

Container Analysis Environments Workshop

Overview

The [Container Analysis Environments Workshop](#) was sponsored by NDS and DataExpLab to bring together a variety of groups leveraging container technology in research computing and data analysis access. Groups represented a wide variety of projects including Blue Waters, LIGO, LSST/DES, CyVerse, TERRA-REF, SciServer, Whole Tale, yt.Hub, NDS Labs, TACC (Agave, BioContainers), SDSC (JupyterHub), XSEDE Gateways, CyberGIS, CRC, and Brown Dog.

Following a set of [presentations](#), the group was asked to [prioritize a list of topics](#) for breakout groups and deep dive discussions. The topics were discussed as follows:

1. Integration with HPC environments
 - a. Singularity/Shifter
 - b. Launching jobs with data from interactive environments
 - c. Agave API
2. Shared storage across containers
 - a. How groups are supporting data access
 - b. Performance/reliability/scalability
 - c. Deep dive on Whole Tale data management architecture
 - d. Security (i.e., HIPAA compliance)
3. Archiving/management/preservation of images
 - a. Centralized registry for Docker and singularity images used in research environments
 - b. Best-practices for image preservation
 - c. Allowing users to dynamically compose images
4. Interactive analysis environments
 - Figuring out "gotchas" -- how are people solving problems. What do they include? Jupyter vs Rstudio vs X environments
 - Profiling load/capping access
5. Opportunities for interoperability/collaboration between systems
6. Container orchestration (scheduling) systems

Several topics could not be discussed in detail due to a lack of expertise including workflow systems and authentication.

Major takeaways from the workshop include:

- Similarity between architectures (SciServer, CyVerse, WT, yt.Hub, Workbench). Some differences, but lots of core similarities. Presents clear opportunity for interoperability.
- Similarities between these services and science gateways, particularly as we look at integration with HPC (but ig differences between Singularity and Docker)
- Jupyter notebooks are a portable research product -- everyone is using them.
 - But RStudio, Shiny, Matlab are still used.
 - Distinction between the developer and the researcher
- Lots of components that we might be able to use at least for design patterns or possibly as implementations
 - Whole Tale data management architecture
 - Agave API

Actions:

- Encapsulating iPython notebooks (make the notebook sharable, not the container)
- Integration prototypes:
 - TERRA-REF -> CyVerse
 - SciServer -> yt.hub
 - Whole Tale -> CyVerse
- Workbench to HPC via Agave
- Follow-up about authentication
- Report at NDS8 (Craig)