ex1

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]: [33, 290, 235, 753, 395]

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

In [11]: mapRdd = rDD.map(lambda X:X*2)

In [12]: redd = mapRdd.reduce(lambda a,b:a+b)

In [13]: print (mapRdd.collect())

[2, 4, 6, 8, 10, 12]


In [15]: print (redd)

42


In [ ]:
```