

```
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 totalDS = partsDS.groupBy($"name").sum("total")
totalDS: org.apache.spark.sql.DataFrame = [name: string, sum(total):
bigint]

scala> totalDS.show()
+-----+-----+
| name|sum(total)|
+-----+-----+
| bolt|         50|
|drill|         2|
|screw|         6|
+-----+-----+
```