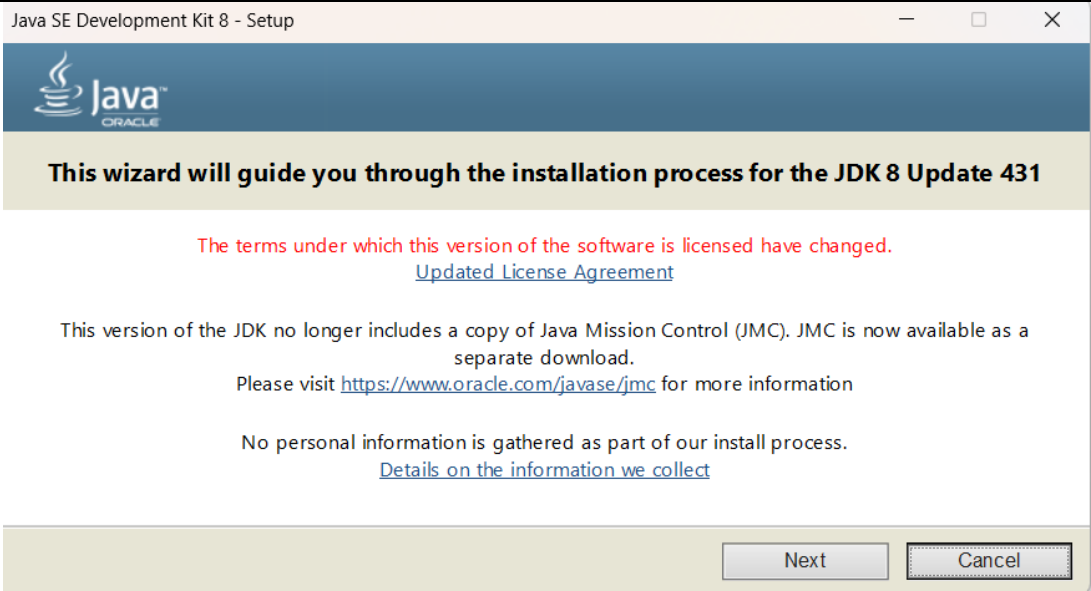





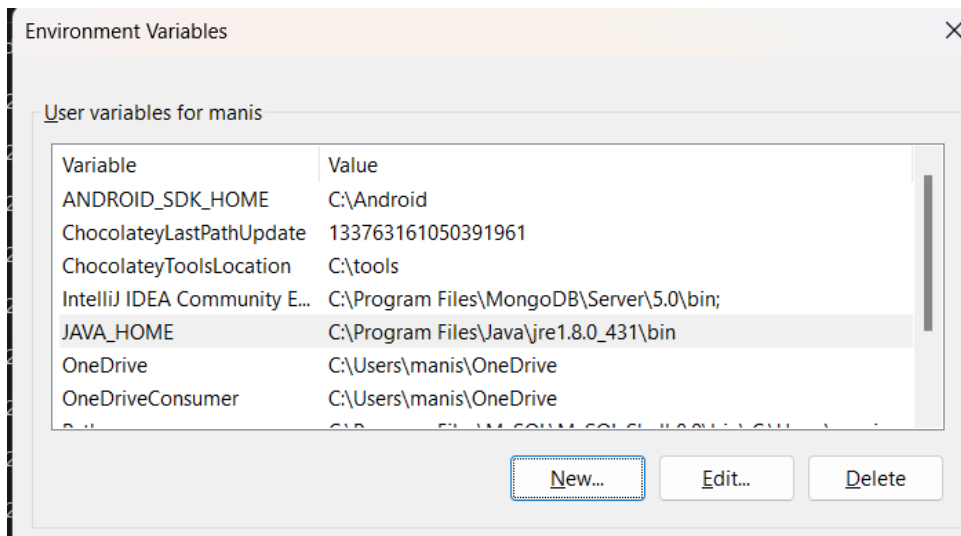
**BHARATIYA VIDYA BHAVAN'S**  
**SARDAR PATEL INSTITUTE OF TECHNOLOGY**  
(Empowered Autonomous Institute Affiliated to Mumbai University)  
**Department Of Computer Engineering**

<b>Name</b>	Manish Shashikant Jadhav
<b>UID</b>	2023301005
<b>Subject</b>	Big Data Analysis and Visualization Lab
<b>Experiment No.</b>	1
<b>Aim</b>	To setup and install Hadoop in Pseudo-Distributed Mode and monitoring Hadoop
<b>1. Install Java JDK 8:</b>	<div><p>Java SE Development Kit 8 - Setup</p><p>Already installed</p></div>



**BHARATIYA VIDYA BHAVAN'S**  
**SARDAR PATEL INSTITUTE OF TECHNOLOGY**  
(Empowered Autonomous Institute Affiliated to Mumbai University)  
**Department Of Computer Engineering**

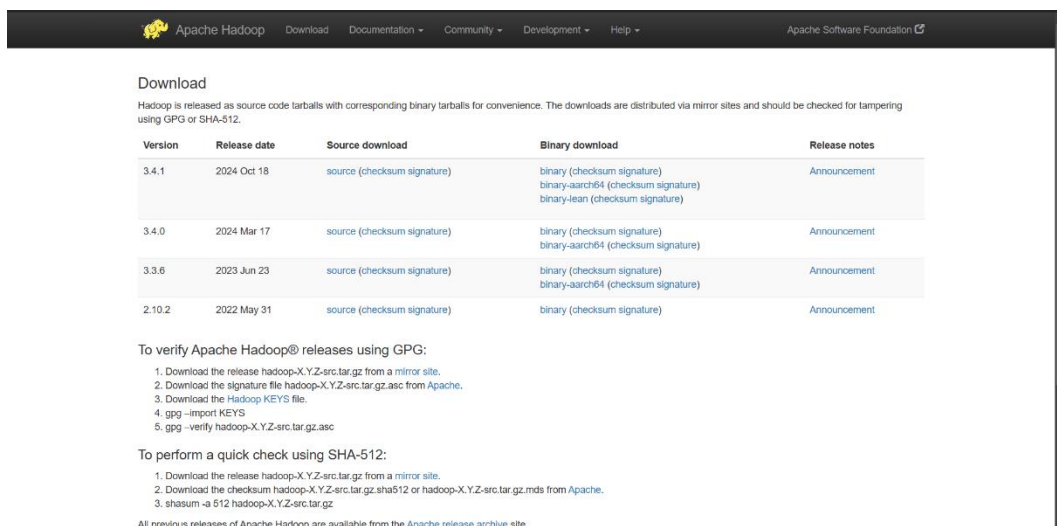
## 2. Setting up Environment Variables



```
C:\Users\manis>java -version
java version "1.8.0_431"
Java(TM) SE Runtime Environment (build 1.8.0_431-b10)
Java HotSpot(TM) 64-Bit Server VM (build 25.431-b10, mixed mode)

C:\Users\manis>
```

## 3. Install Hadoop





**BHARATIYA VIDYA BHAVAN'S**  
**SARDAR PATEL INSTITUTE OF TECHNOLOGY**  
(Empowered Autonomous Institute Affiliated to Mumbai University)  
**Department Of Computer Engineering**



We suggest the following location for your download:

<https://dlcdn.apache.org/hadoop/common/hadoop-3.4.1/hadoop-3.4.1.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-3.4.1/hadoop-3.4.1.tar.gz>

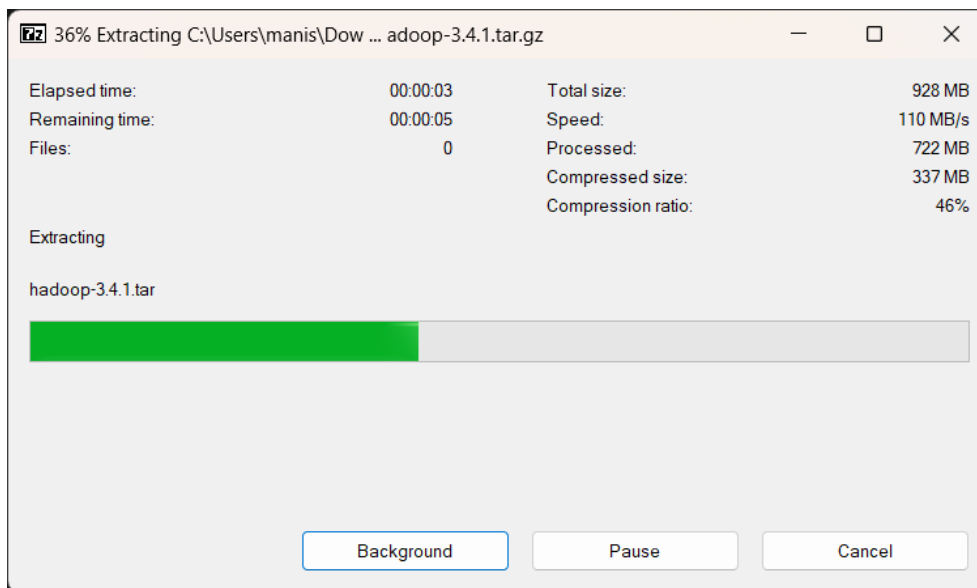
## BACKUP SITES

<https://dlcdn.apache.org/hadoop/common/hadoop-3.4.1/hadoop-3.4.1.tar.gz>

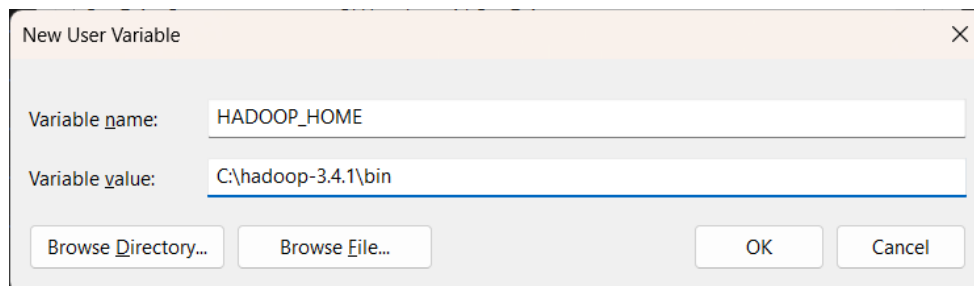
## VERIFY THE INTEGRITY OF THE FILES

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 Apache Software Foundation Releases](#) for more information on why you should verify our releases.

The PGP signature can be verified using PGP or GPG. First download the [KEYS](#) as well as the `.asc` signature file for the relevant distribution. Make sure you get these files from the main distribution site, rather than from a mirror. Then verify the signatures using



#### 4. Set Environment variable for Hadoop





**BHARATIYA VIDYA BHAVAN'S**  
**SARDAR PATEL INSTITUTE OF TECHNOLOGY**  
(Empowered Autonomous Institute Affiliated to Mumbai University)  
**Department Of Computer Engineering**

### 5. Check Hadoop is Installed or not

```
Command Prompt
Microsoft Windows [Version 10.0.22631.4602]
(c) Microsoft Corporation. All rights reserved.

C:\Users\manis>hadoop
Usage: hadoop [--config confdir] [--loglevel loglevel] COMMAND
where COMMAND is one of:
  fs                run a generic filesystem user client
  version           print the version
  jar <jar>         run a jar file
                    note: please use "yarn jar" to launch
                      YARN applications, not this command.
  checknative [-a|-h] check native hadoop and compression libraries availability
  conftest         validate configuration XML files
  distch path:owner:group:permission distributed metadata changer
  distcp <srcurl> <desturl> copy file or directories recursively
  archive -archiveName NAME -p <parent path> <src>* <dest> create a hadoop archive
  classpath        prints the class path needed to get the
                    Hadoop jar and the required libraries
  credential       interact with credential providers
  jnipath          prints the java.library.path
  kerbname         show auth_to_local principal conversion
  kdiag           diagnose kerberos problems
  key             manage keys via the KeyProvider
  daemonlog       get/set the log level for each daemon
  or
  CLASSNAME        run the class named CLASSNAME

Most commands print help when invoked w/o parameters.

C:\Users\manis>hadoop version
Hadoop 3.4.1
Source code repository https://github.com/apache/hadoop.git -r 4d7825309348956336b8f06a08322b78422849b1
Compiled by mthakur on 2024-10-09T14:57Z
Compiled on platform linux-x86_64
Compiled with protoc 3.23.4
From source with checksum 7292fe9dba5e2e44e3a9f763fce3e680
This command was run using /C:/hadoop-3.4.1/share/hadoop/common/hadoop-common-3.4.1.jar

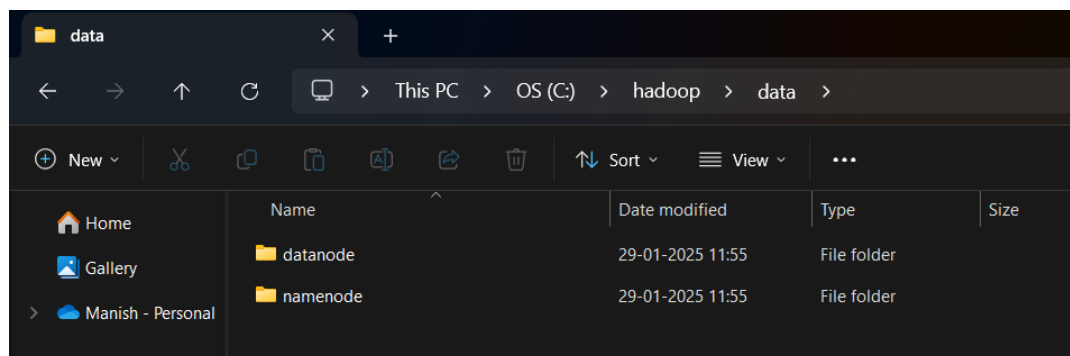
C:\Users\manis>
```

### 6. Starting Hadoop

Configuration of Hadoop cluster is done in below 4 files:

1. %HADOOP\_HOME%\etc\hadoop\hdfs-site.xml
2. %HADOOP\_HOME%\etc\hadoop\core-site.xml
3. %HADOOP\_HOME%\etc\hadoop\mapred-site.xml
4. %HADOOP\_HOME%\etc\hadoop\yarn-site.xml

After the configuration is performed, separate directory is made for datanode and namenode.



Core site, mapreduce and yarn configuration is done. After that we can run datanode, namenode and yarn to start hadoop.



# 7. Hadoop Web UI

Hadoop

Overview

Datanodes

Datanode Volume Failures

Snapshot

Startup Progress

Utilities

## Overview 'localhost:9000' (✓active)

Started:	Wed Jan 29 12:29:57 +0530 2025
Version:	3.4.1, r4d7825309348956339b8f06a08322b78422849b1
Compiled:	Wed Oct 09 20:27:00 +0530 2024 by mthakur from branch:3.4.1
Cluster ID:	CID-bcd779e6-4fd3-4445-bc0c-04a0527d359a
Block Pool ID:	BP-1924174722-192.168.219.1-1738133304370

## 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 173.74 MB of 273 MB Heap Memory. Max Heap Memory is 889 MB.  
Non Heap Memory used 54.27 MB of 56.31 MB Committed Non Heap Memory. Max Non Heap Memory is <unbounded>.

Configured Capacity:	219.07 GB
Configured Remote Capacity:	0 B
DFS Used:	323 B (0%)
Non DFS Used:	215.23 GB
DFS Remaining:	3.84 GB (1.75%)



**BHARATIYA VIDYA BHAVAN'S**  
**SARDAR PATEL INSTITUTE OF TECHNOLOGY**  
(Empowered Autonomous Institute Affiliated to Mumbai University)  
**Department Of Computer Engineering**



**All Applications**

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	Used Resources	Total Resources
0	0	0	0		<memory 0 B, vCores 0>	<memory 0 B, vCores 0>

Cluster Nodes Metrics

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

Scheduler Metrics

Scheduler Type	Scheduling Resource Type	Minimum Allocation	Maximum Allocation	Maximum Cluster Application Priority	Sch
Capacity Scheduler	[memory-mb (unit-M), vcores]	<memory 1024, vCores 1>	<memory 8192, vCores 4>	0	0

Show 20 entries

ID	User	Name	Application Type	Application Tags	Queue	Application Priority	StartTime	LaunchTime	FinishTime	State	FinalStatus	Running Containers	Allocated CPU Vcores	Allocated Memory MB
----	------	------	------------------	------------------	-------	----------------------	-----------	------------	------------	-------	-------------	--------------------	----------------------	---------------------

No data available in table

Showing 0 to 0 of 0 entries

**Conclusion**

Hence by completing this experiment we came to know to setup and install Hadoop in Pseudo-Distributed Mode and monitoring Hadoop.