

```
scala> val data = sc.textFile("/user/acadgild/hadoop/19_Dataset.docx")  
data: org.apache.spark.rdd.RDD[String] = /user/acadgild/hadoop/19_Dataset.docx MapPartitionsRDD[13] at textFile at <console>:24
```

```
scala> val row = data.map(line=>line.split(",")).map(array=>(array(0),array(1),array(2),array(3),array(4)))  
row: org.apache.spark.rdd.RDD[(String, String, String, String, String)] = MapPartitionsRDD[15] at map at <console>:26
```

```
scala> █
```