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 05_1

Тема:

Introduction to HDFS.

Выполнила: Соколова М. С., группа: АДЭУ-201

Преподаватель: Босенко Т. М.

Москва

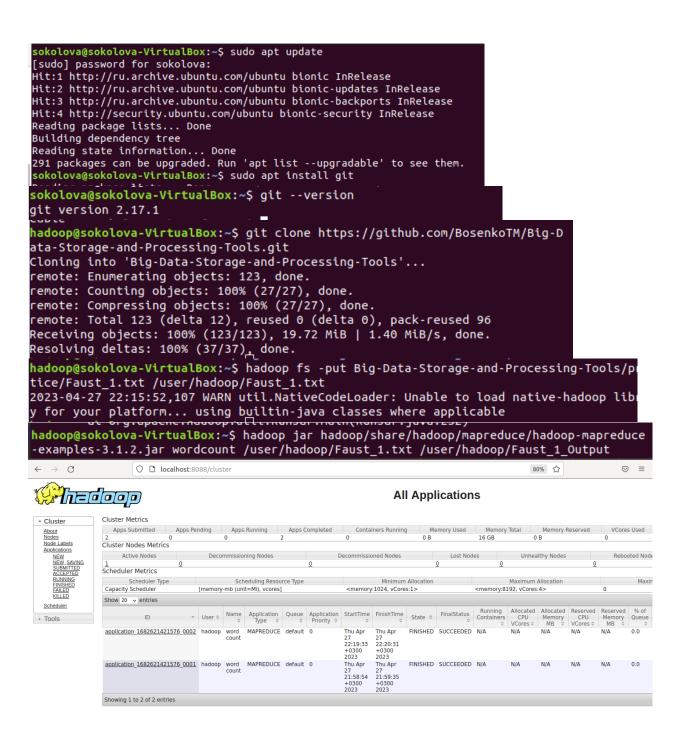
```
marina@marina-VirtualBox: ~
narina@marina-VirtualBox:~$ sudo apt-get install openjdk-8-jdk
marina@marina-VirtualBox:~$ java -version
openjdk version "1.8.0_362"
OpenJDK Runtime Environment (build 1.8.0_362-8u362-ga-0ubuntu1~18.04.1-b09)
OpenJDK 64-Bit Server VM (build 25.362-b09, mixed mode)
 marina@marina-VirtualBox:~$ sudo nano /etc/environment
GNU nano 2.9.3
                                /etc/environment
PATH="/usr/local/sbin:/usr/local/bin:/usr/sbin:/usr/bin:/sbin:/bin:/usr/ga$
JAVA_HOME="/usr/lib/jvm/java-8-openjdk-amd64"
JRE_HOME="/usr/lib/jvm/java-8-openjdk-amd64/jre"
marina@marina-VirtualBox:~$ 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'
Enter new UNIX password:
Retype new UNIX 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
marina@marina-VirtualBox:~$ sudo su hadoop
hadoop@marina-VirtualBox:/home/marina$ exit
exit
marina@marina-VirtualBox:~$
marina@marina-VirtualBox:~$ sudo apt-get install ssh pdsh
Reading package lists... Done
Building dependency tree
Reading state information... Done
The following additional packages will be installed:
   genders libgenders0 ncurses-term openssh-client openssh-server
   openssh-sftp-server ssh-import-id
Suggested packages:
   rdist keychain libpam-ssh monkeysphere ssh-askpass molly-guard rssh
```

```
marina@marina-VirtualBox:~$ sudo su hadoop
hadoop@marina-VirtualBox:/home/marina$ cd
hadoop@marina-VirtualBox:~$ 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:dJY/tN+D1WjHfTqnQg5bQfk+Jltm7qADDRRJ6t7zjmY hadoop@marina-VirtualBox
The key's randomart image is:
+---[RSA 2048]--
        .00
        .0 .0
      . ..0 0...00
       . So +oo.*
      . .. o ++X+.
       . o. *.X+oo
         E+0.+..+.|
        0..+. 00
  ---[SHA256]----+
hadoop@sokolova-VirtualBox:~$ cat /home/hadoop/.ssh/id_rsa.pub >> /home/ha
doop/.ssh/authorized_keys
hadoop@sokolova-VirtualBox:~$ chmod 0600 /home/hadoop/.ssh/authorized_keys
hadoop@marina-VirtualBox:~$ exit
выход
Connection to localhost closed.
hadoop@marina-VirtualBox:~$ wget https://archive.apache.org/dist/hadoop/com
mon/hadoop-3.1.2/hadoop-3.1.2.tar.gz
--2023-04-22 14:16:01-- https://archive.apache.org/dist/hadoop/common/hado
op-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'
in 3m 44s
2023-04-22 14:19:46 (1,41 MB/s) - 'hadoop-3.1.2.tar.gz' saved [332433589/33
2433589]
hadoop@marina-VirtualBox:~$
```

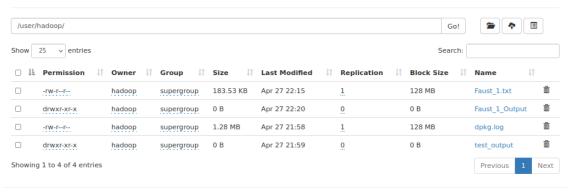
```
GNU nano 2.9.3
                                      bashrc
if ! shopt -oq posix; then
 if [ -f /usr/share/bash-completion/bash_completion ]; then
     /usr/share/bash-completion/bash_completion
 elif [ -f /etc/bash_completion ]; then
    . /etc/bash_completion
  fi
export HADOOP_HOME=/home/hadoop/hadoop
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:$HADOOP_HOME/bin
export PDSH RCMD TYPE=ssh
 GNU nano 2.9.3 /home/hadoop/hadoop/etc/hadoop/hadoop-env.sh
                                                                 Modified
# export HDFS DFSROUTER OPTS=""
# export HADOOP_ENABLE_BUILD_PATHS=<mark>"true"</mark>
export JAVA_HOME=/usr/lib/jvm/java-8-openjdk-amd64
                              ^W Where Is
^G Get Help
               ^O Write Out
                                              ^K Cut Text
                                                             ^J Justify
               ^R Read File
                                 Replace
                                              ^U Uncut Text
                                                             ^Т
^X Exit
                                                                To Linter
 GNU nano 2.9.3
                 /home/hadoop/hadoop/etc/hadoop/core-site.xml
<name>mapreduce.framework.name</name>
<value>yarn</value>
</property>
<name>yarn.app.mapreduce.am.env
<value>HADOOP MAPRED HOME=${HADOOP HOME}</value>
<name>mapreduce.map.env</name>
<value>HADOOP MAPRED HOME=${HADOOP_HOME}</value>
<name>mapreduce.reduce.env</name>
<value>HADOOP_MAPRED_HOME=${HADOOP_HOME}</value>
```

```
GNU nano 2.9.3 /home/hadoop/hadoop/etc/hadoop/yarn-site.xml
  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 -->
<name>yarn.nodemanager.aux-services</name>
<value>mapreduce_shuffle</value>
<name>varn.nodemanager.resource.memory-mb</name>
<value>16384</value>
*************************************
hadoop@marina-VirtualBox:~$ start-dfs.sh
Starting namenodes on [localhost]
Starting datanodes
Starting secondary namenodes [marina-VirtualBox]
2023-04-22 14:42:49,573 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your
platform... using builtin-java classes where applicable hadoop@marina-VirtualBox:~$ start-yarn.sh
Starting resourcemanager
resourcemanager is running as process 12627. Stop it first.
Starting nodemanagers
localhost: nodemanager is running as process 12768. Stop it first.
pdsh@marina-VirtualBox: localhost: ssh exited with exit code 1
hadoop@marina-VirtualBox:~$ hdfs dfsadmin -report
2023-04-22 14:43:39,308 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your
platform... using builtin-java classes where applicable
Configured Capacity: 10499674112 (9.78 GB)
Present Capacity: 2112827392 (1.97 GB)
DFS Remaining: 2112802816 (1.97 GB)
DFS Used: 24576 (24 KB)
DFS Used%: 0.00%
                       -----
Live datanodes (1):
Name: 127.0.0.1:9866 (localhost)
Hostname: marina-VirtualBox
Decommission Status : Normal
Configured Capacity: 10499674112 (9.78 GB)
DFS Used: 24576 (24 KB)
Non DFS Used: 7833305088 (7.30 GB)
DFS Remaining: 2112802816 (1.97 GB)
DFS Used%: 0.00%
DFS Remaining%: 20.12%
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: Sat Apr 22 14:43:40 MSK 2023
Last Block Report: Sat Apr 22 14:42:46 MSK 2023
Num of Blocks: 0
```

```
hadoop@marina-VirtualBox:~$ hadoop fs -mkdir /user
2023-04-22 14:44:46,507 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable hadoop@marina-VirtualBox:~$ hadoop fs -mkdir /user/hadoop
2023-04-22 14:45:02,502 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable hadoop@marina-VirtualBox:~$ hadoop fs -ls /
2023-04-22 14:45:29,184 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your
 platform... using builtin-java classes where applicable
Found 1 items
drwxr-xr-x - hadoop supergroup 0 2023-04-22 14:45 /user
hadoop@marina-VirtualBox:~$ hadoop fs -put /var/log/dpkg.log /user/hadoop/dpkg.log
2023-04-22 14:45:48,958 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your
platform... using builtin-java classes where applicable
hadoop@marina-VirtualBox:~$ hadoop jar hadoop/share/hadoop/mapreduce/hadoop-mapreduce-examples-3.
1.2.jar wordcount /user/hadoop/dpkg.log /user/hadoop/test_output
2023-04-22 14:47:07,174 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your
 platform... using builtin-java classes where applicable
  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=1340963
  File Output Format Counters
              Bytes Written=63683
```



Browse Directory



```
hadoop@sokolova-VirtualBox:~$ hadoop fs -get /user/hadoop/Faust_1_Output/part-
r-00000 Faust 1 Output.csv
2023-04-27 22:21:42,293 WARN util.NativeCodeLoader: Unable to load native-hado
op library for your platform... using builtin-java classes where applicable
hadoop@sokolova-VirtualBox:~$ head -10 Faust_1_Output.csv
"Allein,
'Alles 1
'Als
          1
 'Der
          1
 'Die
          2
 'Er
 Ich
          4
 Ιm
'Mein
 'Nur
          1
hadoop@sokolova-VirtualBox:~$
  Open ▼
             Æ
"Allein,
                     1
 'Alles 1
 "Als
          1
"Der
          1
"Die
          2
"Er
"Ich
"Im
          1
"Mein
"Nur
"Rette 1
"So
          1
"Uhu!
          1
"Wenn
          1
"Wie"
hadoop@sokolova-VirtualBox:~$ hadoop fs -put Big-Data-Storage-and-Processing-Tools/prac
tice/Faust_1.txt /user/hadoop/Faust_1.txt
2023-04-27 22:25:55,998 WARN util.NativeCodeLoader: Unable to load native-hadoop librar
y for your platform... using builtin-java classes where applicable
put: `/user/hadoop/Faust_1.txt': File exists
hadoop@sokolova-VirtualBox:~$ hadoop jar hadoop/share/hadoop/mapreduce/hadoop-mapreduce
examples-3.1.2.jar grep /user/hadoop/Faust_1.txt /user/hadoop/Faust_1_Count_Output 'Fa-
80% 🏠
                                                                                                        ⊘ ≡
ووواهالي
                                                                   All Applications

    Cluster

             Apps Submitted Apps Pending Apps Running Apps Completed Containers Running Memory Used Memory Total Memory Reserved VCores
4 0 1 3 1 2 GB 16 GB 0 B 1
            Cluster Nodes Metrics
             Active Nodes Decommissioning Nodes Decommissioned Nodes Lost Nodes Unhealthy Nodes Rebot
                                                  Minimum Allocation
<memory:1024, vCores:1>
                                                                       0
             Scheduler Metrics
  Scheduler
             → Tools
                                                       Thu Apr N/A
27
22:27:26
+0300
             application 1682621421576 0004 hadoop grep-sort MAPREDUCE default 0
                                                       22:27:26
+03:00
2023
Thu Apr FINISHED SUCCEEDED N/A N/A
27
27
27
22:26:57
22:26:57
2023
2023
2023
                                                                                                         0.0
             application_1682621421576_0003 hadoop grep- MAPREDUCE default 0 search
                                                                                                N/A N/A
                                                        Thu Apr Thu Apr 27 27 22:19:33 22:20:31 +0300 +0300
             application_1682621421576_0002 hadoop word MAPREDUCE default 0 count
                                                                   FINISHED SUCCEEDED N/A
                                                                                                N/A
                                                                                                           0.0
                                                       .. N/A 0.0

2023

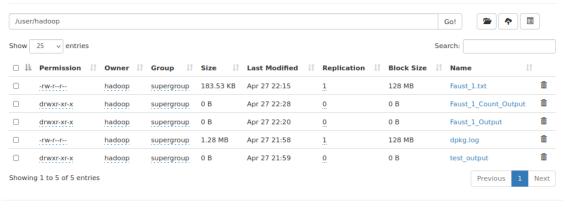
Thu Apr Thu Apr FINISHED SUCCEEDED N/A N/A N/A N/A N/A 0.0
21:58:54 21:59:35
+0300 +0300
2023 2023
             application_1682621421576_0001 hadoop word count MAPREDUCE default 0
```

Showing 1 to 4 of 4 entries

hadoop@sokolova-VirtualBox:~\$ hadoop fs -get /user/hadoop/Faust_1_Count_Output/part-r-0
0000 Faust_1_Count_Output.csv
2023-04-27 22:29:23,655 WARN util.NativeCodeLoader: Unable to load native-hadoop librar
y for your platform... using builtin-java classes where applicable
hadoop@sokolova-VirtualBox:~\$ cat Faust_1_Count_Output.csv
50 Faust



Browse Directory



Hadoop, 2018.