

Data storing using HDFS and Processing using MapReduce:

Step-1:

Open cmd using run as administrator and change directory to sbin folder

Step-2:

Write command : `hdfs namenode -format`

Step-3:

Start dfs using command: `start-dfs`

Step-4:

Start Yarn using command: `start-yarn`

Step-5:

Check all demons are started using command: `jps`

If 1. Datanode 2.namenode 3.nodemanager 4. Resourcemanager is started Than

Step-5:

Create Directory where file is store using HDFS command:

`hadoop fs -mkdir /input` //where input is folder name

Step-6:

Now Upload file using command: `hadoop fs -put C:/data.txt /input`

Step-7: Check file listing using command: `hadoop fs -ls /input/`

Step-8: Read File using command: `hadoop dfs -cat /input/data.txt`

Step-9: Run MapReduce Job using command:

`hadoop jar C:/hadoop-3.3.0/share/hadoop/mapreduce/hadoop-mapreduce-examples-3.3.0.jar wordcount /input /out`

where out is output, directory created by user is not system define. And wordcount is program name.

Step-10:

Open browser and write: `localhost:9870`

Click on utilites _> brows the file system->check input and output directory for output using click head the file(32k).