**Sudha Amarnath**



**Graduate and Extended Studies**

**FA19: CMPE-297 Sec 01 - Special Topics**

**Prof. Chandrasekar Vuppalapati**

**Apache Hadoop: Setting up a Single Node Cluster.**

## Installation on Ubuntu

1. 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)

1. Add a dedicated Hadoop user
   * sudo addgroup hadoop
   * sudo adduser --ingroup hadoop hduser
2. 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

1. 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

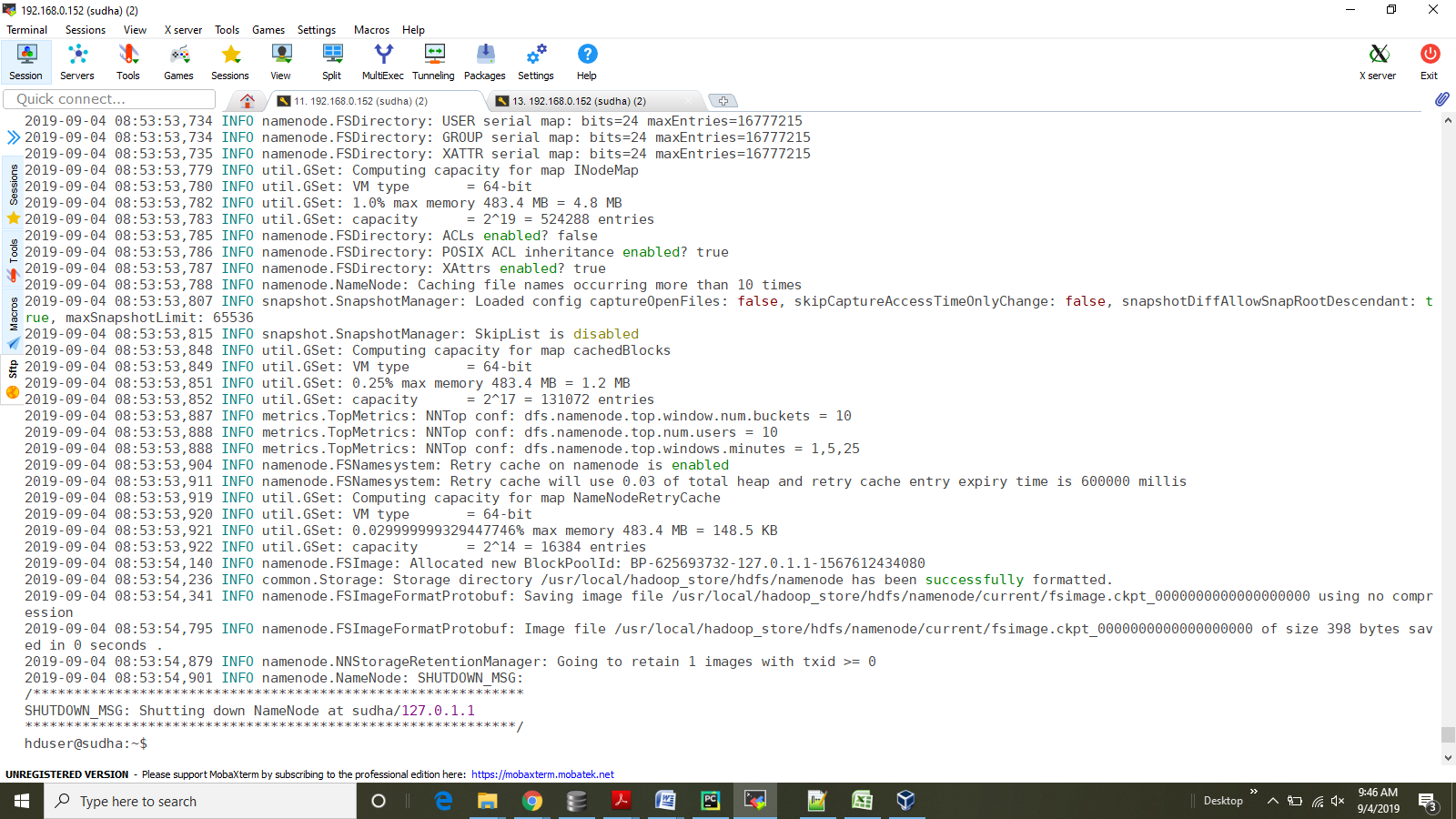
6039 Jps

4536 NameNode

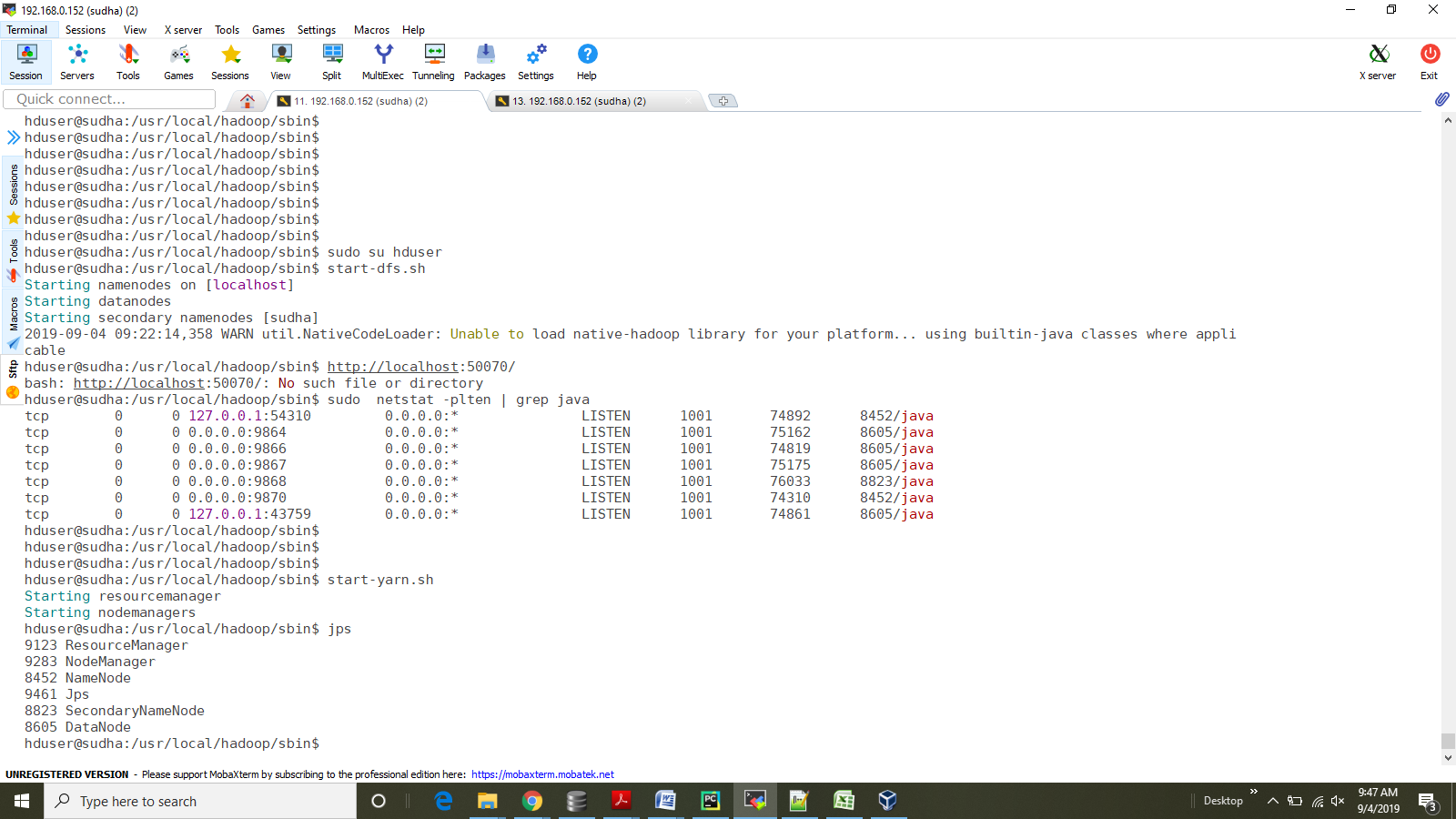
* Check URL - <http://localhost:9870/dfshealth.html#tab-overview>

1. **Screenshots**

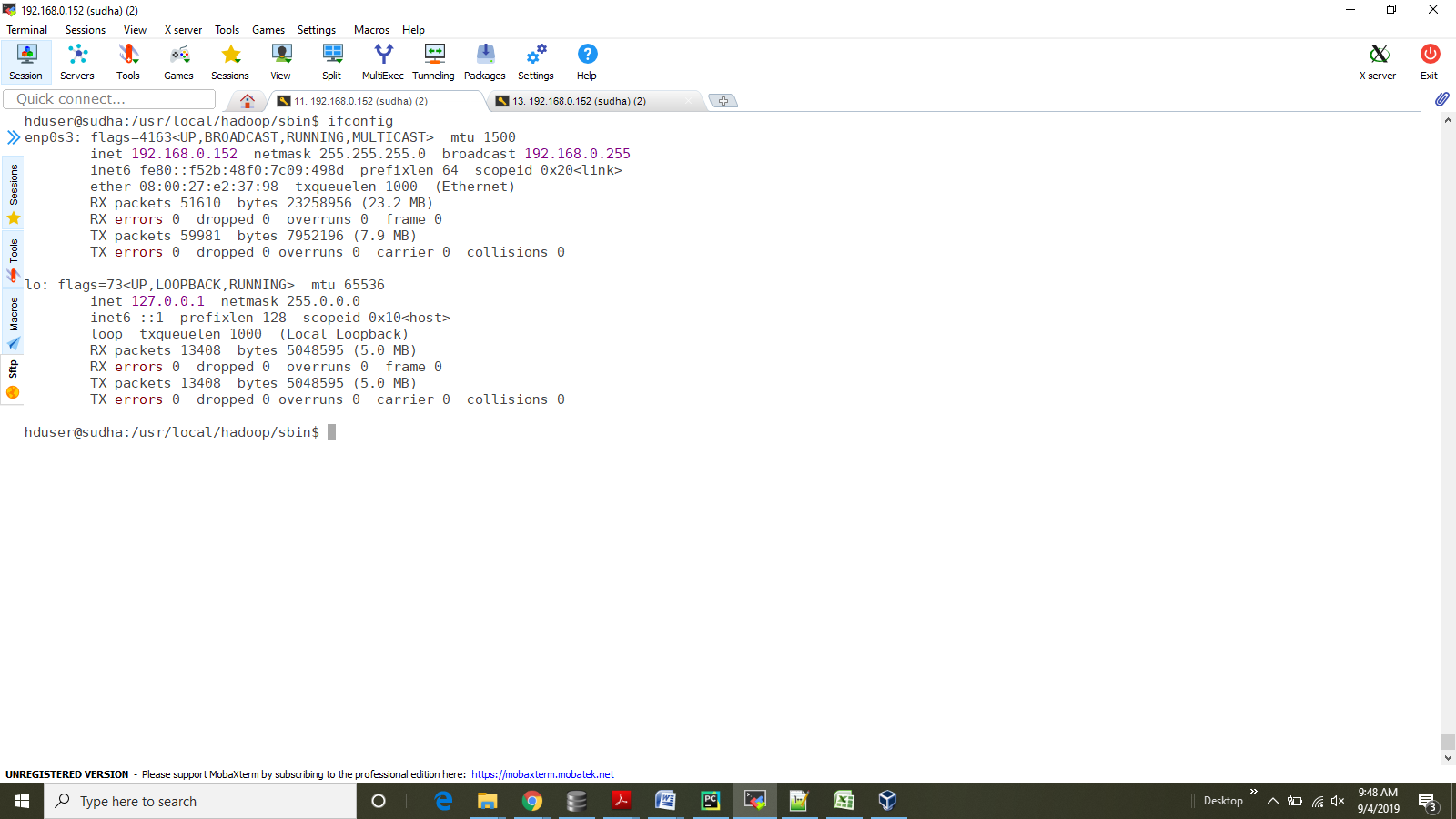
**a. Format Hadoop namenode for initial usage**

****

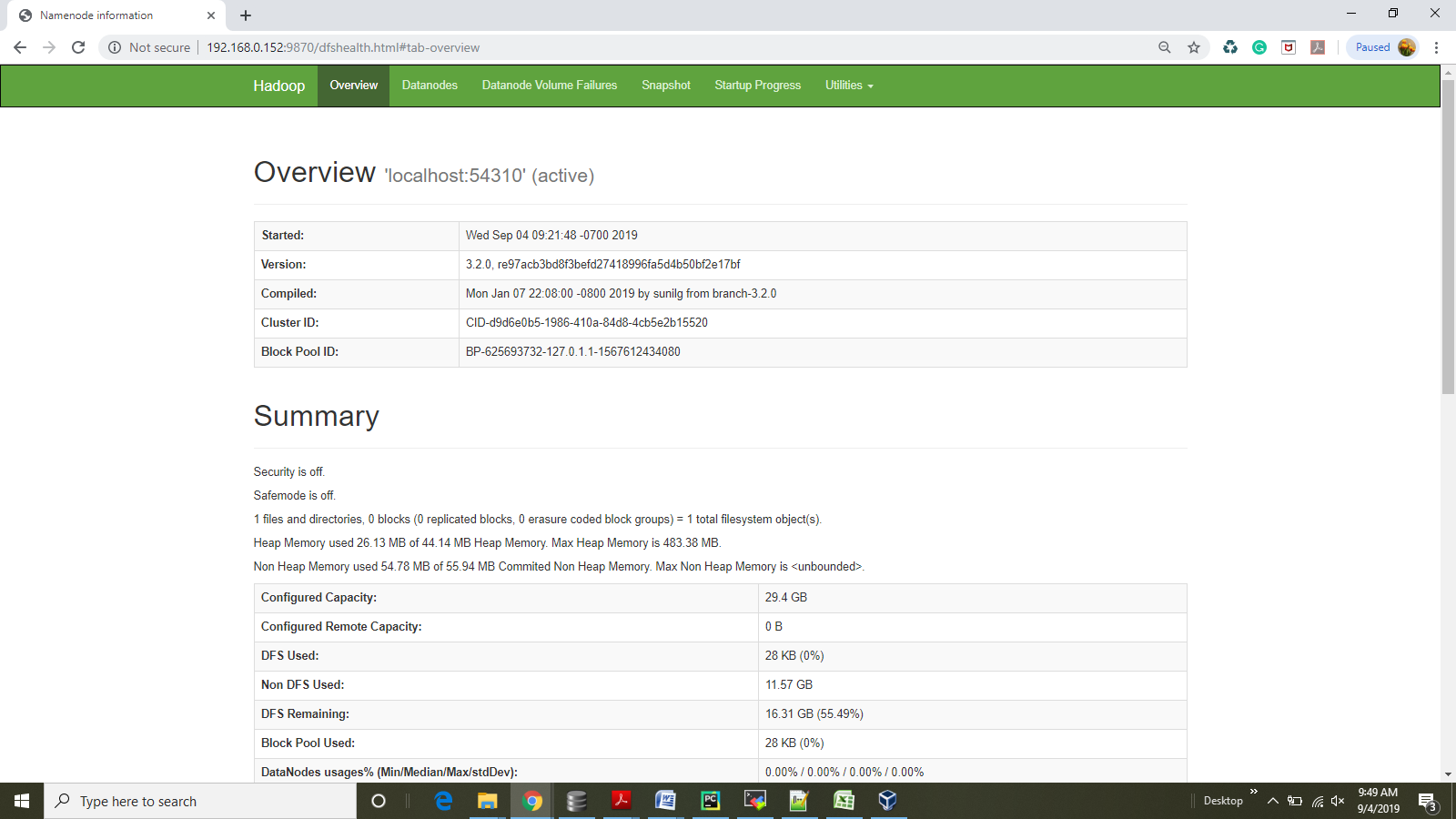
1. **Starting Hadoop using dfs and yarn**

****

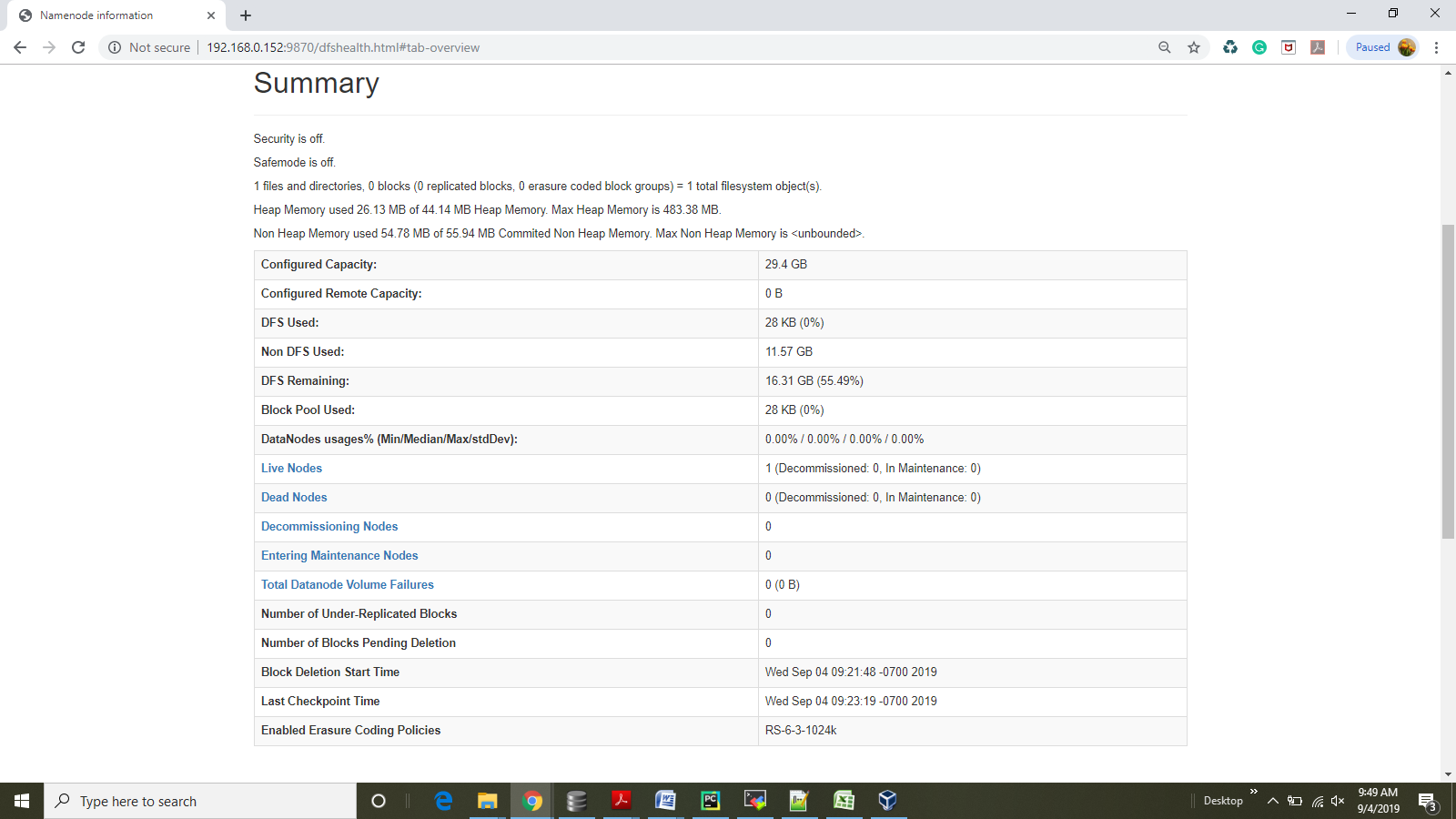
1. **Ubuntu local machine IP**

****

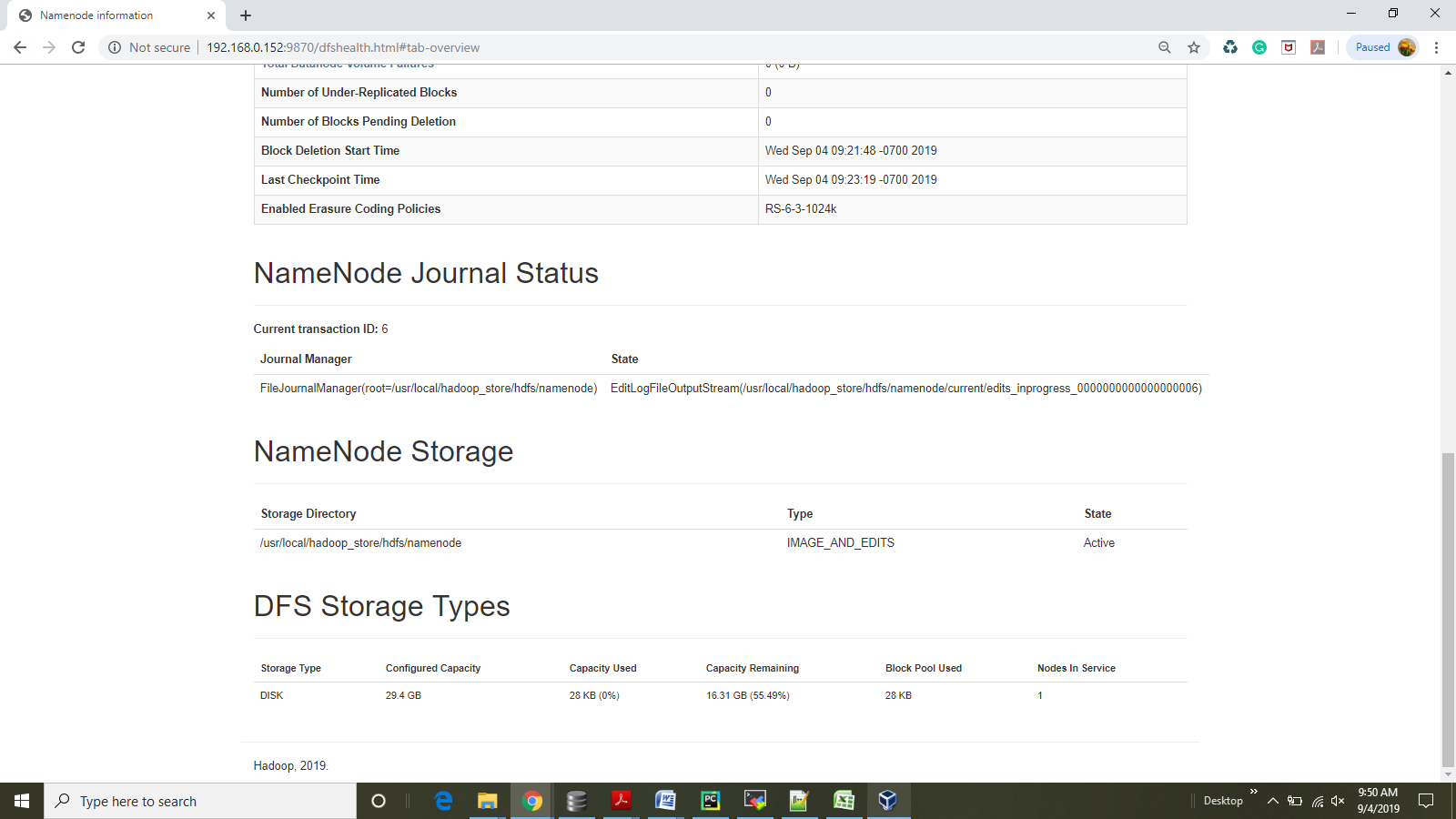
1. **Hadoop cluster overview**

****

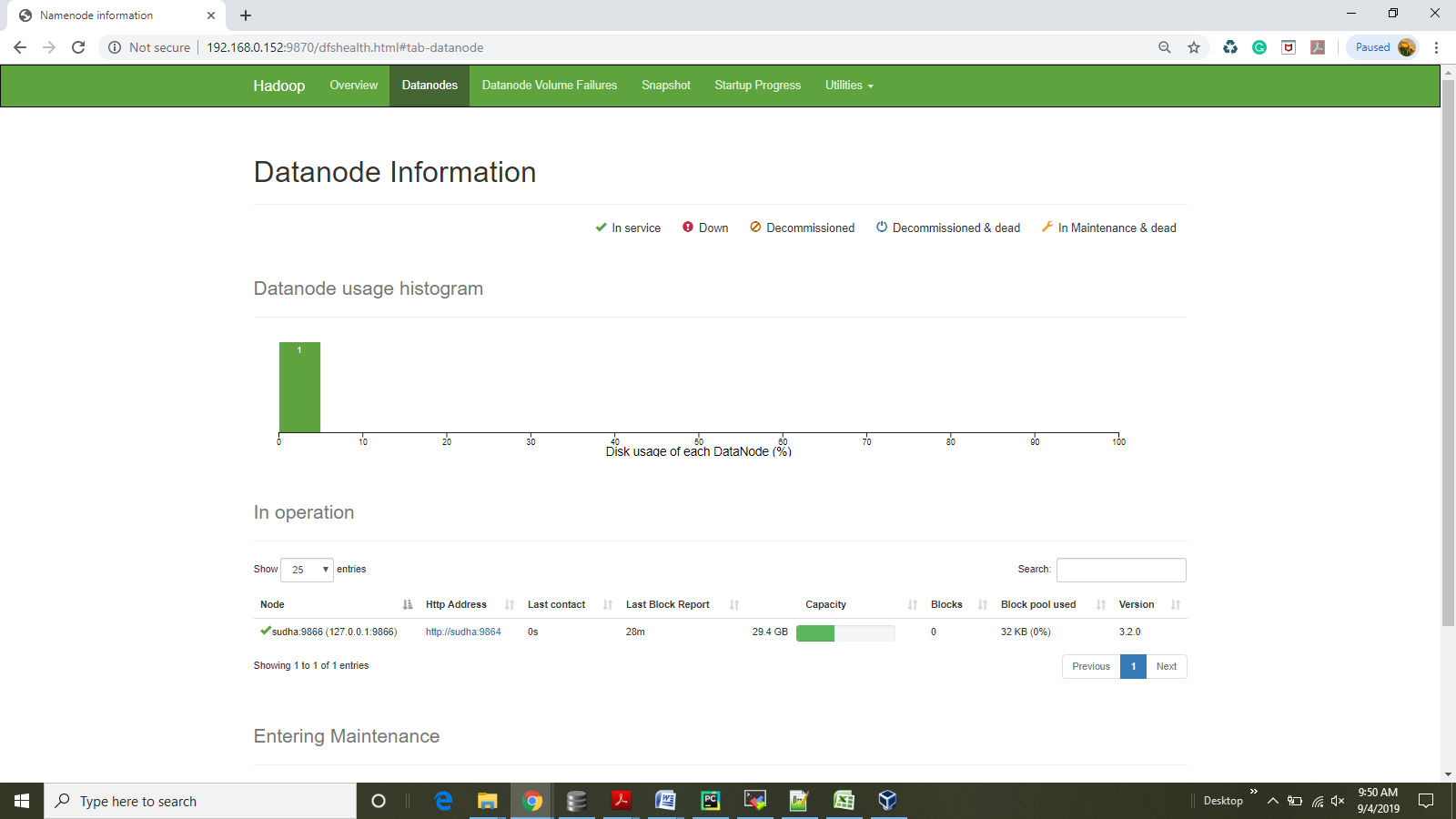
1. **Hadoop cluster summary showing a live node**

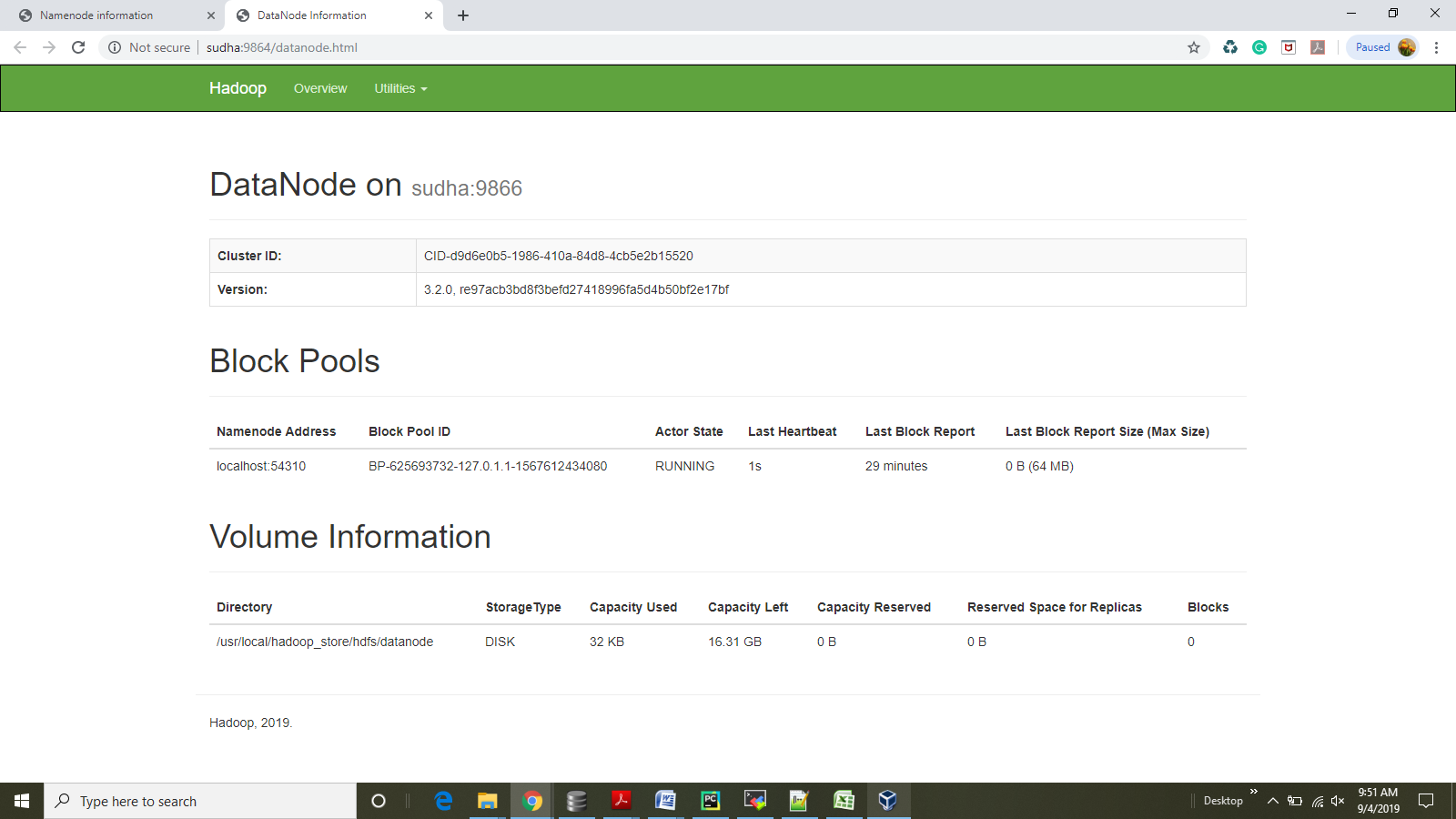
****

1. **NameNode staus**

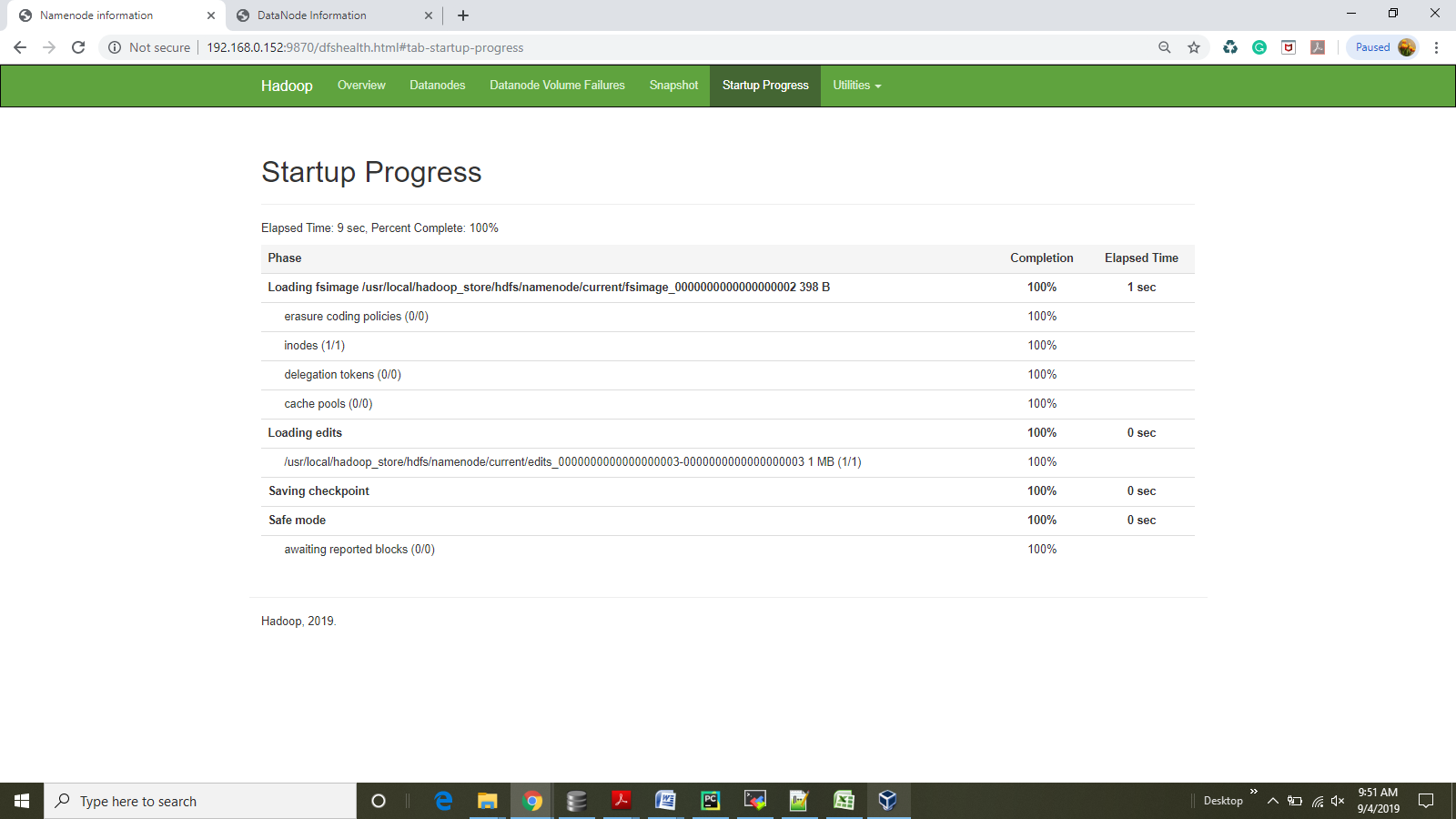
****

1. **Information of the Datanode**

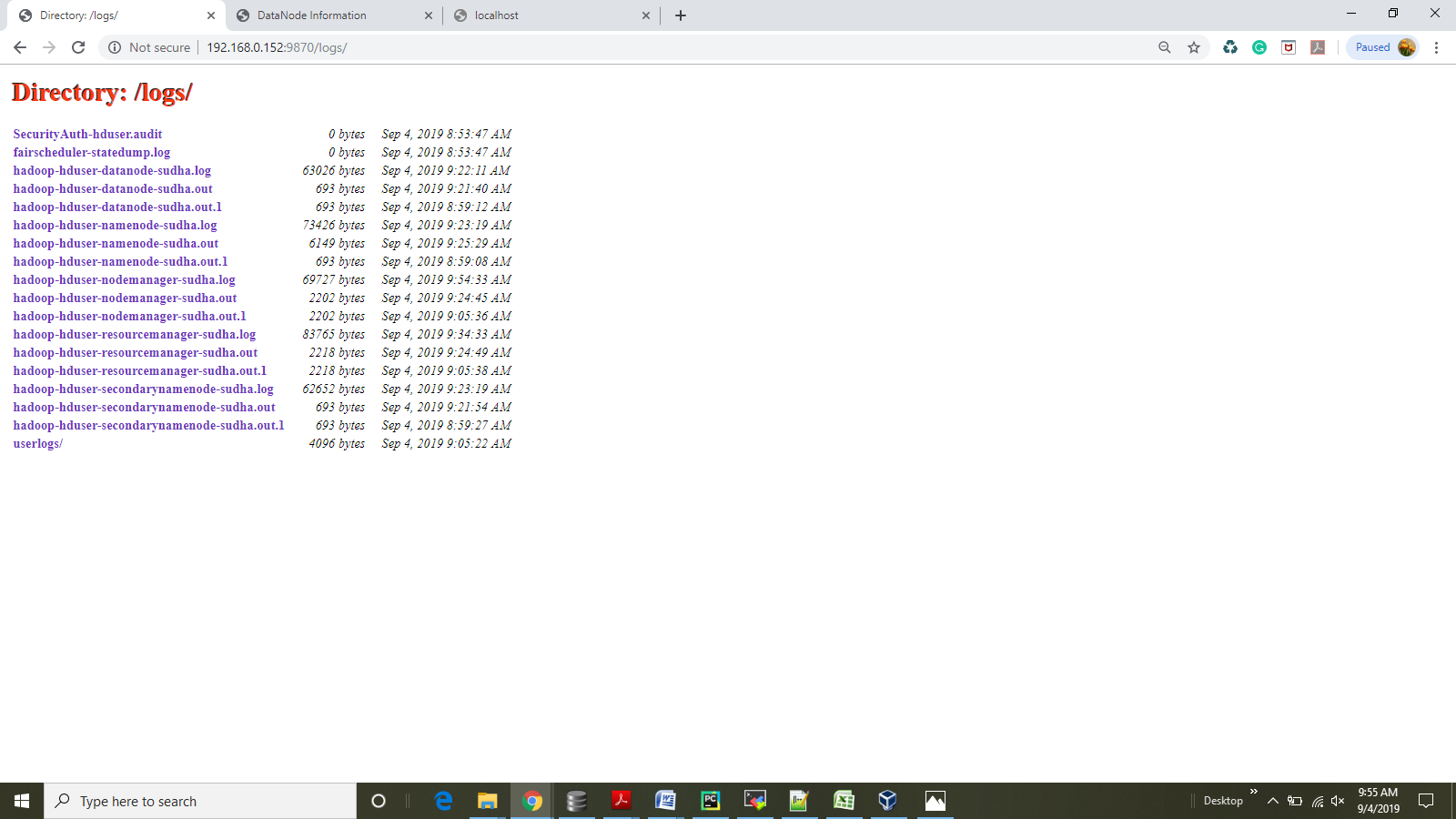
****

****

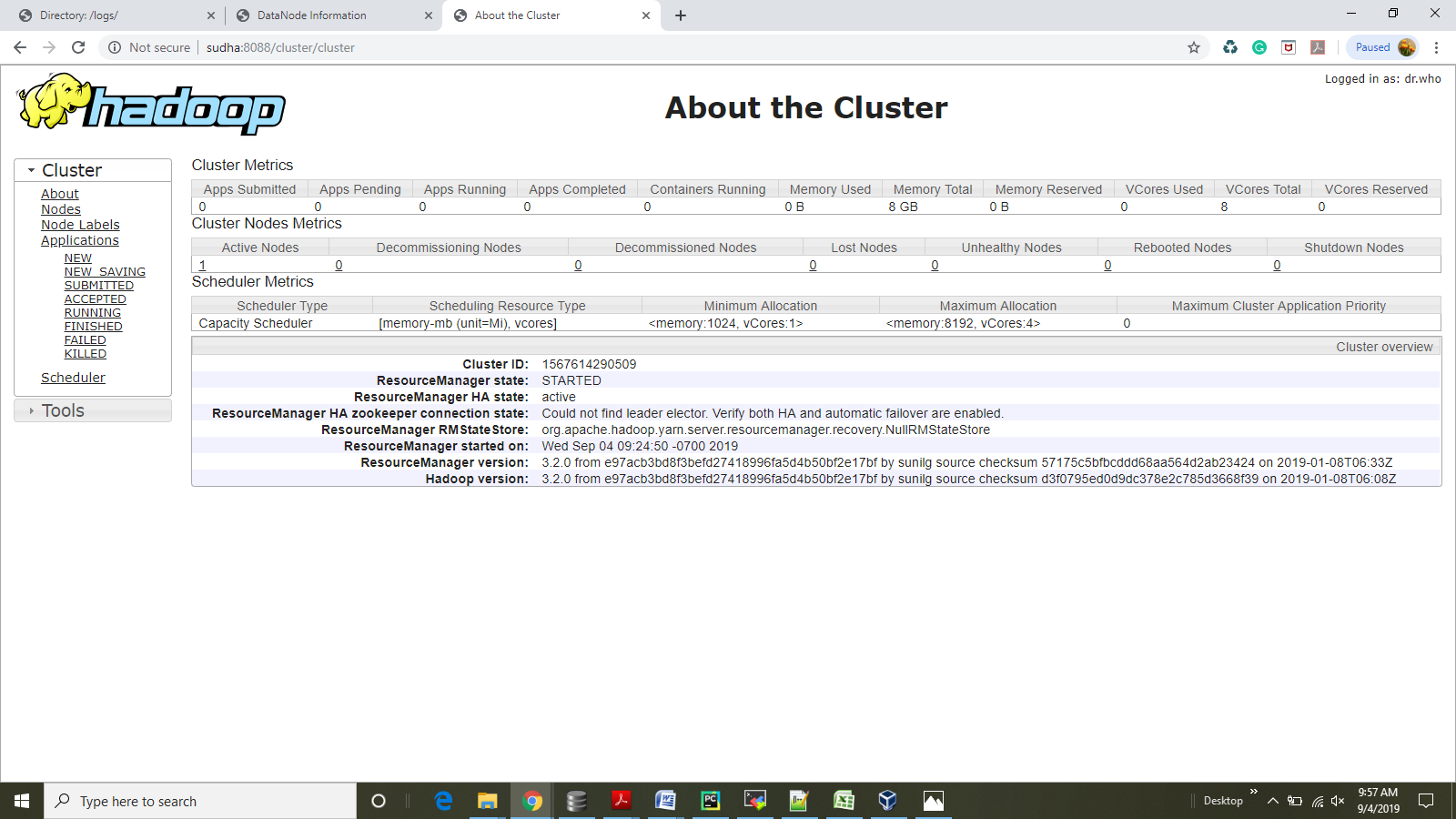
1. **Hadoop cluster startup progress**

****

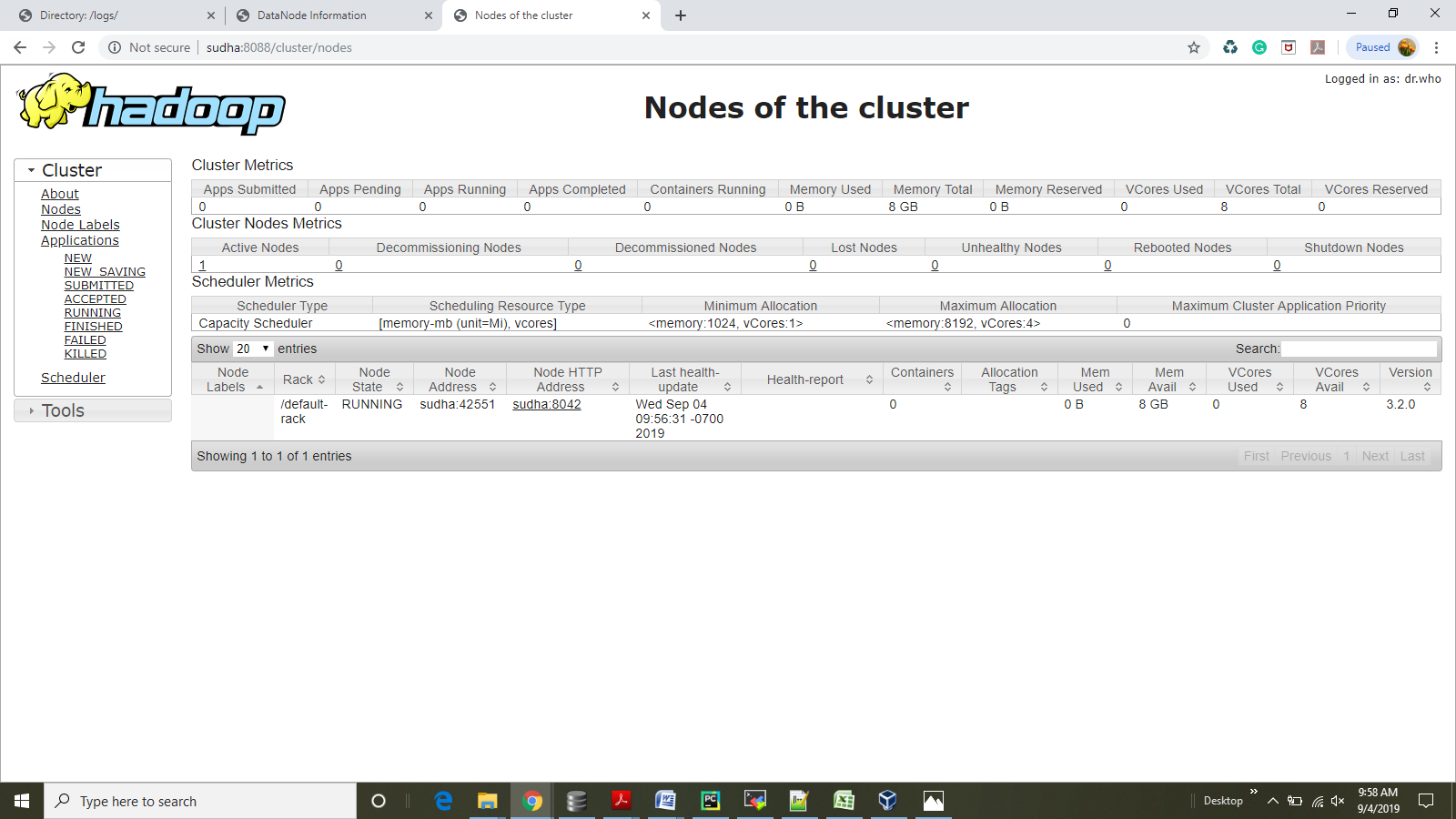
1. **Logs directory**

****

1. **Details of the Hadoop cluster**

****

1. **Information of the nodes of the cluster**

****