Date:-

Name:- Jadhav Prathamesh Arvind
Roll No:- 3014

Experiment No:8

- **Title:-** Install Hadoop Single Node Cluster And Run Simple Application.

- **Aim:-** To Install Hadoop Single Node Cluster And Run Simple Application.

- **Objective:-**

  1. To learn and understand about Hadoop and HDFS.
  2. Downloading and installing the Latest Version of Hadoop.
  3. Downloading and installing the Latest Version of Apache.
  4. Downloading and installing the Latest Version of JDK and JRE.
  5. Configuring environment variables.
  6. Configuring cloud cluster.
  7. Formatting the node names.
  8. To run a sample application in Hadoop for basic word count.

- **Theory:-**

1. **Learn and understand about Hadoop and HDFC –**

   Apache Hadoop is an **open source**, **Scalable**, and **Fault tolerant** framework written in **Java**. It efficiently processes large volumes of data on a cluster of commodity hardware. Hadoop is not only a storage system but it is a platform for large data storage as well as processing. **Hadoop** is an open-source tool from the **ASF** Apache Software Foundation. Open source project means it is freely available and we can even change its source code as per the requirements. If certain functionality does not fulfill your need then you can change it according to your need. Most of Hadoop code is written by *Yahoo, IBM, Facebook, and Cloudera.* It provides an efficient framework for running jobs on multiple nodes of clusters. **Cluster** means a group of systems connected via LAN. Apache Hadoop provides parallel processing of data as it works on multiple machines simultaneously.

   Hadoop HDFS or Hadoop Distributed File System is a distributed file system which provides storage in Hadoop in a distributed fashion. In Hadoop Architecture on the master node, a daemon called *name node* runs for HDFS. On all the slaves a daemon called *data node* runs. Hence slaves are also called as data node. Name node stores meta-data and manages the data nodes. On the other hand, Data nodes stores the data and do the actual task. HDFS is a highly **fault tolerant, distributed, reliable** and **scalable** file system for data storage. First follow this guide to **learn more about features of HDFS** and then proceed further with the Hadoop tutorial.

Date:-

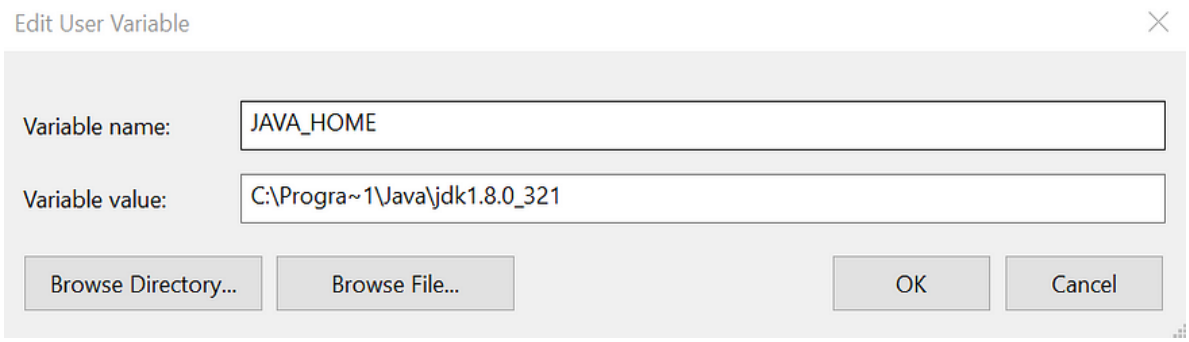
Name:- Jadhav Prathamesh Arvind
Roll No:- 3014

2. JDK and JRE in apache Hadoop –

- **JDK –** This includes the Java **compiler**, which turns their source code into bytecode that can be run on any machine with a JVM.
- **JRE –** The JRE, on the other hand, is like a delivery person who brings the completed table to your doorstep. When you download and install a Java application, the JRE comes along with it. It contains the JVM, which is responsible for executing Java bytecode.

Step 1: Downloading Hadoop binaries.

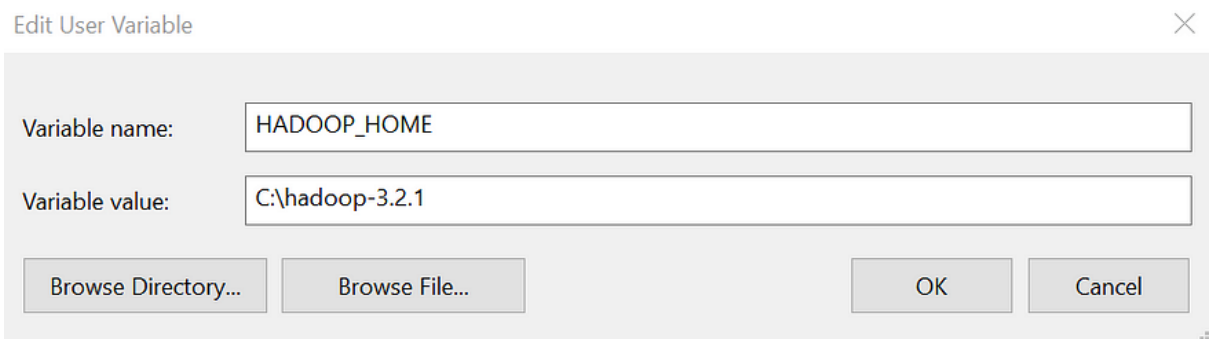
Step 2: Configuring environment variables.



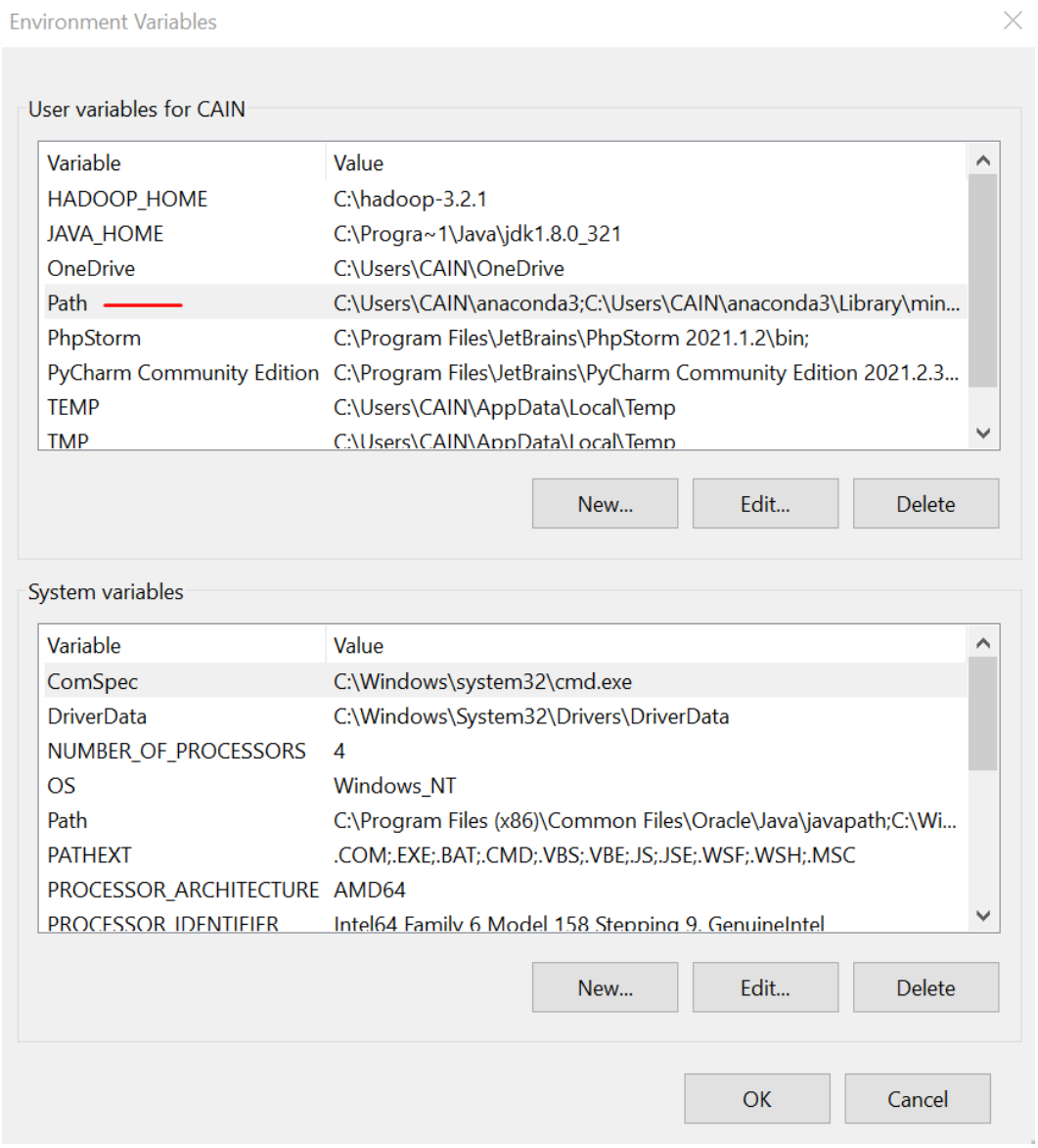
- HADOOP_HOME: This should aim to the Hadoop folder we created on C:

Date:-

Name:- Jadhav Prathamesh Arvind
Roll No:- 3014



- Now, on the environment variable window, we need to select the path variable and click the “Edit” button.



Step 3:- Configuring Hadoop cluster.

Date:-

Name:- Jadhav Prathamesh Arvind
Roll No:- 3014

- We must configure 4 internal files:-

1.- C:\Program_files\hadoop-3.2.1\etc\hadoop\hdfs-site.xml

```
<?xml version="1.0" encoding="UTF-8"?>
<?xml-stylesheet type="text/xsl" href="configuration.xsl"?>
<!--
  Licensed under the Apache License, Version 2.0 (the "License");
  you may not use this file except in compliance with the License.
  You may obtain a copy of the License at

    http://www.apache.org/licenses/LICENSE-2.0

  Unless required by applicable law or agreed to in writing, software
  distributed under the License is distributed on an "AS IS" BASIS,
  WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied.
  See the License for the specific language governing permissions and
  limitations under the License. See accompanying LICENSE file.
-->

<!-- Put site-specific property overrides in this file. -->

<configuration>
    <property>
        <name>dfs.replication</name>
        <value>1</value>
    </property>
    <property>
        <name>dfs.namenode.name.dir</name>
        <value>file:///C:/hadoop-3.2.1/data/dfs/namenode</value>
    </property>
    <property>
        <name>dfs.datanode.data.dir</name>
        <value>file:///C:/hadoop-3.2.1/data/dfs/datanode</value>
    </property>
</configuration>
```

2.- C:\Program_files\hadoop-3.2.1\etc\hadoop\core-site.xml

Date:-

Name:- Jadhav Prathamesh Arvind
Roll No:- 3014

```
<?xml version="1.0" encoding="UTF-8"?>
<?xml-stylesheet type="text/xsl" href="configuration.xsl"?>
<!--
  Licensed under the Apache License, Version 2.0 (the "License");
  you may not use this file except in compliance with the License.
  You may obtain a copy of the License at

    http://www.apache.org/licenses/LICENSE-2.0

  Unless required by applicable law or agreed to in writing, software
  distributed under the License is distributed on an "AS IS" BASIS,
  WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied.
  See the License for the specific language governing permissions and
  limitations under the License. See accompanying LICENSE file.
-->

<!-- Put site-specific property overrides in this file. -->

<configuration>
    <property>
        <name>fs.default.name</name>
        <value>hdfs://localhost:9820</value>
    </property>
</configuration>
```

3.- C:\Program_files\hadoop-3.2.1\etc\hadoop\mapred-site.xml

```
<?xml version="1.0"?>
<?xml-stylesheet type="text/xsl" href="configuration.xsl"?>
<!--
  Licensed under the Apache License, Version 2.0 (the "License");
  you may not use this file except in compliance with the License.
  You may obtain a copy of the License at

    http://www.apache.org/licenses/LICENSE-2.0

  Unless required by applicable law or agreed to in writing, software
  distributed under the License is distributed on an "AS IS" BASIS,
  WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied.
  See the License for the specific language governing permissions and
  limitations under the License. See accompanying LICENSE file.
-->

<!-- Put site-specific property overrides in this file. -->

<configuration>
    <property>
        <name>mapreduce.framework.name</name>
        <value>yarn</value>
        <description>MapReduce framework name</description>
    </property>
</configuration>
```

4.- C:\Program_files\hadoop-3.2.1\etc\hadoop\yarn-site.xml

```
<?xml version="1.0"?>
<!--
  Licensed under the Apache License, Version 2.0 (the "License");
  you may not use this file except in compliance with the License.
  You may obtain a copy of the License at

    http://www.apache.org/licenses/LICENSE-2.0

  Unless required by applicable law or agreed to in writing, software
  distributed under the License is distributed on an "AS IS" BASIS,
  WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied.
  See the License for the specific language governing permissions and
  limitations under the License. See accompanying LICENSE file.
-->
<configuration>

<!-- Site specific YARN configuration properties -->
    <property>
        <name>yarn.nodemanager.aux-services</name>
        <value>mapreduce_shuffle</value>
        <description>Yarn Node Manager Aux Service</description>
    </property>

</configuration>
```

Step 4:- Formatting the Name node

```
hdfs namenode -format
```

Step 5:- Starting Hadoop

```
.\start-dfs.cmd
```

```
./start-yarn.cmd
```

```
jps
```

- **Basic word count example using Hadoop:-**

Date:-

Name:- Jadhav Prathamesh Arvind
Roll No:- 3014

- **Outcome:-**

  Successfully learn and Install Hadoop Single Node Cluster And Run Simple Application.

- **Result:-**

  Thus, we completed the experiment on Install Hadoop Single Node Cluster And Run Simple Application.