CASE STUDY – III
WORKING WITH SENSOR DATA

Spark Session object is created, using log4j ErrorLevel is set to ERROR.

```
package com.CaseStudySensorData.Assignment

import org.apache.spark.sql.SparkSession
import org.apache.log4j.{Level, Logger}
import org.apache.spark.sql.types.{LongType, StringType, StructField, StructType}

object SparkSensorData {

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

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

To load csv file to spark we have to perform the following steps:

- Create a manual schema for both csv files which would provide the schema while Loading both CSV's as shown below

```
//Schema for HVAC.csv
val Manual_schema_HVAC = new StructType(Array(new StructField( name = "Date", StringType,  nullable = true),
  new StructField( name = "Time", StringType,  nullable = false),
  new StructField( name = "TargetTemp", LongType,  nullable = true),
  new StructField( name = "ActualTemp", LongType,  nullable = false),
  new StructField( name = "System", LongType,  nullable = false),
  new StructField( name = "SystemAge", LongType,  nullable = false),
  new StructField( name = "BuildingID", LongType,  nullable = false)))

//Schema for building.csv
val Manual_schema_Building = new StructType(Array(new StructField( name = "BuildingID", LongType,  nullable = true),
  new StructField( name = "BuildingMgr", StringType,  nullable = false),
  new StructField( name = "BuildingAge", LongType,  nullable = true),
  new StructField( name = "HVACproduct", StringType,  nullable = false),
  new StructField( name = "Country", StringType,  nullable = false)))
```

- Loading the CSV files from local file system to spark as shown below
- Registering temp table out of them

CASE STUDY – III
WORKING WITH SENSOR DATA

Objective- 1

ACADGILD

- Load HVAC.csv file into temporary table
- Add a new column, tempchange - set to 1, if there is a change of greater than +/-5 between actual and target temperature

Objective- 2

ACADGILD

Load building.csv file into temporary table

```
//Reading the HVAC csv file
val HVAC = spark.read.format( source = "CSV").option("header",true).schema(Manual_schema_HVAC)
  .load( path = "C:\\Users\\Shruthi\\Downloads\\HVAC.csv")

//Displaying the dataframe contents
HVAC.show()
//Registering the temporary table
HVAC.registerTempTable( tableName = "HVAC_table")
println("HVAC table registered!")

//Reading the building.csv file and creating the temp table
val buildings = spark.read.format( source = "CSV").option("header", true).schema(Manual_schema_Building)
  .load( path = "C:\\Users\\Shruthi\\Downloads\\building.csv")

buildings.show()
buildings.registerTempTable( tableName = "building_table")
println("buildings table registered!")
```

O/P:-

CASE STUDY – III
WORKING WITH SENSOR DATA

```
18/08/12 23:54:42 INFO BlockManager: Initialized BlockManager: BlockManagerId(driver, LAPTOP-UCU4FHT
spark session object is created
+-------+--------+----------+----------+------+---------+----------+
|   Date|    Time|TargetTemp|ActualTemp|System|SystemAge|BuildingID|
+-------+--------+----------+----------+------+---------+----------+
| 6/1/13| 0:00:01|        66|        58|    13|       20|         4|
| 6/2/13| 1:00:01|        69|        68|     3|       20|        17|
| 6/3/13| 2:00:01|        70|        73|    17|       20|        18|
| 6/4/13| 3:00:01|        67|        63|     2|       23|        15|
| 6/5/13| 4:00:01|        68|        74|    16|        9|         3|
| 6/6/13| 5:00:01|        67|        56|    13|       28|         4|
| 6/7/13| 6:00:01|        70|        58|    12|       24|         2|
| 6/8/13| 7:00:01|        70|        73|    20|       26|        16|
| 6/9/13| 8:00:01|        66|        69|    16|        9|         9|
|6/10/13| 9:00:01|        65|        57|     6|        5|        12|
|6/11/13|10:00:01|        67|        70|    10|       17|        15|
|6/12/13|11:00:01|        69|        62|     2|       11|         7|
|6/13/13|12:00:01|        69|        73|    14|        2|        15|
|6/14/13|13:00:01|        65|        61|     3|        2|         6|
|6/15/13|14:00:01|        67|        59|    19|       22|        20|
|6/16/13|15:00:01|        65|        56|    19|       11|         8|
|6/17/13|16:00:01|        67|        57|    15|        7|         6|
|6/18/13|17:00:01|        66|        57|    12|        5|        13|
|6/19/13|18:00:01|        69|        58|     8|       22|         4|
|6/20/13|19:00:01|        67|        55|    17|        5|         7|
+-------+--------+----------+----------+------+---------+----------+
only showing top 20 rows
```

CASE STUDY – III
WORKING WITH SENSOR DATA

```
HVAC table registered!
+----------+-----------+-----------+-----------+------------+
|BuildingID|BuildingMgr|BuildingAge|HVACproduct|     Country|
+----------+-----------+-----------+-----------+------------+
|         1|         M1|         25|     AC1000|         USA|
|         2|         M2|         27|     FN39TG|      France|
|         3|         M3|         28|     JDNS77|      Brazil|
|         4|         M4|         17|     GG1919|     Finland|
|         5|         M5|          3|    ACMAX22|   Hong Kong|
|         6|         M6|          9|     AC1000|   Singapore|
|         7|         M7|         13|     FN39TG|South Africa|
|         8|         M8|         25|     JDNS77|   Australia|
|         9|         M9|         11|     GG1919|      Mexico|
|        10|        M10|         23|    ACMAX22|       China|
|        11|        M11|         14|     AC1000|     Belgium|
|        12|        M12|         26|     FN39TG|     Finland|
|        13|        M13|         25|     JDNS77|Saudi Arabia|
|        14|        M14|         17|     GG1919|     Germany|
|        15|        M15|         19|    ACMAX22|      Israel|
|        16|        M16|         23|     AC1000|      Turkey|
|        17|        M17|         11|     FN39TG|       Egypt|
|        18|        M18|         25|     JDNS77|   Indonesia|
|        19|        M19|         14|     GG1919|      Canada|
|        20|        M20|         19|    ACMAX22|   Argentina|
+----------+-----------+-----------+-----------+------------+

buildings table registered!
```

- Now we will add a new column, tempchange - set to 1, if there is a change of greater than +/-5 between actual and target temperature

```
//Approach : Using Spark SQL
val filterHVAC = spark.sql( sqlText = """select *, IF((TargetTemp-ActualTemp)> 5 ,'1',
    IF((TargetTemp-ActualTemp)< -5 ,'1',0)) as Temp_change_diff from HVAC_table""")
    filterHVAC.show()
```

Output:-

CASE STUDY – III
WORKING WITH SENSOR DATA

```
+-------+--------+----------+----------+------+---------+----------+----------------+
|   Date|    Time|TargetTemp|ActualTemp|System|SystemAge|BuildingID|Temp_change_diff|
+-------+--------+----------+----------+------+---------+----------+----------------+
| 6/1/13| 0:00:01|        66|        58|    13|       20|         4|               1|
| 6/2/13| 1:00:01|        69|        68|     3|       20|        17|               0|
| 6/3/13| 2:00:01|        70|        73|    17|       20|        18|               0|
| 6/4/13| 3:00:01|        67|        63|     2|       23|        15|               0|
| 6/5/13| 4:00:01|        68|        74|    16|        9|         3|               1|
| 6/6/13| 5:00:01|        67|        56|    13|       28|         4|               1|
| 6/7/13| 6:00:01|        70|        58|    12|       24|         2|               1|
| 6/8/13| 7:00:01|        70|        73|    20|       26|        16|               0|
| 6/9/13| 8:00:01|        66|        69|    16|        9|         9|               0|
|6/10/13| 9:00:01|        65|        57|     6|        5|        12|               1|
|6/11/13|10:00:01|        67|        70|    10|       17|        15|               0|
|6/12/13|11:00:01|        69|        62|     2|       11|         7|               1|
|6/13/13|12:00:01|        69|        73|    14|        2|        15|               0|
|6/14/13|13:00:01|        65|        61|     3|        2|         6|               0|
|6/15/13|14:00:01|        67|        59|    19|       22|        20|               1|
|6/16/13|15:00:01|        65|        56|    19|       11|         8|               1|
|6/17/13|16:00:01|        67|        57|    15|        7|         6|               1|
|6/18/13|17:00:01|        66|        57|    12|        5|        13|               1|
|6/19/13|18:00:01|        69|        58|     8|       22|         4|               1|
|6/20/13|19:00:01|        67|        55|    17|        5|         7|               1|
+-------+--------+----------+----------+------+---------+----------+----------------+
only showing top 20 rows
```

Objective - 3

ACADGILD

Figure out the number of times, temperature has changed by 5 degrees or more for each country:

- Join both the tables.
- Select tempchange and country column
- Filter the rows where tempchange is 1 and count the number of occurrence for each country

- We will Join both the tables based on BuildingID, i.e. we will join HVAC and Buildings tables and register the output of join to a temptable called “HVACJBUILD”
- select Temp_hange_diff and country column, then filter the rows where Temp_hange_diff is 1 and count the number of occurrence for each country

```
val joinExpression = filterHVAC.col( colName = "BuildingID") === buildings.toDF().col( colName = "BuildingID")
val HVACJOBUILD = filterHVAC.join(buildings,joinExpression)
    HVACJOBUILD.show()
    HVACJOBUILD.registerTempTable( tableName = "HVACJBUILD")

val selective = spark.sql( sqlText = """select Temp_change_diff, Country from HVACJBUILD WHERE Temp_change_diff = 1""").toDF()
selective.registerTempTable( tableName = "newselective")
spark.sql( sqlText = """select Country, count(Temp_change_diff) from newselective group by Country""").show()
```

CASE STUDY – III
WORKING WITH SENSOR DATA

Output1:-

Joined table:

Date	Time	TargetTemp	ActualTemp	System	SystemAge	BuildingID	Temp_change_diff	BuildingID	BuildingMgr	BuildingAge	HVACproduct	Country
6/1/13	0:00:01	66	58	13	20	4	1	4	M4	17	GG1919	Finland
6/2/13	1:00:01	69	68	3	20	17	0	17	M17	11	FN39TG	Egypt
6/3/13	2:00:01	70	73	17	20	18	0	18	M18	25	JDNS77	Indonesia
6/4/13	3:00:01	67	63	2	23	15	0	15	M15	19	ACMAX22	Israel
6/5/13	4:00:01	68	74	16	9	3	1	3	M3	28	JDNS77	Brazil
6/6/13	5:00:01	67	56	13	28	4	1	4	M4	17	GG1919	Finland
6/7/13	6:00:01	70	58	12	24	2	1	2	M2	27	FN39TG	France
6/8/13	7:00:01	70	73	20	26	16	0	16	M16	23	AC1000	Turkey
6/9/13	8:00:01	66	69	16	9	9	0	9	M9	11	GG1919	Mexico
6/10/13	9:00:01	65	57	6	5	12	1	12	M12	26	FN39TG	Finland
6/11/13	10:00:01	67	70	10	17	15	0	15	M15	19	ACMAX22	Israel
6/12/13	11:00:01	69	62	2	11	7	1	7	M7	13	FN39TG	South Africa
6/13/13	12:00:01	69	73	14	2	15	0	15	M15	19	ACMAX22	Israel
6/14/13	13:00:01	65	61	3	2	6	0	6	M6	9	AC1000	Singapore
6/15/13	14:00:01	67	59	19	22	20	1	20	M20	19	ACMAX22	Argentina
6/16/13	15:00:01	65	56	19	11	8	1	8	M8	25	JDNS77	Australia
6/17/13	16:00:01	67	57	15	7	6	1	6	M6	9	AC1000	Singapore
6/18/13	17:00:01	66	57	12	5	13	1	13	M13	25	JDNS77	Saudi Arabia
6/19/13	18:00:01	69	58	8	22	4	1	4	M4	17	GG1919	Finland
6/20/13	19:00:01	67	55	17	5	7	1	7	M7	13	FN39TG	South Africa

only showing top 20 rows

Output2:-

Count of number of occurrence for each country, where temp_diff is “1”.

Country	count(Temp_change_diff)
Singapore	230
Turkey	243
Germany	196
France	251
Argentina	230
Belgium	199
Finland	473
China	241
Hong Kong	248
Israel	232
USA	213
Mexico	228
Indonesia	243
Saudi Arabia	233
Canada	232
Brazil	226
Australia	225
Egypt	236
South Africa	237

Process finished with exit code 0