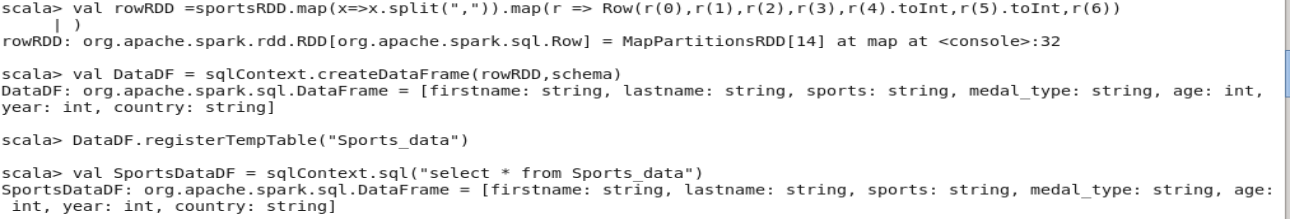
**Assignment 01**

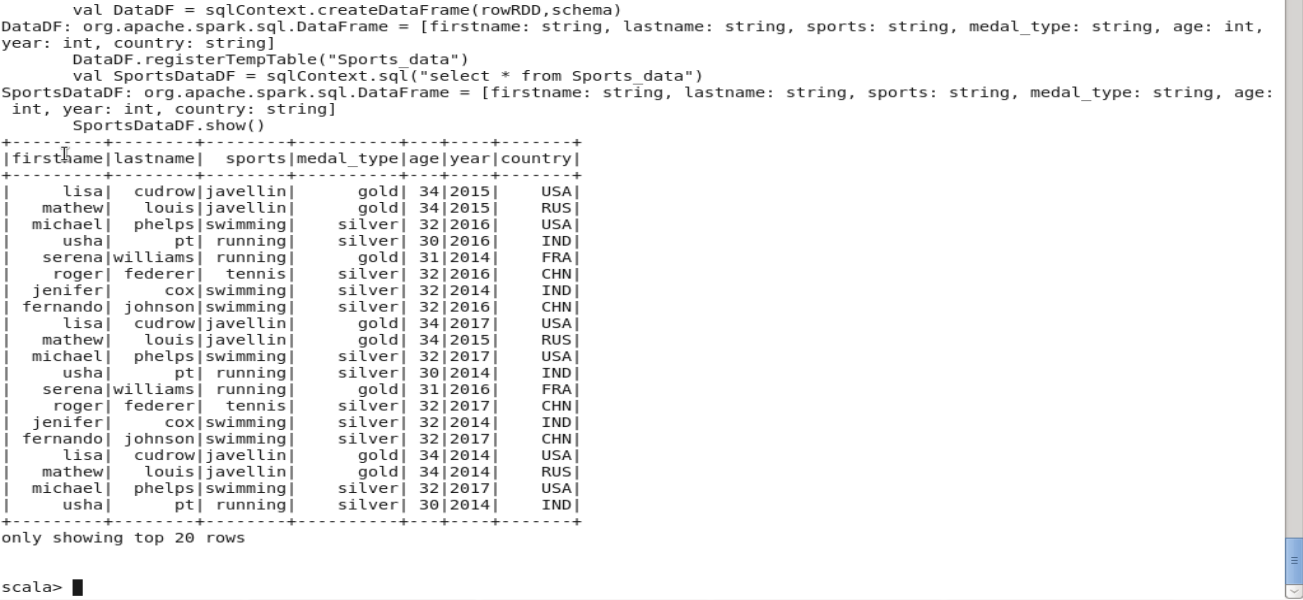
**Data Frame Creation:**

1. Import the spark sql Row, DataFrame and types objects.
2. Create an RDD that will hold the contents of the Sports Dataset(text file format) sportsRDD
3. Create the schema (column name and type) for the data. schemaString
4. Create the schema structure based on the schemaString. schema
5. Create an RDD that will split the data in sportsRDD by ‘,’ and map columns in rows as in schema. rowRDD
6. Create the DataFrame from the rowRDD and schema created before using the “sqlContext” object. SportsDataDF
7. Create a temporary table (Sports\_data) from the DataFrame SportsDataDF



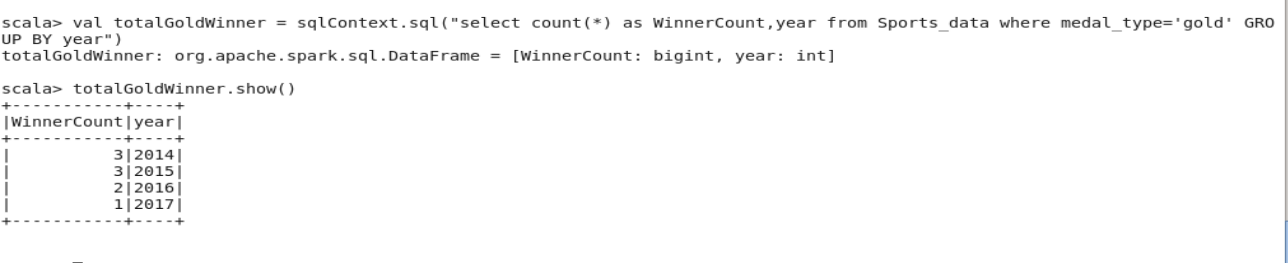


To view the contents of the temp table Sports\_Data, use a sql SELECT Statement with the sqlContext



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

Using the DataFrame that was created above, Get count of the rows of data that have the medal\_type as gold for every year.



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

Using the DataFrame that was created above, Get **sport** and **count** of the rows of data that have the **country** as **USA** and **medal\_type** as **silver** for every **sport** type.

