gedit tempmapper.py

#!/usr/bin/python

import sys

def mapper():

for line in sys.stdin:

year,temperature=line.strip().split()

print "{0}\t{1}".format(year,temperature)

mapper()

gedit temperature.txt

2020 32

2021 40

2020 34

2021 26

2022 28 -

2023 30

2022 42

2023 40

#!/usr/bin/python

import sys

def reducer():

current\_year =None

maxtemp=float('-inf')

for line in sys.stdin:

y,t=line.strip().split("\t")

t=float(t)

if y==current\_year:

maxtemp=max(maxtemp,t)

else:

if current\_year:

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

current\_year=y

maxtemp=t

if current\_year:

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

reducer()

cat temperature.txt | python tempmapper.py | sort | python tempreducer.py

hdfs dfs -put temperature.txt temperature.txt