**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**](http://data-flair.training/blogs/comprehensive-hdfs-guide-introduction-architecture-data-read-write-tutorial/) 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**](http://data-flair.training/blogs/learn-hadoop-hdfs-fault-tolerance/)**, distributed, reliable**and**scalable** file system for data storage. First follow this guide to[**learn more about features of HDFS**](http://data-flair.training/blogs/features-hadoop-hdfs-overview-beginners/)and then proceed further with the Hadoop tutorial.

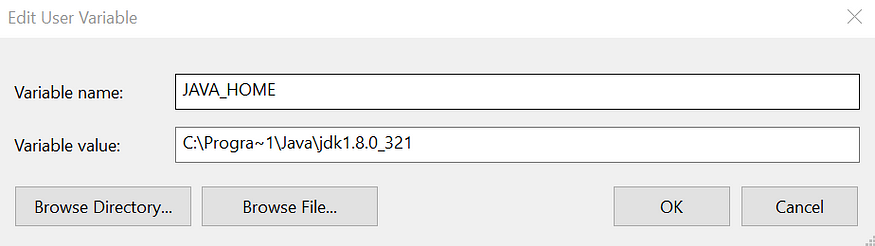
1. **JDK and JRE in apache Hadoop –**

* **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.

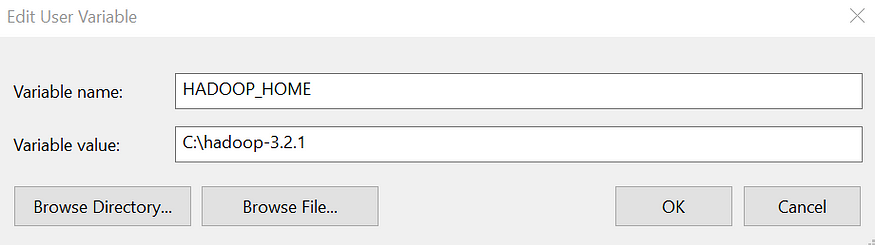
**Step 1:** Downloading Hadoop binaries.



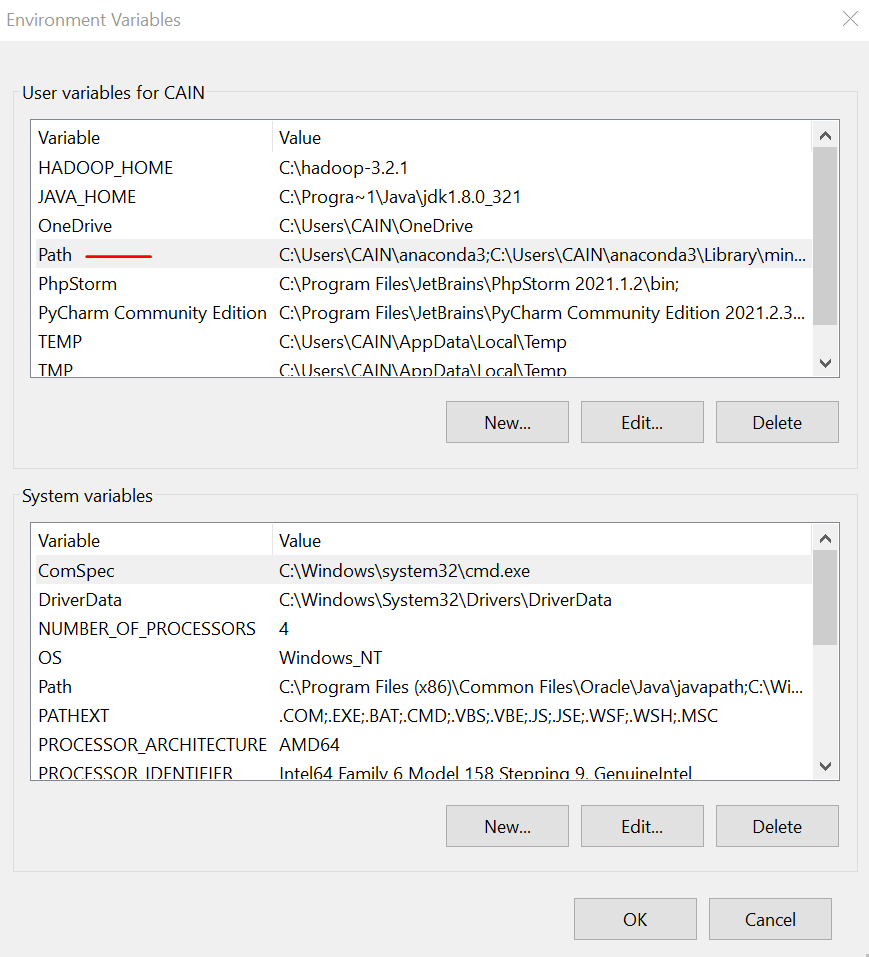
**Step 2:** Configuring environment variables.



* HADOOP\_HOME: This should aim to the Hadoop folder we created on C:



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



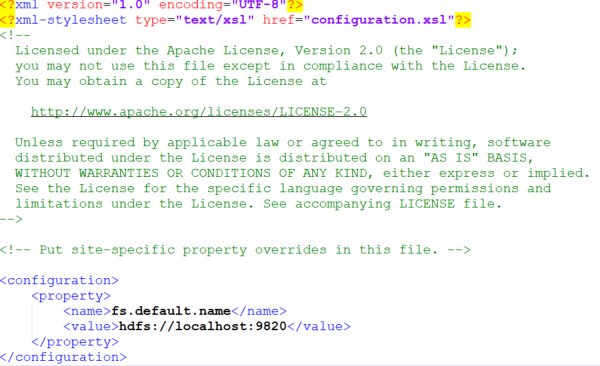
**Step 3:-** Configuring Hadoop cluster.

* We must configure 4 internal files:-

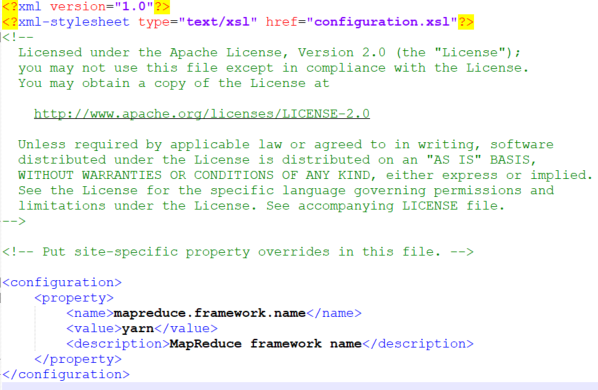
1.- C:\Program\_files\hadoop-3.2.1\etc\hadoop\hdfs-site.xml



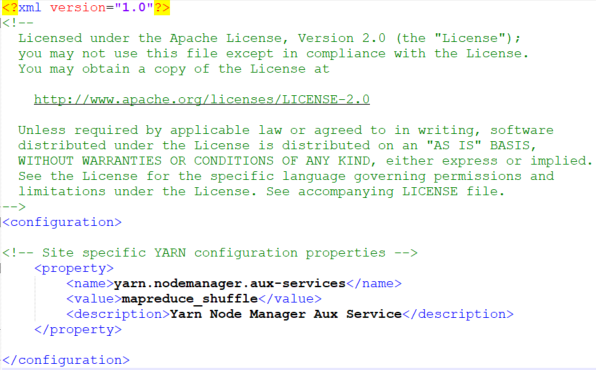
2.- C:\Program\_files\hadoop-3.2.1\etc\hadoop\core-site.xml



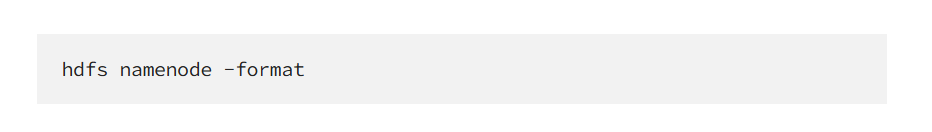
3.- C:\Program\_files\hadoop-3.2.1\etc\hadoop\mapred-site.xml



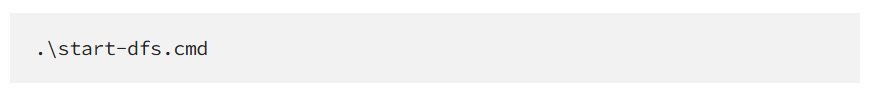
4.- C:\Program\_files\hadoop-3.2.1\etc\hadoop\yarn-site.xml

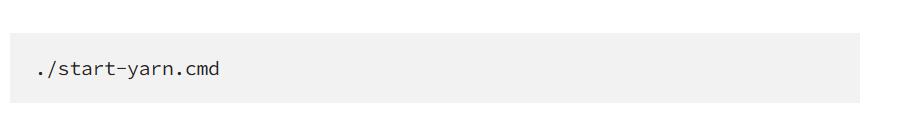


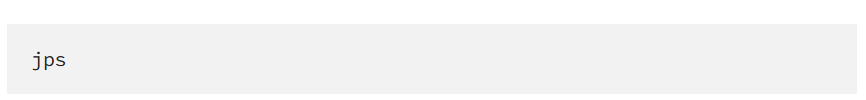
**Step 4:-** Formatting the Name node



**Step 5:-** Starting Hadoop

****

****



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

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

* **Result:-**

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