A Reproducible and Automated Deployment of an HPC Application on a Private Cloud

Nam Pho and Josh Miller
Why HPC as an OSDC tool?

Increase Job Throughput
Via Intelligent Scheduling
Premise

1. How can we **build plugin structures** so that “data commons” can be extended and connected to other users and communities?

2. **What tools and applications can we build to facilitate data intensive discovery and/or cross-disciplinary discoveries?**

3. What open datasets can we add to public data commons like the OSDC that will facilitate data intensive cross-disciplinary discoveries?

4. How can we **enable persistent scientific research reproducibility** and do so in a sustainable way?

5. What sort of **apps/services** (ID services, metadata services) can we build over “data commons”?

6. **How can we make it easier for scientists (postdocs and grad students) to access and analyze large datasets?**

7. How can we help data scientists (**data engineers**, data analysts, and data stewards)?
Technology Stack

- Ansible
- OpenStack
- OpenLava
- Nature
Cluster Layout

Sullivan

[root@hn1 ~]# bsub ...