Big Data For Engineers – Exercises

Spring 2020 – Week 2 – ETH Zurich

Debug

TL;DR You might encounter a bug with some package, if you are already familiar with debugging you can just go to solution 1 and fix it.

Did you get an ugly error?
Let's try to debug it together!

In Python when something goes wrong an exception is thrown.
If the programmer predicts that this might happen, a solution or a workaround can be implemented to fix the issue (this is called "catching an exception").
If the program is not able to deal with this unexpected situation, it will stop and print a traceback of the exception, so we can usually find which point of the code caused the exception.
At the very bottom of it we can find the exception name and usually a small string describing the problem.

In our case we should have something like
`AzureException: Could not find a suitable TLS CA certificate bundle, invalid path: /home/nbuser/anaconda3_420/lib/python3.5/site-packages/requests/cacert.pem`

So basically something somewhere is not finding something else.

Now what to do?
1) Go bother the computer scientist in your company (or your nerdy friend) and ask for his help **nice option :)**
2) Google the error, find people that had the same or similar problem (stack overflow is your best friend) and come up with a solution **computer scientist option :(**

Only for today we'll try the **computer scientist option :(**

If we google the first part of the error string we can find a Stack Overflow post with an explanation here.
This might be promising, so we go back to our traceback to see if it make sense.

```
~/anaconda3_420/lib/python3.5/site-packages/requests/adapters.py in cert_verify(self, conn, url, verify, cert)
    224                 cert_loc = extract_zipped_paths(DEFAULT_CA_BUNDLE_PATH)
--> 225
```

We see that this might actually be exactly our problem.
Apparently a required file saved under the DEFAULT_CA_BUNDLE_PATH path is missing because of some bug.
Combining the solution we can deduct where our wanted file is and tell it to the program.

Solution 1

We can copy the missing file into the right location:

1. Go to the start page of your Azure Notebooks project (the one from which you can upload and open the notebooks)
2. Click on "Terminal" in the top right part of the page
3. This will open a bash terminal on the virtual machine in which your notebook is running
4. From the previous error string and the Stack Overflow post we can find that the missing file is `cacert.pem`
5. Move to the site-packages folder with `cd anaconda3_420/lib/python3.5/site-packages`
6. Copy the file from its current location to the required one with `cp certifi/cacert.pem requests`
7. Try again the following code

Solution 2

Sometimes it may happen that you don't have the permission to see or move around a file in the terminal, so you can modify the path used by the module using Python code as shown in the first solution on Stack Overflow. Feel free to try out this solution as well.
The required absolute path is something similar to `/home/nbuser/anaconda3_420/lib/python3.5/site-packages/certifi/cacert.pem`.