1. What are the total number of gold medal winners every year

Create a case class Sports

**case class Sports(firstname:String,lastname:String,sports:String,medal\_type:String,age:Int,year:Int,country:String)**

Create RDD from the file

**val sportsRDD = sc.textFile("/home/acadgild/inputdir/Sports\_data.txt")**

**val firstline = sportsRDD.first**

Filter out the first line from RDD as it contains headers

**val dataRDD = sportsRDD.filter(line=>(line!=firstline)).map(line=>line.split(","))**

map the RDD with case class Sports

**val datamapRDD = dataRDD.map(s=>Sports(s(0),s(1),s(2),s(3),s(4).toInt,s(5).toInt,s(6)))**

Convert to dataframe

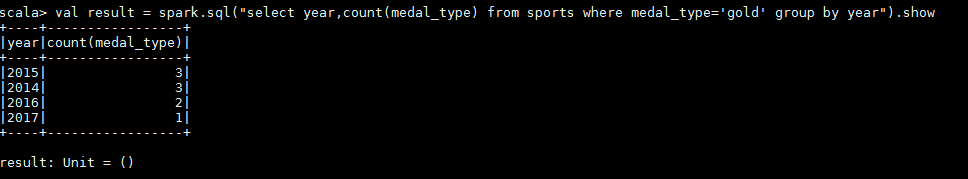
**val sportsDF = datamapRDD.toDF**

Register Dataframe as Temp Table

**sportsDF.registerTempTable("sports")**

Write the spark sql query to get the desired results

**val result = spark.sql("select year,count(medal\_type) from sports where medal\_type='gold' group by year").show**



1. How many silver medals have been won by USA in each sport.

Create a case class Sports

**case class Sports(firstname:String,lastname:String,sports:String,medal\_type:String,age:Int,year:Int,country:String)**

Create RDD from the file

**val sportsRDD = sc.textFile("/home/acadgild/inputdir/Sports\_data.txt")**

**val firstline = sportsRDD.first**

Filter out the first line from RDD as it contains headers

**val dataRDD = sportsRDD.filter(line=>(line!=firstline)).map(line=>line.split(","))**

map the RDD with case class Sports

**val datamapRDD = dataRDD.map(s=>Sports(s(0),s(1),s(2),s(3),s(4).toInt,s(5).toInt,s(6)))**

Convert to dataframe

**val sportsDF = datamapRDD.toDF**

Register Dataframe as Temp Table

**sportsDF.registerTempTable("sports")**

Write the spark sql query to get the desired results

**val result = spark.sql("select country,sports,count(medal\_type) from sports where medal\_type='silver' and country='USA' group by country,sports").show**

