

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 семинарскому занятию № 6

Тема:
«Hadoop»

Выполнил: Григорьев П. К.
группа: ТП-191

Москва
2022

```
lemp@u20-03: ~  
lemp@u20-03:~$ sudo apt-get update  
[sudo] password for lemp:  
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 http://security.ubuntu.com/ubuntu focal-security InRelease  
Hit:6 http://ppa.launchpad.net/serge-rider/dbeaver-ce/ubuntu focal InRelease  
Hit:7 https://dl.google.com/linux/chrome/deb stable InRelease  
Hit:8 https://download.docker.com/linux/ubuntu focal InRelease  
Reading package lists... Done  
lemp@u20-03:~$ sudo apt-get install openjdk-8-jdk  
Reading package lists... Done  
Building dependency tree  
Reading state information... Done
```

```
lemp@u20-03:~$ 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)
```

```
lemp@u20-03:~$ sudo nano /etc/environment
```

```
GNU nano 4.8 /etc/environment Modified  
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"  
  
^G Get Help ^O Write Out ^W Where Is ^K Cut Text ^J Justify ^C Cur Pos M-U Undo  
^X Exit ^R Read File ^\ Replace ^U Paste Text ^T To Spell ^_ Go To Line M-E Redo
```

```
lemp@u20-03:~$ sudo nano /etc/environment
[sudo] password for lemp:
lemp@u20-03:~$ 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
```

```
lemp@u20-03:~$ sudo su hadoop
hadoop@u20-03:/home/lemp$
```

```
hadoop@u20-03:/home/lemp$ exit
exit
lemp@u20-03:~$
```

```
lemp@u20-03:~$ sudo apt-get install ssh pdsh
Reading package lists... Done
Building dependency tree
Reading state information... Done
The following packages were automatically installed:
```

```

lemp@u20-03:~$ sudo su hadoop
hadoop@u20-03:/home/lemp$ cd
hadoop@u20-03:~$ 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:ePhFDmHcFkmHCTGiTVDzck/wcWtHDvPlwE6uroAubuM hadoop@u20-03
The key's randomart image is:
+---[RSA 3072]---+
|      .o=.B+=+=.o  |
|      + =.*+ Bo+   |
|      . o +.+ o++  |
|      = * . .o     |
|      o S + .      |
|      + . .        |
|      . o .        |
|      +. . .       |
|      +Eo. . .     |
+-----[SHA256]-----+

```

```

hadoop@u20-03:~$ cat /home/hadoop/.ssh/id_rsa.pub >> /home/hadoop/.ssh/authorized_keys
hadoop@u20-03:~$ chmod 0600 /home/hadoop/.ssh/authorized_keys

```

```

hadoop@u20-03:~$ ssh localhost
The authenticity of host 'localhost (127.0.0.1)' can't be established.
ECDSA key fingerprint is SHA256:8rWGDYqscFWzXZhMxX6t8jLQHGajC4I5sYSubZwSA30.
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 5.15.0-46-generic x86_64)

```

```

* Documentation:  https://help.ubuntu.com
* Management:    https://landscape.canonical.com
* Support:       https://ubuntu.com/advantage

```

```

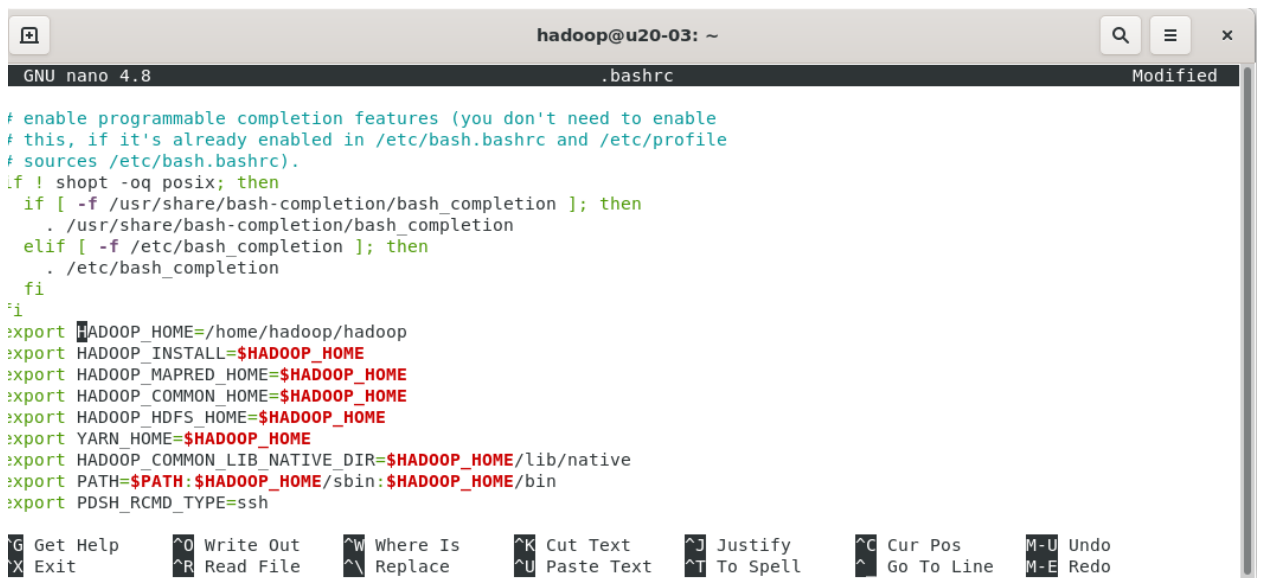
hadoop@u20-03:~$ exit
logout
Connection to localhost closed.
hadoop@u20-03:~$ wget https://archive.apache.org/dist/hadoop/common/hadoop-3.1.2/hadoop-3.1.2.tar.gz
--2022-10-25 10:40:56-- https://archive.apache.org/dist/hadoop/common/hadoop-3.1.2/hadoop-3.1.2.tar.gz
Resolving archive.apache.org (archive.apache.org)... 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/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@u20-03:~$
```

```
hadoop@u20-03:~$ mv hadoop-3.1.2 hadoop
```

```
hadoop@u20-03:~$ nano .bashrc
```



```
hadoop@u20-03: ~
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

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

```
hadoop@u20-03:~$ source .bashrc
```

```
hadoop@u20-03:~$ nano /home/hadoop/hadoop/etc/hadoop/hadoop-env.sh
```

```
hadoop@u20-03: ~
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
^G Get Help ^O Write Out ^W Where Is ^K Cut Text ^J Justify ^C Cur Pos M-U Undo
^X Exit ^R Read File ^\ Replace ^U Paste Text ^T To Spell ^_ Go To Line M-E Redo
```

hadoop@u20-03:~\$ nano /home/hadoop/hadoop/etc/hadoop/core-site.xml

```
hadoop@u20-03: ~
GNU nano 4.8 /home/hadoop/hadoop/etc/hadoop/core-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.
-->

<!-- 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 M-U Undo
^X Exit ^R Read File ^\ Replace ^U Paste Text ^T To Spell ^_ Go To Line M-E Redo
```

hadoop@u20-03:~\$ nano /home/hadoop/hadoop/etc/hadoop/hdfs-site.xml

```
hadoop@u20-03: ~
GNU nano 4.8 /home/hadoop/hadoop/etc/hadoop/hdfs-site.xml Modified
limitations under the License. See accompanying LICENSE file.
-->

<!-- 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 M-U Undo
^X Exit ^R Read File ^\ Replace ^U Paste Text ^T To Spell ^_ Go To Line M-E Redo
```

hadoop@u20-03: ~\$ nano /home/hadoop/hadoop/etc/hadoop/mapred-site.xml

```
hadoop@u20-03: ~
GNU nano 4.8 /home/hadoop/hadoop/etc/hadoop/mapred-site.xml Modified
<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 M-U Undo
^X Exit ^R Read File ^\ Replace ^U Paste Text ^T To Spell ^_ Go To Line M-E Redo
```

hadoop@u20-03: ~\$ nano /home/hadoop/hadoop/etc/hadoop/yarn-site.xml



hadoop@u20-03: ~

GNU nano 4.8

/home/hadoop/hadoop/etc/hadoop/yarn-site.xml

```
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>
```

```
</configuration>
```

```
2022-10-25 11:04:53,419 INFO namenode.FSImageFormatProtobuf: Saving image file /home/hadoop/hadoopdata/hdfs/namenode/
/current/fsimage.ckpt_00000000000000000000 using no compression
2022-10-25 11:04:53,568 INFO namenode.FSImageFormatProtobuf: Image file /home/hadoop/hadoopdata/hdfs/namenode/curren
t/fsimage.ckpt_00000000000000000000 of size 393 bytes saved in 0 seconds .
2022-10-25 11:04:53,586 INFO namenode.NNStorageRetentionManager: Going to retain 1 images with txid >= 0
2022-10-25 11:04:53,599 INFO namenode.NameNode: SHUTDOWN_MSG:
/*****
SHUTDOWN_MSG: Shutting down NameNode at u20-03/172.26.35.123
*****/
hadoop@u20-03:~$
```

```
hadoop@u20-03:~$ start-dfs.sh
```

```
Starting namenodes on [localhost]
```

```
Starting datanodes
```

```
Starting secondary namenodes [u20-03]
```

```
u20-03: Warning: Permanently added 'u20-03,172.26.35.123' (ECDSA) to the list of known hosts.
```

```
2022-10-25 11:05:36,616 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using
builtin-java classes where applicable
```

```
builtin java classes where applica
```

```
hadoop@u20-03:~$ start-yarn.sh
```

```
Starting resourcemanager
```

```
Starting nodemanagers
```



```
hadoop@u20-03:~$ hdfs dfsadmin -report
2022-10-25 11:06:42,131 WARN util.NativeCodeLoader: Unable to load nat:
builtin-java classes where applicable
Configured Capacity: 78143852544 (72.78 GB)
Present Capacity: 42877931520 (39.93 GB)
DFS Remaining: 42877906944 (39.93 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: u20-03
Decommission Status : Normal
Configured Capacity: 78143852544 (72.78 GB)
DFS Used: 24576 (24 KB)
Non DFS Used: 31249612800 (29.10 GB)
DFS Remaining: 42877906944 (39.93 GB)
DFS Used%: 0.00%
DFS Remaining%: 54.87%
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 Oct 25 11:06:40 MSK 2022
Last Block Report: Tue Oct 25 11:05:31 MSK 2022
```

172.26.35.123 — Подключение к удаленному рабочему столу

Activities Google Chrome OKT 25 11:08 en

ds_practice/Exercise_ x All Applications x +

localhost:8088/cluster

All Applications

100% - + Reset Update

Logged in as: dr who

Cluster

- About
- Nodes
- Node Labels
- Applications
- NEW
- NEW SAVING
- SUBMITTED
- ACCEPTED
- RUNNING
- FINISHED
- FAILED
- KILLED
- Scheduler

Tools

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	0 B	16 GB	0 B	0	8	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]	<memory:1024, vCores:1>	<memory:8192, vCores:4>	0

Show: 20 entries

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

Showing 0 to 0 of 0 entries

172.26.35.123 — Подключение к удаленному рабочему столу

Activities Google Chrome OKT 25 11:11 en

ds_practice/Exercise_ x All Applications x 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 Oct 25 11:05:27 +0300 2022
Version:	3.1.2, r1019dde65bc12e05e148ac71e84550d589e5d9a
Compiled:	Tue Jan 29 04:39:00 +0300 2019 by sunlig from branch-3.1.2
Cluster ID:	CID-dbab03cf-ed74-462a-96e9-c26e164bac89
Block Pool ID:	BP-166653375-172.26.35.123-1666685093359

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 90.54 MB of 298 MB Heap Memory. Max Heap Memory is 1.69 GB.

Non Heap Memory used 47.38 MB of 48.4 MB Committed Non Heap Memory. Max Non Heap Memory is <unbounded>.

Configured Capacity:	72.78 GB
Configured Remote Capacity:	0 B
DFS Used:	24 KB (0%)
Non DFS Used:	29.11 GB
DFS Remaining:	39.93 GB (54.87%)
Block Pool Used:	24 KB (0%)

```
hadoop@u20-03:~$ hadoop fs -mkdir /user
2022-10-25 11:13:27,957 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform...
builtin-java classes where applicable
hadoop@u20-03:~$ hadoop fs -mkdir /user/hadoop
2022-10-25 11:13:48,929 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform...
builtin-java classes where applicable
hadoop@u20-03:~$
```

```
hadoop@u20-03:~$ hadoop fs -ls /
2022-10-25 11:14:35,365 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform...
builtin-java classes where applicable
Found 1 items
drwxr-xr-x - hadoop supergroup 0 2022-10-25 11:13 /user
```

Browse Directory

Go!

Show

25

 entries

Search:

<input type="checkbox"/>	Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name	
<input type="checkbox"/>	drwxr-xr-x	hadoop	supergroup	0 B	Oct 25 11:13	0	0 B	user	

Showing 1 to 1 of 1 entries

Previous

1

Next

```
hadoop@u20-03:~$ hadoop fs -put /var/log/dpkg.log /user/hadoop/dpkg.log
2022-10-25 11:15:50,795 WARN util.NativeCodeLoader: Unable to load native-hadoop
builtin-java classes where applicable
```

```
hadoop@u20-03:~$ hadoop jar hadoop/share/hadoop/mapreduce/hadoop-mapreduce-examples-3.1.2.jar wordcount /user/hadoop
/dpkg.log /user/hadoop/test_output
```

ds_practice/Exercise_1

All Applications

Browsing HDFS

+

localhost:8088/cluster

All Applications

Cluster

About

Nodes

Node Labels

Applications

NEW

NEW SAVING

SUBMITTED

ACCEPTED

RUNNING

FINISHED

FAILED

KILLED

Scheduler

Tools

Cluster Metrics

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

Cluster Nodes Metrics

Active Nodes	Decommissioning Nodes	Decommissioned Nodes	Lost Nodes	Unhealthy 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:4>

Show 20 entries

ID	User	Name	Application Type	Queue	Application Priority	StartTime	FinishTime	State	FinalStatus	Running Containers	Allocated CPU VCores	Allocated Memory MB	Reserved CPU VCores
application_1666685168877_0001	hadoop	word count	MAPREDUCE	default	0	Tue Oct 25 11:18:59 +0300 2022	Tue Oct 25 11:19:14 +0300 2022	FINISHED	SUCCEEDED	N/A	N/A	N/A	N/A

Showing 1 to 1 of 1 entries

```

hadoop@u20-03:~$ hadoop jar hadoop/share/hadoop/mapreduce/hadoop-mapreduce-examples-3.1.2.jar wordcount /user/hadoop
/dpkg.log /user/hadoop/test output
2022-10-25 11:18:57,477 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using
builtin-java classes where applicable
2022-10-25 11:18:58,007 INFO client.RMPProxy: Connecting to ResourceManager at /0.0.0.0:8032
2022-10-25 11:18:58,373 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/stag
ing/hadoop/.staging/job_1666685168877_0001
2022-10-25 11:18:58,611 INFO input.FileInputFormat: Total input files to process : 1
2022-10-25 11:18:58,703 INFO mapreduce.JobSubmitter: number of splits:1
2022-10-25 11:18:58,862 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1666685168877_0001
2022-10-25 11:18:58,863 INFO mapreduce.JobSubmitter: Executing with tokens: []
2022-10-25 11:18:59,086 INFO conf.Configuration: resource-types.xml not found
2022-10-25 11:18:59,086 INFO resource.ResourceUtils: Unable to find 'resource-types.xml'.
2022-10-25 11:18:59,537 INFO impl.YarnClientImpl: Submitted application application_1666685168877_0001
2022-10-25 11:18:59,609 INFO mapreduce.Job: The url to track the job: http://u20-03:8088/proxy/application_166668516
8877_0001/
2022-10-25 11:18:59,609 INFO mapreduce.Job: Running job: job_1666685168877_0001
2022-10-25 11:19:05,753 INFO mapreduce.Job: Job job_1666685168877_0001 running in uber mode : false
2022-10-25 11:19:05,755 INFO mapreduce.Job: map 0% reduce 0%
2022-10-25 11:19:10,830 INFO mapreduce.Job: map 100% reduce 0%
2022-10-25 11:19:14,875 INFO mapreduce.Job: map 100% reduce 100%
2022-10-25 11:19:15,896 INFO mapreduce.Job: Job job_1666685168877_0001 completed successfully
2022-10-25 11:19:15,992 INFO mapreduce.Job: Counters: 53
    File System Counters
        FILE: Number of bytes read=9536
        FILE: Number of bytes written=451119
        FILE: Number of read operations=0
        FILE: Number of large read operations=0
        FILE: Number of write operations=0
        HDFS: Number of bytes read=162540
        HDFS: Number of bytes written=7977
        HDFS: Number of read operations=8
        HDFS: Number of large read operations=0
        HDFS: Number of write operations=2
    Job Counters
        Launched map tasks=1
        Launched reduce tasks=1
        Data-local map tasks=1




```

```


-----
5.15.0.48.54~20.04.18 12
5.15.0.50.56~20.04.19 30
5.15.0.52.58~20.04.20 18
5.30.0-9ubuntu0.2 16
5.30.0-9ubuntu0.3 24
5.4.0-126.142 11
5.4.0-128.144 10
5.4.0-131.147 6
6.0-25ubuntu1 4
6.0-25ubuntu1.1 6
6.2-0ubuntu2 7
7.7.0+dfsg-1ubuntu1 36
7.7.0+dfsg-1ubuntu1.1 54
8u342-b07-0ubuntu1~20.04 28
9.4.0-1ubuntu1~20.04.1 35
<none> 376
archives 66
binutils-common:amd64 7
binutils-x86-64-linux-gnu:amd64 7
binutils:amd64 7
build-essential:amd64 7
ca-certificates-java:all 8
ca-certificates:all 4
config-files 10
configure 252
desktop-file-utils:amd64 24
dpkg-dev:all 7
fakeroot:amd64 7
fontconfig:amd64 12
fonts-dejavu-extra:all 7
fonts-opensymbol:all 18
g++-9:amd64 7
g++:amd64 7

```

Browse Directory

//user/hadoop/test_output/part-r-00000 Go!   




Show 25 entries Search:


<input type="checkbox"/>	Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name
<input type="checkbox"/>	-rw-r--r--	hadoop	supergroup	7.79 KB	Oct 25 11:19	1	128 MB	




Showing 1 to 1 of 1 entries Previous 1 Next

Hadoop, 2018.

Домашняя работа

 main  1 branch  0 tags Go to file Add file Code


 Honaht01 Add files via upload 65c475a 12 seconds ago 22 commits

 practice	Add files via upload	20 days ago
 CaP.txt	Add files via upload	12 seconds ago
 README.md	Update README.md	20 days ago

```
hadoop@u20-03:~$ git clone https://github.com/Honaht01/pds-practice
Клонирование в «pds-practice»...
remote: Enumerating objects: 71, done.
remote: Counting objects: 100% (71/71), done.
remote: Compressing objects: 100% (61/61), done.
remote: Total 71 (delta 12), reused 0 (delta 0), pack-reused 0
Распаковка объектов: 100% (71/71), 17.48 Миб | 9.08 Миб/с, готово.
```

```
hadoop@u20-03:~$ hadoop fs -put pds-practice/CaP.txt /user/hadoop/CaP.txt
2022-10-31 20:43:52,091 WARN util.NativeCodeLoader: Unable to load native-hadoop
library for your platform... using builtin-java classes where applicable
hadoop@u20-03:~$ hadoop jar hadoop/share/hadoop/mapreduce/hadoop-mapreduce-examp
les-3.1.2.jar wordcount /user/hadoop/CaP.txt /user/hadoop/CaP_out
2022-10-31 20:45:00,074 WARN util.NativeCodeLoader: Unable to load native-hadoop
library for your platform... using builtin-java classes where applicable
2022-10-31 20:45:00,571 INFO client.RMProxy: Connecting to ResourceManager at /0
.0.0.0:8032
2022-10-31 20:45:00,905 INFO mapreduce.JobResourceUploader: Disabling Erasure Co
```

localhost:8088/cluster/apps



Cluster

About
Nodes
Node Labels
Applications

NEW
NEW_SAVING
SUBMITTED
ACCEPTED
RUNNING
FINISHED
FAILED
KILLED

Scheduler

Tools

Cluster Metrics

Apps Submitted	2	Apps Pending	0	Apps Running	1	Apps Completed	1	Containers Running	1
----------------	---	--------------	---	--------------	---	----------------	---	--------------------	---

Cluster Nodes Metrics

Active Nodes	1	Decommissioning Nodes	0	Decommissioned Nodes	0
--------------	---	-----------------------	---	----------------------	---

Scheduler Metrics

Scheduler Type	Capacity Scheduler	Scheduling Resource Type	[memory-mb (unit-M), vcores]	Memory Limit	<memory:1024, vcores:1
----------------	--------------------	--------------------------	------------------------------	--------------	------------------------

Show: 20 entries

ID	User	Name	Application Type	Queue	Application Priority	StartTime	FinishTime
application_1667232468386_0004	hadoop	word count	MAPREDUCE	default	0	Mon Oct 31 20:45:01 +0300 2022	Mon Oct 31 20:45:11 +0300 2022
application_1667232468386_0003	hadoop	word count	MAPREDUCE	default	0	Mon Oct 31 19:49:02 +0300 2022	Mon Oct 31 19:49:11 +0300 2022

```
hadoop@u20-03: ~
hadoop@u20-03:~$ hadoop fs -get /user/hadoop/CaP_out/part-r-00000 CaP_out.csv
2022-10-31 20:47:30,023 WARN util.NativeCodeLoader: Unable to load native-hadoop
library for your platform... using builtin-java classes where applicable
hadoop@u20-03:~$ head -100 CaP_out.csv | sort -k 2 -n
! 1
& 1
(1862), 1
(1881) 1
(1889) 1
(1891) 1
(1914) 1
(1928), 1
(1938) 1
(1942), 1
(1943) 1
(1989) 1
(a 1
(Allen's 1
(Amalia 1
&amp;c. 1
(And 1
(Aniska 1
(at 1
(but 1
```

```
hadoop@u20-03:~$ head -100 CaP_out.csv | sort -k 2 -n -r
(he 13
(I 10
(as 7
(for 6
(though 5
(she 5
(and 5
(which 4
(Cough-cough-cough.) 4
(She 3
(He 3
!" 3
(whom 2
(The 2
(the 2
(Sonia 2
(so 2
(Raskolnikov 2
(more 2
(I've 2
```

