

Cloud native compute and data infrastructure for the Center for Bioenergy Innovation

Background

- The variety, velocity and volume of data generated by CBI researchers continues to grow at an exponential pace. Additionally, new requirements from DOE with regard to data sharing and access pose challenges for data management. We recognized the need for a cloud native solution to replace CADES, the cloud enabled VM based solution that has been the status quo. The shift to cloud native infrastructure provided by ORNL's Office of Leadership Computing (OLCF) represents a paradigm shift for data management and analytical workflows.

Approach

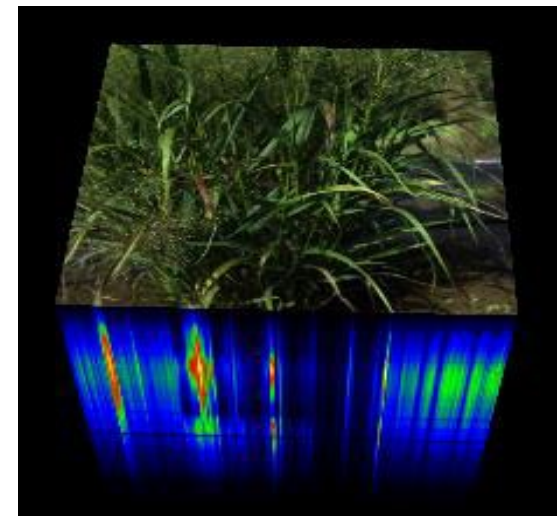
- We utilized the Slate platform provided by OLCF to build a cloud-native data and compute infrastructure for CBI. Slate is an Open-Shift Kubernetes cluster that allows for containerization of applications and data.

Results

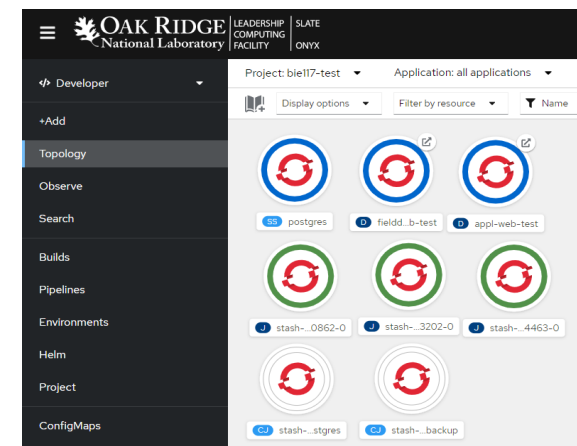
- Approximately 23 terabytes of data were migrated to the OLCF system.
- Two web services, Field Data and APPL-PI, were deployed to manage data access for CBI activities.
 - Field Data is from CBI's extensive GWAS field trials for poplar and switchgrass
 - ORNL's new Advanced Phenotyping Laboratory (APPL) is providing massive amounts of phenotype x genotype data from greenhouse experiments.
 - Data is federated in Kubernetes pods using PostgreSQL, a relational database management system.
- One new analytical pipeline was deployed to analyze hyperspectral data and released on DOE-Code.
 - HyperKube software by Yoon and Martin, 2023. [10.11578/dc.20231101.1](https://doi.org/10.11578/dc.20231101.1)
- A developer's guide was produced to enable computational biologists to generate value-added data products.

Significance

- The switch to cloud-native computing is the next stage of evolution in scientific data management and analysis. This architecture will allow massive parallelization of efforts. It democratizes access to compute resources across CBI. The open-source HyperKube software will increase the speed of analysis for hyperspectral data products coming out of APPL.
- This data system will assist in analyses, data release, and inter-BRC collaborations.



A hyper spectral data cube of Switchgrass from CBI's common garden. Used to develop the HyperKube Kubernetes software



HyperKube deployment on SLATE