Ex. No. 8

SET UP A SINGLE HADOOP CLUSTER AND SHOW THE PROCESS USING WEB UI

AIM:

To set up a single hadoop cluster and show the process using web UI.

PROCEDURE:

Installation steps:

Step 1: Download and install Java

Hadoop is built on Java, so you must have Java installed on your PC. You can get the most recent version of Java from the official website. After downloading, follow the installation wizard to install Java on your system.

JDK: https://www.oracle.com/java/technologies/javase-downloads.html

Step 2: Download Hadoop

Hadoop can be downloaded from the Apache Hadoop website. Make sure to have the latest stable release of Hadoop. Once downloaded, extract the contents to a convenient location.

Hadoop: https://hadoop.apache.org/releases.html

Step 3: Set Environment Variables

You must configure environment variables after downloading and unpacking Hadoop. Launch the Start menu, type “Edit the system environment variables,” and select the result. This will launch the System Properties dialogue box. Click on “Environment Variables” button to open.

210701112

Click “New” under System Variables to add a new variable. Enter the variable name “HADOOP_HOME” and the path to the Hadoop folder as the variable value. Then press “OK.”

Then, under System Variables, locate the “Path” variable and click “Edit.” Click “New” in the Edit

Environment Variable window and enter “%HADOOP_HOME%bin” as the variable value. To close all the windows, use the “OK” button.

Step 4: Setup Hadoop

You must configure Hadoop in this phase by modifying several configuration files. Navigate to the “etc/hadoop” folder in the Hadoop folder. You must make changes to three files:

core-site.xml

hdfs-site.xml

mapred-site.xml

Open each file in a text editor and edit the following properties:

In core-site.xml

```
<configuration>
 <property>
  <name>fs.default.name</name>
  <value>hdfs://localhost:9000</value>
 </property>
</configuration>
```

210701112

In hdfs-site.xml

<configuration>

<property>

<name>dfs.replication</name>

<value>1</value>

</property>

<property>

<name>dfs.namenode.name.dir</name>

<value>file:/hadoop-3.3.1/data/namenode</value>

</property>

<property>

<name>dfs.datanode.data.dir</name>

<value>file:/hadoop-3.3.1/data/datanode</value>

</property>

</configuration>

In mapred-site.xml

<configuration>

<property>

<name>mapred.job.tracker</name>

<value>localhost:54311</value>

</property>

</configuration>

Save the changes in each file.

Step 5: Format Hadoop NameNode

You must format the NameNode before you can start Hadoop. Navigate to the Hadoop bin folder using a command prompt. Execute this command:

210701112

hdfs namenode -format

Step 6: Start Hadoop

To start Hadoop, open a command prompt and navigate to the Hadoop bin folder. Run the following command:

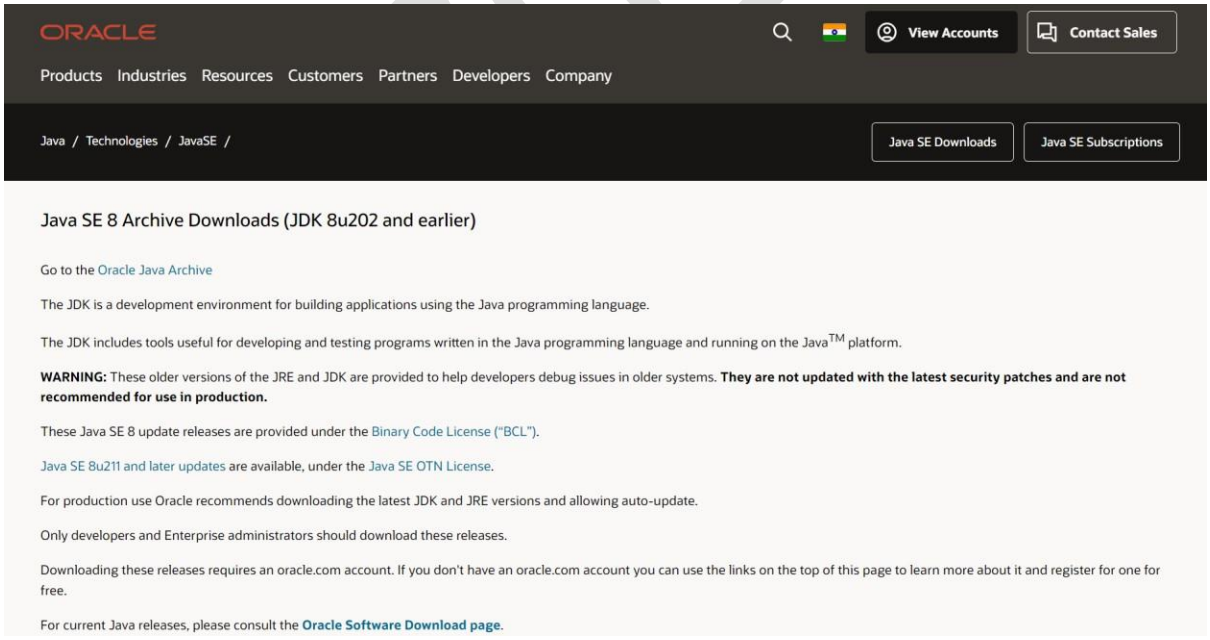
start-dfs.cmd start-yarn.cmd

This command will start all the required Hadoop services, including the NameNode, DataNode, and JobTracker. Wait for a few minutes until all the services are started.

Step 7: Verify Hadoop Installation

To ensure that Hadoop is properly installed, open a web browser and go to http://localhost:9870. This will launch the web interface for the Hadoop NameNode. You should see a page with Hadoop cluster information.

OUTPUT:



Apache Hadoop Download Documentation Community Development Help Apache Software Foundation

Download

Hadoop is released as source code tarballs with corresponding binary tarballs for convenience. The downloads are distributed via mirror sites and should be checked for tampering using GPG or SHA-512.

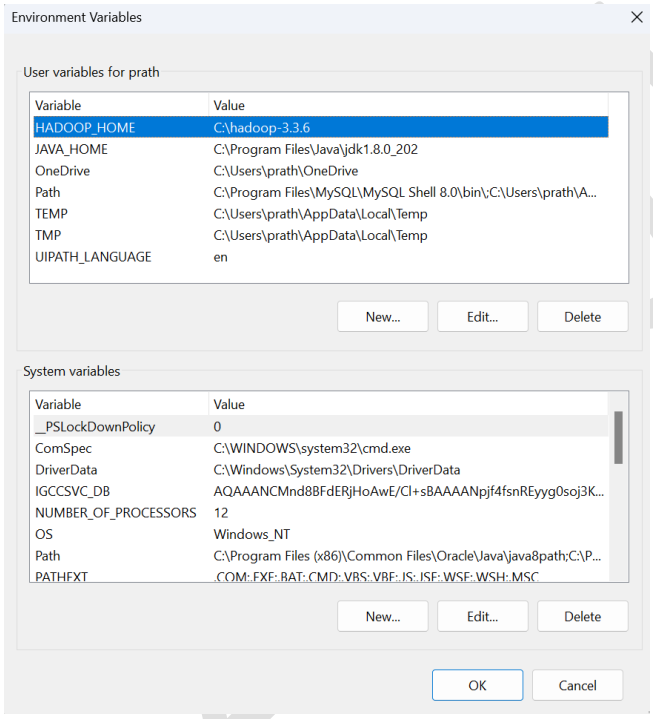
Version	Release date	Source download	Binary download	Release notes
3.4.0	2024 Mar 17	source (checksum signature)	binary (checksum signature) binary-aarch64 (checksum signature)	Announcement
3.3.6	2023 Jun 23	source (checksum signature)	binary (checksum signature) binary-aarch64 (checksum signature)	Announcement
2.10.2	2022 May 31	source (checksum signature)	binary (checksum signature)	Announcement

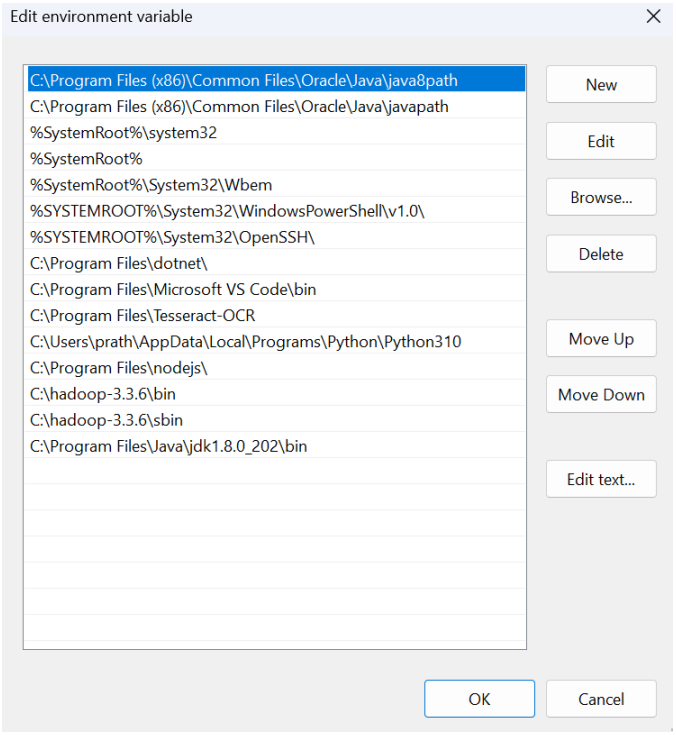
To verify Apache Hadoop® releases using GPG:

1. Download the release hadoop-X.Y.Z-src.tar.gz from a mirror site.
2. Download the signature file hadoop-X.Y.Z-src.tar.gz.asc from Apache.
3. Download the Hadoop KEYS file.
4. gpg --import KEYS
5. gpg --verify hadoop-X.Y.Z-src.tar.gz.asc

To perform a quick check using SHA-512:

1. Download the release hadoop-X.Y.Z-src.tar.gz from a mirror site.
2. Download the checksum hadoop-X.Y.Z-src.tar.gz.sha512 or hadoop-X.Y.Z-src.tar.gz.mds from Apache.
3. shasum -a 512 hadoop-X.Y.Z-src.tar.gz



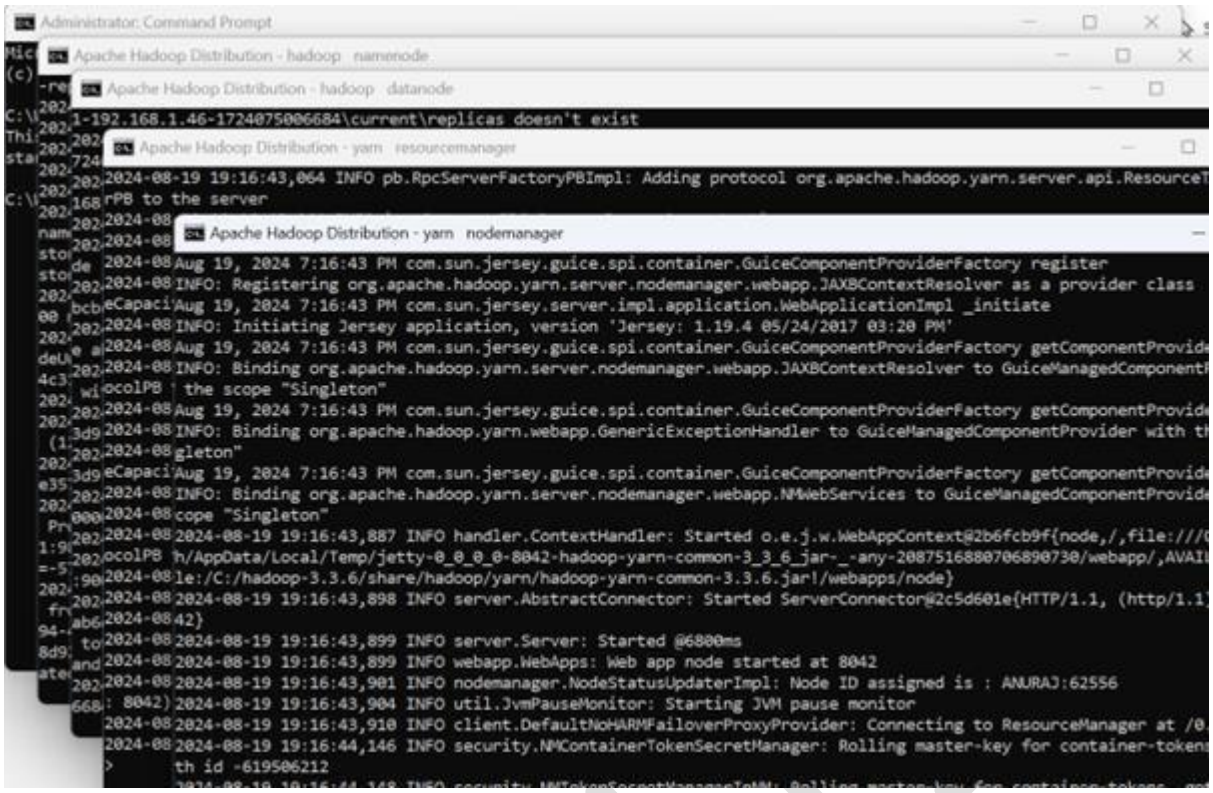



```
C:\Users\prath>hadoop version
Hadoop 3.3.6
Source code repository https://github.com/apache/hadoop.git -r 1be78238728da9266a4f88195058f08fd012bf9c
Compiled by ubuntu on 2023-06-18T08:22Z
Compiled on platform linux-x86_64
Compiled with protoc 3.7.1
From source with checksum 5652179ad55f76cb287d9c633bb53bbd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

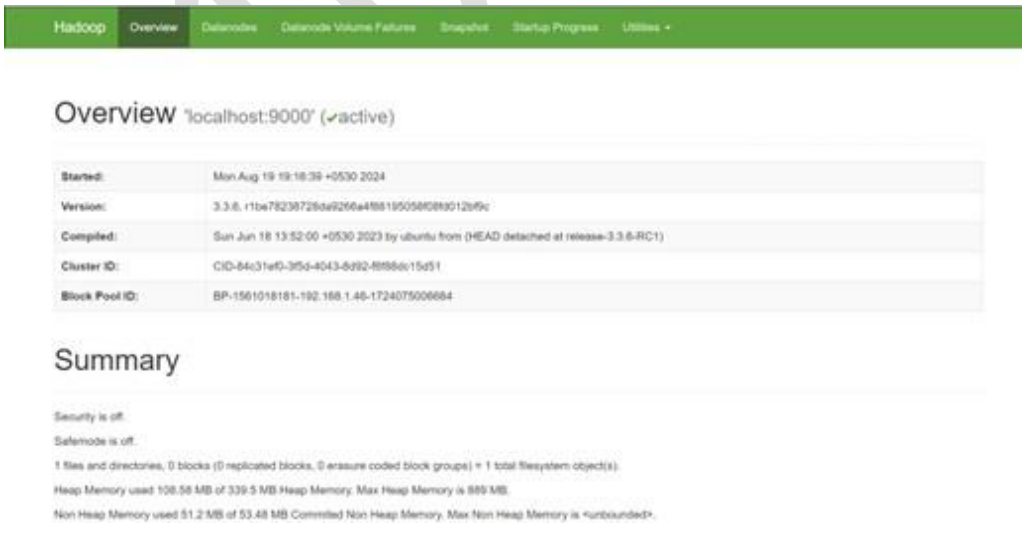
```
C:\Users\prath>java -version
java version "1.8.0_421"
Java(TM) SE Runtime Environment (build 1.8.0_421-b09)
Java HotSpot(TM) Client VM (build 25.421-b09, mixed mode, sharing)
```

```
C:\Users\prath>hdfs namenode -format
2024-08-19 19:13:25,285 INFO namenode.NameNode: STARTUP_MSG:
/************************************************************
STARTUP_MSG: Starting NameNode
```

```
2024-08-19 19:13:26,652 INFO util.GSet: Computing capacity for map NameNodeRetryCache
2024-08-19 19:13:26,652 INFO util.GSet: VM type       = 64-bit
2024-08-19 19:13:26,652 INFO util.GSet: 0.029999999329447746% max memory 889 MB = 273.1 KB
2024-08-19 19:13:26,652 INFO util.GSet: capacity      = 2^15 = 32768 entries
2024-08-19 19:13:26,684 INFO namenode.FSImage: Allocated new BlockPoolId: BP-1561018181-192.168.1.46-1724075006684
2024-08-19 19:13:26,746 INFO common.Storage: Storage directory C:\hadoop-3.3.6\data\namenode has been successfully forma
tted.
```




```
C:\Windows\System32>jps
14256 NodeManager
21760 Jps
14100 DataNode
20552 ResourceManager
22584 NameNode
```



210701112

RESULT:

Thus, to set up a single hadoop cluster and show the process using web UI was completed successfully.