

MapReduce Practical

🕒 Created	@July 17, 2022 1:56 PM
🏷️ Tags	
📁 Property	

Open the terminal by pressing the Ctrl+Alt+T keys and run the following commands

```
$sudo apt update
```

```
$sudo apt upgrade
```

```
$sudo apt install software-properties-common
```

```
$sudo add-apt-repository ppa:deadsnakes/ppa
```

```
$sudo apt install python3.7
```

```
$ sudo update-alternatives --install /usr/bin/python python /usr/bin/python3.7 2
```

```
$python3.7 --version
```

```
$cd ~/hadoopMyFiles
```

```
$sudo gedit mapper.py
```

Add the following lines in the file

```
#!/usr/bin/python
```

```
"""mapper.py"""
```

```
import sys
```

```
for line in sys.stdin:
```

```
    line=line.strip()
```

```
    words=line.split()
```

```
    for word in words:
```

```
        print('%s\t%s'%(word,1))
```

Save and close the file.

```
$sudo gedit reducer.py
```

Add the following lines in the file:

```
#!/usr/bin/python
"""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 += count
    else:
        if prev_word:
            print('s%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))
```

Save and close the file.

```
$ls
```

```
$cat Sample_File.txt
```

```
$cat Sample_File.txt |python mapper.py
```

```
$cat Sample_File.txt |python mapper.py |sort |python reducer.py
```

Commands to run Python programs on Hadoop cluster

```
$ hadoop jar /usr/local/hadoop/share/hadoop/tools/lib/hadoop-streaming-
.jar
```

```
-file /home/hduser/mapper.py -mapper mapper.py  
-file /home/hduser/reducer.py -reducer reducer.py  
-input /user/input.txt  
-output /user/output.txt
```