Sudha Amarnath

SAN JOSÉ STATE
UNIVERSITY

Graduate and Extended Studies

FA19: CMPE-297 Sec 01 - Special Topics

Prof. Chandrasekar Vuppalapati

Apache Hadoop: Setting up a Single Node Cluster.

1. Installation on Ubuntu

a. Install JAVA open JDK
   - sudha@sudha:~$ java –version
     openjdk version "1.8.0_212"
     OpenJDK Runtime Environment (build 1.8.0_212-8u212-b03-0ubuntu1.18.04.1-b03)
     OpenJDK 64-Bit Server VM (build 25.212-b03, mixed mode)

b. Add a dedicated Hadoop user
   - sudo addgroup hadoop
   - sudo adduser --ingroup hadoop hduser

c. Create and setup SSH certificates for password less login
   - sudha@sudha:~$ which ssh
     /usr/bin/ssh
   - su hduser
   - cat $HOME/.ssh/id_rsa.pub >> $HOME/.ssh/authorized_key

d. Install Hadoop
   - Wget hadoop package from the apache library (latest hadoop-3.2.0.tar.gz)
   - tar xvzf hadoop-3.2.0.tar.gz
   - sudo mkdir -p /usr/local/hadoop
   - sudo adduser hduser sudo
   - sudo chown -R hduser:hadoop /usr/local/hadoop
     hduser@sudha:~/hadoop-3.2.0$ ls
     bin etc include lib libexec LICENSE.txt NOTICE.txt README.txt sbin share
   - sudo mv * /usr/local/hadoop
   - Set path for environment variables for JAVA and Hadoop in .bashrc

```
#HADOOP VARIABLES START
export JAVA_HOME=/usr/lib/jvm/java-8-openjdk-amd64
export HADOOP_INSTALL=/usr/local/hadoop
export PATH=$PATH:$HADOOP_INSTALL/bin
export PATH=$PATH:$HADOOP_INSTALL/sbin
export HADOOP_MAPRED_HOME=$HADOOP_INSTALL
```

```
export HADOOP_COMMON_HOME=$HADOOP_INSTALL
export HADOOP_HDFS_HOME=$HADOOP_INSTALL
export YARN_HOME=$HADOOP_INSTALL
export
HADOOP_COMMON_LIB_NATIVE_DIR=$HADOOP_INSTALL/lib/native
export HADOOP_OPTS="-Djava.library.path=$HADOOP_INSTALL/lib"
export PDSH_RCMD_TYPE=ssh
#HADOOP VARIABLES END
```

- Update core-site.xml file with following contents

```
<configuration>
 <property>
  <name>hadoop.tmp.dir</name>
  <value>/app/hadoop/tmp</value>
  <description>A base for other temporary directories.</description>
 </property>
 <property>
  <name>fs.default.name</name>
  <value>hdfs://localhost:54310</value>
  <description>The name of the default file system.  A URI whose
  scheme and authority determine the FileSystem implementation.  The
  uri's scheme determines the config property (fs.SCHEME.impl) naming
  the FileSystem implementation class.  The uri's authority is used to
  determine the host, port, etc. for a filesystem.</description>
 </property>
</configuration>
```

- Update mapred-site.xml file with following contents

```
<configuration>
 <property>
  <name>mapred.job.tracker</name>
  <value>localhost:54311</value>
  <description>The host and port that the MapReduce job tracker runs
  at.  If "local", then jobs are run in-process as a single map
  and reduce task.
  </description>
 </property>
</configuration>
```

- Create namenode and datanode specific directories

```
hduser@sudha:~$ sudo mkdir -p
/usr/local/hadoop_store/hdfs/namenode
hduser@sudha:~$ sudo mkdir -p /usr/local/hadoop_store/hdfs/datanode
hduser@sudha:~$ sudo chown -R hduser:hadoop
/usr/local/hadoop_store
```

- Update hdfs-site.xml file with following contents

```
<configuration>
 <property>
  <name>dfs.replication</name>
  <value>1</value>
  <description>Default block replication.
  The actual number of replications can be specified when the file is
created.
  The default is used if replication is not specified in create time.
  </description>
 </property>
 <property>
   <name>dfs.namenode.name.dir</name>
   <value>file:/usr/local/hadoop_store/hdfs/namenode</value>
 </property>
 <property>
   <name>dfs.datanode.data.dir</name>
   <value>file:/usr/local/hadoop_store/hdfs/datanode</value>
 </property>
</configuration>
```

- Format hadoop namenode for initial usage

```
 hduser@sudha:~$ hadoop namenode -format
```

- Start Hadoop daemons – dfs and yarn

```
hduser@sudha:~$ /usr/local/hadoop/sbin/start-dfs.sh
hduser@sudha:~$ /usr/local/hadoop/sbin/start-yarn.sh
```

- Verify the started processes

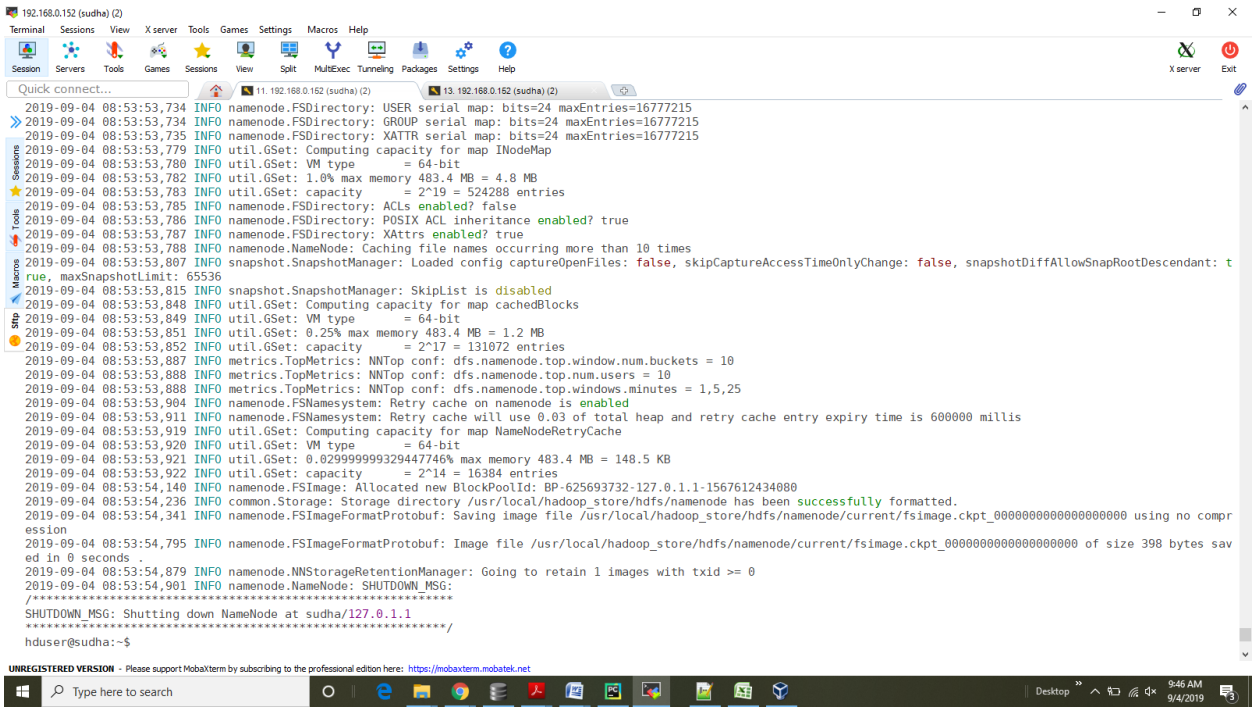
```
hduser@sudha:/usr/local/hadoop/sbin$ jps
5680 NodeManager
```

4688 DataNode
5525 ResourceManager
4902 SecondaryNameNode
6039 Jps
4536 NameNode

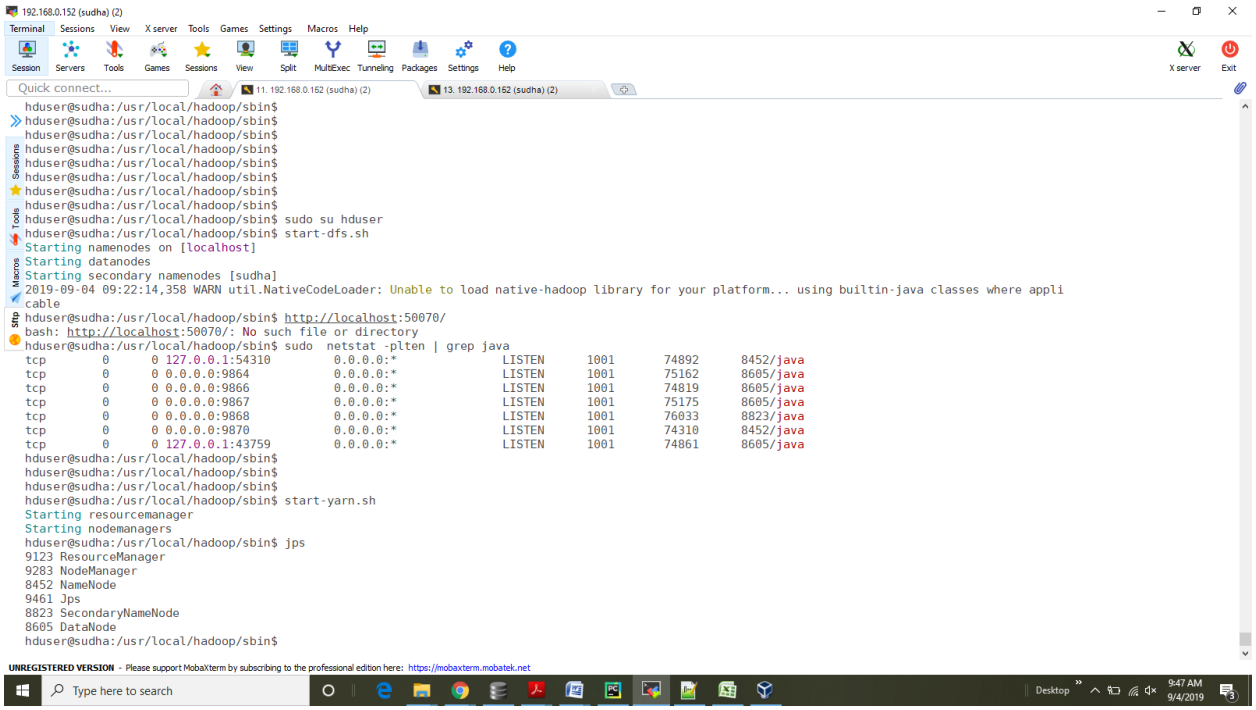
- Check URL - http://localhost:9870/dfshealth.html#tab-overview

2. Screenshots

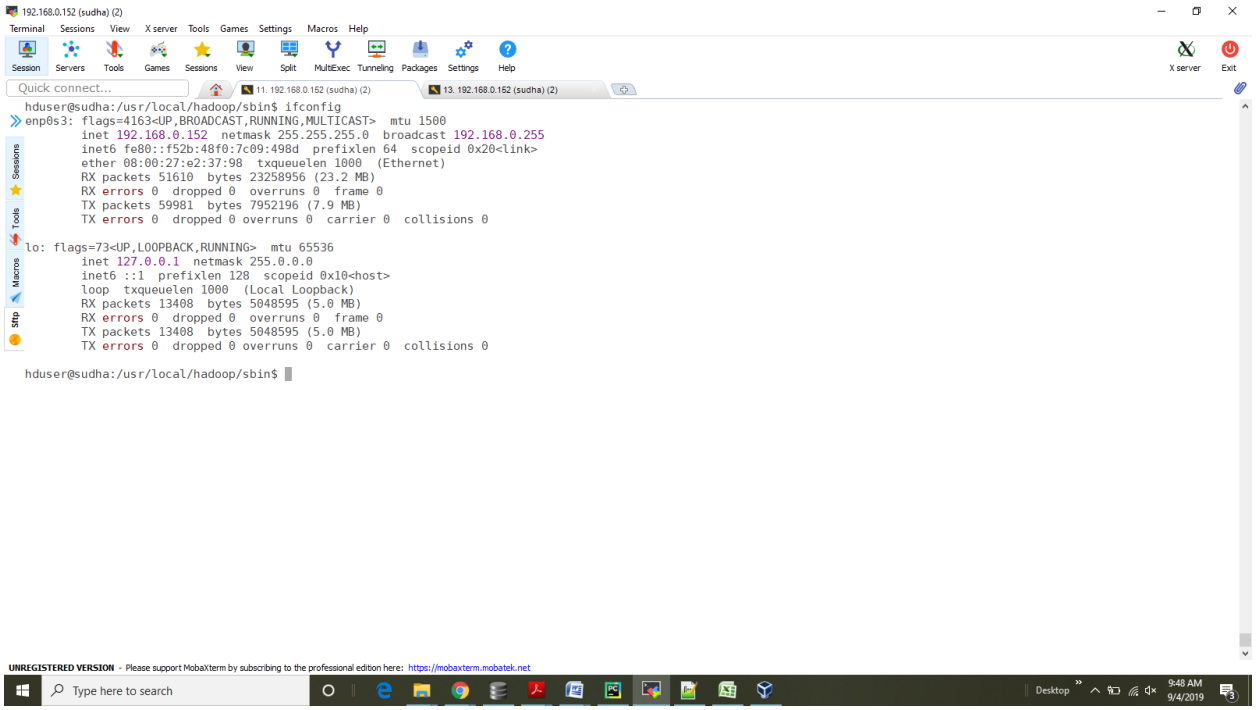
a. Format Hadoop namenode for initial usage



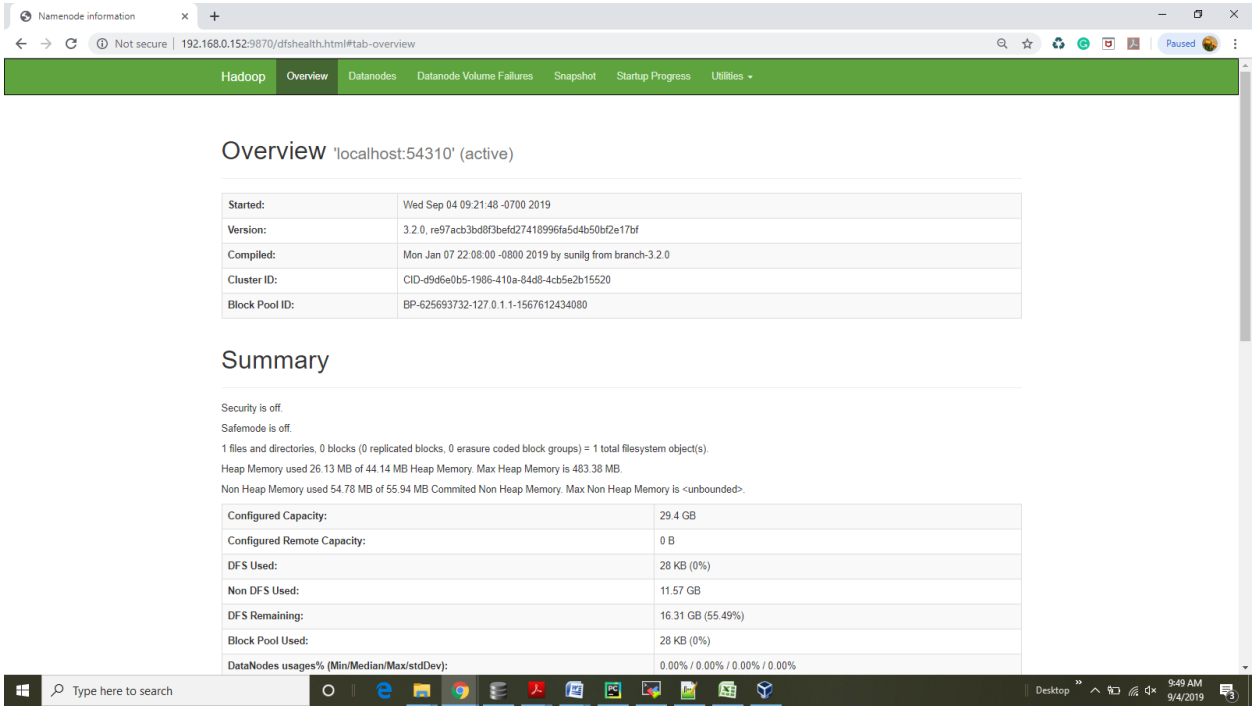
b. Starting Hadoop using dfs and yarn



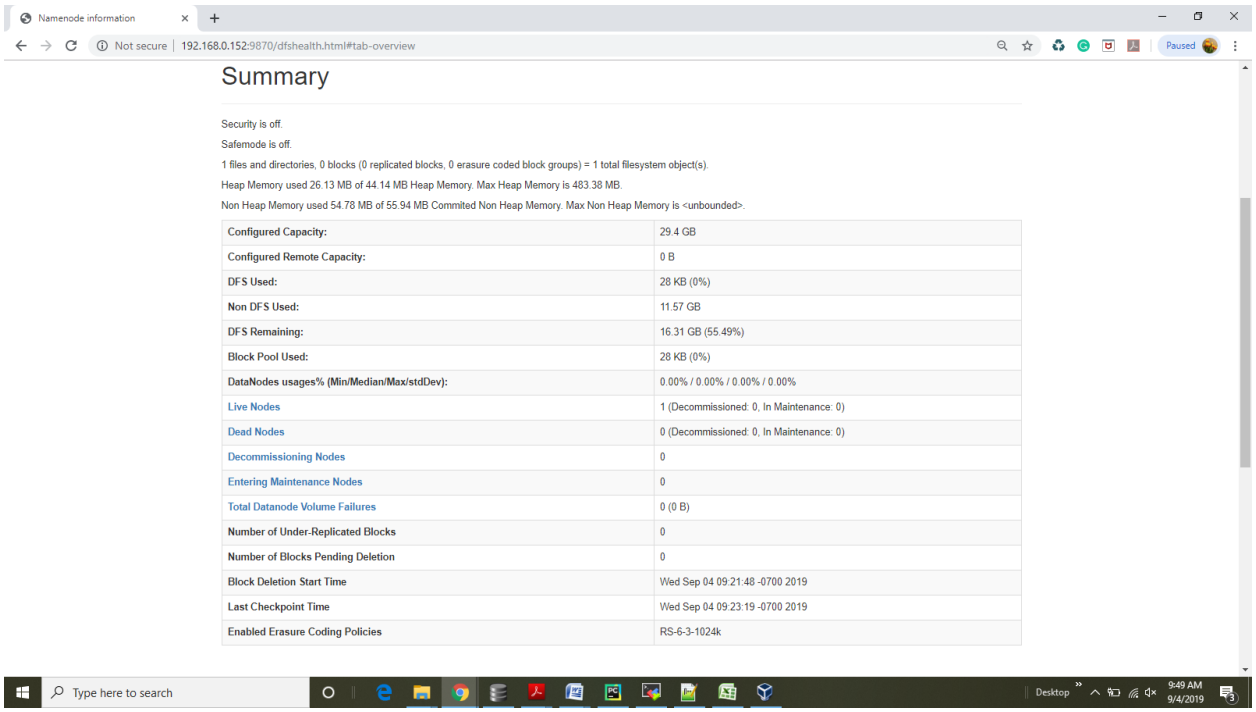
c. Ubuntu local machine IP



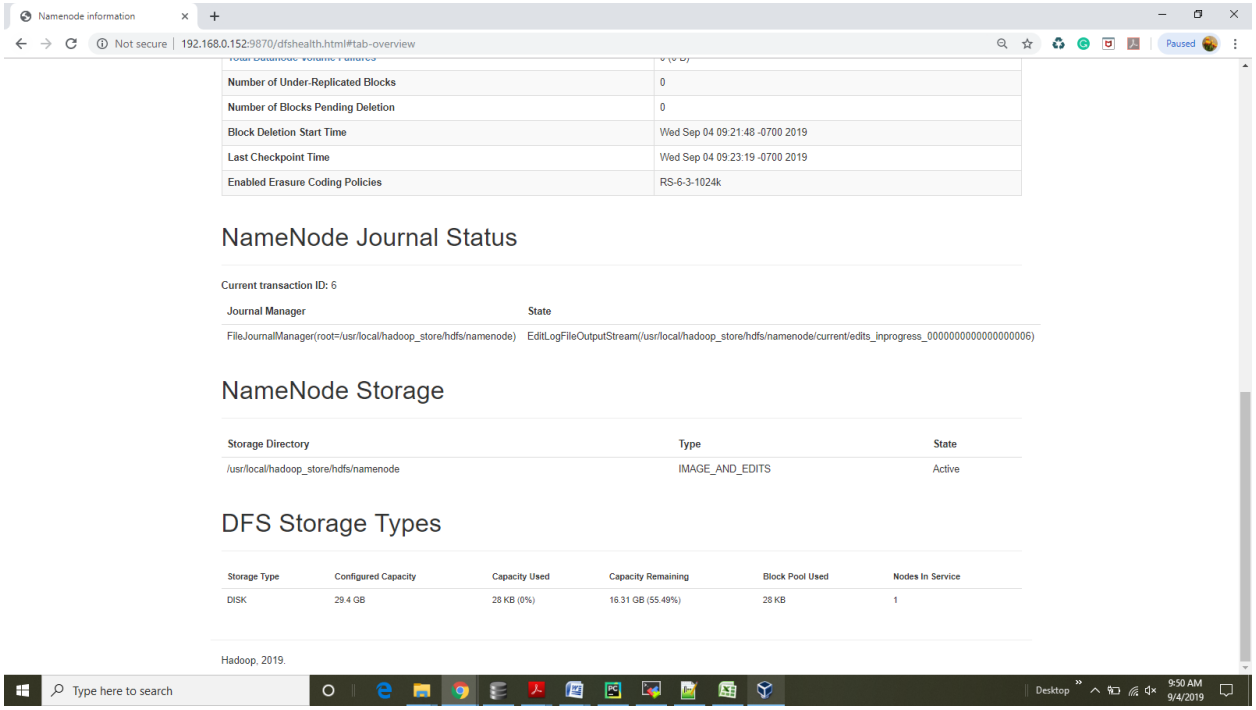
d. Hadoop cluster overview



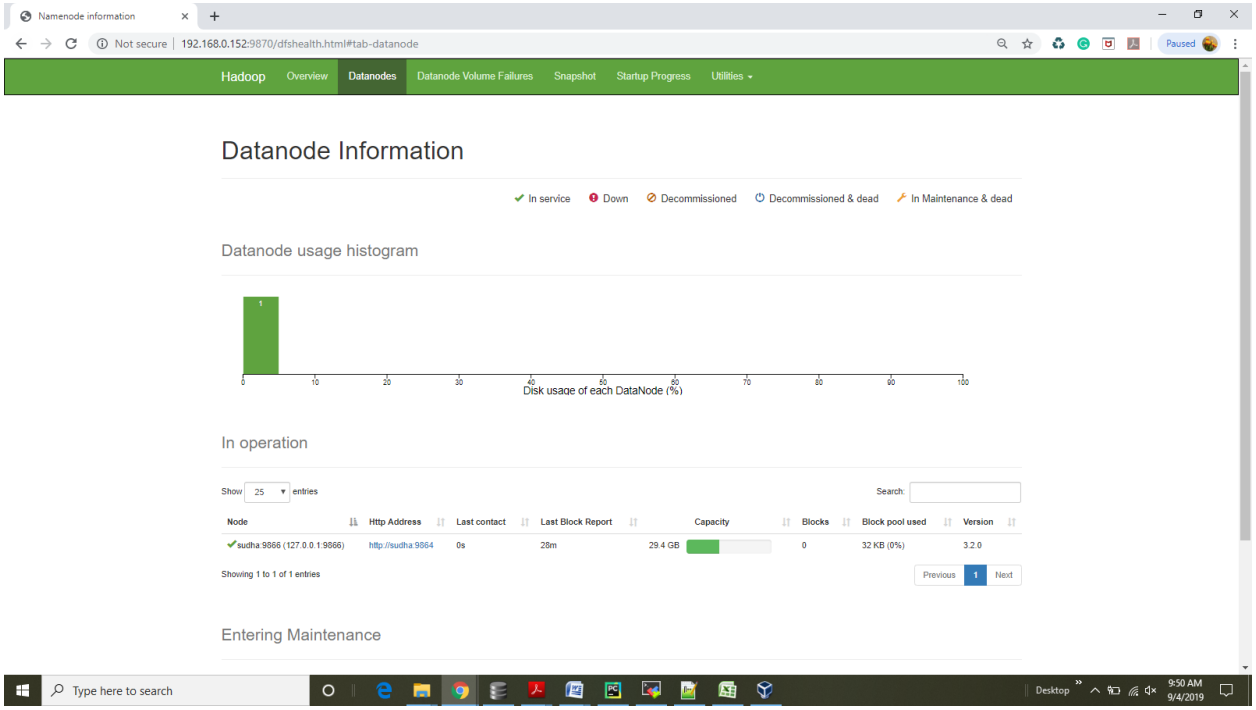
e. Hadoop cluster summary showing a live node

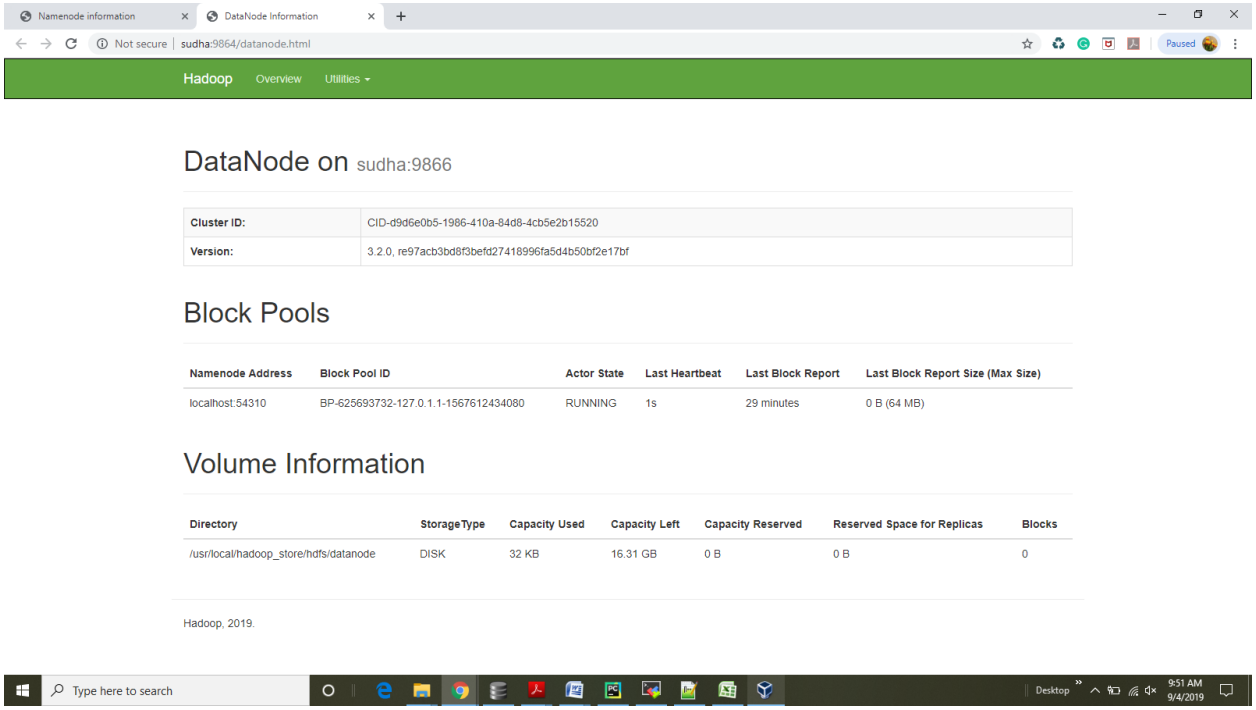


f. NameNode staus

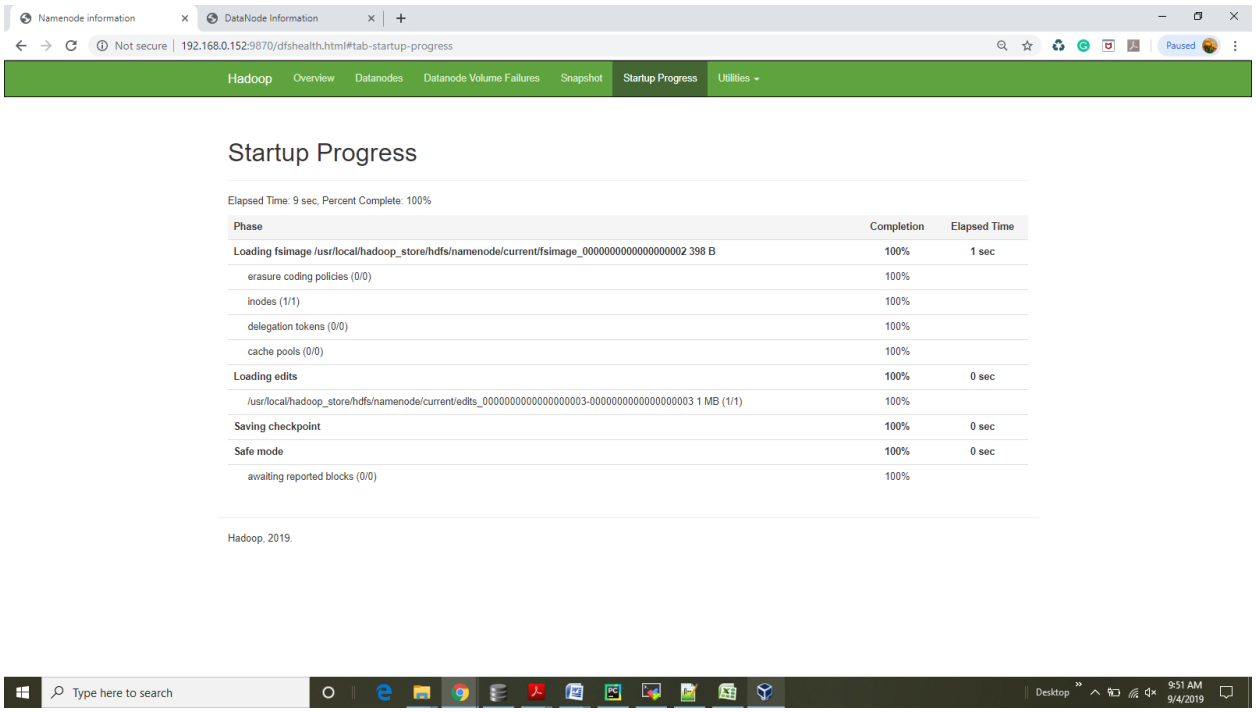


g. Information of the Datanode

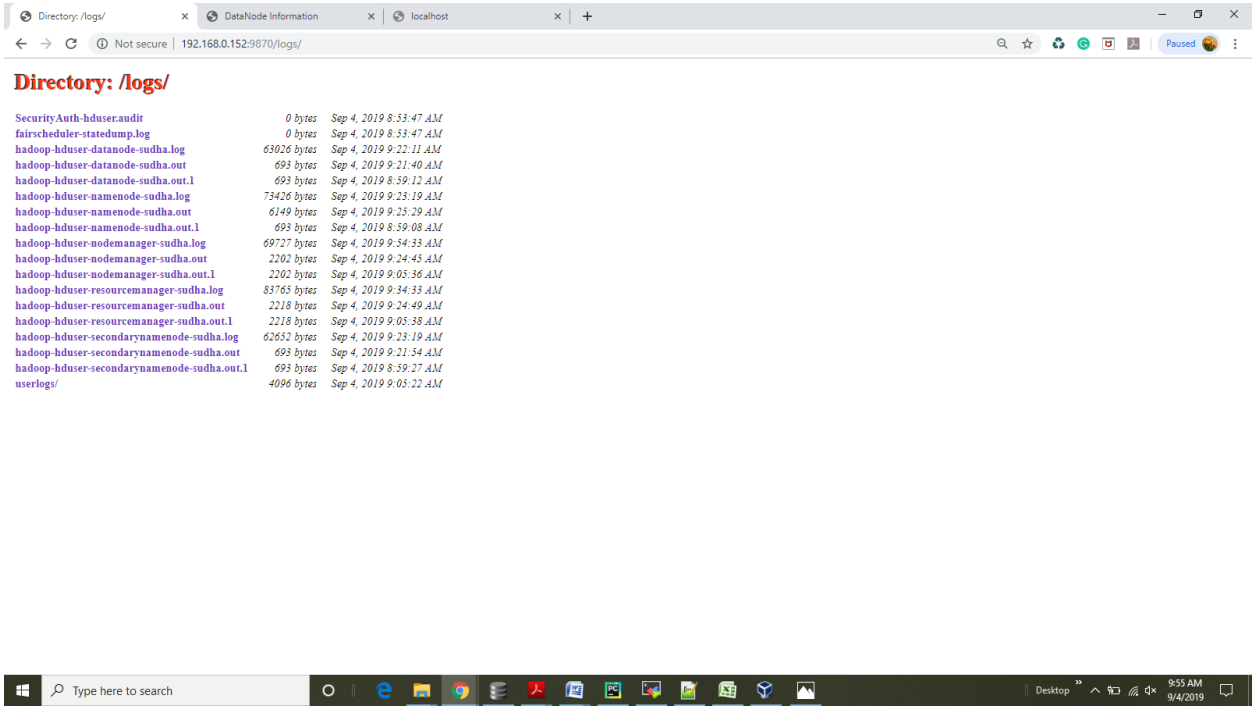




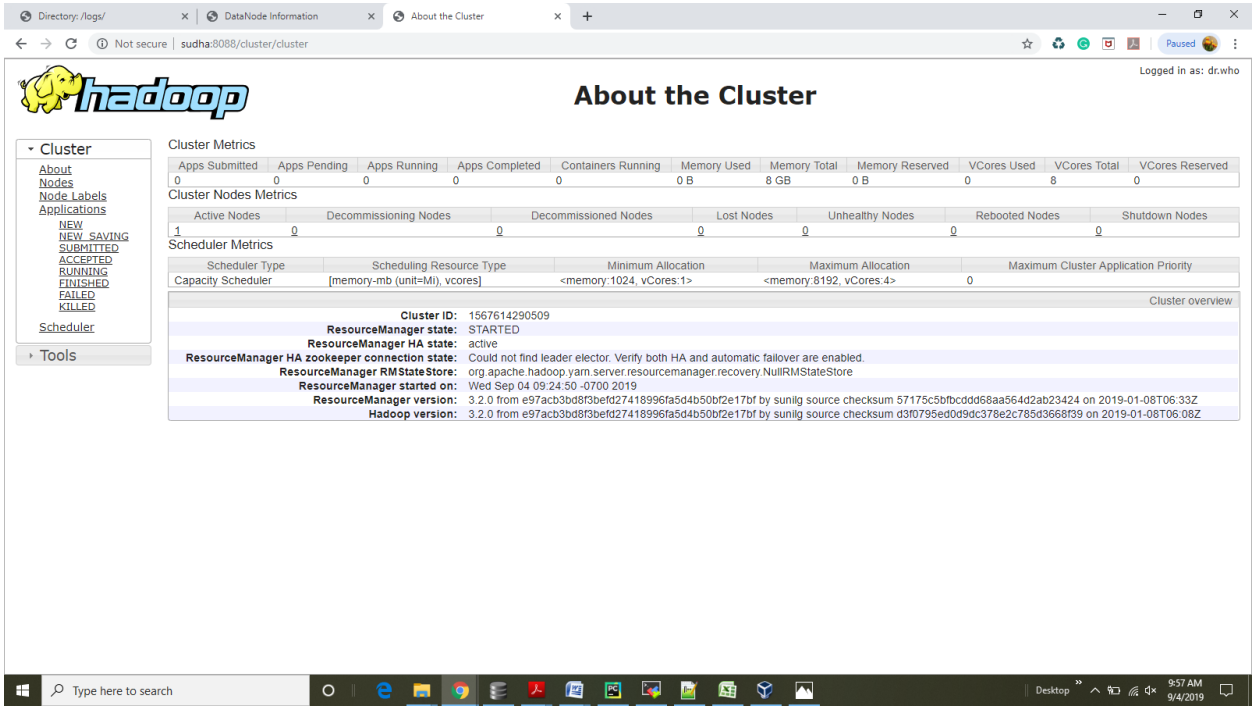
h. Hadoop cluster startup progress



i. Logs directory



j. Details of the Hadoop cluster



k. Information of the nodes of the cluster