

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:  
«Инструменты для хранения и обработки больших данных»  
**пр\_05-01**  
Тема:  
**«Introduction to HDFS»**

Выполнила: Николаева С.Г., АДЭУ-201  
Преподаватель: Босенко Тимур Муртазович

Москва  
2023

```
lemp001@u20-04: ~  
lemp001@u20-04:~$ sudo apt-get install openjdk-8-jdk  
[sudo] password for lemp001:  
Reading package lists... Done  
Building dependency tree  
Reading state information... Done  
lemp001@u20-04:~$ java -version  
openjdk version "1.8.0_362"  
OpenJDK Runtime Environment (build 1.8.0_362-8u362-ga-0ubuntu1~20.04.1-b09)  
OpenJDK 64-Bit Server VM (build 25.362-b09, mixed mode)
```

```
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"  
  
File Name to Write: /etc/environment  
^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
```

```
lemp001@u20-04: ~
Setting up libxt-dev:amd64 (1:1.1.5-1) ...
lemp001@u20-04:~$ java -version
openjdk version "1.8.0_362"
OpenJDK Runtime Environment (build 1.8.0_362-8u362-ga-0ubuntu1~20.04.1-b09)
OpenJDK 64-Bit Server VM (build 25.362-b09, mixed mode)
lemp001@u20-04:~$ sudo nano /etc/environment
lemp001@u20-04:~$ 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]
lemp001@u20-04:~$
```

```
lemp001@u20-04: ~
lemp001@u20-04:~$ sudo su hadoop
hadoop@u20-04:/home/lemp001$ exit
exit
lemp001@u20-04:~$ 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 85 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-
```

```
hadoop@u20-04: ~
Processing triggers for libc-bin (2.31-0ubuntu9.9) ...
Processing triggers for ufw (0.36-6ubuntu1) ...
lemp001@u20-04:~$ sudo su hadoop
hadoop@u20-04:/home/lemp001$ cd
hadoop@u20-04:~$ 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:TtV5FgL+PYREnvF80skbxrsHoIqv6z203mqgeCK2ug hadoop@u20-04
The key's randomart image is:
+---[RSA 3072]---+
|      .oB..      |
|      . + Boo.   |
|      o = B*o    |
|      . o *.o+   |
|      . S . . +o |
|      . . . + . 0.|
|      . . o o.   ..|
|      .+      .+o .|
|      BE.      .+*0*|
+-----[SHA256]-----+
hadoop@u20-04:~$
```

```
hadoop@u20-04: ~
| BE.      .+*0*      |
+-----[SHA256]-----+
hadoop@u20-04:~$ cat /home/hadoop/.ssh/id_rsa.pub >> /home/hadoop/.ssh/authorized_keys
hadoop@u20-04:~$ chmod 0600 /home/hadoop/.ssh/authorized_keys
hadoop@u20-04:~$ ssh localhost
The authenticity of host 'localhost (127.0.0.1)' can't be established.
ECDSA key fingerprint is SHA256:9nmZ5vcfVQAFcKCmaDa8AsvQLbBRcTFQ1T+tjRnUrog.
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

86 updates can be applied immediately.
To see these additional updates run: apt list --upgradable

Your Hardware Enablement Stack (HWE) is supported until April 2025.
*** System restart required ***

The programs included with the Ubuntu system are free software;
the exact distribution terms for each program are described in the
```

```
hadoop@u20-04: ~  
  
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@u20-04:~$ exit  
logout  
Connection to localhost closed.  
hadoop@u20-04:~$ wget https://archive.apache.org/dist/hadoop/common/hadoop-3.1.2/hadoop-3.1.2.tar.gz  
--2023-04-21 13:51:50-- 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 19%[==>] 63,16M 795KB/s eta 4m 43s
```

```
hadoop@u20-04: ~
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@u20-04: ~
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@u20-04:~$ mv hadoop-3.1.2 hadoop
hadoop@u20-04:~$ nano .bashrc
hadoop@u20-04:~$ source .bashrc
hadoop@u20-04:~$
```

```
hadoop@u20-04: ~
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

hadoop@u20-04:~$ mv hadoop-3.1.2 hadoop
hadoop@u20-04:~$ nano .bashrc
hadoop@u20-04:~$ source .bashrc
hadoop@u20-04:~$ nano /home/hadoop/hadoop/etc/hadoop/hadoop-env.sh
hadoop@u20-04:~$
```



```
hadoop@u20-04: ~
GNU nano 4.8 /home/hadoop/hadoop/etc/hadoop/hdfs-site.xml Modified
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>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>

^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@u20-04: ~
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@u20-04: ~$ source hadoop-env.sh
hadoop@u20-04:~$ nano /home/hadoop/hadoop/etc/hadoop/hadoop-env.sh
hadoop@u20-04:~$ nano /home/hadoop/hadoop/etc/hadoop/core-site.xml
hadoop@u20-04:~$ nano /home/hadoop/hadoop/etc/hadoop/hdfs-site.xml
hadoop@u20-04:~$ nano /home/hadoop/hadoop/etc/hadoop/mapred-site.xml
hadoop@u20-04:~$ nano /home/hadoop/hadoop/etc/hadoop/yarn-site.xml
```


hadoop@u20-04: ~

```
total heap and retry cache entry expiry time is 600000 millis
2023-04-21 14:24:20,842 INFO util.GSet: Computing capacity for map NameNodeRet
Cache
2023-04-21 14:24:20,842 INFO util.GSet: VM type          = 64-bit
2023-04-21 14:24:20,842 INFO util.GSet: 0.029999999329447746% max memory 1.9 G
= 584.4 KB
2023-04-21 14:24:20,842 INFO util.GSet: capacity         = 2^16 = 65536 entries
2023-04-21 14:24:20,894 INFO namenode.FSImage: Allocated new BlockPoolId: BP-1
4279312-172.26.35.24-1682076260884
2023-04-21 14:24:20,927 INFO common.Storage: Storage directory /home/hadoop/ha
opdata/hdfs/namenode has been successfully formatted.
2023-04-21 14:24:20,947 INFO namenode.FSImageFormatProtobuf: Saving image file
home/hadoop/hadoopdata/hdfs/namenode/current/fsimage.ckpt_000000000000000000
ing no compression
2023-04-21 14:24:21,094 INFO namenode.FSImageFormatProtobuf: Image file /home/
doop/hadoopdata/hdfs/namenode/current/fsimage.ckpt_000000000000000000 of size
90 bytes saved in 0 seconds .
2023-04-21 14:24:21,116 INFO namenode.NNStorageRetentionManager: Going to reta
1 images with txid >= 0
2023-04-21 14:24:21,120 INFO namenode.NameNode: SHUTDOWN_MSG:
/*****
SHUTDOWN_MSG: Shutting down NameNode at u20-04/172.26.35.24
*****/
hadoop@u20-04:~$ █

.....,
hadoop@u20-04:~$ start-dfs.sh
Starting namenodes on [localhost]
Starting datanodes
Starting secondary namenodes [u20-04]
u20-04: Warning: Permanently added 'u20-04,172.26.35.24' (ECDSA) to the list of
known hosts.
2023-04-21 14:25:22,422 WARN util.NativeCodeLoader: Unable to load native-hadoop
library for your platform... using builtin-java classes where applicable
hadoop@u20-04:~$ █
```

```
hadoop@u20-04: ~  
-----  
Live datanodes (1):  
  
Name: 127.0.0.1:9866 (localhost)  
Hostname: u20-04  
Decommission Status : Normal  
Configured Capacity: 52044496896 (48.47 GB)  
DFS Used: 24576 (24 KB)  
Non DFS Used: 14471852032 (13.48 GB)  
DFS Remaining: 34898489344 (32.50 GB)  
DFS Used%: 0.00%  
DFS Remaining%: 67.06%  
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: Fri Apr 21 14:27:21 MSK 2023  
Last Block Report: Fri Apr 21 14:25:18 MSK 2023  
Num of Blocks: 0  
  
hadoop@u20-04:~$
```

ActivitiesGoogle Chrome  
BigDataAnalytic\_Practic x Exercise\_1.pdf x Входящие (62) - sofia x All Applications x +  
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	0	Apps Pending	0	Apps Running	0	Apps Completed	0	Containers Running	0 B	Memory Used	16 GB	Memory Total	0 B	Memory Reserved	0	VCores Used	8	VCores Total	0
----------------	---	--------------	---	--------------	---	----------------	---	--------------------	-----	-------------	-------	--------------	-----	-----------------	---	-------------	---	--------------	---

Cluster Nodes Metrics

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

Scheduler Metrics

Scheduler Type	Capacity Scheduler	Scheduling Resource Type	[memory-mb (unit=M), vcores]	Minimum Allocation	<memory:1024, vCores:1>	Maximum Allocation	<memory:8192, vCores:4>	Maximum Cluster Application Pri	0
----------------	--------------------	--------------------------	------------------------------	--------------------	-------------------------	--------------------	-------------------------	---------------------------------	---

Show 20 entries

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

Showing 0 to 0 of 0 entries

```
hadoop@u20-04:~$ hadoop fs -mkdir /user  
2023-04-21 14:34:48,307 WARN util.NativeCodeLoader: Unable to load native-hadoop  
library for your platform... using builtin-java classes where applicable  
hadoop@u20-04:~$ hadoop fs -mkdir /user/hadoop  
2023-04-21 14:35:08,857 WARN util.NativeCodeLoader: Unable to load native-hadoop  
library for your platform... using builtin-java classes where applicable  
hadoop@u20-04:~$
```

```
hadoop@u20-04:~$ hadoop fs -put /var/log/dpkg.log /user/hadoop/dpkg.log  
2023-04-21 14:37:15,591 WARN util.NativeCodeLoader: Unable to load native-hadoop  
library for your platform... using builtin-java classes where applicable  
hadoop@u20-04:~$
```

Activities Google Chrome anp 21 14:41

BigDataAnalytic\_Practic x Exercise\_1.pdf x Входящие (62) - sofia x All Applications x Browsing HDFS x +

localhost:8088/cluster Update

hadoop

## All Applications

Cluster Metrics

Apps Submitted	Apps Pending	Apps Running	Apps Completed	Containers Running	Memory Used	Memory Total	Memory Reserved	VCoers Used	VCoers Total	VCoers Reserved
1	0	1	0	1	2 GB	16 GB	0 B	1	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
application_1682076364451_0001	hadoop	word count	MAPREDUCE	default	0	Fri Apr 21 14:40:56 +0300 2023	Fri Apr 21 14:41:21 +0300 2023	FINISHED	SUCCEEDED	1	1	2048	0	0	12.5	12.5		History	0

Showing 1 to 1 of 1 entries

hadoop@u20-04: ~

```

ufw:all 4
uno-lib-private:amd64 18
unpack 182
unpacked 899
unzip:amd64 9
upgrade 275
ure:amd64 18
vim-common:all 18
vim-tiny:amd64 18
wireless-regdb:all 9
x11proto-core-dev:all 7
x11proto-dev:all 7
xorg-sgml-doctools:all 8
xorgxrdp:amd64 7
xrdp:amd64 7
xserver-common:all 9
xserver-xephyr:amd64 9
xserver-xorg-core:amd64 9
xserver-xorg-legacy:amd64 9
xtrans-dev:all 7
xwayland:amd64 9
xxd:amd64 18
zlib1g:amd64 9
hadoop@u20-04:~$

```

Доп задания

```

st1@st1-vbox:~$ git clone https://github.com/BosenkoTM/Big-Data-Storage-and-Processing-Tools
Cloning into 'Big-Data-Storage-and-Processing-Tools'...
remote: Enumerating objects: 123, done.
remote: Counting objects: 100% (27/27), done.
remote: Compressing objects: 100% (27/27), done.
remote: Total 123 (delta 12), reused 0 (delta 0), pack-reused 96
Receiving objects: 100% (123/123), 19.72 MiB | 1.90 MiB/s, done.
Resolving deltas: 100% (37/37), done.

```

```
st1@st1-vbox:~$ $HADOOP_HOME/bin/hadoop fs -mkdir -p /user/st1/
st1@st1-vbox:~$ $HADOOP_HOME/bin/hadoop fs -put Big-Data-Storage-and-Processing
-Tools/practice/Faust_1.txt /user/st1
st1@st1-vbox:~$
```

Activities Google Chrome 4T 16:30




Big-Data-S x Browsing x NEW,NEW x file-How x +

localhost:9870/explorer.html#/user/st1 Update

## Hadoop


Overview Datanodes Datanode Volume Failures Snapshot Startup Progress Utilities

### Browse Directory

/user/st1 Go!   

Search:

Show 25 entries

<input type="checkbox"/>	Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name
<input type="checkbox"/>	-rw-r--r--	st1	supergroup	183.53 KB	Apr 27 16:28	1	128 MB	Faust_1.txt 

Showing 1 to 1 of 1 entries

Previous 1 Next

Hadoop, 2023.

Activities Terminal 4T 16:40

st1@st1-vbox: ~/hadoop-3.3.5/share/hadoop/mapreduce

File Edit View Search Terminal Help

jdifflib-examplessources

st1@st1-vbox:~/hadoop-3.3.5/share/hadoop/mapreduce\$ \$HADOOP\_HOME/bin/hadoop jar \$HADOOP\_HOME/share/hadoop/mapreduce/hadoop-mapreduce-examples-3.3.5.jar wordcount /user/st1/Faust\_1.txt /user/st1/Faust\_1\_Output

2023-04-27 16:38:41,585 INFO client.DefaultNoHARMFailoverProxyProvider: Connecting to ResourceManager at /0.0.0.0:8032

2023-04-27 16:38:43,621 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/st1/.staging/job\_1682597900921\_0001

2023-04-27 16:38:44,781 INFO input.FileInputFormat: Total input files to process : 1

2023-04-27 16:38:45,077 INFO mapreduce.JobSubmitter: number of splits:1

2023-04-27 16:38:45,649 INFO mapreduce.JobSubmitter: Submitting tokens for job: job\_1682597900921\_0001

2023-04-27 16:38:45,650 INFO mapreduce.JobSubmitter: Executing with tokens: []

2023-04-27 16:38:46,265 INFO conf.Configuration: resource-types.xml not found

2023-04-27 16:38:46,266 INFO resource.ResourceUtils: Unable to find 'resource-types.xml'.

2023-04-27 16:38:52,551 INFO impl.YarnClientImpl: Submitted application application\_1682597900921\_0001

2023-04-27 16:38:52,694 INFO mapreduce.Job: The url to track the job: http://st1-vbox:8088/proxy/application\_1682597900921\_0001/

2023-04-27 16:38:52,696 INFO mapreduce.Job: Running job: job\_1682597900921\_0001

2023-04-27 16:39:42,326 INFO mapreduce.Job: Job job\_1682597900921\_0001 running in uber mode : false

2023-04-27 16:39:42,334 INFO mapreduce.Job: map 0% reduce 0%

Activities

Google Chrome

16:40


GitHub - Bosenko x

Browsing HDFS x

All Applications x

localhost:8088/cluster/apps

Update



Cluster

Ubuntu Software

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
1	0	1	0

Cluster Nodes Metrics

Active Nodes	Decommissioning Nodes
1	0

Scheduler Metrics

Scheduler Type	Scheduling Resource Type
Capacity Scheduler	[memory-mb (unit=Mi), vcores]

Show 20 entries

ID	User	Name	Application Type	Application Tags	Queue
application_1682597900921_0001	st1	word count	MAPREDUCE		default

Showing 1 to 1 of 1 entries



```
Activities Terminal 4T 16:42
st1@st1-vbox: ~/hadoop-3.3.5/share/hadoop/mapreduce
File Edit View Search Terminal Help

Total vcore-milliseconds taken by all reduce tasks=20321
Total megabyte-milliseconds taken by all map tasks=28781568
Total megabyte-milliseconds taken by all reduce tasks=20808704
Map-Reduce Framework
Map input records=6943
Map output records=30654
Map output bytes=314631
Map output materialized bytes=138684
Input split bytes=107
Combine input records=30654
Combine output records=9551
Reduce input groups=9551
Reduce shuffle bytes=138684
Reduce input records=9551
Reduce output records=9551
Spilled Records=19102
Shuffled Maps =1
Failed Shuffles=0
Merged Map outputs=1
GC time elapsed (ms)=409
CPU time spent (ms)=5390
Physical memory (bytes) snapshot=474775552
Virtual memory (bytes) snapshot=5196558336
Total committed heap usage (bytes)=372768768
Peak Map Physical memory (bytes)=259264512
Peak Map Virtual memory (bytes)=2597183488
Peak Reduce Physical memory (bytes)=215511040
Peak Reduce Virtual memory (bytes)=2599374848
Shuffle Errors
```



Activities Google Chrome 4T 16:42

GitHub - Bosenko x Browsing HDFS x All Applications x +

localhost:8088/cluster/apps

Logged in as: dr.who

Reserved Resources	Physical Mem Used %	Physical VCores Used %
<memory:0 B, vCores:0>	63	0

Healthy Nodes	Rebooted Nodes	Shutdown Nodes
0	0	

Maximum Cluster Application Priority	Scheduler Busy %
0	0

Search:

Allocated GPUs	Reserved CPU VCores	Reserved Memory MB	Reserved GPUs	% of Queue	% of Cluster	Progress	Tracking UI	Blacklisted Nodes
N/A	N/A	N/A	N/A	0.0	0.0	<div></div>	<a href="#">History</a>	0

First Previous **1** Next Last

```
st1@st1-vbox:~/hadoop-3.3.5/share/hadoop/mapreduce$ $HADOOP_HOME/bin/hadoop fs
-get /user/st1/Faust_1_Output/part-r-00000 Faust_1_Output.csv
st1@st1-vbox:~/hadoop-3.3.5/share/hadoop/mapreduce$ 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
st1@st1-vbox:~/hadoop-3.3.5/share/hadoop/mapreduce$
```

Activities Google Chrome 16:48

GitHub - Bosenko x Browsing HDFS x All Applications x +

localhost:9870/explorer.html#/user/st1 Update

### Hadoop

Overview Datanodes Datanode Volume Failures Snapshot Startup Progress Utilities

## Browse Directory

/user/st1 Go!

Show 25 entries Search:

Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name
-rw-r--r--	st1	supergroup	183.53 KB	Apr 27 16:28	1	128 MB	Faust_1.txt
drwxr-xr-x	st1	supergroup	0 B	Apr 27 16:40	0	0 B	Faust_1_Output

Showing 1 to 2 of 2 entries Previous 1 Next

```
st1@st1-vbox:~$ $HADOOP_HOME/bin/hadoop fs -put Big-Data-Storage-and-Processing
-Tools/practice/Faust_1.txt /user/st1
put: `/user/st1/Faust_1.txt': File exists
st1@st1-vbox:~$ $HADOOP_HOME/bin/hadoop fs -mkdir -p /user/nik/
st1@st1-vbox:~$ $HADOOP_HOME/bin/hadoop fs -put Big-Data-Storage-and-Processing
-Tools/practice/Faust_1.txt /user/nik
```

```


st1@st1-vbox:~$ $HADOOP_HOME/bin/hadoop jar $HADOOP_HOME/share/hadoop/mapreduce
/hadoop-mapreduce-examples-3.3.5.jar grep /user/nik/Faust_1.txt /user/nik/Faust
_1_Count_Output 'Faust'
2023-04-27 16:59:14,070 INFO client.DefaultNoHARMAFailoverProxyProvider: Connect
ing to ResourceManager at /0.0.0.0:8032
2023-04-27 16:59:15,128 INFO mapreduce.JobResourceUploader: Disabling Erasure C
oding for path: /tmp/hadoop-yarn/staging/st1/.staging/job_1682597900921_0002
2023-04-27 16:59:15,714 INFO input.FileInputFormat: Total input files to proces
s : 1
2023-04-27 16:59:16,032 INFO mapreduce.JobSubmitter: number of splits:1
2023-04-27 16:59:16,570 INFO mapreduce.JobSubmitter: Submitting tokens for job:
job_1682597900921_0002
2023-04-27 16:59:16,570 INFO mapreduce.JobSubmitter: Executing with tokens: []
2023-04-27 16:59:16,979 INFO conf.Configuration: resource-types.xml not found
2023-04-27 16:59:16,980 INFO resource.ResourceUtils: Unable to find 'resource-t
ypes.xml'.
2023-04-27 16:59:17,149 INFO impl.YarnClientImpl: Submitted application applica
tion_1682597900921_0002
2023-04-27 16:59:17,249 INFO mapreduce.Job: The url to track the job: http://st
1-vbox:8088/proxy/application_1682597900921_0002/
2023-04-27 16:59:17,251 INFO mapreduce.Job: Running job: job_1682597900921_0002
2023-04-27 16:59:31,126 INFO mapreduce.Job: Job job_1682597900921_0002 running
in uber mode : false
2023-04-27 16:59:31,137 INFO mapreduce.Job:  map 0% reduce 0%
2023-04-27 16:59:52,841 INFO mapreduce.Job:  map 100% reduce 0%

```

Activities Google Chrome 4T 17:00

GitHub - Bosenko x Browsing HDFS x All Applications x

localhost:8088/cluster/apps



Ubuntu Software

- 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
3	0	1	2

### Cluster Nodes Metrics

Active Nodes	Decommissioning Nodes
1	0

### Scheduler Metrics

Scheduler Type	Scheduling Resource Type
Capacity Scheduler	[memory-mb (unit=Mi), vcores]

Show 20 entries

ID	User	Name	Application Type	Application Tags	Queue
<a href="#">application_1682597900921_0003</a>	st1	grep-sort	MAPREDUCE		default
<a href="#">application_1682597900921_0002</a>	st1	grep-search	MAPREDUCE		default
<a href="#">application_1682597900921_0001</a>	st1	word count	MAPREDUCE		default

Showing 1 to 3 of 3 entries

```
Activities Terminal 4T 17:01 ● st1@st1-vbox: ~
Files File Edit View Search Terminal Help
2023-04-27 17:00:34,064 INFO mapreduce.Job: map 0% reduce 0%
2023-04-27 17:00:44,628 INFO mapreduce.Job: map 100% reduce 0%
2023-04-27 17:00:59,030 INFO mapreduce.Job: map 100% reduce 100%
2023-04-27 17:01:01,069 INFO mapreduce.Job: Job job_1682597900921_0003 completed successfully
2023-04-27 17:01:01,209 INFO mapreduce.Job: Counters: 54
    File System Counters
        FILE: Number of bytes read=22
        FILE: Number of bytes written=552879
        FILE: Number of read operations=0
        FILE: Number of large read operations=0
        FILE: Number of write operations=0
        HDFS: Number of bytes read=236
        HDFS: Number of bytes written=9
        HDFS: Number of read operations=9
        HDFS: Number of large read operations=0
        HDFS: Number of write operations=2
        HDFS: Number of bytes read erasure-coded=0
    Job Counters
        Launched map tasks=1
        Launched reduce tasks=1
        Data-local map tasks=1
        Total time spent by all maps in occupied slots (ms)=8142
        Total time spent by all reduces in occupied slots (ms)=10510
        Total time spent by all map tasks (ms)=8142
        Total time spent by all reduce tasks (ms)=10510
        Total vcore-milliseconds taken by all map tasks=8142
        Total vcore-milliseconds taken by all reduce tasks=10510
        Total megabyte-milliseconds taken by all map tasks=8337408
```

Activities Google Chrome 4T 17:02

GitHub - Bosenko x Browsing HDFS x All Applications x

localhost:8088/cluster/apps

LibreOffice Writer Logged in as: dr.who

Reserved Resources	Physical Mem Used %	Physical VCores Used %
memory:0 B, vCores:0>	59	0

Idle Nodes	Rebooted Nodes	Shutdown Nodes
0	0	

Maximum Cluster Application Priority	Scheduler Busy %
	0

Search:

Allocated GPUs	Reserved CPU VCores	Reserved Memory MB	Reserved GPUs	% of Queue	% of Cluster	Progress	Tracking UI	Blacklisted Nodes
N/A	N/A	N/A	N/A	0.0	0.0		<a href="#">History</a>	0
N/A	N/A	N/A	N/A	0.0	0.0		<a href="#">History</a>	0
N/A	N/A	N/A	N/A	0.0	0.0		<a href="#">History</a>	0

First Previous 1 Next Last

```
st1@st1-vbox:~$ $HADOOP_HOME/bin/hadoop fs -get /user/nik/Faust_1_Count_Output/
part-r-000000 Faust_1_Count_Output.csv
st1@st1-vbox:~$ cat Faust_1_Count_Output.csv
50      Faust
```

Activities

Google Chrome

17:05

GitHub - Bosenko x

Browsing HDFS x

All Applications x

+

localhost:9870/explorer.html#/user/nik

Update

Hadoop

OverviewDatanodesDatanode Volume FailuresSnapshotStartup ProgressUtilities

# Browse Directory

Go!

Show 25 entriesSearch:

	Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name
<input type="checkbox"/>	<a href="#">-rw-r--r--</a>	<a href="#">st1</a>	<a href="#">supergroup</a>	183.53 KB	Apr 27 16:54	<a href="#">1</a>	128 MB	<a href="#">Faust_1.txt</a>
<input type="checkbox"/>	<a href="#">drwxr-xr-x</a>	<a href="#">st1</a>	<a href="#">supergroup</a>	0 B	Apr 27 17:00	<a href="#">0</a>	0 B	<a href="#">Faust_1_Count_Ou</a>

Showing 1 to 2 of 2 entries

Previous1Next