Ex No 1 Roll No: 210701173

## 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:**

 Install Java 8: Download Java 8 from the link: <a href="http://www.oracle.com/technetwork/java/javase/downloads/jdk8-downloads-2133151.html">http://www.oracle.com/technetwork/java/javase/downloads/jdk8-downloads-2133151.html</a>

- 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:

```
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: <a href="http://www.apache.org/dyn/closer.cgi/hadoop/common/hadoop-3.3.6/hadoop-3.3.6.tar.gz">http://www.apache.org/dyn/closer.cgi/hadoop/common/hadoop-3.3.6/hadoop-3.3.6.tar.gz</a>
  - 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: <a href="https://github.com/amihalik/hadoop-common-3.3.6-bin/tree/master/bin">https://github.com/amihalik/hadoop-common-3.3.6-bin/tree/master/bin</a>. 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 \square etc \square 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>
  <name>dfs.datanode.data.dir</name>
  <value>file:///C:/hadoop-3.3.6/data/datanode/value>
 </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>
 </configuration>
yarn-site.xml:
<?xml version="1.0"?>
<configuration>
property>
  <name>yarn.nodemanager.aux-services</name>
  <value>mapreduce shuffle</value>
 </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 2
                             02
         h. Set environmental variables: Do: My computer \square Properties \square
            Advance system settings □
            Advanced 

Environmental variables
                  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".
- 1. Go to your web browser and search "localhost:9870" to access Hadoop NameNode. For Resource Manager, search "localhost:8088".

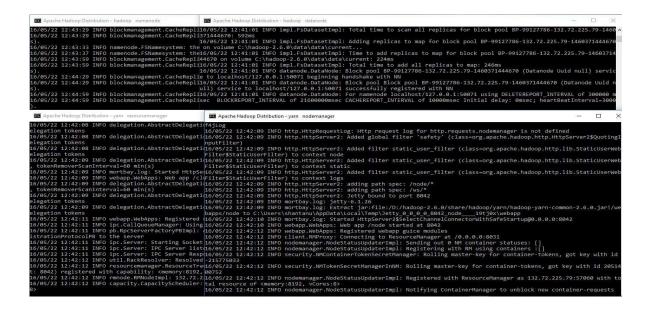
```
C. Windows (System 12) bifs namenode - forward 12) bifs na
```

```
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>
```





## **RESULT:**

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