

# Week 09

## How Does The Internet Work? Playing with Python3 and cURL

Melvyn Ian Drag

October 27, 2020

### 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 Goal today

- add GET and POST to your vocabulary
- know that http requests have headers and bodies
- know how to do GET and POST with curl
- know how to send data with curl and POST
- know that Python is whitespace sensitive
- be able to complete a mildly interesting homework assignment.
- know that localhost and 127.0.0.1 are synonyms ( pseudonyms? )

## 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 Set up machine

Install Python3.

```
1 root@machine$ apt install python3-dev
```

Also install a couple of other things, create a new sudoer, and download the class notes.

```
1 root@machine$ apt install git curl # we'll need git and curl tonight
2 root@machine$ adduser yourname
3 root@machine$ usermod -a -G yourname sudo # or sudo yourname, I can't remember
4 the order
5 root@machine$ su - yourname
6 yourname@machine$ git clone https://github.com/melvyniandrag/LinuxClassRepo.git
```

The code we'll use tonight is in the class repo.

## 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 I Hate This Lecture Because This is a Linux Class, Not a Python Class

*Why are you wasting time teaching Python in a Linux class!!!!!!? I didn't sign up for this!!*

We're using Python in this class for this webserver activity because I think it is hand's down the easiest language to read and write. Since I don't have time to teach you Python right now, I'm just using Python code because you can more or less skim it and understand what it does. We're only using Python for about the next three hours, so I'm just going to give you the bare minimum information needed to understand what's happening.

One thing that is important to understand is that Python is a whitespace sensitive language. Whereas other languages use curly braces to denote logical blocks in code - for example, ask class where the *main* and *sayHi* functions , and the *MyClass* class, start and ends in this Java program:

```
1 public class MyClass{
2     public static void sayHi() [
3         System.out.println("Hi");
4     ]
5
6     public static void main(String[] args){
7         sayHi();
8     }
9 }
```

In python, the situation is different:

```
1 def sayHi():
2     s = "Hi"
3     print(s)
4
5 if __name__ == "__main__":
6     sayHi()
```

Note the lack of curly braces. Write this code on your machine and run it as;

```
1 user@machine$ python3 whateverYouNamedTheFile.py
```

Then, change the space before the 's' to 4 spaces and the space before the 'print' to a tab. Try to run the code again and it will fail.

They will look the same, but they are not the same on the byte level. more on that at the end of lecture if there is time.

Python expects either all tabs or all spaces. If you mix tabs and spaces in your code, Python get's mad.

As I've said before, there is a lot of bickering in the tech community about everything. Some people hate python because of the whitespace <https://news.ycombinator.com/item?id=1463451>

I really like Python. I think it's the easiest language to write, there are many great libraries for math, science, system administration, web development, graphics, etc.. The community is great! There are conferences all year long all around the world. And I like not having to waste space in my code with curly braces. Different strokes.

**TL;DR You should learn Python. Not now - you're too busy. Next year you should learn Python.**

## 2.6 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 = "<html>\n\t<body>\n\t\t<h1>{}\n\t\t</h1>\n\t\t<p> Hi
19                  </p>\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 = "127.0.0.1"
35      #addr = "localhost"
36      port = 8000
37      server_address = (addr, port)
38      httpd = HTTPServer(server_address, MyFirstServer)
39
40      print("Starting httpd server on {}:{}".format(addr, port))
41      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
```

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 web servers 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

## 2.7 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>

```

## 2.8 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:

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.

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.

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.

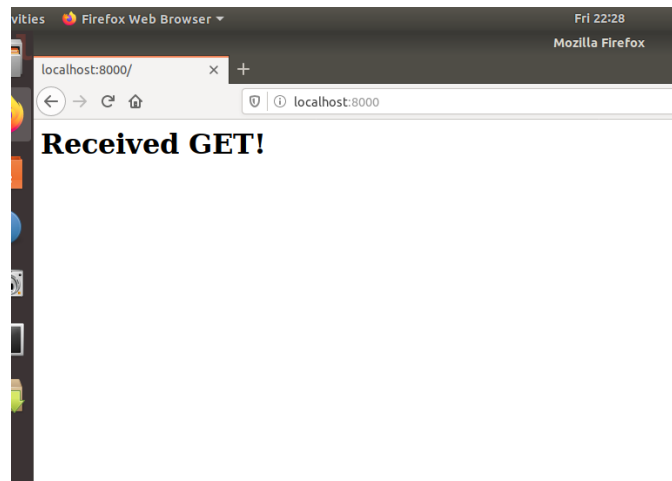


Figure 1: GET request in browser to localhost:8000

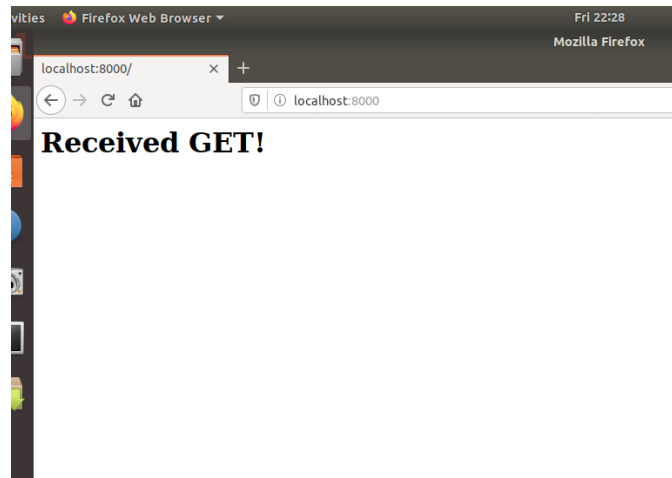


Figure 2: GET request in browser to 127.0.0.1:8000

### 3 Two Exercises

**#1 : Change port 8000 in the code to port 80 and try to rerun.**

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.

**#2 : Change 'localhost' in the code to '127.0.0.1'. Note that it runs as expected - localhost and 127.0.0.1 are the same thing.**

If ip addresses interest you, like, why was '127.0.0.1' selected as 'localhost'? Are there other ip addresses that were designated as local ip addresses? Then you should learn more about computer networking! I admit that I'm very interested in the subject, but haven't studied it enough due to other responsibilities and ambitions.

### 4 10 Minute Break

We'll take a 10 minute break now in case you need a coffee or to go to the bathroom. If you don't mind continuing to work, I encourage you to change the GET request in your little website so that your browser displays the HTML from section 2.7. If you can do that, try to modify it some more with some other HTML tags.

```
melvyn$ python3 Server.py
Starting httpd server on localhost:8000
127.0.0.1 - - [25/Oct/2019 22:28:36] "GET / HTTP/1.1" 200 -
127.0.0.1 - - [25/Oct/2019 22:34:10] "GET / HTTP/1.1" 200 -
```

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

```
melvyn$ python3 Server.py
Traceback (most recent call last):
  File "Server.py", line 36, in <module>
    httpd = HTTPServer(server_address, MyFirstServer)
  File "/usr/lib/python3.6/socketserver.py", line 456, in __init__
    self.server_bind()
  File "/usr/lib/python3.6/http/server.py", line 136, in server_bind
    socketserver.TCPServer.server_bind(self)
  File "/usr/lib/python3.6/socketserver.py", line 470, in server_bind
    self.socket.bind(self.server_address)
PermissionError: [Errno 13] Permission denied
melvyn$ sudo python3 Server.py
Starting httpd server on localhost:8000
^[]
```

Figure 4: Need Root Access for Port 80

## 5 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 ).

### 5.1 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.
15 user@machine$ curl -X POST -H "Content-Type: application/text" -H "Content-Length: 10"
    -d"myname=melvyn" localhost:8000
16 # Look what happened due to Content-Length, play around with the length to see what else
    happens
17 # also check and see if anything happens if you change Content-Type to application/json
    or application/x-www-form-urlencoded
18 user@machine$
```

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.

### 5.2 Headers

As with most other data transmission protocols, file formats, etc. - most computer data formats - there is a header, followed by a body. Whenever you send an HTTP request, some header data goes along with it. Two of the more

common ones you'll see are:

1. Content-Length - how long the data being passed is
2. Content-Type - what type of data is being sent? Json? Text? HTML? Other?

### 5.3 HTTP Return Codes

There are many different return codes. When you send a request to a webserver, it responds with some header data, a body, and also, a return code. Common return codes that you may have seen already are:

1. 200 (OK) - The 200 return code is when you HTTP transaction succeeds
2. 404 (Not Found) - Happens when you request something from the internet that is not there
3. 500 ( Internal Server Error ) - Generic message for when something has gone wrong with the server

To see a 200, type

```
1 curl --verbose https://www.google.com 2>&1 | grep "^< HTTP"
```

To see a 404 type

```
1 icurl --verbose https://www.google.com/doesntExist 2>&1 | grep "^< HTTP"
```

To see a 500 - I'm not sure how to force this error off the top of my head.

More information about return codes can be found here: [https://en.wikipedia.org/wiki/List\\_of\\_HTTP\\_status\\_codes](https://en.wikipedia.org/wiki/List_of_HTTP_status_codes)

I said I'm not sure how to force the 500 error because I'm not an experience web programmer - but I do know a thing or two about Linux. Did you all understnad the piped commands I used to generate the status codes?

Two interesting things:

- 2>&1 means pass stderr to stdout. Did this because curl seems to dump to stderr.
- grep "**<HTTP**" means find a line starting with "**<HTTP**"

### 5.4 Congratulations!

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

As I said, the purpose of this assignment is not for you to become a webdeveloper, but just for you to know a couple of things so that when we set up an Apache webserver, it's not totally foreign to you.

## 6 A Real World Curl Example

### 6.1 Notes to self:

Need to find a good example from in here: <https://docs.github.com/en/free-pro-team@latest/rest> as the example from previous semesters doesn't work anymore. They have deprecated the api.

In fact, here are some great activities: <https://docs.github.com/en/free-pro-team@latest/rest/guides/getting-started-with-the-rest-api>

In FACT I think homework will be to create an issue on a github repo using the API. Have a look here <https://docs.github.com/en/free-pro-team@latest/rest/guides/getting-started-with-the-rest-api> and note the curl command does not look like a post, it looks like a get. Weird!

## 6.2 Some exercises in class

You will encounter REST APIs in your career. Many popular web platforms like twitter, reddit, facebook and github offer REST APIs that we can use as programmers to get data from and send data to the website without having to use a browser. You might use this in scripts.

For example, in my short career as a datascientist I spent a lot of time writing bash scripts that used curl to fetch data from certain platforms and then some other bash tools to send the data off to some server or other for further mathematical analysis. You might well do something like this.

Today we'll just see a real world REST API and do some stuff to it with curl.

List all my repos - note you can only see the public ones. (homework idea - have students write a script to count how many repos are returned by this command)

```
1 $ curl -i https://api.github.com/users/melvyniandrag/repos
```

Create an oauth token with class. More specifically create a personal access token <https://docs.github.com/en/free-pro-team@latest/github/authenticating-to-github/creating-a-personal-access-token> Then run this command to create a repo:

## 7 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



## 8 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. add a gitlab front end to the git server
4. Linux and Text encodings
5. xxd, everyone's favorite binary packet analyzer
6. The AWK programming language
7. The hows and whys of formatting harddrives, usb sticks, solid state drives, etc.
8. Encryption with gpg + pgp. How it relates to ssh and other pub/priv key schemes.
9. sed