과목 명: 시스템프로그래밍

담당 교수 명: 소 정 민

<<Assignment 6>>

**서강대학교 컴퓨터공학과**

**[학번 ] 20141515**

**[이름] 김연후**

목 차

1. 프로그램 개요 3

2. 프로그램 설명 3

2.1 프로그램 흐름도 3

3. 모듈 정의 3

3.1 모듈 이름 : main() 3

3.2 모듈 이름: display\_instructions() 3

4. 전역 변수 정의 3

5. 코드 4

# 프로그램 개요

실수들의 list를 받아와서 group별로 최댓값을 갖는 값을 찾는다.

# 프로그램 설명

## 프로그램 흐름도



그림 1> 프로그램 흐름도

우선 generate.py를 실행하여 0~10000의 100000000개의 실수 값을 random하게 받아 각각 group을 배정하여 나열해놓은 input.data 파일을 만든다. 그 다음 mapper를 이용하여 ‘group \t number’ 형식으로 나열해준다. 그 다음 reducer를 이용하여 각 그룹 별 최댓값을 갖는 값을 maxNum list에 저장해두고 이후 출력해서 결과를 보인다.

# 모듈 정의

## Mapper.py

for문을 이용하여 file을 line별로 분류하여 각 line에서 group index와 random 생성된 숫자를 받아와서 group \t random real number로 출력해준다.

## Reducer.py

for문을 이용하여 mapper에 의해 group \t random real number 형식으로 정리된 파일을 받아와서 line 별로 group과 random real number를 비교해본다. 해당 group을 index로 하는 maxNum list의 최댓값과 현재 line의 random real number와 비교해서 현재 line의 number 가 더 크면 update를 위해 값을 바꿔준다. for문의 line 읽어오는 것이 모두 끝나면 마지막에 각 group들의 최댓값을 출력해준다.

# 전역 변수 정의

# 코드

Mapper.py

#!/usr/bin/env python

# Use the sys module

import sys

# 'file' in this case is STDIN

def read\_input(file):

# Split each line into words

for line in file:

yield line.split()

def main(separator='\t'):

# Read the data using read\_input

data = read\_input(sys.stdin)

# Process each word returned from read\_input

for nums in data:

# Process each word

for num in nums:

d = num.split(',')

print('%s%s%s' % (d[0], separator, d[1]))

if \_\_name\_\_ == "\_\_main\_\_":

main()

Reducer.py

#!/usr/bin/env python

# import modules

from itertools import groupby

from operator import itemgetter

import sys

# 'file' in this case is STDIN

def read\_mapper\_output(file, separator='\t'):

# Go through each line

for line in file:

# Strip out the separator character

yield line.rstrip().split(separator)

def main(separator='\t'):

# Read the data using read\_mapper\_output

data = read\_mapper\_output(sys.stdin, separator=separator)

# Group words and counts into 'group'

# Since MapReduce is a distributed process, each word

# may have multiple counts. 'group' will have all counts

# which can be retrieved using the word as the key.

maxNum = [0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0]

for group, current\_num in data:

try:

if float(maxNum[int(group)]) < float(current\_num):

maxNum[int(group)] = float(current\_num)

except ValueError:

# Count was not a number, so do nothing

pass

for i in range(0,16):

print("%s%s%f" % (i, separator, maxNum[i]))

if \_\_name\_\_ == "\_\_main\_\_":

main()