SPARK SQL

* Load data from input file.

val data=sc.textFile("C:\\Users\\Parab\\Desktop\\acadglid\\new\\s19\_sparl\_sql\\Sports\_data.txt")

* Identify the header

val header =data.first()

* Separate out the header

val data1=data.filter(row=>row!=header)

* Define class according to row.

case class sport(firstName:String,lastName:String,sport:String,medalType:String,age:Int,year:Int,country:String)

* Load data using class.

val sportData=data1.map(x=>x.split(",")).filter(x=>x.length>=7).map(x=>sport(x(0),x(1),x(2),x(3),x(4).toInt,x(5).toInt,x(6))).toDF

* Register DF as table sport\_data

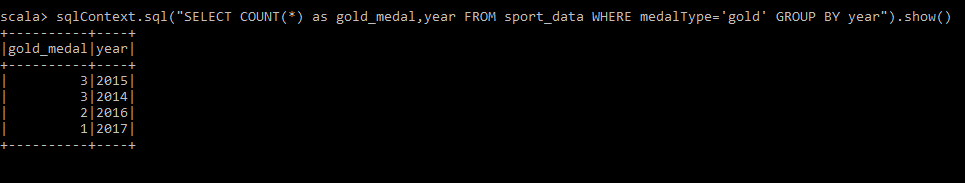
sportData.registerTempTable("sport\_data")

* Intialize sql Context

val sqlContext = new org.apache.spark.sql.SQLContext(sc)

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

sqlContext.sql("SELECT COUNT(\*) as gold\_medal,year FROM sport\_data WHERE medalType='gold' GROUP BY year").show()



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

sqlContext.sql("SELECT COUNT(\*) as silver\_medal, sport FROM sport\_data WHERE medalType='silver' and country = 'USA' group by sport").show()

