

Ex 1 INSTALL CONFIGURE AND RUN HADOOP AND HDFS

Aim:

To install configure and run hadoop and hdfs.

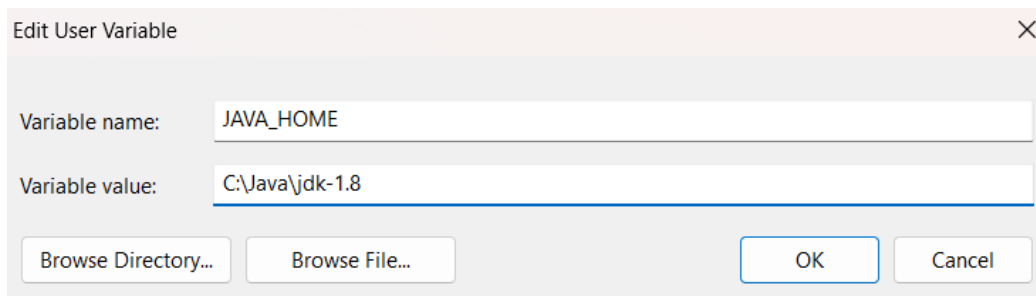
Procedure:

1. To install Java

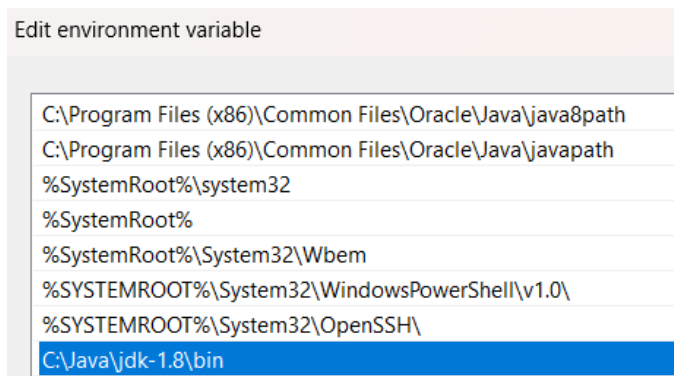
- 1) Check if java is available in the system

```
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



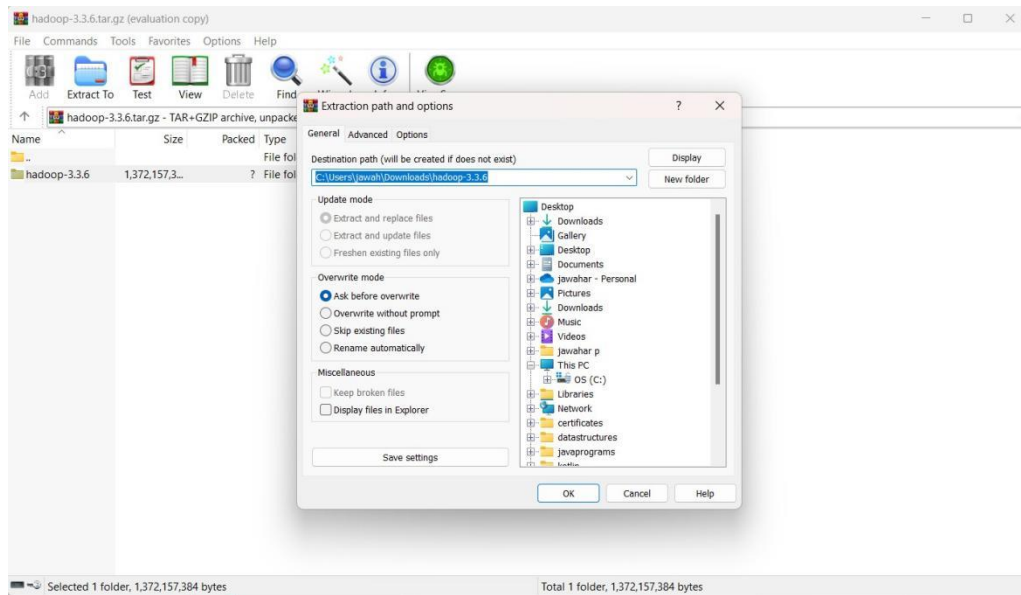
2. Hadoop Installation

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

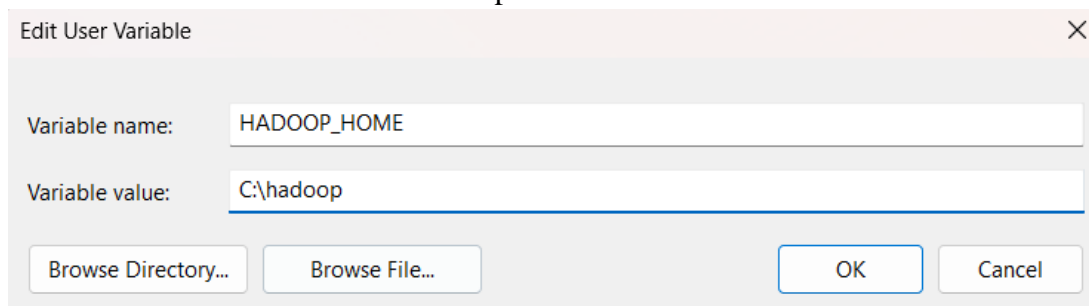
| | | | | |
|-------|-------------|-----------------------------|--|--------------|
| 3.3.6 | 2023 Jun 23 | source (checksum signature) | binary (checksum signature) binary-aarch64 (checksum signature) | Announcement |
|-------|-------------|-----------------------------|--|--------------|

Download the binary(checksum signature)

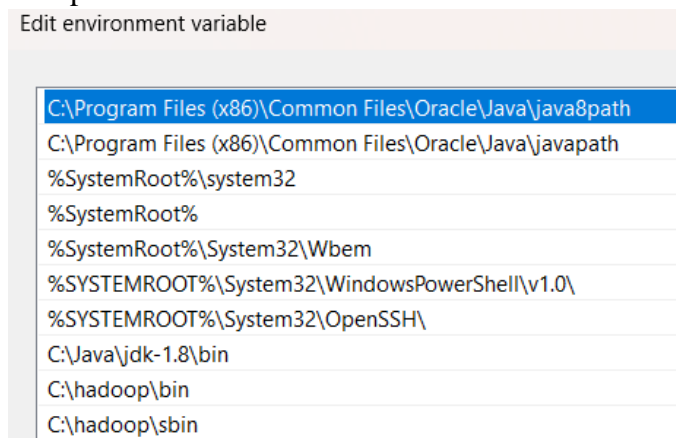
2. Extract the jar files to C://Hadoop



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>
```
3. Configure mapred-site.xml file by adding the following xml code

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

4. Configure yarn-site.xml file by adding the following xml code

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

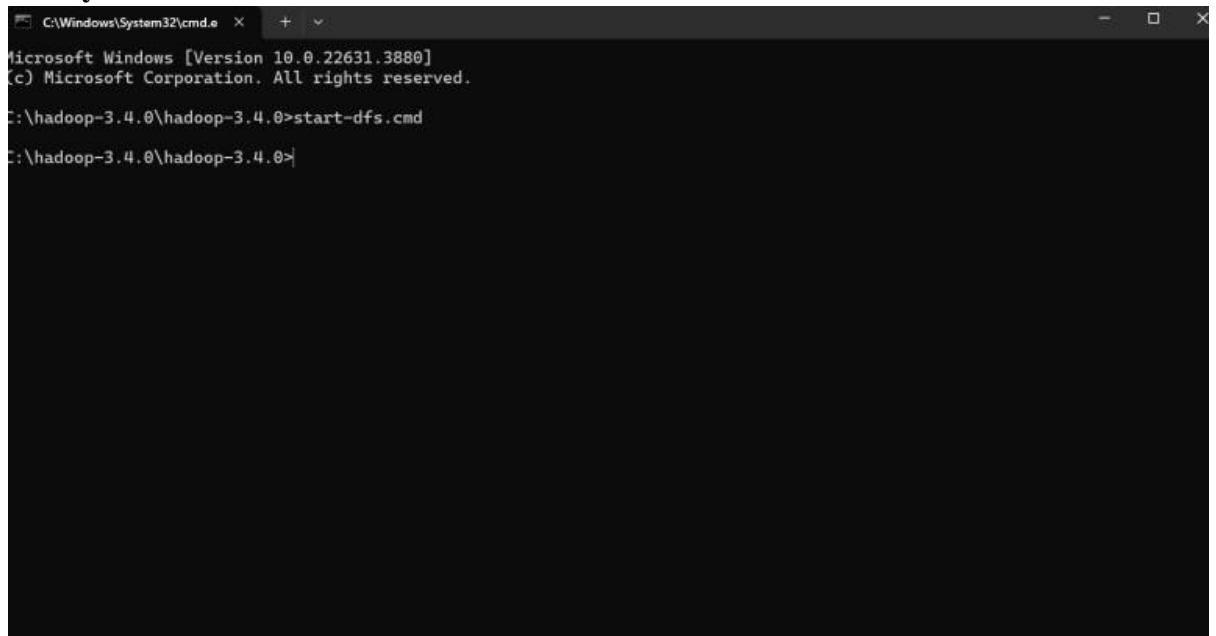
5. Change the bin shell command files.
6. Thus hadoop is configured.

4. Hadoop execution

1. To check whether hadoop is running we must start the hadoop. To start hadoop we must use the command

start-dfs.cmd

start-yarn.cmd



```
C:\Windows\System32\cmd.exe
Microsoft Windows [Version 10.0.22631.3880]
(c) Microsoft Corporation. All rights reserved.

C:\hadoop-3.4.0\hadoop-3.4.0>start-dfs.cmd

C:\hadoop-3.4.0\hadoop-3.4.0>
```

```
C:\Windows\System32\cmd.exe
Apache Hadoop Distribution
-hadoop-3.4.0\share\hadoop\hdfs\lib\curator-framework-5.2.0.jar;C:\hadoop-3.4.0\share\hadoop\hdfs\lib\dnsjava-3.4.0.jar;C:\hadoop-3.4.0\share\hadoop\hdfs\lib\failureaccess-1.0.jar;C:\hadoop-3.4.0\share\hadoop\hdfs\lib\gson-2.9.0.jar;C:\hadoop-3.4.0\share\hadoop\hdfs\lib\guava-27.0-jre.jar;C:\hadoop-3.4.0\share\hadoop\hdfs\lib\hadoop-auth-3.4.0.jar;C:\hadoop-3.4.0\share\hadoop\hdfs\lib\hadoop-shaded-guava-1.2.0.jar;C:\hadoop-3.4.0\share\hadoop\hdfs\lib\HikariCP-4.0.3.jar;C:\hadoop-3.4.0\share\hadoop\hdfs\lib\httpclient-4.5.13.jar;C:\hadoop-3.4.0\share\hadoop\hdfs\lib\httpcore-4.4.13.jar;C:\hadoop-3.4.0\share\hadoop\hdfs\lib\jackson-annotations-2.12.7.jar;C:\hadoop-3.4.0\share\hadoop\hdfs\lib\jackson-core-2.12.7.jar;C:\hadoop-3.4.0\share\hadoop\hdfs\lib\jackson-databind-2.12.7.1.jar;C:\hadoop-3.4.0\share\hadoop\hdfs\lib\jakarta.activation-api-1.2.1.jar;C:\hadoop-3.4.0\share\hadoop\hdfs\lib\javax.servlet-api-3.1.0.jar;C:\hadoop-3.4.0\share\hadoop\hdfs\lib\jaxb-api-2.2.11.jar;C:\hadoop-3.4.0\share\hadoop\hdfs\lib\jcip-annotations-1.0-1.jar;C:\hadoop-3.4.0\share\hadoop\hdfs\lib\jersey-core-1.19.4.jar;C:\hadoop-3.4.0\share\hadoop\hdfs\lib\jersey-json-1.20.jar;C:\hadoop-3.4.0\share\hadoop\hdfs\lib\jersey-server-1.19.4.jar;C:\hadoop-3.4.0\share\hadoop\hdfs\lib\jersey-servlet-1.19.4.jar;C:\hadoop-3.4.0\share\hadoop\hdfs\lib\jetty-io-9.4.53.v20231009.jar;C:\hadoop-3.4.0\share\hadoop\hdfs\lib\jetty-security-9.4.53.v20231009.jar;C:\hadoop-3.4.0\share\hadoop\hdfs\lib\jetty-server-9.4.53.v20231009.jar;C:\hadoop-3.4.0\share\hadoop\hdfs\lib\jetty-util-9.4.53.v20231009.jar;C:\hadoop-3.4.0\share\hadoop\hdfs\lib\jetty-webapp-9.4.53.v20231009.jar;C:\hadoop-3.4.0\share\hadoop\hdfs\lib\jetty-xml-9.4.53.v20231009.jar;C:\hadoop-3.4.0\share\hadoop\hdfs\lib\jline-3.9.0.jar;C:\hadoop-3.4.0\share\hadoop\hdfs\lib\jsch-0.1.55.jar;C:\hadoop-3.4.0\share\hadoop\hdfs\lib\jsr305-3.0.2.jar;C:\hadoop-3.4.0\share\hadoop\hdfs\lib\kerb-admin-2.0.3.jar;C:\hadoop-3.4.0\share\hadoop\hdfs\lib\kerb-client-2.0.3.jar;C:\hadoop-3.4.0\share\hadoop\hdfs\lib\kerb-
```

- To check whether both namenode and datanode are running using the command **jps**

```
C:\>jps
17248 Jps
23168 NameNode
26468 ResourceManager
22104 DataNode
16412 NodeManager
```

- Check if hadoop runs in localhost.

To check this go to browser and type localhost:9870

Overview 'localhost:9000' (✓active)

| | |
|----------------|--|
| Started: | Sun Sep 08 00:07:23 +0530 2024 |
| Version: | 3.3.6, r1be78238728da9266a4f8195058f08fd012bffc |
| Compiled: | Sun Jun 18 13:52:00 +0530 2023 by ubuntu from (HEAD detached at release-3.3.6-RC1) |
| Cluster ID: | CID-ec9f7235-3052-4054-8c6a-e321b866745f |
| Block Pool ID: | BP-195161732-192.168.29.226-1724686865039 |

Summary

Security is off.
Safemode is off.

124 files and directories, 60 blocks (60 replicated blocks, 0 erasure coded block groups) = 184 total filesystem object(s).

Heap Memory used 132.02 MB of 238.5 MB Heap Memory. Max Heap Memory is 889 MB.

Non Heap Memory used 69.74 MB of 71.13 MB Committed Non Heap Memory. Max Non Heap Memory is <unbounded>.

| | |
|-----------------------------|--------------------|
| Configured Capacity: | 475.29 GB |
| Configured Remote Capacity: | 0 B |
| DFS Used: | 101.26 MB (0.02%) |
| Non DFS Used: | 74.64 GB |
| DFS Remaining: | 400.56 GB (84.28%) |

Thus hadoop runs successfully

Result:

Thus hadoop is installed, configured and run successfully.

