

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:



```

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 C drive, 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 files there. Paste all these files into the “bin” folder of Hadoop3.3.6.
- c. Create a “data” folder inside Hadoop-3.3.6, and also create two more folders in the “data” folder as “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 writing set
 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.

- [illegible]

```
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>
```


Namenode information

localhost:9870/dfshealth.html#tab-overview

Hadoop

Overview

Datanodes

Datanode Volume Failures

Snapshot

Startup Progress

Utilities

Overview 'localhost:9000' (✓active)

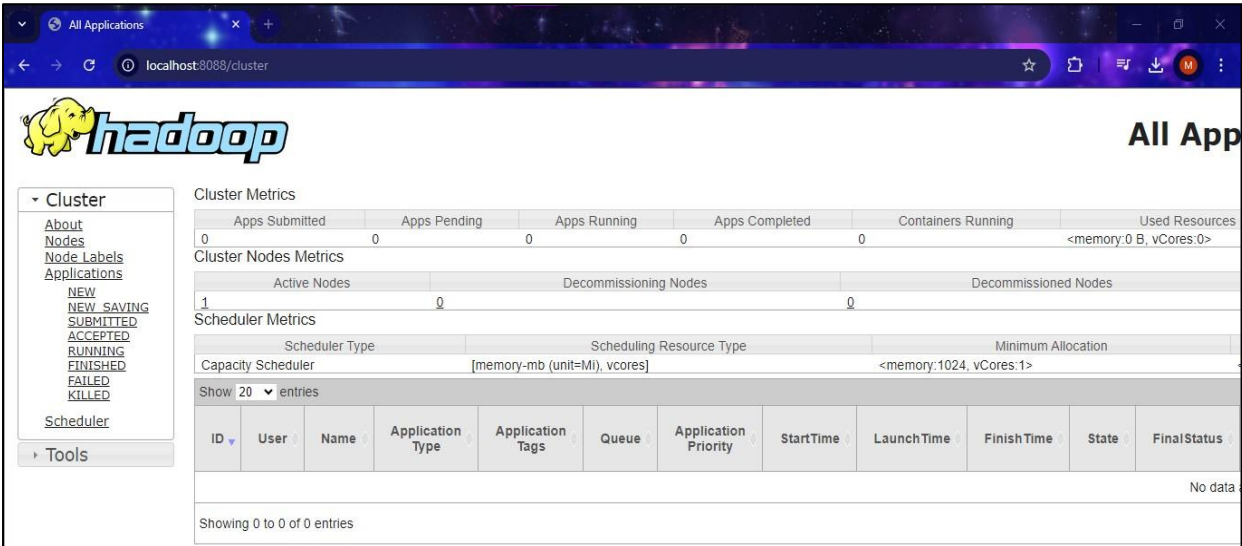
Started:	Sun Sep 08 00:06:16 +0530 2024
Version:	3.3.6, r1be78238728da9266a4f88195058f08fd012bf9c
Compiled:	Sun Jun 18 13:52:00 +0530 2023 by ubuntu from (HEAD detached at release-3.3.6-RC1)
Cluster ID:	CID-50e55097-947b-42ab-ac22-401506263472
Block Pool ID:	BP-1769990285-192.168.56.1-1724916758896

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).



The screenshot displays the Hadoop web interface in a browser window. The address bar shows 'localhost:8088/cluster'. The page features the Hadoop logo and a navigation menu on the left with options like 'Cluster', 'About', 'Nodes', 'Node Labels', 'Applications', and 'Scheduler'. The main content area is titled 'All App' and contains several sections:

- Cluster Metrics:** A table showing 'Apps Submitted' (0), 'Apps Pending' (0), 'Apps Running' (0), 'Apps Completed' (0), 'Containers Running' (0), and 'Used Resources' (<memory:0 B, vCores:0>).
- Cluster Nodes Metrics:** A table showing 'Active Nodes' (1), 'Decommissioning Nodes' (0), and 'Decommissioned Nodes' (0).
- Scheduler Metrics:** A table showing 'Scheduler Type' (Capacity Scheduler), 'Scheduling Resource Type' ([memory-mb (unit=Mi), vcores]), and 'Minimum Allocation' (<memory:1024, vCores:1>).
- Applications Table:** A table with columns: ID, User, Name, Application Type, Application Tags, Queue, Application Priority, StartTime, LaunchTime, FinishTime, State, and FinalStatus. It shows 'No data' and 'Showing 0 to 0 of 0 entries'.

RESULT:

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