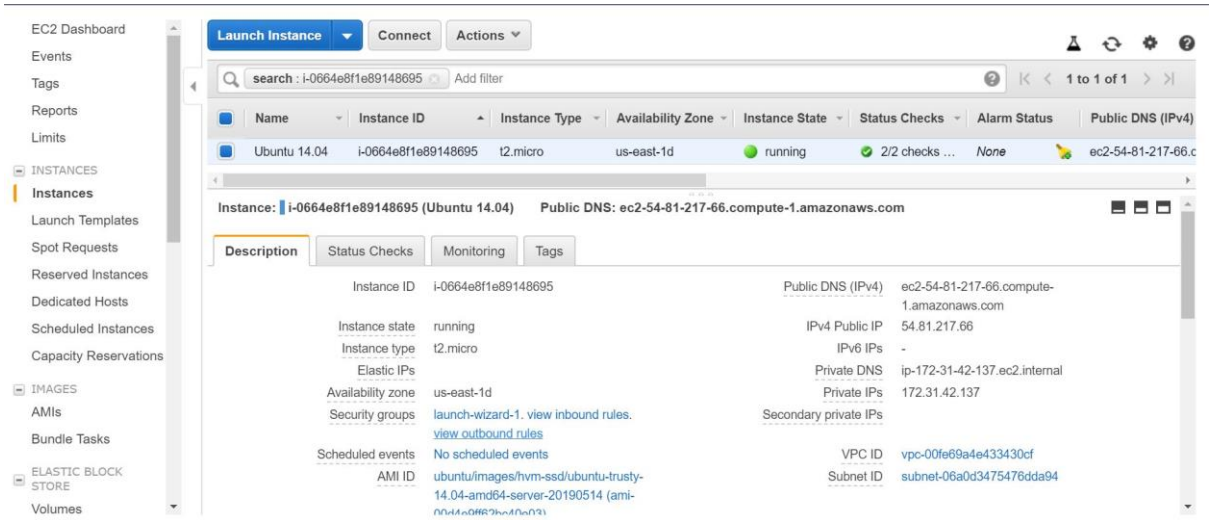
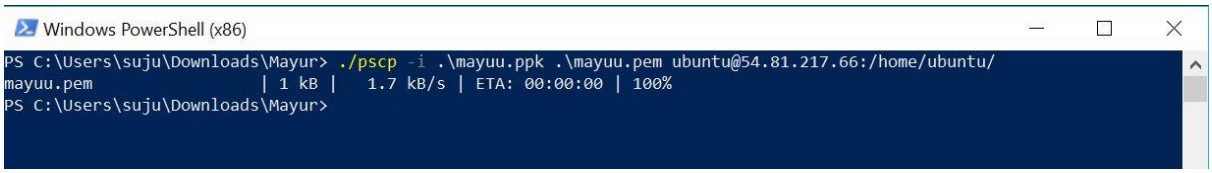
Single Node Installation

1. Take an Ubuntu 14.04 EC2 instance.

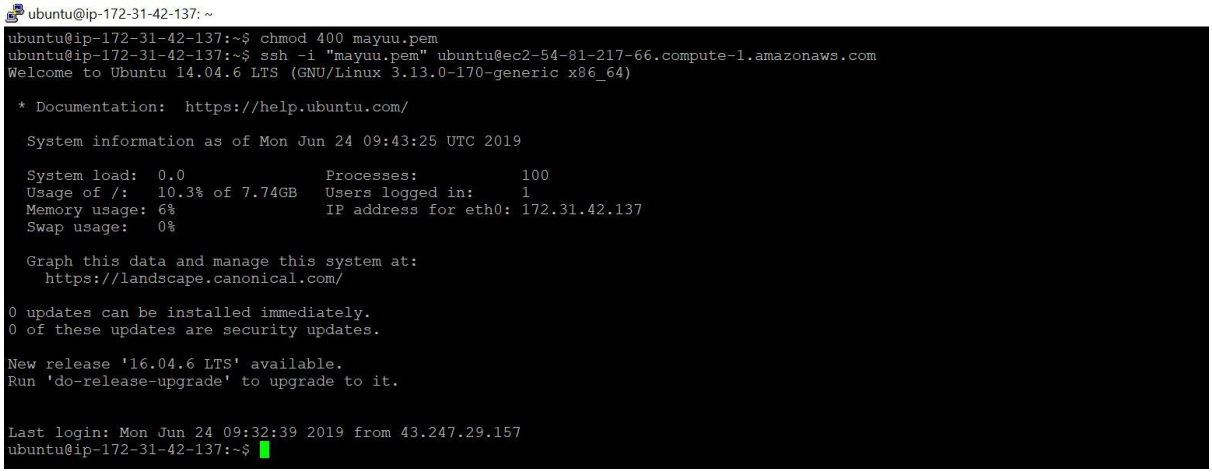


2. Connect to this Ubuntu machine using public IP
    - In Mac/ Ubuntu using terminal
    - In Windows using Putty
3. While using Putty we have to import our key to the server using pscp in Powershell/ WinScp.



4. Chmod 400 mayuu.pem
5. Connect to the machine

    ssh -i "mayuu.pem" ubuntu@ec2-54-81-217-66.compute-1.amazonaws.com



Now we are done connecting, now we have to install a cluster on single node.

6. sudo apt-get update

```
Hit http://us-east-1.ec2.archive.ubuntu.com trusty/main amd64 Packages
Hit http://us-east-1.ec2.archive.ubuntu.com trusty/restricted amd64 Packages
Hit http://us-east-1.ec2.archive.ubuntu.com trusty/universe amd64 Packages
Hit http://us-east-1.ec2.archive.ubuntu.com trusty/multiverse amd64 Packages
Hit http://us-east-1.ec2.archive.ubuntu.com trusty/main Translation-en
Hit http://us-east-1.ec2.archive.ubuntu.com trusty/multiverse Translation-en
Hit http://us-east-1.ec2.archive.ubuntu.com trusty/restricted Translation-en
Hit http://us-east-1.ec2.archive.ubuntu.com trusty/universe Translation-en
Ign http://us-east-1.ec2.archive.ubuntu.com trusty/main Translation-en_US
Ign http://us-east-1.ec2.archive.ubuntu.com trusty/multiverse Translation-en_US
Ign http://us-east-1.ec2.archive.ubuntu.com trusty/restricted Translation-en_US
Ign http://us-east-1.ec2.archive.ubuntu.com trusty/universe Translation-en_US
Get:24 http://security.ubuntu.com trusty-security/universe Sources [102 kB]
Get:25 http://security.ubuntu.com trusty-security/main amd64 Packages [835 kB]
Get:26 http://security.ubuntu.com trusty-security/universe amd64 Packages [294 kB]
Get:27 http://security.ubuntu.com trusty-security/main Translation-en [448 kB]
Get:28 http://security.ubuntu.com trusty-security/universe Translation-en [162 kB]
Fetched 13.1 MB in 5s (2,578 kB/s)
Reading package lists... Done
ubuntu@ip-172-31-42-137:~$
```

7. Downloading and installing java (Hadoop requires a working Java 1.6+)

sudo apt-get install openjdk-7-jdk -y

8. Downloading Hadoop

   wget https://archive.apache.org/dist/hadoop/common/hadoop-1.2.1/hadoop-1.2.1.tar.gz

```
ubuntu@ip-172-31-42-137:~$ wget https://archive.apache.org/dist/hadoop/common/hadoop-1.2.1/hadoop-1.2.1.tar.gz
--2019-06-24 10:07:12--  https://archive.apache.org/dist/hadoop/common/hadoop-1.2.1/hadoop-1.2.1.tar.gz
Resolving archive.apache.org (archive.apache.org)... 163.172.17.199
Connecting to archive.apache.org (archive.apache.org)|163.172.17.199|:443... connected.
HTTP request sent, awaiting response... 200 OK
Length: 63851630 (61M) [application/x-gzip]
Saving to: 'hadoop-1.2.1.tar.gz'

100%[=======================================================================================================>] 63,851,630  11.1MB/s   in 5.9s

2019-06-24 10:07:18 (10.4 MB/s) - 'hadoop-1.2.1.tar.gz' saved [63851630/63851630]
```

```
ubuntu@ip-172-31-42-137:~$ nano ~/.bashrc
ubuntu@ip-172-31-42-137:~$ nano /usr/local/hadoop/conf/hadoop-env.sh
ubuntu@ip-172-31-42-137:~$ nano /usr/local/hadoop/conf/hadoop-env.sh
ubuntu@ip-172-31-42-137:~$
ubuntu@ip-172-31-42-137:~$ cd hadoop-1.2.1/
ubuntu@ip-172-31-42-137:~/hadoop-1.2.1$ ls
bin         CHANGES.txt  docs                 hadoop-core-1.2.1.jar         hadoop-test-1.2.1.jar   ivy.xml  LICENSE.txt  sbin   webapps
build.xml   conf         hadoop-ant-1.2.1.jar hadoop-examples-1.2.1.jar     hadoop-tools-1.2.1.jar  lib      NOTICE.txt   share
c++         contrib      hadoop-client-1.2.1.jar hadoop-minicluster-1.2.1.jar ivy                     libexec  README.txt   src
ubuntu@ip-172-31-42-137:~/hadoop-1.2.1$
```

   tar -xzvf hadoop-1.2.1.tar.gz

   sudo mv hadoop-1.2.1 /usr/local/Hadoop

9. Configuring bashrc linux env setup
   nano ~/.bashrc

```
# 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 HADOOP_PREFIX=/usr/local/hadoop/
export PATH=$PATH:$HADOOP_PREFIX/bin
export JAVA_HOME=/usr/lib/jvm/java-7-openjdk-amd64
export PATH=$PATH:$JAVA_HOME
```

10. Setting hadoop env

nano /usr/local/hadoop/conf/hadoop-env.sh

```
# A string representing this instance of hadoop. $USER by default.
# export HADOOP_IDENT_STRING=$USER

# The scheduling priority for daemon processes.  See 'man nice'.
# export HADOOP_NICENESS=10
export JAVA_HOME=/usr/lib/jvm/java-7-openjdk-amd64
export HADOOP_OPTS=-Djava.net.preferIPV4Stack=true
```

11. Configuring xml's

- Configuring core-site.xml

  nano /usr/local/hadoop/conf/core-site.xml

```
<?xml version="1.0"?>
<?xml-stylesheet type="text/xsl" href="configuration.xsl"?>

<!-- Put site-specific property overrides in this file. -->

<configuration>
<property>
<name>fs.default.name</name>
<value>hdfs://localhost:9000</value>
</property>

<property>
<name>hadoop.tmp.dir</name>
<value>/usr/local/hadoop/tmp</value>
</property>
</configuration>
```

- Configuring hdfs-site.xml

nano /usr/local/hadoop/conf/hdfs-site.xml

```
<?xml version="1.0"?>
<?xml-stylesheet type="text/xsl" href="configuration.xsl"?>

<!-- Put site-specific property overrides in this file. -->

<configuration>
<property>
<name>dfs.replication</name>
<value>1</value>
</property>
</configuration>
```

- Configuring mapred-site.xml

nano /usr/local/hadoop/conf/mapred-site.xml]

```
<?xml version="1.0"?>
<?xml-stylesheet type="text/xsl" href="configuration.xsl"?>

<!-- Put site-specific property overrides in this file. -->

<configuration>
<property>
<name>mapred.job.tracker</name>
<value>hdfs://localhost:9001</value>
</property>
</configuration>
```

12. Making tmp dir
    mkdir /usr/local/hadoop/tmp

13. **Exec bash**
    exec bash
14. We have to create rsa-key pair to communicate to localhost
    ssh-keygen

```
ubuntu@ip-172-31-42-137:~$ ssh-keygen
Generating public/private rsa key pair.
Enter file in which to save the key (/home/ubuntu/.ssh/id_rsa):
Enter passphrase (empty for no passphrase):
Enter same passphrase again:
Your identification has been saved in /home/ubuntu/.ssh/id_rsa.
Your public key has been saved in /home/ubuntu/.ssh/id_rsa.pub.
The key fingerprint is:
af:9e:b9:55:8f:dc:c8:e8:13:9f:db:b3:96:25:aa:3f ubuntu@ip-172-31-42-137
The key's randomart image is:
+--[ RSA 2048]----+
|                 |
|                 |
|                 |
|                 |
|        S    .   |
|         ..= =. .|
|          +o=oo+ |
|         *. E.+  |
|        .*.o+o+oo |
+-----------------+
ubuntu@ip-172-31-42-137:~$
```

>> Authorisation

cat ~/.ssh/id_rsa.pub >> ~/.ssh/authorized_keys

15. Configuring DNS address for local

```
ubuntu@ip-172-31-42-137:~$ cat /etc/hosts
127.0.0.1 localhost

# The following lines are desirable for IPv6 capable hosts
::1 ip6-localhost ip6-loopback
fe00::0 ip6-localnet
ff00::0 ip6-mcastprefix
ff02::1 ip6-allnodes
ff02::2 ip6-allrouters
ff02::3 ip6-allhosts
ubuntu@ip-172-31-42-137:~$
```

16. ssh localhost

connecting to the localhost

17. Formatting hadoop namenode

hadoop namenode -format

```
19/06/24 10:58:15 INFO namenode.FSNamesystem: dfs.block.invalidate.limit=100
19/06/24 10:58:15 INFO namenode.FSNamesystem: isAccessTokenEnabled=false accessKeyUpdateInterval=0 min(s), accessTokenLifetime=0 min(s)
19/06/24 10:58:15 INFO namenode.FSEditLog: dfs.namenode.edits.toleration.length = 0
19/06/24 10:58:15 INFO namenode.NameNode: Caching file names occuring more than 10 times
19/06/24 10:58:15 INFO common.Storage: Image file /usr/local/hadoop/tmp/dfs/name/current/fsimage of size 112 bytes saved in 0 seconds.
19/06/24 10:58:15 INFO namenode.FSEditLog: closing edit log: position=4, editlog=/usr/local/hadoop/tmp/dfs/name/current/edits
19/06/24 10:58:15 INFO namenode.FSEditLog: close success: truncate to 4, editlog=/usr/local/hadoop/tmp/dfs/name/current/edits
19/06/24 10:58:15 INFO common.Storage: Storage directory /usr/local/hadoop/tmp/dfs/name has been successfully formatted.
19/06/24 10:58:15 INFO namenode.NameNode: SHUTDOWN_MSG:
/************************************************************
SHUTDOWN_MSG: Shutting down NameNode at ip-172-31-42-137/172.31.42.137
************************************************************/
ubuntu@ip-172-31-42-137:~$
```

18. Starting physical daemons

start-dfs.sh

```
ubuntu@ip-172-31-42-137:~$ start-dfs.sh
starting namenode, logging to /usr/local/hadoop/libexec/../logs/hadoop-ubuntu-namenode-ip-172-31-42-137.out
localhost: starting datanode, logging to /usr/local/hadoop/libexec/../logs/hadoop-ubuntu-datanode-ip-172-31-42-137.out
localhost: starting secondarynamenode, logging to /usr/local/hadoop/libexec/../logs/hadoop-ubuntu-secondarynamenode-ip-172-31-42-137.out
ubuntu@ip-172-31-42-137:~$
```

19. Starting logical daemons

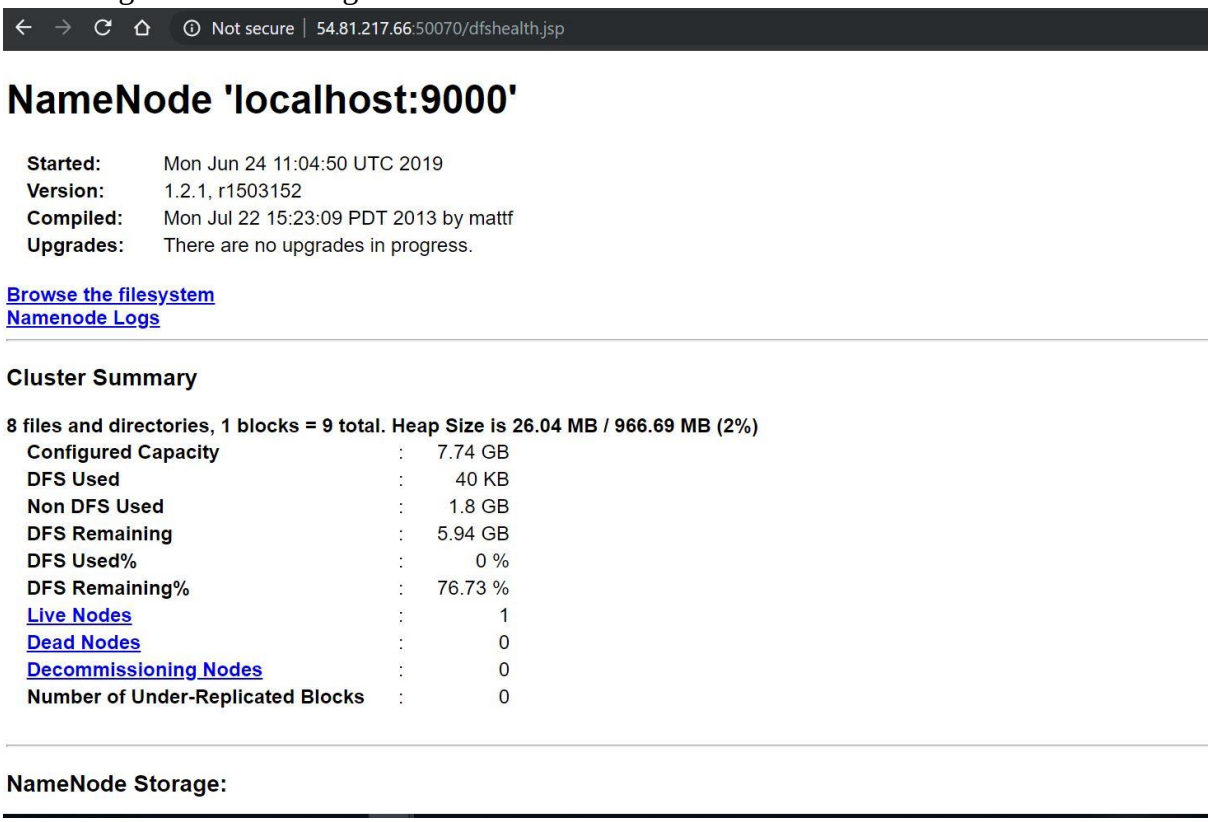
Start-mapred.sh

```
ubuntu@ip-172-31-42-137:~$ start-mapred.sh
starting jobtracker, logging to /usr/local/hadoop/libexec/../logs/hadoop-ubuntu-jobtracker-ip-172-31-42-137.out
localhost: starting tasktracker, logging to /usr/local/hadoop/libexec/../logs/hadoop-ubuntu-tasktracker-ip-172-31-42-137.out
ubuntu@ip-172-31-42-137:~$
```

20. jps

```
ubuntu@ip-172-31-42-137:~$ jps
8863 TaskTracker
8960 Jps
8592 SecondaryNameNode
8407 DataNode
8233 NameNode
8692 JobTracker
ubuntu@ip-172-31-42-137:~$
```

21. Accessing Namenode using Web UI



22. Now creating a file & uploading data on the cluster.

```
ubuntu@ip-172-31-42-137:/usr/local/hadoop/tmp$ ls
dfs  mapred
ubuntu@ip-172-31-42-137:/usr/local/hadoop/tmp$ cd dfs/
ubuntu@ip-172-31-42-137:/usr/local/hadoop/tmp/dfs$ ls
data  name  namesecondary
ubuntu@ip-172-31-42-137:/usr/local/hadoop/tmp/dfs$ cd data/
ubuntu@ip-172-31-42-137:/usr/local/hadoop/tmp/dfs/data$ ls
blocksBeingWritten  current  detach  in_use.lock  storage  tmp
ubuntu@ip-172-31-42-137:/usr/local/hadoop/tmp/dfs/data$ cd current/
ubuntu@ip-172-31-42-137:/usr/local/hadoop/tmp/dfs/data/current$ ls
blk_3594589285843858810  blk_3594589285843858810_1001.meta  dncp_block_verification.log.curr  VERSION
ubuntu@ip-172-31-42-137:/usr/local/hadoop/tmp/dfs/data/current$ nano cloudage
ubuntu@ip-172-31-42-137:/usr/local/hadoop/tmp/dfs/data/current$ ls
blk_3594589285843858810  blk_3594589285843858810_1001.meta  cloudage  dncp_block_verification.log.curr  VERSION
ubuntu@ip-172-31-42-137:/usr/local/hadoop/tmp/dfs/data/current$ hadoop fs -put cloudage .
ubuntu@ip-172-31-42-137:/usr/local/hadoop/tmp/dfs/data/current$
```

hadoop fs -put cloudage .

Contents of directory /user/ubuntu

Goto : /user/ubuntu go

Go to parent directory

Name	Type	Size	Replication	Block Size	Modification Time	Permission	Owner	Group
cloudage	file	1.67 KB	1	64 MB	2019-06-24 11:28	rw-r--r--	ubuntu	supergroup

23. Benchmarking
    Running a pi Job

```
Job Finished in 40.85 seconds
Estimated value of Pi is 3.14080000000000000000
ubuntu@ip-172-31-42-137:~$
```