gedit salemapper.py

#!/usr/bin/python

import sys

def mapper():

for line in sys.stdin:

fields=line.strip().split('\t')

p=fields[1]

c=fields[7]

print "{0}\t{1}".format(c,p)

mapper()

gedit sales.txt

cat sales.txt | python salemapper.py

gedit salereducer.py

#!/usr/bin/env python

import sys

def reducer():

cc=None

pc=0

For line in sys.stdin:

c,p=line.strip().split('\t')

if c=='country':continue

if c==cc:

pc+=1

else:

if cc:

print "{0}\t{1}".format(cc,pc)

cc=c

pc=1

if cc:

print "{0}\t{1}".format(cc,pc)

reducer()

cat sales.txt | python salemapper.py | sort | python salereducer.py

hdfs dfs -put sales.txt sales.txt