**PRATHEEK P NAYAK  
1NT19IS117**

**EXERCISE 3:**

**Word Count Program:**

package wcsample;

import java.io.IOException;

import java.util.Iterator;

import java.util.StringTokenizer;

import org.apache.hadoop.fs.Path;

import org.apache.hadoop.io.IntWritable;

import org.apache.hadoop.io.LongWritable;

import org.apache.hadoop.io.Text;

import org.apache.hadoop.mapred.FileInputFormat;

import org.apache.hadoop.mapred.FileOutputFormat;

import org.apache.hadoop.mapred.JobClient;

import org.apache.hadoop.mapred.JobConf;

import org.apache.hadoop.mapred.MapReduceBase;

import org.apache.hadoop.mapred.Mapper;

import org.apache.hadoop.mapred.OutputCollector;

import org.apache.hadoop.mapred.Reducer;

import org.apache.hadoop.mapred.Reporter;

import org.apache.hadoop.mapred.TextInputFormat;

import org.apache.hadoop.mapred.TextOutputFormat;

public class wordcount {

public static class Map extends MapReduceBase implements Mapper<LongWritable, Text, Text, IntWritable> {

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

private Text word = new Text();

@Override

public void map(LongWritable key, Text value, OutputCollector<Text, IntWritable> output, Reporter arg3)

throws IOException {

String line = value.toString();

StringTokenizer tokenizer = new StringTokenizer(line);

while (tokenizer.hasMoreTokens()) {

word.set(tokenizer.nextToken());

output.collect(word, one);

}

}

}

public static class Reduce extends MapReduceBase implements Reducer<Text, IntWritable, Text, IntWritable> {

@Override

public void reduce(Text key, Iterator<IntWritable> value, OutputCollector<Text, IntWritable> output,

Reporter arg3) throws IOException {

// TODO Auto-generated method stub

int sum = 0;

while (value.hasNext()) {

sum += value.next().get();

}

output.collect(key, new IntWritable(sum));

}

}

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

// TODO Auto-generated method stub

JobConf conf = new JobConf(wordcount.class);

conf.setJobName("wordcount");

conf.setOutputKeyClass(Text.class);

conf.setOutputValueClass(IntWritable.class);

//Providing the mapper and reducer class names

conf.setMapperClass(Map.class);

conf.setCombinerClass(Reduce.class); //set theCombiner class

conf.setReducerClass(Reduce.class);

conf.setInputFormat(TextInputFormat.class);

conf.setOutputFormat(TextOutputFormat.class);

FileInputFormat.setInputPaths(conf, new Path(args[0]));

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

JobClient.runJob(conf);

}

}

OUTPUT:

hdoop@admin1-HP-280-G4-MT-Business-PC:~$ cd $HADOOP\_HOME

hdoop@admin1-HP-280-G4-MT-Business-PC:~/hadoop-3.2.1$ cd sbin

hdoop@admin1-HP-280-G4-MT-Business-PC:~/hadoop-3.2.1/sbin$ jps

11174 Jps

10646 NodeManager

10311 ResourceManager

9688 NameNode

9867 DataNode

10123 SecondaryNameNode

9342 org.eclipse.equinox.launcher\_1.5.600.v20191014-2022.jar

hdoop@admin1-HP-280-G4-MT-Business-PC:~/hadoop-3.2.1/sbin$ start-all.sh

WARNING: Attempting to start all Apache Hadoop daemons as hdoop in 10 seconds.

WARNING: This is not a recommended production deployment configuration.

WARNING: Use CTRL-C to abort.

Starting namenodes on [localhost]

localhost: namenode is running as process 9688. Stop it first.

Starting datanodes

localhost: datanode is running as process 9867. Stop it first.

Starting secondary namenodes [admin1-HP-280-G4-MT-Business-PC]

admin1-HP-280-G4-MT-Business-PC: secondarynamenode is running as process 10123. Stop it first.

Starting resourcemanager

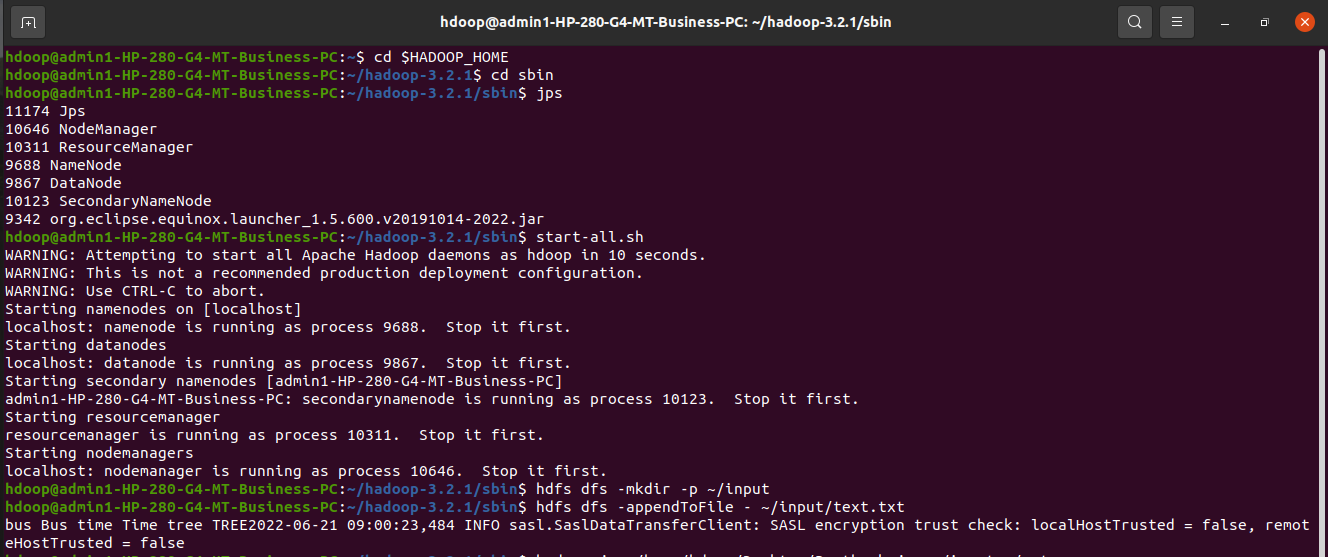
resourcemanager is running as process 10311. Stop it first.

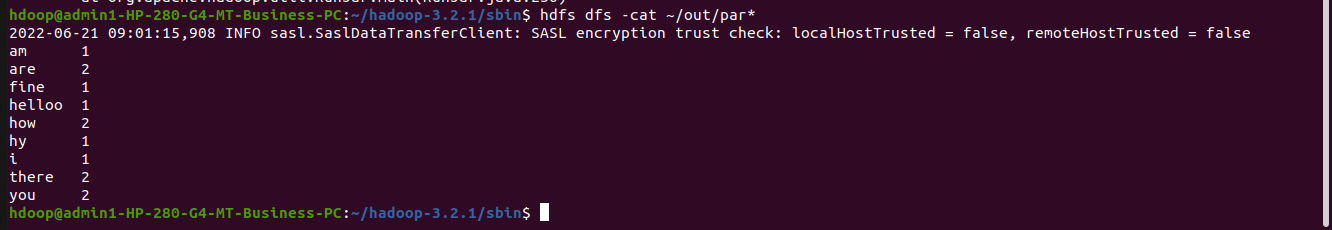
Starting nodemanagers

localhost: nodemanager is running as process 10646. Stop it first.

hdoop@admin1-HP-280-G4-MT-Business-PC:~/hadoop-3.2.1/sbin$ hdfs dfs -mkdir -p ~/input

hdoop@admin1-HP-280-G4-MT-Business-PC:~/hadoop-3.2.1/sbin$ hdfs dfs -appendToFile - ~/input/text.txt

bus Bus time Time tree TREE2022-06-21 09:00:23,484 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false

hdoop@admin1-HP-280-G4-MT-Business-PC:~/hadoop-3.2.1/sbin$ hdfs dfs -cat ~/out/par\*