



**S. B. JAIN INSTITUTE OF TECHNOLOGY,
MANAGEMENT & RESEARCH, NAGPUR.**

Practical No. 1 (Pre-Lab)

Aim: Analyze and Demonstrate the Installation Process of Big Data Tool Hadoop 3.3.6 and JDK 1.8 on Windows Platform.

Name of Student: Shruti Pradeep Bagdi

Roll No.: CS22130

Semester/Year: 7th / 4th

Academic Session: 2025-2026

Date of Performance:

Date of Submission:

AIM: Analyze and Demonstrate the Installation Process of Big Data Tool Hadoop 3.3.6 and JDK 1.8 on Windows Platform.

OBJECTIVE/EXPECTED LEARNING OUTCOME:

The objectives and expected learning outcome of this practical are:

- Able to understand the Basics of Big Data and Hadoop
- Able to understand of Hadoop's core components: HDFS (Hadoop Distributed File System) and YARN (Yet Another Resource Negotiator)
- Able to Work with Hadoop Ecosystem Tools
- Keep up-to-date with the latest trends and advancements in the Hadoop ecosystem

HARDWARE AND SOFTWARE REQUIRMENTS:

Hardware Requirement: High Configuration computer

Software Requirement: Hadoop-3.3.6, jdk1.8, notepad++, 7zip.

THEORY:

Hadoop software can be installed in three modes of

Hadoop is a Java-based programming framework that supports the processing and storage of extremely large datasets on a cluster of inexpensive machines. It was the first major open source project in the big data playing field and is sponsored by the Apache Software Foundation.

Hadoop-3.3.6 3 is comprised of four main layers:

- **Hadoop Common** is the collection of utilities and libraries that support other Hadoop modules.
- **HDFS**, which stands for Hadoop Distributed File System, is responsible for persisting data to disk.
- **YARN**, short for Yet Another Resource Negotiator, is the "operating system" for HDFS.
- **MapReduce** is the original processing model for Hadoop clusters. It distributes work within the cluster or map, then organizes and reduces the results from the nodes into a response to a query. Many other processing models are available for the 2.x version of Hadoop.

Hadoop clusters are relatively complex to set up, so the project includes a stand-alone mode which is suitable for learning about Hadoop, performing simple operations, and debugging.

Procedure:

we'll install Hadoop in stand-alone mode and run one of the example MapReduce programs it includes to verify the installation.

Step1: Installing Java 8 version.

Java JDK Link to download

<https://www.oracle.com/java/technologies/javase-jdk8-downloads.html>

extract and install Java in C:\Java

– open cmd and type -> javac -version

```
Microsoft Windows [Version 10.0.19041.572]
(c) 2020 Microsoft Corporation. All rights reserved.

C:\Users\asus>javac -version
javac 1.8.0_241
```

Note: To set the path for environment variables. i.e. JAVA_HOME

Step2: Installing Hadoop

With Java in place, we'll visit the Apache Hadoop Releases page to find the most recent stable release.

Follow the binary for the current release:

Download Hadoop from www.hadoop.apache.org

The screenshot shows the Apache Hadoop releases page. At the top, there's a navigation bar with links for Download, Documentation, Community, Development, and Help. Below the navigation, there's a heading 'Download' and a note about source and binary tarballs. A table lists three releases: 3.3.6 (Jun 23, 2023), 3.2.4 (Jul 22, 2022), and 2.10.2 (May 31, 2022). Each row includes links for source and binary downloads and a 'Release notes' link. Below the table, there's a section for verifying releases using GPG, which includes a numbered list of steps. Further down, there's a section for performing a quick check using SHA-512, also with a numbered list of steps. At the bottom of the page, a note states: 'All previous releases of Hadoop are available from the Apache release archive site.'

Version	Release date	Source download	Binary download	Release notes
3.3.6	2023 Jun 23	source (checksum signature)	binary (checksum signature) binary-aarch64 (checksum signature)	Announcement
3.2.4	2022 Jul 22	source (checksum signature)	binary (checksum signature)	Announcement
2.10.2	2022 May 31	source (checksum signature)	binary (checksum signature)	Announcement

Big Data Analysis (PECCS702P)

We suggest the following site for your download:
<https://dlcdn.apache.org/hadoop/common/hadoop-3.3.6/hadoop-3.3.6.tar.gz>

Alternate download locations are suggested below.

It is essential that you [verify the integrity](#) of the downloaded file using the PGP signature (`.asc` file) or a hash (`.md5` or `.sha*` file).

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

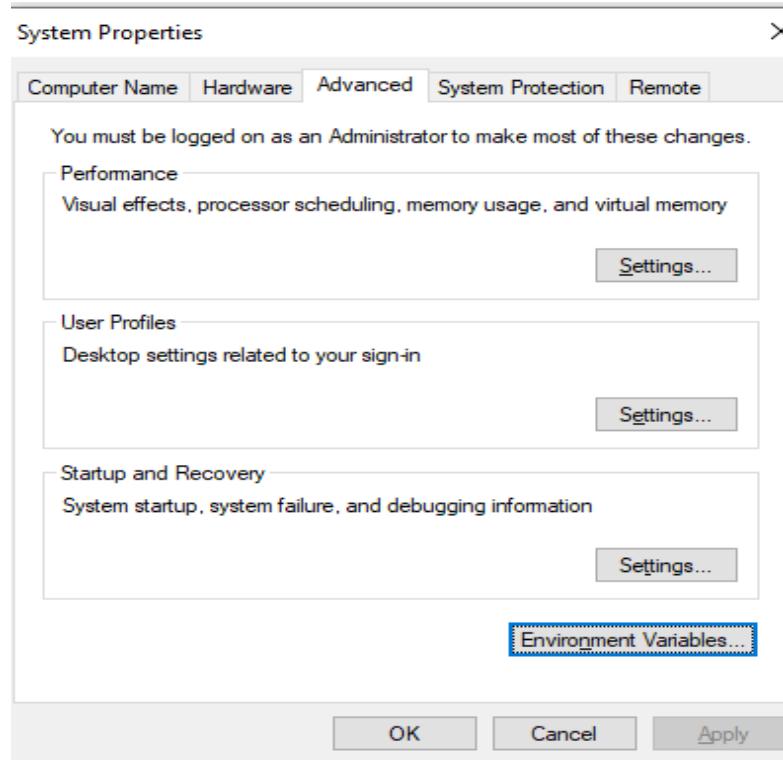
BACKUP SITE
<https://dlcdn.apache.org/hadoop/common/hadoop-3.3.6/hadoop-3.3.6.tar.gz>

VERIFY THE INTEGRITY OF THE FILES

It is essential that you verify the integrity of the downloaded file using the PGP signature (`.asc` file) or a hash (`.md5` or `.sha*` file). Please read [Verifying](#)



1. Set the path JAVA_HOME Environment variable
2. Set the path HADOOP_HOME Environment variable



User variables for Aaditya	
Variable	Value
HADOOP_HOME	C:\hadoop-3.3.6\bin
JAVA_HOME	C:\Java\jdk1.8.0_202\bin
OneDrive	C:\Users\Aaditya\OneDrive
Path	C:\Users\Aaditya\AppData\Local\Microsoft\WindowsApps;
TEMP	C:\Users\Aaditya\AppData\Local\Temp
TMP	C:\Users\Aaditya\AppData\Local\Temp

New...

Edit...

Delete

Edit User Variable

Variable name:	<input type="text" value="JAVA_HOME"/>
Variable value:	<input type="text" value="C:\Java\jdk1.8.0_202\bin"/>
Browse Directory...	Browse File...
OK Cancel	
PATHEXT	.COM;.EXE;.BAT;.CMD;.VBS;.VBE;.JS;.JSE;.WSF;.WSH;.MSC
PROCESSOR_ARCHITECTURE	AMD64
New... Edit... Delete	
OK Cancel	

Edit environment variable

<input type="text" value="C:\Program Files (x86)\Common Files\Oracle\Java\javapath"/> <input type="text" value="%SystemRoot%\system32"/> <input type="text" value="%SystemRoot%"/> <input type="text" value="%SystemRoot%\System32\Wbem"/> <input type="text" value="%SYSTEMROOT%\System32\WindowsPowerShell\v1.0\"/> <input type="text" value="%SYSTEMROOT%\System32\OpenSSH\"/> <input style="background-color: #0070C0; color: white; font-weight: bold; font-style: italic;" type="text" value="C:\Java\jdk1.8.0_202\bin"/> <input type="text" value="C:\hadoop-3.3.6\bin"/> <input type="text" value="C:\hadoop-3.3.6\sbin"/>	New Edit Browse... Delete Move Up Move Down Edit text...
OK Cancel	

Configurations: -

a) File C:/Hadoop-3.3.6/etc/hadoop/core-site.xml, paste below xml paragraph and save this file.

```
<configuration>
<property>
<name>fs.defaultFS</name>
<value>hdfs://localhost:9000</value>
</property>
</configuration>
```

b) C:/Hadoop-3.3.6/etc/hadoop/mapred-site.xml, paste below xml paragraph and save this file.

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

c) Create folder "data" under "C:\Hadoop-3.2.1"

- 1) Create folder "datanode" under "C:\Hadoop-3.2.1\data"
- 2) Create folder "namenode" under "C:\Hadoop-3.2.1\data" data

d) Edit file C:\Hadoop-3.3.6/etc/hadoop/dfs-site.xml, paste below xml paragraph and save this file.

```
<configuration>
<property>
<name>dfs.replication</name>
<value>1</value>
</property>
<property>
<name>dfs.namenode.name.dir</name>
<value>C:\hadoop-3.3.6\data\namenode</value>
</property>
<property>
<name>dfs.datanode.data.dir</name>
<value>C:\hadoop-3.3.6\data\datanode</value>
</property>
</configuration>
```

e) Edit file C:/Hadoop-3.3.6/etc/hadoop/yarn-site.xml, paste below xml paragraph and save this file.

```
<configuration>
<property>
<name>yarn.nodemanager.aux-services</name>
```

```
<value>mapreduce_shuffle</value>
</property>
<property>
<name>yarn.nodemanager.auxservices.mapreduce.shuffle.class</name>
<value>org.apache.hadoop.mapred.ShuffleHandler</value>
</property>
</configuration>
```

f) Edit file C:/Hadoop-3.3.6/etc/hadoop/hadoop-env.cmd

set the path for

```
set JAVA_HOME=C:\java\jdk1.8.0_28\
```

Testing:

Procedure to Run Hadoop

1. Install Apache Hadoop 3.3.6 in Microsoft Windows OS

If Apache Hadoop 3.3.6 is not already installed then follow the post Build, Install, Configure and Run Apache Hadoop 3.3.6 in Microsoft Windows OS.

2. Start HDFS (Namenode and Datanode) and YARN (Resource Manager and Node Manager)

Run following commands. *Command Prompt*

```
C:\Users\abhijitg>cd c:\hadoop
```

```
c:\hadoop>sbin\start-dfs
```

```
c:\hadoop>sbin\start-yarn
```

starting yarn daemons

Start namenode and datanode with this command

- type start-dfs.cmd
- Start yarn through this command
- type start-yarn.cmd

Make sure these apps are running

- Hadoop Namenode
- Hadoop datanode
- YARN Resource Manager
- YARN Node Manager

Namenode, Datanode, Resource Manager and Node Manager will be started in few minutes and ready to execute Hadoop **MapReduce** job in the Single Node (pseudo-distributed mode) cluster.

OUTPUT (SCREENSHOTS)

The image shows two terminal windows side-by-side, both titled "Apache Hadoop Distribution - hadoop".

Terminal 1 (Top): This window displays logs from the "namenode" process. The logs show the startup of the NameNode, the initialization of the FSNamesystem, and the successful quota initialization. It also includes configuration details like storage types and space usage.

```
ber of blocks processed: 0/0
2025-07-31 21:01:24,201 INFO ipc.Server: IPC Server Responder: starting
2025-07-31 21:01:24,202 INFO ipc.Server: IPC Server listener on 9000: starting
2025-07-31 21:01:24,203 INFO namenode.NameNode: NameNode RPC up at: localhost/127.0.0.1:9000.
2025-07-31 21:01:24,204 INFO namenode.FSNamesystem: Starting services required for active state
2025-07-31 21:01:24,205 INFO namenode.FSDirectory: Initializing quota with 12 thread(s)
2025-07-31 21:01:24,212 INFO namenode.FSDirectory: Quota initialization completed in 6 milliseconds
name space=1
storage space=0
storage types=RAM_DISK=0, SSD=0, DISK=0, ARCHIVE=0, PROVIDED=0, NVDIMM=0
2025-07-31 21:01:24,215 INFO blockmanagement.CacheReplicationMonitor: Starting CacheReplicationMonitor with interval 300
00 milliseconds
2025-07-31 21:01:24,801 INFO hdfs.StateChange: BLOCK* registerDatanode: from DatanodeRegistration(127.0.0.1:9866, datano
deUuid=0e139bba-97f4-4c4e-96aa-c02ebd6fc59c, infoPort=9864, infoSecurePort=0, ipcPort=9867, storageInfo=lv=-57;cid=CID-4
1c0d2bc-68d6-44c6-92f4-44c5bec4766e;nsid=1808835015;c=1753975636063) storage 0e139bba-97f4-4c4e-96aa-c02ebd6fc59c
2025-07-31 21:01:24,803 INFO net.NetworkTopology: Adding a new node: /default-rack/127.0.0.1:9866
2025-07-31 21:01:24,804 INFO blockmanagement.BlockReportLeaseManager: Registered DN 0e139bba-97f4-4c4e-96aa-c02ebd6fc59c
(127.0.0.1:9866).
2025-07-31 21:01:24,887 INFO blockmanagement.DatanodeDescriptor: Adding new storage ID DS-feca884c-fc5f-441d-a9c6-b56162
7a9fcf for DN 127.0.0.1:9866
2025-07-31 21:01:24,931 INFO BlockStateChange: BLOCK* processReport 0xa9fafaf75265de6c0 with lease ID 0xb000b22a8f8835e0:
Processing first storage report for DS-feca884c-fc5f-441d-a9c6-b561627a9fcf from datanode DatanodeRegistration(127.0.
0.1:9866, datanodeUuid=0e139bba-97f4-4c4e-96aa-c02ebd6fc59c, infoPort=9864, infoSecurePort=0, ipcPort=9867, storageInfo=lv
=-57;cid=CID-41c0d2bc-68d6-44c6-92f4-44c5bec4766e;nsid=1808835015;c=1753975636063)
2025-07-31 21:01:24,932 INFO BlockStateChange: BLOCK* processReport 0xa9fafaf75265de6c0 with lease ID 0xb000b22a8f8835e0:
from storage DS-feca884c-fc5f-441d-a9c6-b561627a9fcf node DatanodeRegistration(127.0.0.1:9866, datanodeUuid=0e139bba-97
f4-4c4e-96aa-c02ebd6fc59c, infoPort=9864, infoSecurePort=0, ipcPort=9867, storageInfo=lv=-57;cid=CID-41c0d2bc-68d6-44c6-
92f4-44c5bec4766e;nsid=1808835015;c=1753975636063), blocks: 0, hasStaleStorage: false, processing time: 1 msecs, invalid
atedBlocks: 0
```

Terminal 2 (Bottom): This window displays logs from the "datanode" process. It shows the datanode performing volume checks, scanning for blocks, and successfully registering with the NameNode. It also handles block reports and heartbeat responses.

```
68.1.106-1753975636063: 5ms
2025-07-31 21:01:24,718 INFO checker.ThrottledAsyncChecker: Scheduling a check for C:\hadoop\data\datanode
2025-07-31 21:01:24,729 INFO checker.DatasetVolumeChecker: Scheduled health check for volume C:\hadoop\data\datanode
2025-07-31 21:01:24,733 INFO datanode.VolumeScanner: Now scanning bpid BP-948558045-192.168.1.106-1753975636063 on volum
e C:\hadoop\data\datanode
2025-07-31 21:01:24,735 INFO datanode.VolumeScanner: VolumeScanner(C:\hadoop\data\datanode, DS-feca884c-fc5f-441d-a9c6-b
561627a9fcf): finished scanning block pool BP-948558045-192.168.1.106-1753975636063
2025-07-31 21:01:24,739 WARN datanode.DirectoryScanner: dfs.datanode.directoryscan.throttle.limit.ms.per.sec set to valu
e above 1000 ms/sec. Assuming default value of -1
2025-07-31 21:01:24,740 INFO datanode.DirectoryScanner: Periodic Directory Tree Verification scan starting in 15237089ms
with interval of 2160000ms and throttle limit of -1ms/s
2025-07-31 21:01:24,746 INFO datanode.DataNode: Block pool BP-948558045-192.168.1.106-1753975636063 (Datanode Uuid 0e139
bba-97f4-4c4e-96aa-c02ebd6fc59c) service to localhost/127.0.0.1:9000 beginning handshake with NN: localhost/127.0.0.1:90
00.
2025-07-31 21:01:24,749 INFO datanode.VolumeScanner: VolumeScanner(C:\hadoop\data\datanode, DS-feca884c-fc5f-441d-a9c6-b
561627a9fcf): no suitable block pools found to scan. Waiting 1814399984 ms.
2025-07-31 21:01:24,817 INFO datanode.DataNode: Block pool BP-948558045-192.168.1.106-1753975636063 (Datanode Uuid 0e139
bba-97f4-4c4e-96aa-c02ebd6fc59c) service to localhost/127.0.0.1:9000 successfully registered with NN: localhost/127.0.
1:9000.
2025-07-31 21:01:24,818 INFO datanode.DataNode: For namenode localhost/127.0.0.1:9000 using BLOCKREPORT_INTERVAL of 2160
000msecs CACHEREPORT_INTERVAL of 10000msecs Initial delay: 0msecs; heartBeatInterval=3000
2025-07-31 21:01:24,905 INFO datanode.DataNode: After receiving heartbeat response, updating state of namenode localhost
:9000 to active
2025-07-31 21:01:24,962 INFO datanode.DataNode: Successfully sent block report 0xa9fafaf75265de6c0 with lease ID 0xb000b2
2a8f8835e0 to namenode: localhost/127.0.0.1:9000, containing 1 storage report(s), of which we sent 1. The reports had 0
total blocks and used 1 RPC(s). This took 4 msecs to generate and 52 msecs for RPC and NN processing. Got back one comm
and: FinalizeCommand/5.
2025-07-31 21:01:24,964 INFO datanode.DataNode: Got finalize command for block pool BP-948558045-192.168.1.106-17539756
36063
```

Big Data Analysis (PECCS702P)

```
Apache Hadoop Distribution - yarn nodemanager
INFO: Registering org.apache.hadoop.server.nodemanager.webapp.JAXBContextResolver as a provider class
Jul 31, 2025 9:04:43 PM com.sun.jersey.server.impl.application.WebApplicationImpl _initiate
INFO: Initiating Jersey application, version 'Jersey: 1.19.4 05/24/2017 03:20 PM'
Jul 31, 2025 9:04:43 PM com.sun.jersey.spi.container.GuiceComponentProviderFactory getComponentProvider
INFO: Binding org.apache.hadoop.yarn.server.nodemanager.webapp.JAXBContextResolver to GuiceManagedComponentProvider with the scope "Singleton"
Jul 31, 2025 9:04:43 PM com.sun.jersey.spi.container.GuiceComponentProviderFactory getComponentProvider
INFO: Binding org.apache.hadoop.yarn.webapp.GenericExceptionHandler to GuiceManagedComponentProvider with the scope "Singleton"
Jul 31, 2025 9:04:43 PM com.sun.jersey.spi.container.GuiceComponentProviderFactory getComponentProvider
INFO: Binding org.apache.hadoop.yarn.server.nodemanager.webapp.NMWebServices to GuiceManagedComponentProvider with the scope "Singleton"
2025-07-31 21:04:44,048 INFO handler.ContextHandler: Started o.e.j.w.WebAppContext@4703c998{node/,file:///C:/Users/user/AppData/Local/Temp/jetty-0_0_0-8042-hadoop-yarn-common-3_4_0_jar-_any-180424751160214467/webapp/,AVAILABLE}{jar:file:///C:/hadoop/share/hadoop/yarn/hadoop-yarn-common-3.4.0.jar!/webapps/node}
2025-07-31 21:04:44,070 INFO server.AbstractConnector: Started ServerConnector@49fe3142[HTTP/1.1, (http/1.1)]{0.0.0.0:8042}
2025-07-31 21:04:44,071 INFO server.Server: Started @6186ms
2025-07-31 21:04:44,074 INFO webapp.WebApps: Web app node started at 8042
2025-07-31 21:04:44,076 INFO nodemanager.NodeStatusUpdaterImpl: Node ID assigned is : DESKTOP-6AT72I8:63086.
2025-07-31 21:04:44,079 INFO util.JvmPauseMonitor: Starting JVM pause monitor
2025-07-31 21:04:44,095 INFO client.DefaultNoHARMFailoverProxyProvider: Connecting to ResourceManager at /0.0.0.0:8031
2025-07-31 21:04:44,162 INFO nodemanager.NodeStatusUpdaterImpl: Running Applications Size : 0.
2025-07-31 21:04:44,509 INFO security.NMContainerTokenSecretManager: Rolling master-key for container-tokens, got key with id 1088830966
2025-07-31 21:04:44,510 INFO security.NMTokenSecretManagerInNM: Rolling master-key for container-tokens, got key with id 708500551
2025-07-31 21:04:44,511 INFO nodemanager.NodeStatusUpdaterImpl: Registered with ResourceManager as DESKTOP-6AT72I8:63086 with total resource of <memory:8192, vCores:8>
[...]
Apache Hadoop Distribution - yarn resourcemanager
2025-07-31 21:04:44,005 INFO util.JvmPauseMonitor: Starting JVM pause monitor
2025-07-31 21:04:44,022 INFO ipc.CallQueueManager: Using callQueue: class java.util.concurrent.LinkedBlockingQueue, queueCapacity: 5000, scheduler: class org.apache.hadoop.ipc.DefaultRpcScheduler, ipcBackoff: false, ipcFailOver: false.
2025-07-31 21:04:44,027 INFO ipc.Server: Listener at 0.0.0.0:8030
2025-07-31 21:04:44,029 INFO ipc.Server: Starting Socket Reader #1 for port 8030
2025-07-31 21:04:44,039 INFO pb.RpcServerFactoryPBImpl: Adding protocol org.apache.hadoop.yarn.api.ApplicationMasterProtocolPB to the server
2025-07-31 21:04:44,041 INFO ipc.Server: IPC Server Responder: starting
2025-07-31 21:04:44,041 INFO ipc.Server: IPC Server listener on 8030: starting
2025-07-31 21:04:44,168 INFO ipc.CallQueueManager: Using callQueue: class java.util.concurrent.LinkedBlockingQueue, queueCapacity: 5000, scheduler: class org.apache.hadoop.ipc.DefaultRpcScheduler, ipcBackoff: false, ipcFailOver: false.
2025-07-31 21:04:44,168 INFO ipc.Server: Listener at 0.0.0.0:8032
2025-07-31 21:04:44,170 INFO ipc.Server: Starting Socket Reader #1 for port 8032
2025-07-31 21:04:44,174 INFO pb.RpcServerFactoryPBImpl: Adding protocol org.apache.hadoop.yarn.api.ApplicationClientProtocolPB to the server
2025-07-31 21:04:44,175 INFO ipc.Server: IPC Server Responder: starting
2025-07-31 21:04:44,178 INFO ipc.Server: IPC Server listener on 8032: starting
2025-07-31 21:04:44,487 INFO resourcemanager.ResourceTrackerService: NodeManager from node DESKTOP-6AT72I8(cmPort: 63086 httpPort: 8042) registered with capability: <memory:8192, vCores:8>, assigned nodeId DESKTOP-6AT72I8:63086
2025-07-31 21:04:44,489 INFO rmnode.RMNodeImpl: DESKTOP-6AT72I8:63086 Node Transitioned from NEW to RUNNING
2025-07-31 21:04:44,516 INFO capacity.AbstractLeafQueue: LeafQueue: root.default update max app related, maxApplications =10000, maxApplicationsPerUser=10000, Abs Cap:1.0, Cap: 1.0, MaxCap : 1.0
2025-07-31 21:04:44,517 INFO capacity.CapacityScheduler: Added node DESKTOP-6AT72I8:63086 clusterResource: <memory:8192, vCores:8>
2025-07-31 21:04:44,517 INFO capacity.AbstractLeafQueue: LeafQueue: root.default update max app related, maxApplications =10000, maxApplicationsPerUser=10000, Abs Cap:1.0, Cap: 1.0, MaxCap : 1.0
2025-07-31 21:04:44,689 INFO webproxy.ProxyCA: Created Certificate for OU=YARN-ce017825-1377-417e-a93d-30b1db72d2dd
2025-07-31 21:04:44,749 INFO recovery.RMStateStore: Storing CA Certificate and Private Key
2025-07-31 21:04:44,750 INFO resourcemanager.ResourceManager: Transitioned to active state
```

Big Data Analysis (PECCS702P)

Started: Thu Jul 31 21:01:23 +0530 2025
Version: 3.4.0, rbd8b77f398fb26bb7791783192ee7a5dfaeecc760
Compiled: Mon Mar 04 12:05:00 +0530 2024 by root from (HEAD detached at release-3.4.0-RC3)
Cluster ID: CID-41c0d2bc-68d6-44c6-92f4-44c5b5bec4766e
Block Pool ID: BP-948558045-192.168.1.106-1753975636063

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 142.27 MB of 247 MB Heap Memory. Max Heap Memory is 889 MB.
Non Heap Memory used 55.77 MB of 58.25 MB Committed Non Heap Memory. Max Non Heap Memory is <unbounded>.

Configured Capacity:	241.45 GB
----------------------	-----------

http://localhost:9870/explorer.html#/

All Applications

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

No data available in table

http://localhost:8080/cluster

CONCLUSION:

We've installed Hadoop in stand-alone mode and verified it by running an example program it provided.

DISCUSSION AND VIVA VOCE:

- What is Hadoop, and why is it used in the context of big data?
- Describe the steps involved in setting up a multi-node Hadoop cluster.

Big Data Analysis (PECCS702P)

- What is YARN? How does it manage resources in a Hadoop cluster?
- Describe the role of the Resource Manager in YARN.
- Describe the steps involved in upgrading a Hadoop cluster to a newer version.

REFERENCE:

- <https://www.apache.org/dyn/closer.cgi/hadoop/common/hadoop-3.3.6/hadoop-3.3.6.tar.gz>
- <https://ubuntu.com/tutorials/how-to-run-ubuntu-desktop-on-a-virtual-machine-using-virtualbox#2-create-a-new-virtual-machine>

Observation book: (3)	Viva-Voce (3)	Quality of Submission and timely Evaluation (4)
Total:		Sign with date: