

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

Практическая работа №5.1

Тема:

Introduction to HDFS

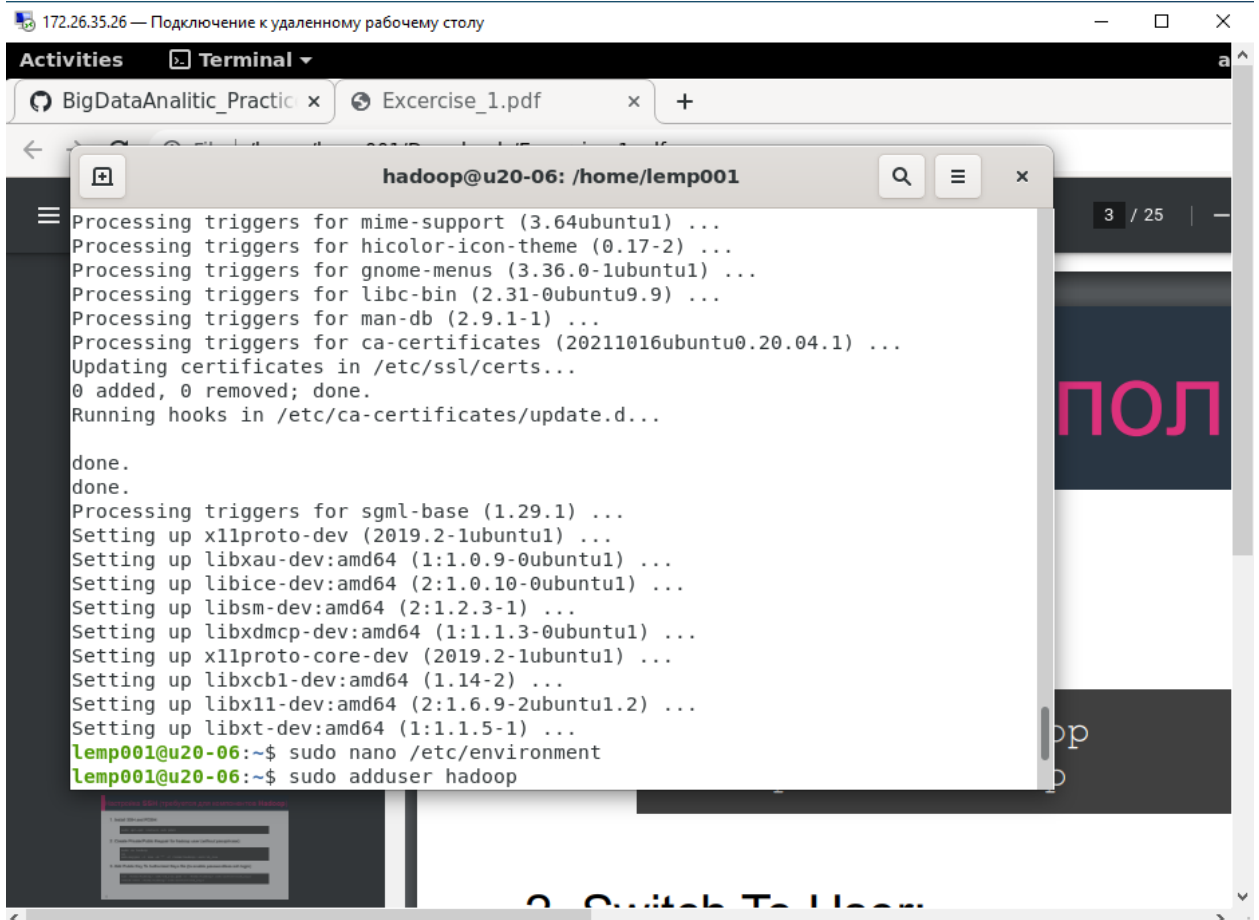
Выполнила: Шепелева Е. В., группа: АДЭУ-201

Преподаватель: Т. М. Босенко

Москва

2022

Задание в классе



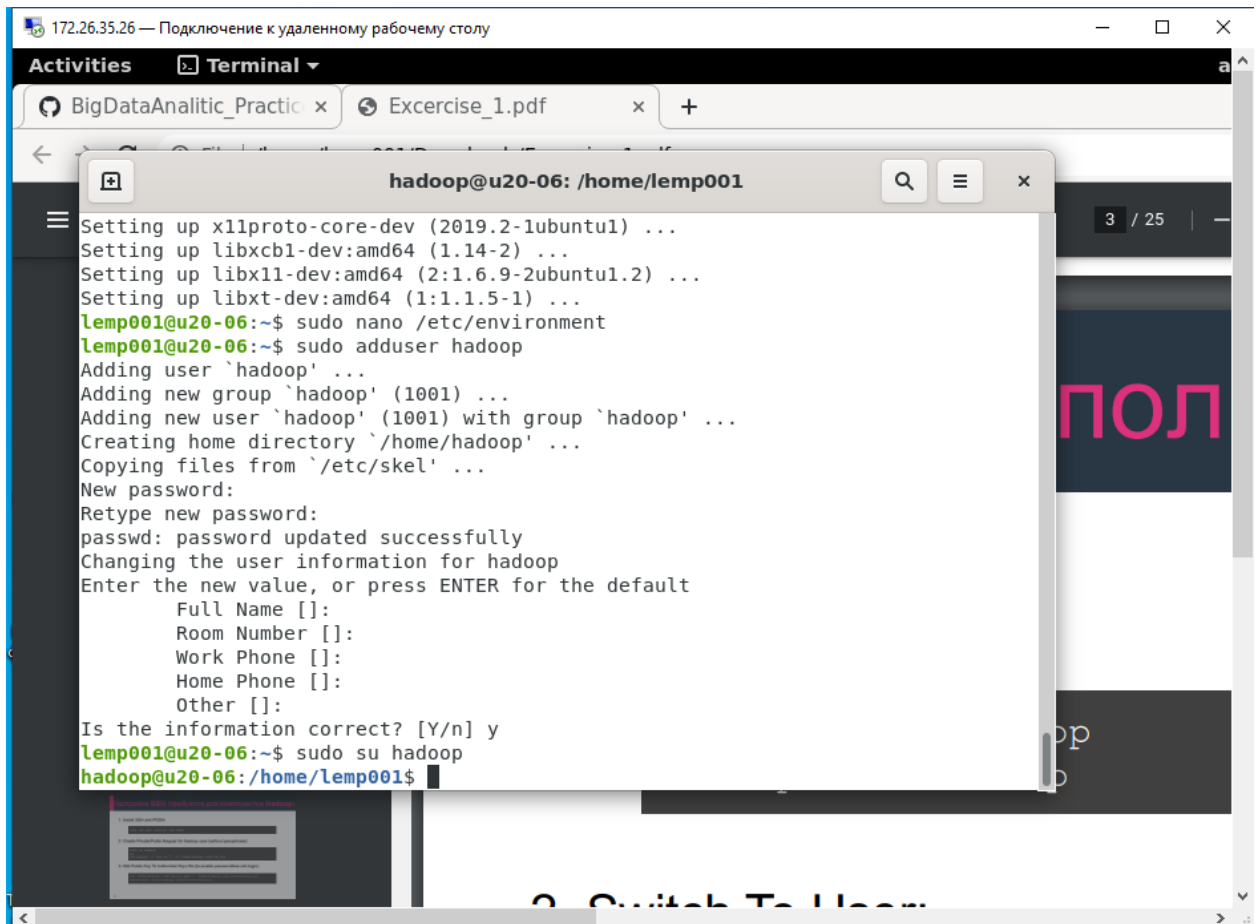
172.26.35.26 — Подключение к удаленному рабочему столу

Activities Terminal

BigDataAnalytic_Practic Exercise_1.pdf

hadoop@u20-06: /home/lemp001

```
Processing triggers for mime-support (3.64ubuntu1) ...
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 (20211016ubuntu0.20.04.1) ...
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@u20-06:~$ sudo nano /etc/environment
lemp001@u20-06:~$ sudo adduser hadoop
```



172.26.35.26 — Подключение к удаленному рабочему столу

Activities Terminal

BigDataAnalytic_Practic Exercise_1.pdf

hadoop@u20-06: /home/lemp001

```
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@u20-06:~$ sudo nano /etc/environment
lemp001@u20-06:~$ 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
lemp001@u20-06:~$ sudo su hadoop
hadoop@u20-06: /home/lemp001$
```

```
lemp001@u20-06: ~  
Creating config file /etc/ssh/sshd_config with new version  
Creating SSH2 RSA key; this may take some time ...  
3072 SHA256:j/bbgkcucTd8o9ifDzxt0NWRPMPD0xVNIQmDsPlZw0c root@u20-06 (RSA)  
Creating SSH2 ECDSA key; this may take some time ...  
256 SHA256:f/ZgYnhEXHnec9aXvQ9uGrsDoaaCh3a4IuB7L+MDq04 root@u20-06 (ECDSA)  
Creating SSH2 ED25519 key; this may take some time ...  
256 SHA256:zIkr3Q80Pmhd0PP0rKKSZhuPYqh3Pz6hm6+ikK1rXw root@u20-06 (ED25519)  
Created symlink /etc/systemd/system/ssh.service → /lib/systemd/system/ssh.service.  
Created symlink /etc/systemd/system/multi-user.target.wants/ssh.service → /lib/systemd/system/ssh.service.  
rescue-ssh.target is a disabled or a static unit, not starting it.  
Setting up ssh-import-id (5.10-0ubuntu1) ...  
Attempting to convert /etc/ssh/ssh_import_id  
Setting up genders (1.22-1build2) ...  
Setting up ncurses-term (6.2-0ubuntu2) ...  
Setting up ssh (1:8.2p1-4ubuntu0.5) ...  
Setting up pdsh (2.31-3build2) ...  
Processing triggers for systemd (245.4-4ubuntu3.20) ...  
Processing triggers for man-db (2.9.1-1) ...  
Processing triggers for libc-bin (2.31-0ubuntu9.9) ...  
Processing triggers for ufw (0.36-6ubuntu1) ...  
lemp001@u20-06:~$
```

```
hadoop@u20-06: ~  
Processing triggers for libc-bin (2.31-0ubuntu9.9) ...  
Processing triggers for ufw (0.36-6ubuntu1) ...  
lemp001@u20-06:~$ sudo su hadoop  
hadoop@u20-06:/home/lemp001$ cd  
hadoop@u20-06:~$ 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:j+BQVA2VnyDU6KQdIITCwFbp3NLX0LU5V06YTLlamba hadoop@u20-06  
The key's randomart image is:  
+---[RSA 3072]-----+  
|=. .++ o+=*o+.. |  
|.+. . o .+. * = |  
|.. o o .+= o o |  
| + +.. o. = @ |  
| o o SE o . |  
| o . oo |  
| . . . . |  
+---[SHA256]-----+  
hadoop@u20-06:~$
```

172.26.35.26 — Подключение к удаленному рабочему столу

Activities Terminal anp 21

BigDataAnalytic_Practic x Exercise_1.pdf +

hadoop@u20-06: ~

```
Processing triggers for ufw (0.36-6ubuntu1) ...
lemp001@u20-06:~$ sudo su hadoop
hadoop@u20-06:~/home/lemp001$ cd
hadoop@u20-06:~$ 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:j+BQVA2VnyDU6KQdIITCwFbp3NLX0LU5V06YTLlambda hadoop@u20-06
The key's randomart image is:
+---[RSA 3072]-----+
| = .++ o+=*o+.. |
| .. o .=. * = |
| .. o ..=o+ 0 o |
| + +..o.= @ |
| o o SE 0 . |
| o . oo |
| . . . |
+-----+

hadoop@u20-06:~$ cat /home/hadoop/.ssh/id_rsa.pub >> /home/hadoop/.ssh/authorized_keys
hadoop@u20-06:~$ chmod 0600 /home/hadoop/.ssh/authorized_keys
hadoop@u20-06:~$
```

4. Check if SSH is working

172.26.35.26 — Подключение к удаленному рабочему столу

Activities Terminal anp 21

BigDataAnalytic_Practic x Exercise_1.pdf +

hadoop@u20-06: ~

```
hadoop@u20-06:~$ chmod 0600 /home/hadoop/.ssh/authorized_keys
hadoop@u20-06:~$ ssh localhost
The authenticity of host 'localhost (127.0.0.1)' can't be established.
ECDSA key fingerprint is SHA256:f/ZgYnhEXHnec9aXvQ9uGrSDoaaCh3a4IuB7L+MDq04.
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
individual files in /usr/share/doc/*/copyright.

Ubuntu comes with ABSOLUTELY NO WARRANTY, to the extent permitted by
applicable law.

hadoop@u20-06:~$ exit
logout
Connection to localhost closed.
hadoop@u20-06:~$
```

4

Memory usage: 2% IP address

Activities Terminal

BigDataAnalytic_Practic x Exercice_1.pdf

hadoop@u20-06: ~

```
hadoop-3.1.2/share/doc/hadoop/images/icon_success_sml.gif
hadoop-3.1.2/share/doc/hadoop/images/icon_info_sml.gif
hadoop-3.1.2/share/doc/hadoop/images/breadcrumbs.jpg
hadoop-3.1.2/share/doc/hadoop/images/expanded.gif
hadoop-3.1.2/share/doc/hadoop/images/apache-maven-project-2.png
hadoop-3.1.2/share/doc/hadoop/images/icon_error_sml.gif
hadoop-3.1.2/share/doc/hadoop/images/logo_apache.jpg
hadoop-3.1.2/share/doc/hadoop/images/icon_warning_sml.gif
hadoop-3.1.2/share/doc/hadoop/images/newwindow.png
hadoop-3.1.2/share/doc/hadoop/images/maven-logo-2.gif
hadoop-3.1.2/share/doc/hadoop/images/banner.jpg
hadoop-3.1.2/share/doc/hadoop/images/logo_maven.jpg
hadoop-3.1.2/share/doc/hadoop/images/collapsed.gif
hadoop-3.1.2/share/doc/hadoop/images/external.png
hadoop-3.1.2/share/doc/hadoop/images/h5.jpg
hadoop-3.1.2/share/doc/hadoop/images/logos/
hadoop-3.1.2/share/doc/hadoop/images/logos/build-by-maven-white.png
hadoop-3.1.2/share/doc/hadoop/images/logos/build-by-maven-black.png
hadoop-3.1.2/share/doc/hadoop/images/logos/maven-feather.png
hadoop-3.1.2/share/doc/hadoop/index.html
hadoop-3.1.2/share/doc/hadoop/project-reports.html
hadoop-3.1.2/include/
hadoop-3.1.2/include/hdfs.h
hadoop-3.1.2/include/Pipes.hh
hadoop-3.1.2/include/SerialUtils.hh
hadoop-3.1.2/include/StringUtils.hh
hadoop-3.1.2/include/TemplateFactory.hh
hadoop@u20-06:~$ mv hadoop-3.1.2 hadoop
hadoop@u20-06:~$
```

4

tar -xvzf hadoop-3.1.2.tar.gz

Activities Terminal

BigDataAnalytic_Practic x Exercice_1.pdf

hadoop@u20-06: ~

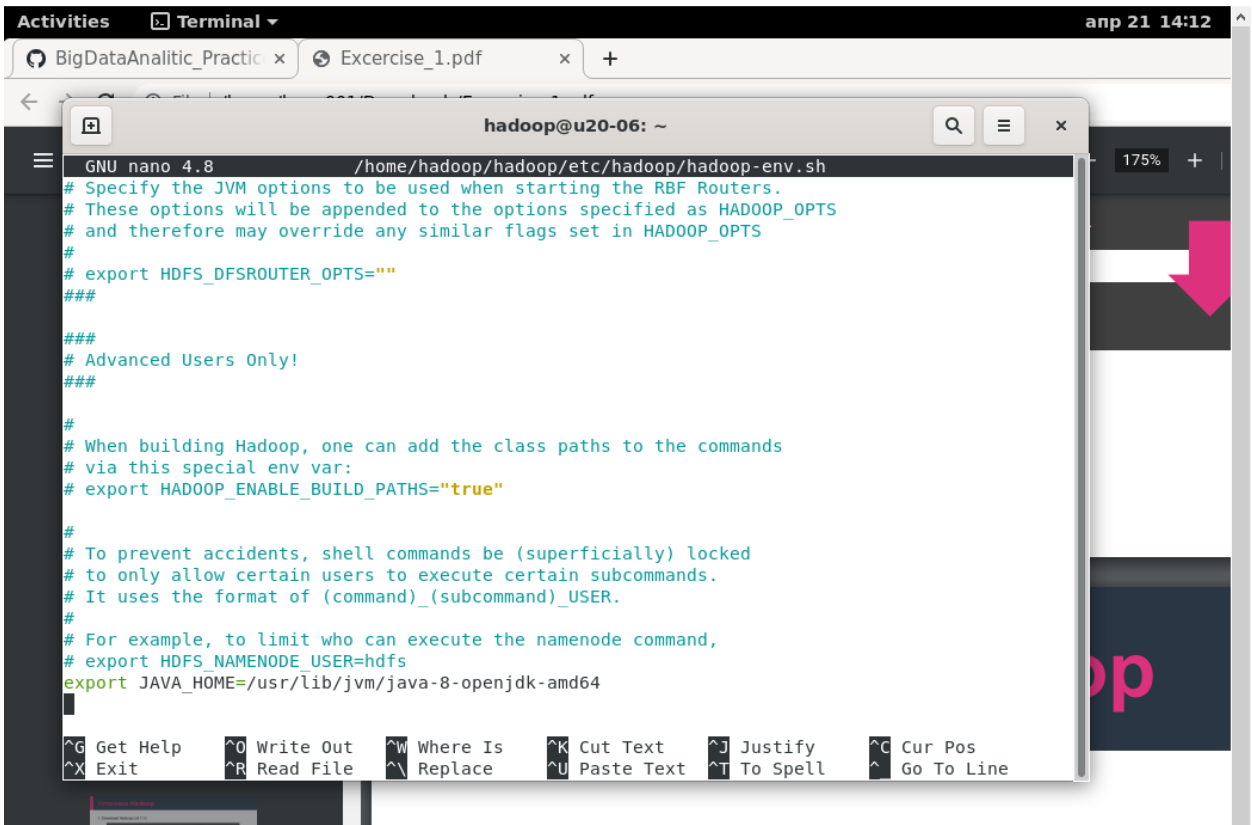
```
GNU nano 4.8 .bashrc Modified
# See /usr/share/doc/bash-doc/examples in the bash-doc package.

if [ -f ~/.bash_aliases ]; then
    . ~/.bash_aliases
fi

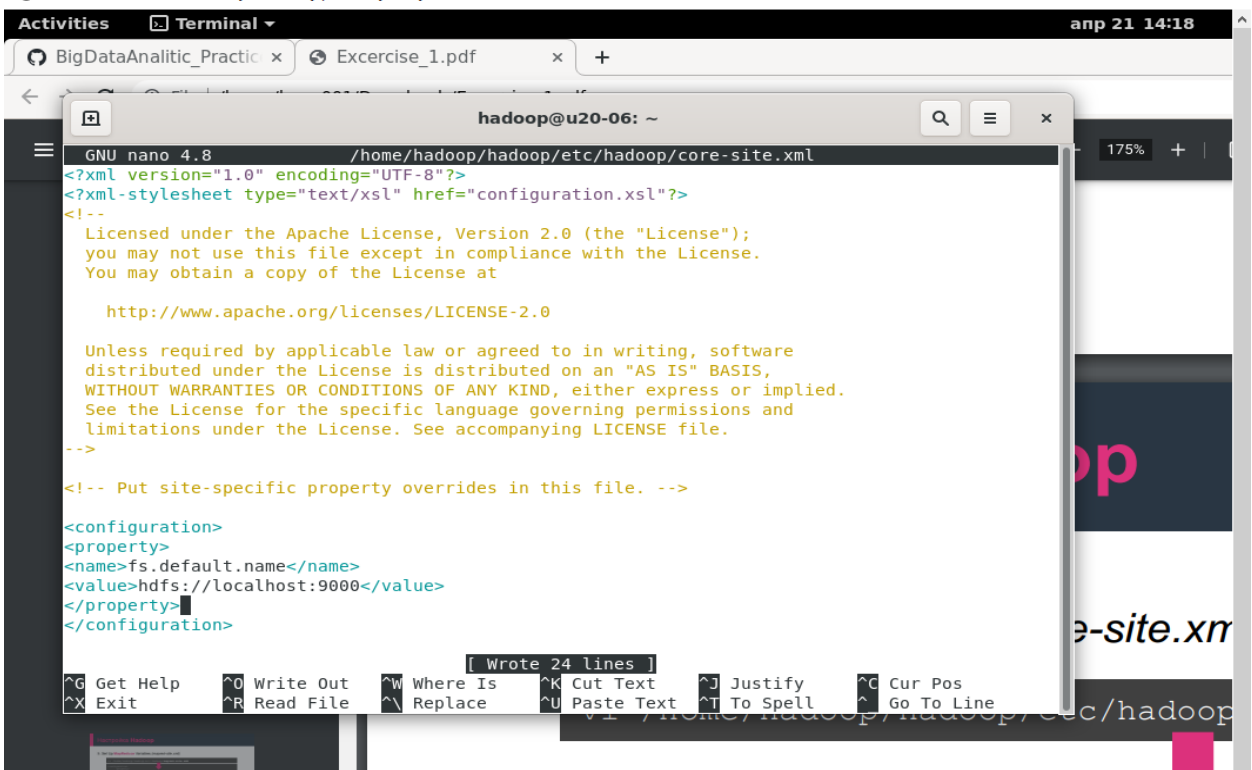
# enable programmable completion features (you don't need to enable
# this, if it's already enabled in /etc/bash.bashrc and /etc/profile
# sources /etc/bash.bashrc).
if ! shopt -oq posix; then
    if [ -f /usr/share/bash-completion/bash_completion ]; then
        . /usr/share/bash-completion/bash_completion
    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
^X Exit ^R Read File ^\ Replace ^U Paste Text ^T To Spell ^_ Go To Line
```

export YARN_HOME=\$HADOOP_HOME



```
GNU nano 4.8 /home/hadoop/hadoop/etc/hadoop/hadoop-env.sh
# Specify the JVM options to be used when starting the RBF Routers.
# These options will be appended to the options specified as HADOOP_OPTS
# and therefore may override any similar flags set in HADOOP_OPTS
#
# 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
```



```
GNU nano 4.8 /home/hadoop/hadoop/etc/hadoop/core-site.xml
<?xml version="1.0" encoding="UTF-8"?>
<?xml-stylesheet type="text/xsl" href="configuration.xsl"?>
<!--
Licensed under the Apache License, Version 2.0 (the "License");
you may not use this file except in compliance with the License.
You may obtain a copy of the License at

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

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

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

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

172.26.35.26 — Подключение к удаленному рабочему столу

Activities Terminal anp 21 14:20

BigDataAnalytic_Practic x Exercise_1.pdf x +

hadoop@u20-06: ~

GNU nano 4.8 /home/hadoop/hadoop/etc/hadoop/hdfs-site.xml

```
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>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>
```

[Wrote 32 lines]

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

<property>

172.26.35.26 — Подключение к удаленному рабочему столу

Activities Terminal anp 21 14:21

BigDataAnalytic_Practic x Exercise_1.pdf x +

hadoop@u20-06: ~

GNU nano 4.8 /home/hadoop/hadoop/etc/hadoop/mapred-site.xml

```
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>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>
```

[Wrote 36 lines]

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

<name>mapreduce.map.env</name>
<value>HADOOP_MAPRED_HOME=\${HADOOP_HOME}</value>


```
hadoop@u20-06: ~  
GNU nano 4.8 /home/hadoop/hadoop/etc/hadoop/yarn-site.xml  
<?xml version="1.0"?>  
<!--  
Licensed under the Apache License, Version 2.0 (the "License");  
you may not use this file except in compliance with the License.  
You may obtain a copy of the License at  
  
http://www.apache.org/licenses/LICENSE-2.0  
  
Unless required by applicable law or agreed to in writing, software  
distributed under the License is distributed on an "AS IS" BASIS,  
WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied.  
See the License for the specific language governing permissions and  
limitations under the License. See accompanying LICENSE file.  
-->  
<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>
```

172.26.35.26 — Подключение к удаленному рабочему столу

Activities Terminal anp 21 14:24

BigDataAnalytic_Practic x Exercise_1.pdf x +

```
hadoop@u20-06: ~  
2023-04-21 14:24:16,666 INFO util.GSet: Computing capacity for map cachedBlocks  
2023-04-21 14:24:16,667 INFO util.GSet: VM type = 64-bit  
2023-04-21 14:24:16,667 INFO util.GSet: 0.25% max memory 1.9 GB = 4.8 MB  
2023-04-21 14:24:16,667 INFO util.GSet: capacity = 2^19 = 524288 entries  
2023-04-21 14:24:16,677 INFO metrics.TopMetrics: NNTop conf: dfs.namenode.top.window.num.buckets = 10  
2023-04-21 14:24:16,677 INFO metrics.TopMetrics: NNTop conf: dfs.namenode.top.num.users = 10  
2023-04-21 14:24:16,677 INFO metrics.TopMetrics: NNTop conf: dfs.namenode.top.window.minutes = 1,5,25  
2023-04-21 14:24:16,710 INFO namenode.FSNamesystem: Retry cache on namenode is enabled  
2023-04-21 14:24:16,710 INFO namenode.FSNamesystem: Retry cache will use 0.03 of total heap and retry cache entry expiry time is 600000 millis  
2023-04-21 14:24:16,711 INFO util.GSet: Computing capacity for map NameNodeRetryCache  
2023-04-21 14:24:16,713 INFO util.GSet: VM type = 64-bit  
2023-04-21 14:24:16,715 INFO util.GSet: 0.0299999999329447746% max memory 1.9 GB = 584.4 KB  
2023-04-21 14:24:16,715 INFO util.GSet: capacity = 2^16 = 65536 entries  
2023-04-21 14:24:16,790 INFO namenode.FSImage: Allocated new BlockPoolId: BP-1412464720-172.26.35.26-1682076256761  
2023-04-21 14:24:16,819 INFO common.Storage: Storage directory /home/hadoop/hadoopdata/hdfs/namenode has been successfully formatted.  
2023-04-21 14:24:16,858 INFO namenode.FSImageFormatProtobuf: Saving image file /home/hadoop/hadoopdata/hdfs/namenode/current/fsimage.ckpt_000000000000000000 using no compression  
2023-04-21 14:24:17,021 INFO namenode.FSImageFormatProtobuf: Image file /home/hadoop/hadoopdata/hdfs/namenode/current/fsimage.ckpt_000000000000000000 of size 393 bytes saved in 0 seconds  
2023-04-21 14:24:17,046 INFO namenode.NNStorageRetentionManager: Going to retain 1 images with txid >= 0  
2023-04-21 14:24:17,062 INFO namenode.NameNode: SHUTDOWN MSG:  
/*****  
SHUTDOWN MSG: Shutting down NameNode at u20-06/172.26.35.26  
*****/  
hadoop@u20-06:~$
```

7 Clear HDFS


```

hadoop@u20-06: ~
2023-04-21 14:24:16,710 INFO namenode.FSNamesystem: Retry cache will use 0.03 of total heap and
d retry cache entry expiry time is 600000 millis
2023-04-21 14:24:16,711 INFO util.GSet: Computing capacity for map NameNodeRetryCache
2023-04-21 14:24:16,713 INFO util.GSet: VM type = 64-bit
2023-04-21 14:24:16,715 INFO util.GSet: 0.029999999329447746% max memory 1.9 GB = 584.4 KB
2023-04-21 14:24:16,715 INFO util.GSet: capacity = 2^16 = 65536 entries
2023-04-21 14:24:16,790 INFO namenode.FSImage: Allocated new BlockPoolId: BP-1412464720-172.26
.35.26-1682076256761
2023-04-21 14:24:16,819 INFO common.Storage: Storage directory /home/hadoop/hadoopdata/hdfs/na
menode has been successfully formatted.
2023-04-21 14:24:16,858 INFO namenode.FSImageFormatProtobuf: Saving image file /home/hadoop/ha
dooopdata/hdfs/namenode/current/fsimage.ckpt_000000000000000000 using no compression
2023-04-21 14:24:17,021 INFO namenode.FSImageFormatProtobuf: Image file /home/hadoop/hadoopdat
a/hdfs/namenode/current/fsimage.ckpt_000000000000000000 of size 393 bytes saved in 0 seconds
.
2023-04-21 14:24:17,046 INFO namenode.NNStorageRetentionManager: Going to retain 1 images with
.txid >= 0
2023-04-21 14:24:17,062 INFO namenode.NameNode: SHUTDOWN_MSG:
/*****
SHUTDOWN_MSG: Shutting down NameNode at u20-06/172.26.35.26
*****/
hadoop@u20-06:~$ start-dfs.sh
Starting namenodes on [localhost]
Starting datanodes
Starting secondary namenodes [u20-06]
u20-06: Warning: Permanently added 'u20-06,172.26.35.26' (ECDSA) to the list of known hosts.
2023-04-21 14:25:26,641 WARN util.NativeCodeLoader: Unable to load native-hadoop library for y
our platform... using builtin-java classes where applicable
hadoop@u20-06:~$ start-yarn.sh
Starting resourcemanager
Starting nodemanagers
hadoop@u20-06:~$

```

```

hadoop@u20-06: ~
Erasure Pending deletion blocks: 0
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: u20-06
Decommission Status : Normal
Configured Capacity: 52044496896 (48.47 GB)
DFS Used: 24576 (24 KB)
Non DFS Used: 14557724672 (13.56 GB)
DFS Remaining: 34812616704 (32.42 GB)
DFS Used%: 0.00%
DFS Remaining%: 66.89%
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 14:27:25 MSK 2023
Last Block Report: Fri Apr 21 14:25:22 MSK 2023
Num of Blocks: 0

hadoop@u20-06:~$

```

```
hadoop@u20-06: ~  
Name: 127.0.0.1:9866 (localhost)  
Hostname: u20-06  
Decommission Status : Normal  
Configured Capacity: 52044496896 (48.47 GB)  
DFS Used: 24576 (24 KB)  
Non DFS Used: 14557724672 (13.56 GB)  
DFS Remaining: 34812616704 (32.42 GB)  
DFS Used%: 0.00%  
DFS Remaining%: 66.89%  
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 14:27:25 MSK 2023  
Last Block Report: Fri Apr 21 14:25:22 MSK 2023  
Num of Blocks: 0  
  
hadoop@u20-06:~$ hadoop fs -mkdir /user  
2023-04-21 14:34:48,869 WARN util.NativeCodeLoader: Unable to load native-hadoop library for y  
our platform... using builtin-java classes where applicable  
hadoop@u20-06:~$ hadoop fs -mkdir /user/hadoop  
2023-04-21 14:35:08,101 WARN util.NativeCodeLoader: Unable to load native-hadoop library for y  
our platform... using builtin-java classes where applicable  
hadoop@u20-06:~$ hadoop fs -put /var/log/dpkg.log /user/hadoop/dpkg.log  
2023-04-21 14:37:08,752 WARN util.NativeCodeLoader: Unable to load native-hadoop library for y  
our platform... using builtin-java classes where applicable
```

Browse Directory

/user/hadoop

Go!

Show

25

entries

Search:

| <input type="checkbox"/> | <div><div></div></div> Permission | <div><div></div></div> Owner | <div><div></div></div> Group | <div><div></div></div> Size | <div><div></div></div> Last Modified | <div><div></div></div> Replication | <div><div></div></div> Block Size | <div><div></div></div> Name | <div><div></div></div> |
|--------------------------|-----------------------------------|------------------------------|------------------------------|-----------------------------|--------------------------------------|------------------------------------|-----------------------------------|-----------------------------|------------------------|
| <input type="checkbox"/> | -rw-r--r-- | hadoop | supergroup | 315.94 KB | Apr 21 14:37 | 1 | 128 MB | dpkg.log | <div><div></div></div> |

Showing 1 to 1 of 1 entries

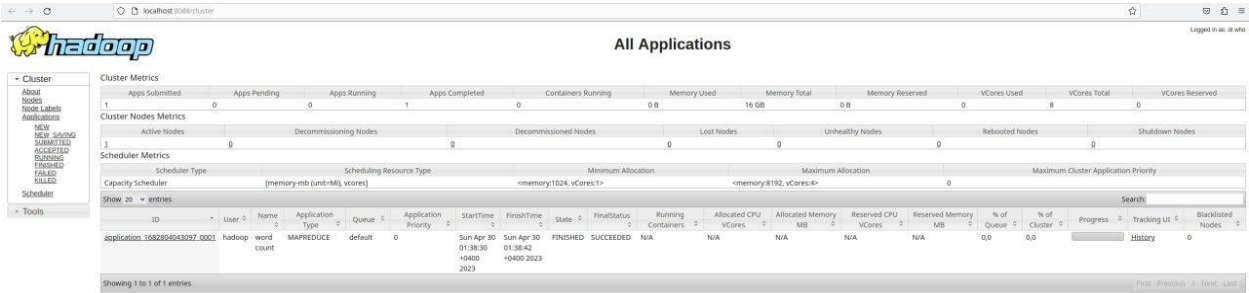
Previous

1

Next

```
lemp001@u20-06: ~  
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
```

Задание дома



```
File System Counters
  FILE: Number of bytes read=79326
  FILE: Number of bytes written=590691
  FILE: Number of read operations=0
  FILE: Number of large read operations=0
  FILE: Number of write operations=0
  HDFS: Number of bytes read=1184444
  HDFS: Number of bytes written=66792
  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
  Total time spent by all maps in occupied slots (ms)=1551
  Total time spent by all reduces in occupied slots (ms)=1443
  Total time spent by all map tasks (ms)=1551
  Total time spent by all reduce tasks (ms)=1443
  Total vcore-milliseconds taken by all map tasks=1551
  Total vcore-milliseconds taken by all reduce tasks=1443
  Total megabyte-milliseconds taken by all map tasks=1588224
  Total megabyte-milliseconds taken by all reduce tasks=1477632

Map-Reduce Framework
  Map input records=16100
  Map output records=96482
  Map output bytes=1570265
  Map output materialized bytes=79326
  Input split bytes=107
  Combine input records=96482
  Combine output records=3350
  Reduce input groups=3350
  Reduce shuffle bytes=79326
  Reduce input records=3350
  Reduce output records=3350
  Spilled Records=6700
  Shuffled Maps =1
  Failed Shuffles=0
  Merged Map outputs=1
  GC time elapsed (ms)=52
  CPU time spent (ms)=1550
  Physical memory (bytes) snapshot=505683068
  Virtual memory (bytes) snapshot=5156057088
  Total committed heap usage (bytes)=522190848
  Peak Map Physical memory (bytes)=311803904
  Peak Map Virtual memory (bytes)=2576543744
  Peak Reduce Physical memory (bytes)=193880064
  Peak Reduce Virtual memory (bytes)=2579513344

Shuffle Errors
  BAD_ID=0
  CONNECTION=0
  IO_ERROR=0
  WRONG_LENGTH=0
  WRONG_MAP=0
  WRONG_REDUCE=0

File Input Format Counters
  Bytes Read=1184337

File Output Format Counters
  Bytes Written=66792
```

| <div> <div></div> <div> <div>Файл</div> <div>Правка</div> <div>Вид</div> <div>Закладки</div> <div>Модули</div> <div>Настройка</div> <div>Справка</div> </div> </div> | | | |
|--|----|--|--|
| vlc-plugin-access-extra:amd64 | 7 | | |
| vlc-plugin-base:amd64 | 7 | | |
| vlc-plugin-notify:amd64 | 7 | | |
| vlc-plugin-qt:amd64 | 7 | | |
| vlc-plugin-samba:amd64 | 7 | | |
| vlc-plugin-skins2:amd64 | 7 | | |
| vlc-plugin-video-output:amd64 | 7 | | |
| vlc-plugin-video-splitter:amd64 | 7 | | |
| vlc-plugin-visualization:amd64 | 7 | | |
| vlc:amd64 | 7 | | |
| vulkan-tools:amd64 | 7 | | |
| wamerican:all | 7 | | |
| wbritish:all | 7 | | |
| wget:amd64 | 7 | | |
| whiptail:amd64 | 7 | | |
| whoopsie-preferences:amd64 | 7 | | |
| whoopsie:amd64 | 7 | | |
| wireless-regdb:all | 7 | | |
| wireless-tools:amd64 | 7 | | |
| wpasupplicant:amd64 | 7 | | |
| x11-apps:amd64 | 7 | | |
| x11-common:all | 7 | | |
| x11-session-utils:amd64 | 7 | | |
| x11-utils:amd64 | 7 | | |
| x11-xkb-utils:amd64 | 7 | | |
| x11-xserver-utils:amd64 | 7 | | |
| x11proto-dev:all | 7 | | |
| xauth:amd64 | 7 | | |
| xbitmaps:all | 7 | | |
| xcvt:amd64 | 7 | | |
| xdg-desktop-portal-gtk:amd64 | 7 | | |
| xdg-desktop-portal-kde:amd64 | 7 | | |
| xdg-desktop-portal:amd64 | 7 | | |
| xdg-user-dirs:amd64 | 7 | | |
| xdg-utils:all | 7 | | |
| xfonts-base:all | 7 | | |
| xfonts-encodings:all | 7 | | |
| xfonts-scalable:all | 7 | | |
| xfonts-utils:amd64 | 7 | | |
| xfsprogs:amd64 | 7 | | |
| xinit:amd64 | 7 | | |
| xinput:amd64 | 7 | | |
| xkb-data:all | 7 | | |
| xml-core:all | 8 | | |
| xorg-docs-core:all | 7 | | |
| xorg-sgml-doctools:all | 8 | | |
| xorg:amd64 | 7 | | |
| xserver-common:all | 16 | | |
| xserver-xorg-core:amd64 | 16 | | |
| xserver-xorg-input-all:amd64 | 7 | | |
| xserver-xorg-input-libinput:amd64 | 7 | | |
| xserver-xorg-input-wacom:amd64 | 7 | | |
| xserver-xorg-legacy:amd64 | 16 | | |
| xserver-xorg-video-all:amd64 | 7 | | |
| xserver-xorg-video-amdgpu:amd64 | 7 | | |
| xserver-xorg-video-ati:amd64 | 7 | | |
| xserver-xorg-video-fbdev:amd64 | 7 | | |
| xserver-xorg-video-intel:amd64 | 7 | | |
| xserver-xorg-video-nouveau:amd64 | 7 | | |
| xserver-xorg-video-qxl:amd64 | 7 | | |
| xserver-xorg-video-radeon:amd64 | 7 | | |
| xserver-xorg-video-vesa:amd64 | 7 | | |
| xserver-xorg-video-vmware:amd64 | 7 | | |
| xserver-xorg:amd64 | 7 | | |
| xsettingsd:amd64 | 7 | | |
| xterm:amd64 | 7 | | |
| xtrans-dev:all | 7 | | |
| xxd:amd64 | 25 | | |
| xz-utils:amd64 | 7 | | |
| zip:amd64 | 7 | | |
| zlib1g:amd64 | 16 | | |
| zstd:amd64 | 7 | | |


```
hadoop hadoop-3.1.2.tar.gz hadoopdata
hadoop@ubuntu:~$ 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 | 6.90 MiB/s, done.
Resolving deltas: 100% (37/37), done.
hadoop@ubuntu:~$
```

```
hadoop@ubuntu:~$ hadoop fs -mkdir -p /user/hadoop/
2023-04-30 01:54:07,756 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform...
using builtin-java classes where applicable
```

```
hadoop@ubuntu:~$ hadoop jar hadoop/share/hadoop/mapreduce/hadoop-mapreduce-examples-3.1.2.jar wordcount /user/
hadoop/Faust_1.txt /user/hadoop/Faust_1_Output
2023-04-30 01:56:22,975 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform...
using builtin-java classes where applicable
2023-04-30 01:56:23,350 INFO client.RMProxy: Connecting to ResourceManager at /0.0.0.0:8032
2023-04-30 01:56:23,558 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yar
n/staging/hadoop/.staging/job_1682804043097_0002
2023-04-30 01:56:23,680 INFO input.FileInputFormat: Total input files to process : 1
2023-04-30 01:56:23,721 INFO mapreduce.JobSubmitter: number of splits:1
2023-04-30 01:56:23,795 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1682804043097_0002
2023-04-30 01:56:23,795 INFO mapreduce.JobSubmitter: Executing with tokens: []
2023-04-30 01:56:23,895 INFO conf.Configuration: resource-types.xml not found
2023-04-30 01:56:23,895 INFO resource.ResourceUtils: Unable to find 'resource-types.xml'.
2023-04-30 01:56:23,931 INFO impl.YarnClientImpl: Submitted application application_1682804043097_0002
2023-04-30 01:56:23,954 INFO mapreduce.Job: The url to track the job: http://ubuntu:8088/proxy/application_168
```

All Applications

| ID | Name | Application Type | Queue | Application Priority | Start Time | Finish Time | State | Final Status | Running Containers | Allocated CPU | Allocated Memory | Reserved CPU | Reserved Memory | % of Queue | % of Cluster | Progress | Tracking UI | Blacklisted Nodes |
|--------------------------------|-------------------|------------------|---------|----------------------|--------------------------------|--------------------------------|----------|--------------|--------------------|---------------|------------------|--------------|-----------------|------------|--------------|----------|-------------|-------------------|
| application_1682804043097_0002 | hadoop word count | MAPREDUCE | default | 0 | Sun Apr 30 01:56:23 +0400 2023 | Sun Apr 30 01:56:34 +0400 2023 | FINISHED | SUCCEEDED | N/A | N/A | N/A | N/A | N/A | 0.0 | 0.0 | History | 0 | |
| application_1682804043097_0001 | hadoop word count | MAPREDUCE | default | 0 | Sun Apr 30 01:38:30 +0400 2023 | Sun Apr 30 01:38:42 +0400 2023 | FINISHED | SUCCEEDED | N/A | N/A | N/A | N/A | N/A | 0.0 | 0.0 | History | 0 | |

```
hadoop@ubuntu:~$ hadoop fs -get /user/hadoop/Faust_1_Output/part-r-000000 Faust_1_Output.csv
2023-04-30 01:58:52,653 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform...
using builtin-java classes where applicable
```

```
hadoop@ubuntu:~$ 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
```

```
hadoop@ubuntu:~$ hadoop jar hadoop/share/hadoop/mapreduce/hadoop-mapreduce-examples-3.1.2.jar grep /user/hadoop
p/Faust_1.txt /user/hadoop/Faust_1_Count_Output 'Faust'
2023-04-30 02:00:38,402 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform...
using builtin-java classes where applicable
2023-04-30 02:00:38,777 INFO client.RMProxy: Connecting to ResourceManager at /0.0.0.0:8032
2023-04-30 02:00:39,002 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yar
n/staging/hadoop/.staging/job_1682804043097_0003
2023-04-30 02:00:39,134 INFO input.FileInputFormat: Total input files to process : 1
2023-04-30 02:00:39,576 INFO mapreduce.JobSubmitter: number of splits:1
2023-04-30 02:00:40,054 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1682804043097_0003
2023-04-30 02:00:40,055 INFO mapreduce.JobSubmitter: Executing with tokens: []
2023-04-30 02:00:40,155 INFO conf.Configuration: resource-types.xml not found
2023-04-30 02:00:40,155 INFO resource.ResourceUtils: Unable to find 'resource-types.xml'.
2023-04-30 02:00:40,191 INFO impl.YarnClientImpl: Submitted application application_1682804043097_0003
2023-04-30 02:00:40,218 INFO mapreduce.Job: The url to track the job: http://ubuntu:8088/proxy/application_168
2804043097_0003/
```

Cluster

Cluster Metrics

Cluster Nodes Metrics

Scheduler Metrics

Capacity Scheduler

Tools

hadoop

All Applications

Cluster Metrics

Cluster Nodes Metrics

Scheduler Metrics

Capacity Scheduler

Tools

```
hadoop@ubuntu:~$ hadoop fs -get /user/hadoop/Faust_1_Count_Output/part-r-00000 Faust_1_Count_Output.csv
2023-04-30 02:03:41,739 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform...
using builtin-java classes where applicable
hadoop@ubuntu:~$ cat Faust_1_Count_Output.csv
50      Faust
```