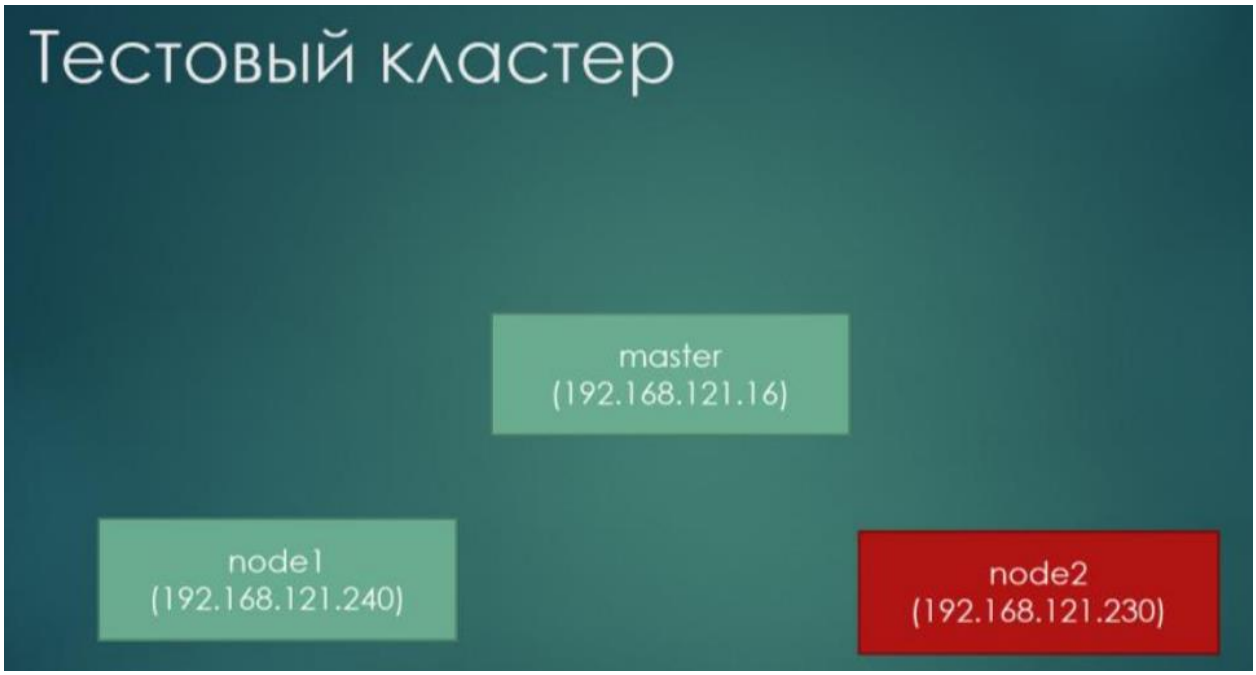
Лабораторная работа №2. Кокарев Д. В. РИМ-201211

Реализовать схему тестового кластера:



Установка ssh:

sudo apt-get install ssh

Установка java:

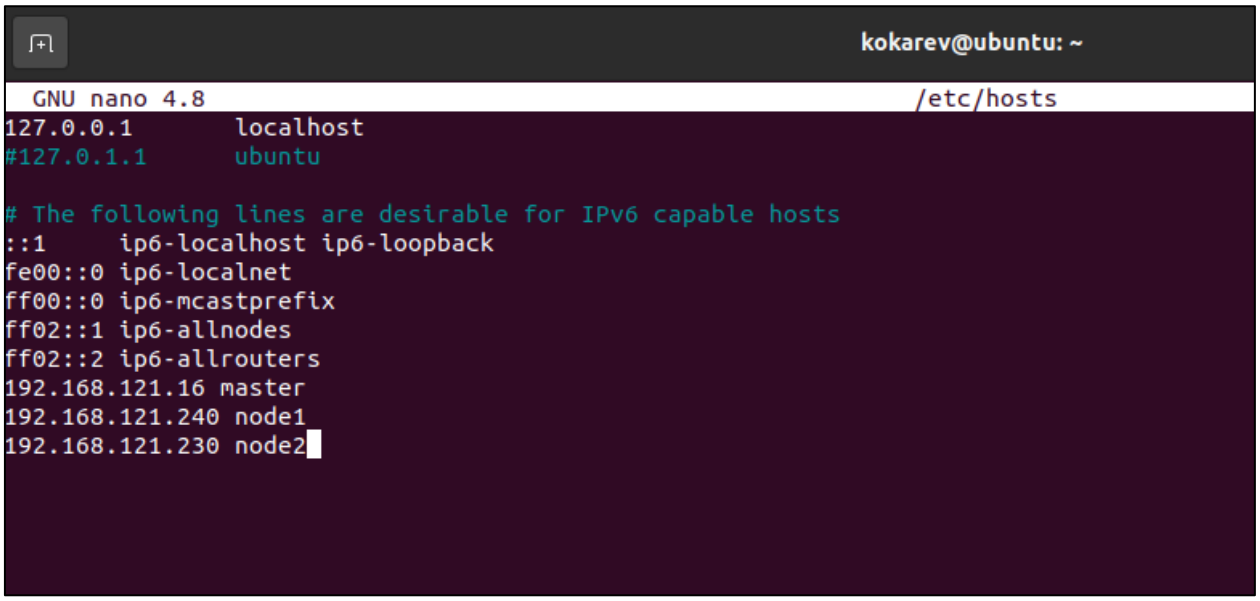
sudo apt-get install openjdk-8-jdk

Создание отдельной учетной записи для запуска hadoop:

sudo addgroup hadoop
sudo adduser --ingroup hadoop hduser
sudo usermod -aG sudo hduser

Редактирование файла /etc/hosts

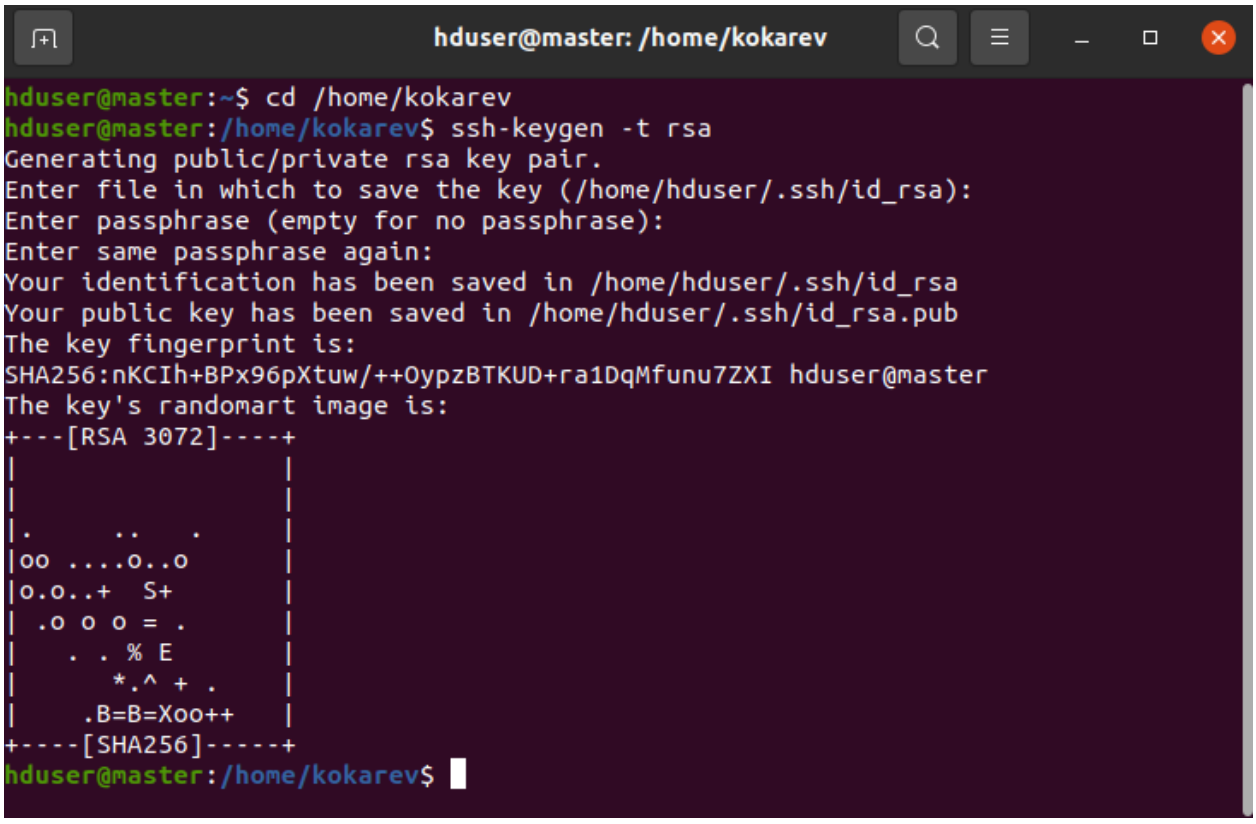
sudo nano /etc/hosts



Далее заходим и продолжаем под hduser

Получение SSH ключей

ssh-keygen -t rsa



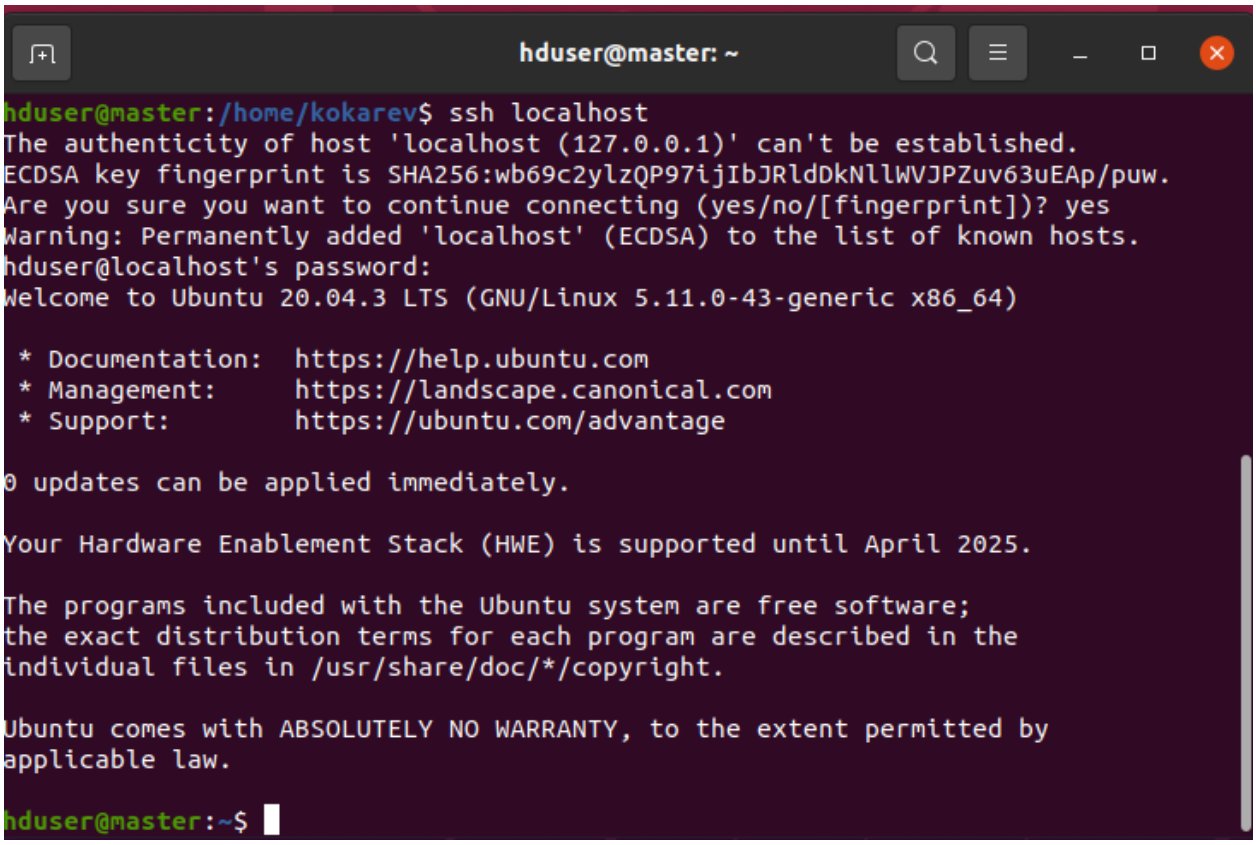
Перемещаем полученные ключи:

cat ~/.ssh/id_rsa.pub >> ~/.ssh/authorized_keys

chmod 0600 ~/.ssh/authorized_keys

Проверка подключения по ssh (ssh localhost)

ssh localhost



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

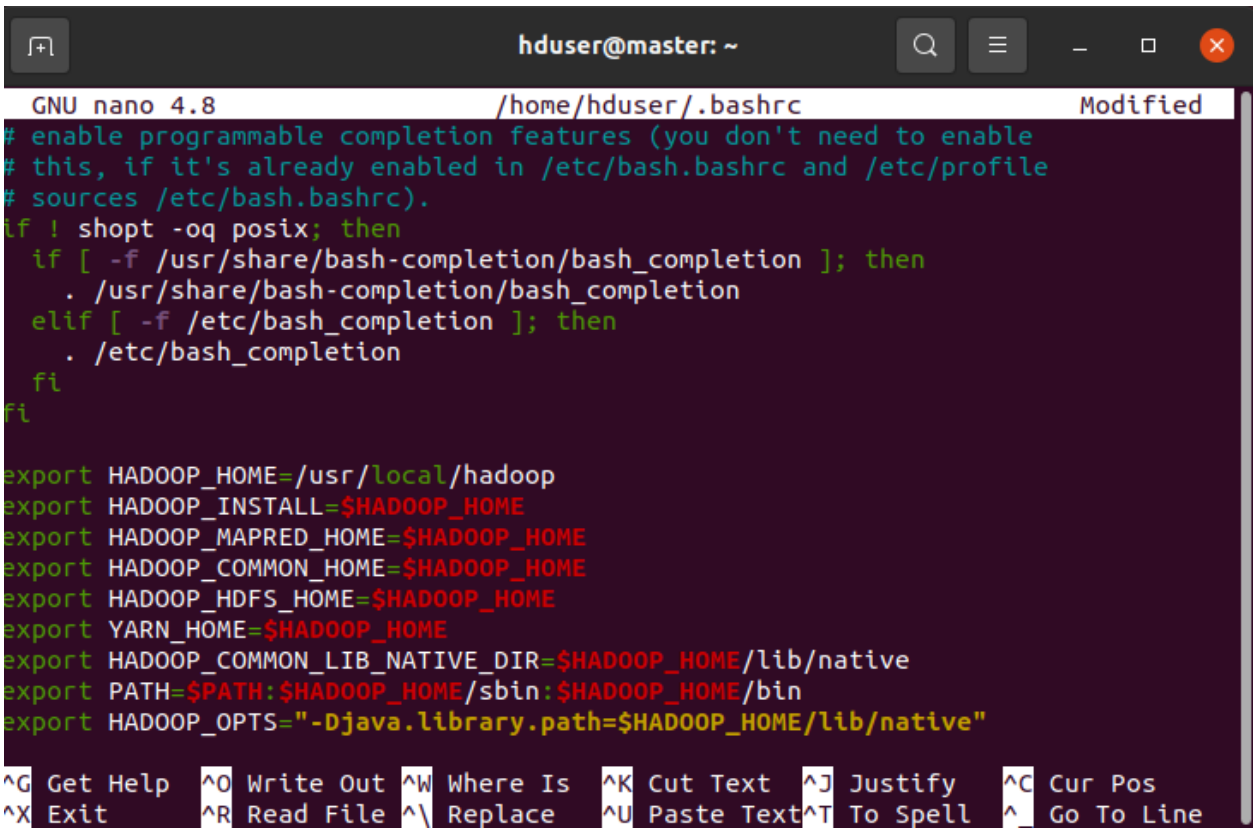
sudo mkdir /usr/local/hadoop/logs

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

Настройтка Apache Hadoop:

Настройка переменных среды в файле ~/.bashrc

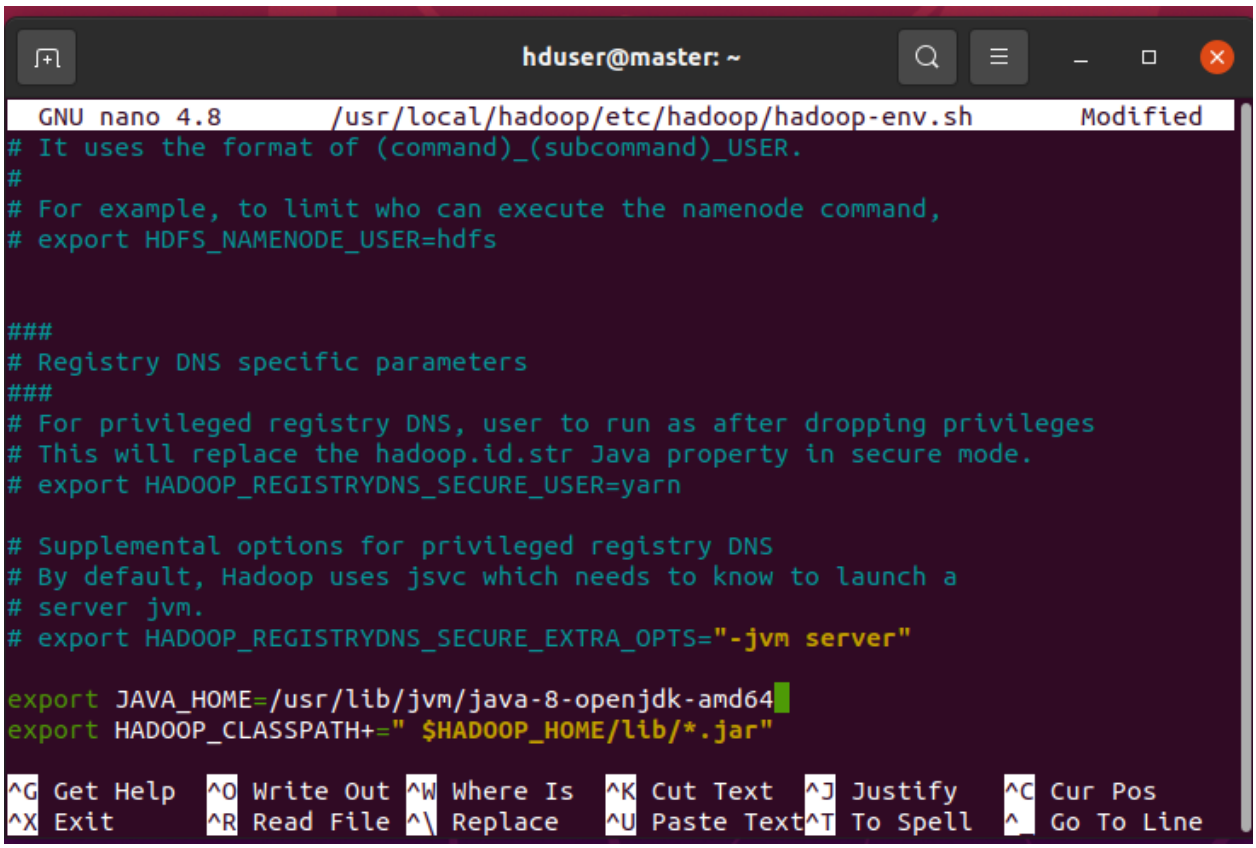
sudo nano ~/.bashrc



source ~/.bashrc

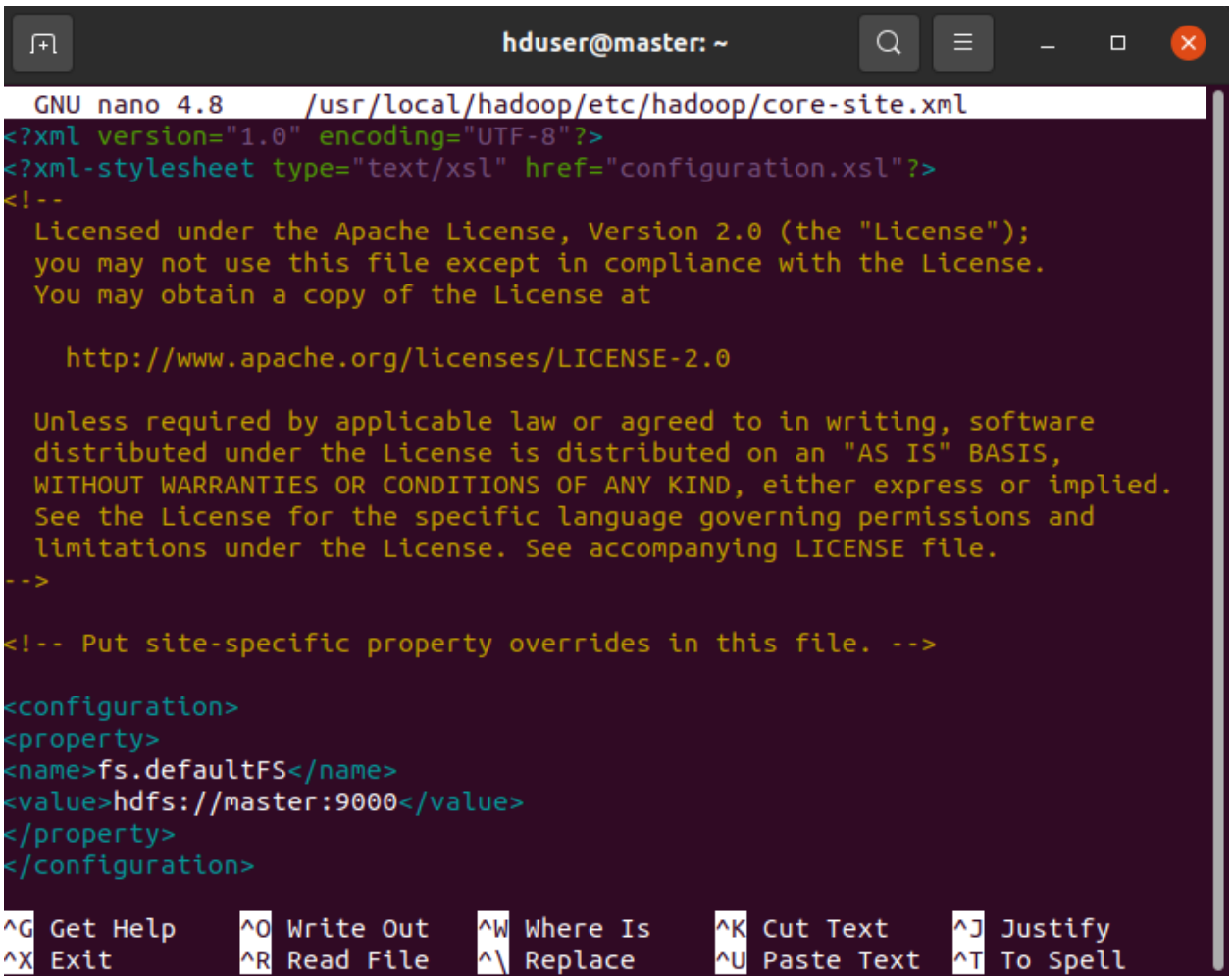
Затем нужно определить переменные среды Java, чтобы настроить параметры проекта, связанные с YARN, HDFS, MapReduce и Hadoop:hadoopenv.sh

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

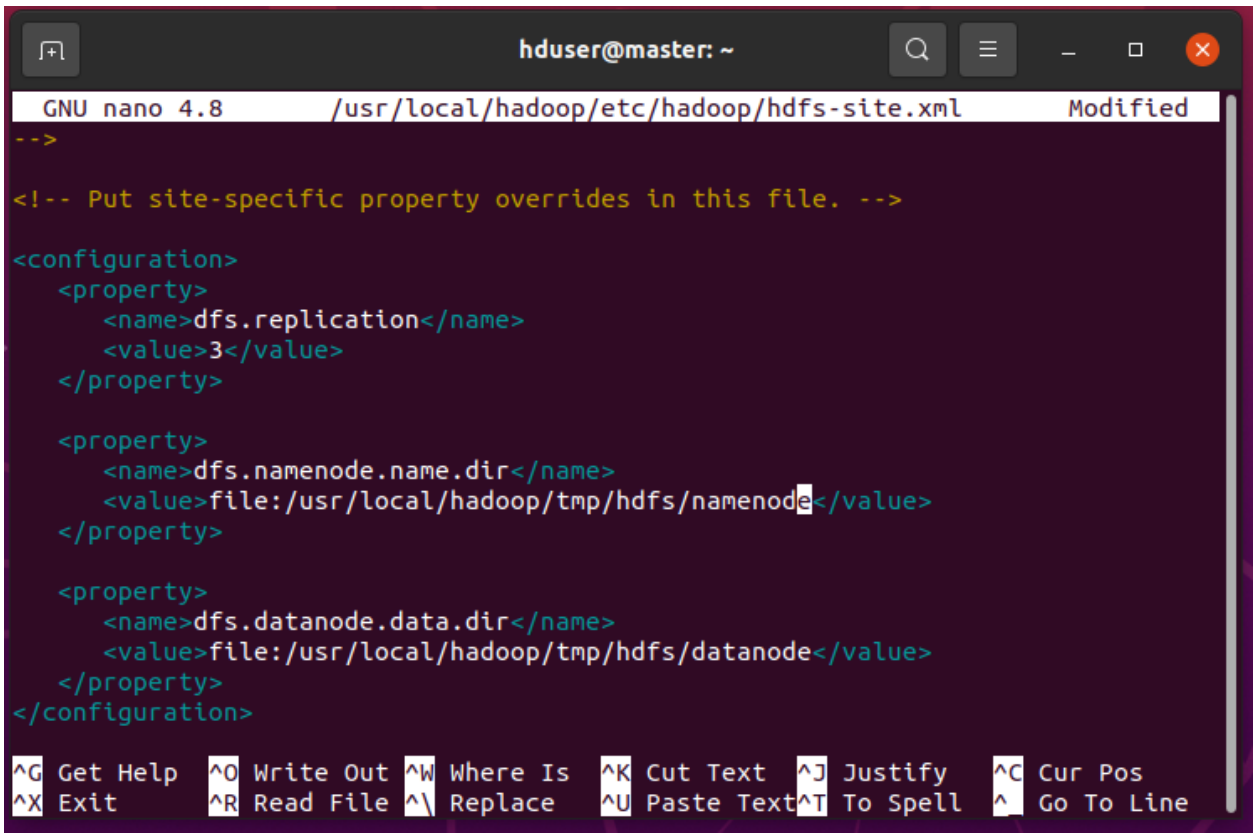


Настройка файлов

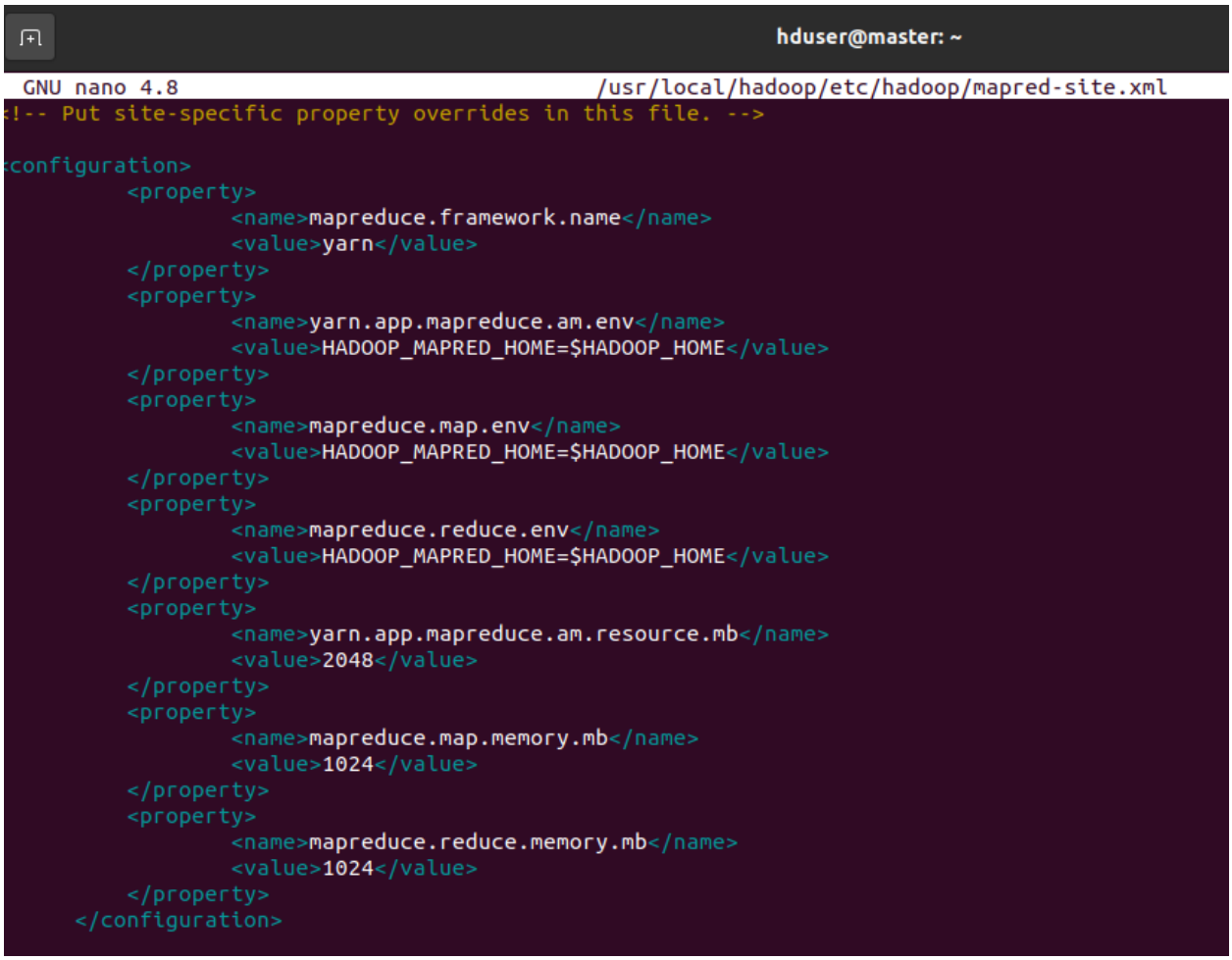
sudo nano /usr/local/hadoop/etc/hadoop/core-site.xml



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



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

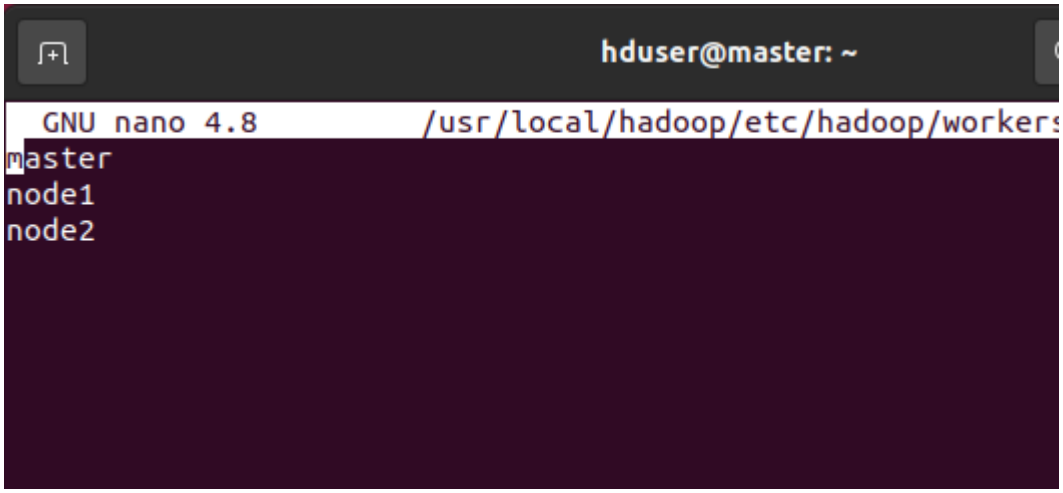


sudo nano /usr/local/hadoop/etc/hadoop/yarn-site.xml

```
<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



Форматирование HDFS NameNode

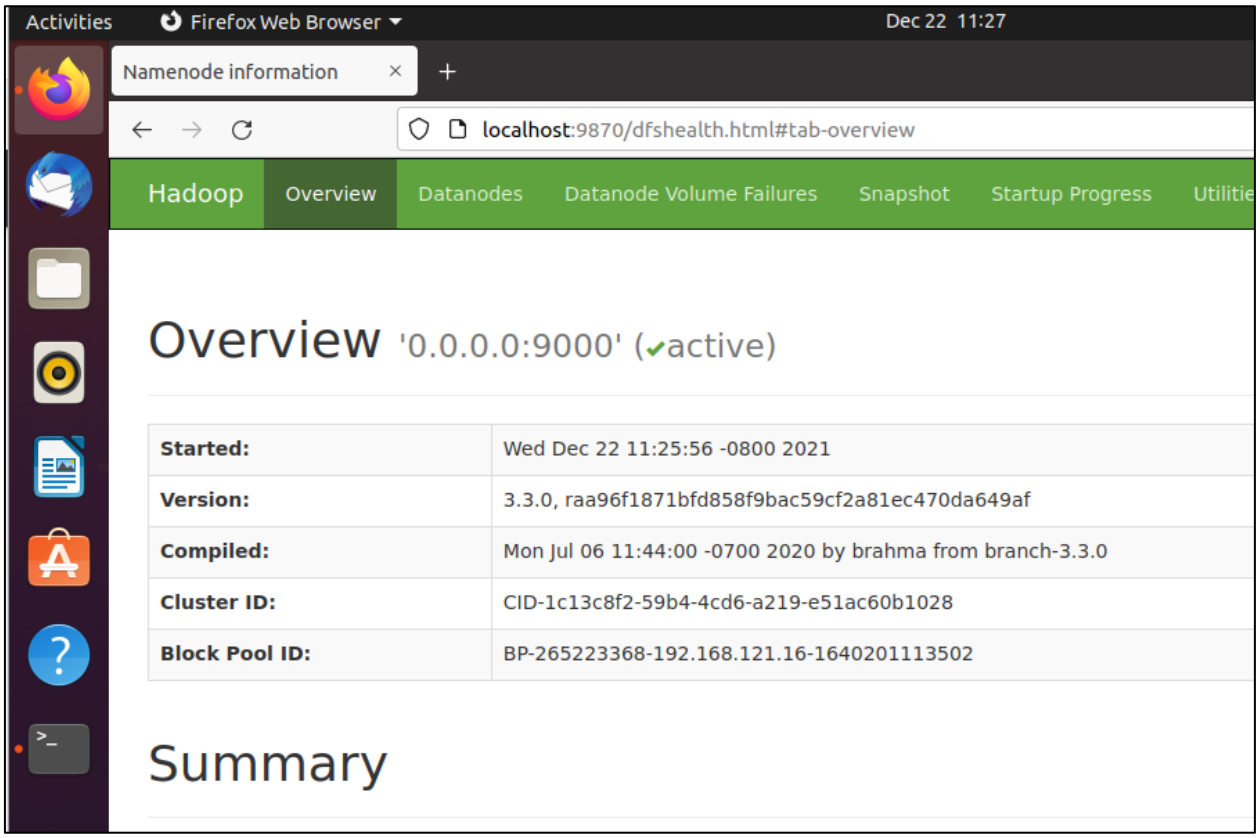
su - hduser

hdfs namenode -format

Запуск кластера Hadoop

start-dfs.sh

start-yarn.sh



Задание 1:

На мастере:

Каталог namenode создался 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

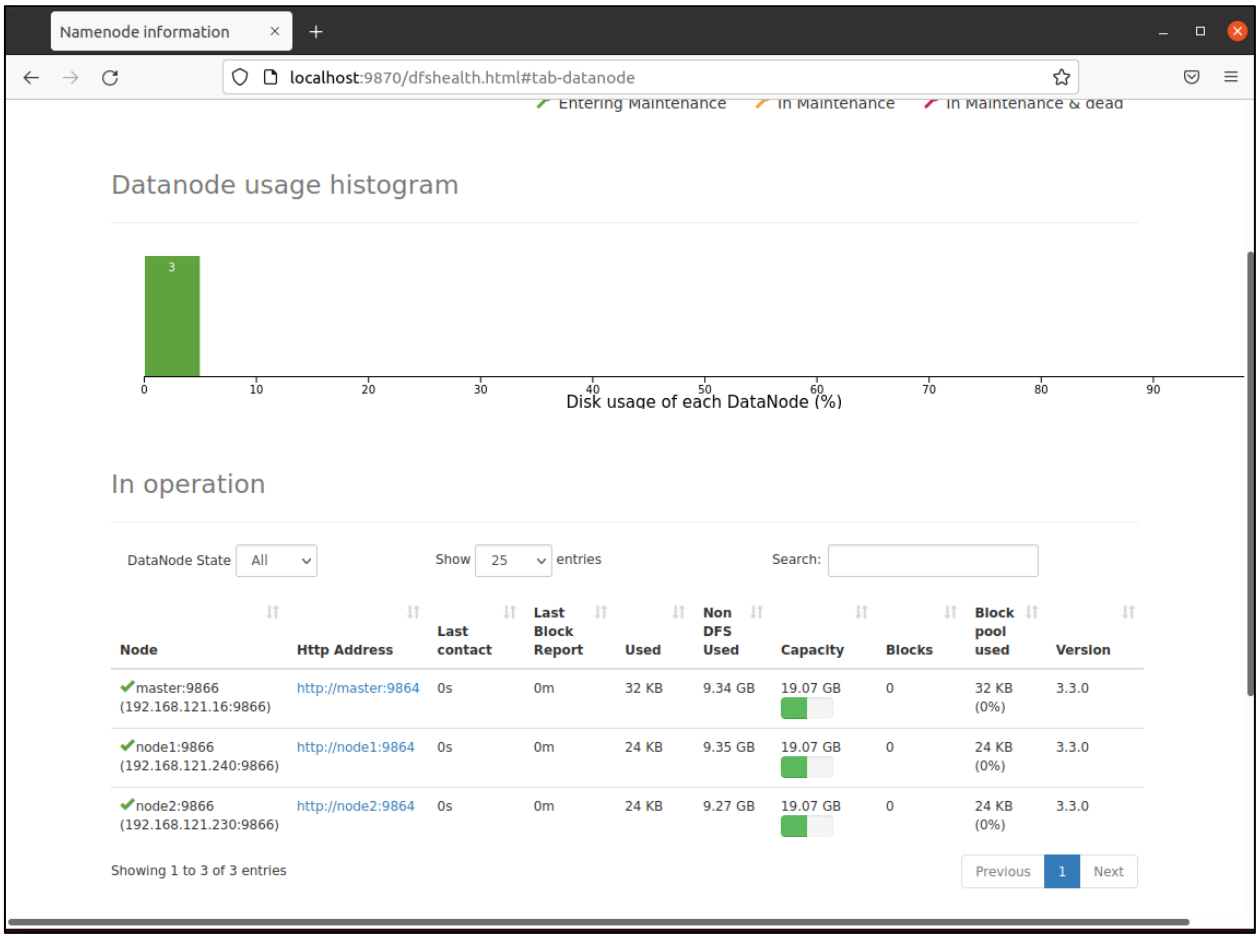
Запускаем start-dfs.sh и start-yarn.sh на мастере

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

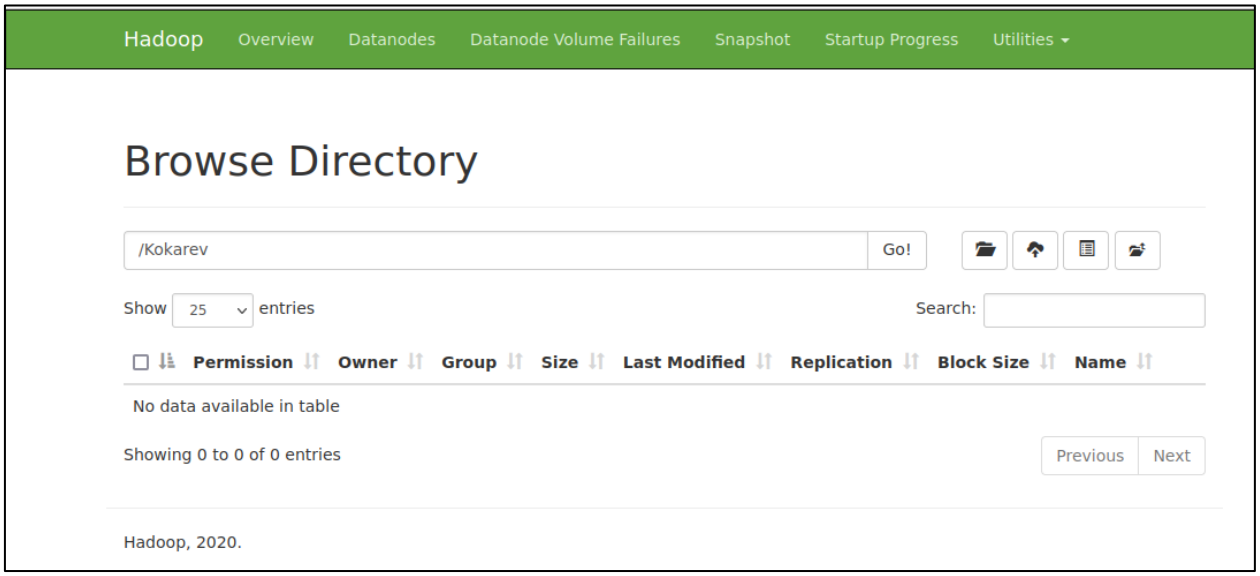
ssh node1

```
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
```



Создаем файл и кладем в ранее созданную папку:

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