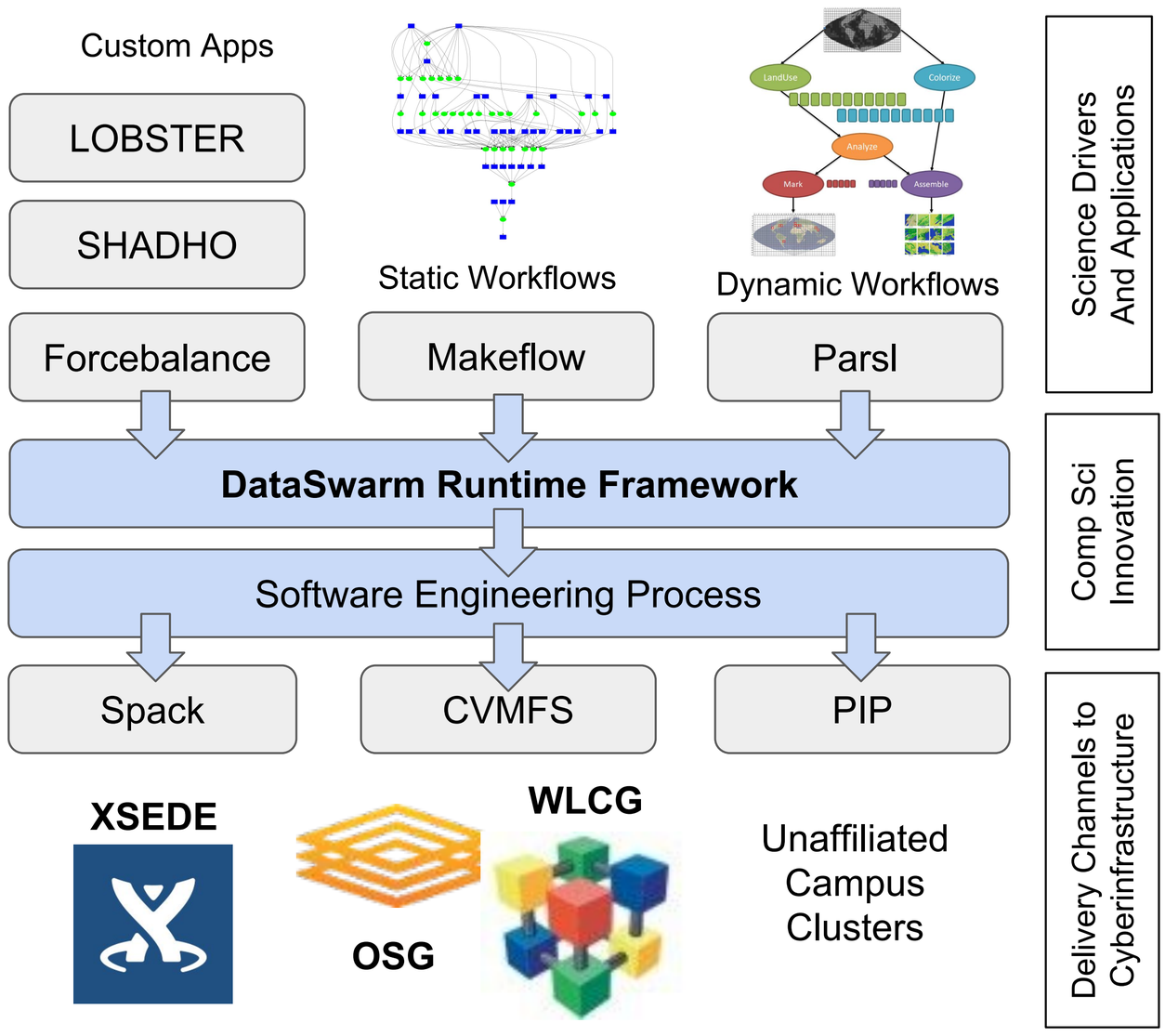
CSSI Element: DataSwarm: A User-Level Framework for Data Intensive Scientific Applications

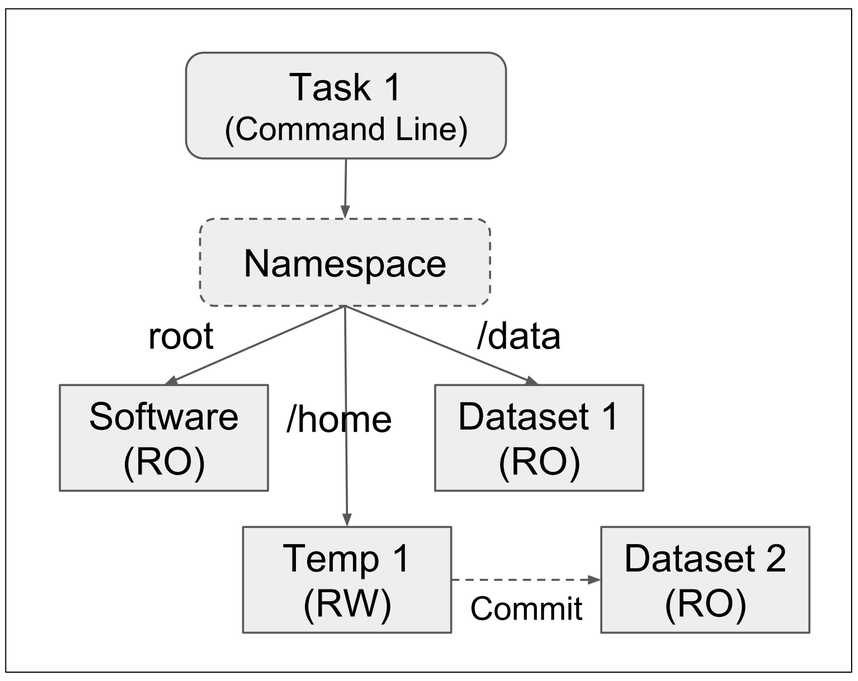
PI: Douglas Thain, University of Notre Dame

Award #: 1931348

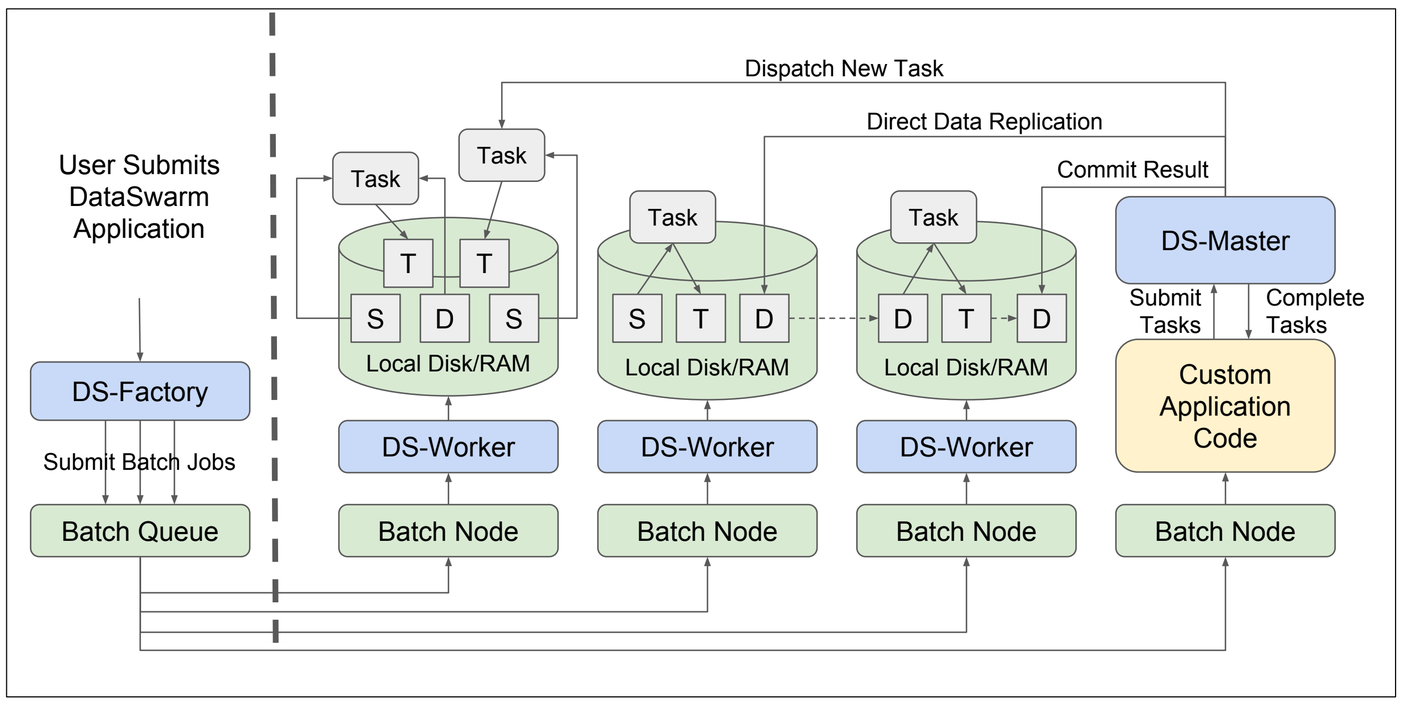
DataSwarm is a new application framework that will facilitate the construction of large, data intensive, scientific applications that must run on top of national cyberinfrastructure (OSG, XSEDE, campus clusters) . A key challenge of deploying large scale applications is that they often have complex data *requirements* and software dependencies that are not pre-installed on the target systems. Rather, the framework must bring these items along, deploy them into the system as needed, and schedule tasks relative to the current deployment. DataSwarm builds upon our technology and experience with prior CSSI projects **Makeflow** and **Work Queue.**



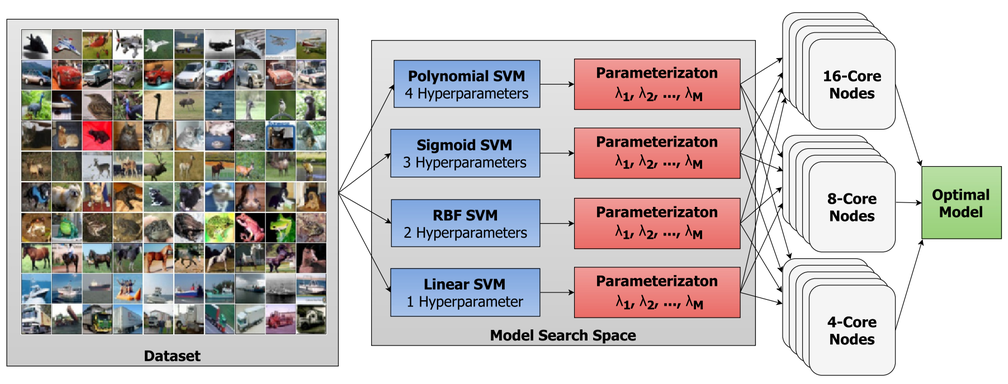
Molecular Task Composition



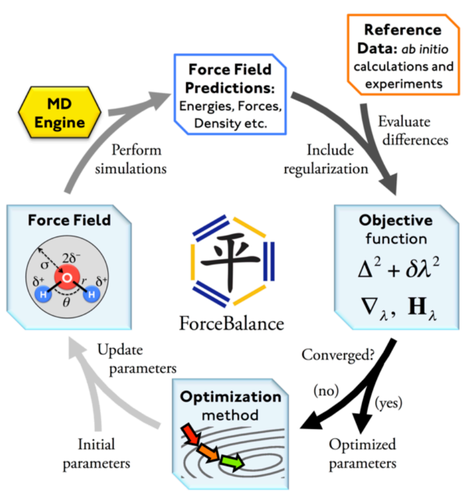
System Architecture



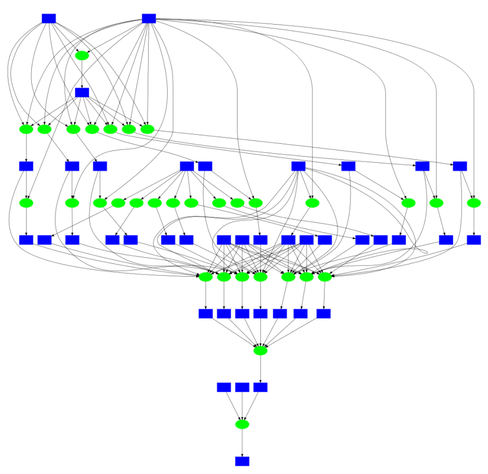
SHADHO (Hyperparameter Opt)



ForceBalance (Mol. Dynamics)



Bioinformatics Workflows



(Very) Early Results: Managed Software Delivery

Python: Import tensorflow