Discovery Environment

Extensible Data Science workbench and data-centric collaboration platform powered by iRODS

CyVerse Discovery Environment Development Team
University of Arizona
Agenda

• Discovery Environment (DE) Overview
  • Overview
  • Features
  • Technology Choices
• Terrain API
  • Overview
  • Available Documentation
  • Brief Demonstration
• Visual Interactive Computing Environment (VICE)
  • Overview
  • Architecture
  • Demonstration
Motivation

- Two types of tasks.
- Make the mundane stuff easier.
  - Data management
  - Software management
- Make the interesting stuff easier if possible.
- Make it reproducible
Requirements

- Scalable
- Extensible
- Low barrier of entry
- High productivity
Overview - Usage Statistics

Jobs, Users & Sessions

- Logins: 426802
- Jobs: 362610
- Users: 21490
Data Management

- CyVerse Data Store
- Share data sets
- Search all data that is accessible
- Automatically detect format and type of data in files
- Third party and built-in data visualization tools
  - Genome browsers (Ensembl, UCSC, JBrowse, etc) with byte-range service
  - Tabular data view
- Metadata management, tags and comments
Tools and Apps

- Graphical user interface to apps
- Apps can target several different platforms
- Add your own tools and apps
- Apps can be chained together in a pipeline
- GPU support for VICE
- More than 500 apps with documentation and example data sets
- Almost 300 distinct Docker images
Analyses

• What’s an analysis?
• Control resource allocation (CPU, RAM, Disk)
• Output files are uploaded to the CyVerse Data Store
• Parameters are recorded
• You’ll be notified when an analysis completes
• Batch processing
Terrain - Goals

- Avoid the monolith trap
- Scalability
- Extensibility
- Customization
Terrain - Documentation

- Avoid the monolith trap
- [https://de.cyverse.org/terrain/docs](https://de.cyverse.org/terrain/docs)
  - Interactive console
  - Dynamic documentation
  - Work in progress
- [https://cyverse-de.github.io/api](https://cyverse-de.github.io/api)
  - Ignore the authentication instructions
  - Mostly complete
  - Documentation for some endpoints is outdated
Terrain Demonstration

Documentation, Console, Command Line
VICE - Overview

- Work with software and data interactively
- Visualize
- Experiment
- Discover
Demonstration

Discovery Environment Inception
Future

• Integrate 3rd Party storage providers
• BYOC
• Improved UX
• Singularity support
• 3rd party install