

Distributed and Cloud Computing

CMT202



Student Number: 1766249

Date Set: 27th February 2018

E-Submission Date: 9:30 am 17th April 2018

Table of Contents

How to run and test code	3
Part1.....	3
Part2.....	6
Part3.....	7
Part4.....	9

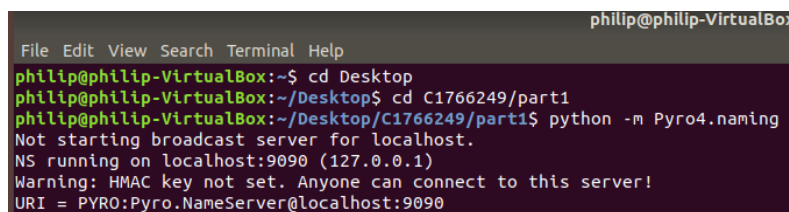
How to run and test code

Part1

For the first exercise we were asked to build a distributed data storage system using a remote object paradigm that will allow their employees to access, store and update book information.

In order to run this distributed system, you should follow the following steps:

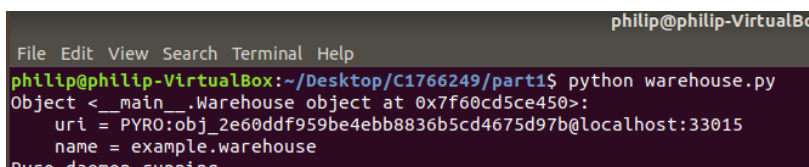
1. Open your terminal/cmd in Linux and download/install the uuid package by writing “***sudo apt-get install uuid-dev***”. The uuid package (universally unique identifier) as the its name implies it will allow the program to generate unique ids for each book. When the installation completes close that terminal.
2. Then Open a new terminal in the directory containing folder ‘part1’ either by using the command ‘***cd ../C1766249/part1***’ or by right-clicking inside folder ‘part1’ and then left-click ‘Open in Terminal’. Once you open the terminal in the appropriate directory execute the following command to start Pyro name server: “***python -m Pyro4.naming***” (or simply: ***pyro4-ns***) The name server will start and it prints something like:



```
philip@philip-VirtualBox:~$ cd Desktop
philip@philip-VirtualBox:~/Desktop$ cd C1766249/part1
philip@philip-VirtualBox:~/Desktop/C1766249/part1$ python -m Pyro4.naming
Not starting broadcast server for localhost.
NS running on localhost:9090 (127.0.0.1)
Warning: HMAC key not set. Anyone can connect to this server!
URI = PYRO:Pyro.NameServer@localhost:9090
```

Figure 1.0

3. Then, Open a new terminal in the same directory as before. This time execute the command “***python warehouse.py***” to run the Pyro daemon server. This will allow Pyro to listen for remote method call, dispatch them to the appropriate actual objects and return the results to the caller. This will print something like this:



```
philip@philip-VirtualBox:~/Desktop/C1766249/part1$ python warehouse.py
Object <__main__.Warehouse object at 0x7f60cd5ce450>:
  uri = PYRO:obj_2e60ddf959be4ebb8836b5cd4675d97b@localhost:33015
  name = example.warehouse
Pyro daemon running
```

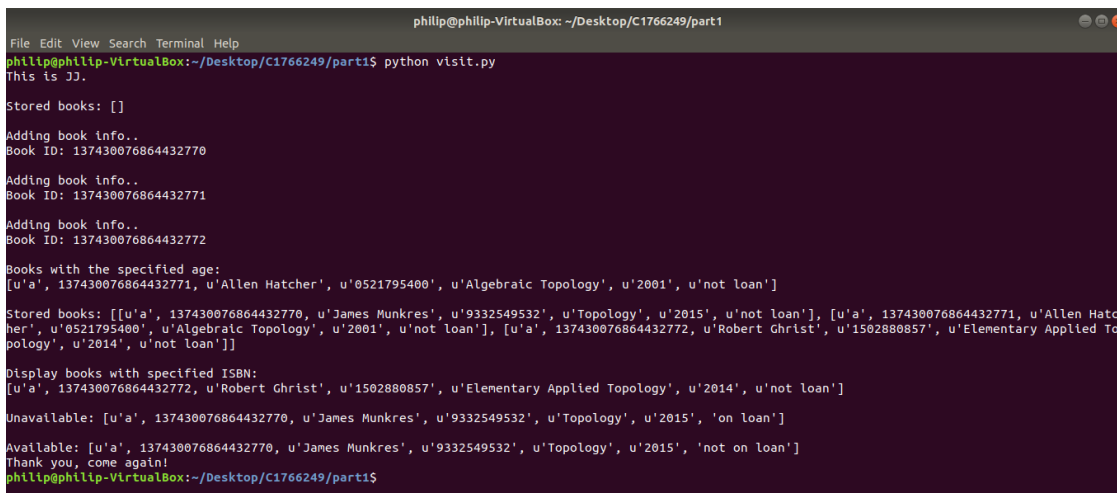
Figure 1.1

- Similarly, open another terminal in the same directory but this time execute the command ***“python visit.py”*** which will run the program with the following csv file (C1766249/part1/tasks.txt):

```
d
a,James Munkres,9332549532,Topology,2015
a,Allen Hatcher,0521795400,Algebraic Topology,2001
a,Robert Ghrist,1502880857,Elementary Applied Topology,2014
sy,2000,2002
d
si,1502880857
ol,9332549532

nol,933254953
```

In addition it will print the following answer:

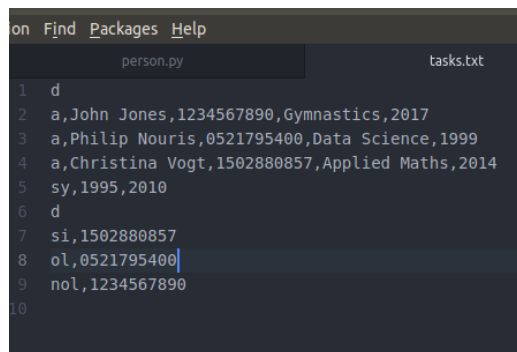


```
phillip@phillip-VirtualBox: ~/Desktop/C1766249/part1
File Edit View Search Terminal Help
phillip@phillip-VirtualBox:~/Desktop/C1766249/part1$ python visit.py
This is JJ.
Stored books: []
Adding book info..
Book ID: 137430076864432770
Adding book info..
Book ID: 137430076864432771
Adding book info..
Book ID: 137430076864432772
Books with the specified age:
[u'a', 137430076864432771, u'Allen Hatcher', u'0521795400', u'Algebraic Topology', u'2001', u'not loan']
Stored books: [[u'a', 137430076864432770, u'James Munkres', u'9332549532', u'Topology', u'2015', u'not loan'], [u'a', 137430076864432771, u'Allen Hatcher', u'0521795400', u'Algebraic Topology', u'2001', u'not loan'], [u'a', 137430076864432772, u'Robert Ghrist', u'1502880857', u'Elementary Applied Topology', u'2014', u'not loan']]
Display books with specified ISBN:
[u'a', 137430076864432772, u'Robert Ghrist', u'1502880857', u'Elementary Applied Topology', u'2014', u'not loan']
Unavailable: [u'a', 137430076864432770, u'James Munkres', u'9332549532', u'Topology', u'2015', u'on loan']
Available: [u'a', 137430076864432770, u'James Munkres', u'9332549532', u'Topology', u'2015', u'not on loan']
Thank you, come again!
phillip@phillip-VirtualBox:~/Desktop/C1766249/part1$
```

Figure 1.2

- Finally, to test the code is correct you can edit the tasks.txt file appropriately by using different book information and save it (**Figure 1.3**), then repeat the steps starting from *step 3* as daemon server needs to be reset every time a change to the code or the csv is made:

New text file –



```
on Find Packages Help
person.py tasks.txt
1 d
2 a,John Jones,1234567890,Gymnastics,2017
3 a,Philip Nouris,0521795400,Data Science,1999
4 a,Christina Vogt,1502880857,Applied Maths,2014
5 sy,1995,2010
6 d
7 si,1502880857
8 ol,0521795400
9 nol,1234567890
10
```

Figure 1.3

New answer –

```
philip@philip-VirtualBox: ~/Desktop/C1766249/part1
File Edit View Search Terminal Help
philip@philip-VirtualBox:~/Desktop/C1766249/part1$ python visit.py
This is JJ.

Stored books: []

Adding book info..
Book ID: 137430116258620080

Adding book info..
Book ID: 137430116258620081

Adding book info..
Book ID: 137430116258620082

Books with the specified age:
[u'a', 137430116258620081, u'Philip Nouris', u'0521795400', u'Data Science', u'1999', u'not loan']

Stored books: [[u'a', 137430116258620080, u'John Jones', u'1234567890', u'Gymnastics', u'2017', u'not loan'], [u'a', 137430116258620081, u'Philip Nouris', u'0521795400', u'Data Science', u'1999', u'not loan'], [u'a', 137430116258620082, u'Christina Vogt', u'1502880857', u'Applied Maths', u'2014', u'not loan']]

Display books with specified ISBN:
[u'a', 137430116258620082, u'Christina Vogt', u'1502880857', u'Applied Maths', u'2014', u'not loan']

Unavailable: [u'a', 137430116258620081, u'Philip Nouris', u'0521795400', u'Data Science', u'1999', 'on loan']

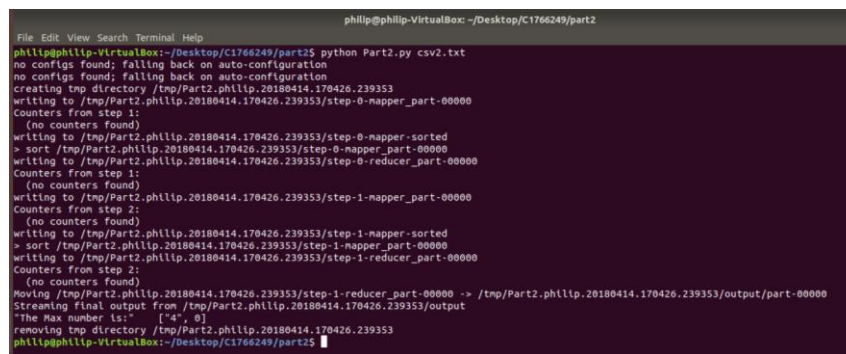
Available: [u'a', 137430116258620080, u'John Jones', u'1234567890', u'Gymnastics', u'2017', 'not on loan']
Thank you, come again!
philip@philip-VirtualBox:~/Desktop/C1766249/part1$
```

Figure 1.4

Part2

In part 2 we were asked to write a MapReduce program which determines the maximum of all numbers in a CSV file.

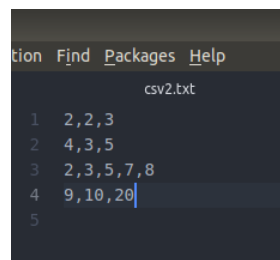
1. First open the terminal/cmd in the directory containing folder '**part2**' either by using the command '**cd ../C1766249/part2**' or by right-clicking inside folder '**part2**' and then left-click '**Open in Terminal**'.
2. Then write and execute "**python Part2.py csv2.txt**". Always have the python file first, followed by the desired text file when running a MapReduce program. This will print something like this:



```
philip@philip-VirtualBox: ~/Desktop/C1766249/part2
File Edit View Search Terminal Help
philip@philip-VirtualBox:~/Desktop/C1766249/part2$ python Part2.py csv2.txt
no configs found; falling back on auto-configuration
no configs found; falling back on auto-configuration
creating tmp directory /tmp/Part2.philip.20180414.170426.239353
writing to /tmp/Part2.philip.20180414.170426.239353/step-0-mapper_part-00000
Counters from step 1:
(no counters found)
writing to /tmp/Part2.philip.20180414.170426.239353/step-0-mapper-sorted
> sort /tmp/Part2.philip.20180414.170426.239353/step-0-mapper_part-00000
writing to /tmp/Part2.philip.20180414.170426.239353/step-0-reducer_part-00000
Counters from step 1:
(no counters found)
writing to /tmp/Part2.philip.20180414.170426.239353/step-1-mapper_part-00000
Counters from step 2:
(no counters found)
writing to /tmp/Part2.philip.20180414.170426.239353/step-1-mapper-sorted
> sort /tmp/Part2.philip.20180414.170426.239353/step-1-mapper_part-00000
writing to /tmp/Part2.philip.20180414.170426.239353/step-1-reducer_part-00000
Counters from step 2:
(no counters found)
Moving /tmp/Part2.philip.20180414.170426.239353/step-1-reducer_part-00000 -> /tmp/Part2.philip.20180414.170426.239353/output/part-00000
Streaming final output from /tmp/Part2.philip.20180414.170426.239353/output
"The max number is:" [4, 0]
removing tmp directory /tmp/Part2.philip.20180414.170426.239353
philip@philip-VirtualBox:~/Desktop/C1766249/part2$
```

Figure 2.0

3. To test the code edit the "**csv2.txt**" file in a text editor save, and then repeat steps **1&2**.
Example:



```
tion Find Packages Help
csv2.txt
1 2, 2, 3
2 4, 3, 5
3 2, 3, 5, 7, 8
4 9, 10, 20
5
```

Figure 2.1



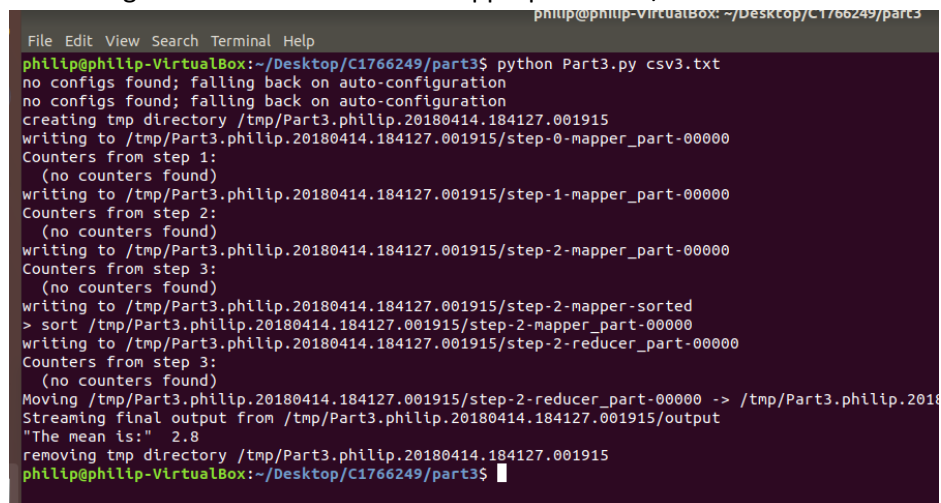
```
writing to /tmp/Part2.philip.20180414.182225.400217/step-1-mapper_part-00000
Counters from step 2:
(no counters found)
writing to /tmp/Part2.philip.20180414.182225.400217/step-1-mapper-sorted
> sort /tmp/Part2.philip.20180414.182225.400217/step-1-mapper_part-00000
writing to /tmp/Part2.philip.20180414.182225.400217/step-1-reducer_part-00000
Counters from step 2:
(no counters found)
Moving /tmp/Part2.philip.20180414.182225.400217/step-1-reducer_part-00000 -> /tmp/Part2.philip.20180414.182225.400217/output
Streaming final output from /tmp/Part2.philip.20180414.182225.400217/output
"The max number is:" [20, 0]
removing tmp directory /tmp/Part2.philip.20180414.182225.400217
philip@philip-VirtualBox:~/Desktop/C1766249/part2$
```

Figure 2.2

Part3

Part 3 was similar to part2, we had to write a MapReduce program which would determine the mean of all numbers in a file.

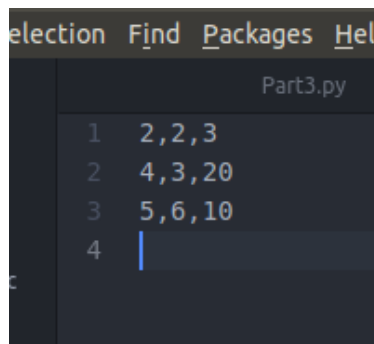
1. In the same way as in part 2 open the terminal/cmd, but in the directory containing folder '**part3**' either by using the command '**cd ./C1766249/part3**' or by right-clicking inside folder '**part3**' and then left-click '**Open in Terminal**'.
2. Then write and execute "**python Part3.py csv3.txt**". Always have the python file first, followed by the desired text file when running a MapReduce program. Executing that command should print the average of all the numbers in the appropriate text/csv file:



```
philip@philip-VirtualBox: ~/Desktop/C1766249/part3
File Edit View Search Terminal Help
philip@philip-VirtualBox:~/Desktop/C1766249/part3$ python Part3.py csv3.txt
no configs found; falling back on auto-configuration
no configs found; falling back on auto-configuration
creating tmp directory /tmp/Part3.philip.20180414.184127.001915
writing to /tmp/Part3.philip.20180414.184127.001915/step-0-mapper_part-00000
Counters from step 1:
(no counters found)
writing to /tmp/Part3.philip.20180414.184127.001915/step-1-mapper_part-00000
Counters from step 2:
(no counters found)
writing to /tmp/Part3.philip.20180414.184127.001915/step-2-mapper_part-00000
Counters from step 3:
(no counters found)
writing to /tmp/Part3.philip.20180414.184127.001915/step-2-mapper-sorted
> sort /tmp/Part3.philip.20180414.184127.001915/step-2-mapper_part-00000
writing to /tmp/Part3.philip.20180414.184127.001915/step-2-reducer_part-00000
Counters from step 3:
(no counters found)
Moving /tmp/Part3.philip.20180414.184127.001915/step-2-reducer_part-00000 -> /tmp/Part3.philip.20180414.184127.001915/output
Streaming final output from /tmp/Part3.philip.20180414.184127.001915/output
"The mean is:" 2.8
removing tmp directory /tmp/Part3.philip.20180414.184127.001915
philip@philip-VirtualBox:~/Desktop/C1766249/part3$
```

Figure 3.0

3. In order to test the code, change the content of the "**csv3.txt**" file by using a text editor as shown (**Figure 3.1**), and then repeat steps 1 & 2.



```
Part3.py
1 2,2,3
2 4,3,20
3 5,6,10
4
```

Figure 3.2

New answer -

```
philip@philip-VirtualBox: ~/Desktop
File Edit View Search Terminal Help
philip@philip-VirtualBox:~/Desktop/C1766249/part3$ python Part3.py csv3.txt
no configs found; falling back on auto-configuration
no configs found; falling back on auto-configuration
creating tmp directory /tmp/Part3.philip.20180414.184607.327333
writing to /tmp/Part3.philip.20180414.184607.327333/step-0-mapper_part-00000
Counters from step 1:
(no counters found)
writing to /tmp/Part3.philip.20180414.184607.327333/step-1-mapper_part-00000
Counters from step 2:
(no counters found)
writing to /tmp/Part3.philip.20180414.184607.327333/step-2-mapper_part-00000
Counters from step 3:
(no counters found)
writing to /tmp/Part3.philip.20180414.184607.327333/step-2-mapper-sorted
> sort /tmp/Part3.philip.20180414.184607.327333/step-2-mapper_part-00000
writing to /tmp/Part3.philip.20180414.184607.327333/step-2-reducer_part-00000
Counters from step 3:
(no counters found)
Moving /tmp/Part3.philip.20180414.184607.327333/step-2-reducer_part-00000 -> /tmp/P
Streaming final output from /tmp/Part3.philip.20180414.184607.327333/output
"The mean is:" 6.111111111111111
removing tmp directory /tmp/Part3.philip.20180414.184607.327333
philip@philip-VirtualBox:~/Desktop/C1766249/part3$
```

Figure 3.2

Part4

In Part 4, we were asked to write a MapReduce program which would find all paths of length two in the corresponding URLs in the text file. The answer should have been a triples of URLs, like (u,v,w) where there was a link from u to v and v to w.

1. In the same way as in previous parts open the terminal/cmd, in the directory containing folder **'part4'** either by using the command **'cd ./C1766249/part4'** or by right-clicking inside folder **'part4'** and then left-click **'Open in Terminal'**.
2. Therefore, write and execute **"python Part4.py csv4.txt"**. Always have the python file first, followed by the desired text file when running a MapReduce program.