

### Installation of Ubuntu on Virtual Box.

1. Download the .iso image for desktop for Ubuntu.
2. Open Virtual box and click on the new button.
3. Give a name to the virtual machine and select the location for it to install.
4. Assign RAM size to your virtual machine.
5. Create a virtual hard disk for the machine to store files.
6. Select the type of hard disk. Using VDI type is recommended.
7. Either of physical storage type can be selected using dynamically allocated disk is by default.
8. Select disk size and provide the destination folder to install.
9. After the disk creation is done, boot the virtual machine and begin installing.
10. If the installation disk is not automatically detected. Browse the file location and select the iso file for Ubuntu.
11. Proceed with installation file and wait for future option.
12. Click on install Ubuntu and select keyboard layout.
13. Select installation type and click on the install now.
14. Select location, time zone, choose name for computer and set password.
15. Once the installation process is done, reboot VM.



## Installing Hadoop on Ubuntu.

### 1. Install Java Development Kit (JDK)

- > sudo apt update
- > sudo apt install openjdk-8-jdk

### 2. Verify java version.

- > java -version.

### 3. Install SSH:

- > sudo apt install ssh.

Secure shell (SSH) is essential to Hadoop's operation since it secures communication between clustered nodes.

### 4. Create Hadoop user.

- > sudo adduser hadoop.

Create a dedicated user specifically for hadoop operation.

### 5. Switch user

- > su - hadoop.

### 6. Configure SSH

- > ssh-keygen -t rsa

### 7. Set permission

```
cat ~/.ssh/id_rsa.pub >> ~/.ssh/authorized_keys.  
chmod 640 ~/.ssh/authorized-keys.
```

### 8. SSH to localhost:

- > ssh localhost

### 9. Switch user:

- > su - hadoop.

### 10. Install hadoop.



> wget https://dlcdn.apache.org/hadoop/common/  
hadoop-3.3.6/hadoop-3.3.6.tar.gz

extract the contents using,

```
tar -xvzf hadoop-3.3.6.tar.gz  
rm -rf *.hashrc
```

> to setup the environment variables:

```
export JAVA_HOME = /usr/lib/jvm/java-8-openjdk-amd64
```

```
export HADOOP_HOME = /home/hadoop/hadoop
```

```
export HADOOP_INSTALL = $HADOOP_HOME
```

```
export HADOOP_MAPRED_HOME = $HADOOP_HOME
```

```
export HADOOP_COMMON_HOME = $HADOOP_HOME
```

```
export HADOOP_HDFS_HOME = $HADOOP_HOME
```

```
export HADOOP_YARN_HOME = $HADOOP_HOME
```

```
export HADOOP_COMMON_LIB_NATIVE_DIR = $HADOOP_HOME/lib/native
```

```
export PATH = $PATH:$HADOOP_HOME/sbin:$HADOOP_HOME/bin
```

```
export HADOOP_OPTS="-Djava.library.path=$HADOOP_HOME/lib/native"
```

> source v/.bashrc.

> open hadoop environment configuration file.

```
nano $HADOOP_HOME/etc/hadoop/hadoop-env.sh
```

> set Java path:

```
export JAVA_HOME = /usr/lib/jvm/java-8-openjdk-amd64
```

II. Configure Hadoop:

> create necessary directories.

```
cd /hadoop/
```

```
mkdir -p /hadoopdata/hdfs/{namenode, datanode}
```

> Edit core-site.xml:

```
nano $HADOOP_HOME/etc/hadoop/core-site.xml
```



> Edit hdfs-site.xml

```
nano $ HADOOP_HOME/etc/hadoop/hdfs-site.xml
```

> Edit mapred-site.xml

```
nano $ Hadoop-Home/etc/hadoop/mapred-site.xml
```

> Edit yarn-site.xml

```
nano $ Hadoop-home/etc/hadoop/yarn-site.xml
```

12. Start Hadoop cluster:

> Format the name node:

```
hdfs namenode-format
```

> Start the hadoop cluster

```
start-all.sh.
```



varsh@ubuntu: ~/Desktop



```
varsh@ubuntu:~/Desktop$ java -version
openjdk version "1.8.0_422"
OpenJDK Runtime Environment (build 1.8.0_422-8u422-b05-1~22.04-b05)
OpenJDK 64-Bit Server VM (build 25.422-b05, mixed mode)
varsh@ubuntu:~/Desktop$
```



```
hadoop-3.2.3/lib/native/libnativetask.so
hadoop-3.2.3/lib/native/libhdfspp.a
hadoop-3.2.3/LICENSE.txt
varsh@ubuntu:~$ cd hadoop-3.2.3/
varsh@ubuntu:~/hadoop-3.2.3$ ls
bin  etc  include  lib  libexec  LICENSE.txt  NOTICE.txt  README.txt  sbin  share
varsh@ubuntu:~/hadoop-3.2.3$ cd etc/hadoop/
varsh@ubuntu:~/hadoop-3.2.3/etc/hadoop$ ls
capacity-scheduler.xml      httpfs-log4j.properties      mapred-site.xml
configuration.xsl           httpfs-signature.secret      shellprofile.d
container-executor.cfg      httpfs-site.xml              ssl-client.xml.example
core-site.xml               kms-acls.xml                  ssl-server.xml.example
hadoop-env.cmd              kms-env.sh                    user_ec_policies.xml.template
hadoop-env.sh               kms-log4j.properties         workers
hadoop-metrics2.properties kms-site.xml                  yarn-env.cmd
hadoop-policy.xml           log4j.properties             yarn-env.sh
hadoop-user-functions.sh.example mapred-env.cmd                yarnservice-log4j.properties
hdfs-site.xml               mapred-env.sh                 yarn-site.xml
httpfs-env.sh               mapred-queues.xml.template
varsh@ubuntu:~/hadoop-3.2.3/etc/hadoop$ sudo nano hadoop-env.sh
[sudo] password for varsh:
varsh@ubuntu:~/hadoop-3.2.3/etc/hadoop$ sudo nano core-site.xml
varsh@ubuntu:~/hadoop-3.2.3/etc/hadoop$ sudo nano hdfs-site.xml
varsh@ubuntu:~/hadoop-3.2.3/etc/hadoop$ sudo nano mapred-site.xml
varsh@ubuntu:~/hadoop-3.2.3/etc/hadoop$ sudo nano yarn-site.xml
varsh@ubuntu:~/hadoop-3.2.3/etc/hadoop$ ssh localhost
The authenticity of host 'localhost (127.0.0.1)' can't be established.
ED25519 key fingerprint is SHA256:pyydpR3wKFJgIoa0wnRP5WDyI8ayuqT5/GbGzLwCCJ0.
This key is not known by any other names
Are you sure you want to continue connecting (yes/no/[fingerprint])? yes
Warning: Permanently added 'localhost' (ED25519) to the list of known hosts.
varsh@localhost's password:
Welcome to Ubuntu 22.04.4 LTS (GNU/Linux 6.5.0-41-generic x86_64)

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

Expanded Security Maintenance for Applications is not enabled.
```

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

Expanded Security Maintenance for Applications is not enabled.

252 updates can be applied immediately.  
195 of these updates are standard security updates.  
To see these additional updates run: `apt list --upgradable`

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

The programs included with the Ubuntu system are free software;  
the exact distribution terms for each program are described in the  
individual files in `/usr/share/doc/*/copyright`.

Ubuntu comes with ABSOLUTELY NO WARRANTY, to the extent permitted by  
applicable law.

```
varsh@ubuntu:~$ ssh-keygen -t rsa -P '' -f ~/.ssh/id_rsa
cat ~/.ssh/id_rsa.pub >> ~/.ssh/authorized_keys
Generating public/private rsa key pair.
Your identification has been saved in /home/varsh/.ssh/id_rsa
Your public key has been saved in /home/varsh/.ssh/id_rsa.pub
The key fingerprint is:
SHA256:e7xxP6SvY7cfR0IPLNhqA6GQW09QnQtT6wJNBqd2gTQ varsh@ubuntu
The key's randomart image is:
```

```
+---[RSA 3072]-----+
|  .oE=Bo..          |
|  ..o0+oo+  .       |
|  o=o+oo.o  +       |
|  .. o.o.. o o      |
|      S =   . o      |
|      =   .  .o      |
|      . +  .o ..     |
|      . ++.o o       |
|      ...=+++.      |
+---[SHA256]-----+
```

```
varsh@ubuntu:~$
```



```
varsh@ubuntu:~$ ls /home/varsh/hadoop-3.2.3/bin
container-executor  hadoop  hadoop.cmd  hdfs  hdfs.cmd  mapred  mapred.cmd  oom-listener  test-container-executor  yarn  yarn.cmd
varsh@ubuntu:~$ nano ~/.bashrc
varsh@ubuntu:~$ source ~/.bashrc
varsh@ubuntu:~$ hadoop version
Hadoop 3.2.3
Source code repository https://github.com/apache/hadoop -r abe5358143720085498613d399be3bbf01e0f131
Compiled by ubuntu on 2022-03-20T01:18Z
Compiled with protoc 2.5.0
From source with checksum 39bb14faec14b3aa25388a6d7c345fe8
This command was run using /home/varsh/hadoop-3.2.3/share/hadoop/common/hadoop-common-3.2.3.jar
varsh@ubuntu:~$ hadoop fs -mkdir /user
varsh@ubuntu:~$ hadoop fs -mkdir /user/hadoopEg
varsh@ubuntu:~$ touch demo.csv
varsh@ubuntu:~$ hadoop fs -put demo.csv /user/hadoopEg
varsh@ubuntu:~$
```



Activities

Firefox Web Browser

Sep 1 13:24

Install Hadoop on Ubuntu

Gemini

ChatGPT

Namenode information

localhost:9870/dfshealth.html#tab-overview

Hadoop

Overview

Datanodes

Datanode Volume Failures

Snapshot

Startup Progress

Utilities

# Overview 'localhost:9000' (active)

Started:	Sun Sep 01 13:08:41 +0530 2024
Version:	3.2.3, rabe5358143720085498613d399be3bbf01e0f131
Compiled:	Sun Mar 20 06:48:00 +0530 2022 by ubuntu from branch-3.2.3
Cluster ID:	CID-acd79c72-8c9c-4d4f-8a9e-11dbefe1d229
Block Pool ID:	BP-1864208724-127.0.1.1-1725173939730

## Summary

Security is off.

Safemode is off.

4 files and directories, 0 blocks (0 replicated blocks, 0 erasure coded block groups) = 4 total filesystem object(s).

Heap Memory used 104.76 MB of 187 MB Heap Memory. Max Heap Memory is 654.5 MB.

Non Heap Memory used 60.48 MB of 61.71 MB Committed Non Heap Memory. Max Non Heap Memory is <unbounded>.

Configured Capacity:	23.94 GB
----------------------	----------



Activities

Firefox Web Browser

Sep 1 13:23

Install Hadoop on Ubuntu

Gemini

ChatGPT

Browsing HDFS

localhost:9870/explorer.html

Hadoop

Overview

Datanodes

Datanode Volume Failures

Snapshot

Startup Progress

Utilities

Files

Browse Directory

/

Go!

Show 25 entries

Search:

	Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name	
<input type="checkbox"/>	<u>drwxr-xr-x</u>	<u>varsh</u>	<u>supergroup</u>	0 B	Sep 01 13:19	0	0 B	<u>user</u>	

Showing 1 to 1 of 1 entries

Previous1Next

Hadoop, 2022.



```
varsh@ubuntu:~$ /home/varsh/hadoop-3.2.3/sbin/start-all.sh
WARNING: Attempting to start all Apache Hadoop daemons as varsh 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]
Starting resourcemanager
resourcemanager is running as process 9810. Stop it first and ensure /tmp/hadoop-varsh-resourcemanager.pid file is empty before retry.
Starting nodemanagers
```



```
varsh@ubuntu:~$ /home/varsh/hadoop-3.2.3/sbin/stop-all.sh
WARNING: Stopping all Apache Hadoop daemons as varsh in 10 seconds.
WARNING: Use CTRL-C to abort.
Stopping namenodes on [localhost]
Stopping datanodes
Stopping secondary namenodes [ubuntu]
Stopping nodemanagers
Stopping resourcemanager
varsh@ubuntu:~$
```