

# Single Node Installation

1. Take an Ubuntu 14.04 EC2 instance.

The screenshot shows the AWS EC2 Instances page. On the left, there's a sidebar with navigation links like EC2 Dashboard, Events, Tags, Reports, Limits, Instances (selected), Launch Templates, Spot Requests, Reserved Instances, Dedicated Hosts, Scheduled Instances, Capacity Reservations, Images (AMIs selected), Bundle Tasks, and Elastic Block Store (Volumes). The main pane displays a table of instances. One instance is selected: "Ubuntu 14.04" with Instance ID "i-0664e8f1e89148695". The details shown include:

Attribute	Value
Instance ID	i-0664e8f1e89148695
Instance state	running
Instance type	t2.micro
Availability zone	us-east-1d
Security groups	launch-wizard-1, view inbound rules, view outbound rules
Scheduled events	No scheduled events
AMI ID	ubuntu/images/hvm-ssd/ubuntu-trusty-14.04-amd64-server-20190514 (ami-001a00f69a4e433430cf)
Public DNS (IPv4)	ec2-54-81-217-66.compute-1.amazonaws.com
IPv4 Public IP	54.81.217.66
IPv6 IPs	-
Private DNS	ip-172-31-42-137.ec2.internal
Private IPs	172.31.42.137
Secondary private IPs	-
VPC ID	vpc-00fe69a4e433430cf
Subnet ID	subnet-06a0d3475476dda94

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.

```
PS C:\Users\suju\Downloads\Mayur> ./pscp -i .\mayuu.ppk .\mayuu.pem ubuntu@54.81.217.66:/home/ubuntu/mayuu.pem
mayuu.pem | 1 kB | 1.7 kB/s | ETA: 00:00:00 | 100%
```

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

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

```
ubuntu@ip-172-31-42-137:~$ chmod 400 mayuu.pem
ubuntu@ip-172-31-42-137:~$ ssh -i "mayuu.pem" ubuntu@ec2-54-81-217-66.compute-1.amazonaws.com
Welcome to Ubuntu 14.04.6 LTS (GNU/Linux 3.13.0-170-generic x86_64)

 * Documentation: https://help.ubuntu.com/
 System information as of Mon Jun 24 09:43:25 UTC 2019

 System load: 0.0          Processes:           100
 Usage of /: 10.3% of 7.74GB   Users logged in:    1
 Memory usage: 6%           IP address for eth0: 172.31.42.137
 Swap usage: 0%

 Graph this data and manage this system at:
 https://landscape.canonical.com/

 0 updates can be installed immediately.
 0 of these updates are security updates.

 New release '16.04.6 LTS' available.
 Run 'do-release-upgrade' to upgrade to it.

 Last login: Mon Jun 24 09:32:39 2019 from 43.247.29.157
ubuntu@ip-172-31-42-137:~$
```

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: 63051630 (61M) [application/x-gzip]
Saving to: 'hadoop-1.2.1.tar.gz'

100%[=====] 63,051,630  11.1MB/s   in 5.9s
2019-06-24 10:07:18 (10.4 MB/s) - 'hadoop-1.2.1.tar.gz' saved [63051630/63051630]
ubuntu@ip-172-31-42-137:~$ █
```

```
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 -xvf 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. --&gt;

&lt;configuration&gt;
&lt;property&gt;
&lt;name&gt;fs.default.name&lt;/name&gt;
&lt;value&gt;hdfs://localhost:9000&lt;/value&gt;
&lt;/property&gt;
&lt;property&gt;
&lt;name&gt;hadoop.tmp.dir&lt;/name&gt;
&lt;value&gt;/usr/local/hadoop/tmp&lt;/value&gt;
&lt;/property&gt;
&lt;/configuration&gt;</pre>
```

- 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. --&gt;

&lt;configuration&gt;
&lt;property&gt;
&lt;name&gt;dfs.replication&lt;/name&gt;
&lt;value&gt;1&lt;/value&gt;
&lt;/property&gt;
&lt;/configuration&gt;</pre>
```

- 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. --&gt;

&lt;configuration&gt;
&lt;property&gt;
&lt;name&gt;mapred.job.tracker&lt;/name&gt;
&lt;value&gt;hdfs://localhost:9001&lt;/value&gt;
&lt;/property&gt;
&lt;/configuration&gt;</pre>
```

## 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 accessTokenUpdateInterval=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 occurring 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

← → C ⌂ ⓘ Not secure | 54.81.217.66:50070/dfshealth.jsp

### NameNode 'localhost:9000'

**Started:** Mon Jun 24 11:04:50 UTC 2019  
**Version:** 1.2.1, r1503152  
**Compiled:** Mon Jul 22 15:23:09 PDT 2013 by mattf  
**Upgrades:** There are no upgrades in progress.

[Browse the filesystem](#)  
[Namenode Logs](#)

#### Cluster Summary

8 files and directories, 1 blocks = 9 total. Heap Size is 26.04 MB / 966.69 MB (2%)

Configured Capacity	:	7.74 GB
DFS Used	:	40 KB
Non DFS Used	:	1.8 GB
DFS Remaining	:	5.94 GB
DFS Used%	:	0 %
DFS Remaining%	:	76.73 %
<a href="#">Live Nodes</a>	:	1
<a href="#">Dead Nodes</a>	:	0
<a href="#">Decommissioning Nodes</a>	:	0
Number of Under-Replicated Blocks	:	0

#### NameNode Storage:

## 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 dnncp_block_verification.log.curr VERSION
ubuntu@ip-172-31-42-137:/usr/local/hadoop/tmp/dfs/data/current$ nano clouddage
ubuntu@ip-172-31-42-137:/usr/local/hadoop/tmp/dfs/data/current$ ls
blk_3594589285843858810 blk_3594589285843858810_1001.meta clouddage dnncp_block_verification.log.curr VERSION
ubuntu@ip-172-31-42-137:/usr/local/hadoop/tmp/dfs/data/current$ hadoop fs -put clouddage .
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 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