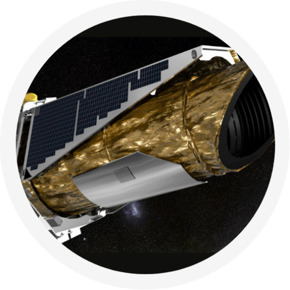
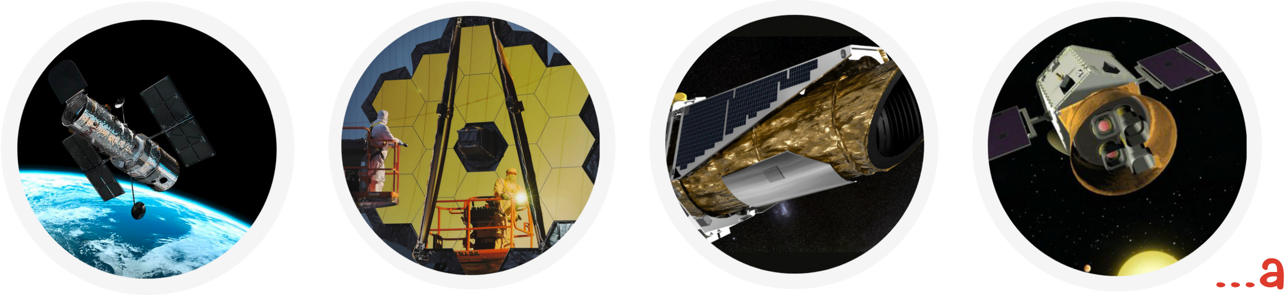


An Update on the MAST Science Platform

Jonathan Hargis

Deputy Branch Manager, Archive Sciences Branch
Mikulski Archive for Space Telescopes / STScI

+ Ivelina Momcheva, Arfon Smith, Josh Peek, Mike Fox
+ Jacob Matuskey, Christian Mesh, Erik Tollerud, Steve Crawford



...and now TESS!

An Update on the MAST Science Platform

Jonathan Hargis

Deputy Branch Manager, Archive Sciences Branch
Mikulski Archive for Space Telescopes / STScI

+ Ivelina Momcheva, Arfon Smith, Josh Peek, Mike Fox
+ Jacob Matuskey, Christian Mesh, Erik Tollerud, Steve Crawford

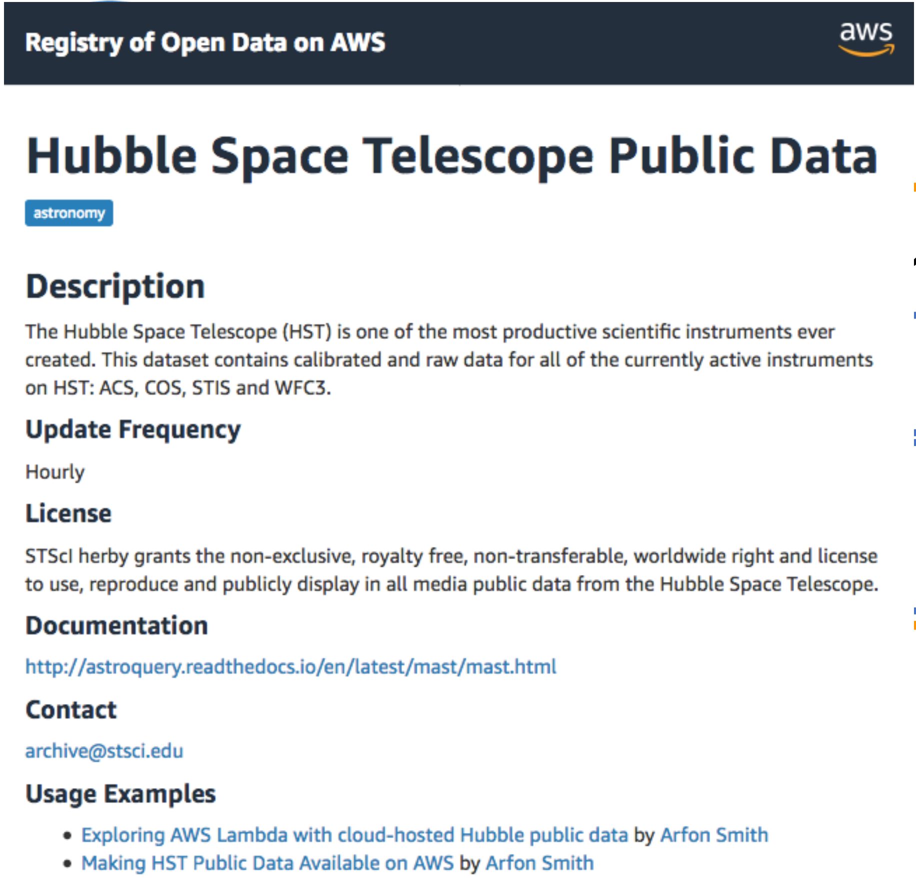
Talk Overview

- Brief overview of technology stack
- Demo and walkthrough
- Deploying your own Science Platform
- Challenges and Future Directions

Common technologies, many implementations

docker
jupyterhub
astropy
kubernetes
amazon web services™
IVOA
GitHub
esa
ipac
SciServer
NOAO DATALAB
STScI | SPACE TELESCOPE SCIENCE INSTITUTE
LSST
Large Synoptic Survey Telescope
NATIONAL DATA SERVICE PARTICIPANT
binder

STScI Science Platform



Registry of Open Data on AWS

Hubble Space Telescope Public Data

astronomy

Description

The Hubble Space Telescope (HST) is one of the most productive scientific instruments ever created. This dataset contains calibrated and raw data for all of the currently active instruments on HST: ACS, COS, STIS and WFC3.

Update Frequency

Hourly

License

STScI herby grants the non-exclusive, royalty free, non-transferable, worldwide right and license to use, reproduce and publicly display in all media public data from the Hubble Space Telescope.

Documentation

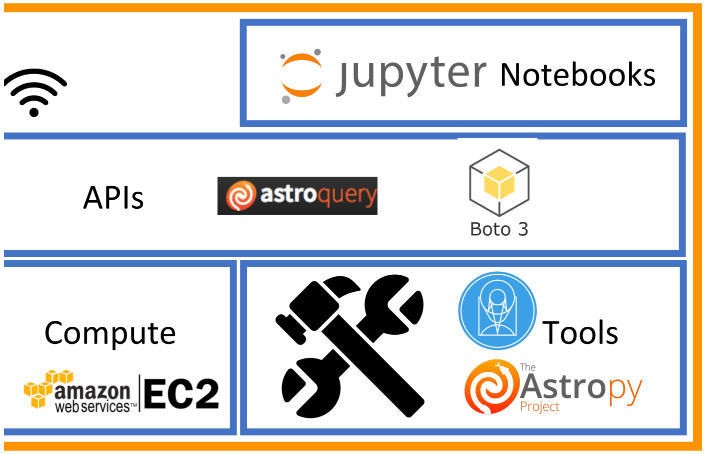
http://astroquery.readthedocs.io/en/latest/mast/mast.html

Contact

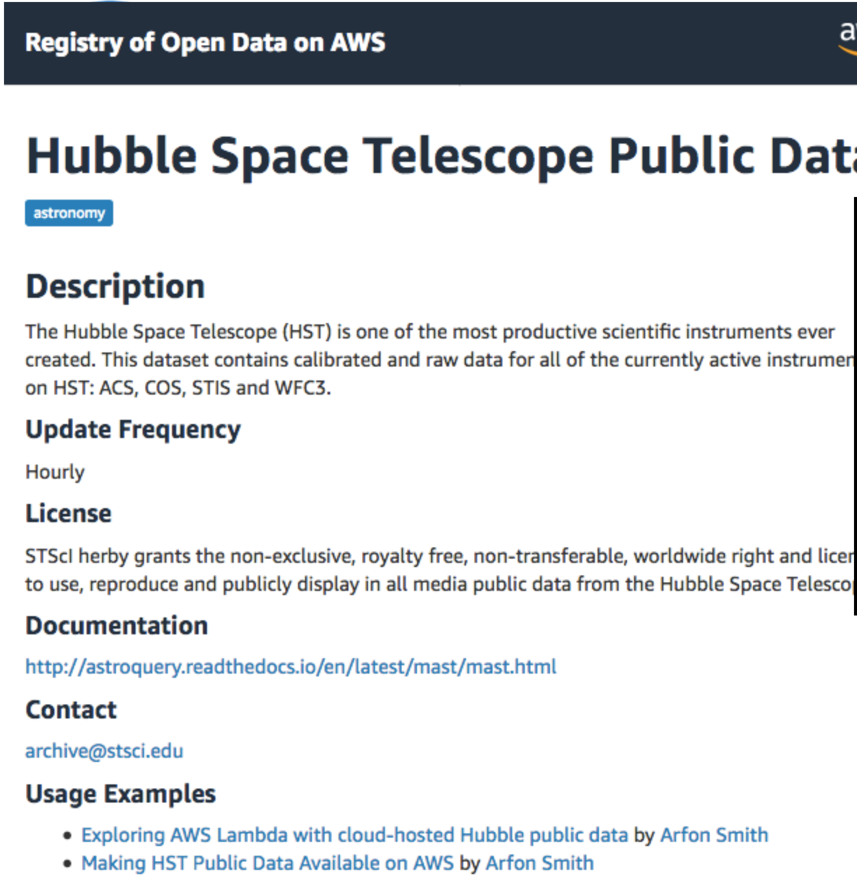
archive@stsci.edu

Usage Examples

- Exploring AWS Lambda with cloud-hosted Hubble public data by Arfon Smith
- Making HST Public Data Available on AWS by Arfon Smith



STScI Science Platform



Registry of Open Data on AWS

Hubble Space Telescope Public Data ...and TESS!

astronomy

Description

The Hubble Space Telescope (HST) is one of the most productive scientific instruments ever created. This dataset contains calibrated and raw data for all of the currently active instrumen on HST: ACS, COS, STIS and WFC3.

Update Frequency

Hourly

License

STScI herby grants the non-exclusive, royalty free, non-transferable, worldwide right and licen to use, reproduce and publicly display in all media public data from the Hubble Space Telesco

Documentation

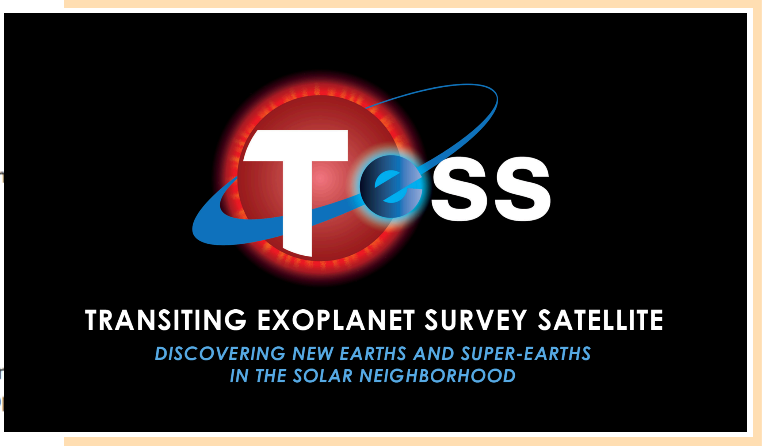
http://astroquery.readthedocs.io/en/latest/mast/mast.html

Contact

archive@stsci.edu

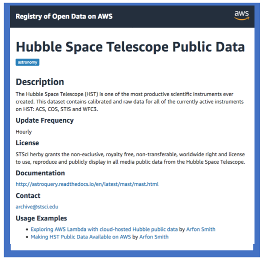
Usage Examples

- Exploring AWS Lambda with cloud-hosted Hubble public data by Arfon Smith
- Making HST Public Data Available on AWS by Arfon Smith


TESS
TRANSITING EXOPLANET SURVEY SATELLITE
DISCOVERING NEW EARTHS AND SUPER-EARTHS
IN THE SOLAR NEIGHBORHOOD

Science Platform Architecture

STScI



Registry of Open Data on AWS

Hubble Space Telescope Public Data

astronomy

Description

The Hubble Space Telescope (HST) is one of the most productive scientific instruments ever created. This dataset contains calibrated and raw data for all of the currently active instruments on HST: ACS, COS, STIS and WFC3.

Update Frequency

Hourly

License

STScI herby grants the non-exclusive, royalty free, non-transferable, worldwide right and license to use, reproduce and publicly display in all media public data from the Hubble Space Telescope.

Documentation

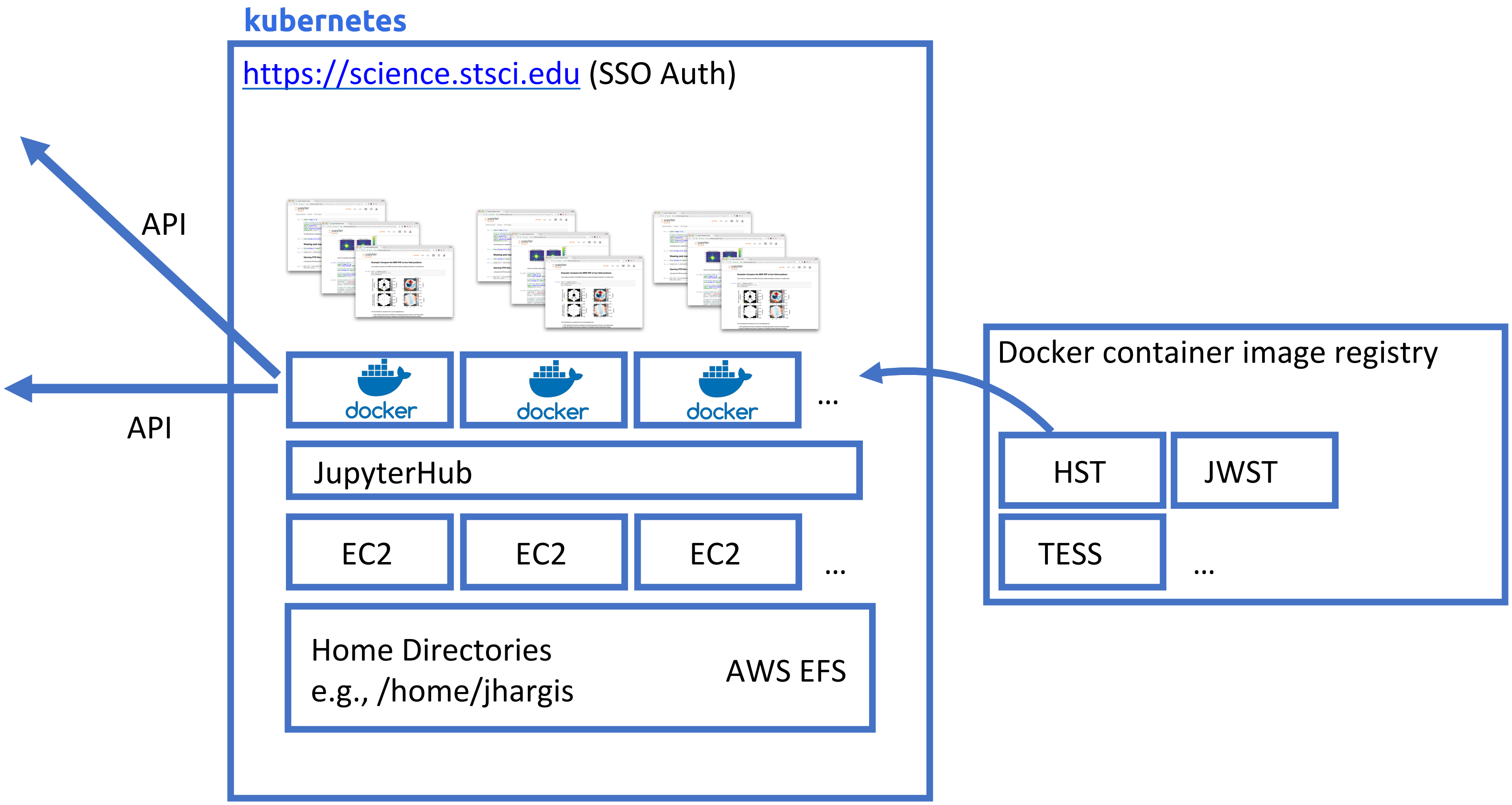
http://astroquery.readthedocs.io/en/latest/mast/mast.html

Contact

archive@stsci.edu

Usage Examples

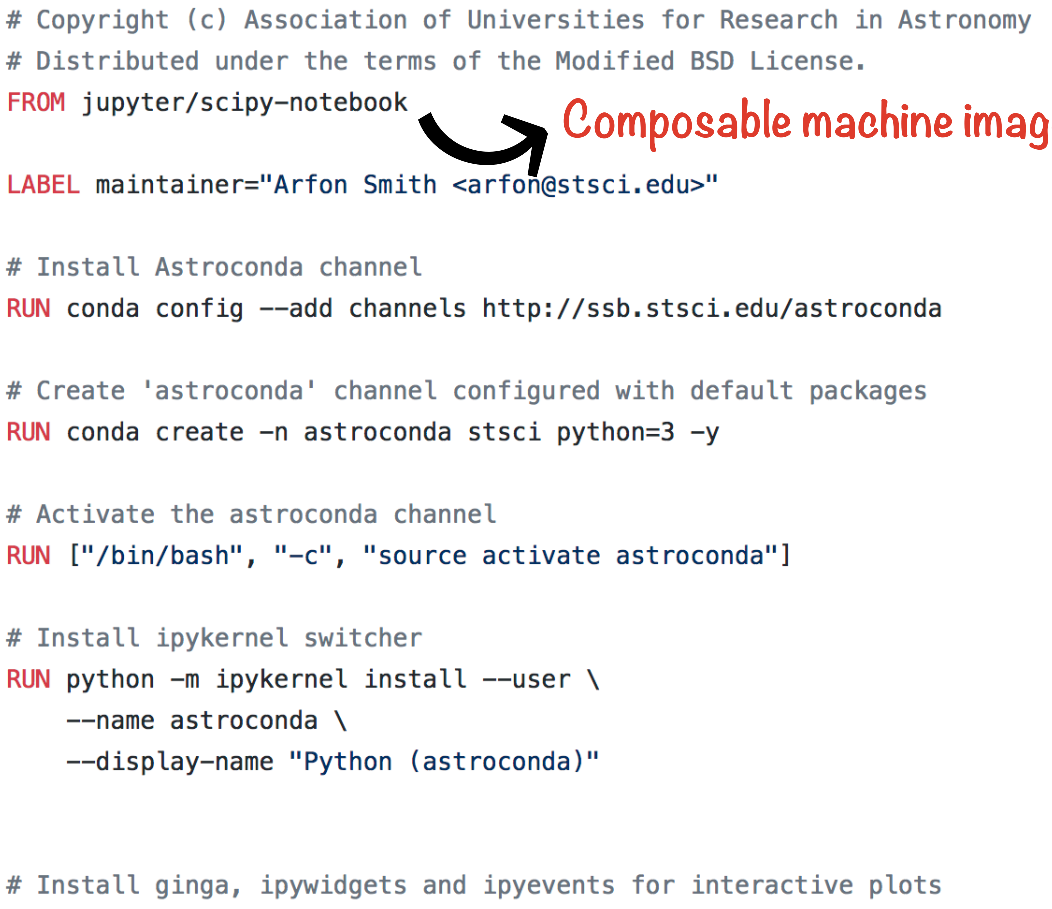
- Exploring AWS Lambda with cloud-hosted Hubble public data by Arfon Smith
- Making HST Public Data Available on AWS by Arfon Smith



kubernetes

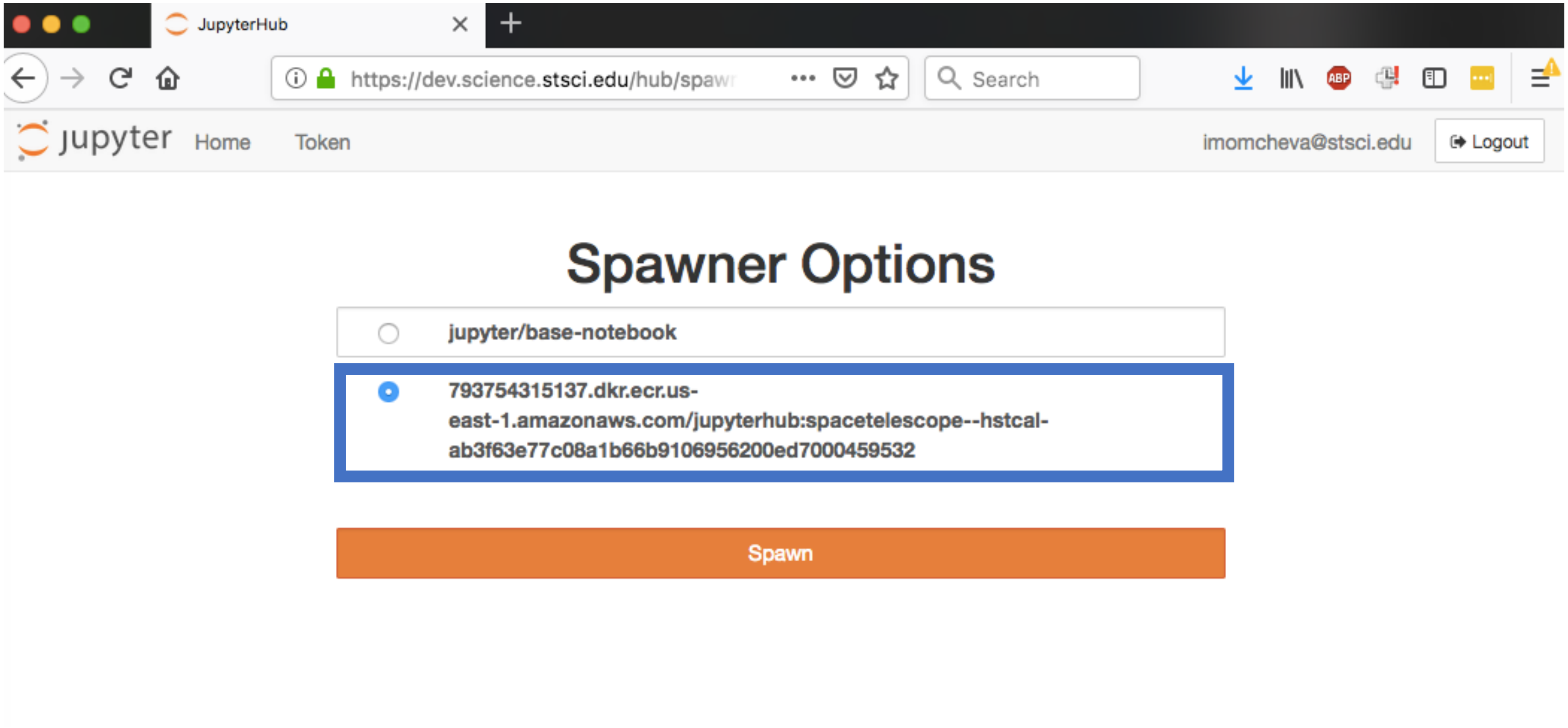
https://science.stsci.edu (SSO Auth)

Docker (Container Files)



```
# Copyright (c) Association of Universities for Research in Astronomy
# Distributed under the terms of the Modified BSD License.
FROM jupyter/scipy-notebook

LABEL maintainer="Arfon Smith <arfon@stsci.edu>"

# Install Astroconda channel
RUN conda config --add channels http://ssb.stsci.edu/astroconda

# Create 'astroconda' channel configured with default packages
RUN conda create -n astroconda stsci python=3 -y

# Activate the astroconda channel
RUN ["/bin/bash", "-c", "source activate astroconda"]

# Install ipykernel switcher
RUN python -m ipykernel install --user \
    --name astroconda \
    --display-name "Python (astroconda)"


# Install ginga, ipywidgets and ipyevents for interactive plots
```

Composable machine images: FROM lsstsqre/pipeline



Spawner Options

○ jupyter/base-notebook

● 793754315137.dkr.ecr.us-east-1.amazonaws.com/jupyterhub:spacetelescope--hstcal-ab3f63e77c08a1b66b9106956200ed7000459532

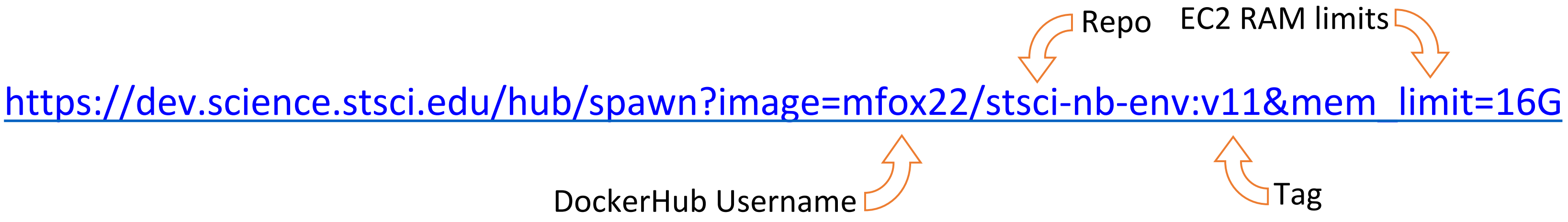
Simplifying Access: Magic Links

We have created URLs of the form:

https://dev.science.stsci.edu/hub/spawn?image=mfox22/stsci-nb-env:v11&mem_limit=16G

Simplifying Access: Magic Links

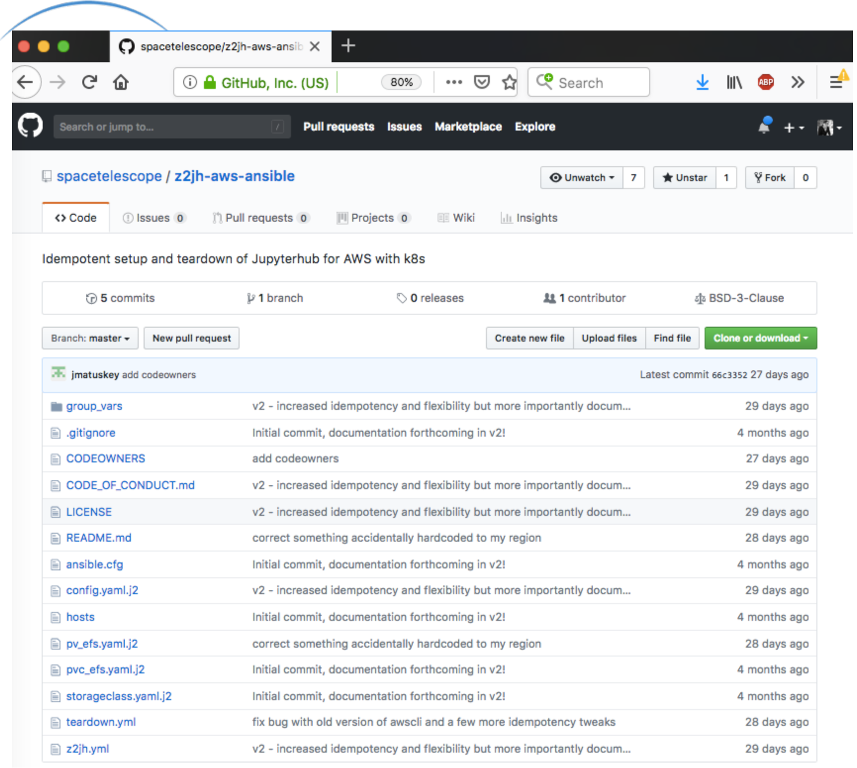
We have created URLs of the form:



https://dev.science.stsci.edu/hub/spawn?image=mfox22/stsci-nb-env:v11&mem_limit=16G

Reference Deployment

https://github.com/spacetelescope/z2jh-aws-ansible



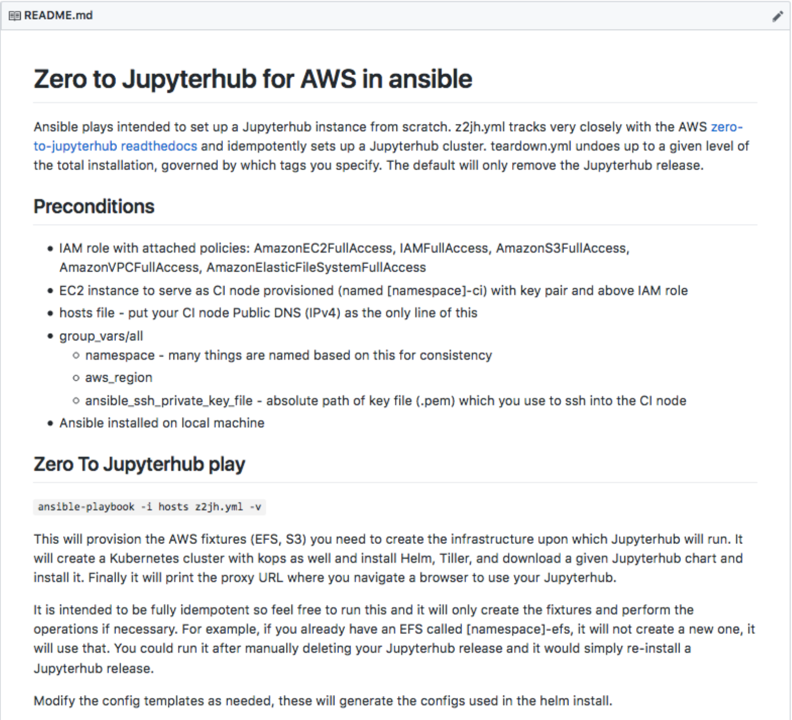
spacetelescope / z2jh-aws-ansible

Idempotent setup and teardown of Jupyterhub for AWS with k8s

5 commits | 1 branch | 0 releases | 1 contributor | BSD-3-Clause

jmatuskey add codeowners — Latest commit 66c3352 27 days ago

group_vars	v2 - increased idempotency and flexibility but more importantly docum...	29 days ago
.gitignore	Initial commit, documentation forthcoming in v2!	4 months ago
CODEOWNERS	add codeowners	27 days ago
CODE_OF_CONDUCT.md	v2 - increased idempotency and flexibility but more importantly docum...	29 days ago
LICENSE	v2 - increased idempotency and flexibility but more importantly docum...	29 days ago
README.md	correct something accidentally hardcoded to my region	28 days ago
ansible.cfg	Initial commit, documentation forthcoming in v2!	4 months ago
config.yaml.j2	v2 - increased idempotency and flexibility but more importantly docum...	29 days ago
hosts	Initial commit, documentation forthcoming in v2!	4 months ago
pv_efs.yaml.j2	correct something accidentally hardcoded to my region	28 days ago
pvc_efs.yaml.j2	Initial commit, documentation forthcoming in v2!	4 months ago
storageclass.yaml.j2	Initial commit, documentation forthcoming in v2!	4 months ago
teardown.yml	fix bug with old version of awscli and a few more idempotency tweaks	28 days ago
z2jh.yml	v2 - increased idempotency and flexibility but more importantly docum...	29 days ago



Zero to Jupyterhub for AWS in ansible

Ansible plays intended to set up a Jupyterhub instance from scratch. z2jh.yml tracks very closely with the AWS zero-to-jupyterhub readthedocs and idempotently sets up a Jupyterhub cluster. teardown.yml undoes up to a given level of the total installation, governed by which tags you specify. The default will only remove the Jupyterhub release.

Preconditions

- IAM role with attached policies: AmazonEC2FullAccess, IAMFullAccess, AmazonS3FullAccess, AmazonVPCFullAccess, AmazonElasticFileSystemFullAccess
- EC2 instance to serve as CI node provisioned (named [namespace]-ci) with key pair and above IAM role
- hosts file - put your CI node Public DNS (IPv4) as the only line of this
- group_vars/all
  - namespace - many things are named based on this for consistency
  - aws_region
  - ansible_ssh_private_key_file - absolute path of key file (.pem) which you use to ssh into the CI node
- Ansible installed on local machine

Zero To Jupyterhub play

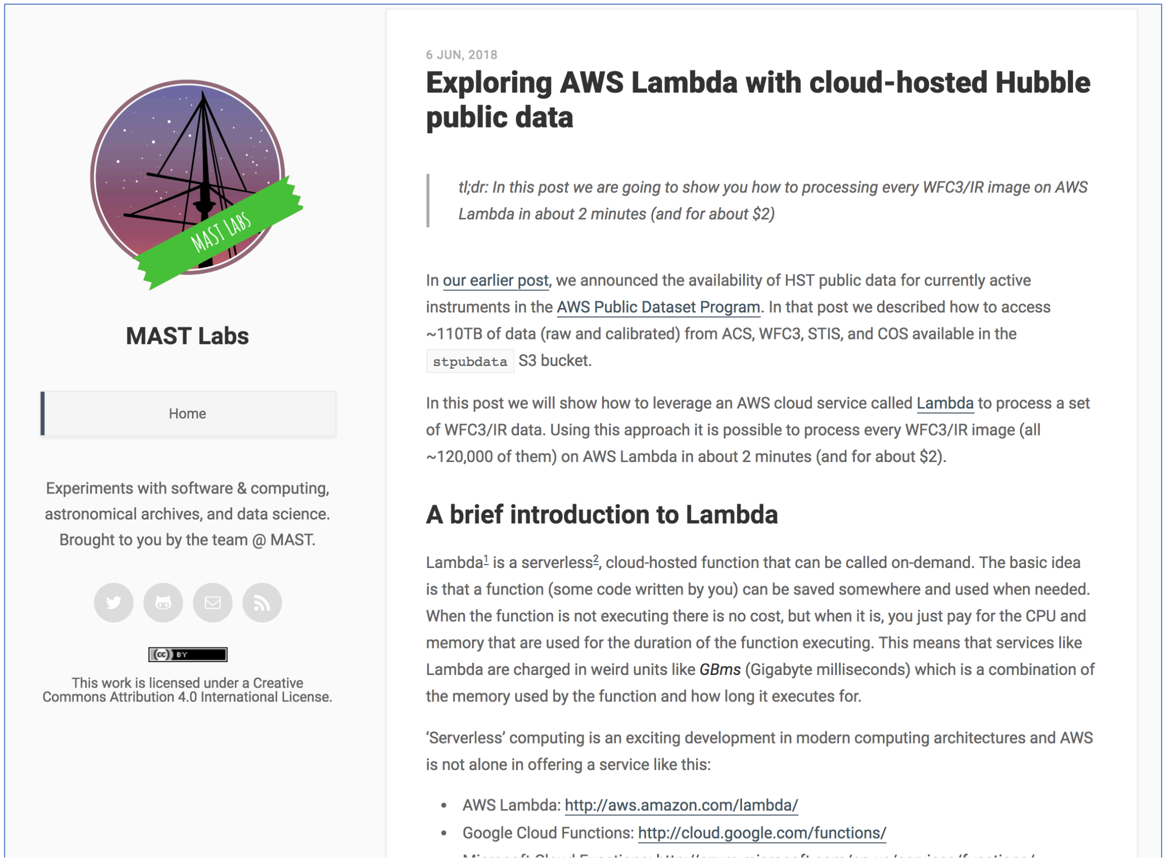
```
ansible-playbook -i hosts z2jh.yml -v
```

This will provision the AWS fixtures (EFS, S3) you need to create the infrastructure upon which Jupyterhub will run. It will create a Kubernetes cluster with kops as well and install Helm, Tiller, and download a given Jupyterhub chart and install it. Finally it will print the proxy URL where you navigate a browser to use your Jupyterhub.

It is intended to be fully idempotent so feel free to run this and it will only create the fixtures and perform the operations if necessary. For example, if you already have an EFS called [namespace]-efs, it will not create a new one, it will use that. You could run it after manually deleting your Jupyterhub release and it would simply re-install a Jupyterhub release.

Modify the config templates as needed, these will generate the configs used in the helm install.

MAST Labs Blog



MAST Labs

Home

Experiments with software & computing, astronomical archives, and data science. Brought to you by the team @ MAST.

This work is licensed under a Creative Commons Attribution 4.0 International License.

6 JUN, 2018

Exploring AWS Lambda with cloud-hosted Hubble public data

> *tl;dr: In this post we are going to show you how to processing every WFC3/IR image on AWS Lambda in about 2 minutes (and for about $2)*

In our earlier post, we announced the availability of HST public data for currently active instruments in the AWS Public Dataset Program. In that post we described how to access ~110TB of data (raw and calibrated) from ACS, WFC3, STIS, and COS available in the `stpubdata` S3 bucket.

In this post we will show how to leverage an AWS cloud service called Lambda to process a set of WFC3/IR data. Using this approach it is possible to process every WFC3/IR image (all ~120,000 of them) on AWS Lambda in about 2 minutes (and for about $2).

A brief introduction to Lambda

Lambda[1] is a serverless[2], cloud-hosted function that can be called on-demand. The basic idea is that a function (some code written by you) can be saved somewhere and used when needed. When the function is not executing there is no cost, but when it is, you just pay for the CPU and memory that are used for the duration of the function executing. This means that services like Lambda are charged in weird units like *GBms* (Gigabyte milliseconds) which is a combination of the memory used by the function and how long it executes for.

'Serverless' computing is an exciting development in modern computing architectures and AWS is not alone in offering a service like this:

- AWS Lambda: http://aws.amazon.com/lambda/
- Google Cloud Functions: http://cloud.google.com/functions/
- Microsoft Cloud Functions: http://azure.microsoft.com/en-us/services/functions/

https://mast-labs.stsci.io/

- How to access HST public data set
- How to use AWS Lambda to do source detection at-scale

Challenges and Future Directions

Challenges and Future Directions

Future Directions

- Bringing the user / code to the data *processing pipeline*
- Use of science platform for data processing pipeline operations
- Enable integration with simulations and community contributed high-level science products
- Aiming to have a beta version of the platform available for TESS Data Workshop @ STScI

Challenges

- Bringing the user / code to the data *processing pipeline*
- A science platform is only as good as the Notebooks you provide
- Notebooks are not a one-sized-fits-all solution
- Collaborative workflows currently not possible
- Billing model & user quotas
- User management: privacy vs. security
- Running batch compute

STScI | SPACE TELESCOPE SCIENCE INSTITUTE

Thank you!

Jonathan Hargis
jhargis@stsci.edu