In [1]:
```
from pyspark import SparkContext, SparkConf
```

In [2]:
```
conf = SparkConf().setMaster("local").setAppName("temperature")
sc = SparkContext(conf= conf)
```

```
24/10/13 20:16:20 WARN Utils: Your hostname, sundharakumar-HP-348-G7 resolves to a loopb
ack address: 127.0.1.1; using 192.168.0.114 instead (on interface wlo1)
24/10/13 20:16:20 WARN Utils: Set SPARK_LOCAL_IP if you need to bind to another address
Setting default log level to "WARN".
To adjust logging level use sc.setLogLevel(newLevel). For SparkR, use setLogLevel(newLev
el).
24/10/13 20:16:21 WARN NativeCodeLoader: Unable to load native-hadoop library for your p
latform... using builtin-java classes where applicable
```

In [10]:
```
lines = sc.textFile("file:///home/sundharakumar/Downloads/temp.csv")
```

In [11]:
```
def parseLine(line):
    fields = line.split(',')
    stationID = fields[0]
    typeID =fields[2]
    temp = float(fields[3])
    return (stationID, typeID, temp)
```

In [12]:
```
parsedLines = lines.map(parseLine)
```

In [13]:
```
minTemp = parsedLines.filter(lambda x: "TMIN" in x[1])
```

In [14]:
```
minTemp1 = minTemp.map( lambda x: (x[0],x[2]))
```

In [15]:
```
minTemp1 = minTemp1.reduceByKey(lambda x,y: min(x,y))
```

In [16]:
```
results = minTemp1.collect();
```

In [17]:
```
results
```

Out[17]:
```
[('ITE00100554', -148.0), ('EZE00100082', -135.0)]
```

In [ ]:
```

```