lab2_1

```
// TODO: Replace <FILL IN> with appropriate code
val data = 1 to 10000
print(data)
```

Took: 2 seconds 148 milliseconds, at 2016-9-25 16:54

```
// TODO: Replace <FILL IN> with appropriate code
val distData = sc.parallelize(data)
print(distData)
```

Took: 2 seconds, at 2016-9-25 16:54

```
val x = distData.filter(s=>s<10)
x.collect()
```

9 entries total

Show:

10 ▼

1	2
0	1
1	2
2	3
3	4
4	5
5	6
6	7
7	8
8	9

Showing 3 of 3 records

```
val pagecounts = sc.textFile("/home/akash/kth/data_intensive/labs/id2221/lab2/data/pag
print(pagecounts)
```

Took: 2 seconds 407 milliseconds, at 2016-9-25 20:37

```
for (x <-pagecounts.take(2)) {
    println(x)
}
```

Took: 3 seconds 103 milliseconds, at 2016-9-25 20:37

```
pagecounts.count()
```

1398882

Took: 2 seconds 383 milliseconds, at 2016-9-25 20:38

```
val enPages = pagecounts.filter(s => s.split("\\s")(1)=="en").cache()
```

Took: 813 milliseconds, at 2016-9-25 20:41

```
enPages.count()
```

970545

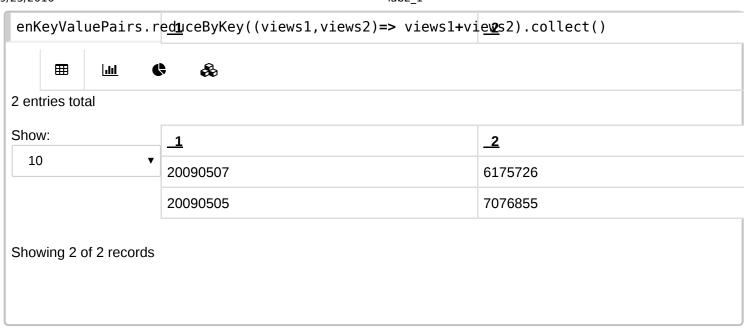Took: 1 second 59 milliseconds, at 2016-9-25 20:41

Took: 844 milliseconds, at 2016-9-25 21:29

```
val enTuples = enPages map(s=>{
  var fields = s.split("\\s")
  (fields(0).substring(0,8),fields(3).toInt)
})

//check with Amir 2 unique dates only
val enKeyValuePairs = enTuples
```

Took: 705 milliseconds, at 2016-9-25 23:3

```
enKeyValuePairs.reduceByKey((views1,views2)=> views1+views2).collect()
```

⊞ |ᎪᎢ| ◓   ⬡

2 entries total

Show:

| 10   ▾ |

_1_	_2_
20090507	6175726
20090505	7076855

Showing 2 of 2 records

```
enPages.map(s=>{
  var fields = s.split("\\s")
  (fields(0).substring(0,8),fields(3).toInt)
}).reduceByKey((views1,views2)=> views1+views2, 1).collect()
```

⊞ |ᎪᎢ| ◓   ⬡

2 entries total

Show:

| 10   ▾ |

_1_	_2_
20090507	6175726
20090505	7076855

Showing 2 of 2 records

```
/1check With Amir order of f2lter and reduce
enPages.map(x => x.split(" ")).map(s=> (s(2),s(3).toInt))
 .reduceByKey((views1,views2)=> views1+views2, 40).filter(viewTuple=> viewTuple._2 > 20(
```

⊞   ⛁

3 entries total

Show:

10 ▼

Search:

_1	_2
468159	Special:Search
451126	Main_Page
1066734	404_error/

Showing 3 of 3 records

Pages: Previous   1   Next

Took: 6 seconds 278 milliseconds, at 2016-9-25 23:23

◀ ▮▮▮▮▮▮▮▮   ▶

Build: | **buildTime**-*Sun Sep 25 12:54:11 UTC 2016* | **formattedShaVersion**-*0.7.0-SNAPSHOT-f6cd60a95cfc19cbae80285f187da9baec04e436* | **sbtVersion**-*0.13.8* | **scalaVersion**-*2.10.6* | **sparkNotebookVersion**-*0.7.0-SNAPSHOT* | **hadoopVersion**-*2.2.0* | **jets3tVersion**-*0.7.1* | **jlineDef**-*(org.scala-lang,2.10.6)* | **sparkVersion**-*2.0.0* | **withHive**-*false* |.