ex2

May 10, 2017

```
In [1]: import pyspark

In [2]: sc = pyspark.SparkContext('local[*]')

In [3]: rdd = sc.parallelize(range(1000))
        rdd.takeSample(False,5)

Out[3]: [524, 649, 427, 195, 321]

In [4]: import scipy

In [5]: data = []

In [6]: for xx in range (1,7):
            data = data+[xx]

In [7]: print(data)

[1, 2, 3, 4, 5, 6]


In [8]: rDD = sc.parallelize(data,4)

In [9]: mapRdd = rDD.filter(lambda X:X%5==0)

In [10]: redd = mapRdd.reduce(lambda a,b:a+b)

In [11]: print (mapRdd.collect())

[5]


In [12]: print (redd)

5


In [ ]:
```