**Practical 6**

Hadoop Installation

**Aim:**

Install, configure and run Hadoop and HDFS and explore HDFS on Windows

**Code:**

**Steps to Install Hadoop**

1. Install Java JDK 1.8
2. Download Hadoop and extract and place under C drive
3. Set Path in Environment Variables
4. Config files under Hadoop directory
5. Create folder datanode and namenode under data directory
6. Edit HDFS and YARN files
7. Set Java Home environment in Hadoop environment
8. Setup Complete. Test by executing start-all.cmd

**There are two ways to install Hadoop, i.e.**

1. Single node
2. Multi node

Here, we use multi node cluster.

1. **Install Java**
2. – Java JDK Link to download

<https://www.oracle.com/java/technologies/javase-jdk8-downloads.html>

1. – extract and install Java in C:\Java
2. – open cmd and type -> javac -version

Graphical user interface, text

Description automatically generated

1. **Download Hadoop**

<https://www.apache.org/dyn/closer.cgi/hadoop/common/hadoop-3.3.0/hadoop-3.3.0.tar.gz>

* right click .rar.gz file -> show more options -> 7-zip->and extract to C:\Hadoop-3.3.0\

Graphical user interface, text, application

Description automatically generated

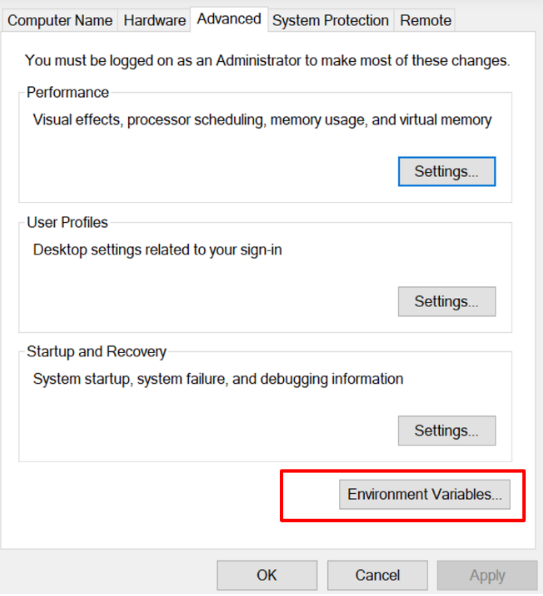
1. **Set the path JAVA\_HOME Environment variable**
2. **Set the path HADOOP\_HOME Environment variable**

Graphical user interface, application

Description automatically generated

Graphical user interface, text, application, chat or text message

Description automatically generated



Click on **New to both user variables and system variables.**

Graphical user interface, text, application, email

Description automatically generated

Graphical user interface, text, application, email

Description automatically generated

**Click on user variable -> path -> edit-> add path for Hadoop and java upto ‘bin’** Graphical user interface, text, application

Description automatically generated

Click Ok, Ok, Ok.

1. **Configurations**

**Edit file C:/Hadoop-3.3.0/etc/hadoop/core-site.xml,**

paste the xml code in folder and save

**======================================================**

<configuration>

<property>

<name>fs.defaultFS</name>

<value>hdfs://localhost:9000</value>

</property>

</configuration>

**======================================================**

**Rename “mapred-site.xml.template” to “mapred-site.xml” and edit this file C:/Hadoop-3.3.0/etc/hadoop/mapred-site.xml, paste xml code and save this file.**

**======================================================**

<configuration>

<property>

<name>mapreduce.framework.name</name>

<value>yarn</value>

</property>

</configuration>

**======================================================**

**Create folder “data” under “C:\Hadoop-3.3.0”**

**Create folder “datanode” under “C:\Hadoop-3.3.0\data”**

**Create folder “namenode” under “C:\Hadoop-3.3.0\data”**

**======================================================**

**Edit file C:\Hadoop-3.3.0/etc/hadoop/hdfs-site.xml,**

**paste xml code and save this file.**

<configuration>

<property>

<name>dfs.replication</name>

<value>1</value>

</property>

<property>

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

<value>/hadoop-3.3.0/data/namenode</value>

</property>

<property>

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

<value>/hadoop-3.3.0/data/datanode</value>

</property>

</configuration>

**======================================================**

**Edit file C:/Hadoop-3.3.0/etc/hadoop/yarn-site.xml,**

**paste xml code and save this file.**

<configuration>

<property>

<name>yarn.nodemanager.aux-services</name>

<value>mapreduce\_shuffle</value>

</property>

<property>

<name>yarn.nodemanager.auxservices.mapreduce.shuffle.class</name>

<value>org.apache.hadoop.mapred.ShuffleHandler</value>

</property>

<property>

<name>yarn.resourcemanager.address</name>

<value>127.0.0.1:8032</value>

</property>

<property>

<name>yarn.resourcemanager.scheduler.address</name>

<value>127.0.0.1:8030</value>

</property>

<property>

<name>yarn.resourcemanager.resource-tracker.address</name>

<value>127.0.0.1:8031</value>

</property>

</configuration>

**======================================================**

1. **Edit file C:/Hadoop-3.3.0/etc/hadoop/hadoop-env.cmd**

Find “JAVA\_HOME=%JAVA\_HOME%” and replace it as

set JAVA\_HOME="C:\Java\jdk1.8.0\_361"

**======================================================**

1. **Download “redistributable” package**

**Download and run VC\_redist.x64.exe**

This is a “redistributable” package of the Visual C runtime code for 64-bit applications, from Microsoft. It contains certain shared code that every application written with Visual C expects to have available on the Windows computer it runs on.

1. **Hadoop Configurations**

**Download bin folder from**

[**https://github.com/s911415/apache-hadoop-3.1.0-winutils**](https://github.com/s911415/apache-hadoop-3.1.0-winutils)

**– Copy the bin folder to c:\hadoop-3.3.0. Replace the existing bin folder.**

1. **copy "hadoop-yarn-server-timelineservice-3.0.3.jar" from ~\hadoop-3.0.3\share\hadoop\yarn\timelineservice to ~\hadoop-3.0.3\share\hadoop\yarn folder.**
2. **Format the NameNode**

**– Open cmd ‘Run as Administrator’ and type command “hdfs namenode –format”**

Text

Description automatically generated

Text

Description automatically generated

**11. Testing**

**– Open cmd ‘Run as Administrator’ and change directory to C:\Hadoop-3.3.0\sbin**

**– type start-all.cmd**

**OR**

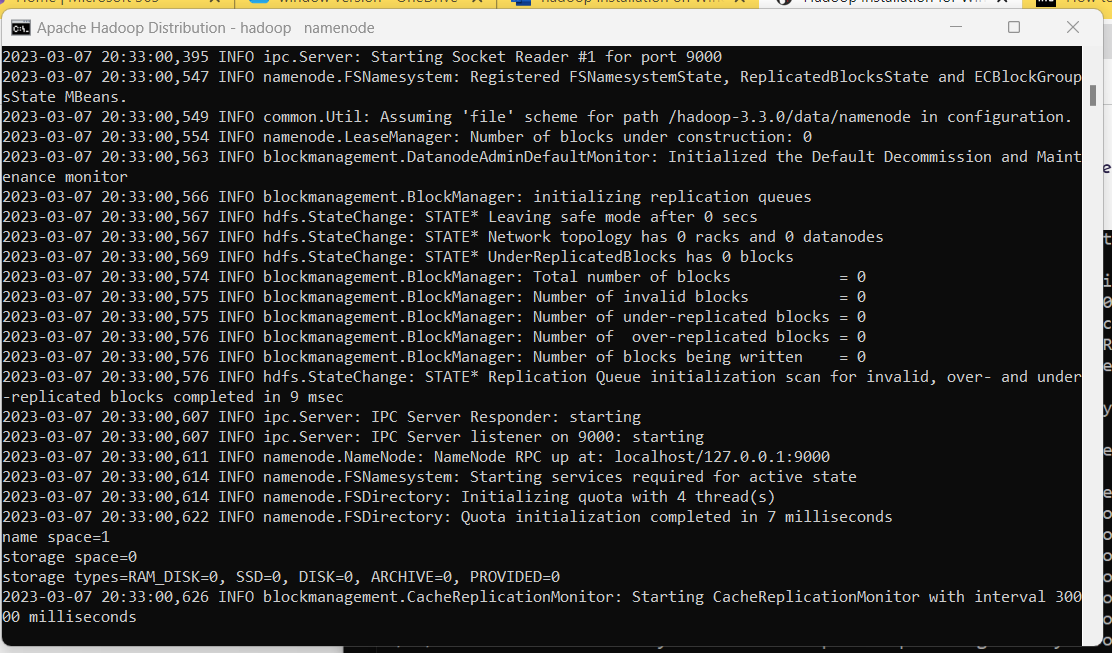
**- type start-dfs.cmd**

**– type start-yarn.cmd**

Text

Description automatically generated

**– You will get 4 more running threads for Datanode, namenode, resouce manager and node manager**



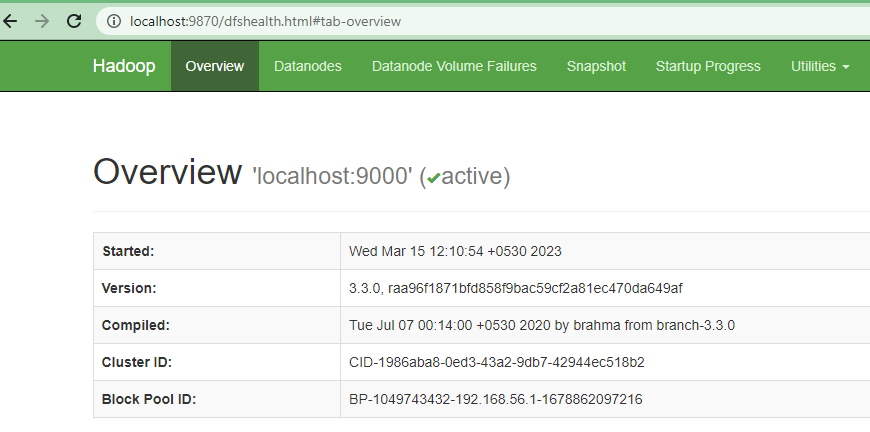
**Output:**

12. Type JPS command to start-all.cmd command prompt, you will get following output.

Text

Description automatically generated

13. Run <http://localhost:9870/> from any browser



Graphical user interface, text, application, table, email, Excel

Description automatically generated