

```
scala> case class Part( name: String, total:Long)
defined class Part
```

```
scala> val lines = sc.textFile("sales.txt")
lines: org.apache.spark.rdd.RDD[String] = sales.txt MapPartitionsRDD[3]
at textFile at <console>:24
```

```
scala> val partsDS = lines.map(_._split("
")).map(attributes=>Part(attributes(0),attributes(1).toInt)).toDS()
partsDS: org.apache.spark.sql.Dataset[Part] = [name: string, total:
bigint]
```

```
scala> partsDS.show()
```

```
+-----+-----+
| name|total|
+-----+-----+
| bolt|    45|
| bolt|     5|
|drill|     1|
|drill|     1|
|screw|     1|
|screw|     2|
|screw|     3|
+-----+-----+
```

```
scala> val partsDF = partsDS.toDF();
partsDF: org.apache.spark.sql.DataFrame = [name: string, total: bigint]
```

```
scala> partsDF.show()
```

```
+-----+-----+
| name|total|
+-----+-----+
| bolt|    45|
| bolt|     5|
|drill|     1|
|drill|     1|
|screw|     1|
|screw|     2|
|screw|     3|
+-----+-----+
```

```
scala> partsDF.createOrReplaceTempView("partsTable")
```

```
scala> spark.sql("SELECT name, sum(total) FROM partsTable GROUP BY name")
res3: org.apache.spark.sql.DataFrame = [name: string, sum(total): bigint]
```

```
scala> res3.show()
```

```
+-----+-----+
| name|sum(total)|
+-----+-----+
| bolt|          50|
|drill|           2|
|screw|           6|
+-----+-----+
```

```
scala> val total = spark.sql("SELECT name, sum(total) FROM partsTable  
GROUP BY name")  
total: org.apache.spark.sql.DataFrame = [name: string, sum(total):  
bigint]
```

```
scala> total.show();
```

```
+-----+-----+  
| name|sum(total)|  
+-----+-----+  
| bolt|          50|  
|drill|           2|  
|screw|           6|  
+-----+-----+
```