LINCC Science Platform

Release 0.1

Aidan, Aditi, and Paolo

Jul 03, 2022

CONTENTS:

An initiative of the LSST Corporation aiming to help the LSST user community prepare for LSST science.

Key Goals:

1. Open the discovery space to a larger and more inclusive community (LINCC Frameworks/Infrastructure
2. Foster the creation and sharing of analysis tools and data for key LSST science goals (LINCC Frameworks)
3. Train and mentor future science leaders (LINCC Hubs, Catalyst Fellowships)

LINCC Hub Collaborators: Carnegie-Mellon University, Northwestern University (CIERA), University of Arizona, University of Washington (DiRAC Institute)

Note: This documentation site is under active development

CHAPTER

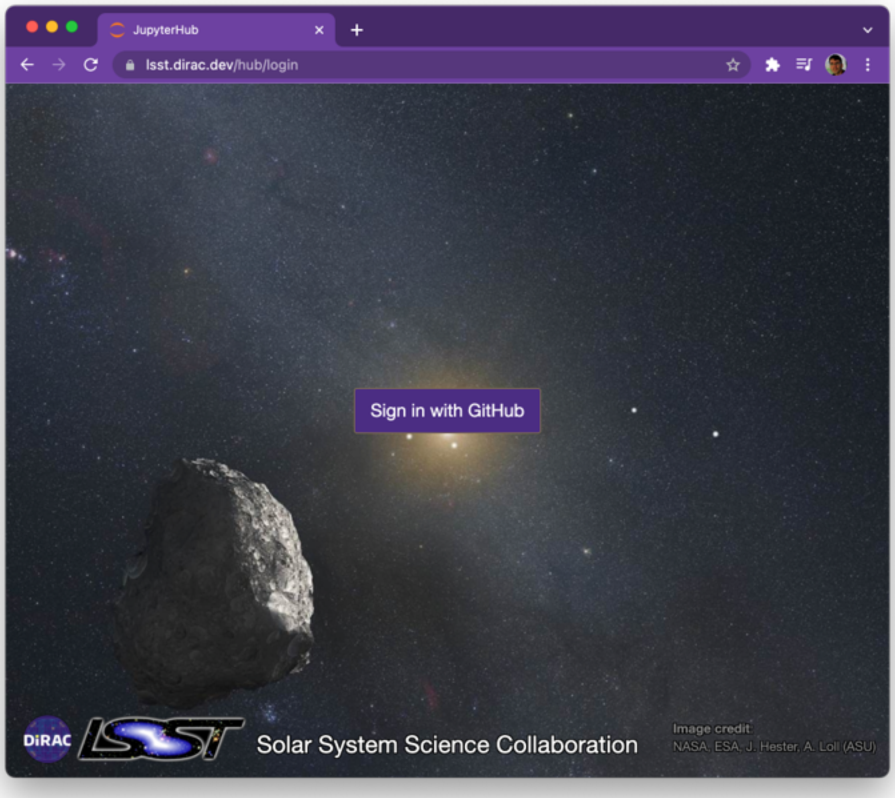
ONE

ACCESSING LINCC JUPYTERHUB

- **To log in, you will need:**
    1. A GitHub account (https://github.com).
    2. Membership in the `lincc-hub` GitHub organization (https://github.com/lincc-hub)
- **To get an invitation to join the lincc-hub organization:**
    - Join the `#lincc-hub` channel on LSSTC Slack
    - **Post your github username, and one of our admins will send you an invitation in the next few minutes.**
        * Example message: "Hi, I'd like to get access to LINCC JupyterHub. My github ID is mjuric. Thanks!"

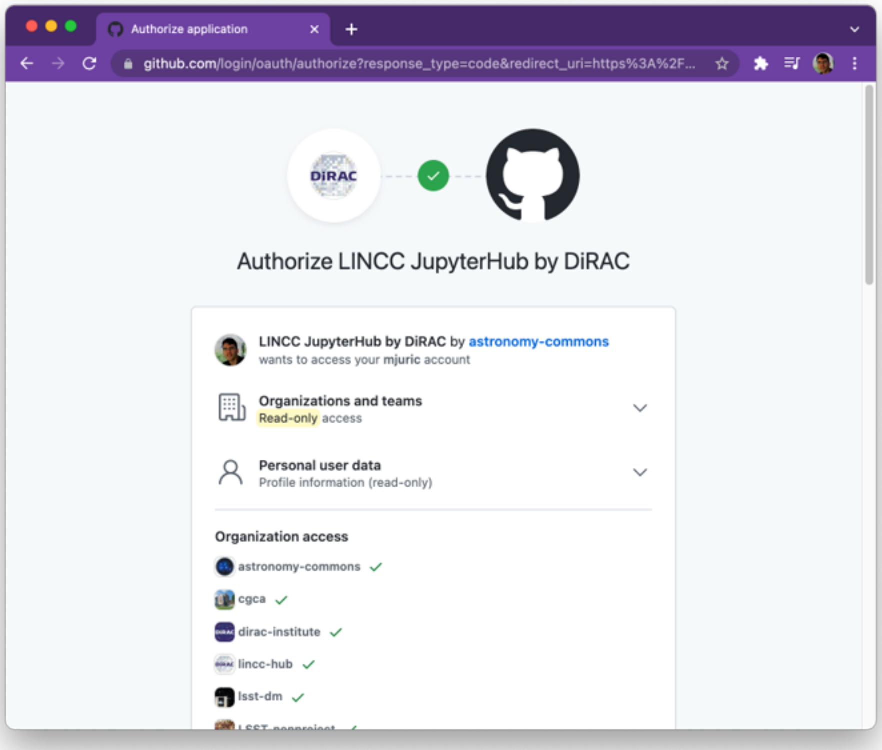
1.1 Logging In

1. Make sure to accept an invitation to join lincc-hub on github!
2. Go to https://lsst.dirac.dev



3. The first time you log in, GitHub will ask you to allow access*[0].

[0] Grant access to organizations you wish to collaborate with within the hub


Authorize application
github.com/login/oauth/authorize?response_type=code&redirect_uri=https%3A%2F...
DiRAC

Authorize LINCC JupyterHub by DiRAC

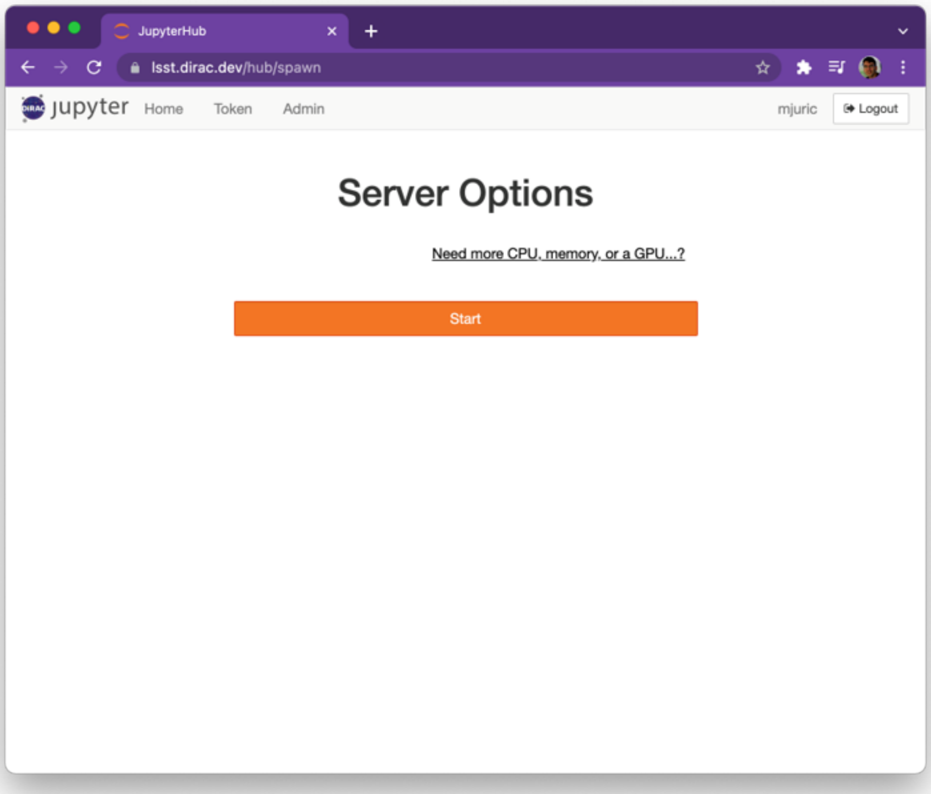
LINCC JupyterHub by DiRAC by astronomy-commons
wants to access your mjuric account

Organizations and teams
Read-only access

Personal user data
Profile information (read-only)

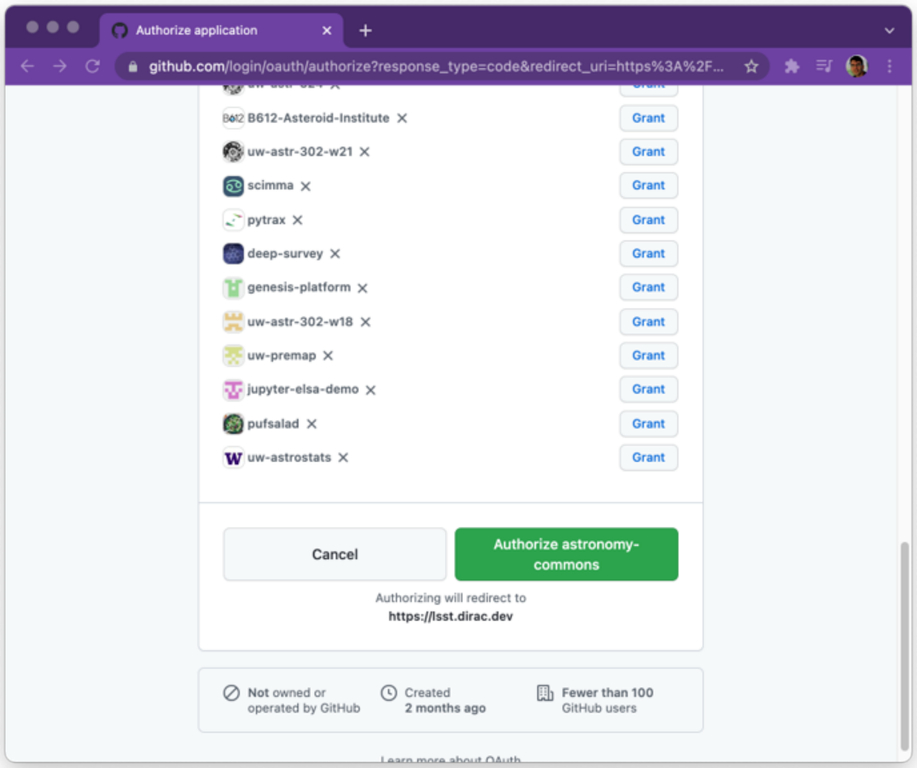
Organization access
astronomy-commons ✓
cgca ✓
dirac-institute ✓
lincc-hub ✓
lsst-dm ✓

4. Upon clicking on the green button (left figure), you will be redirected and logged into the Hub



Server Options

Need more CPU, memory, or a GPU...?


Authorize application
github.com/login/oauth/authorize?response_type=code&redirect_uri=https%3A%2F...

B612-Asteroid-Institute ✕	Grant
uw-astr-302-w21 ✕	Grant
scimma ✕	Grant
pytrax ✕	Grant
deep-survey ✕	Grant
genesis-platform ✕	Grant
uw-astr-302-w18 ✕	Grant
uw-premap ✕	Grant
jupyter-elsa-demo ✕	Grant
pufsalad ✕	Grant
uw-astrostats ✕	Grant

Cancel
Authorize astronomy-commons

Authorizing will redirect to
https://lsst.dirac.dev

Not owned or operated by GitHub
Created 2 months ago
Fewer than 100 GitHub users

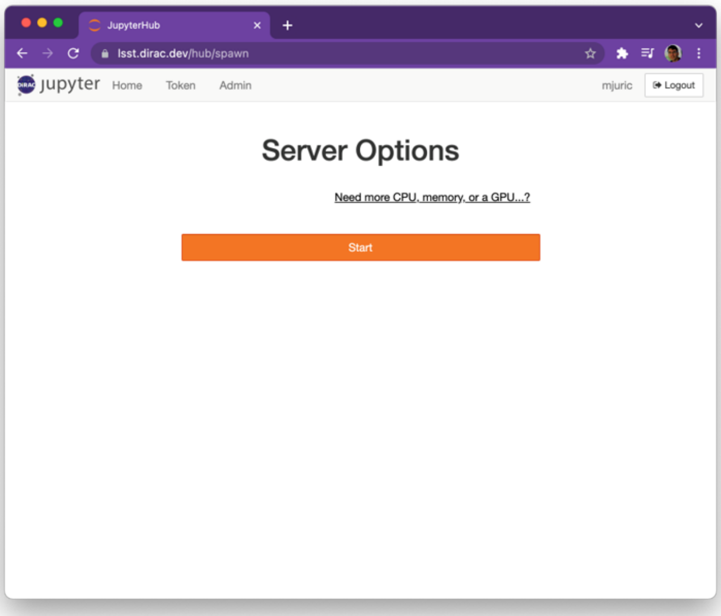
CHAPTER

TWO

INTRODUCTION TO JUPYTER

Your "Home in the Clouds"

- When you logged in and clicked "Start" (big orange button), Jupyter asked Amazon Web Services to allocate a dedicated machine* for your use.

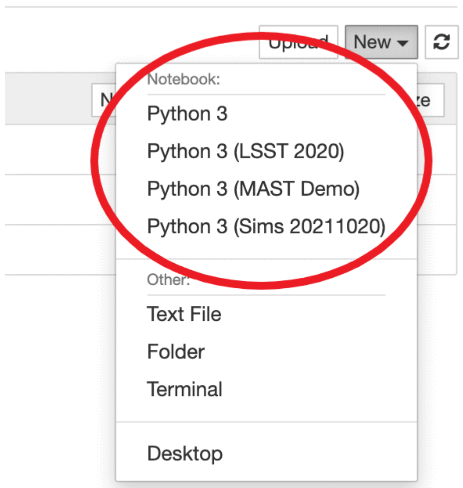


- While this machine is just yours to use, it has access to the shared directory with software and files. This is not unlike how office computers in many institutions see the same filesystem.
- This machine can only be accessed through Jupyter (Notebook, Lab, Terminal). Note: direct SSH access is in the works.

Important: If you don't use it for more than an hour (close the browser with Jupyter), it will be shut down to conserve resources. Disk data will be preserved, but all running Jupyter notebooks will be shut down.

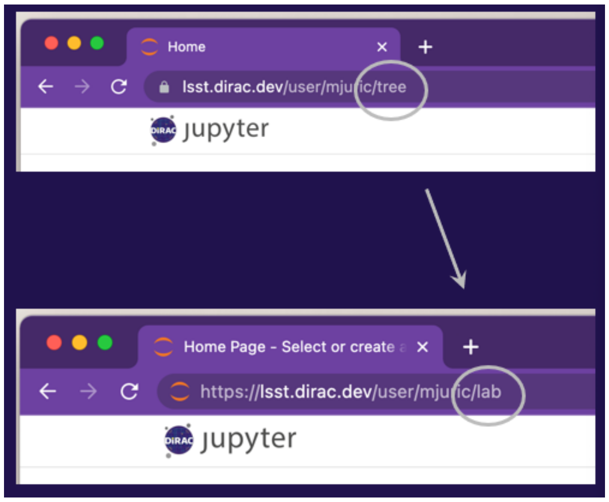
2.1 Jupyter Notebooks

You can access a notebook from the Jupyter homepage by clicking the "New" button and selecting a virtual environment for your notebook.



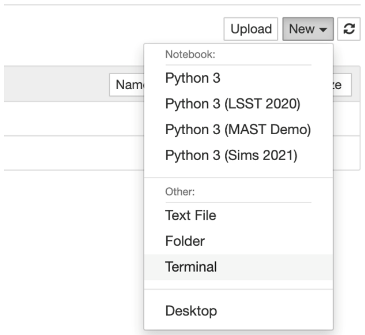
2.1.1 Accessing JupyterLab

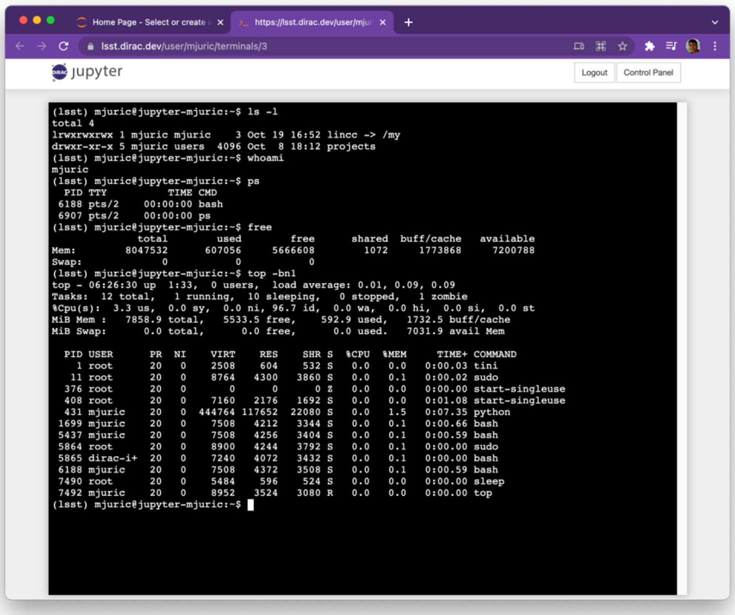
If you prefer the JupyterLab interface, replace the 'tree' in your URL with 'lab'.


Home
lsst.dirac.dev/user/mjuric/tree
jupyter
Home Page - Select or create a
https://lsst.dirac.dev/user/mjuric/lab
jupyter

2.2 Terminal

To access the Terminal, click the "New" button and select the "Terminal" option. This will open a new terminal instance for your server.

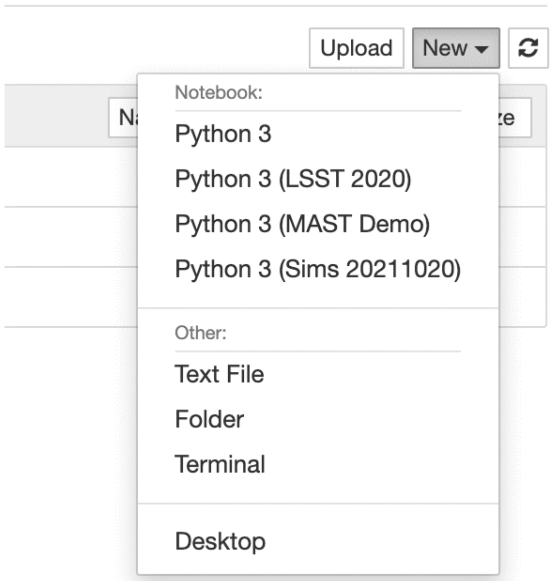



Home Page - Select or create a
https://lsst.dirac.dev/user/mjur
lsst.dirac.dev/user/mjuric/terminals/3
jupyter
Logout
Control Panel

```
(lsst) mjuric@jupyter-mjuric:~$ ls -l
total 4
lrwxrwxrwx 1 mjuric mjuric    3 Oct 19 16:52 lincc -> /my
drwxr-xr-x 5 mjuric users  4096 Oct  8 18:12 projects
(lsst) mjuric@jupyter-mjuric:~$ whoami
mjuric
(lsst) mjuric@jupyter-mjuric:~$ ps
  PID TTY          TIME CMD
 6188 pts/2    00:00:00 bash
 6907 pts/2    00:00:00 ps
(lsst) mjuric@jupyter-mjuric:~$ free
              total        used        free      shared  buff/cache   available
Mem:        8047532      607056     5666608        1072     1773868     7200788
Swap:             0           0           0
(lsst) mjuric@jupyter-mjuric:~$ top -bn1
top - 06:26:30 up  1:33,  0 users,  load average: 0.01, 0.09, 0.09
Tasks:  12 total,   1 running,  10 sleeping,   0 stopped,   1 zombie
%Cpu(s):  3.3 us,  0.0 sy,  0.0 ni, 96.7 id,  0.0 wa,  0.0 hi,  0.0 si,  0.0 st
MiB Mem :   7858.9 total,   5533.5 free,    592.9 used,   1732.5 buff/cache
MiB Swap:      0.0 total,      0.0 free,      0.0 used.   7031.9 avail Mem

  PID USER      PR  NI    VIRT    RES    SHR S  %CPU  %MEM     TIME+ COMMAND
    1 root      20   0    2508    604    532 S   0.0   0.0   0:00.03 tini
   11 root      20   0    8764   4300   3860 S   0.0   0.1   0:00.02 sudo
  376 root      20   0       0      0      0 Z   0.0   0.0   0:00.00 start-singleuse
  408 root      20   0    7160   2176   1692 S   0.0   0.0   0:01.08 start-singleuse
  431 mjuric    20   0  444764 117652  22080 S   0.0   1.5   0:07.35 python
 1699 mjuric    20   0    7508   4212   3344 S   0.0   0.1   0:00.66 bash
 5437 mjuric    20   0    7508   4256   3404 S   0.0   0.1   0:00.59 bash
 5864 root      20   0    8900   4244   3792 S   0.0   0.1   0:00.00 sudo
 5865 dirac-i+  20   0    7240   4072   3432 S   0.0   0.1   0:00.00 bash
 6188 mjuric    20   0    7508   4372   3508 S   0.0   0.1   0:00.59 bash
 7490 root      20   0    5484    596    524 S   0.0   0.0   0:00.00 sleep
 7492 mjuric    20   0    8952   3524   3080 R   0.0   0.0   0:00.00 top
(lsst) mjuric@jupyter-mjuric:~$ 
```

You can use this to manage your directories, edit files, install and run codes, etc.



2.3 Available Software

- Python 3: Python 3 with the current LSST Stack Release.
- Python 3 (Sims NNNNNN): Python 3 with a recent install of the rubin_sims tools (https://github.com/lsst/rubin_sim)
- Python 3 (xxx): Older LSST Stack releases and custom environments prepared for events.

CHAPTER

THREE

CUSTOMIZING YOUR HUB

3.1 Adding software: Conda Environments

You can create any number of personal conda environments where you have complete control over the software available. To create and activate a new conda environment, run the following in a terminal:

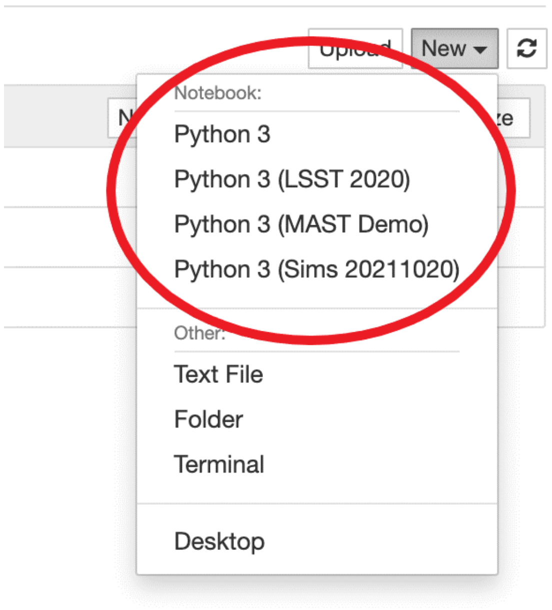
```
$ conda create -n my_environment python=3
$ conda activate my_environment
```

Feel free to change "my_environment" to any environment name you desire. After activating the environment, you can use `conda` and `pip` commands to install any software you desire.

It may be helpful to install the `mamba` software, a faster version of `conda`:

```
$ conda install mamba
```

If the `ipykernel` package is installed in the environment, it will automatically show up in your list of kernels and be available for use in a Jupyter notebook.



To install `ipykernel` use this command,

```
$ conda install ipykernel
```

or

```
$ mamba install ipykernel
```

Important: These environments are only accessible to you (they’re personal).

CHAPTER

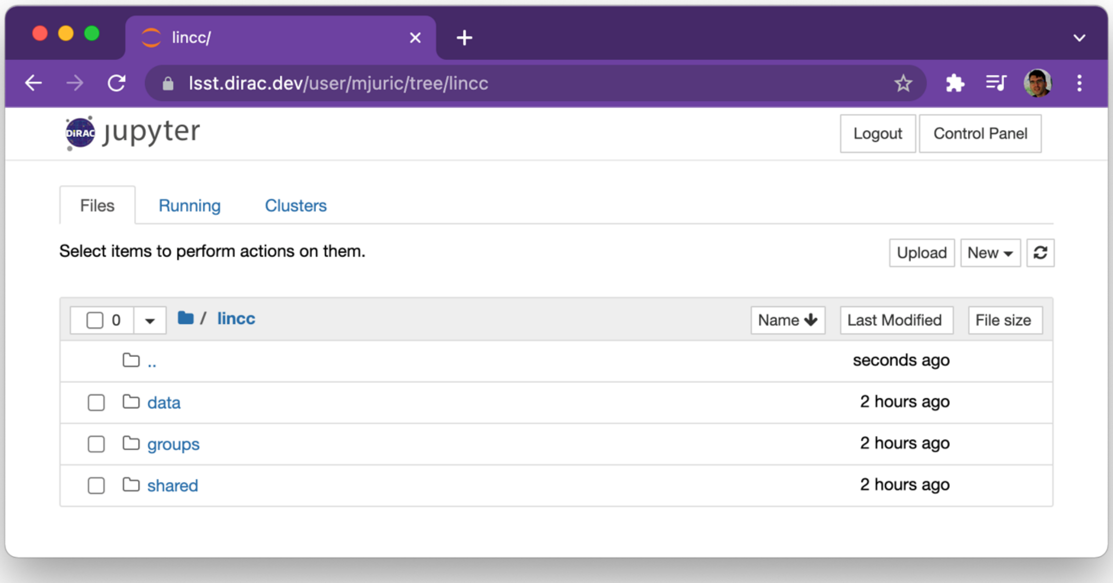
FOUR

COLLABORATING USING OUR PLATFORM

4.1 Collaboration Facilities

Every user has a 'lincc' folder in their home directory, with three subdirectories:

- **data:** shared, centrally managed, data.
- **groups:** group-owned, group-managed, files.
- **shared:** a folder anyone can write to, and readable by everyone on LINCC JupyterHub.

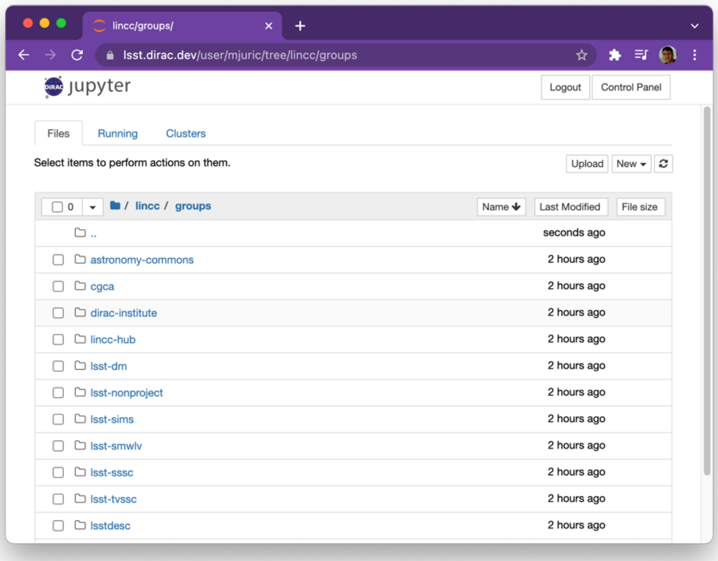


Select items to perform actions on them.

Name	Last Modified
..	seconds ago
data	2 hours ago
groups	2 hours ago
shared	2 hours ago

4.2 Group file spaces

- When you log in, any organization that you're a member of on github (and to which you've granted LINCC JupyterHub access; see slide #7), is given:
  - a UNIX group on LINCC JupyterHub
  - a directory in `lincc/groups`, readable and writable only by members of that group.



Select items to perform actions on them.

/ lincc / groups

Name	Last Modified
..	seconds ago
astronomy-commons	2 hours ago
cgca	2 hours ago
dirac-institute	2 hours ago
lincc-hub	2 hours ago
lsst-dm	2 hours ago
lsst-nonproject	2 hours ago
lsst-sims	2 hours ago
lsst-smwlv	2 hours ago
lsst-sssc	2 hours ago
lsst-tvssc	2 hours ago
lsstdesc	2 hours ago

- You can place shared notebooks, datafiles, software here…
- Works for any github organization, and is fully automatic.
- Enables seamless collaboration!

4.3 Group-controlled environments and software

- A frequent use for shared directories is to install and maintain common software environments.
- It's good to protect such environments from accidental corruption or deletion.
  - E.g., it's too easy to accidentally run 'pip install some_package' and ruin a carefully curated conda environment.
  - Especially true for large groups (e.g. science collaborations).
- To defend against this, each group has an admin account, designed to own its software and important files.

Here is an example of how to use an admin account in the terminal:

```
# login as admin for dirac-institute
$sudo -u dirac-institute -s
$cd

# install a new environment
```

(continues on next page)

(continued from previous page)

```
$conda create -n solarsystem openorb
exit

# check that we have access to the env
$conda env list
$conda activate solarsystem
$oorb

# make it accessible in Jupyter
$sudo -u dirac-institute -s
$cd
$conda activate solarsystem
$mamba install ipykernel
```

Note: Only GitHub organization owners have access to the admin account.

Tip:

- Your **home directory** is visible only by you. Keep any private files there.
- A **group directory** is visible only to eponymous GitHub organization's members. Keep common files there.
- The **shared directory** is visible (readable and writable) to anyone. Place files there that you want to share with the entire hub (that are public).

CHAPTER

FIVE

AVAILABLE RESOURCES

5.1 Disk space

1TB of space shared by everyone. We can add more if (when) needed.

5.2 Computational resources

- Machines are available with different combinations of CPU, RAM, and GPU.
- Please use the smallest machine that can fit your workload (to help us conserve $$).
- Also, larger machines may take more time to start up (up to ~5 minutes).

You can access different server options by clicking *Need more CPU, memory, or a GPU… ?* on the Server Options page.

images/picture11.png

CHAPTER

SIX

SSH

A SSH "jump host" is deployed with this chart to allow users to use *scp* and *sftp* to copy files to the NFS. Each notebook server starts its own SSH service, allowing users to access their notebook servers using *ssh* via the jump host.

If a user wants to utilize this, they should perform the following steps:

1. Launch their notebook server.
2. Start a terminal and open the file `~/.ssh/authorized_keys` with a text editor.
3. On their local machine, the user should generate a new or reuse an old SSH public key. Existing public/private key pairs may be named `~/.ssh/id_rsa` and `~/.ssh/id_rsa.pub`. New public/private key pairs can be generated by running `ssh-keygen`. The user should copy their public key from their local machine and add it as a new line in the file from (2) on the remote machine.
4. The user should edit the file `~/.ssh/config` on their local machine to include the following, replacing `<username>` with their username on the JupyterHub:

```
Host lsst-hub-ssh
    User <username>
    Hostname ssh.lsst.dirac.dev

Host lsst-hub
    User <username>
    Hostname lincc-<username>.notebooks
    ProxyJump lsst-hub-ssh
```

For example:

```
Host lsst-hub-ssh
    User stevenstetzler
    Hostname ssh.lsst.dirac.dev

Host lsst-hub
    User stevenstetzler
    Hostname lincc-stevenstetzler.notebooks
    ProxyJump lsst-hub-ssh
```

5. On their local machine, the user can ssh to their running notebook server using:

```
$ ssh lsst-hub
```

Alternatively, they can specify the jump host explicitly:

```
$ ssh -J <username>@ssh.lsst.dirac.dev <username>@lincc-<username>.notebooks
```

CHAPTER SEVEN

SUPPORT AND DEVELOPMENT

`#lincc-hub` on LSSTC Slack is the main communication channel. Join and give us comments, questions, or feature requests there!

Support team on Slack:

- @Aditi Chauhan, @Aidan Berres, @Paulo Andres Stevens Barrera

Developers and researchers:

- @Steven Stetzler @mjuric @colin
- A larger software development team coming in 2022 (follow us on https://dirac.us/linkedin to be notified of job openings)

Warning: LINCC JupyterHub is a new system, still under heavy development. We apologize for the occasional glitch!

CHAPTER

EIGHT

INDICES AND TABLES

- genindex
- modindex
- search