



VIT[®]
Vellore Institute of Technology
(Deemed to be University under section 3 of UGC Act, 1956)

DIGITAL ASSIGNMENT 1

Name: Soumyadeep Ganguly

Reg No: 24MDT0082

Course: Big Data Analytics Lab

Course Code: PMDS507P

Step 1: Check java version

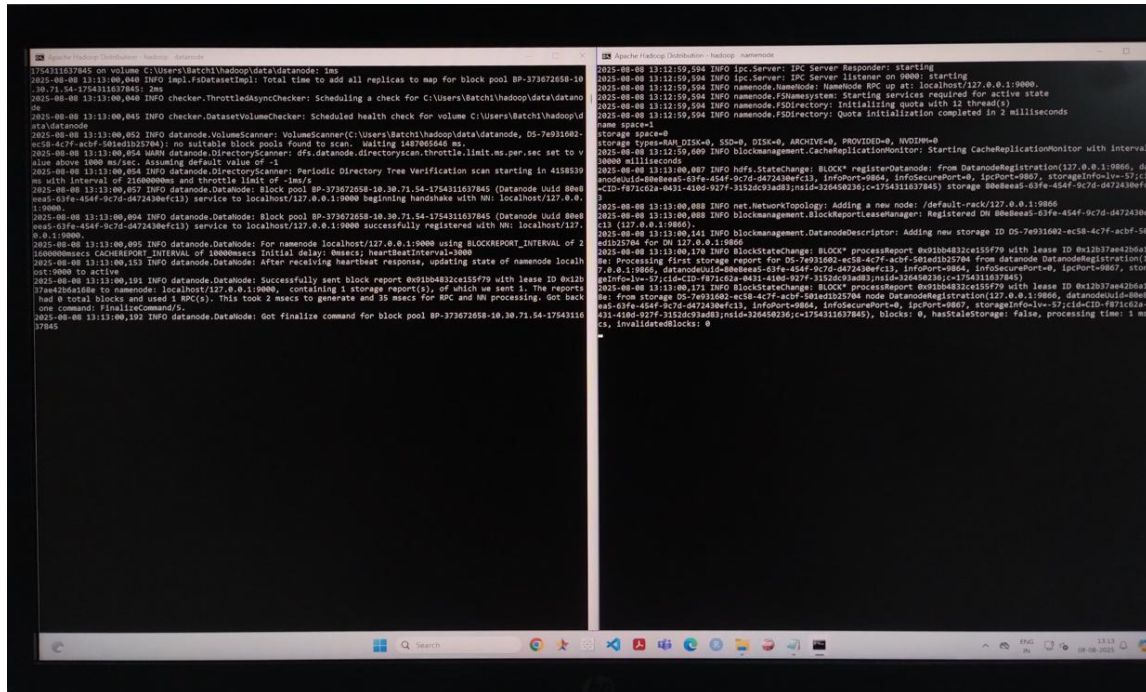
```
C:\Users\Batch1>java -version
java version "1.8.0_461"
Java(TM) SE Runtime Environment (build 1.8.0_461-b11)
Java HotSpot(TM) 64-Bit Server VM (build 25.461-b11, mixed mode)
```

Step 2: Check Hadoop version

```
C:\Users\Batch1>hadoop version
Hadoop 3.4.1
Source code repository https://github.com/
Compiled by mthakur on 2024-10-09T14:57Z
```

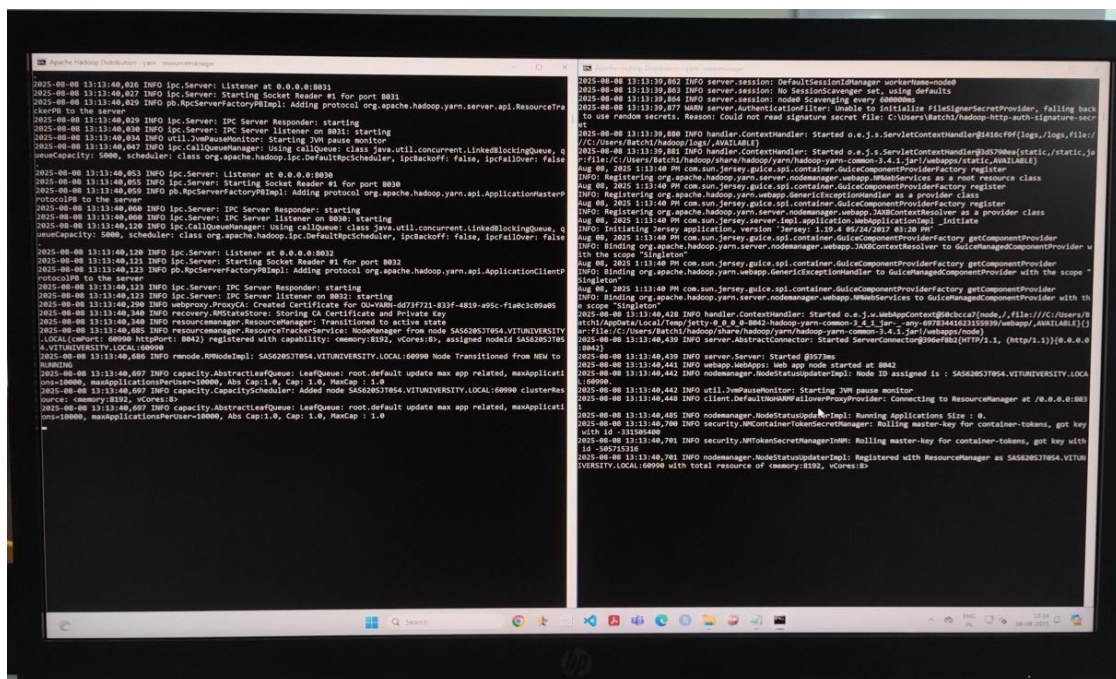
Step 3: Go to Command Prompt and type: start-dfs.cmd

This command will start the “NameNode” and “DataNode” in separate windows.



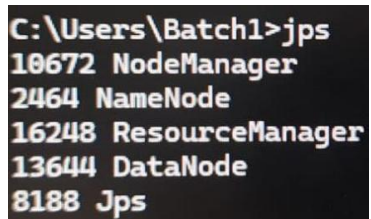
Step 4: Now type: start-yarn.cmd

It will start the “NodeManager” and “ResourceManager” in 2 separate windows.



Step 5: Now type: jps

It will display a list of all currently running Java processes on the local system.



```
C:\Users\Batch1>jps
10672 NodeManager
2464 NameNode
16248 ResourceManager
13644 DataNode
8188 Jps
```

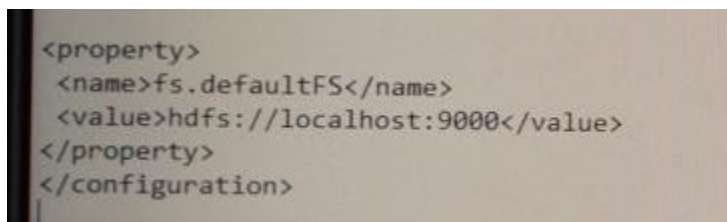
Step 6: Configure **core-site.xml** under the Hadoop folder

```
<property>
```

```
<name>fs.defaultFS</name>
```

```
<value>hdfs://localhost:9000</value>
```

```
</property>
```



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

Step 7: Configuration for **hdfs-site.xml** file.

```
<property>
```

```
<name>dfs.replication</name>
```

```
<value>1</value>
```

```
</property>
```

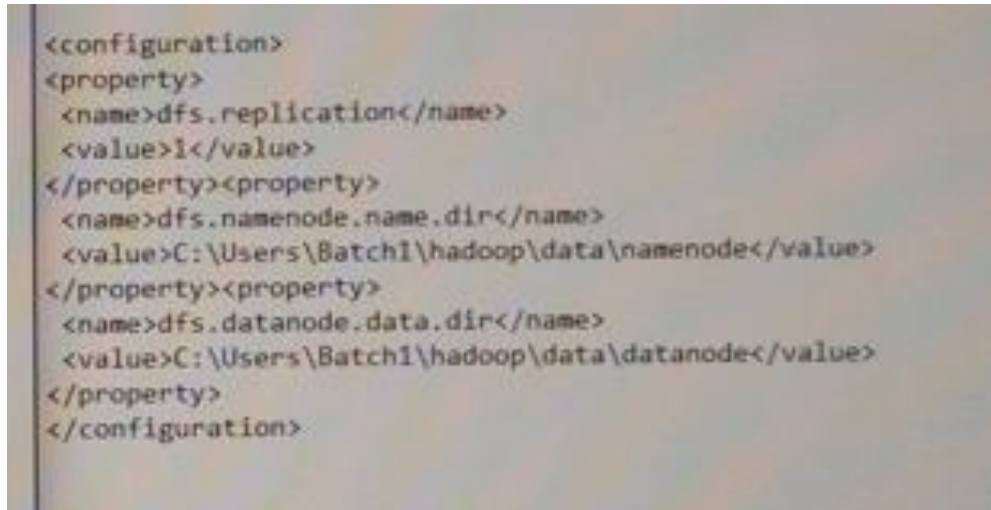
```
<property>
```

```
<name>dfs.namenode.name.dir</name>
```

```
<value>C:\users\batch1\hadoop\data\namenode</value>
```

```
</property>
```

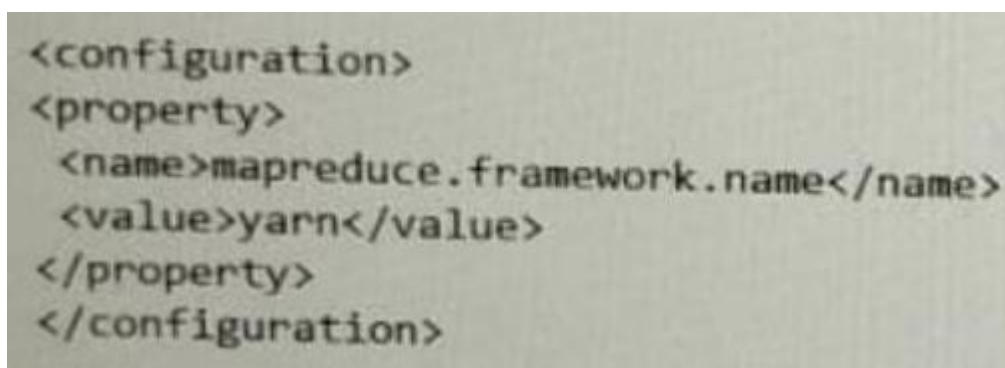
```
<property>
<name>dfs.namenode.name.dir</name>
<value>C:\users\batch1\hadoop\data\datanode</value>
</property>
```

A screenshot of a code editor showing XML configuration for Hadoop DFS. The code is as follows:

```
<configuration>
  <property>
    <name>dfs.replication</name>
    <value>1</value>
  </property>
  <property>
    <name>dfs.namenode.name.dir</name>
    <value>C:\Users\Batch1\hadoop\data\namenode</value>
  </property>
  <property>
    <name>dfs.datanode.data.dir</name>
    <value>C:\Users\Batch1\hadoop\data\datanode</value>
  </property>
</configuration>
```

Step 8: Configuration for **mapred-site.xml**

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

A screenshot of a code editor showing XML configuration for mapred-site.xml. The code is as follows:

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

Step 9: Configuration for **yarn-site.xml** file

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

```
<configuration>
<property>
  <name>yarn.tmp.dir</name>
  <value>C:\Users\Batch1\tmp\yarn-${user.name}</value>
  <description>A base for other temporary directories.</description>
</property>
<property>
  <name>yarn.nodemanager.aux-services</name>
  <value>mapreduce_shuffle</value>
</property>
<property>
  <name>yarn.nodemanager.aux-services.mapreduce.shuffle.class</name>
  <value>org.apache.hadoop.mapred.ShuffleHandler</value>
</property>
</configuration>
```