

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:~$

```

```

lemp001@u20-09:~$ 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
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
Suggested packages:
  default-jre libice-doc libsm-doc libx11-doc libxcb-doc libxt-doc
  openjdk-8-demo openjdk-8-source visualvm fonts-ipafont-gothic
  fonts-ipafont-mincho fonts-wqy-microhei fonts-wqy-zenhei
The following NEW 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
  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 (0.24-1ubuntu3) ...
Processing triggers for mime-support (3.04ubuntu1) ...
Processing triggers for hicolor-icon-theme (0.17-2) ...
Processing triggers for gnome-menus (3.36.0-1ubuntu1) ...
Processing triggers for libc-bin (2.31-0ubuntu9.9) ...
Processing triggers for man-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 (2019.2-1ubuntu1) ...
Setting up libxau-dev:amd64 (1:1.0.9-0ubuntu1) ...
Setting up libice-dev:amd64 (2:1.0.10-0ubuntu1) ...
Setting up libsm-dev:amd64 (2:1.2.3-1) ...
Setting up libxdmcp-dev:amd64 (1:1.1.3-0ubuntu1) ...
Setting up x11proto-core-dev (2019.2-1ubuntu1) ...
Setting up libxcb1-dev:amd64 (1.14-2) ...
Setting up libx11-dev:amd64 (2:1.6.9-2ubuntu1.2) ...
Setting up libxt-dev:amd64 (1:1.1.5-1) ...
lemp001@u20-09:~$

```

```

Setting up libxt-dev:amd64 (1:1.1.5-1) ...
lemp001@u20-09:~$ 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)
lemp001@u20-09:~$

```

```

GNU nano 4.8 /etc/environment
PATH="/usr/local/sbin:/usr/local/bin:/usr/sbin:/usr/bin:/sbin:/bin:/usr/games:/usr/local/games:/snap/bin"
JAVA_HOME="/usr/lib/jvm/java-8-openjdk-amd64"
JRE_HOME="/usr/lib/jvm/java-8-openjdk-amd64/jre"

```

```

lemp001@u20-09:~$ sudo adduser hadoop
Adding new 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]
lemp001@u20-09:~$

```

```

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 ncurses-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 1:8.2p1-4ubuntu0.5 [51,5 kB]
Get:2 http://ru.archive.ubuntu.com/ubuntu focal-updates/main amd64 openssh-server amd64 1:8.2p1-4ubuntu0.5 [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/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: +pVz3BJB2wll0/zl+J6L5XG3bW+eYy5J07xNCv3QoxI hadoop@u20-09
The key's randomart image is:
+---[RSA 3072]---+
|                 o.+ |
|                 o * o |
|                 = o . |
|                 o + |
|                $ . o . |
|                . E.+.. |
|                . +.Bo+*+ |
|                . . +.B@+% |
|                . =*8O |
+-----[SHA256]-----+
hadoop@u20-09:~$

```



```
hadoop@u20-09:~$ cat /home/hadoop/.ssh/id_rsa.pub >> /home/hadoop/.ssh/authorized_keys
hadoop@u20-09:~$ 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
 * Support:       https://ubuntu.com/advantage

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:~$
```

```
hadoop@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
z
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=$HADOOP_HOME
export HADOOP_MAPRED_HOME=$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
```

```
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
###
###
# 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

File Name to Write: /home/hadoop/hadoop/etc/hadoop/hadoop-env.sh
^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
```

```
GNU nano 4.8 /home/hadoop/hadoop/etc/hadoop/core-site.xml Modified
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>
```

```
<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>
```



```
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>
```

```
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>

<!-- Site specific YARN configuration properties -->
<property>
<name>yarn.nodemanager.aux-services</name>
<value>mapreduce_shuffle</value>
</property>
<property>
<name>yarn.nodemanager.resource.memory-mb</name>
<value>16384</value>
</property>

^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
```

```
hadoop@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 = u20-09/172.26.35.129
STARTUP_MSG: args = [-format]
STARTUP_MSG: version = 3.1.2
STARTUP_MSG: 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-l
ogging-1.1.3.jar:/home/hadoop/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/hadoop/share/hadoop/commo
n/lib/jersey-server-1.19.jar:/home/hadoop/hadoop/share/hadoop/common/lib/woodstox-core-5.0.3.jar:/hom
e/hadoop/hadoop/share/hadoop/common/lib/kerb-crypto-1.0.1.jar:/home/hadoop/hadoop/share/hadoop/commo
n/lib/kerb-util-1.0.1.jar:/home/hadoop/hadoop/share/hadoop/common/lib/guava-11.0.2.jar:/home/hadoop/ha
doo/share/hadoop/common/lib/kerb-identity-1.0.1.jar:/home/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:~$


```

```

hadoop@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%

```



Cluster

[About](#)
[Nodes](#)
[Node Labels](#)
[Applications](#)
[NEW](#)
[STOP SAVING](#)
[SUBMITTED](#)
[PENDING](#)
[FINISHED](#)
[FAILED](#)
[KILLED](#)

Scheduler

Tools

All Applications

Cluster Metrics

Apps Submitted	Apps Pending	Apps Running	Apps Completed	Containers Running	Memory Used	Memory Total	Memory Reserved	VCores Used	VCores Total	VCores Reserved
0	0	0	0	0 B	10 GB	0 B	0 B	0	0	0

Cluster Nodes Metrics

Active Nodes	Decommissioning Nodes	Decommissioned Nodes	Lost Nodes	Unhealthy Nodes	Rebooted Nodes	Shutdown Nodes
1	0	0	0	0	0	0

Scheduler Metrics

Scheduler Type	Scheduling Resource Type	Minimum Allocation	Maximum Allocation	Maximum Cluster Application Priority
Capacity Scheduler	[memory-mb (unit=M), vcores]	<memory1024, vCores1>	<memory8192, vCores4>	0

Show 20 entries

ID	User	Name	Application Type	Queue	Application Priority	StartTime	FinishTime	State	FinalStatus	Running Containers	Allocated CPU V-Cores	Allocated Memory MB	Reserved CPU V-Cores	Reserved Memory MB	% of Queue	% of Cluster	Progress	Tracking UI	Blacklisted Nodes
No data available in table																			

Showing 0 to 0 of 0 entries

First Previous Next Last

Overview 'localhost:9000' (active)

Started:	Tue Nov 01 09:54:47 +0300 2022
Version:	3.1.2, r1019d0e5b0cd12e05ef48ac71e84550d58e525a
Compiled:	Tue Jan 29 04:39:06 +0300 2019 by sunlig from branch-3.1.2
Cluster ID:	CID-dfc45740-c4ca-4b86-911c-421d2af7b60a
Block Pool ID:	BP-1940099704-172.20.35.129-1667285598398

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 43.18 MB of 119.94 MB Heap Memory. Max Heap Memory is 1.86 GB.




Non Heap Memory used 47.85 MB of 46.77 MB Committed Non Heap Memory. Max Non Heap Memory is «unbounded».

Configured Capacity:	48.47 GB
Configured Remote Capacity:	0 B
DFS Used:	24 KB (0%)
Non DFS Used:	12.31 GB
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



Browse Directory

/ Go!   

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

Hadoop, 2018.

```
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 wordcount
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
doo-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.InterruptedExceptio
    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

Metrics

Submitted	Apps Pending	Apps Running	Apps Completed	Containers Running	Memory Used	Memory Total	Memory Reserved
0	0	1	0	0 B	16 GB	0 B	

Nodes Metrics

Active Nodes	Decommissioning Nodes	Decommissioned Nodes	Lost Nodes	Unhealthy Nodes
0	0	0	0	0

Job Metrics

Scheduler Type	Scheduling Resource Type	Minimum Allocation	Maximum Allocation
Scheduler	[memory-mb (unit=Mi), vcores]	<memory:1024, vCores:1>	<memory:8192, vCores:4>

▼ entries

ID	User	Name	Application Type	Queue	Application Priority	StartTime	FinishTime	State	FinalStatus	Running Containers	Allocated CPU VCoers	Allocated Memory MB	Reserved CPU VCore
job_1667285747739_0001	hadoop	word count	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

1 to 1 of 1 entries

Namenode information

+

localhost:9870/dfshealth.html#tab-datanode

Update

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

In operation

Show

25

entries

Search:

Node	Http Address	Last contact	Last Block Report	Capacity	Blocks	Block pool used	Version
<div>In service</div> <div>u20-09:9866 (127.0.0.1:9866)</div>	http://u20-09:9866	2s	13m	48.47 GB	5	302.78 KB (0%)	3.1.2

Showing 1 to 1 of 1 entries

Previous

1

Next

Hadoop

Overview

Datanodes

Datanode Volume Failures

Snapshot

Startup Progress

Utilities

Browse Directory

/

Go!

Folder

Upload

Refresh

Show

25

entries

Search:

<input type="checkbox"/>	Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name	
<input type="checkbox"/>	drwx-----	hadoop	supergroup	0 B	Nov 01 10:05	0	0 B	tmp	<div>Trash</div>
<input type="checkbox"/>	drwxr-xr-x	hadoop	supergroup	0 B	Nov 01 10:02	0	0 B	user	<div>Trash</div>

Showing 1 to 2 of 2 entries

Previous

1

Next

Hadoop, 2018.

Browse Directory

Show entries

Search:

<input type="checkbox"/>	Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name	
<input type="checkbox"/>	-rw-r--r--	hadoop	supergroup	0 B	Nov 01 10:05	1	128 MB	_SUCCESS	
<input type="checkbox"/>	-rw-r--r--	hadoop	supergroup	2.27 KB	Nov 01 10:05	1	128 MB	part-r-00000	

Showing 1 to 2 of 2 entries

Hadoop, 2018.

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 directory.
hadoop@u20-09:~$ hadoop fs -put /home/hadoop/ds_practice/exercises/winter_semester_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.RMPProxy: Connecting to ResourceManager at /0.0.0.0:8032
2022-11-01 15:07:05,030 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding for path: /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_1667304364494_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 15:07:05,939 INFO resource.ResourceUtils: Unable to find 'resource-types.xml'.
2022-11-01 15:07:06,039 INFO impl.YarnClientImpl: Submitted application application_1667304364494_0002
2022-11-01 15:07:06,113 INFO mapreduce.Job: The url to track the job: http://u20-09:8088/proxy/application_1667304364494_0002/
2022-11-01 15:07:06,115 INFO mapreduce.Job: Running job: job_1667304364494_0002
2022-11-01 15:07:14,307 INFO mapreduce.Job: Job job_1667304364494_0002 running in uber mode : false
2022-11-01 15:07:14,311 INFO mapreduce.Job: map 0% reduce 0%
2022-11-01 15:07:20,404 INFO mapreduce.Job: map 100% reduce 0%
2022-11-01 15:07:25,447 INFO mapreduce.Job: map 100% reduce 100%
2022-11-01 15:07:26,482 INFO mapreduce.Job: Job job_1667304364494_0002 completed successfully
2022-11-01 15:07:26,644 INFO mapreduce.Job: Counters: 53
```

```

Bytes Written=100850
hadoop@u20-09:~$ 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
der      440
lich     435
Und      400
nicht    346
Desk     319
lein     291
das      252
in       243
sich     225
Ich      223
den      221
MEPHISTOPHELES: 219
FAUST:    200
hadoop@u20-09:~$

```

ds_practice/exercises/wir x Browsing HDFS x +

localhost:9870/explorer.html#/user/hadoop/Faust_output1

Hadoop Overview Details Info Diagnostics Metrics Failures Resources Status Diagnostics Utilization

Browse Directory

/user/hadoop/Faust_output1

Show 25 entries

Permission

Showing 1 to 2 of 2 entries

Hadoop, 2016.

File information - part-r-00000

[Download](#) [Head the file \(first 32K\)](#) [Tail the file \(last 32K\)](#)

Block information -- Block 0

Block ID: 1073741853

Block Pool ID: BP-1940039704-172.26.35.129-1667285598398

Generation Stamp: 1029

Size: 100850

Availability:

- u20-09

Close

Name

_SUCCESS

part-r-00000

Previous 1 Next

```
Open  part-r-00000  Save  -  x
~/Downloads

1 "Allein, 1
2 "Alles 1
3 "Als 1
4 "Der 1
5 "Die 2
6 "Er 2
7 "Ich 4
8 "Im 1
9 "Mein 1
10 "Nur 1
11 "Rette 1
12 "So 1
13 "Uhu! 1
14 "Wenn 1
15 "Wie", 1
16 "muß 1
17 "nicht 1
18 "ungerechtes 1
19 'raus! 1
20 's 4
21 (Ab.) 10
22 (Alle 1
23 (Das 2
24 (Der 2
25 (Die 4
26 (Ein 1
27 (Er 24
28 (Es 2
29 (Faust 8
30 (Fortissimo): 1
31 (Gehn 4
32 (I) 1
33 (II) 1
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/0vchinnikova111/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/0vchinnikova111/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
```



```

hadoop@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 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.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 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.JobSubmitter: 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

```

```

Bytes Written=229185
hadoop@u20-09:~$ hadoop fs -cat /user/hadoop/cap_output2/part-r-000000 | 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
and      5991
to       5092
a        4374
of       3669
I        3412
he       3352
in       2919
was      2713
you      2691
that     2534
his      1964
at       1912
it       1742
with     1648
not      1619
hadoop@u20-09:~$

```