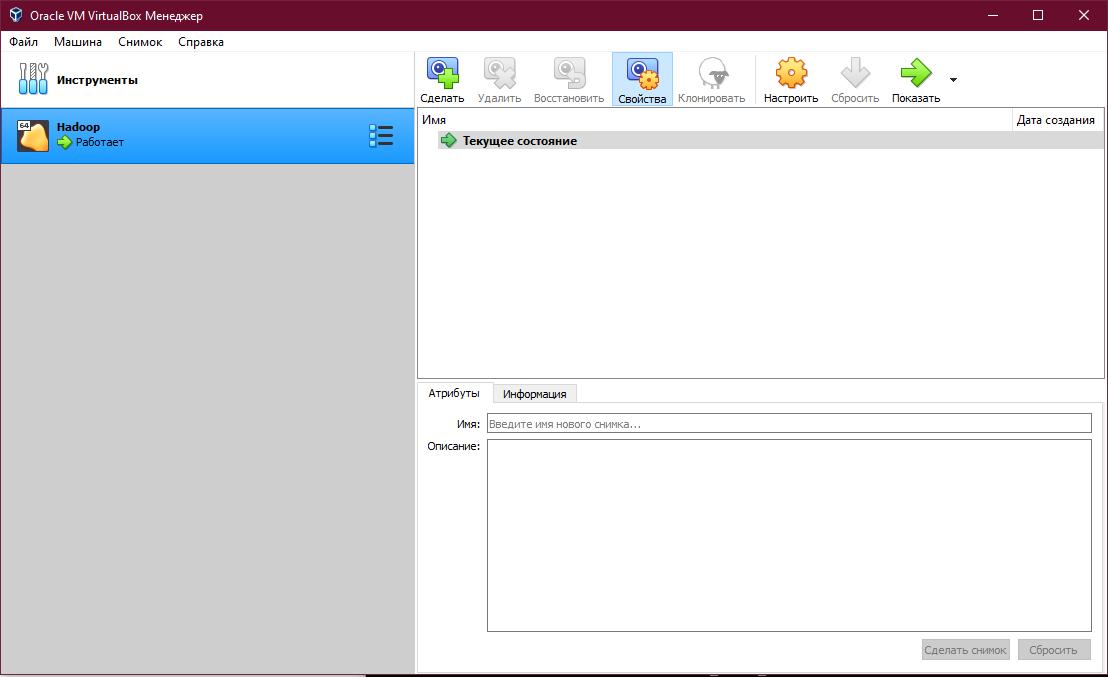
1. Создать новую виртуальную машину в Virtual Box

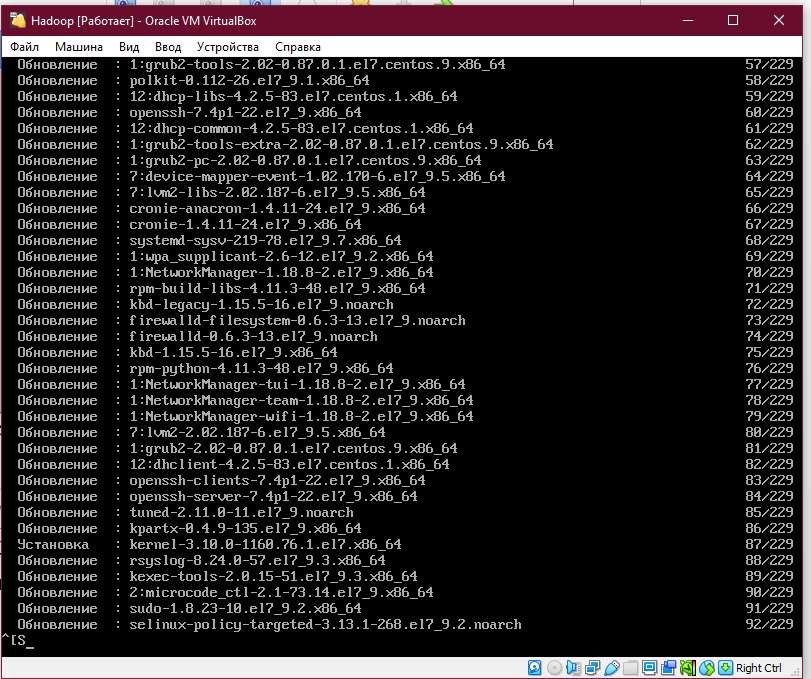


При помощи гуи создал новую машину

1. Уcтановить Linux (рекомендуется CentOS 7)

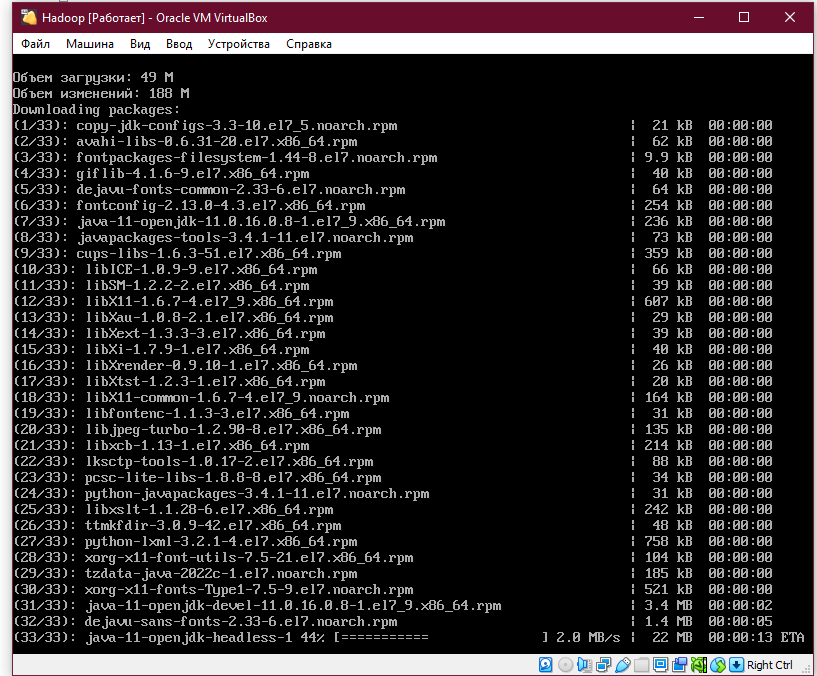
Скасал минимал образ CentOS 7 с оф сайта, установил, переконфигурировал /etc/sysconfig/network-scripts/ifcfg-enp0s3 при помощи vi, чтобы интерфейс запускиался при старте и сразу его запустил вручную: ifup enp0s3

Сразу же выполнил yum update



1. Установить JDK актулальной версии

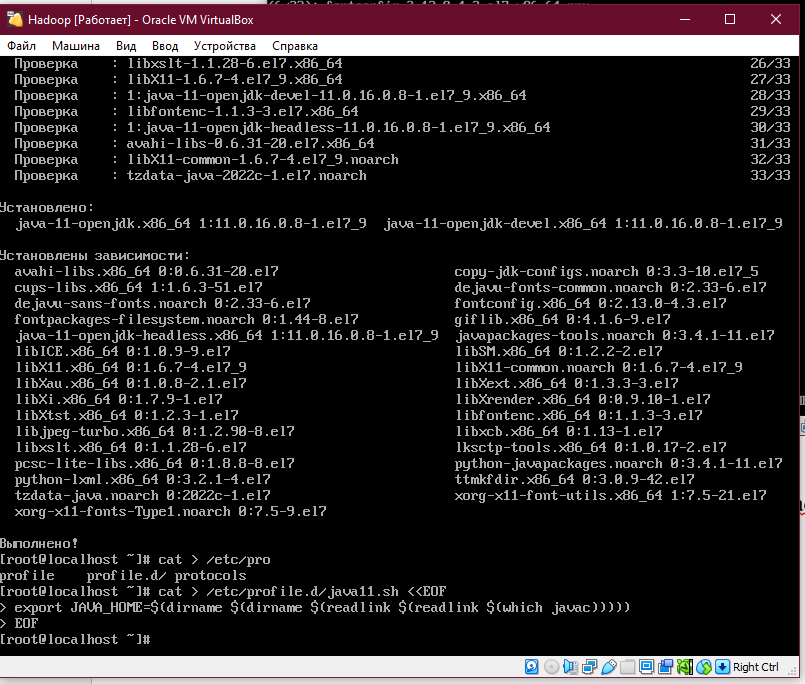
yum -y install java-11-openjdk java-11-openjdk-devel



cat > /etc/profile.d/java11.sh <<EOF

export JAVA\_HOME=$(readlink $(which java))

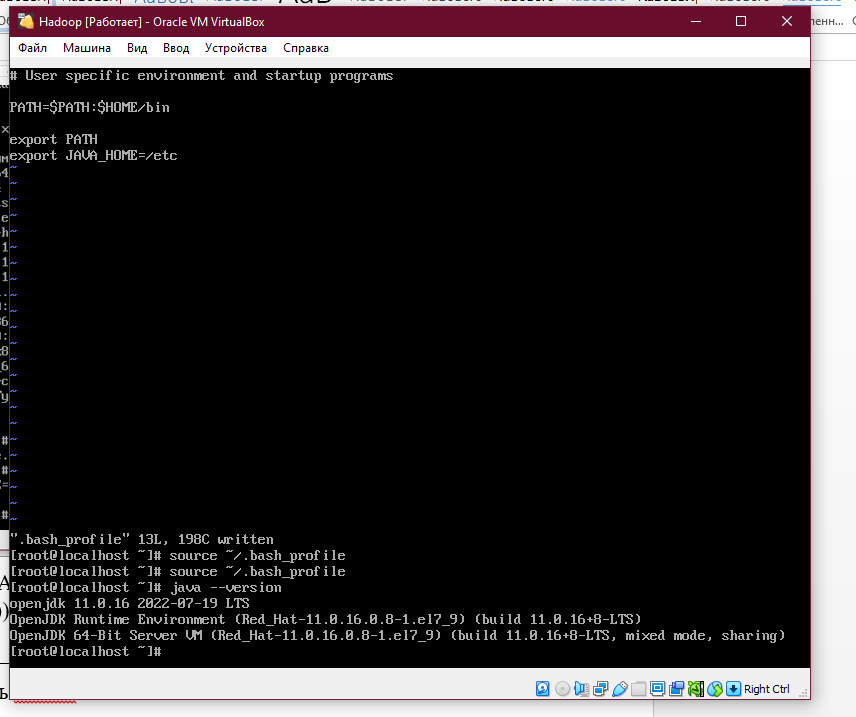
EOF



echo "export JAVA\_HOME=readlink –f (which java)" >> ~/.bash\_profile

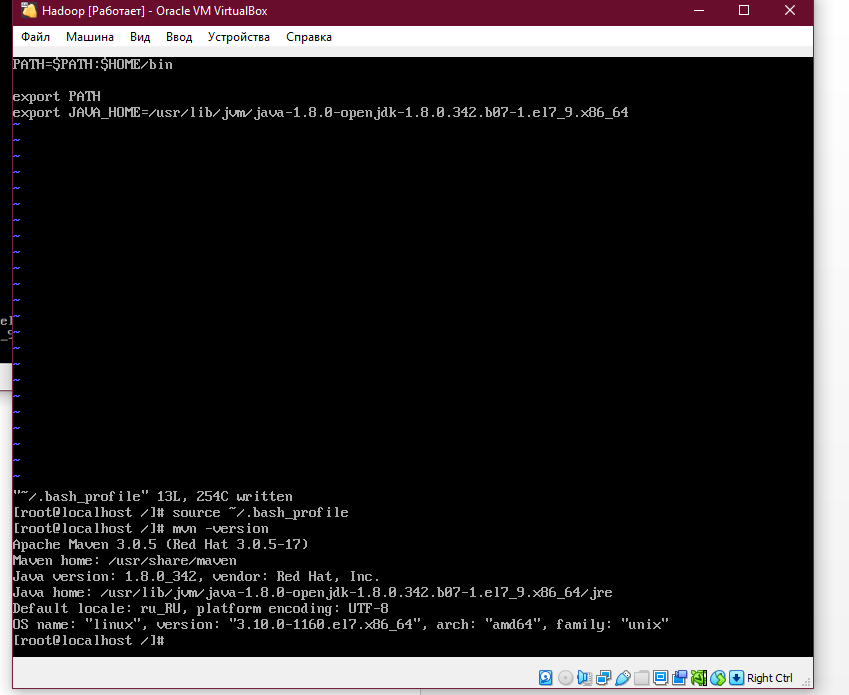
через vi чуть подправил

source ~/.bash\_profile



1. Установить Maven

yum install maven



1. Установить компилятор С++, а также необходимые инструменты: gcc gcc-c++ autoconf automake libtool cmake

yum group install "Development Tools"

yum install cmake

1. Установить доп. библиотеки zlib-devel openssl openssl-devel snappy snappy-devel bzip2 bzip2-devel protobuf protobuf-devel

Yum install –y x where x lib name

1. Скачать, собрать и устанавливить Hadoop

cd ~

wget http://apache.rediris.es/hadoop/common/hadoop-3.3.4/hadoop-3.3.4-src.tar.gz

tar -xvf ~/hadoop-3.3.4-src.tar.gz

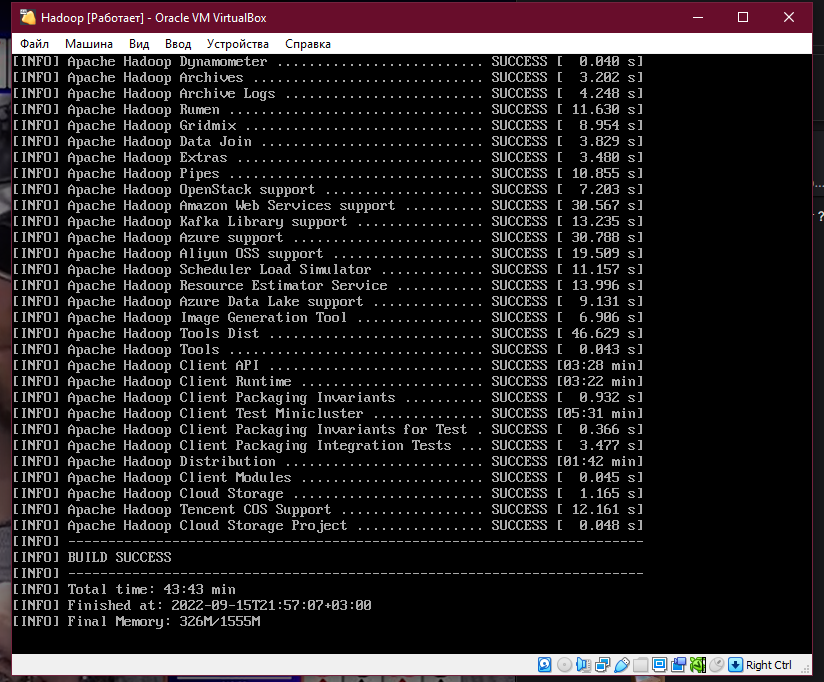
mv ~/hadoop-3.3.4-src ~/hadoop-src

cd ~/hadoop-src

mvn package -Pdist,native,docs -DskipTests –Dtar

(+ поставил более новые версии мавена и cmake)

(+ понял что мало места а у центос xfs )) пришлось увеличивать диск и моунтить новую партицию, сразу на 24 гигабайта, могли бы и предупредить)



tar -C/haddop/opt -xvf /hadoop/hadoop-src/hadoop-dist/target/hadoop-3.3.4.tar.gz

mv /hadoop/opt/hadoop-\* /hadoop/opt/hadoop

echo "PATH=\"/hadoop/opt/hadoop/bin:\$PATH\"" >> ~/.bashrc

source ~/.bashrc

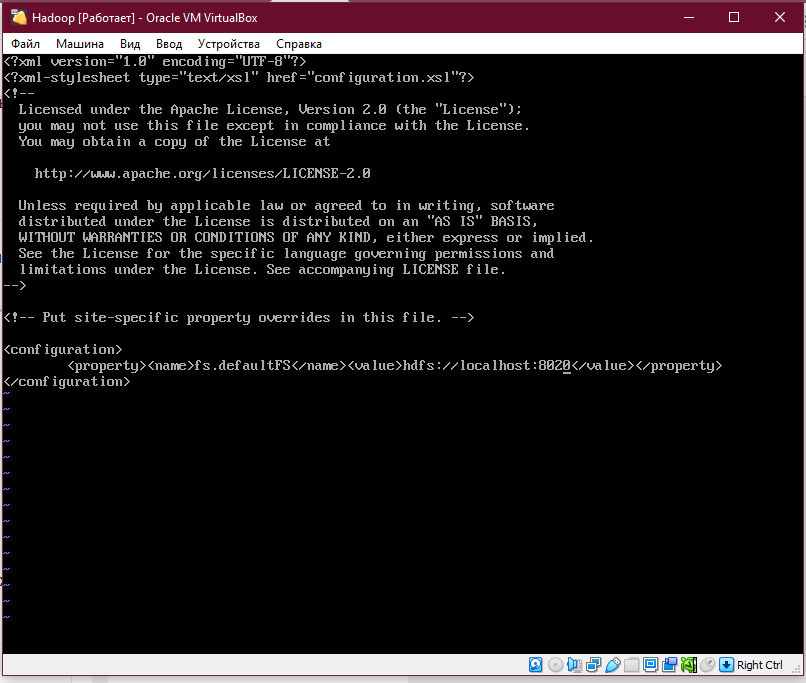
1. Прописать путь к JAVA\_HOME в настроечных файлах Hadoop

sed -i “1iJAVA\_HOME=$JAVA\_HOME” /opt/hadoop/etc/hadoop/hadoop-env.sh

sed -i “1iJAVA\_HOME=$JAVA\_HOME” /opt/hadoop/etc/hadoop/yarn-env.sh

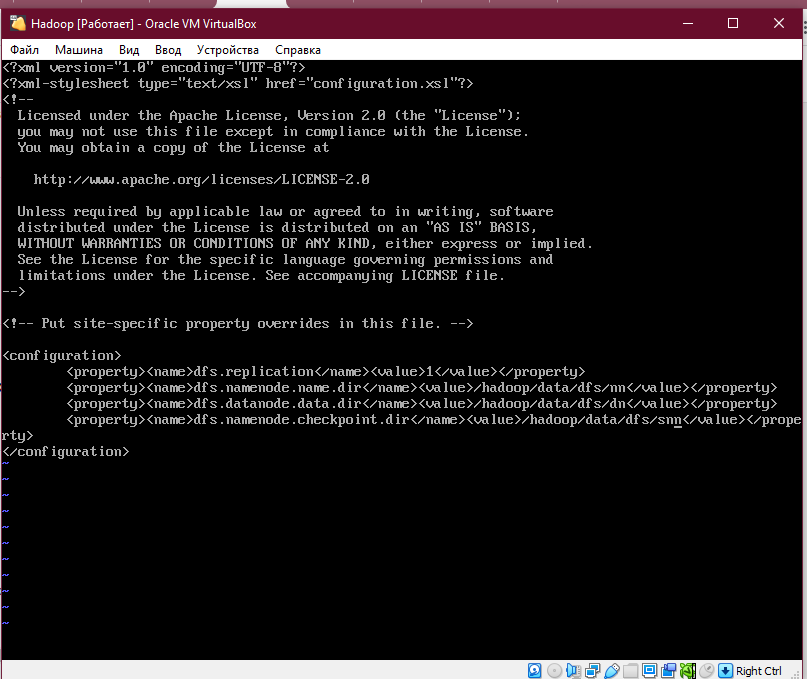
1. Сконфигурировать URL для обращения к Hadoop

vi /hadoop/opt/hadoop/etc/hadoop/core-site.xml



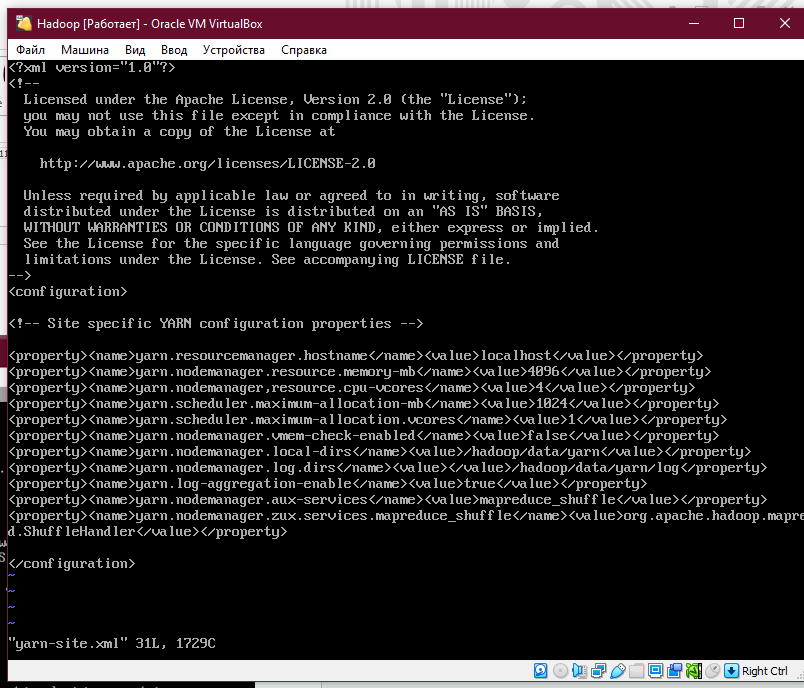
1. Настроить пораметры Hadoop: dfs.replication, dfs.namenode.name.dir, dfs.datanode.data.dir, dfs.namenode.checkpoint.dir

vi /hadoop/opt/hadoop/etc/hadoop/hdfs-site.xml



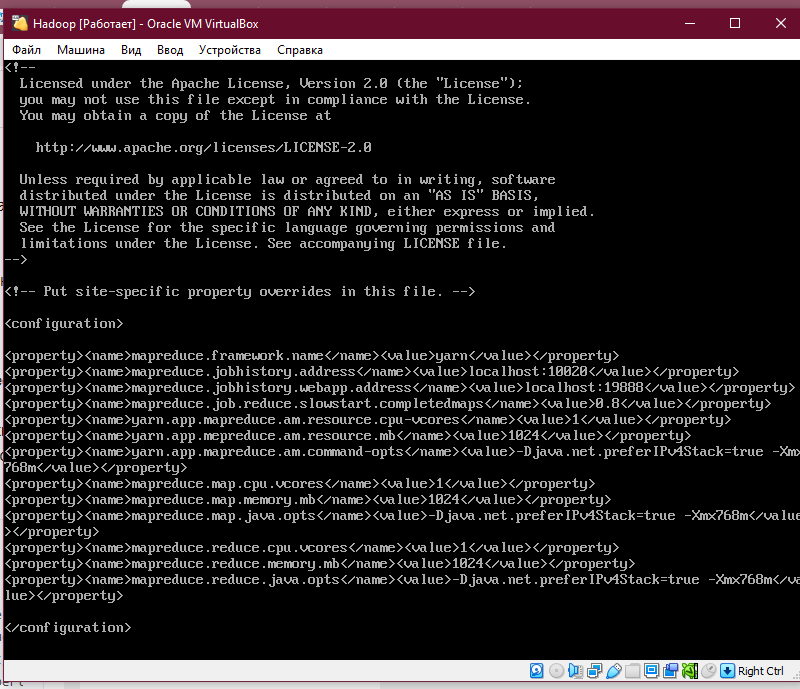
1. Настроить параметры YARN

vi /hadoop/opt/hadoop/etc/hadoop/yarn-site.xml



1. Настроить параметры фреймворка MapReduce

vi /hadoop/opt/hadoop/etc/hadoop/mapred-site.xml



1. Создать директорию для данных HDFS и параметров YARN

mkdir hadoop/data

1. Отформатировать HDFS

hadoop namenode –format

1. Запустить сервисы кластера:

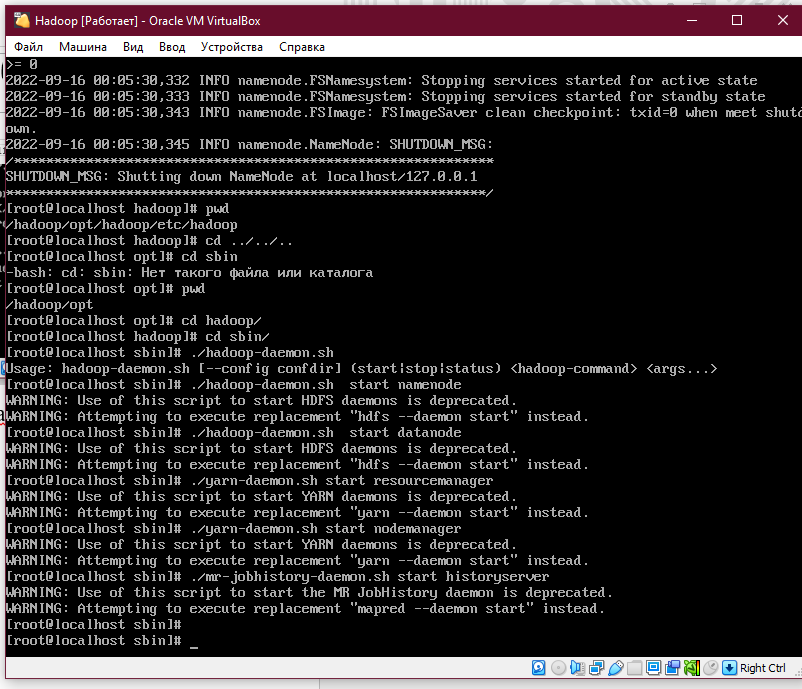
/opt/hadoop/sbin/hadoop-daemon.sh start namenode

/opt/hadoop/sbin/hadoop-daemon.sh start datanode

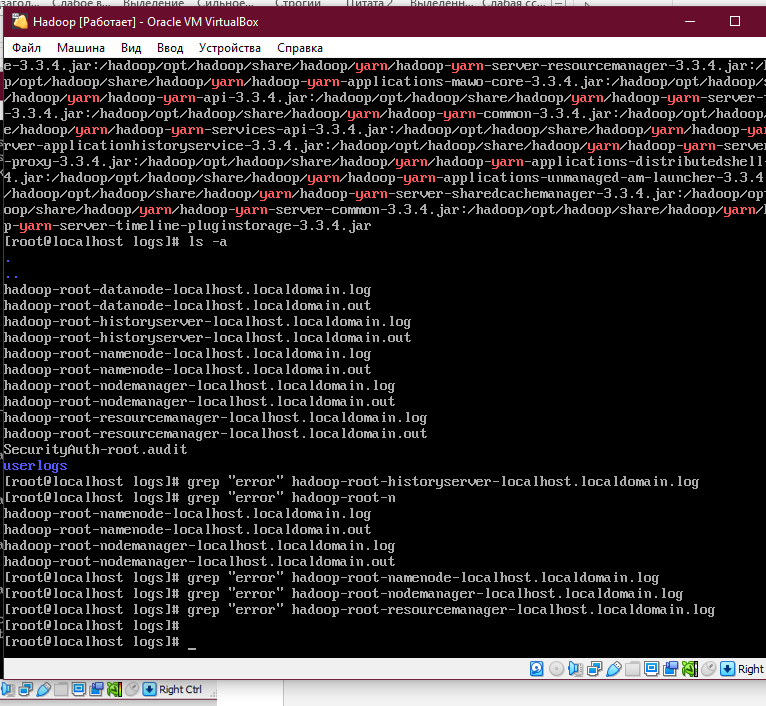
/opt/hadoop/sbin/yarn-daemon.sh start resourcemanager

/opt/hadoop/sbin/yarn-daemon.sh start nodemanager

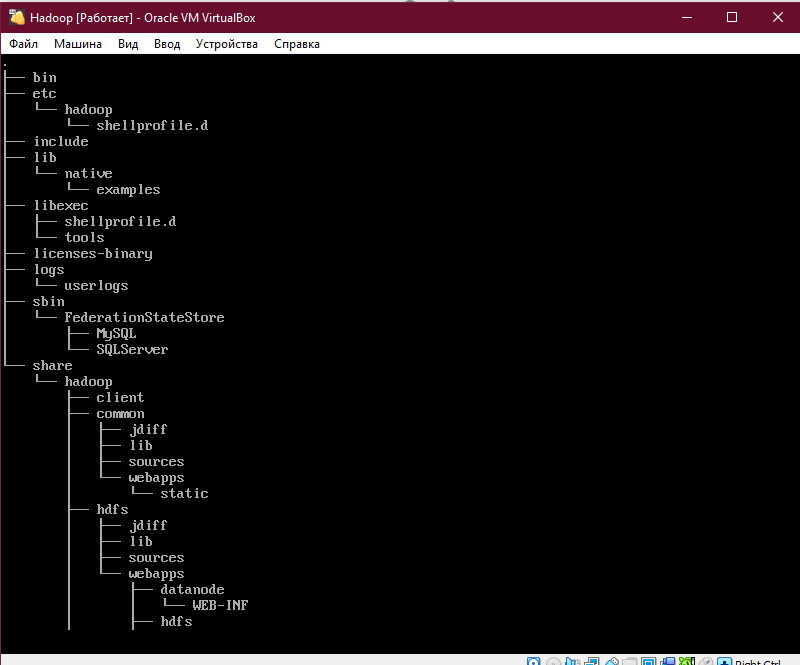
/opt/hadoop/sbin/mr-jobhistory-daemon.sh start historyserver



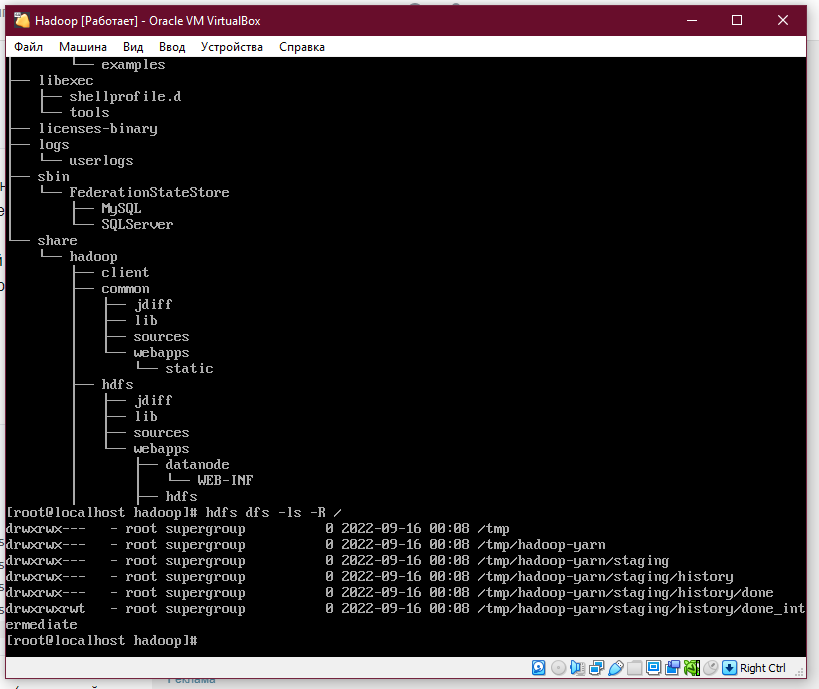
1. Убедиться в отсутствии ошибок в файлах log



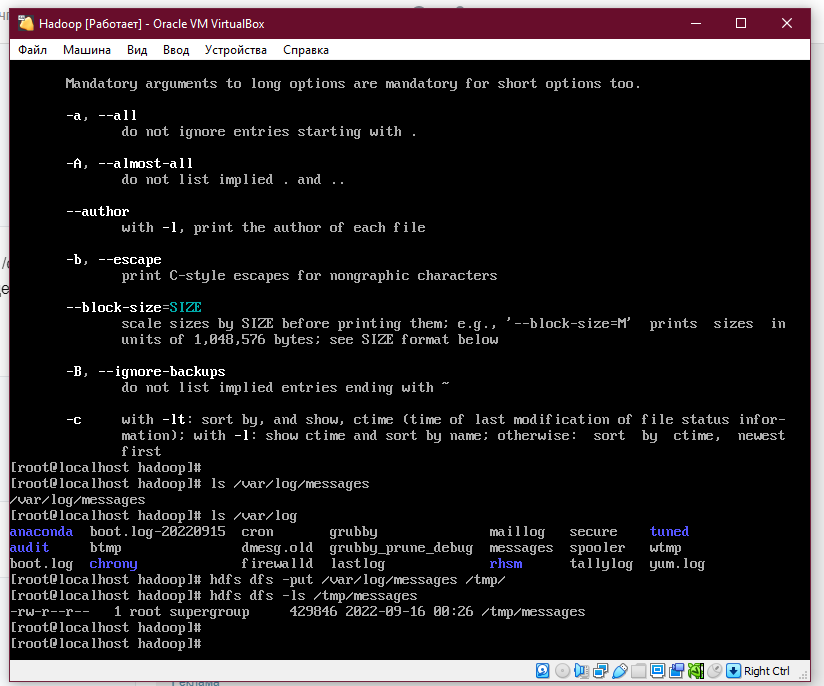
1. Отобразить содержимое директории Hadoop



1. Отобразить содержимое HDFS

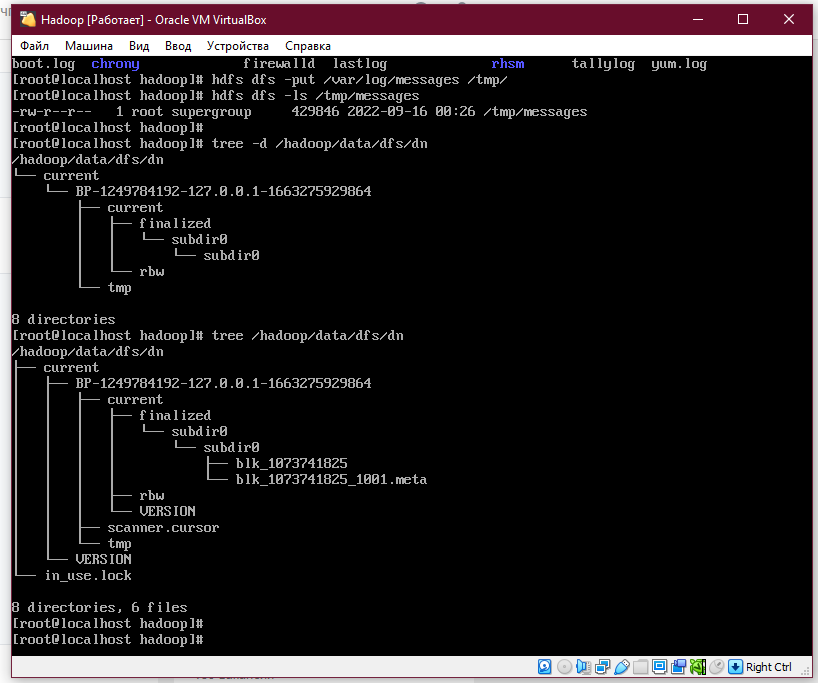


1. Скопировать несколько файлов из фаловой системы компьютера в HDFS

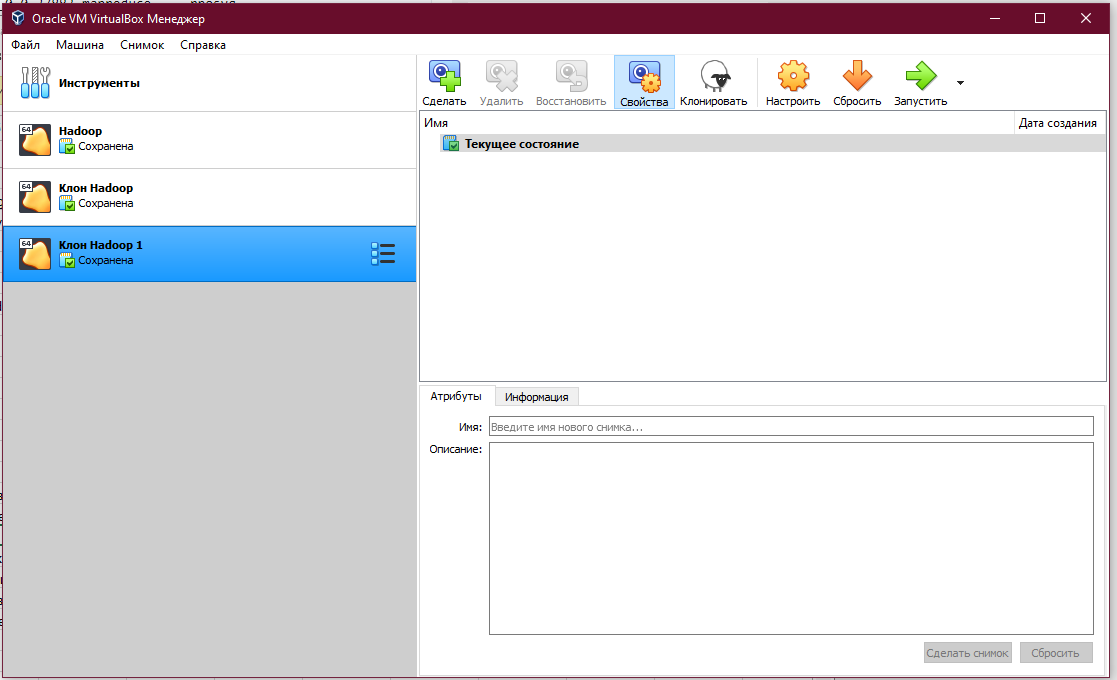


1. Изучить изменение физической структуры директорий при добавлении файлов в HDFS

Появился первый блок и его КС



1. Клонировать созданные виртуальные машины (мин. 3 экз)

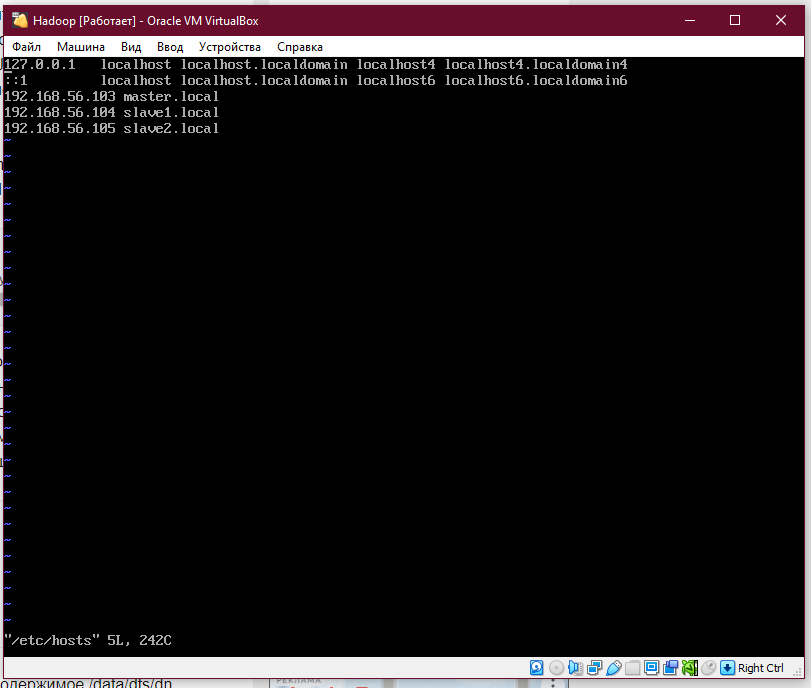


1. Изменить конфигурации Hadoop таким образом, чтобы одна из них стала NameNode, а остальные – DataNode

cd /hadoop/opt/hadoop/etc/hadoop

sed -i 's/localhost/master.local/' core-site.xml hdfs-site.xml yarn-site.xml mapred-site.xml

Также на всех трех нодах нашего кластера сконфигурируем /etc/hosts, чтоб каждая машина кластера могла обращаться к другим по имени хоста.



Дополнительно на нодах slave1.local и slave2.local нужно очистить содержимое /data/dfs/dn

Все готово. На master.local запускаем все сервисы:

/opt/hadoop/sbin/hadoop-daemon.sh start namenode

/opt/hadoop/sbin/hadoop-daemon.sh start datanode

/opt/hadoop/sbin/yarn-daemon.sh start resourcemanager

/opt/hadoop/sbin/yarn-daemon.sh start nodemanager

/opt/hadoop/sbin/mr-jobhistory-daemon.sh start historyserver

На slave1.local и slave2.local запускаем только DataNode и NodeManager:

/opt/hadoop/sbin/hadoop-daemon.sh start datanode

/opt/hadoop/sbin/yarn-daemon.sh start nodemanager

Проверим, что наш кластер теперь состоит из трех нод.

Для HDFS посмотрим на вывод команды dfsadmin -report и убедимся, что все три машины включены в список Live datanodes:

hdfs dfsadmin -report

Live datanodes (3):

Name: 192.168.122.70:50010 (master.local)

Name: 192.168.122.59:50010 (slave1.local)

Name: 192.168.122.217:50010 (slave2.local)

Или зайдем на веб страничку NameNode:

master.local:50070/dfshealth.html#tab-datanode

Для YARN посмотрим на вывод команды node -list:

yarn node -list -all

17/01/06 06:17:52 INFO client.RMProxy: Connecting to ResourceManager at master.local/192.168.122.70:8032

Total Nodes:3

Node-Id Node-State Node-Http-Address Number-of-Running-Containers

slave2.local:39694 RUNNING slave2.local:8042 0

slave1.local:36880 RUNNING slave1.local:8042 0

master.local:44373 RUNNING master.local:8042 0

Или зайдем на веб страничку ResourceManager

master.local:8088/cluster/nodes

Все ноды должны быть в списке со статусом RUNNING.

Напоследок убедимся, что запускаемые приложения MapReduce используют ресурсы на всех трех нодах. Запустим уже знакомое нам приложение Pi из hadoop-mapreduce-examples.jar:

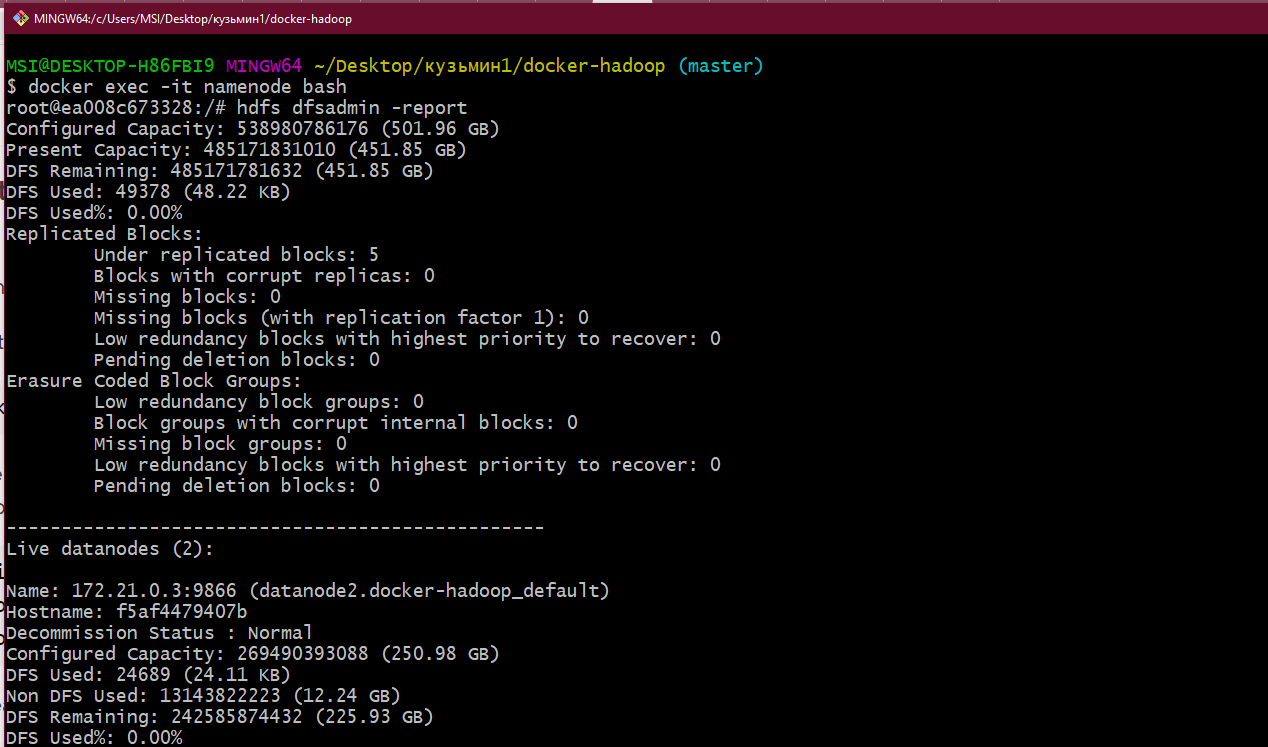
yarn jar /opt/hadoop/share/hadoop/mapreduce/hadoop-mapreduce-examples-2.7.3.jar pi 30 1000

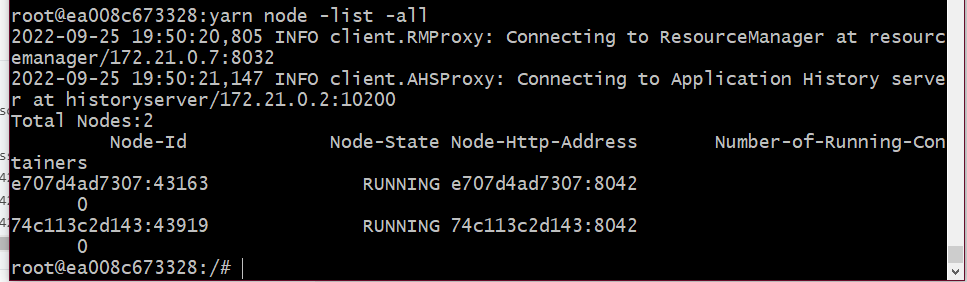
Во время выполнения приложения еще раз посмотрим вывод yarn node -list -all:

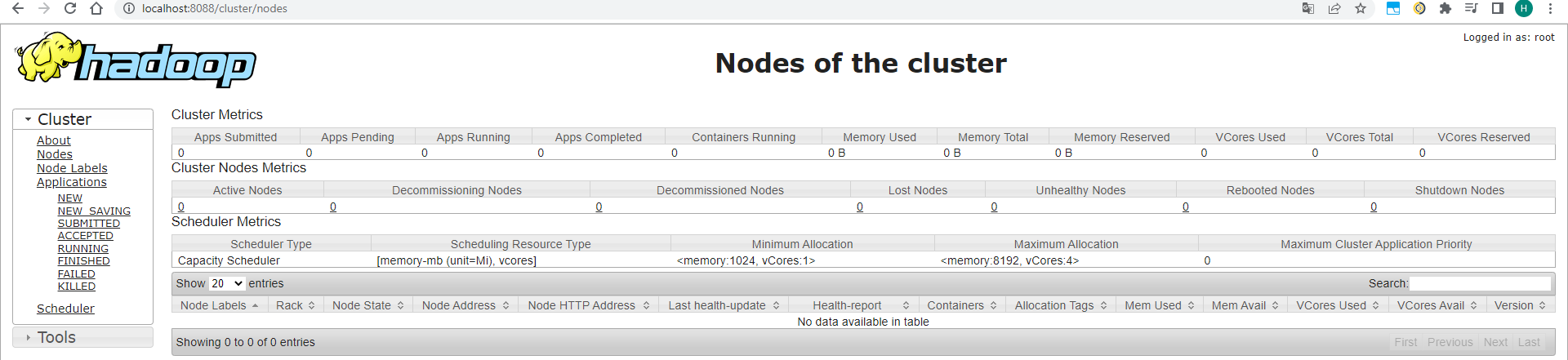
Number-of-Running-Containers — 4 на каждой ноде.

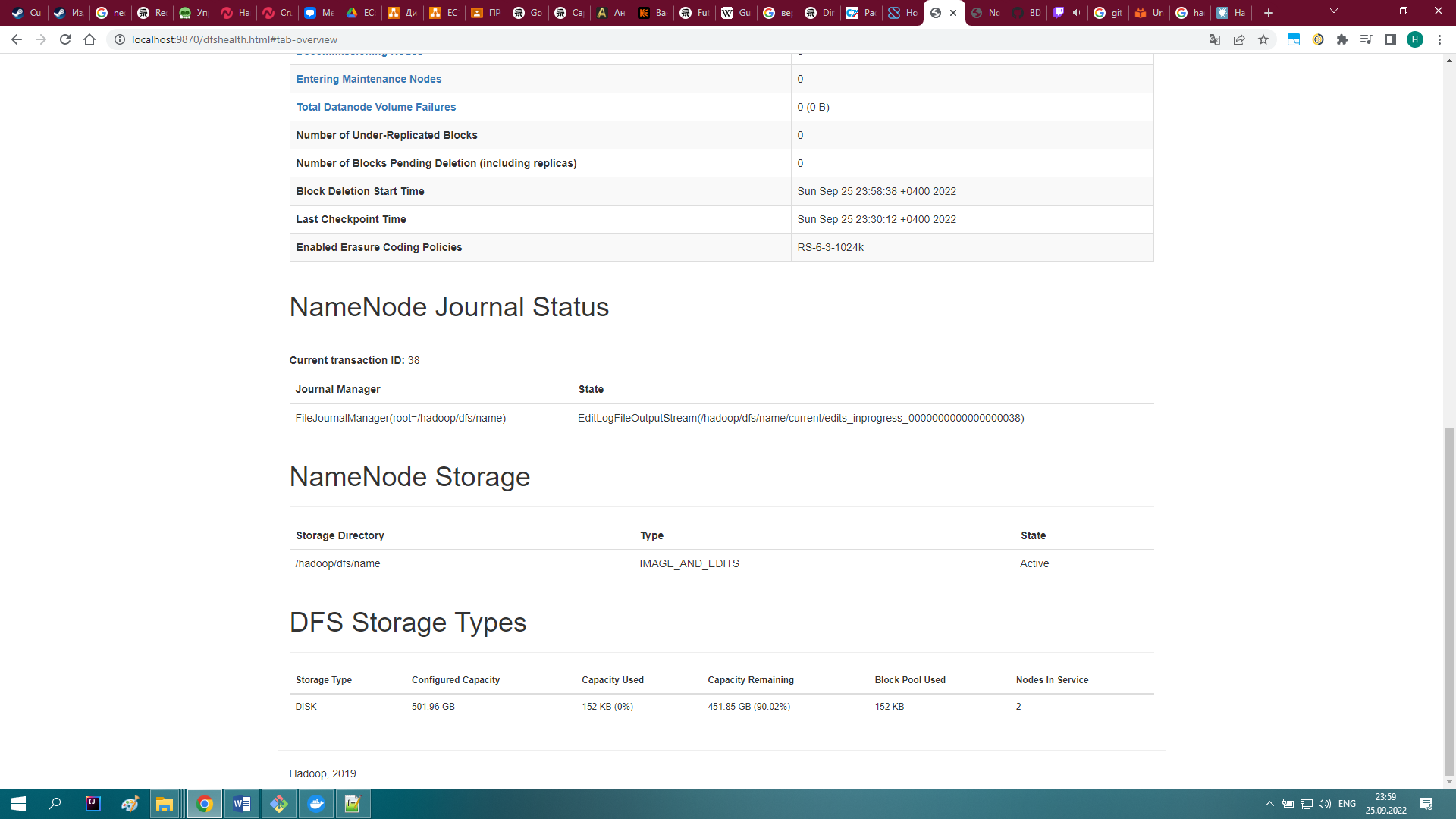
Также мы можем зайти на master.local:8088/cluster/nodes и посмотреть, сколько ядер и памяти используется всеми приложениями в общей сложности на каждой ноде.

1. Вывести на экран список узлов, из которых состоит кластер. Убедиться, что все из них запущены

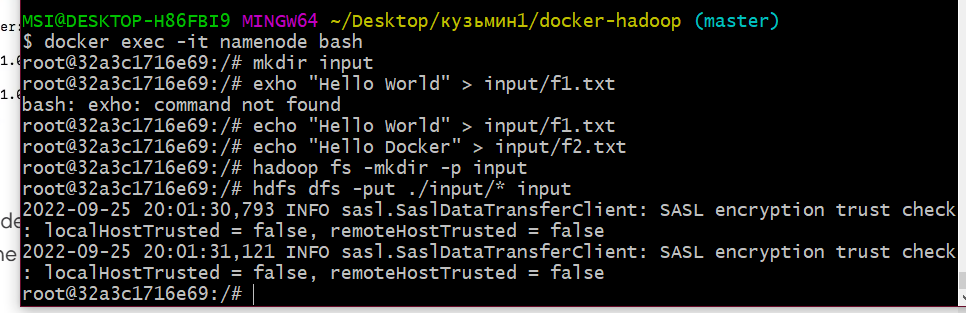




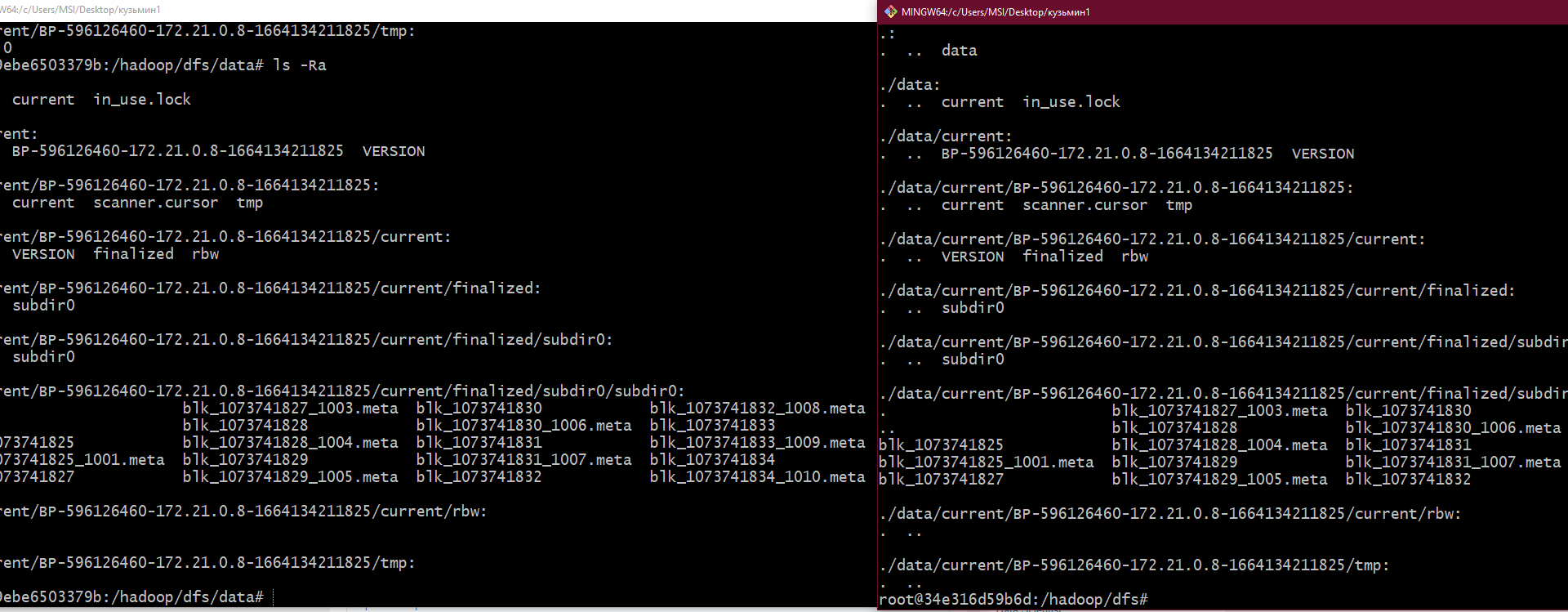


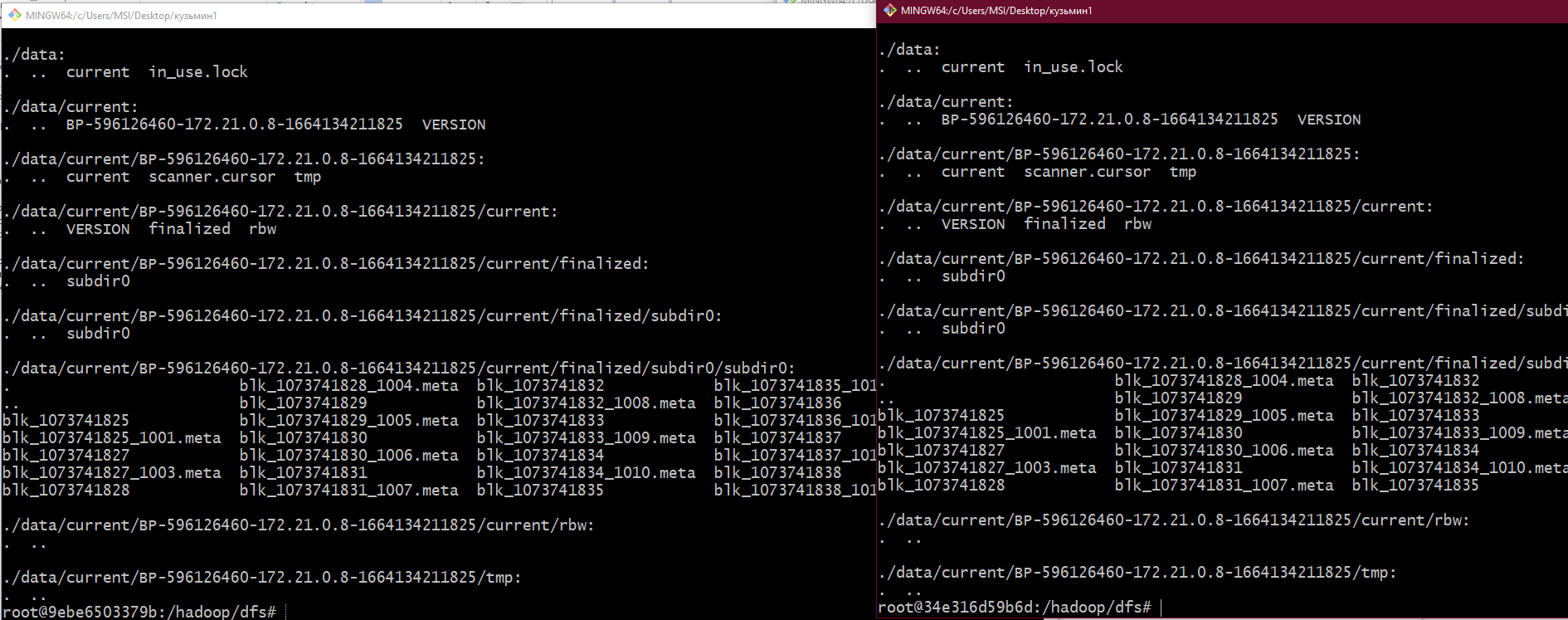


1. Скопировать несколько файлов из файловой системы компьютера в HDFS

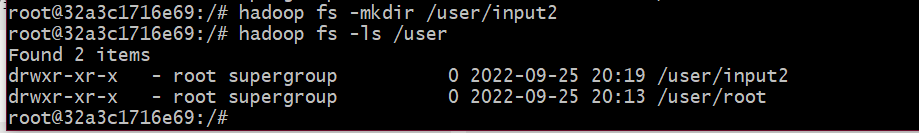


1. Изучить изменение физической структуры директорий при добавлении файлов в HDFS

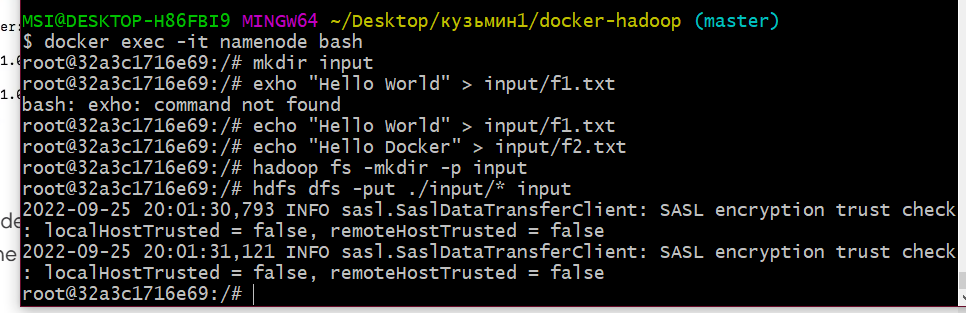




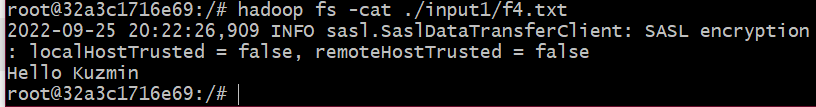
1. Выполнить базовые операции работы с файлами в HDFS



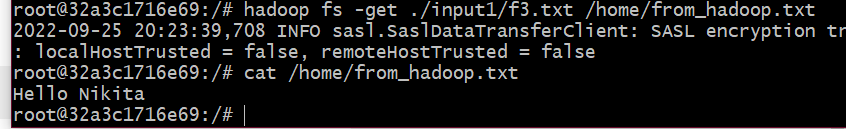
Mkdir + ls



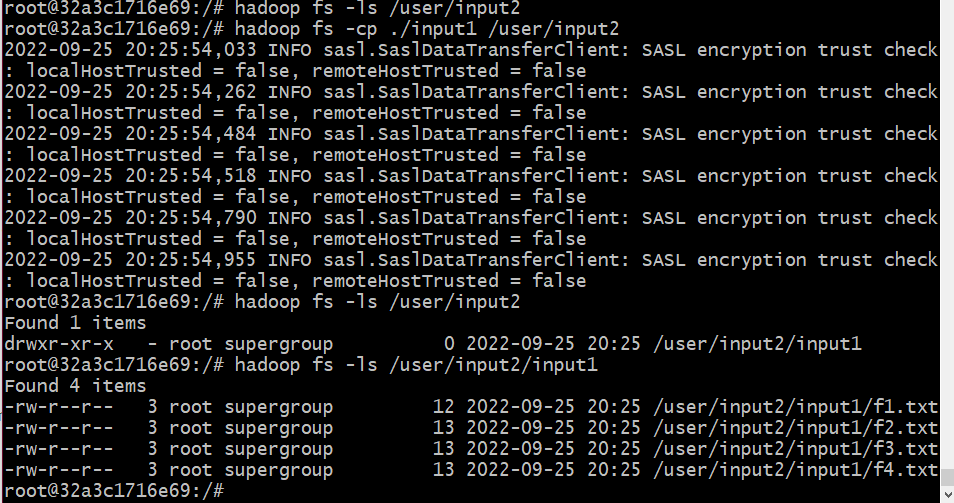
Put



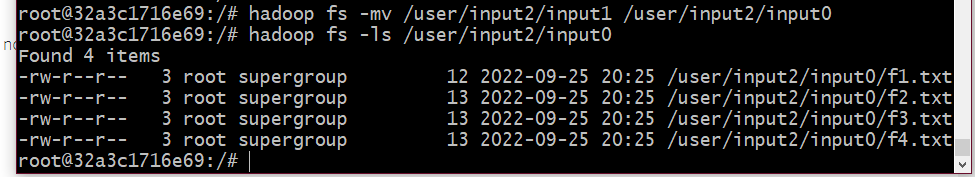
Cat



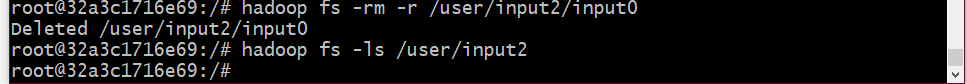
Get



Cp



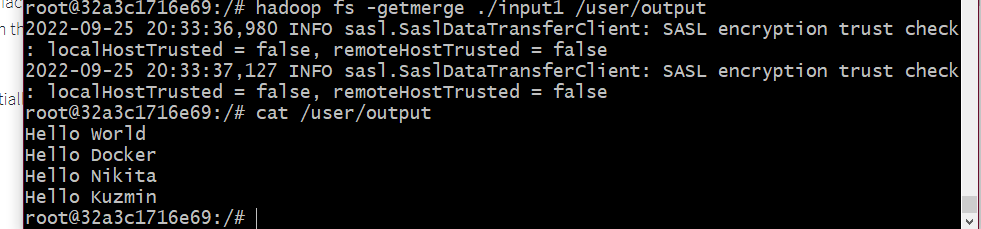
Mv



Rm



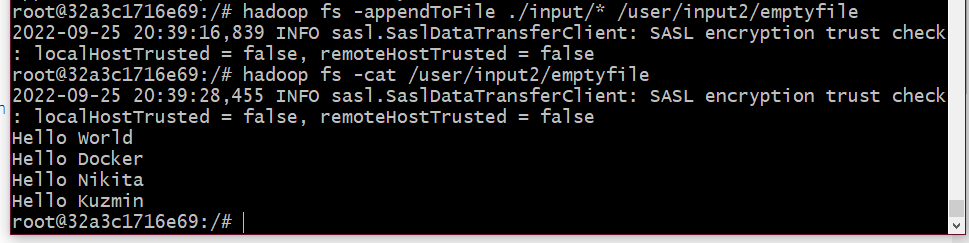
Touchz



Getmerge

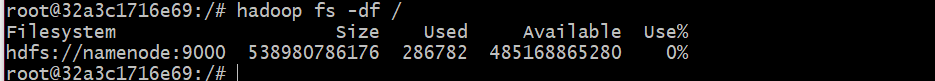


Setrep (replication of single file)

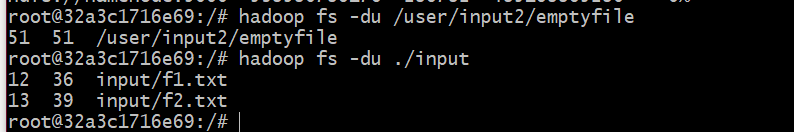


Append to file

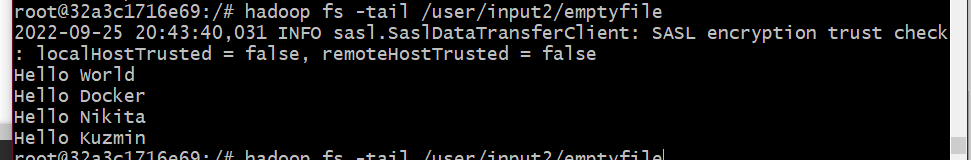
Сложно придумать нормальное изображение на chown chmod особено на chown но типа есть и меняет разрешения и владельца



Df



Du



Tail (последний 1кб файла, я знаю, надо было логи выводить, один фиг не заскриншочу)