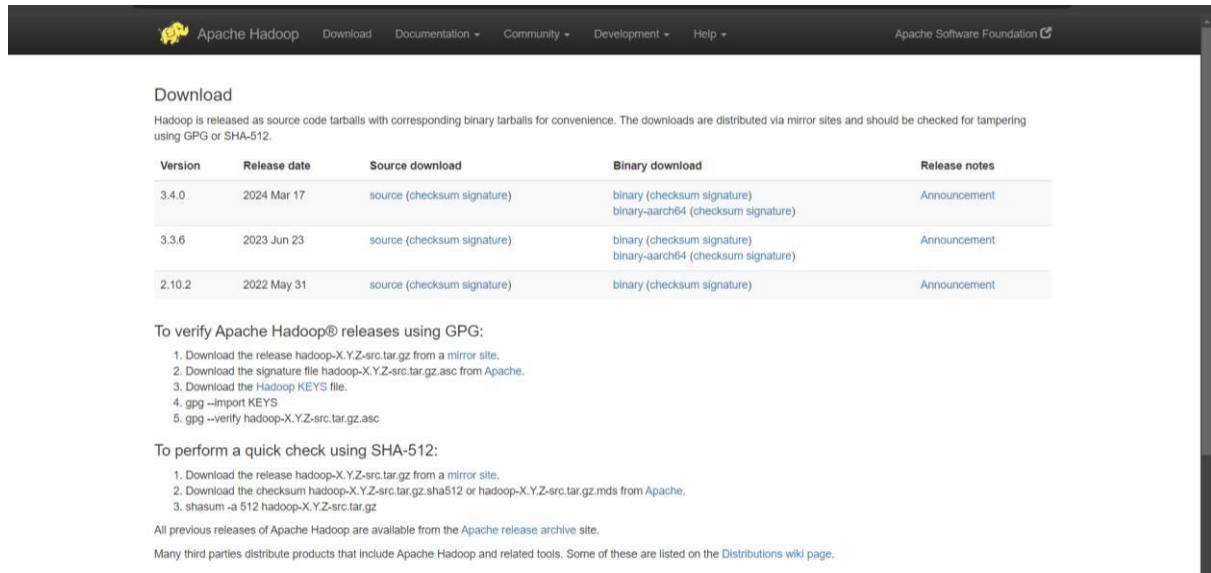


Exp. No : 1

Installation of Hadoop

1. Download Hadoop Binary file



The screenshot shows the Apache Hadoop website's download section. It includes a navigation bar with links for Download, Documentation, Community, Development, and Help. The main content area is titled 'Download' and explains that Hadoop is released as source code tarballs with corresponding binary tarballs. It provides a table of releases with columns for Version, Release date, Source download, Binary download, and Release notes. Below the table, there are instructions on how to verify releases using GPG and SHA-512, and a note about previous releases being available from the Apache release archive site.

Version	Release date	Source download	Binary download	Release notes
3.4.0	2024 Mar 17	source (checksum signature)	binary (checksum signature) binary-aarch64 (checksum signature)	Announcement
3.3.6	2023 Jun 23	source (checksum signature)	binary (checksum signature) binary-aarch64 (checksum signature)	Announcement
2.10.2	2022 May 31	source (checksum signature)	binary (checksum signature)	Announcement

To verify Apache Hadoop® releases using GPG:

1. Download the release `hadoop-X.Y.Z-src.tar.gz` from a [mirror site](#).
2. Download the signature file `hadoop-X.Y.Z-src.tar.gz.asc` from Apache.
3. Download the Hadoop `KEYS` file.
4. `gpg --import KEYS`
5. `gpg --verify hadoop-X.Y.Z-src.tar.gz.asc`

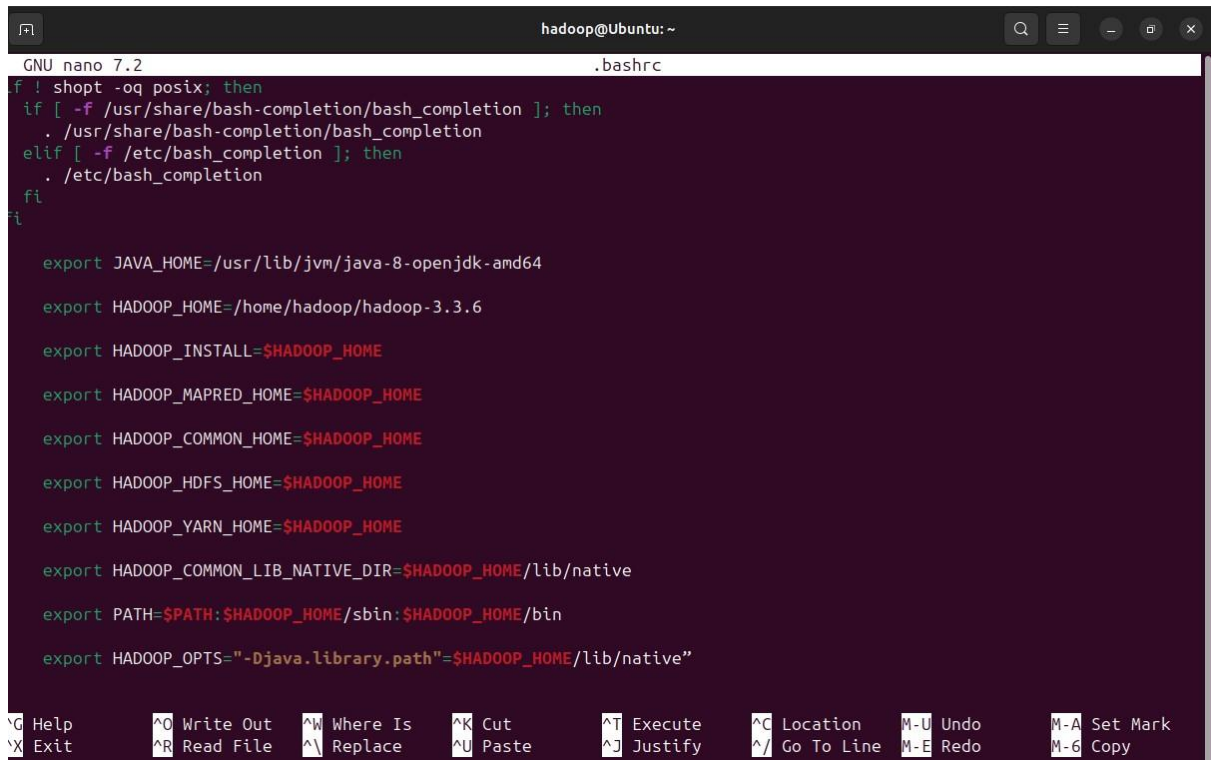
To perform a quick check using SHA-512:

1. Download the release `hadoop-X.Y.Z-src.tar.gz` from a [mirror site](#).
2. Download the checksum `hadoop-X.Y.Z-src.tar.gz.sha512` or `hadoop-X.Y.Z-src.tar.gz.mds` from Apache.
3. `shasum -a 512 hadoop-X.Y.Z-src.tar.gz`

All previous releases of Apache Hadoop are available from the [Apache release archive site](#).

Many third parties distribute products that include Apache Hadoop and related tools. Some of these are listed on the [Distributions wiki page](#).

2. Modifying .bashrc file



The screenshot shows a terminal window with the nano text editor open to the `.bashrc` file. The file contains configuration for Hadoop, including setting `JAVA_HOME`, `HADOOP_HOME`, and various `HADOOP_*_HOME` variables. The `PATH` variable is updated to include the Hadoop bin directory. The `HADOOP_OPTS` variable is also set.

```
GNU nano 7.2 .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 JAVA_HOME=/usr/lib/jvm/java-8-openjdk-amd64

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

3. Updating JAVA_HOME in \$HADOOP_HOME/etc/hadoop/hadoop-env.sh

```

hadoop@Ubuntu: ~/hadoop-3.3.6/etc/hadoop
GNU nano 7.2                                hadoop-env.sh
## {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:
#
export JAVA_HOME=/usr/lib/jvm/java-8-openjdk-amd64
#
# 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=

# Location of Hadoop. By default, Hadoop will attempt to determine
# this location based upon its execution path.

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

```

4. Modifying \$HADOOP_HOME/etc/hadoop/core-site.xml

```

hadoop@Ubuntu: ~/hadoop-3.3.6/etc/hadoop
GNU nano 7.2                                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>fs.defaultFS</name>

    <value>hdfs://localhost:9000</value>

  </property>

</configuration>

[ Read 29 lines ]
^G Help      ^O Write Out  ^W Where Is   ^K Cut        ^T Execute    ^C Location   ^U Undo       ^A Set Mark
^X Exit      ^R Read File  ^\ Replace    ^U Paste      ^J Justify    ^_ Go To Line  ^E Redo       ^M Copy

```

5. Modifying \$HADOOP_HOME/etc/hadoop/hdfs-site.xml

```

hadoop@Ubuntu: ~/hadoop-3.3.6/etc/hadoop
GNU nano 7.2 hdfs-site.xml
</configuration>
->
<configuration>

<property>

<name>dfs.replication</name>

<value>1</value>

</property>

<property> <name>dfs.namenode.name.dir</name>

<value>file:///home/hadoop/hadoopdata/hdfs/namenode</value>

</property>

<property>

<name>dfs.datanode.data.dir</name>

<value>file:///home/hadoop/hadoopdata/hdfs/datanode</value>

</property>

</configuration>

^G Help      ^O Write Out  ^W Where Is   ^K Cut        ^T Execute    ^C Location   M-U Undo      M-A Set Mark
^X Exit      ^R Read File  ^\ Replace    ^U Paste      ^J Justify    ^/_ Go To Line M-E Redo      M-6 Copy

```

6. Modifying \$HADOOP_HOME/etc/hadoop/mapred-site.xml

```

hadoop@Ubuntu: ~/hadoop-3.3.6/etc/hadoop
GNU nano 7.2 mapred-site.xml
<configuration>

<property>

<name>yarn.app.mapreduce.am.env</name>

<value>HADOOP_MAPRED_HOME=$HADOOP_HOME/home/hadoop/hadoop/bin/hadoop</value>

</property>

<property>

<name>mapreduce.map.env</name>

<value>HADOOP_MAPRED_HOME=$HADOOP_HOME/home/hadoop/hadoop/bin/hadoop</value>

</property>

<property>

<name>mapreduce.reduce.env</name>

<value>HADOOP_MAPRED_HOME=$HADOOP_HOME/home/hadoop/hadoop/bin/hadoop</value>

</property>

</configuration>

^G Help      ^O Write Out  ^W Where Is   ^K Cut        ^T Execute    ^C Location   M-U Undo      M-A Set Mark
^X Exit      ^R Read File  ^\ Replace    ^U Paste      ^J Justify    ^/_ Go To Line M-E Redo      M-6 Copy

```

7. Modifying \$HADOOP_HOME/etc/hadoop/yarn-site.xml


```

hadoop@Ubuntu: ~/hadoop-3.3.6/etc/hadoop
GNU nano 7.2                                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>

  </configuration>

[ Read 24 lines ]
^G Help      ^O Write Out ^W Where Is  ^K Cut       ^T Execute   ^C Location  M-U Undo     M-A Set Mark
^X Exit      ^R Read File ^_ Replace   ^U Paste     ^J Justify   ^_ Go To Line M-E Redo     M-6 Copy

```

8. Execute hdfs namenode -format

```

hadoop@Ubuntu: ~/hadoop-3.3.6/sbin
hadoop@Ubuntu:~/hadoop-3.3.6/sbin$ hdfs namenode -format
2024-08-15 16:44:36,005 INFO namenode.NameNode: STARTUP_MSG:
/*****
STARTUP_MSG: Starting NameNode
STARTUP_MSG:   host = Ubuntu/127.0.1.1
STARTUP_MSG:   args = [-format]
STARTUP_MSG:   version = 3.3.6
STARTUP_MSG:   classpath = /home/hadoop/hadoop-3.3.6/etc/hadoop:/home/hadoop/hadoop-3.3.6/share/hadoop/common/lib/common
s-beanutils-1.9.4.jar:/home/hadoop/hadoop-3.3.6/share/hadoop/common/lib/kerby-asn1-1.0.1.jar:/home/hadoop/hadoop-3.3.6/s
hare/hadoop/common/lib/httpclient-4.5.13.jar:/home/hadoop/hadoop-3.3.6/share/hadoop/common/lib/netty-resolver-dns-native
-macos-4.1.89.Final.jar:/home/hadoop/hadoop-3.3.6/share/hadoop/common/lib/jsp-api-2.1.jar:/home/hadoop/hadoop-3.3.6/sha
re/hadoop/common/lib/zookeeper-jute-3.6.3.jar:/home/hadoop/hadoop-3.3.6/share/hadoop/common/lib/netty-transport
-sctp-4.1.89.Final.jar:/home/hadoop/hadoop-3.3.6/share/hadoop/common/lib/listenablefuture-9999.0-empty-to-avoid-conflic
t-with-guava.jar:/home/hadoop/hadoop-3.3.6/share/hadoop/common/lib/kerb-identity-1.0.1.jar:/home/hadoop/hadoop-3.3.6/sha
re/hadoop/common/lib/jetty-util-ajax-9.4.51.v20230217.jar:/home/hadoop/hadoop-3.3.6/share/hadoop/common/lib/netty-codec
redis-4.1.89.Final.jar:/home/hadoop/hadoop-3.3.6/share/hadoop/common/lib/jaxb-impl-2.2.3-1.jar:/home/hadoop/hadoop-3.3.6
/share/hadoop/common/lib/jackson-core-2.12.7.jar:/home/hadoop/hadoop-3.3.6/share/hadoop/common/lib/netty-common-4.1.89.F
inal.jar:/home/hadoop/hadoop-3.3.6/share/hadoop/common/lib/checker-qual-2.5.2.jar:/home/hadoop/hadoop-3.3.6/share/hadoop
/common/lib/netty-all-4.1.89.Final.jar:/home/hadoop/hadoop-3.3.6/share/hadoop/common/lib/kerb-admin-1.0.1.jar:/home/hado
op/hadoop-3.3.6/share/hadoop/common/lib/jetty-servlet-9.4.51.v20230217.jar:/home/hadoop/hadoop-3.3.6/share/hadoop/common
/lib/netty-handler-4.1.89.Final.jar:/home/hadoop/hadoop-3.3.6/share/hadoop/common/lib/javax.servlet-api-3.1.0.jar:/home
hadoop/hadoop-3.3.6/share/hadoop/common/lib/jackson-databind-2.12.7.1.jar:/home/hadoop/hadoop-3.3.6/share/hadoop/common
/lib/jersey-json-1.20.jar:/home/hadoop/hadoop-3.3.6/share/hadoop/common/lib/token-provider-1.0.1.jar:/home/hadoop/hadoop-
3.3.6/share/hadoop/common/lib/stax2-api-4.2.1.jar:/home/hadoop/hadoop-3.3.6/share/hadoop/common/lib/curator-recipes-5.2.
0.jar:/home/hadoop/hadoop-3.3.6/share/hadoop/common/lib/animal-sniffer-annotations-1.17.jar:/home/hadoop/hadoop-3.3.6/sh
are/hadoop/common/lib/jettison-1.5.4.jar:/home/hadoop/hadoop-3.3.6/share/hadoop/common/lib/jackson-core-asl-1.9.13.jar:/
home/hadoop/hadoop-3.3.6/share/hadoop/common/lib/netty-codec-memcache-4.1.89.Final.jar:/home/hadoop/hadoop-3.3.6/share/h
adoop/common/lib/metrics-core-3.2.4.jar:/home/hadoop/hadoop-3.3.6/share/hadoop/common/lib/hadoop-shaded-protobuf_3_7-1.1
.1.jar:/home/hadoop/hadoop-3.3.6/share/hadoop/common/lib/netty-handler-ssl-ocsp-4.1.89.Final.jar:/home/hadoop/hadoop-3.3
.6/share/hadoop/common/lib/netty-transport-rxtx-4.1.89.Final.jar:/home/hadoop/hadoop-3.3.6/share/hadoop/common/lib/kerb
-util-1.0.1.jar:/home/hadoop/hadoop-3.3.6/share/hadoop/common/lib/commons-compress-1.21.jar:/home/hadoop/hadoop-3.3.6/sha
re/hadoop/common/lib/netty-codec-4.1.89.Final.jar:/home/hadoop/hadoop-3.3.6/share/hadoop/common/lib/jetty-security-9.4.5
1.v20230217.jar:/home/hadoop/hadoop-3.3.6/share/hadoop/common/lib/oson-2.9.0.jar:/home/hadoop/hadoop-3.3.6/share/hadoop/

```

```

hadoop@Ubuntu: ~/hadoop-3.3.6/sbin
2024-08-15 16:44:40,084 INFO metrics.TopMetrics: NNTop conf: dfs.namenode.top.num.users = 10
2024-08-15 16:44:40,084 INFO metrics.TopMetrics: NNTop conf: dfs.namenode.top.windows.minutes = 1,5,25
2024-08-15 16:44:40,121 INFO namenode.FSNamesystem: Retry cache on namenode is enabled
2024-08-15 16:44:40,121 INFO namenode.FSNamesystem: Retry cache will use 0.03 of total heap and retry cache entry expiry
time is 600000 millis
2024-08-15 16:44:40,146 INFO util.GSet: Computing capacity for map NameNodeRetryCache
2024-08-15 16:44:40,150 INFO util.GSet: VM type = 64-bit
2024-08-15 16:44:40,153 INFO util.GSet: 0.029999999329447746% max memory 871.5 MB = 267.7 KB
2024-08-15 16:44:40,153 INFO util.GSet: capacity = 2^15 = 32768 entries
Re-format filesystem in Storage Directory root= /home/hadoop/hadoopdata/hdfs/namenode; location= null ? (Y or N) Y
2024-08-15 16:44:43,736 INFO namenode.FSImage: Allocated new BlockPoolId: BP-1847114295-127.0.1.1-1723720483718
2024-08-15 16:44:43,736 INFO common.Storage: Will remove files: [/home/hadoop/hadoopdata/hdfs/namenode/current/edits_inp
rogress_00000000000000000002, /home/hadoop/hadoopdata/hdfs/namenode/current/fsimage_00000000000000000001.md5, /home/hadoop
/hadoopdata/hdfs/namenode/current/fsimage_00000000000000000000.md5, /home/hadoop/hadoopdata/hdfs/namenode/current/VERSION
, /home/hadoop/hadoopdata/hdfs/namenode/current/fsimage_00000000000000000001, /home/hadoop/hadoopdata/hdfs/namenode/curre
nt/seen_txid, /home/hadoop/hadoopdata/hdfs/namenode/current/fsimage_00000000000000000000, /home/hadoop/hadoopdata/hdfs/na
menode/current/edits_00000000000000000001-00000000000000000001]
2024-08-15 16:44:43,917 INFO common.Storage: Storage directory /home/hadoop/hadoopdata/hdfs/namenode has been successfu
ly formatted.
2024-08-15 16:44:44,225 INFO namenode.FSImageFormatProtobuf: Saving image file /home/hadoop/hadoopdata/hdfs/namenode/cur
rent/fsimage.ckpt_00000000000000000000 using no compression
2024-08-15 16:44:44,577 INFO namenode.FSImageFormatProtobuf: Image file /home/hadoop/hadoopdata/hdfs/namenode/current/fs
image.ckpt_00000000000000000000 of size 401 bytes saved in 0 seconds .
2024-08-15 16:44:44,630 INFO namenode.NNStorageRetentionManager: Going to retain 1 images with txid >= 0
2024-08-15 16:44:44,661 INFO namenode.FSNamesystem: Stopping services started for active state
2024-08-15 16:44:44,661 INFO namenode.FSNamesystem: Stopping services started for standby state
2024-08-15 16:44:44,675 INFO namenode.FSImage: FSImageSaver clean checkpoint: txid=0 when meet shutdown.
2024-08-15 16:44:44,685 INFO namenode.NameNode: SHUTDOWN_MSG:
/*****
SHUTDOWN_MSG: Shutting down NameNode at Ubuntu/127.0.1.1
*****/
hadoop@Ubuntu:~/hadoop-3.3.6/sbin$

```

9. Execute start-dfs.sh

```

hadoop@Ubuntu:~/hadoop-3.3.6/sbin$ start-dfs.sh
Starting namenodes on [localhost]
Starting datanodes
Starting secondary namenodes [Ubuntu]
2024-08-15 16:47:24,236 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using buil
tin-java classes where applicable

```

10. Execute start-yarn.sh

```

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

```

11. Execute jps command

```

hadoop@Ubuntu:~/hadoop-3.3.6/sbin$ jps
4758 NodeManager
4390 SecondaryNameNode
5111 Jps
4632 ResourceManager
4015 NameNode

```

12. Open <http://localhost:9870>

Overview 'localhost:9000' (✓active)

Started:	Thu Aug 15 16:47:15 +0530 2024
Version:	3.3.6, r1be78238728da9266a4f88195058f08fd012bf9c
Compiled:	Sun Jun 18 13:52:00 +0530 2023 by ubuntu from (HEAD detached at release-3.3.6-RC1)
Cluster ID:	CID-4797107d-bd90-4038-afed-e5c495ecd59b
Block Pool ID:	BP-1847114295-127.0.1.1-1723720483718

Summary

Security is off.
 Safemode is off.
 1 files and directories, 0 blocks (0 replicated blocks, 0 erasure coded block groups) = 1 total filesystem object(s).
 Heap Memory used 79.95 MB of 261 MB Heap Memory. Max Heap Memory is 871.5 MB.
 Non Heap Memory used 47.19 MB of 48.65 MB Committed Non Heap Memory. Max Non Heap Memory is <unbounded>.

Configured Capacity:	0 B
-----------------------------	-----

13. Open <http://localhost:8088>

Cluster

- About
- Nodes
- Node Labels
- Applications
 - NEW
 - NEW SAVING
 - SUBMITTED
 - ACCEPTED
 - RUNNING
 - FINISHED
 - FAILED
 - KILLED
- Scheduler
- Tools

Cluster Metrics

Apps Submitted	Apps Pending	Apps Running	Apps Completed	Containers Running
0	0	0	0	0

Cluster Nodes Metrics

Active Nodes	Decommissioning Nodes	Decommissioned Nodes
1	0	0

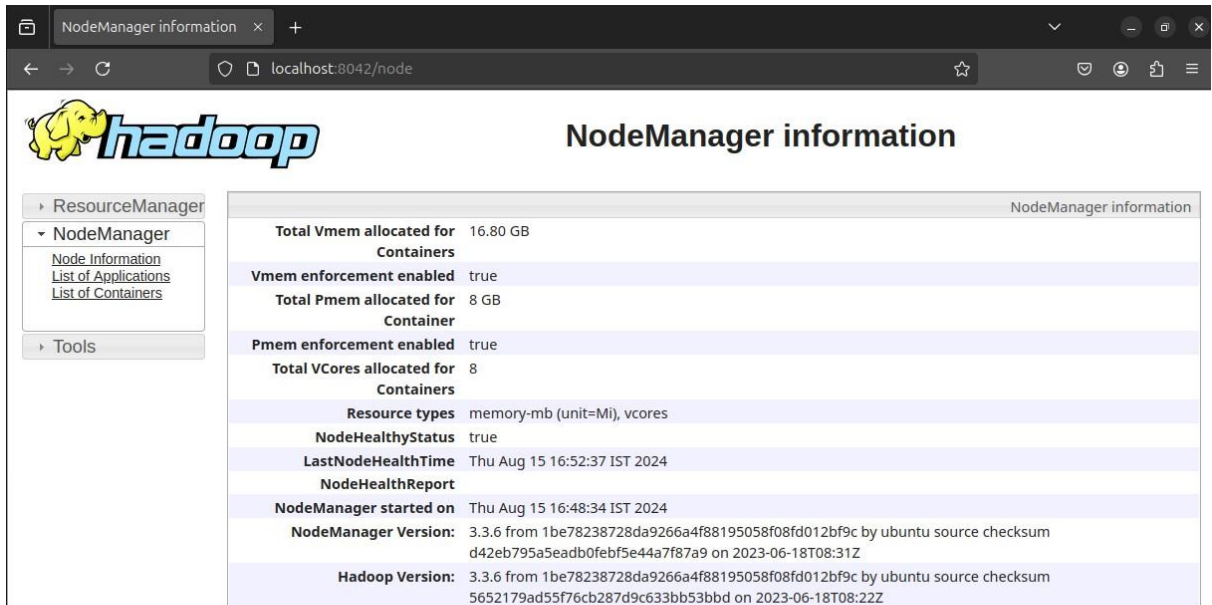
Scheduler Metrics

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

Show 20 entries

ID	User	Name	Application Type	Application Tags	Queue	Application Priority	StartTime	LaunchTime	FinishTime
Showing 0 to 0 of 0 entries									

14. Open <http://localhost:8042>



The screenshot shows a web browser window with the address bar set to `localhost:8042/node`. The page title is "NodeManager information" and it features the Hadoop logo. On the left, a sidebar contains a navigation menu with the following items: "ResourceManager", "NodeManager" (selected), "Node Information", "List of Applications", "List of Containers", and "Tools". The main content area displays a table of NodeManager information.

NodeManager information	
Total Vmem allocated for Containers	16.80 GB
Vmem enforcement enabled	true
Total Pmem allocated for Container	8 GB
Pmem enforcement enabled	true
Total VCoers allocated for Containers	8
Resource types	memory-mb (unit=Mi), vcores
NodeHealthyStatus	true
LastNodeHealthTime	Thu Aug 15 16:52:37 IST 2024
NodeHealthReport	
NodeManager started on	Thu Aug 15 16:48:34 IST 2024
NodeManager Version:	3.3.6 from 1be78238728da9266a4f88195058f08fd012bf9c by ubuntu source checksum d42eb795a5eadb0febf5e44a7f87a9 on 2023-06-18T08:31Z
Hadoop Version:	3.3.6 from 1be78238728da9266a4f88195058f08fd012bf9c by ubuntu source checksum 5652179ad55f76cb287d9c633bb53bbd on 2023-06-18T08:22Z