

Ex No: 1

Installation of Hadoop Framework,

AIM:

Installation of Hadoop Framework, its components and study the HADOOP ecosystem

Hadoop is an open-source framework that allows to store and process big data in a distributed environment across clusters of computers using simple programming models. It is designed to scale up from single servers to thousands of machines, each offering local computation and storage.

Hadoop Architecture:

The Apache Hadoop framework includes following four modules:

Hadoop Common:

Contains Java libraries and utilities needed by other Hadoop modules. These libraries give file system and OS level abstraction and comprise of the essential Java files and scripts that are required to start Hadoop.

Hadoop Distributed File System (HDFS): A distributed file-system that provides highthroughput access to application data on the community machines thus providing very high aggregate bandwidth across the cluster.

Hadoop YARN: A resource-management framework responsible for job scheduling and cluster resource management.

Hadoop MapReduce: This is a YARN-based programming model for parallel processing of large data sets.

Hadoop Installation procedure:

Step 1: Download and install Java

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

Step 2: Download Hadoop

<https://hadoop.apache.org/releases.html>

Step 3: Set Environment Variables

Step 4: Setup Hadoop

You must configure Hadoop in this phase by modifying several configuration files. Navigate to the “etc/hadoop” folder in the Hadoop folder. You must make changes to three files:

core-site.xml

```
<configuration>
  <property>
    <name>fs.default.name</name>
    <value>hdfs://localhost:9000</value>
  </property>
</configuration>
```

hdfs-site.xml

```
<configuration>
  <property>
    <name>dfs.replication</name>
    <value>1</value>
  </property>
  <property>
    <name>dfs.namenode.name.dir</name>
    <value>file:/hadoop-3.3.1/data/namenode</value>
  </property>
  <property>
    <name>dfs.datanode.data.dir</name>
    <value>file:/hadoop-3.3.1/data/datanode</value>
  </property>
</configuration>
```

mapred-site.xml

```
<configuration>
  <property>
    <name>mapred.job.tracker</name>
    <value>localhost:54311</value>
  </property>
</configuration>
```

Step 5: Format Hadoop NameNode

```
hadoop namenode -format
```

Step 6: Start Hadoop

```
start-all.cmd
```

Step 7: Verify Hadoop Installation

```
http://localhost:50070/.
```