

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

Практическая работа № 6

Тема: «Hadoop»

Выполнила: Петешова С. В., ТП-191

Руководитель: Босенко Т. М.

Москва
2022

Установка и настройка JAVA

```
lemp001@u20-11: ~  
lemp001@u20-11:~$ sudo apt-get update  
[sudo] password for lemp001:  
Hit:1 http://ru.archive.ubuntu.com/ubuntu focal InRelease  
Get:2 http://ru.archive.ubuntu.com/ubuntu focal-updates InRelease [114 kB]  
Hit:3 http://deb.anydesk.com all InRelease  
Get:4 http://ru.archive.ubuntu.com/ubuntu focal-backports InRelease [108 kB]  
Hit:5 https://dl.google.com/linux/chrome/deb stable InRelease  
Hit:6 http://security.ubuntu.com/ubuntu focal-security InRelease  
Hit:7 https://download.docker.com/linux/ubuntu focal InRelease  
Fetched 222 kB in 1s (316 kB/s)  
Reading package lists... Done  
lemp001@u20-11:~$ sudo apt-get install openjdk-8-jdk  
Reading package lists... Done  
Building dependency tree  
Reading state information... Done  
The following packages were automatically installed and are no longer required:  
  chromium-codecs-ffmpeg-extra gstreamer1.0-vaapi libfwupdplugin1  
  libgstreamer-plugins-bad1.0-0 libva-wayland2  
Use 'sudo apt autoremove' to remove them.  
The following additional packages will be installed:  
  ca-certificates-java fonts-dejavu-extra java-common libatk-wrapper-java  
  libatk-wrapper-java-jni libice-dev libpthread-stubs0-dev libsm-dev  
  libx11-dev libxau-dev libxcb1-dev libxdmcp-dev libxt-dev  
  openjdk-8-jdk-headless openjdk-8-jre openjdk-8-jre-headless  
lemp001@u20-11:~$ java -version  
openjdk version "1.8.0_342"  
OpenJDK Runtime Environment (build 1.8.0_342-8u342-b07-0ubuntu1~20.04-b07)  
OpenJDK 64-Bit Server VM (build 25.342-b07, mixed mode)  
lemp001@u20-11:~$  
  
lemp001@u20-11: /  
GNU nano 4.8 /etc/enviroment  
JAVA_HOME="/usr/lib/jvm/java-8-openjdk-amd64"  
JRE_HOME="/usr/lib/jvm/java-8-openjdk-amd64/jre"  
  
[ Read 2 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
```

Настройка пользователя Hadoop

```
lemp001@u20-11:/$ 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@u20-11:/$
```

```
lemp001@u20-11:/$ sudo passwd hadoop
New password:
Retype new password:
passwd: password updated successfully
lemp001@u20-11:/$ sudo su hadoop
hadoop@u20-11:/$ exit
exit
lemp001@u20-11:/$
```

Настройка SSH


```
lemp001@u20-11:/$ 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
```

```
hadoop@u20-11:~$ 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:NBV8RLUPHbP8Xmp6Q/64wl2bpcCojrmIejD0gWEmbok hadoop@u20-11
The key's randomart image is:
+---[RSA 3072]-----+
|      oo...o.o. |
|. +      +.   +.o |
|* +      .+.   + .|
|E* .    o .   . +.|
|o . .   S     + o|
| o .      .+   .|
|  o      +.o+.o|
|   .. . o . =..B |
| .o. . +oo  o=..|
+-----[SHA256]-----+
hadoop@u20-11:~$
```

```
hadoop@u20-11:~$ cat /home/hadoop/.ssh/id_rsa.pub >> /home/hadoop/.ssh/authorized_keys
hadoop@u20-11:~$ chmod 0600 /home/hadoop/.ssh/authorized_keys
```

```
hadoop@u20-11:~$ ssh localhost
The authenticity of host 'localhost (127.0.0.1)' can't be established.
ECDSA key fingerprint is SHA256:UMX/Ye1dB6sxbRJ+caL5IhYWW04n3FdzmYXFikL3geo.
Are you sure you want to continue connecting (yes/no/[fingerprint])? y
Please type 'yes', 'no' or the fingerprint: yes
Warning: Permanently added 'localhost' (ECDSA) to the list of known hosts.
Welcome to Ubuntu 20.04.4 LTS (GNU/Linux 5.13.0-51-generic x86_64)
```

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

```
219 updates can be applied immediately.
165 of these updates are standard security updates.
To see these additional updates run: apt list --upgradable
```

Your Hardware Enablement Stack (HWE) is supported until April 2025.

The programs included with the Ubuntu system are free software;
the exact distribution terms for each program are described in the
individual files in /usr/share/doc/*/copyright.

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

```
hadoop@u20-11:~$ exit
logout
Connection to localhost closed.
hadoop@u20-11:~$
```

Установка Hadoop

```
hadoop@u20-11:~$ wget https://archive.apache.org/dist/hadoop/common/hadoop-3.1.2/hadoop-3.1.2.tar.gz
--2022-11-01 10:15:45-- 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      100%[=====] 317,03M  1,87MB/s   in 2m 45s
```

```
2022-11-01 10:18:31 (1,92 MB/s) - 'hadoop-3.1.2.tar.gz' saved [332433589/332433589]
```

```
hadoop@u20-11:~$
```

```
hadoop-3.1.2/share/doc/hadoop/hadoop-yarn-project/images/collapsed.gif
hadoop-3.1.2/share/doc/hadoop/hadoop-yarn-project/images/external.png
hadoop-3.1.2/share/doc/hadoop/hadoop-yarn-project/images/h5.jpg
hadoop-3.1.2/share/doc/hadoop/hadoop-yarn-project/images/logos/
hadoop-3.1.2/share/doc/hadoop/hadoop-yarn-project/images/logos/build-by-maven-white.png
hadoop-3.1.2/share/doc/hadoop/hadoop-yarn-project/images/logos/build-by-maven-black.png
hadoop-3.1.2/share/doc/hadoop/hadoop-yarn-project/images/logos/maven-feather.png
hadoop-3.1.2/share/doc/hadoop/images/
hadoop-3.1.2/share/doc/hadoop/images/bg.jpg
hadoop-3.1.2/share/doc/hadoop/images/h3.jpg
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@u20-11:~$
```

```
hadoop@u20-11:~$ mv hadoop-3.1.2 hadoop
```

Настройка Hadoop

```
hadoop@u20-11:~$ nano .bashrc
hadoop@u20-11:~$ source .bashrc
```

```
export HADOOP_HOME=/home/hadoop/hadoop
export HADOOP_INSTALL=$HADOOP_HOME
export HADOOP_MAPRED_HOME=$HADOOP_HOME
export HADOOP_COMMON_HOME=$HADOOP_HOME
export HADOOP_HDFS_HOME=$HADOOP_HOME
export YARN_HOME=$HADOOP_HOME
export HADOOP_COMMON_LIB_NATIVE_DIR=$HADOOP_HOME/lib/native
export PATH=$PATH:$HADOOP_HOME/sbin:$HADOOP_HOME/bin
export PDSH_RCMD_TYPE=ssh
```

```
hadoop@u20-11:~$ nano /home/hadoop/hadoop/etc/hadoop/hadoop-env.sh
#export JAVA_HOME=/usr/lib/jvm/java-8-openjdk-amd64
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>
```

```
<?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>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>
  <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>
</configuration>

```

[Read 39 lines]

```

<configuration>

  <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>

```

```

hadoop@u20-11:~$ hdfs namenode -format
WARNING: /home/hadoop/hadoop/logs does not exist. Creating.
2022-11-01 15:51:38,701 INFO namenode.NameNode: STARTUP_MSG:
/*****
STARTUP_MSG: Starting NameNode
STARTUP_MSG:   host = u20-11/172.26.35.131
STARTUP_MSG:   args = [-format]
STARTUP_MSG:   version = 3.1.2
STARTUP_MSG:   classpath = /home/hadoop/hadoop/etc/hadoop:/home/hadoop/hadoop/share/hadoop/common/lib/ne
tty-3.10.5.Final.jar:/home/hadoop/hadoop/share/hadoop/common/lib/log4j-1.2.17.jar:/home/hadoop/hadoop/sh
are/hadoop/common/lib/kerb-core-1.0.1.jar:/home/hadoop/hadoop/share/hadoop/common/lib/commons-logging-1.
1.3.jar:/home/hadoop/hadoop/share/hadoop/common/lib/jackson-core-asl-1.9.13.jar:/home/hadoop/hadoop/shar
e/hadoop/common/lib/audience-annotations-0.5.0.jar:/home/hadoop/hadoop/share/hadoop/common/lib/jersey-se
rver-1.19.jar:/home/hadoop/hadoop/share/hadoop/common/lib/woodstox-core-5.0.3.jar:/home/hadoop/hadoop/sh
are/hadoop/common/lib/kerb-crypto-1.0.1.jar:/home/hadoop/hadoop/share/hadoop/common/lib/kerb-util-1.0.1.
jar:/home/hadoop/hadoop/share/hadoop/common/lib/guava-11.0.2.jar:/home/hadoop/hadoop/share/hadoop/common
/lib/kerb-identity-1.0.1.jar:/home/hadoop/hadoop/share/hadoop/common/lib/paranamer-2.3.jar:/home/hadoop/
hadoop/share/hadoop/common/lib/jsr305-3.0.0.jar:/home/hadoop/hadoop/share/hadoop/common/lib/jetty-server
-9.3.24.v20180605.jar:/home/hadoop/hadoop/share/hadoop/common/lib/kerb-server-1.0.1.jar:/home/hadoop/had
oop/share/hadoop/common/lib/commons-beanutils-1.9.3.jar:/home/hadoop/hadoop/share/hadoop/common/lib/http
client-4.5.2.jar:/home/hadoop/hadoop/share/hadoop/common/lib/zookeeper-3.4.13.jar:/home/hadoop/hadoop/sh
are/hadoop/common/lib/commons-math3-3.1.1.jar:/home/hadoop/hadoop/share/hadoop/common/lib/jackson-databi
nd-2.7.8.jar:/home/hadoop/hadoop/share/hadoop/common/lib/commons-configuration2-2.1.1.jar:/home/hadoop/h
adoop/share/hadoop/common/lib/commons-compress-1.18.jar:/home/hadoop/hadoop/share/hadoop/common/lib/jack
son-core-2.7.8.jar:/home/hadoop/hadoop/share/hadoop/common/lib/htrace-core4-4.1.0-incubating.jar:/home/h
adoop/hadoop/share/hadoop/common/lib/jetty-security-9.3.24.v20180605.jar:/home/hadoop/hadoop/share/hadoo
p/common/lib/jsr311-api-1.1.1.jar:/home/hadoop/hadoop/share/hadoop/common/lib/javax.servlet-api-3.1.0.ja
r:/home/hadoop/hadoop/share/hadoop/common/lib/commons-net-3.6.jar:/home/hadoop/hadoop/share/hadoop/commo
n/lib/gson-2.2.4.jar:/home/hadoop/hadoop/share/hadoop/common/lib/re2j-1.1.jar:/home/hadoop/hadoop/share/
hadoop/common/lib/commons-lang3-3.4.jar:/home/hadoop/hadoop/share/hadoop/common/lib/json-smart-2.3.jar:/
home/hadoop/hadoop/share/hadoop/common/lib/jackson-annotations-2.7.8.jar:/home/hadoop/hadoop/share/hadoo
p/common/lib/jersey-json-1.19.jar:/home/hadoop/hadoop/share/hadoop/common/lib curator-framework-2.13.0.j
ar:/home/hadoop/hadoop/share/hadoop/common/lib/jsch-0.1.54.jar:/home/hadoop/hadoop/share/hadoop/common/l
ib/kerb-identity-1.0.1.jar:/home/hadoop/hadoop/share/hadoop/common/lib/metrics-core-3.2.4.jar:/home/hadoop

```



```

hadoop@u20-11:~$ start-dfs.sh
Starting namenodes on [localhost]
Starting datanodes
Starting secondary namenodes [u20-11]
u20-11: Warning: Permanently added 'u20-11,172.26.35.131' (ECDSA) to the list of known hosts.
2022-11-01 15:54:14,280 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platfo
rm... using builtin-java classes where applicable

```

```

hadoop@u20-11:~$ start-yarn.sh
Starting resourcemanager
Starting nodemanagers

```

Проверка Hadoop/HDFS

```

hadoop@u20-11:~$ hdfs dfsadmin -report
2022-11-01 15:57:55,389 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platfo
rm... using builtin-java classes where applicable
Configured Capacity: 52044496896 (48.47 GB)
Present Capacity: 36063883264 (33.59 GB)
DFS Remaining: 36063858688 (33.59 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: u20-11
Decommission Status : Normal
Configured Capacity: 52044496896 (48.47 GB)
DFS Used: 24576 (24 KB)
Non DFS Used: 13306482688 (12.39 GB)
DFS Remaining: 36063858688 (33.59 GB)
DFS Used%: 0.00%
DFS Remaining%: 69.29%
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 Nov 01 15:57:55 MSK 2022
Last Block Report: Tue Nov 01 15:54:10 MSK 2022
Num of Blocks: 0

```


hadoop

All Applications
Logged in as: db-who

Cluster

About NodesNode LabelsApplications

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	10 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 V-Cores	Allocated Memory MB	Reserved CPU V-Cores	Reserved Memory MB	% of Queue	% of Cluster	Progress	Tracking UI	Blacklisted Nodes
No data available in table																			

Showing 0 to 0 of 0 entries

First Previous Next Last

Activities

Google Chrome

HOR 1 16:02

Namenode information

+

localhost:9870/dfshealth.html#tab-overview

Update

Hadoop

Overview

Datanodes

Datanode Volume Failures

Snapshot

Startup Progress

Utilities

Overview 'localhost:9000' (active)

Started:	Tue Nov 01 15:54:02 +0300 2022
Version:	3.1.2, r1019ddcf0b0c12e05e48ac71e84500e589e5d5a
Compiled:	Tue Jan 29 04:39:00 +0300 2019 by sunlig from branch-3.1.2
Cluster ID:	CID-46444e98-a785-4f56-8c16-3ce7696a40ee
Block Pool ID:	BP-1891624118-172.26.35.131-1667307100278

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 48.41 MB of 119.94 MB Heap Memory. Max Heap Memory is 1.86 GB.
Non Heap Memory used 48.13 MB of 48.9 MB Committed Non Heap Memory. Max Non Heap Memory is <unbounded>.

Configured Capacity:	48.47 GB
Configured Remote Capacity:	0 B
DFS Used:	24 KB (0%)
Non DFS Used:	12.33 GB
DFS Remaining:	33.65 GB (69.42%)
Block Pool Used:	24 KB (0%)
DataNodes usages% (Min/Median/Max/stdDev):	0.00% / 0.00% / 0.00% / 0.00%
Live Nodes	1 (Decommissioned: 0, In Maintenance: 0)
Dead Nodes	0 (Decommissioned: 0, In Maintenance: 0)
Decommissioning Nodes	0
Entering Maintenance Nodes	0
Total Datanode Volume Failures	0 (0 B)
Number of Under-Replicated Blocks	0
Number of Blocks Pending Deletion	0
Block Deletion Start Time	Tue Nov 01 15:54:02 +0300 2022
Last Checkpoint Time	Tue Nov 01 15:55:14 +0300 2022

NameNode Journal Status

Datanode Information

 In service  Down  Decommissioned  Decommissioned & dead  In Maintenance  In Maintenance & dead

Datanode usage histogram



In operation

Show entries

Search:

Node	Http Address	Last contact	Last Block Report	Capacity	Blocks	Block pool used	Version
 rs20-11-9866 (127.0.0.1:9866)	http://rs20-11-9866:4	2s	24m	48.47 GB <div><div></div></div>	0	28 KB (0%)	3.1.2

Showing 1 to 1 of 1 entries

Previous **1** Next

Entering Maintenance

No nodes are entering maintenance.

Decommissioning

No nodes are decommissioning.

Hadoop, 2018.

Hadoop

Overview

Datanodes




Datanode Volume Failures

Snapshot

Startup Progress









Utilties

Browse Directory

Show entries

Search:

☐  Permission  Owner  Group  Size  Last Modified  Replication  Block Size  Name

No data available in table

Showing 0 to 0 of 0 entries

Previous Next

Hadoop, 2018.

Работа в HDFS

```
hadoop@u20-11:~$ hadoop fs -mkdir /user
2022-11-01 16:09:03,679 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platfo
rm... using builtin-java classes where applicable
hadoop@u20-11:~$ hadoop fs -mkdir /user/hadoop
2022-11-01 16:09:43,400 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platfo
rm... using builtin-java classes where applicable
hadoop@u20-11:~$
```

```
Found 1 items
drwxr-xr-x  - hadoop supergroup          0 2022-11-01 16:09 /user
hadoop@u20-11:~$
```

Запуск примера

```
hadoop@u20-11:~$ hadoop jar hadoop/share/hadoop/mapreduce/hadoop-mapreduce-examples-3.1.2.jar wordcount
/user/hadoop/dpkg.log /user/hadoop/test_output
2022-11-01 16:35:53,563 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
2022-11-01 16:35:54,428 INFO client.RMProxy: Connecting to ResourceManager at /0.0.0.0:8032
2022-11-01 16:35:54,957 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/hadoop/.staging/job_1667307395011_0001
2022-11-01 16:35:55,296 INFO input.FileInputFormat: Total input files to process : 1
2022-11-01 16:35:55,427 INFO mapreduce.JobSubmitter: number of splits:1
2022-11-01 16:35:55,663 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1667307395011_0001
2022-11-01 16:35:55,664 INFO mapreduce.JobSubmitter: Executing with tokens: []
2022-11-01 16:35:55,926 INFO conf.Configuration: resource-types.xml not found
2022-11-01 16:35:55,927 INFO resource.ResourceUtils: Unable to find 'resource-types.xml'.
2022-11-01 16:35:56,265 INFO impl.YarnClientImpl: Submitted application application_1667307395011_0001
2022-11-01 16:35:56,348 INFO mapreduce.Job: The url to track the job: http://u20-11:8088/proxy/application_1667307395011_0001/
2022-11-01 16:35:56,351 INFO mapreduce.Job: Running job: job_1667307395011_0001
2022-11-01 16:36:07,516 INFO mapreduce.Job: Job job_1667307395011_0001 running in uber mode : false
2022-11-01 16:36:07,518 INFO mapreduce.Job:  map 0% reduce 0%
2022-11-01 16:36:12,600 INFO mapreduce.Job:  map 100% reduce 0%
2022-11-01 16:36:19,670 INFO mapreduce.Job:  map 100% reduce 100%
2022-11-01 16:36:19,687 INFO mapreduce.Job: Job job_1667307395011_0001 completed successfully
2022-11-01 16:36:19,830 INFO mapreduce.Job: Counters: 53
    File System Counters
        FILE: Number of bytes read=2834
        FILE: Number of bytes written=437715
        FILE: Number of read operations=0
        FILE: Number of large read operations=0
        FILE: Number of write operations=0
        HDFS: Number of bytes read=26253
        HDFS: Number of bytes written=2314
        HDFS: Number of read operations=8
        HDFS: Number of large read operations=0
        HDFS: Number of write operations=2
    Job Counters
```

[illegible]

```




hadoop@u20-11:~$ hadoop fs -cat /user/hadoop/test_output/part-r-00000
2022-11-01 16:39:03,638 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platfo
rm... using builtin-java classes where applicable
0.0.2-Subuntu1 7
0.17-2 8
0.24-1ubuntu3 8
0.36-6ubuntu1 4
0.37.1-1 14
0.4-1 7
0.72 7
0.9.12-1 7
03:58:03 2
03:58:04 3
03:58:08 6
03:58:13 6
03:58:20 5
03:58:21 6
03:58:24 3
04:03:15 4
04:03:16 10
04:03:20 10
04:03:22 4
04:06:35 10
04:06:36 5
04:06:37 12
04:06:39 6
04:06:40 10
09:21:32 7
09:21:35 21
09:21:36 22
09:21:37 19
09:21:38 1
09:21:39 35
09:21:40 4
09:21:42 33
09:21:46 4
09:21:47 41
1.14-2 7
1.20-2 7

```

Browse Directory



/

Go!



Show 25 entries

Search:

<input type="checkbox"/>	Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name	<input type="checkbox"/>
<input type="checkbox"/>	drwx-----	hadoop	supergroup	0 B	Nov 01 16:35	0	0 B	tmp	
<input type="checkbox"/>	drwxr-xr-x	hadoop	supergroup	0 B	Nov 01 16:09	0	0 B	user	

Showing 1 to 2 of 2 entries

Previous

1

Next

Hadop, 2018.

Hadoop, 2018.

Анна Каренина (Самостоятельное задание)

```
hadoop@u20-11:~$ git clone https://github.com/Svetlana16P/practicds1.git
Cloning into 'practicds1'...
remote: Enumerating objects: 56, done.
remote: Counting objects: 100% (56/56), done.
remote: Compressing objects: 100% (53/53), done.
remote: Total 56 (delta 14), reused 0 (delta 0), pack-reused 0
Unpacking objects: 100% (56/56), 5.00 MiB | 6.82 MiB/s, done.
hadoop@u20-11:~$
```

```
hadoop@u20-11:~$ git clone https://github.com/Svetlana16P/practicds2.git
Cloning into 'practicds2'...
remote: Enumerating objects: 59, done.
remote: Counting objects: 100% (59/59), done.
remote: Compressing objects: 100% (56/56), done.
remote: Total 59 (delta 14), reused 0 (delta 0), pack-reused 0
Unpacking objects: 100% (59/59), 5.00 MiB | 2.26 MiB/s, done.
hadoop@u20-11:~$ hadoop fs -put practicds2/AK.txt /user/hadoop/AK.txt
2022-11-07 23:25:29,405 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platfo
rm... using builtin-java classes where applicable
hadoop@u20-11:~$ hadoop jar hadoop/share/hadoop/mapreduce/hadoop-mapreduce-examples-3.1.2.jar wordcount
/user/hadoop/AK.txt /user/hadoop/AK_Output
2022-11-07 23:31:17,501 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platfo
rm... using builtin-java classes where applicable
2022-11-07 23:31:18,383 INFO client.RMProxy: Connecting to ResourceManager at /0.0.0.0:8032
2022-11-07 23:31:18,933 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoo
p-yarn/staging/hadoop/.staging/job_1667307395011_0006
2022-11-07 23:31:19,249 INFO input.FileInputFormat: Total input files to process : 1
2022-11-07 23:31:19,391 INFO mapreduce.JobSubmitter: number of splits:1
2022-11-07 23:31:19,640 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1667307395011_0006
2022-11-07 23:31:19,641 INFO mapreduce.JobSubmitter: Executing with tokens: []
2022-11-07 23:31:19,923 INFO conf.Configuration: resource-types.xml not found
2022-11-07 23:31:19,924 INFO resource.ResourceUtils: Unable to find 'resource-types.xml'.
2022-11-07 23:31:20,061 INFO impl.YarnClientImpl: Submitted application application_1667307395011_0006
2022-11-07 23:31:20,136 INFO mapreduce.Job: The url to track the job: http://u20-11:8088/proxy/applicati
on_1667307395011_0006/
2022-11-07 23:31:20,139 INFO mapreduce.Job: Running job: job_1667307395011_0006
2022-11-07 23:31:28,329 INFO mapreduce.Job: Job job_1667307395011_0006 running in uber mode : false
2022-11-07 23:31:28,331 INFO mapreduce.Job: map 0% reduce 0%
2022-11-07 23:31:35,557 INFO mapreduce.Job: map 100% reduce 0%
2022-11-07 23:31:41,610 INFO mapreduce.Job: map 100% reduce 100%
2022-11-07 23:31:41,632 INFO mapreduce.Job: Job job_1667307395011_0006 completed successfully
2022-11-07 23:31:41,800 INFO mapreduce.Job: Counters: 53
```



Application application_1667307395011_0006

Logged in as: drwho

Cluster

[About](#)
[Nodes](#)
[Node Labels](#)
[Applications](#)
[NEW](#)
[NEW SAVING](#)
[SUBMITTED](#)
[ACCEPTED](#)
[RUNNING](#)
[FINISHED](#)
[FAILED](#)
[KILLED](#)
[Scheduler](#)

Tools

Application Overview

User: **hadoop**
 Name: **word count**
 Application Type: **MAPREDUCE**
 Application Tags:
 Application Priority: **0** (Higher integer value indicates higher priority)
 YarnApplicationState: **FINISHED**
 Queue: **default**
 FinalStatus Reported by AM: **SUCCEEDED**
 Started: **Mon Nov 07 23:31:19 +0300 2022**
 Elapsed: **20sec**
 Tracking URL: **History**
 Log Aggregation Status: **DISABLED**
 Application Timeout (Remaining Time): **Unlimited**
 Diagnostics:
 Unmanaged Application: **false**
 Application Node Label expression: **<Not set>**
 AM container Node Label expression: **<DEFAULT_PARTITION>**

Application Metrics

Total Resource Preempted: **<memory:0, vCores:0>**
 Total Number of Non-AM Containers Preempted: **0**
 Total Number of AM Containers Preempted: **0**
 Resource Preempted from Current Attempt: **<memory:0, vCores:0>**
 Number of Non-AM Containers Preempted from Current Attempt: **0**
 Aggregate Resource Allocation: **63494 MB-seconds, 34 vcore-seconds**
 Aggregate Preempted Resource Allocation: **0 MB-seconds, 0 vcore-seconds**

Show 20 entries

Search:

Attempt ID	Started	Node	Logs	Nodes blacklisted by the app	Nodes blacklisted by the system
attempt_1667307395011_0006_000001	Mon Nov 7 23:31:20 +0300 2022	http://u20-11:8042	Logs	0	0

Showing 1 to 1 of 1 entries

First Previous 1 Next Last


```
Bytes Written=318561
hadoop@u20-11:~$ hadoop fs -get /user/hadoop/AK_Output/part-r-00000 AK_Output.csv
2022-11-07 23:35:02,896 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platfo
rm... using builtin-java classes where applicable
hadoop@u20-11:~$ had -10 AK_Output.csv
```

```
hadoop@u20-11:~$ head -10 AK_Output.csv
```

```
"      1
"'And    1
"'Blessed      1
"'Eh,      1
"'I        3
"'Joinest      1
"'Sergei      1
"'There 1
"'When, 1
"'You 1
```

```
hadoop@u20-11:~$
```

```
hadoop@u20-11:~$ hadoop jar hadoop/share/hadoop/mapreduce/hadoop-mapreduce-examples-3.1.2.jar grep /use
r/hadoop/AK.txt /user/hadoop/AK_Count_Output 'Anna'
2022-11-07 23:47:08,731 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platfo
rm... using builtin-java classes where applicable
2022-11-07 23:47:09,576 INFO client.RMProxy: Connecting to ResourceManager at /0.0.0.0:8032
2022-11-07 23:47:10,115 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding for path: /tmp/hado
op-yarn/staging/hadoop/.staging/job_1667307395011_0011
2022-11-07 23:47:10,455 INFO input.FileInputFormat: Total input files to process : 1
2022-11-07 23:47:10,590 INFO mapreduce.JobSubmitter: number of splits:1
2022-11-07 23:47:10,832 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1667307395011_0011
2022-11-07 23:47:10,833 INFO mapreduce.JobSubmitter: Executing with tokens: []
2022-11-07 23:47:11,091 INFO conf.Configuration: resource-types.xml not found
2022-11-07 23:47:11,092 INFO resource.ResourceUtils: Unable to find 'resource-types.xml'.
2022-11-07 23:47:11,232 INFO impl.YarnClientImpl: Submitted application application_1667307395011_0011
2022-11-07 23:47:11,389 INFO mapreduce.Job: The url to track the job: http://u20-11:8088/proxy/applicati
on_1667307395011_0011/
2022-11-07 23:47:11,390 INFO mapreduce.Job: Running job: job_1667307395011_0011
2022-11-07 23:47:18,510 INFO mapreduce.Job: Job job_1667307395011_0011 running in uber mode : false
2022-11-07 23:47:18,512 INFO mapreduce.Job:  map 0% reduce 0%
2022-11-07 23:47:23,598 INFO mapreduce.Job:  map 100% reduce 0%
hadoop@u20-11:~$ hadoop fs -get /user/hadoop/AK_Count_Output/part-r-00000 AK_Count_Output.csv
2022-11-07 23:48:35,650 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platfo
rm... using builtin-java classes where applicable
hadoop@u20-11:~$ cat AK_Count_Output.csv
827      Anna
hadoop@u20-11:~$
```