

Part 1)

```
scala> val rdd = sc.textFile("sales.txt")
rdd: org.apache.spark.rdd.RDD[String] = sales.txt MapPartitionsRDD[1] at textFile at <console>:24

scala> val pairs = rdd.map( line => {
|   val p = line.split(" ")
|   (p(0), p(1).toInt)
| })
pairs: org.apache.spark.rdd.RDD[(String, Int)] = MapPartitionsRDD[2] at map at <console>:25

scala> val totals = pairs.reduceByKey(_ + _)
totals: org.apache.spark.rdd.RDD[(String, Int)] = ShuffledRDD[3] at reduceByKey at <console>:25

scala> totals.collect()
res0: Array[(String, Int)] = Array((screw,168), (nail,5), (washer,66), (bolt,51), (drill,2))
```

Part 2)

```
scala> case class Sale(part: String, qty: Int)
defined class Sale

scala> import spark.implicits._
import spark.implicits._

scala> val ds = spark.read.textFile("sales.txt").map(line => {
|   val p = line.split("\s+")
|   Sale(p(0), p(1).toInt)
| })
ds: org.apache.spark.sql.Dataset[Sale] = [part: string, qty: int]

scala> ds.printSchema()
root
|-- part: string (nullable = true)
|-- qty: integer (nullable = false)

scala> val result = ds.groupBy("part").sum("qty")
result: org.apache.spark.sql.DataFrame = [part: string, sum(qty): bigint]

scala> result.show()
+-----+-----+
| part|sum(qty)|
+-----+-----+
| washer|    66|
| bolt|    51|
| nail|     5|
| screw|   168|
| drill|     2|
+-----+-----+
```

Task 3)

```
scala> import spark.implicits._  
import spark.implicits._  
  
scala> val df = spark.read.textFile("sales.txt").map(line => {  
| val p = line.split("\\s+")  
| (p(0), p(1).toInt)  
| }).toDF("part", "qty")  
df: org.apache.spark.sql.DataFrame = [part: string, qty: int]  
  
scala> df.createOrReplaceTempView("sales")  
  
scala> val result = spark.sql("""  
| SELECT part, SUM(qty) AS total_qty  
| FROM sales  
| GROUP BY part  
| """)  
result: org.apache.spark.sql.DataFrame = [part: string, total_qty: bigint]  
  
scala> result.show()  
+-----+-----+  
| part|total_qty|  
+-----+-----+  
| washer|      66|  
| bolt|      51|  
| nail|       5|  
| screw|    168|  
| drill|       2|  
+-----+-----+
```