

PRACTICAL - 8

Aim : Installing Hadoop in EC2

Step 1: Launch EC2 Instance

EC2 > Instances > i-045ba2b60826b86fc > Connect to instance

Connect to instance Info

Connect to your instance i-045ba2b60826b86fc (namenode) using any of these options

EC2 Instance Connect

Session Manager

SSH client

EC2 serial console

Instance ID
i-045ba2b60826b86fc (namenode)

1. Open an SSH client.
2. Locate your private key file. The key used to launch this instance is keyhadoop.pem
3. Run this command, if necessary, to ensure your key is not publicly viewable.
chmod 400 "keyhadoop.pem"
4. Connect to your instance using its Public DNS:
ec2-54-146-187-238.compute-1.amazonaws.com

Example:
ssh -i "keyhadoop.pem" ubuntu@ec2-54-146-187-238.compute-1.amazonaws.com

Note: In most cases, the guessed username is correct. However, read your AMI usage instructions to check if the AMI owner has changed the default AMI username.

Cancel

Step 2: Connect to the EC2 Instance

```
ubuntu@43.204.214.164:22 - Bitvise xterm - ubuntu@ip-172-31-1-43: -
ubuntu@ip-172-31-1-43:~$
ubuntu@ip-172-31-1-43:~$ hostname
ip-172-31-1-43
ubuntu@ip-172-31-1-43:~$ localhost
localhost: command not found
ubuntu@ip-172-31-1-43:~$
```

Step 3: Install Java

```
ubuntu@43.204.214.164:22 - Bitvise xterm - ubuntu@ip-172-31-1-43: -
Welcome to Ubuntu 24.04.1 LTS (GNU/Linux 6.8.0-1016-aws x86_64)

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

System information as of Mon Oct 21 10:45:00 UTC 2024

System load:  0.0               Processes:            116
Usage of /:   46.5% of 6.71GB   Users logged in:     1
Memory usage: 60%              IPv4 address for enx0: 172.31.1.43
Swap usage:   0%

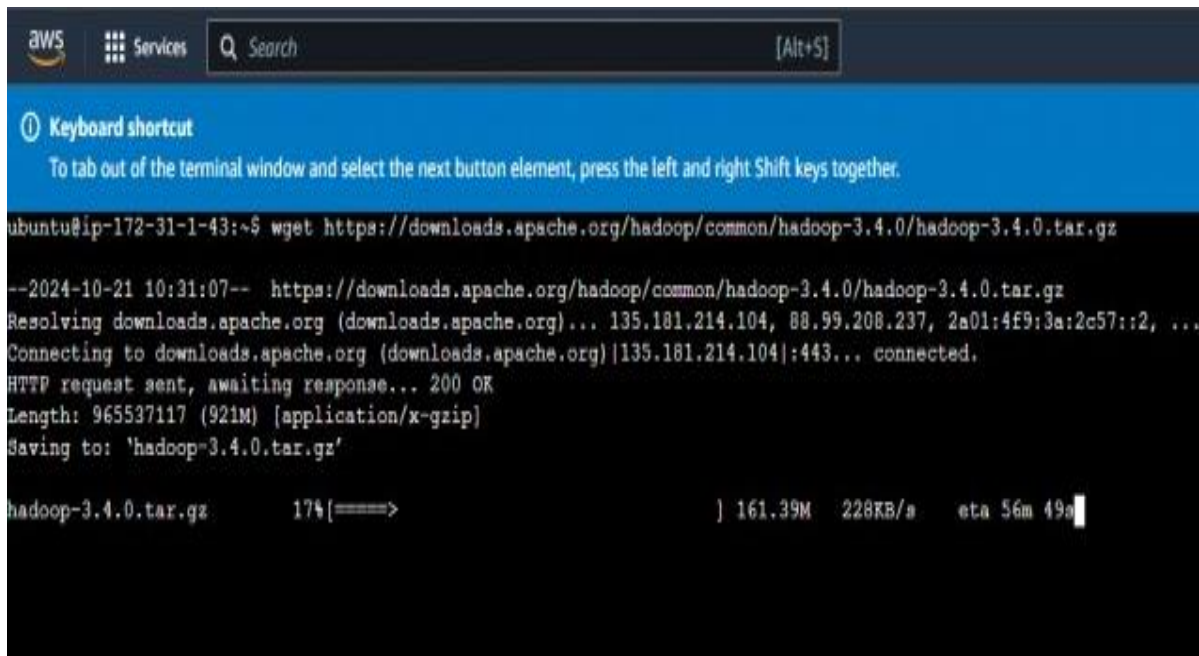
Expanded Security Maintenance for Applications is not enabled.

0 updates can be applied immediately.

Enable ESM Apps to receive additional future security updates.
See https://ubuntu.com/esm or run: sudo pro status

*** System restart required ***
Last login: Mon Oct 21 10:20:08 2024 from 13.233.177.4
ubuntu@ip-172-31-1-43:~$ sudo apt update
sudo apt install openjdk-8-jdk -y
Hit:1 http://ap-south-1.ec2.archive.ubuntu.com/ubuntu noble InRelease
Hit:2 http://ap-south-1.ec2.archive.ubuntu.com/ubuntu noble-updates InRelease
Hit:3 http://ap-south-1.ec2.archive.ubuntu.com/ubuntu noble-backports InRelease
Hit:4 http://security.ubuntu.com/ubuntu noble-security InRelease
Reading package lists... Done
Building dependency tree... Done
Reading state information... Done
All packages are up to date.
ubuntu@ip-172-31-1-43:~$
```

Step 4: Download Hadoop

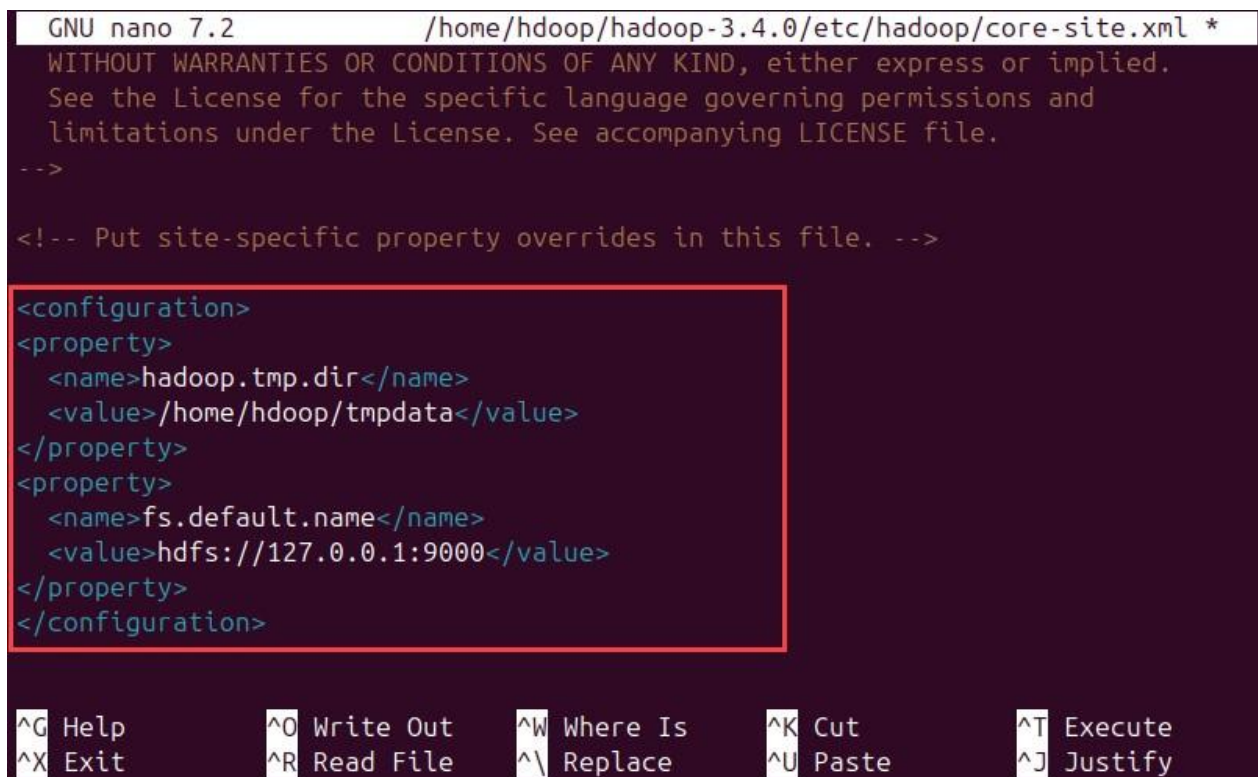


The screenshot shows a terminal window with the AWS logo and a search bar at the top. A blue banner at the top contains a keyboard shortcut tip: "Keyboard shortcut: To tab out of the terminal window and select the next button element, press the left and right Shift keys together." The terminal output shows the command `wget https://downloads.apache.org/hadoop/common/hadoop-3.4.0/hadoop-3.4.0.tar.gz` being executed. The output indicates the file is being resolved, connected to, and downloaded. The progress bar shows 17% completion, with a total size of 161.39M and a download speed of 228KB/s.

```
ubuntu@ip-172-31-1-43:~$ wget https://downloads.apache.org/hadoop/common/hadoop-3.4.0/hadoop-3.4.0.tar.gz
--2024-10-21 10:31:07-- https://downloads.apache.org/hadoop/common/hadoop-3.4.0/hadoop-3.4.0.tar.gz
Resolving downloads.apache.org (downloads.apache.org)... 135.181.214.104, 88.99.208.237, 2a01:4f9:3a:2c57::2, ...
Connecting to downloads.apache.org (downloads.apache.org)|135.181.214.104|:443... connected.
HTTP request sent, awaiting response... 200 OK
Length: 965537117 (921M) [application/x-gzip]
Saving to: 'hadoop-3.4.0.tar.gz'

hadoop-3.4.0.tar.gz      17%[====>                ] 161.39M  228KB/s  eta 56m 49s
```

Step 5: Configure Hadoop



The screenshot shows the nano text editor editing the file `/home/hdoop/hadoop-3.4.0/etc/hadoop/core-site.xml`. The editor displays the standard GNU nano 7.2 header and a red box highlights the configuration section. The configuration includes two properties: `hadoop.tmp.dir` set to `/home/hdoop/tmpdata` and `fs.default.name` set to `hdfs://127.0.0.1:9000`. The bottom of the screen shows the nano editor's command palette with various shortcuts.

```
GNU nano 7.2 /home/hdoop/hadoop-3.4.0/etc/hadoop/core-site.xml *
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>hadoop.tmp.dir</name>
  <value>/home/hdoop/tmpdata</value>
</property>
<property>
  <name>fs.default.name</name>
  <value>hdfs://127.0.0.1:9000</value>
</property>
</configuration>

^G Help      ^O Write Out ^W Where Is  ^K Cut       ^T Execute
^X Exit      ^R Read File ^\ Replace   ^U Paste     ^J Justify
```

Step 6: Set Environment Variables

```
GNU nano 7.4.2 /home/hadoop/hadoop-3.4.0/etc/hadoop/yarn-site.xml *
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.aux-services.mapreduce.shuffle.class</name>
  <value>org.apache.hadoop.mapred.ShuffleHandler</value>
</property>
<property>
  <name>yarn.resourcemanager.hostname</name>
  <value>127.0.0.1</value>
</property>
<property>
  <name>yarn.acl.enable</name>
  <value>0</value>
</property>
<property>
  <name>yarn.nodemanager.env-whitelist</name>
  <value>JAVA_HOME,HADOOP_COMMON_HOME,HADOOP_HDFS_HOME,HADOOP_CONF_DIR,CLASSPATH_PERPEND_DISTCACHE,HADOOP_YARN_HOME,HADOOP_MAPRED_HOME</value>
</property>
</configuration>
```

^G Help ^O Write Out ^M Where Is ^K Cut ^T Execute ^C Location ^U Undo ^-A Set Mark ^_] To Bracket
 ^X Exit ^R Read File ^\ Replace ^V Paste ^J Justify ^_ Go To Line ^E Redo ^-G Copy ^_ Where Was

Step 7: Format the HDFS

```
hadoop@phoenixnap:~$ hdfs namenode -format
WARNING: /home/hadoop/hadoop-3.4.0/logs does not exist. Creating.
2024-09-09 13:08:42,739 INFO namenode.NameNode: STARTUP_MSG:
/******
STARTUP_MSG: Starting NameNode
STARTUP_MSG: host = phoenixnap/127.0.1.1
STARTUP_MSG: args = [-format]
STARTUP_MSG: version = 3.4.0
STARTUP_MSG: classpath = /home/hadoop/hadoop-3.4.0/etc/hadoop:/home/hadoop/hadoc
3.4.0/share/hadoop/common/lib curator-client-5.2.0.jar:/home/hadoop/hadoop-3.4.0,
2024-09-09 13:08:45,012 INFO namenode.FSNamesystem: Stopping services started fo
standby state
2024-09-09 13:08:45,018 INFO namenode.FSImage: FSImageSaver clean checkpoint: tx
=0 when meet shutdown.
2024-09-09 13:08:45,019 INFO namenode.NameNode: SHUTDOWN_MSG:
/******
SHUTDOWN_MSG: Shutting down NameNode at phoenixnap/127.0.1.1
```

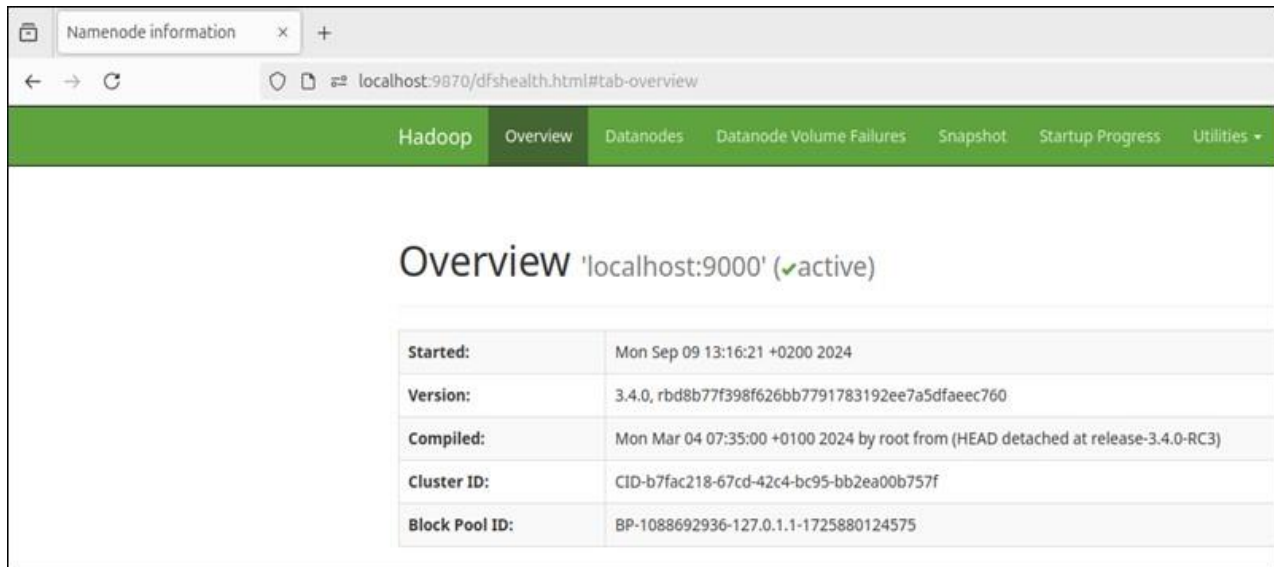
Step 8: Start Hadoop

```
hadoop@phoenixnap:~/hadoop-3.4.0/sbin$ ./start-dfs.sh
Starting namenodes on [localhost] ←
Starting datanodes ←
Starting secondary namenodes [phoenixnap] ←
```

```
hadoop@phoenixnap:~/hadoop-3.4.0/sbin$ ./start-yarn.sh
Starting resourcemanager ←
Starting nodemanagers ←
```

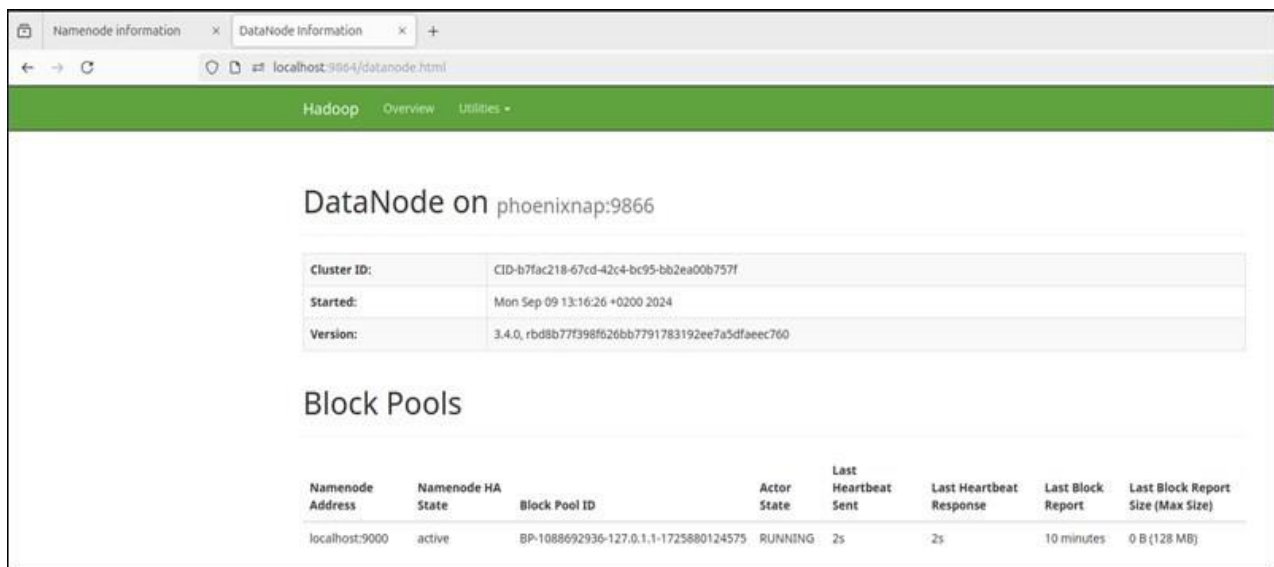
```
hadoop@phoenixnap:~/hadoop-3.4.0/sbin$ jps
45169 DataNode
46355 ResourceManager
45033 NameNode
46476 NodeManager
45373 SecondaryNameNode
47390 Jps
```


Step 9: Access Hadoop



The screenshot shows a web browser window with the address bar displaying `localhost:9870/dfshealth.html#tab-overview`. The page has a green header bar with navigation tabs: **Hadoop**, **Overview**, **Datanodes**, **Datanode Volume Failures**, **Snapshot**, **Startup Progress**, and **Utilities**. The main content area is titled **Overview 'localhost:9000' (✓active)**. Below the title is a table with the following information:

Started:	Mon Sep 09 13:16:21 +0200 2024
Version:	3.4.0, rbd8b77f398f626bb7791783192ee7a5dfaec760
Compiled:	Mon Mar 04 07:35:00 +0100 2024 by root from (HEAD detached at release-3.4.0-RC3)
Cluster ID:	CID-b7fac218-67cd-42c4-bc95-bb2ea00b757f
Block Pool ID:	BP-1088692936-127.0.1.1-1725880124575



The screenshot shows a web browser window with the address bar displaying `localhost:9864/datanode.html`. The page has a green header bar with navigation tabs: **Hadoop**, **Overview**, and **Utilities**. The main content area is titled **DataNode on phoenixnap:9866**. Below the title is a table with the following information:

Cluster ID:	CID-b7fac218-67cd-42c4-bc95-bb2ea00b757f
Started:	Mon Sep 09 13:16:26 +0200 2024
Version:	3.4.0, rbd8b77f398f626bb7791783192ee7a5dfaec760

Below this table is a section titled **Block Pools**. It contains a table with the following information:

Namenode Address	Namenode HA State	Block Pool ID	Actor State	Last Heartbeat Sent	Last Heartbeat Response	Last Block Report	Last Block Report Size (Max Size)
localhost:9000	active	BP-1088692936-127.0.1.1-1725880124575	RUNNING	2s	2s	10 minutes	0 B (128 MB)

Conclusion : Hence in this practical, we have successfully studied to install Hadoop in EC2 system.