**Experiment-2**

**Aim:** Installation of Hadoop package on Windows Operating System.

**Objective:** To install Hadoop on the system.

**Software Requirements:**

1. JDK 8

**Theory:**

**Hadoop Installation on Windows 10**

As a beginner, you might feel reluctant in performing cloud computing which requires subscriptions. While you can install a virtual machine as well in your system, it requires allocation of a large amount of RAM for it to function smoothly else it would hang constantly.

You can install Hadoop in your system as well which would be a feasible way to learn Hadoop.

We will be installing single node pseudo-distributed hadoop cluster on windows 10.

**Prerequisite**:

To install Hadoop, you should have Java version 1.8 in your system.

Check your java version through this command on command prompt

|  |  |
| --- | --- |
| 1. java-version |  |

If java is not installed in your system, then –

Go this link –

[https://www.oracle.com/technetwork/java/javase/downloads/jdk8-downl...](https://www.oracle.com/technetwork/java/javase/downloads/jdk8-downloads-2133151.html)

Accept the license,

Download the file according to your operating system. Keep the java folder directly under the local disk directory (C:\Java\jdk1.8.0\_152) rather than in Program Files (C:\Program Files\Java\jdk1.8.0\_152) as it can create errors afterwards.

After downloading java version 1.8, download hadoop version 3.1 from this link –

[https://archive.apache.org/dist/hadoop/common/hadoop-3.1.0/hadoop-3...](https://archive.apache.org/dist/hadoop/common/hadoop-3.1.0/hadoop-3.1.0.tar.gz)

Extract it to a folder.

1. Create a new user variable. Put the Variable\_name as HADOOP\_HOME and Variable\_value as the path of the bin folder where you extracted hadoop.

Likewise, create a new user variable with variable name as JAVA\_HOME and variable value as the path of the bin folder in the Java directory.

Now we need to set Hadoop bin directory and Java bin directory path in system variable path.

Edit Path in system variable

Click on New and add the bin directory path of Hadoop and Java in it.

### ****Configurations****

Now we need to edit some files located in the hadoop directory of the etc folder where we installed hadoop

1. Edit the file core-site.xml in the hadoop directory. Copy this xml property in the configuration in the file

<configuration>

<property>

<name>fs.defaultFS</name>

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

</property>

</configuration>

1. Edit mapred-site.xml and copy this property in the cofiguration

<configuration>

<property>

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

<value>yarn</value>

</property>

</configuration>

1. Create a folder ‘data’ in the hadoop directory

Create a folder with the name ‘datanode’ and a folder ‘namenode’ in this data directory

1. Edit the file hdfs-site.xml and add below property in the configuration

Note: The path of namenode and datanode across value would be the path of the datanode and namenode folders you just created.

<configuration>

<property>

<name>dfs.replication</name>

<value>1</value>

</property>

<property>

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

<value>/hadoop-3.2.2/Nodes/NameNode</value>

</property>

<property>

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

<value>/hadoop-3.2.2/Nodes/DataNode</value>

</property>

</configuration>

1. Edit the file yarn-site.xml and add below property in the configuration

<property>

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

<value>mapreduce\_shuffle</value>

</property>

<property>

<name>yarn.nodemanager.aux-services.mapreduce.shuffle.class</name>

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

</property>

1. Edit hadoop-env.cmd and replace %JAVA\_HOME% with the path of the java folder where your jdk 1.8 is installed











