

Week 09

How Does The Internet Work? Python, cURL and a REST API

Melvyn Ian Drag

October 28, 2019

Abstract

This Lecture is preamble to the next one in which we will set up an apache webserver. We'll take a peek at Python, revisit cURL, and make some web requests.

1 Exam Make Up

Folks who missed the exam last week can come early and work until 7:15 to get it done. Lecture starts promptly at 7:15.

2 Python Webserver

2.1 What is Python?

Python is a scripting language kind of like bash or javascript or perl - but different. It's the favorite language of legions of programmers. Python is used everywhere from scientific computing - I used it to solve PDEs in grad school - to game development, to system administration, to excel spreadsheet modifying, to a million other use cases. Today we're going to look at Python as a language for making websites and serving webpages.

2.2 What is a webserver?

As we discussed before, when type a website into a browser (like Chrome or Firefox, or the like) and hit *ENTER*, you ask a webserver for some information. The webserver responds with the information you requested and you see a webpage. This type of transaction is called a **GET** request when performed over the **Hyper Text Transmission Protocol (HTTP)** or the related **HTTPS**, which is the secure flavor of HTTP. We'll discuss what we mean by 'secure' in a few weeks.

Another thing you do when talking to a website is give some data to a webserver. This happens when you type your First Name, Last Name, User Name, phone number, password, etc. into a form on a website and click *Submit*. When you click the button, the browser sends a **POST** request to the webserver that is listening to you from some other corner of the earth.

In the case of either a GET or POST request, the request goes to a webserver, which is a computer far far away from you that has some code running on it that will respond to your GETs and your POSTs. That code is very often written in Java. It is frequently written in Python. Many write that code in Ruby. Many others use Scala to write the web server code. Recently, people started using Javascript to write the server code.

In this class we will write the server code in Python.

2.3 Install Python 3.

Install it with

```
1 user@machine$ sudo apt install python3-dev
```

2.4 Don't install Python 2

Python 2 is being deprecated next year. The Python language maintainers have been telling the Python community to move to Python3 for about 10 years now. If you are new to Python, use Python 3. If you are old to Python, hurry up and port your code to Python 3!

2.5 Simple Python Webserver

Okay, so now that you have Python installed, we can set up a simple web server and begin to make some HTTP GET and SET requests!

Here is the code:

```
1  """
2  Code stolen here:
3
4  https://gist.github.com/bradmontgomery/2219997#file-dummy-web-server-py
5
6  and then modified to be more directly relatable to our lecture.
7  """
8
9  from http.server import HTTPServer, BaseHTTPRequestHandler
10
11 class MyFirstServer(BaseHTTPRequestHandler):
12     def _set_headers(self):
13         self.send_response(200)
14         self.send_header("Content-type", "text/html")
15         self.end_headers()
16
17     def _html(self, message):
18         content =
19             "<html>\n\t<body>\n\t\t<h1>{}</h1>\n\t</body>\n</html>\n".format(message)
20         return content.encode("utf8") # NOTE: must return a bytes object!
21
22     def do_GET(self):
23         self._set_headers()
24         self.wfile.write(self._html("Received GET!"))
25
26     def do_POST(self):
27         content_length = int(self.headers['Content-Length'])
28         post_data = self.rfile.read(content_length)
29
30         self._set_headers()
31         self.wfile.write(self._html("Received POST!" + post_data.decode("utf8")))
32
33 if __name__ == "__main__":
34     addr = "localhost"
35     port = 80
36     server_address = (addr, port)
37     httpd = HTTPServer(server_address, MyFirstServer)
38
39     print("Starting httpd server on {}:{}".format(addr, port))
40     httpd.serve_forever()
```

This code is available in the class repository under the name *Server.py*. From just this simple little example, we can learn a lot about how the internet works. As we survey the example, we can issue some curl requests and verify that they work.

We can run this code from the command line with:

```
1 user@machine$ python3 Server.py
2 Starting httpd server on localhost:8000
```

So we see that our little baby website is running on "localhost" at port 8000. Standard HTTP applications run on port 80, but for this demo app we are running on port 8000. Just for reference - HTTPS applications run (by default) on port 443. And ssh is over port 22. And postgresql is on port 5432. So now you know 4 special ports!

A summary of what we know about ports so far:

- port 22 - ssh
- port 80 - HTTP
- port 443 - HTTPS
- port 5432 - PostgreSQL

In the webdevelopment world, we make **GET** and **POST** requests to webserver that have IP Addresses. But we rarely refer to them by their ip address - typically we use a domain name like 'www.google.com' instead of a number address. We are doing the same thing here - we name our local webserver "localhost" instead of referring to it by its numeric name - 127.0.0.1

2.6 A quick look at the code

Note that there are methods for handling **POST** and **GET** requests. This is characteristic of code running on a webserver. The code will have methods for handling different types of HTTP requests. You'll also see a few lines of code for setting up an HTTP server to run on 'localhost:8000'. There's a line that does something or other with "headers", and we'll look into those in a minute. And there's a function that does something with html. If you've never seen HTML before, that's what it looks like. It's a markup language. If you don't know what HTML is, you can learn it in just a few minutes, it's pretty simple. It structures a document between some brackets. HTML always looks something like this:

```
1 <html>
2 <body>
3   <h1> This is a title </h1>
4   <h2> This is a subtitle </h2>
5   <div> This is a paragraph about something </div>
6 </body>
7 </html>
```

2.7 Verify Website Works With A Browser

To verify that our website is running, open a browser window (chrome, firefox, etc) and type "localhost:8000". You should get a response like this:

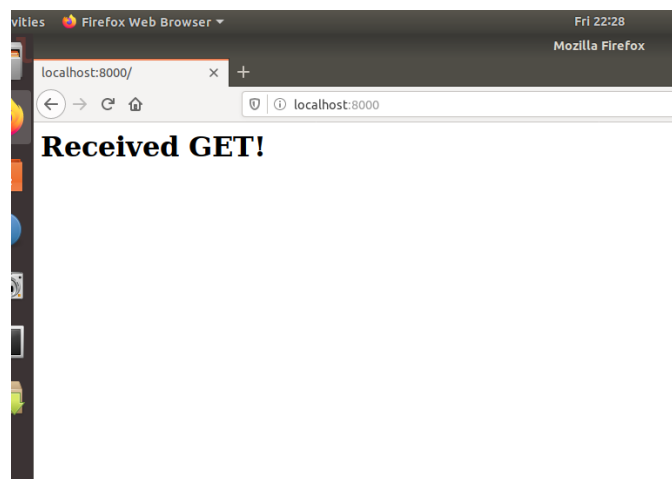


Figure 1: GET request in browser to localhost:8000

So we know our website "localhost:8000" is running.

And as I mentioned before, “localhost” is an alias for “127.0.0.1”, which we can readily verify by going to “127.0.0.1:8000” while our little application is running.

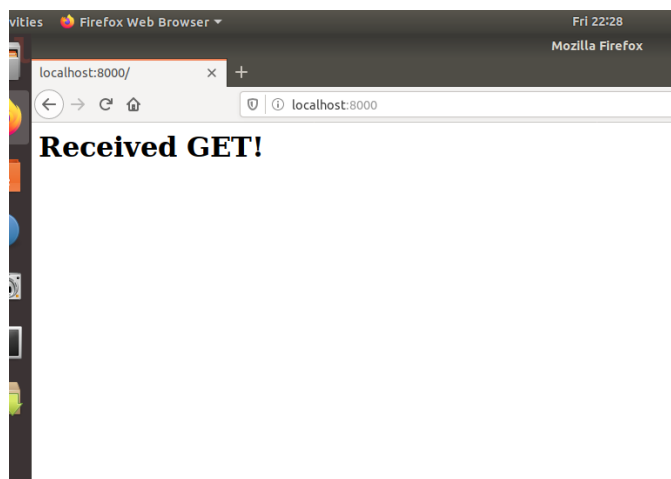


Figure 2: GET request in browser to 127.0.0.1:8000

One more thing of interest - in case you’ve never seen this happen before - is you can see the Server code handling requests on the server side. Go back to the terminal window where the webserver is running and you’ll see it logging output after handling every request.

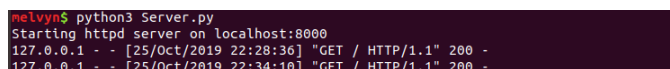


Figure 3: See the webserver handling the GET request traffic.

You may be frustrated because something doesn’t match up with your expectations here. We see a website in a browser, but how come we don’t type “www.something.com”??? What is this “localhost” nonsense? What do you mean it’s running on port 8000??? I thought you said something or other about port 80 being for http??? All this understanding will come with time. For now we just need to know that websites handle GET and POST requests (and other types of requests too). And we’ve seen a little bit of code that features a get and post handler, and we’ve verified that (the GET at least) works as expected.

Now enough fooling around! Let’s use cURL to test out our little website.

3 More About Ports

Some ports are specified for a specific purpose by the os and may require special permissions to use. For example, port 80 requires root access to use. You can try to change the port 8000 in the sample code to port 80 and you will see a permission denied error unless you use sudo.

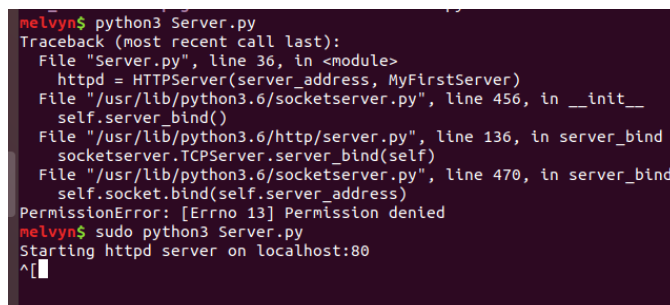


Figure 4: Need Root Access for Port 80

4 cURL

cURL is a Linux tool that has been around for about 20 years. It's used for making requests. It can be thought of as a command line browser, in some sense. The things you usually do in a browser, like typing an address into a search bar, entering data into a form on a web page, clicking a button - these things can be done on the command line with cURL (mostly).

4.1 Exercise #1

Open another terminal window in addition to the one running your Python webserver. Then enter these commands:

```
1 user@machine$ curl localhost:8000
2 # returns some html from the website
3 user@machine$ curl -X GET localhost:8000
4 # returns the same thing, because the default curl action is GET
5 user@machine$ curl -X POST -d"myname=melvyn" localhost:8000
6 # returns the POST response, along with the data you passed in
7 user@machine$ curl -X POST --verbose -d"myname=melvyn" localhost:8000
8 # returns the same thing, but this time with lots of extra information!
9 # this time you can see what you sent over via the POST in detail
10 # along with a detailed response from the server
11 # The lines with a > are what you sent.
12 # The lines with a < are what came back.
13 user@machine$ curl -X POST --verbose -d"myname=melvyn&mylastname=drag" localhost:8000
14 # Try sending more and more data values by separating them with an ampersand.
```

The '-X' flag is used for specifying the HTTP verb associated with your request. As you've seen, the webserver will contain code for handling different types of requests.

The '-d' flag is used for passing data. Typically we associate the '-d' flag with POST requests and not GET requests.

4.2 Headers

As with most other data transmission protocols, file formats, etc. - most computer data formats - there is a header, followed by a body. What are HTTP headers?

4.3 Congratulations!

You may have just run your first webserver! If this wasn't your first webserver project, I hope it was fun anyway.

4.4 HTTP Return Codes

There are many different return codes. When you send a request to a webserver, it responds

5 A Python Programming Interlude

5.1 Intro and Using Github to Test

What is HTTP/HTTPS?

A protocol used on the internet. You send some data to a server, the server responds with some data. As you will see with many data transmission technologies with computers, there is a header and a body. The header typically has meta data and bytes required by certain bits of computer hardware to process the body. Then the body contains the actual information.

HTTP/HTTPS work with a few verbs. We'll just focus on two of them for now, but there are (at least) 4.

Make some get, post requests.

We're going to use this tutorial here: <https://gist.github.com/joyrexus/85bf6b02979d8a7b0308>

Do one with authentication and a REST API.

5.2 What kind of headers does curl take?

What kind of header does it take?

Can we do two factor authentication with the REST API?

6 Recap of what you know

1. A bit about the Bash programming language
2. grep & regular expressions
3. What a user is on Linux
4. What a group is on Linux
5. What is git?
6. You're comfortable using a command line interface now
7. There are a few different languages on the command line - we've used bash and dash
8. What permissions are and how to modify them
9. What is a root user
10. What is a process
11. What is a job
12. How to send signals to processes (kill, CTRL+C, CTRL+Z)
13. How to make code ignore or block signals
14. How to install software on Linux
15. A cool trick for using setuid / setgid to make a non-root user do some root stuff
16. basics of relational dbs and how to configure one on Linux
17. What is cron
18. curl
19. some vocabulary like API, REST, regex, SQL, DB
20. A bit about python programming

7 Coming Up In This Class

We have a few more important things to cover

1. set up apache webserver
2. set up a git server
3. Linux and Text encodings
4. xxd, everyone's favorite binary packet analyzer
5. The AWK programming language
6. The hows and whys of formatting harddrives, usb sticks, solid state drives, etc.
7. Encryption with gpg + pgp. How it relates to ssh and other pub/priv key schemes.