

Министерство науки и высшего образования Российской Федерации
Федеральное государственное автономное образовательное учреждение
высшего образования
УРАЛЬСКИЙ ФЕДЕРАЛЬНЫЙ УНИВЕРСИТЕТ
ИМЕНИ ПЕРВОГО ПРЕЗИДЕНТА РОССИИ Б.Н. ЕЛЬЦИНА
(УрФУ имени первого Президента России Б.Н. Ельцина)
Институт радиоэлектроники и информационных технологий — РТФ

ОТЧЁТ

по лабораторной работе №2

по дисциплине «Методы и инструменты анализа больших данных»

Преподаватель	_____	_____	С.Г. Мирвода
	(дата)	(подпись)	
Студент	_____	_____	А.М. Белоусов
	(дата)	(подпись)	
Студент	_____	_____	А.В. Жиденко
	(дата)	(подпись)	

Группа: РИМ-201211

Екатеринбург 2021

Цель работы: знакомство с распределённой файловой системой HDFS.

Задание 0

Задача: подготовить полигон.

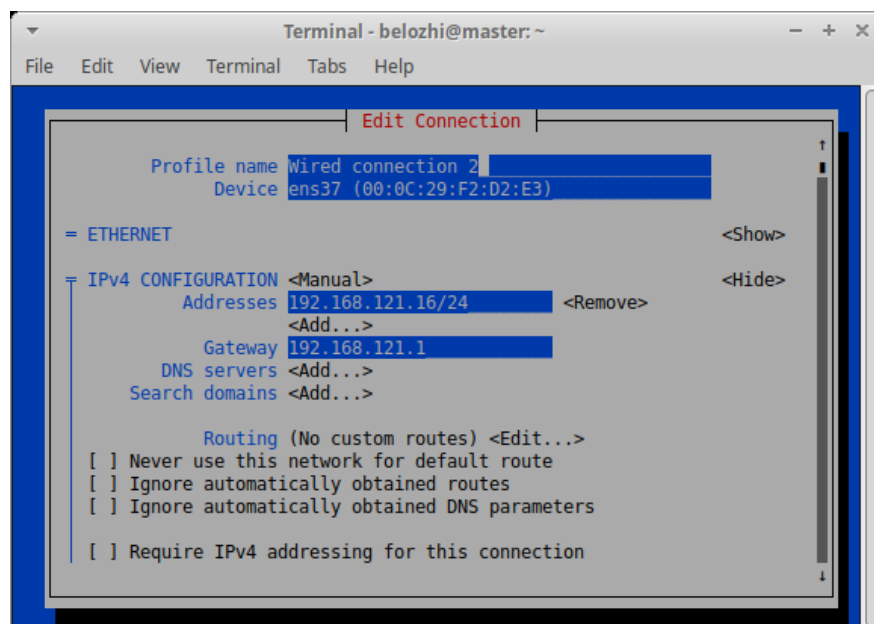
1. Установить и настроить кластер HDFS согласно инструкции Cluster и примеру

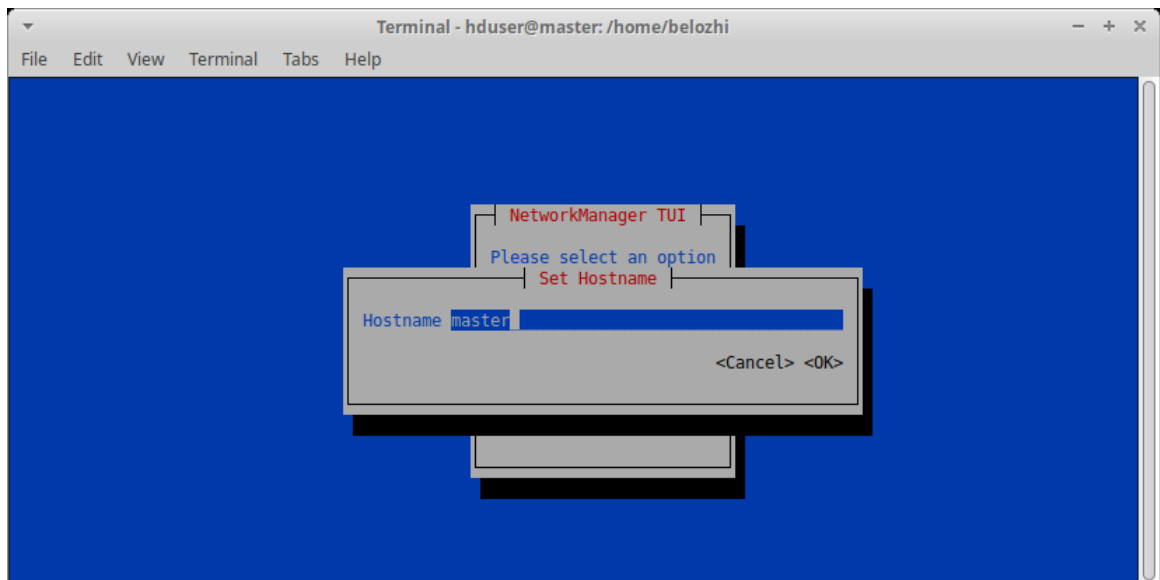
Настройка ip адреса согласно схеме задания



Настроим имя и ip-адрес VM Master с помощью утилиты nmtui.

Nmtui





```

Terminal - belozhi@master: ~
File Edit View Terminal Tabs Help

su: user hadoop does not exist
belozhi@master:~$ nmtui
belozhi@master:~$ ip addr
1: lo: <LOOPBACK,UP,LOWER_UP> mtu 65536 qdisc noqueue state UNKNOWN group default qlen 1000
    link/loopback 00:00:00:00:00:00 brd 00:00:00:00:00:00
    inet 127.0.0.1/8 scope host lo
        valid_lft forever preferred_lft forever
    inet6 ::1/128 scope host
        valid_lft forever preferred_lft forever
2: ens33: <BROADCAST,MULTICAST,UP,LOWER_UP> mtu 1500 qdisc fq_codel state UP group default qlen 1000
    link/ether 00:0c:29:f2:d2:d9 brd ff:ff:ff:ff:ff:ff
    altname enp2s1
    inet 192.168.128.128/24 brd 192.168.128.255 scope global dynamic noprefixroute ens33
        valid_lft 1338sec preferred_lft 1338sec
3: ens37: <BROADCAST,MULTICAST,UP,LOWER_UP> mtu 1500 qdisc fq_codel state UP group default qlen 1000
    link/ether 00:0c:29:f2:d2:e3 brd ff:ff:ff:ff:ff:ff
    altname enp2s5
    inet 192.168.121.16/24 brd 192.168.121.255 scope global noprefixroute ens37
        valid_lft forever preferred_lft forever
belozhi@master:~$

```

Установка ssh

sudo apt-get update && sudo apt-get upgrade

sudo apt-get install ssh

Установка Java

\$ sudo apt-get install openjdk-8-jdk

Создание отдельной учетной записи для запуска Hadoop

sudo addgroup hadoop

sudo adduser --ingroup hadoop hduser

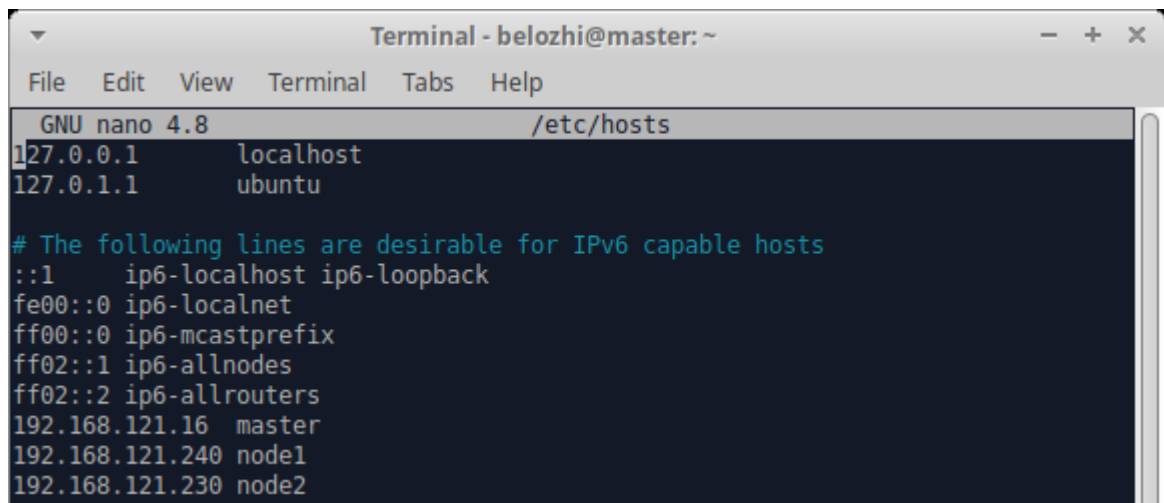
sudo usermod -aG sudo hduser

```

root@ubuntu:/home/belozhi# sudo adduser --ingroup hadoop hduser
Adding user `hduser' ...
Adding new user `hduser' (1001) with group `hadoop' ...
Creating home directory `/home/hduser' ...
Copying files from `/etc/skel' ...
New password:
Retype new password:
Sorry, passwords do not match.
passwd: Authentication token manipulation error
passwd: password unchanged
Try again? [y/N] y
New password:
Retype new password:
passwd: password updated successfully
Changing the user information for hduser
Enter the new value, or press ENTER for the default
    Full Name []: master
    Room Number []: 1
    Work Phone []:
    Home Phone []:
    Other []:
Is the information correct? [Y/n] y
root@ubuntu:/home/belozhi# sudo usermod -aG sudo hduser

```

Редактирование файла /etc/hosts



```

Terminal - belozhi@master: ~
File Edit View Terminal Tabs Help
GNU nano 4.8 /etc/hosts
1 127.0.0.1 localhost
2 127.0.1.1 ubuntu
# The following lines are desirable for IPv6 capable hosts
::1 ip6-localhost ip6-loopback
fe00::0 ip6-localnet
ff00::0 ip6-mcastprefix
ff02::1 ip6-allnodes
ff02::2 ip6-allrouters
192.168.121.16 master
192.168.121.240 node1
192.168.121.230 node2

```

Получение SSH ключей

```
Terminal - hduser@master: ~
File Edit View Terminal Tabs Help

hduser@master:/home/belozhi$ ssh-keygen -t rsa
Generating public/private rsa key pair.
Enter file in which to save the key (/home/hduser/.ssh/id_rsa):
Enter passphrase (empty for no passphrase):
Enter same passphrase again:
Your identification has been saved in /home/hduser/.ssh/id_rsa
Your public key has been saved in /home/hduser/.ssh/id_rsa.pub
The key fingerprint is:
SHA256:fWSbWGR33BTj7kDGX42btdk+GHzkj0GNckf3BIjKuXQ hduser@master
The key's randomart image is:
+---[RSA 3072]-----+
|      .0.0==|
|      .00..+=|
|      . 0  +=,+*|
|      =.E==0%=B|
|      .Soo.+0o@o|
|      .  .0 B. |
|      .  0. |
|      .  . |
+---[SHA256]-----+
hduser@master:/home/belozhi$ cat ~/.ssh/id_rsa.pub >> ~/.ssh/authorized_keys
hduser@master:/home/belozhi$ chmod 0600 ~/.ssh/authorized_keys
hduser@master:/home/belozhi$ ssh localhost
Welcome to Ubuntu 20.04.3 LTS (GNU/Linux 5.11.0-41-generic x86_64)

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

3 updates can be applied immediately.
```

Проверка подключения по ssh (*ssh localhost*)

```
hduser@master:/home/belozhi$ ssh localhost
Welcome to Ubuntu 20.04.3 LTS (GNU/Linux 5.11.0-41-generic x86_64)

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

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

Your Hardware Enablement Stack (HWE) is supported until April 2025.
Last login: Sat Dec 11 15:32:39 2021 from 192.168.121.240
```

Установка Apache Hadoop в Ubuntu 20.04

su - hduser

wget https://downloads.apache.org/hadoop/common/hadoop-3.3.0/hadoop-3.3.0.tar.gz

tar -xvzf hadoop-3.3.0.tar.gz

Затем переместим извлеченный каталог в `:/usr/local/`

sudo mv hadoop-3.3.0 /usr/local/hadoop

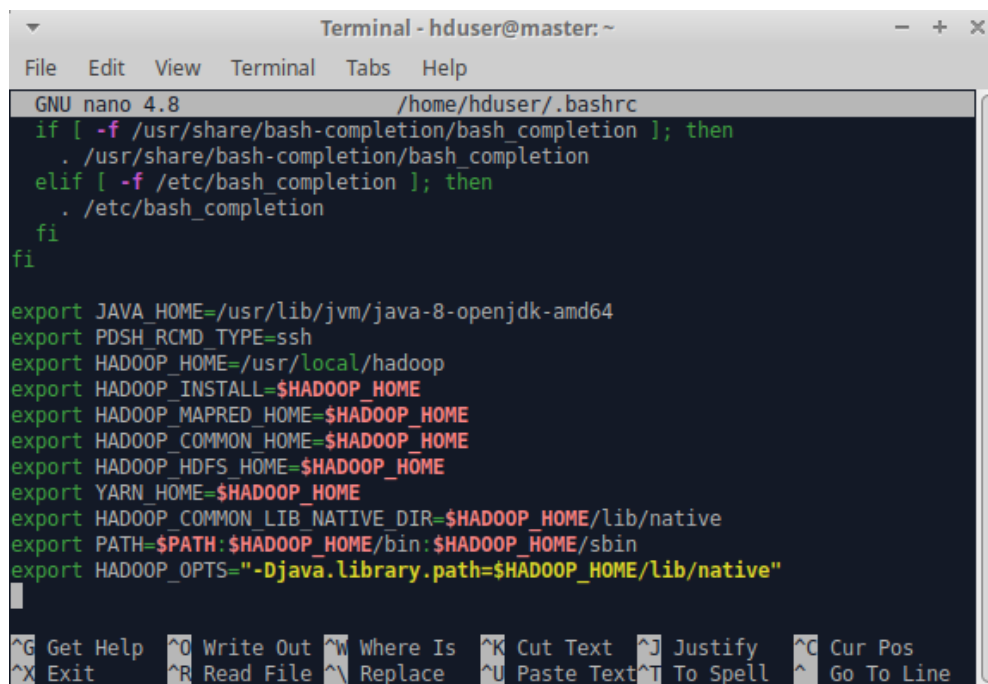
sudo mkdir /usr/local/hadoop/logs

chown -R hduser:hadoop /usr/local/hadoop

Настройте Apache Hadoop

Настройка переменных среды в файле ~/.bashrc

nano ~/.bashrc



```
Terminal - hduser@master: ~
File Edit View Terminal Tabs Help
GNU nano 4.8 /home/hduser/.bashrc
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 JAVA_HOME=/usr/lib/jvm/java-8-openjdk-amd64
export PDSH_RCMD_TYPE=ssh
export HADOOP_HOME=/usr/local/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/bin:$HADOOP_HOME/sbin
export HADOOP_OPTS="-Djava.library.path=$HADOOP_HOME/lib/native"
```

source ~/.bashrc

Затем нужно определить переменные среды Java, чтобы настроить параметры проекта, связанные с YARN, HDFS, MapReduce и Hadoop:hadoop-env.sh

sudo nano \$HADOOP_HOME/etc/hadoop/hadoop-env.sh

export JAVA_HOME=/usr/lib/jvm/java-8-openjdk-amd64

export HADOOP_CLASSPATH+=" \$HADOOP_HOME/lib/* .jar"

```
Terminal - hduser@master: ~
File Edit View Terminal Tabs Help
GNU nano 4.8 /usr/local/hadoop/etc/hadoop/hadoop-env.sh
# export HDFS_NAMENODE_USER=hdfs

###
# Registry DNS specific parameters
###
# For privileged registry DNS, user to run as after dropping privileges
# This will replace the hadoop.id.str Java property in secure mode.
# export HADOOP_REGISTRYDNS_SECURE_USER=yarn

# Supplemental options for privileged registry DNS
# By default, Hadoop uses jsvc which needs to know to launch a
# server jvm.
# export HADOOP_REGISTRYDNS_SECURE_EXTRA_OPTS="-jvm server"

export JAVA_HOME=/usr/lib/jvm/java-8-openjdk-amd64
export HADOOP_CLASSPATH+=" $HADOOP_HOME/lib/*.jar"

^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
```

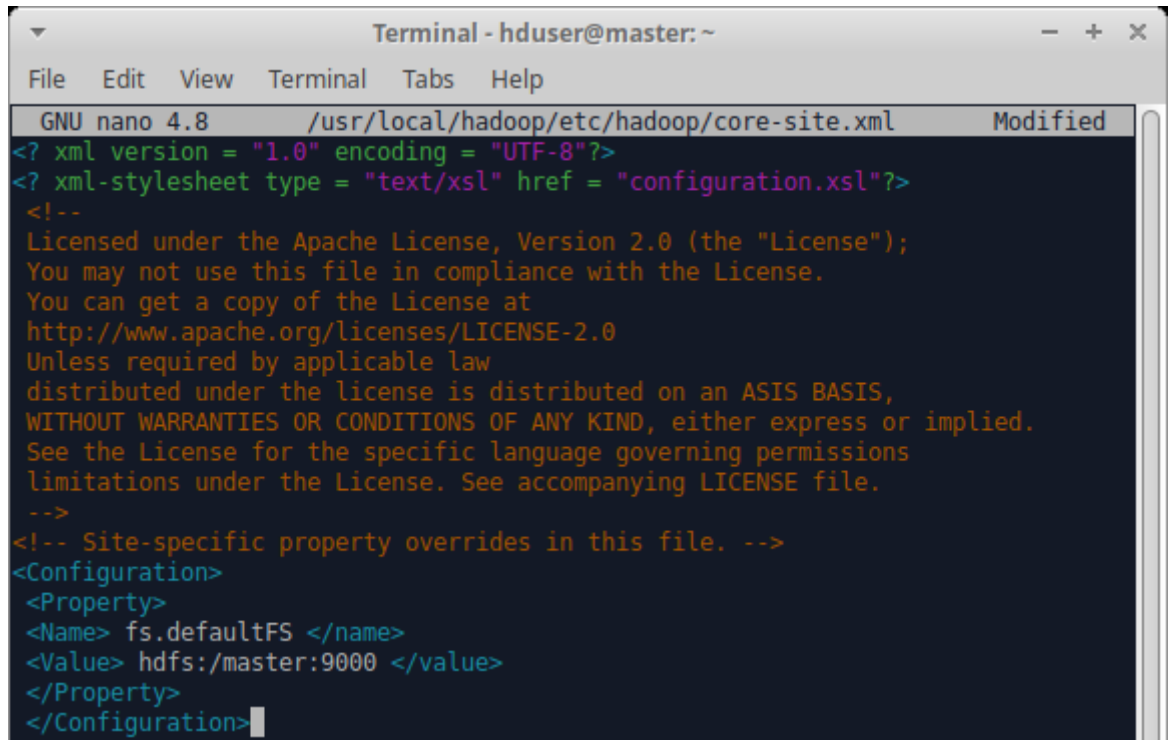
Проверка версии Hadoop

hadoop version

```
hduser@master:~$ hadoop version
Hadoop 3.3.0
Source code repository https://gitbox.apache.org/repos/asf/hadoop.git -r aa96f18
71bfd858f9bac59cf2a81ec470da649af
Compiled by brahma on 2020-07-06T18:44Z
Compiled with protoc 3.7.1
From source with checksum 5dc29b802d6ccd77b262ef9d04d19c4
This command was run using /usr/local/hadoop/share/hadoop/common/hadoop-common-3
.3.0.jar
```

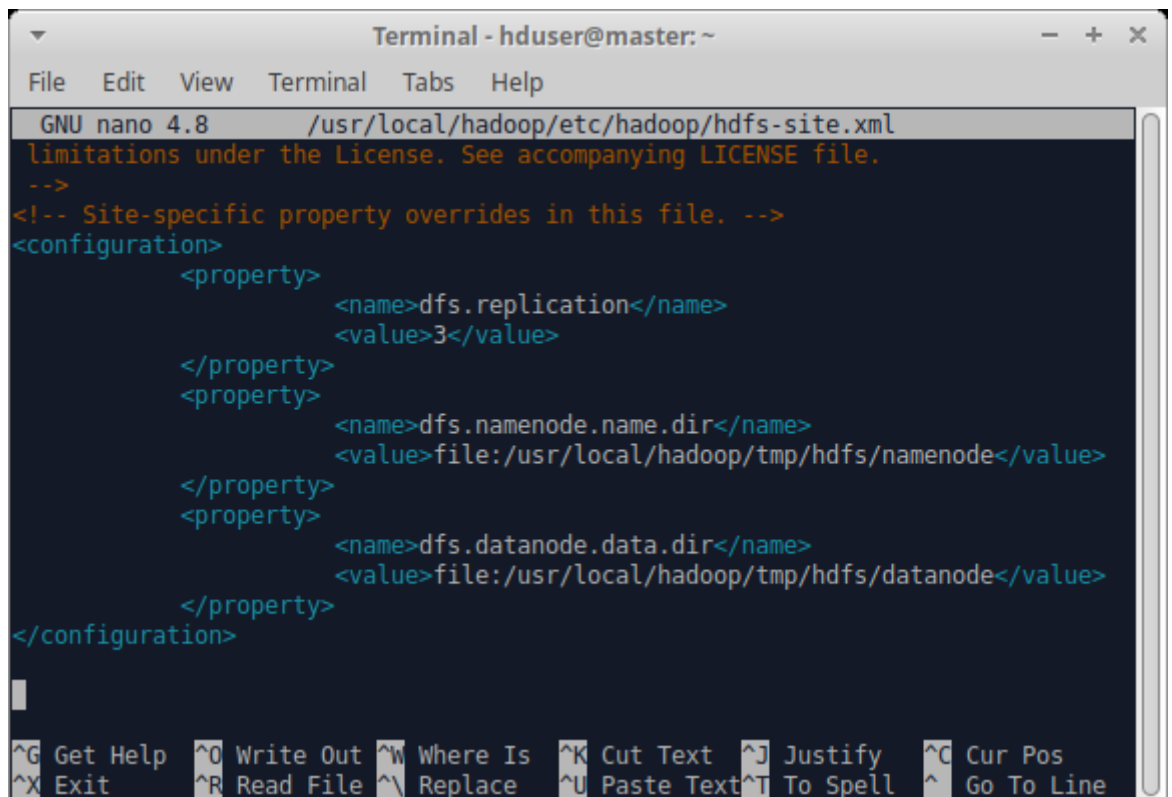
Настройка файлов

Настроить файл core-site.xml



```
Terminal - hduser@master: ~
File Edit View Terminal Tabs Help
GNU nano 4.8 /usr/local/hadoop/etc/hadoop/core-site.xml Modified
<? 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 in compliance with the License.
You can get a copy of the License at
http://www.apache.org/licenses/LICENSE-2.0
Unless required by applicable law
distributed under the license is distributed on an ASIS BASIS,
WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied.
See the License for the specific language governing permissions
limitations under the License. See accompanying LICENSE file.
-->
<!-- Site-specific property overrides in this file. -->
<Configuration>
  <Property>
    <Name> fs.defaultFS </name>
    <Value> hdfs://master:9000 </value>
  </Property>
</Configuration>
```

Настроить файл hdfs-site.xml



```
Terminal - hduser@master: ~
File Edit View Terminal Tabs Help
GNU nano 4.8 /usr/local/hadoop/etc/hadoop/hdfs-site.xml
limitations under the License. See accompanying LICENSE file.
-->
<!-- Site-specific property overrides in this file. -->
<configuration>
  <property>
    <name>dfs.replication</name>
    <value>3</value>
  </property>
  <property>
    <name>dfs.namenode.name.dir</name>
    <value>file:/usr/local/hadoop/tmp/hdfs/namenode</value>
  </property>
  <property>
    <name>dfs.datanode.data.dir</name>
    <value>file:/usr/local/hadoop/tmp/hdfs/datanode</value>
  </property>
</configuration>

^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
```

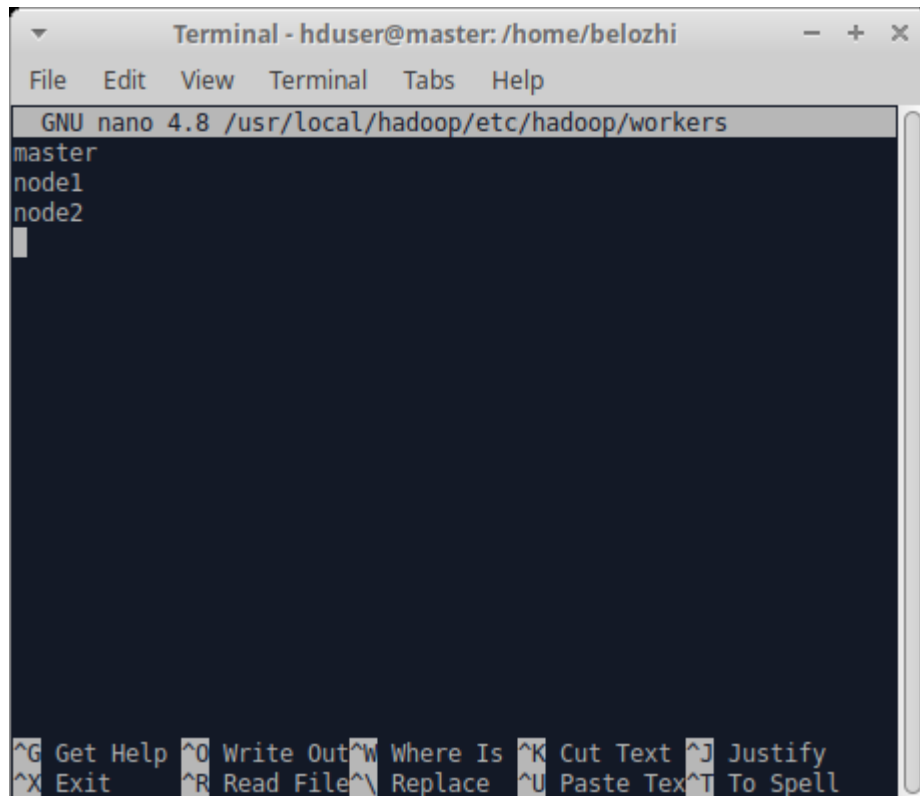

Настроить файл mapred-site.xml

```
Terminal - hduser@master: /home/belozhi
File Edit View Terminal Tabs Help
GNU nano 4.8 /usr/local/hadoop/etc/hadoop/mapred-site.xml
See the License for the specific language governing permissions and
limitations under the License. See accompanying LICENSE 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>
  <property>
    <name>yarn.app.mapreduce.am.resource.mb</name>
    <value>2048</value>
  </property>
  <property>
    <name>mapreduce.map.memory.mb</name>
    <value>1024</value>
  </property>
  <property>
    <name>mapreduce.reduce.memory.mb</name>
    <value>1024</value>
  </property>
</configuration>
```

Настроить файл yarn-site.xml

```
Terminal - hduser@master: /home/belozhi
File Edit View Terminal Tabs Help
GNU nano 4.8 /usr/local/hadoop/etc/hadoop/yarn-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.
-->
<configuration>
<!-- Site specific YARN configuration properties -->
  <property>
    <name>yarn.acl.enable</name>
    <value>0</value>
  </property>
  <property>
    <name>yarn.resourcemanager.hostname</name>
    <value>master</value>
  </property>
  <property>
    <name>yarn.nodemanager.aux-services</name>
    <value>mapreduce_shuffle</value>
  </property>
  <property>
    <name>yarn.nodemanager.resource.memory-mb</name>
    <value>2048</value>
  </property>
  <property>
    <name>yarn.scheduler.maximum-allocation-mb</name>
    <value>2048</value>
  </property>
  <property>
    <name>yarn.scheduler.minimum-allocation-mb</name>
    <value>1024</value>
  </property>
  <property>
    <name>yarn.nodemanager.vmem-check-enabled</name>
    <value>false</value>
  </property>
</configuration>
```

Добавим на узле master все рабочие узлы в файл etc/hadoop/workers.



```
Terminal - hduser@master: /home/belozhi
File Edit View Terminal Tabs Help
GNU nano 4.8 /usr/local/hadoop/etc/hadoop/workers
master
node1
node2
^G Get Help ^O Write Out ^W Where Is ^K Cut Text ^J Justify
^X Exit ^R Read File ^\ Replace ^U Paste Text ^T To Spell
```

Форматирование HDFS NameNode

su - hduser

hdfs namenode -format

Запуск кластера Hadoop

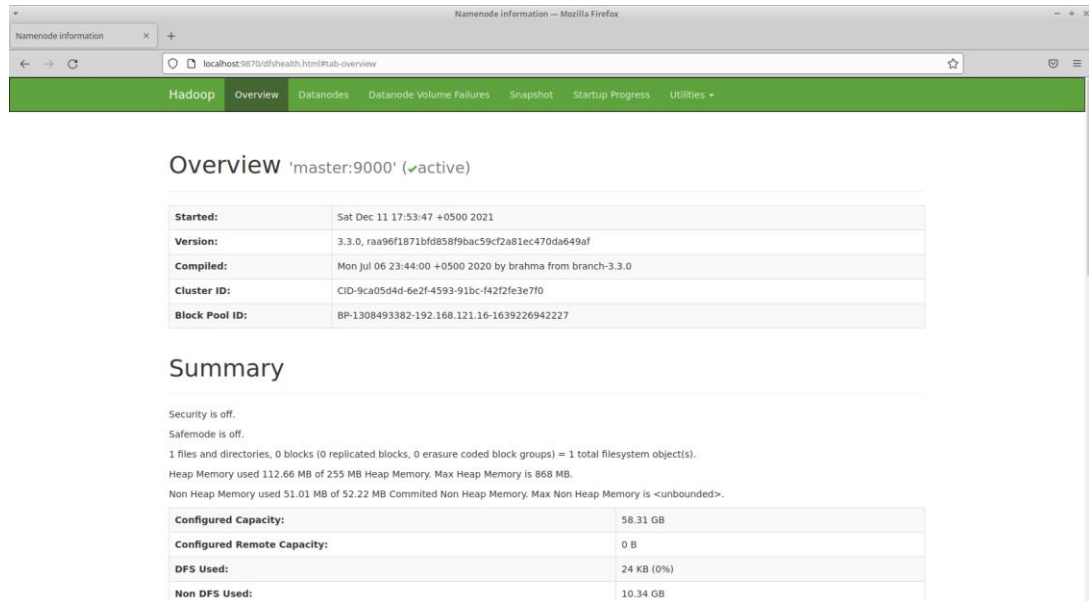
start-dfs.sh

```
hduser@master:~$ start-dfs.sh
Starting namenodes on [master]
Starting datanodes
Starting secondary namenodes [master]
```

start-yarn.sh

```
hduser@master:~$ start-yarn.sh
Starting resourcemanager
Starting nodemanagers
```

Проверка доступа к Apache Hadoop



The screenshot shows the 'Overview' page for the NameNode in a web browser. The page title is 'NameNode Information - Mozilla Firefox'. The browser address bar shows 'localhost:9870/dfshealth.html#tab-overview'. The page has a green header with 'Hadoop' and a navigation menu with 'Overview', 'Datanodes', 'Datanode Volume Failures', 'Snapshot', 'Startup Progress', and 'Utilities'. The main content area is titled 'Overview 'master:9000' (active)'. It contains a table with the following information:

Started:	Sat Dec 11 17:53:47 +0500 2021
Version:	3.3.0, raa96f1871bfd85f9bac59cf2a81ec470da649af
Compiled:	Mon Jul 06 23:44:00 +0500 2020 by brahma from branch-3.3.0
Cluster ID:	CID-9ca05d4d-6e2f-4593-91bc-f42f2fe3e7f0
Block Pool ID:	BP-1308493382-192.168.121.16-1639226942227

Below the table is a 'Summary' section. It states: '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 112.66 MB of 255 MB Heap Memory. Max Heap Memory is 868 MB.', 'Non Heap Memory used 51.01 MB of 52.22 MB Committed Non Heap Memory. Max Non Heap Memory is <unbounded>.' Below this is another table:

Configured Capacity:	58.31 GB
Configured Remote Capacity:	0 B
DFS Used:	24 KB (0%)
Non DFS Used:	10.34 GB

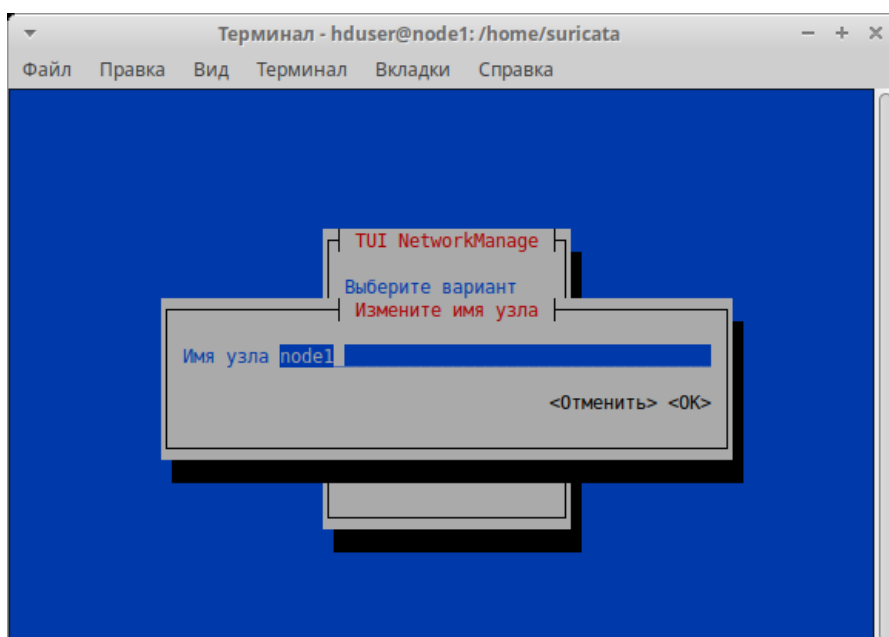
Задание 1

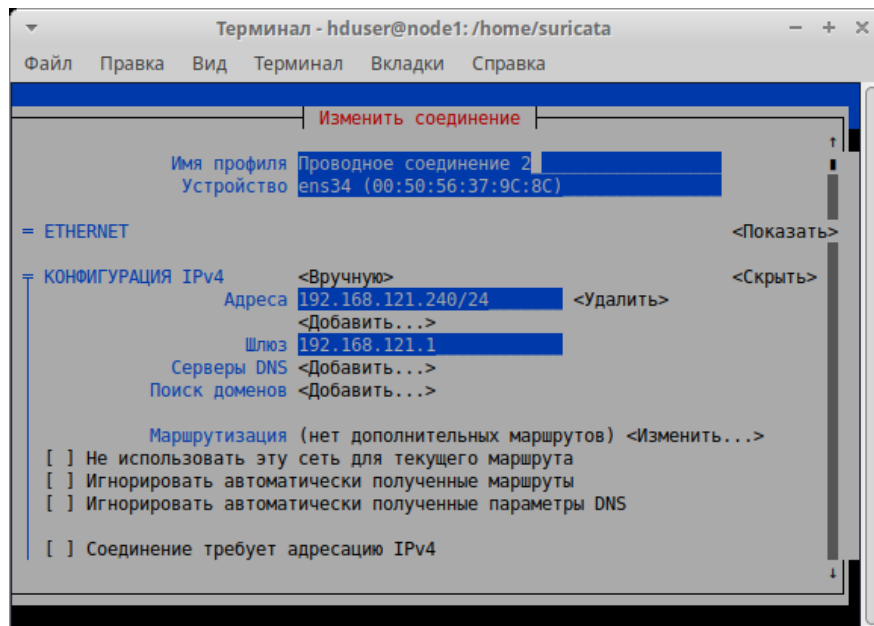
Задача: познакомиться с процессом добавления новых узлов в существующий кластер.

1. Воспользовавшись знаниями из Задания 0 и схемой кластера из презентации (Лабораторная №2), установите и настройте DataNode на узлы кластера node1 и node2.

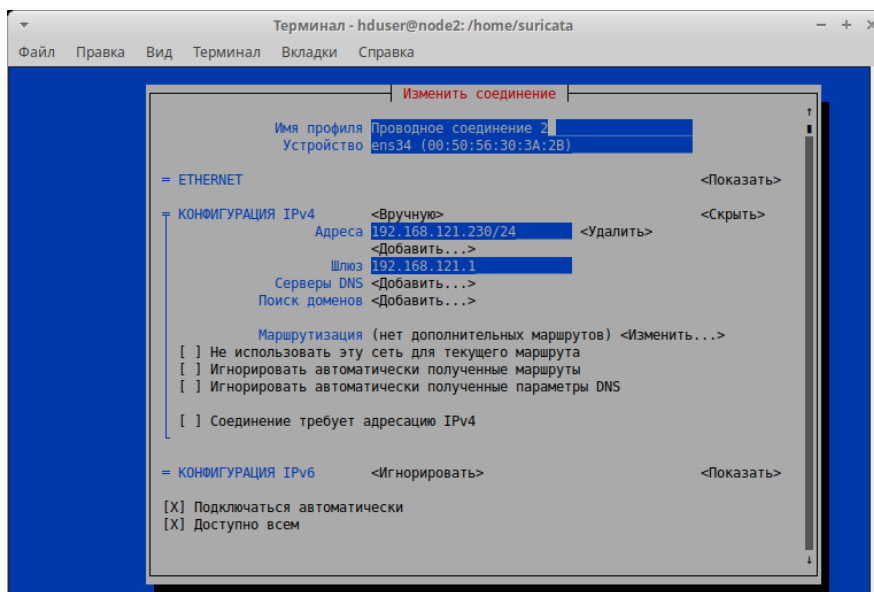
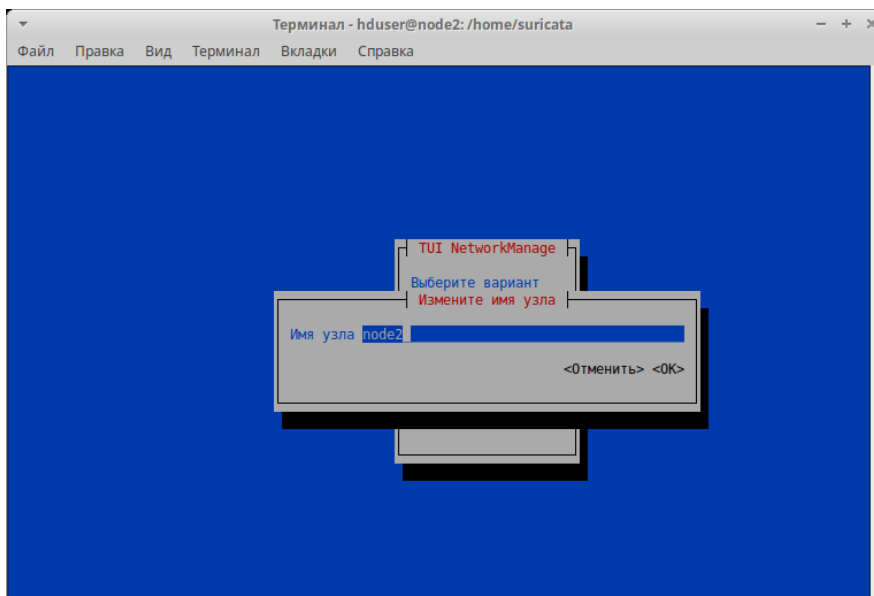
Настройка сетевых имён и ip-адресов на VM node1 и node2.

node1





node2



Далее аналогичным образом (как в в Задании 0) выполняется:

- Настройка учетной записи hduser
- Получение SSH ключей
- Установка и настройка Hadoop

Копируем ключи SSH на подчиненные машины:

```
hduser@master: ~$ ssh-copy-id -i $HOME/.ssh/id_rsa.pub hduser@node1
```

и на вторую:

```
hduser@master: ~$ ssh-copy-id -i $HOME/.ssh/id_rsa.pub hduser@ node2
```

Последний шаг настройки - удаление старых каталогов HDFS.

Для этого на мастере и подчиненных узлах нужно удалить каталоги `/usr/local/hadoop/hadoop_tmp/`:

```
sudo rm -rf /usr/local/hadoop/hadoop_tmp/
```

Далее на **мастере** создать каталоги namenode и datanode, так как мастер-машина у нас одновременно мастер и подчиненная.

Команды:

```
sudo mkdir -p /usr/local/hadoop/hadoop_tmp/hdfs/namenode
```

```
sudo mkdir -p /usr/local/hadoop/hadoop_tmp/hdfs/datanode
```

Затем назначить владельцем каталога пользователя hduser группы hadoop:

```
sudo chown hduser:hadoop -R /usr/local/hadoop/hadoop_tmp/
```

На подчиненных машинах сделать тоже самое, но создать только каталог datanode:

```
sudo mkdir -p /usr/local/hadoop/hadoop_tmp/hdfs/datanode
```

```
sudo chown hduser:hadoop -R /usr/local/hadoop/hadoop_tmp/
```

Теперь все готово, отформатируем файловую систему, выполнив команду на Мастере:

```
hdfs namenode -format
```

Если нет ошибок, запустим hdfs и yarn с помощью скрипта из первой части, либо командами:

start-dfs.sh

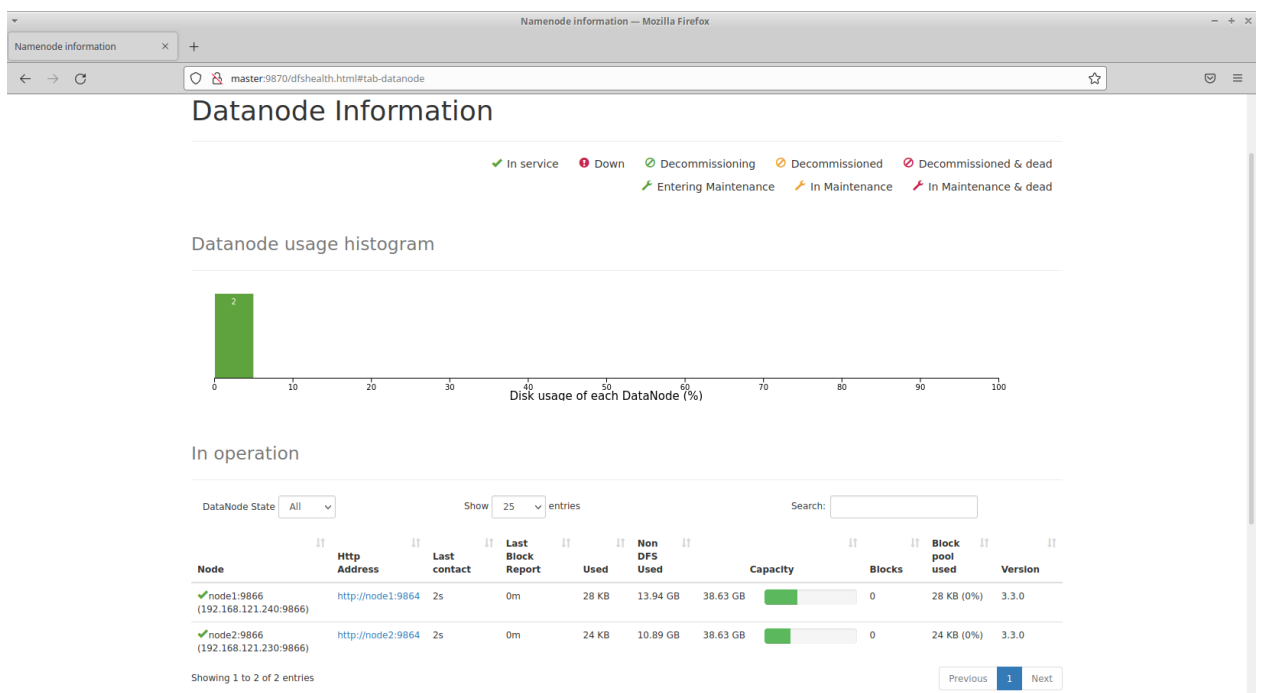
start-yarn.sh

```
hduser@master:/home/belozhi$ start-dfs.sh
Starting namenodes on [master]
Starting datanodes
Starting secondary namenodes [master]
hduser@master:/home/belozhi$ start-yarn.sh
Starting resourcemanager
Starting nodemanagers
node2: ERROR: Cannot set priority of nodemanager process 2023
```

С помощью команды `jps` можно посмотреть запущенные java-процессы на мастере.

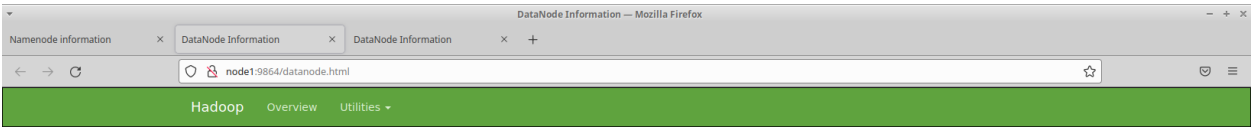
```
hduser@master:/home/belozhi$ jps
15136 Jps
5205 NodeManager
4758 SecondaryNameNode
4971 ResourceManager
4285 NameNode
```

Проверка работы кластера в web-интерфейсе.



In operation

Datanode State: All									
Show 25 entries									
Search:									
Node	Http Address	Last contact	Last Block Report	Used	Non DFS Used	Capacity	Blocks	Block pool used	Version
node1:9866 (192.168.121.240:9866)	http://node1:9864	1s	31m	32.06 KB	13.95 GB	38.63 GB	1	32.06 KB (0%)	3.3.0
node2:9866 (192.168.121.230:9866)	http://node2:9864	0s	31m	28.06 KB	10.89 GB	38.63 GB	1	28.06 KB (0%)	3.3.0



DataNode on node1:9866

Cluster ID:	CID-5ba5b554-c819-470b-b7ec-4c6734423567
Version:	3.3.0, raa96f1871bfd858f9bac59cf2a81ec470da649af

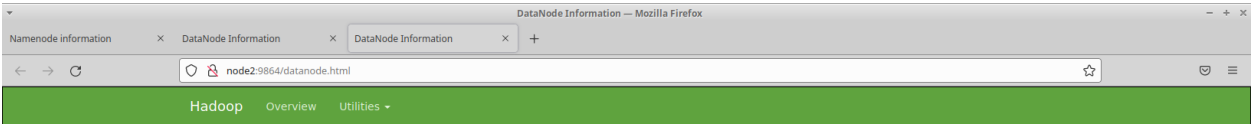
Block Pools

Namenode Address	Block Pool ID	Actor State	Last Heartbeat	Last Block Report	Last Block Report Size (Max Size)
master:9000	BP-939744533-192.168.121.16-1639248653348	RUNNING	0s	9 minutes	0 B (128 MB)

Volume Information

Directory	StorageType	Capacity Used	Capacity Left	Capacity Reserved	Reserved Space for Replicas	Blocks
/usr/local/hadoop/tmp/hdfs/datanode	DISK	28 KB	22.7 GB	0 B	0 B	0

Hadoop, 2020.



DataNode on node2:9866

Cluster ID:	CID-5ba5b554-c819-470b-b7ec-4c6734423567
Version:	3.3.0, raa96f1871bfd858f9bac59cf2a81ec470da649af

Block Pools

Namenode Address	Block Pool ID	Actor State	Last Heartbeat	Last Block Report	Last Block Report Size (Max Size)
master:9000	BP-939744533-192.168.121.16-1639248653348	RUNNING	0s	9 minutes	0 B (128 MB)

Volume Information

Directory	StorageType	Capacity Used	Capacity Left	Capacity Reserved	Reserved Space for Replicas	Blocks
/usr/local/hadoop/tmp/hdfs/datanode	DISK	24 KB	25.75 GB	0 B	0 B	0

Hadoop, 2020.

2. Подключитесь к NameNode с помощью ssh и выполните команду `hdfs dfsadmin -report` в отчёте вы должны увидеть число узлов (Live datanodes) и свой свежедобавленный узел.

```
hduser@master:/home/belozhi$ ssh node2
Welcome to Ubuntu 20.04.3 LTS (GNU/Linux 5.11.0-41-generic x86_64)

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

17 updates can be applied immediately.
Чтобы просмотреть дополнительные обновления выполните: apt list --upgradable

Your Hardware Enablement Stack (HWE) is supported until April 2025.
Last login: Sat Dec 11 23:58:01 2021 from 192.168.121.16
```

```
hduser@node2:~$ hdfs dfsadmin -report
Configured Capacity: 82951274496 (77.25 GB)
Present Capacity: 52018753536 (48.45 GB)
DFS Remaining: 52018700288 (48.45 GB)
DFS Used: 53248 (52 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 (2):

Name: 192.168.121.230:9866 (node2)
Hostname: node2
Decommission Status : Normal
Configured Capacity: 41475637248 (38.63 GB)
DFS Used: 24576 (24 KB)
Non DFS Used: 11692052480 (10.89 GB)
DFS Remaining: 27646300160 (25.75 GB)
DFS Used%: 0.00%
DFS Remaining%: 66.66%
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: Sun Dec 12 00:06:38 YEKT 2021
Last Block Report: Sun Dec 12 00:03:20 YEKT 2021
Num of Blocks: 0
```



```
Name: 192.168.121.240:9866 (node1)
Hostname: node1
Decommission Status : Normal
Configured Capacity: 41475637248 (38.63 GB)
DFS Used: 28672 (28 KB)
Non DFS Used: 14965948416 (13.94 GB)
DFS Remaining: 24372400128 (22.70 GB)
DFS Used%: 0.00%
DFS Remaining%: 58.76%
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: Sun Dec 12 00:06:37 YEKT 2021
Last Block Report: Sun Dec 12 00:03:19 YEKT 2021
Num of Blocks: 0
```

Задание 2

Задача: работа с HDFS с помощью консольных утилит.

1. Разделитесь на команды и с помощью утилиты `hadoop` создайте в кластере:

1.1 Папку с идентификатором своей команды

```
hduser@node2:~$ hadoop fs -mkdir hdfs://master:9000/Belousov_Zhidenko
mkdir: `hdfs://master:9000/Belousov_Zhidenko': File exists
```

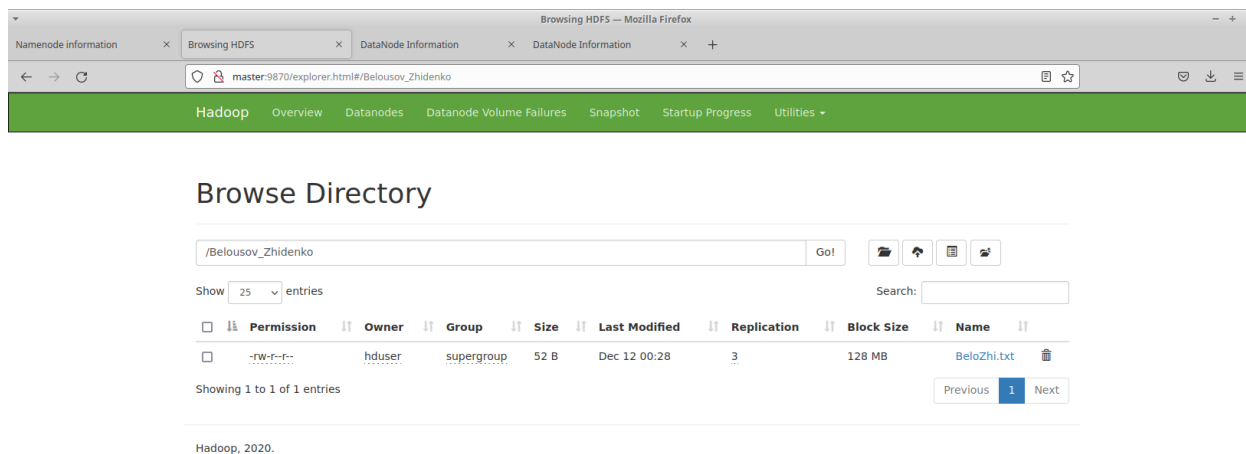
```
hduser@node2:~$ hadoop fs -ls /
Found 1 items
drwxr-xr-x - hduser supergroup          0 2021-12-12 00:17 /Belousov_Zhidenko
```

1.2 Внутри папки положите файл с фамилиями участников команды

```
hduser@node2:~$ hadoop fs -ls /
Found 1 items
drwxr-xr-x - hduser supergroup          0 2021-12-12 00:28 /Belousov_Zhidenko
hduser@node2:~$ hadoop fs -ls /Belousov_Zhidenko
Found 1 items
-rw-r--r-- 3 hduser supergroup          52 2021-12-12 00:28 /Belousov_Zhidenko/BeloZhi.txt
```

```
hduser@node2:~$ hadoop fs -cat /Belousov_Zhidenko/BeloZhi.txt
1. Belousov Andrey
2. Zhidenko Alexander
RIM-201211
```

Посмотреть файлы в каталоге Hadoop можно также через web-интерфейс.



The screenshot shows the Hadoop web interface in a Mozilla Firefox browser. The address bar displays the URL `master:9870/explorer.html#/Belousov_Zhidenko`. The navigation bar includes links for Hadoop, Overview, Datanodes, Datanode Volume Failures, Snapshot, Startup Progress, and Utilities. The main content area is titled "Browse Directory" and shows the path `/Belousov_Zhidenko`. Below the path is a search bar and a "Go!" button. The interface displays a table of file entries with columns for Permission, Owner, Group, Size, Last Modified, Replication, Block Size, and Name. The table shows one entry: `BelouZhi.txt` with a size of 128 MB and 3 replications. The interface also includes a "Showing 1 to 1 of 1 entries" message and a pagination bar with "Previous", "1", and "Next" buttons.

Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name
-rw-r--r--	hduser	supergroup	52 B	Dec 12 00:28	3	128 MB	BelouZhi.txt

Showing 1 to 1 of 1 entries

Previous 1 Next

Hadoop, 2020.