


1. INSTALLATION OF HADOOP

a) Install Apache Hadoop

Download the latest version of Apache Hadoop from official website.

 Apache Hadoop Download Documentation Community Development Help Apache Software Foundation

Download

Hadoop is released as source code tarballs with corresponding binary tarballs for convenience. The downloads are distributed via mirror sites and should be checked for tampering using GPG or SHA-512.

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:

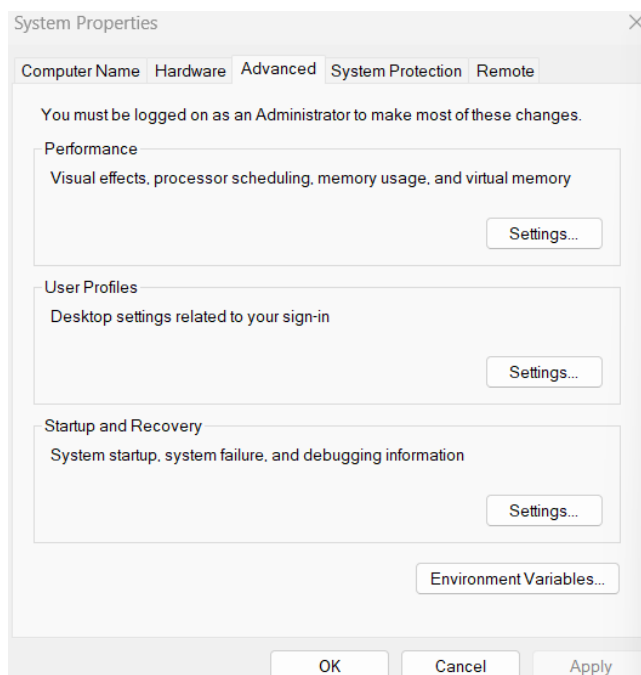
1. Download the release `hadoop-X.Y.Z-src.tar.gz` from a [mirror site](#).
2. Download the checksum `hadoop-X.Y.Z-src.tar.gz.sha512` or `hadoop-X.Y.Z-src.tar.gz.mds` from [Apache](#).
3. `shasum -a 512 hadoop-X.Y.Z-src.tar.gz`

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.

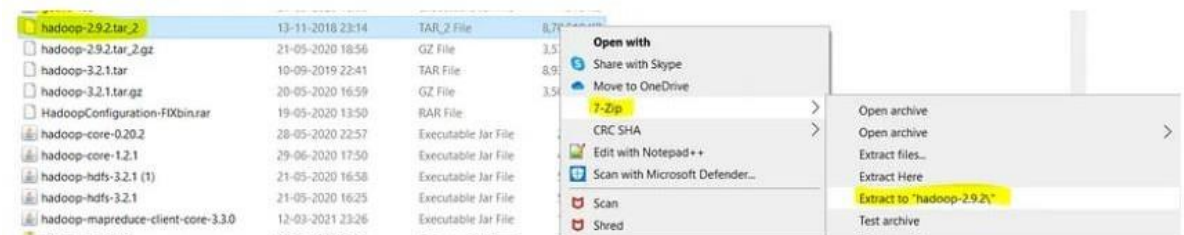
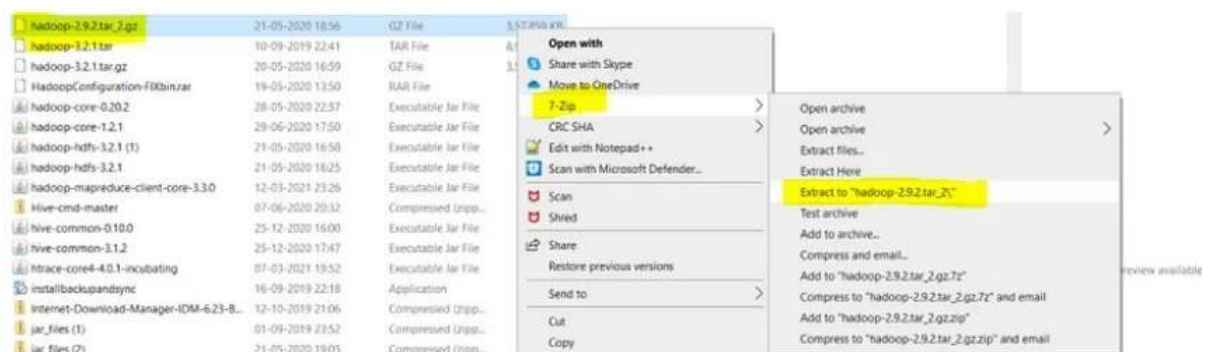
b) Setting Up Environment Variables:

Go to edit environment variables, and select environment variables to set up the environment variables



Setting JAVA_HOME

- Open environment Variable and click on “New” in “User Variable”
- Now , add JAVA_HOME in variable name and path of Java(jdk) in Variable Value.
- Click OK and we are half done with setting JAVA_HOME.



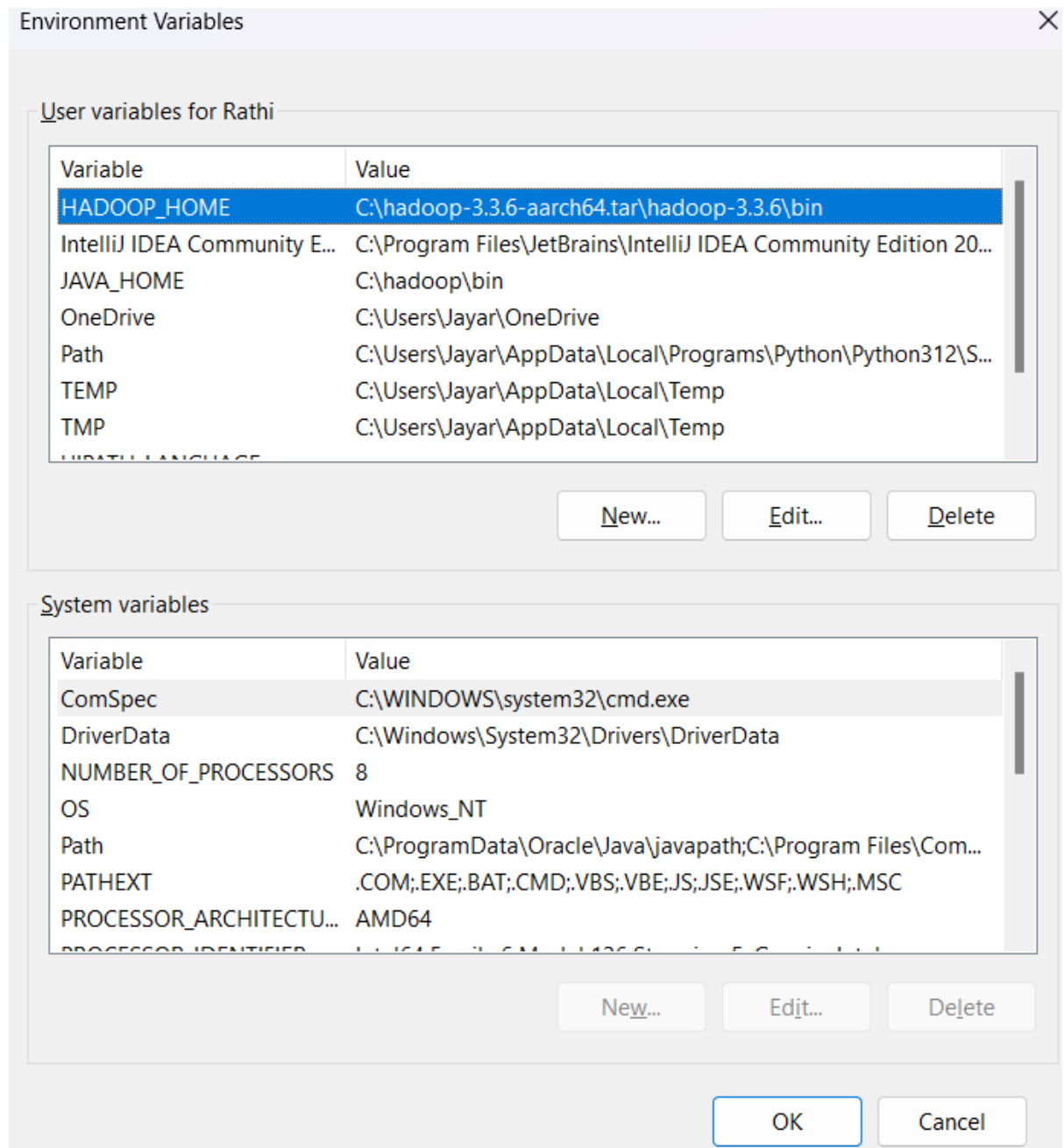
This PC > Shashank (D:) > Shashank > Study > hadoop-2.9.2

Name	Date modified	Type	Size
bin	20-09-2020 16:35	File folder	
data	20-09-2020 16:31	File folder	
etc	13-11-2018 20:45	File folder	
include	13-11-2018 20:45	File folder	
lib	13-11-2018 20:45	File folder	
libexec	13-11-2018 20:45	File folder	
logs	20-09-2020 16:41	File folder	
sbin	13-11-2018 20:45	File folder	
share	13-11-2018 20:45	File folder	
LICENSE	13-11-2018 20:45	TXT File	104 KB
NOTICE	13-11-2018 20:45	TXT File	16 KB
README	13-11-2018 20:45	TXT File	2 KB

Setting HADOOP_HOME

- Open environment Variable and click on “New” in “User Variable”

- Now, add HADOOP_HOME in variable name and path of Hadoop folder in Variable Value.
- Click OK and we are half done with setting HADOOP_HOME.



Setting Path Variable

- Last step in setting Environment variable is setting Path in System Variable.
- Select Path variable in the system variables and click on “Edit”.
- Now we need to add these paths to Path Variable one by one:-
 - * %JAVA_HOME%\bin
 - * %HADOOP_HOME%\bin
 - * %HADOOP_HOME%\sbin
- Click OK and OK. & we are done with Setting Environment Variables.

Verify the Paths

- Open a **NEW** Command Window
- Run following commands.

```
echo %JAVA_HOME%
```

```
echo %HADOOP_HOME%
```

```
echo %PATH%
```

Creating Folders

- Create **DATA folder** in the Hadoop directory
- These folders are important because files on HDFS resides inside the datanode.

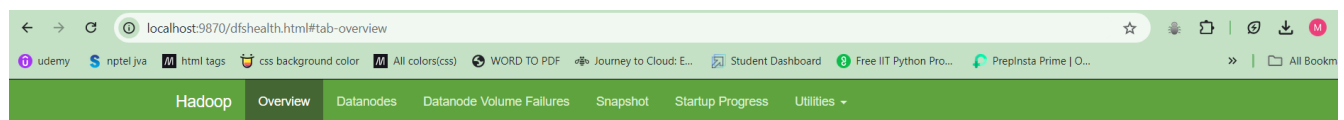
OUTPUT:

```
Administrator: Command Prompt

Microsoft Windows [Version 10.0.22631.4169]
(c) Microsoft Corporation. All rights reserved.

C:\Windows\System32>start-all.cmd
This script is Deprecated. Instead use start-dfs.cmd and start-yarn.cmd
starting yarn daemons

C:\Windows\System32>jps
10620 NameNode
13612 NodeManager
15500 Jps
7612 DataNode
9244 ResourceManager
```

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

Started:	Thu Sep 12 23:21:07 +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-4f9db0b0-e529-4beb-bf7e-2234452a9923
Block Pool ID:	BP-2132870559-192.168.56.1-1723145592918

Summary

Security is off.

Safe mode is ON. The reported blocks 93 has reached the threshold 0.9990 of total blocks 93. The minimum number of live datanodes is not required. In safe mode extension, Safe mode will be turned off automatically in 9 seconds.

191 files and directories, 93 blocks (93 replicated blocks, 0 erasure coded block groups) = 284 total filesystem object(s).

Heap Memory used 58.1 MB of 261.5 MB Heap Memory. Max Heap Memory is 889 MB.

Non Heap Memory used 63.02 MB of 64.56 MB Committed Non Heap Memory. Max Non Heap Memory is <unbounded>.

Result:

Thus the Apache Hadoop is installed on windows Successfully.