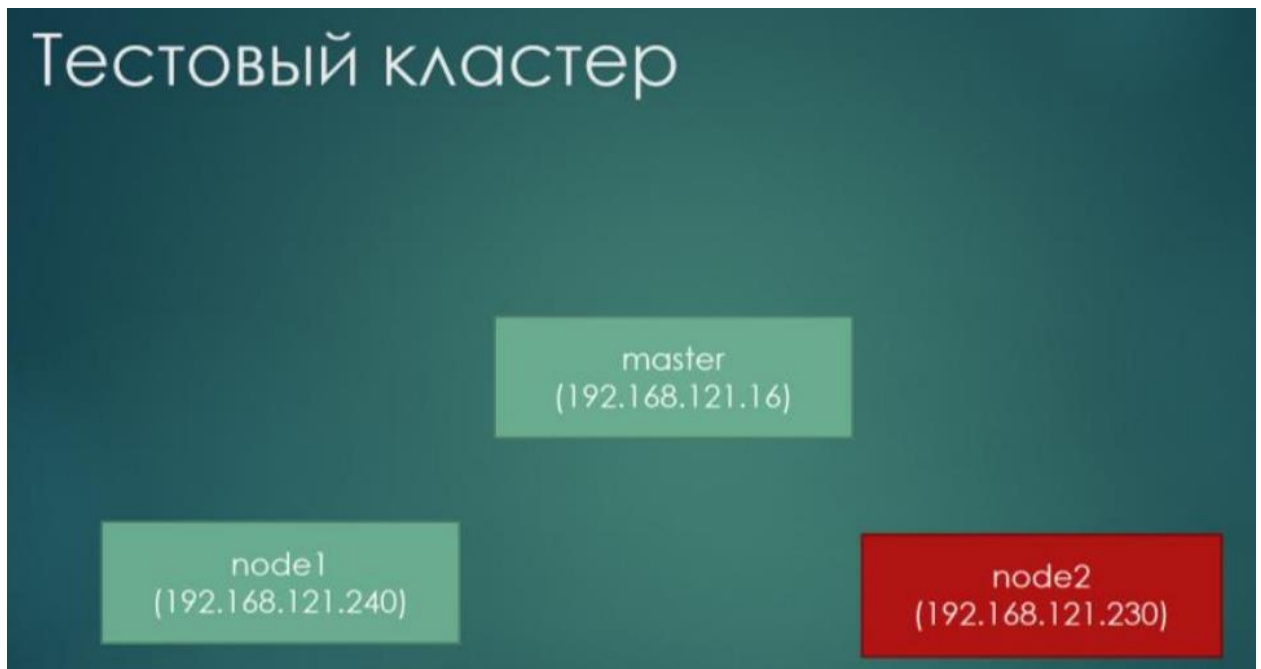


## Лабораторная работа №2. Кокарев Д. В. РИМ-201211

Реализовать схему тестового кластера:



Установка ssh:

```
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
```

Редактирование файла /etc/hosts

```
sudo nano /etc/hosts
```

```
kokarev@ubuntu: ~  
GNU nano 4.8 /etc/hosts  
127.0.0.1    localhost  
#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
```

Далее заходим и продолжаем под hduser

Получение SSH ключей

ssh-keygen -t rsa

```
hduser@master: /home/kokarev  
hduser@master:~$ cd /home/kokarev  
hduser@master:/home/kokarev$ 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:nKCIh+BPx96pXtuw/++0ypzBTKUD+ra1DqMfunu7ZXI hduser@master  
The key's randomart image is:  
+---[RSA 3072]-----+  
|  
|.      ..      .  
|oo  ....o..o  
|o.o..+  S+  
|.o o o = .  
|  . . % E  
|    *.^ + .  
|.B=B=Xoo++  
+----[SHA256]-----+  
hduser@master:/home/kokarev$
```

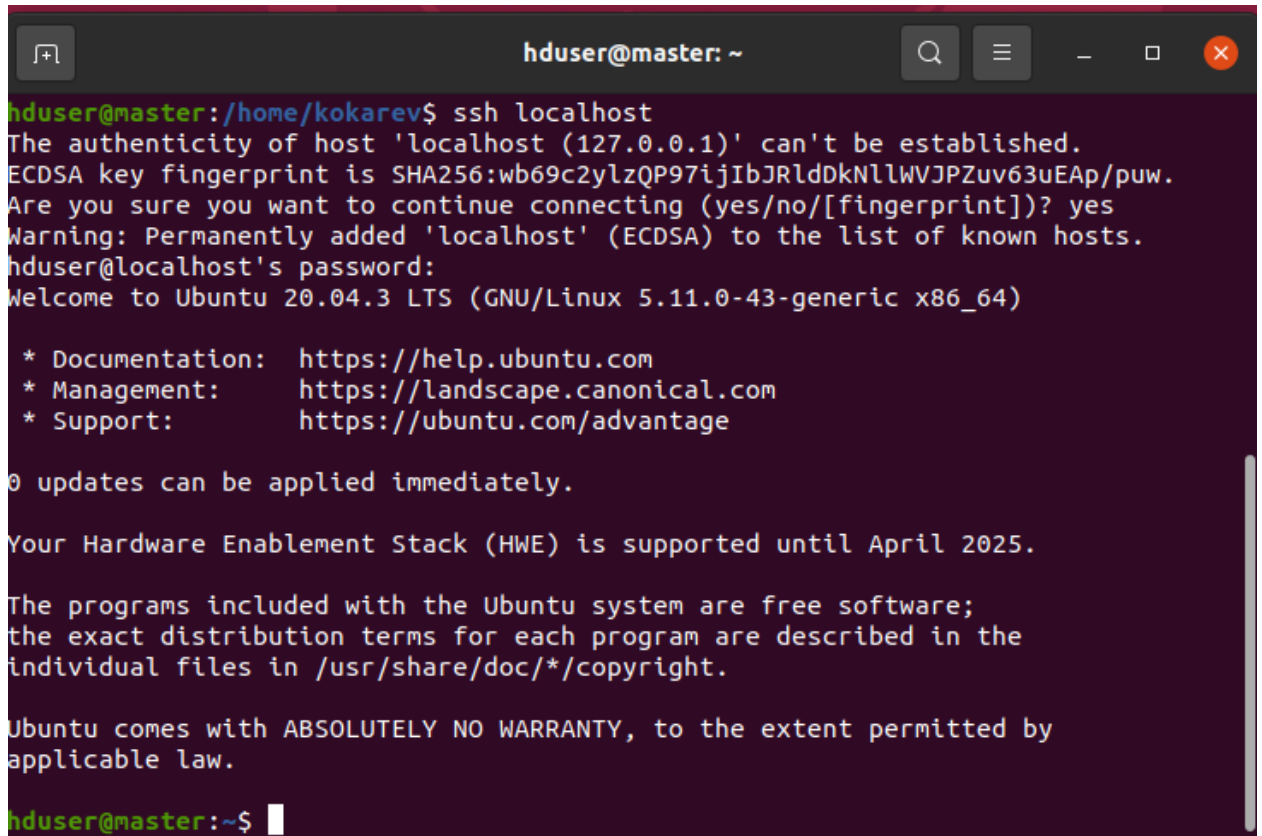
Перемещаем полученные ключи:

cat ~/.ssh/id\_rsa.pub >> ~/.ssh/authorized\_keys

chmod 0600 ~/.ssh/authorized\_keys

Проверка подключения по ssh (ssh localhost)

ssh localhost



```
hduser@master: ~  
hduser@master:/home/kokarev$ ssh localhost  
The authenticity of host 'localhost (127.0.0.1)' can't be established.  
ECDSA key fingerprint is SHA256:wb69c2ylzQP97ijIbJRldDkNllWVJPZuv63uEAp/puw.  
Are you sure you want to continue connecting (yes/no/[fingerprint])? yes  
Warning: Permanently added 'localhost' (ECDSA) to the list of known hosts.  
hduser@localhost's password:  
Welcome to Ubuntu 20.04.3 LTS (GNU/Linux 5.11.0-43-generic x86_64)  
  
* Documentation:  https://help.ubuntu.com  
* Management:    https://landscape.canonical.com  
* Support:       https://ubuntu.com/advantage  
  
0 updates can be applied immediately.  
  
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.  
hduser@master:~$
```

Установка Apache Hadoop

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

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

Настройка Apache Hadoop:

Настройка переменных среды в файле ~/.bashrc

sudo nano ~/.bashrc

```
GNU nano 4.8 /home/hduser/.bashrc Modified
# enable programmable completion features (you don't need to enable
# this, if it's already enabled in /etc/bash.bashrc and /etc/profile
# sources /etc/bash.bashrc).
if ! shopt -oq posix; then
  if [ -f /usr/share/bash-completion/bash_completion ]; then
    . /usr/share/bash-completion/bash_completion
  elif [ -f /etc/bash_completion ]; then
    . /etc/bash_completion
  fi
fi

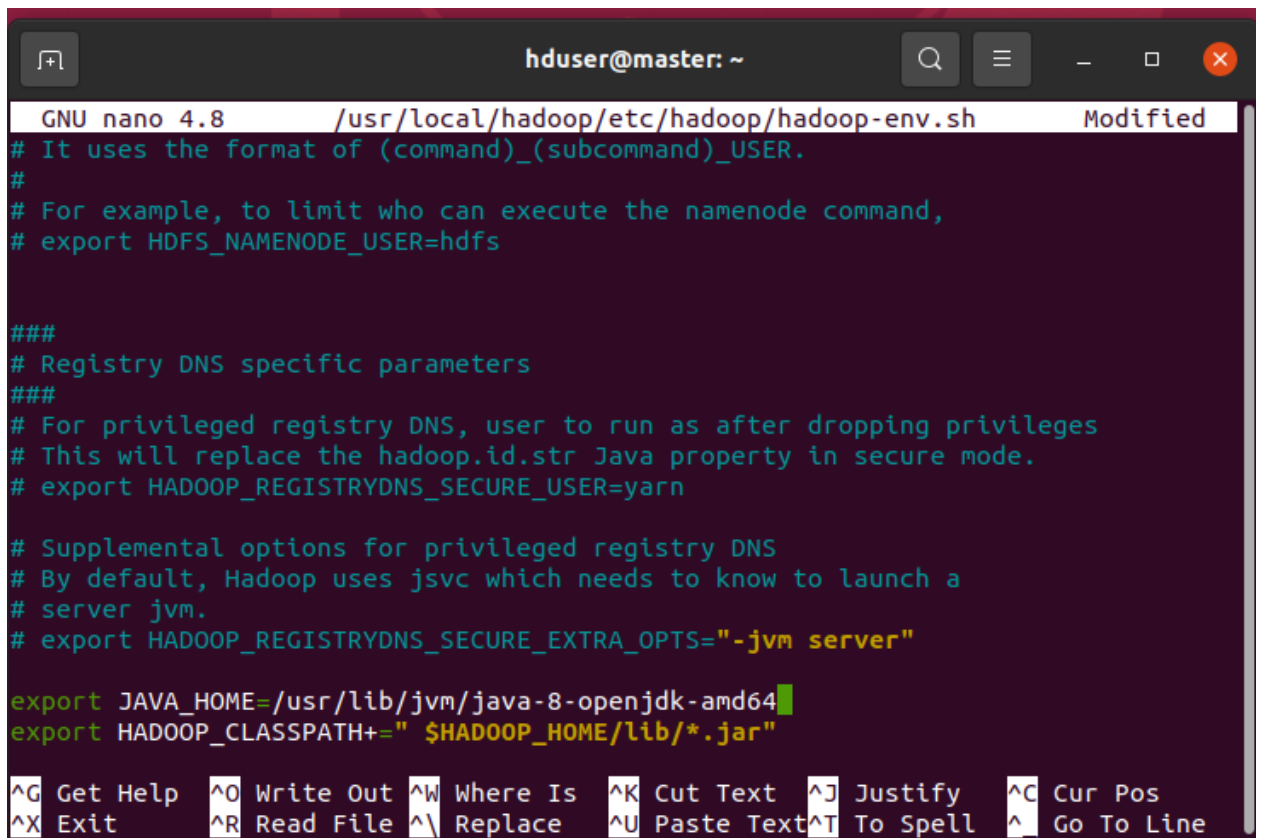
export HADOOP_HOME=/usr/local/hadoop
export HADOOP_INSTALL=$HADOOP_HOME
export HADOOP_MAPRED_HOME=$HADOOP_HOME
export HADOOP_COMMON_HOME=$HADOOP_HOME
export HADOOP_HDFS_HOME=$HADOOP_HOME
export YARN_HOME=$HADOOP_HOME
export HADOOP_COMMON_LIB_NATIVE_DIR=$HADOOP_HOME/lib/native
export PATH=$PATH:$HADOOP_HOME/sbin:$HADOOP_HOME/bin
export HADOOP_OPTS="-Djava.library.path=$HADOOP_HOME/lib/native"

^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
```

source ~/.bashrc

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

sudo nano \$HADOOP\_HOME/etc/hadoop/hadoop-env.sh



```
GNU nano 4.8 /usr/local/hadoop/etc/hadoop/hadoop-env.sh Modified
# It uses the format of (command)_(subcommand)_USER.
#
# For example, to limit who can execute the namenode command,
# 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
```

Настройка файлов

`sudo nano /usr/local/hadoop/etc/hadoop/core-site.xml`

```
hduser@master: ~  
GNU nano 4.8 /usr/local/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.defaultFS</name>  
<value>hdfs://master:9000</value>  
</property>  
</configuration>  
  
^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
```

```
sudo nano /usr/local/hadoop/etc/hadoop/hdfs-site.xml
```

```
hduser@master: ~
GNU nano 4.8 /usr/local/hadoop/etc/hadoop/hdfs-site.xml Modified
-->

<!-- Put 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
```

sudo nano /usr/local/hadoop/etc/hadoop/mapred-site.xml

```
hduser@master: ~
GNU nano 4.8 /usr/local/hadoop/etc/hadoop/mapred-site.xml
<!-- Put site-specific property overrides in this file. -->

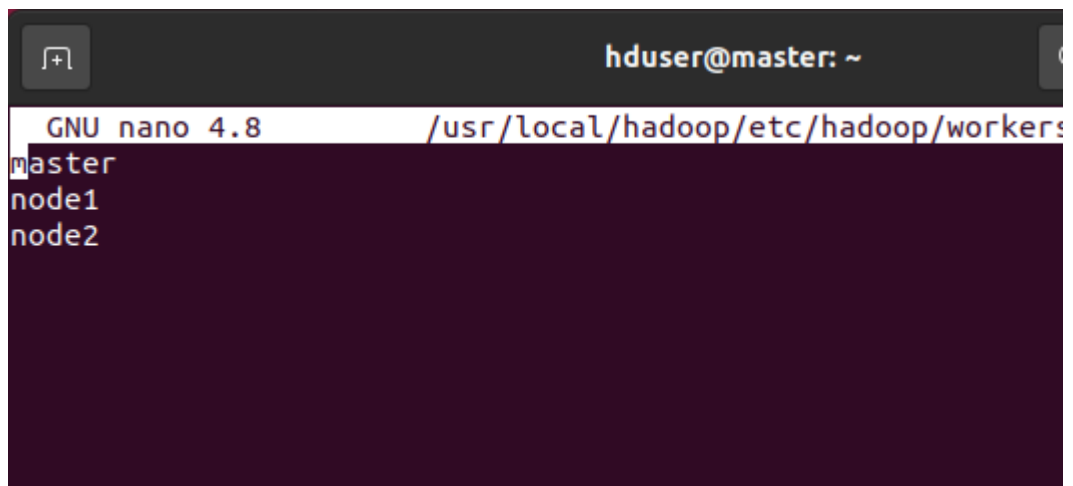
<configuration>
  <property>
    <name>mapreduce.framework.name</name>
    <value>yarn</value>
  </property>
  <property>
    <name>yarn.app.mapreduce.am.env</name>
    <value>HADOOP_MAPRED_HOME=$HADOOP_HOME</value>
  </property>
  <property>
    <name>mapreduce.map.env</name>
    <value>HADOOP_MAPRED_HOME=$HADOOP_HOME</value>
  </property>
  <property>
    <name>mapreduce.reduce.env</name>
    <value>HADOOP_MAPRED_HOME=$HADOOP_HOME</value>
  </property>
  <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>
```

sudo nano /usr/local/hadoop/etc/hadoop/yarn-site.xml

```
<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

sudo nano /usr/local/hadoop/etc/hadoop/workers



```
GNU nano 4.8 /usr/local/hadoop/etc/hadoop/workers
master
node1
node2
```



## Форматирование HDFS NameNode

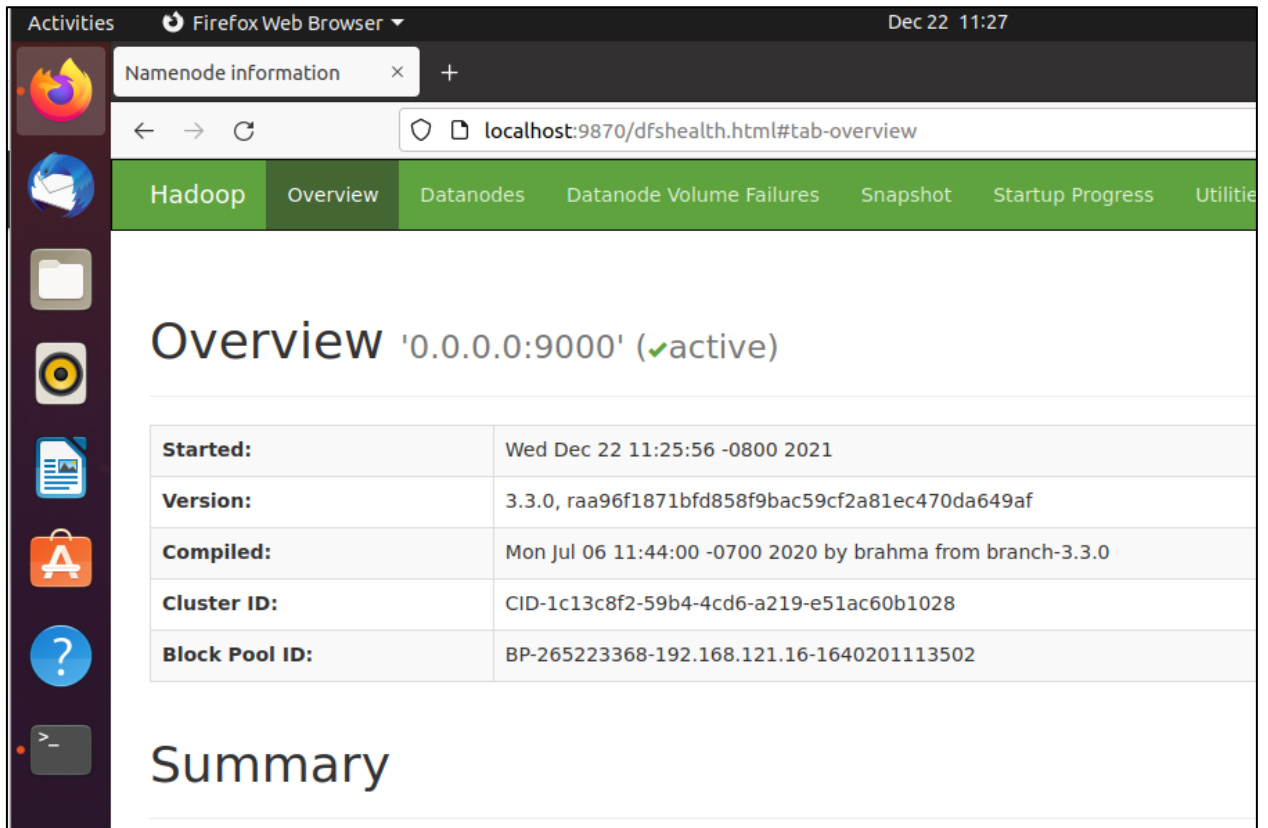
su - hduser

hdfs namenode -format

## Запуск кластера Hadoop

start-dfs.sh

start-yarn.sh



## Задание 1:

На мастере:

Каталог namenode создан автоматически

Создаем каталог datanode

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

форматируем файловую систему:

```
hdfs namenode -format
```

Копируем виртуальную машину мастера два раза – это будут узлы node1 и node2

Настраиваем имена узлов и ip адреса

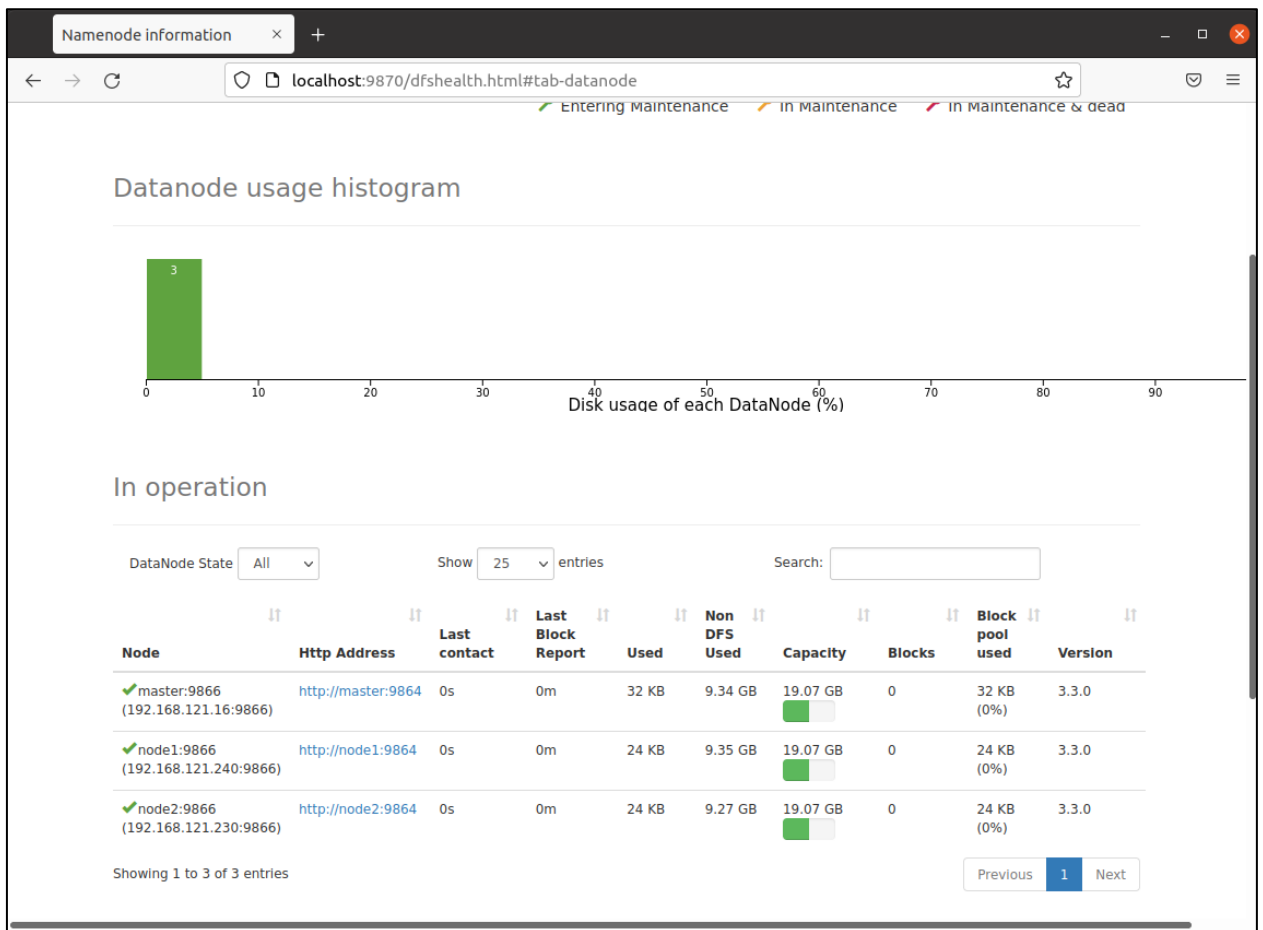
Запускаем start-dfs.sh и start-yarn.sh на мастере

Проверяем количество узлов:

hdfs dfsadmin -report

```
hduser@master:~$ hdfs dfsadmin -report
Configured Capacity: 61415964672 (57.20 GB)
Present Capacity: 28200980480 (26.26 GB)
DFS Remaining: 28200890368 (26.26 GB)
DFS Used: 90112 (88 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 (3):
```

Переходим в браузер:



## Задание 2.

Подключаемся по ssh с мастера к node1

ssh node1

```
hduser@master:~$ ssh node1
Welcome to Ubuntu 20.04.3 LTS (GNU/Linux 5.11.0-43-generic x86_64)

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

0 updates can be applied immediately.

Your Hardware Enablement Stack (HWE) is supported until April 2025.
Last login: Wed Dec 22 12:25:13 2021 from 192.168.121.16
```

Создаем папку

`hadoop fs -mkdir hdfs://master:9000/Kokarev`

```
hduser@node1:~$ hadoop fs -ls /
Found 1 items
drwxr-xr-x  - hduser supergroup          0 2021-12-22 12:55 /Kokarev
```

Hadoop
Overview
Datanodes
Datanode Volume Failures
Snapshot
Startup Progress
Utilities

## Browse Directory

Show  entries
Search:

<input type="checkbox"/>	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, 2020.

Создаем файл и кладем в ранее созданную папку:

```
hadoop fs -put /home/hduser/Desktop/ttt/test.txt hdfs://master:9000/Kokarev
```

Browsing HDFS
+

localhost:9870/explorer.html#/Kokarev

Hadoop
Overview
Datanodes
Datanode Volume Failures
Snapshot
Startup Progress
Utilities

## Browse Directory

Show  entries
Search:

<input type="checkbox"/>	Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name
<input type="checkbox"/>	-rw-r--r--	hduser	supergroup	16 B	Dec 25 07:20	3	128 MB	test.txt

Showing 1 to 1 of 1 entries

Previous
1
Next

Hadoop, 2020.