**Group A**

**Lab 3B**

**Scala**

**Participants:**

**1. Aadarsha chapagain**

**2.Jyoti shukla**

**3.Rishi Phaneendra Varma**

**4.Priti Bhale**

**5.Sreya Treesa Johny**

**6.Piyush Bhatia**

**Code:**

**package org.pedram.tarining**

**import org.apache.spark.SparkConf**

**import org.apache.spark.SparkContext**

**object ReportGenerator {**

**def main(args: Array[String]) {**

**val conf = (new SparkConf)**

**.setAppName("Sales Report Generator")**

**.setMaster("local[\*]")**

**val sc = new SparkContext(conf)**

**val rdd = sc.textFile("C:/Users/aadar/Downloads/SalesJan2009.csv")**

**val pairRdd = rdd.map(line => {**

**val elements = line.split(",")**

**(elements(2).trim, elements(1).trim.toInt)**

**})**

**val aggRdd = pairRdd.reduceByKey((x, y) => x + y)**

**aggRdd.foreach(element => println(element.\_1 + " = " + element.\_2))**

**}**

**}**

**Slide 54:**

**Graphical user interface, text, application

Description automatically generated**

**Slide 55:**

**A screenshot of a computer

Description automatically generated**

**Slide 56:**

**A screenshot of a computer

Description automatically generated**

**Code To run in sandbox Hortonworks:**

package org.pedram.tarining

import org.apache.spark.SparkConf

import org.apache.spark.SparkContext

object ReportGenerator {

def main(args: Array[String]) {

val conf = (new SparkConf)

.setAppName("Sales Report Generator")

val sc = new SparkContext(conf)

val rdd = sc.textFile("/user/aadarsha/SalesJan2009.csv")

val pairRdd = rdd.map(line => {

val elements = line.split(",")

(elements(2).trim, elements(1).trim.toInt)

})

val aggRdd = pairRdd.reduceByKey((x, y) => x + y)

aggRdd.foreach(element => println(element.\_1 + " = " + element.\_2))

}

}

**Slide 63:**

**Text

Description automatically generated**

**Slide 64:**

**Text

Description automatically generated**