

Academic Workflow for Research Repositories Using iRODS and Object Storage

2016 iRODS User's Group Meeting

9 June 2016 Randall Splinter, Ph.D.

HPC Research Computing Solutions Architect RSplinter@ddn.com 770.633.2994

Agenda

Introduction to the Problem

- HPC Workflows
- The Problem of Long Term Archiving
- Object Store to the Rescue
- How iRODS Enables Object Storage
- Why iRODS with DDN WOS is a Superior Solution for Research Repositories
- A Case Study



Introduction to the Problem

The Problem of Collaboration

- NAS technologies (NFS, CIFS) are local
 - They don't tend to scale well over WAN distances
 - -But collaborators are frequently widely separated in distance
 - How to enable researchers to share data without administrative overhead – securely
 - -Typically, only system administrators can control the ACLs on NFS/CIFS mounts
- FTP
 - Security
- Tape
 - Yech





HPC Workflows

Ingest data from a source (Analysis of data)

- Pre-analysis on low-end storage
- Move cleaned up data to a PFS and compute
- After full analysis data is moved to long term storage
 - Can this be automated? Yes, with iRODS.

No ingest (Pure simulation)

- Compute models are run on a compute cluster with PFS
- After full analysis data is moved to long term storage

Again automation is key.

Data sets are exploding!





The Problem of Long Term Archiving

Data must be secure

- From deletion (accident or deliberate)
- Loss from theft
 - Security hacks
 - Faculty or students leave and take IP with them

Must satisfy regulatory restrictions

- HIPAA, for instance
- Changing hardware standards
 - In particular tape standards
- Hardware availability
 - Spinning media is mechanical and will not last forever





Object Store to the Rescue



- All Object stores provide a way to replicate data over large distances
 - Some more effectively than others
 - Provides a way to effectively share data over WAN scales
 - Most object stores were designed for cloud storage
 - Security has always been important
 - Ease of data sharing has been important
- This now enables more effective data sharing and data security than with traditional storage solutions at price points that traditional NAS systems cannot approach.



6

How iRODS Enables Object Storage

- ▶ iRODS is a very effective middleware layer for accessing multiple storage resources
- IRODS handles the security and database management of the ingested data
- Provides a powerful metadata search capability for ingested data
- Provides a rules engine for the processing of incoming data and the manipulation of data on the back-end
 - For instance,
 - Data can be moved to slower storage resources as they age or another criteria is met (Essentially HSM)
 - Data can be secured from removal, editing or modification based upon criteria using the "null" chmod – Retention policies!
 - Anything else?





DDN WOS: Key Features

Exabyte Scalability

Create virtually limitless data repositories, non-disruptively seamlessly scale to over an exabyte of capacity



Fully-Integrated Object Storage Appliance

WOS7000, 60 Drives in 4U, with 1 or 2 object storage servers per appliance

Pure Object Storage

Formats drives with custom WOS disk file system, no Linux file I/Os, no fragmentation, fully contiguous object read and write operations for maximum disk efficiency

User Defined Metadata and Metadata Search

Applications can assign their own metadata via object storage API, WOS now also supports parallel search of WOS user metadata





- Ease of Scalability with WOS
- Ease of administration Once rules are tested and in place the system can be managed with a minimum of administrative overhead
- Automating workflows to guarantee consistency and reproducibility in the science that is produced





9

© 2016 DataDirect Networks, Inc. * Other names and brands may be claimed as the property of others. Any statements or representations around future events are subject to change.

Why iRODS with DDN WOS is a Superior Solution for Repositories

- Ease of auditing for both usage and back charging and for maintaining adequate data security compliance
- DDN WOS makes remote replication simple and provides a straightforward way to manage DR systems





© 2016 DataDirect Networks, Inc. * Other names and brands may be claimed as the property of others. Any statements or representations around future events are subject to change.



Why iRODS with WOS is a Superior Solution for Repositories

- Central to any repository is the ability to add metadata tags and search metadata.
 - iRODS has extensive abilities to do that Significantly better than any competing options





© 2016 DataDirect Networks, Inc. * Other names and brands may be claimed as the property of others. Any statements or representations around future events are subject to change.







© 2016 DataDirect Networks, Inc. * Other names and brands may be claimed as the property of others. Any statements or representations around future events are subject to change.

ddn.com

Thank You!

Keep in touch with us

Questions?



sales@ddn.com

@ddn_limitless



9351 Deering Avenue Chatsworth, CA 91311

1.800.837.2298 1.818.700.4000



company/datadirect-networks





