

## EXP.1: Downloading and installing Hadoop, Understanding different Hadoop modes, Startup scripts, Configuration files.

### AIM:


To Download and install Hadoop, Understanding different Hadoop modes, Startup scripts, Configuration files.

### PROCEDURE:

#### 1. Install Java 8: Download Java 8 from the link:

<http://www.oracle.com/technetwork/java/javase/downloads/jdk8-downloads-2133151.html>

- a. Set environmental variables:
  - i. User variable:
    - Variable: JAVA\_HOME
    - Value: C:/java
  - ii. System variable:
    - Variable: PATH
    - Value: C:/java/bin
- b. Check on cmd, see below:



```
ca Command Prompt
Microsoft Windows [Version 10.0.19045.4780]
(c) Microsoft Corporation. All rights reserved.

C:\Users\user>cd \

C:\>java -version
java version "1.8.0_202"
Java(TM) SE Runtime Environment (build 1.8.0_202-b08)
Java HotSpot(TM) 64-Bit Server VM (build 25.202-b08, mixed mode)

C:\>
```

#### 2. Download Hadoop-3.3.6: download Hadoop 3.3.6 from the link:

<http://www.apache.org/dyn/closer.cgi/hadoop/common/hadoop-3.3.6/hadoop-3.3.6.tar.gz>

- a. Put extracted Hadoop-3.3.6 files into C drive. Note that do not put these extracted files into Cdrive, where you installed your Windows.
- b. Download “*hadoop-common-3.3.6-bin-master*” from the link: <https://github.com/amihalik/hadoop-common-3.3.6-bin/tree/master/bin>. You will see 11 filesthere. Paste all these files into the “bin” folder of Hadoop-3.3.6.
- c. Create a “data” folder inside Hadoop-3.3.6, and also create two more folders in the “data” folderas “data” and “name.”
- d. Create a folder to store temporary data during execution of a project, such as “D:/hadoop/temp.”
- e. Create a log folder, such as “D:/hadoop/userlog”
- f. Go to Hadoop-3.3.6 →etc → Hadoop and edit four files:
  - i. core-site.xml
  - ii. hdfs-site.xml
  - iii. mapred-site.xml
  - iv. yarn-site.xml

**core-site.xml:**

```
<?xml version="1.0" encoding="UTF-8"?>
<?xml-stylesheet type="text/xsl" href="configuration.xsl"?>
<configuration>
  <property>
    <name>fs.defaultFS</name>
    <value>hdfs://localhost:9000</value>
  </property>
</configuration>
```

**hdfs-site.xml:**

```
<?xml version="1.0" encoding="UTF-8"?>
<?xml-stylesheet type="text/xsl" href="configuration.xsl"?>
<configuration>
  <property>
    <name>dfs.replication</name>
    <value>1</value>
  </property>
  <property>
    <name>dfs.namenode.name.dir</name>
    <value>file:///C:/hadoop-3.3.6/data/namenode</value>
  </property>
  <property>
    <name>dfs.datanode.data.dir</name>
    <value>file:///C:/hadoop-3.3.6/data/datanode</value>
  </property>
</configuration>
```

**mapred-site.xml:**

```
<?xml version="1.0"?>
<?xml-stylesheet type="text/xsl" href="configuration.xsl"?>
<configuration>
  <property>
    <name>mapreduce.framework.name</name>
    <value>yarn</value>
  </property>
</configuration>
```

**yarn-site.xml:**

```
<?xml version="1.0"?>
<configuration>
  <property>
    <name>yarn.nodemanager.aux-services</name>
    <value>mapreduce_shuffle</value>
  </property>
</configuration>
```

- g. Go to the location: “Hadoop-3.3.6→etc→hadoop,” and edit “hadoop-env.cmd” by writingset  
JAVA\_HOME=C:/java/jdk1.8.0\_202
- h. Set environmental variables: Do: My computer → Properties → Advance  
system settings →  
Advanced → Environmental variables
  - i. User variables:
    - Variable: HADOOP\_HOME
    - Value: C:/hadoop-3.3.6
  - ii. System variable
    - Variable: Path
    - Value: C:/hadoop-3.3.6/bin  
C:/hadoop-3.3.6/sbin  
C:/hadoop-3.3.6/share/hadoop/common/\*  
C:/hadoop-3.3.6/share/hadoop/hdfs  
C:/hadoop-3.3.6/share/hadoop/hdfs/lib/\*  
C:/hadoop-3.3.6/share/hadoop/hdfs/\*  
C:/hadoop-3.3.6/share/hadoop/yarn/lib/\*  
C:/hadoop-3.3.6/share/hadoop/yarn/\*  
C:/hadoop-3.3.6/share/hadoop/mapreduce/lib/\*  
C:/hadoop-3.3.6/share/hadoop/mapreduce/\*  
C:/hadoop-3.3.6/share/hadoop/common/lib/\*
- i. Check on cmd; see below.

```
C:\Windows\System32>hadoop version
Hadoop 3.3.6
Source code repository https://github.com/apache/hadoop.git -r 1be78238728da9266a4f88195058f08fd012bf9c
Compiled by ubuntu on 2023-06-18T08:22Z
Compiled on platform linux-x86_64
Compiled with protoc 3.7.1
From source with checksum 5652179ad55f76cb287d9c633bb53bbd
This command was run using /C:/hadoop-3.3.6/share/hadoop/common/hadoop-common-3.3.6.jar
```

- j. **Format name-node:** On cmd go to the location “Hadoop-2.6.0→bin” by  
writing on cmd “cdhadoop-2.6.0.\bin” and then “hdfs namenode –format”
- k. Start Hadoop. Go to the location: “D:\hadoop-2.6.0\sbin.” Run the following  
files as administrator “start-all.cmd”.
- l. Go to your web browser and search “localhost:9870” to access Hadoop  
NameNode. For Resource Manager, search “localhost:8088”.

```
C:\> Administrator: Command Prompt
Microsoft Windows [Version 10.0.19045.4780]
(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
11104 Jps
12868 DataNode
11288 ResourceManager
12456 NodeManager
5596 NameNode

C:\WINDOWS\system32>
```

[illegible]



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

Started:	Tue Sep 17 12:39:58 +0530 2024
Version:	3.3.6, r1be78238728da9266e4f88195058f08fd012bf9c
Compiled:	Sun Jun 18 13:52:00 +0530 2023 by ubuntu from (HEAD detached at release-3.3.6-RC1)
Cluster ID:	CID-756fa2e5-b7de-42f0-9099-9d114f91cbb7
Block Pool ID:	BP-1161759510-192.168.56.1-1723520378615

**Summary**

Security is off.  
 Safemode is off.  
 70 files and directories, 37 blocks (37 replicated blocks, 0 erasure coded block groups) = 107 total filesystem object(s).  
 Heap Memory used 60.4 MB of 274.5 MB Heap Memory. Max Heap Memory is 889 MB.  
 Non Heap Memory used 64.77 MB of 66.38 MB Committed Non Heap Memory. Max Non Heap Memory is <unbounded>.

Configured Capacity:	460.59 GB
Configured Remote Capacity:	0 B
DFS Used:	42.98 MB (0.01%)
Non DFS Used:	298.35 GB
DFS Remaining:	162.2 GB (35.21%)
Block Pool Used:	42.98 MB (0.01%)

**All Applications**

Cluster Metrics

Apps Submitted	Apps Pending	Apps Running	Apps Completed	Containers Running	Used Resources
1	0	0	1	0	<memory 0 B, vCores 0>

Cluster Nodes Metrics

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

Scheduler Metrics

Scheduler Type	Scheduling Resource Type	Minimum Allocation	Maximum Allocation
Capacity Scheduler	[memory-mb (unit=Mi), vcores]	<memory 1024, vCores 1>	<memory 8192, vCores 4>

Show 20 entries

ID	User	Name	Application Type	Application Tags	Queue	Application Priority	StartTime	LaunchTime	FinishTime	State	FinalStatus
application_1726557002070_0001	Navneeth	streamjob7904707133888847456 jar	MAPREDUCE		default	0	Tue Sep 17 12:44:44 +0550 2024	Tue Sep 17 12:44:54 +0550 2024	Tue Sep 17 12:47:22 +0550 2024	FINISHED	FAILED

Showing 1 to 1 of 1 entries

## RESULT:

The step-by-step installation and configuration of Hadoop on Windows system have been successfully completed.