EMC MetaLnx - An iRODS Admin & Metadata UI

Steve Worth

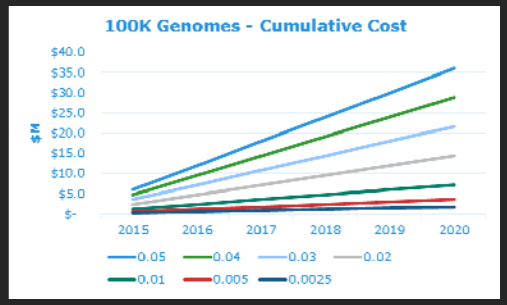
Director, Global New Business Development Operations

6/10/2015
iRODS User Group Meeting

EMC²

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?



© Copyright 2015 EMC Corporation. All rights reserved.

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

© Copyright 2015 EMC Corporation. All rights reserved.

The Next Important IT Skill…

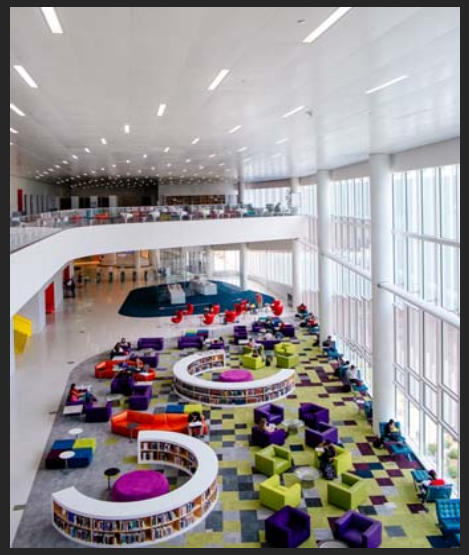
- Advanced metadata will be a core information tool
- With data growing at over 2 Exabytes / year the need to catalogue, structure, and organize data will increase
- The Internet of things will make long term curation of data more critical (100+ years)

(In past times this was called Library Science)

© Copyright 2015 EMC Corporation. All rights reserved.

Want Proof?

This is the modern day collegiate library….



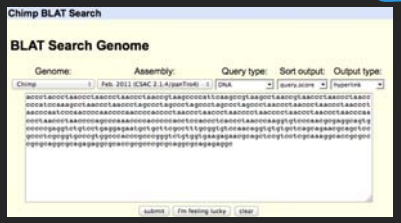
See any books?

James B. Hunt Library – North Carolina State Univ.

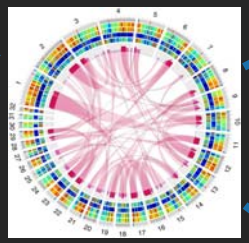
© Copyright 2015 EMC Corporation. All rights reserved.

IT World According to Genomes

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



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



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

Analytics:
- Population trending
- Reference genome
- Pattern matching

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

© Copyright 2015 EMC Corporation. All rights reserved.

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

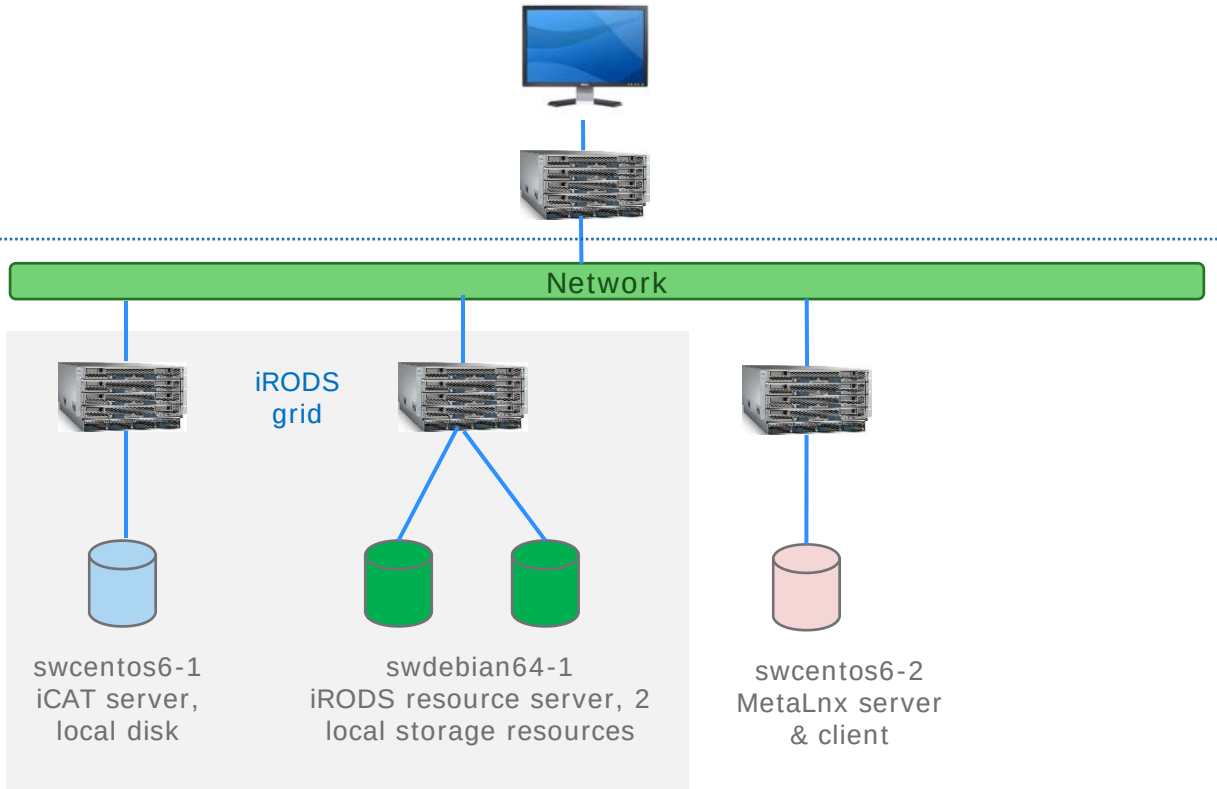
© Copyright 2015 EMC Corporation. All rights reserved.

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

© Copyright 2015 EMC Corporation. All rights reserved.

EMC MetaLnx Demo Setup

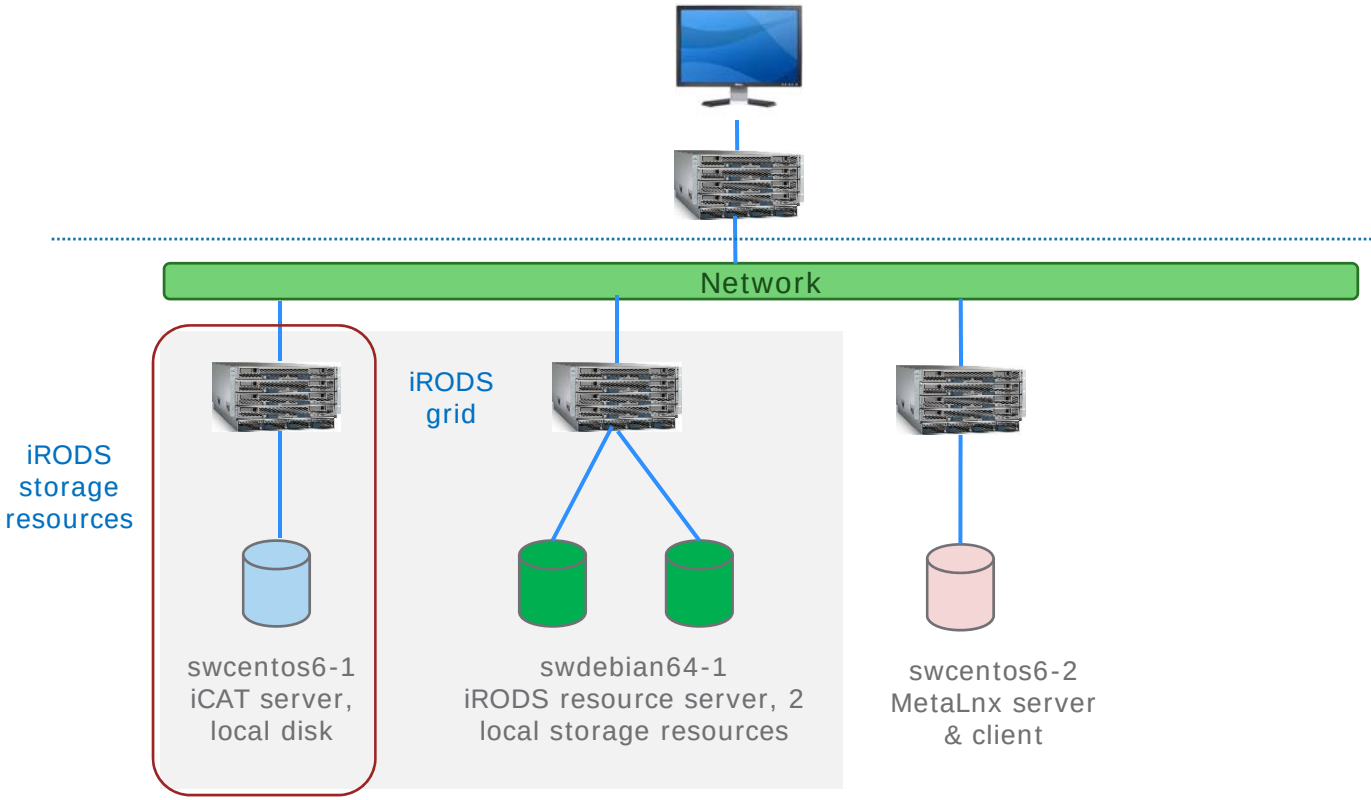


Physical System:
Intel i7, 4 cores,
12 GB memory

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

© Copyright 2014 EMC Corporation. All rights reserved.

EMC MetaLnx Demo Setup

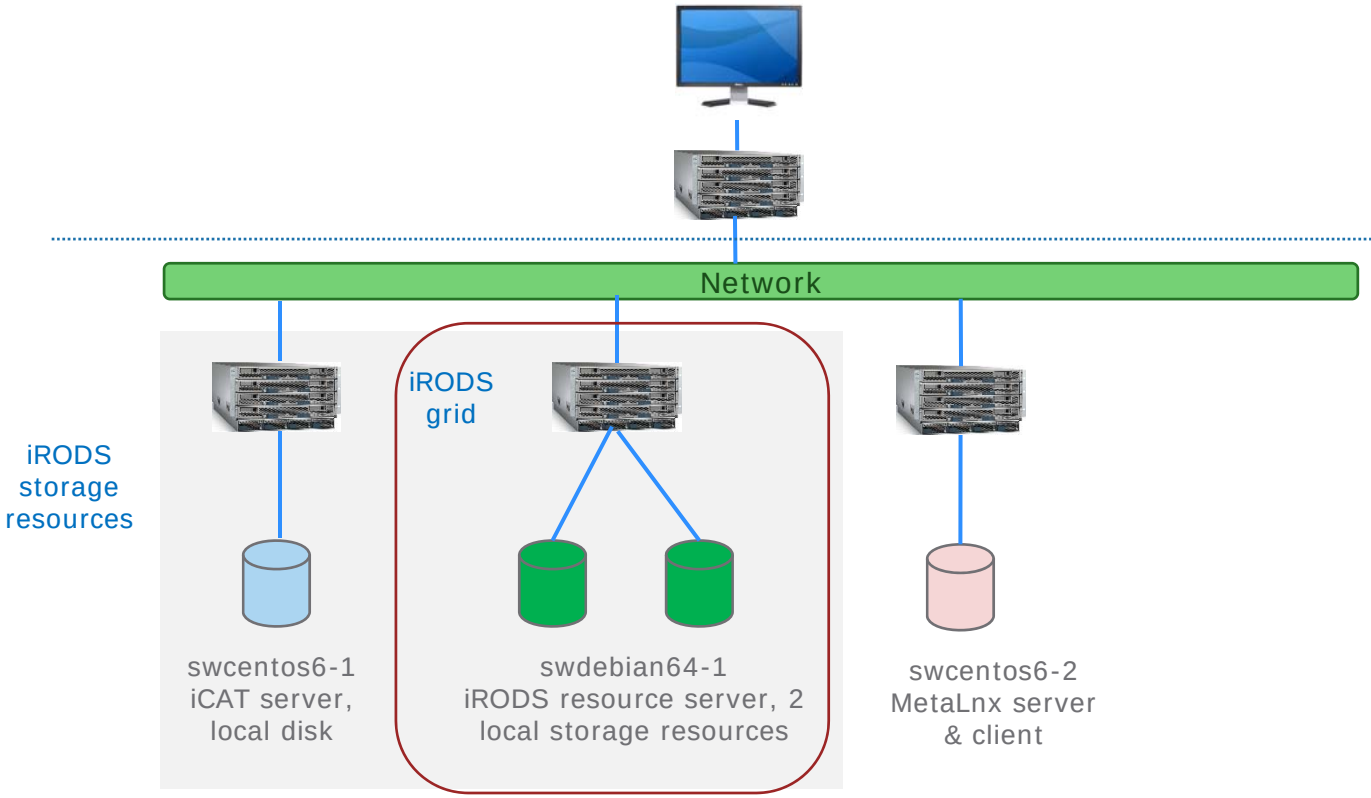


Physical System:
Intel i7, 4 cores,
12 GB memory

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

© Copyright 2014 EMC Corporation. All rights reserved.

EMC MetaLnx Demo Setup

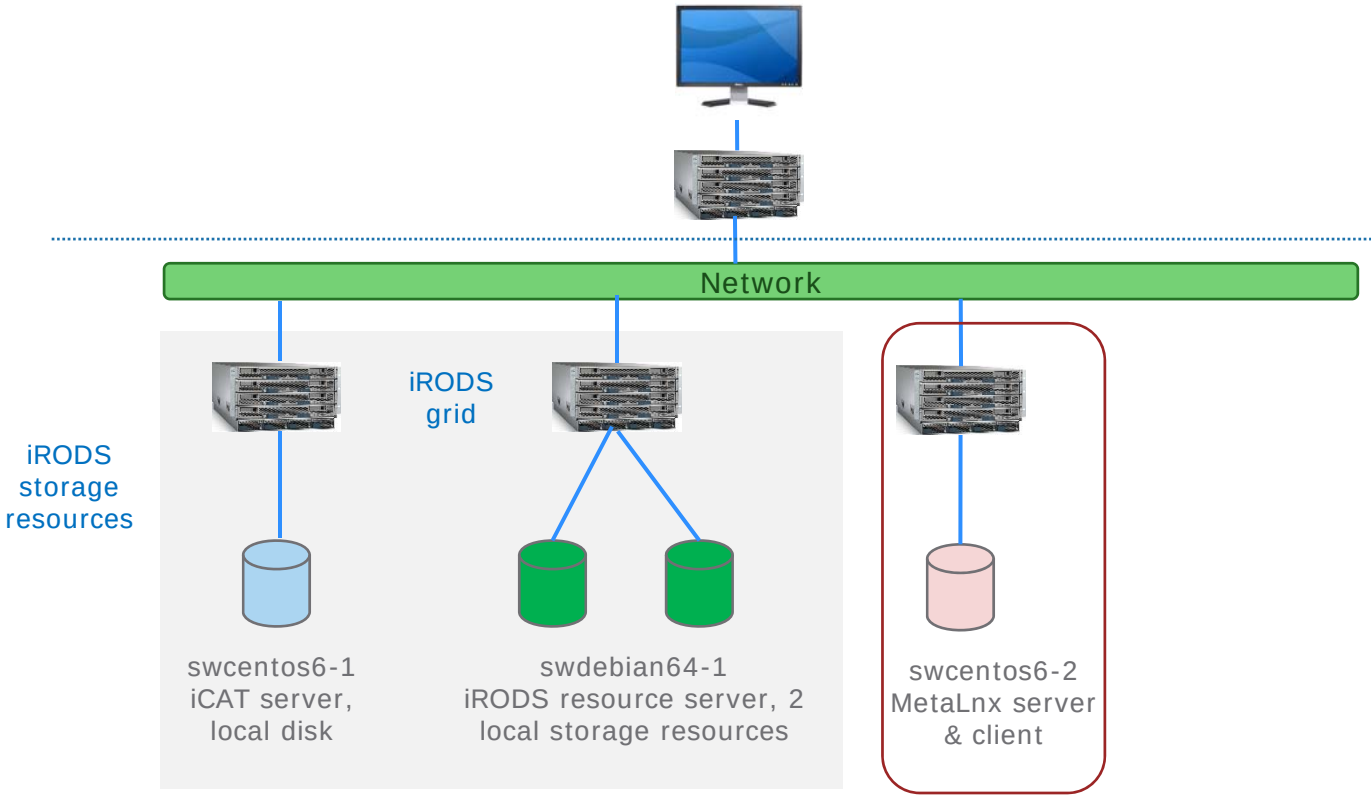


Physical System:
Intel i7, 4 cores,
12 GB memory

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

© Copyright 2014 EMC Corporation. All rights reserved.

EMC MetaLnx Demo Setup

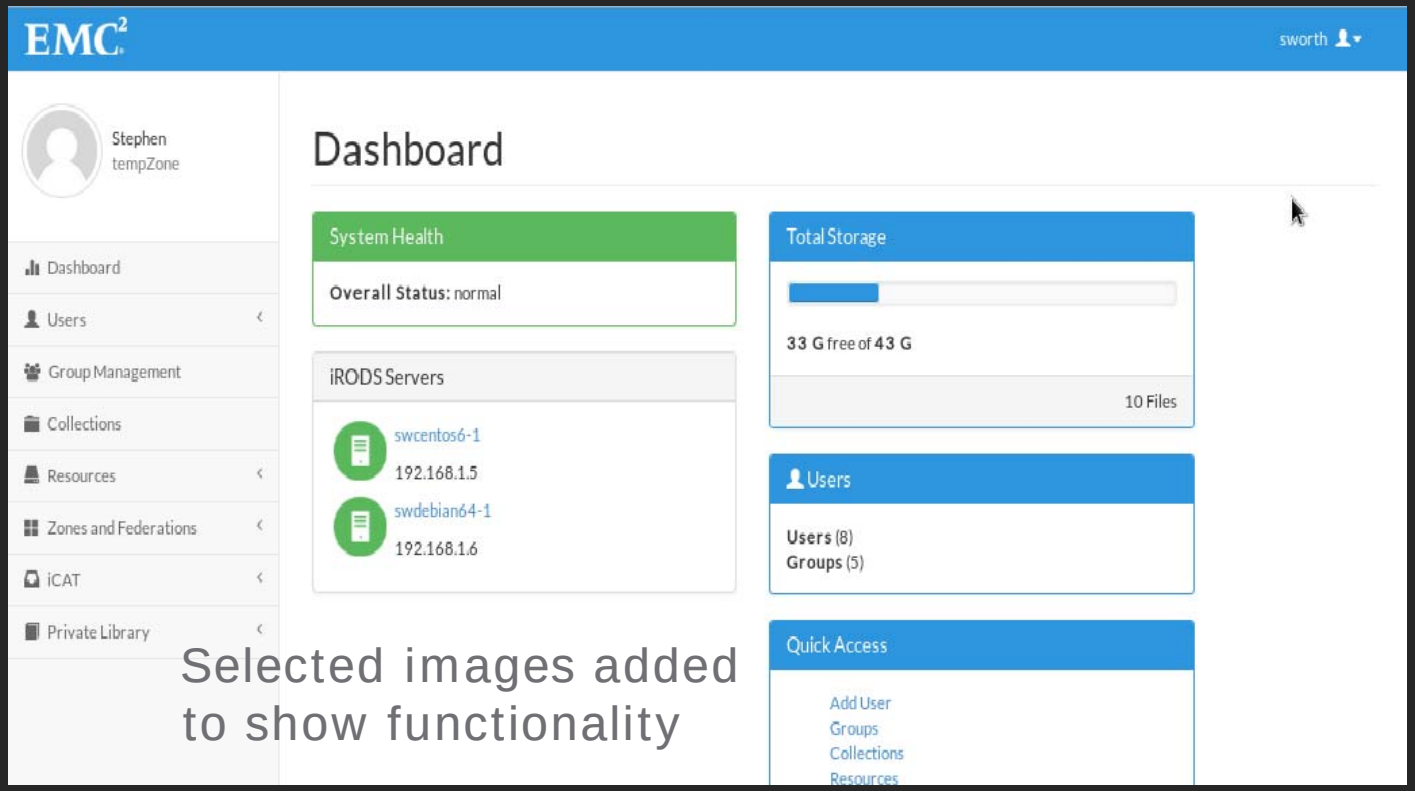


Physical System:
Intel i7, 4 cores,
12 GB memory

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

© Copyright 2014 EMC Corporation. All rights reserved.

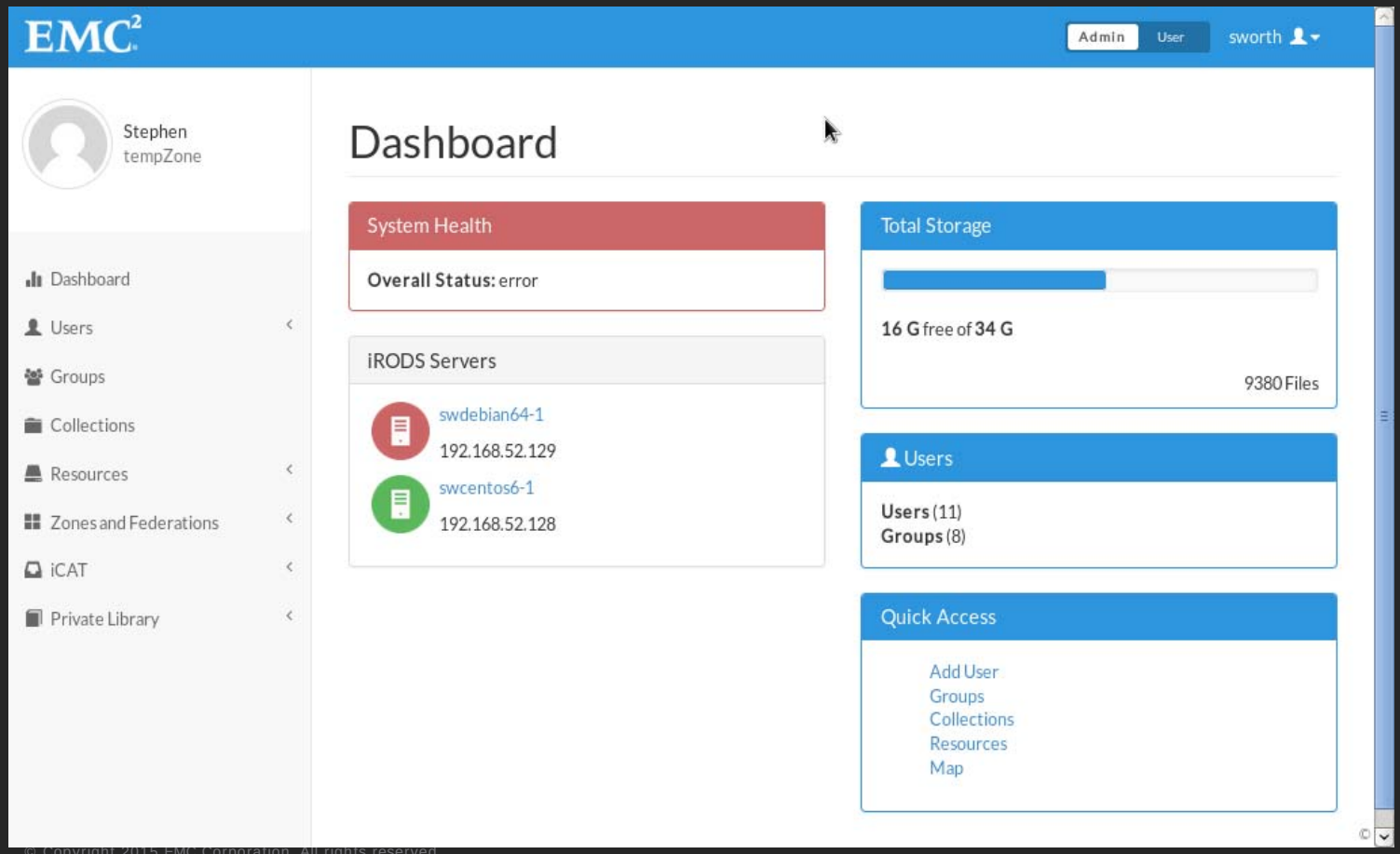
MetaLnx Live Demo



Selected images added to show functionality

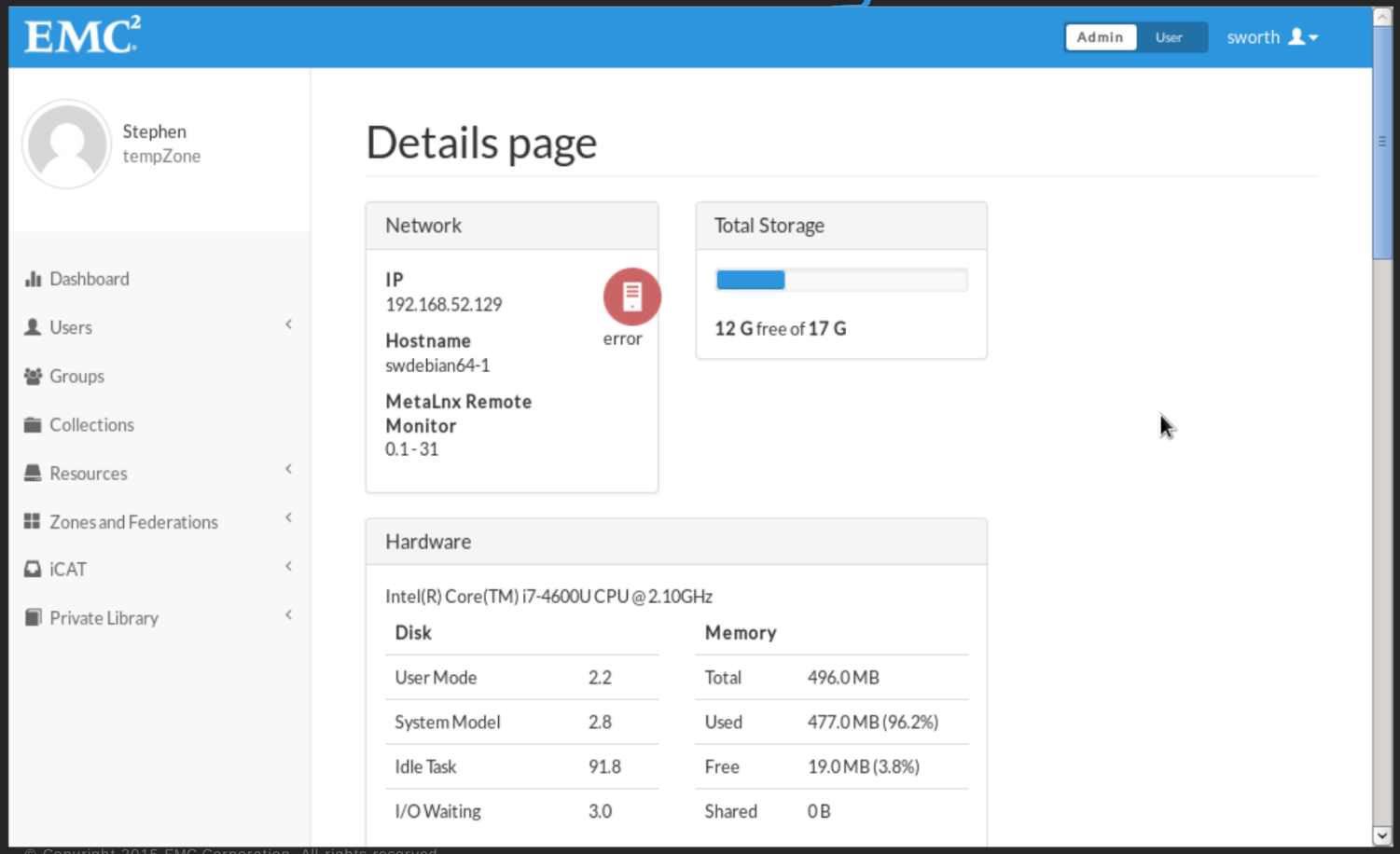
© Copyright 2015 EMC Corporation. All rights reserved.

Dashboard - iRODS Server Down



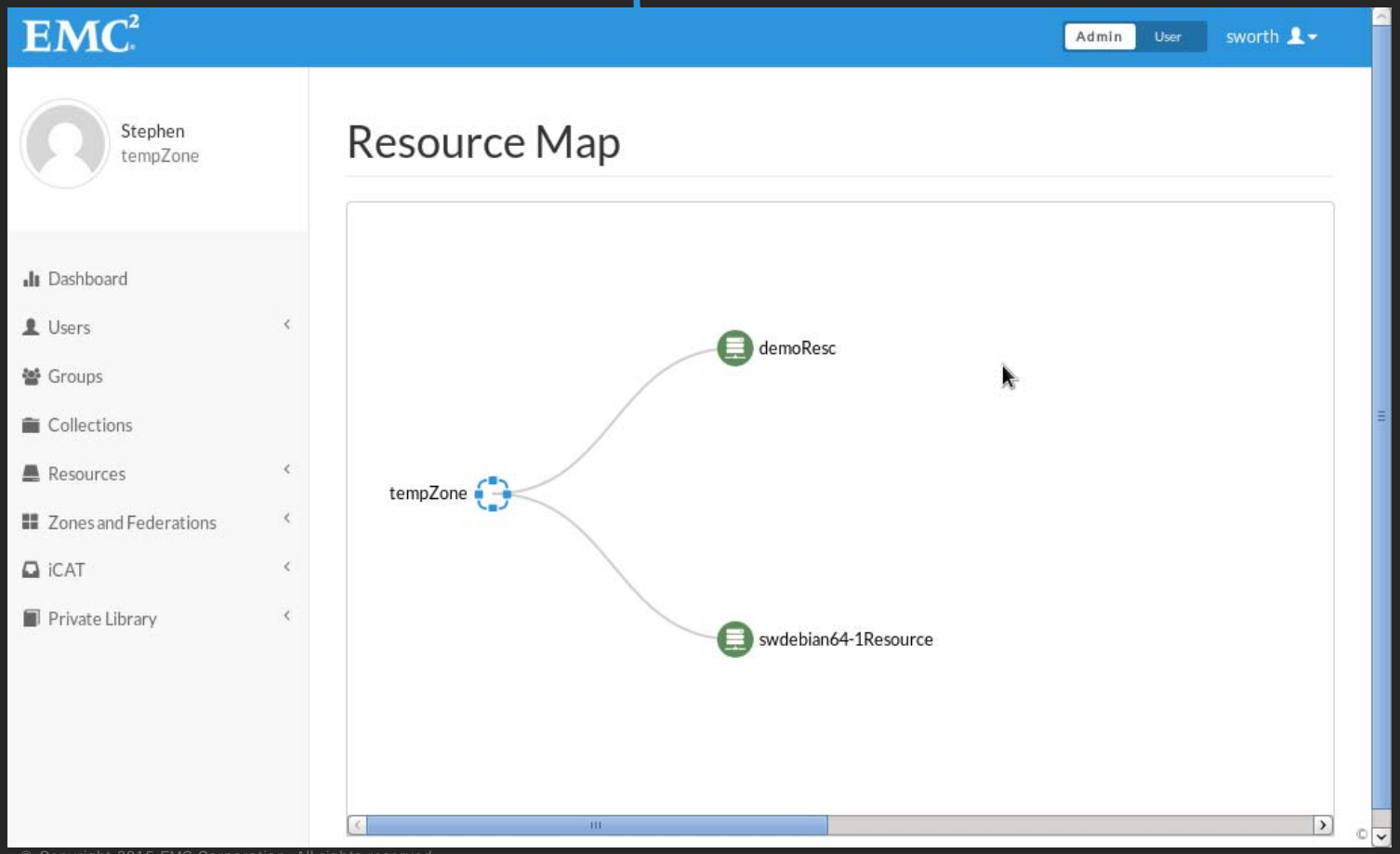
© Copyright 2015 EMC Corporation. All rights reserved.

Server Details Page



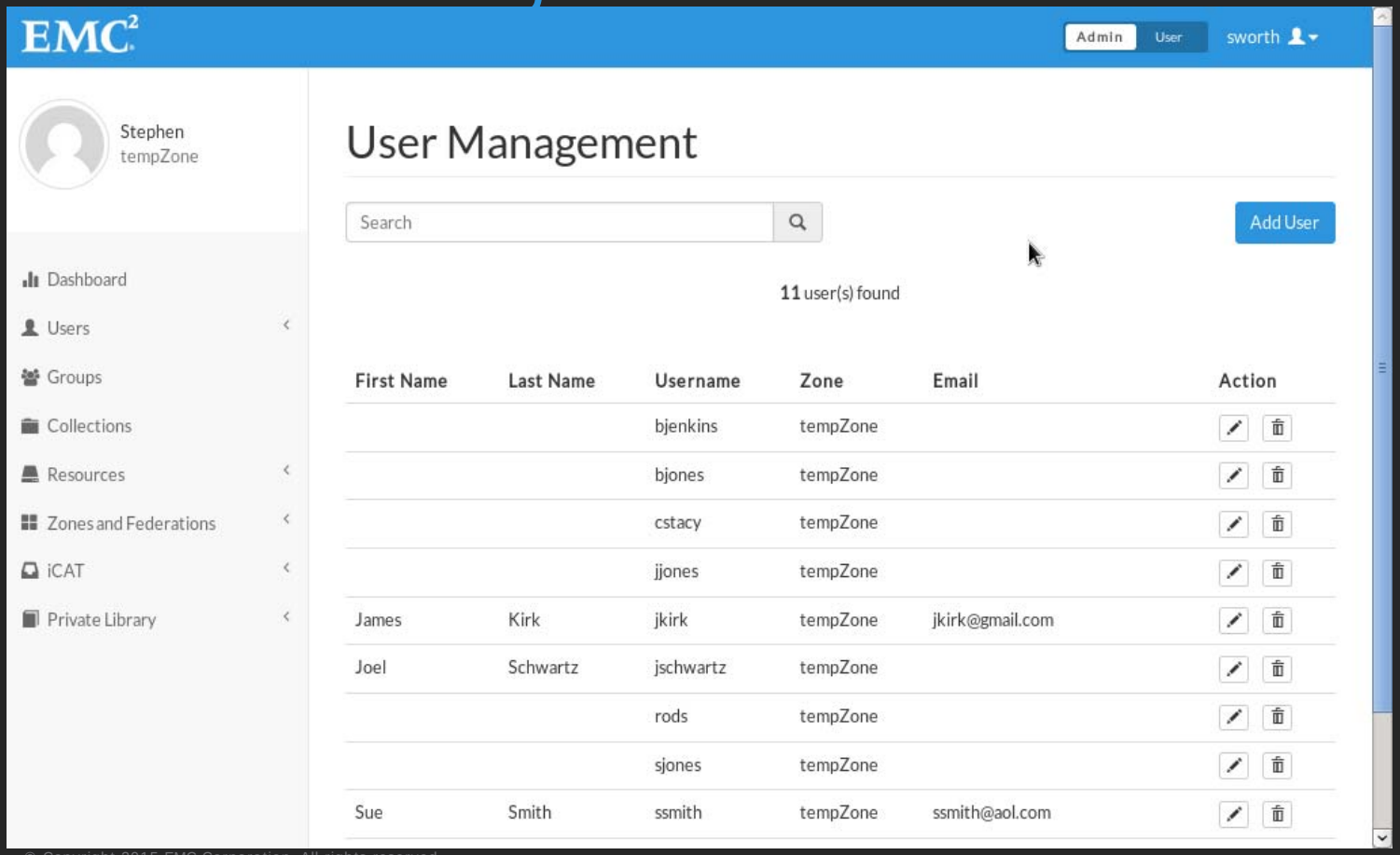
© Copyright 2015 EMC Corporation. All rights reserved.

Resource Map



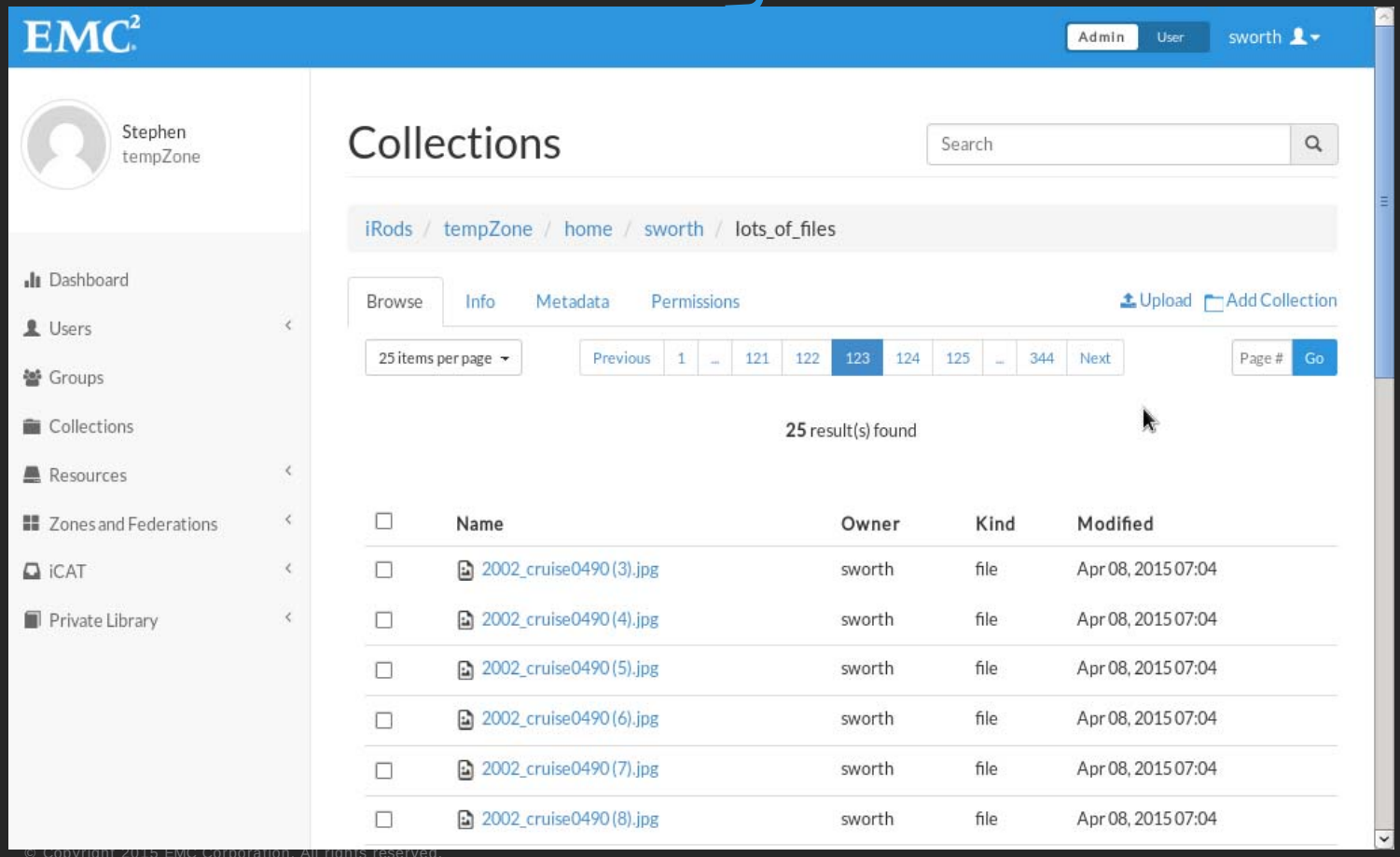
© Copyright 2015 EMC Corporation. All rights reserved.

User Management



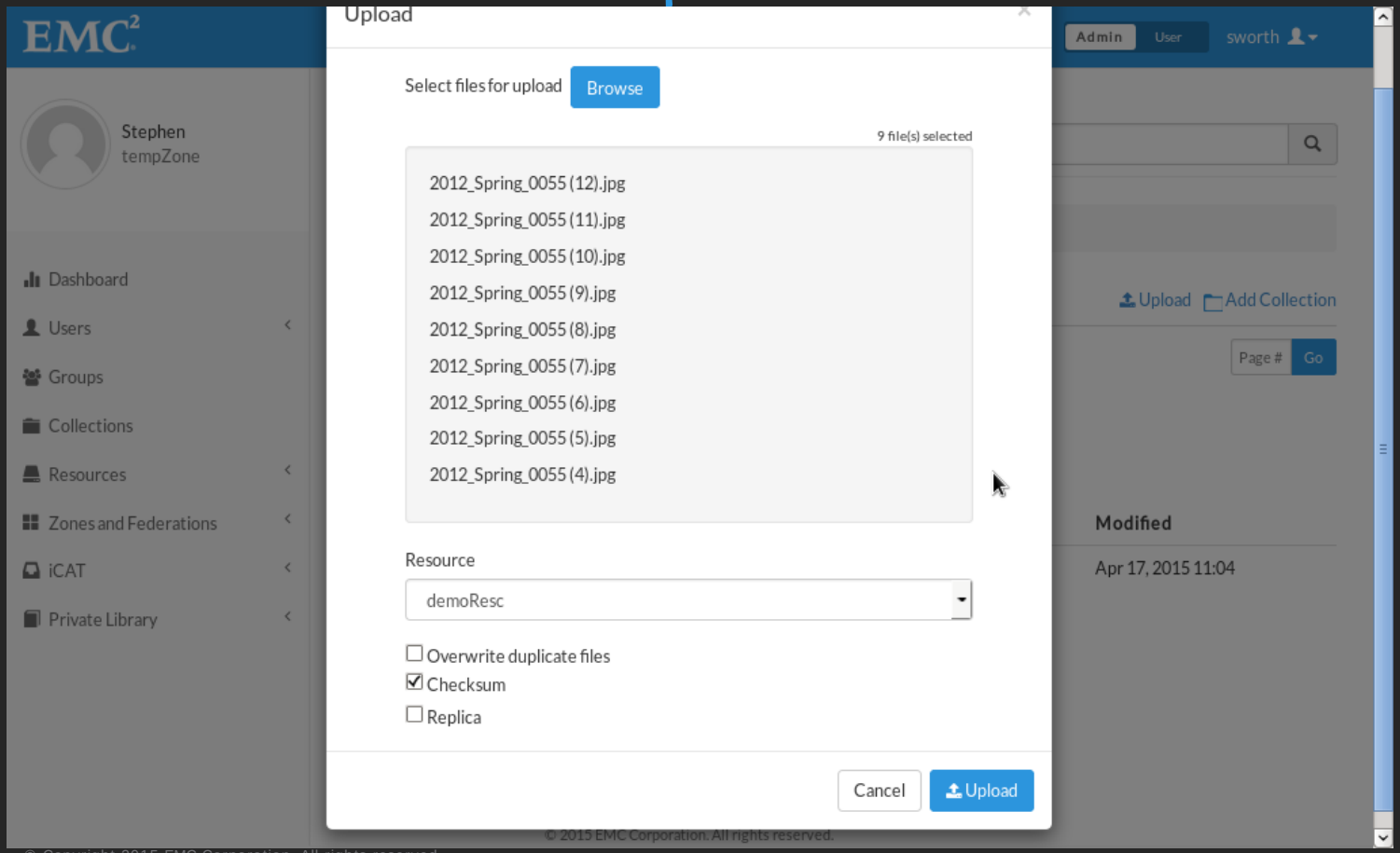
© Copyright 2015 EMC Corporation. All rights reserved.

Collections - Pagination



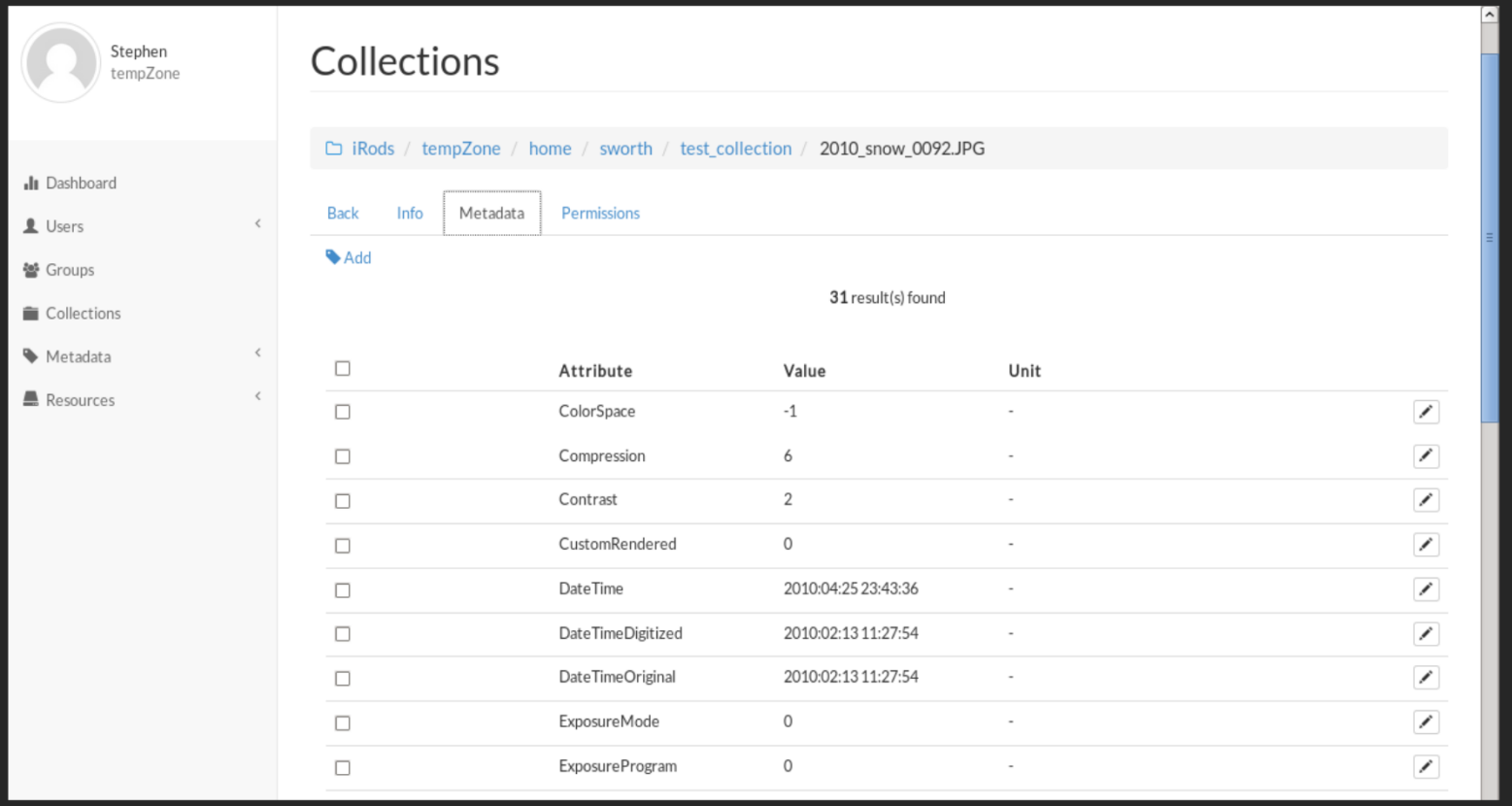
© Copyright 2015 EMC Corporation. All rights reserved.

Collections - Upload



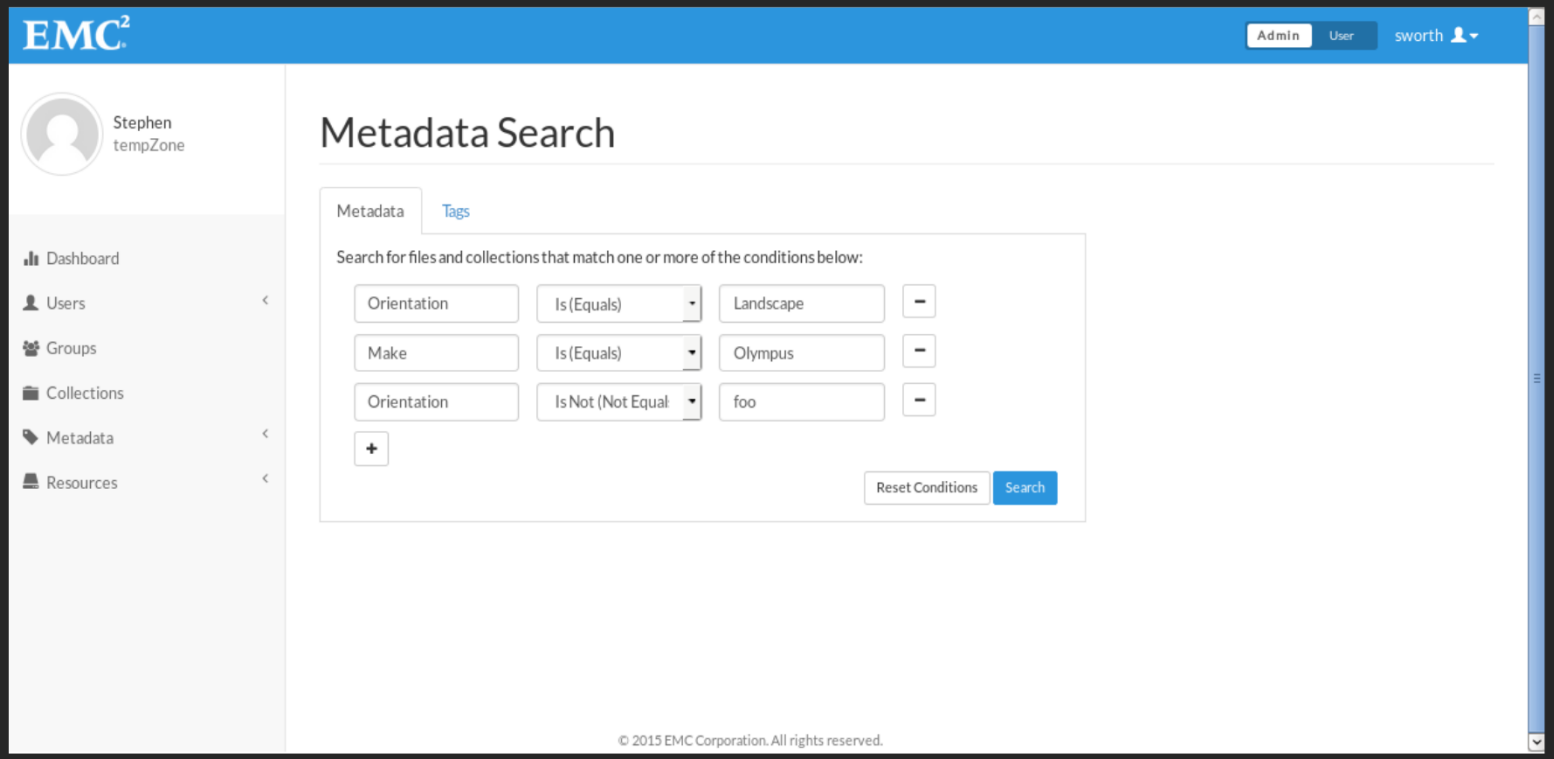
© Copyright 2015 EMC Corporation. All rights reserved.

Collections - Extracted Metadata



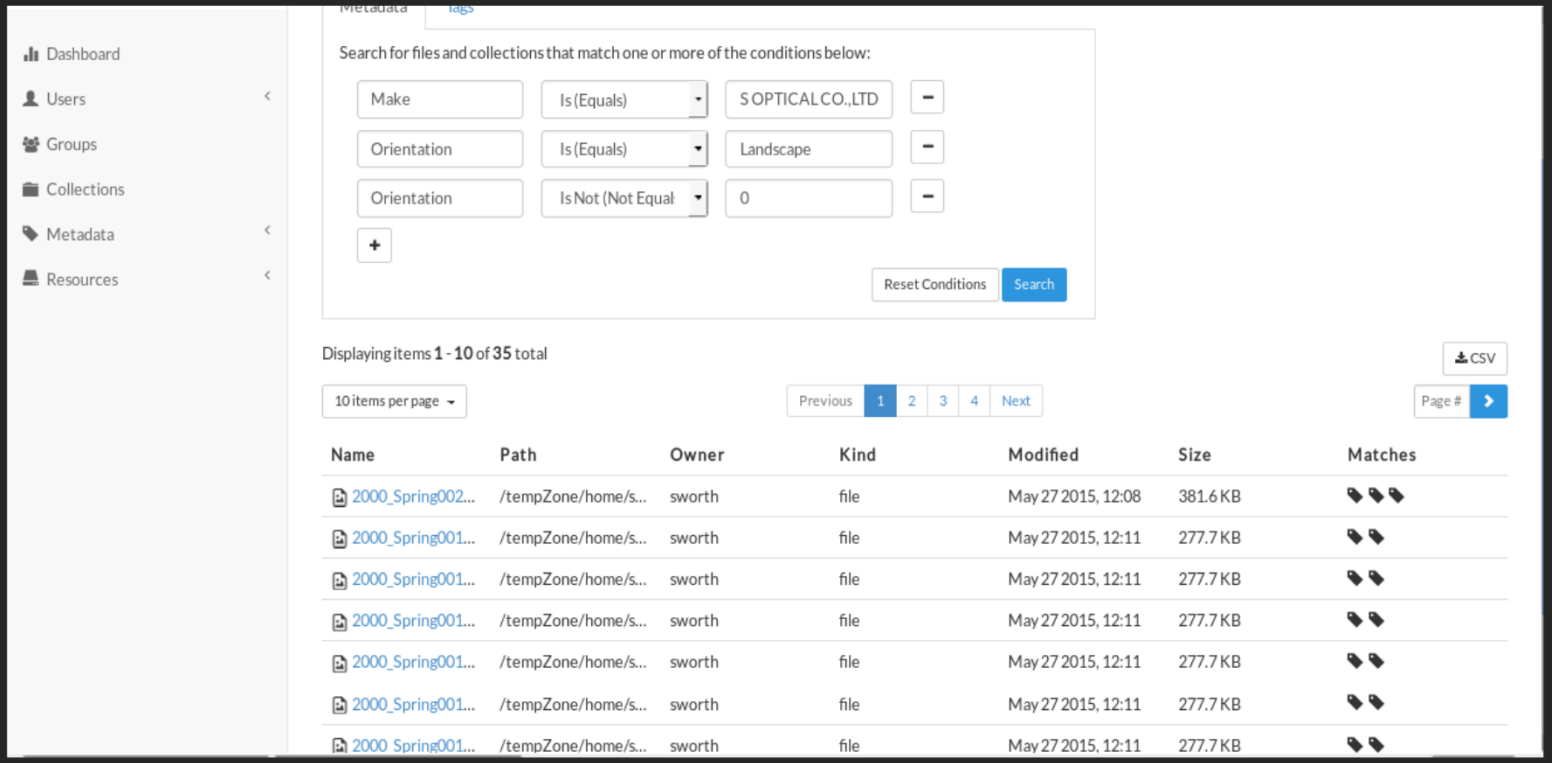
© Copyright 2015 EMC Corporation. All rights reserved.

Metadata Search



© Copyright 2015 EMC Corporation. All rights reserved.

Metadata Search Results



© Copyright 2015 EMC Corporation. All rights reserved.

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

© Copyright 2015 EMC Corporation. All rights reserved.

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

© Copyright 2015 EMC Corporation. All rights reserved.

Questions ?

Contacts:

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

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

© Copyright 2015 EMC Corporation. All rights reserved.