gedit countmapper.py

#!/usr/bin/python

import sys

def mapper():

for line in sys.stdin:

words=line.strip().split()

for word in words:

print "{0}\t1".format(word)

mapper()

gedit words.txt

**words.txt**

abcd

dbca

dcba

adcb

cat words.txt | python countmapper.py

gedit counreducer.py

**countreducer.py**

#!/usr/bin/python

import sys

def reducer():

current\_word,current\_count=None,0

for line in sys.stdin:

word,count=line.strip().split("\t")

count=int(count)

if word==current\_word:

current\_count+=count

else:

if current\_word:

print "{0}\t{1}".format(current\_word,current\_count)

current\_word=word

current\_count=count

if current\_word:

print "{0}\t{1}".format(current\_word,current\_count)

reducer()

cat words.txt | python countmapper.py | sort | python countreducer.py

hdfs dfs -put words.txt words.txt