EMC MetaLnx - An iRODS Admin & Metadata UI

Steve Worth

Director, Global New Business Development Operations

6/10/2015
iRODS User Group Meeting
Motivation

• Observations
  – A data deluge is upon us – will surpass 44 Zetabytes by 2020
    • Big data, IOT, and mobile computing are changing how we interact with data
  – Cost dynamics are changing
    • Cloud computing / open source feed the “race to zero” in computing
    • Opex costs begin to dominate over Capex costs

• Implication
  – “Big Data” is now IP
  – Advanced metadata facilitates IT data control

• Investigation
  – iRODS enables next generation metadata
  – How will advanced metadata facilitate IT operations?
Advanced Metadata is Key

- Maximize storage assets - find what’s valuable, no matter where it’s located
  - Eliminate data that is not valuable (so called “dark data”)
- Metadata supports indexing / fast search methods
  - Ideal for non-SQL techniques
- Automate movement and processing of data
- Metadata can be stored/transported more cheaply
  - An object with 1000 tags, 4096 character limit, 3 tag AVU model requires < 12MB for the tag set.
- Securely share data with collaborators
  - Easier to separate PHI from non-PHI data
  - Can be used to ensure data is authentic and unaltered
The Next Important IT Skill...

- Advanced metadata will be a core information tool
- With data growing at over 2 Exabytes / year the need to catalogue, structure, and organize data will increase
- The Internet of things will make long term curation of data more critical (100+ years)

Information Science

(In past times this was called Library Science)
Want Proof?

This is the modern day collegiate library....

See any books?

James B. Hunt Library – North Carolina State Univ.
IT World According to Genomes

Analysis:
- Sequence generation
- Variation processing
- Protein analysis
- Palindrome analysis

Workflow:
- Sample sequencing
- Data workflow
- Analysis workflow
- Variant workflow
- Reports

Analytics:
- Population trending
- Reference genome
- Pattern matching

Data Management:
- Data collection
- Archiving
- Backup / protection
- Access / Rights mgmt.
- Auditing
- Providence
- Federation
- Data workflow
- Perpetual curation

Metadata Management:
- Research Annotations
- Phenotype Data Tagging
- Subject Data Tagging
- Structural Tagging

Genomics leverages all aspects of computing, with big data sets, and interesting curation needs
Genome Management Design Approach

- **Administration**
  - System / account management
  - User access and security
  - Sharing/federation
  - Data organization
  - Data archiving

- **Users (Curation/Research)**
  - Data access/organization
  - Metadata identification
  - Template utilities
  - Searching
  - Sharing (Internal/External)
  - IT is a tool only
Working With iRODS…

- **User commands** (subset)
  - iinit
  - iput
  - iget
  - imkdir
  - ichmod
  - icp
  - irm
  - ils
  - ipwd
  - icd
  - irepl
  - iexit
  - ipasswd
  - ichksum
  - imv
  - iphymv
  - ireg
  - irmtrash
  - irsync

- **Iadmin commands** (subset)
  - lt
  - lr
  - ls
  - lz
  - lg
  - mkuser
  - moduser
  - aua
  - rua
  - rpp
  - rmuser
  - mkdir
  - rmdir
  - mkresc
  - modresc
  - modrescdatapaths
  - rmresc
  - mkzone
  - modzone
EMC MetaLnx Demo Setup

Physical System:
Intel i7, 4 cores,
12 GB memory

Virtual Machine Layout:
1 core, 1 GB memory, 10 GB disk / each

iRODS grid

swcentos6-1
iCAT server, local disk

swdebian64-1
iRODS resource server, 2 local storage resources

swcentos6-2
MetaLnx server & client

iRODS storage resources

Network
EMC MetaLnx Demo Setup

Physical System:
Intel i7, 4 cores, 12 GB memory

Virtual Machine Layout:
1 core, 1 GB memory, 10 GB disk / each

Network

iRODS grid

swcentos6-1
iCAT server, local disk

swdebian64-1
iRODS resource server, 2 local storage resources

swcentos6-2
MetaLnx server & client

iRODS storage resources

© Copyright 2014 EMC Corporation. All rights reserved.
EMC MetaLnx Demo Setup

Physical System:
Intel i7, 4 cores,
12 GB memory

Virtual Machine Layout:
1 core, 1 GB memory, 10
GB disk / each

© Copyright 2014 EMC Corporation. All rights reserved.
EMC MetaLnx Demo Setup

Physical System:
Intel i7, 4 cores, 12 GB memory

Virtual Machine Layout:
1 core, 1 GB memory, 10 GB disk / each
MetaLnx Live Demo

Dashboard

System Health

- Overall Status: normal

iRODS Servers

- irods06-1: 192.168.1.5
- irods06-1: 192.168.1.6

Total Storage

- 33 GB free of 43 GB
- 10 Files

Users

- Users (6)
- Groups (5)

Quick Access

Add User
Groups
Collections
Resources

Selected images added to show functionality
Dashboard - iRODS Server Down
Server Details Page

Details page

Network
- **IP**: 192.168.52.129
- **Hostname**: swdebian64-1
- **MetaLnx Remote Monitor**: 0.1-31

Total Storage
- 12 G free of 17 G

Hardware
- **Intel(R) Core(TM) i7-4600U CPU @ 2.10GHz**
  - **Disk**
    - User Mode: 2.2
    - System Model: 2.8
    - Idle Task: 91.8
    - I/O Waiting: 3.0
  - **Memory**
    - Total: 496.0 MB
    - Used: 477.0 MB (96.2%)
    - Free: 19.0 MB (3.8%)
    - Shared: 0 B
## User Management

### User Management

**Search**

11 user(s) found

<table>
<thead>
<tr>
<th>First Name</th>
<th>Last Name</th>
<th>Username</th>
<th>Zone</th>
<th>Email</th>
<th>Action</th>
</tr>
</thead>
<tbody>
<tr>
<td>bjenkins</td>
<td></td>
<td></td>
<td>tempZone</td>
<td></td>
<td></td>
</tr>
<tr>
<td>bjones</td>
<td></td>
<td></td>
<td>tempZone</td>
<td></td>
<td></td>
</tr>
<tr>
<td>cstacy</td>
<td></td>
<td></td>
<td>tempZone</td>
<td></td>
<td></td>
</tr>
<tr>
<td>jones</td>
<td></td>
<td></td>
<td>tempZone</td>
<td></td>
<td></td>
</tr>
<tr>
<td>James</td>
<td>Kirk</td>
<td>jkirk</td>
<td>tempZone</td>
<td><a href="mailto:jkirk@gmail.com">jkirk@gmail.com</a></td>
<td></td>
</tr>
<tr>
<td>Joel</td>
<td>Schwartz</td>
<td>jschwartz</td>
<td>tempZone</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rods</td>
<td></td>
<td></td>
<td>tempZone</td>
<td></td>
<td></td>
</tr>
<tr>
<td>sjones</td>
<td></td>
<td></td>
<td>tempZone</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sue</td>
<td>Smith</td>
<td>ssmith</td>
<td>tempZone</td>
<td><a href="mailto:ssmith@aol.com">ssmith@aol.com</a></td>
<td></td>
</tr>
</tbody>
</table>

© Copyright 2015 EMC Corporation. All rights reserved.
Collections - Pagination

Collections

View: Info, Metadata, Permissions

25 items per page

25 result(s) found

<table>
<thead>
<tr>
<th>Name</th>
<th>Owner</th>
<th>Kind</th>
<th>Modified</th>
</tr>
</thead>
<tbody>
<tr>
<td>2002_cruise0490(3).jpg</td>
<td>sworth</td>
<td>file</td>
<td>Apr 08, 2015 07:04</td>
</tr>
<tr>
<td>2002_cruise0490(4).jpg</td>
<td>sworth</td>
<td>file</td>
<td>Apr 08, 2015 07:04</td>
</tr>
<tr>
<td>2002_cruise0490(5).jpg</td>
<td>sworth</td>
<td>file</td>
<td>Apr 08, 2015 07:04</td>
</tr>
<tr>
<td>2002_cruise0490(6).jpg</td>
<td>sworth</td>
<td>file</td>
<td>Apr 08, 2015 07:04</td>
</tr>
<tr>
<td>2002_cruise0490(7).jpg</td>
<td>sworth</td>
<td>file</td>
<td>Apr 08, 2015 07:04</td>
</tr>
<tr>
<td>2002_cruise0490(8).jpg</td>
<td>sworth</td>
<td>file</td>
<td>Apr 08, 2015 07:04</td>
</tr>
</tbody>
</table>
Collections - upload
Collections - Extracted Metadata
Metadata Search

Search for files and collections that match one or more of the conditions below:

- Orientation: Is (Equals) Landscape
- Make: Is (Equals) Olympus
- Orientation: Is Not (Not Equal) foo

Reset Conditions  Search
# Metadata Search Results

### Search Criteria

- **Make**: S OPTICAL CO., LTD
- **Orientation**: Landscape
- **Orientation**: Not Equal to 0

### Search Results

<table>
<thead>
<tr>
<th>Name</th>
<th>Path</th>
<th>Owner</th>
<th>Kind</th>
<th>Modified</th>
<th>Size</th>
<th>Matches</th>
</tr>
</thead>
<tbody>
<tr>
<td>2000_Spring002...</td>
<td>/tempZone/home/s...</td>
<td>swort</td>
<td>file</td>
<td>May 27 2015, 12:08</td>
<td>381.6 KB</td>
<td>❌</td>
</tr>
<tr>
<td>2000_Spring001...</td>
<td>/tempZone/home/s...</td>
<td>swort</td>
<td>file</td>
<td>May 27 2015, 12:11</td>
<td>277.7 KB</td>
<td>❌</td>
</tr>
<tr>
<td>2000_Spring001...</td>
<td>/tempZone/home/s...</td>
<td>swort</td>
<td>file</td>
<td>May 27 2015, 12:11</td>
<td>277.7 KB</td>
<td>❌</td>
</tr>
<tr>
<td>2000_Spring001...</td>
<td>/tempZone/home/s...</td>
<td>swort</td>
<td>file</td>
<td>May 27 2015, 12:11</td>
<td>277.7 KB</td>
<td>❌</td>
</tr>
<tr>
<td>2000_Spring001...</td>
<td>/tempZone/home/s...</td>
<td>swort</td>
<td>file</td>
<td>May 27 2015, 12:11</td>
<td>277.7 KB</td>
<td>❌</td>
</tr>
</tbody>
</table>
MetaLnx Value Proposition: Simplified Metadata Management

• Simple, easy iRODS grid administration
• Simple, intuitive tool for non-IT researchers:
  – Collection management
  – Automated embedded metadata extraction
  – Metadata searching / editing
  – Metadata templates – consistent tagging of data
• Scale out design supports PCs and tablets
• Extensible to any data vertical
EMC iRODS Efforts

• MetaLnx
  – An iRODS Administration and Metadata Interface tool

• Storage Resource Drivers for iRODS
  – Isilon – HDFS interface
  – ECS – Atmos object interface

Both are in active development and beta testing
Questions?

Contacts:

Sasha Paegle  
Senior Business Development Manager  
Sasha.Paegle@isilon.com

Stephen Worth  
Director, Global New Business Development Operations  
Stephen.Worth@emc.com