Working with Sensor Data

Obj-1

Load HVAC.csv file into temporary table

```
import org.apache.spark.sql.SparkSession

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

    val sparkSession = SparkSession.builder.master( master = "local")
      .appName( name = "spark session example")
      .getOrCreate()

    val hvacDF = sparkSession.read.format( source = "csv").option("header", "true").option("inferSchema", "true")
      .load( path = "E:\\HVAC.csv")
    hvacDF.show( numRows = 10)
```

output

```
18/08/08 19:55:22 INFO DAGScheduler: ResultStage 2 (show at SensorCaseStudy.scala:12) finished in 0
18/08/08 19:55:22 INFO DAGScheduler: Job 2 finished: show at SensorCaseStudy.scala:12, took 0.26485
18/08/08 19:55:22 INFO TaskSchedulerImpl: Removed TaskSet 2.0, whose tasks have all completed, from
+-------+-------+----------+----------+------+---------+----------+
|   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|
+-------+-------+----------+----------+------+---------+----------+
only showing top 10 rows

18/08/08 19:55:23 INFO FileSourceStrategy: Pruning directories with:
```

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

```
import sparkSession.implicits._

hvacDF.createOrReplaceTempView( viewName = "HVAC_Data")
val newhvacDF = hvacDF.select( cols = $"Date", $"Time", $"TargetTemp".cast( to = "Int"), $"ActualTemp".cast( to = "Int"), $"Syst
val newcolhvacDF = sparkSession.sql( sqlText = "select *,IF((targettemp - actualtemp) > 5, '1', IF" + "((targettemp - actualtem
newcolhvacDF.createOrReplaceTempView( viewName = "newcolhvacDF")
newcolhvacDF.show()
```

OUTPUT

```
18/08/08 19:55:23 INFO DAGScheduler: ResultStage 3 (show at SensorCaseStudy.scala:19)
18/08/08 19:55:23 INFO DAGScheduler: Job 3 finished: show at SensorCaseStudy.scala:19
+-------+--------+----------+----------+------+---------+----------+----------+
|   Date|    Time|TargetTemp|ActualTemp|System|SystemAge|BuildingID|tempchange|
+-------+--------+----------+----------+------+---------+----------+----------+
| 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|
+-------+--------+----------+----------+------+---------+----------+----------+
```

OBJ-2

Load building.csv file into temporary table

```
  val buildingData = sparkSession.read.format( source = "csv").option("header", "true").option("inferSchema", "true").load( path
  buildingData.createOrReplaceTempView( viewName = "Building_Data")
  buildingData.show( numRows = 10)

  val joinedDF = newcolhvacDF.as( alias = "HD").join(buildingData.as( alias = "BD"),
    joinExprs = $"BD.BuildingID" === $"HD.BuildingID").filter( condition = $"tempchange" === 1).groupBy( col1 = "Country").count().
  }
}
```

```
18/08/08 19:55:24 INFO DAGScheduler: ResultStage 6 (show at SensorCaseStudy.s
18/08/08 19:55:24 INFO DAGScheduler: Job 6 finished: show at SensorCaseStudy.
+----------+-----------+-----------+-----------+------------+
|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|
+----------+-----------+-----------+-----------+------------+
only showing top 10 rows

18/08/08 19:55:24 INFO ContextCleaner: Cleaned accumulator 99
18/08/08 19:55:24 INFO ContextCleaner: Cleaned accumulator 136
```

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

OUTPut

```
18/08/08 19:55:28 INFO DAGScheduler: Job 12 finished: show at Sens
+------------+-----+
|     Country|count|
+------------+-----+
|   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|
+------------+-----+

18/08/08 19:55:28 INFO SparkContext: Invoking stop() from shutdown
```

Whole code

Code for CASEstudy 3

```
import org.apache.spark.sql.SparkSession

object SensorCaseStudy {
  def main(args: Array[String]) {

    val sparkSession = SparkSession.builder.master("local")
      .appName("spark session example")
      .getOrCreate()

    val hvacDF = sparkSession.read.format("csv").option("header",
"true").option("inferSchema", "true")
      .load("E:\\HVAC.csv")
    hvacDF.show(10)
    import sparkSession.implicits._

    hvacDF.createOrReplaceTempView("HVAC_Data")
    val newhvacDF = hvacDF.select($"Date", $"Time", $"TargetTemp".cast("Int"),
$"ActualTemp".cast("Int"), $"System".cast("Int"), $"SystemAge".cast("Int"),
$"BuildingID")
    val newcolhvacDF = sparkSession.sql("select *,IF((targettemp - actualtemp) > 5,
'1', IF" + "((targettemp - actualtemp) < -5, '1', 0)) AS tempchange from
HVAC_Data").toDF()
```

```
    newcolhvacDF.createOrReplaceTempView("newcolhvacDF")
    newcolhvacDF.show()

    val buildingData = sparkSession.read.format("csv").option("header",
"true").option("inferSchema", "true").load("E:\\building.csv").toDF()
    buildingData.createOrReplaceTempView("Building_Data")
    buildingData.show(10)

    val joinedDF = newcolhvacDF.as("HD").join(buildingData.as("BD"),
      $"BD.BuildingID" === $"HD.BuildingID").filter($"tempchange" ===
1).groupBy("Country").count().show()
  }
}
```