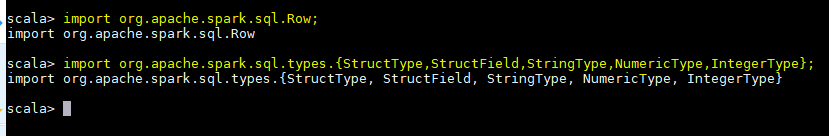
# Using spark-sql, Find:

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

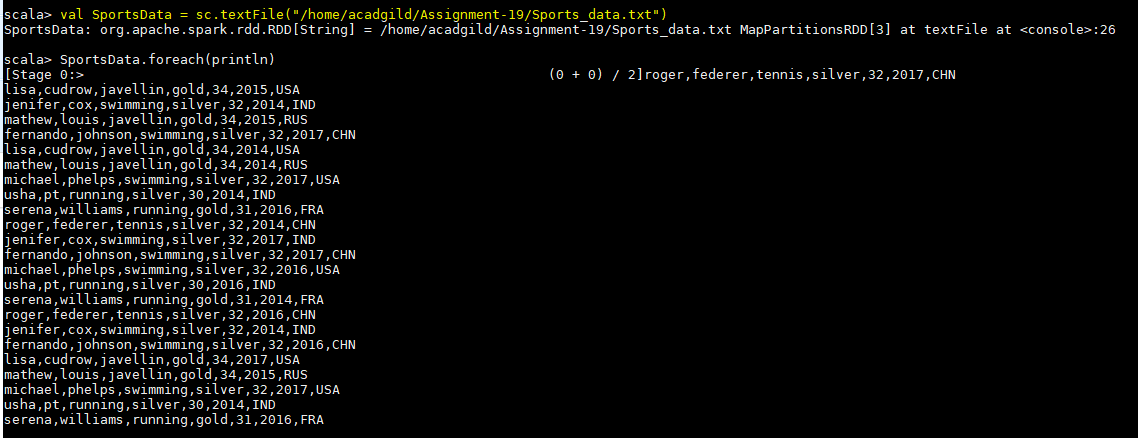
Below is the command used to find the result-

* val SportsData = sc.textFile("/home/acadgild/Assignment-19/Sports\_data.txt")
* val schemaString = "firstname:string,lastname:string,sports:string,medal:string,age:integer,year:integer,country:s tring"
* val schema = StructType(schemaString.split(",").map(x => StructField(x.split(":")(0), if (x.split(":")(1).equals("string")) StringType else IntegerType, true)))
* val rowRDD = SportsData.map(\_.split(",")).map(r => Row(r(0), r(1), r(2), r(3), r(4).toInt, r(5).toInt, r(6)))
* val SportsDataDF = spark.createDataFrame(rowRDD, schema)
* SportsDataDF.createOrReplaceTempView("Sports\_Data")
* val result1DF = spark.sql("SELECT year,COUNT(\*) FROM Sports\_Data WHERE medal = 'gold' GROUP BY year")
* result1DF.show()

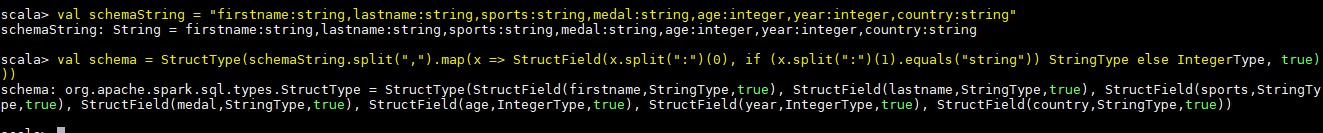
In order to proceed we need to import some dependencies as shown below-

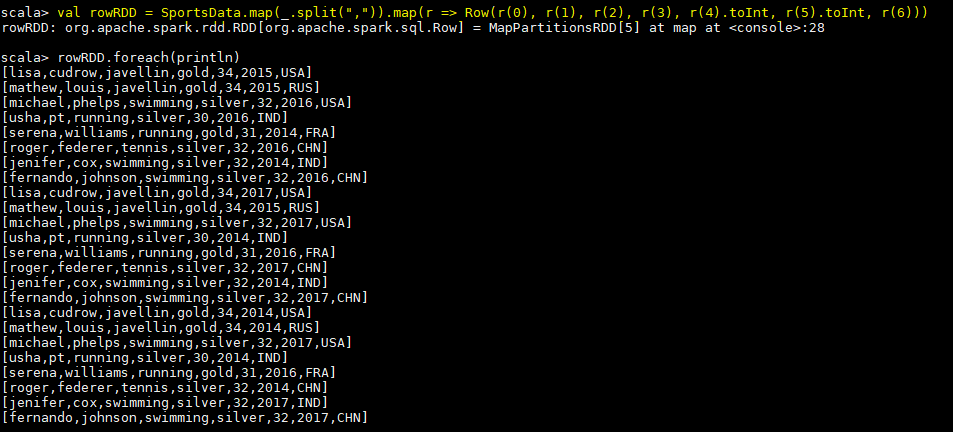


Now we are creating a RDD which reads from the input file-

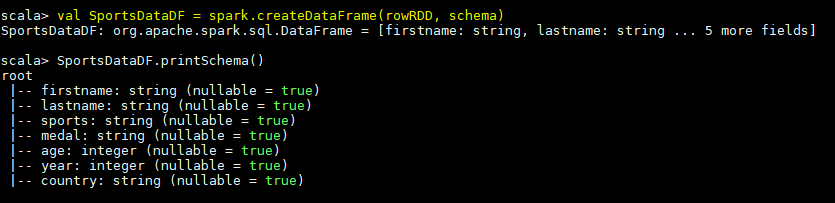


Since it is a text file we need to define schema too. Below screenshot shows the same-

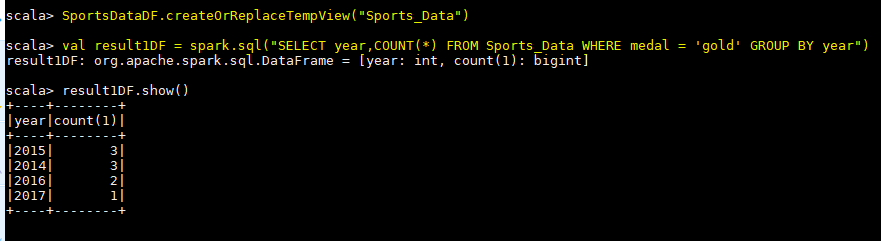




Now we are creating the dataframe by passing the RDD which reads the file and schema to spark session object-



Here we are creating a temporary table first from the dataframe. Finally we can execute our SQL query on the temporary table to find the result-

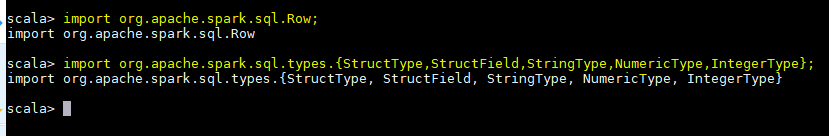


# How many silver medals have been won by USA in each sport?

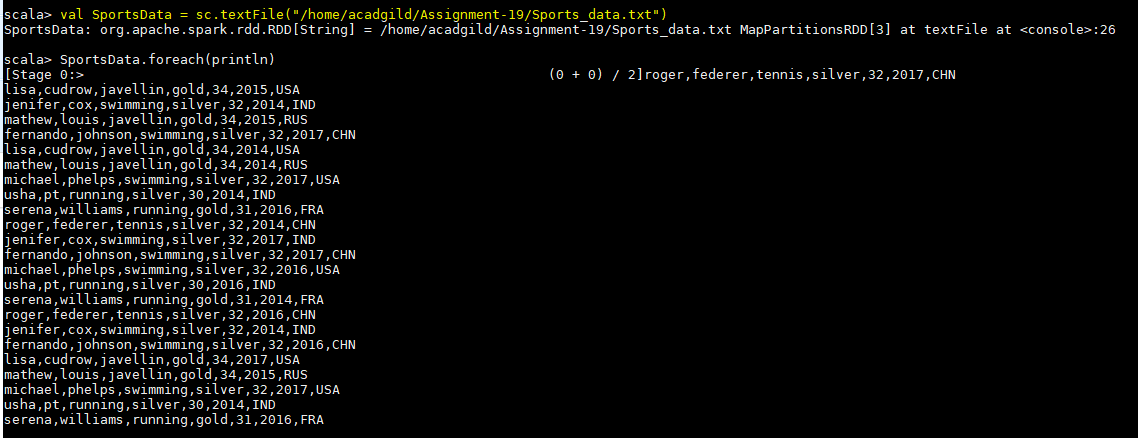
Below is the code used to find the result-

* val SportsData = sc.textFile("/home/acadgild/Assignment-19/Sports\_data.txt")
* val schemaString = "firstname:string,lastname:string,sports:string,medal:string,age:integer,year:integer,country:s tring"
* val schema = StructType(schemaString.split(",").map(x => StructField(x.split(":")(0), if (x.split(":")(1).equals("string")) StringType else IntegerType, true)))
* val rowRDD = SportsData.map(\_.split(",")).map(r => Row(r(0), r(1), r(2), r(3), r(4).toInt, r(5).toInt, r(6)))
* val SportsDataDF = spark.createDataFrame(rowRDD, schema)
* SportsDataDF.createOrReplaceTempView("Sports\_Data")
* val result2DF = spark.sql("SELECT sports,COUNT(\*) FROM Sports\_Data WHERE medal = 'silver' and country ='USA' GROUP BY sports")
* result2DF.show()

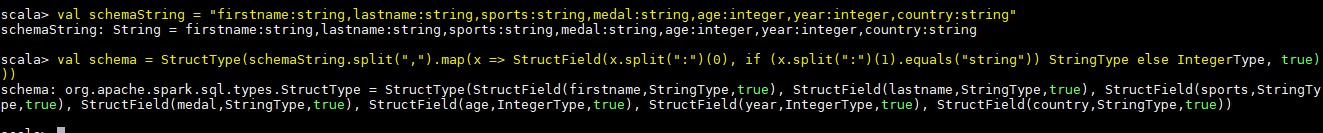
In order to proceed we need to import some dependencies as shown below-

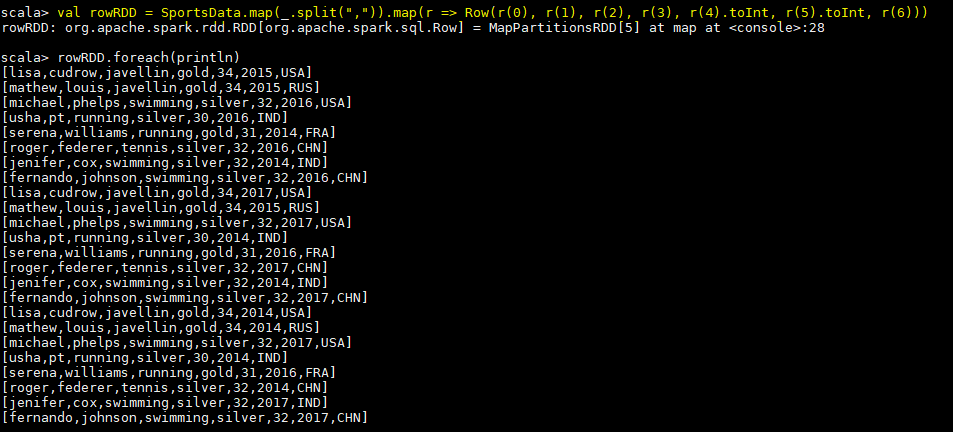


Now we are creating a RDD which reads from the input file-

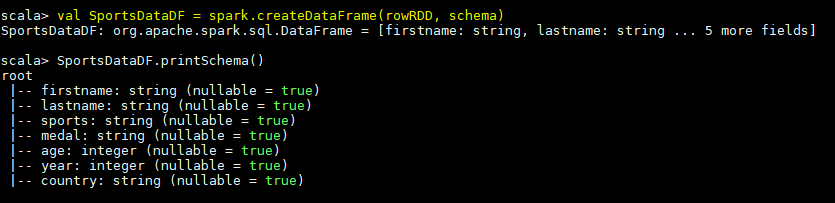


Since it is a text file we need to define schema too. Below screenshot shows the same-





Now we are creating the dataframe by passing the RDD which reads the file and schema to spark session object-



Finally we can execute our query by applying it on the temporary table created-

