

Session 21 - Case Study 3 Sensor

Task 1:

- 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.

Code:

```
package SQL

import org.apache.spark.sql.SparkSession

object SparkSQLUseCase1 {
  case class
  hvac_cls(Date:String,Time:String,TargetTemp:Int,ActualTemp:Int,System:Int,SystemAge:Int,B
uildingId:Int)
  case class
  building(buildid:Int,buildmgr:String,buildAge:Int,hvacproduct:String,Country:String)

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

    println("hey scala")

    val spark = SparkSession
      .builder()
      .master(master= "local")
      .appName( name= "Spark SQL Use Case 1")
      .config("spark.some.config.option", "some-value")
      .getOrCreate()

    println("Spark Session Object Created")

    spark.sparkContext.setLogLevel("WARN")

    val data = spark.sparkContext.textFile("C:/Users/Gaurav/Desktop/HVAC.csv")
    println("HVAC Data -->" + data.count())

    val header = data.first()
    val data1 = data.filter(row => row != header)
    println("Header removed from the data !!")

    import spark.implicits._

    val hvac = data1.map(x=>x.split(",")).map(x =>
    hvac_cls(x(0),x(1),x(2).toInt,x(3).toInt,x(4).toInt,x(5).toInt,x(6).toInt)).toDF()
    hvac.show()
    println("HVAC Dataframe created !!")
```

```

hvac.registerTempTable("HVAC")
println("Dataframe registered as table !!")

val hvac1 = spark.sql("select *,IF((targettemp - actualtemp) > 5, '1', IF((targettemp -
actualtemp) < -5, '1', 0)) AS tempchange from HVAC")
hvac1.registerTempTable("HVAC1")
println("Dataframe registered as HVAC1 table !!")

hvac.registerTempTable("HVAC")
println("Dataframe registered as table !!")

```

Output:

```

ScalaSparkProject [C:\Users\Gaurav\IdeaProjects\ScalaSparkProject] - ...\src\main\scala\SQL\SparkSQLUseCase1.scala [ScalaSparkProject] - IntelliJ IDEA
File Edit View Navigate Code Analyze Refactor Build Run Tools VCS Window Help
ScalaSparkProject / src / main / scala / SQL / SparkSQLUseCase1.scala
Run: SparkSQLUseCase1 x
"C:\Program Files\Java\jdk1.8.0_101\bin\java.exe" ...
hey scala
Using Spark's default log4j profile: org/apache/spark/log4j-defaults.properties
19/01/07 22:35:11 INFO SparkContext: Running Spark version 2.1.0
19/01/07 22:35:12 INFO SecurityManager: Changing view acls to: Gaurav
19/01/07 22:35:12 INFO SecurityManager: Changing modify acls to: Gaurav
19/01/07 22:35:12 INFO SecurityManager: Changing view acls groups to:
19/01/07 22:35:12 INFO SecurityManager: Changing modify acls groups to:
19/01/07 22:35:12 INFO SecurityManager: SecurityManager: authentication disabled; ui acls disabled; users with view permissions: Set(Gaurav); groups with view permissions: Set()
19/01/07 22:35:14 INFO Utils: Successfully started service 'sparkDriver' on port 51503.
19/01/07 22:35:14 INFO SparkEnv: Registering MapOutputTracker
19/01/07 22:35:14 INFO SparkEnv: Registering BlockManagerMaster
19/01/07 22:35:14 INFO BlockManagerMasterEndpoint: Using org.apache.spark.storage.DefaultTopologyMapper for getting topology information
19/01/07 22:35:14 INFO BlockManagerMasterEndpoint: BlockManagerMasterEndpoint up
19/01/07 22:35:14 INFO DiskBlockManager: Created local directory at C:\Users\Gaurav\AppData\Local\Temp\blockmgr-6ed75306-0c16-4b70-86a3-ala3f5f15613
19/01/07 22:35:14 INFO MemoryStore: MemoryStore started with capacity 1170.6 MB
19/01/07 22:35:14 INFO SparkEnv: Registering OutputCommitCoordinator
19/01/07 22:35:15 WARN Utils: Service 'SparkUI' could not bind on port 4040. Attempting port 4041.
19/01/07 22:35:15 INFO Utils: Successfully started service 'SparkUI' on port 4041.
19/01/07 22:35:15 INFO SparkUI: Bound SparkUI to 0.0.0.0, and started at http://192.168.213.1:4041
19/01/07 22:35:15 INFO Executor: Starting executor ID driver on host localhost
19/01/07 22:35:15 INFO Utils: Successfully started service 'org.apache.spark.network.netty.NettyBlockTransferService' on port 51512.
19/01/07 22:35:15 INFO NettyBlockTransferService: Server created on 192.168.213.1:51512
19/01/07 22:35:15 INFO BlockManager: Using org.apache.spark.storage.RandomBlockReplicationPolicy for block replication policy
19/01/07 22:35:15 INFO BlockManagerMaster: Registering BlockManager BlockManagerId(driver, 192.168.213.1, 51512, None)
19/01/07 22:35:15 INFO BlockManagerMasterEndpoint: Registering block manager 192.168.213.1:51512 with 1170.6 MB RAM, BlockManagerId(driver, 192.168.213.1, 51512, None)
19/01/07 22:35:15 INFO BlockManagerMaster: Registered BlockManager BlockManagerId(driver, 192.168.213.1, 51512, None)
19/01/07 22:35:15 INFO BlockManager: Initialized BlockManager: BlockManagerId(driver, 192.168.213.1, 51512, None)
19/01/07 22:35:15 INFO WarehouseState: Warehouse path is 'file:/C:/Users/Gaurav/IdeaProjects/ScalaSparkProject/spark-warehouse/'.
Spark Session Object Created
HVAC Data -->>>8001
Header removed from the data !!
+-----+-----+-----+-----+-----+-----+
| Date | Time | TargetTemp | ActualTemp | System | SystemAge | BuildingId |
+-----+-----+-----+-----+-----+-----+

```

```

ScalaSparkProject [C:\Users\Gaurav\IdeaProjects\ScalaSparkProject] - ...\src\main\scala\SQL\SparkSQLUseCase1.scala [ScalaSparkProject] - IntelliJ IDEA
File Edit View Navigate Code Analyze Refactor Build Run Tools VCS Window Help
ScalaSparkProject / src / main / scala / SQL / SparkSQLUseCase1.scala
Run: SparkSQLUseCase1 x
Spark Session Object Created
HVAC Data -->>>8001
Header removed from the data !!
+-----+-----+-----+-----+-----+-----+
| 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
HVAC Dataframe created !!
Dataframe registered as table !!
Dataframe registered as HVAC1 table !!
Dataframe registered as table !!

```

Task 2: Load building.csv file into temporary table

Code:

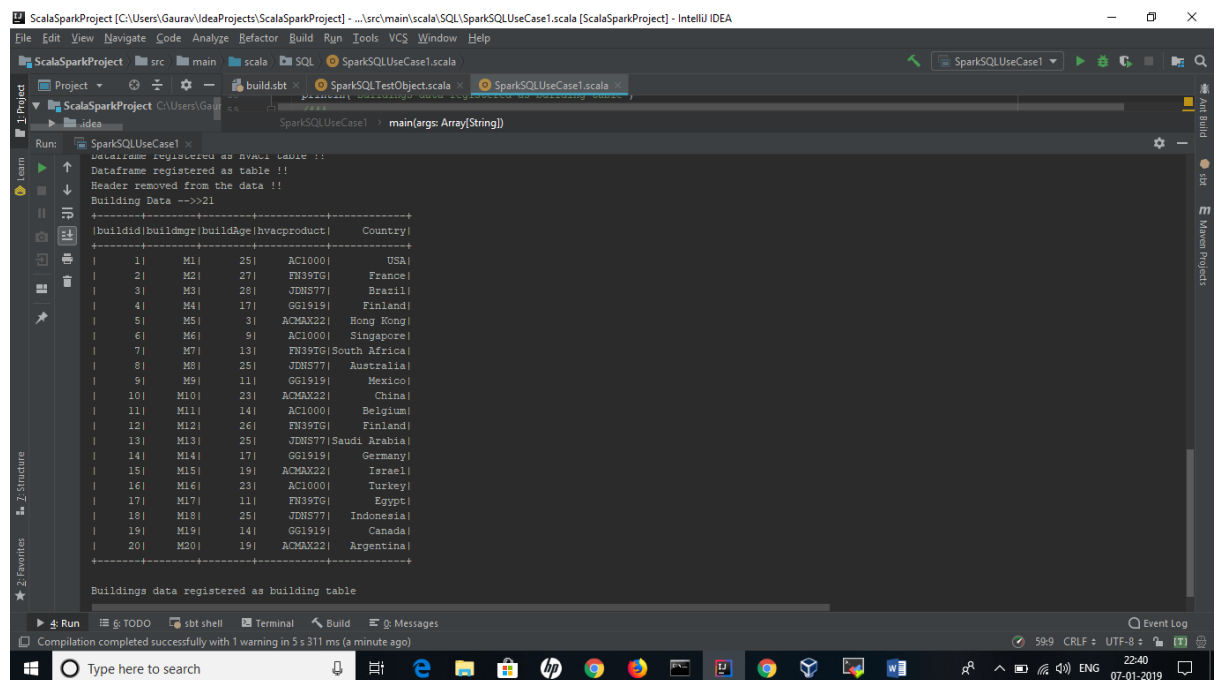
```
val data2 = spark.sparkContext.textFile("C:/Users/Gaurav/Desktop/building.csv")
val header1 = data2.first()
val data3 = data2.filter(row => row != header1)

println("Header removed from the data !!")
println("Building Data -->" + data2.count())

val build = data3.map(x => x.split(",")).map(x =>
  building(x(0).toInt, x(1), x(2).toInt, x(3), x(4))).toDF
build.show()
build.registerTempTable("building")

println("Buildings data registered as building table")
```

Output:



```
ScalaSparkProject [C:\Users\Gaurav\IdeaProjects\ScalaSparkProject] - ...\src\main\scala\SQL\SparkSQLUseCase1.scala [ScalaSparkProject] - IntelliJ IDEA
File Edit View Navigate Code Analyze Refactor Build Run Tools VCS Window Help
ScalaSparkProject | src | main | scala | SQL | SparkSQLUseCase1.scala | SparkSQLUseCase1.scala
Project | build.sbt | SparkSQLTestObject.scala | SparkSQLUseCase1.scala
ScalaSparkProject C:\Users\Gaurav\IdeaProjects\ScalaSparkProject
SparkSQLUseCase1 | main(args: Array[String])
Run: SparkSQLUseCase1
Dataframe registered as table !!
Header removed from the data !!
Building Data -->21
+-----+-----+-----+-----+-----+
|buildid|buildmgr|buildAge|hvacproduct|Country|
+-----+-----+-----+-----+-----+
| 1 | M1 | 25 | AC1000 | USA |
| 2 | M2 | 27 | FN39TG | France |
| 3 | M3 | 28 | JDN577 | Brazil |
| 4 | M4 | 17 | GG1919 | Finland |
| 5 | M5 | 31 | ACMA22 | Hong Kong |
| 6 | M6 | 9 | AC1000 | Singapore |
| 7 | M7 | 13 | FN39TG | South Africa |
| 8 | M8 | 25 | JDN577 | Australia |
| 9 | M9 | 11 | GG1919 | Mexico |
| 10 | M10 | 23 | ACMA22 | China |
| 11 | M11 | 14 | AC1000 | Belgium |
| 12 | M12 | 26 | FN39TG | Finland |
| 13 | M13 | 25 | JDN577 | Saudi Arabia |
| 14 | M14 | 17 | GG1919 | Germany |
| 15 | M15 | 19 | ACMA22 | Israel |
| 16 | M16 | 23 | AC1000 | Turkey |
| 17 | M17 | 11 | FN39TG | Egypt |
| 18 | M18 | 25 | JDN577 | Indonesia |
| 19 | M19 | 14 | GG1919 | Canada |
| 20 | M20 | 19 | ACMA22 | Argentina |
+-----+-----+-----+-----+-----+
Buildings data registered as building table
Run | Run | sbt shell | Terminal | Build | Messages
Compilation completed successfully with 1 warning in 5 s 311 ms (a minute ago)
Type here to search
```

Task 3: 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.

Code:

```

val build1 = spark.sql("select h.*, b.country, b.hvacproduct from building b join hvac1 h on
buildid = buildingid")
build1.show()

val tempCountry = build1.map(x => (new Integer(x(7).toString),x(8).toString))
tempCountry.show()

val tempCountryOnes = tempCountry.filter(x=> {if(x._1==1) true else false})
tempCountryOnes.show()

tempCountryOnes.groupBy("_2").count.show

```

Output:

```

ScalaSparkProject [C:\Users\Gaurav\IdeaProjects\ScalaSparkProject] - ...src\main\scala\SQL\SparkSQLUseCase1.scala [ScalaSparkProject] - IntelliJ IDEA
File Edit View Navigate Code Analyze Refactor Build Run Tools VCS Window Help
ScalaSparkProject src main scala SQL SparkSQLUseCase1.scala
Project Run SparkSQLUseCase1
Build.sbt SparkSQLTestObject.scala SparkSQLUseCase1.scala
Buildings data registered as building table
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
| Date | Time | TargetTemp | ActualTemp | System | SystemAge | BuildingId | tempchange | country | hvacproduct |
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
| 6/10/13 | 9:00:01 | 65 | 57 | 6 | 5 | 12 | 1 | Finland | FN39TG |
| 6/18/13 | 23:13:19 | 66 | 75 | 1 | 13 | 12 | 1 | Finland | FN39TG |
| 6/2/13 | 13:43:51 | 65 | 72 | 20 | 26 | 12 | 1 | Finland | FN39TG |
| 6/13/13 | 0:13:20 | 67 | 77 | 8 | 19 | 12 | 1 | Finland | FN39TG |
| 6/16/13 | 3:13:20 | 67 | 55 | 11 | 16 | 12 | 1 | Finland | FN39TG |
| 6/30/13 | 17:13:20 | 65 | 57 | 17 | 9 | 12 | 1 | Finland | FN39TG |
| 6/1/13 | 18:13:20 | 68 | 65 | 7 | 21 | 12 | 0 | Finland | FN39TG |
| 6/25/13 | 18:33:07 | 70 | 66 | 20 | 20 | 12 | 0 | Finland | FN39TG |
| 6/17/13 | 16:00:01 | 69 | 68 | 16 | 4 | 12 | 0 | Finland | FN39TG |
| 6/5/13 | 16:43:51 | 69 | 69 | 19 | 15 | 12 | 0 | Finland | FN39TG |
| 6/23/13 | 10:13:20 | 65 | 61 | 1 | 1 | 12 | 0 | Finland | FN39TG |
| 6/29/13 | 16:13:20 | 67 | 80 | 12 | 8 | 12 | 1 | Finland | FN39TG |
| 6/4/13 | 21:13:20 | 66 | 72 | 7 | 1 | 12 | 1 | Finland | FN39TG |
| 6/3/13 | 2:00:01 | 69 | 72 | 7 | 21 | 12 | 0 | Finland | FN39TG |
| 6/16/13 | 15:00:01 | 67 | 77 | 4 | 22 | 12 | 1 | Finland | FN39TG |
| 6/22/13 | 21:00:01 | 70 | 77 | 13 | 12 | 12 | 1 | Finland | FN39TG |
| 6/26/13 | 7:43:51 | 65 | 62 | 6 | 6 | 12 | 0 | Finland | FN39TG |
| 6/26/13 | 13:13:20 | 65 | 63 | 20 | 9 | 12 | 0 | Finland | FN39TG |
| 6/30/13 | 17:13:20 | 66 | 62 | 14 | 26 | 12 | 0 | Finland | FN39TG |
| 6/10/13 | 3:33:07 | 70 | 78 | 5 | 9 | 12 | 1 | Finland | FN39TG |
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
only showing top 20 rows
+-----+-----+
| _1 | _2 |
+-----+-----+
| 1 | Finland |
+-----+-----+
Run TODO sbt shell Terminal Build Messages
Compilation completed successfully with 1 warning in 5 s 25 ms (a minute ago)
Type here to search
109/100 CRLF UTF-8 ENG 22:44 07-01-2019

```

ScalaSparkProject [C:\Users\Gaurav\IdeaProjects\ScalaSparkProject] - ...src\main\scala\SQL\SparkSQLUseCase1.scala [ScalaSparkProject] - IntelliJ IDEA

File Edit View Navigate Code Analyze Refactor Build Run Tools VCS Window Help

ScalaSparkProject | src | main | scala | SQL | SparkSQLUseCase1.scala

Project | SparkSQLUseCase1.scala | SparkSQLTestObject.scala | SparkSQLUseCase1.scala

Run: SparkSQLUseCase1 x

```
-----+
|_1|_2|
-----+
|1|Finland|
|1|Finland|
|1|Finland|
|1|Finland|
|1|Finland|
|1|Finland|
|0|Finland|
|0|Finland|
|0|Finland|
|0|Finland|
|1|Finland|
|1|Finland|
|1|Finland|
|0|Finland|
|0|Finland|
|0|Finland|
|1|Finland|
|1|Finland|
|0|Finland|
|0|Finland|
|1|Finland|
-----+
only showing top 20 rows
-----+
|_1|_2|
-----+
|1|Finland|
|1|Finland|
|1|Finland|
|1|Finland|
|1|Finland|
|1|Finland|
|0|Finland|
|0|Finland|
|0|Finland|
|1|Finland|
|1|Finland|
|0|Finland|
|0|Finland|
|1|Finland|
|1|Finland|
|0|Finland|
|0|Finland|
|1|Finland|
-----+
```

Run | TODO | sbt shell | Terminal | Build | Messages

Compilation completed successfully with 1 warning in 5 s 25 ms (a minute ago)

Type here to search

109:100 CRLF UTF-8 ENG 22:44 07-01-2019

ScalaSparkProject [C:\Users\Gaurav\IdeaProjects\ScalaSparkProject] - ...src\main\scala\SQL\SparkSQLUseCase1.scala [ScalaSparkProject] - IntelliJ IDEA

File Edit View Navigate Code Analyze Refactor Build Run Tools VCS Window Help

ScalaSparkProject | src | main | scala | SQL | SparkSQLUseCase1.scala

Project | SparkSQLUseCase1.scala | SparkSQLTestObject.scala | SparkSQLUseCase1.scala

Run: SparkSQLUseCase1 x

```
-----+
|_1|_2|
-----+
|1|Finland|
|1|Finland|
|1|Finland|
|1|Finland|
|1|Finland|
|1|Finland|
|0|Finland|
|0|Finland|
|0|Finland|
|0|Finland|
|1|Finland|
|1|Finland|
|1|Finland|
|0|Finland|
|0|Finland|
|0|Finland|
|1|Finland|
|1|Finland|
|0|Finland|
|0|Finland|
|1|Finland|
-----+
only showing top 20 rows
-----+
|_1|_2|
-----+
|1|Finland|
|1|Finland|
|1|Finland|
|1|Finland|
|1|Finland|
|1|Finland|
|0|Finland|
|0|Finland|
|0|Finland|
|1|Finland|
|1|Finland|
|0|Finland|
|0|Finland|
|1|Finland|
|1|Finland|
|0|Finland|
|0|Finland|
|1|Finland|
-----+
```

Run | TODO | sbt shell | Terminal | Build | Messages

Compilation completed successfully with 1 warning in 5 s 25 ms (a minute ago)

Type here to search

109:100 CRLF UTF-8 ENG 22:44 07-01-2019

ScalaSparkProject [C:\Users\Gaurav\IdeaProjects\ScalaSparkProject] - ...src\main\scala\SQL\SparkSQLUseCase1.scala [ScalaSparkProject] - IntelliJ IDEA

File Edit View Navigate Code Analyze Refactor Build Run Tools VCS Window Help

ScalaSparkProject src main scala SQL SparkSQLUseCase1.scala

Project SparkSQLUseCase1 x build.sbt x SparkSQLTestObject.scala x SparkSQLUseCase1 x

Run: SparkSQLUseCase1 x

```
1 | 1|Finland|
2 | 1|Finland|
+-----+
only showing top 20 rows
+-----+
1 | _2|count|
+-----+
1 | Singapore| 230|
1 | Turkey| 243|
1 | Germany| 196|
1 | France| 251|
1 | Argentina| 230|
1 | Belgium| 199|
1 | Finland| 473|
1 | China| 241|
1 | Hong Kong| 248|
1 | Israel| 232|
1 | USA| 213|
1 | Mexico| 228|
1 | Indonesia| 243|
1 | Saudi Arabia| 233|
1 | Canada| 232|
1 | Brazil| 226|
1 | Australia| 225|
1 | Egypt| 236|
1 | South Africa| 237|
+-----+

Process finished with exit code 0
```

Run | TODO | sbt shell | Terminal | Build | Messages

Compilation completed successfully with 1 warning in 5 s 25 ms (a minute ago)

Type here to search

195.1 CRLF UTF-8 ENG 22:44 07-01-2019