

# EMC MetaLnx - An iRODS Admin & Metadata UI

*Steve Worth*

Director, Global New Business Development  
Operations

6/10/2015  
iRODS User Group Meeting

**EMC<sup>2</sup>**

# Motivation

- Observations

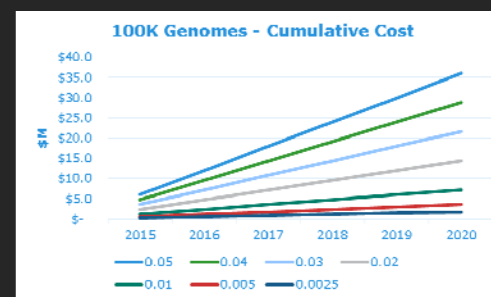
- A data deluge is upon us – will surpass 44 Zetabytes by 2020
  - Big data, IOT, and mobile computing are changing how we interact with data
- Cost dynamics are changing
  - Cloud computing / open source feed the “race to zero” in computing
- Opex costs begin to dominate over Capex costs

- 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)

(In past times this was called Library Science)

# Want Proof?

This is the modern day collegiate library....



See any  
books?



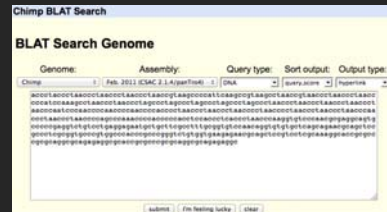
James B. Hunt Library – North Carolina State Univ.

EMC<sup>2</sup>

# IT World According to Genomes

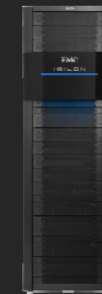
## Analysis:

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



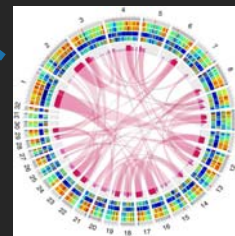
## Data Management:

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



## Workflow:

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



## 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*

EMC<sup>2</sup>

# 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)

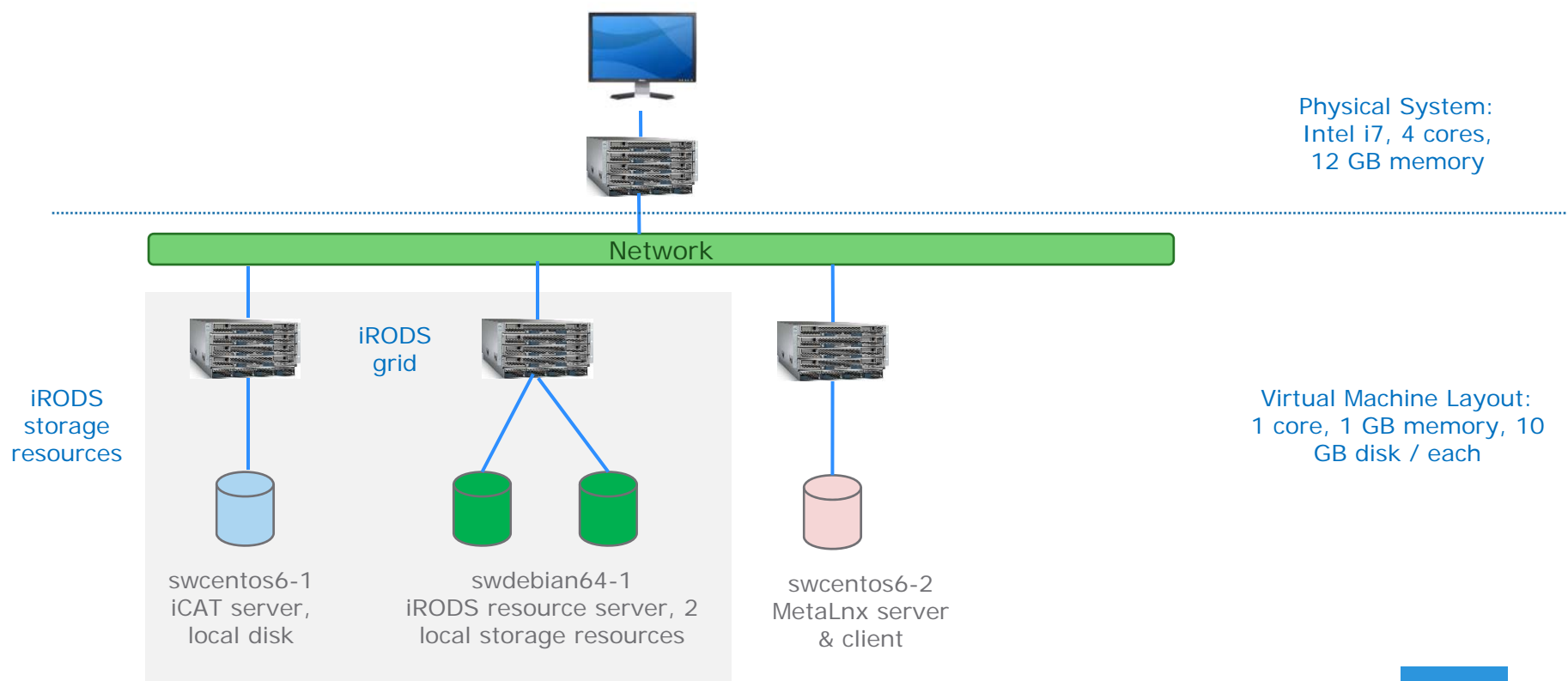
- iinit
- iput
- iget
- imkdir
- ichmod
- icp
- irm
- ils
- ipwd
- icd
- irepl
- iexit
- ipasswd
- ichksum
- imv
- iphymv
- ireg
- irmtrash
- irsync

- **Admin commands** (subset)

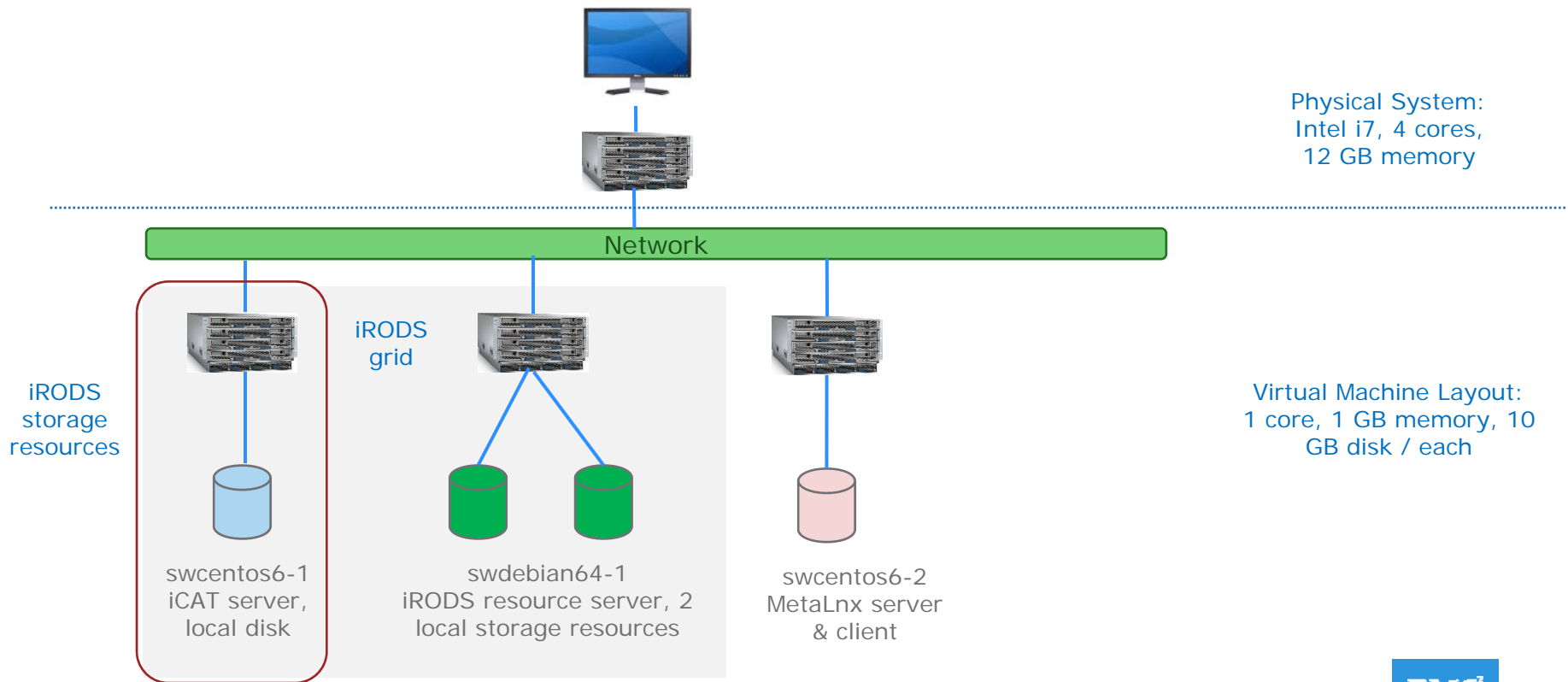
- lt
- lr
- ls
- lz
- lg
- mkuser
- moduser
- aua
- rua
- rpp
- rmuser
- mkdir
- rmdir
- mkresc
- modresc
- modrescdatapaths
- rmresc
- mkzone
- modzone



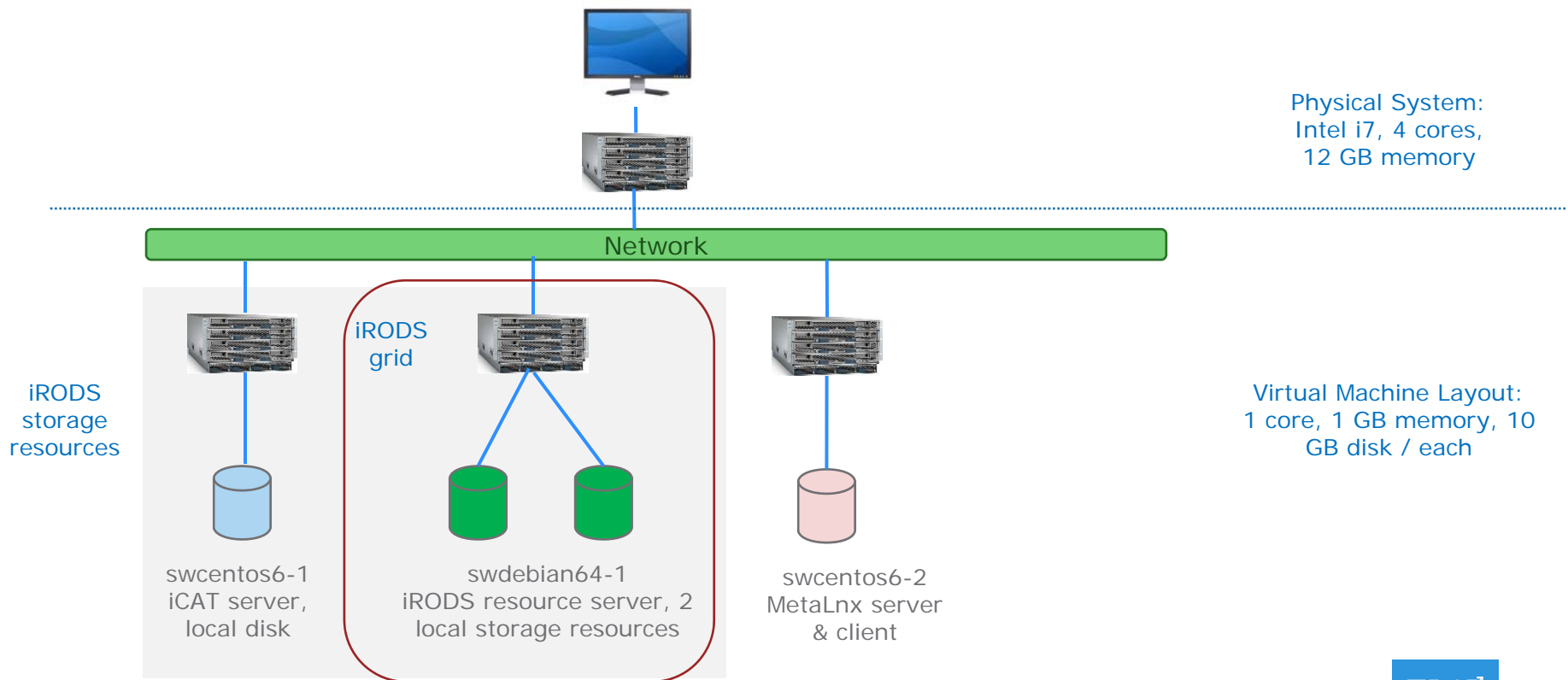
# EMC MetaLnx Demo Setup



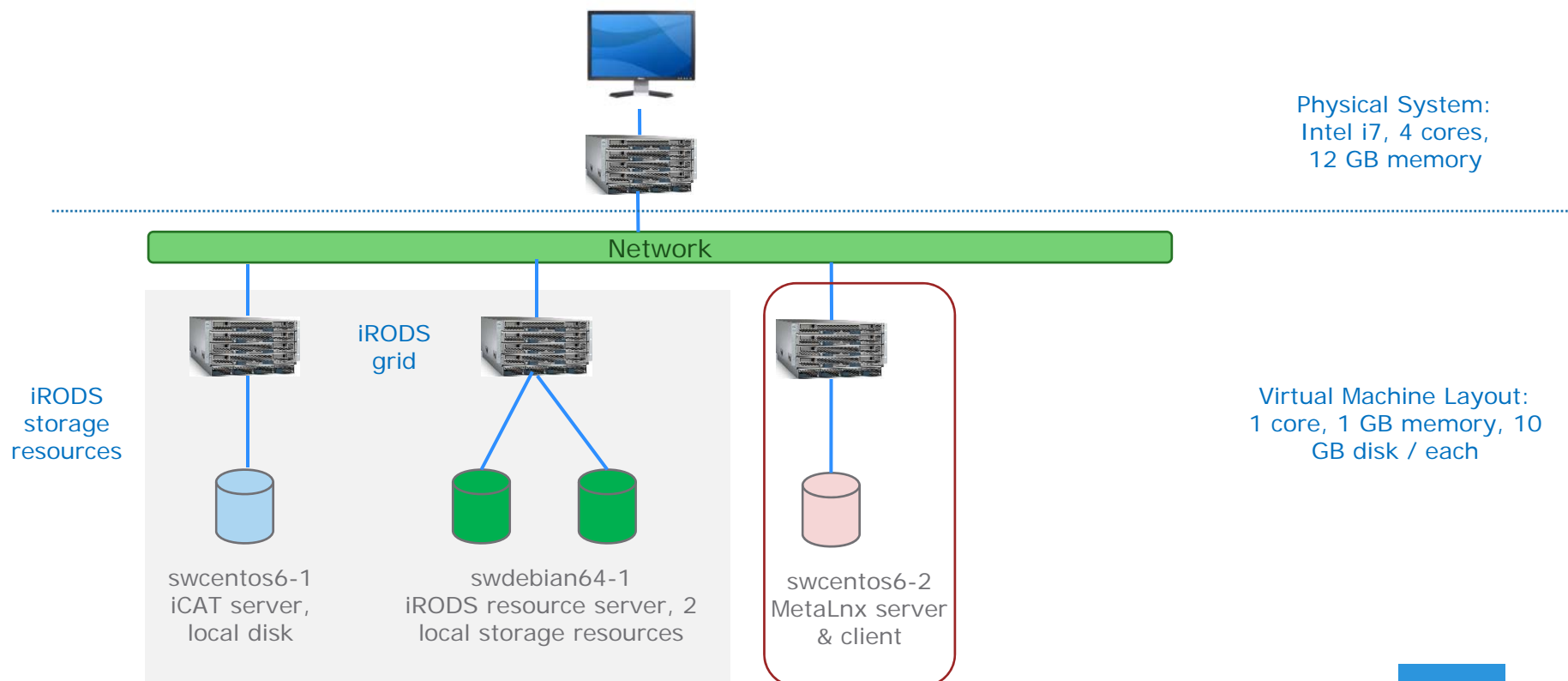
# EMC MetaLnx Demo Setup



# EMC MetaLnx Demo Setup



# EMC MetaLnx Demo Setup



# MetaLnx Live Demo

The screenshot displays the EMC² MetaLnx Dashboard. The interface features a blue header with the EMC² logo and a user profile for 'sworth'. A left sidebar contains navigation links: Dashboard, Users, Group Management, Collections, Resources, Zones and Federations, ICAT, and Private Library. The main content area is titled 'Dashboard' and includes several widgets: 'System Health' showing 'Overall Status: normal', 'iRODS Servers' listing 'swcentos6-1' (192.168.1.5) and 'swdebian64-1' (192.168.1.6), 'Total Storage' showing '33 G free of 43 G' and '10 Files', 'Users' showing 'Users (8)' and 'Groups (5)', and 'Quick Access' with links to 'Add User', 'Groups', 'Collections', and 'Resources'. A text overlay in the bottom left of the dashboard area reads: 'Selected images added to show functionality'.

# Dashboard - iRODS Server Down

The screenshot displays the EMC<sup>2</sup> iRODS dashboard interface. At the top, a blue header bar contains the EMC<sup>2</sup> logo on the left and navigation links for 'Admin', 'User', and a user profile 'sworth' on the right. A left-hand sidebar lists navigation options: Dashboard, Users, Groups, Collections, Resources, Zones and Federations, iCAT, and Private Library. The main content area is titled 'Dashboard' and features several widgets. The 'System Health' widget, highlighted with a red header, shows an 'Overall Status: error'. Below it, the 'iRODS Servers' widget lists two servers: 'swdebian64-1' at IP 192.168.52.129 and 'swcentos6-1' at IP 192.168.52.128. To the right, the 'Total Storage' widget shows a progress bar indicating '16 G free of 34 G' and '9380 Files'. Below that, the 'Users' widget shows 'Users (11)' and 'Groups (8)'. At the bottom right, a 'Quick Access' widget provides links to 'Add User', 'Groups', 'Collections', 'Resources', and 'Map'. The bottom of the dashboard includes a copyright notice and a small EMC<sup>2</sup> logo in the corner.

EMC<sup>2</sup>

Admin User sworth

Stephen tempZone

## Dashboard

### System Health

**Overall Status:** error

### iRODS Servers

swdebian64-1	192.168.52.129
swcentos6-1	192.168.52.128

### Total Storage

16 G free of 34 G

9380 Files

### Users

Users (11)  
Groups (8)

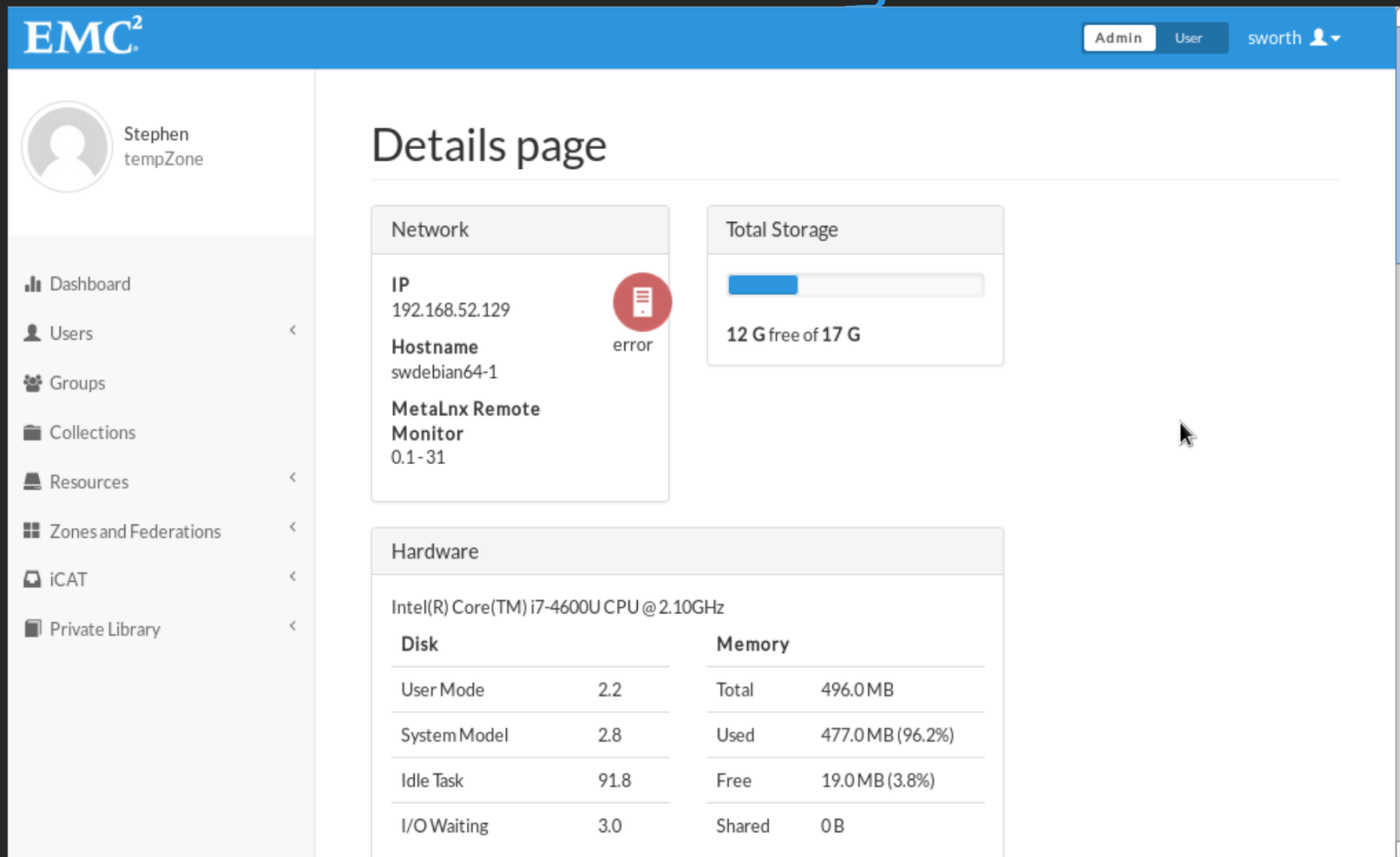
### Quick Access

- Add User
- Groups
- Collections
- Resources
- Map

© Copyright 2015 EMC Corporation. All rights reserved.

EMC<sup>2</sup>

# Server Details Page



**EMC<sup>2</sup>** Admin User sworth

**Stephen tempZone**

## Details page

### Network

**IP**  
192.168.52.129

**Hostname**  
swdebian64-1

**MetaLnx Remote Monitor**  
0.1-31

error

### Total Storage

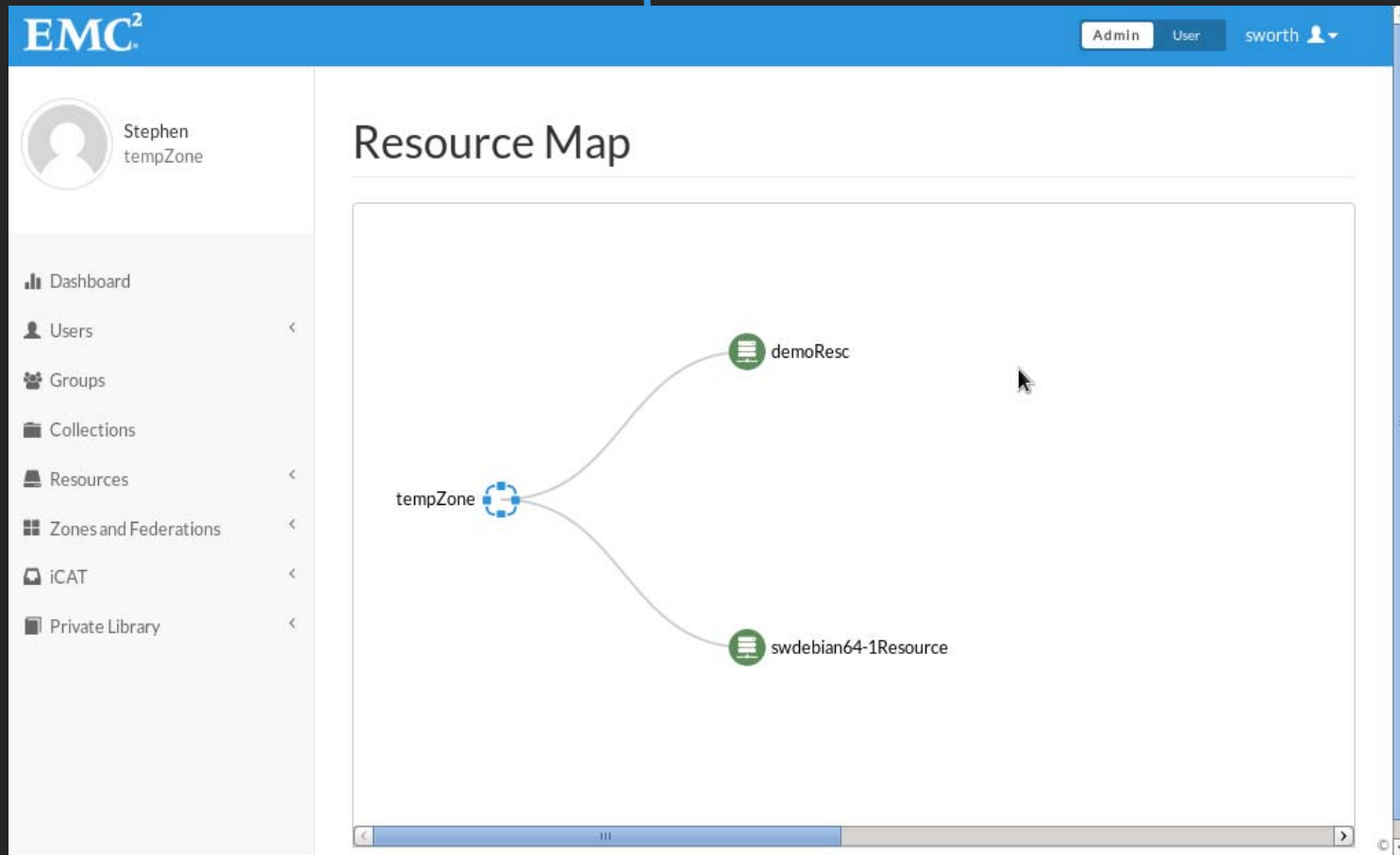
12 G free of 17 G

### Hardware

Intel(R) Core(TM) i7-4600U CPU @ 2.10GHz

Disk		Memory	
User Mode	2.2	Total	496.0 MB
System Model	2.8	Used	477.0 MB (96.2%)
Idle Task	91.8	Free	19.0 MB (3.8%)
I/O Waiting	3.0	Shared	0 B

# Resource Map





# User Management

EMC<sup>2</sup> Admin User sworth

Stephen tempZone

Dashboard  
Users  
Groups  
Collections  
Resources  
Zones and Federations  
iCAT  
Private Library

## User Management

Search

11 user(s) found

First Name	Last Name	Username	Zone	Email	Action
		bjenkins	tempZone		<input type="button" value="Edit"/> <input type="button" value="Delete"/>
		bjones	tempZone		<input type="button" value="Edit"/> <input type="button" value="Delete"/>
		cstacy	tempZone		<input type="button" value="Edit"/> <input type="button" value="Delete"/>
		jjones	tempZone		<input type="button" value="Edit"/> <input type="button" value="Delete"/>
James	Kirk	jkirk	tempZone	jkirk@gmail.com	<input type="button" value="Edit"/> <input type="button" value="Delete"/>
Joel	Schwartz	jschwartz	tempZone		<input type="button" value="Edit"/> <input type="button" value="Delete"/>
		rods	tempZone		<input type="button" value="Edit"/> <input type="button" value="Delete"/>
		sjones	tempZone		<input type="button" value="Edit"/> <input type="button" value="Delete"/>
Sue	Smith	ssmith	tempZone	ssmith@aol.com	<input type="button" value="Edit"/> <input type="button" value="Delete"/>

# Collections - Pagination

The screenshot displays the EMC2 Collections web interface. The top navigation bar is blue with the EMC2 logo on the left and user controls (Admin, User, sworth) on the right. A left sidebar contains a navigation menu with items like Dashboard, Users, Groups, Collections, Resources, Zones and Federations, ICAT, and Private Library. The main content area is titled 'Collections' and includes a search bar. Below the title is a breadcrumb trail: iRods / tempZone / home / sworth / lots\_of\_files. There are tabs for Browse, Info, Metadata, and Permissions, along with links for Upload and Add Collection. The pagination section shows '25 items per page' and a list of page numbers from 1 to 344, with page 123 currently selected. Below this, it states '25 result(s) found'. A table lists the first six items, each with a checkbox, a file icon, a name, an owner, a kind, and a modified date.

EMC<sup>2</sup>

Admin User sworth

Stephen tempZone

## Collections

Search

iRods / tempZone / home / sworth / lots\_of\_files

Browse Info Metadata Permissions Upload Add Collection

25 items per page Previous 1 ... 121 122 123 124 125 ... 344 Next Page # Go

25 result(s) found

<input type="checkbox"/>	Name	Owner	Kind	Modified
<input type="checkbox"/>	2002_cruise0490 (3).jpg	sworth	file	Apr 08, 2015 07:04
<input type="checkbox"/>	2002_cruise0490 (4).jpg	sworth	file	Apr 08, 2015 07:04
<input type="checkbox"/>	2002_cruise0490 (5).jpg	sworth	file	Apr 08, 2015 07:04
<input type="checkbox"/>	2002_cruise0490 (6).jpg	sworth	file	Apr 08, 2015 07:04
<input type="checkbox"/>	2002_cruise0490 (7).jpg	sworth	file	Apr 08, 2015 07:04
<input type="checkbox"/>	2002_cruise0490 (8).jpg	sworth	file	Apr 08, 2015 07:04

# Collections - Upload

The screenshot shows the EMC² Collections Upload dialog box. The dialog has a title bar 'Upload' and a close button. Inside, there's a 'Select files for upload' section with a 'Browse' button. Below this, a list of 9 files is shown, all named '2012\_Spring\_0055' followed by a number in parentheses and '.jpg'. A status bar indicates '9 file(s) selected'. Below the file list is a 'Resource' dropdown menu currently set to 'demoResc'. At the bottom, there are three checkboxes: 'Overwrite duplicate files' (unchecked), 'Checksum' (checked), and 'Replica' (unchecked). The dialog has 'Cancel' and 'Upload' buttons at the bottom right.

EMC<sup>2</sup>

Stephen  
tempZone

Dashboard  
Users  
Groups  
Collections  
Resources  
Zones and Federations  
iCAT  
Private Library

Admin User sworth


Upload Add Collection

Page # Go

Modified  
Apr 17, 2015 11:04

© 2015 EMC Corporation. All rights reserved.

# Collections - Extracted Metadata



Stephen  
tempZone

Dashboard

Users

Groups

Collections

Metadata

Resources










## Collections

iRods / tempZone / home / sworth / test\_collection / 2010\_snow\_0092.JPG

BackInfoMetadataPermissions

Add

31 result(s) found

<input type="checkbox"/>	Attribute	Value	Unit	
<input type="checkbox"/>	ColorSpace	-1	-	
<input type="checkbox"/>	Compression	6	-	
<input type="checkbox"/>	Contrast	2	-	
<input type="checkbox"/>	CustomRendered	0	-	
<input type="checkbox"/>	DateTime	2010:04:25 23:43:36	-	
<input type="checkbox"/>	DateTimeDigitized	2010:02:13 11:27:54	-	
<input type="checkbox"/>	DateTimeOriginal	2010:02:13 11:27:54	-	
<input type="checkbox"/>	ExposureMode	0	-	
<input type="checkbox"/>	ExposureProgram	0	-	

# Metadata Search

EMC<sup>2</sup>

Admin User sworth

Stephen tempZone

Dashboard  
Users  
Groups  
Collections  
Metadata  
Resources

## Metadata Search

Metadata Tags

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

© 2015 EMC Corporation. All rights reserved.

# Metadata Search Results

Dashboard

Users

Groups

Collections

Metadata

Resources

Metadata

Tags

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

Make

Is (Equals)

SOPTICAL CO.,LTD

—

Orientation

Is (Equals)

Landscape

—

Orientation

Is Not (Not Equal)

0

—

+

Reset Conditions

Search

Displaying items 1 - 10 of 35 total

10 items per page

Previous

1

2

3























4

Next

Page #

>

CSV

Name	Path	Owner	Kind	Modified	Size	Matches
 2000_Spring002...	/tempZone/home/s...	sworth	file	May 27 2015, 12:08	381.6 KB	  
 2000_Spring001...	/tempZone/home/s...	sworth	file	May 27 2015, 12:11	277.7 KB	 
 2000_Spring001...	/tempZone/home/s...	sworth	file	May 27 2015, 12:11	277.7 KB	 
 2000_Spring001...	/tempZone/home/s...	sworth	file	May 27 2015, 12:11	277.7 KB	 
 2000_Spring001...	/tempZone/home/s...	sworth	file	May 27 2015, 12:11	277.7 KB	 
 2000_Spring001...	/tempZone/home/s...	sworth	file	May 27 2015, 12:11	277.7 KB	 
 2000_Spring001...	/tempZone/home/s...	sworth	file	May 27 2015, 12:11	277.7 KB	 

# 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](mailto:Sasha.Paegle@isilon.com)

Stephen Worth  
Director, Global New Business  
Development Operations  
[Stephen.Worth@emc.com](mailto:Stephen.Worth@emc.com)