

## CS5542 Big Data Apps and Analytics LAB ASSIGNMENT #2 report

**Spark Programming:** Write a spark program with an interesting use case using text data as the input and program should have at least Two Spark Transformations and Two Spark Actions.

### Use case introduction:

I designed a small spark use case which described as a company log system, employees will do certain tasks every day, each action that employee made will be recorded in a log file, and each event will be given a date automatically, base on this case, i designed a small spark project that can read the log file and recognize the date then recognize and calculate how many days that the employee attended.

1. Following is the key **code** of the system:

```
val wc = input.flatMap(line=>{line.split(" ")}).map(word=>(word,1)).cache()

val output=wc.reduceByKey(_+_ )

val date = output.collect().filter(f =>
  f._1.length() == 8 && f._1.charAt(2) == '/' && f._1.charAt(5) == '/'
)

var number = 0

date.foreach(f => number += 1)

println("Number of days logged" + number)
```

2. and the **map reduce diagram** is showing below:

