

Experiment-3

1) Aim/Objective:

To install and configure Hadoop in Windows OS.

2) Theory-

Following steps are to followed for installing Hadoop in Windows:

(I) Installing Java:

- Download and install Java (JDK 8 or higher)
- Set Environment variables (JAVA_HOME: JDK/bin path)
- Run cmd and use the command “java -version” to check whether Java is installed properly or not.

(II) Downloading Hadoop:

- Download Hadoop binaries for windows platform from official website.
- The files are compressed in “.tar” format and needs to be extracted first.
- Ignore some error/warning messages that are displayed at the end of extraction process.

(III) Hadoop configuration:

- Hadoop can be configured in three different modes: Local (Standalone), pseudo-distributed, and fully-distributed mode.
- We will use the pseudo distributed mode and thus need to change settings in four configuration files located in the folder “etc/hadoop” (these changes are explained in “3) Program Code” section)

(IV) Running Hadoop:

- Set Environment variables (HADOOP_HOME: Hadoop/bin path)
- Run cmd and use the command “hdfs namenode -format” to check whether hadoop is installed properly or not.
- Now change directory to hadoop/sbin in cmd and type “start-all” to run all the daemons in JVMs.

3) Program Code:

Following is the list of files and their respecting content to be configured (changes to be made under respective tags as mentioned):

(a) Configuration for mapred-site.xml

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

Shubham Mathur (JDAT)

```
</configuration>
```

(b) Configuration for core-site.xml

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

(c) Configuration for hdfs-site.xml

```
<configuration>
  <property>
    <name>dfs.replication</name>
    <value>1</value>
  </property>
  <property>
    <name>dfs.namenode.name.dir</name>
    <value>PATH TO NAMENODE</value>
  </property>
  <property>
    <name>dfs.datanode.data.dir</name>
    <value>PATH TO DATANODE</value>
  </property>
</configuration>
```

(d) Configuration for yarn-site.xml

```
<configuration>
  <property>
    <name>yarn.nodemanager.aux-services</name>
    <value>mapreduce_shuffle</value>
  </property>
  <property>
    <name>yarn.nodemanager.auxservices.mapreduce.shuffle.class</name>
    <value>org.apache.hadoop.mapred.ShuffleHandler</value>
  </property>
</configuration>
```

Shubham Mathur (JDAT)

4) Sample output of a program:

N/A

5) Conclusion: Hadoop has been installed and configured in Windows successfully.