Jupyter Notebooks at scale for Gateways and Workshops

Andrea Zonca SDSC

Slides and links: bit.ly/ecss_zonca
Jupyter Notebooks

LIGO Open Science Center Gravitational Waves detection:

- View Notebook
- Execute Notebook in the cloud
Jupyter notebook local

Browser tabs

Laptop

http

zeroMQ

Notebook
Local Web Application

Py3 Kernel

Py2.7 Kernel

.jupyter JSON files

Kernel
Jupyter notebook remote

Laptop

Server

https + password

Jupyter Notebook

Web Application

Py3 Kernel

Py2.7 Kernel

Kernel

.jupyter JSON files
Jupyterhub multi user

Laptop 1

Laptop 2

Server

http proxy

https + local accounts

Jupyterhub

Notebook

Python 3

Python 2.7

Julia

Python 3

Jupyterhub

https + local accounts

Server

http proxy

Jupyterhub

Notebook

Python 3

Python 2.7

Julia

Python 3
Default JupyterHub deployment

- Automatically deploy with Ansible: [github.com/jupyterhub/jupyterhub-deploy-teaching](https://github.com/jupyterhub/jupyterhub-deploy-teaching)
- 1 server with Ubuntu
- NGINX proxy + HTTPS with letsencrypt
- Authentication with Local Unix accounts
- Launch Notebooks as normal processes
Authenticators

- XSEDE credentials
- Globus
- Github
- CILogon (Campus credentials)
- Google Accounts
- many more...
Spawners

- batchspawner: launch via SLURM/PBS/SGE...
- SwarmSpawner: launch in Docker containers via Swarm
- KubeSpawner: launch on Kubernetes pods
Spawn Notebooks on Supercomputers

- Authentication with XSEDE or Globus
- batchspawner
  - SSH with community account or GSISSH with XSEDE token
  - Submit job to SLURM
  - Wait for job to start
  - Connect back to JupyterHub with SSH tunnel
- Link to tutorial
Cloud deployment JetStream/Swarm

- **Master node, Docker containers:**
  - NGINX with letsencrypt
  - Jupyterhub
  - NFS server - persistent homes, user quota

- **SwarmSpawner**
  - Managed by Docker Swarm over many servers

- **Scales to tens of servers, <100 users**

- [Link to tutorial](#)
Laptop 1

Google/Github

Master Node

Kubernetes

Rook Ceph

JS volumes: /vol_b

Worker Node

Kubernetes

Rook Ceph

/vol_b

Worker Node

Kubernetes

Rook Ceph

/vol_b

Laptop 2
JetStream Master Node

NGINX

JupyterHub

JHub Proxy

Worker Node

Notebook

Worker Node

Notebook

Worker Node

Notebook
JetStream with Kubernetes

- Kubernetes services (zero-to-jupyterhub):
  - NGINX with letsencrypt HTTPS
  - Jupyterhub
  - Jupyterhub Proxy
  - Automatically restarted if die

- KubeSpawner
  - Launches user Notebooks containers
  - Persistent data with Ceph distributed File System

- Scales to hundreds of servers, ~1000 users

- Link to tutorial
Thanks

- Contact me for anything related to Jupyter on XSEDE
- zonca@sdsc.edu