

Progress Report for PHY190047, November 2020

Purpose of allocation

This allocation was primarily requested to test XSEDE resources for a JupyterHub analysis environment for the CDMS dark-matter search collaboration. The intent is to use this environment both for analysis within CDMS and for outreach efforts. I am requesting additional SUs because I would like to ramp up user testing on the prototype.

Findings

Andrea Zonca has worked extensively with the PI to determine appropriate JupyterHub deployment on JetStream. There is now a working deployment and based on current use we estimate it will consume 20,000 CPU hours per month for six users. The system provides a ready-to-use environment for analyzers and provides a shared data store for collaboration data and default settings that allow easy connections to the collaboration git repositories. We tested the platform for light analysis use over the summer and fall and are beginning to ramp up to more data-intensive analysis tasks on the platform. We are requesting support from Andrea Zonca during this ramp-up.

As expected, commitment from the CDMS collaboration is critical for the popularity of this resource. The most critical support is making analysis data available. Two students have been working since the summer to make this data available and are making steady progress. Another critical support is user documentation. The collaboration has developed clear guides to logging in to the resource and has developed several tutorials that can be used on the system that walk through common data analysis tasks. More of these tutorials are under development.

In addition, some workflows would potentially benefit from Dask. Efforts to deploy Dask on the system are underway.

Documentation for JupyterHub deployment

- Docs about deployment: https://github.com/det-lab/jupyterhub-deploy-kubernetes-jetstream/tree/cdms_zonca
- Tutorial: <https://zonca.dev/2020/02/cvmfs-kubernetes.html>
- Link to the deployment: <https://supercdms.jetstream-cloud.org/hub/login>

Documentation for Dask development

- <https://github.com/det-lab/jupyterhub-deploy-kubernetes-jetstream/issues/7>