
Group B

Assignment No: 1

Aim: Write a code in JAVA for a simple WordCount application that counts the number of occurrences of each word in a given input set using the Hadoop MapReduce framework on local-standalone set-up.

Pre-requisite:

i)Java Installation

ii)Hadoop Installation

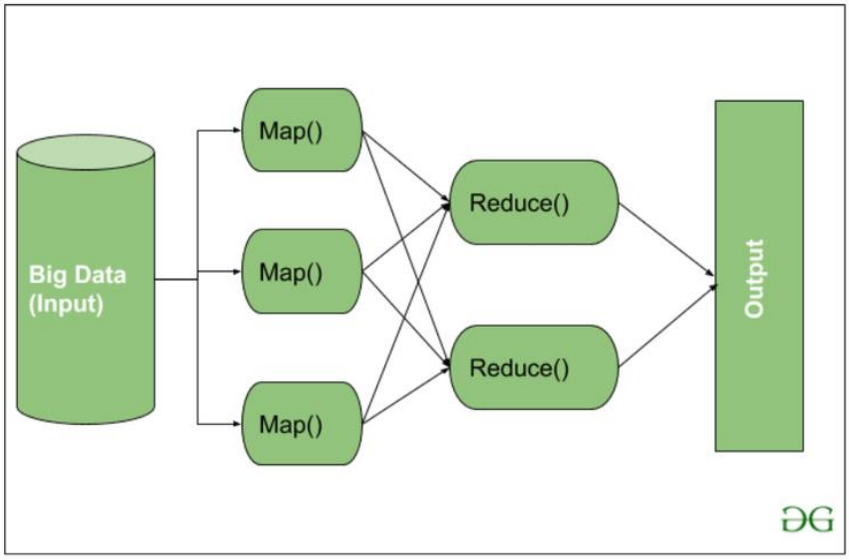
Theory:

Hadoop:

Hadoop is a framework written in Java programming language that works over the collection of commodity hardware. Before Hadoop, we are using a single system for storing and processing data. Also, we are dependent on RDBMS which only stores the structured data. To solve the problem of such huge complex data, Hadoop provides the best solution.

Map Reduce in Hadoop :

One of the three components of Hadoop is Map Reduce. The first component of Hadoop that is, Hadoop Distributed File System (HDFS) is responsible for storing the file. The second component that is, Map Reduce is responsible for processing the file. MapReduce has mainly 2 tasks which are divided phase-wise.In first phase, Map is utilised and in next phase Reduce is utilised



Wordcount Program steps

Step 1) mkdir words

Step 2) Download hadoop-core-1.2.1.jar, which is used to compile and execute the MapReduce program. Visit the following
link
http://mvnrepository.com/artifact/org.apache.hadoop/hadoop-core/1.2.1

Step 3) Put that downloaded jar file into words folder.

Step 4) Implement WordCount.java program.

Step 5) Create input1.txt on home directory with some random text

Step 6) go on words path then compile

javac -classpath /home/vijay/words/hadoop-core-1.2.1.jar /home/vijay/words/WordCount.java

javac -classpath $HADOOP_HOME/share/hadoop/mapreduce/hadoop-mapreduce-client-core-3.2.4.jar:$HADOOP_HOME/share/hadoop/mapreduce/hadoop-mapreduce-client-common-3.2.4.jar:$HADOOP_HOME/share/hadoop/common/hadoop-common-3.2.4.jar /home/gurukul/WordCount.java

Step 7) jar -cvf words.jar -c words/ .

Step 8) cd .. then use following commands

hadoop fs -mkdir /input

hadoop fs -put input1.txt /input

hadoop fs -ls /input

hadoop jar /home/vijay/words/words12.jar WordCount /input/input1.txt /out321

hadoop fs -ls /out321

hadoop fs -cat /out321/part-r-00000

(Otherwise check in Browsing HDFS -> Utilities -> Browse the file System -> /)

Assignment Questions

1. **What is the map reduce explain with a small example?**
2. **Write down steps to install hadoop.**