

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.

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>
```

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:

```
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:

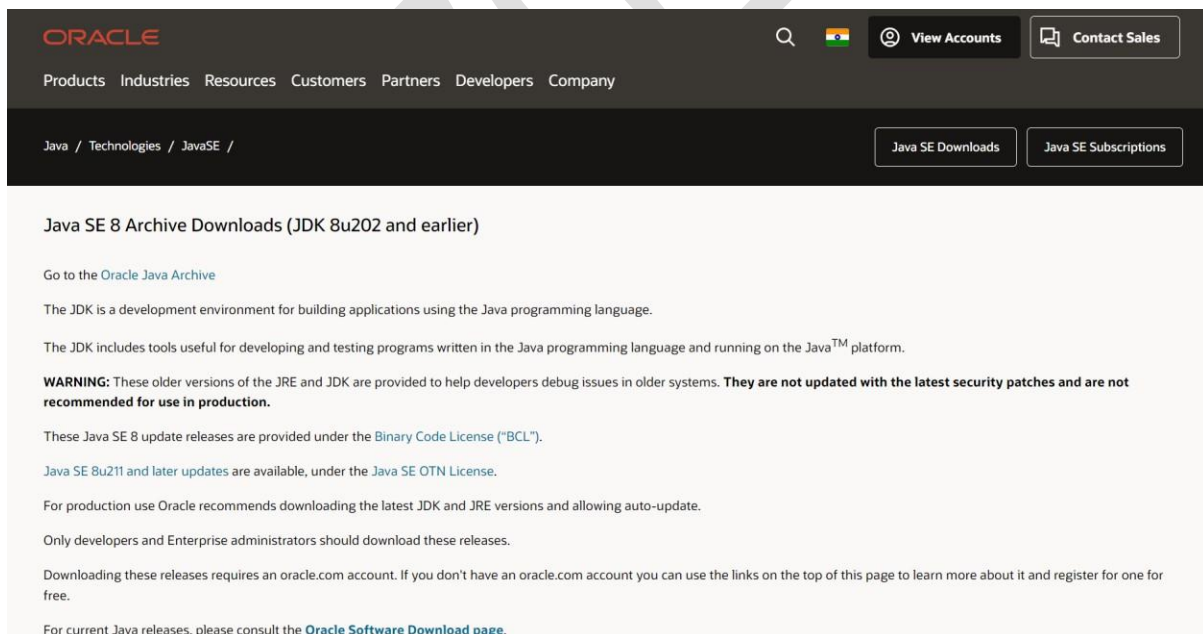
```
start-dfs.cmd start-yarn.cmd
```

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:



The screenshot shows the Oracle website's navigation bar with the Oracle logo, a search icon, a flag for India, and links for 'View Accounts' and 'Contact Sales'. Below the navigation bar is a breadcrumb trail: 'Java / Technologies / JavaSE /'. To the right of the breadcrumb are two buttons: 'Java SE Downloads' and 'Java SE Subscriptions'. The main content area is titled 'Java SE 8 Archive Downloads (JDK 8u202 and earlier)'. It includes a link to 'Go to the Oracle Java Archive', a description of the JDK as a development environment, and a warning that these older versions are not updated with the latest security patches and are not recommended for use in production. It also mentions that these releases are provided under the Binary Code License (BCL) and that Java SE 8u211 and later updates are available under the Java SE OTN License. A note states that for production use, Oracle recommends downloading the latest JDK and JRE versions and allowing auto-update. It also mentions that only developers and Enterprise administrators should download these releases. A final note states that downloading these releases requires an oracle.com account, and if you don't have one, you can use the links on the top of this page to learn more about it and register for one for free. At the bottom, it says 'For current Java releases, please consult the Oracle Software Download page.'

 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`

Environment Variables

User variables for prath

Variable	Value
HADOOP_HOME	C:\hadoop-3.3.6
JAVA_HOME	C:\Program Files\Java\jdk1.8.0_202
OneDrive	C:\Users\prath\OneDrive
Path	C:\Program Files\MySQL\MySQL Shell 8.0\bin\;C:\Users\prath\A...
TEMP	C:\Users\prath\AppData\Local\Temp
TMP	C:\Users\prath\AppData\Local\Temp
UIPATH_LANGUAGE	en

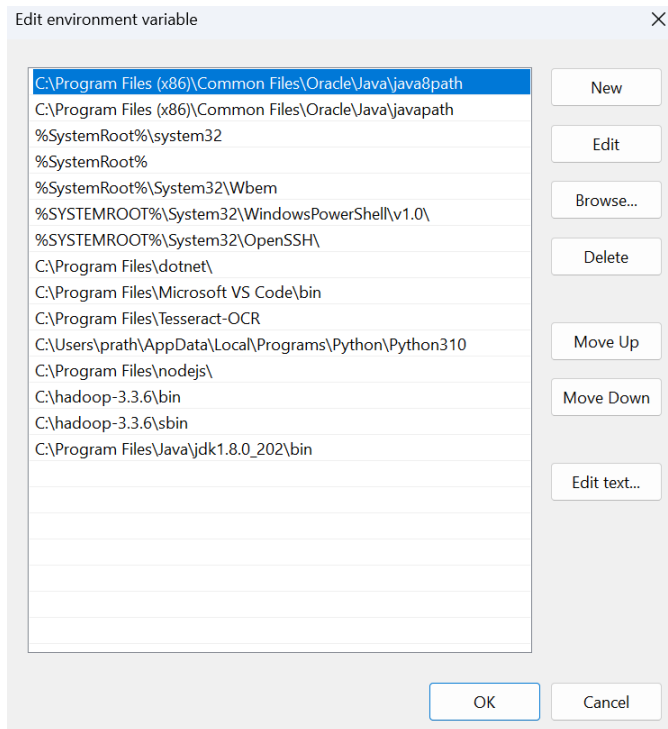
New... Edit... Delete

System variables

Variable	Value
_PSLockDownPolicy	0
ComSpec	C:\WINDOWS\system32\cmd.exe
DriverData	C:\Windows\System32\Drivers\DriverData
IGCCSVC_DB	AQAAANCMnd8BFdERjHoAwE/CI+sBAAAAAnpjf4fsnREyyg0soj3K...
NUMBER_OF_PROCESSORS	12
OS	Windows_NT
Path	C:\Program Files (x86)\Common Files\Oracle\Java\java8path;C\...
PATHEXT	.COM;.EXE;.BAT;.CMD;.VBS;.VBE;.JS;.JSE;.WSF;.WSH;.MSC

New... Edit... Delete

OK Cancel



```
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 formatted.
```

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

[Hadoop](#)
[Overview](#)
[DataNodes](#)
[DataNode Volume Failures](#)
[Snapshots](#)
[Startup Progress](#)
[Utilities](#)

Overview `"localhost:9000"` (✓active)

Started:	Mon Aug 19 19:18:39 +0530 2024
Version:	3.3.6, r1be78238728da9266a488195058089012b99c
Compiled:	Sun Jun 18 13:52:00 +0530 2023 by ubuntu from (HEAD detached at release-3.3.6-RC1)
Cluster ID:	CID-64c31e0-3f5d-4043-8992-8f88dc15d51
Block Pool ID:	BP-1561018181-192.168.1.48-1724075006684

Summary

Summary

Security is off.
 Safe mode is off.
 1 files and directories, 0 blocks (0 replicated blocks, 0 erasure coded block groups) = 1 total filesystem object(s).
 Heap Memory used 108.58 MB of 339.5 MB Heap Memory. Max Heap Memory is 889 MB.
 Non Heap Memory used 51.2 MB of 53.48 MB Committed Non Heap Memory. Max Non Heap Memory is "unbounded".



The screenshot displays the Hadoop web UI interface. On the left, a sidebar menu includes 'Cluster' (with sub-items: About, Nodes, Node Labels, Associations, NEW, NEW SAVING, SUBMITTED, ACCEPTED, RUNNING, FINISHED, FAILED, KILLED) and 'Scheduler'. The main content area shows 'Cluster Metrics' with a table of application states (Submitted, Pending, Running, Completed, Containers Running, Used) all at 0. Below this, 'Cluster Nodes Metrics' shows 1 Active Node, 0 Decommissioning Nodes, and 0 Decommissioned Nodes. The 'Scheduler Metrics' section indicates the Scheduler Type is 'Capacity Scheduler' and the Scheduling Resource Type is '(memory-mb (unit-MB), vcores)'. A table of application entries is shown with columns: ID, User, Name, Application Type, Application Tags, Queue, Application Priority, StartTime, LaunchTime, FinishTime, State, and Progress. The table is currently empty, showing 'Showing 0 to 0 of 0 entries'.

RESULT:

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