Roll No: 20BCE204

Course Name and Course Code: 2CS702 Big Data Analytics

Practical No: 4

Aim: Write a Java Code and Execute on Hadoop to count Words of a File.

Code:

import *java.io.IOException*;

import *java.util.StringTokenizer*;

import *org.apache.hadoop.conf.Configuration*;

import *org.apache.hadoop.fs.Path*;

import *org.apache.hadoop.io.IntWritable*;

import *org.apache.hadoop.io.Text*;

import *org.apache.hadoop.mapreduce.Job*;

import *org.apache.hadoop.mapreduce.Mapper*;

import *org.apache.hadoop.mapreduce.Reducer*;

import *org.apache.hadoop.mapreduce.lib.input.FileInputFormat*;

import *org.apache.hadoop.mapreduce.lib.output.FileOutputFormat*;

*public* *class* WordCount {

*public* *static* *class* TokenizerMapper

*extends* Mapper<Object, Text, Text, IntWritable>{

*private* *final* *static* IntWritable one = new IntWritable(1);

*private* Text word = new Text();

*public* void map(Object key, Text value, Context context

) *throws* IOException, InterruptedException {

StringTokenizer itr = new StringTokenizer(value.toString());

while (itr.hasMoreTokens()) {

word.set(itr.nextToken());

context.write(word, one);

}

}

}

*public* *static* *class* IntSumReducer

*extends* Reducer<Text,IntWritable,Text,IntWritable> {

*private* IntWritable result = new IntWritable();

*public* void reduce(Text key, Iterable<IntWritable> values,

Context context

) *throws* IOException, InterruptedException {

int sum = 0;

for (IntWritable val : values) {

sum += val.get();

}

result.set(sum);

context.write(key, result);

}

}

*public* *static* void main(String[] args) *throws* Exception {

Configuration conf = new Configuration();

Job job = Job.getInstance(conf, "word count");

job.setJarByClass(WordCount.class);

job.setMapperClass(TokenizerMapper.class);

job.setCombinerClass(IntSumReducer.class);

job.setReducerClass(IntSumReducer.class);

job.setOutputKeyClass(Text.class);

job.setOutputValueClass(IntWritable.class);

FileInputFormat.addInputPath(job, new Path(args[0]));

FileOutputFormat.setOutputPath(job, new Path(args[1]));

System.exit(job.waitForCompletion(true) ? 0 : 1);

}

}

Steps in Terminal:

% hadoop fs -mkdir /Practical4

% hadoop fs -mkdir /Practical4/Input

% hadoop fs -put /Users/dhyan/Desktop/Practical4/input.txt /Practical4/Input

% export JAVA\_HOME=`/usr/libexec/java\_home -v 1.8.0\_333`

% export HADOOP\_CLASSPATH=$(hadoop classpath)

% javac -classpath ${HADOOP\_CLASSPATH} -d /Users/dhyan/Desktop/Practical4/Classes /Users/dhyan/Desktop/Practical4/WordCount.java

% jar -cvf practical4.jar -C /Users/dhyan/Desktop/Practical4/Classes .

% hadoop jar /Users/dhyan/Desktop/Practical4/practical4.jar WordCount /Practical4/Input /Practical4/Output

% hadoop dfs -cat /Practical4/Output/part-r-00000

Output:







