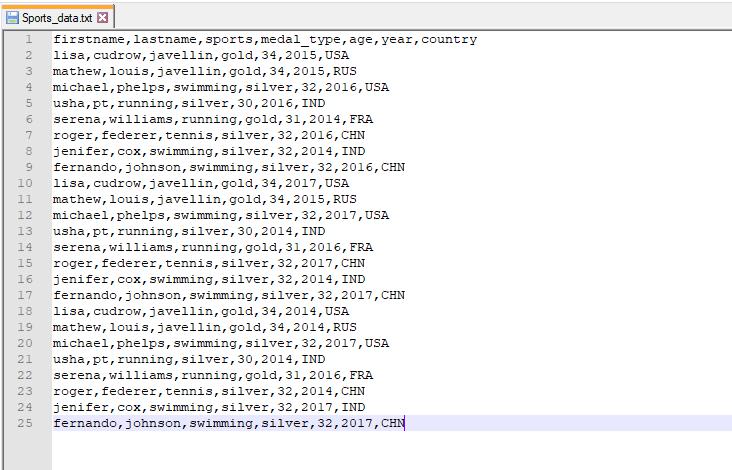
ASSIGNMENT 19.1

**Input Dataset:**

We have sports related data with us:

**Sports\_data**

firstname, lastname, sports, medal\_type, age, year, country



**Problem Statement:**

Using spark-SQL, find:

1. What are the total number of gold medal winners every year?
2. How many silver medals have been won by USA in each sport?

**Solution:**

1. Here is the Spark code snippet to find the total number of gold medal winners every year: // import required Spark packages

import org.apache.spark.sql.SparkSession

object Assignment18\_1 {

def main(args: Array[String]): Unit = {

* create a SparkSession object that can be used to create various contexts of Spark such as sqlContext val spark = SparkSession

.builder()

.config("spark.sql.warehouse.dir", "file:///c:/tmp/spark-warehouse")

.master("local[\*]")

.getOrCreate()

* initialize sqlContext

val sqlContext = spark.sqlContext

// load input data file – Sports\_data.txt

val sports\_DF = sqlContext.read.option("header", "true").csv("E:\\Acadgild\\Session 19\\ Sports\_data.txt")

//create a temporary view – Sports\_data

sports\_DF.createOrReplaceTempView("Sports\_data")

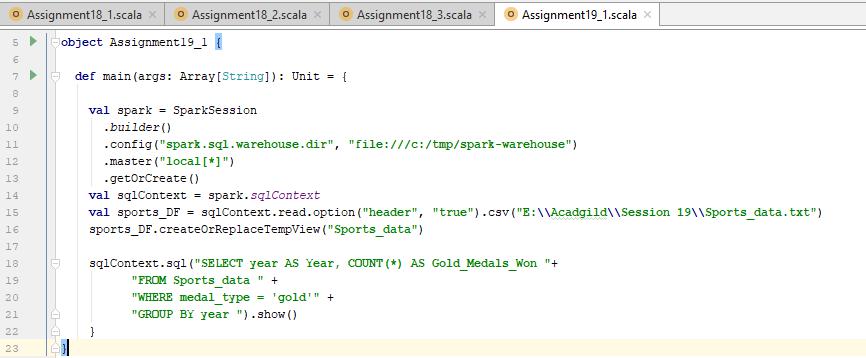
* SQL query to find the total number of gold medal winners every year sqlContext.sql("SELECT year AS Year, COUNT(\*) AS Gold\_Medals\_Won "+

"FROM Sports\_data " + "WHERE medal\_type = 'gold' " +

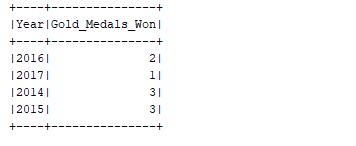
"GROUP BY year ").show()

}

}



**Output:**



1. Here is the Spark code to find the total number of silver medals won by USA in each sport: // import required Spark packages

import org.apache.spark.sql.SparkSession

object Assignment18\_1 {

def main(args: Array[String]): Unit = {

* create a SparkSession object that can be used to create various contexts of Spark such as sqlContext val spark = SparkSession

.builder()

.config("spark.sql.warehouse.dir", "file:///c:/tmp/spark-warehouse")

.master("local[\*]")

.getOrCreate()

// initialize sqlContext

val sqlContext = spark.sqlContext

// load input data file – Sports\_data.txt

val sports\_DF = sqlContext.read.option("header", "true").csv("E:\\Acadgild\\Session 19\\ Sports\_data.txt")

//create a temporary view – Sports\_data

sports\_DF.createOrReplaceTempView("Sports\_data")

* SQL query to find the total number of gold medal winners every year sqlContext.sql("SELECT sports AS Sport, COUNT(\*) AS Silver\_Medals\_Won\_By\_USA "+

"FROM Sports\_data " +

"WHERE medal\_type = 'silver' AND country = 'USA' " +

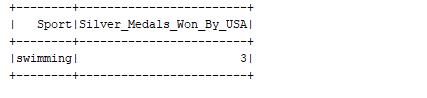
"GROUP BY sports ").show()

}

}



**Output:**



To summarize, USA has won **3** silver medals in swimming category.