

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

```
scala> val result = spark.sql("select year,count(medal_type) from sports where medal_type='gold' group by year").show
+---+-----+
|year|count(medal_type)|
+---+-----+
|2015|                3|
|2014|                3|
|2016|                2|
|2017|                1|
+---+-----+

result: Unit = ()
```

2. 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

```
scala> val result = spark.sql("select country,sports,count(medal_type) from sports where medal_type='silver' and country='USA' group by country,sports").show
+-----+-----+-----+
|country| sports|count(medal_type)|
+-----+-----+-----+
|   USA|swimming|          3|
+-----+-----+-----+

result: Unit = ()

scala> |
```