gedit elecmapper.py

#!/usr/bin/python

import sys

def mapper():

for line in sys.stdin:

line=line.strip()

y,consume=line.split(',')

consume=consume.split()

for v in consume:

print "{0}\t{1}".format(y,v) ELECTRICAL

mapper()

gedit electricity.txt

1979,23 23 43 24 25 26 26 26 26 25 26 25

1980,26 27 28 28 28 30 31 31 31 30 30 30

1981,31 32 32 32 33 34 35 36 36 34 34 35

1984,39 38 39 39 39 41 42 43 40 39 38 40

1985,38 39 39 39 41 41 41 00 40 39 39 45

cat electricity.txt | python elecmapper.py

gedit elecreducer.py

#!/usr/bin/env python

import sys

def reducer():

current\_year=None

maxc=float('-inf')

for line in sys.stdin:

y,c=map(int,line.strip().split('\t'))

if current\_year is not None and current\_year!=y:

print "{0}\t{1}".format(current\_year,maxc)

maxc=float('-inf')

maxc=max(maxc,c)

current\_year=y

if current\_year is not None:

print "{0}\t{1}".format(current\_year,maxc)

reducer()

cat electricity.txt | python elecmapper.py | sort | python elecreducer.py

hdfs dfs -put electricity.txt electricity.txt