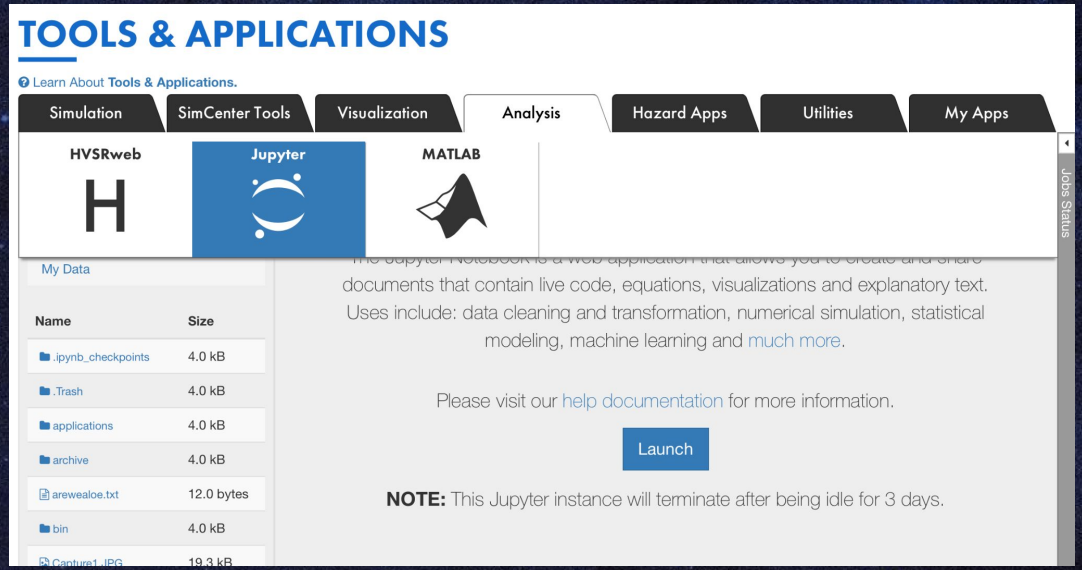
JupyterHub update
(DS Team meeting 2/23/2023)

Anagha Jamthe
Cloud and Interactive Computing
TACC

Background

- Current JupyterHub design utilizes a Kubernetes cluster running on Cyclone VMs, per project (tenant).
- This cluster approximately contains 220 CPU cores, 1200 GB of memory across 20 servers.
- JupyterHub Instances currently running at TACC support:
  - DesignSafe-CI
  - TACC [Hobby Eberly Telescope for Dark Energy Experiments (HETDEX), and UT courses]
  - DARPA SHADE Project

Launch Designsafe JupyterHub

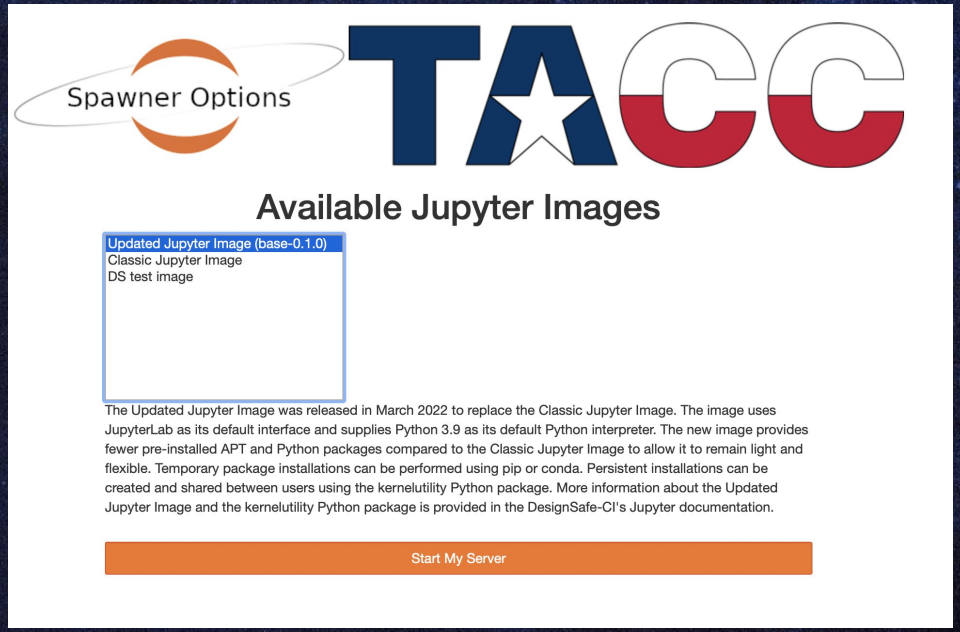


Launch from the Designsafe website https://www.designsafe-ci.org/

OR

Launch from the Designsafe JupyterHub Login Page https://jupyter.designsafe-ci.org/hub/login

Spawner Screen > Image selection



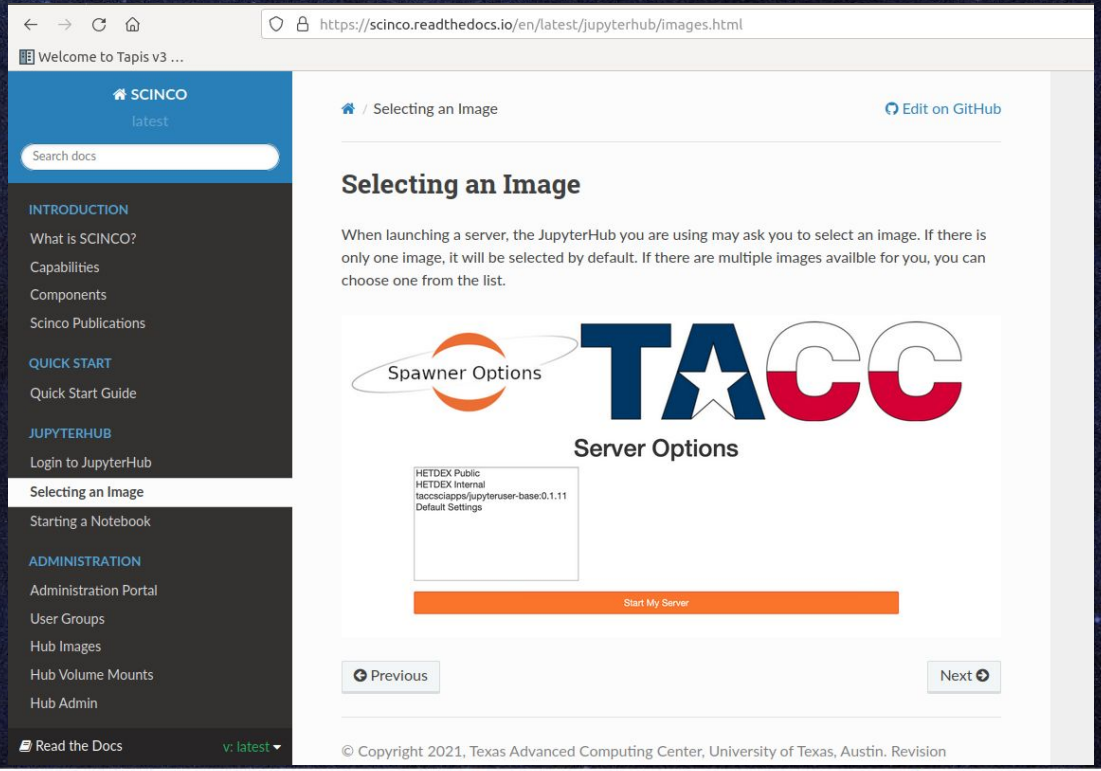
- Updated image uses JupyterLab as its default interface.
- Python 3.9 interpreter
- Light weight as compared to the classic image
- 3 Kernels: Python, R, Julia

Package Installation

- Updated Image has basic packages installed: scipy, numpy, pandas,matplotlib, agave, etc.
- Ephemeral User Installations (Recommended): users can pip or conda install any packages. Packages will not persist after the session closes
- For packages to persist between kernel sessions use the kernel utility developed by Joe V. More details can be found on this webinar
  https://www.youtube.com/watch?v=wpbRGdFzIFw

  !pip install kernelutility

Full Documentation Site Now Available



Next in the pipeline..

Reporting Usage Metrics

- Reporting Metrics – Targeting first release in early April
  - Total number of authentications (logins) to Jupyter Hub
    - By username and time stamp
    - Can derive total number of new users using jupyterhub and similar metrics (e.g., number of unique users for a time period)
  - Notebook files created
    - By username, file path, and time stamp
    - Can derive metrics such as total number of files created in a time period
  - Working on additional metrics such as files renamed, but this is still work in progress.
- Monitoring Dashboard – In place now, but making operational improvements
  - Memory and CPU usage for individual nodes in the cluster
  - Number of users per node
  - Memory usage per user
- Weekly reports will be generated and emailed to a set of project administrators

Administrative Portal

Using the Admin Portal administrators can:

- Start and stop servers on behalf of users
- Add volume mounts to a Hub
- Add custom images to a Hub
- See the list of users on the Hub
- Configure user groups for the Hub

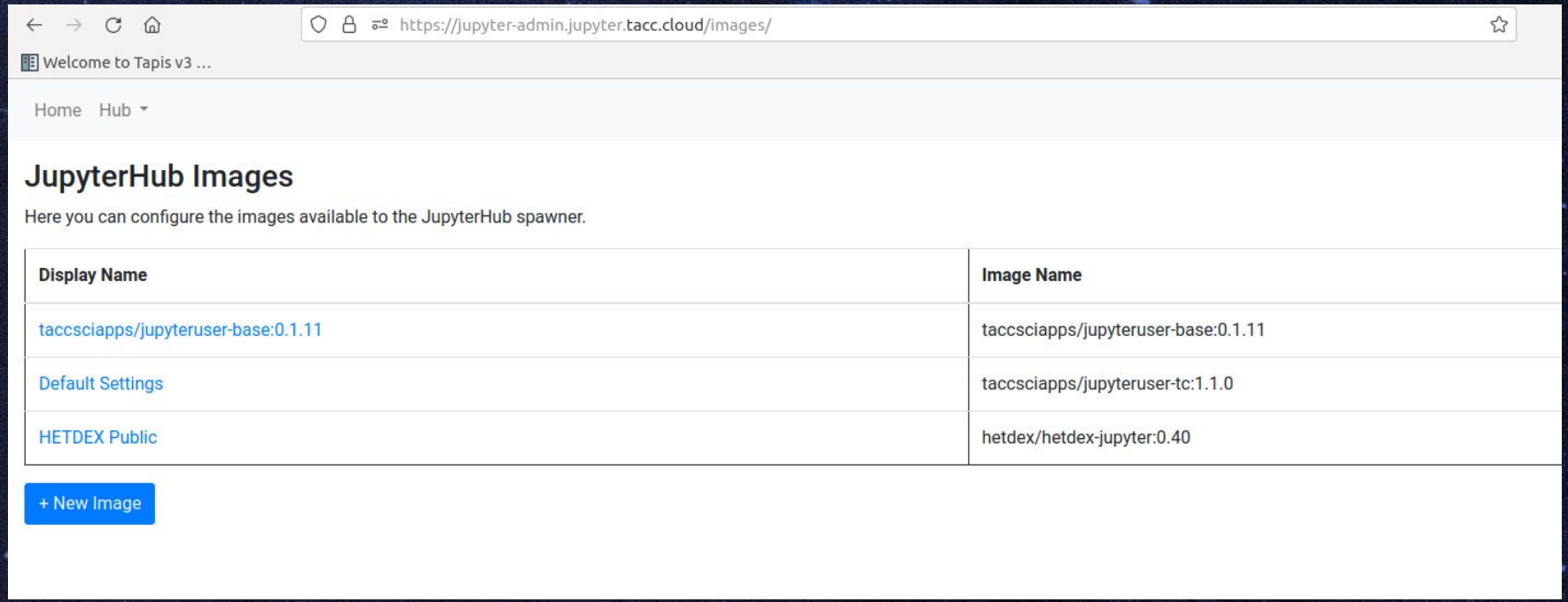
More information about Administrative portal can be found here
https://scinco.readthedocs.io/en/latest/admin/index.html

Currently, the Admin Portal depends on Tapis v3.

Portal is already deployed in TACC JupyterHub.

Will be made available for Designsafe JupyterHub after transitioning of the Hub to Tapis version 3.

Administrative Portal Screenshot



Additional Efforts

- Upgrading to Jupyter Hub v3.x
  - Deployed to our develop environment
  - Will deploy TACC instance to Prod this month
  - Will upgrade DS after that
- Updating the Jupyter HPC Tapis application
  - This is working in Tapis v3 now
- Adding more compute capacity (RAM and CPU) as more servers are added to K8s
  - Goal: support 32GB servers for all users

Thank You!