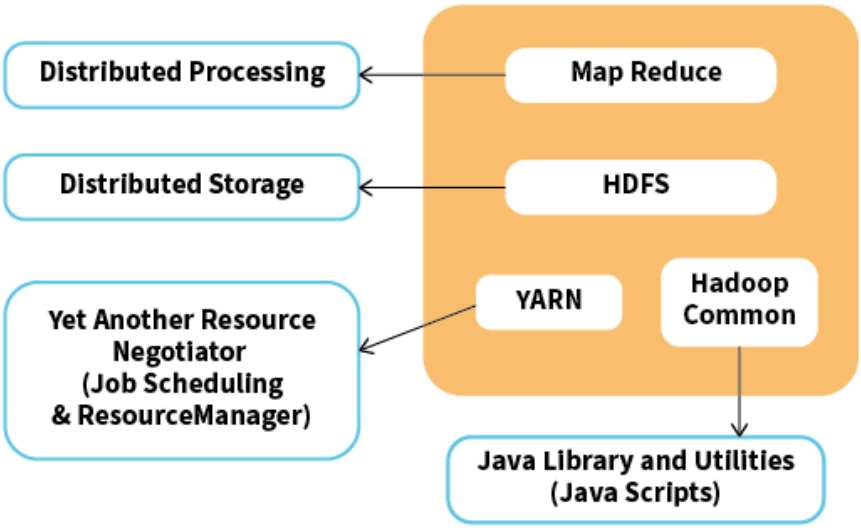
Academic Year: 2022-2023

Name:	Prerna Sunil Jadhav
Sap Id:	60004220127
Class:	T. Y. B. Tech (Computer Engineering)
Course:	Big Data Infrastructure Laboratory
Course Code:	DJ19CEEL6011
Experiment No.:	02

AIM: Install Hadoop on a Single Node Cluster.

WHAT IS HADOOP & WHY IS IT IMPORTANT?

- Hadoop is an open-source software programming framework for storing a large amount of data and performing the computation. Its framework is based on Java programming with some native code in C and shell scripts.
- Hadoop is an open-source software framework that is used for storing and processing large amounts of data in a distributed computing environment. It is designed to handle big data and is based on the MapReduce programming model, which allows for the parallel processing of large datasets.
- Hadoop has two main components:
  - HDFS (Hadoop Distributed File System): This is the storage component of Hadoop, which allows for the storage of large amounts of data across multiple machines. It is designed to work with commodity hardware, which makes it cost-effective.
  - YARN (Yet Another Resource Negotiator): This is the resource management component of Hadoop, which manages the allocation of resources (such as CPU and memory) for processing the data stored in HDFS.
  - Hadoop also includes several additional modules that provide additional functionality, such as Hive (a SQL-like query language), Pig (a high-level platform for creating MapReduce programs), and HBase (a non-relational, distributed database).
  - Hadoop is commonly used in big data scenarios such as data warehousing, business intelligence, and machine learning. It's also used for data processing, data analysis, and data mining. It enables the distributed processing of large data sets across clusters of computers using a simple programming model.
- Hadoop is important as one of the primary tools to store and process huge amounts of data quickly. It does this by using a distributed computing model which enables the fast processing of data that can be rapidly scaled by adding computing nodes.
- Hadoop Architecture
  - Hadoop stands as a robust platform for storing and processing vast amounts of data. It serves as a key solution for storing and analysing data from diverse sources, including databases, web servers, and file systems.
  - Built on the MapReduce programming algorithm, Hadoop architecture comprises four key components, each playing a crucial role in managing and processing extensive datasets.
    - HDFS (Hadoop Distributed File System)
    - MapReduce
    - YARN (Yet Another Resource Negotiator)
    - Common Utilities or Hadoop Common



INSTALLATION:

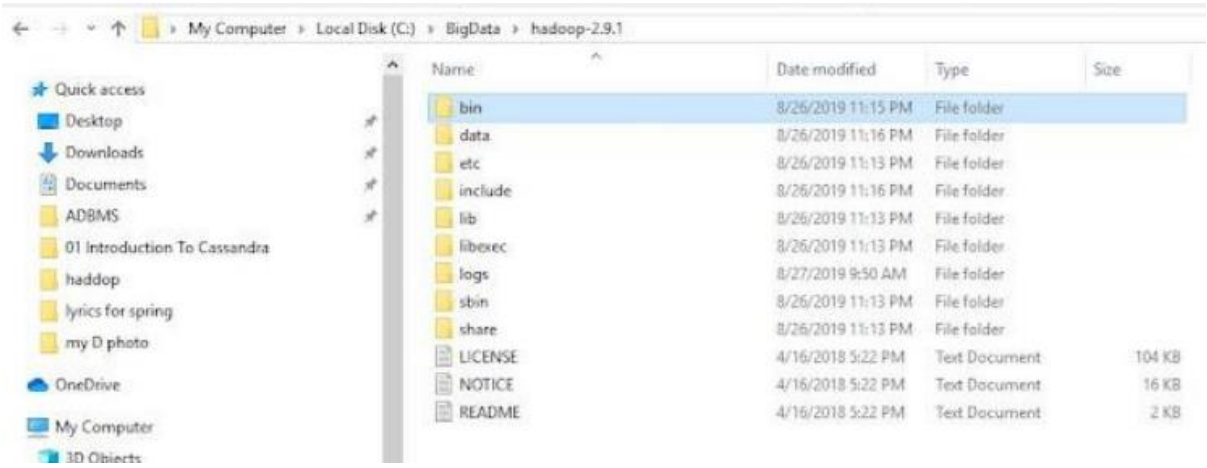
Install Hadoop 2.9.1 on Windows 10

First download the Hadoop 2.9.1 from the below link.

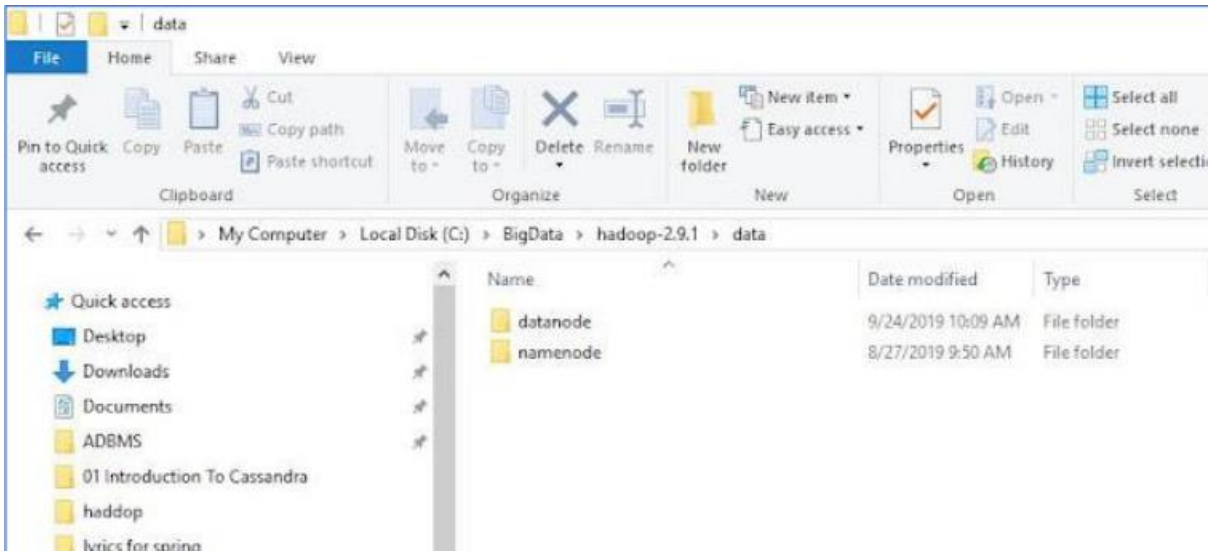
https://www.apache.org/dyn/closer.cgi/hadoop/common/hadoop-2.9.1/hadoop-2.9.1.tar.gz

Create a folder path as below and copy the downloaded msi into this folder.

Path:- 'C:/BigData/hadoop-2.9.1'



Go to C:/BigData/3adoop-2.9.1 and create a folder 'data'. Inside the 'data' folder create two folders 'datanode' and 'namenode'.



Then Set Hadoop Environment Variables

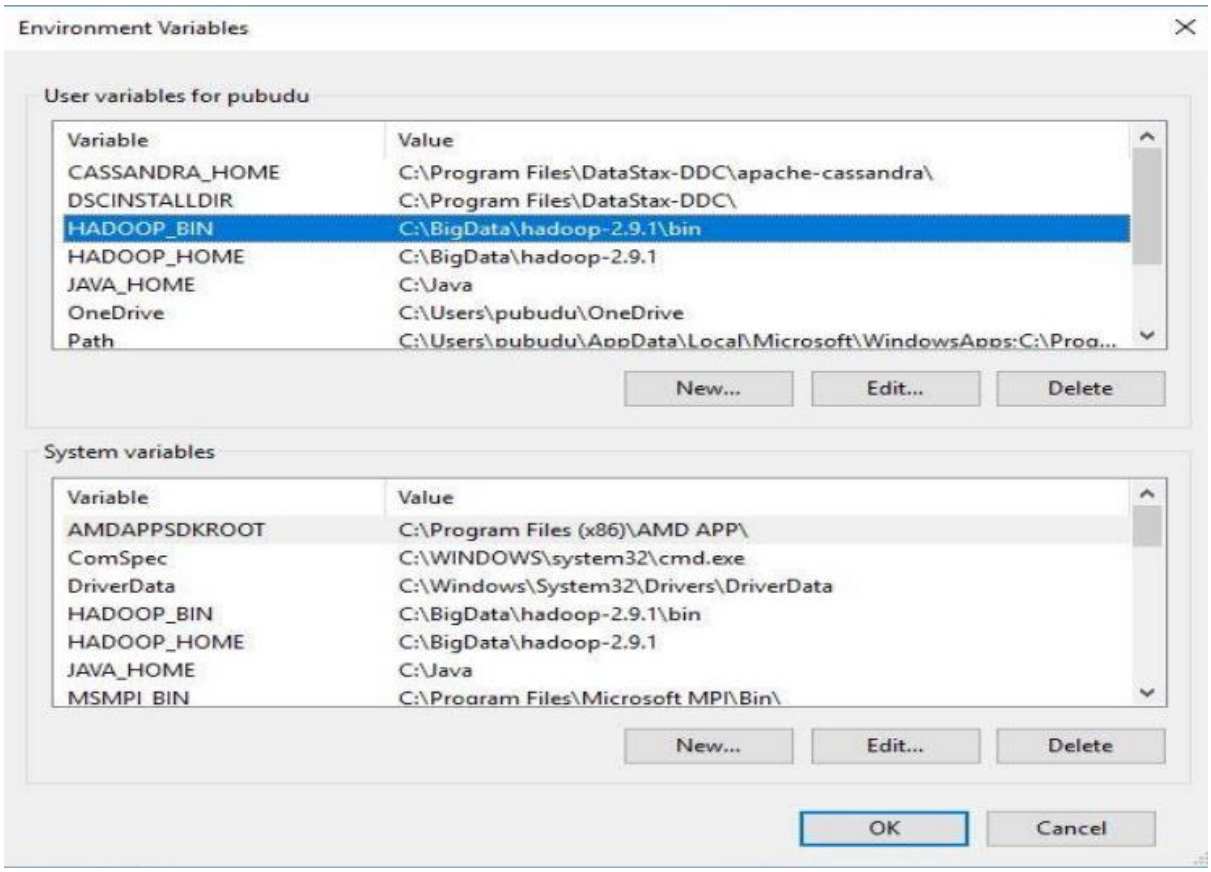
HADOOP_HOME="C:\BigData\hadoop-2.9.1"

HADOOP_BIN="C:\BigData\hadoop-2.9.1\bin"

JAVA_HOME=<JDK installation location>"

To set these variables, go to My Computer or This PC. Right click --> Properties --> Advanced

System settings --> Environment variables. Click New to create a new environment variables

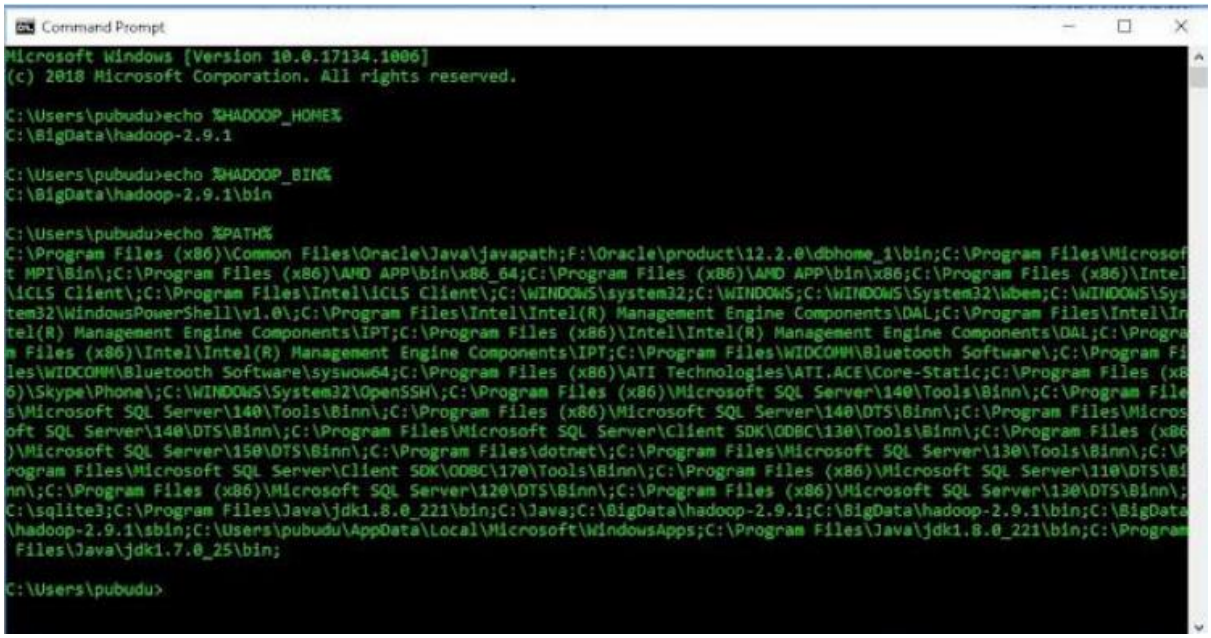


To validate the above setting, open new cmd and check the output.

echo %HADOOP_HOME%

echo %HADOOP_BIN%

echo %PATH%



To configure the Hadoop on windows we have to edit below mention files in the extracted location.

1. hadoop-env.cmd

2. core-site.xml

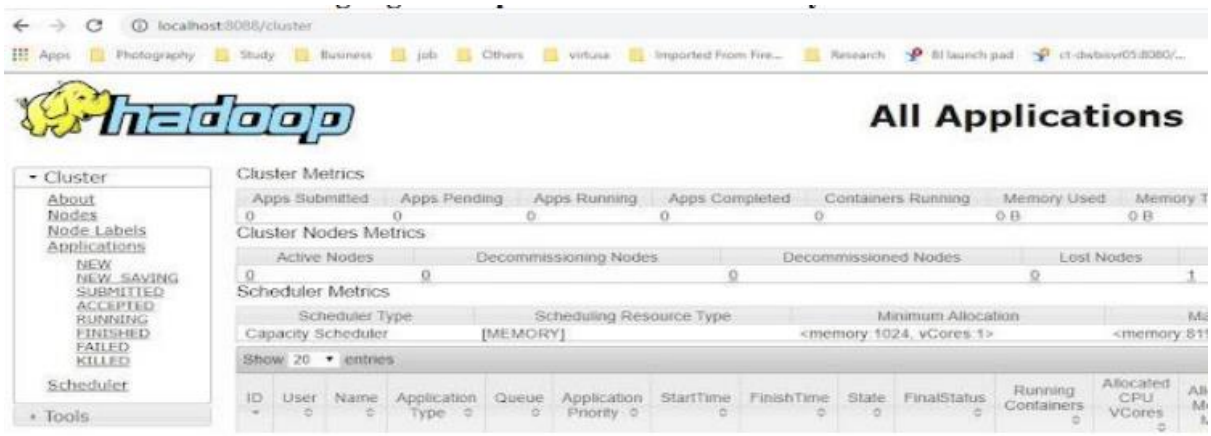
3. hdfs-site.xml

4. mapred-site.xml

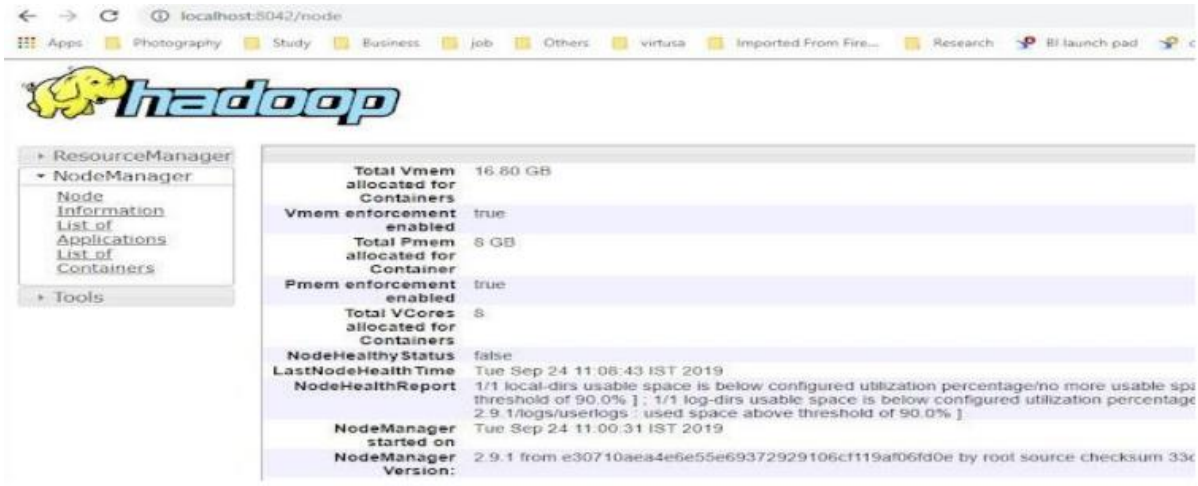
5. yarn-site.xml

Now you can access all the Hadoop components via web urls.

To access Resource Manager go to http://localhost:8088 from your web browser.

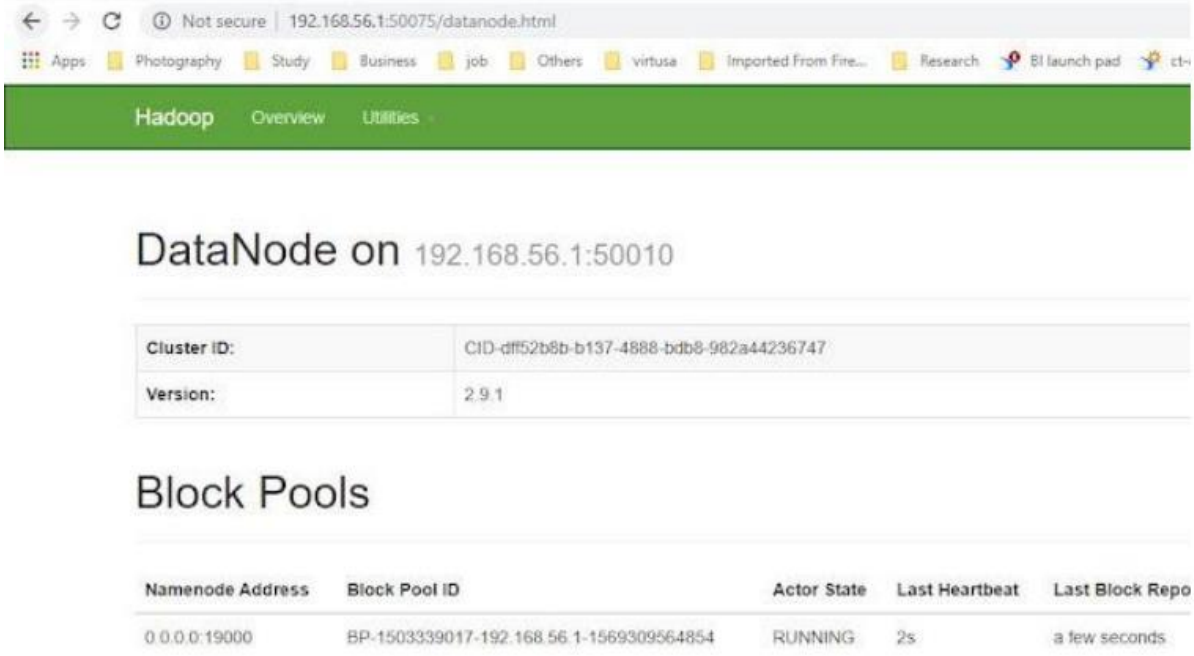


To access Node Manager go to http://localhost:8042 from your web browser.



To access Name Node go to http://localhost:50070 from your web browser

To access Data Node go to http://localhost:50075 from your web browser.



CONCLUSION: Hence, we successfully installed Hadoop