

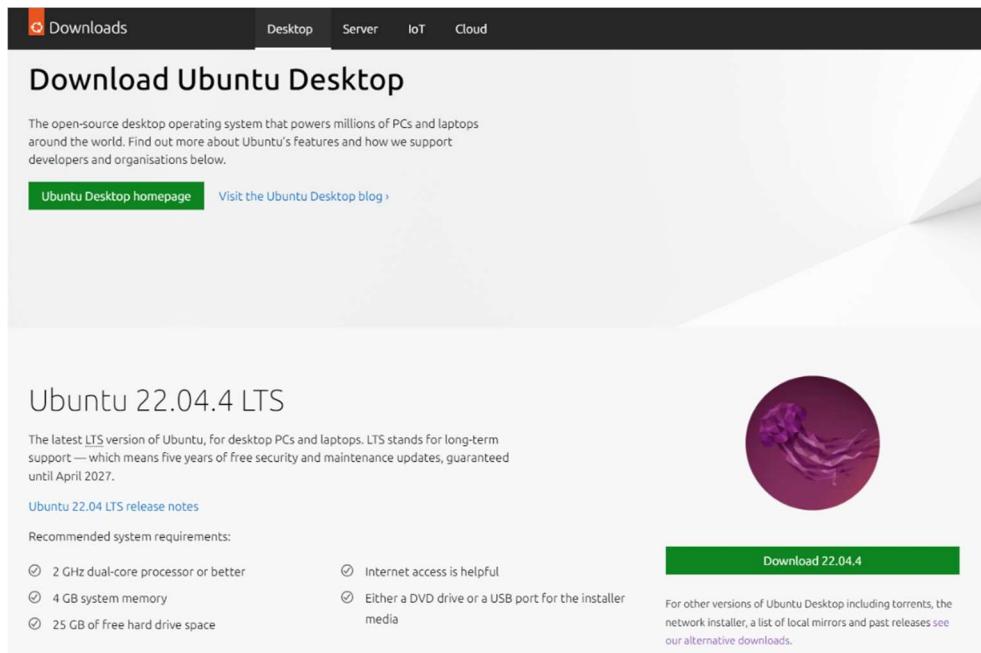
S.No	Date	Name of the Practical	Sign
1		Installation of Ubuntu	
2		Installation of Hadoop on Ubuntu	
3		Single node Hadoop deployment (pseudodistributed mode) <ul style="list-style-type: none"> • bashrc • hadoop-env.sh • core-site.xml • hdfs-site.xml • mapred-site-xml • yarn-site.xml 	
4		Access Hadoop UI from Browser	
5		Perform Hands-on basic Hadoop commands	

PRACTICAL - 1 Aim:

Installation of Ubuntu

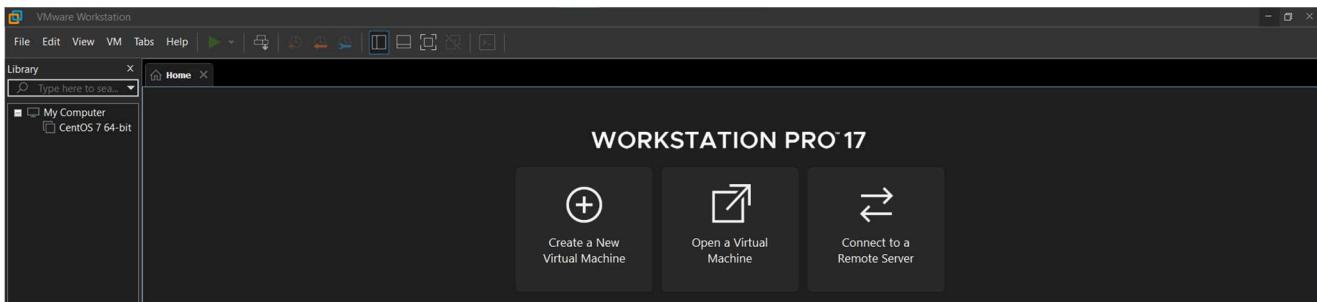
Source Code & Output:

1. Go to <https://ubuntu.com/download/desktop>



The screenshot shows the Ubuntu Desktop download page. At the top, there's a navigation bar with 'Downloads' (highlighted in orange), 'Desktop', 'Server', 'IoT', and 'Cloud'. Below the navigation bar, the title 'Download Ubuntu Desktop' is displayed. A brief description follows: 'The open-source desktop operating system that powers millions of PCs and laptops around the world. Find out more about Ubuntu's features and how we support developers and organisations below.' Below this, there are two buttons: 'Ubuntu Desktop homepage' and 'Visit the Ubuntu Desktop blog'. The main content area features a large image of a purple and white abstract pattern. To the left, there's a section for 'Ubuntu 22.04.4 LTS' with a brief description: 'The latest LTS version of Ubuntu, for desktop PCs and laptops. LTS stands for long-term support — which means five years of free security and maintenance updates, guaranteed until April 2027.' Below this is a link to 'Ubuntu 22.04 LTS release notes'. To the right, there's a 'Download 22.04.4' button. On the far right, there's a note: 'For other versions of Ubuntu Desktop including torrents, the network installer, a list of local mirrors and past releases see our alternative downloads.'

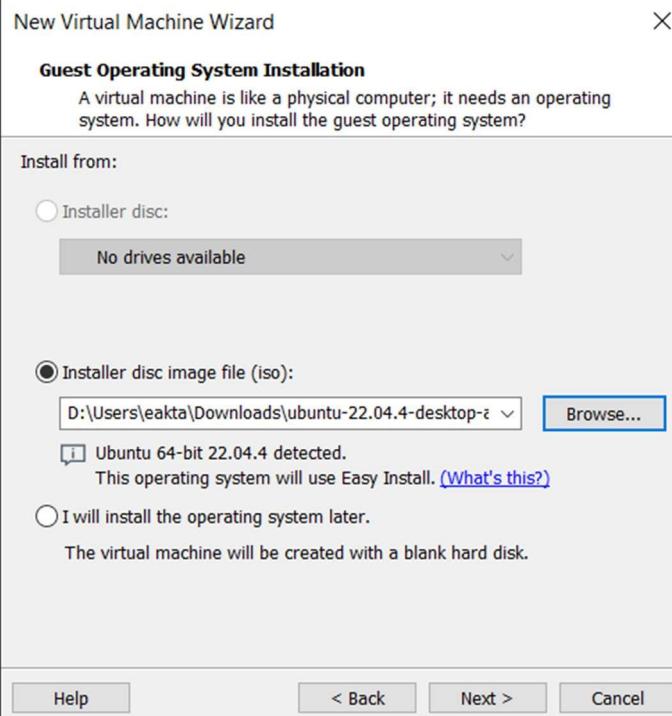
2. Open VMware Workstation. Click on 'Create a New Virtual Machine'.



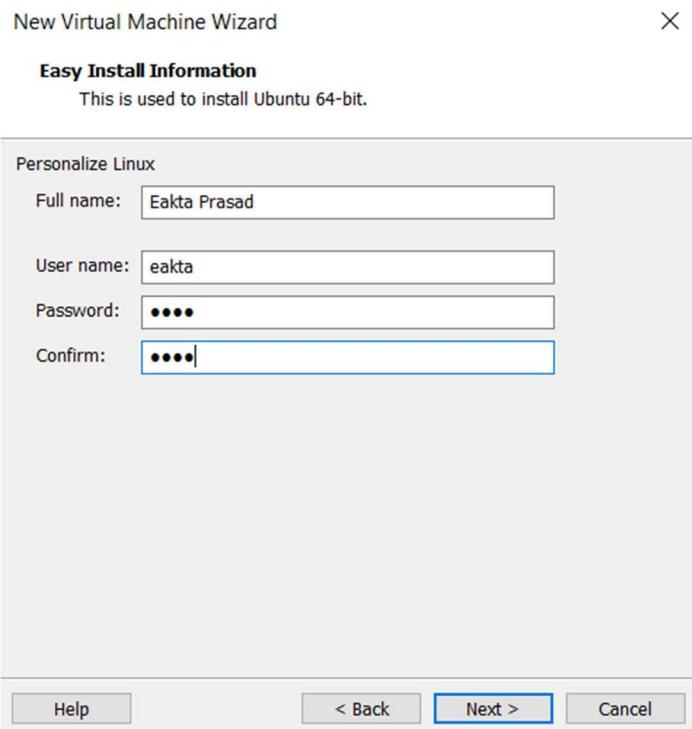
3. New Virtual Machine Wizard opens. Click on Next.



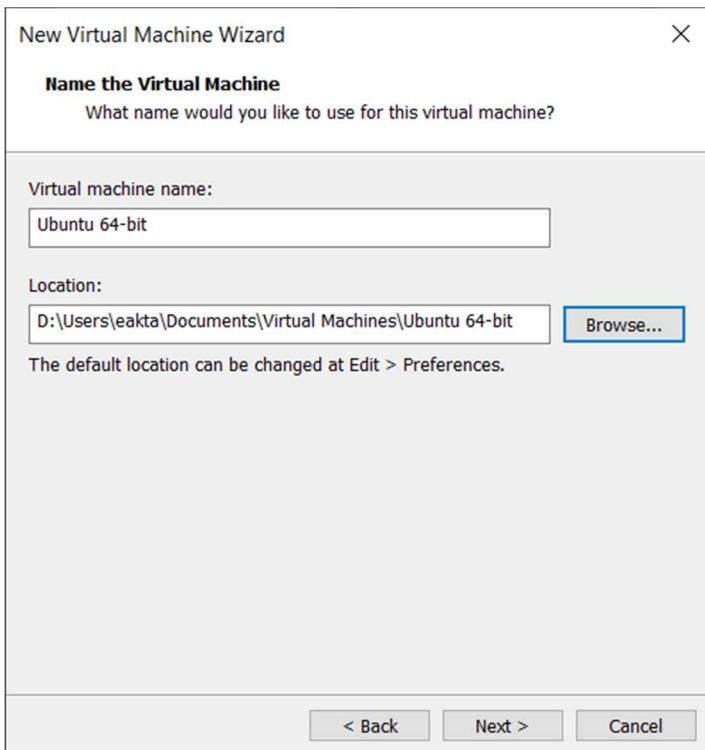
4. Browse the path where the ubuntu's iso file is downloaded. And click on Next.



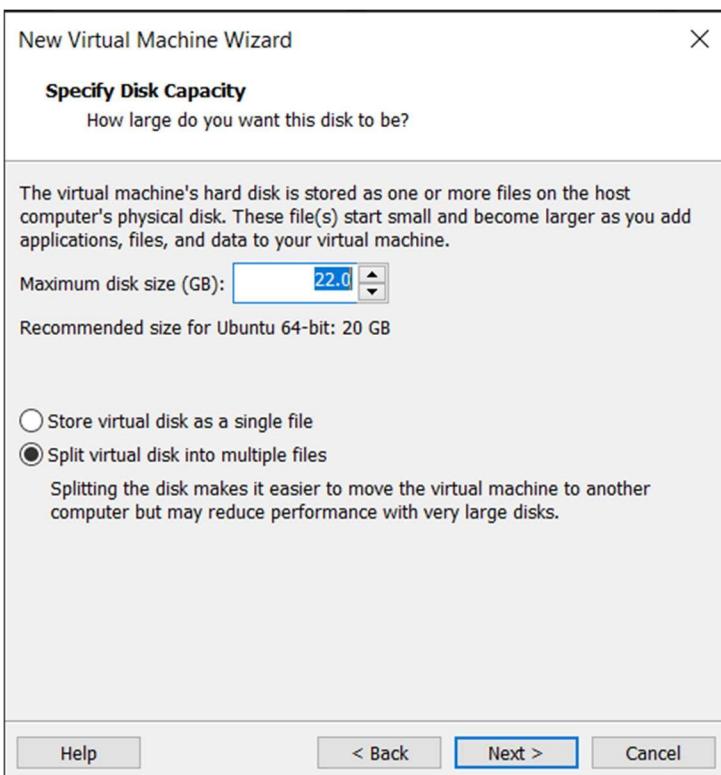
5. Fill the details and click on Next.



6. Provide a name for your virtual machine and also the location. Then click on Next.



7. Click on Next.



8. Click on Finish.

Ready to Create Virtual Machine

Click Finish to create the virtual machine and start installing Ubuntu 64-bit and then VMware Tools.

The virtual machine will be created with the following settings:

Name:	Ubuntu 64-bit
Location:	D:\Users\eakta\Documents\Virtual Machines\Ubuntu 64...
Version:	Workstation 17.x
Operating System:	Ubuntu 64-bit
Hard Disk:	22 GB, Split
Memory:	2048 MB
Network Adapter:	NAT
Other Devices:	2 CPU cores, CD/DVD, USB Controller, Printer, Sound C...

[Customize Hardware...](#)

Power on this virtual machine after creation

< Back

[Finish](#)

Cancel

9. Ubuntu has been successfully created in VMWare. Now select Ubuntu 64-bit which appears on the left panel and then click on the Play virtual machine option. Select the keyboard layout and click on next.

Keyboard layout

Choose your keyboard layout:

English (Australian)
English (Cameroon)
English (Ghana)
English (Nigeria)
English (South Africa)
English (UK)
English (US)
Esperanto
Estonian
Faroese
Filipino
Finnish
French
French (Canada)

English (US)
English (US) - Cherokee
English (US) - English (Colemak)
English (US) - English (Colemak-DH ISO)
English (US) - English (Colemak-DH)
English (US) - English (Dvorak)
English (US) - English (Dvorak, alt. intl.)
English (US) - English (Dvorak, intl., with dead keys)
English (US) - English (Dvorak, left-handed)
English (US) - English (Dvorak, right-handed)
English (US) - English (Macintosh)
English (US) - English (Norman)
English (US) - English (US, Symbolic)
English (US) - English (US, alt.intl.)

Type here to test your keyboard

[Detect Keyboard Layout](#)

Quit

Back

Continue

› Verifying the installation configuration...

10. Select Normal installation and also both the Other options. Then click on Continue.

Updates and other software

What apps would you like to install to start with?

Normal installation

Web browser, utilities, office software, games, and media players.

Minimal installation

Web browser and basic utilities.

Other options

Download updates while installing Ubuntu

This saves time after installation.

Install third-party software for graphics and Wi-Fi hardware and additional media formats

This software is subject to license terms included with its documentation. Some is proprietary.

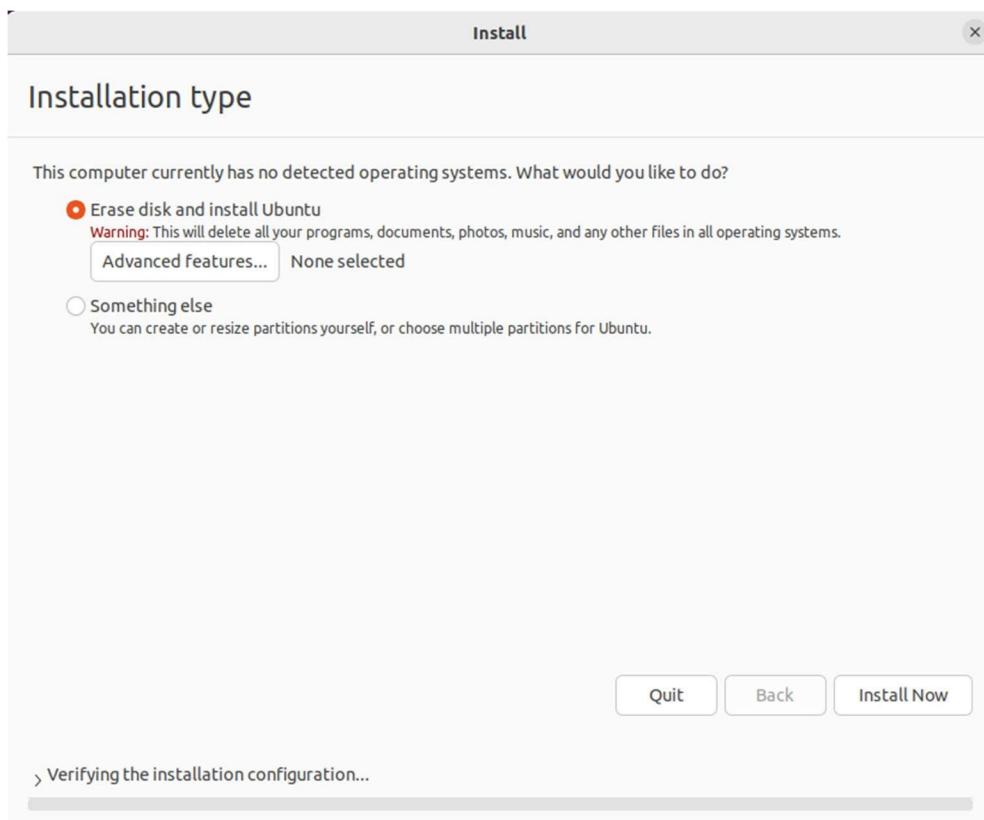
Quit

Back

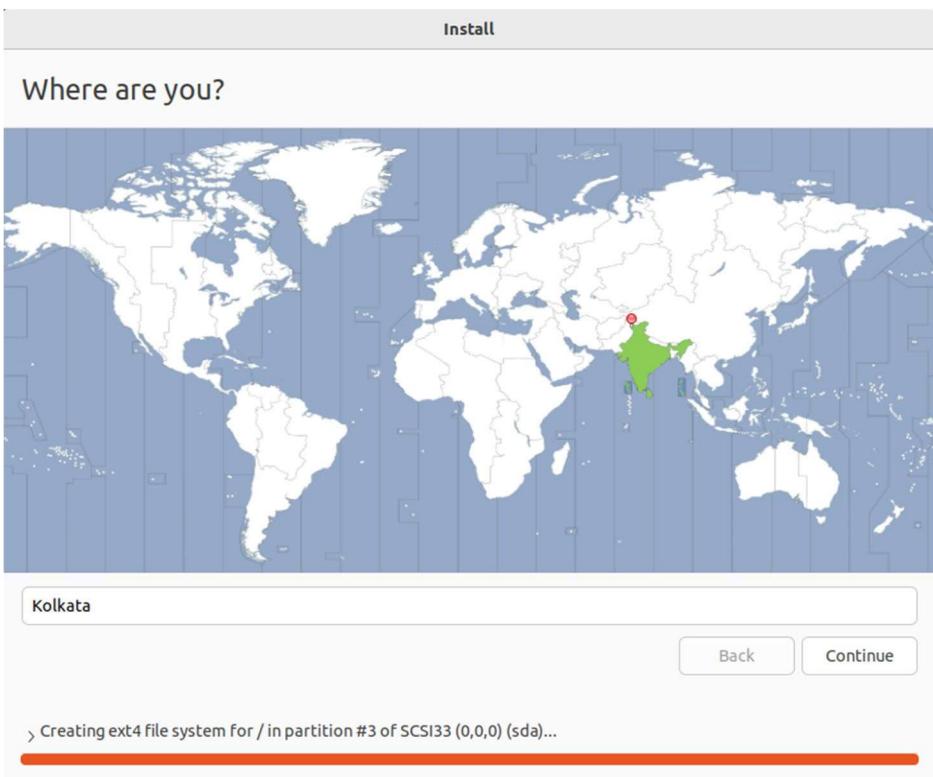
Continue

> Verifying the installation configuration...

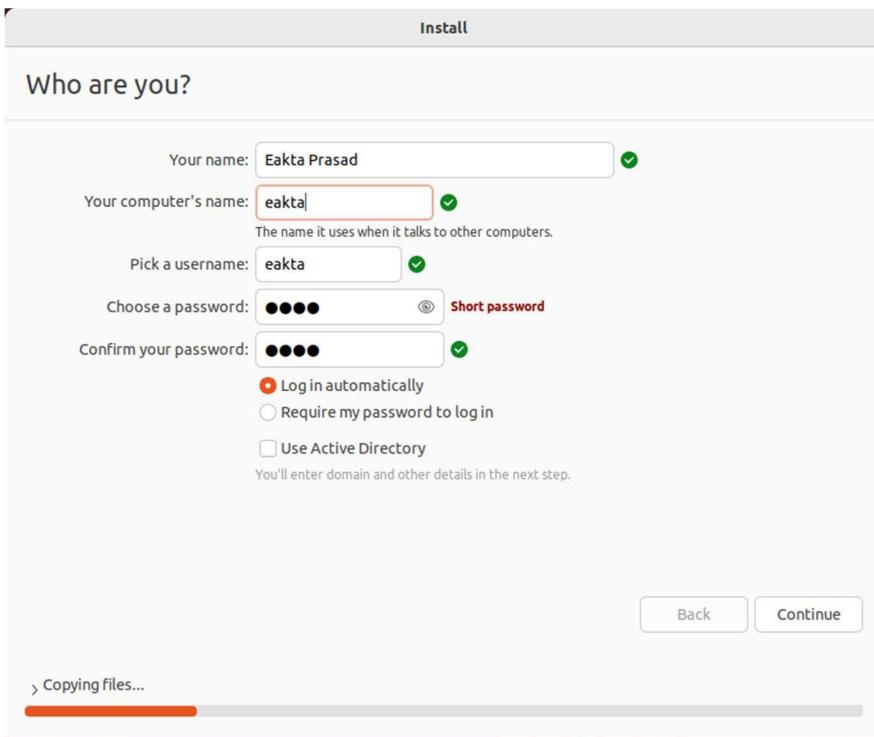
11. Select 'Erase disk and install Ubuntu' and click on Install Now.



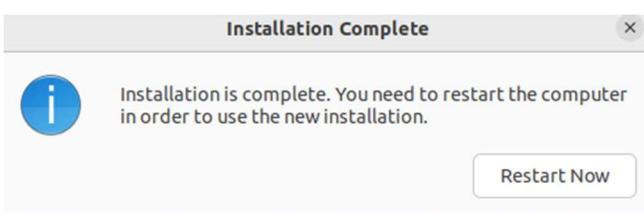
12. Select the zone(city) and click on Continue.



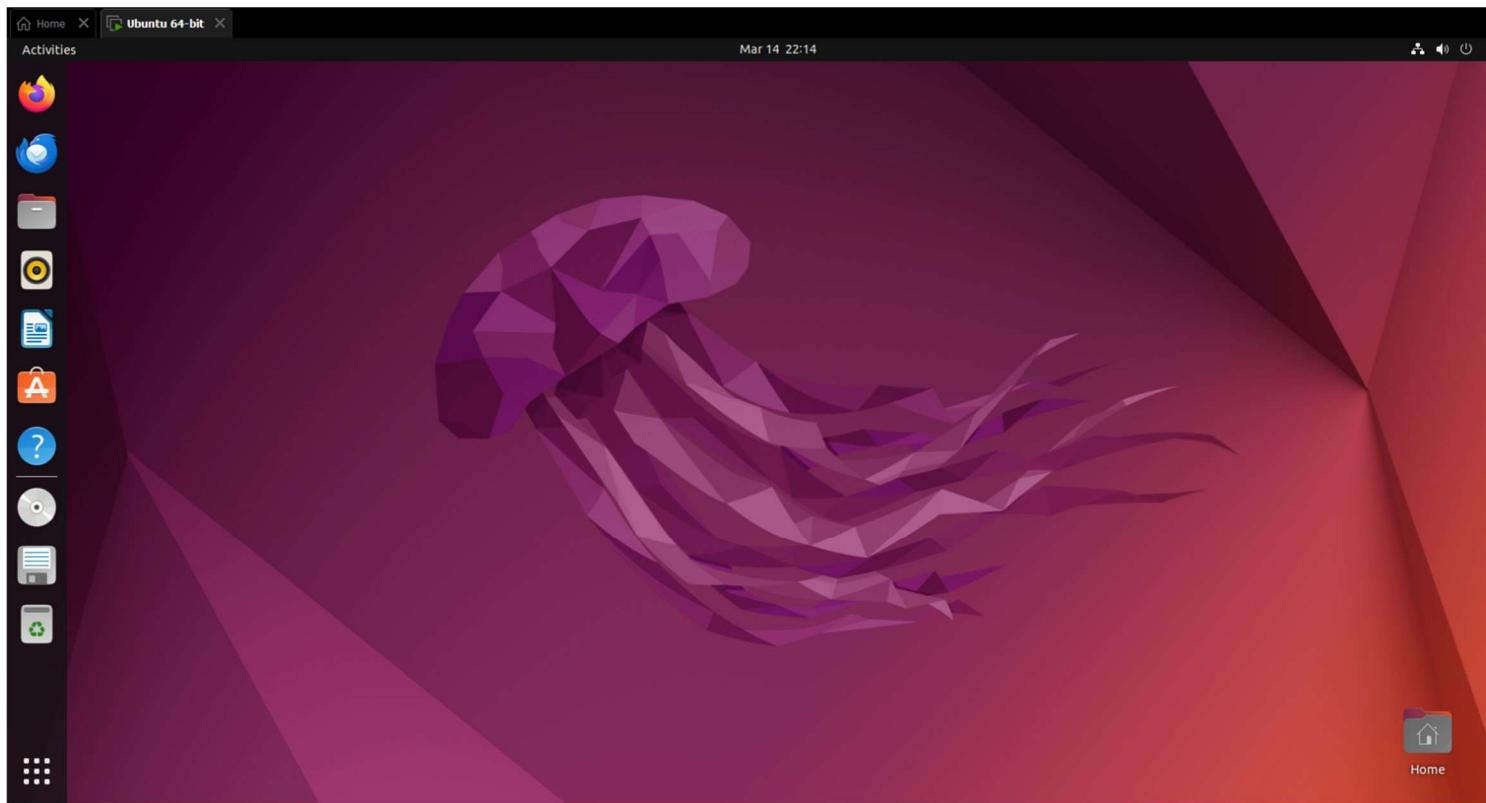
13. Fill the details and click on Continue.



14. The Ubuntu installer will begin copying files on the virtual machine disk Once the installation completes, click on the Restart Now button.



15. Ubuntu is successfully installed.



PRACTICAL - 2 Aim:

Installation of Hadoop on Ubuntu

Source Code & Output:

Command 1: update your system before initiating a new installation

```
$ sudo apt update
```

```
eakta@eakta:~$ sudo apt update
[sudo] password for eakta:
Hit:1 http://in.archive.ubuntu.com/ubuntu jammy InRelease
Get:2 http://in.archive.ubuntu.com/ubuntu jammy-updates InRelease [119 kB]
Hit:3 http://in.archive.ubuntu.com/ubuntu jammy-backports InRelease
Get:4 http://security.ubuntu.com/ubuntu jammy-security InRelease [110 kB]
Fetched 229 kB in 5s (43.6 kB/s)
Reading package lists... Done
Building dependency tree... Done
Reading state information... Done
2 packages can be upgraded. Run 'apt list --upgradable' to see them.
```

Command 2: install OpenJDK 8

```
$ sudo apt install openjdk-8-jdk -y
```

```
eakta@eakta:~$ sudo apt install openjdk-8-jdk -y
Reading package lists... Done
Building dependency tree... Done
Reading state information... Done
The following additional packages will be installed:
  ca-certificates-java fonts-dejavu-extra java-common libatk-wrapper-java libatk-wrapper-java-jni libice-dev libpthread-stubs0-dev libsm-dev libx11-dev libxau-dev libxcb1-dev libxdmcp-dev libxt-dev
  openjdk-8-jdk-headless openjdk-8-jre openjdk-8-jre-headless x11proto-dev xorg-sgml-doctools xtrans-dev
Suggested packages:
  default-jre libice-doc libsm-doc libxcb1-doc libxt-doc openjdk-8-demo openjdk-8-source visualvm fonts-nanum fonts-ipafont-gothic fonts-ipafont-mincho fonts-wqy-microhei fonts-wqy-zenhei
The following NEW packages will be installed:
  ca-certificates-java fonts-dejavu-extra java-common libatk-wrapper-java libatk-wrapper-java-jni libice-dev libpthread-stubs0-dev libsm-dev libx11-dev libxau-dev libxcb1-dev libxdmcp-dev libxt-dev
  openjdk-8-jdk openjdk-8-jdk-headless openjdk-8-jre openjdk-8-jre-headless x11proto-dev xorg-sgml-doctools xtrans-dev
0 upgraded, 20 newly installed, 0 to remove and 2 not upgraded.
```

Command 3: Once the installation process is complete, verify the current Java version

```
$ java -version; javac -version
```

```
eakta@eakta:~$ java -version; javac -version
openjdk version "1.8.0_392"
OpenJDK Runtime Environment (build 1.8.0_392-8u392-ga-1~22.04-b08)
OpenJDK 64-Bit Server VM (build 25.392-b08, mixed mode)
javac 1.8.0_392
```

Command 4: Install the OpenSSH server and client

```
$ sudo apt install openssh-server openssh-client -y
```

```
eakta@eakta:~$ sudo apt install openssh-server openssh-client -y
Reading package lists... Done
Building dependency tree... Done
Reading state information... Done
openssh-client is already the newest version (1:8.9p1-3ubuntu0.6).
openssh-client set to manually installed.
The following additional packages will be installed:
  ncurses-term openssh-sftp-server ssh-import-id
Suggested packages:
  molly-guard monkeysphere ssh-askpass
The following NEW packages will be installed:
  ncurses-term openssh-server openssh-sftp-server ssh-import-id
0 upgraded, 4 newly installed, 0 to remove and 2 not upgraded.
```

Command 5: Generate an SSH key pair and define the location is to be stored in -

```
$ ssh-keygen -t rsa -P '' -f ~/.ssh/id_rsa
```

```
eakta@eakta:~$ ssh-keygen -t rsa -P '' -f ~/.ssh/id_rsa
Generating public/private rsa key pair.
Created directory '/home/eakta/.ssh'.
Your identification has been saved in /home/eakta/.ssh/id_rsa
Your public key has been saved in /home/eakta/.ssh/id_rsa.pub
The key fingerprint is:
SHA256:W23BFEEZyhNBjTkUYVsZkFL4Q+cpvTxpjw7g3ISTo0 eakta@eakta
The key's randomart image is:
+---[RSA 3072]---+
| oB@@B= |
| *XBo |
| .+%.o |
| . + B . |
| oS o * |
| E .+ + |
| o ...+ |
| + o oo |
| + .o |
+---[SHA256]---+
```

Command 6: Use the cat command to store the public key as authorized_keys in the ssh directory

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

Command 7: Set the permissions for your user with the chmod command

```
$ chmod 0600 ~/.ssh/authorized_keys
```

Command 8: The new user is now able to SSH without needing to enter a password every time. Verify everything is set up correctly by using the user to SSH to localhost. After an initial prompt, the Hadoop user is now able to establish an SSH connection to the localhost seamlessly.

```
eakta@eakta:~$ cat ~/.ssh/id_rsa.pub >> ~/.ssh/authorized_keys
eakta@eakta:~$ chmod 0600 ~/.ssh/authorized_keys
eakta@eakta:~$ ssh localhost
The authenticity of host 'localhost (127.0.0.1)' can't be established.
ED25519 key fingerprint is SHA256:nAf1gBuYk8jQ0uGpxAdyqSqm5VJLKS04BPGIJkl95fg.
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.
Welcome to Ubuntu 22.04.4 LTS (GNU/Linux 6.5.0-25-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.

0 updates can be applied immediately.

7 additional security updates can be applied with ESM Apps.
Learn more about enabling ESM Apps service at https://ubuntu.com/esm
```

Command 9: download the Hadoop package with the wget command

```
$ wget https://downloads.apache.org/hadoop/common/hadoop-3.3.6/hadoop-3.3.6-aarch64.tar.gz
eakta@eakta:~$ wget https://downloads.apache.org/hadoop/common/hadoop-3.3.6/hadoop-3.3.6-aarch64.tar.gz
--2024-03-14 22:48:23-- https://downloads.apache.org/hadoop/common/hadoop-3.3.6/hadoop-3.3.6-aarch64.tar.gz
Resolving downloads.apache.org (downloads.apache.org)... 88.99.208.237, 2a01:4f8:10a:39da::2, 2a01:4f9:3a:2c57::2
Connecting to downloads.apache.org (downloads.apache.org)|88.99.208.237|:443... connected.
HTTP request sent, awaiting response... 200 OK
Length: 744905783 (710M) [application/x-gzip]
Saving to: 'hadoop-3.3.6-aarch64.tar.gz'

hadoop-3.3.6-aarch64.tar.gz          100%[=====] 710.40M  2.18MB/s   in 12m 6s

2024-03-14 23:00:29 (1003 KB/s) - 'hadoop-3.3.6-aarch64.tar.gz' saved [744905783/744905783]
```

Command 10: Once the download is complete, extract the files to initiate the Hadoop installation. The Hadoop binary files are now located within the *hadoop-3.3.6* directory.

```
$ tar xzf hadoop-3.3.6-aarch64.tar.gz
```

```
eakta@eakta:~$ tar xzf hadoop-3.3.6-aarch64.tar.gz
eakta@eakta:~$ ls
Desktop  Documents  Downloads  hadoop-3.3.6  hadoop-3.3.6-aarch64.tar.gz  Music  Pictures  Public  snap  Templates  Videos
```

PRACTICAL - 3 Aim:

Single node Hadoop deployment (pseudo-distributed mode)

Source Code & Output:

[Configure Hadoop Environment Variables (bashrc)]

Command 1: Edit the `.bashrc` shell configuration file using a text editor of your choice (we will be using nano)

```
$ sudo nano .bashrc
```

```
eakta@eakta:~$ ls
Desktop Documents Downloads hadoop-3.3.6 hadoop-3.3.6-aarch64.tar.gz Music Pictures Public snap Templates Videos
eakta@eakta:~$ cd hadoop-3.3.6
eakta@eakta:~/hadoop-3.3.6$ pwd
/home/eakta/hadoop-3.3.6
eakta@eakta:~/hadoop-3.3.6$ sudo nano .bashrc
```

Define the Hadoop environment variables by adding the following content to the end of the file. Once you add the variables, save(ctrl+s) and exit(ctrl+x) the `.bashrc` file.

```
#Hadoop Related Options
export HADOOP_HOME=/home/eakta/hadoop-3.3.6
export HADOOP_INSTALL=$HADOOP_HOME
export HADOOP_MAPRED_HOME=$HADOOP_HOME
export HADOOP_COMMON_HOME=$HADOOP_HOME
export HADOOP_HDFS_HOME=$HADOOP_HOME
export 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"
```

```
GNU nano 6.2                                .bashrc *
alias ll='ls -alF'
alias la='ls -A'
alias l='ls -CF'

# Add an "alert" alias for long running commands.  Use like so:
# sleep 10; alert
alias alert='notify-send --urgency=low -i "$( [ $? = 0 ] && echo terminal || echo error)" "$(history|tail -n1|sed -e '\''s/^\s*/'/g'\'' -e '\''s/\s*$/\''')"

# Alias definitions.
# You may want to put all your additions into a separate file like
# ~/.bash_aliases, instead of adding them here directly.
# See /usr/share/doc/bash-doc/examples in the bash-doc package.

if [ -f ~/.bash_aliases ]; then
  . ~/.bash_aliases
fi

# 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

#Hadoop Related Options
export HADOOP_HOME=/home/eakta/hadoop-3.3.6
export HADOOP_INSTALL=$HADOOP_HOME
export HADOOP_MAPRED_HOME=$HADOOP_HOME
export HADOOP_COMMON_HOME=$HADOOP_HOME
export HADOOP_HDFS_HOME=$HADOOP_HOME
export 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"
```

Command 2: It is vital to apply the changes to the current running environment

```
$ source ~/.bashrc
```

```
eakta@eakta:~$ source ~/.bashrc
```

[Edit.hadoop-env.sh File]

Command 3: locate the correct Java path. The resulting output provides the path to the Java binary directory.

```
$ which javac
```

Command 4: Use the provided path to find the OpenJDK directory with the following command. The section of the path just before the /bin/javac directory needs to be assigned to the **\$JAVA_HOME** variable.

```
$ readlink -f /usr/bin/javac
```

Command 5: The *hadoop-env.sh* file serves as a master file to configure YARN, HDFS, MapReduce, and Hadoop-related project settings. When setting up a **single node Hadoop cluster**, you need to define which Java implementation is to be utilized. Use the previously created **\$HADOOP_HOME** variable to access the *hadoopenv.sh* file:

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

```
eakta@eakta:~$ which javac  
/usr/bin/javac  
eakta@eakta:~$ readlink -f /usr/bin/javac  
/usr/lib/jvm/java-8-openjdk-amd64/bin/javac  
eakta@eakta:~$ sudo nano $HADOOP_HOME/etc/hadoop/hadoop-env.sh
```

Uncomment the **\$JAVA_HOME** variable (i.e., remove the # sign) and add the full path to the OpenJDK installation on your system. If you have installed the same version as presented in the first part of this tutorial, add the following line: `export JAVA_HOME=/usr/lib/jvm/java-8-openjdk-amd64`

```
GNU nano 6.2                               /home/eakta/hadoop-3.3.6/etc/hadoop/hadoop-env.sh  
## Precedence rules:  
##  
## {yarn-env.sh|hdfs-env.sh} > hadoop-env.sh > hard-coded defaults  
##  
## {YARN_xyz|HDFS_xyz} > HADOOP_xyz > hard-coded defaults  
##  
  
# Many of the options here are built from the perspective that users  
# may want to provide OVERWRITING values on the command line.  
# For example:  
#  
#   JAVA_HOME=/usr/java/testing hdfs dfs -ls  
#  
# Therefore, the vast majority (BUT NOT ALL!) of these defaults  
# are configured for substitution and not append. If append  
# is preferable, modify this file accordingly.  
  
###  
# Generic settings for HADOOP  
###  
  
# Technically, the only required environment variable is JAVA_HOME.  
# All others are optional. However, the defaults are probably not  
# preferred. Many sites configure these options outside of Hadoop,  
# such as in /etc/profile.d  
  
# The java implementation to use. By default, this environment  
# variable is REQUIRED on ALL platforms except OS X!  
export JAVA_HOME=/usr/lib/jvm/java-8-openjdk-amd64
```

[Edit core-site.xml File]

Command 6: Displays information about files in the current directory.

```
$ ls
```

Command 7: Creates a new directory tmpdata

```
$ mkdir tmpdata
```

Command 8: Changes the working directory to tmpdata

```
$ cd tmpdata
```

Command 9: Shows the current working directory's path

```
$ pwd
```

Command 10: will bring the user above one directory.

```
$ cd ..
```

Command 11: by itself will always bring the user to the home directory.

```
$ cd ~
```

Command 12: The *core-site.xml* file defines HDFS and Hadoop core properties. To set up Hadoop in a pseudodistributed mode, you need to **specify the URL** for your NameNode, and the temporary directory Hadoop uses for the map and reduce process. Open the *core-site.xml* file in a text editor:

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

```
eakta@eakta:~$ ls
Desktop  Documents  Downloads  hadoop-3.3.6  hadoop-3.3.6-aarch64.tar.gz  Music  Pictures  Public  snap  Templates  Videos
eakta@eakta:~$ mkdir tmpdata
eakta@eakta:~$ cd tmpdata
eakta@eakta:~/tmpdata$ pwd
/home/eakta/tmpdata
eakta@eakta:~/tmpdata$ cd ..
eakta@eakta:~$ sudo nano $HADOOP_HOME/etc/hadoop/core-site.xml
```

Add the following configuration to override the default values for the temporary directory and add your HDFS URL to replace the default local file system setting:

```
<configuration>
<property>
  <name>hadoop.tmp.dir</name>
  <value>/home/eakta/tmpdata</value>
</property>
<property>
  <name>fs.default.name</name>
  <value>hdfs://127.0.0.1:9000</value>
</property>
</configuration>
```

```

GNU nano 6.2                               /home/eakta/hadoop-3.3.6/etc/hadoop/core-site.xml
<?xml version="1.0" encoding="UTF-8"?>
<?xml-stylesheet type="text/xsl" href="configuration.xsl"?>
<!--
 Licensed under the Apache License, Version 2.0 (the "License");
 you may not use this file except in compliance with the License.
 You may obtain a copy of the License at

 http://www.apache.org/licenses/LICENSE-2.0

 Unless required by applicable law or agreed to in writing, software
 distributed under the License is distributed on an "AS IS" BASIS,
 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/eakta/tmpdata</value>
</property>
<property>
  <name>fs.default.name</name>
  <value>hdfs://127.0.0.1:9000</value>
</property>
</configuration>

```

[Edit hdfs-site.xml File]

Command 13: Creates a new directory dfsdata

```
$ mkdir dfsdata
```

Command 14: Changes the working directory to dfsdata directory

```
$ cd dfsdata
```

Command 15: Creates a new directory namenode

```
$ mkdir namenode
```

Command 16: Creates a new directory datanode

```
$ mkdir datanode
```

Command 17: will bring the user above one directory.

```
$ cd ..
```

Command 18: The properties in the *hdfs-site.xml* file govern the location for storing node metadata, fsimage file, and edit log file. Configure the file by defining the **NameNode** and **DataNode storage directories**. Additionally, the default **dfs.replication** value of **3** needs to be changed to **1** to match the single node setup. Use the following command to open the *hdfs-site.xml* file for editing:

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

```

eakta@eakta:~$ ls
Desktop  Downloads      hadoop-3.3.6-aarch64.tar.gz  Pictures  snap      tmpdata
Documents  hadoop-3.3.6  Music                      Public    Templates  Videos
eakta@eakta:~$ mkdir dfsdata
eakta@eakta:~$ cd dfsdata
eakta@eakta:~/dfsdata$ mkdir namenode
eakta@eakta:~/dfsdata$ mkdir datanode
eakta@eakta:~/dfsdata$ ls
datanode  namenode
eakta@eakta:~/dfsdata$ cd ..
eakta@eakta:~$ sudo nano $HADOOP_HOME/etc/hadoop/hdfs-site.xml
[sudo] password for eakta:

```

Add the following configuration to the file and, if needed, adjust the NameNode and DataNode directories to your custom locations:

```

<configuration>
<property>
  <name>dfs.data.dir</name>
  <value>/home/eakta/dfsdata/namenode</value>
</property>
<property>
  <name>dfs.data.dir</name>
  <value>/home/eakta/dfsdata/datanode</value>
</property>
<property>
  <name>dfs.replication</name>
  <value>1</value>
</property>
</configuration>

```

```

GNU nano 6.2                               /home/eakta/hadoop-3.3.6/etc/hadoop/hdfs-site.xml
<?xml version="1.0" encoding="UTF-8"?>
<?xml-stylesheet type="text/xsl" href="configuration.xsl"?>
<!--
Licensed under the Apache License, Version 2.0 (the "License");
you may not use this file except in compliance with the License.
You may obtain a copy of the License at

  http://www.apache.org/licenses/LICENSE-2.0

Unless required by applicable law or agreed to in writing, software
distributed under the License is distributed on an "AS IS" BASIS,
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>dfs.data.dir</name>
  <value>/home/eakta/dfsdata/namenode</value>
</property>
<property>
  <name>dfs.data.dir</name>
  <value>/home/eakta/dfsdata/datanode</value>
</property>
<property>
  <name>dfs.replication</name>
  <value>1</value>
</property>
</configuration>

```

[Edit mapred-site.xml File]

Command 18: Use the following command to access the *mapred-site.xml* file and **define MapReduce values**:

```
$ sudo nano $HADOOP_HOME/etc/hadoop/mapred-site.xml
```

```
eakta@eakta:~$ sudo nano $HADOOP_HOME/etc/hadoop/mapred-site.xml  
[sudo] password for eakta:
```

Add the following configuration to change the default MapReduce framework name value to **yarn**:

```
<configuration>  
<property>  
<name>mapreduce.framework.name</name>  
<value>yarn</value>  
</property>  
</configuration>
```

```
GNU nano 6.2                               /home/eakta/hadoop-3.3.6/etc/hadoop/mapred-site.xml  
<?xml version="1.0"?>  
<?xml-stylesheet type="text/xsl" href="configuration.xsl"?>  
<!--  
 Licensed under the Apache License, Version 2.0 (the "License");  
 you may not use this file except in compliance with the License.  
 You may obtain a copy of the License at  
  
 http://www.apache.org/licenses/LICENSE-2.0  
  
 Unless required by applicable law or agreed to in writing, software  
 distributed under the License is distributed on an "AS IS" BASIS,  
 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>mapreduce.framework.name</name>  
<value>yarn</value>  
</property>  
</configuration>
```

[Edit yarn-site.xml File]

Command 19: The *yarn-site.xml* file is used to define settings relevant to **YARN**. It contains configurations for the **Node Manager**, **Resource Manager**, **Containers**, and **Application Master**. Open the *yarn-site.xml* file in a text editor:

```
$ sudo nano $HADOOP_HOME/etc/hadoop/yarn-site.xml
```

```
eakta@eakta:~$ sudo nano $HADOOP_HOME/etc/hadoop/yarn-site.xml
```

Append the following configuration to the 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>

```

```

GNU nano 6.2                               /home/eakta/hadoop-3.3.6/etc/hadoop/yarn-site.xml
<?xml version="1.0"?>
<!--
  Licensed under the Apache License, Version 2.0 (the "License");
  you may not use this file except in compliance with the License.
  You may obtain a copy of the License at

    http://www.apache.org/licenses/LICENSE-2.0

  Unless required by applicable law or agreed to in writing, software
  distributed under the License is distributed on an "AS IS" BASIS,
  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.
-->

<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>

```

[Format HDFS NameNode]

Command 20: It is important to **format the NameNode** before starting Hadoop services for the first time:

```
$ hdfs namenode -format
```

```

eakta@eakta:~$ hdfs namenode -format
WARNING: /home/eakta/hadoop-3.3.6/logs does not exist. Creating.
2024-03-14 23:43:10,569 INFO namenode.NameNode: STARTUP_MSG:
*****STARTUP_MSG: Starting NameNode
STARTUP_MSG: host = eakta/127.0.1.1
STARTUP_MSG: args = [-format]
STARTUP_MSG: version = 3.3.6
STARTUP_MSG: classpath = /home/eakta/hadoop-3.3.6/etc/hadoop:/home/eakta/hadoop-3.3.6/share/hadoop/common/lib/jetty-servlet-9.4.51.v20230217.jar
:/home/eakta/hadoop-3.3.6/share/hadoop/common/lib/netty-codec-stomp-4.1.89.Final.jar:/home/eakta/hadoop-3.3.6/share/hadoop/common/lib/stax2-api-4.

```

The shutdown notification signifies the end of the NameNode format process.

```
2024-03-14 23:43:13,503 INFO namenode.NNStorageRetentionManager: Going to retain 1 images with txid >= 0
2024-03-14 23:43:13,526 INFO namenode.FSNamesystem: Stopping services started for active state
2024-03-14 23:43:13,526 INFO namenode.FSNamesystem: Stopping services started for standby state
2024-03-14 23:43:13,536 INFO namenode.FSIImage: FSImageSaver clean checkpoint: txid=0 when meet shutdown.
2024-03-14 23:43:13,537 INFO namenode.NameNode: SHUTDOWN_MSG:
*****SHUTDOWN_MSG: Shutting down NameNode at eakta/127.0.1.1
*****
```

[Start Hadoop Cluster]

Command 21: Navigate to the *hadoop-3.2.1/sbin* directory and execute the following commands to start the NameNode and DataNode:

```
$ ls
$ cd hadoop-3.3.6
$ cd sbin
$ ./start-dfs.sh
```

The system takes a few moments to initiate the necessary nodes.

```
eakta@eakta: $ ls
Desktop  dfssdata  Documents  Downloads  hadoop-3.3.6  hadoop-3.3.6-aarch64.tar.gz  Music  Pictures  Public  snap  Templates  tmpdata  Videos
eakta@eakta: $ cd hadoop-3.3.6
eakta@eakta:~/hadoop-3.3.6$ cd sbin
eakta@eakta:~/hadoop-3.3.6/sbin$ ./start-dfs.sh
Starting namenodes on [localhost]
Starting datanodes
Starting secondary namenodes [eakta]
2024-03-15 00:14:10,938 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
```

Command 22: Once the namenode, datanodes, and secondary namenode are up and running, start the YARN resource and nodemanagers by typing:

```
$ ./start-yarn.sh
```

The output informs you that the processes are starting.

```
eakta@eakta:~/hadoop-3.3.6/sbin$ ./start-yarn.sh
Starting resourcemanager
Starting nodemanagers
```

Command 23: to check if all the daemons are active and running as Java processes:

```
$ jps
```

If everything is working as intended, the resulting list of running Java processes contains all the HDFS and YARN daemons.

```
eakta@eakta:~/hadoop-3.3.6/sbin$ jps
8851 DataNode
9716 Jps
9352 NodeManager
9032 SecondaryNameNode
9242 ResourceManager
8733 NameNode
```

PRACTICAL - 4 Aim:

Access Hadoop UI from Browser

Source Code & Output:

1. [Use FireFox](#) and navigate to <http://localhost:9870> [Hadoop NameNode UI - provides a comprehensive overview of the entire cluster.]

The screenshot shows the Hadoop NameNode UI Overview page. At the top, there's a navigation bar with tabs: Hadoop (selected), Overview, Datanodes, Datanode Volume Failures, Snapshot, Startup Progress, and Utilities. Below the navigation bar, the main title is "Overview 'localhost:9000' (✓active)". Underneath, there's a table with cluster configuration details:

Started:	Fri Mar 15 00:14:00 +0530 2024
Version:	3.3.6, r1be78238728da9266a4f88195058f08fd012bf9c
Compiled:	Mon Jun 19 04:45:00 +0530 2023 by ubuntu from (HEAD detached at release-3.3.6-RC1)
Cluster ID:	CID-ff06f6d2-42fe-4640-b118-1ba8e715431
Block Pool ID:	BP-1234804171-127.0.1.1-1710439992993

Go to Utilities >> Browse Directory

The screenshot shows the Hadoop NameNode UI Browse Directory page. At the top, there's a navigation bar with tabs: Hadoop (selected), Overview, Datanodes, Datanode Volume Failures, Snapshot, Startup Progress, and Utilities. Below the navigation bar, the main title is "Browse Directory". The page features a table header with columns: Permission, Owner, Group, Size, Last Modified, Replication, Block Size, and Name. A message "No data available in table" is displayed. At the bottom, it says "Showing 0 to 0 of 0 entries".

2. DataNodes → <http://localhost:9864>

DataNode Information × +

localhost:9864/datanode.html

Hadoop Overview Utilities ▾

DataNode on eakta:9866

Cluster ID:	CID-ff06f6d2-42fe-4640-b118-1baf8e715431
Started:	Fri Mar 15 00:14:06 +0530 2024
Version:	3.3.6, r1be78238728da9266a4f88195058f08fd012bf9c

Block Pools

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-1234804171-127.0.1.1-1710439992993	RUNNING	0s	0s	9 minutes	0 B (128 MB)

Volume Information

Directory	StorageType	Capacity Used	Capacity Left	Capacity Reserved	Reserved Space for Replicas	Blocks
/home/eakta/dfsdata/datanode	DISK	24 KB	5.89 GB	0 B	0 B	0

3. YARN Resource Manager → <http://localhost:8088>

All Applications × +

localhost:8088/cluster

90% ☆

 All Applications

Cluster Metrics

Apps Submitted	Apps Pending	Apps Running	Apps Completed	Containers Running	Used Resources	Total
0	0	0	0	0	<memory:0 B, vCores:0>	<memory:8 GB, vCores:0>

Cluster Nodes Metrics

Active Nodes	Decommissioning Nodes	Decommissioned Nodes	Lost Nodes
1	0	0	0

Scheduler Metrics

Scheduler Type	Scheduling Resource Type	Minimum Allocation	Maximum Allocation
Capacity Scheduler	[memory-mb (unit=Mi), vcores]	<memory:1024, vCores:1>	<memory:8192, vCores:4>

Show 20 entries

ID	User	Name	Application Type	Application Tags	Queue	Application Priority	StartTime	LaunchTime	FinishTime	State	FinalStatus	Running Containers	Allocated CPU Vcores
No data available in table													

Showing 0 to 0 of 0 entries

PRACTICAL - 5 Aim:

Perform Hands-on basic Hadoop commands

Source Code & Output:

Command 1: mkdir (typically used to create directories / folders)
\$ ls

\$ hadoop fs -mkdir /first (Creates a directory in Hadoop File System)

hadoop fs -mkdir <HDFS directory>

\$ hdfs dfs -mkdir /second (Creates a directory in specified HDFS location)

hdfs dfs -mkdir <HDFS directory>

```
eakta@eakta:~$ ls
Desktop   Downloads          Music    snap      Videos
dfsdta   hadoop-3.3.6       Pictures  Templates
Documents  hadoop-3.3.6-aarch64.tar.gz  Public    tmpdata
eakta@eakta:~$ hadoop fs -mkdir /first
2024-03-22 15:32:09,661 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your p
latform... using builtin-java classes where applicable
eakta@eakta:~$ hdfs dfs -mkdir /second
2024-03-22 15:32:41,648 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your p
latform... using builtin-java classes where applicable
```

Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name
drwxr-xr-x	eakta	supergroup	0 B	Mar 22 15:32	0	0 B	first
drwxr-xr-x	eakta	supergroup	0 B	Mar 22 15:32	0	0 B	second

Showing 1 to 2 of 2 entries

Hadoop, 2023.

Command 2: rmdir (used to remove empty directories)

\$ hadoop fs -rmdir /first (Delete a directory in Hadoop File System)

hadoop fs -rmdir <HDFS directory>

```
eakta@eakta:~$ hadoop fs -rmdir /first
2024-03-22 15:35:07,300 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your p
latform... using builtin-java classes where applicable
eakta@eakta:~$
```

[localhost:9870/explorer.html#/](#)

Hadoop Overview Datanodes Datanode Volume Failures Snapshot Startup Progress Utilities ▾

Browse Directory

/

Show 25 entries Search:

Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name
drwxr-Xr-X	eakta	supergroup	0 B	Mar 22 15:32	0	0 B	second

Showing 1 to 1 of 1 entries Previous 1 Next

Hadoop, 2023.

Command 3: put (to copy files from the local system to a remote location)

```
$ hadoop fs -put tmpdata / ( Copies the file from local file system to HDFS )
```

```
hadoop fs -put <local directory> <HDFS directory>
```

```
eakta@eakta:~$ hadoop fs -put tmpdata /
2024-03-22 15:47:28,270 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your p
latform... using builtin-java classes where applicable
eakta@eakta:~$
```

[localhost:9870/explorer.html#/](#)

Hadoop Overview Datanodes Datanode Volume Failures Snapshot Startup Progress Utilities ▾

Browse Directory

/

Show 25 entries Search:

Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name
drwxr-Xr-X	eakta	supergroup	0 B	Mar 22 15:32	0	0 B	second
drwxr-Xr-X	eakta	supergroup	0 B	Mar 22 15:47	0	0 B	tmpdata

Showing 1 to 2 of 2 entries Previous 1 Next

Hadoop, 2023.

```
$ hdfs fs -put tmpdata /second/
```

```
eakta@eakta:~$ hadoop fs -put tmpdata /second/
2024-03-22 15:49:34,840 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your p
latform... using builtin-java classes where applicable
eakta@eakta:~$
```

[localhost:9870/explorer.html#/second](#)

Hadoop Overview Datanodes Datanode Volume Failures Snapshot Startup Progress Utilities ▾

Browse Directory

/second/

Show 25 entries Search:

Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name
drwxr-Xr-X	eakta	supergroup	0 B	Mar 22 15:49	0	0 B	tmpdata

Showing 1 to 1 of 1 entries Previous 1 Next

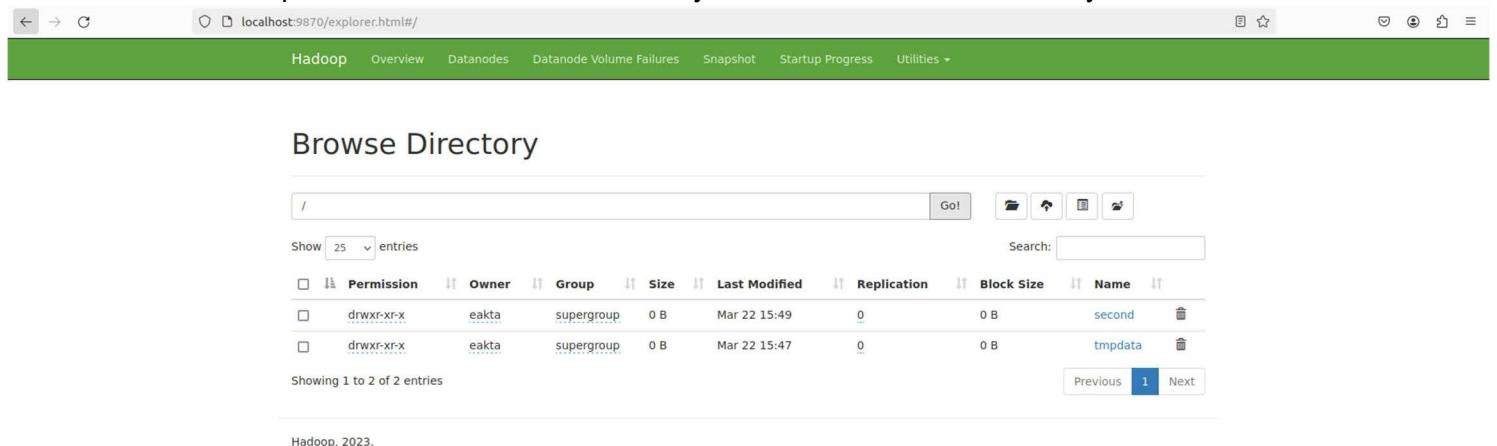
Hadoop, 2023.

Command 4: mv (used to move files or directories from one location to another)

```
$ hadoop fs -mv /second /tmpdata ( used to move a file from one location in HDFS to another location in
```

HDFS)

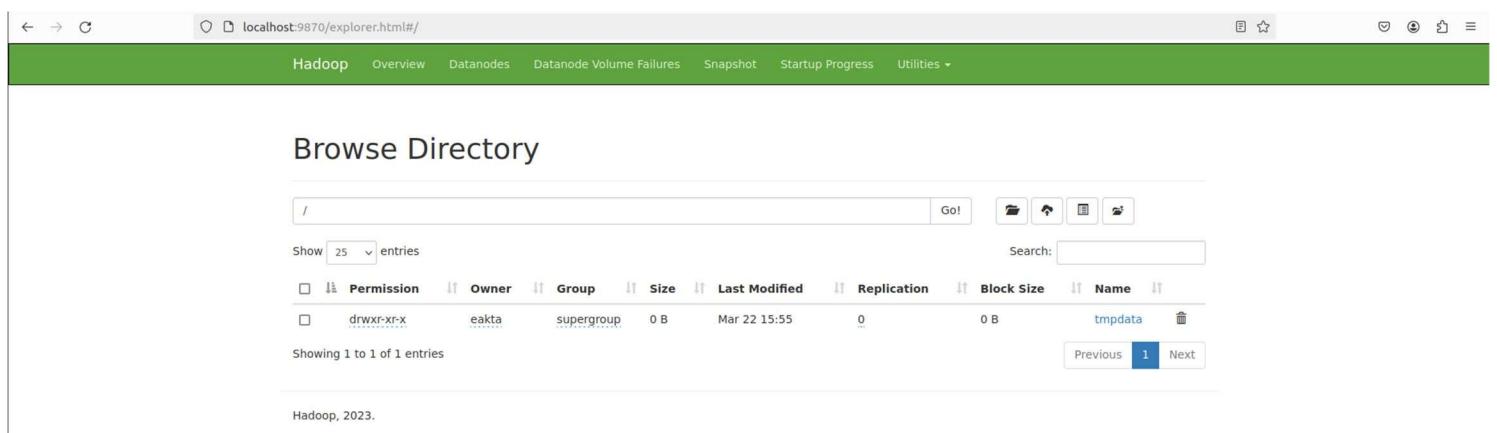
hadoop fs -mv <source HDFS directory> <destination HDFS directory>



The screenshot shows the HDFS browser interface at the root path (/). It displays two entries: 'second' and 'tmpdata'. Both entries have a permission of 'drwxr-xr-x', owned by 'eakta' and belong to the 'supergroup'. The size is 0 B and they were last modified on Mar 22 15:47. The replication factor is 0 and the block size is 0 B. The 'Name' column lists 'second' and 'tmpdata' respectively. Below the table, it says 'Showing 1 to 2 of 2 entries'. At the bottom, there is a note: 'Hadoop, 2023.'

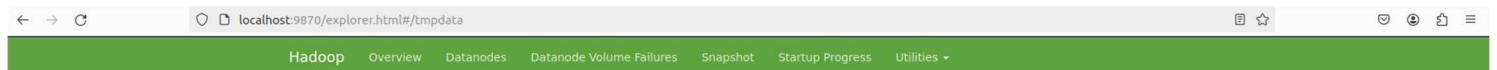
	Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name
□	drwxr-xr-x	eakta	supergroup	0 B	Mar 22 15:49	0	0 B	second
□	drwxr-xr-x	eakta	supergroup	0 B	Mar 22 15:47	0	0 B	tmpdata

```
eakta@eakta:~$ hadoop fs -mv /second /tmpdata
2024-03-22 15:55:14,593 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
eakta@eakta:~$
```



The screenshot shows the HDFS browser interface at the root path (/). It now displays only one entry: 'tmpdata'. This entry has a permission of 'drwxr-xr-x', owned by 'eakta' and belongs to the 'supergroup'. The size is 0 B and it was last modified on Mar 22 15:55. The replication factor is 0 and the block size is 0 B. The 'Name' column lists 'tmpdata'. Below the table, it says 'Showing 1 to 1 of 1 entries'. At the bottom, there is a note: 'Hadoop, 2023.'

	Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name
□	drwxr-xr-x	eakta	supergroup	0 B	Mar 22 15:55	0	0 B	tmpdata



The screenshot shows the HDFS browser interface at the '/tmpdata' path. It displays three entries: 'dfs', 'nm-local-dir', and 'second'. All three entries have a permission of 'drwxr-xr-x', owned by 'eakta' and belong to the 'supergroup'. The size is 0 B and they were last modified on Mar 22 15:47 or 15:49. The replication factor is 0 and the block size is 0 B. The 'Name' column lists 'dfs', 'nm-local-dir', and 'second' respectively. Below the table, it says 'Showing 1 to 3 of 3 entries'. At the bottom, there is a note: 'Hadoop, 2023.'

	Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name
□	drwxr-xr-x	eakta	supergroup	0 B	Mar 22 15:47	0	0 B	dfs
□	drwxr-xr-x	eakta	supergroup	0 B	Mar 22 15:47	0	0 B	nm-local-dir
□	drwxr-xr-x	eakta	supergroup	0 B	Mar 22 15:49	0	0 B	second

Browse Directory

/tmpdata	Go!					
Show 25 entries	Search:					
	Permission	Owner	Group	Size	Last Modified	Replication
□	drwxr-xr-x	eakta	supergroup	0 B	Mar 22 15:47	0
□	drwxr-xr-x	eakta	supergroup	0 B	Mar 22 15:47	0
□	drwxr-xr-x	eakta	supergroup	0 B	Mar 22 15:49	0

Showing 1 to 3 of 3 entries

Previous 1 Next

Hadoop, 2023.

Command 5: get (used to copy files from a remote location to the local system)

\$ hadoop fs -ls / (Lists all the files and directories given for the hdfs destination path)

hadoop fs -ls <HDFS directory>

\$ hadoop fs -get /five ~/ (Copies the file from HDFS to local file system)

hadoop fs -get <HDFS directory> <local directory>

```

eakta@eakta:~$ ls
Desktop Documents eakta      hadoop-3.3.6-aarch64.tar.gz Pictures Public Templates Videos
dfsdata Downloads hadoop-3.3.6 Music          poem.txt snap    tmpdata
eakta@eakta:~$ hadoop fs -ls /
2024-03-22 16:32:36,435 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java class
es where applicable
Found 3 items
drwxr-xr-x  - eakta supergroup          0 2024-03-22 15:59 /eakta
-rw-r--r--  1 eakta supergroup        625 2024-03-22 16:28 /poem.txt
drwxr-xr-x  - eakta supergroup          0 2024-03-22 15:55 /tmpdata
eakta@eakta:~$ hadoop fs -mkdir /four
2024-03-22 16:38:24,095 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java class
es where applicable
eakta@eakta:~$ hadoop fs -mkdir /five
2024-03-22 16:38:41,825 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java class
es where applicable
eakta@eakta:~$ hadoop fs -ls /
2024-03-22 16:38:55,990 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java class
es where applicable
Found 5 items
drwxr-xr-x  - eakta supergroup          0 2024-03-22 15:59 /eakta
drwxr-xr-x  - eakta supergroup          0 2024-03-22 16:38 /five
drwxr-xr-x  - eakta supergroup          0 2024-03-22 16:38 /four
-rw-r--r--  1 eakta supergroup        625 2024-03-22 16:28 /poem.txt
drwxr-xr-x  - eakta supergroup          0 2024-03-22 15:55 /tmpdata
eakta@eakta:~$ hadoop fs -get /five ~/
2024-03-22 16:39:33,948 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java class
es where applicable
eakta@eakta:~$ ls
Desktop Documents eakta hadoop-3.3.6           Music     poem.txt  snap    tmpdata
dfsdata Downloads five   hadoop-3.3.6-aarch64.tar.gz Pictures  Public   Templates Videos
eakta@eakta:~$
```

Name	Block Size	Replication	Last Modified	Size	Group	Owner	Permission
eakta	0 B	0	Mar 22 15:59	0 B	supergroup	ekta	drwxr-xr-x
five	0 B	0	Mar 22 16:38	0 B	supergroup	ekta	drwxr-xr-x
four	0 B	0	Mar 22 16:38	0 B	supergroup	ekta	drwxr-xr-x
poem.txt	128 MB	1	Mar 22 16:28	625 B	supergroup	ekta	-rw-r--r--
tmpdata	0 B	0	Mar 22 15:55	0 B	supergroup	ekta	drwxr-xr-x

Browse Directory

/									Go!				
Show 25 entries		Search: <input type="text"/>											
	Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name					
<input type="checkbox"/>	drwxr-xr-x	ekta	supergroup	0 B	Mar 22 15:59	0	0 B	eakta					
<input type="checkbox"/>	drwxr-xr-x	ekta	supergroup	0 B	Mar 22 16:38	0	0 B	five					
<input type="checkbox"/>	drwxr-xr-x	ekta	supergroup	0 B	Mar 22 16:38	0	0 B	four					
<input type="checkbox"/>	-rw-r--r--	ekta	supergroup	625 B	Mar 22 16:28	1	128 MB	poem.txt					
<input type="checkbox"/>	drwxr-xr-x	ekta	supergroup	0 B	Mar 22 15:55	0	0 B	tmpdata					

Showing 1 to 5 of 5 entries

Hadoop, 2023.

Command 6: copyFromLocal (used in Hadoop's File System (HDFS) to copy files or directories from the local file system to the Hadoop distributed file system)

```
$ hadoop fs -copyFromLocal eakta /
```

```
hadoop fs -copyFromLocal <local directory> <HDFS directory>
```

```

eakta@eakta:~$ ls
Desktop Documents hadoop-3.3.6           Music     Public  Templates  Videos
dfsdata Downloads hadoop-3.3.6-aarch64.tar.gz Pictures snap    tmpdata
eakta@eakta:~$ mkdir eakta
eakta@eakta:~$ ls
Desktop Documents eakta      hadoop-3.3.6-aarch64.tar.gz Pictures snap    tmpdata
dfsdata Downloads hadoop-3.3.6 Music          Public    Templates Videos
eakta@eakta:~$ hadoop fs -copyFromLocal eakta /
2024-03-22 15:59:22,696 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your p
latform... using builtin-java classes where applicable
eakta@eakta:~$
```

Browse Directory

/											<input type="button" value="Go!"/>				
Show 25 entries											Search:				
	Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name							
<input type="checkbox"/>	drwxr-xr-x	eakta	supergroup	0 B	Mar 22 15:59	0	0 B	eakta							
<input type="checkbox"/>	drwxr-xr-x	eakta	supergroup	0 B	Mar 22 15:55	0	0 B	tmpdata							

Showing 1 to 2 of 2 entries

[Previous](#) [1](#) [Next](#)

Hadoop, 2023.

Command 7: cat (used to concatenate and display the contents of one or more files)

\$ gedit poem.txt (used to open the file named "poem.txt" in the default text editor)

Then Enter the contents for poem.txt

gedit <filename>

\$ hadoop fs -put poem.txt / (used to copy the file named "poem.txt" from the local file system to the root directory ("/") of the Hadoop distributed file system (HDFS))

hadoop fs -put <local directory> <HDFS directory>

\$ hadoop fs -cat /poem.txt (used to display the contents of files stored in HDFS)

hadoop fs -cat <HDFS directory>

```
eakta@eakta:~$ hadoop fs -ls /
2024-03-22 16:24:13,043 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Found 2 items
drwxr-xr-x - eakta supergroup 0 2024-03-22 15:59 /eakta
drwxr-xr-x - eakta supergroup 0 2024-03-22 15:55 /tmpdata
eakta@eakta:~$ hadoop fs -put poem.txt /
2024-03-22 16:27:56,393 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
eakta@eakta:~$ hadoop fs -ls /
2024-03-22 16:28:27,013 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Found 3 items
drwxr-xr-x - eakta supergroup 0 2024-03-22 15:59 /eakta
-rw-r--r-- 1 eakta supergroup 625 2024-03-22 16:28 /poem.txt
drwxr-xr-x - eakta supergroup 0 2024-03-22 15:55 /tmpdata
eakta@eakta:~$ hadoop fs -cat /poem.txt
2024-03-22 16:28:41,243 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Twinkle, twinkle, little star,
How I wonder what you are.
Up above the world so high,
Like a diamond in the sky.

When the blazing sun is gone,
When he nothing shines upon,
Then you show your little light,
Twinkle, twinkle, all the night.

Then the traveller in the dark,
Thanks you for your tiny spark,
He could not see which way to go,
If you did not twinkle so.

In the dark blue sky you keep,
And often through my curtains peep,
For you never shut your eye,
'Till the sun is in the sky.

As your bright and tiny spark,
Lights the traveller in the dark,-
Though I know not what you are,
Twinkle, twinkle, little star.
eakta@eakta:~$
```

Command 8: copyToLocal (used in Hadoop's File System (HDFS) to copy files or directories from the Hadoop distributed file system to the local file system)

```
$ hadoop fs -copyToLocal / four
```

```
hadoop fs -copyToLocal <HDFS directory> <Local directory>
```

```
eakta@eakta:~$ ls
Desktop Documents eakta hadoop-3.3.6          Music   poem.txt snap     tmpdata
dfsdata Downloads five hadoop-3.3.6-aarch64.tar.gz Pictures Public   Templates Videos
eakta@eakta:~$ hadoop fs -ls /
2024-03-22 16:43:57,226 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Found 5 items
drwxr-xr-x  - eakta supergroup      0 2024-03-22 15:59 /eakta
drwxr-xr-x  - eakta supergroup      0 2024-03-22 16:38 /five
drwxr-xr-x  - eakta supergroup      0 2024-03-22 16:38 /four
-rw-r--r--  1 eakta supergroup    625 2024-03-22 16:28 /poem.txt
drwxr-xr-x  - eakta supergroup      0 2024-03-22 15:55 /tmpdata
eakta@eakta:~$ hadoop fs -copyToLocal / four
2024-03-22 16:48:02,567 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
eakta@eakta:~$ ls
Desktop Documents eakta four      hadoop-3.3.6-aarch64.tar.gz Pictures Public   Templates Videos
dfsdata Downloads five    hadoop-3.3.6 Music       poem.txt snap     tmpdata
eakta@eakta:~$
```

Command 9: moveFromLocal (used to move files or directories from the local file system to the Hadoop distributed file system (HDFS))

```
$ hadoop fs -moveFromLocal geet /
```

```
hadoop fs -moveFromLocal <Local directory> <HDFS directory>
```

```
eakta@eakta:~$ ls
'big data'  Downloads  geet                      Music   snap
Desktop     eakta      hadoop-3.3.6              Pictures Templates
dfsdata     five       hadoop-3.3.6-aarch64.tar.gz poem.txt tmpdata
Documents   four      hello                     Public   Videos
eakta@eakta:~$ hadoop fs -ls /
2024-03-23 11:24:26,826 WARN util.NativeCodeLoader: Unable to load native-hadoop
library for your platform... using builtin-java classes where applicable
Found 5 items
drwxr-xr-x  - eakta supergroup      0 2024-03-22 15:59 /eakta
drwxr-xr-x  - eakta supergroup      0 2024-03-22 16:38 /five
drwxr-xr-x  - eakta supergroup      0 2024-03-22 16:38 /four
-rw-r--r--  1 eakta supergroup    625 2024-03-22 16:28 /poem.txt
drwxr-xr-x  - eakta supergroup      0 2024-03-22 15:55 /tmpdata
eakta@eakta:~$ hadoop fs -moveFromLocal geet /
2024-03-23 11:25:10,473 WARN util.NativeCodeLoader: Unable to load native-hadoop
library for your platform... using builtin-java classes where applicable
eakta@eakta:~$
```

Hadoop Overview Datanodes Datanode Volume Failures Snapshot Startup Progress Utilities ▾

Browse Directory

/

Show 25 entries Search:

<input type="checkbox"/>	Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name	<input type="button" value="Delete"/>
<input type="checkbox"/>	drwxr-xr-x	eakta	supergroup	0 B	Mar 22 15:59	0	0 B	eakta	<input type="button" value="Delete"/>
<input type="checkbox"/>	drwxr-xr-x	eakta	supergroup	0 B	Mar 22 16:38	0	0 B	five	<input type="button" value="Delete"/>
<input type="checkbox"/>	drwxr-xr-x	eakta	supergroup	0 B	Mar 22 16:38	0	0 B	four	<input type="button" value="Delete"/>
<input type="checkbox"/>	drwxr-xr-x	eakta	supergroup	0 B	Mar 23 11:25	0	0 B	geet	<input type="button" value="Delete"/>
<input type="checkbox"/>	-rw-r--r--	eakta	supergroup	625 B	Mar 22 16:28	1	128 MB	poem.txt	<input type="button" value="Delete"/>
<input type="checkbox"/>	drwxr-xr-x	eakta	supergroup	0 B	Mar 22 15:55	0	0 B	tmpdata	<input type="button" value="Delete"/>

Showing 1 to 6 of 6 entries

Previous 1 Next

Hadoop, 2023.

Command 10: cp (copy command - used to copy file from one directory to another directory)

\$ hadoop fs -cp /poem.txt /geet/ (copies the file from the source directory in HDFS to the destination directory in HDFS)

hadoop fs -cp <source HDFS directory> <destination HDFS directory>

```
eakta@eakta:~$ hadoop fs -cp /poem.txt /geet/
2024-03-23 11:30:55,982 WARN util.NativeCodeLoader: Unable to load native-hadoop
library for your platform... using builtin-java classes where applicable
eakta@eakta:~$
```

Hadoop Overview Datanodes Datanode Volume Failures Snapshot Startup Progress Utilities ▾

Browse Directory

/geet

Show 25 entries Search:

<input type="checkbox"/>	Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name	<input type="button" value="Delete"/>
<input type="checkbox"/>	-rw-r--r--	eakta	supergroup	625 B	Mar 23 11:31	1	128 MB	poem.txt	<input type="button" value="Delete"/>

Showing 1 to 1 of 1 entries

Previous 1 Next

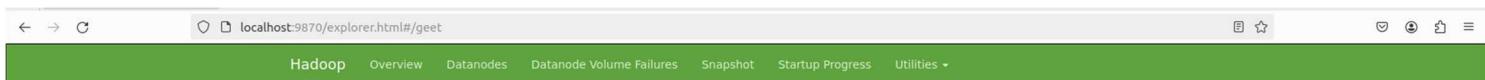
Hadoop, 2023.

Command 11: du (estimate file or directory space usage)

```
$ hadoop fs -du /geet ( estimate the space usage of files and directories stored in HDFS )
```

```
hadoop fs -du <HDFS directory>
```

```
eakta@eakta:~$ hadoop fs -ls /
2024-03-23 11:47:05,602 WARN util.NativeCodeLoader: Unable to load native-hadoop
library for your platform... using builtin-java classes where applicable
Found 6 items
drwxr-xr-x  - eakta supergroup      0 2024-03-22 15:59 /eakta
drwxr-xr-x  - eakta supergroup      0 2024-03-22 16:38 /five
drwxr-xr-x  - eakta supergroup      0 2024-03-22 16:38 /four
drwxr-xr-x  - eakta supergroup      0 2024-03-23 11:31 /geet
-rw-r--r--  1 eakta supergroup    625 2024-03-22 16:28 /poem.txt
drwxr-xr-x  - eakta supergroup      0 2024-03-22 15:55 /tmpdata
eakta@eakta:~$ hadoop fs -du /geet
2024-03-23 11:47:36,205 WARN util.NativeCodeLoader: Unable to load native-hadoop
library for your platform... using builtin-java classes where applicable
625  625  /geet/poem.txt
```



```
$ hadoop fs -du /
```

```
eakta@eakta:~$ hadoop fs -du /
2024-03-23 11:51:21,476 WARN util.NativeCodeLoader: Unable to load native-hadoop
library for your platform... using builtin-java classes where applicable
0      0      /eakta
0      0      /five
0      0      /four
625    625    /geet
625    625    /poem.txt
6296190 6296190 /tmpdata
```

Command 12: stat (displays detailed information about the specified file, including its permissions, size, and timestamps)

```
$ hadoop fs -stat /geet ( provides the stat of the file or directory in HDFS )
```

```
hadoop fs -stat <HDFS directory>
```

It can provide data in the following formats. By default, it uses '%y'.

Format	Description
%b	File size in bytes
%g	Group name of owner
%n	File name

%o	Block size
%r	Replication
%u	User name of owner
%y	Modification date

\$ hadoop fs -stat "%u" /geet

hadoop fs -stat "format" <HDFS directory>

Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name
drwxr-xr-x	eakta	supergroup	0 B	Mar 22 15:59	0	0 B	ekta
drwxr-xr-x	eakta	supergroup	0 B	Mar 22 16:38	0	0 B	five
drwxr-xr-x	eakta	supergroup	0 B	Mar 22 16:38	0	0 B	four
drwxr-xr-x	eakta	supergroup	0 B	Mar 23 11:31	0	0 B	geet
-rw-r--r--	eakta	supergroup	625 B	Mar 22 16:28	1	128 MB	poem.txt
drwxr-xr-x	eakta	supergroup	0 B	Mar 22 15:55	0	0 B	tmpdata

Hadoop, 2023.

```
eakta@eakta:~$ hadoop fs -stat /geet
2024-03-23 12:07:17,053 WARN util.NativeCodeLoader: Unable to load native-hadoop
library for your platform... using builtin-java classes where applicable
2024-03-23 06:01:09
eakta@eakta:~$ hadoop fs -stat "%u" /geet
2024-03-23 12:10:06,089 WARN util.NativeCodeLoader: Unable to load native-hadoop
library for your platform... using builtin-java classes where applicable
eakta
eakta@eakta:~$ hadoop fs -stat "%b" /geet
2024-03-23 12:10:41,825 WARN util.NativeCodeLoader: Unable to load native-hadoop
library for your platform... using builtin-java classes where applicable
0
eakta@eakta:~$
```