NANDHAKUMAR S
210701172

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:

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>
```

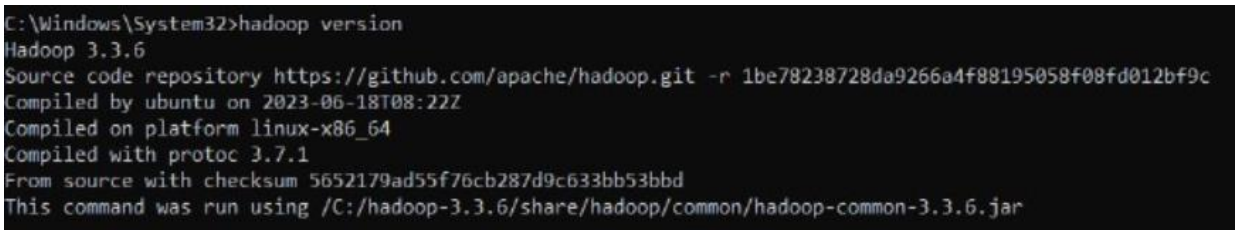
yarn-site.xml:

```
<?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.

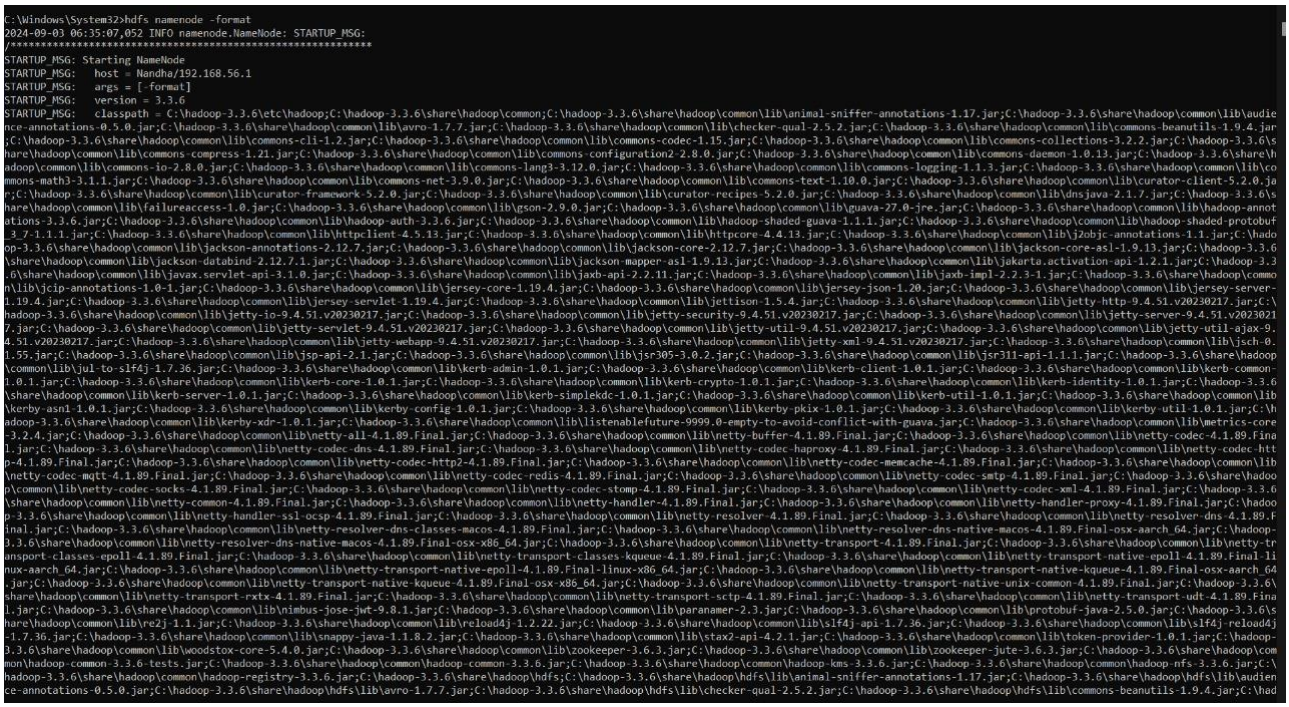


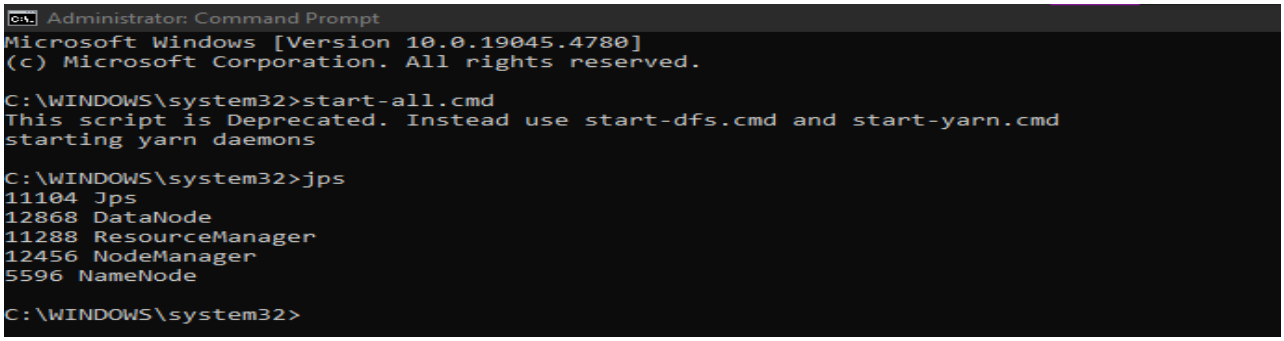
j. **Format name-node**: On cmd go to the location "Hadoop-2.6.0→bin" by writing on cmd "cdhadoop-2.6.0.\bin" and then "hdfs namenode –format"

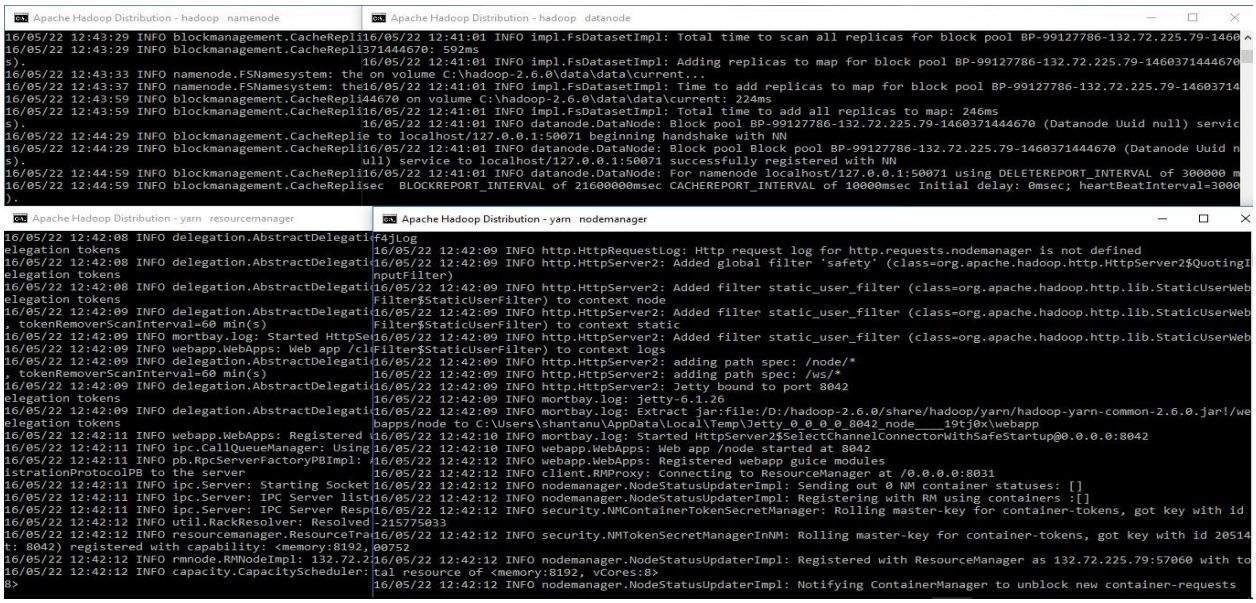
k. Start Hadoop. Go to the location: "D:\hadoop-2.6.0\sbin." Run the following

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”.

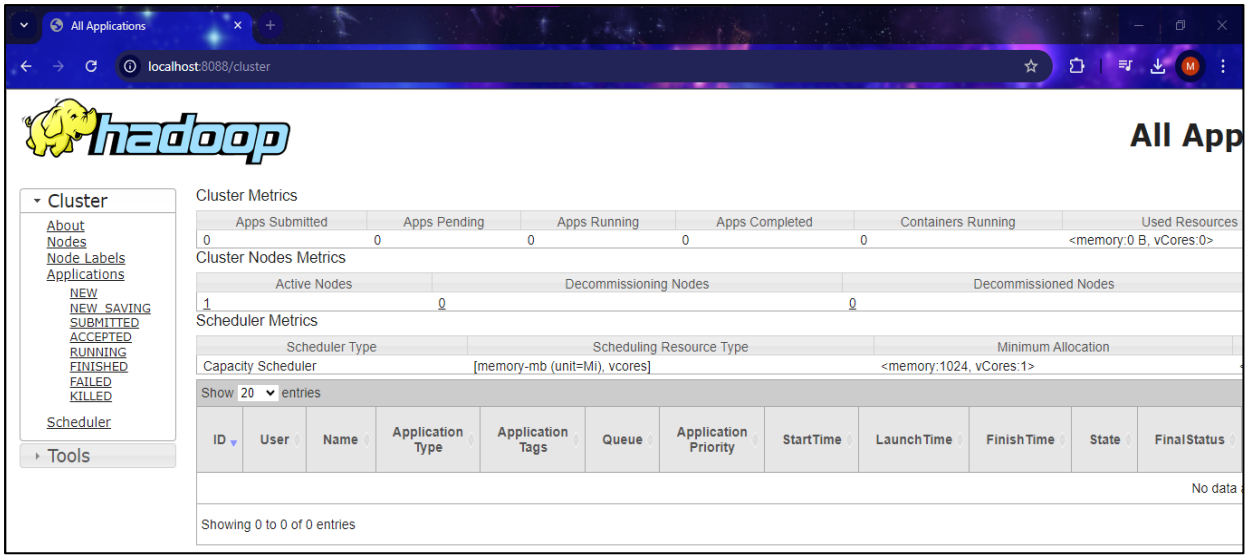





Apache Hadoop Distribution - hadoop namenode
Apache Hadoop Distribution - hadoop datanode
Apache Hadoop Distribution - yarn resourcemanager
Apache Hadoop Distribution - yarn nodemanager


Namenode information
localhost:9870/dfshealth.html#tab-overview
Hadoop
Overview
Datanodes
Datanode Volume Failures
Snapshot
Startup Progress
Utilities
Overview 'localhost:9000' (active)
Started: Sun Sep 08 00:06:16 +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-50e55097-947b-42ab-ac22-401506263472
Block Pool ID: BP-1769990285-192.168.56.1-1724916758896
Summary
Security is off.
Safemode is off.
75 files and directories, 27 blocks (27 replicated blocks, 0 erasure coded block groups) = 102 total filesystem object(s).

NANDHAKUMAR S
210701172



RESULT:

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