



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

Practical No. 0

Aim: : Apply Study and implement the installation of Big Data tool HADOOP-3.3.6 and jdk1.8 for java framework in Windows.

Name of Student:

Roll No.:

Semester/Year: 7th sem/ 4th Year

Academic Session: 2024-2025

Date of Performance:

Date of Submission:

AIM: Apply Study and implement the installation of Big Data tool HADOOP-3.3.6 and jdk1.8 for java framework in Windows.

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 understanding 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 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

```
C:\Users\pc>javac -version
javac 17.0.9
```

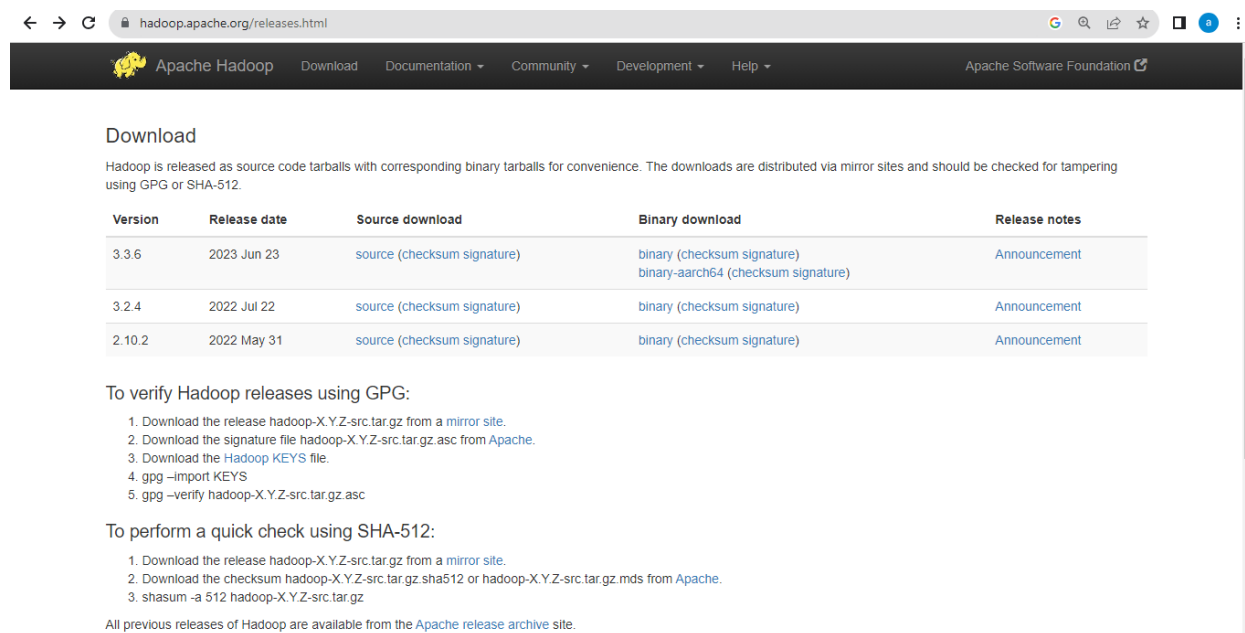
Note: To set the path for environment variables. i.e. JAVA_HOME

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. The header includes the Apache Hadoop logo and navigation links: Download, Documentation, Community, Development, and Help. The main content area is titled "Download" and contains a table of releases. Below the table, there are instructions on how to verify releases using GPG and SHA-512.

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

To verify 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:

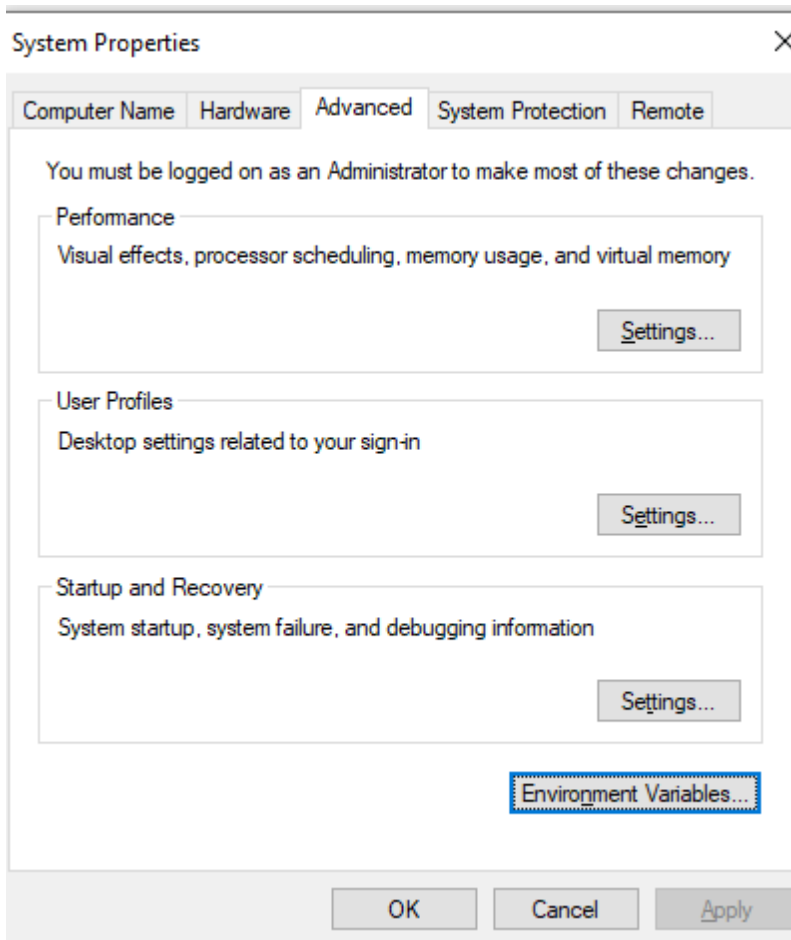
1. Download the release `hadoop-X.Y.Z-src.tar.gz` from a [mirror site](#).
2. Download the checksum `hadoop-X.Y.Z-src.tar.gz.sha512` or `hadoop-X.Y.Z-src.tar.gz.mds` from [Apache](#).
3. `shasum -a 512 hadoop-X.Y.Z-src.tar.gz`

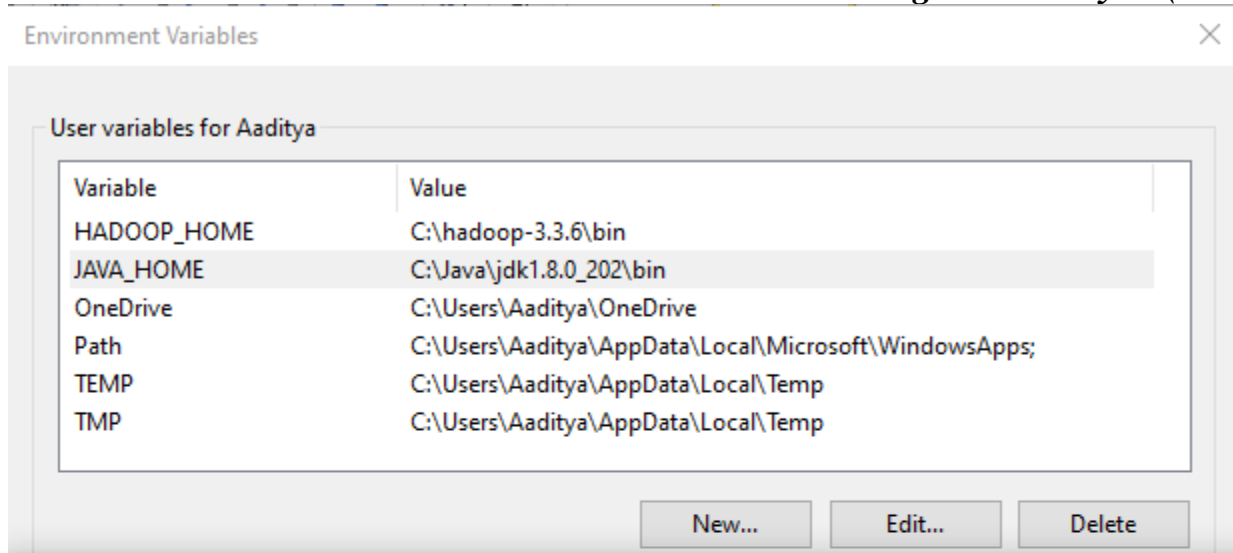
All previous releases of Hadoop are available from the [Apache release archive](#) site.



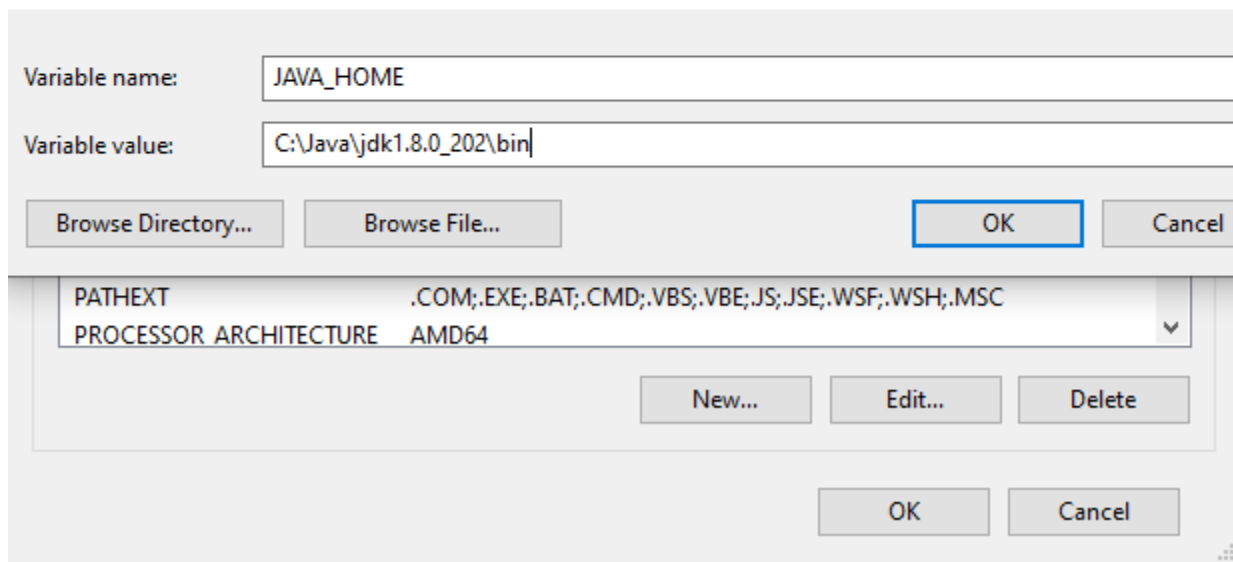
The screenshot shows the Apache Hadoop download page in a web browser. The address bar shows the URL: <https://dlcdn.apache.org/hadoop/common/hadoop-3.3.6/hadoop-3.3.6.tar.gz>. The page features the Apache Software Foundation logo and navigation links like News, About, Make a Donation, The Apache Way, Join Us, and Downloads. The main content area suggests the download site and provides the URL: <https://dlcdn.apache.org/hadoop/common/hadoop-3.3.6/hadoop-3.3.6.tar.gz>. It also mentions alternate download locations and the importance of verifying the integrity of the downloaded file using PGP signatures or hashes. The page is titled "COMMUNITY-LED DEVELOPMENT 'THE APACHE WAY'" and includes sections for "HTTP", "BACKUP SITE", and "VERIFY THE INTEGRITY OF THE FILES". The Windows taskbar at the bottom shows the search bar, task view, and various application icons, along with the system clock indicating 3:10 PM on 8/22/2023.

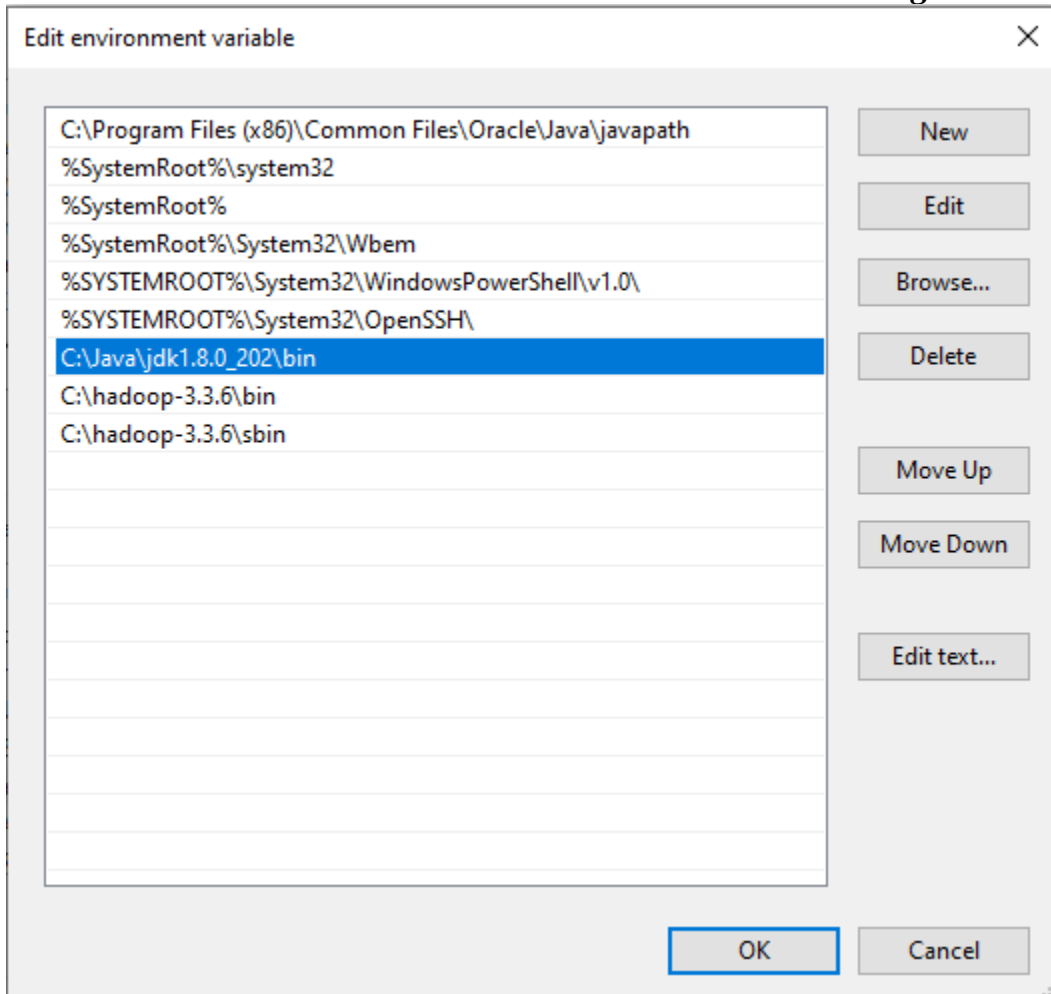
1. Set the path JAVA_HOME Environment variable
2. Set the path HADOOP_HOME Environment variable





Edit User Variable





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/hdfs-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

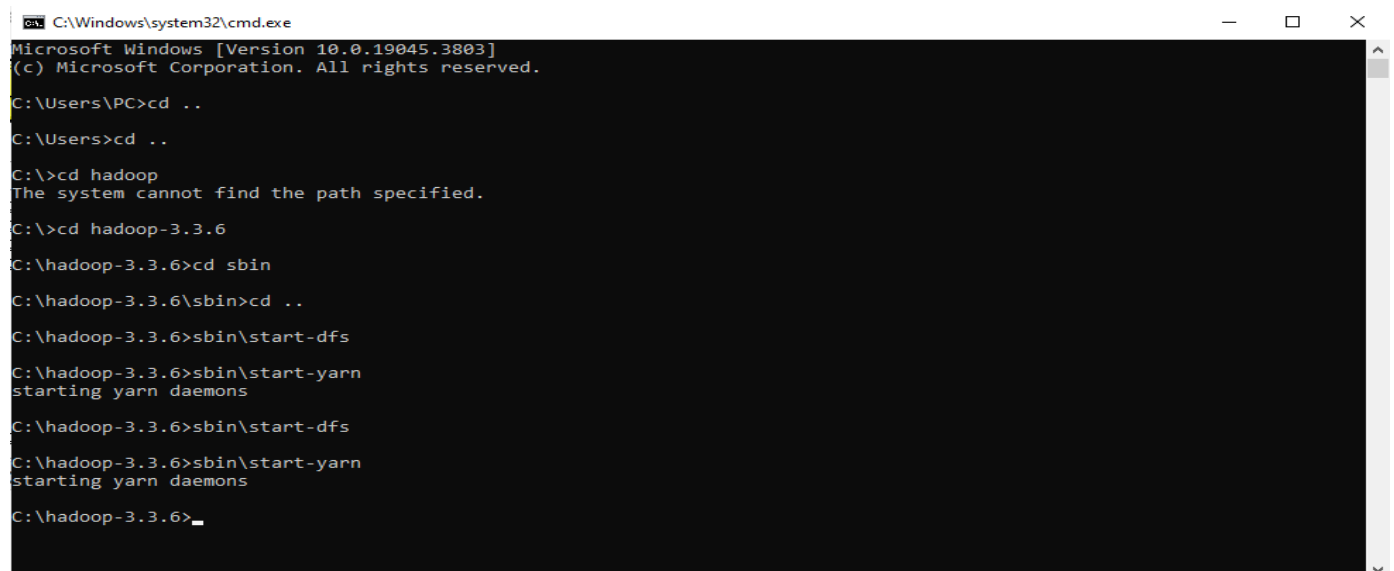
– Hadoop datanode

– 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)




```
C:\Windows\system32\cmd.exe
Microsoft Windows [Version 10.0.19045.3803]
(c) Microsoft Corporation. All rights reserved.

C:\Users\PC>cd ..
C:\Users>cd ..
C:\>cd hadoop
The system cannot find the path specified.
C:\>cd hadoop-3.3.6
C:\hadoop-3.3.6>cd sbin
C:\hadoop-3.3.6\sbin>cd ..
C:\hadoop-3.3.6>sbin\start-dfs
C:\hadoop-3.3.6>sbin\start-yarn
starting yarn daemons
C:\hadoop-3.3.6>sbin\start-dfs
C:\hadoop-3.3.6>sbin\start-yarn
starting yarn daemons
C:\hadoop-3.3.6>
```


 Apache Hadoop Distribution - hadoop namenode



```
NamenodeWebHdfsMethods.putRoot(org.apache.hadoop.security.UserGroupInformation,org.apache.hadoop.hdfs.web.resources.Dele
gationParam,org.apache.hadoop.hdfs.web.resources.UserParam,org.apache.hadoop.hdfs.web.resources.DoAsParam,org.apache.ha
doop.hdfs.web.resources.PutOpParam,org.apache.hadoop.hdfs.web.resources.DestinationParam,org.apache.hadoop.hdfs.web.resou
rces.OwnerParam,org.apache.hadoop.hdfs.web.resources.GroupParam,org.apache.hadoop.hdfs.web.resources.PermissionParam,org
.apache.hadoop.hdfs.web.resources.UnmaskedPermissionParam,org.apache.hadoop.hdfs.web.resources.OverwriteParam,org.apach
e.hadoop.hdfs.web.resources.BufferSizeParam,org.apache.hadoop.hdfs.web.resources.ReplicationParam,org.apache.hadoop.hdfs
.web.resources.BlockSizeParam,org.apache.hadoop.hdfs.web.resources.ModificationTimeParam,org.apache.hadoop.hdfs.web.resou
rces.AccessTimeParam,org.apache.hadoop.hdfs.web.resources.RenameOptionSetParam,org.apache.hadoop.hdfs.web.resources.Crea
teParentParam,org.apache.hadoop.hdfs.web.resources.TokenArgumentParam,org.apache.hadoop.hdfs.web.resources.AclPermission
Param,org.apache.hadoop.hdfs.web.resources.XAttrNameParam,org.apache.hadoop.hdfs.web.resources.XAttrValueParam,org.apach
e.hadoop.hdfs.web.resources.XAttrSetFlagParam,org.apache.hadoop.hdfs.web.resources.SnapshotNameParam,org.apache.hadoop.h
dfs.web.resources.OldSnapshotNameParam,org.apache.hadoop.hdfs.web.resources.ExcludeDatanodesParam,org.apache.hadoop.hdfs
.web.resources.CreateFlagParam,org.apache.hadoop.hdfs.web.resources.NoRedirectParam,org.apache.hadoop.hdfs.web.resources
.StoragePolicyParam,org.apache.hadoop.hdfs.web.resources.ECPolicyParam,org.apache.hadoop.hdfs.web.resources.NamespaceQuo
taParam,org.apache.hadoop.hdfs.web.resources.StorageSpaceQuotaParam,org.apache.hadoop.hdfs.web.resources.StorageTypePara
m) throws java.io.IOException,java.lang.InterruptedException, with URI template, "/", is treated as a resource method
WARNING: A sub-resource method, public javax.ws.rs.core.Response org.apache.hadoop.hdfs.server.namenode.web.resources.
NamenodeWebHdfsMethods.postRoot(org.apache.hadoop.security.UserGroupInformation,org.apache.hadoop.hdfs.web.resources.Del
egationParam,org.apache.hadoop.hdfs.web.resources.UserParam,org.apache.hadoop.hdfs.web.resources.DoAsParam,org.apache.ha
doop.hdfs.web.resources.PostOpParam,org.apache.hadoop.hdfs.web.resources.ConcatSourcesParam,org.apache.hadoop.hdfs.web.r
esources.BufferSizeParam,org.apache.hadoop.hdfs.web.resources.ExcludeDatanodesParam,org.apache.hadoop.hdfs.web.resources
.NewLengthParam,org.apache.hadoop.hdfs.web.resources.NoRedirectParam) throws java.io.IOException,java.lang.InterruptedException, with URI template, "/", is treated as a resource method
WARNING: A sub-resource method, public javax.ws.rs.core.Response org.apache.hadoop.hdfs.server.namenode.web.resources.
NamenodeWebHdfsMethods.deleteRoot(org.apache.hadoop.security.UserGroupInformation,org.apache.hadoop.hdfs.web.resources.D
elegationParam,org.apache.hadoop.hdfs.web.resources.UserParam,org.apache.hadoop.hdfs.web.resources.DoAsParam,org.apache
.hadoop.hdfs.web.resources.DeleteOpParam,org.apache.hadoop.hdfs.web.resources.RecursiveParam,org.apache.hadoop.hdfs.web.r
esources.SnapshotNameParam) throws java.io.IOException,java.lang.InterruptedException, with URI template, "/", is treat
ed as a resource method
```

 Apache Hadoop Distribution - hadoop datanode

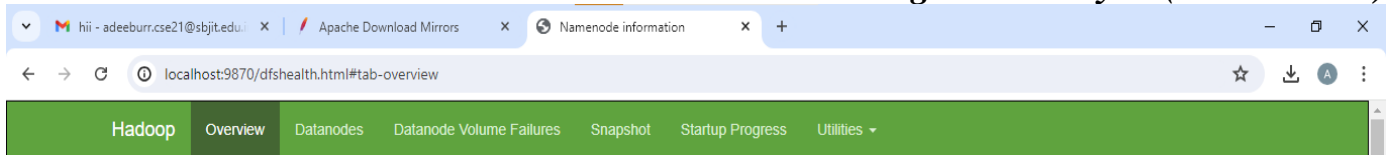


```
168.8.178-1695195980952: 3ms
2024-07-12 12:38:38,966 INFO checker.ThrottledAsyncChecker: Scheduling a check for C:\hadoop-3.3.6\data\datanode
2024-07-12 12:38:38,976 INFO checker.DatasetVolumeChecker: Scheduled health check for volume C:\hadoop-3.3.6\data\datanode
2024-07-12 12:38:38,995 INFO datanode.VolumeScanner: Now rescanning bpid BP-1554588795-192.168.8.178-1695195980952 on vo
lume C:\hadoop-3.3.6\data\datanode, after more than 504 hour(s)
2024-07-12 12:38:38,996 INFO datanode.VolumeScanner: VolumeScanner(C:\hadoop-3.3.6\data\datanode, DS-f74823ea-9d4e-4c01-
aef6-839f6e832f3b): finished scanning block pool BP-1554588795-192.168.8.178-1695195980952
2024-07-12 12:38:39,000 WARN datanode.DirectoryScanner: dfs.datanode.directoryscan.throttle.limit.ms.per.sec set to valu
e above 1000 ms/sec. Assuming default value of -1
2024-07-12 12:38:39,000 INFO datanode.DirectoryScanner: Periodic Directory Tree Verification scan starting in 18528444ms
with interval of 21600000ms and throttle limit of -1ms/s
2024-07-12 12:38:39,008 INFO datanode.DataNode: Block pool BP-1554588795-192.168.8.178-1695195980952 (Datanode Uuid 91c2
72e7-34f0-4619-a0b7-8fb76bdaf8f7) service to localhost/127.0.0.1:9000 beginning handshake with NN
2024-07-12 12:38:39,009 INFO datanode.VolumeScanner: VolumeScanner(C:\hadoop-3.3.6\data\datanode, DS-f74823ea-9d4e-4c01-
aef6-839f6e832f3b): no suitable block pools found to scan. Waiting 1814399986 ms.
2024-07-12 12:38:39,128 INFO datanode.DataNode: Block pool BP-1554588795-192.168.8.178-1695195980952 (Datanode Uuid 91c2
72e7-34f0-4619-a0b7-8fb76bdaf8f7) service to localhost/127.0.0.1:9000 successfully registered with NN
2024-07-12 12:38:39,129 INFO datanode.DataNode: For namenode localhost/127.0.0.1:9000 using BLOCKREPORT_INTERVAL of 2160
0000msecs CACHEREPORT_INTERVAL of 10000msecs Initial delay: 0msecs; heartbeatInterval=3000
2024-07-12 12:38:39,129 INFO datanode.DataNode: Starting IBR Task Handler.
2024-07-12 12:38:39,222 INFO datanode.DataNode: After receiving heartbeat response, updating state of namenode localhost
:9000 to active
2024-07-12 12:38:39,270 INFO datanode.DataNode: Successfully sent block report 0x3c10929e6494fdb0 with lease ID 0x6ee893
f4a7b1e41e 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 5 msec to generate and 42 msec for RPC and NN processing. Got back one comm
and: FinalizeCommand/5.
2024-07-12 12:38:39,271 INFO datanode.DataNode: Got finalize command for block pool BP-1554588795-192.168.8.178-16951959
80952
```

```
Apache Hadoop Distribution - yarn resourcemanager
2024-07-12 12:38:48,840 INFO placement.MultiNodeSortingManager: Starting NodeSortingService=MultiNodeSortingManager
2024-07-12 12:38:48,858 INFO ipc.CallQueueManager: Using callQueue: class java.util.concurrent.LinkedBlockingQueue, queueCapacity: 5000, scheduler: class org.apache.hadoop.ipc.DefaultRpcScheduler, ipcBackoff: false.
2024-07-12 12:38:48,874 INFO ipc.Server: Listener at 0.0.0.0:8031
2024-07-12 12:38:48,886 INFO ipc.Server: Starting Socket Reader #1 for port 8031
2024-07-12 12:38:48,894 INFO pb.RpcServerFactoryPBImpl: Adding protocol org.apache.hadoop.yarn.server.api.ResourceTrackerPB to the server
2024-07-12 12:38:48,895 INFO ipc.Server: IPC Server listener on 8031: starting
2024-07-12 12:38:48,896 INFO ipc.Server: IPC Server Responder: starting
2024-07-12 12:38:48,913 INFO util.JvmPauseMonitor: Starting JVM pause monitor
2024-07-12 12:38:48,929 INFO ipc.CallQueueManager: Using callQueue: class java.util.concurrent.LinkedBlockingQueue, queueCapacity: 5000, scheduler: class org.apache.hadoop.ipc.DefaultRpcScheduler, ipcBackoff: false.
2024-07-12 12:38:48,953 INFO ipc.Server: Listener at 0.0.0.0:8030
2024-07-12 12:38:48,963 INFO ipc.Server: Starting Socket Reader #1 for port 8030
2024-07-12 12:38:48,974 INFO pb.RpcServerFactoryPBImpl: Adding protocol org.apache.hadoop.yarn.api.ApplicationMasterProtocolPB to the server
2024-07-12 12:38:48,974 INFO ipc.Server: IPC Server listener on 8030: starting
2024-07-12 12:38:48,975 INFO ipc.Server: IPC Server Responder: starting
2024-07-12 12:38:49,061 INFO ipc.CallQueueManager: Using callQueue: class java.util.concurrent.LinkedBlockingQueue, queueCapacity: 5000, scheduler: class org.apache.hadoop.ipc.DefaultRpcScheduler, ipcBackoff: false.
2024-07-12 12:38:49,076 INFO ipc.Server: Listener at 0.0.0.0:8032
2024-07-12 12:38:49,088 INFO ipc.Server: Starting Socket Reader #1 for port 8032
2024-07-12 12:38:49,096 INFO pb.RpcServerFactoryPBImpl: Adding protocol org.apache.hadoop.yarn.api.ApplicationClientProtocolPB to the server
2024-07-12 12:38:49,097 INFO ipc.Server: IPC Server Responder: starting
2024-07-12 12:38:49,098 INFO ipc.Server: IPC Server listener on 8032: starting
2024-07-12 12:38:49,547 INFO webproxy.ProxyCA: Created Certificate for OU=YARN-67057bd1-b042-4563-b1b8-9f762eef30b6
2024-07-12 12:38:49,567 INFO recovery.RMStateStore: Storing CA Certificate and Private Key
2024-07-12 12:38:49,567 INFO resourcemanager.ResourceManager: Transitioned to active state
```

```
Apache Hadoop Distribution
at org.apache.hadoop.service.AbstractService.init(AbstractService.java:164)
at org.apache.hadoop.service.CompositeService.serviceInit(CompositeService.java:109)
at org.apache.hadoop.yarn.server.nodemanager.containermanager.ContainerManagerImpl.serviceInit(ContainerManagerImpl.java:327)
at org.apache.hadoop.service.AbstractService.init(AbstractService.java:164)
at org.apache.hadoop.service.CompositeService.serviceInit(CompositeService.java:109)
at org.apache.hadoop.yarn.server.nodemanager.NodeManager.serviceInit(NodeManager.java:494)
at org.apache.hadoop.service.AbstractService.init(AbstractService.java:164)
at org.apache.hadoop.yarn.server.nodemanager.NodeManager.initAndStartNodeManager(NodeManager.java:962)
at org.apache.hadoop.yarn.server.nodemanager.NodeManager.main(NodeManager.java:1042)

at org.apache.hadoop.fs.RawLocalFileSystem$DeprecatedRawLocalFileStatus.loadPermissionInfoByNonNativeIO(RawLocalFileSystem.java:1040)
at org.apache.hadoop.fs.RawLocalFileSystem$DeprecatedRawLocalFileStatus.loadPermissionInfo(RawLocalFileSystem.java:991)
at org.apache.hadoop.fs.RawLocalFileSystem$DeprecatedRawLocalFileStatus.getPermission(RawLocalFileSystem.java:952)
at org.apache.hadoop.yarn.server.nodemanager.containermanager.localizer.ResourceLocalizationService.initializeLocalDir(ResourceLocalizationService.java:1451)
at org.apache.hadoop.yarn.server.nodemanager.containermanager.localizer.ResourceLocalizationService.initializeLocalDirs(ResourceLocalizationService.java:1421)
at org.apache.hadoop.yarn.server.nodemanager.containermanager.localizer.ResourceLocalizationService.serviceInit(ResourceLocalizationService.java:261)
... 9 more
2024-07-12 12:39:23,702 INFO nodemanager.NodeManager: SHUTDOWN_MSG:
/*****
SHUTDOWN_MSG: Shutting down NodeManager at DESKTOP-7AGANVA/192.168.8.178
*****/
C:\hadoop-3.3.6>
```



The screenshot shows the Hadoop DFS Health Overview page in a web browser. The browser tabs include 'hii - adeeburr.cse21@sbjit.edu...', 'Apache Download Mirrors', and 'Namenode information'. The address bar shows 'localhost:9870/dfshealth.html#tab-overview'. The page has a green header with navigation links: 'Hadoop', 'Overview', 'Datanodes', 'Datanode Volume Failures', 'Snapshot', 'Startup Progress', and 'Utilities'. The 'Overview' tab is selected.

Overview 'localhost:9000' (✓active)

Started:	Fri Jul 12 12:38:37 +0530 2024
Version:	3.3.6, r1be78238728da9266a4f88195056f08fd012bf9c
Compiled:	Sun Jun 18 13:52:00 +0530 2023 by ubuntu from (HEAD detached at release-3.3.6-RC1)
Cluster ID:	CID-aa649808-9e00-4e18-8f1c-d6adce77ab66
Block Pool ID:	BP-1554588795-192.168.8.178-1695195980952

Summary

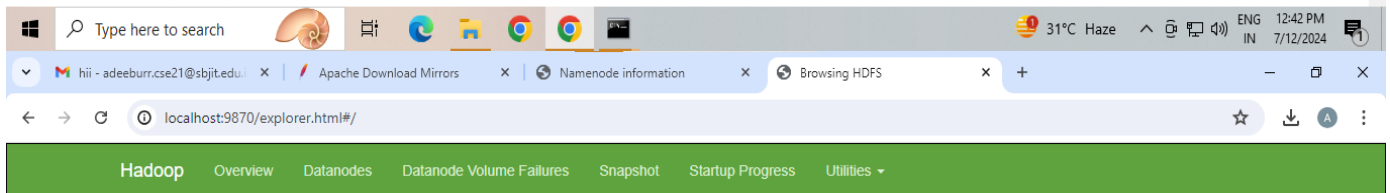
Security is off.

Safemode is off.

1 files and directories, 0 blocks (0 replicated blocks, 0 erasure coded block groups) = 1 total filesystem object(s).

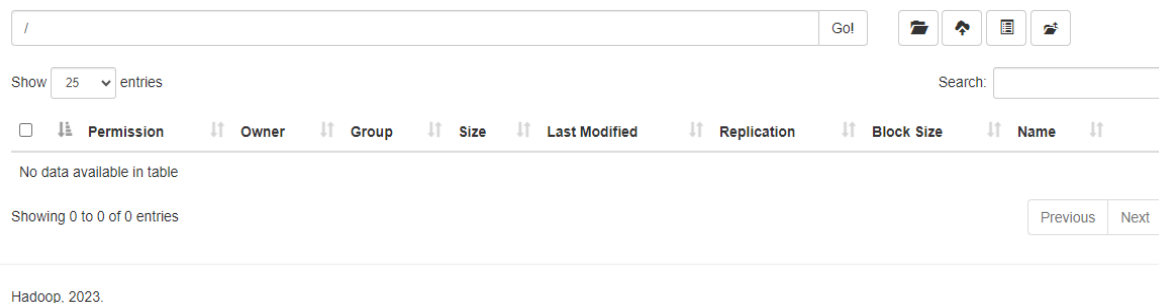
Heap Memory used 100.2 MB of 276.5 MB Heap Memory. Max Heap Memory is 889 MB.

Non Heap Memory used 50.13 MB of 52.36 MB Committed Non Heap Memory. Max Non Heap Memory is <unbounded>.



The screenshot shows the Hadoop Explorer page in a web browser. The browser tabs include 'hii - adeeburr.cse21@sbjit.edu...', 'Apache Download Mirrors', 'Namenode information', and 'Browsing HDFS'. The address bar shows 'localhost:9870/explorer.html#/'. The page has a green header with navigation links: 'Hadoop', 'Overview', 'Datanodes', 'Datanode Volume Failures', 'Snapshot', 'Startup Progress', and 'Utilities'. The 'Overview' tab is selected.

Browse Directory



The screenshot shows the Hadoop Browse Directory page. It features a search bar with the text '/', a 'Go!' button, and icons for file operations. Below the search bar, there is a 'Show 25 entries' dropdown and a 'Search:' input field. A table header is visible with columns: 'Permission', 'Owner', 'Group', 'Size', 'Last Modified', 'Replication', 'Block Size', and 'Name'. The table content is empty, displaying 'No data available in table' and 'Showing 0 to 0 of 0 entries'. At the bottom, there are 'Previous' and 'Next' buttons and the text 'Hadoop, 2023.'



The screenshot shows the Windows taskbar at the bottom of the screen. It includes the Start button, a search bar with the text 'Type here to search', and several application icons. The system tray on the right shows the date and time as '12:44 PM 7/12/2024' and the language as 'ENG IN'. There is also a notification icon with a number '1'.

The screenshot displays the Hadoop web interface at `localhost:8088/cluster`. The interface includes a sidebar with navigation links for Cluster, About, Nodes, Node Labels, Applications, and Scheduler. The main content area shows various metrics and a table of applications.

Cluster Metrics

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

Cluster Nodes Metrics

Active Nodes	Decommissioning Nodes	Decommissioned Nodes
0	0	0

Scheduler Metrics

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

Applications

NEW
NEW SAVING
SUBMITTED
ACCEPTED
RUNNING
FINISHED
FAILED
KILLED

Scheduler

Tools

Showing 0 to 0 of 0 entries

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.
- 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>