Assignment Solution

Week9: Apache Spark - General Purpose Cluster Computing Framework

Solution 1 :

```
val rdd1 = spark.sparkContext.textFile("/Users/trendytech/Desktop/dataset1")
    rdd1.collect().foreach(println)

    val rdd2 = rdd1.map(line => {
      val fields = line.split(",")
      if (fields(1).toInt > 18)
        (fields(0),fields(1),fields(2),"Y")
      else
        (fields(0),fields(1),fields(2),"N")
     })

    rdd2.collect().foreach(println)
```

Solution 2 :

```
import org.apache.spark._
import org.apache.spark.SparkContext._
import org.apache.log4j._
import scala.math.min

/** Find the minimum temperature by weather station */
object MinTemperatures {

  def parseLine(line:String)= {
    val fields = line.split(“,”)
    val stationID = fields(0)
    val entryType = fields(2)
    val temperature = fields(3)
    (stationID, entryType, temperature)
  }
```

Cont...

```
/** Our main function where the action happens */
def main(args: Array[String]) {

  // Set the log level to only print errors
  Logger.getLogger("org").setLevel(Level.ERROR)

  // Create a SparkContext using every core of the local machine
  val sc = new SparkContext("local[*]", "MinTemperatures")

  // Read each line of input data
  val lines = sc.textFile("/Users/sumitm/Desktop/spark-data/temp-data.csv")

  // Convert to (stationID, entryType, temperature) tuples
  val parsedLines = lines.map(parseLine)

  // Filter out all but TMIN entries
  val minTemps = parsedLines.filter(x => x._2 == "TMIN")
```

Cont...

```
    // Convert to (stationID, temperature)
    val stationTemps = minTemps.map(x => (x._1, x._3.toFloat))

    // Reduce by stationID retaining the minimum temperature found
    val minTempsByStation = stationTemps.reduceByKey( (x,y) => min(x,y))

    // Collect, format, and print the results
    val results = minTempsByStation.collect()

    for (result <- results.sorted) {
       val station = result._1
       val temp = result._2
       val formattedTemp = f"$temp%.2f F"
       println(s"$station minimum temperature: $formattedTemp")
    }
  }
}
```

Google REVIEWS

5 Star Google Rated
Big Data Course

LEARN FROM THE EXPERT

9108179578

Call for more details