Session 19 Assignment 1

Problem 1 – What are the total number of gold medals every year

Import the required package.

```
scala> import org.apache.spark.sql.SQLContext;
import org.apache.spark.sql.SQLContext

scala> val sqlContext=new SQLContext(sc);
warning: there was one deprecation warning; re-run with -deprecation for details
sqlContext: org.apache.spark.sql.SQLContext = org.apache.spark.sql.SQLContext@10939bc1

scala>
```

Read the input data file and convert it into a temp table.

val df=sqlContext.read.format("csv").option("header","true").load("/home/bigdata/deepak/docs/Acadgild/Session19Assignment1/Sports_data.txt").registerTempTable("sportsdata")

```
scala> val df=sqlContext.read.format("csv").option("header","true").load("/home/bigdata/deepak/docs/Acadgild/Session19Assignment1/Sports_data.txt").registerTem
pTable("sportsdata")
warning: there was one deprecation warning; re-run with -deprecation for details
df: Unit = ()

scala>
```

Now, execute the following SQL query to get the results.

val result=sqlContext.sql("select year,medal_type,count(*) from sportsdata where medal_type='gold' group by year,medal_type")

```
scala> val result=sqlContext.sql("select year,medal_type,count(*) from sportsdata where medal_type='gold' group by year,medal_type")
result: org.apache.spark.sql.DataFrame = [year: string, medal_type: string ... 1 more field]

scala>
```

Display the output

```
scala> val result=sqlContext.sql("select year,medal_type,count(*) from sportsdata where medal_type='gold' group by year,medal_type")
result: org.apache.spark.sql.DataFrame = [year: string, medal_type: string ... 1 more field]

scala> result.show
+----+----------+--------+
|year|medal_type|count(1)|
+----+----------+--------+
|2016|      gold|       2|
|2014|      gold|       3|
|2015|      gold|       3|
|2017|      gold|       1|
+----+----------+--------+
```

Problem 2 – What are the total number of silver medals have been won by USA in each sport?

Import the required package.

```
scala> import org.apache.spark.sql.SQLContext;
import org.apache.spark.sql.SQLContext

scala> val sqlContext=new SQLContext(sc);
warning: there was one deprecation warning; re-run with -deprecation for details
sqlContext: org.apache.spark.sql.SQLContext = org.apache.spark.sql.SQLContext@10939bc1

scala>
```

Read the input data file and convert it into a temp table.

val df=sqlContext.read.format("csv").option("header","true").load("/home/bigdata/deepak/docs/Acadgild/Session19Assignment1/Sports_data.txt").registerTempTable("sportsdata")

```
scala> val df=sqlContext.read.format("csv").option("header","true").load("/home/bigdata/deepak/docs/Acadgild/Session19Assignment1/Sports_data.txt").registerTem
pTable("sportsdata")
warning: there was one deprecation warning; re-run with -deprecation for details
df: Unit = ()

scala>
```

Execute the following query to get the results.

val result=sqlContext.sql("select sports,count(medal_type) from sportsdata where medal_type='silver' and country='USA' group by sports")

Display the results.

```
scala> val result=sqlContext.sql("select sports,count(medal_type) from sportsdata where medal_type='silver' and country='USA' group by sports")
result: org.apache.spark.sql.DataFrame = [sports: string, count(medal_type): bigint]

scala> result.show
+--------+-----------------+
|  sports|count(medal_type)|
+--------+-----------------+
|swimming|                3|
+--------+-----------------+
```