

**Exp No: 8**

**Date: 03.09.2024**

## **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:

Installation Screenshots:

Step 1: Download and install Java

ORACLE

Products Industries Resources Customers Partners Developers Company

🔍 🇮🇳

👤 View Accounts

📄 Contact Sales

Java downloads Tools and resources Java archive

Java 8 Java 8 Enterprise Performance Pack Java 11

### Java SE Development Kit 8u421

Java SE subscribers will receive JDK 8 updates until at least **December 2030**.

Manual update required for some Java 8 users on macOS.

**The Oracle JDK 8 license changed in April 2019**

The [Oracle Technology Network License Agreement](#) for Oracle Java SE is substantially different from prior Oracle JDK 8 licenses. This license permits certain uses, such as personal use and development use, at no cost -- but other uses authorized under prior Oracle JDK licenses may no longer be available. Please review the terms carefully before downloading and using this product. FAQs are available [here](#).

Commercial license and support are available for a low cost with [Java SE Universal Subscription](#).

JDK 8 software is licensed under the [Oracle Technology Network License Agreement](#) for Oracle Java SE.

Java SE 8u421 checksums and OL 8 GPG Keys for RPMs

Linux macOS Solaris **Windows**

Product/file description	File size	Download
x86 Installer	141.01 MB	<a href="#">jdk-8u421-windows-i586.exe</a>
x64 Installer	150.83 MB	<a href="#">jdk-8u421-windows-x64.exe</a>

<https://www.oracle.com/in/java/technologies/downloads/#java8-windows>

Edit User Variable

✕

Variable name:

Variable value:

Browse Directory...

Browse File...

OK

Cancel

# Edit environment variable



C:\Program Files (x86)\Common Files\Oracle\Java\java8path  
C:\Program Files (x86)\Common Files\Oracle\Java\javapath  
C:\Program Files\Python311\Scripts\  
C:\Program Files\Python311\  
%SystemRoot%\system32  
%SystemRoot%  
%SystemRoot%\System32\Wbem  
%SYSTEMROOT%\System32\WindowsPowerShell\v1.0\  
%SYSTEMROOT%\System32\OpenSSH\  
C:\Users\Admin\AppData\Roaming\Python\Python311\Scripts  
C:\Program Files\nodejs\  
D:\Admin\Git\cmd  
C:\Java\jdk-1.8\bin  
C:\Hadoop\bin  
C:\Hadoop\sbin

New

Edit

Browse...

Delete

Move Up

Move Down

Edit text...

OK


Cancel


```


C:\Users\Admin>java -version
java version "1.8.0_421"
Java(TM) SE Runtime Environment (build 1.8.0_421-b09)
Java HotSpot(TM) 64-Bit Server VM (build 25.421-b09, mixed mode)


```


## Step 2: Download Hadoop


 Apache Hadoop

 Download

 Documentation

 Community

 Development

 Help

Apache Software Foundation

 Apache Hadoop

The Apache® Hadoop® project develops open-source software for reliable, scalable, distributed computing.

The Apache Hadoop software library is a framework that allows for the distributed processing of large data sets across clusters of computers using simple programming models. It is designed to scale up from single servers to thousands of machines, each offering local computation and storage. Rather than rely on hardware to deliver high-availability, the library itself is designed to detect and handle failures at the application layer, so delivering a highly-available service on top of a cluster of computers, each of which may be prone to failures.

Learn more »

Download »

Getting started »

### Latest news

Release 3.4.0 available2024 Mar 17

This is the first release of Apache Hadoop 3.4 line. It contains 2888 bug fixes, improvements and enhancements since 3.3.

Users are encouraged to read the [overview of main changes](#). For details of release check

### Modules

The project includes these modules:

- **Hadoop Common:** The common utilities that support the other Hadoop modules.
- **Hadoop Distributed File System (HDFS™):** A distributed file system that provides high-throughput access to application data.
- **Hadoop YARN:** A framework for job scheduling and cluster resource management.

### Related projects

Other Hadoop-related projects at Apache include:

- **Ambari™:** A web-based tool for provisioning, managing, and monitoring Apache Hadoop clusters which includes support for Hadoop HDFS, Hadoop MapReduce, Hive, HCatalog, HBase, ZooKeeper, Oozie, Pig and Sqoop. Ambari also provides a dashboard for viewing cluster health such as hostnames and ability to view MapReduce, Pig and

## Step 3: Set Environment Variables

Edit User Variable

Variable name:

HADOOP\_HOME

Variable value:

C:\Hadoop\bin

Browse Directory...

Browse File...

OK

Cancel

# Edit environment variable



C:\Program Files (x86)\Common Files\Oracle\Java\java8path  
C:\Program Files (x86)\Common Files\Oracle\Java\javapath  
C:\Program Files\Python311\Scripts\  
C:\Program Files\Python311\  
%SystemRoot%\system32  
%SystemRoot%  
%SystemRoot%\System32\Wbem  
%SYSTEMROOT%\System32\WindowsPowerShell\v1.0\  
%SYSTEMROOT%\System32\OpenSSH\  
C:\Users\Admin\AppData\Roaming\Python\Python311\Scripts  
C:\Program Files\nodejs\  
D:\Admin\Git\cmd  
C:\Java\jdk-1.8\bin  
C:\Hadoop\bin  
C:\Hadoop\sbin

New

Edit

Browse...

Delete

Move Up

Move Down

Edit text...

OK

Cancel

```

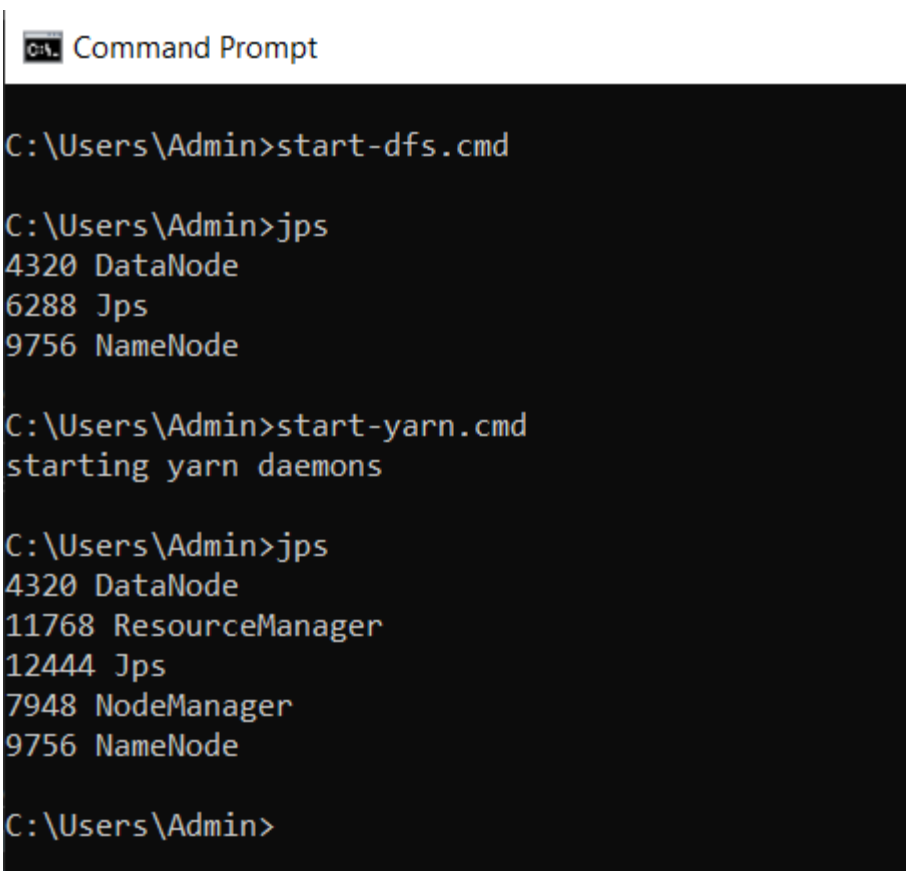
C:\Users\Admin>hadoop version
Hadoop 3.3.6
Source code repository https://github.com/apache/hadoop.git -r 1be78238728da9266a4f88195058f08fd012bf9c
Compiled by ubuntu on 2023-06-18T08:22Z
Compiled on platform linux-x86_64
Compiled with protoc 3.7.1
From source with checksum 5652179ad55f76cb287d9c633bb53bbd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

```

## Step 4: Format Hadoop NameNode

```
C:\Users\Admin>hdfs namenode -format
2024-08-07 09:22:38,053 INFO namenode.NameNode: STARTUP_MSG:
/*****
STARTUP_MSG: Starting NameNode
STARTUP_MSG: host = DESKTOP-TF65P79/192.168.56.1
STARTUP_MSG: args = [-format]
STARTUP_MSG: version = 3.3.6
STARTUP_MSG: classpath = C:\Hadoop\etc\hadoop;c:\Hadoop\share\hadoop\common;c:\Hadoop\share\hadoop\common\lib\animal-s
niffer-annotations-1.17.jar;c:\Hadoop\share\hadoop\common\lib\audience-annotations-0.5.0.jar;c:\Hadoop\share\hadoop\commo
n\lib\avro-1.7.7.jar;c:\Hadoop\share\hadoop\common\lib\checker-qual-2.5.2.jar;c:\Hadoop\share\hadoop\common\lib\commons
-beanutils-1.9.4.jar;c:\Hadoop\share\hadoop\common\lib\commons-cli-1.2.jar;c:\Hadoop\share\hadoop\common\lib\commons-cod
ec-1.15.jar;c:\Hadoop\share\hadoop\common\lib\commons-collections-3.2.2.jar;c:\Hadoop\share\hadoop\common\lib\commons-co
mpress-1.21.jar;c:\Hadoop\share\hadoop\common\lib\commons-configuration2-2.8.0.jar;c:\Hadoop\share\hadoop\common\lib\com
mons-daemon-1.0.13.jar;c:\Hadoop\share\hadoop\common\lib\commons-io-2.8.0.jar;c:\Hadoop\share\hadoop\common\lib\commons-
lang3-3.12.0.jar;c:\Hadoop\share\hadoop\common\lib\commons-logging-1.1.3.jar;c:\Hadoop\share\hadoop\common\lib\commons-m
ath3-3.1.1.jar;c:\Hadoop\share\hadoop\common\lib\commons-net-3.9.0.jar;c:\Hadoop\share\hadoop\common\lib\commons-text-1.
10.0.jar;c:\Hadoop\share\hadoop\common\lib\curator-client-5.2.0.jar;c:\Hadoop\share\hadoop\common\lib\curator-framework
5.2.0.jar;c:\Hadoop\share\hadoop\common\lib\curator-recipes-5.2.0.jar;c:\Hadoop\share\hadoop\common\lib\dnsjava-2.1.7.jar
```

## Step 5: Start Hadoop





Step 6: Verify Hadoop Installation

localhost:9870/dfshealth.html#tab-overview

HadoopOverviewDatanodesDatanode Volume FailuresSnapshotStartup ProgressUtilities

Overview

'localhost:9000' (active)

Started:	Wed Aug 07 10:33:32 +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-b4721dc1-9cf4-4238-8c79-a87f7ea9e95e
Block Pool ID:	BP-1278284451-192.168.56.1-1723006987378

Summary

Security is off.

Safemode is off.

1 files and directories, 0 blocks (0 replicated blocks, 0 erasure coded block groups) = 1 total filesystem object(s).

Heap Memory used 56.23 MB of 196.5 MB Heap Memory. Max Heap Memory is 889 MB.

Non Heap Memory used 52.04 MB of 53.27 MB Committed Non Heap Memory. Max Non Heap Memory is <unbounded>.

Configured Capacity:	118.63 GB
----------------------	-----------

Configured Capacity:	118.63 GB
Configured Remote Capacity:	0 B
DFS Used:	148 B (0%)
Non DFS Used:	97.34 GB
DFS Remaining:	21.29 GB (17.95%)
Block Pool Used:	148 B (0%)
DataNodes usages% (Min/Median/Max/stdDev):	0.00% / 0.00% / 0.00% / 0.00%
Live Nodes	1 (Decommissioned: 0, In Maintenance: 0)
Dead Nodes	0 (Decommissioned: 0, In Maintenance: 0)
Decommissioning Nodes	0
Entering Maintenance Nodes	0
Total Datanode Volume Failures	0 (0 B)
Number of Under-Replicated Blocks	0
Number of Blocks Pending Deletion (including replicas)	0
Block Deletion Start Time	Wed Aug 07 10:33:32 +0530 2024
Last Checkpoint Time	Wed Aug 07 10:33:07 +0530 2024
Enabled Erasure Coding Policies	RS-6-3-1024k

# NameNode Journal Status

Current transaction ID: 1	
Journal Manager	State
FileJournalManager(root=C:\Hadoop\data\namenode)	EditLogFileOutputStream(C:\Hadoop\data\namenode\current\edits_inprogress_0000000000000000001)

# NameNode Storage

Storage Directory	Type	State
C:\Hadoop\data\namenode	IMAGE_AND_EDITS	Active

# DFS Storage Types

Storage Type	Configured Capacity	Capacity Used	Capacity Remaining	Block Pool Used	Nodes In Service
DISK	118.63 GB	148 B (0%)	21.29 GB (17.95%)	148 B	1

Hadoop, 2023.

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