Monthly PaNGalaxy Community Meeting

4 November 2022

ORNL is managed by UT-Battelle LLC for the US Department of Energy
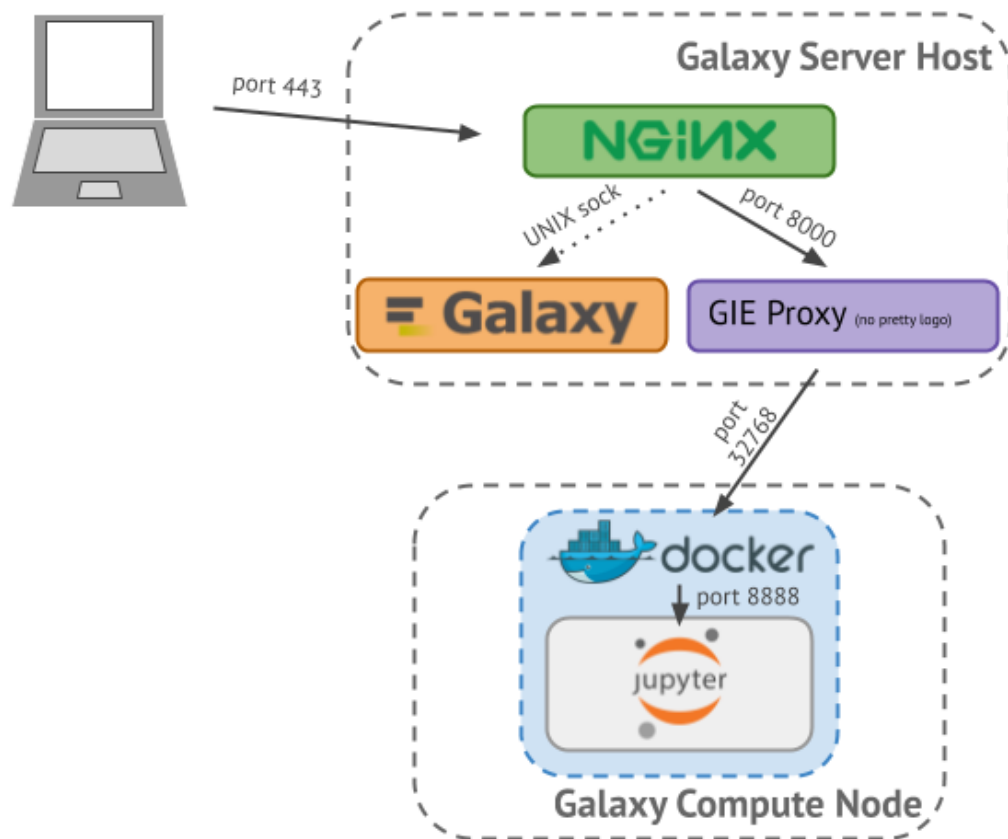
U.S. DEPARTMENT OF ENERGY

Agenda

- Welcome for our colleagues from Muon Spectroscopy!
- Prototyping tools using Jupyter notebooks (ORNL)
- Deploying Galaxy on ORNL infrastructure (ORNL)
- Community topics:
  - Common tools/visualizations/datatypes
  - Event monitoring service
  - Muon Spectroscopy community
- Other topics
- Next meeting

OAK RIDGE
National Laboratory

Jupyter notebooks – our experience

Galaxy Server Host
- port 443
- NGINX
- UNIX sock
- port 8000
- Galaxy
- GIE Proxy (no pretty logo)
- port 32768

docker
port 8888
jupyter

Galaxy Compute Node

- We have only used Jupyter notebooks so far
  - Rstudio, other environments?

- Created own Docker image (Jupyter Lab + Mantid)

- Created own tools
  - generic one and a specific neutron tool
  - pushed to https://github.com/PaNGalaxy/neutron-tools

- Gained some experience in how to use them
  - no user feedback yet

https://training.galaxyproject.org/training-material/topics/admin/tutorials/interactive-tools/tutorial.html

OAK RIDGE
National Laboratory

Jupyter notebooks – integration into a workflow

Different modes

- Use as a standalone interactive tool
  - input/output via Bioblend
  - input/output via the tool

- Use as a standalone non-interactive tool

- Use as a part of a workflow
  - non-interactive
  - interactive?

- Reuse a notebook

- Create a regular tool out of a notebook

🔧 **Interactive Jupyter Notebook ORNL** (Galaxy Version 0.1)

Jupyter notebook

Start with a fresh notebook

Select 'Start with a fresh notebook' to use a default one.

Execute notebook noninteractively

⦿ No

This option is useful in workflows when you just want to execute a notebook and not

Use dataset collection for input
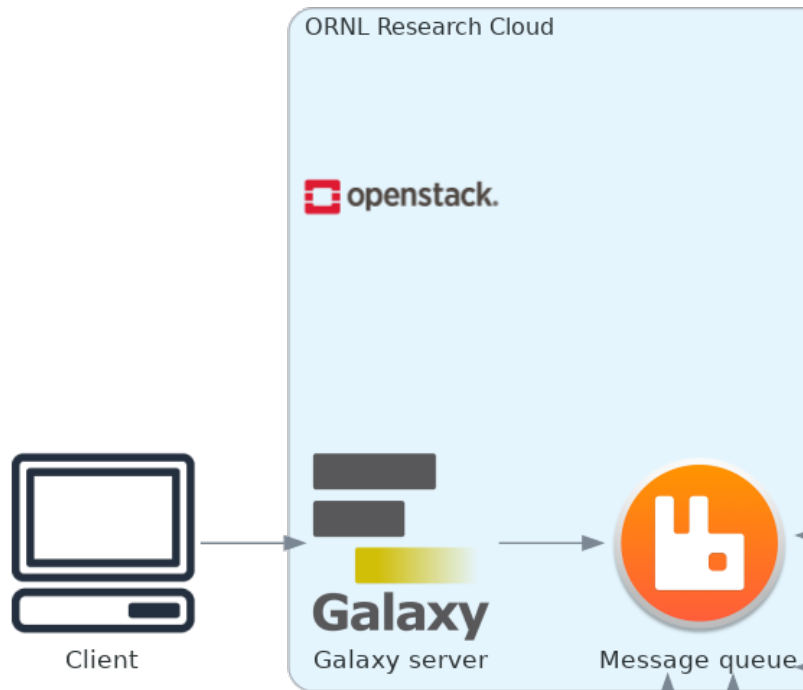
⦿ No

Datasets

➕ Insert Datasets

Type of output

Single Dataset

✔ Execute

OAK RIDGE
National Laboratory

# ORNL Neutrons Galaxy – Current Architecture

- ORNL Research Cloud
  - On-premise OpenStack cloud instance

- Galaxy Server
  - Virtual machine (VM) in ORNL Research Cloud

- Message Queue
  - RabbitMQ
  - VM in ORNL Research Cloud
  - Used to queue up work for compute resources

OAK RIDGE
National Laboratory

ORNL Neutrons Galaxy – Current Deployment
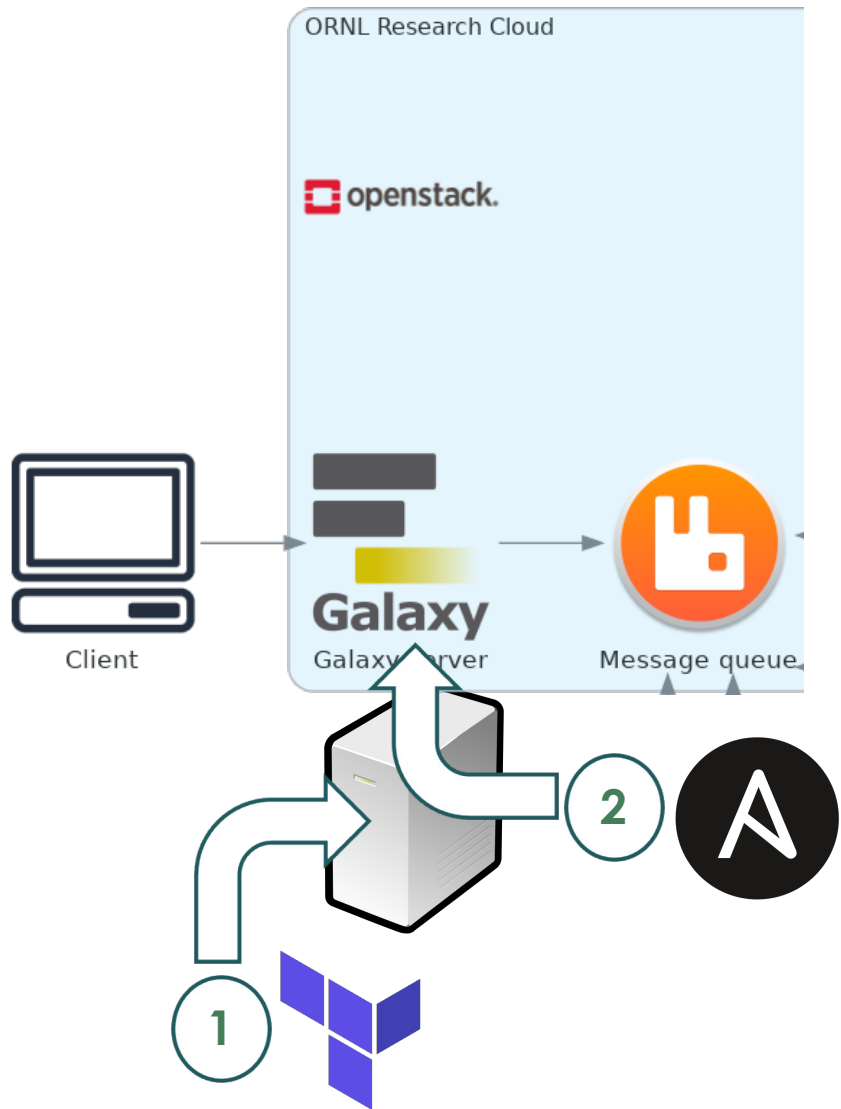
1. Provision Infrastructure
   - Use Terraform, an Infrastructre-as-Code tool
   - Creates the cloud resource (I.e. VM + network rules / security group rules)

2. Configure / Install Galaxy
   - Configured via modified Galaxy Ansible playbooks
     - Examples:
       - https://github.com/galaxyproject/ansible-galaxy
       - https://github.com/galaxyproject/ansible-nginx
   - Modifications include...
     - Adding wildcard TLS/SSL certificates to server for the Jupyter Notebooks capabilities
     - Fixes for file permissions and ownership
     - Still others needed for more reliable installs!

- Galaxy containers for Development
  - Maintain Galaxy containers we use for some development
  - Not currently used in our deployments

OAK RIDGE
National Laboratory

ORNL Neutrons Galaxy – Planned Architecture Setup
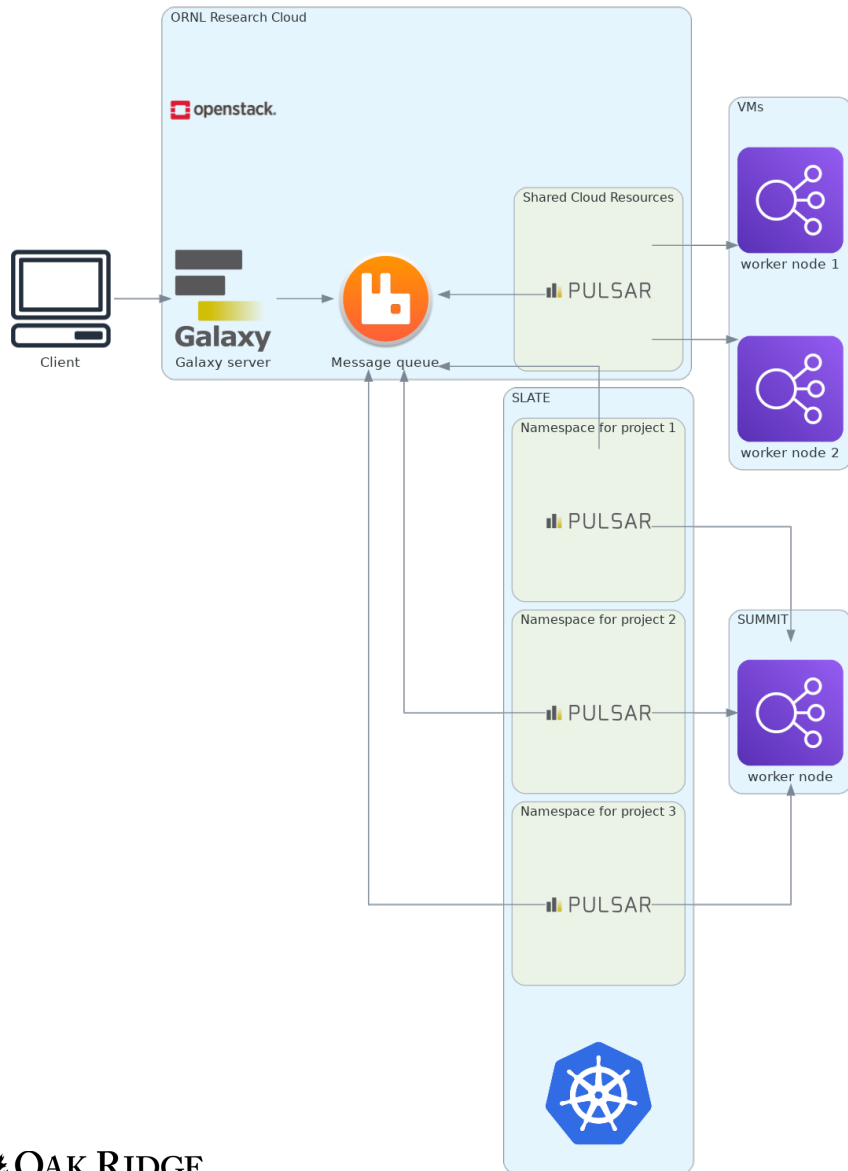
- Pulsar
  - Manage work issued by Galaxy to compute resources
  - Matches requirements of hardware to parts of the workflows

- Compute in the Cloud
  - Easy setup, just VMs
  - CPU and CPU + GPU available in our cloud
  - Good for small to moderate jobs

- Compute on HPC
  - Slate Cluster
    - Kubernetes (well, OpenShift...) to host containers able to submit jobs to HPC resources (i.e. Summit for now, Frontier eventually)
    - Ingress / egress of commands and data to shared HPC User Facility
  - Summit
    - Where actual compute will run
    - Behind a job scheduler
  - Good for large jobs; User must setup the Pulsar containers in their Namespace of the Project

OAK RIDGE
National Laboratory

ORNL Neutrons Galaxy – Future Deployment Discussions
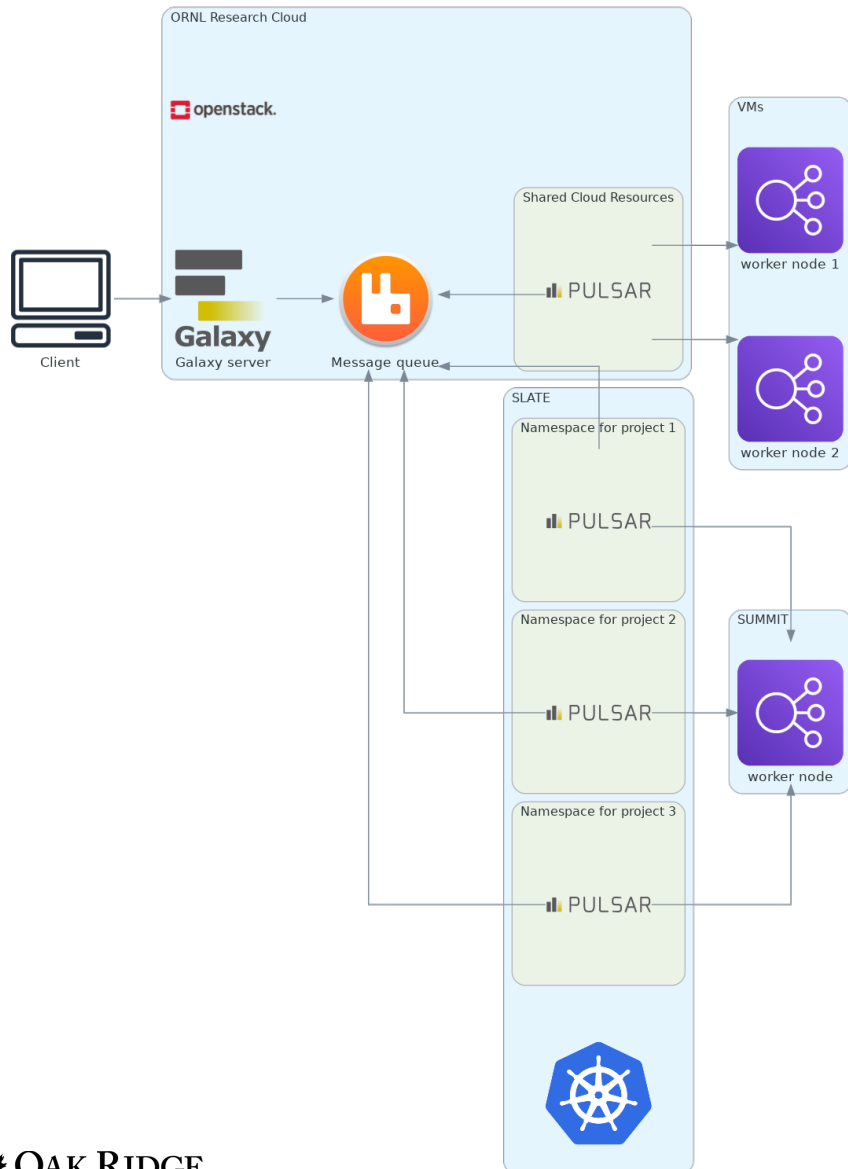
- Containerization?
  - We use containers in development. Can we use those for deployment?
  - Have container orchestration setup in ORNL Research Cloud (i.e. Kubernetes clusters)
  - Unknowns:
    - Accessing the Neutrons Sciences NFS, which we don't own via a Pod in a Kubernetes cluster (persistent volume with the NFS?) Permissions?
    - Tools run on central server and spawn containers. Docker-in-docker scenario?
  - Jupyter Notebooks
    - Still able to spawn inside container?
    - Handling the wildcard certificate: How to set that up with ingress controllers and dynamic domain names for notebooks?
- Scaling?
  - Galaxy is a monolith in many ways
    - Frontend + backend
    - Central server is not distributed
- Monitoring?
  - Maybe Prometheus + Node Exporter (+ Grafana)