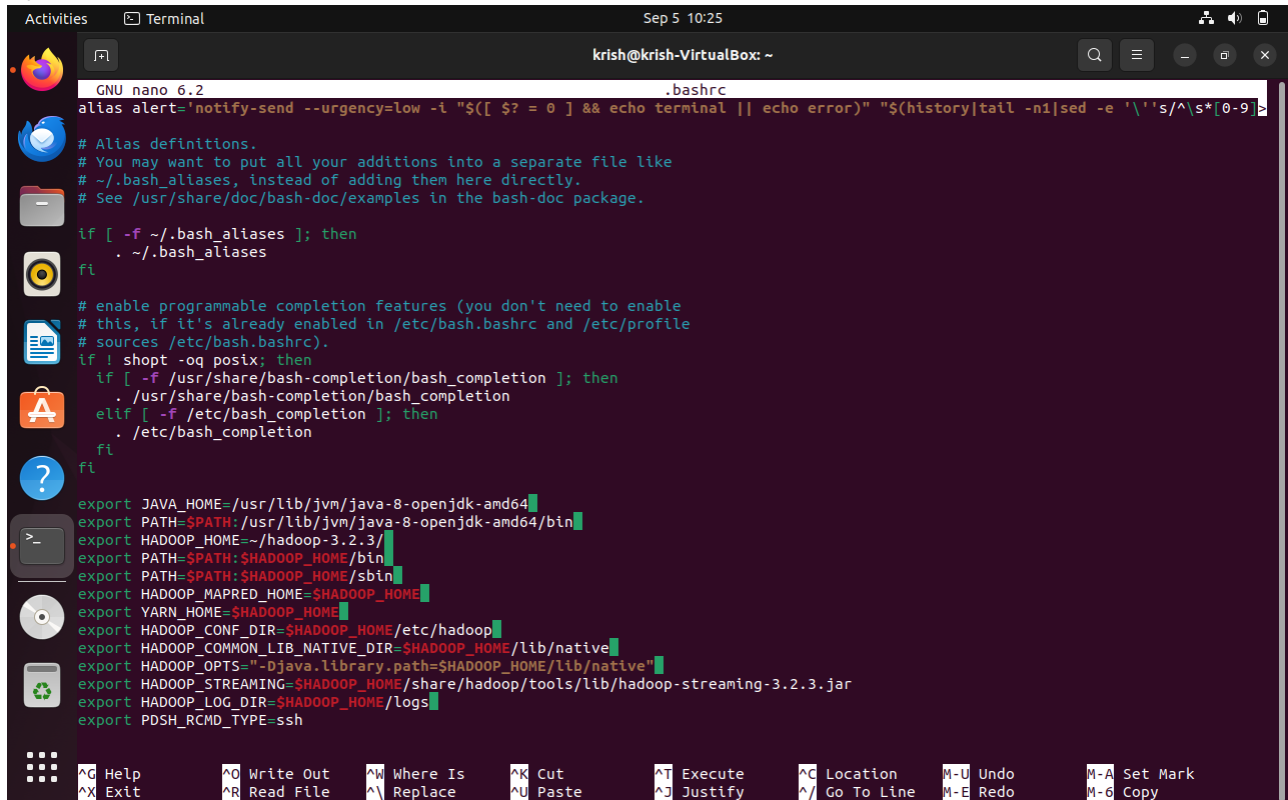


## Exp No: 1 Downloading and installing Hadoop, Understanding different Hadoop modes, Startup scripts, Configuration files.

\$ nano ~/.bashrc



```

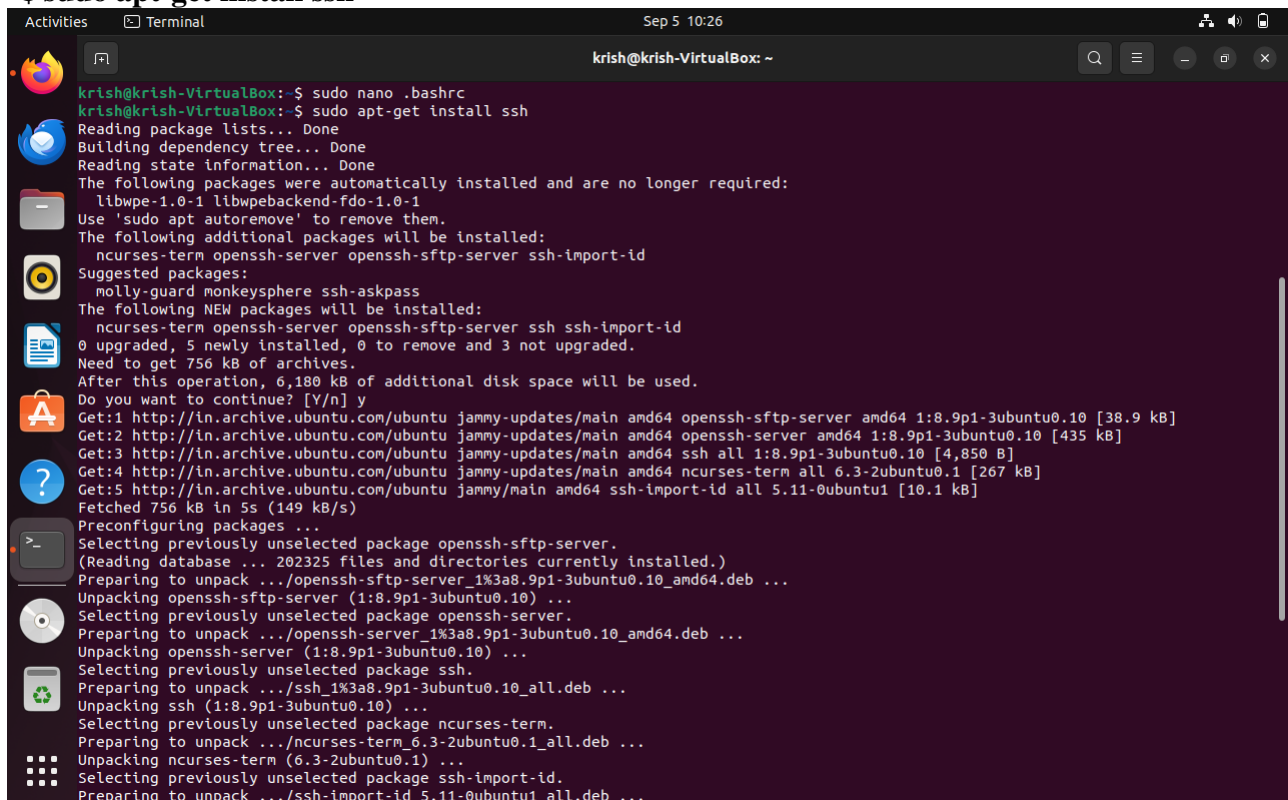
GNU nano 6.2                               .bashrc
alias alert='notify-send --urgency=low -i "${?} = 0" && echo terminal || echo error" "${history|tail -n1|sed -e '\''s/\s*[0-9]>
# Alias definitions.
# You may want to put all your additions into a separate file like
# ~/.bash_aliases, instead of adding them here directly.
# See /usr/share/doc/bash-doc/examples in the bash-doc package.

if [ -f ~/.bash_aliases ]; then
. ~/.bash_aliases
fi

# enable programmable completion features (you don't need to enable
# this, if it's already enabled in /etc/bash.bashrc and /etc/profile
# sources /etc/bash.bashrc).
if ! shopt -oq posix; then
if [ -f /usr/share/bash-completion/bash_completion ]; then
. /usr/share/bash-completion/bash_completion
elif [ -f /etc/bash_completion ]; then
. /etc/bash_completion
fi
fi

export JAVA_HOME=/usr/lib/jvm/java-8-openjdk-amd64
export PATH=$PATH:/usr/lib/jvm/java-8-openjdk-amd64/bin
export HADOOP_HOME=~/hadoop-3.2.3/
export PATH=$PATH:$HADOOP_HOME/bin
export PATH=$PATH:$HADOOP_HOME/sbin
export HADOOP_MAPRED_HOME=$HADOOP_HOME
export YARN_HOME=$HADOOP_HOME
export HADOOP_CONF_DIR=$HADOOP_HOME/etc/hadoop
export HADOOP_COMMON_LIB_NATIVE_DIR=$HADOOP_HOME/lib/native
export HADOOP_OPTS="-Djava.library.path=$HADOOP_HOME/lib/native"
export HADOOP_STREAMING=$HADOOP_HOME/share/hadoop/tools/lib/hadoop-streaming-3.2.3.jar
export HADOOP_LOG_DIR=$HADOOP_HOME/logs
export PDSH_RCMD_TYPE=ssh
  
```

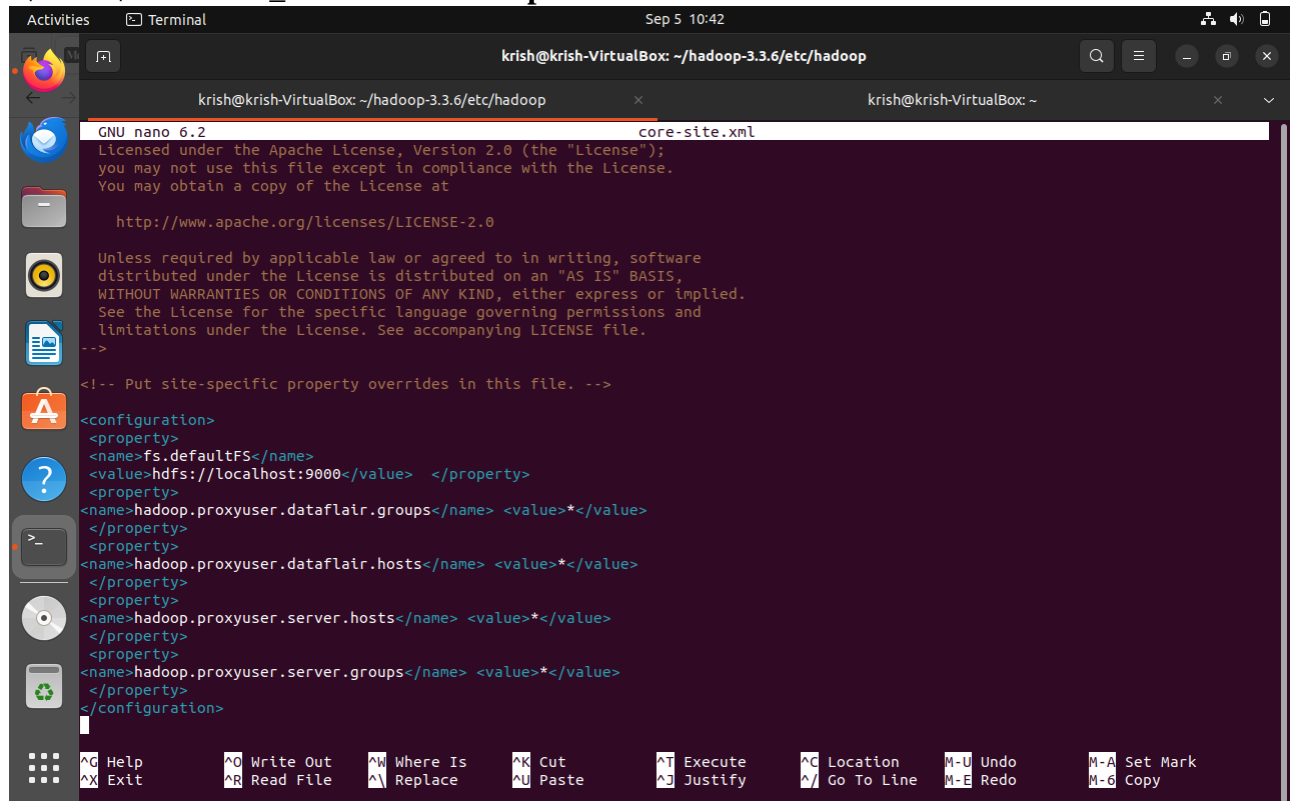
\$ sudo apt-get install ssh



```

krish@krish-VirtualBox:~$ sudo nano .bashrc
krish@krish-VirtualBox:~$ sudo apt-get install ssh
Reading package lists... Done
Building dependency tree... Done
Reading state information... Done
The following packages were automatically installed and are no longer required:
  libwp6-1.0-1 libwp6backend-fdo-1.0-1
Use 'sudo apt autoremove' to remove them.
The following additional packages will be installed:
  ncurses-term openssh-server openssh-sftp-server ssh-import-id
Suggested packages:
  molly-guard monkeysphere ssh-askpass
The following NEW packages will be installed:
  ncurses-term openssh-server openssh-sftp-server ssh ssh-import-id
0 upgraded, 5 newly installed, 0 to remove and 3 not upgraded.
Need to get 756 kB of archives.
After this operation, 6,180 kB of additional disk space will be used.
Do you want to continue? [Y/n] y
Get:1 http://in.archive.ubuntu.com/ubuntu jammy-updates/main amd64 openssh-sftp-server amd64 1:8.9p1-3ubuntu0.10 [38.9 kB]
Get:2 http://in.archive.ubuntu.com/ubuntu jammy-updates/main amd64 openssh-server amd64 1:8.9p1-3ubuntu0.10 [435 kB]
Get:3 http://in.archive.ubuntu.com/ubuntu jammy-updates/main amd64 ssh all 1:8.9p1-3ubuntu0.10 [4,850 B]
Get:4 http://in.archive.ubuntu.com/ubuntu jammy-updates/main amd64 ncurses-term all 6.3-2ubuntu0.1 [267 kB]
Get:5 http://in.archive.ubuntu.com/ubuntu jammy/main amd64 ssh-import-id all 5.11-0ubuntu1 [10.1 kB]
Fetched 756 kB in 5s (149 kB/s)
Preconfiguring packages ...
Selecting previously unselected package openssh-sftp-server.
(Reading database ... 202325 files and directories currently installed.)
Preparing to unpack .../openssh-sftp-server_1%3a8.9p1-3ubuntu0.10_amd64.deb ...
Unpacking openssh-sftp-server (1:8.9p1-3ubuntu0.10) ...
Selecting previously unselected package openssh-server.
Preparing to unpack .../openssh-server_1%3a8.9p1-3ubuntu0.10_amd64.deb ...
Unpacking openssh-server (1:8.9p1-3ubuntu0.10) ...
Selecting previously unselected package ssh.
Preparing to unpack .../ssh_1%3a8.9p1-3ubuntu0.10_all.deb ...
Unpacking ssh (1:8.9p1-3ubuntu0.10) ...
Selecting previously unselected package ncurses-term.
Preparing to unpack .../ncurses-term_6.3-2ubuntu0.1_all.deb ...
Unpacking ncurses-term (6.3-2ubuntu0.1) ...
Selecting previously unselected package ssh-import-id.
Preparing to unpack .../ssh-import-id_5.11-0ubuntu1_all.deb ...
  
```

## \$nano \$HADOOP\_HOME/etc/hadoop/core-site.xml



```
krish@krish-VirtualBox: ~/hadoop-3.3.6/etc/hadoop
GNU nano 6.2 core-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

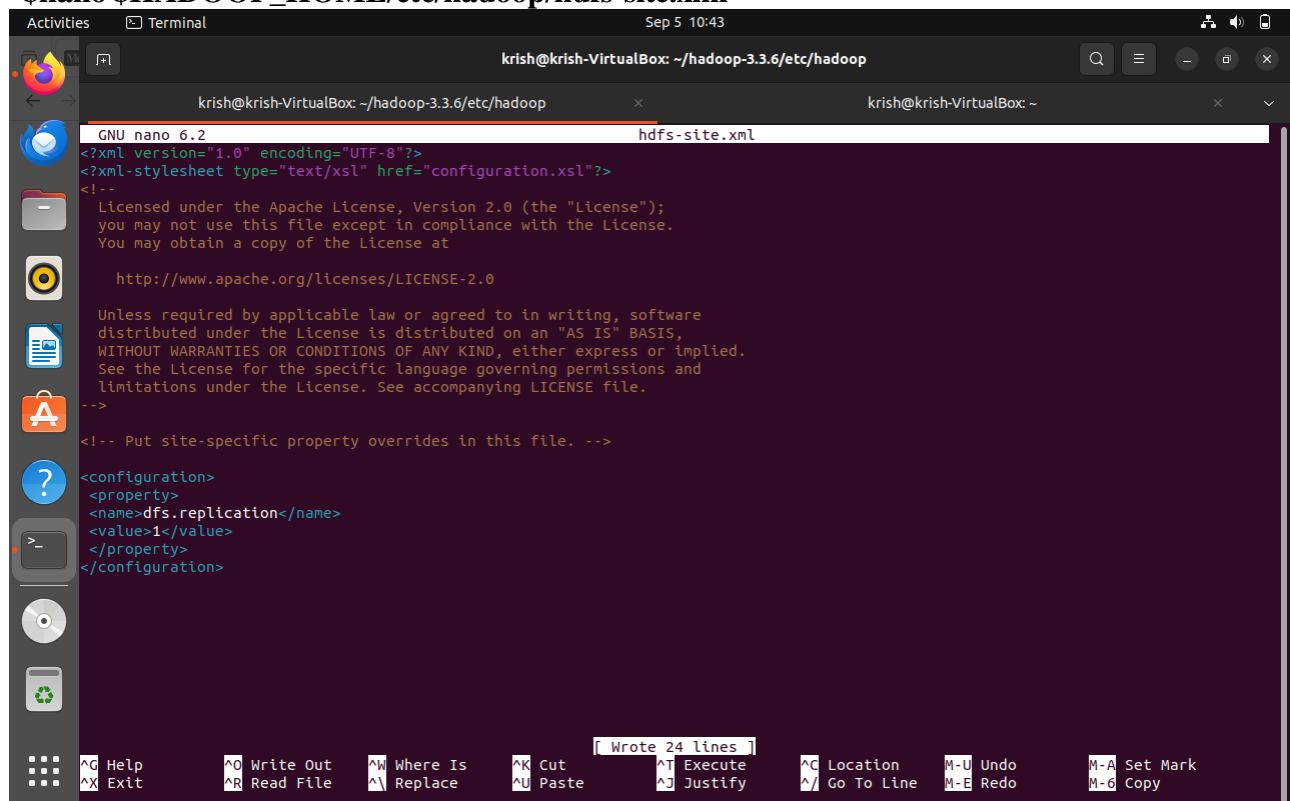
    http://www.apache.org/licenses/LICENSE-2.0

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>fs.defaultFS</name>
    <value>hdfs://localhost:9000</value>  </property>
  <property>
    <name>hadoop.proxyuser.dataflair.groups</name> <value>*</value>
  </property>
  <property>
    <name>hadoop.proxyuser.dataflair.hosts</name> <value>*</value>
  </property>
  <property>
    <name>hadoop.proxyuser.server.hosts</name> <value>*</value>
  </property>
  <property>
    <name>hadoop.proxyuser.server.groups</name> <value>*</value>
  </property>
</configuration>
```

## \$nano \$HADOOP\_HOME/etc/hadoop/hdfs-site.xml



```
krish@krish-VirtualBox: ~/hadoop-3.3.6/etc/hadoop
GNU nano 6.2 hdfs-site.xml
<?xml version="1.0" encoding="UTF-8"?>
<?xml-stylesheet type="text/xsl" href="configuration.xsl"?>
<!--
  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>dfs.replication</name>
    <value>1</value>
  </property>
</configuration>
```

Wrote 24 lines

**\$nano \$HADOOP\_HOME/etc/hadoop/mapred-site.xml**

```

krish@krish-VirtualBox: ~/hadoop-3.3.6/etc/hadoop
GNU nano 6.2 mapred-site.xml
<?xml version="1.0"?>
<?xml-stylesheet type="text/xsl" href="configuration.xsl"?>
<!--
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>

```

Wrote 28 lines

Help Exit Write Out Read File Where Is Replace Cut Paste Execute Justify Location Go To Line Undo Redo Set Mark Copy

**\$nano \$HADOOP\_HOME/etc/hadoop/yarn-site.xml**

```

krish@krish-VirtualBox: ~/hadoop-3.3.6/etc/hadoop
GNU nano 6.2 yarn-site.xml
<?xml version="1.0"?>
<!--
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.
-->

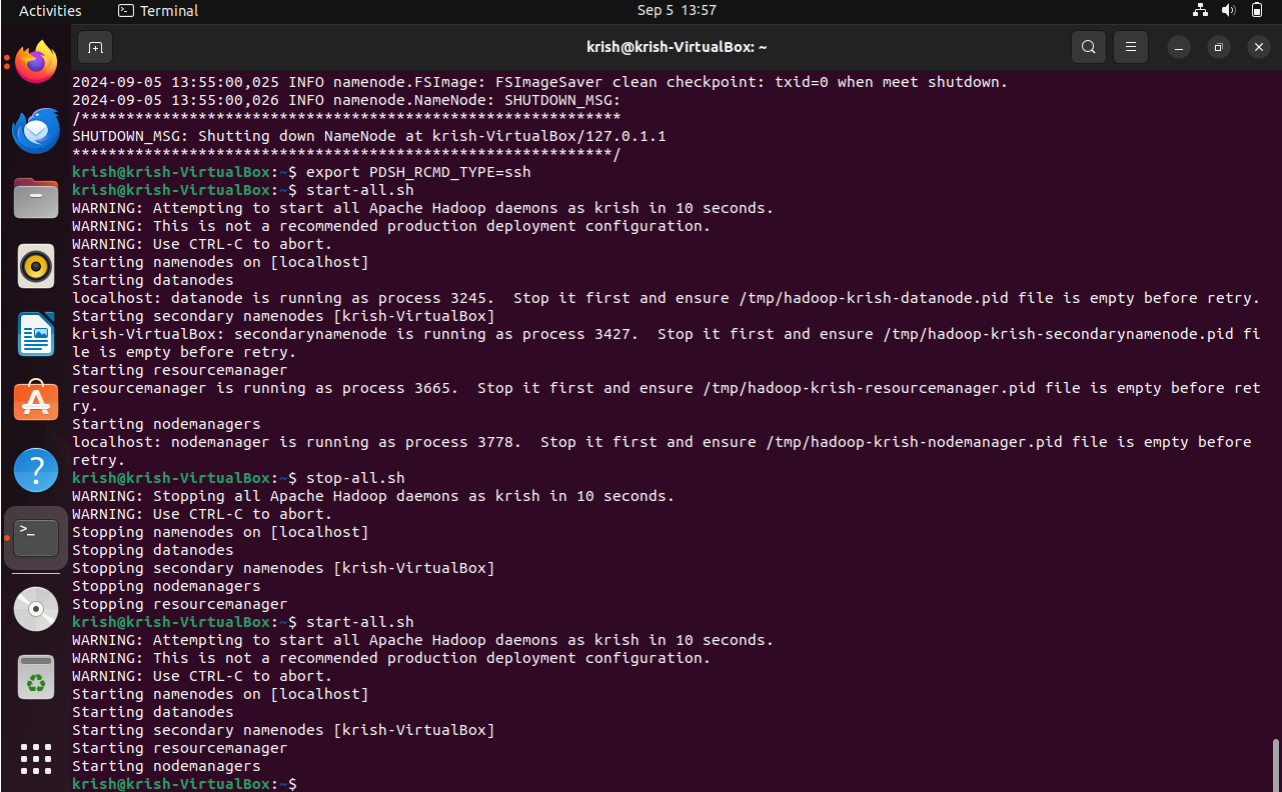
<configuration>
  <property>
    <name>yarn.nodemanager.aux-services</name>
    <value>mapreduce_shuffle</value>
  </property>
  <property>
    <name>yarn.nodemanager.env-whitelist</name>
    <value>JAVA_HOME,HADOOP_COMMON_HOME,HADOOP_HDFS_HOME,HADOOP_CONF_DIR,CLASSPATH_PREP END_DISTCACHE,HADOOP_YARN_HOME,HADOOP_MAPRED_HO
  </property>
</configuration>

```

Wrote 26 lines

Help Exit Write Out Read File Where Is Replace Cut Paste Execute Justify Location Go To Line Undo Redo Set Mark Copy

\$ start-all.sh



A terminal window titled "krish@krish-VirtualBox: ~" with a search bar and window controls. The terminal displays a sequence of Hadoop startup and shutdown logs. The startup sequence begins with a warning about the production configuration, followed by the starting of namenodes on localhost, datanodes on localhost, secondary namenodes on krish-VirtualBox, and resource managers on localhost. The shutdown sequence follows, stopping all components in reverse order. The terminal also shows a manual export of the PDSH\_RCMD\_TYPE variable and a final startup attempt.

```
2024-09-05 13:55:00,025 INFO namenode.FSImage: FSImageSaver clean checkpoint: txid=0 when meet shutdown.
2024-09-05 13:55:00,026 INFO namenode.NameNode: SHUTDOWN_MSG:
/*****
krish@krish-VirtualBox:~$ export PDSH_RCMD_TYPE=ssh
krish@krish-VirtualBox:~$ start-all.sh
WARNING: Attempting to start all Apache Hadoop daemons as krish in 10 seconds.
WARNING: This is not a recommended production deployment configuration.
WARNING: Use CTRL-C to abort.
Starting namenodes on [localhost]
Starting datanodes
localhost: datanode is running as process 3245. Stop it first and ensure /tmp/hadoop-krish-datanode.pid file is empty before retry.
Starting secondary namenodes [krish-VirtualBox]
krish-VirtualBox: secondarynamenode is running as process 3427. Stop it first and ensure /tmp/hadoop-krish-secondarynamenode.pid file is empty before retry.
Starting resource manager
resource manager is running as process 3665. Stop it first and ensure /tmp/hadoop-krish-resource manager.pid file is empty before retry.
Starting nodemanagers
localhost: nodemanager is running as process 3778. Stop it first and ensure /tmp/hadoop-krish-nodemanager.pid file is empty before retry.
krish@krish-VirtualBox:~$ stop-all.sh
WARNING: Stopping all Apache Hadoop daemons as krish in 10 seconds.
WARNING: Use CTRL-C to abort.
Stopping namenodes on [localhost]
Stopping datanodes
Stopping secondary namenodes [krish-VirtualBox]
Stopping nodemanagers
Stopping resource manager
krish@krish-VirtualBox:~$ start-all.sh
WARNING: Attempting to start all Apache Hadoop daemons as krish in 10 seconds.
WARNING: This is not a recommended production deployment configuration.
WARNING: Use CTRL-C to abort.
Starting namenodes on [localhost]
Starting datanodes
Starting secondary namenodes [krish-VirtualBox]
Starting resource manager
Starting nodemanagers
krish@krish-VirtualBox:~$
```

localhost:9870

Activities Firefox Web Browser Sep 5 13:58

Install Hadoop on U x Problem loading pa x ChatGPT x All Applications x Namenode information x

localhost:9870/dfshealth.html#tab-overview

Hadoop Overview Datanodes Datanode Volume Failures Snapshot Startup Progress Utilities

## Overview 'localhost:9000' (✓active)

Started:	Thu Sep 05 13:57:31 +0530 2024
Version:	3.3.6, r1be78238728da9266a4f88195058f08fd012bf9c
Compiled:	Sun Jun 18 13:52:00 +0530 2023 by ubuntu from (HEAD detached at release-3.3.6-RC1)
Cluster ID:	CID-80ff24a7-0b44-44fc-be3c-8aa854eeae66
Block Pool ID:	BP-545190007-127.0.1.1-1725524699282

## Summary

Security is off.  
Safemode is off.  
1 files and directories, 0 blocks (0 replicated blocks, 0 erasure coded block groups) = 1 total filesystem object(s).  
Heap Memory used 110.19 MB of 237 MB Heap Memory. Max Heap Memory is 868 MB.  
Non Heap Memory used 48.81 MB of 50.77 MB Committed Non Heap Memory. Max Non Heap Memory is <unbounded>.


Configured Capacity:	23.95 GB
----------------------	----------

localhost:8088

Activities Firefox Web Browser Sep 5 13:58

Install Hadoop on U x Problem loading pa x ChatGPT x All Applications x Namenode information x

localhost:8088/cluster



Cluster

- About
- Nodes
- Node Labels
- Applications
  - NEW
  - NEW SAVING
  - SUBMITTED
  - ACCEPTED
  - RUNNING
  - FINISHED
  - FAILED
  - KILLED
- Scheduler

Tools

### Cluster Metrics

Apps Submitted	Apps Pending	Apps Running	Apps Completed	Containers Running
0	0	0	0	0

### Cluster Nodes Metrics

Active Nodes	Decommissioning Nodes	Decom
1	0	0

### Scheduler Metrics

Scheduler Type	Scheduling Resource Type	Min
Capacity Scheduler	[memory-mb (unit=Mi), vcores]	<memory:1024, vCore:

Show 20 entries

ID	User	Name	Application Type	Application Tags	Queue	Application Priority	StartTime	LaunchTime	Finis
----	------	------	------------------	------------------	-------	----------------------	-----------	------------	-------

Showing 0 to 0 of 0 entries