Департамент образования и науки города Москвы
Государственное автономное образовательное учреждение
высшего образования города Москвы
«Московский городской педагогический университет»
Институт цифрового образования
Департамент информатики, управления и технологий

### ДИСЦИПЛИНА:

«Распределенные системы»

Отчет по практической работе №5

Тема:

«Hadoop»

Выполнила:

Овчинникова А.А.

группа: ТП-191

Преподаватель:

Босенко Т.М.

Москва

2022

```
lemp001@u20-09:-$ 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@u20-09:-5
```

```
lenp001@u20-09:-$ sudo apt-get install openjdk-8-jdk
Reading package lists... Done
Building dependency tree
Reading state infornation... Done
The following packages were automatically installed and are no longer required:
    chromium-codecs-ffnpeg-extra gstreamer1.0-vaapi libfwupdplugin1
    libgstreamer-plugins-badi.0-0 libva-wayland2
Use 'sudo apt autoremove' to remove them.
The following additional packages will be installed:
    ca-certificates-java fonts-dejavu-extra java-connon libatk-wrapper-java
    libatk-wrapper-java-jni libice-dev libpthread-stubs0-dev libsm-dev
    libx11-dev libxau-dev libxch-dev libxdncp-dev libxt-dev
    openjdk-8-jdk-headless openjdk-8-jre openjdk-8-jre-headless
    xiiproto-core-dev xiiproto-dev xorg-sgml-doctools xtrans-dev
Suggested packages:
    default-jre libice-doc libsm-doc libxi1-doc libxcb-doc libxt-doc
    openjdk-8-demo openjdk-8-source visualvm fonts-ipafont-gothic
    fonts-ipafont-mincho fonts-way-microhel fonts-way-zenhel
The following NEW packages will be installed:
    ca-certificates-java fonts-dejavu-extra java-connon libatk-wrapper-java
    libatk-wrapper-java-jni libice-dev libxdncp-dev libxt-dev openjdk-8-jdk
    openjdk-8-jdk-headless openjdk-8-jre openjdk-8-jre-headless
```

```
Processing triggers for fontconfig (2.13.1-2ubuntu3) ...

Processing triggers for desktop-file-utils (0.24-lubuntu3) ...

Processing triggers for nine-support (3.64ubuntu1) ...

Processing triggers for hicolor-icon-theme (0.17-2) ...

Processing triggers for libe-bin (2.31-0ubuntu9.9) ...

Processing triggers for nan-db (2.9.1-1) ...

Processing triggers for ca-certificates (20211016-20.04.1) ...

Updating certificates in /etc/ssl/certs...

0 added, 0 removed; done.

Running hooks in /etc/ca-certificates/update.d...

done.

done.

Processing triggers for sgml-base (1.29.1) ...

Setting up x11proto-dev (2019.2-lubuntu1) ...

Setting up libiae-dev:and64 (1:1.0.9-0ubuntu1) ...

Setting up libiae-dev:and64 (2:1.0.10-0ubuntu1) ...

Setting up libxahcp-dev:and64 (1:1.1.3-0ubuntu1) ...

Setting up libxahcp-dev:and64 (1:1.1.3-0ubuntu1) ...

Setting up x11proto-core-dev (2019.2-lubuntu1) ...

Setting up libxahcp-dev:and64 (1:1.4-2) ...

Setting up libx11-dev:and64 (2:1.6.9-2ubuntu1.2) ...

Setting up libx11-dev:and64 (2:1.6.9-2ubuntu1.2) ...

Setting up libxt-dev:and64 (1:1.1.5-1) ...

lemp061@u20-09:-$
```

```
lemp801@u20-09:~$ java -version
openjdk version "1.8.0_342"
OpenJDK Runtime Environment (build 1.8.0_342-8u342-b67-8ubuntu1-20.04-b67)
OpenJDK 64-Bit Server VM (build 25.342-b67, mixed mode)
lemp801@u20-09:~$
```

GNU nano 4.8

PATH="/usr/local/sbin:/usr/local/bin:/usr/sbin:/sbin:/bin:/usr/gane
JAVA\_HOME="/usr/lib/jvn/java-8-openjdk-and64"

JRE\_HOME="/usr/lib/jvn/java-8-openjdk-and64"

```
lemp001gu20-09: $ sudo adduser hadoop

/*din="ser hadoop" ...
AnyDesk ew 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]
lemp001gu20-09: $
```

```
hadoop@u20-09:/home/lemp001$ exit exit lemp001@u20-09:~$
```

```
lemp001@u20-09:-$ 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
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 nourses-term openssh-server openssh-sftp-server pdsh ssh ssh-import-id
0 upgraded, 8 newly installed, 0 to remove and 214 not upgraded.
Need to get 860 kB of archives.
After this operation, 6 649 kB of additional disk space will be used.
Do you want to continue? [Y/n] y
Get:1 http://ru.archive.ubuntu.com/ubuntu focal-updates/main amd64 openssh-sftp-server amd64 1:8.2p1-4
ubuntu0.5 [51,5 kB]
Get:2 http://ru.archive.ubuntu.com/ubuntu focal-updates/main amd64 openssh-server amd64 1:8.2p1-4ubunt
u0.5 [377 kB]
Get:3 http://ru.archive.ubuntu.com/ubuntu focal-updates/main amd64 ssh all 1:8.2p1-4ubuntu0.5 [5 084 B
Get:4 http://ru.archive.ubuntu.com/ubuntu focal/universe amd64 libgenders0 amd64 1.22-1build2 [29,2 kB,
Get:5 http://ru.archive.ubuntu.com/ubuntu focal/universe amd64 genders amd64 1.22-1build2 [29,7 kB]
Get:6 http://ru.archive.ubuntu.com/ubuntu focal/universe amd64 genders amd64 1.22-1build2 [29,7 kB]
Get:6 http://ru.archive.ubuntu.com/ubuntu focal/main amd64 ncurses-term all 6.2-0ubuntu2 [249 kB]
Get:7 http://ru.archive.ubuntu.com/ubuntu focal/universe amd64 pdsh amd64 2.31-3build2 [108 kB]
Get:8 http://ru.archive.ubuntu.com/ubuntu focal/main amd64 ssh-import-id all 5.10-0ubuntu1 [10,0 kB]
Fetched 860 kB in 0s (4 675 kB/s)
```

```
lemp001@u20-09:-$ sudo su hadoop
hadoop@u20-09:/home/lemp001$ cd
hadoop@u20-09:-$ 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:+pVz3BJB2wllo/zl+J6L5XG3bW+eYy5J07xNCv3QoxI hadoop@u20-09
The key's randomart image is:
+---[RSA 3072]----+
            0.+
            0 * 0
             = 0 .
             0 +
             . 0 .
           E.+..
           +.Bo+*+1
         . +.B@+%|
             .=*&0|
+----[SHA256]----+
hadoop@u20-09:-$
```

```
hadoop@u20-09:~$ cat /home/hadoop/.ssh/id_rsa.pub >> /home/hadoop/.ssh/authorized_keys
hadoop@u20-09:-S chmod 0600 /home/hadoop/.ssh/authorized keys
hadoop@u20-09:-$
hadoop@u20-09:-$ ssh localhost
The authenticity of host 'localhost (127.0.0.1)' can't be established. ECDSA key fingerprint is SHA256:k7sgkXvWUhmswyGv0g56fTaj8+H/PnNANN3LmcUwz4w.
Are you sure you want to continue connecting (yes/no/[fingerprint])? yes Warning: Permanently added 'localhost' (ECDSA) to the list of known hosts.
Welcome to Ubuntu 20.04.4 LTS (GNU/Linux 5.13.0-51-generic x86_64)
 * Documentation: https://help.ubuntu.com
 * Management:
                     https://landscape.canonical.com
                     https://ubuntu.com/advantage
 * Support:
219 updates can be applied immediately.
165 of these updates are standard security updates.
To see these additional updates run: apt list --upgradable
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@u20-09:-$ exit
logout
Connection to localhost closed.
hadoop@u20-09:-$
       u20-09:~$ wget https://archive.apache.org/dist/hadoop/common/hadoop-3.1.2/hadoop-3.1.2.tar.gz
--2022-11-01 09:32:35-- https://archive.apache.org/dist/hadoop/common/hadoop-3.1.2/hadoop-3.1.2.tar.g
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 9,04MB/S
                                                                                         in 31s
2022-11-01 09:33:06 (10,3 MB/s) - 'hadoop-3.1.2.tar.gz' saved [332433589/332433589]
hadoop@u20-09:-$
hadoop@u20-09:~$ mv hadoop-3.1.2 hadoop
hadoop@u20-09:~$
export HADOOP_HOME=/home/hadoop/hadoop
export HADOOP INSTALL
export HADOOP_MAPRED_HOME=SHA
export HADOOP COMMON HOME=
export HADOOP HDFS HOME=
export YARN HOME=
                                     __DIK_SMADOOP_HOME/lib/native
/sbin:$HADOOP_HOME/bi-
export HADOOP COMMON LIB NATIVE DIR=
export PATH=
export PDSH_RCMD_TYPE=ssh
hadoop@u20-09:~$ nano .bashrc
```

hadoop@u20-09:~\$ source .bashrc hadoop@u20-09:~\$

```
GNU nano 4.8
                    /home/hadoop/hadoop/etc/hadoop/hadoop-env.sh
                                                                      Modified
 export HADOOP_ENABLE_BUILD_PATHS="true"
 xport JAVA_HOME=/usr/lib/jvm/java-8-openjdk-amd64
File Name to Write: /home/hadoop/hadoop/etc/hadoop/hadoop-env.sh
                                        M-A Append
                    M-D DOS Format
                                                            M-B Backup File
  Get Help
  Cancel
                    M-M
                       Mac Format
                                        M-P
                                            Prepend
                                                             ^T To Files
                    /home/hadoop/hadoop/etc/hadoop/core-site.xml
                                                                      Modified
 GNU nano 4.8
 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. ->
name>fs.default.name</name>
value>hdfs://localhost:9000</value>
/configuration>
<name>dfs.replication</name>
<value>1</value>
<name>dfs.name.dir</name>
<value>file:///home/hadoop/hadoopdata/hdfs/namenode</value>
property>
<name>dfs.data.dir</name>
```

<value>file:///home/hadoop/hadoopdata/hdfs/datanode</value>

```
!-- Put site-specific property overrides in this file. -->
:name>mapreduce.framework.name</name>
value>yarn</value>
property>
:name>yarn.app.mapreduce.am.env</name>
:value>HADOOP_MAPRED_HOME=${HADOOP_HOME}
cname>mapreduce.map.env</name>
:value>HADOOP MAPRED HOME=${HADOOP HOME}</value>
cname>mapreduce.reduce.env</name>
cvalue>HADOOP_MAPRED_HOME=${HADOOP_HOME}
/property>
```

```
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.
<! Site specific YARN configuration properties ->
<name>yarn.nodemanager.aux-services</name>
<value>mapreduce shuffle</value>
<name>yarn.nodemanager.resource.memory-mb</name>
<value>16384</value>
  Get Help
            ^O Write Out ^W Where Is
                                      ^K Cut Text ^J Justify
                                                                 ^C Cur Pos
             ^R Read File ^\ Replace
                                       ^U Paste Text^T To Spell
  Exit
                                                                   Go To Line
```

```
nadoop@u20-09:-$ hdfs namenode -format
WARNING: /home/hadoop/hadoop/logs does not exist. Creating.
2022-11-01 09:53:16,982 INFO namenode.NameNode: STARTUP_MSG:
STARTUP_MSG: Starting NameNode
STARTUP_MSG: host = u20-09/1
                  host = u20-09/172.26.35.129
STARTUP MSG:
                   args = [-format]
STARTUP_MSG:
STARTUP_MSG:
                   version = 3.1.2
                   classpath = /home/hadoop/hadoop/etc/hadoop:/home/hadoop/hadoop/share/hadoop/common/lib
/netty-3.10.5.Final.jar:/home/hadoop/hadoop/share/hadoop/common/lib/log4j-1.2.17.jar:/home/hadoop/had
oop/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/share/hadoop/common/lib/jackson-core-asl-1.9.13.jar:/home/hadoop
/hadoop/share/hadoop/common/lib/audience-annotations-0.5.0.jar:/home/hadoop/share/hadoop/commo
n/lib/jersey-server-1.19.jar:/home/hadoop/hadoop/share/hadoop/common/lib/woodstox-core-5.0.3.jar:/home/hadoop/share/hadoop/share/hadoop/common/lib/kerb-crypto-1.0.1.jar:/home/hadoop/hadoop/share/hadoop/common
/llb/kerb-util-1.0.1.jar:/home/hadoop/hadoop/share/hadoop/common/llb/guava-11.0.2.jar:/home/hadoop/hadoop/hadoop/share/hadoop/common/lib/para
```

```
hadoop@u20-09:~$ start-dfs.sh

Starting namenodes on [localhost]

Starting datanodes

Starting secondary namenodes [u20-09]

u20-09: Warning: Permanently added 'u20-09,172.26.35.129' (ECDSA) to the list of known hosts.

2022-11-01 09:54:59,315 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your pla

tform... using builtin-java classes where applicable

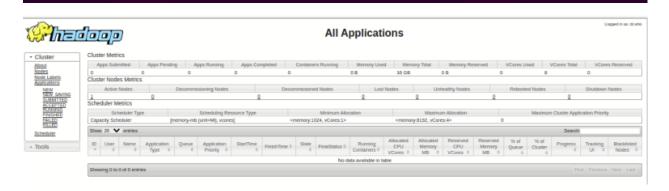
hadoop@u20-09:~$ start-yarn.sh

Starting resourcemanager

Starting nodemanagers

hadoop@u20-09:~$
```

```
adoop@u20-09: $ hdfs dfsadmin -report
2022-11-01 09:56:51,112 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your pla
tform... using builtin-java classes where applicable
Configured Capacity: 52044496896 (48.47 GB)
Present Capacity: 36087705600 (33.61 GB)
DFS Remaining: 36087681024 (33.61 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: u20-09
Decommission Status : Normal
Configured Capacity: 52044496896 (48.47 GB)
DFS Used: 24576 (24 KB)
Non DFS Used: 13282660352 (12.37 GB)
DFS Remaining: 36087681024 (33.61 GB)
DFS Used%: 0.00%
DFS Remaining%: 69.34%
```



Hadoop Overview Datanodes Datanode Volume Failures Snapshot Startup Progress Utilities

### Overview 'localhost:9000' (active)

Started:	Tue Nov 01 09:54:47 +0300 2022
Version:	3.1.2,r1019d0e05bcf12e05ef48ac71e84350d589e5d9a
Compiled:	Tue Jan 29 04:39:00 =0300 2019 by suring from branch-3.1.2
Cluster ID:	CID-dtx45740-c4ca-4b86-911c-421d2af7b80a
Block Pool ID:	BP-1940039704-172.26.35.129-1667285096398

## Summary

Security is off.

Salemode is of

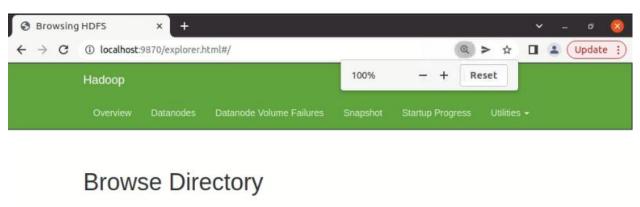
1 files and directories, 0 blocks (0 replicated blocks, 0 erasure coded block groups) = 1 total filesystem object(s).

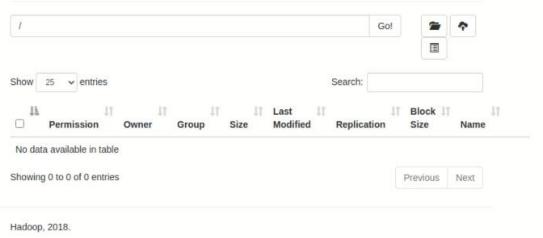
Heap Memory used 45.18 MB of 119.94 MB Heap Memory. Max Heap Memory is 1.86 GB.

Non Heap Memory used 47.85 MB of 48.77 MB Committed Non Heap Memory. Max Non Heap Memory is <urbox

Configured Capacity:	48.47 OB
Configured Remote Capacity:	0.8
DFS Used:	24 KB (0%)
Non DFS Used:	12.31 08
DFS Remaining:	33.67 GB (69.47%)
Block Pool Used:	24 KB (0%)
DataNodes usages% (Min/Median/Max/stdDev):	0.00% / 0.00% / 0.00% / 0.00%
Live Nodes	1 (Decommissioned: 0, in Maintenance: 0)
Dead Nodes	0 (Decommissioned: 0, in Maintenance: 0)
Decommissioning Nodes	0
Entering Maintenance Nodes	0
Total Datanode Volume Failures	0 (0 B)
Number of Under-Replicated Blocks	0
Number of Blocks Pending Deletion	0
Block Deletion Start Time	Tue Nov 01 09:54:47 +0300 2022
Last Checkpoint Time	Tue Nov 01 09:55:59 +0300 2022

NameNode Journal Status





hadoop@u20-09:-\$ hadoop fs -mkdir /user
2022-11-01 10:01:49,907 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your plat
form... using builtin-java classes where applicable
hadoop@u20-09:-\$ hadoop fs -mkdir /user/hadoop
2022-11-01 10:02:57,023 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your plat
form... using builtin-java classes where applicable
hadoop@u20-09:-\$

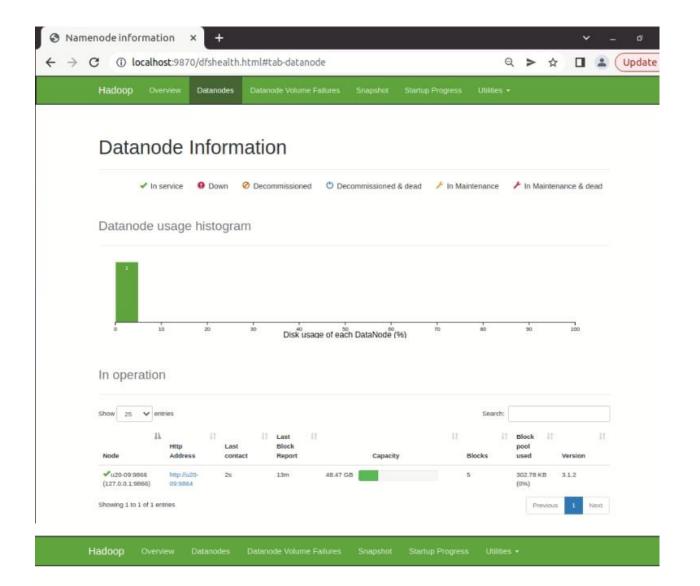
hadoop@u20-09:-\$ hadoop fs -ls /
2022-11-01 10:03:52,345 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your plat
form... using builtin-java classes where applicable
Found 1 items
drwxr-xr-x - hadoop supergroup 0 2022-11-01 10:02 /user
hadoop@u20-09:-\$

```
hadoop@u20-09:~$ hadoop fs -ls /
2022-11-01 10:03:52,345 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your plat
form... using builtin-java classes where applicable
Found 1 items
drwxr-xr-x
                - hadoop supergroup
                                                        0 2022-11-01 10:02 /user
hadoop@u20-09:-$ hadoop fs -put /var/log/dpkg.log /user/hadoop/dpkg.log
2022-11-01 10:04:43,791 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your plat
form... using builtin-java classes where applicable
hadoop@u20-09:-$ hadoop jar hadoop/share/hadoop/mapreduce/hadoop-mapreduce-examples-3.1.2.jar wordcoun t /user/hadoop/dpkg.log /user/hadoop/test_output 2022-11-01 10:05:18,294 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your plat
form... using builtin-java classes where applicable
2022-11-01 10:05:19,234 INFO client.RMProxy: Connecting to ResourceManager at /0.0.0.0:8032
2022-11-01 10:05:19,778 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding for path: /tmp/ha
doop-yarn/staging/hadoop/.staging/job_1667285747739_0001
2022-11-01 10:05:20,144 INFO input.FileInputFormat: Total input files to process : 1
2022-11-01 10:05:20,259 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)
2022-11-01 10:05:20,314 INFO mapreduce.JobSubmitter: number of splits:1
2022-11-01 10:05:20,582 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1667285747739_0001
2022-11-01 10:05:20,583 INFO mapreduce.JobSubmitter: Executing with tokens: []
2022-11-01 10:05:20,861 INFO conf.Configuration: resource-types.xml not found
2022-11-01 10:05:20,864 INFO resource.ResourceUtils: Unable to find 'resource-types.xml'.
2022-11-01 10:05:21,182 INFO impl.YarnClientImpl: Submitted application application_1667285747739_0001
2022-11-01 10:05:21,303 INFO mapreduce.Job: The url to track the job: http://u20-09:8088/proxy/applica
tion_1667285747739_0001/
2022-11-01 10:05:21,308 INFO mapreduce.Job: Running job: job_1667285747739_0001
```

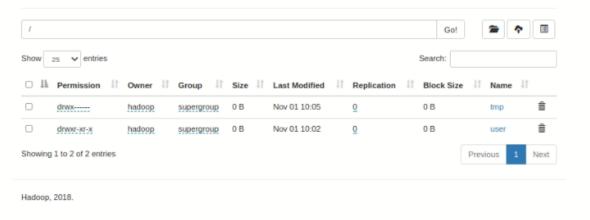


### **All Applications**

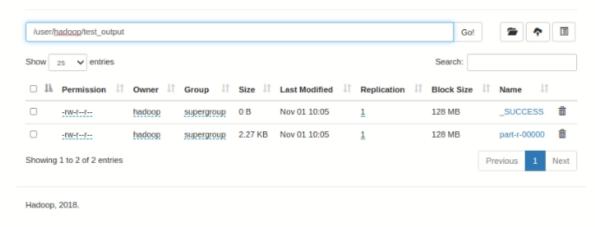
Submitted	Apps P	ending	nding Apps Running			Apps Completed		Containers Running		emory Used	Memory Total		Memory Reserved		
0 0		1	1		0	0 B		-	16 GB		0 B				
lodes Metrics															
tive Nodes		Dec	Decommissioning Nodes				Decommissioned Nodes			Lost Nodes		Unhealthy Nodes			
	0					0	<u>0</u>			0	0				
er Metrics															
Scheduler Type			Scheduling Resource Type					Minimum Allocation				Maximum Allocation			
Scheduler		[memo	[memory-mb (unit=Mi), vcores]					<memory:1024, vcores:1=""></memory:1024,>			<memory:8192, vcores:4=""></memory:8192,>				
✓ entries															
ID	·	User ¢	Name o	Application Type \$	Queue	Application Priority 0	StartTime	FinishTime	State 0	FinalStatus ◊	Running Containers	Allocated CPU VCores	Allocated Memory MB 0	Reserv CPU VCore	
n 166728574773	9 0001	hadoop	word	MAPREDUCE	default	0	Tue Nov 1 10:05:21 +0300 2022	Tue Nov 1 10:05:42 +0300 2022	FINISHED	SUCCEEDED	N/A	N/A	N/A	N/A	



# **Browse Directory**



# **Browse Directory**

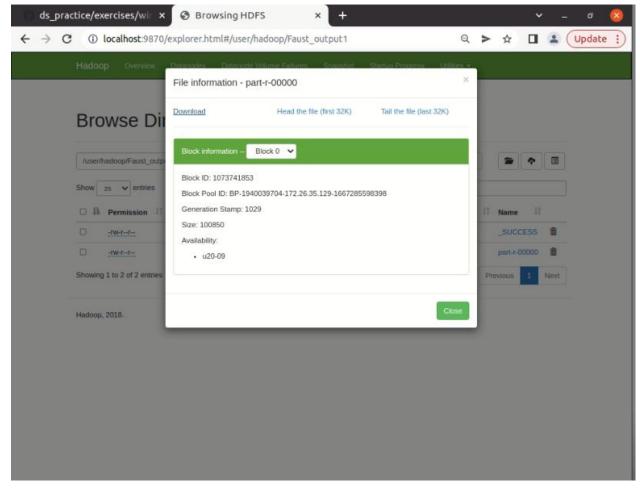


### Самостоятельная работа

```
lemp001@u20-09:-$ sudo su hadoop
[sudo] password for lemp001:
hadoop@u20-09:/home/lemp001$ cd
hadoop@u20-09:-$ git clone https://github.com/BosenkoTM/ds_practice.git
fatal: destination path 'ds_practice' already exists and is not an empty directo
ry.
hadoop@u20-09:-$ hadoop fs -put /home/hadoop/ds_practice/exercises/winter_semest
er_2021-2022/05_hadoop/sample_data/Faust_1.txt
2022-11-01 14:56:39,359 WARN util.NativeCodeLoader: Unable to load native-hadoop
library for your platform... using builtin-java classes where applicable
hadoop@u20-09:-$
```

```
hadoop@u20-09:-$ hadoop jar hadoop/share/hadoop/mapreduce/hadoop-mapreduce-examples-3.1.2.jar wordcount /user/hadoop/Faust_1.txt /user/hadoop/Faust_output1
2022-11-01 15:07:03,664 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
2022-11-01 15:07:04,537 INFO client.RMProxy: Connecting to ResourceManager at /0.0.0.0:8032
2022-11-01 15:07:05,030 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding for pat h: /tmp/hadoop-yarn/staging/hadoop/.staging/job_1667304364494_0002
2022-11-01 15:07:05,310 INFO input.FileInputFormat: Total input files to process: 1
2022-11-01 15:07:05,430 INFO mapreduce.JobSubmitter: number of splits:1
2022-11-01 15:07:05,664 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_16673043
64494_0002
2022-11-01 15:07:05,665 INFO mapreduce.JobSubmitter: Executing with tokens: []
2022-11-01 15:07:05,933 INFO conf.Configuration: resource-types.xml not found
2022-11-01 15:07:06,039 INFO resource.ResourceUtils: Unable to find 'resource-types.xml'.
2022-11-01 15:07:06,039 INFO impl.YarnClientImpl: Submitted application application_16673043
64494_0002
2022-11-01 15:07:06,113 INFO mapreduce.Job: The url to track the job: http://u20-09:8088/pro
xy/application_1667304364494_0002
2022-11-01 15:07:06,113 INFO mapreduce.Job: Running job: job_1667304364494_0002
2022-11-01 15:07:14,307 INFO mapreduce.Job: Map 0% reduce 0%
2022-11-01 15:07:26,482 INFO mapreduce.Job: map 100% reduce 0%
2022-11-01 15:07:26,482 INFO mapreduce.Job: Job job_1667304364494_0002 completed successfull
y
2022-11-01 15:07:26,644 INFO mapreduce.Job: Job job_1667304364494_0002 completed successfull
y
2022-11-01 15:07:26,644 INFO mapreduce.Job: Job job_1667304364494_0002 completed successfull
y
2022-11-01 15:07:26,644 INFO mapreduce.Job: Counters: 53
```

```
Bytes Written=100850
 ıadoop@u20-09:~Ş hadoop fs -cat /user/hadoop/Faust_output1/part-r-00000 | sort -k z -n -r |
head -n 16
sort: invalid number at field start: invalid count at start of 'z'
2022-11-01 15:09:58,207 WARN util.NativeCodeLoader: Unable to load native-hadoop library for
your platform... using builtin-java classes where applicable
cat: Unable to write to output stream.
hadoop@u20-09:~$ hadoop fs -cat /user/hadoop/Faust_output1/part-r-00000 | sort -k 2 -n -r |
head -n 16
2022-11-01 15:10:33,243 WARN util.NativeCodeLoader: Unable to load native-hadoop library for
 your platform... using builtin-java classes where applicable
und
           509
die
           463
           440
der
ich
           435
Und
           400
nicht
           346
           319
Desk
           291
etn
           284
das
           252
in
           243
stch
           225
Ich
           223
den
MEPHISTOPHELES: 219
FAUST: 200
hadoop@u20-09:-$
```



```
part-r-00000
Open
      ▼ 11
                                                                                    Save
 1 "Allein,
 2 "Alles 1
 3 "Als
 4 "Der
 5 "Die
          2
 6 "Er
          2
 7 "Ich
          4
 8 "Im
 9 "Mein
10 "Nur
11 "Rette 1
12 "So
13 "Uhu!
14 "Wenn
15 "Wie",
16 "MU♦
17 "nicht 1
18 "ungerechtes
19 'raus! 1
20 's
21 (Ab.)
          10
22 (Alle
          1
23 (Das
          2
24 (Der
          2
25 (Die
          4
26 (Ein
27 (Er
          24
28 (Es
29 (Faust 8
30 (Fortissimo): 1
31 (Gehn 4
32 (I)
          1
33 (II)
34 (Ihn
          1
35 (Indem 1
36 (Indessen
                   1
37 (Laut.)
                   2
                                                          Plain Text ▼ Tab Width: 8 ▼
                                                                                        Ln 1, Col 1
                                                                                                         INS
```

### Домашнее задание

```
hadoop@u20-09:-$ git clone https://github.com/Ovchinnikova111/practicds.git
Cloning into 'practicds'...
remote: Enumerating objects: 62, done.
remote: Counting objects: 100% (62/62), done.
remote: Compressing objects: 100% (56/56), done.
remote: Total 62 (delta 13), reused 0 (delta 0), pack-reused 0
Unpacking objects: 100% (62/62), 5.69 MiB | 8.70 MiB/s, done.
hadoop@u20-09:-$

hadoop@u20-09:-$ git clone https://github.com/Ovchinnikova111/practicds.git
cloning into 'practicds'...
remote: Enumerating objects: 62, done.
remote: Counting objects: 100% (62/62), done.
remote: Compressing objects: 100% (56/56), done.
remote: Total 62 (delta 13), reused 0 (delta 0), pack-reused 0
Unpacking objects: 100% (62/62), 5.69 MiB | 8.70 MiB/s, done.

hadoop@u20-09:-$ hadoop fs -put /home/hadoop/practicds/practice/cap.txt
2022-11-01 15:57:36,097 WARN util.NativeCodeLoader: Unable to load native-hadoop library for
your platform... using builtin-java classes where applicable
put: 'cap.txt': File exists
```

```
adoop@u20-09:~$ hadoop jar hadoop/share/hadoop/mapreduce/hadoop-mapreduce-examples-3.1.2.jar
wordcount /user/hadoop/cap.txt /user/hadoop/cap_output2
2022-11-01 15:58:35,968 WARN util.NativeCodeLoader: Unable to load native-hadoop library for
your platform... using builtin-java classes where applicable
2022-11-01 15:58:36,883 INFO client.RMProxy: Connecting to ResourceManager at /0.0.0.0:8032
2022-11-01 15:58:37,382 INFO tetent.km/roxy: connecting to Resourcemanager at 70.0.0.0.0032
2022-11-01 15:58:37,382 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding for path
: /tmp/hadoop-yarn/staging/hadoop/.staging/job_1667304364494_0007
2022-11-01 15:58:37,694 INFO input.FileInputFormat: Total input files to process : 1
2022-11-01 15:58:37,788 WARN hdfs.DataStreamer: Caught exception
java.lang.InterruptedException
           at java.lang.Object.wait(Native Method)
           at java.lang.Thread.joln(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)
2022-11-01 15:58:37,837 INFO mapreduce.JobSubmitter: number of splits:1
2022-11-01 15:58:38,084 INFO mapreduce.JobSubmitter: Submitting tokens for job: job 166730436
4494_0007
2022-11-01 15:58:38,086 INFO mapreduce.JobSubmltter: Executing with tokens: []
2022-11-01 15:58:38,357 INFO conf.Configuration: resource-types.xml not found
2022-11-01 15:58:38,358 INFO resource.ResourceUtils: Unable to find 'resource-types.xml'.
2022-11-01 15:58:38,454 INFO impl.YarnClientImpl: Submitted application application 166730436
4494_0007
2022-11-01 15:58:38,532 INFO mapreduce.Job: The url to track the job: http://u20-09:8088/prox
y/application_1667304364494_0007/
2022-11-01_15:58:38,537_INFO_mapreduce.Job: Running_job: job_1667304364494_0007
```

```
hadoop@u20-09:~$ hadoop fs -cat /user/hadoop/cap_output2/part-r-00000 | sort -k 2 -n -r | hea
d -n 16
2022-11-01 16:00:24,091 WARN util.NativeCodeLoader: Unable to load native-hadoop library for
your platform... using builtin-java classes where applicable
the
        7214
        5991
and
to
        5092
        4374
of
        3669
        3412
        3352
he
1n
        2919
was
        2713
you
        2691
that
        2534
        1964
hts
at
        1912
it
with
        1742
        1648
not
        1619
 adoop@u20-09:~$
```