

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

TOOLS & APPLICATIONS

Learn About Tools & Applications.

Simulation SimCenter Tools Visualization Analysis Hazard Apps Utilities My Apps

HVSweb **Jupyter** **MATLAB**

My Data

Name	Size
.pynb_checkpoints	4.0 kB
.Trash	4.0 kB
applications	4.0 kB
archive	4.0 kB
arewealoe.txt	12.0 bytes
bin	4.0 kB
Capture1.JPG	19.3 kB

The Jupyter Notebook is a web application that allows you to create and share documents that contain live code, equations, visualizations and explanatory text. Uses include: data cleaning and transformation, numerical simulation, statistical modeling, machine learning and [much more](#).

Please visit our [help documentation](#) for more information.

[Launch](#)

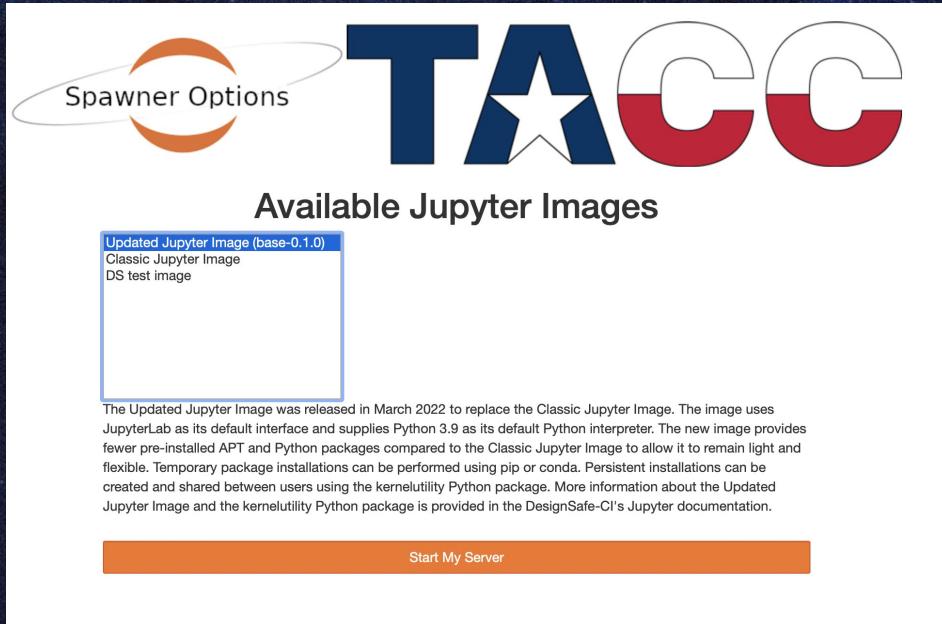
NOTE: This Jupyter instance will terminate after being idle for 3 days.

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



The screenshot shows the TACC Spawner Screen. At the top, there is a "Spawner Options" button with an orange circular icon. Below it is the TACC logo, which consists of a blue star inside a blue "T" and the letters "ACC" in red. The main title is "Available Jupyter Images". Below this, there is a list of images:

- Updated Jupyter Image (base-0.1.0) (selected, highlighted with a blue border)
- Classic Jupyter Image
- DS test image

A detailed description of the "Updated Jupyter Image" follows:

The Updated Jupyter Image was released in March 2022 to replace the Classic Jupyter Image. The image uses JupyterLab as its default interface and supplies Python 3.9 as its default Python interpreter. The new image provides fewer pre-installed APT and Python packages compared to the Classic Jupyter Image to allow it to remain light and flexible. Temporary package installations can be performed using pip or conda. Persistent installations can be created and shared between users using the kernelutility Python package. More information about the Updated Jupyter Image and the kernelutility Python package is provided in the DesignSafe-CI's Jupyter documentation.

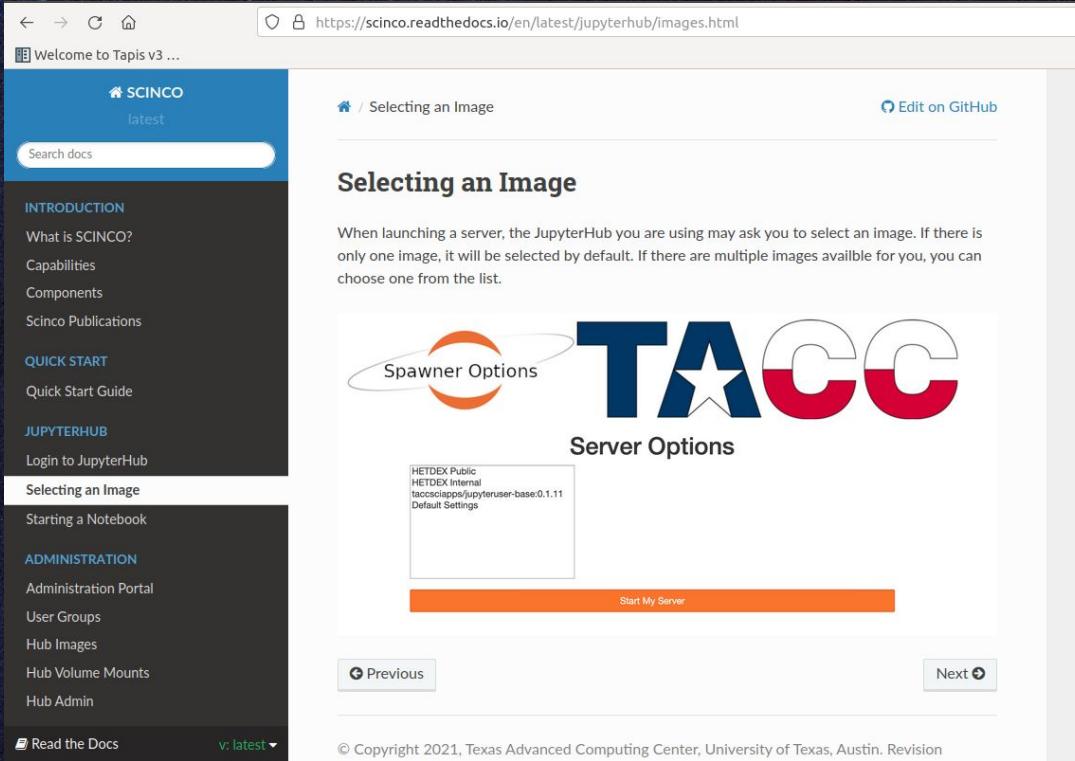
[Start My Server](#)

- 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=wpbRGdFzlFw>
!pip install kernelutility

Full Documentation Site Now Available



The screenshot shows a web browser displaying the SCINCO documentation site at <https://scinco.readthedocs.io/en/latest/jupyterhub/images.html>. The page title is "Selecting an Image". The left sidebar contains a navigation menu with sections: INTRODUCTION (What is SCINCO?, Capabilities, Components, Scinco Publications), QUICK START (Quick Start Guide), JUPYTERHUB (Login to JupyterHub, Selecting an Image, Starting a Notebook), and ADMINISTRATION (Administration Portal, User Groups, Hub Images, Hub Volume Mounts, Hub Admin). The "Selecting an Image" link in the JUPYTERHUB section is highlighted with a red box. The main content area shows the "Selecting an Image" page, which includes the TACC logo, a "Spawner Options" section with a dropdown menu showing "HETDEX Public", "HETDEX Internal", "taccslapps/jupyteruser-base:0.1.11", and "Default Settings", and a large orange "Start My Server" button. Navigation buttons for "Previous" and "Next" are also present.

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

The screenshot shows a web browser window with the URL <https://jupyter-admin.jupyter.tacc.cloud/images/>. The page title is "Welcome to Tapis v3 ...". The main content area is titled "JupyterHub Images" and contains the following text: "Here you can configure the images available to the JupyterHub spawner." Below this, there is a table with two columns: "Display Name" and "Image Name". The table rows are as follows:

Display Name	Image Name
taccsciapps/jupyteruser-base:0.1.11	taccsciapps/jupyteruser-base:0.1.11
Default Settings	taccsciapps/jupyteruser-tc:1.1.0
HETDEX Public	hetdex/hetdex-jupyter:0.40

At the bottom left of the table is a blue button labeled "+ New Image".

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!