STScI | SPACE TELESCOPE SCIENCE INSTITUTE

EXPANDING THE FRONTIERS OF SPACE ASTRONOMY

Data Reduction with Jupyterhub

Steve Crawford

9 October 2019

James Webb Space Telescope

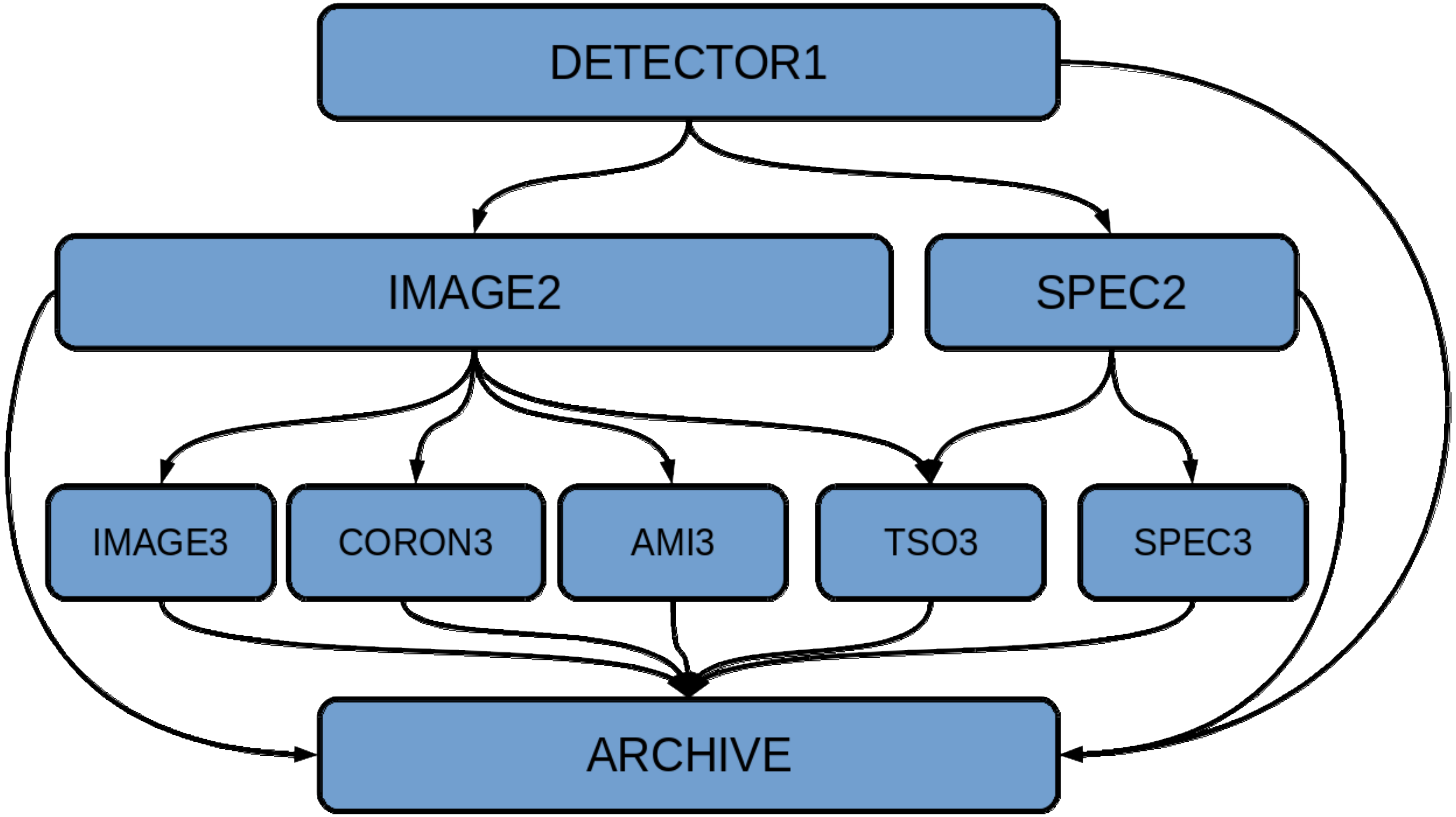
JWST is 6.5m space telescope expected to launch in March 2021.

Suite of near- and mid-infrared imaging and spectroscopic instruments:

- FGS
- MIRI
- NIRCAM
- NIRISS
- NIRSPEC

A fully assembled JWST (August 2019)

JWST Calibration software: A single package to reduce them all



P10.19 H. Bushouse The JWST Science Calibration Pipeline

https://github.com/spacetelescope/jwst

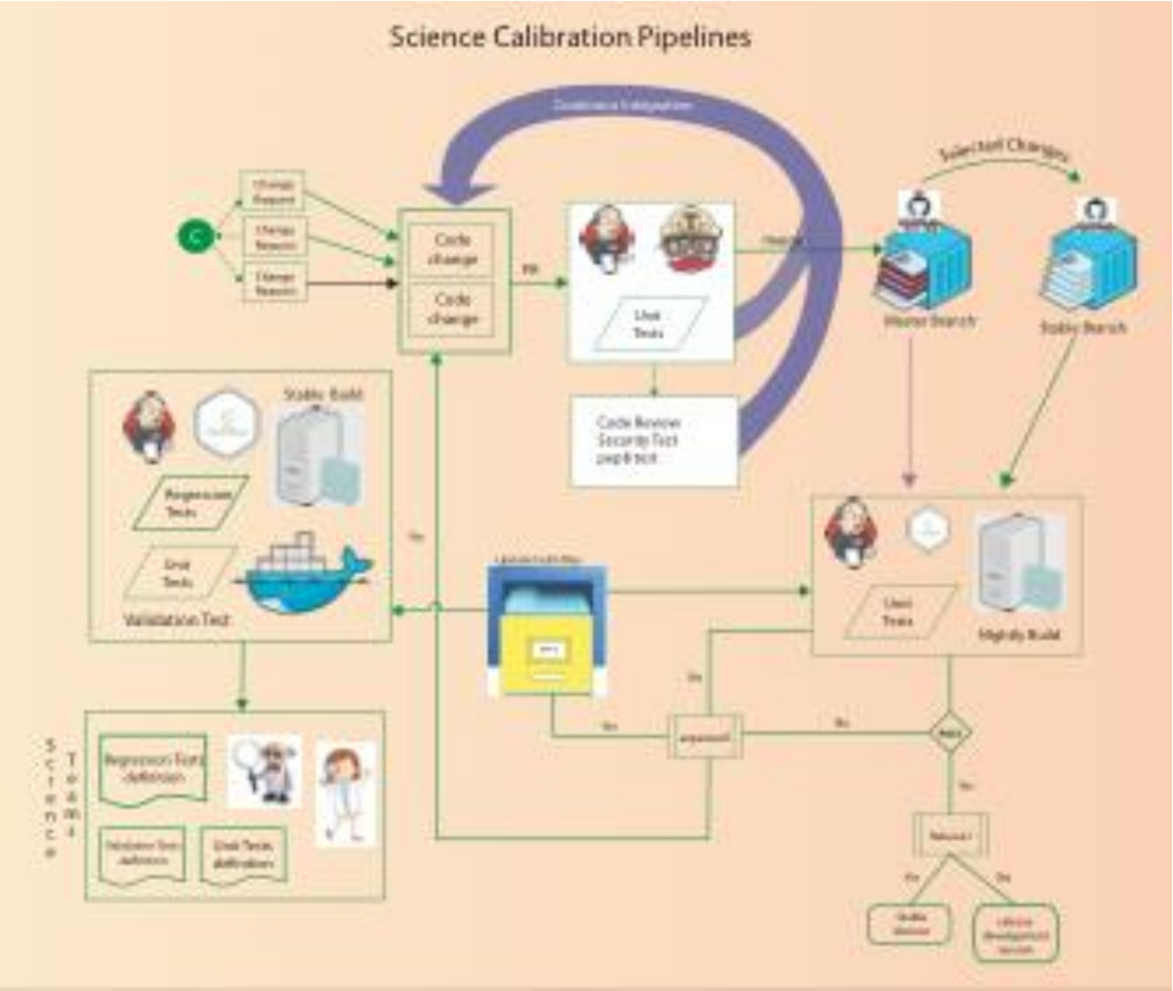
Testing JWST Calibration Software

The JWST calibration software testing includes:

- Verification through Integration and Testing with Science Operations Center software
- Validation by Instrument team Scientists

Along with the automated tests, Astronomers are writing Jupyter notebooks for manual testing of the calibration software

P9.16 R. Diaz End-to-end validation framework

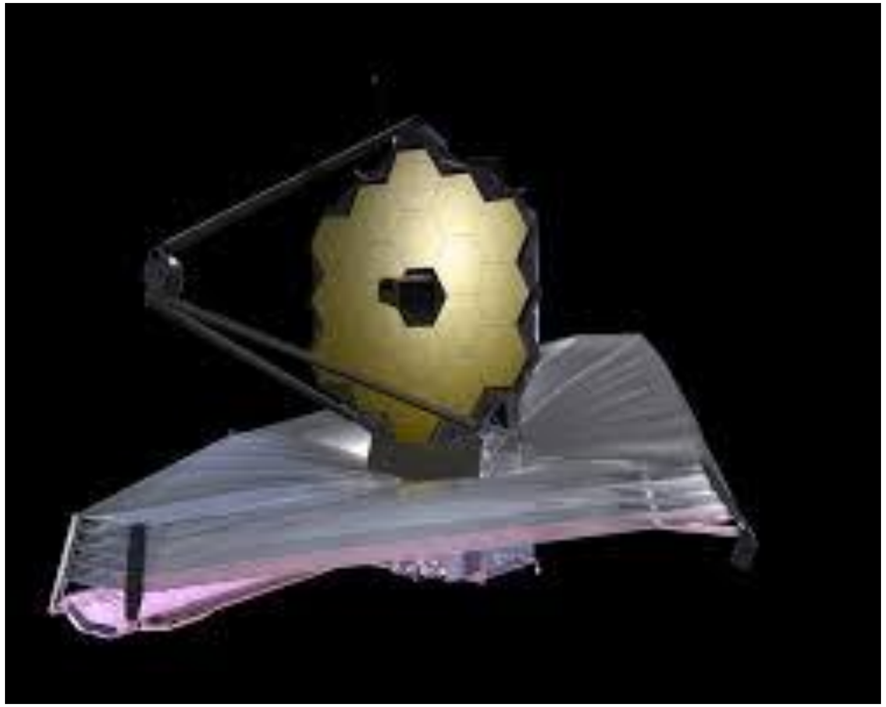


Motivation

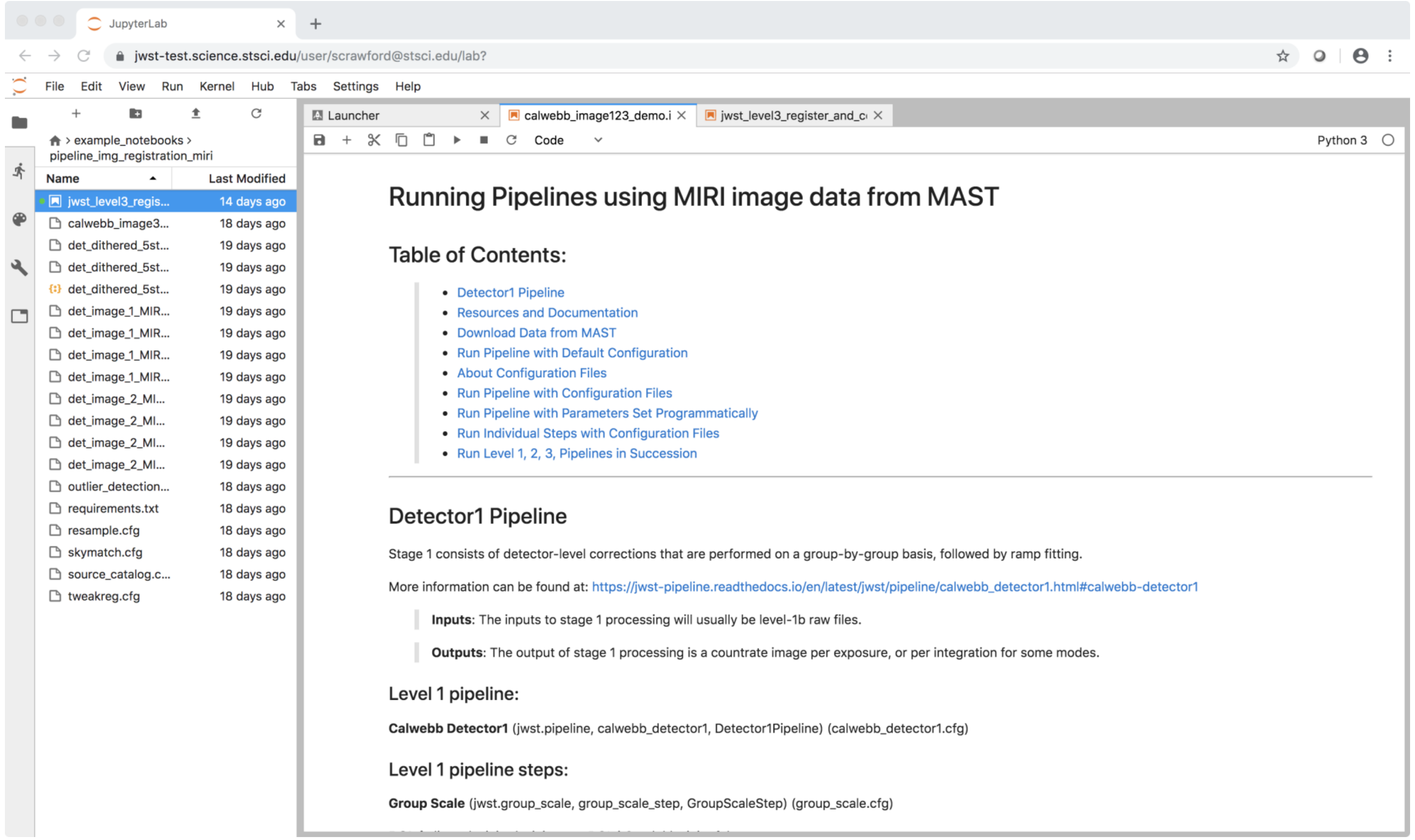
How do we provide a reproducible environment for testing?

How can we provide an environment that can be seamlessly updated?

How will we enable investigators to quickly extract the scientific results from complex observations?



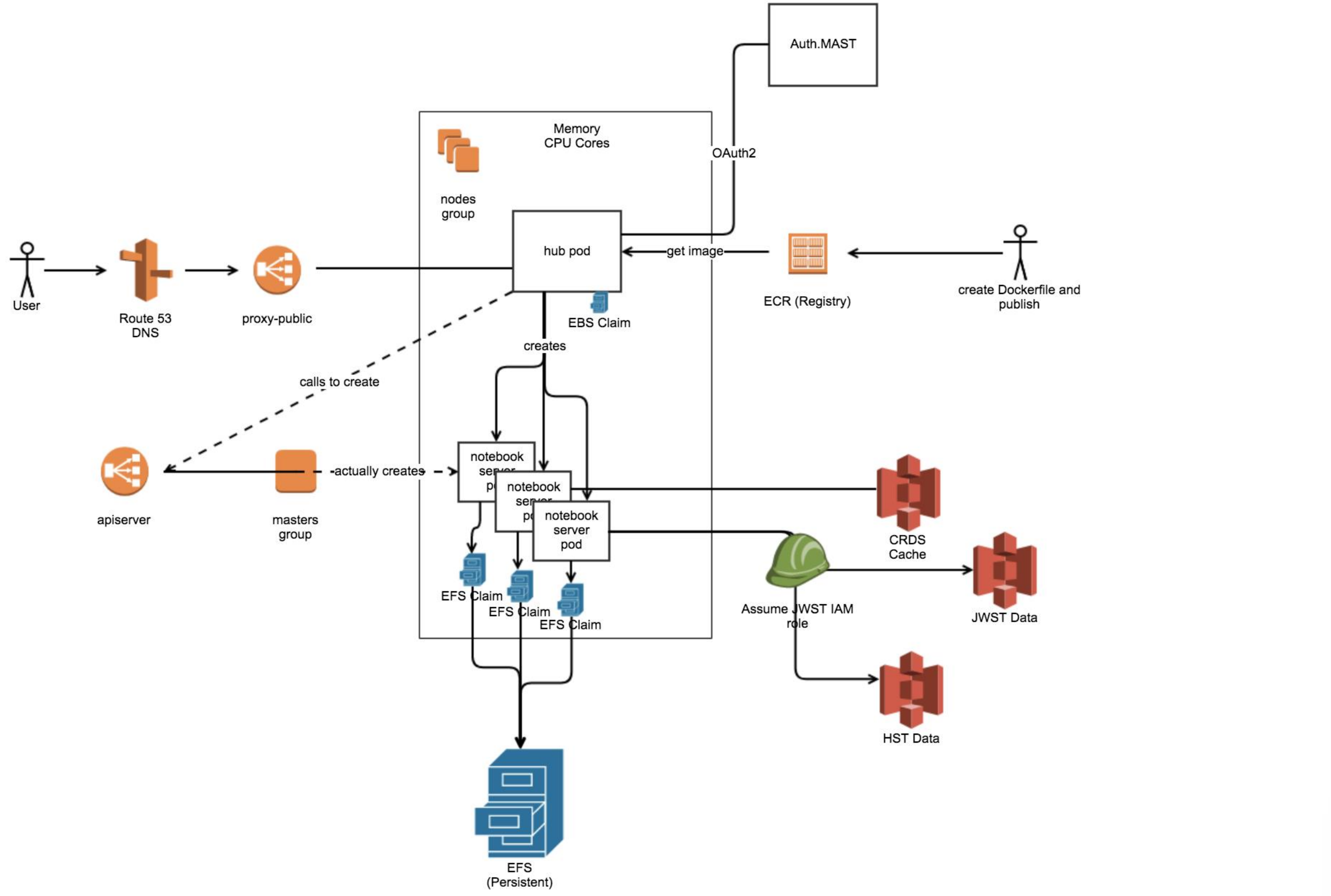
Jupyterhub: a scalable science platform



Jupyterhub provides a browser-based science platform.

- Running on AWS provides flexibility and scalability
- Used previously for TESS and WFIRST workshops

Components in Jupyterhub



https://mast-labs.stsci.io/2019/02/zero-to-jupyterhub-with-ansible

Cloud based Calibration Reference Data System

CRDS ingests, stores, associates, and serves the references files for JWST. A cloud based cache has been enabled to allow users on the Juptyerhub system to directly read from S3 storage

Home | Login | Guide

JWST Calibration Reference Data System (CRDS)

Obtain Best Reference Files

1. Using the Command Line
2. From Dataset ID or FITS Header Upload
3. Exploring with Instrument Parameters

Reference File Database Services

1. Browse Database
2. Recent Activity

Operational References (under context jwst_0541.pmap)

- fgs
- miri
- nircam
- niriss
- nirspec
- system

Context History (more history, all contexts)

Start Date	Context	Status	Description
2019-07-30	jwst_0541.pmap	operational	As per Jira issues CRDS-287, REDCAT-52, and CRDS-153, the rmaps for the following NIRSpec reference types were updated to include READPATT values of NRSRAPIDD1, NRSRAPIDD2, and NRSRAPIDD6. This rmaps submission includes changes to the following DARK, READNOISE, REFPIX and SUPERBIAS.
2019-07-23	jwst_0540.pmap	archived	New MRS REGIONS files for all channel-band combinations are being delivered to fix an issue where a few slices were incorrectly a fraction of an arcsec too long. There are no header format changes, no necessary pipeline revisions, and

Access to Data

Astroquery plus MAST authentication

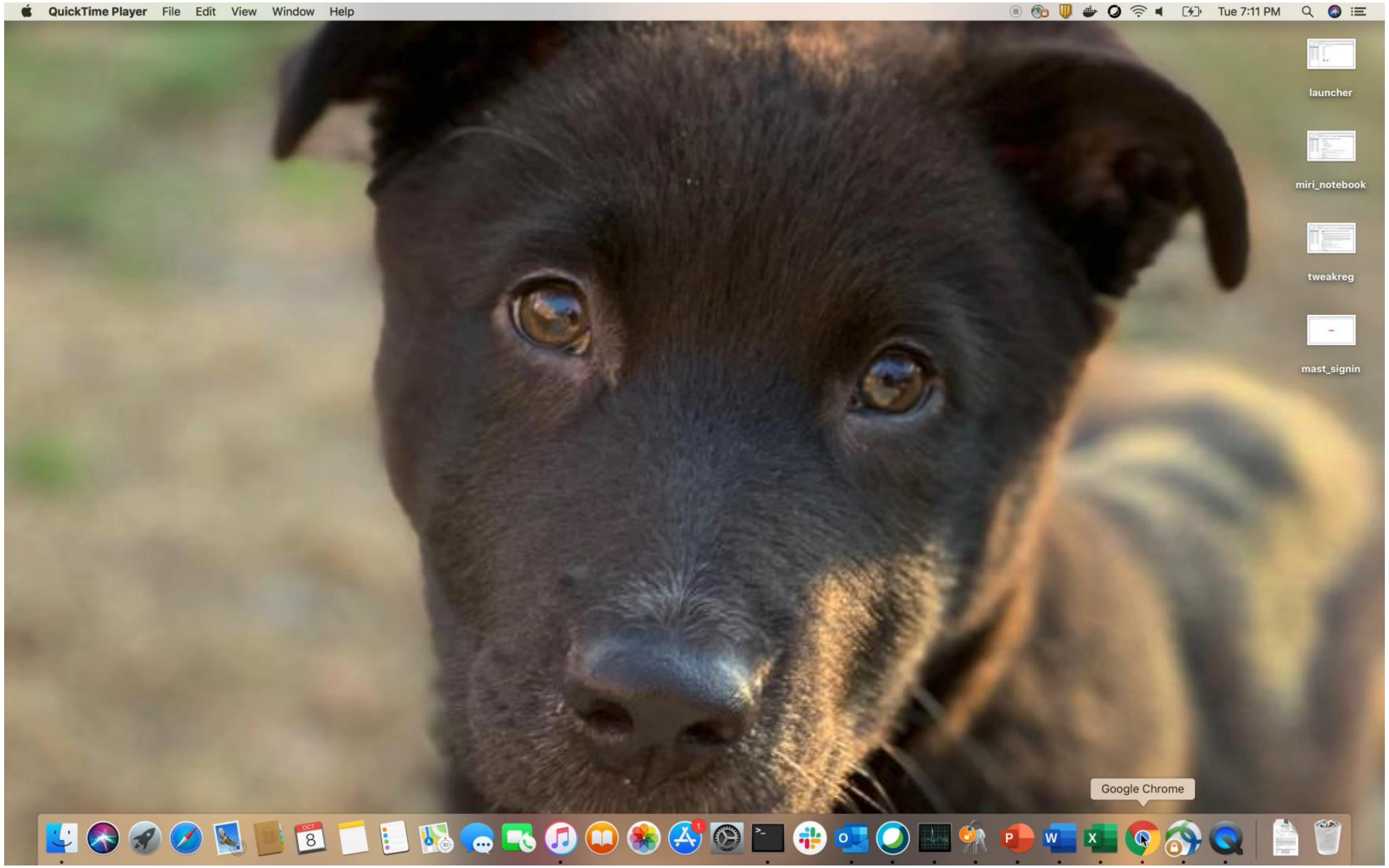
```
>>> from astroquery.mast import Observations

>>> obs_table = Observations.query_object("M8",radius=".02 deg")
>>> data_products_by_obs = Observations.get_product_list(obs_table[0:2])
>>> print(data_products_by_obs)
```

Data Redirector

```
base_url = 'https://data.science.stsci.edu/redirect/JWST/jwst-data_analysis_tools/miri_simulated_data/'
association_filename = "det_dithered_5stars.json"
urlretrieve(base_url + association_filename, association_filename)
print("Downloaded: {}".format(association_filename))
```

Demo of the Environment



Challenges in the Jupyterhub Environment

- Security
  - How do we secure the environment from external and internal sources?
  - How do we control exclusive access data in the environment?
- Collaboration
  - How do we provide an environment that enables collaboration while being secure?
- Observability
  - What are the important metrics to monitor?
  - How do we have real time insight into the system?
- Cost
  - How do we control costs for different users and different use cases?
  - How do we provide access for different users?
- Adoption
  - As a new tool, how do we introduce it to the community?

Future Opportunities

- Authenticated Cloud service for direct access to data
- Providing the full suite of JWST Scientific tools in a single, browser based user interface
    - P11.8 J Taylor: JWST Data Simulations and How to Use Them
    - P10.5 M. Bourque: The James Webb Space Telescope Quicklook Application (JWQL)
- Integrating with addition astronomer and data science software to enable a range of scientific use cases
- Jupyter-widget Glueviz based data analysis tools
- Integrate with batch processing software for job control and monitoring
- Enable cloud based reprocessing and analysis for HST public data set on AWS S3
- The data size of WFIRST will require *bringing users to the data* and this is an early prototype of the High Level Processing Partition

Thank you

Jacob Matuskey Mary Romelfanger Andrew Cortese Brian Hayden Todd Miller Edward Slavich Christine Slocum Michael Gough Pey-Lian Lim Joe Hunkeler Erik Tollerud Michael Fox Clara Brasseur Jonathan Hargis Yuvi Panda

Howard Bushouse Jonathan Eisenhamer Nadia Dencheva James Davies David Grumm Philip Hodge Jane Morrison Megan Sosey Robert Jedrzejewski David Davis

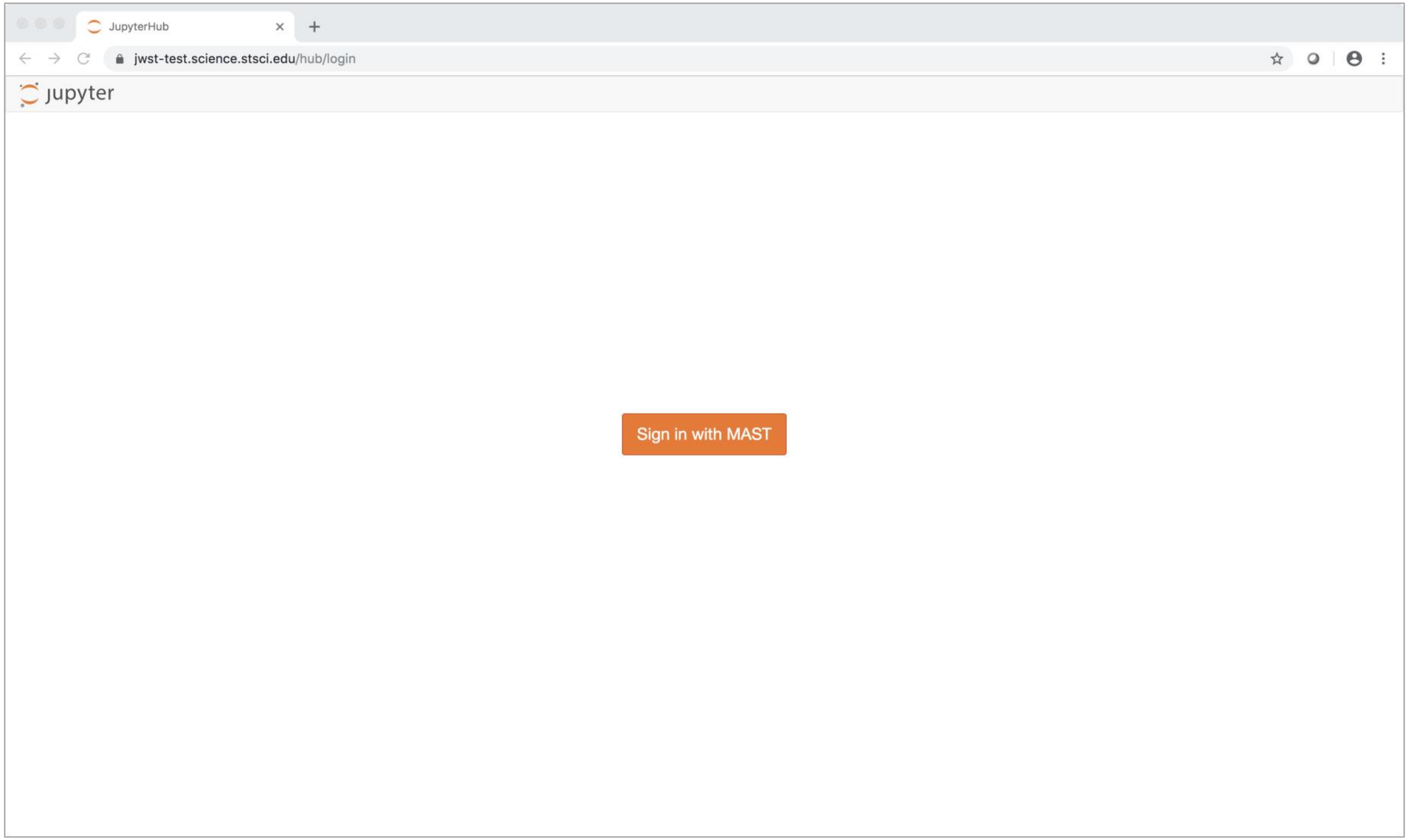
Alicia Canipe Julien Girard Bryan Misty Cracraft Rosa Diaz Anton Koekemoer Karl Gordon Sarah Kendrew Mees Fix

Arfon Smith Ivelina Momcheva Joshua Peek

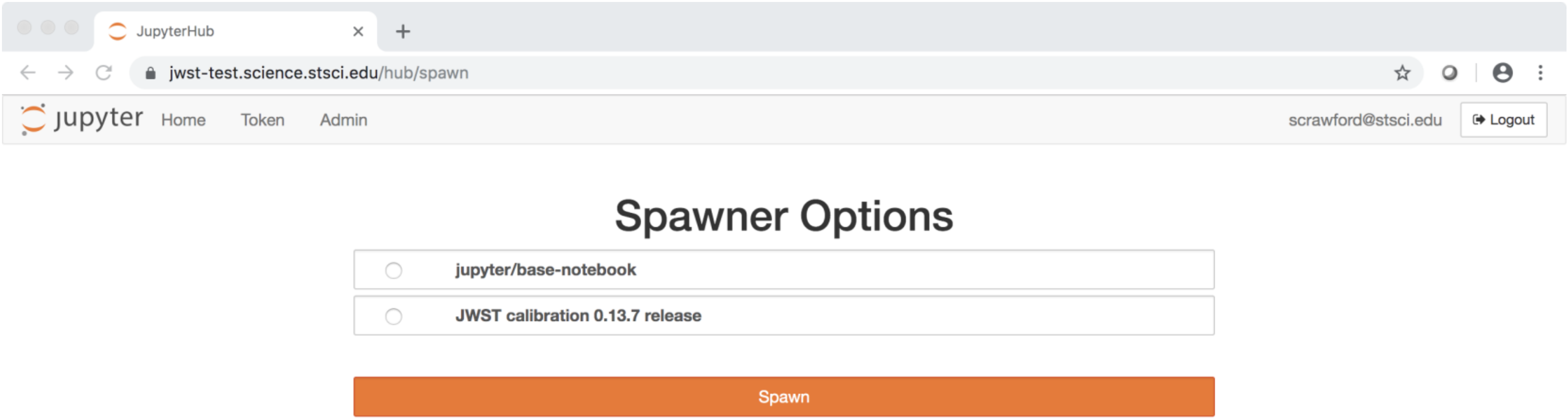
Software available at: https://github.com/spacetelescope

Interested in joining the team?
http://www.stsci.edu/opportunities

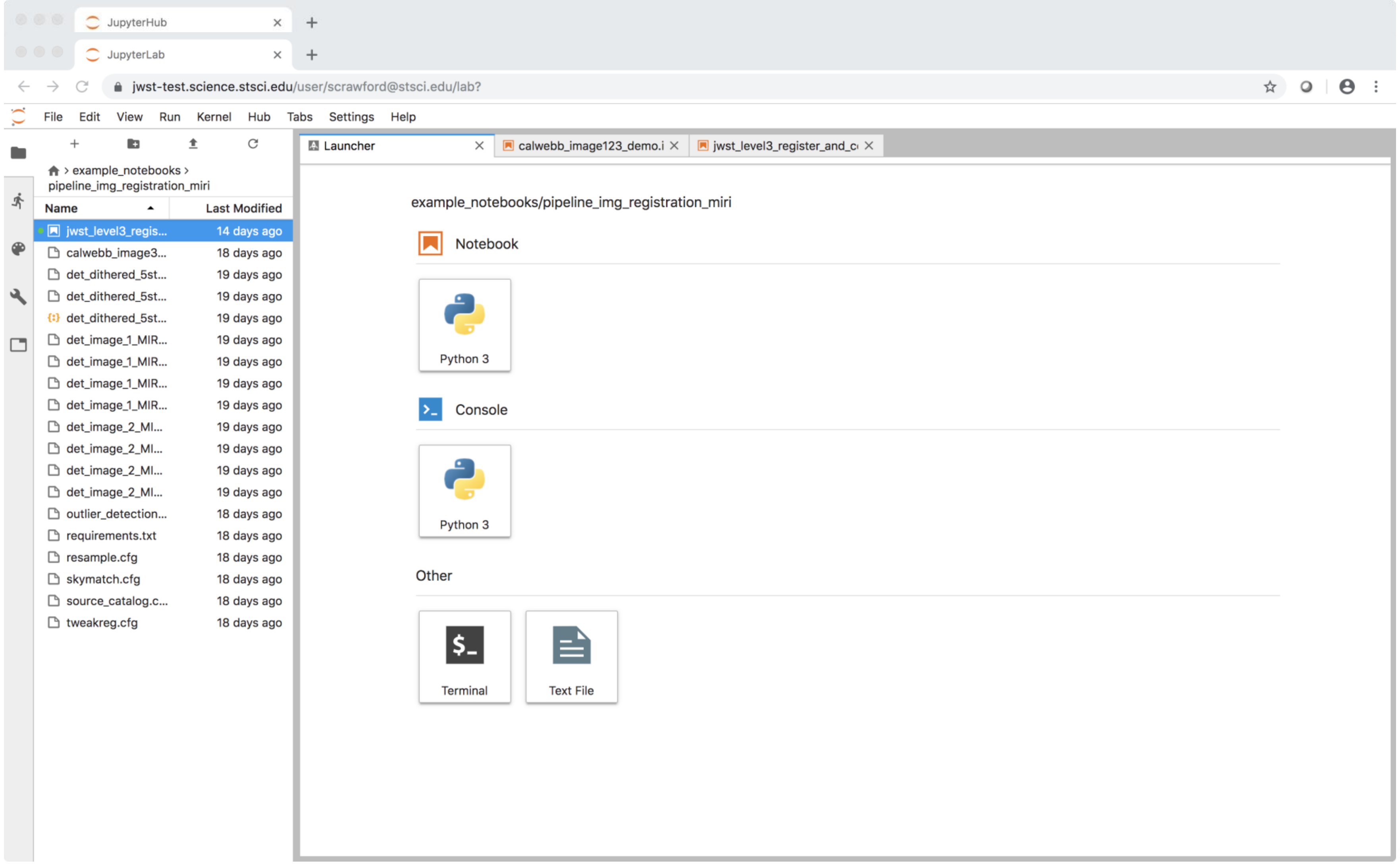
Demo of the environment



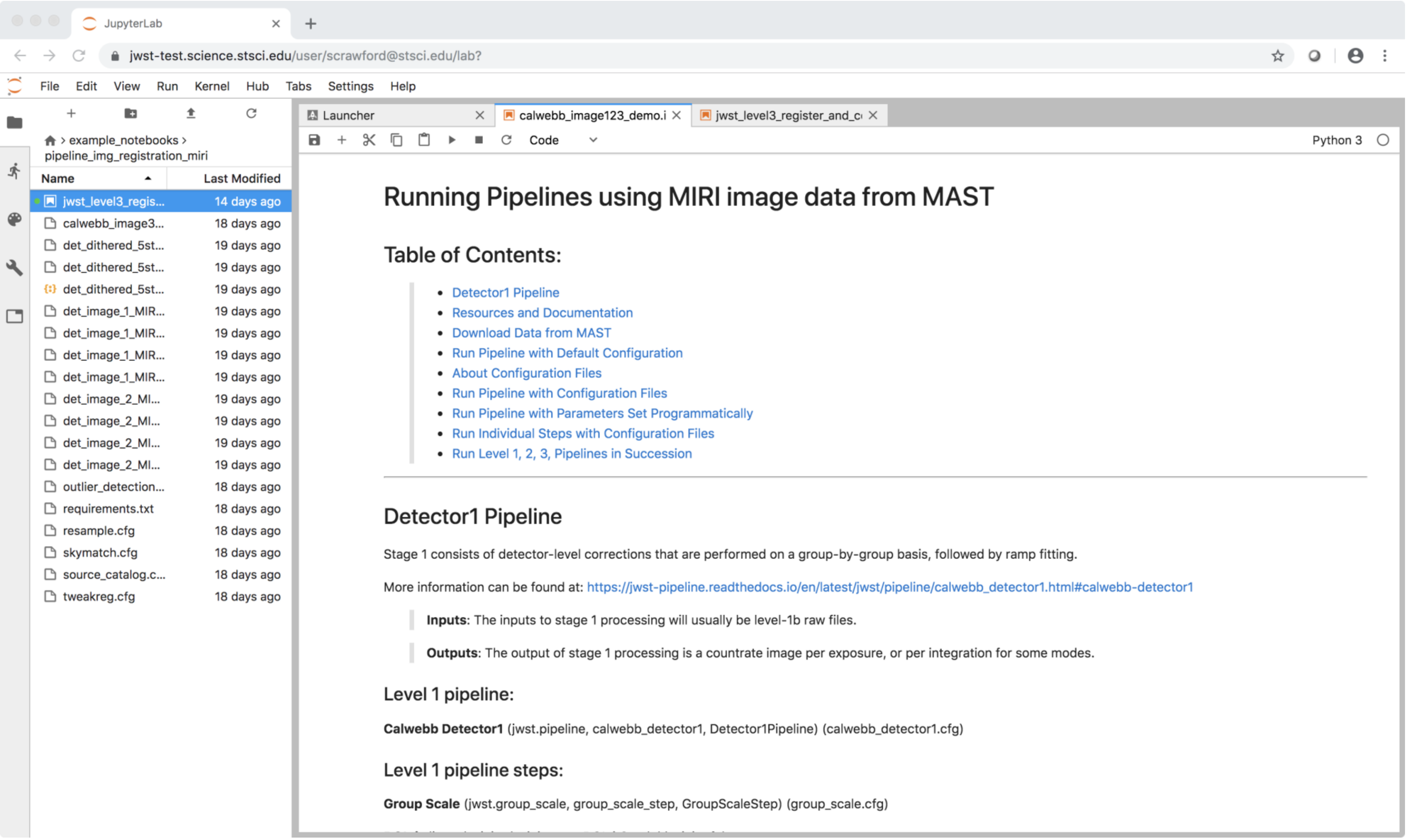
Demo of the environment



Demo of the environment



Demo of the environment



Demo of the environment