1. Spark Session Object is created as below
2. ErrorLevel set using SetLevel(Level.ERROR) – imported log4j jar as well

```scala
package com.CaseStudyIV.HospData

import org.apache.spark.sql.SparkSession
import org.apache.log4j.{Level, Logger}
import org.apache.spark.sql.types._


object SparkHospData {

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

    //spark session object
    val spark = SparkSession
      .builder()
      .master( master = "local")
      .appName( name = "Spark SQL - IV")
      .getOrCreate()

    // Removing all INFO logs in consol printing only result sets
    val rootLogger = Logger.getRootLogger()
    rootLogger.setLevel(Level.ERROR)

    println("spark session object is created")
```

```
spark session object is created
  Hospital Schema is created
```

# Objective - 1

ACAD**GILD**

### Load file into spark

Before Loading .CSV File into Spark we need to define schema as below

```scala
val Manual_schema = new StructType(Array( new StructField( name = "DRGDefinition",StringType,
  nullable = true),
  new StructField( name = "ProviderId", LongType, nullable = false),
  new StructField( name = "ProviderName", StringType, nullable = true),
  new StructField( name = "ProviderStreetAddress", StringType, nullable = false),
  new StructField( name = "ProviderCity", StringType, nullable = false),
  new StructField( name = "ProviderState", StringType, nullable = false),
  new StructField( name = "ProviderZipCode", LongType, nullable = false),
  new StructField( name = "HospitalReferralRegionDescription", StringType, nullable = true),
  new StructField( name = "TotalDischarges", LongType, nullable = false),
  new StructField( name = "AverageCoveredCharges", DoubleType, nullable = false),
  new StructField( name = "AverageTotalPayments", DoubleType, nullable = false),
  new StructField( name = "AverageMedicarePayments", DoubleType, nullable = false)))

println(" Hospital Schema is created ")
```

```
spark session object is created
  Hospital Schema is created
```

## CASE STUDY - IV
## WORKIGN WITH HOSPITAL DATA

.CSV file is loaded as below into Hospital_data

```scala
val Hospital_data = spark.read.format( source = "csv")
  .option("header", "true")
  .schema(Manual_schema)
  .load( path = "C:\\Users\\Shruthi\\Downloads\\inpatientCharges.csv").toDF()
println("Hospitale Data top 20 rows ")
  Hospital_data.show()
```

o/p:-

```
Hospitale Data top 20 rows
+--------------------+----------+--------------------+--------------------+------------+-------------+--------------+------------------------------------+--------------+----------------------+--------
|       DRGDefinition|ProviderId|        ProviderName|ProviderStreetAddress|ProviderCity|ProviderState|ProviderZipCode|HospitalReferralRegionDescription|TotalDischarges|AverageCoveredCharges|Averag
+--------------------+----------+--------------------+--------------------+------------+-------------+--------------+------------------------------------+--------------+----------------------+--------
|039 - EXTRACRANIA...|     10001|SOUTHEAST ALABAMA...| 1108 ROSS CLARK C...|      DOTHAN|           AL|         36301|                          AL - Dothan|            91|             32963.07|
|039 - EXTRACRANIA...|     10005|MARSHALL MEDICAL ...| 2505 U S HIGHWAY ...|        BOAZ|           AL|         35957|                      AL - Birmingham|            14|             15131.85|
|039 - EXTRACRANIA...|     10006|ELIZA COFFEE MEMO...|   205 MARENGO STREET|    FLORENCE|           AL|         35631|                      AL - Birmingham|            24|             37560.37|
|039 - EXTRACRANIA...|     10011|   ST VINCENT'S EAST| 50 MEDICAL PARK E...|  BIRMINGHAM|           AL|         35235|                      AL - Birmingham|            25|             13998.28|
|039 - EXTRACRANIA...|     10016|SHELBY BAPTIST ME...| 1000 FIRST STREET...|   ALABASTER|           AL|         35007|                      AL - Birmingham|            18|             31633.27|
|039 - EXTRACRANIA...|     10023|BAPTIST MEDICAL C...| 2105 EAST SOUTH B...|  MONTGOMERY|           AL|         36116|                      AL - Montgomery|            67|             16920.79|
|039 - EXTRACRANIA...|     10029|EAST ALABAMA MEDI...| 2000 PEPPERELL PA...|     OPELIKA|           AL|         36801|                      AL - Birmingham|            51|             11977.13|
|039 - EXTRACRANIA...|     10033|UNIVERSITY OF ALA...| 619 SOUTH 19TH ST...|  BIRMINGHAM|           AL|         35233|                      AL - Birmingham|            32|             35841.09|
|039 - EXTRACRANIA...|     10039|  HUNTSVILLE HOSPITAL|      101 SIVLEY RD|  HUNTSVILLE|           AL|         35801|                      AL - Huntsville|           135|             28523.39|
|039 - EXTRACRANIA...|     10040|GADSDEN REGIONAL ...| 1007 GOODYEAR AVENUE|     GADSDEN|           AL|         35903|                      AL - Birmingham|            34|             75233.38|
|039 - EXTRACRANIA...|     10046|RIVERVIEW REGIONA...|  600 SOUTH THIRD S...|     GADSDEN|           AL|         35901|                      AL - Birmingham|            14|             67327.92|
|039 - EXTRACRANIA...|     10055|    FLOWERS HOSPITAL| 4370 WEST MAIN ST...|      DOTHAN|           AL|         36305|                          AL - Dothan|            45|             39607.28|
|039 - EXTRACRANIA...|     10056|ST VINCENT'S BIRM...| 810 ST VINCENT'S ...|  BIRMINGHAM|           AL|         35205|                      AL - Birmingham|            43|             22862.23|
|039 - EXTRACRANIA...|     10078|NORTHEAST ALABAMA...|    400 EAST 10TH STREET|    ANNISTON|           AL|         36207|                      AL - Birmingham|            21|             31110.85|
|039 - EXTRACRANIA...|     10083|SOUTH BALDWIN REG...| 1613 NORTH MCKENZ...|       FOLEY|           AL|         36535|                          AL - Mobile|            15|             25411.33|
|039 - EXTRACRANIA...|     10085|DECATUR GENERAL H...|   1201 7TH STREET SE|     DECATUR|           AL|         35609|                      AL - Huntsville|            27|              9234.51|
|039 - EXTRACRANIA...|     10090| PROVIDENCE HOSPITAL| 6801 AIRPORT BOUL...|      MOBILE|           AL|         36608|                          AL - Mobile|            27|             15895.85|
|039 - EXTRACRANIA...|     10092|D C H REGIONAL ME...| 809 UNIVERSITY BO...|  TUSCALOOSA|           AL|         35401|                      AL - Tuscaloosa|            31|             19721.16|
|039 - EXTRACRANIA...|     10100|     THOMAS HOSPITAL|   750 MORPHY AVENUE|    FAIRHOPE|           AL|         36532|                          AL - Mobile|            18|             10710.88|
|039 - EXTRACRANIA...|     10103|BAPTIST MEDICAL C...| 701 PRINCETON AVE...|  BIRMINGHAM|           AL|         35211|                      AL - Birmingham|            33|             51343.75|
+--------------------+----------+--------------------+--------------------+------------+-------------+--------------+------------------------------------+--------------+----------------------+--------
only showing top 20 rows
```

Created view out of Hospital_data using below registering in name Hospital_view

```scala
//Created View out of Hospital data to run Spark SQL queries
Hospital_data.createOrReplaceTempView( viewName = "Hospital_view")
```

# Objective - 2                                    ACAD**GILD**

➢ What is the average amount of **AverageCoveredCharges** per state

➢ find out the **AverageTotalPayments** charges per state

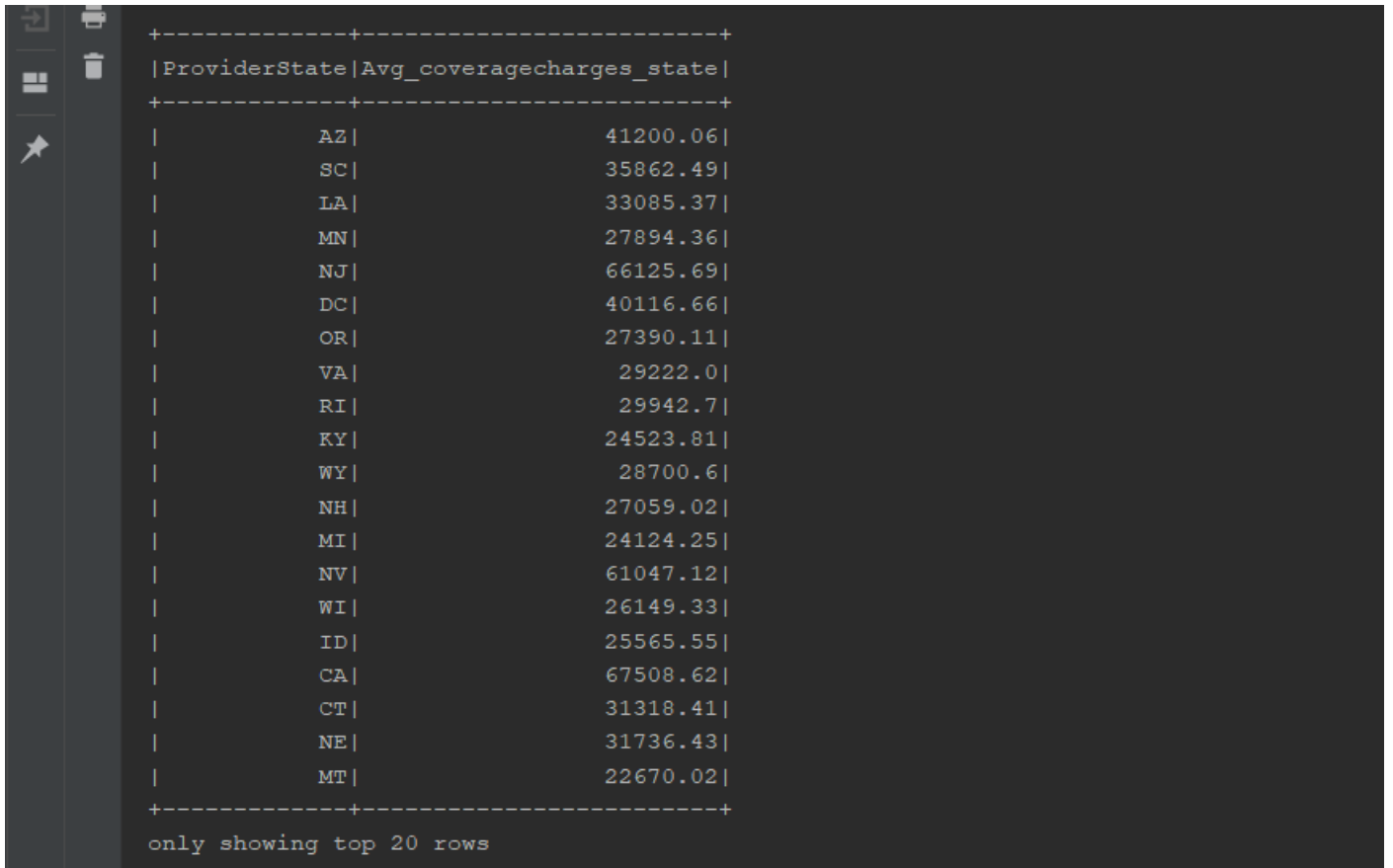➢ find out the **AverageMedicarePayments** charges per state.

Using Spark SQL query Average Covered Charges are processed from Hospital_view temporary view

## What is the average amount of AverageCoveredCharges per state

```scala
// What is the average amount of AverageCoveredCharges per state
spark.sql(
  """select ProviderState, round(avg(AverageCoveredCharges),2) as
    |Avg_coveragecharges_state from Hospital_view group by ProviderState""".stripMargin).show()
```

# CASE STUDY - IV
## WORKIGN WITH HOSPITAL DATA

O/P:-

```
+------------+----------------------+
|ProviderState|Avg_coveragecharges_state|
+------------+----------------------+
|          AZ|              41200.06|
|          SC|              35862.49|
|          LA|              33085.37|
|          MN|              27894.36|
|          NJ|              66125.69|
|          DC|              40116.66|
|          OR|              27390.11|
|          VA|               29222.0|
|          RI|               29942.7|
|          KY|              24523.81|
|          WY|               28700.6|
|          NH|              27059.02|
|          MI|              24124.25|
|          NV|              61047.12|
|          WI|              26149.33|
|          ID|              25565.55|
|          CA|              67508.62|
|          CT|              31318.41|
|          NE|              31736.43|
|          MT|              22670.02|
+------------+----------------------+
only showing top 20 rows
```

# Find out the AverageTotalPayments charges per state

Average Total payments using Spark SQL

```
//find out the AverageTotalPayments charges per state
spark.sql( sqlText = """select ProviderState, round(sum(cast(AverageTotalPayments as
        decimal)/cast(pow(10,2) as decimal)),2) as Avg_tot_payment_state from Hospital_view
        group by ProviderState""").show()
```

O/P:-

```
+-------------+--------------------+
|ProviderState|Avg_tot_payment_state|
+-------------+--------------------+
|           AZ|           289506.28|
|           SC|           260000.46|
|           LA|           261492.69|
|           MN|           224034.81|
|           NJ|           515368.48|
|           DC|            60051.02|
|           OR|           135566.46|
|           VA|           385017.99|
|           RI|            61796.34|
|           KY|           267315.91|
|           WY|            28154.28|
|           NH|            76454.02|
|           MI|           528592.49|
|           NV|           123706.69|
|           WI|           262732.20|
|           ID|            54147.78|
|           CA|          1649942.14|
|           CT|           228559.40|
|           NE|            99102.66|
|           MT|            46819.20|
+-------------+--------------------+
only showing top 20 rows
```

Average Total payments using Spark SQL

# find out the AverageMedicarePayments charges per state.

Using Spark SQL, Temporary View

By executing following code we weill get Average Medicare Payments

```
//find out the AverageMedicarePayments charges per state.
spark.sql( sqlText = """select ProviderState, round(sum(cast(AverageMedicarePayments as
          decimal)/cast(pow(10,2) as decimal)),2) as Avg_Medi_payment_state from Hospital_view
          group by ProviderState""").show()
```

O/P :-

```
+-------------+---------------------+
|ProviderState|Avg_Medi_payment_state|
+-------------+---------------------+
|           AZ|            251621.67|
|           SC|            224239.61|
|           LA|            223626.26|
|           MN|            194104.94|
|           NJ|            462666.61|
|           DC|             54571.33|
|           OR|            117368.35|
|           VA|            326583.58|
|           RI|             54789.50|
|           KY|            232011.54|
|           WY|             23562.37|
|           NH|             66864.97|
|           MI|            469403.22|
|           NV|            105146.54|
|           WI|            226794.06|
|           ID|             46625.52|
|           CA|           1501628.09|
|           CT|            203203.55|
|           NE|             84881.92|
|           MT|             40384.33|
+-------------+---------------------+
only showing top 20 rows
```

# Objective - 3

ACAD**GILD**

➢ Find out the total number of **Discharges** per state and for each disease

➢ Sort the output in descending order of **totalDischarges**

Find out the total number of **Discharges** per state and for each disease

Sort the output in descending order of **totalDischarges**

- Using below query we will able to find out Number of discharges per state
- Sort will be done by including order by Desc clause in SQL query

```
// Find out the total number of Discharges per state and for each disease
// Sort the output in descending order of totalDischarges
spark.sql( sqlText = """select DRGDefinition,ProviderState, sum(TotalDischarges) as
         Total_Discharges from Hospital_view group by DRGDefinition, ProviderState order by
         Total_Discharges desc """).show()
```

O/P:-

```
+--------------------+-------------+----------------+
|       DRGDefinition|ProviderState|Total_Discharges|
+--------------------+-------------+----------------+
|871 - SEPTICEMIA ...|           CA|           34284|
|470 - MAJOR JOINT...|           TX|           30095|
|470 - MAJOR JOINT...|           FL|           29985|
|470 - MAJOR JOINT...|           CA|           29731|
|871 - SEPTICEMIA ...|           TX|           23144|
|871 - SEPTICEMIA ...|           NY|           21970|
|392 - ESOPHAGITIS...|           FL|           21298|
|470 - MAJOR JOINT...|           IL|           20095|
|470 - MAJOR JOINT...|           NY|           19371|
|871 - SEPTICEMIA ...|           FL|           18660|
|690 - KIDNEY & UR...|           TX|           17384|
|392 - ESOPHAGITIS...|           NY|           17337|
|470 - MAJOR JOINT...|           MI|           16847|
|470 - MAJOR JOINT...|           PA|           16712|
|292 - HEART FAILU...|           FL|           16639|
|690 - KIDNEY & UR...|           FL|           16405|
|470 - MAJOR JOINT...|           OH|           16062|
|470 - MAJOR JOINT...|           NC|           15820|
|871 - SEPTICEMIA ...|           IL|           15610|
|871 - SEPTICEMIA ...|           MI|           15548|
+--------------------+-------------+----------------+
only showing top 20 rows


Process finished with exit code 0
```