Національний технічний університет України

«Київський політехнічний інститут імені Ігоря Сікорського»

Факультет інформатики та обчислювальної техніки

Кафедра автоматизованих систем обробки інформації та управління

Методи і засоби обробки великих даних

Лабораторна робота №2

«Програми для MapReduce»

Виконав:

студент групи ІС-12мп

Коноплянка Д. С.

Перевірила:

Тимофєєва Ю. С.

Київ 2021 р.

**Мета роботи:** Написати та виконати програму для MapReduce.

Варіант 6:

Написати програму, що виводить слова, які зустрічаються в тексті від 5 до

10 разів.

Mapper.py

#!/usr/bin/env python

stopwords = ["a", "and", "around", "every", "for", "from", "in",

"is", "it", "not", "on", "one", "the", "to", "under", "of", "at", "are", "as", "or"]

import sys

import re

for line in sys.stdin:

words = line.split()

for word in words:

cleanWord = re.sub('\W+', '', word)

if (cleanWord in stopwords):

continue

print('%s\t%s' % (cleanWord, 1))

reducer.py

#!/usr/bin/env python

from operator import itemgetter

import sys

current\_word = None

current\_count = 0

word = None

for line in sys.stdin:

word, count = line.split('\t', 1)

try:

count = int(count)

except ValueError:

continue

if current\_word == word:

current\_count += count

else:

if current\_word and current\_count >=5 and current\_count <= 10:

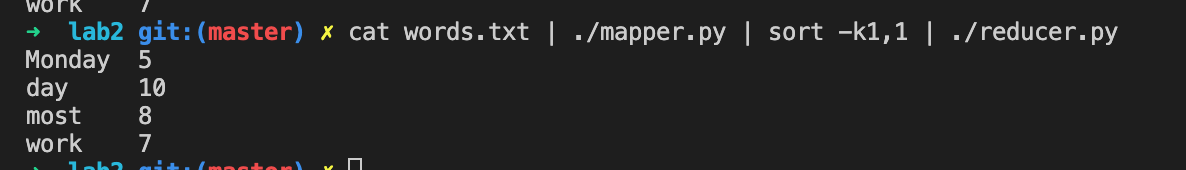
print('%s\t%s' % (current\_word, current\_count))

current\_count = count

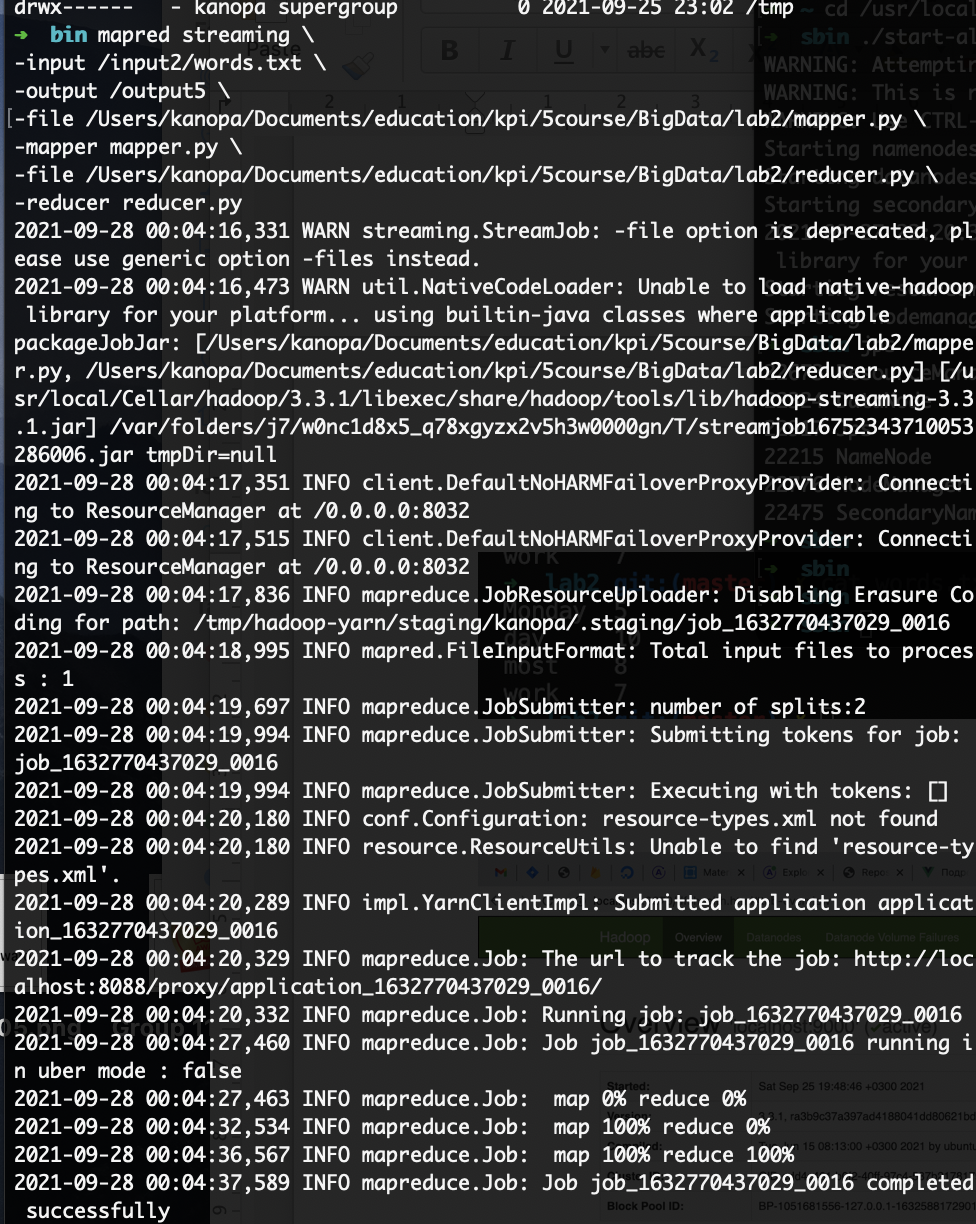
current\_word = word

if current\_word == word and current\_count >=5 and current\_count <= 10:

print('%s\t%s' % (current\_word, current\_count))



Команда для streaming Hadoop





Перевірка запису результатів у hadoop

