

Tutorial: Intro to Git/GitHub

Tutorial Lead



FERNANDO PEREZ

Technical support



JANE KOH



ANTHONY ARENDT

Teaching Assistants



SEBASTIAN ALVIS



SHANE GRIGSBY



DANIEL SHAPERO
PRINCIPAL RESEARCH SCIENTIST
APL



YARA MOHAJERANI



DAVID SHEAN



JESSICA SCHEICK

JupyterHub and Pangeo

What is JupyterHub?

*Host pre-configured data science environments
on shared infrastructure*

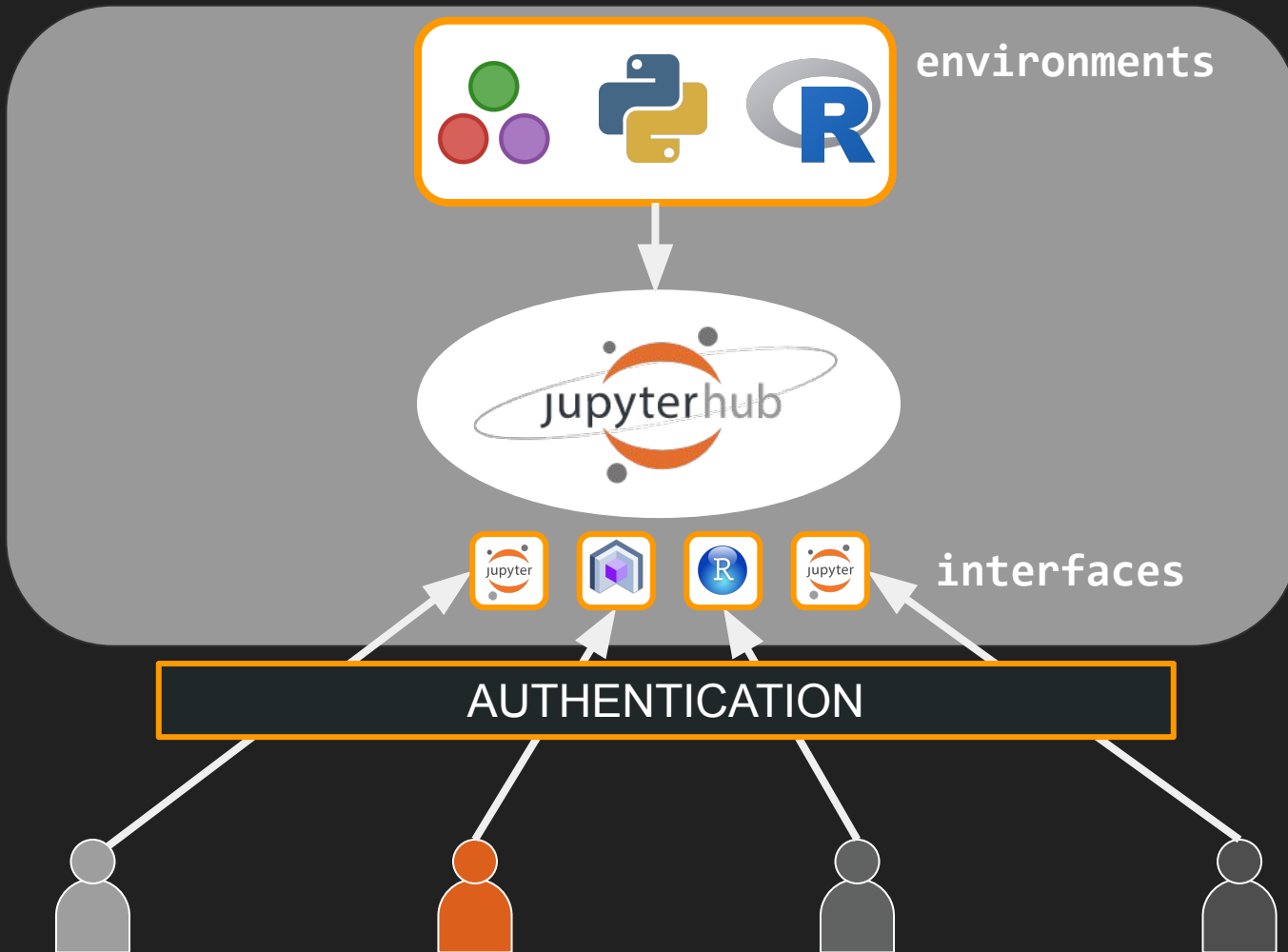


jupyter.org/hub



My fancy machine
in the cloud
or HPC center







Harnessing the power of cloud computing to study the whole Earth interactively

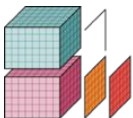


Interactivity



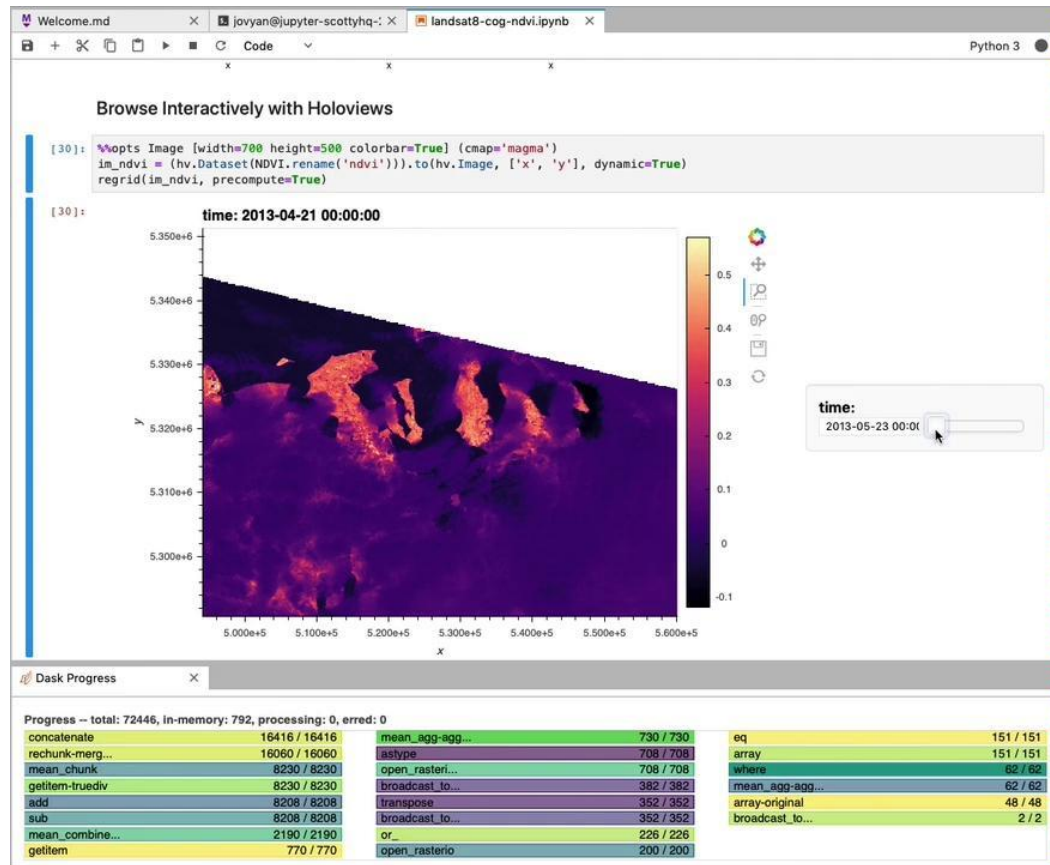
DASK

Distributed computing



xarray

Data models / numerics



Scott Henderson

[Follow](#)

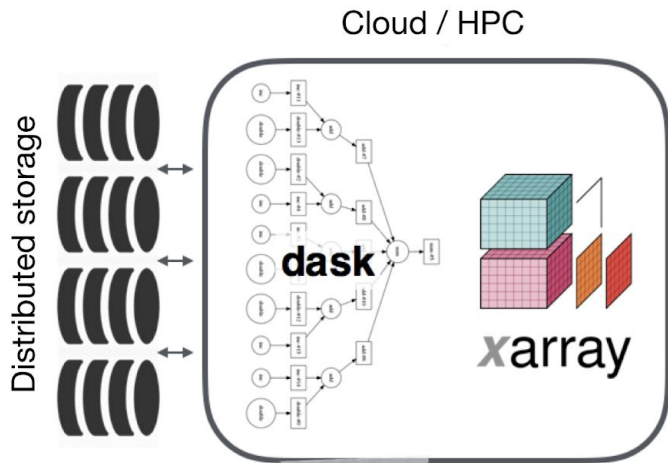
Research geophysicist at University of Washington eScience Institute

Oct 1 · 7 min read



Pangeo Architecture

“Analysis Ready Data”
stored on globally-available
distributed storage.



Parallel computing system allows
users deploy clusters of compute
nodes for data processing.

Dask tells the nodes what to do.

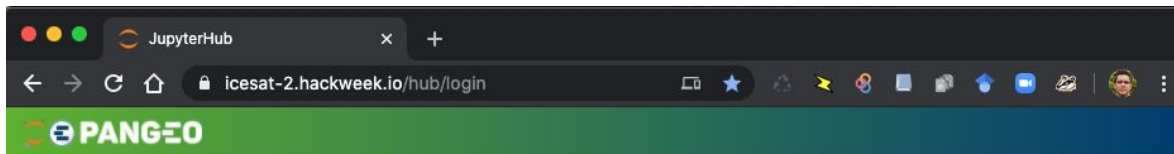
Jupyter for interactive
access remote systems

end user

web browser

Xarray provides data structures
and intuitive interface for
interacting with datasets

Logging into Pangeo: <https://icesat-2.hackweek.io>




A COMMUNITY HUB FOR THE PANGEO PROJECT

Welcome to icesat-2.hackweek.io, the computational environment for [Icesat-2 Hackweek](#)! This hub lives in AWS region [us-west-2](#). It is maintained by the [Pangeo project](#) and is supported by [NASA Grant #17-ACCESS17-0003](#) and cloud credits from Amazon. **This is a prototype and should be treated accordingly. We make no promises that the hub will remain active. Do not store passwords or sensitive data in your home directory** Access is currently limited to members of the [Pangeo GitHub Organization](#) and the [IceSat-2 Hackweek Organization](#). To provide feedback and report any technical problems, please use the [github issue tracker](#).

Sign in with GitHub

Once you authenticate with your GitHub ID

 Home Token

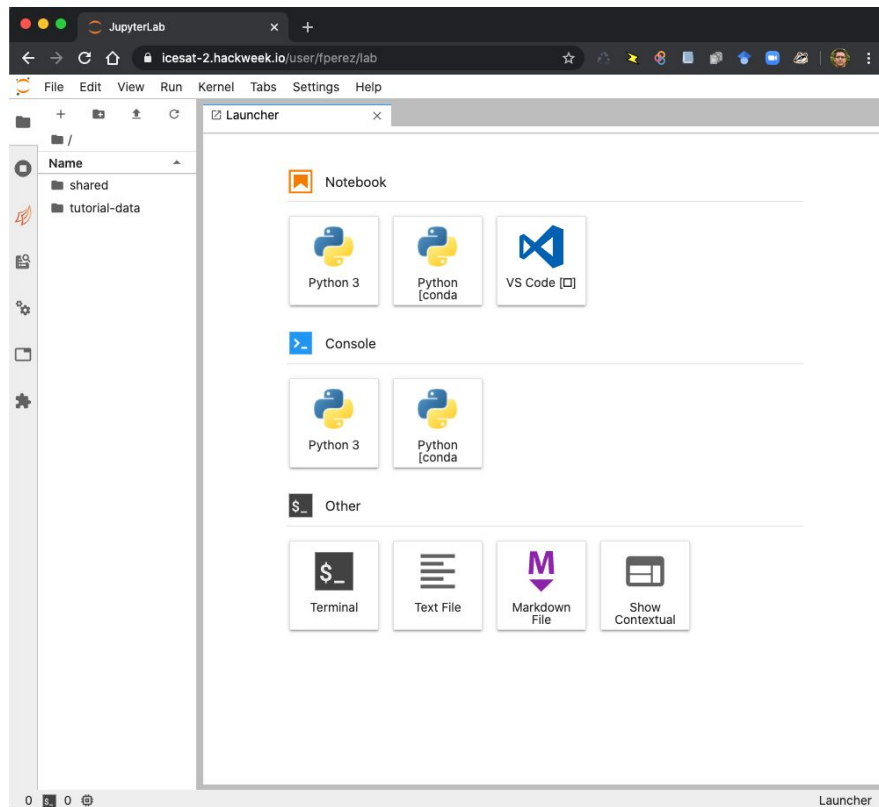
fperez Logout

Server Options

☒ IceSat-2 Hackweek 2020 Image

Start

The default JupyterLab interface



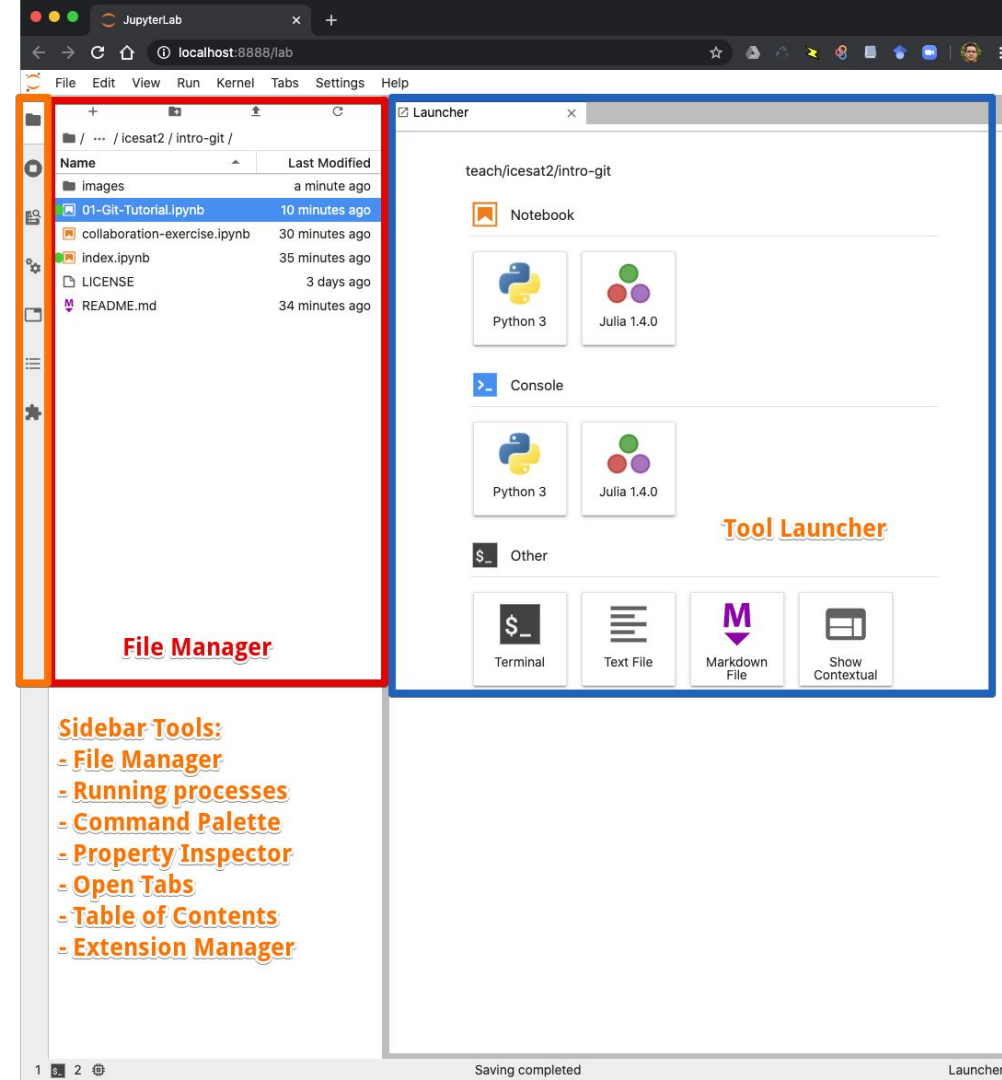
Pause: everyone OK so far?

Ping us on #questions if you hit any snags

JupyterLab

A modular, extensible Data Science Environment

All the tools for everyday scientific computing and data science, locally or remotely.



Name	Last Modified
images	19 minutes ago
01-Git-Tutorial.ipynb	seconds ago
collaboration-exer...	19 minutes ago
index.ipynb	25 minutes ago
LICENSE	3 days ago
README.md	24 minutes ago

01-Git-Tutorial.ipynb Python 3

Notebook with Python 3 kernel

Credit: ProGit book, by Scott Chacon, CC License.

A repository: a group of linked commits

Note: these form a Directed Acyclic Graph (DAG), with nodes identified by their hash.

A hash: a fingerprint of the content of each commit and its parent

```
[1]: from hashlib import sha1

# Our first commit
data1 = b'This is the start of my paper.'
meta1 = b'date: 1/1/17'
hash1 = sha1(data1 + meta1).hexdigest()
print('Hash:', hash1)

Hash: 3b32905baabd5ff22b3832c892078f78f5e5bd3b
```

01-Git-Tutorial.ipynb

Python 3.7.6 | packaged by conda-forge | (default, Jun 1 2020, 18:33:30)
Type 'copyright', 'credits' or 'license' for more information
IPython 7.15.0 -- An enhanced Interactive Python. Type '?' for help.

```
[3]: hash2

[3]: '1c12d2aad51d5fc33e5b83a03b8787dfadde92a4'

[4]: hash1

[4]: '3b32905baabd5ff22b3832c892078f78f5e5bd3b'
```

Console connected to above notebook kernel

index.ipynb README.md

Text editor with Markdown file open

```
6 ScneICK_
7 ## Schedule (all times in PDT)
8
9 *Date:* June 10, 9am
10
11 - 8:45 - 9:00 AM [Anthony]: Getting connected
12 - 9:00 - 9:05 AM [Fernando Pérez]: Introduction and Motivation
13 - 9:05 - 9:15 AM [Fernando Pérez]: JupyterLab Basics
14 - 9:15 - 9:25 AM [Fernando Paolo]: Command Line Basics
15 - 9:25 - 9:40 AM [Fernando Pérez]: Git Fundamentals
16 - 9:40 - 10:00 AM [Anthony]: Git repositories
17 - 10:00 - 10:25 AM [Daniel]: HackWeek Github Workflows
18 - 10:25 - 10:30 AM [Fernando Pérez]: Wrap up
19
20 ## Accessing the Pangeo Hub
21
22 https://icesat-2.hackweek.io/hub/login
23
24
25 ## Credit
26
27 This collection of tutorials was adapted from the [ICESat-2 HackWeek tutorial]
28 (https://github.com/ICESAT-2HackWeek/intro-jupyter-git) by [fperez]
29 (https://github.com/fperez)
```

Terminal 1

```
(base) (master)alpmayo[intro-git]> ls -la
total 240
drwxr-xr-x 11 fperez staff 352 Jun 2 13:13 ./
drwxr-xr-x 9 fperez staff 288 Jun 2 12:17 ../
drwxr-xr-x 14 fperez staff 448 Jun 2 13:13 .git/
-rw-r--r-- 1 fperez staff 20 Jun 2 13:04 .gitignore
drwxr-xr-x 5 fperez staff 160 Jun 2 13:04 .ipynb_checkpoints/
-rw-r--r-- 1 fperez staff 92993 Jun 2 13:06 01-Git-Tutorial.ipynb
-rw-r--r-- 1 fperez staff 1529 May 30 12:07 LICENSE
-rw-r--r-- 1 fperez staff 1006 Jun 2 13:08 README.md
-rw-r--r-- 1 fperez staff 11703 Jun 2 13:13 collaboration-exercise.ipynb
drwxr-xr-x 21 fperez staff 672 Jun 2 13:13 images/
-rw-r--r-- 1 fperez staff 1217 Jun 2 13:07 index.ipynb
(base) (master)alpmayo[intro-git]> git remote -v
origin git@github.com:ICESAT-2HackWeek/intro-git.git (fetch)
origin git@github.com:ICESAT-2HackWeek/intro-git.git (push)
(base) (master)alpmayo[intro-git]> git status
On branch master
Your branch is up to date with 'origin/master'.

Changes not staged for commit:
  (use "git add <file>..." to update what will be committed)
  (use "git restore <file>..." to discard changes in working directory)
        modified:   01-Git-Tutorial.ipynb
        modified:   README.md
        modified:   index.ipynb

no changes added to commit (use "git add" and/or "git commit -a")
(base) (master)alpmayo[intro-git]>
```

Terminal running on the same host

Let's grab our tutorial materials!

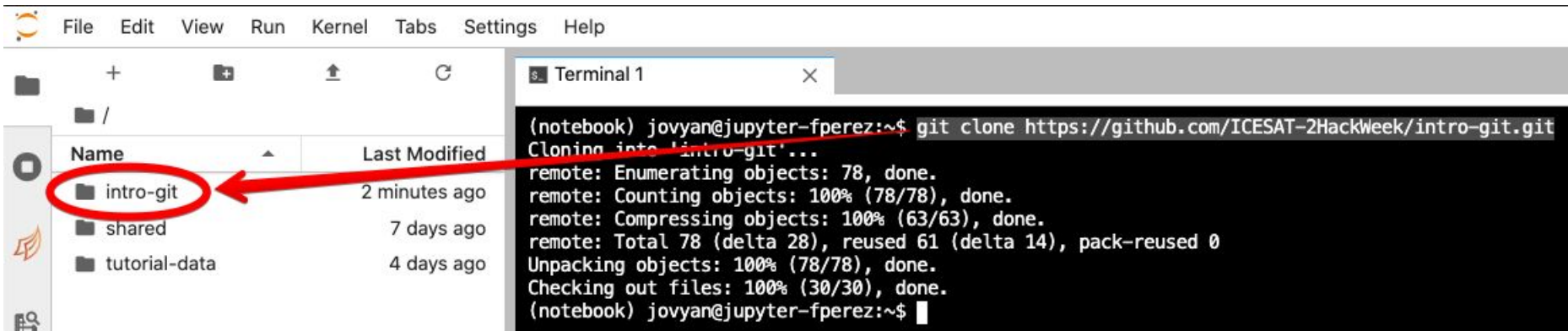
Location: <https://github.com/ICESAT-2HackWeek/intro-git>

In a terminal, type:

```
git clone https://github.com/ICESAT-2HackWeek/intro-git.git
```

If you had already done that, then in the intro-git directory type:

```
git pull
```



The screenshot shows a Jupyter Notebook interface. On the left, a file browser pane displays a directory structure with three items: 'intro-git' (highlighted with a red circle and a red arrow), 'shared', and 'tutorial-data'. The 'intro-git' directory is shown with a folder icon. The main area of the interface is a terminal window titled 'Terminal 1'. It displays the output of the command `git clone https://github.com/ICESAT-2HackWeek/intro-git.git`. The output shows the cloning process, including enumerating objects, counting objects, compressing objects, and unpacking objects, all completed successfully. The terminal prompt is `(notebook) jovyan@jupyter-fperez:~$`.

Name	Last Modified
intro-git	2 minutes ago
shared	7 days ago
tutorial-data	4 days ago

```
(notebook) jovyan@jupyter-fperez:~$ git clone https://github.com/ICESAT-2HackWeek/intro-git.git
Cloning into 'intro-git'...
remote: Enumerating objects: 78, done.
remote: Counting objects: 100% (78/78), done.
remote: Compressing objects: 100% (63/63), done.
remote: Total 78 (delta 28), reused 61 (delta 14), pack-reused 0
Unpacking objects: 100% (78/78), done.
Checking out files: 100% (30/30), done.
(notebook) jovyan@jupyter-fperez:~$
```