

Map Reduce – Word Count

```

[cloudera@quickstart ~]$ ls
222010301016 Downloads lib Public
cloudera-manager eclipse local_222010301016 Template
cm_api.py enterprise-deployment.json Music Videos
Desktop express-deployment.json parcels workspac
Documents kerberos Pictures

[cloudera@quickstart ~]$ cd local_222010301016/
[cloudera@quickstart local_222010301016]$ ls
8th_sem placements projects
[cloudera@quickstart local_222010301016]$ gedit input.txt

```

The terminal window shows the user navigating to the directory `local_222010301016` and opening `input.txt` in the `gedit` editor. The editor window displays the following text:

```

hello world
hello sir
hello madam
madam is madam
sir is sir
madam sir madam

```

```

[cloudera@quickstart local_222010301016]$ python mapper.py
KeyboardInterrupt
>>> exit()
[cloudera@quickstart local_222010301016]$ gedit mapper.py

```

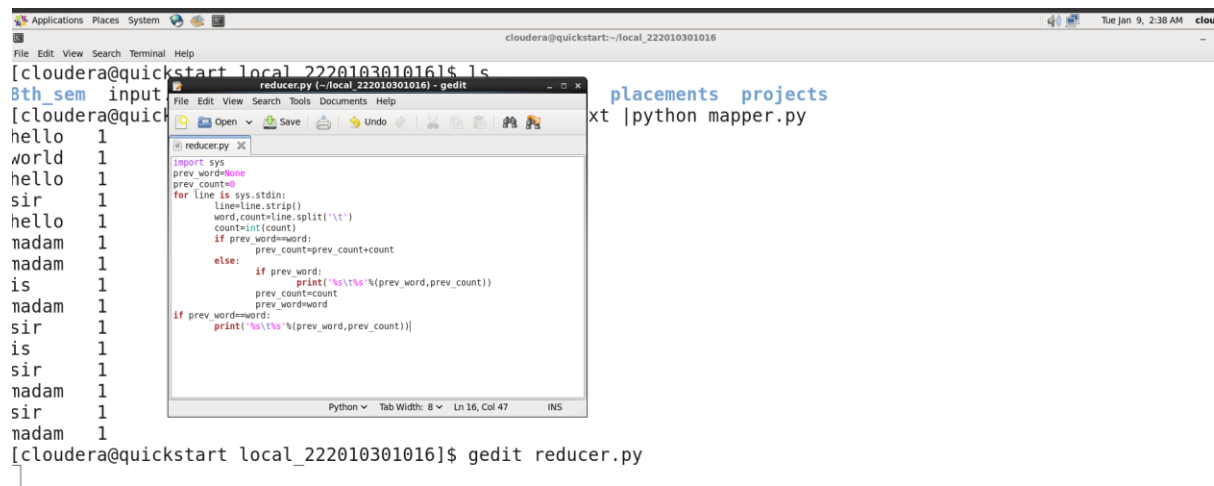
The terminal window shows the user running `python mapper.py`, which results in a `KeyboardInterrupt`. The user then exits the Python shell and opens `mapper.py` in the `gedit` editor. The editor window displays the following Python code:

```

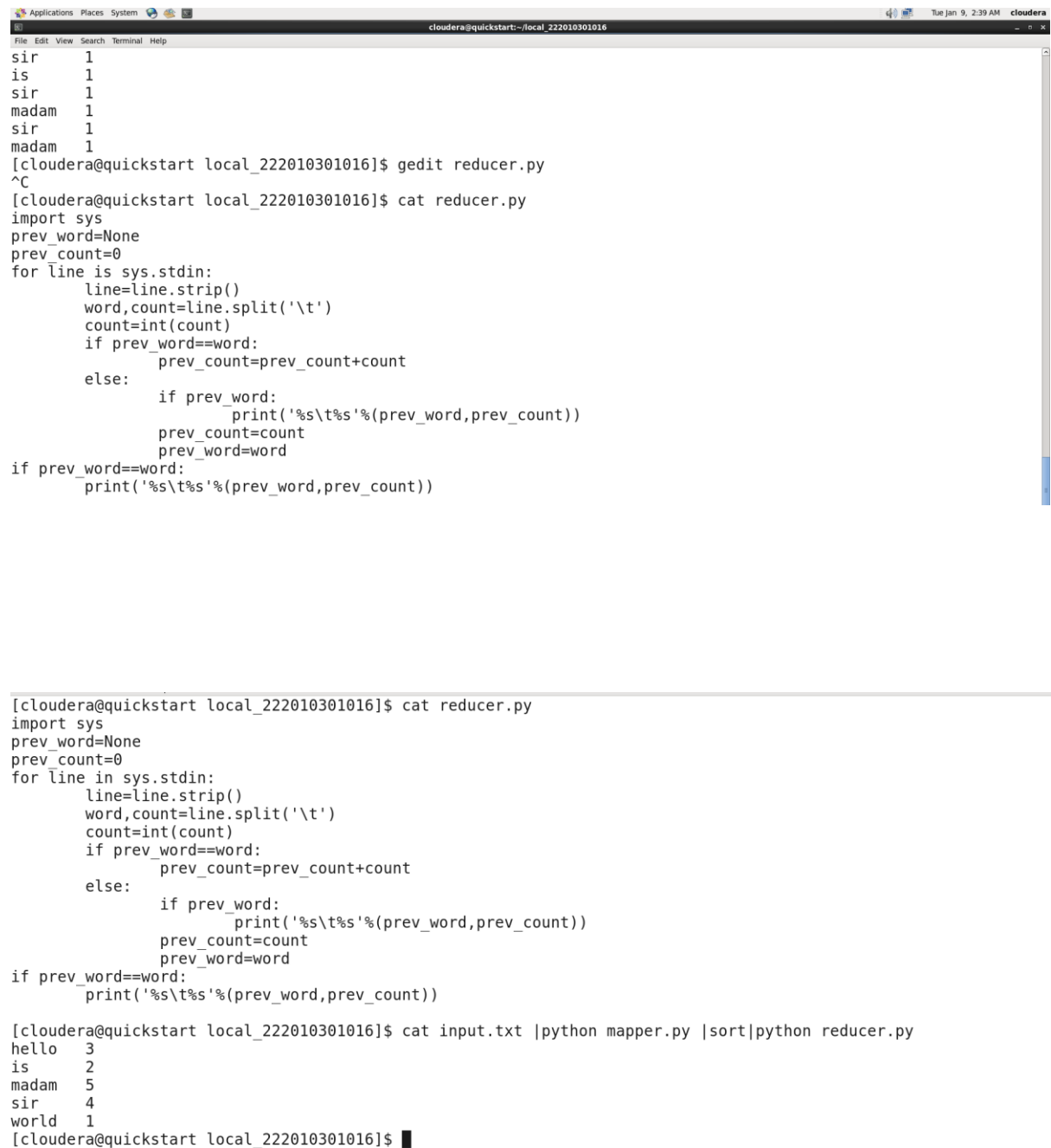
import sys
for line in sys.stdin:
    line = line.strip()
    words = line.split()
    for word in words:
        print('%s\t%s'%(word,1))

```

```
File Edit View Search Terminal Help
[cloudera@quickstart local_222010301016]$ ls
8th_sem input.txt input.txt~ mapper.py mapper.py~ placements projects
[cloudera@quickstart local_222010301016]$ cat input.txt |python mapper.py
hello 1
world 1
hello 1
sir 1
hello 1
madam 1
madam 1
is 1
madam 1
sir 1
is 1
sir 1
madam 1
sir 1
madam 1
[cloudera@quickstart local_222010301016]$
```



```
Applications Places System cloudera@quickstart:~/local_222010301016 Tue Jan 9, 2:38 AM
File Edit View Search Terminal Help
[cloudera@quickstart local_222010301016]$ ls
8th_sem input.txt input.txt~ mapper.py mapper.py~ placements projects
[cloudera@quickstart local_222010301016]$ cat input.txt |python mapper.py
hello 1
world 1
hello 1
sir 1
hello 1
madam 1
madam 1
is 1
madam 1
sir 1
is 1
sir 1
madam 1
sir 1
madam 1
[cloudera@quickstart local_222010301016]$ gedit reducer.py
reducer.py
import sys
prev_word=None
prev_count=0
for line in sys.stdin:
    line=line.strip()
    word,count=line.split('\t')
    count=int(count)
    if prev_word==word:
        prev_count+=prev_count+count
    else:
        if prev_word:
            print("%s\t%s"%(prev_word,prev_count))
            prev_count=count
            prev_word=word
        if prev_word==word:
            print("%s\t%s"%(prev_word,prev_count))
```



```
Applications Places System cloudera@quickstart:~/local_222010301016
File Edit View Search Terminal Help
sir 1
is 1
sir 1
madam 1
sir 1
madam 1
[cloudera@quickstart local_222010301016]$ gedit reducer.py
^C
[cloudera@quickstart local_222010301016]$ cat reducer.py
import sys
prev_word=None
prev_count=0
for line in sys.stdin:
    line=line.strip()
    word,count=line.split('\t')
    count=int(count)
    if prev_word==word:
        prev_count=prev_count+count
    else:
        if prev_word:
            print('%s\t%s'%(prev_word,prev_count))
        prev_count=count
        prev_word=word
if prev_word==word:
    print('%s\t%s'%(prev_word,prev_count))

[cloudera@quickstart local_222010301016]$ cat reducer.py
import sys
prev_word=None
prev_count=0
for line in sys.stdin:
    line=line.strip()
    word,count=line.split('\t')
    count=int(count)
    if prev_word==word:
        prev_count=prev_count+count
    else:
        if prev_word:
            print('%s\t%s'%(prev_word,prev_count))
        prev_count=count
        prev_word=word
if prev_word==word:
    print('%s\t%s'%(prev_word,prev_count))

[cloudera@quickstart local_222010301016]$ cat input.txt |python mapper.py |sort|python reducer.py
hello 3
is 2
madam 5
sir 4
world 1
[cloudera@quickstart local_222010301016]$
```