Community Collaborations:

DE, Syndicate

Nirav Merchant
The University of Arizona
nirav@email.arizona.edu
http://www.cyverse.org Twitter: @CyVerseOrg
DE: Community Edition, Containers ...

- CyVerse Discovery Environment (DE) is available for deployment as collaboration platform for institutions
- Supports data lifecycle management (iRODS) and container lifecycle management (Docker, Singularity)
- Users can select container from any URL (docker hub, quay.io etc.) get web UI for it and connect with iRODS data
- Code is at [https://github.com/cyverse-de/](https://github.com/cyverse-de/) and developer documentation is [https://github.com/cyverse-de/](https://github.com/cyverse-de/) paper [https://f1000research.com/articles/5-1442/v3](https://f1000research.com/articles/5-1442/v3)
- New: Secure Interactive jobs (Jupyter notebooks, Rshiny, Kibana dashboards) via web proxy to running tasks
Syndicate: Edge computing with iRODS

• Significant amount of data in our Data Store from large projects (Astronomy, Climate models, Genomics, Images)
• Users wanting to work with these data sets (many files and large files), but only needing few files from the collection
• Computational resources utilized are highly distributed (laptop to cloud and HPC centers)
• Some projects have data in Institutional repositories, cloud resources not allowing easy access (scale) for analysis
• Users cannot readily modify paths to file/directories in their analysis workflow
The real DM challenge

Does this look like a Data Management Experts

Distributed Set of Collaborators

Commodity Cloud Storage

S3

DropBox/Box

Pre-Stage

Write-Back

Institutional Resources

XSEDE

CyVerse
Syndicate Solution

Manages data consistency and key distribution

Treats cloud storage as a block device

Bridges application workflow and HTTP transport; e.g.,
  – Jupyter
  – Hadoop

Acquires data from existing data stores; e.g.,
  – CyVerse
  – XSEDE
Syndicate: Edge computing with iRODS

• What have we built so far: Consortium of Universities with CDN locally, Docker containers with popular datasets (mainly iRODS from CyVerse), Hadoop integration

• We are in early stage (beta) and are focusing on performance and scalability

• If you would like to participate visit website or email nirav@email.arizona.edu

• Details at: http://www.syndicate-storage.org/

• For more information about taking advantage of Syndicate's capabilities, see the User Guide and watch the tutorial videos and demos