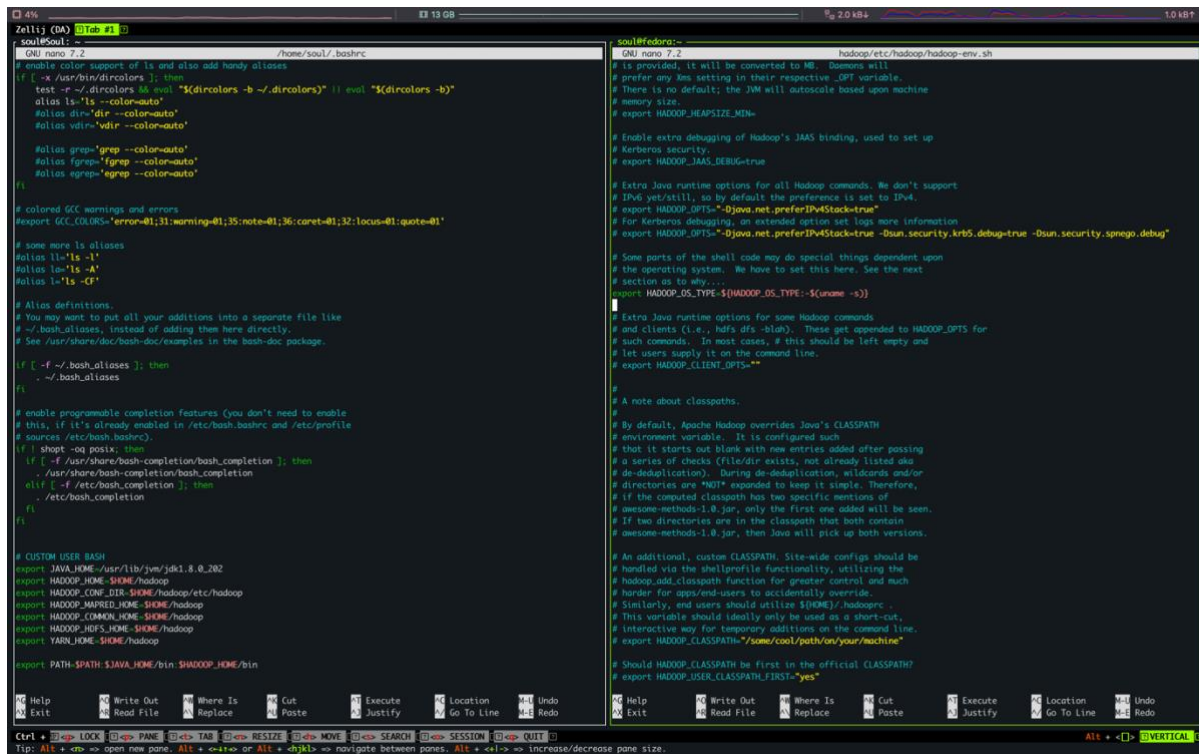


Exp No: 1

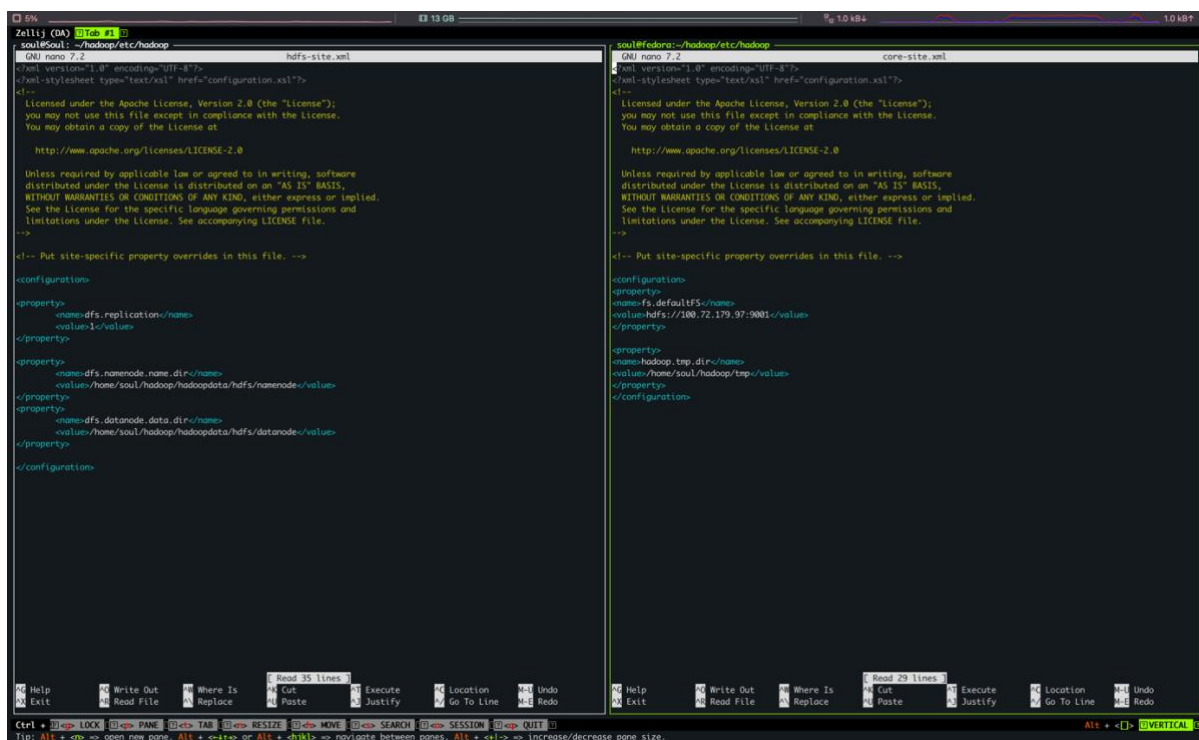
Downloading And Installing Hadoop, Understanding Different Hadoop Modes, Startup Scripts, Configuration Files.

Bash Profile (.bashrc) and Hadoop Profile (hadoop-env.sh):



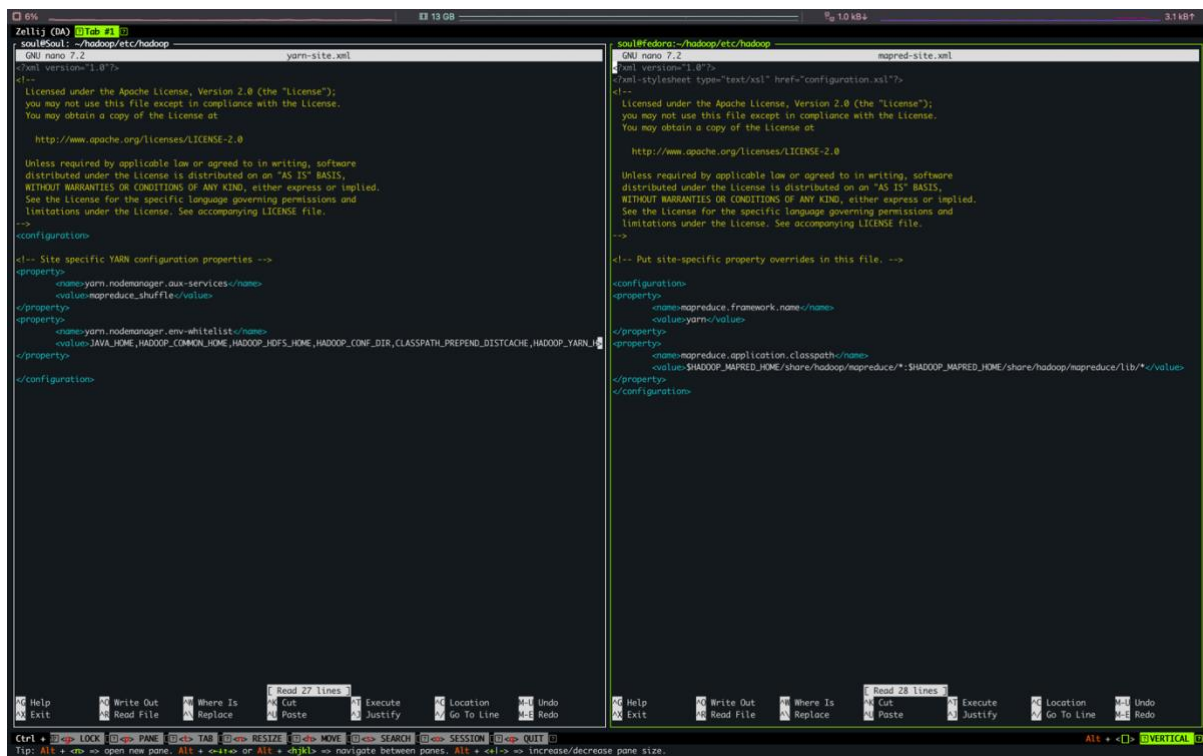
The image shows two terminal windows side-by-side. The left window is editing the `~/.bashrc` file in nano. It contains various shell aliases for `ls`, `grep`, and `find`, as well as colored GCC warnings and error messages. It also sets up Hadoop environment variables like `JAVA_HOME`, `HADOOP_HOME`, `HADOOP_CONF_DIR`, `HADOOP_MAPRED_HOME`, `HADOOP_COMMON_HOME`, `HADOOP_HDFS_HOME`, and `YARN_HOME`. The right window is editing the `hadoop/etc/hadoop/hadoop-env.sh` file. It contains comments about Java runtime options, Kerberos security, and Hadoop debugging options. It also sets `HADOOP_OPTS` and `HADOOP_CLIENT_OPTS` for different Hadoop commands.

HDFS Configuration (hdfs-site.xml, core-site.xml):



The image shows two terminal windows side-by-side. The left window is editing the `hdfs-site.xml` file in nano. It contains XML configuration for HDFS, including `dfs.replication` and `dfs.namenode.name.dir`. The right window is editing the `core-site.xml` file in nano. It contains XML configuration for the core Hadoop filesystem, including `fs.defaultFS` and `hadoop.tmp.dir`.

Resource Manager Configuration (yarn-site.xml, mapred-site.xml):



```
GNU nano 7.2 yarn-site.xml
<!--
Licensed under the Apache License, Version 2.0 (the "License");
you may not use this file except in compliance with the License.
You may obtain a copy of the License at

    http://www.apache.org/licenses/LICENSE-2.0

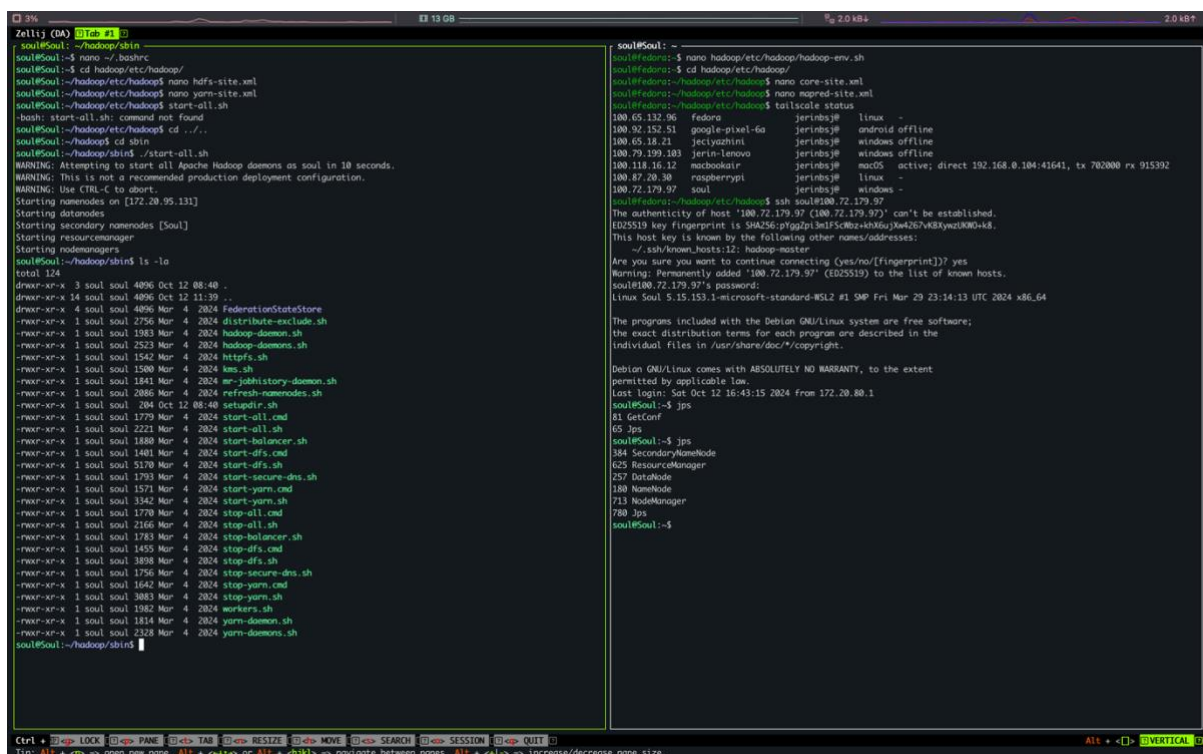
Unless required by applicable law or agreed to in writing, software
distributed under the License is distributed on an "AS IS" BASIS,
WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied.
See the License for the specific language governing permissions and
limitations under the License. See accompanying LICENSE file.
-->
<configuration>
<!-- Site specific YARN configuration properties -->
<property>
<name>yarn.nodemanager.aux-services</name>
<value>mapreduce_shuffle</value>
</property>
<property>
<name>yarn.nodemanager.env-whitelist</name>
<value>JAVA_HOME,HADOOP_COMMON_HOME,HADOOP_HDFS_HOME,HADOOP_CONF_DIR,CLASSPATH_PREPEND_DISTCACHE,HADOOP_YARN_HOME</value>
</property>
</configuration>

GNU nano 7.2 mapred-site.xml
<!--
Licensed under the Apache License, Version 2.0 (the "License");
you may not use this file except in compliance with the License.
You may obtain a copy of the License at

    http://www.apache.org/licenses/LICENSE-2.0

Unless required by applicable law or agreed to in writing, software
distributed under the License is distributed on an "AS IS" BASIS,
WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied.
See the License for the specific language governing permissions and
limitations under the License. See accompanying LICENSE file.
-->
<!-- Put site-specific property overrides in this file. -->
<configuration>
<property>
<name>mapreduce.framework.name</name>
<value>yarn</value>
</property>
<property>
<name>mapreduce.application.classpath</name>
<value>${HADOOP_MAPRED_HOME}/share/hadoop/mapreduce/*:${HADOOP_MAPRED_HOME}/share/hadoop/mapreduce/lib/*</value>
</property>
</configuration>
```

Starting Hadoop Cluster and Verifying (./start-all.sh, jps) :



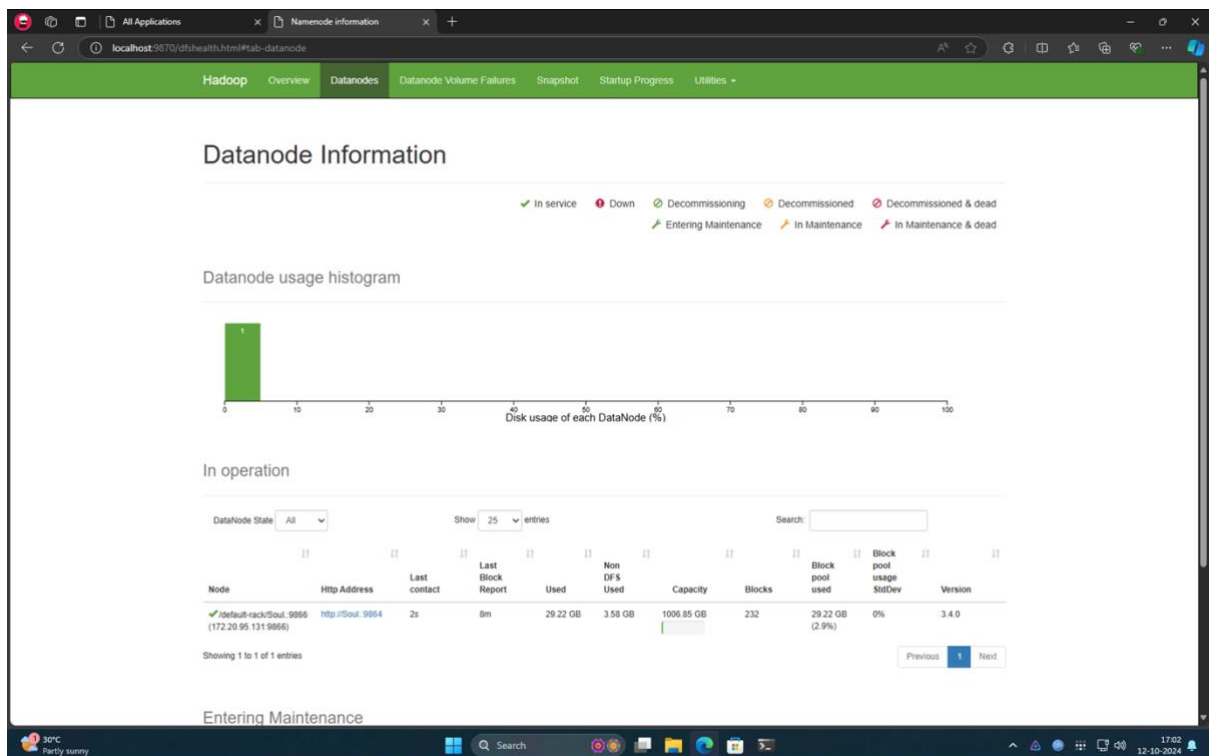
```
GNU nano 7.2 /etc/ssh/ssh_config
soul@Soul: ~/hadoop/sbin
soul@Soul:~$ cd /etc/hadoop/
soul@Soul:~/etc/hadoop$ nano hdfs-site.xml
soul@Soul:~/etc/hadoop$ nano yarn-site.xml
soul@Soul:~/etc/hadoop$ start-all.sh
-bash: start-all.sh: command not found
soul@Soul:~/etc/hadoop$ cd ../..
soul@Soul:~/hadoop$ cd sbin
soul@Soul:~/hadoop/sbin$ ./start-all.sh
WARNING: Attempting to start all Apache Hadoop daemons as soul in 10 seconds.
WARNING: This is not a recommended production deployment configuration.
WARNING: Use CTRL-C to abort.
Starting namenodes on [172.20.95.131]
Starting datanodes
Starting secondary namenodes [soul]
Starting resourcemanager
Starting nodemanager
soul@Soul:~/hadoop/sbin$ ls -la
total 124
drwxr-xr-x 3 soul soul 4096 Oct 12 08:40 .
drwxr-xr-x 14 soul soul 4096 Oct 12 11:39 ..
drwxr-xr-x 4 soul soul 4096 Mar 4 2024 FederationStateStore
-rwxr-xr-x 1 soul soul 2756 Mar 4 2024 distribute-exclude.sh
-rwxr-xr-x 1 soul soul 1983 Mar 4 2024 hadoop-daemon.sh
-rwxr-xr-x 1 soul soul 2523 Mar 4 2024 hadoop-daemons.sh
-rwxr-xr-x 1 soul soul 1542 Mar 4 2024 httpfs.sh
-rwxr-xr-x 1 soul soul 1580 Mar 4 2024 kms.sh
-rwxr-xr-x 1 soul soul 1841 Mar 4 2024 mr-jobhistory-daemon.sh
-rwxr-xr-x 1 soul soul 2086 Mar 4 2024 refresh-namenodes.sh
-rwxr-xr-x 1 soul soul 204 Oct 12 08:40 setupdfr.sh
-rwxr-xr-x 1 soul soul 1770 Mar 4 2024 start-all.cmd
-rwxr-xr-x 1 soul soul 2221 Mar 4 2024 start-all.sh
-rwxr-xr-x 1 soul soul 1880 Mar 4 2024 start-balancer.sh
-rwxr-xr-x 1 soul soul 1481 Mar 4 2024 start-dfs.cmd
-rwxr-xr-x 1 soul soul 5170 Mar 4 2024 start-dfs.sh
-rwxr-xr-x 1 soul soul 1793 Mar 4 2024 start-secure-dns.sh
-rwxr-xr-x 1 soul soul 1571 Mar 4 2024 start-yarn.cmd
-rwxr-xr-x 1 soul soul 3342 Mar 4 2024 start-yarn.sh
-rwxr-xr-x 1 soul soul 1770 Mar 4 2024 stop-all.cmd
-rwxr-xr-x 1 soul soul 2166 Mar 4 2024 stop-all.sh
-rwxr-xr-x 1 soul soul 1783 Mar 4 2024 stop-balancer.sh
-rwxr-xr-x 1 soul soul 1455 Mar 4 2024 stop-dfs.cmd
-rwxr-xr-x 1 soul soul 3898 Mar 4 2024 stop-dfs.sh
-rwxr-xr-x 1 soul soul 1756 Mar 4 2024 stop-secure-dns.sh
-rwxr-xr-x 1 soul soul 1642 Mar 4 2024 stop-yarn.cmd
-rwxr-xr-x 1 soul soul 3083 Mar 4 2024 stop-yarn.sh
-rwxr-xr-x 1 soul soul 1982 Mar 4 2024 workers.sh
-rwxr-xr-x 1 soul soul 1814 Mar 4 2024 yarn-daemon.sh
-rwxr-xr-x 1 soul soul 2328 Mar 4 2024 yarn-daemons.sh
soul@Soul:~/hadoop/sbin$

GNU nano 7.2 /etc/ssh/ssh_config
soul@Soul:~$
soul@Soul:~$ nano /etc/hadoop/hadoop-env.sh
soul@Soul:~$ cd /etc/hadoop/
soul@Soul:~/etc/hadoop$ nano core-site.xml
soul@Soul:~/etc/hadoop$ nano mapred-site.xml
soul@Soul:~/etc/hadoop$ tailcat status
100.65.132.96 Fedora jerinbsj@ linux -
100.92.152.51 google-pixel-6a jerinbsj@ android offline
100.65.18.21 jeciyachini jerinbsj@ windows offline
100.79.199.103 jerin-lenovo jerinbsj@ windows offline
100.118.16.12 macbookair jerinbsj@ macOS active; direct 192.168.0.104:41641, tx 702000 rx 915392
100.87.20.30 raspberry jerinbsj@ linux -
100.72.179.97 soul jerinbsj@ windows -
soul@Soul:~/etc/hadoop$ ssh soul@100.72.179.97
The authenticity of host '100.72.179.97 (100.72.179.97)' can't be established.
ED25519 key fingerprint is SHA256:pGgZpJm1SchtzshhJpJw4E2PK8yWxR0wK8.
This host key is known by the following other names/addresses:
./ssh/known_hosts:12: hadoop-master
Are you sure you want to continue connecting (yes/no/[fingerprint])? yes
Warning: Permanently added '100.72.179.97' (ED25519) to the list of known hosts.
soul@100.72.179.97's password:
Linux Soul 5.15.153.1-microsoft-standard-WSL2 #1 SMP Fri Mar 29 23:14:13 UTC 2024 x86_64

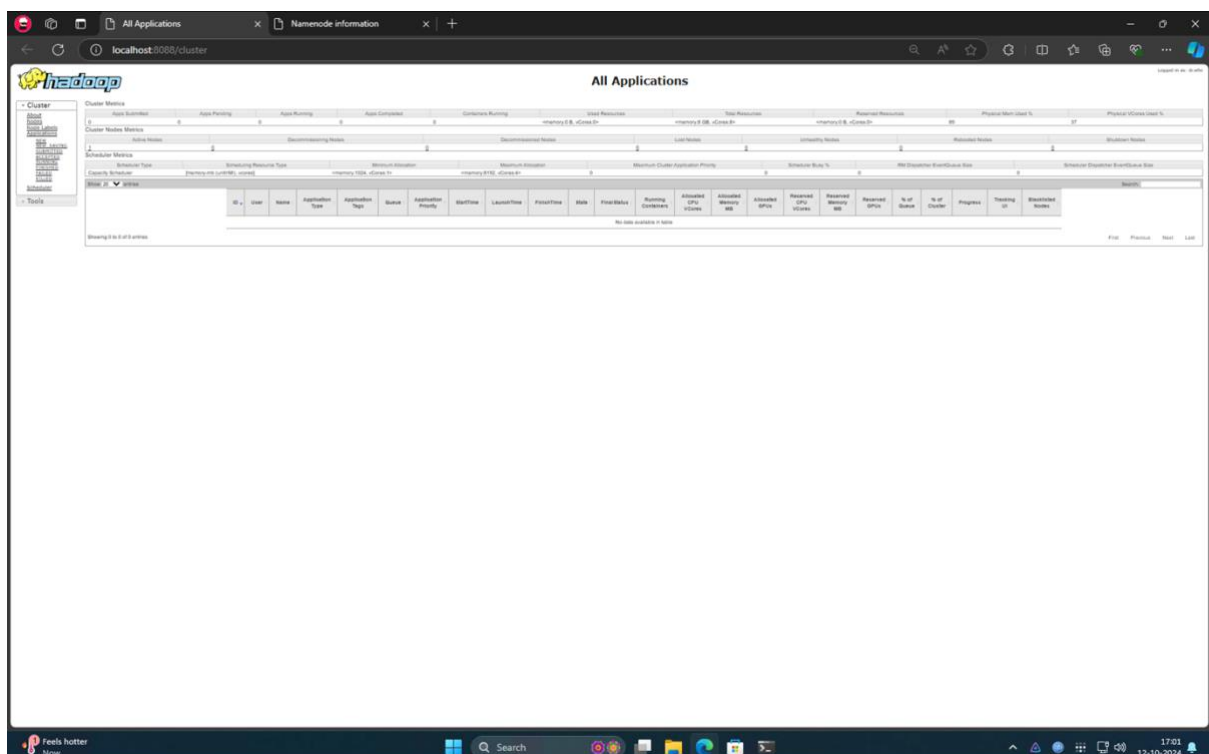
The programs included with the Debian GNU/Linux system are free software;
the exact distribution terms for each program are described in the
individual files in /usr/share/doc/*/*copyright.

Debian GNU/Linux comes with ABSOLUTELY NO WARRANTY, to the extent
permitted by applicable law.
Last login: Sat Oct 12 16:43:15 2024 from 172.20.80.1
soul@Soul:~$ jps
61 GetConf
65 Jps
364 SecondaryNameNode
626 ResourceManager
257 DataNode
180 NameNode
713 NodeManager
780 Jps
soul@Soul:~$
```

```
http://<hadoop-master-ip>:9870
```



```
http://<hadoop-master-ip>:8088
```



Result:

Thus Hadoop HDFS and YARN is successfully configured and started.