과목 명: 시스템프로그래밍

담당 교수 명: 소 정 민

<<Assignment 6>>

**서강대학교 컴퓨터공학과**

**20171601**

**강주형**

**목 차**

1. 프로그램 개요 3

2. 프로그램 설명 3

2.1 프로그램 흐름도 3

3. 모듈 정의 4

3.1 모듈 이름 : Mapper() 4

3.1.1 기능 4

3.2 모듈 이름 : Reducer() 4

3.1.1 기능 4

4. 코드

4.1 mapper.py 4

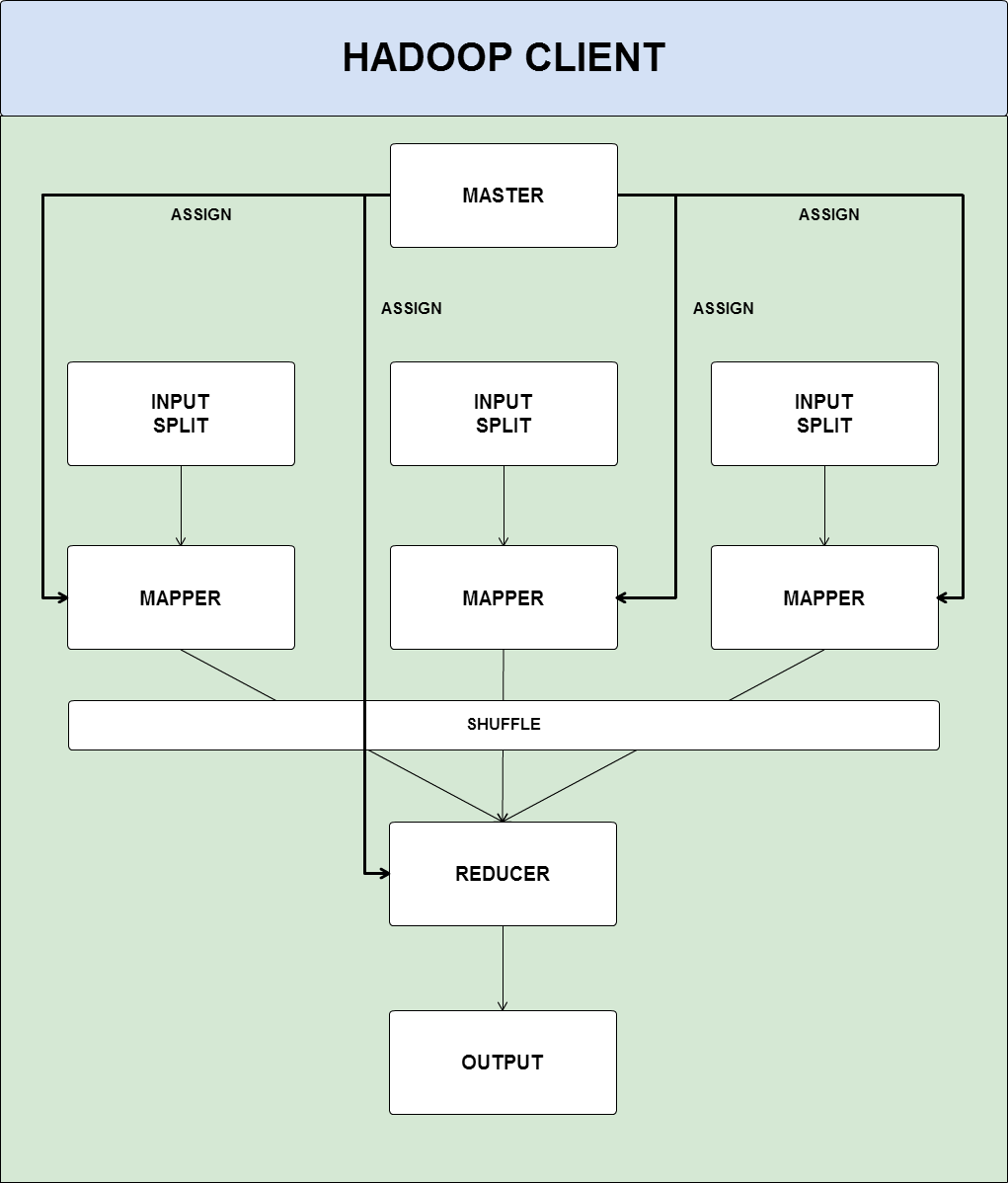
4.2 reducer.py 4

1. **프로그램 개요**

이 프로젝트는 Hadoop과 MapReduce에 대한 기본적인 이해를 향상시키고자 하는 것을 목표로 한다.

1. **프로그램 설명**

이 문제는 각 그룹에서 최댓값을 찾는 것이다. 입력은 한 개의 정수와 한 개의 실제 숫자를 가지며, 각 숫자는 쉼표로 구분된다. 정수 부분은 그룹 번호를 나타내며 실제 숫자 부분은 값을 의미한다. 주어진 데이터에 대해 각 그룹에 대한 최댓값을 찾아야 한다.



[흐름도]

1. **모듈 정의**

3.1 Mapper()

3.1.1 기능

: input의 내용을 한줄 씩 읽고 이를 콤마를 기준으로 나눠서 (그룹 번호) (콤마) (숫자) 의 형태로 나타낸다.

3.2 Reducer()

3.1.1 기능

: 그룹 번호마다 최댓값을 찾아서 그룹별 최댓값을 출력한다.

1. **코드**

4.1 mapper.py

#!/usr/bin/env python

import sys

mapped = {}

def read\_input(file):

for line in file:

yield line.strip().split(',')

def main(separator='\t'):

data = read\_input(sys.stdin)

for li in data:

print("%d%s%f" %(int (li[0]), separator, float(li[1])))

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

main()

4.2 reducer.py

#!/usr/bin/env python

from itertools import groupby

from operator import itemgetter

import sys

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

for line in file:

yield line.strip().split(separator, 1)

def main(separator = '\t'):

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

for key, group in groupby(data, itemgetter[0]):

try:

maxNum = 0

for num\_group, val in group:

if float(val) > maxNum:

maxNum = float(val)

print("%s%s%s" %(num\_group, separator, maxNum))

except ValueError:

pass

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

main()