**NITTE MEENAKSHI INSTITUTE OF TECHNOLOGY**

**(An autonomous institution with A Grade by NAAC /UGC, Affiliated to Visvesvaraya Technological University, Belgaum, Approved by UGC/AICTE/Govt. of Karnataka)**

**Yelahanka, Bengaluru-560064**

**ACADEMIC YEAR 2020-2021**  


**BIG DATA**

REPORT ON

**HADOOP ,MapReduce PROGRAMMING AND HDFS FILE SYSTEMS**

Submitted By

**Yash(1NT17IS202)**

**Yatanvesh Bhardwaj(1NT17IS203)**

Submitted To

**Mr. Prashanth B S**

**Asst.Prof Dept of ISE,NMIT**

**DEPARTMENT OF INFORMATION SCIENCE AND ENGINEERING**

**NITTE MEENAKSHI INSTITUTE OF TECHNOLOGY,BANGALORE 560064**

**Problem statement 1:**

Create a dataset in excel as .csv file and it should contain the following fields with at least 20 sample datasets in it.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| NAME | USN | SUBJECT1 | SUBJECT2 | SUBJECT3 | PASS/FAIL |
| Agasthya | 1NT17IS111 | 78 | 85 | 99 | PASS |

Use the Hadoop MapReduce programming framework to come up with a Program which will take the data from this .csv file and computes the following

1. Total number of students who have scored more than 60 in Subject 1

2. Total number of students who have passed in all the subjects.

package com.yatan.yash;

import java.io.IOException;

import java.util.\*;

import org.apache.hadoop.fs.Path;

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

import org.apache.hadoop.mapred.\*;

public class Counter {

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

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

public void map(LongWritable *key*,Text *value*,OutputCollector<Text,IntWritable> *output*,Reporter *reporter*)throws IOException{

String row = value.toString();

String[] cols = row.split(",");

if(Integer.parseInt(cols[3])>60)

output.collect(new Text("Number of Students having score in subject 1> 60:"),writer);

if(cols[5].equals("PASS"))

output.collect(new Text("Passed Students:"),writer);

}

}

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

public void reduce(Text *key*,Iterator<IntWritable> *values*,OutputCollector<Text,IntWritable> *output*,Reporter *reporter*)throws IOException{

int count=0;

while(values.hasNext()){

IntWritable value =values.next();

count += value.get();

}

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

}

}

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

JobConf conf = new JobConf(Counter.class);

conf.setJobName("counter\_job");

conf.setOutputKeyClass(Text.class);

conf.setOutputValueClass(IntWritable.class);

conf.setMapperClass(Map.class);

conf.setCombinerClass(Reduce.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

# Compile program

$HADOOP\_HOME/bin/hadoop com.sun.tools.javac.Main ~/docker-hadoop/Counter.java -d ~/docker-hadoop/build/

# Create jar

jar -cvf ~/docker-hadoop/Counter.jar .

# Copy jar file into docker namenode process

docker cp ~/docker-hadoop/build/Counter.jar 9c673589c50f:Counter.jar

# Open namenode bash shell

docker exec -it namenode bashnamenode bash

root@9c673589c50f:/# hdfs dfs -mkdir -p input

root@9c673589c50f:/# hdfs dfs -mkdir -p output

# Copy input csv

root@9c673589c50f:/# hdfs dfs -put ~/docker-hadoop/data.csv ~/input

# Run jar inside docker bash shell

root@9c673589c50f:/# hadoop jar ./Counter.jar com.yatan.yash.Counter ~/input.csv ~/output

root@9c673589c50f:/# hdfs dfs -cat ~/output/part\*

Number of Students having score in subject 1 > 60: 11

Passed Students: 16