

## INSTALL CONFIGURE AND RUN HADOOP AND HDFS

### Aim:

To install configure and run hadoop and hdfs.

### Procedure:

#### 1. To install Java

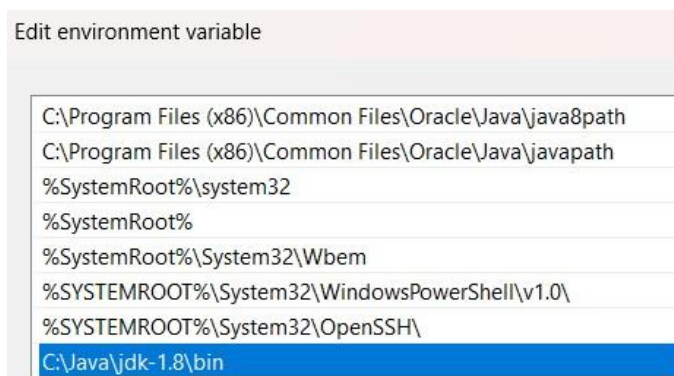
- 1) Check if java is available in the system

```
C:\Windows\System32>java -version
java version "1.8.0_421"
Java(TM) SE Runtime Environment (build 1.8.0_421-b09)
Java HotSpot(TM) 64-Bit Server VM (build 25.421-b09, mixed mode)
```

- 2) If not install java jdk 1.8 and set the environment variables



- 3) Set the path variable



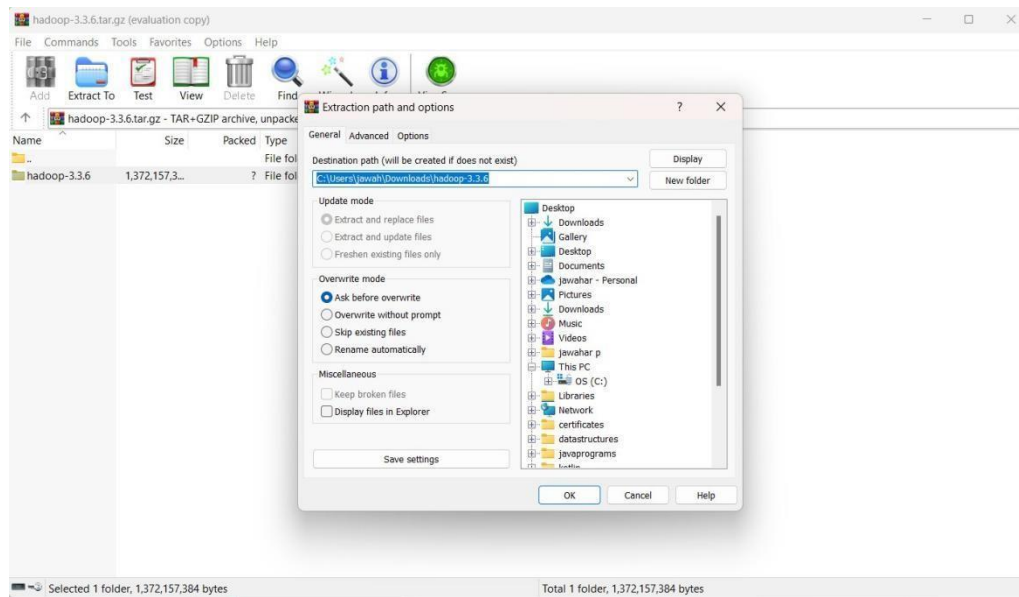
#### 2. Hadoop Installation

- 1) Install Hadoop 3.3.6 from <https://hadoop.apache.org/releases.html>

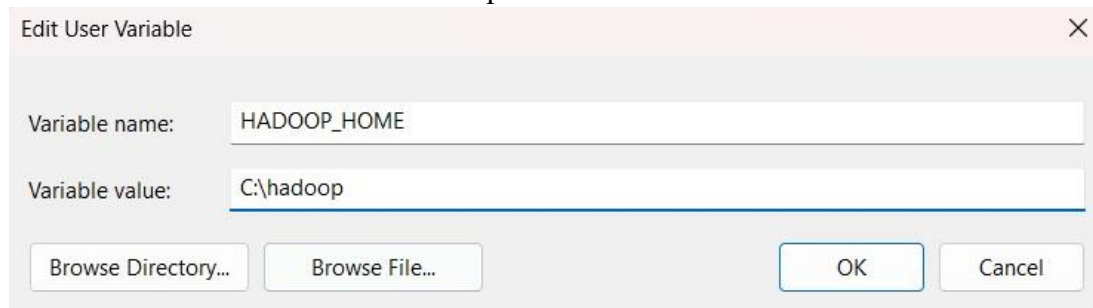
3.3.6	2023 Jun 23	source (checksum signature)	binary (checksum signature) binary-aarch64 (checksum signature)	Announcement
-------	-------------	-----------------------------	--	--------------

Download the binary(checksum signature)

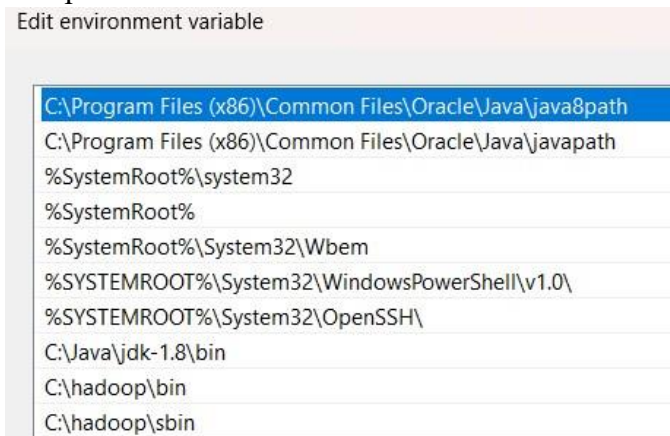
- 2) Extract the jar files to C://Hadoop



### 3) Add environment variables for Hadoop



### Add path variable



### 4) Check if Hadoop is installed successfully using the command prompt

```

C:\Windows\System32>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
    trace              view and modify Hadoop tracing settings
    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.

```

5) Thus Hadoop is installed successfully

### 3. Hadoop Configuration

- 1) Configure core-site.xml in C:\hadoop\etc\hadoop by adding
 

```

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

```
- 2) Configure the httpfs-site.xml file by adding the following xml code
 

```

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

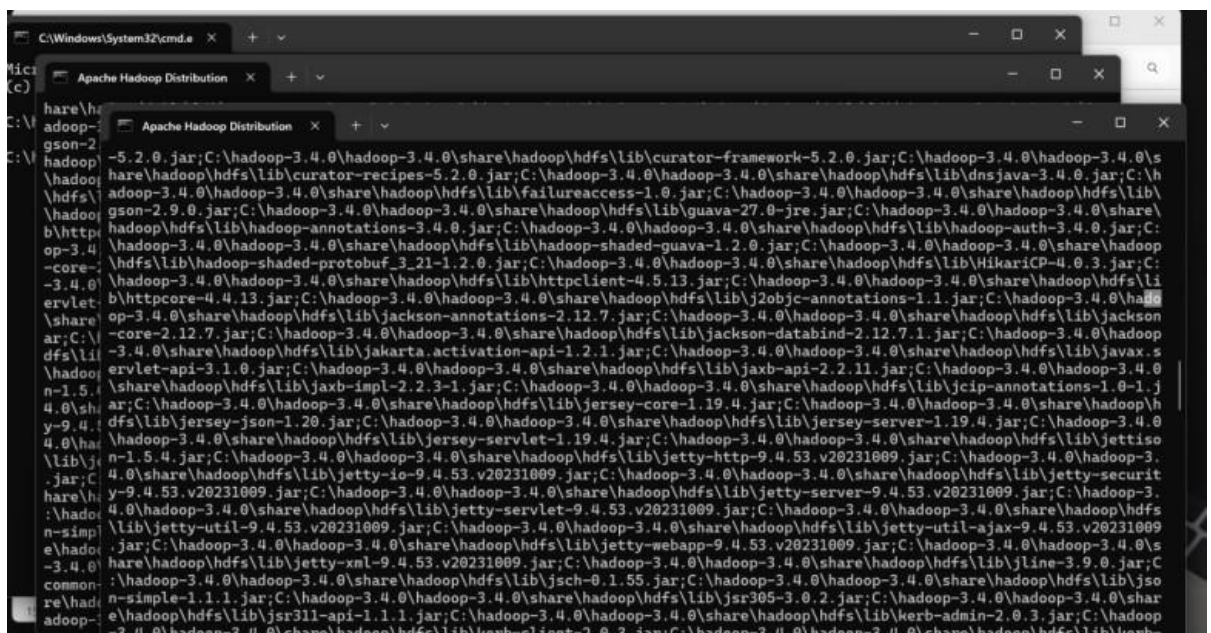
```

- ## 4. Hadoop execution

- ```
C:\Hadoop\sbin>start-dfs.cmd

C:\Hadoop\sbin>start-yarn.cmd
starting yarn daemons

C:\Hadoop\sbin>jps
13120 NameNode
2384 NodeManager
4100 DataNode
7956 ResourceManager
9124 Jps
```



2. Check if hadoop runs in localhost.  
To check this go to browser and type localhost:9870

HadoopOverviewDatanodesDatanode Volume FailuresSnapshotStartup ProgressUtilities

Overview 'localhost:9000' (active)

|                |                                                                                    |
|----------------|------------------------------------------------------------------------------------|
| Started:       | Fri Sep 13 21:19:35 +0530 2024                                                     |
| Version:       | 3.3.1, ra3b9c37a397ad4188041dd80621bdeefc46885f2                                   |
| Compiled:      | Tue Jun 15 10:43:00 +0530 2021 by ubuntu from (HEAD detached at release-3.3.1-RC3) |
| Cluster ID:    | CID-0f72c4f6-d9e3-4f2f-9b48-d38e385aaf7f                                           |
| Block Pool ID: | BP-399902486-192.168.228.238-1724038237583                                         |

Summary

Security is off.

Safemode is off.

203 files and directories, 85 blocks (85 replicated blocks, 0 erasure coded block groups) = 288 total filesystem object(s).

Heap Memory used 35.51 MB of 180.5 MB Heap Memory. Max Heap Memory is 889 MB.

Non Heap Memory used 52.64 MB of 53.69 MB Committed Non Heap Memory. Max Non Heap Memory is <unbounded>.

Non Heap Memory used 52.64 MB of 53.69 MB Committed Non Heap Memory. Max Non Heap Memory is <unbounded>.

|                                                        |                                          |
|--------------------------------------------------------|------------------------------------------|
| Configured Capacity:                                   | 475.5 GB                                 |
| Configured Remote Capacity:                            | 0 B                                      |
| DFS Used:                                              | 177.47 MB (0.04%)                        |
| Non DFS Used:                                          | 143.36 GB                                |
| DFS Remaining:                                         | 331.96 GB (69.81%)                       |
| Block Pool Used:                                       | 177.47 MB (0.04%)                        |
| DataNodes usages% (Min/Median/Max/stdDev):             | 0.04% / 0.04% / 0.04% / 0.00%            |
| <a href="#">Live Nodes</a>                             | 1 (Decommissioned: 0, In Maintenance: 0) |
| <a href="#">Dead Nodes</a>                             | 0 (Decommissioned: 0, In Maintenance: 0) |
| <a href="#">Decommissioning Nodes</a>                  | 0                                        |
| <a href="#">Entering Maintenance Nodes</a>             | 0                                        |
| <a href="#">Total Datanode Volume Failures</a>         | 0 (0 B)                                  |
| Number of Under-Replicated Blocks                      | 15                                       |
| Number of Blocks Pending Deletion (including replicas) | 0                                        |
| Block Deletion Start Time                              | Fri Sep 13 21:19:35 +0530 2024           |
| Last Checkpoint Time                                   | Fri Sep 13 21:19:36 +0530 2024           |
| Enabled Erasure Coding Policies                        | RS-6-3-1024k                             |

Result:

Thus hadoop has been installed, configured and run successfully.