

Homework 3 for Distributed System

Yanjian Zhang, 16300200020

Task 1:

mapper

```
import sys
import io
input_stream = io.TextIOWrapper(sys.stdin.buffer, encoding='gbk')
for line in input_stream:
    lineset = line.strip().split()
    # print file type and its count
    print("\t".join([lineset[-1].split(".")[1], "1"]))
```

reducer

```
import sys
cur_word = None
sum = 0
for line in sys.stdin:
    word, cnt = line.strip().split("\t")
    if cur_word == None:
        cur_word = word
    if cur_word != word:
        # print file type and its count
        print("\t".join([cur_word, str(sum)]))
        cur_word = word
        sum = 0
    sum += int(cnt)
```

result:

```
z@ubuntu:~/workdoc/homework3$ bash run.sh
rmr: DEPRECATED: Please use '-rm -r' instead.
rmr: 'output': No such file or directory
2019-03-31 05:04:48,076 WARN streaming.StreamJob: -file option is deprecated, please use generic option -files instead.
packageJobJar: [./hw3map.py, ./hw3reduce.py, /tmp/hadoop-unjar366701828782491255/] [] /tmp/streamjob1873967643470205880.jar tmpDir=null
2019-03-31 05:04:49,067 INFO client.RMPProxy: Connecting to ResourceManager at /0.0.0.0:8032
2019-03-31 05:04:49,327 INFO client.RMPProxy: Connecting to ResourceManager at /0.0.0.0:8032
2019-03-31 05:04:49,973 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/z/.staging/job_1554033682428_0001
2019-03-31 05:04:50,337 INFO mapred.FileInputFormat: Total input files to process : 1
2019-03-31 05:04:50,490 INFO mapreduce.JobSubmitter: number of splits:2
2019-03-31 05:04:50,629 INFO Configuration.deprecation: yarn.resourcemanager.system-metrics-publisher.enabled is deprecated. Instead, use yarn.system-metrics-publisher.enabled
2019-03-31 05:04:51,287 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1554033682428_0001
2019-03-31 05:04:51,295 INFO mapreduce.JobSubmitter: Executing with tokens: []
2019-03-31 05:04:51,962 INFO conf.Configuration: resource-types.xml not found
2019-03-31 05:04:51,963 INFO resource.ResourceUtils: Unable to find 'resource-types.xml'.
2019-03-31 05:04:52,963 INFO impl.YarnClientImpl: Submitted application application_1554033682428_0001
2019-03-31 05:04:53,036 INFO mapreduce.Job: The url to track the job: http://ubuntu:8088/proxy/application_1554033682428_0001/
2019-03-31 05:04:53,040 INFO mapreduce.Job: Running job: job_1554033682428_0001
2019-03-31 05:06:02,905 INFO mapreduce.Job: Job job_1554033682428_0001 running in uber mode : false
2019-03-31 05:06:02,907 INFO mapreduce.Job: map 0% reduce 0%
2019-03-31 05:07:17,031 INFO mapreduce.Job: map 33% reduce 0%
2019-03-31 05:07:24,490 INFO mapreduce.Job: map 50% reduce 0%
2019-03-31 05:07:37,601 INFO mapreduce.Job: map 100% reduce 0%
2019-03-31 05:07:44,698 INFO mapreduce.Job: map 100% reduce 100%
2019-03-31 05:07:46,721 INFO mapreduce.Job: Job job_1554033682428_0001 completed successfully
2019-03-31 05:07:47,092 INFO mapreduce.Job: Counters: 55
```

docx	1
dwg	32
jpg	9

Task 2

mapper

```
import sys
import io
input_stream = io.TextIOWrapper(sys.stdin.buffer, encoding='gbk')
for line in input_stream:
    lineset = line.strip().split()
    scale = lineset[2].replace(",", "")
    name = lineset[-1]
    # print the name and the file size
    print("\t".join([name, scale]))
```

reducer

```
import sys
dic1={}
for line in sys.stdin:
    name, scale = line.strip().split("\t")
    dic1[scale]=name
dic2=sorted(dic1.items(),key=lambda x: int(x[0]),reverse = True)
for one in dic2:
    print (one[1])
```

result

```
z@ubuntu:~/workdoc/homework3$ bash run.sh
rmr: DEPRECATED: Please use '-rm -r' instead.
Deleted output
2019-03-31 06:01:58,727 WARN streaming.StreamJob: -file option is deprecated, please use generic option -files instead.
packageJobJar: [/hw3map.py, ./hw3reduce.py, /tmp/hadoop-unjar662222662987569721/] [/tmp/streamjob1218275315934597389.jar tmpDir=null]
2019-03-31 06:02:00,299 INFO client.RMPProxy: Connecting to ResourceManager at /0.0.0.0:8032
2019-03-31 06:02:01,903 INFO client.RMPProxy: Connecting to ResourceManager at /0.0.0.0:8032
2019-03-31 06:02:04,630 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/z/.staging/job_1554033682428_0002
2019-03-31 06:02:06,195 INFO mapred.FileInputFormat: Total input files to process : 1
2019-03-31 06:02:06,926 INFO mapreduce.JobSubmitter: number of splits:2
2019-03-31 06:02:07,250 INFO Configuration.deprecation: yarn.resourcemanager.system-metrics-publisher.enabled is deprecated. Instead, use yarn.system-metrics-publisher.enabled
2019-03-31 06:02:09,221 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1554033682428_0002
2019-03-31 06:02:09,222 INFO mapreduce.JobSubmitter: Executing with tokens: []
2019-03-31 06:02:12,070 INFO conf.Configuration: resource-types.xml not found
2019-03-31 06:02:12,071 INFO resource.ResourceUtils: Unable to find 'resource-types.xml'.
2019-03-31 06:02:15,129 INFO impl.YarnClientImpl: Submitted application application_1554033682428_0002
2019-03-31 06:02:15,511 INFO mapreduce.Job: The url to track the job: http://ubuntu:8088/proxy/application_1554033682428_0002/
2019-03-31 06:02:15,512 INFO mapreduce.Job: Running job: job_1554033682428_0002
2019-03-31 06:02:45,701 INFO mapreduce.Job: Job job_1554033682428_0002 running in uber mode : false
2019-03-31 06:02:45,703 INFO mapreduce.Job: map 0% reduce 0%
2019-03-31 06:03:32,423 INFO mapreduce.Job: map 50% reduce 0%
2019-03-31 06:03:42,621 INFO mapreduce.Job: map 100% reduce 0%
2019-03-31 06:03:46,656 INFO mapreduce.Job: map 100% reduce 100%
2019-03-31 06:03:47,675 INFO mapreduce.Job: Job job_1554033682428_0002 completed successfully
2019-03-31 06:03:47,800 INFO mapreduce.Job: Counters: 55
```

总体-弱施-00-02.pdf
总体-弱施-00-01.pdf
典型幕墙详图.dwg
一层平面图.dwg
核心筒详图（一）.dwg
00-04.pdf

00-05.pdf
00-06.pdf
1-1剖面.dwg
00-11.pdf
00-12.pdf
00-08.pdf
00-09.pdf
00-13.pdf
00-03.pdf
N-A立面图.dwg
00-10.pdf
00-14.pdf
03-01.pdf
一层防火分区图.dwg
00-07.pdf
标准层防火分区图.dwg
屋面层平面图.dwg
二层防火分区图.dwg
三-九层、十二-十五层平面图.dwg
二层平面图.dwg
三十层防火分区图.dwg
十八层平面图.dwg
三十层平面图.dwg
十六-十七层平面图.dwg
二十二-二十九层平面图.dwg
十九层平面图.dwg
二十层平面图.dwg
二十一层平面图.dwg
避难层防火分区图.dwg
十层平面图.dwg
02-06.pdf
02-03.pdf
02-05.pdf
02-02.pdf
02-04.pdf
02-07.pdf
02-10.pdf
02-01.pdf
03-02.pdf
总体-弱施-02-202.pdf
01-02.pdf
02-09.pdf
总体-弱施-02-207.pdf
00-01.pdf
01-01.pdf
02-08.pdf
总体-弱施-02-203.pdf
00-02.pdf
总体-弱施-01-01.pdf
总体-弱施-02-204.pdf
02-11.pdf
总体-弱施-02-201.pdf
02-12.pdf

20181006-01_页面_2.jpg
01-03.pdf
20181005-01_页面_3.jpg
20181006-01_页面_1.jpg
总体-弱施-02-206.pdf
总体-弱施-02-205.pdf
20181005-01_页面_4.jpg
20181006-01_页面_3.jpg
20181005-01_页面_5.jpg
20181005-01_页面_1.jpg
20181005-01_页面_2.jpg
20181005-01.pdf
588550239356540541.jpg
20181006-01.pdf
居民意见第2条回复.docx