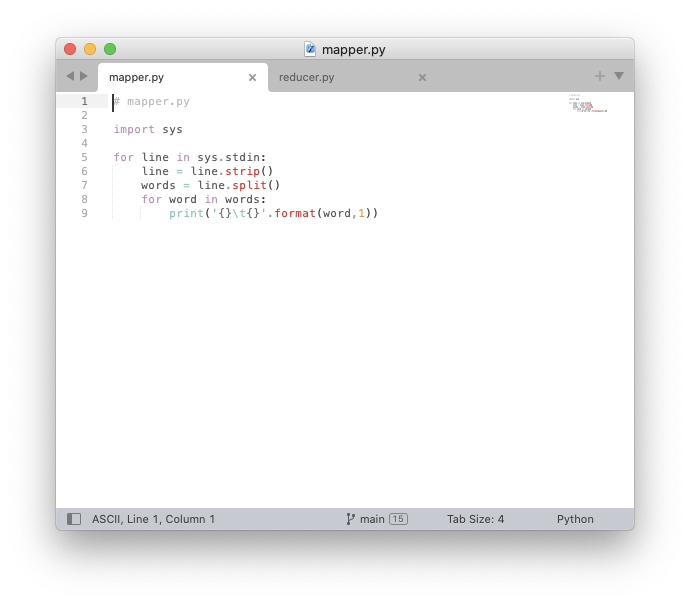
1. 应用相关的代码
2. 运行结果截图
3. 过程中遇到的问题与解决方案
4. 应用相关的代码

Mapper.py



# mapper.py

import sys

for line in sys.stdin:

line = line.strip()

words = line.split()

for word in words:

print('{}\t{}'.format(word,1))

reducer.py



# reducer.py

import sys

current\_word = None

current\_count = 0

word = None

for line in sys.stdin:

line = line.strip()

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

try:

count = int(count)

except ValueError:

continue

if current\_word == word:

current\_count += count

else:

if current\_word is not None:

print('{}\t{}'.format(current\_word,current\_count))

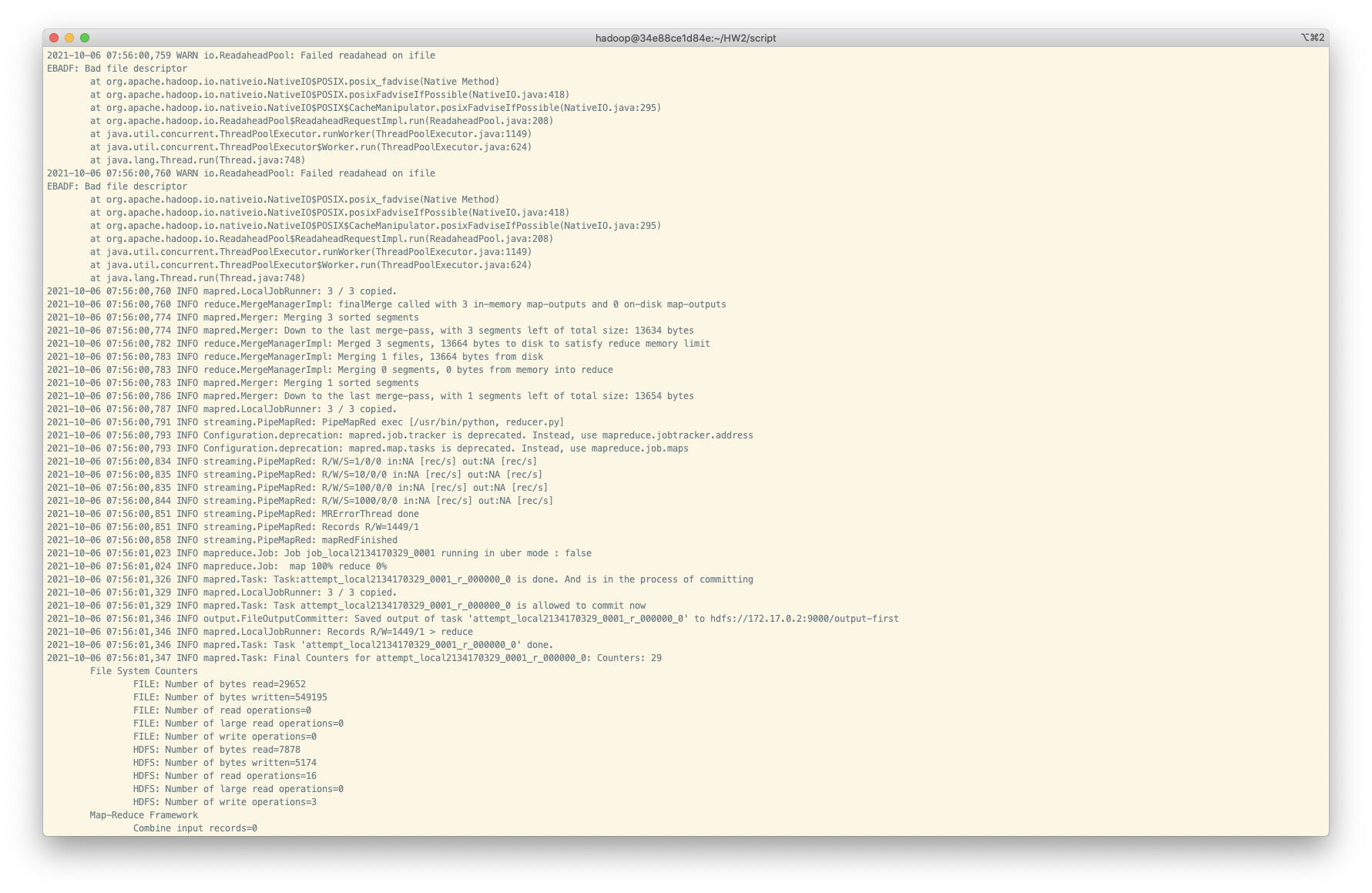
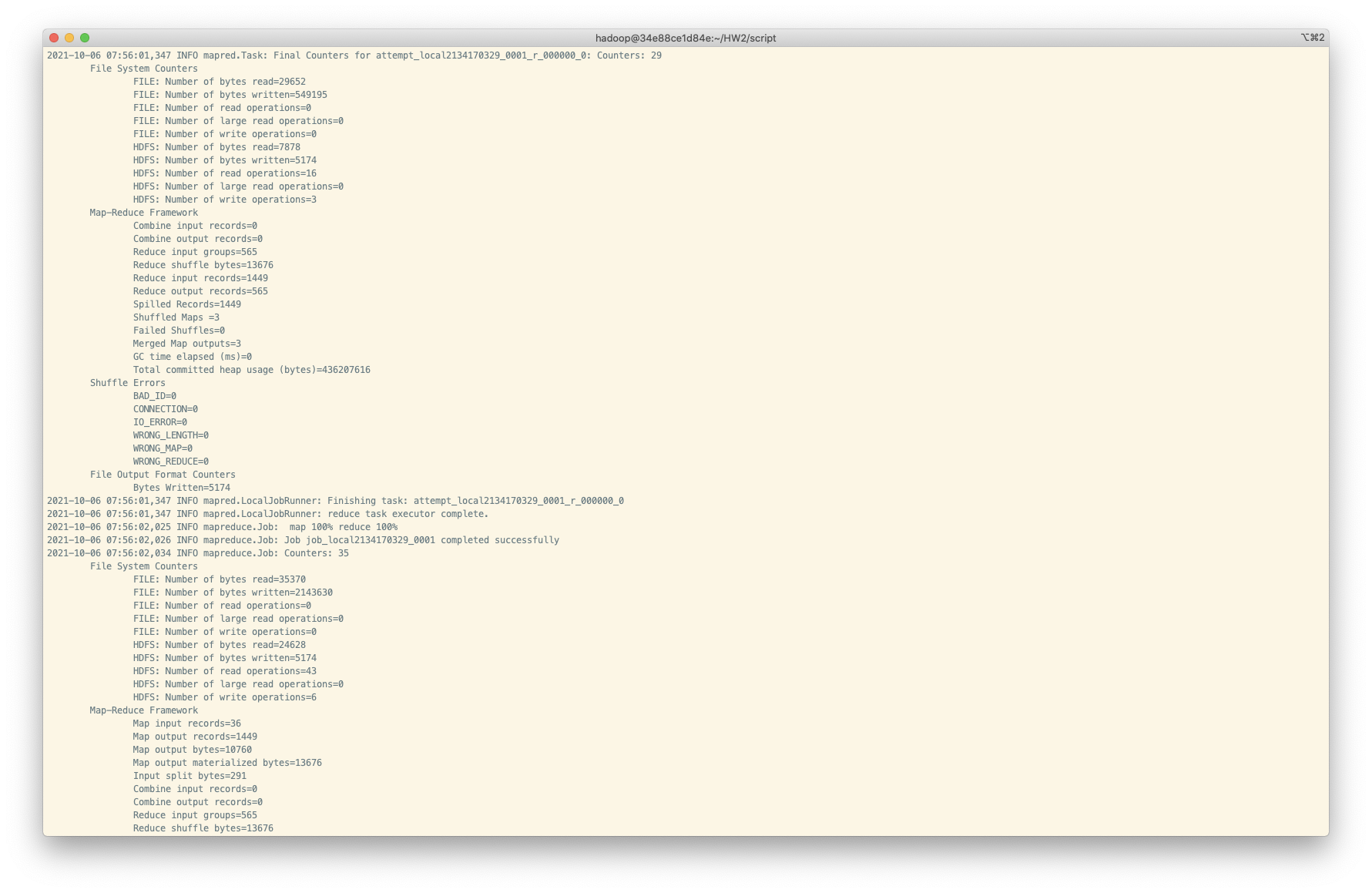
current\_word = word

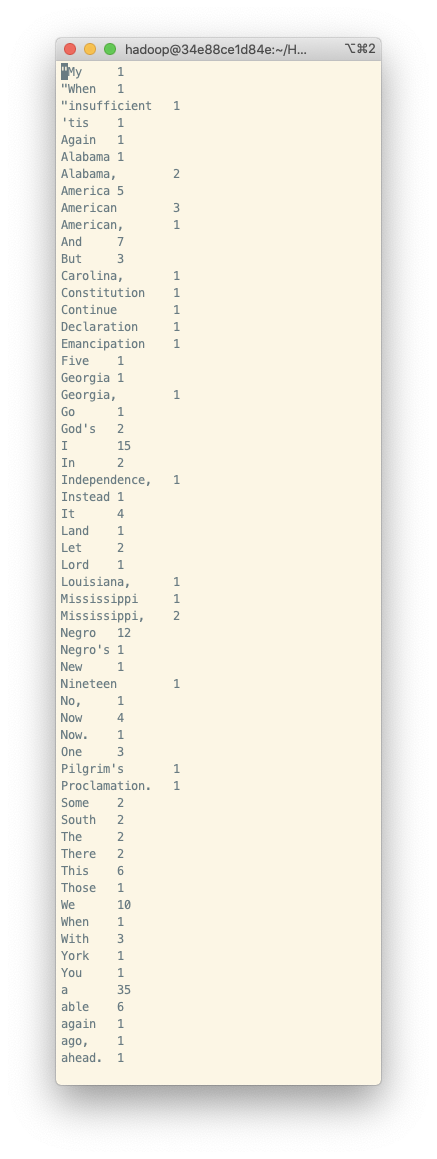
current\_count = count

if current\_word is not None:

print('{}\t{}'.format(current\_word,current\_count))

1. 运行结果截图



1. 过程中遇到的问题与解决方案

# 问题记录

## 实验环境搭建

1. [Hadoop运行环境部署](https://www.runoob.com/w3cnote/hadoop-setup.html)

2. [HDFS配置和使用](https://www.runoob.com/w3cnote/hdfs-setup.html)

3. hadoop运行python代码

- 报错

> /home/hadoop/HW2/./mapper.py: line 3: import: command not found

> /home/hadoop/HW2/./mapper.py: line 6: syntax error near unexpected token 'line'

> /home/hadoop/HW2/./mapper.py: line 6: ' line = line.strip()'

\* 查询后发现，-mapper参数后面应该跟一条命令，即需要指定执行环境和编译器

\* 同时由于使用的容器不带python，因此还需安装python

- 安装python

```sh

sudo yum install zlib-devel bzip2-devel openssl-devel ncurses-devel sqlite-devel readline-devel tk-devel gcc make libffi-devel -y

sudo yum -y install make wget

wget https://www.python.org/ftp/python/3.7.4/Python-3.7.4.tgz

sudo tar -zxvf Python-3.7.4.tgz

cd Python-3.7.4

sudo ./configure --prefix=/usr/local/python3

sudo make && sudo make install

#添加python3的软链接

sudo ln -s /usr/local/python3/bin/python3.7 /usr/bin/python

#测试是否安装成功了

python -V

```

- 修改运行参数

`hadoop jar /usr/local/hadoop/share/hadoop/tools/lib/hadoop-streaming-3.1.4.jar -file mapper.py -mapper "python mapper.py" -file reducer.py -reducer "python reducer.py" -input /input/\* -output /output-first`