

```
spark2-shell --master yarn
```

```
rdd1 = sc.textFile("bigLogNew.txt")
```

```
rdd2 = rdd1.map(lambda x: (x.split(":")[0], x.split(":")[1]))
```

```
rdd3 = rdd2.groupByKey()
```

```
rdd4 = rdd3.map(lambda x: (x[0], len(x[1])))
```

```
rdd4.collect()
```

```
=====
```

```
import random
```

```
def randomGenerator():  
    return random.randint(1,60)
```

```
def myFunction(x):  
    if(x[0][0:4]=="WARN"):  
        return("WARN",x[1])  
    else:  
        return ("ERROR",x[1])
```

```
rdd1 = sc.textFile("bigLogNew.txt")
```

```
rdd2 = rdd1.map(lambda x:(x.split(":")[0] + str(randomGenerator()),x.split(":")[1]))
```

```
rdd3 = rdd2.groupByKey()
```

```
rdd4 = rdd3.map(lambda x: (x[0] , len(x[1])))
```

```
rdd4.cache()
```

```
rdd5 = rdd4.map(lambda x: myFunction(x))
```

```
rdd6 = rdd5.reduceByKey(lambda x,y: x+y)
```

```
rdd6.collect()
```