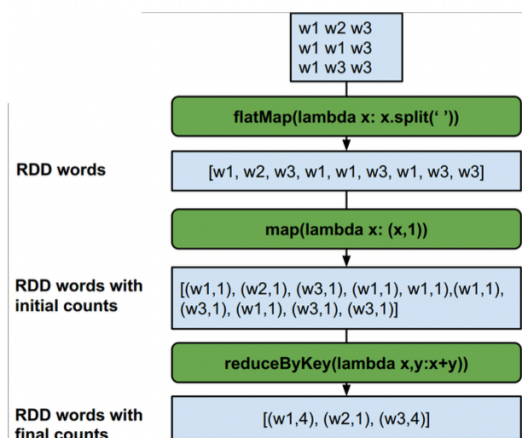


Word Count

```
import sys
from pyspark import SparkContext
sc = SparkContext(appName="WordCountExample")
lines = sc.textFile(sys.argv[1])
counts = lines.flatMap(lambda x: x.split(' ')) \
               .map(lambda x: (x, 1)) \
               .reduceByKey(lambda x,y:x+y)
output = counts.collect()
for (word, count) in output:
    print "%s: %i" % (word, count)
sc.stop()
```



Run using command line

- Turn on docker bash
- spark-submit wordcount.py README.md
- Result will be shown as follows

```
19/05/19 07:56:51 INFO scheduler.TaskSchedulerImpl: Removed
pool
19/05/19 07:56:51 INFO scheduler.DAGScheduler: ResultStage 1
finished in 0.150 s
19/05/19 07:56:51 INFO scheduler.DAGScheduler: Job 0 finished
0, took 2.050066 s
Turks: 1
States,294: 1
Algeria,United: 1
States,2025: 1
States,955: 1
States,Czech: 1
Colombia,United: 1
States,588: 1
States,Dominican: 1
```