

SparkFlight4

October 24, 2015

```
In [55]: from operator import add
data = sc.textFile("/home/matyi/Egyetem/osz2/big_data_eszkozok/hazi/input/2008.csv")
header = data.first()
rows = data.filter(lambda line: line != header) \
.map(lambda line: line.split(",")) \
.filter(lambda line: len(line)>1) \
.map(lambda line: (line[16],line[15])) \
.filter(lambda line: line[1] != "NA") \
.map(lambda line: (line[0], int(line[1]))) \
.filter(lambda line: line[1] > 0) \
.map(lambda line : (line[0], 1)) \
.reduceByKey(add) \
.takeOrdered(10, key=lambda x: -x[1])
#rows = [i[0] for i in rows]
rows

Out[55]: [(u'ATL', 175017),
(u'ORD', 159427),
(u'DFW', 127749),
(u'DEN', 104414),
(u'LAX', 87258),
(u'IAH', 87139),
(u'PHX', 82915),
(u'LAS', 76240),
(u'EWR', 69612),
(u'DTW', 59837)]
```