

Sensor Data

Code:

```
package spark.basic.cl

import org.apache.spark.sql.Session
import org.apache.spark.sql.types.IntegerType
import org.apache.spark.sql.functions._

object IntP extends App {

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

  val sparkcontext = sparkSession.sparkContext

  //OBJECTIVE 1
  //LOAD HVAC.csv

  val hv = sparkSession.read.format("csv").option("header","true")
    .load("/home/acadgild/Downloads/HVAC.csv")

  import sparkSession.implicits._

  val hvdf = hv.select($"Date" ,
    $"Time", $"TargetTemp".cast(IntegerType), $"ActualTemp".cast(IntegerType)
    , $"System".cast(IntegerType), $"SystemAge".cast(IntegerType), $"BuildingID".cast(I
negerType))

  // Add New Coulmn Temp change

  val tempnew = hvdf.withColumnn("Tempchange" ,
    when($"ActualTemp" - $"TargetTemp" > 5,1).otherwise(0))

  tempnew.show()

  //OBJECTIVE 2
  //Load building.csv

  val bu = sparkSession.read.format("csv").option("header","true")
    .load("/home/acadgild/Downloads/building.csv")

  val budf = bu.select($"BuildingID".cast(IntegerType) ,
    $"BuildingMgr", $"BuildingAge",
    $"HVACproduct", $"Country")

  budf.show()

  //OBJECTIVE 3
```

//Join Two tables where Tempchange is 1

```
val buhv = tempnew.join(budf,"BuildingID")
  .select($"Tempchange",$"Country")
  .where($"Tempchange" === 1)
```

```
buhv.show()
```

//Count Country Occurrence

```
val cou = buhv.groupBy($"Country").count().show()
```

```
}
```

OUTPUT:

OBJECTIVE 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

IntelliJ IDEA interface showing a Spark SQL query result in a table format. The table has columns: Date, Time, TargetTemp, ActualTemp, System, SystemAge, BuildingID, and Tempchange. The data shows temperature readings over time for various buildings, with Tempchange values mostly 0 or 1. The bottom of the screenshot shows a terminal window with logs from FileSourceStrategy and MemoryStore.

Date	Time	TargetTemp	ActualTemp	System	SystemAge	BuildingID	Tempchange
6/1/13	0:00:01	66	58	13	20	4	0
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	0
6/7/13	6:00:01	70	58	12	24	2	0
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	0
6/11/13	10:00:01	67	70	10	17	15	0
6/12/13	11:00:01	69	62	2	11	7	0
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	0
6/16/13	15:00:01	65	56	19	11	8	0
6/17/13	16:00:01	67	57	15	7	6	0
6/18/13	17:00:01	66	57	12	5	13	0
6/19/13	18:00:01	69	58	8	22	4	0
6/20/13	19:00:01	67	55	17	5	7	0

only showing top 20 rows

18/08/06 20:35:42 INFO FileSourceStrategy: Pruning directories with:
18/08/06 20:35:42 INFO FileSourceStrategy: Post-Scan Filters: (length(trim(value#67)) > 0)
18/08/06 20:35:42 INFO FileSourceStrategy: Output Data Schema: struct<value: string>
18/08/06 20:35:42 INFO FileSourceScanExec: Pushed Filters:
18/08/06 20:35:42 INFO MemoryStore: Block broadcast_5 stored as values in memory (estimated size 219.6 KB, free 110.2 MB)

Compilation completed successfully in 43 s 393 ms (2 minutes ago)

OBJECTIVE 2:

Load building.csv file into temporary table

IntelliJ IDEA interface showing the project `IntPrep` and the execution of a Scala script `IntP.scala`. The main editor displays the following table:

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

The Run console shows the following logs:

```
18/08/06 20:38:43 INFO executor: Finished task 0.0 in stage 3.0 (TID 3): 2072 bytes result sent to driver
18/08/06 20:38:43 INFO TaskSetManager: Finished task 0.0 in stage 3.0 (TID 3) in 85 ms on localhost (executor driver) (1/1)
18/08/06 20:38:43 INFO DAGScheduler: ResultStage 3 (show at IntP.scala:38) finished in 0.085 s
18/08/06 20:38:43 INFO DAGScheduler: Job 3 finished: show at IntP.scala:38, took 0.113191 s

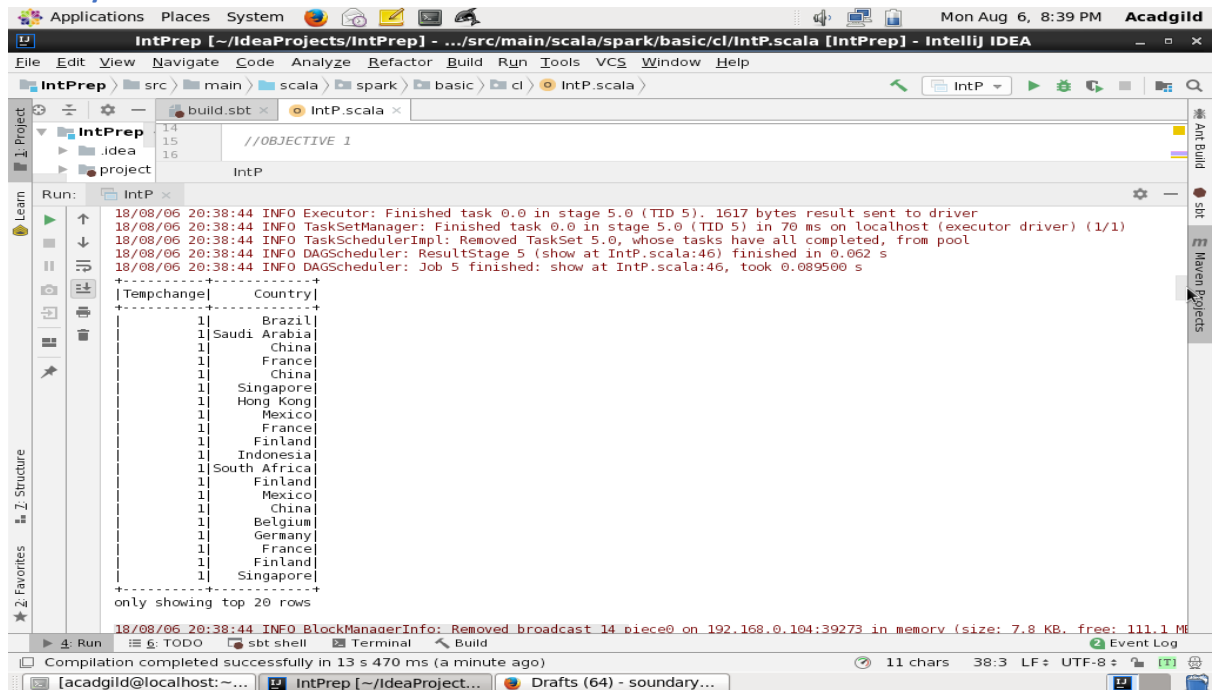
+-----+-----+-----+-----+-----+
|BuildingID|BuildingMgr|BuildingAge|HVACproduct|Country|
+-----+-----+-----+-----+-----+
|1| M1 | 25 | AC1000 | USA |
|2| M2 | 27 | FN39TG | France |
|3| M3 | 28 | JDNS77 | Brazil |
|4| M4 | 17 | GG1919 | Finland |
|5| M5 | 3 | ACMAX22 | Hong Kong |
|6| M6 | 9 | AC1000 | 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 |
+-----+-----+-----+-----+-----+

18/08/06 20:38:43 INFO TaskSchedulerImpl: Removed TaskSet 3.0, whose tasks have all completed, from pool
18/08/06 20:38:43 INFO BlockManagerInfo: Removed broadcast_8_piece0 on 192.168.0.104:39273 in memory (size: 20.6 KB, free: 111.1 MB)
18/08/06 20:38:43 INFO ContextCleaner: Cleaned accumulator 126
```

Compilation completed successfully in 13 s 470 ms (a minute ago)

OBJECTIVE 3:

- ○ 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



```
//OBJECTIVE 1
IntP

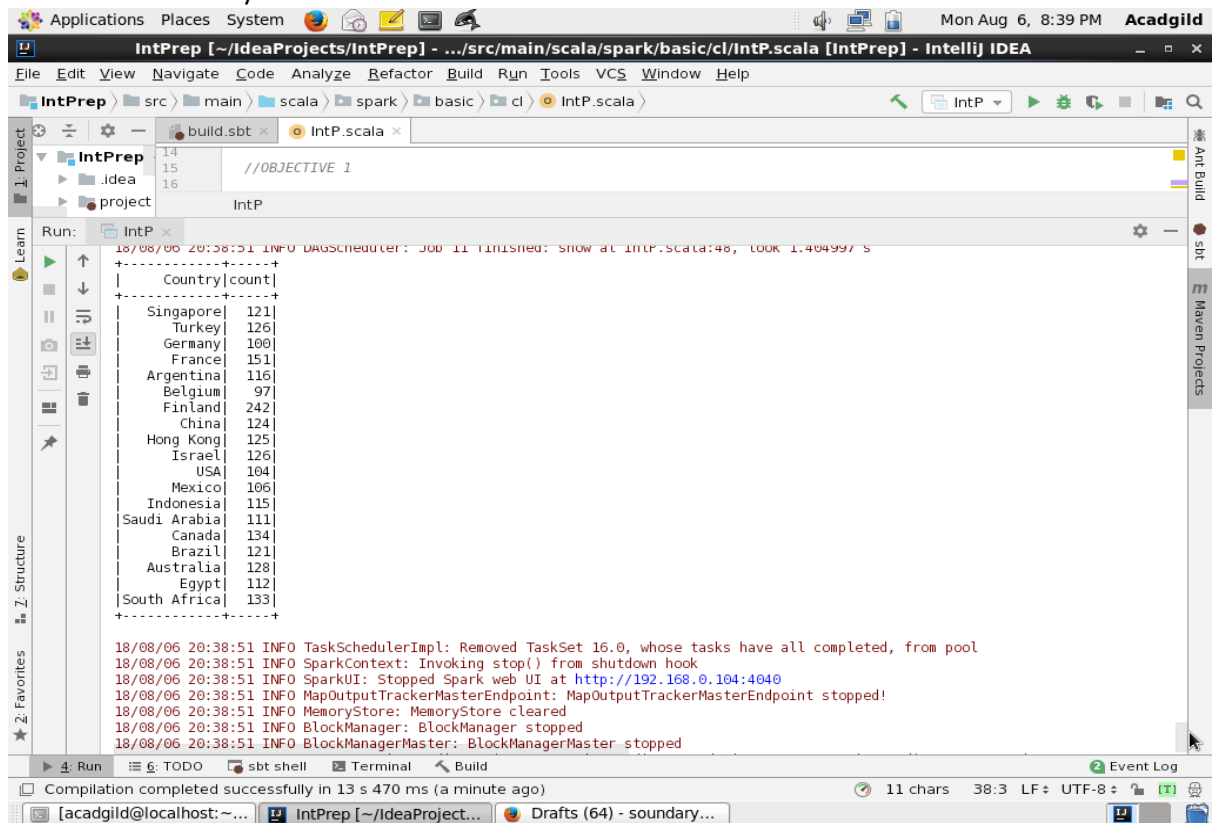
18/08/06 20:38:44 INFO Executor: Finished task 0.0 in stage 5.0 (TID 5). 1617 bytes result sent to driver
18/08/06 20:38:44 INFO TaskSetManager: Finished task 0.0 in stage 5.0 (TID 5) in 70 ms on localhost (executor driver) (1/1)
18/08/06 20:38:44 INFO TaskSchedulerImpl: Removed TaskSet 5.0, whose tasks have all completed, from pool
18/08/06 20:38:44 INFO DAGScheduler: ResultStage 5 (show at IntP.scala:46) finished in 0.062 s
18/08/06 20:38:44 INFO DAGScheduler: Job 5 finished: show at IntP.scala:46, took 0.089500 s

+-----+
|Tempchange|      Country|
+-----+
|1|      Brazil|
|1| Saudi Arabia|
|1|      China|
|1|      France|
|1|      China|
|1| Singapore|
|1| Hong Kong|
|1|      Mexico|
|1|      France|
|1| Finland|
|1| Indonesia|
|1| South Africa|
|1| Finland|
|1|      Mexico|
|1|      China|
|1| Belgium|
|1| Germany|
|1| France|
|1| Finland|
|1| Singapore|
+-----+
only showing top 20 rows

18/08/06 20:38:44 INFO BlockManagerInfo: Removed broadcast 14 piece0 on 192.168.0.104:39273 in memory (size: 7.8 KB, free: 111.1 MB)

Compilation completed successfully in 13 s 470 ms (a minute ago)
```

- Count No of Country:



```
//OBJECTIVE 1
IntP

18/08/06 20:38:51 INFO DAGScheduler: Job 11 finished: show at IntP.scala:48, took 1.404997 s

+-----+
|Country|count|
+-----+
|Singapore|121|
|Turkey|126|
|Germany|100|
|France|151|
|Argentina|116|
|Belgium|97|
|Finland|242|
|China|124|
|Hong Kong|125|
|Israel|126|
|USA|104|
|Mexico|106|
|Indonesia|115|
|Saudi Arabia|111|
|Canada|134|
|Brazil|121|
|Australia|128|
|Egypt|112|
|South Africa|133|
+-----+

18/08/06 20:38:51 INFO TaskSchedulerImpl: Removed TaskSet 16.0, whose tasks have all completed, from pool
18/08/06 20:38:51 INFO SparkContext: Invoking stop() from shutdown hook
18/08/06 20:38:51 INFO SparkUI: Stopped Spark web UI at http://192.168.0.104:4040
18/08/06 20:38:51 INFO MapOutputTrackerMasterEndpoint: MapOutputTrackerMasterEndpoint stopped!
18/08/06 20:38:51 INFO MemoryStore: MemoryStore cleared
18/08/06 20:38:51 INFO BlockManager: BlockManager stopped
18/08/06 20:38:51 INFO BlockManagerMaster: BlockManagerMaster stopped

Compilation completed successfully in 13 s 470 ms (a minute ago)
```