K9/0qY5tdlpMUWhFcmehefn03ndXSW
jEBtwFrLhrVi0fA2GMWEz102URlyeQMqGyTLxCgytMEAano2Xtd+PYAlF4T7gx9yBOXToxiMiNeaf
h8oZ271pT4N68m//AfDPGE7DxApuh9xpGxmVNNdJKPHCK9yFFrRsFUHNhreZAP+vzfAtRWIWGah
M3nVcCmVzy80EFjt19Cc5vmUZ7aaKnP7dmRCcn580fVFb1viWoaoWV4WyPEKRkdjaIy2Bgby9rGm
Ac0YVG86Tt07Lbhbi0+Hea5M3Wqoz/ANG8S39xMrTo4oUbX6epV9ExXkTO6i8+8UIS7xs5SDr06jX
QGTYOGAsuJEoUe4x/uaf527gmo9TpnaFzb9WguQzYTQjh1Ip89jFHMIN3gnd8SU6vrtqLa4dTJ4d
7sxAuBHCDuj84nnng0c8vEWJBlnW1YRKzfGjD4FeRQWxFvXUqyQbj1UDBGnY0/snKTvFhdCiNVmf
N4UTOav32J6ZqsGWHNNoiMCfXxtFGffxGOILY50FJYPJ2GEC14gz068bpImH45YoQY0QWb2LmwQM
AvEA1g1NUB5lidVznrxagL0wZygktPqEKTIcebgE/D+TYTNLCb0+9UOKhu8k1Gi21BT/3ckH13Iv
XgSSksxJtxhadoop@u20-01:~$ hadoop fs -tail /user/mgpu/Brinkovan/VA.gz
XgSSksxJtxhadoop@u20-01:~$ hadoop fs -tail /user/mgpu/Brinkovan/VA.gz
2022-11-14 17:03:44,002 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platfo
rm... using builtin-java classes where applicable
6 ♦♦♦]♦4♦♦♦T.♦
♦*♦♦♦?♦♦♦JRH>6♦♦♦N♦♦[u2♦♦,♦R♦7rW♦[♦♦]♦♦ ♦♦zN=AiS♦n♦♦♦N♦♦♦♦♦>♦♦♦2♦05 ♦♦@♦♦♦v♦♦♦♦)♦gd♦:<K♦♦♦♦
♦h0♦j ♦♦j
♦♦♦♦lH#♦2♦U♦7♦(▯♦a♦0♦j♦♦♦♦7C+♦^♦♦♦0♦AS♦
♦♦♦♦♠SE0X♦+|♦♦♦♦k4#♦♦♦♦$♦♦0♦♦]♦♦j♦♦ē♦♦{:♦♦♦ ♦♦♦$♦♦♦♦♦♦♦.R"W♦♦"^♦♦♦v♦sB♦♦|
g♦♦♦♦♦♦♦▯♦/
♦ ♦
+♦3♦♦♦♦♦♦&冉{♦♦♦r♦♦}F5♦♦2♦♦y♦p<"♦pΞ{$!~-♦♦♦♦♦a♦♦♦vY^♦♦♦e1♦_♦&}♦♦n(♦♦1♦♦♦♦ZPr♦♦%♦Kq♦♦y
♦;♦♦♦0k<rV_♦♦♦♦;♦>
M<♦U;♦♦♦♦♦%♦ ♦N♦♦47♦♦♦Z♦♦♦♦E机♦+♦GVq♦
♦SH\;j♦_=Y♦W0♦♦b]♦>♦♦♦說♦♦♦.T♦|♦F♦♦♦k♦♦3♦$^N{♦♦j<♦♦♦h♦L+♦Y♦♦|g8♦Z♦c♦♦!BiDt♦♦u♦♦!♦$w2♦V♦♦(♦չջ♦♦♦PX♦♦♦Ks1♦
GeZ♦♦qhX♦l♦♦♦♦♦Eh6%▯8V{♦♦:♦M{♦♦♦♦ {♦♦♦{♦♦:♦♦/♦/♦}&z;♦I♦♦♦♦♦~♦OR♦♦♦♦♦θ♦n♦♦K♦♦♦S♦♦♦♦|♦G♦g♦♦♦@fme n
♦♦~`♦N♦♦♦S♦♦♦♦♦b♦♦@jgc!Jhe♦♦;te_♦♦#♦E♦I♦u♦♦3♦v_♦♦N♦♦♦♦r9♦♦-♦♦♦♦/♦)♦♦OQ♦♦♦>Z♦]♦?r♦♦0a♦♦2♦♦r♦H♦CZ♦'♦)ck♦♦
@♦]♦♦♦♦iIZ♦♦♦t♦(N♦+♦!♦♦♦PP♦g[♦♦♦♦♦r'D8♦!l♦ ♦|S♦♦♦♦~♦♦♦♦3♦♦♦ .?♦]♦♦♦♦:2♦'"q♦:♦'M`v♦♦v♦ky♦♦6♦♦;♦♦,♦
♦♦z{♦♦♦♦hadoop@u20-01:~$
```


9. Вывести статистику по директории /user/mgpu/Brinkovan виртуальной машины:

```
hadoop@u20-01:~$ hdfs dfs -du -h /user/mgpu/Brinkovan
2022-11-14 17:06:06,602 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platfo
rm... using builtin-java classes where applicable
7.2 M  7.2 M  /user/mgpu/Brinkovan/VA.gz
9.5 M  9.5 M  /user/mgpu/Brinkovan/VA.txt
```











10. Удалить поддиректорию / Brinkovan со всем содержимым:

```
hadoop@u20-01:~$ hdfs dfs -rm -r /user/mgpu/Brinkovan
2022-11-14 17:06:54,320 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platfo
rm... using builtin-java classes where applicable
Deleted /user/mgpu/Brinkovan
```





Show  entries Search:

									
<input type="checkbox"/>	Permission	Owner	Group	Size	Last Modified	Replication	Block Size		Name
No data available in table									

Showing 0 to 0 of 0 entries