Task 1 : highest temp

Input data :: sensor id -date -temp

s1,2016-01-01,20.5

s2,2016-01-01,30.1

s1,2016-01-02,60.2

s2,2016-01-02,20.4

s1,2016-01-03,55.5

s2,2016-01-03,52.5

Output data

60.2

-----------

Task 2: count of temperature greater than 50 for every sensor

Input data

s1,2016-01-01,20.5

s2,2016-01-01,30.1

s1,2016-01-02,60.2

s2,2016-01-02,20.4

s1,2016-01-03,55.5

s2,2016-01-03,52.5

Output data - count , sensor name

2, s1

1, s2

////////////////////////////////////////////////////////////////////

# build.sbt

ThisBuild/version:="0.1.0-SNAPSHOT"

ThisBuild/scalaVersion:="2.11.10"

//https://mvnrepository.com/artifact/commons-io/commons-io

libraryDependencies+="commons-io"%"commons-io"%"2.6"

//https://mvnrepository.com/artifact/org.apache.spark/spark-core

libraryDependencies+="org.apache.spark"%%"spark-core"%"2.4.3"

//https://mvnrepository.com/artifact/org.apache.spark/spark-sql

libraryDependencies+="org.apache.spark"%%"spark-sql"%"2.4.3"

//https://mvnrepository.com/artifact/org.apache.spark/spark-streaming

libraryDependencies+="org.apache.spark"%%"spark-streaming"%"2.4.3"

//https://mvnrepository.com/artifact/org.apache.hadoop/hadoop-client

libraryDependencies+="org.apache.hadoop"%"hadoop-client"%"2.7.1"

/////////////////////////////////////////////////

package org.itc.com

import org.apache.spark.SparkContext

object Main {

def main(args: Array[String]): Unit = {

println("Hello world!")

val sc = new SparkContext(master = "local[1]", appName = "firstWC")

val rdd1 = sc.textFile(path = "D:/spark\_code/task1/input/temp.txt")

rdd1.collect().foreach(println)

val rdd2 = rdd1.map(x => x.split(",")(2).toDouble)

val max\_temp = rdd2.max()

println(max\_temp)

val rdd3 = rdd1.map(line => {

val Array(sensor\_name, \_, temp) = line.split(",")

(sensor\_name, temp.toDouble)

})

val rdd4 = rdd3.filter { case (\_, temp) => temp > 50 }

.map { case (sensor\_name, \_) => (sensor\_name, 1) }

.reduceByKey( + )

rdd4.collect().foreach(println)

rdd4.saveAsTextFile( path = "D:/spark\_code/task1/output/result")

}

}





