Laboratory Assignment File

for

Big Data Analytics (CS0552)

Master of Technology

in

Computer Science & Engineering

Submitted by

Yugal

(Roll no. 25903053)

Submitted to

Dr. Shveta Mahajan

(Assistant Professor, NIT Jalandhar)

Department of Computer Science & Engineering

Dr. B R Ambedkar National Institute of Technology Jalandhar

Punjab, India-144008

May, 2026

Contents

List of Figures

Assignment 1: Install Apache Hadoop and Setup Single Node Cluster

Objective: To install and configure Apache Hadoop in pseudo-distributed mode and perform basic HDFS operations such as upload, delete, replication check, and permission handling.

1.1 Step 1: Installing Java

Updating System

```
sudo apt update
```

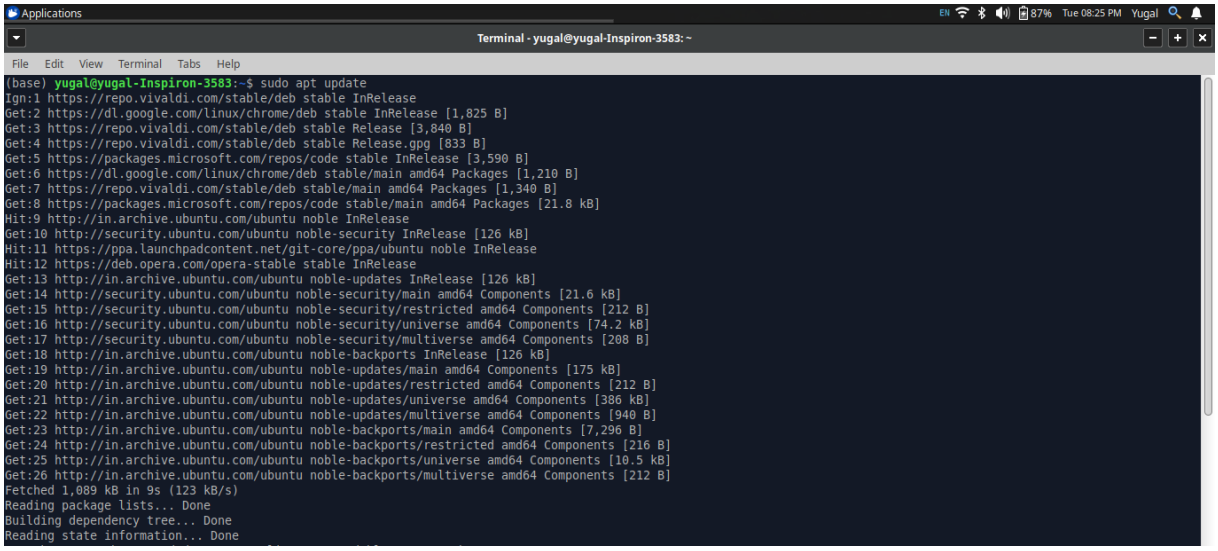

Figure 1: System Update Command

Installing Java

```
sudo apt install openjdk-11-jdk -y
```

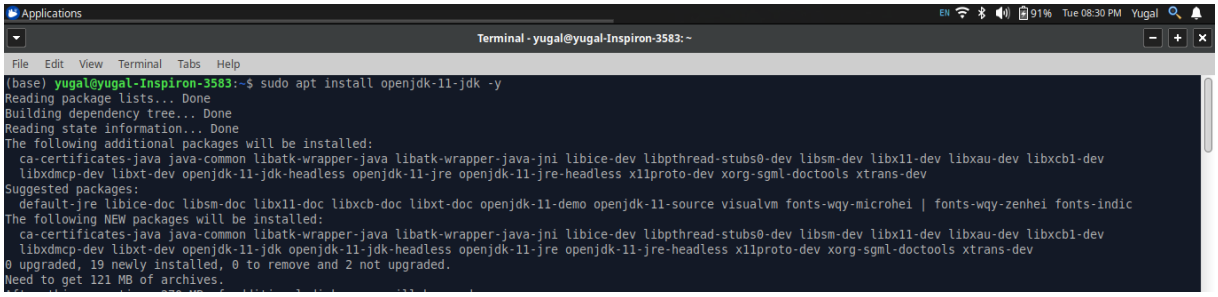

Figure 2: Java Installation

Java Version and Path Verification

```
java -version
readlink -f $(which java)
```

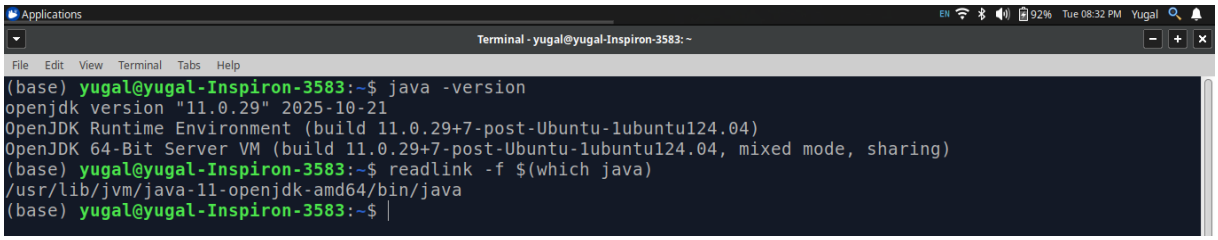

Figure 3: Java Version and Path Verification

1.2 Step 2: Setup SSH

Install SSH Server

```
sudo apt install openssh-server -y
ssh localhost
```

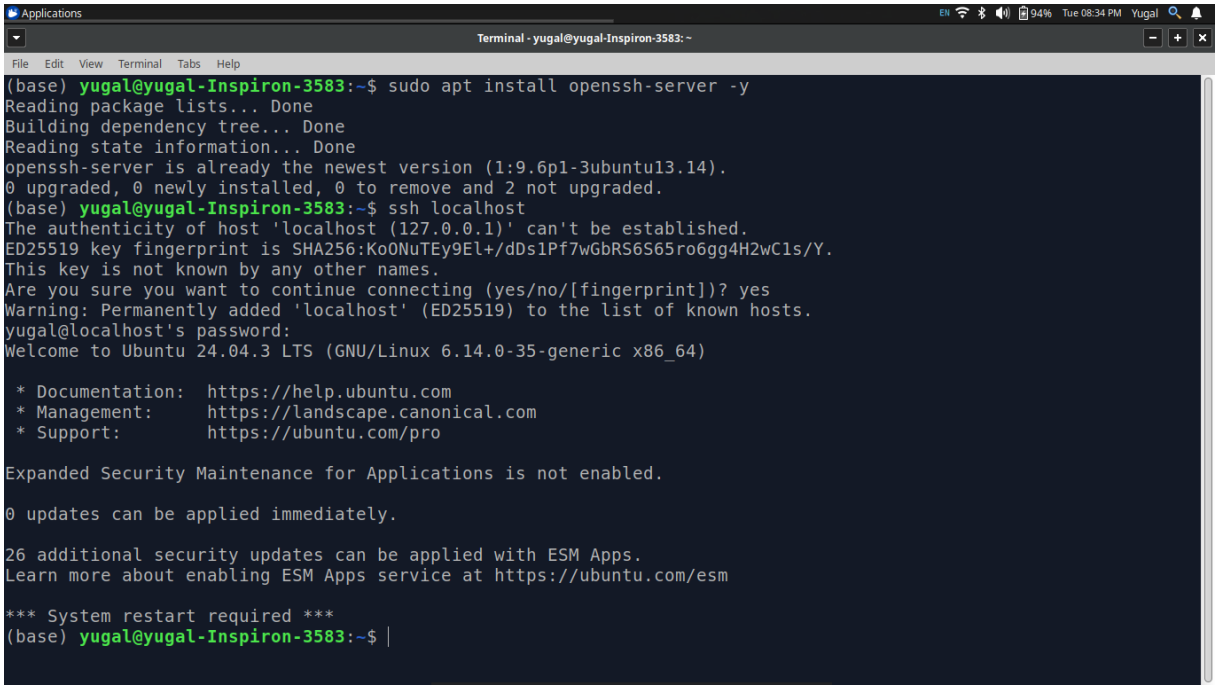

Figure 4: Install SSH Server

Generate SSH Key Pair

```
ssh-keygen -t rsa -P ""
cat ~/.ssh/id_rsa.pub >> ~/.ssh/authorized_keys
```

```
ssh localhost
exit
```

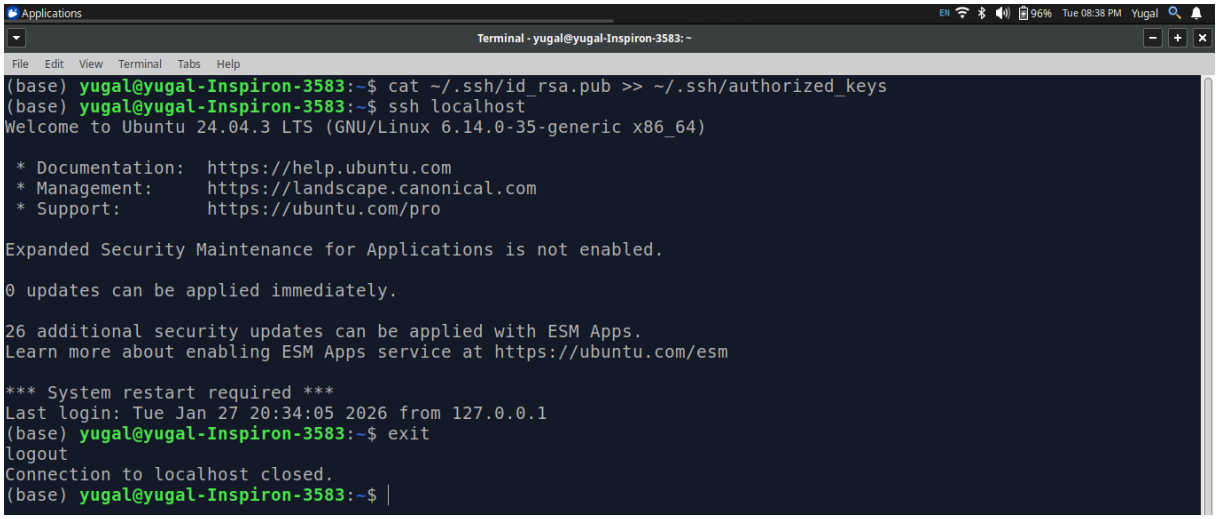

Figure 5: Generate SSH Key Pair

1.3 Step 3: Download and Configure Hadoop

Create Hadoop User

```
sudo addgroup hadoop
sudo adduser --ingroup hadoop hadoop
sudo adduser hadoop sudo
su - hadoop
```

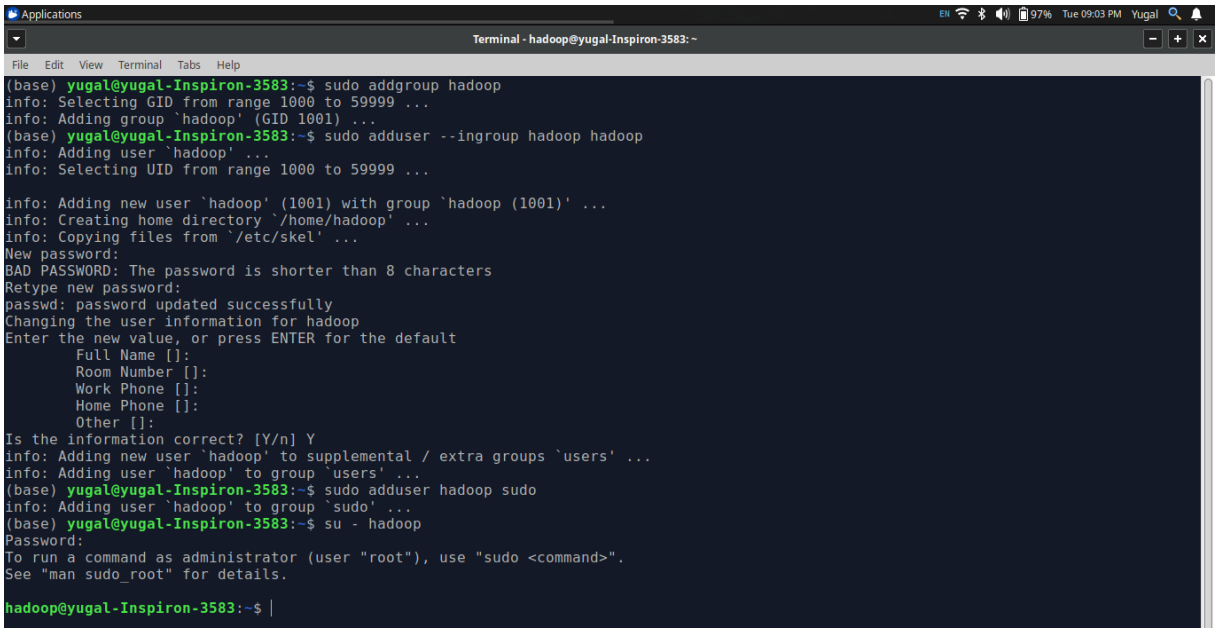

Figure 6: Create Hadoop User

Download and Extract Hadoop

```
exit
sudo cp /home/yugal/Downloads/hadoop-3.3.0.tar.gz /home/hadoop/
sudo chown hadoop:hadoop /home/hadoop/hadoop-3.3.0.tar.gz
su - hadoop
ls
tar -xvzf hadoop-3.3.0.tar.gz
```

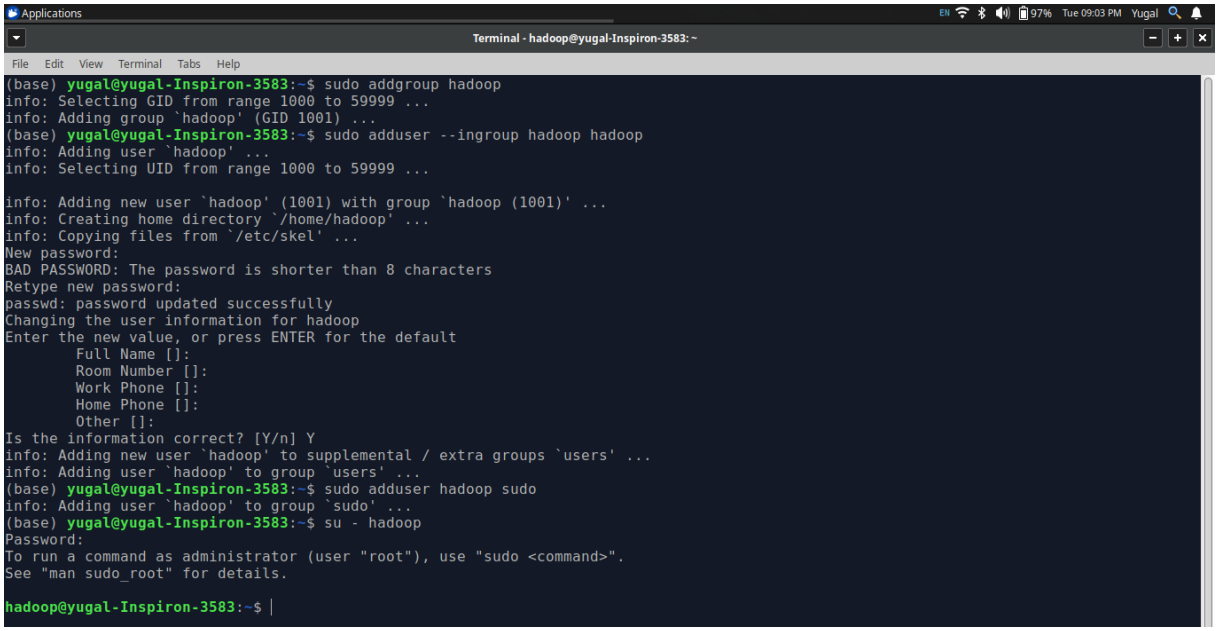

Figure 7: Download and Extract Hadoop

Moving Hadoop

```
mv hadoop-3.3.0 hadoop
ls
```

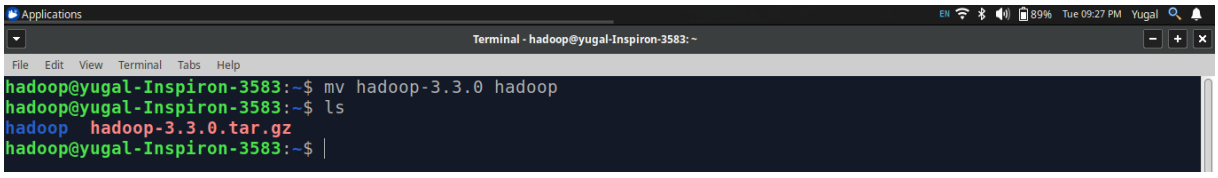

Figure 8: Moving Hadoop

1.4 Step 4: Environment Variable Configuration

.bashrc Modification

```
nano ~/.bashrc
```

```
% Add the following lines at the end of the file
export JAVA_HOME=/usr/lib/jvm/java-11-openjdk-amd64
export HADOOP_HOME=/home/hadoop/hadoop
export PATH=$PATH:$HADOOP_HOME/bin:$HADOOP_HOME/sbin
```

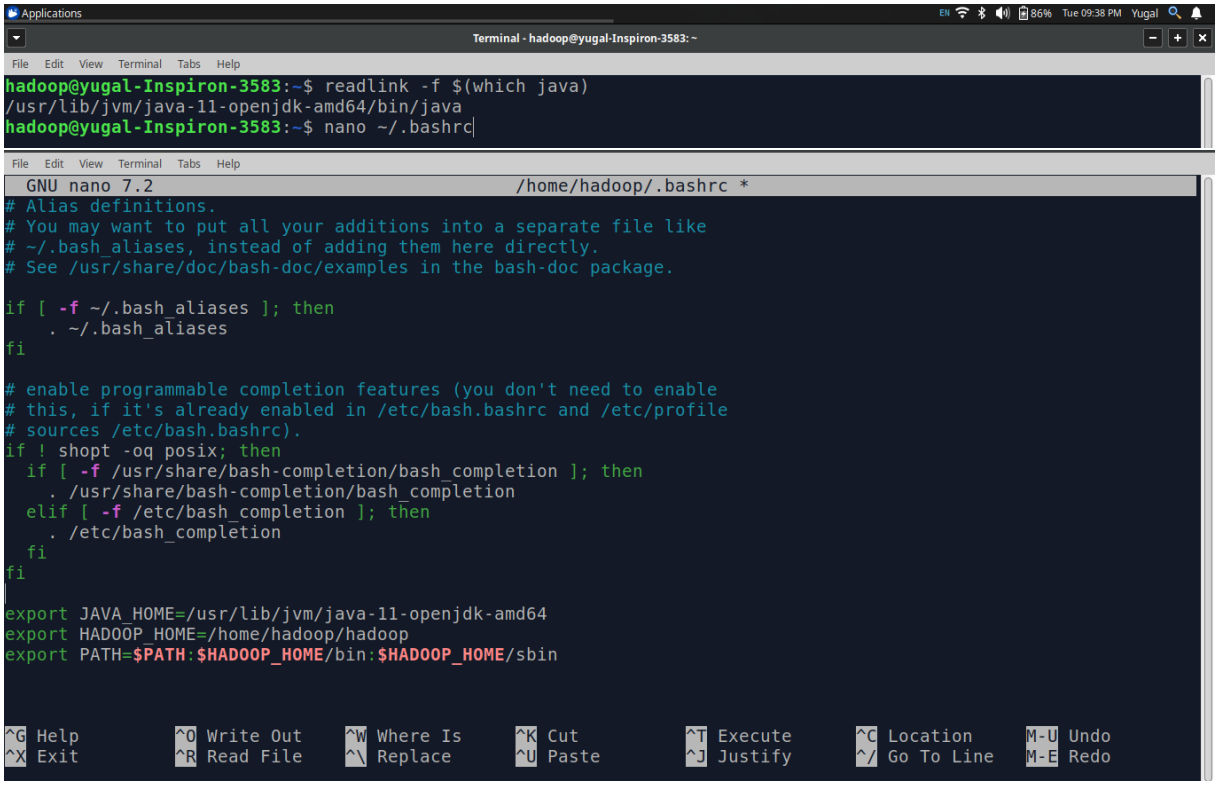

Figure 9: .bashrc Modification

1.5 Step 5: Hadoop XML Configuration Files

hadoop-env.sh Modification

```
source ~/.bashrc
echo $HADOOP_HOME
/home/hadoop/hadoop
cd $HADOOP_HOME/etc/hadoop
ls
nano hadoop-env.sh
% Add the following lines at the path
export JAVA_HOME=/usr/lib/jvm/java-11-openjdk-amd64
```

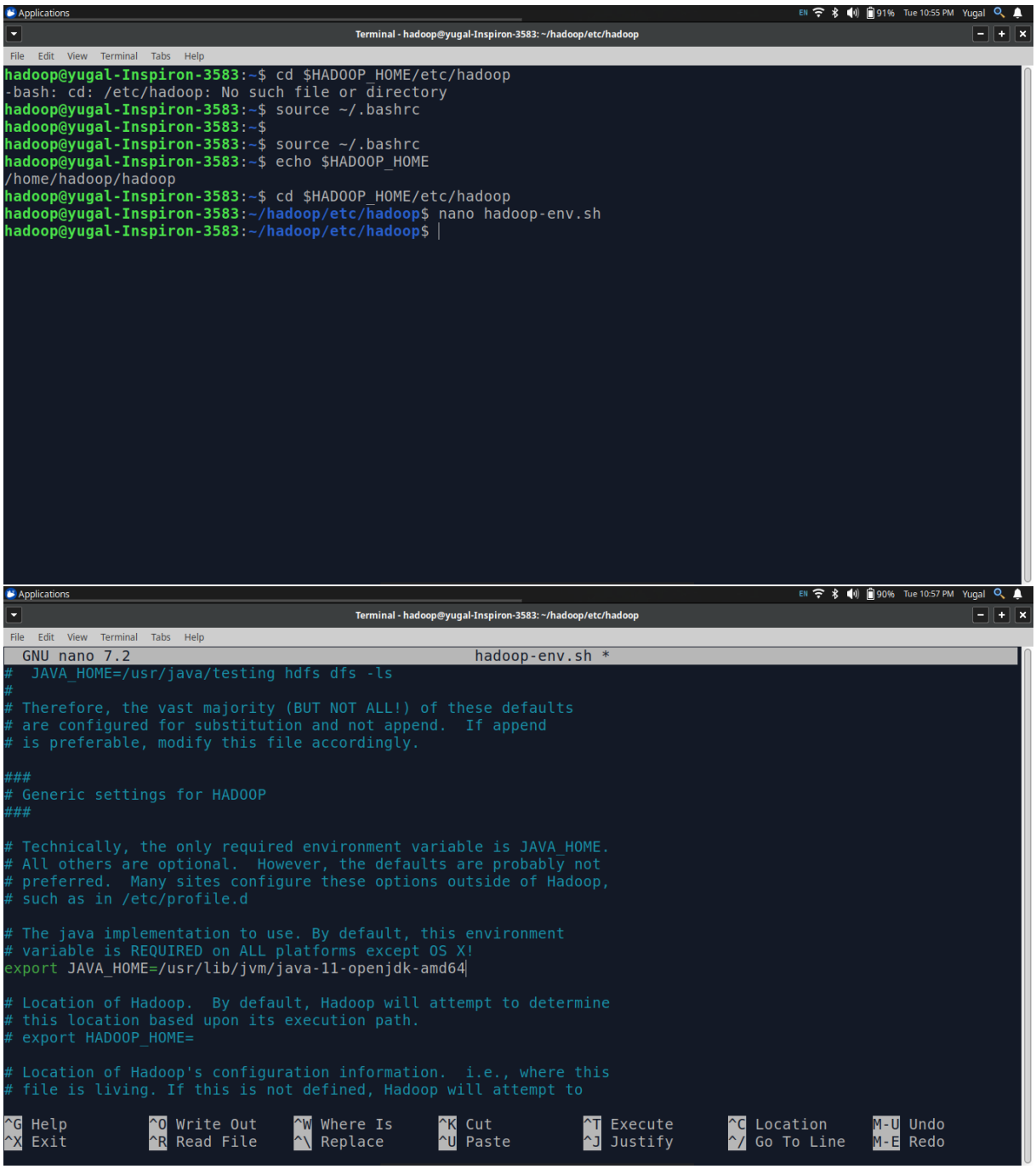

Figure 10: hadoop-env.sh Modification

core-site.xml Modification

```
nano core-site.xml
% Add the following property inside <configuration> tag
<property>
  <name>fs.defaultFS</name>
  <value>hdfs://localhost:9000</value>
```

```
</property>
```

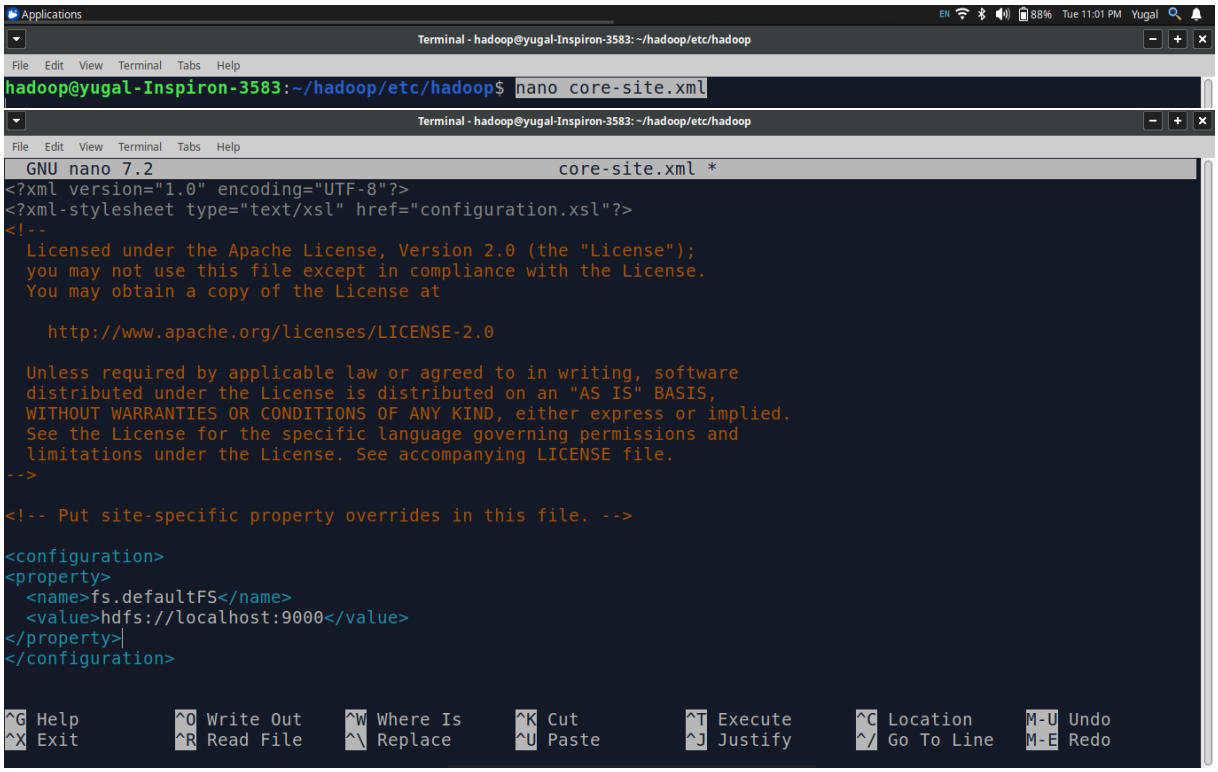

Figure 11: core-site.xml Modification

hdfs-site.xml Modification

```
nano hdfs-site.xml
% Add the following property inside <configuration> tag
<property>
  <name>dfs.replication</name>
  <value>1</value>
</property>

<property>
  <name>dfs.namenode.name.dir</name>
  <value>file:///home/hadoop/hdfs/namenode</value>
</property>

<property>
  <name>dfs.datanode.data.dir</name>
  <value>file:///home/hadoop/hdfs/datanode</value>
```

```
</property>
```

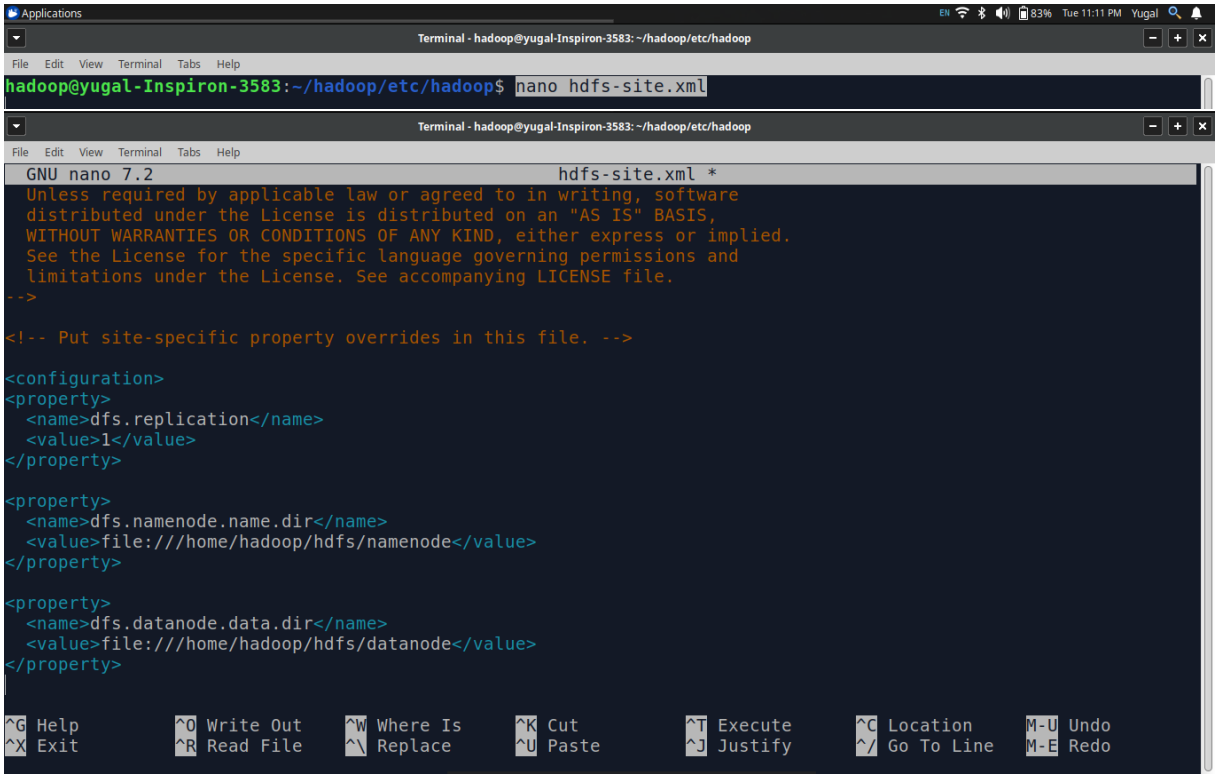

Figure 12: hdfs-site.xml Modification

Creating Folder for NameNode and DataNode

```
mkdir -p ~/hdfs/namenode
mkdir -p ~/hdfs/datanode
```

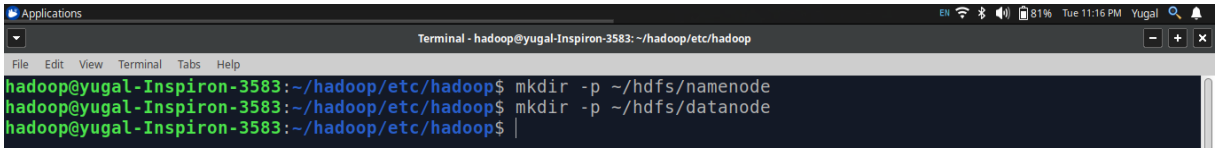

Figure 13: Creating Folder for NameNode and DataNode

mapred-site.xml Modification

```
nano mapred-site.xml
% Add the following property inside <configuration> tag
<property>
    <name>mapreduce.framework.name</name>
    <value>yarn</value>
</property>
```

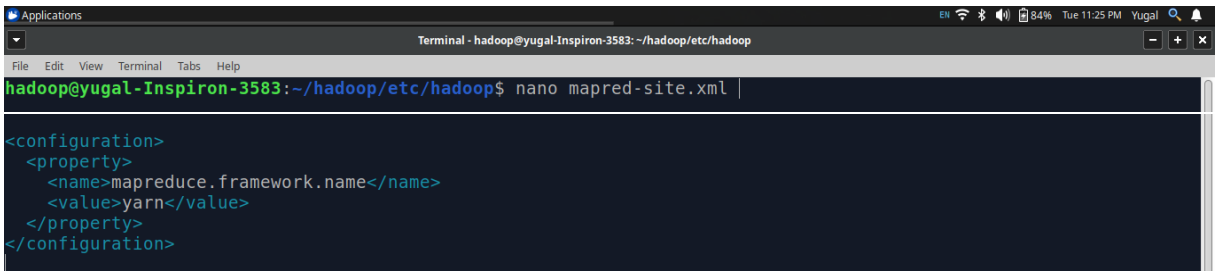

Figure 14: mapred-site.xml Modification

yarn-site.xml Modification

```
nano yarn-site.xml
% Add the following property inside <configuration> tag
<property>
  <name>yarn.nodemanager.aux-services</name>
  <value>mapreduce_shuffle</value>
</property>
```

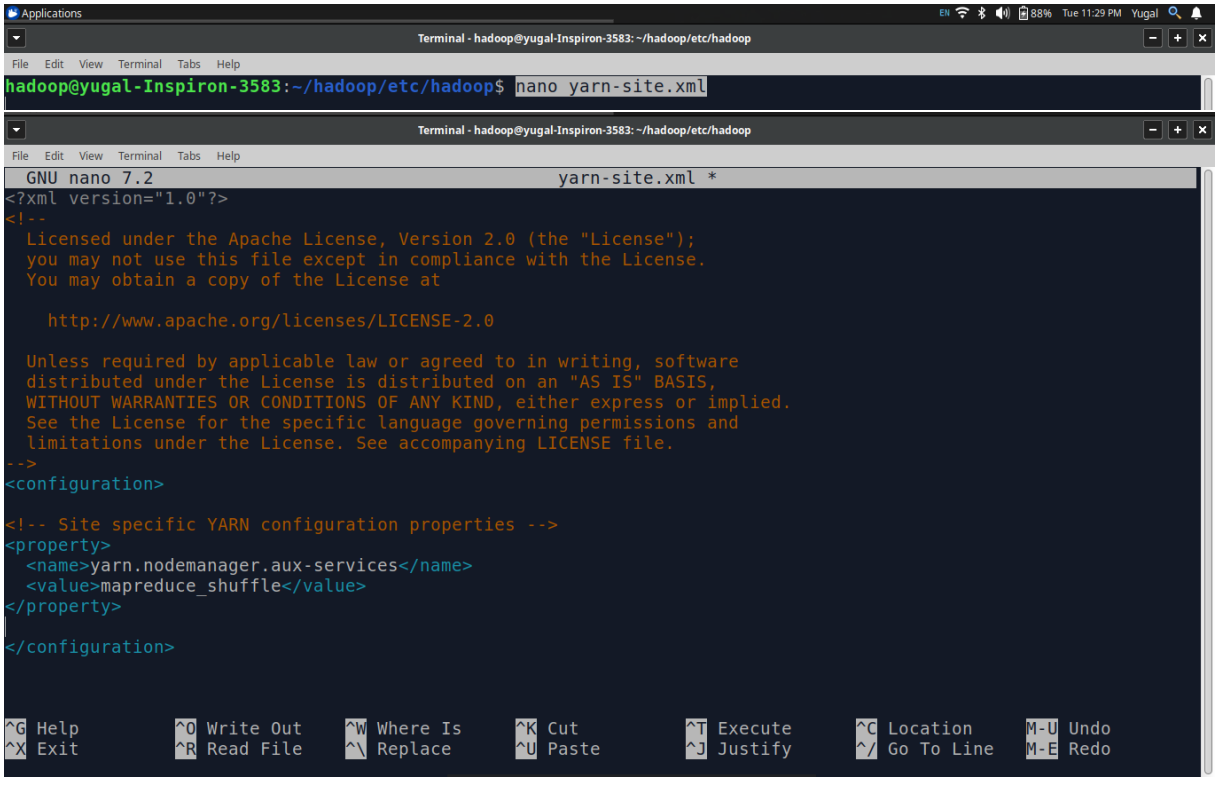

Figure 15: yarn-site.xml Modification

1.6 Step 6: Format NameNode & Start Hadoop Services

Format NameNode

```
hdfs namenode -format
```

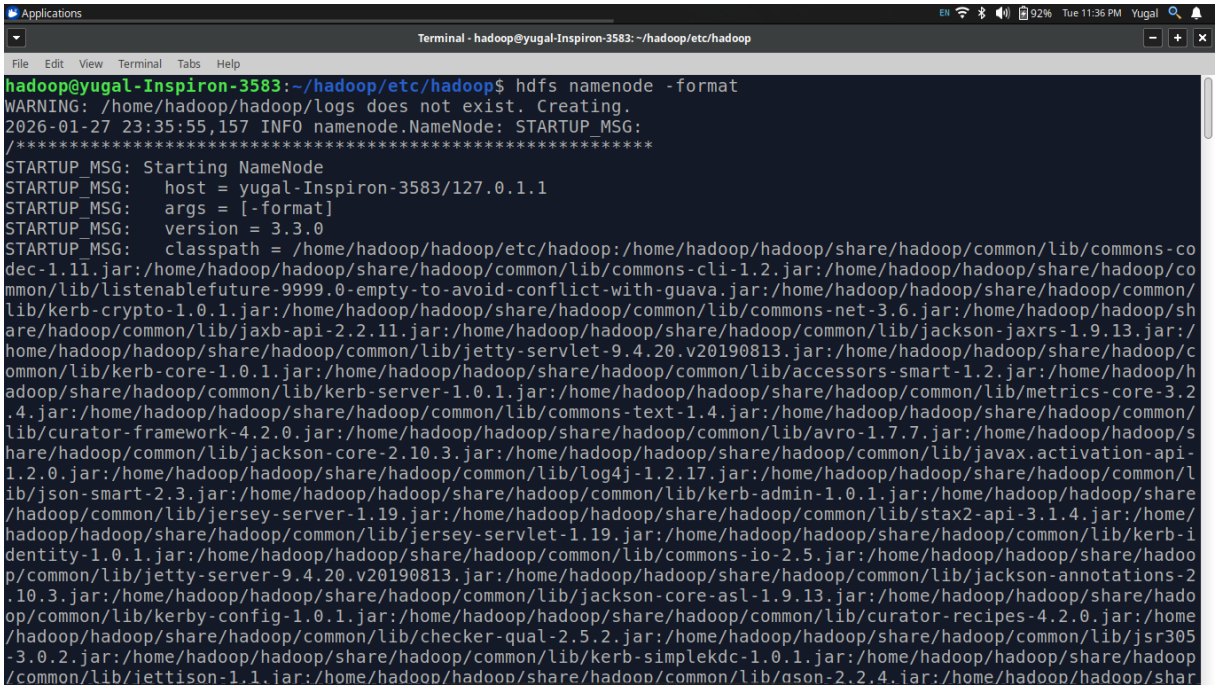

Figure 16: Format NameNode

Start Hadoop Services

```
start-dfs.sh
```

```
jps
```

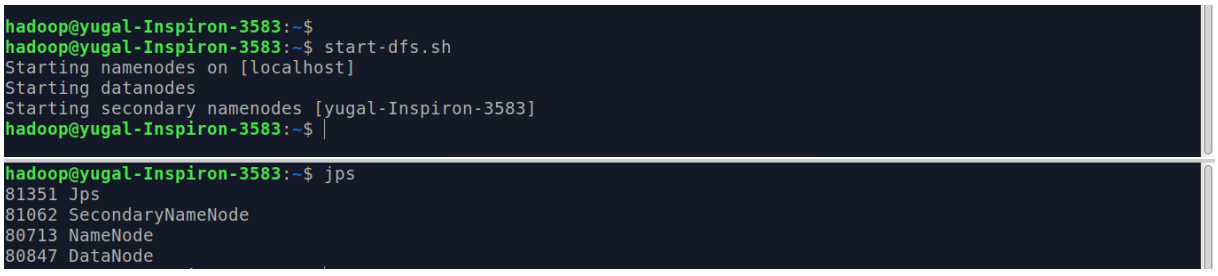

Figure 17: Start Hadoop Services

Start YARN Services

```
start-yarn.sh
```

```
jps
```

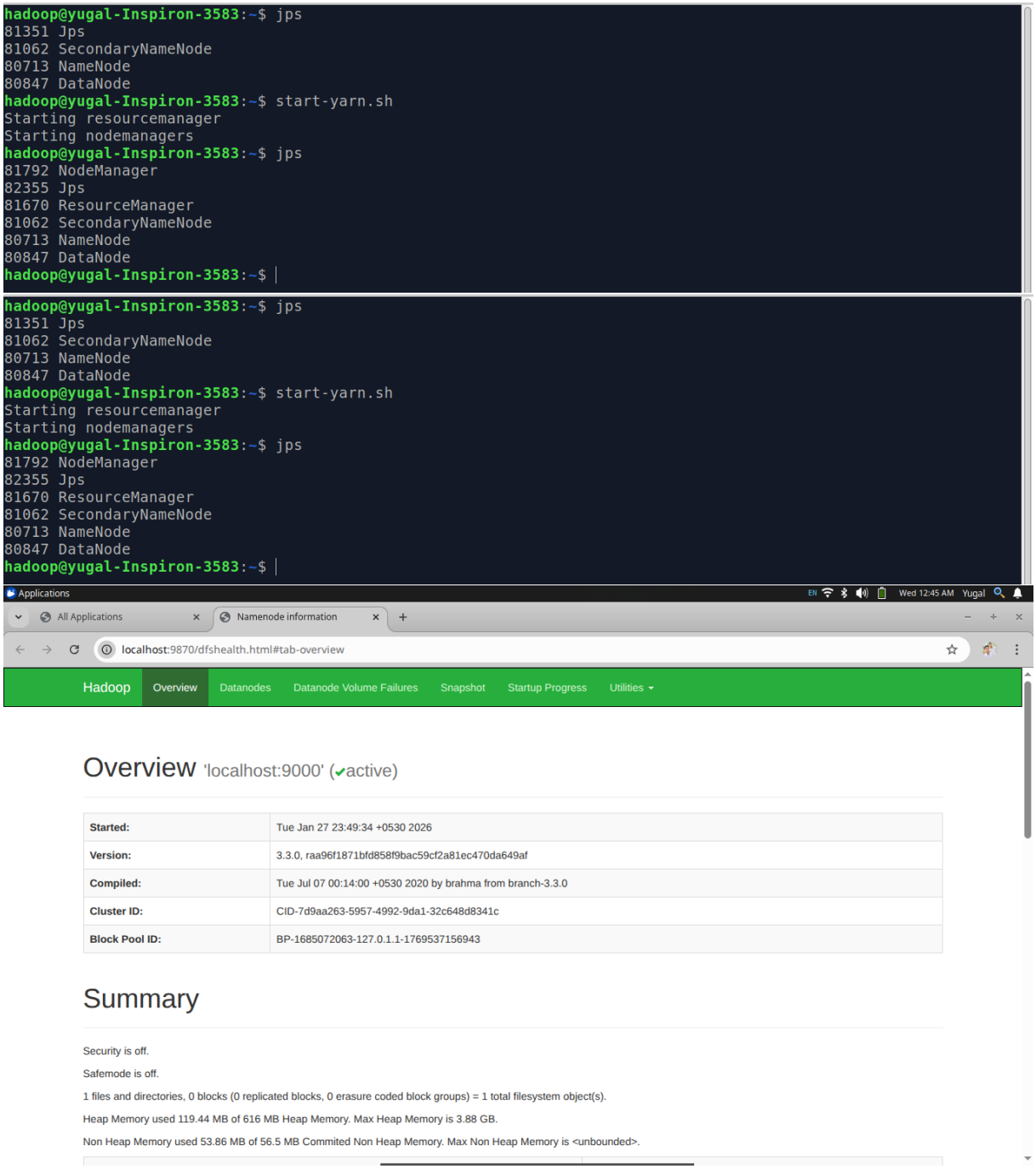

Figure 18: Start YARN Services

1.7 Step 7: HDFS Operations

```
% Create a test file
echo "Hello Big Data" > file.txt
% Make directory in HDFS
hdfs dfs -mkdir /data
```

```
% Upload file to HDFS
hdfs dfs -put file.txt /data
% List files
hdfs dfs -ls /data
% Check blocks & replication
hdfs fsck /data/file.txt -files -blocks
% Change permission
hdfs dfs -chmod 777 /data/file.txt
% Delete file
hdfs dfs -rm /data/file.txt
hdfs dfs -ls /data
```

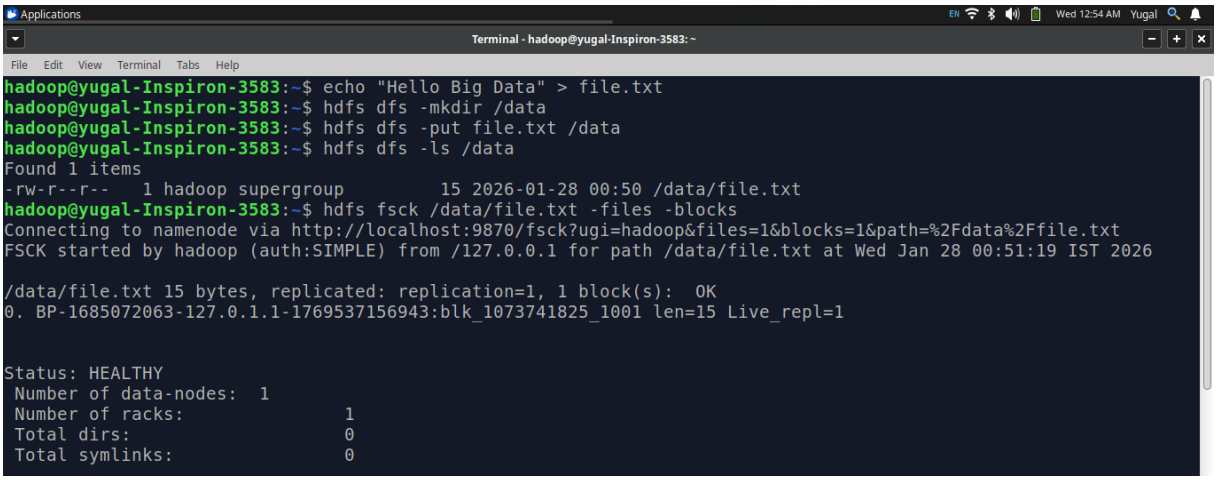

Figure 19: HDFS Operation 1

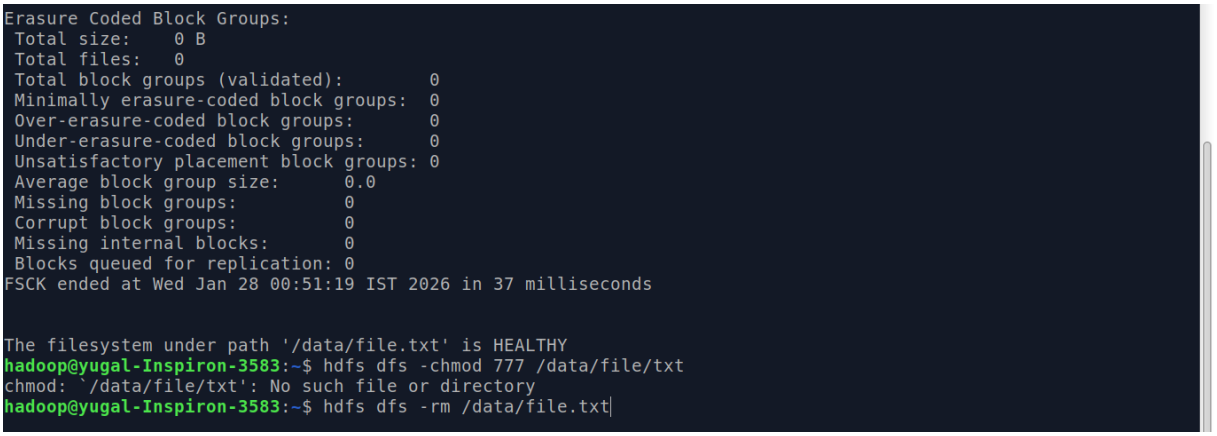

Figure 20: HDFS Operation 2

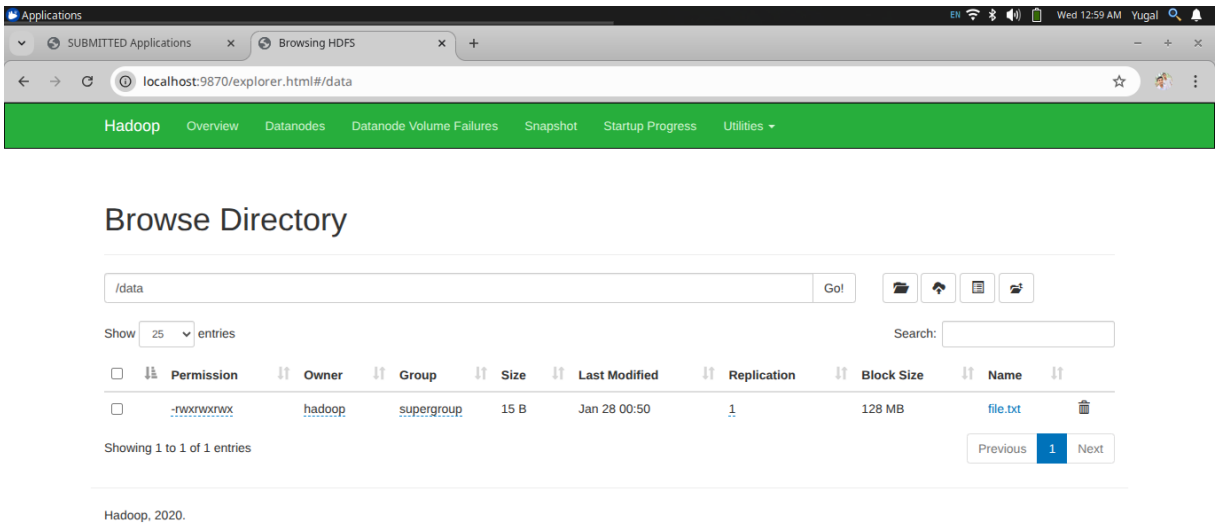

Figure 21: HDFS Result 1

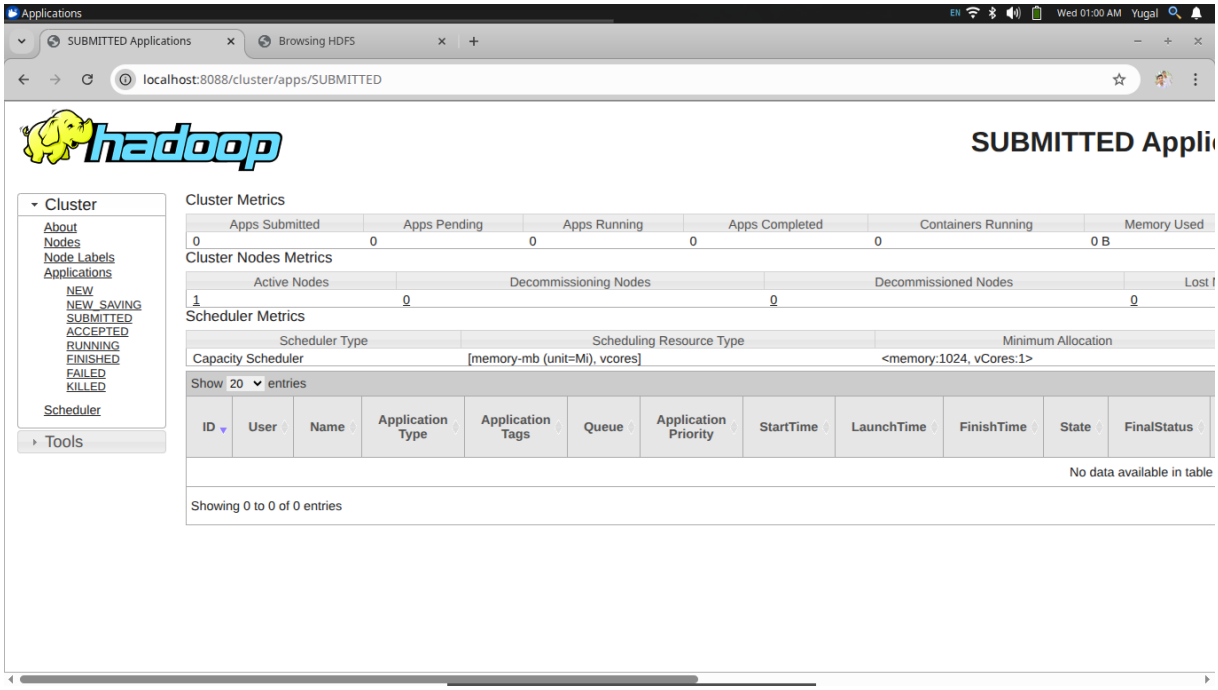

Figure 22: HDFS Result 2

1.8 Result

Hadoop was successfully installed and configured in pseudo-distributed mode, and various HDFS operations were performed successfully.

1.9 Conclusion

This experiment provided hands-on experience in setting up a Big Data environment using Hadoop and understanding the working of HDFS through practical command execution.