

< (../02-ssh/)

Extra Unix Shell Material (../)

> (../04-permissions/)

Transferring Files

? Overview

Teaching: 10 min

Exercises: 0 min

Questions

- How to use wget, curl and lftp to transfer file?

Objectives

- FIXME

There are other ways to interact with remote files other than git.

It is true that we can clone an entire git repository, or even one level of a git repository using: `git clone --depth-1 repository_name .` What about files that do not exist in a git repository? If we wish to download files from the shell we can use tools such as Wget, cURL, and lftp.

Wget

Wget is a simple tool developed for the GNU Project that downloads files with the HTTP, HTTPS and FTP protocols. It is widely used by Unix-like users and is available with most Linux distributions.

To download this lesson (located at <http://swcarpentry.github.io/shell-extras/03-file-transfer.html>) from the web via HTTP we can simply type:

```
$ wget http://swcarpentry.github.io/shell-extras/03-file-transfer.html
```

```
--2014-11-21 09:41:31--
http://swcarpentry.github.io/shell-extras/03-file-transfer.html
Resolving software-carpentry.org (software-carpentry.org)... 174.136.14.108
Connecting to software-carpentry.org (software-carpentry.org)|174.136.14.108|:80... connected.
HTTP request sent, awaiting response... 200 OK
Length: 8901 (8.7K) [text/html]
Saving to: '03-file_transfer.html'

100%[=====>] 8,901      --.-K/s   in 0.05s

2014-11-21 09:41:31 (187 KB/s) - '03-file_transfer.html' saved [8901/8901]
```

Alternatively, you can add more options, which are in the form:

```
wget -r -np -D domain_name target_URL
```

where `-r` means recursively crawl to other files and directories, `-np` means avoid crawling to parent directories, and `-D` means to target only the following domain name

For our URL it would be:

```
$ wget -r -np -D software-carpentry.org http://swcarpentry.github.io/shell-extras/03-file-transfer.html
```

To restrict retrieval to a particular extension(s) we can use the `-A` option followed by a comma separated list:

```
wget -r -np -D software-carpentry.org -A html http://swcarpentry.github.io/shell-extras/03-file-transfer.html
```

We can also clone a webpage with its local dependencies:

```
$ wget -mkq target_URL
```

We could also clone the entire website:

```
$ wget -mkq -np -D domain_name domain_name_URL
```

and add the `-nH` option if we do not want a subdirectory created for the websites content:

e.g.

```
$ wget -mkq -np -nH -D example.com http://example.com
```

where:

`-m` is for mirroring with time stamping, infinite recursion depth, and preservation of FTP directory settings `-k` converts links to make them suitable for local viewing `-q` supresses the output to the screen

The above command can also save the clone the contents of one domain to another if we are using ssh or sshfs to access a webserver.

Please refer to the man page by typing `man wget` in the shell for more information.

cURL

Alternatively, we can use `cURL` . It supports a much larger range of protocols including common mail based protocols like pop3 and smtp.

To download this lesson (located at <http://swcarpentry.github.io/shell-extras/03-file-transfer.html>) from the web via HTTP we can simply type:

```
$ curl -o 10-file_transfer.html http://swcarpentry.github.io/shell-extras/03-file-transfer.html
```

% Total	% Received	% Xferd	Average Speed	Time	Time	Time	Current
			Dload Upload	Total	Spent	Left	Speed
100 14005	100 14005	0 0	35170 0	--:--:--	--:--:--	--:--:--	105k

This input to curl is in the form:

```
curl -o filename_for_local_machine target_url
```

where the `-o` option says write the output to a file instead of the stdout (the screen), and `file_name_for_local_machine` is any file name you choose to save to the local machine, and `target_URL` is where the file is the URL where the file is on the web

Removing the `-o` option, and following the syntax `curl target_URL` outputs the contents of the url to the screen. If we wanted to enhance the functionality we have we could use information from the pipes and filters section, which is lesson 4 from the unix shell session.

For example, we could type `curl http://swcarpentry.github.io/shell-extras/03-file-transfer.html | grep curl` which would tell us that indeed this URL contains the string curl. We could make the output cleaner by limiting the output of curl to just the file contents by using the `-s` option (e.g. `curl -s http://swcarpentry.github.io/shell-extras/03-file-transfer.html | grep curl`).

If we wanted only the text and not the html tags in our output we could use html to text parser such as `html2text`.

```
$ curl -s http://swcarpentry.github.io/shell-extras/03-file-transfer.html | html2text | grep curl
```

With `wget`, we can obtain the same results by typing:

```
$ wget -q -D swcarpentry.github.io -O /dev/stdout http://swcarpentry.github.io/shell-extras/03-file-transfer.html | html2text | grep curl
```

`wget` offers more functionality natively than `curl` for retrieving entire directories. We could use `wget` to first retrieve an entire directory and then run `html2text` and `grep` to find a particular string. `curl` is limited to retrieving one or more specified URLs that cannot be obtained by recursively crawling a directory. The situation may be improved by combining with other unix tools, but is not thought as being as good as `wget`.

Please refer to the man pages by typing `man wget`, `man curl`, and `man html2text` in the shell for more information.

lftp

Another option is `lftp`. It has a lot of capability, and even does simple bittorrent.

If we want to retrieve `03-file-transfer.html` on the website and save it with the filename `03-file-transfer.html` locally:

```
$ lftp -c get http://swcarpentry.github.io/shell-extras/03-file-transfer.html
```

If we want to print `03-file-transfer.html` to the screen instead:

```
$ lftp -c cat http://swcarpentry.github.io/shell-extras/03-file-transfer.html
```

To obtain retrieve all of the files with a particular extension in a directory we can type:

```
$ lftp -c mget {URL for directory}/*.extension_name
```

For example, to retrieve all of the `.html` files in the `extras` folder:

```
$ lftp -c mget http://swcarpentry.github.io/shell-extras/*.html
```

Please refer to the man page by typing `man lftp` in the shell for more information.

❗ Key Points

- FIXME

◀ (../02-ssh/)

➤ (../04-permissions/)

Copyright © 2016 - 2019 Software Carpentry Foundation (<https://software-carpentry.org>)

Edit on GitHub (https://github.com/swcarpentry/shell-extras/edit/gh-pages/_episodes/03-file-transfer.md) / Source (<https://github.com/swcarpentry/shell-extras/>) / Contributing ([../contributing/](https://github.com/swcarpentry/shell-extras/blob/master/CONTRIBUTING.md)) / Cite ([../citation/](https://github.com/swcarpentry/shell-extras/blob/master/CITATION.md)) / Contact (<mailto:lessons@software-carpentry.org>)