Sensor Data

Code:

```
package spark.basic.cl

import org.apache.spark.sql.SparkSession
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
ntegerType))

// Add New Coulmn Temp change

  val tempnew = hvdf.withColumn("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
```

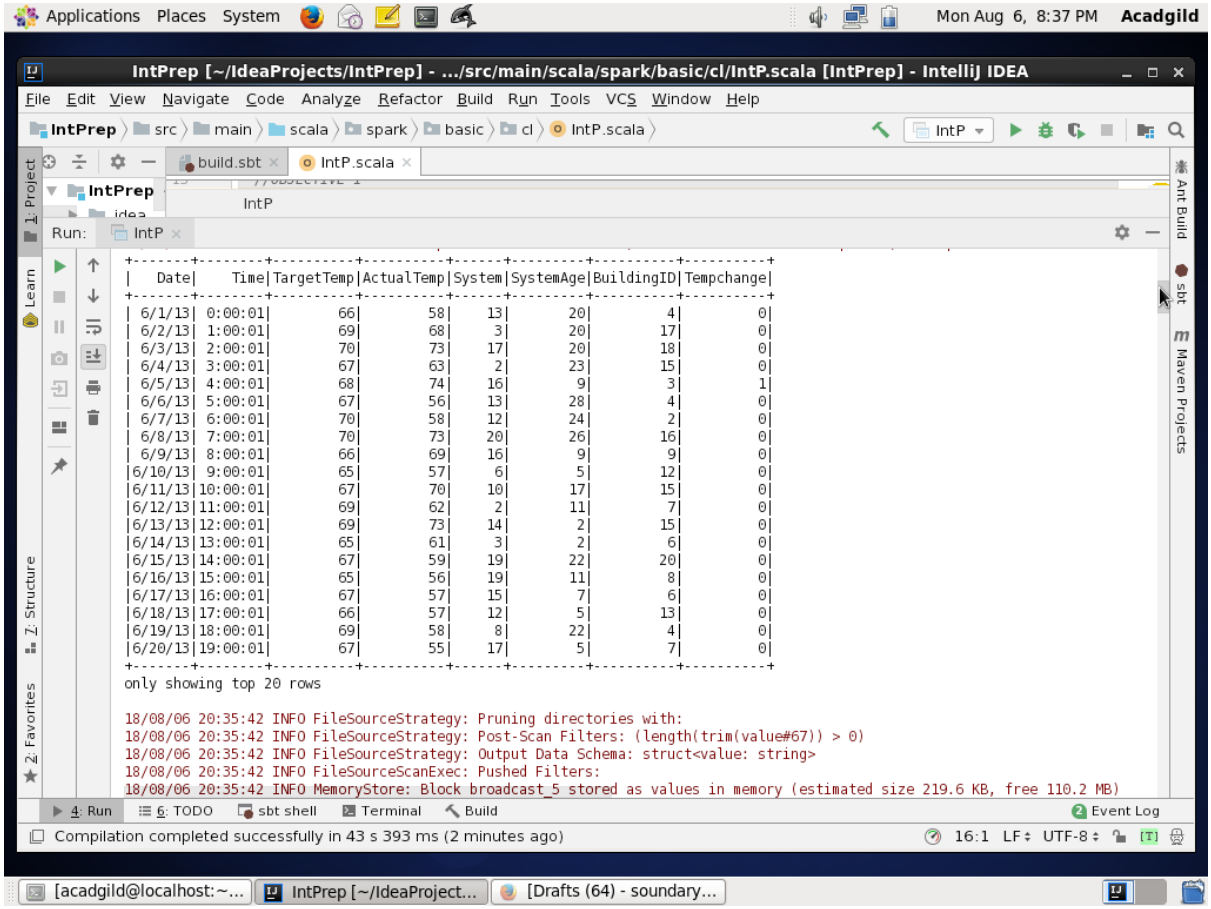
```
//Join Two tables where Tempchange is 1
  val buhv = tempnew.join(budf,"BuildingID")
    .select($"Tempchange",$"Country")
    .where($"Tempchange" === 1)

  buhv.show()
  //Count Country Occurrence
  val cou = buhv.groupBy($"Country").count().show()

}
```

OUTPUT:

OBJECTIVE 1

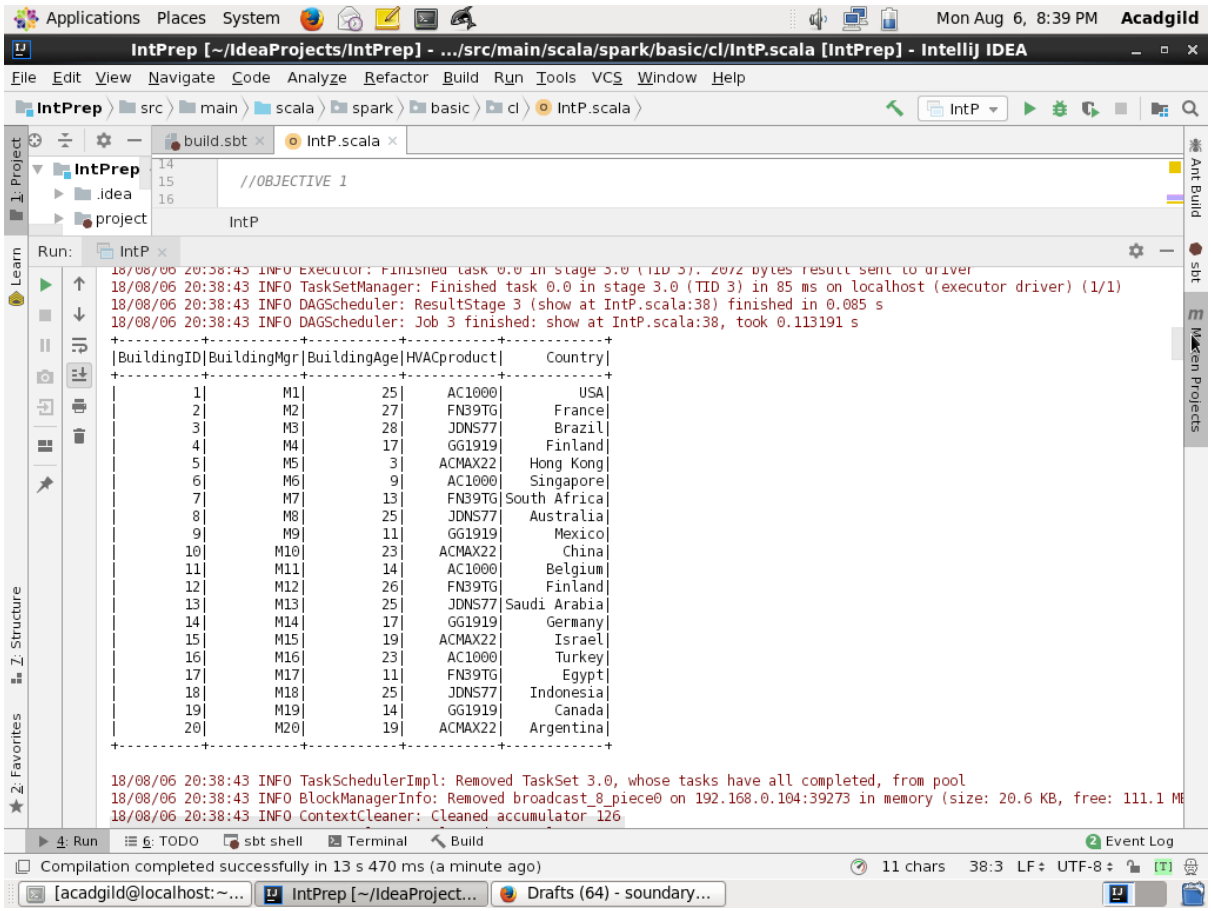
- **Load HVAC.csv file into temporary table**
- **Add a new column, tempchange - set to 1, if there is a change of greater than +/-5 between actual**

and target temperature



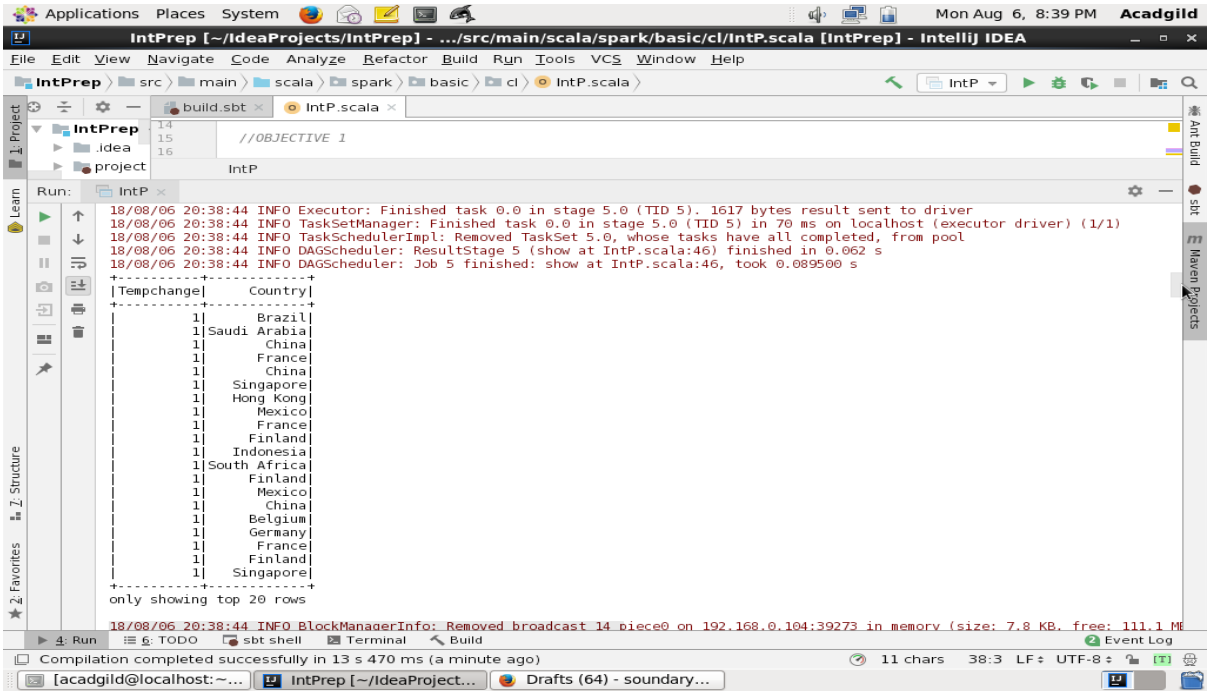
OBJECTIVE 2:

Load building.csv file into temporary table



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



- Count No of Country: