HADOOP INSTALLATION STEP-BY-STEP

```
amigandhi@ubuntu:~$ ssh localhost
ssh: connect to host localhost port 22: Connection refused
amigandhi@ubuntu:~$ sudo apt-get update
[sudo] password for amigandhi:
Hit:1 http://us.archive.ubuntu.com/ubuntu bionic InRelease
Get:2 http://security.ubuntu.com/ubuntu bionic-security InRelease [88.7 kB]
Get:3 http://us.archive.ubuntu.com/ubuntu bionic-updates InRelease [88.7 kB]
Get:4 http://us.archive.ubuntu.com/ubuntu bionic-backports InRelease [74.6 kB]
Get:5 http://security.ubuntu.com/ubuntu bionic-security/universe i386 Packages [617 kB]
Get:6 http://us.archive.ubuntu.com/ubuntu bionic-updates/universe amd64 Packages [1,054 kB]
Get:7 http://security.ubuntu.com/ubuntu bionic-security/universe amd64 Packages [649 kB]
Get:8 http://security.ubuntu.com/ubuntu bionic-security/universe amd64 DEP-11 Metadata [42.1 kB]
Get:9 http://security.ubuntu.com/ubuntu bionic-security/universe DEP-11 48x48 Icons [16.4 kB]
Get:10 http://security.ubuntu.com/ubuntu bionic-security/universe DEP-11 64x64 Icons [111 kB]
Get:11 http://security.ubuntu.com/ubuntu bionic-security/multiverse amd64 DEP-11 Metadata [2,464 B]
Get:12 http://security.ubuntu.com/ubuntu bionic-security/main i386 Packages [442 kB]
Get:13 http://security.ubuntu.com/ubuntu bionic-security/main amd64 Packages [651 kB]
Get:14 http://security.ubuntu.com/ubuntu bionic-security/main Translation-en [211 kB]
Get:15 http://security.ubuntu.com/ubuntu bionic-security/main amd64 DEP-11 Metadata [38.6 kB]
Get:16 http://security.ubuntu.com/ubuntu bionic-security/main DEP-11 48x48 Icons [17.6 kB]
Get:17 http://security.ubuntu.com/ubuntu bionic-security/main DEP-11 64x64 Icons [41.5 kB]
Get:18 http://us.archive.ubuntu.com/ubuntu bionic-updates/universe i386 Packages [1,010 kB]
Get:19 http://apt.postgresql.org/pub/repos/apt bionic-pgdg InRelease [46.3 kB]
Get:20 http://us.archive.ubuntu.com/ubuntu bionic-updates/universe amd64 DEP-11 Metadata [264 kB]
Get:21 http://us.archive.ubuntu.com/ubuntu bionic-updates/universe DEP-11 48x48 Icons [194 kB]
Get:22 http://us.archive.ubuntu.com/ubuntu bionic-updates/universe DEP-11 64x64 Icons [438 kB]
Get:23 http://us.archive.ubuntu.com/ubuntu bionic-updates/multiverse amd64 DEP-11 Metadata [2,468 B]
Get:24 http://us.archive.ubuntu.com/ubuntu bionic-updates/main amd64 Packages [871 kB]
Get:25 http://us.archive.ubuntu.com/ubuntu bionic-updates/main i386 Packages [650 kB]
Get:26 http://apt.postgresql.org/pub/repos/apt bionic-pgdg/main amd64 Packages [179 kB]
Get:27 http://us.archive.ubuntu.com/ubuntu bionic-updates/main Translation-en [303 kB]
Get:28 http://us.archive.ubuntu.com/ubuntu bionic-updates/main amd64 DEP-11 Metadata [307 kB]
```

```
amigandhi@ubuntu:~$ sudo apt-get install openssh-server
Reading package lists... Done
Building dependency tree
Reading state information... Done
The following additional packages will be installed:
ncurses-term openssh-sftp-server ssh-import-id
Suggested packages:
molly-guard monkeysphere rssh ssh-askpass
The following NEW packages will be installed:
ncurses-term openssh-server openssh-sftp-server ssh-import-id
D upgraded, 4 newly installed, O to remove and 145 not upgraded.
Need to get 637 kB of archives.
After this operation, 5,316 kB of additional disk space will be used.
Do you want to continue? [Y/n] y
Get:1 http://us.archive.ubuntu.com/ubuntu bionic-updates/main amd64 ncurses-term_all 6.1-1ubuntu1.18.04 [24
Get:2 http://security.ubuntu.com/ubuntu bionic-security/main amd64 openssh-sftp-server amd64 1:7.6p1-4ubunt
u0.3 [45.6 kB]
Get:3 http://security.ubuntu.com/ubuntu bionic-security/main amd64 openssh-server amd64 1:7.6p1-4ubuntu0.3
[333 kB]
.
Get:4 http://us.archive.ubuntu.com/ubuntu bionic-updates/main amd64 ssh-import-id all 5.7-0ubuntu1.1 [10.9
kB]
Fetched 637 kB in 0s (2,963 kB/s)
Preconfiguring packagès´...
Selecting previously unselected package ncurses-term.
(Reading database ... 174804 files and directories currently installed.)
Preparing to unpack .../ncurses-term_6.1-1ubuntu1.18.04_all.deb ...
Unpacking ncurses-term (6.1-1ubuntu1.18.04) .
```

```
amtgandhi@ubuntu:-/.ssh$ ssh localhost
The authenticity of host 'localhost (127.0.0.1)' can't be established.
ECDSA key fingerprint is SHA256:rSpBS4KI1HMBMMti0PftqFocky1MVBf1LAunqQgfL5w.
Are you sure you want to continue connecting (yes/no)? yes
Warning: Permanently added 'localhost' (ECDSA) to the list of known hosts.
amigandhi@localhost's password:
Permission dented, please try again.
amigandhi@localhost's password:
Permission dented, please try again.
amigandhi@localhost's password:
Welcome to Ubuntu 18.04.3 LTS (GNU/Linux 5.3.0-40-generic x86_64)

* Documentation: https://help.ubuntu.com
* Management: https://landscape.canonical.com
* Support: https://lubuntu.com/advantage

* Multipass 1.0 is out! Get Ubuntu VMs on demand on your Linux, Windows or
Mac. Supports cloud-init for fast, local, cloud devops simulation.

https://multipass.run/

* Latest Kubernetes 1.18 beta is now available for your laptop, NUC, cloud
instance or Raspberry Pi, with automatic updates to the final GA release.

sudo snap install microk8s --channel=1.18/beta --classic

* Canonical Livepatch is available for installation.
- Reduce system reboots and improve kernel security. Activate at:
```

NOTE: Download latest version of Hadoop(3.2.1) from below link –

https://archive.apache.org/dist/hadoop/common/

```
amigandhi@ubuntu:~$ cat ~/.ssh/id_rsa.pub >> ~/.ssh/authorized_keys
amigandhi@ubuntu:~$
 migandhi@ubuntu:~$
amigandhi@ubuntu:~$
amigandhi@ubuntu:~$ cd Downloads/
amigandhi@ubuntu:~/Downloads$ ls
dev hadoop-3.2.1
amigandhi@ubuntu:~/Downloads$ sudo mv hadoop-3.2.1 /usr/local/bin/
[sudo] password for amigandhi:
amigandhi@ubuntu:~/Downloads$ cd /usr/local/bin
amigandhi@ubuntu:/usr/local/bin$ ls
aws aws_completer hadoop-3.2.1
amigandhi@ubuntu:/usr/local/bin$ sudo mkdir hadoop-tmp
amigandhi@ubuntu:/usr/local/bin$ ls
aws aws_completer hadoop-3.2.1 hadoop-tmp
 migandhi@ubuntu:/usr/local/bin$ cd hadoop-
hadoop-3.2.1/ hadoop-tmp/
amigandhi@ubuntu:/usr/local/bin$ cd hadoop-3.2.1/
amigandhi@ubuntu:/usr/local/bin/hadoop-3.2.1$ ls
               libexec NOTICE.txt sbin
LICENSE.txt README.txt share
bin include libexec
etc lib
amigandhi@ubuntu:/usr/local/bin/hadoop-3.2.1$
amigandhi@ubuntu:/usr/local/bin/hadoop-3.2.1$
amigandhi@ubuntu:/usr/local/bin/hadoop-3.2.1$ cd etc/hadoop/
amigandhi@ubuntu:/usr/local/bin/hadoop-3.2.1/etc/hadoop$ ls
capacity-scheduler.xml
                                      kms-log4j.properties
configuration.xsl
                                      kms-site.xml
container-executor.cfg
                                      log4j.properties
core-site.xml
                                      mapred-env.cmd
hadoop-env.cmd
                                      mapred-env.sh
hadoop-env.sh
                                      mapred-queues.xml.template
```

NOTE: Look at the link below for supported Java version for your Hadoop (Hadoop 3.2.1 is compatible with java 7 and 8 only!)

https://cwiki.apache.org/confluence/display/HADOOP/Hadoop+Java+Versions

```
[sudo] password for amigandhi: Reading package lists... Done
Building dependency tree
Reading state information... Done
The following additional packages will be installed:
   fonts-dejavu-extra libatk-wrapper-java libatk-wrapper-java-jni libgif7
libice-dev libpthread-stubs0-dev libsm-dev libx11-dev libx11-doc libxau-dev
libxcb1-dev libxdmcp-dev libxt-dev openjdk-8-jdk-headless
   openjdk-8-jre-headless x11proto-core-dev x11proto-dev xorg-sgml-doctools
   xtrans-dev
Suggested packages:
   libice-doc libsm-doc libxcb-doc libxt-doc openjdk-8-demo openjdk-8-source
   visualvm icedtea-8-plugin fonts-ipafont-gothic fonts-ipafont-mincho fonts-wqy-microhei fonts-wqy-zenhei
The following NEW packages will be installed:
   fonts-dejavu-extra libatk-wrapper-java libatk-wrapper-java-jni libgif7
libice-dev libpthread-stubs0-dev libsm-dev libx11-dev libx11-doc libxau-dev
   libxcb1-dev libxdmcp-dev libxt-dev openjdk-8-jdk openjdk-8-jdk-headless
   openjdk-8-jre openjdk-8-jre-headless x11proto-core-dev x11proto-dev xorg-sgml-doctools xtrans-dev
0 upgraded, 21 newly installed, 0 to remove and 145 not upgraded.
Need to get 43.1 MB of archives.
After this operation, 167 MB of additional disk space will be used.
Do you want to continue? [Y/n] y
Get:1 http://us.archive.ubuntu.com/ubuntu bionic/main amd64 fonts-dejavu-extra all 2.37-1 [1,953 kB]
Get:2 http://security.ubuntu.com/ubuntu bionic-security/main amd64 libgif7 amd64 5.1.4-2ubuntu0.1 [30.9 k
В]
Get:3 http://security.ubuntu.com/ubuntu bionic-security/universe amd64 openjdk-8-jre-headless amd64 8u242
```

NOTE: If you already have java installed with another version, please follow following commands –

Commands:

- 1. sudo update-alternatives --config java
 - Write the number of selection for java 8
- 2. java -version

```
amigandhi@ubuntu:~$ sudo update-alternatives --config java
There are 2 choices for the alternative java (providing /usr/bin/java).
  Selection
               Path
                                                               Priority
                                                                          Status
  0
               /usr/lib/jvm/java-11-openjdk-amd64/bin/java
                                                                1111
                                                                          auto mode
               /usr/lib/jvm/java-11-openjdk-amd64/bin/java
                                                                1111
                                                                          manual mode
  1
               /usr/lib/jvm/java-8-openjdk-amd64/jre/bin/java
                                                                          manual mode
* 2
                                                                1081
Press <enter> to keep the current choice[*], or type selection number:
amigandhi@ubuntu:~$
amigandhi@ubuntu:~$ vi ~/.bash_profile
amigandhi@ubuntu:~$ nano ~/.bash_profile
amigandhi@ubuntu:~$ source ~/.bash_profile
amigandhi@ubuntu:~$ echo $JAVA_HOME
/usr/lib/jvm/java-8-openjdk-amd64/
amigandhi@ubuntu:~$
amigandhi@ubuntu:~$ echo $PATH
/usr/local/sbin:/usr/local/bin:/usr/sbin:/usr/bin:/bin:/usr/games:/usr/local/games:/
snap/bin
amigandhi@ubuntu:~$ which java
/usr/bin/java
amigandhi@ubuntu:~$
```

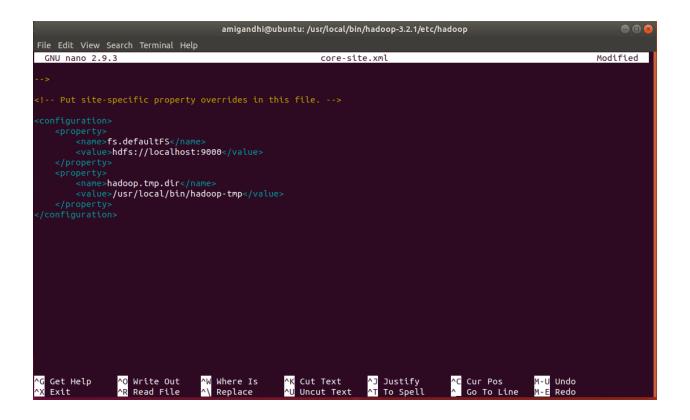
NOTE: Refer to Hadoop docs link below to set up your Hadoop environment to run on single node in pseudo-distributed environment. You can also refer to screenshots below to set up configurations for your hadoop environment.

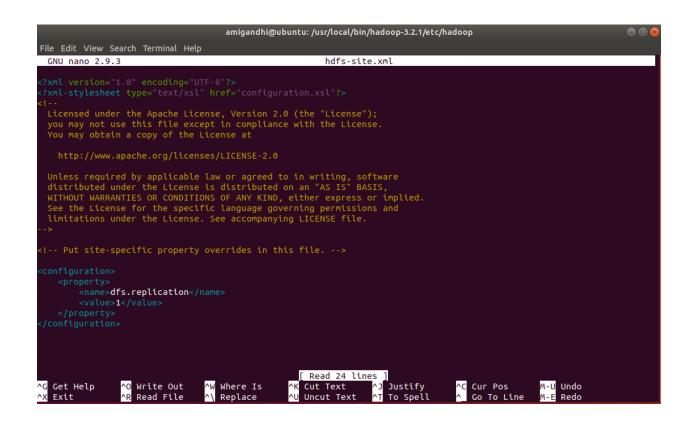
http://hadoop.apache.org/docs/r2.7.3/hadoop-project-dist/hadoop-common/SingleCluster.html

```
amigandhi@ubuntu:/usr/local/bin/hadoop-3.2.1/etc/hadoop$ sudo nano hadoop-env.sh
amigandhi@ubuntu:/usr/local/bin/hadoop-3.2.1/etc/hadoop$
```

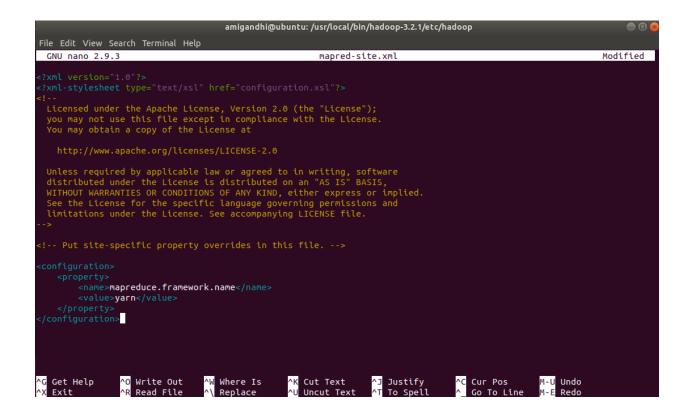
```
GNU nano 2.9.3
                                                                                       hadoop-env.sh
                                                                                                                                                                           Modified
## THIS FILE ACTS AS THE MASTER FILE FOR ALL HADOOP PROJECTS.
## SETTINGS HERE WILL BE READ BY ALL HADOOP COMMANDS. THEREFORE,
## ONE CAN USE THIS FILE TO SET YARN, HDFS, AND MAPREDUCE
 JAVA_HOME=/usr/lib/jvm/java-8-openjdk-amd64/
                                                   ^W Where Is
^\ Replace
^G Get Help
                         ^O Write Out
^R Read File
                                                                            ^K Cut Text
^U Uncut Tex
                                                                                                          Justify
                                                                                                                               ^C Cur Pos
                                                                                                                                                        M-U Undo
^X Exit
                             Read File
                                                                                Uncut Text
                                                                                                          To Linter
                                                                                                                                   Go To Line
```

amigandhi@ubuntu:/usr/local/bin/hadoop-3.2.1/etc/hadoop\$ sudo nano core-site.xml

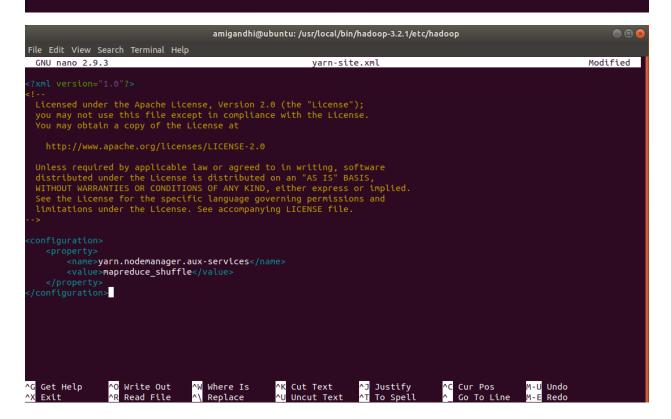




#amigandhi@ubuntu:/usr/local/bin/hadoop-3.2.1/etc/hadoop\$ sudo nano mapred_site.xml



amigandhi@ubuntu:/usr/local/bin/hadoop-3.2.1/etc/hadoop\$ sudo nano yarn-site.xml



NOTE: Configuration is complete for Hadoop setup. Always format the namenode ONLY when you install Hadoop. Then, start your Hadoop. Please refer screenshots below –

```
amigandhi@ubuntu:/usr/local/bin/hadoop-3.2.1/etc/hadoop$ cd ../..
amigandhi@ubuntu:/usr/local/bin/hadoop-3.2.1$ cd bin
amigandhi@ubuntu:/usr/local/bin/hadoop-3.2.1/bin$ ls -la
total 1040
drwxr-xr-x 2 amigandhi amigandhi
                                           4096 Sep 10 12:51
drwxr-xr-x 9 amigandhi amigandhi
                                          4096 Sep 10 12:51 ...
-rwxr-xr-x 1 amigandhi amigandhi 442480 Sep 10 12:33 container-executor
-rwxr-xr-x 1 amigandhi amigandhi 8707 Sep 10 11:57 hadoop
-rwxr-xr-x 1 amigandhi amigandhi
                                         11265 Sep 10 11:57 hadoop.cmd
-rwxr-xr-x 1 amigandhi amigandhi 11274 Sep 10 12:01 hdfs
                                          8081 Sep 10 12:01 hdfs.cmd
6237 Sep 10 12:36 mapred
-rwxr-xr-x 1 amigandhi amigandhi
-rwxr-xr-x 1 amigandhi amigandhi
-rwxr-xr-x 1 amigandhi amigandhi
                                           6311 Sep 10 12:36 mapred.cmd
-rwxr-xr-x 1 amigandhi amigandhi 29184 Sep 10 12:33 <mark>oom-listene</mark>r
-rwxr-xr-x 1 amigandhi amigandhi 484264 Sep 10 12:33 test-container-executor
-rwxr-xr-x 1 amigandhi amigandhi 12112 Sep 10 12:33 yarn
-rwxr-xr-x 1 amigandhi amigandhi 12840 Sep 10 12:33 yarn.cmd
amigandhi@ubuntu:/usr/local/bin/hadoop-3.2.1/bin$ ./hadoop namenode -format
```

```
annganuniwubuncu: /usi/tocat/bin/nauoop-3.2. i/bin
2020-02-29 21:23:52,821 INFO namenode.NameNode: Caching file names occurring more than 10 times
2020-02-29 21:23:52,825 INFO snapshot.SnapshotManager: Loaded config captureOpenFiles: false, skipCaptureAccessTimeOnl
yChange: false, snapshotDiffAllowSnapRootDescendant: true, maxSnapshotLimit: 65536
ychange: Talse, Shapshotottrattowshapkootoescendant: true, MaxshapshottLmit: 55356
2020-02-29 21:23:52,826 INFO snapshot.SnapshotManager: SkipList is disabled
2020-02-29 21:23:52,829 INFO util.GSet: Computing capacity for map cachedBlocks
2020-02-29 21:23:52,829 INFO util.GSet: VM type = 64-bit
2020-02-29 21:23:52,829 INFO util.GSet: 0.25% max memory 1.7 GB = 4.4 MB
2020-02-29 21:23:52,829 INFO util.GSet: capacity = 2^19 = 524288 entries
2020-02-29 21:23:52,836 INFO metrics.TopMetrics: NNTop conf: dfs.namenode.top.window.num.buckets = 10
2020-02-29 21:23:52,836 INFO metrics TopMetrics: NNTop conf: dfs.namenode.top.gum_users = 10
2020-02-29 21:23:52,836 INFO metrics.TopMetrics: NNTop conf: dfs.namenode.top.num.users = 10
2020-02-29 21:23:52,836 INFO metrics.TopMetrics: NNTop conf: dfs.namenode.top.windows.minutes = 1,5,25
2020-02-29 21:23:52,839 INFO namenode.FSNamesystem: Retry cache on namenode is enabled
2020-02-29 21:23:52,839 INFO namenode.FSNamesystem: Retry cache will use 0.03 of total heap and retry cache entry expi
ry time is 600000 míllis
2020-02-29 21:23:52,841 INFO util.GSet: Computing capacity for map NameNodeRetryCache
2020-02-29 21:23:52,841 INFO util.GSet: VM type == 64-bit
2020-02-29 21:23:52,841 INFO util.GSet: 0.02999999329447746% max memory 1.7 GB = 541.9 KB
2020-02-29 21:23:52,841 INFO util.GSet: capacity = 2^16 = 65536 entries
2020-02-29 21:23:52,860 INFO namenode.FSImage: Allocated new BlockPoolId: BP-947238340-127.0.1.1-1583029432855
2020-02-29 21:23:52,873 INFO common.Storage: Storage directory /usr/local/bin/hadoop-tmp/dfs/name has been successfull
v formatted.
.
2020-02-29 21:23:52,895 INFO namenode.FSImageFormatProtobuf: Saving image file /usr/local/bin/hadoop-tmp/dfs/name/curr
ent/fsimage.ckpt_00000000000000000000 using no compression
2020-02-29 21:23:52,963 INFO namenode.FSImageFormatProtobuf: Image file /usr/local/bin/hadoop-tmp/dfs/name/current/fsi
2020-02-29 21:23:52,973 INFO namenode.NNStorageRetentionManager: Going to retain 1 images with txid >= 0 2020-02-29 21:23:52,979 INFO namenode.FSImageS FSImageSaver clean checkpoint: txid=0 when meet shutdown.
2020-02-29 21:23:52,979 INFO namenode.NameNode: SHUTDOWN_MSG:
SHUTDOWN_MSG: Shutting down NameNode at ubuntu/127.0.1.1
 amigandhi@ubuntu:/usr/local/bin/hadoop-3.2.1/bin$
amigandhi@ubuntu:/usr/local/bin/hadoop-3.2.1/bin$
```

```
amigandhi@ubuntu: /usr/local/bin/hadoop-3.2.1/sbin

File Edit View Search Terminal Help

amigandhi@ubuntu:/usr/local/bin/hadoop-3.2.1/bin$

amigandhi@ubuntu:/usr/local/bin/hadoop-3.2.1/bin$

amigandhi@ubuntu:/usr/local/bin/hadoop-3.2.1/bin$

amigandhi@ubuntu:/usr/local/bin/hadoop-3.2.1/bin$ cd ..

amigandhi@ubuntu:/usr/local/bin/hadoop-3.2.1/scd sbin

amigandhi@ubuntu:/usr/local/bin/hadoop-3.2.1.5 cd sbin

amigandhi@ubuntu:/usr/local/bin/hadoop-3.2.1.5 shin$ ./start-all.sh

WARNING: Attempting to start all Apache Hadoop daemons as amigandhi 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 [ubuntu]

ubuntu: Warning: Permanently added 'ubuntu' (ECDSA) to the list of known hosts.

Starting resourcemanager

Starting nodemanagers
```

NOTE: To check if all your nodes are up, use command: 'jps'. Once you are down, gracefully shutdown Hadoop.

```
amigandhi@ubuntu:/usr/local/bin/hadoop-3.2.1/sbin$ jps

11328 Jps

10788 ResourceManager

10309 DataNode

10968 NodeManager

101106 NameNode

10555 SecondaryNameNode

amigandhi@ubuntu:/usr/local/bin/hadoop-3.2.1/sbin$ ./stop-all.sh

WARNING: Stopping all Apache Hadoop daemons as amigandhi in 10 seconds.

WARNING: Use CTRL-C to abort.

Stopping namenodes on [localhost]

Stopping datanodes

Stopping secondary namenodes [ubuntu]

Stopping resourcemanager

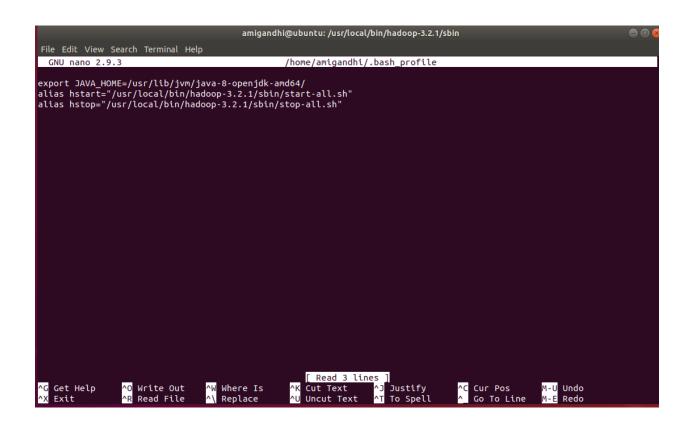
amigandhi@ubuntu:/usr/local/bin/hadoop-3.2.1/sbin$ jps

12312 Jps

amigandhi@ubuntu:/usr/local/bin/hadoop-3.2.1/sbin$
```

NOTE: Every time to start and stop Hadoop, you need to be in /usr/local/bin/hadoop-3.2.1/sbin path only. Hence, its always a good practice to create alias in such conditions to start/stop hadoop from any location.

```
amigandhi@ubuntu:/usr/local/bin/hadoop-3.2.1/sbin$
amigandhi@ubuntu:/usr/local/bin/hadoop-3.2.1/sbin$
amigandhi@ubuntu:/usr/local/bin/hadoop-3.2.1/sbin$
nano ~/.bash_profile
```



```
amigandhi@ubuntu:/usr/local/bin/hadoop-3.2.1/sbin$
amigandhi@ubuntu:/usr/local/bin/hadoop-3.2.1/sbin$ source ~/.bash_profile
amigandhi@ubuntu:/usr/local/bin/hadoop-3.2.1/sbin$ hstart
WARNING: Attempting to start all Apache Hadoop daemons as amigandhi in 10 seconds.
WARNING: This is not a recommended production deployment configuration.
WARNING: Use CTRL-C to abort.
Starting namenodes on [localhost]
Starting secondary namenodes [ubuntu]
Starting resourcemanager
Starting resourcemanagers
amigandhi@ubuntu:/usr/local/bin/hadoop-3.2.1/sbin$ hstop
WARNING: Stopping all Apache Hadoop daemons as amigandhi in 10 seconds.
WARNING: Use CTRL-C to abort.
Stopping namenodes on [localhost]
Stopping datanodes
Stopping secondary namenodes [ubuntu]
Stopping nodemanagers
Stopping resourcemanager
stopping resourcemanager
stopping resourcemanager
stopping resourcemanager
stopping resourcemanager
stopping resourcemanager
amigandhi@ubuntu:/usr/local/bin/hadoop-3.2.1/sbin$ jps
15116 Jps
```

THANK YOU!!