

RMOUG TRAINING DAYS

February 11-13, 2013 Denver, Colorado
Colorado Convention Center



www.rmouug.org

THE COLLEGiate PEAKS. PHOTO COURTESY OF JOHN KONG, RMouug MEMBER

Integrating the Oracle Database Appliance with the Sun ZFS Storage Appliance to Create a Secure Scalable High Availability Environment

Daniel Morgan | damorgan12c@gmail.com | morganslibrary.org

Integrating ODAs with ZFS to Create a Secure Scalable HA Environment

Presented: Rocky Mountain Oracle User Group - 13 January, 2013

Agenda

- Introductions
- Executive Summary
- Puzzle Pieces (HA)
- ODA
 - What and Why
 - Installation
 - Value Adds
- ZFS File System
 - Pooling
 - Redundancy
- ZFS Storage Appliance
- Questions

Daniel A. Morgan

 Oracle ACE Director



Consultant to Harvard University



University of Washington Oracle Instructor, ret.



The Morgan of Morgan's Library on the web



Board Member: Western Washington OUG

- Upcoming Presentations

- Feb 18: Victoria, BC
- Feb 19: Vancouver BC
- Mar 13-14: Utah Oracle Users Group
- Apr 15-17: Oracle User Group Finland
- Apr 17-20: Oracle User Group Norway
- May 15 Oracle User Group Azerbaijan
- May 19 Bulgarian Oracle User Group



Official Beta Site



Morgan's Library: www.morganslibrary.org

 **Morgan's Library**

Morgan's 2012 - 2013 Calendar

Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Jan
[don't censor the web]						[don't be silent]					

[Home](#)

[Community](#)

[Events](#)

[Training](#)

[Evening Workshops](#)

Resources 

[Library](#)

[How Can I?](#)

[Code Samples](#)

[Presentations](#)

[Links](#)

[Book Reviews](#)

[Downloads](#)

[User Groups](#)

[Humor](#)

[General](#)

[Contact](#)

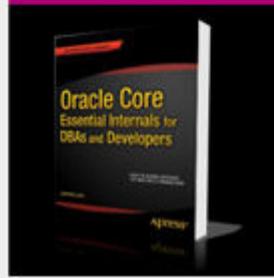
[About](#)

[Services](#)

[Legal Notice & Terms of Use](#)

[Privacy Statement](#)

Presentations Map 

Oracle Core 

Training Events

- [Harvard University - Feb 6 - 10, 2012](#)
- [RMOUG Training Days - Feb 14 - 16, 2012](#)
- [VanOUG Canada - Feb 23, 2012](#)
- [OUGN: Oslo, Norway - Mar 21 - 24, 2012](#)
- [BeNeLux Connect: Maastricht - Apr 24, 2012](#)
- [OUG Harmony Finland - May 30 - 31, 2012](#)
- [OUG Harmony Latvia - Jun 01, 2012](#)

Morgan 
aboard USA-71

Library News

- [Morgan's Notepad vi \(Blog\)](#)
- [Join the Western Washington OUG](#)
- [Morgan's Oracle Podcast](#)
- [US Government STIGs \(Security Checklists\)](#)
- [Bryn Llewellyn's PL/SQl White Paper](#)
- [Bryn Llewellyn's Editioning White Paper](#)
- [Explain Plan White Paper](#)
- [Troubleshooting Performance](#)

Oracle Events 

Next Event: RMOUG Denver, CO Feb 14-16

ACE News

Would you like to become an Oracle ACE? 

[Learn more about becoming an ACE](#)



- [ACE Directory](#)
- [ACE Google Map](#)
- [ACE Program](#)
- [Stanley's Blog](#)

Congratulations to our newest ACEs

Daniel Morgan | damorgan12c@gmail.com | morganslibrary.org

Integrating ODAs with ZFS to Create a Secure Scalable HA Environment

Presented: Rocky Mountain Oracle User Group - 13 January, 2013

Travels



Daniel Morgan | damorgan12c@gmail.com | morganslibrary.org

Integrating ODAs with ZFS to Create a Secure Scalable HA Environment

Presented: Rocky Mountain Oracle User Group - 13 January, 2013

Executive Summary

- Never make a technology argument when a financial argument will suffice
 - Your CFO wants to talk about ROI not IOPS
 - Will this technology support our organization's needs?
 - Can we right-size it today and will it scale for tomorrow?
 - Does it meet our regulatory and compliance requirements?
 - What is involved in migrating current operations to it?
 - Can our existing team deploy and maintain it?
 - Can we find qualified technologists who already know it?
 - Can the vendor(s) involved fully support the tech stack?
 - How will this affect our customers?
 - How will this affect our financial position?
 - capital expense to obtain it
 - operating expense to maintain it
 - future retirement expense

Questions we as IT professionals must answer

- Why does deployment take so long and cost so much?
- Why are we spending so much on support?
- Why does patching so often break something else?
- Why do we spend so much time fighting fires?



IT infrastructure on a good day



Daniel Morgan | damorgan12c@gmail.com | morganslibrary.org

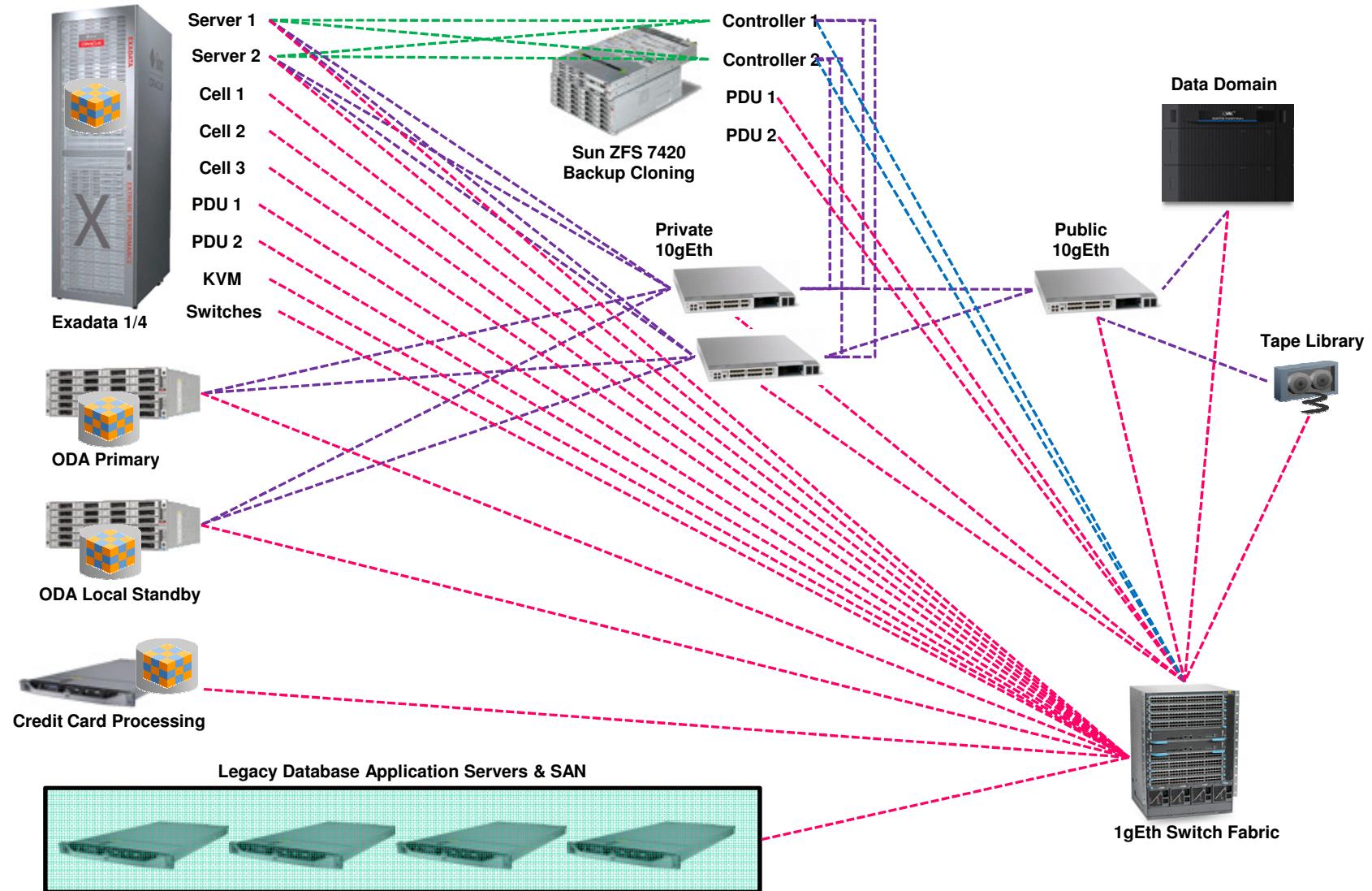
Integrating ODAs with ZFS to Create a Secure Scalable HA Environment

Presented: Rocky Mountain Oracle User Group - 13 January, 2013

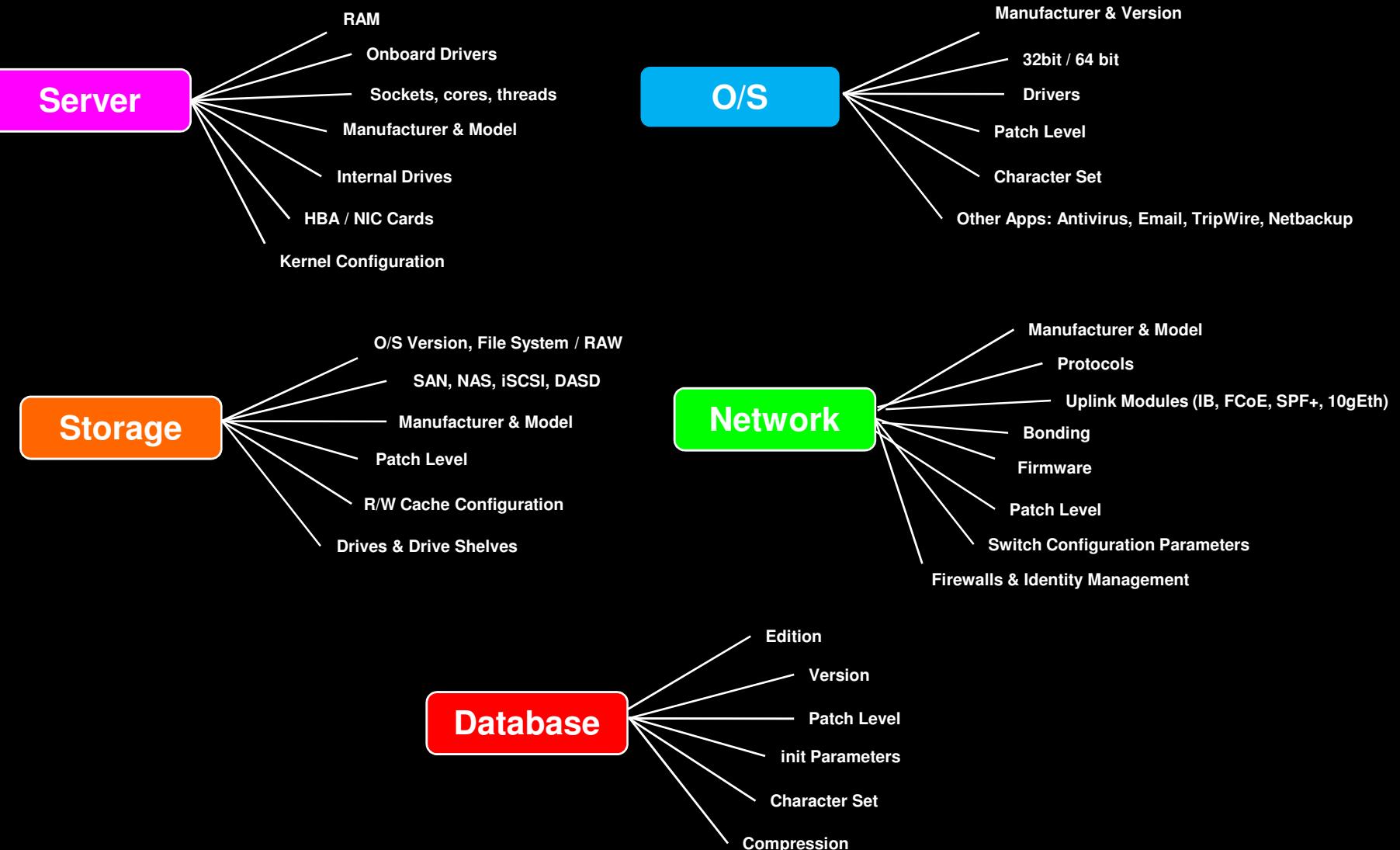
IT infrastructure meets a single point of failure



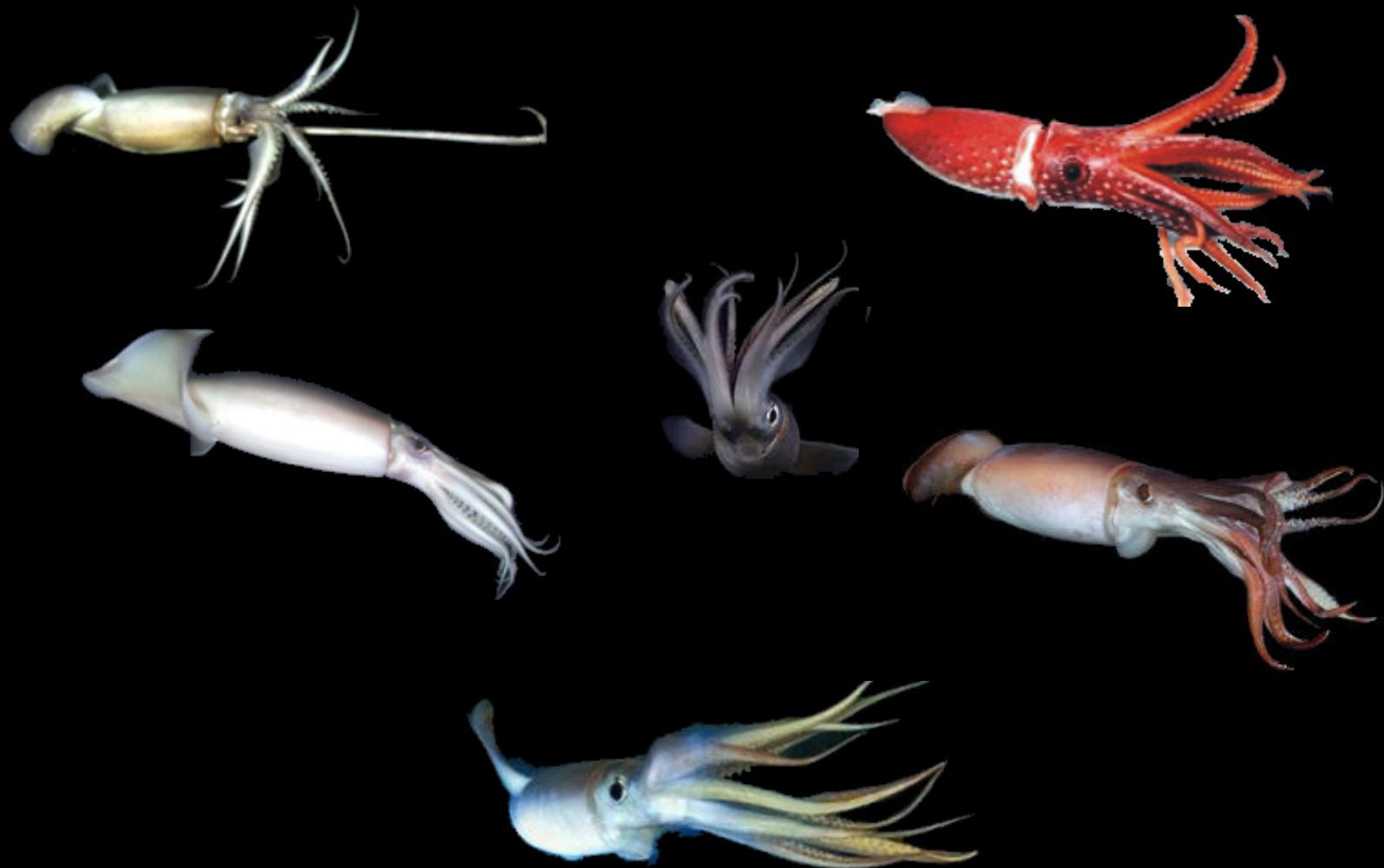
Puzzle Pieces



Static Puzzle Pieces



Animated Puzzle Pieces



Daniel Morgan | damorgan12c@gmail.com | morganslibrary.org

Integrating ODAs with ZFS to Create a Secure Scalable HA Environment

Presented: Rocky Mountain Oracle User Group - 13 January, 2013

It's hard to embrace a barrel of squid



Puzzle Pieces

- The decisions we've made in the past guarantee that
 - No one has ever built a RAC cluster with our configuration
 - No one has ever applied operating system and firmware patches to our configuration
 - No one has ever patched to our configuration
 - Oracle has never tested and certified our configuration
 - No one in support can exactly duplicate our environment



LONELINESS

IF YOU FIND YOURSELF STRUGGLING WITH LONELINESS, YOU'RE NOT ALONE.
AND YET YOU ARE ALONE. SO VERY ALONE.

Daniel Morgan | damorgan12c@gmail.com | morganslibrary.org

Integrating ODAs with ZFS to Create a Secure Scalable HA Environment

Presented: Rocky Mountain Oracle User Group - 13 January, 2013

The Solution

make different, and better, decisions

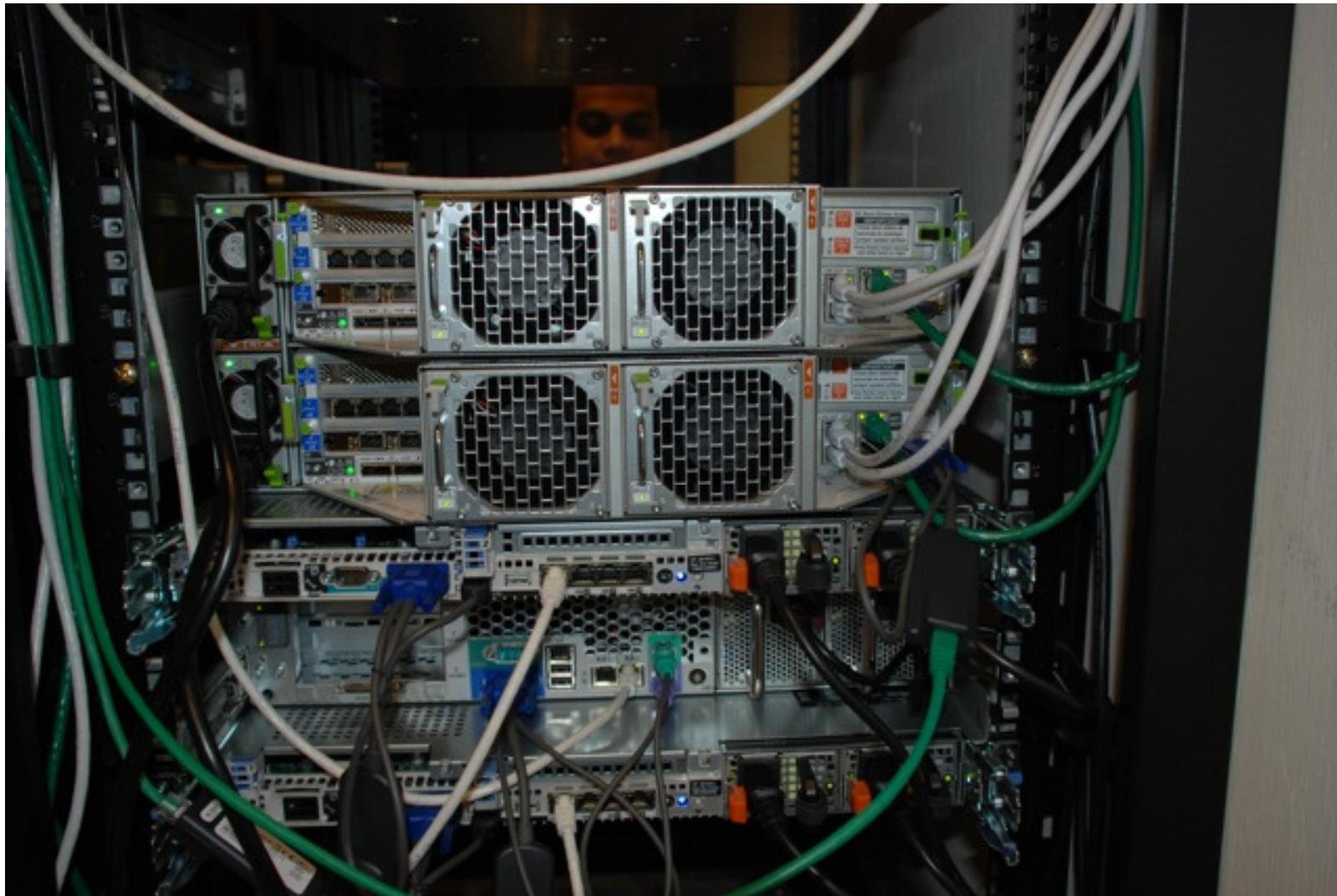
What is an ODA?

- An appliance
 - A single line on the invoice ... plus the power chords
 - But you get root and sys: The customer is in control
- Announced last year at OpenWorld
- Engineered two server RAC cluster in a 4U case
 - 24 CPU cores
 - 192 GB of RAM
 - 12 TB of direct attached storage with ASM mirroring
 - 2TB RAID mirrored disks for O/S and Oracle binaries
- One size fits all ... but ...
- License only the resources you need
- Cores licensed dictate all on-board Oracle licensing

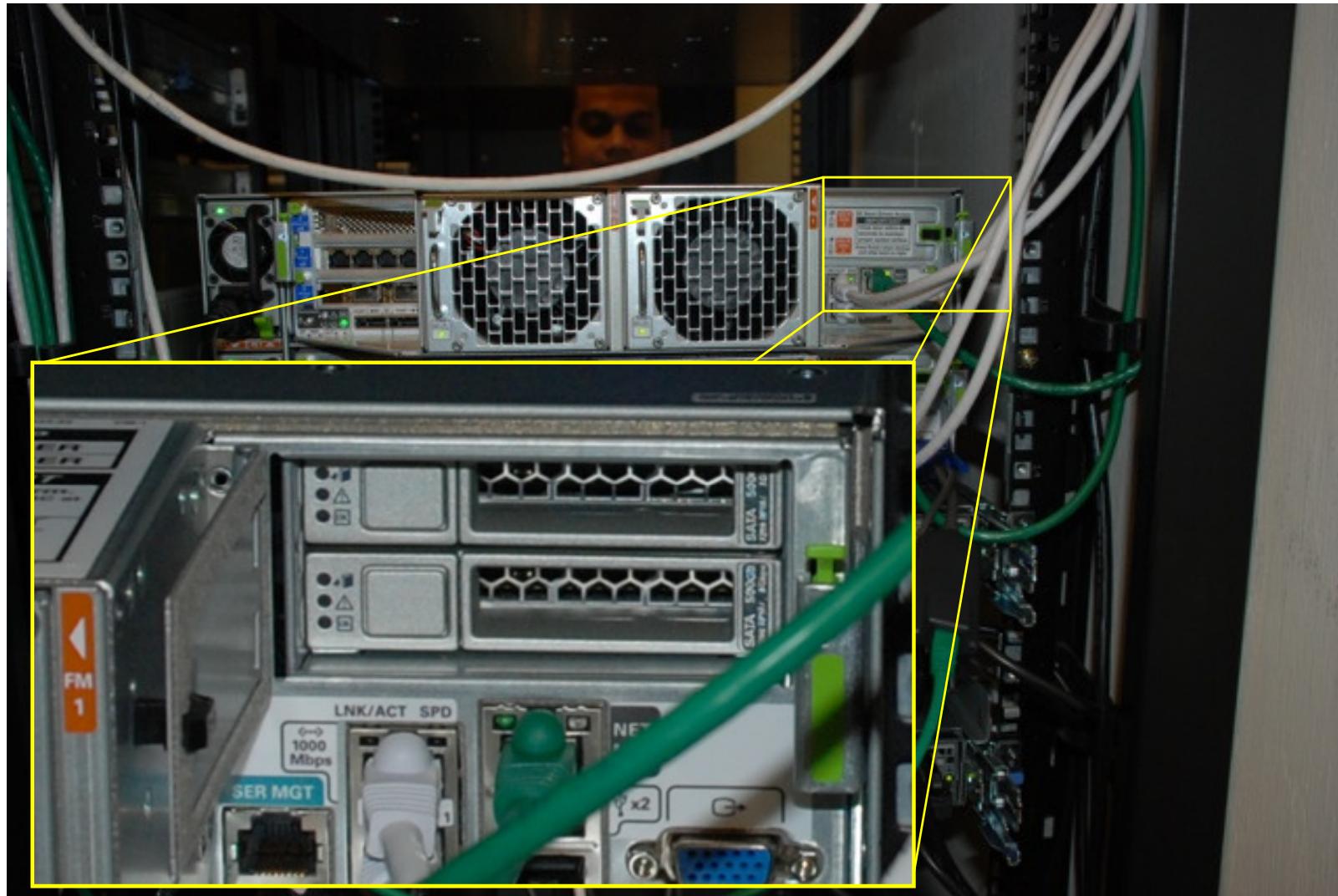
ODA in Pictures



ODA in Pictures



ODA in Pictures



Why an ODA?

- Minimize complexity from rack-and-stack through database deployment
- Fewer resources required to deploy
 - UNIX System Admins: not required
 - Network Admins: not required
 - Storage Admins: not required
- Ease of maintenance and patching
 - One patch combines O/S, drivers, networking, infrastructure
 - One patch database
- Supports multiple Oracle databases
- Petabyte storage available with ZFS
- Can form the basis for deploying HA applications in organizations that lack in-depth technical resources

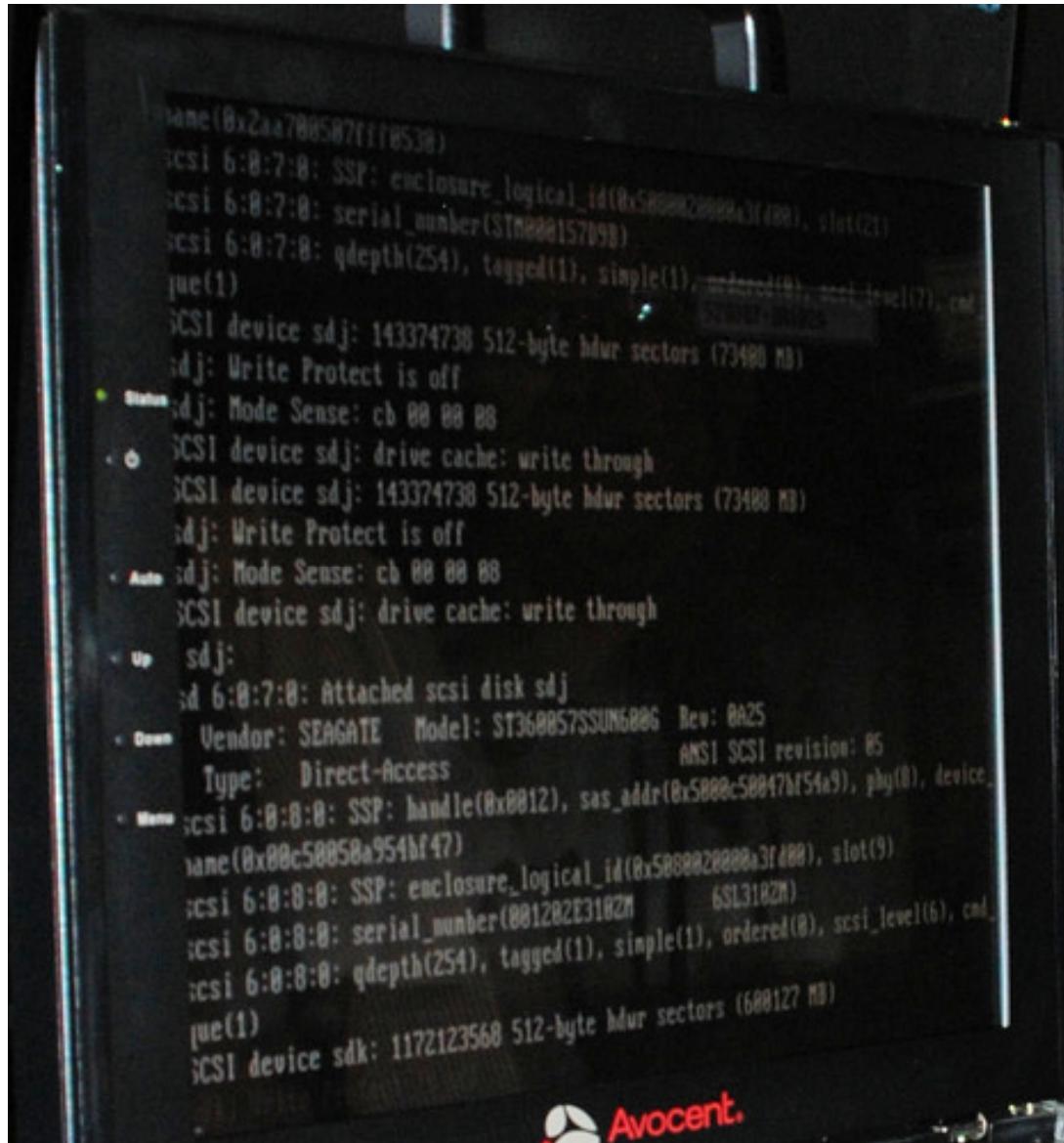
No rolling patches ... and they are not childproof



Discussion

Installation

Step 1: Power On

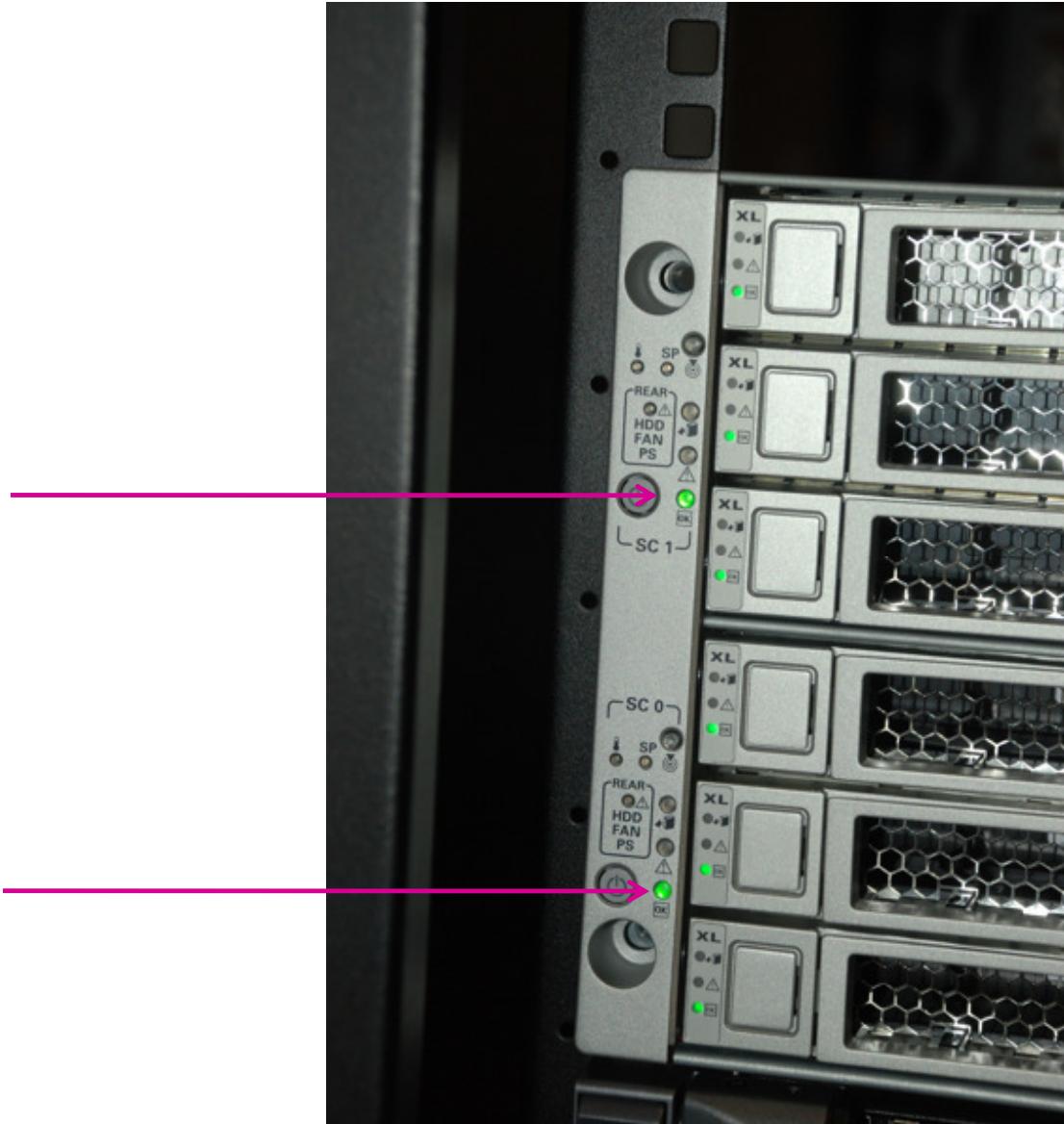


Daniel Morgan | damorgan12c@gmail.com | morganslibrary.org

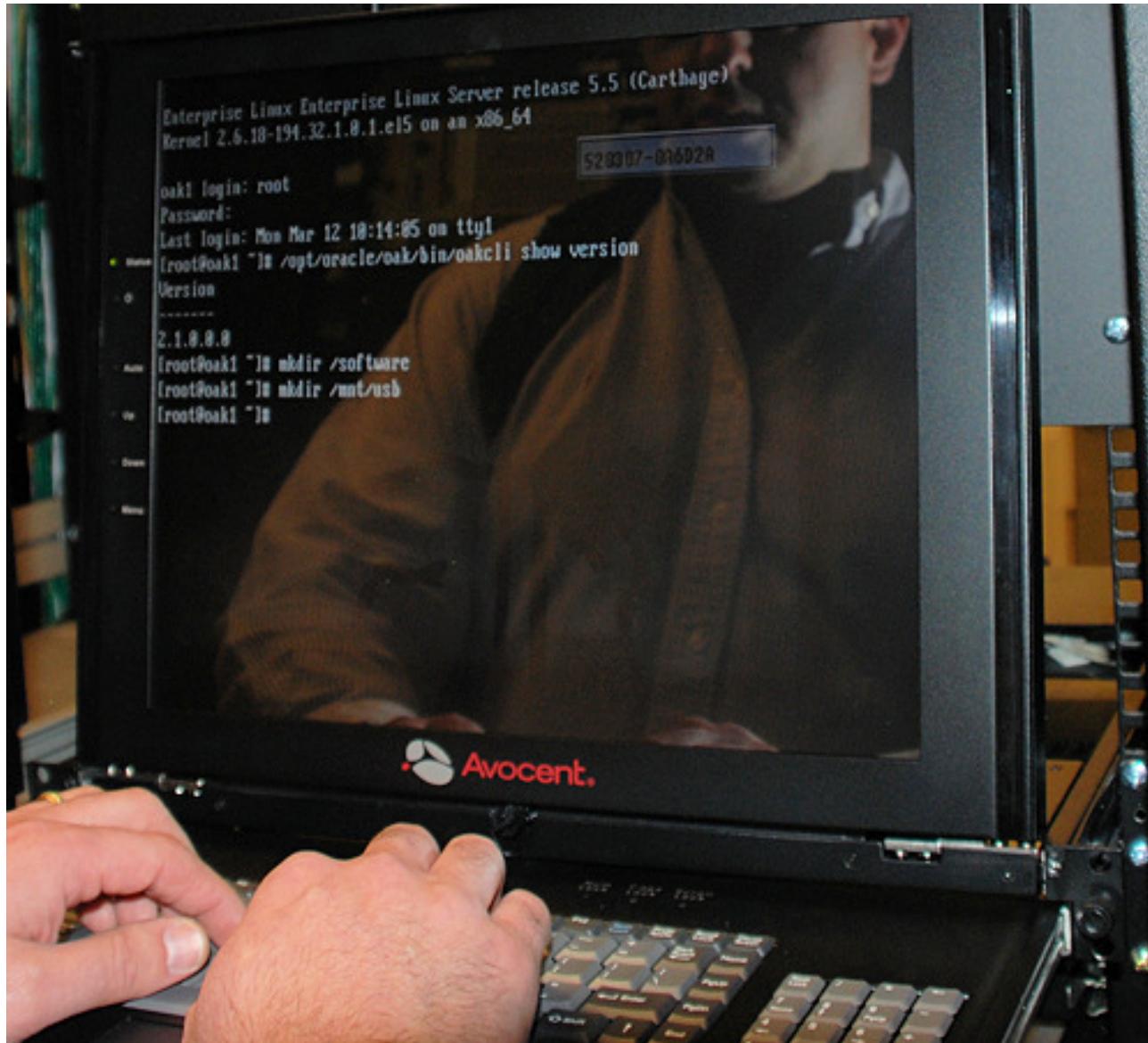
Integrating ODAs with ZFS to Create a Secure Scalable HA Environment

Presented: Rocky Mountain Oracle User Group - 13 January, 2013

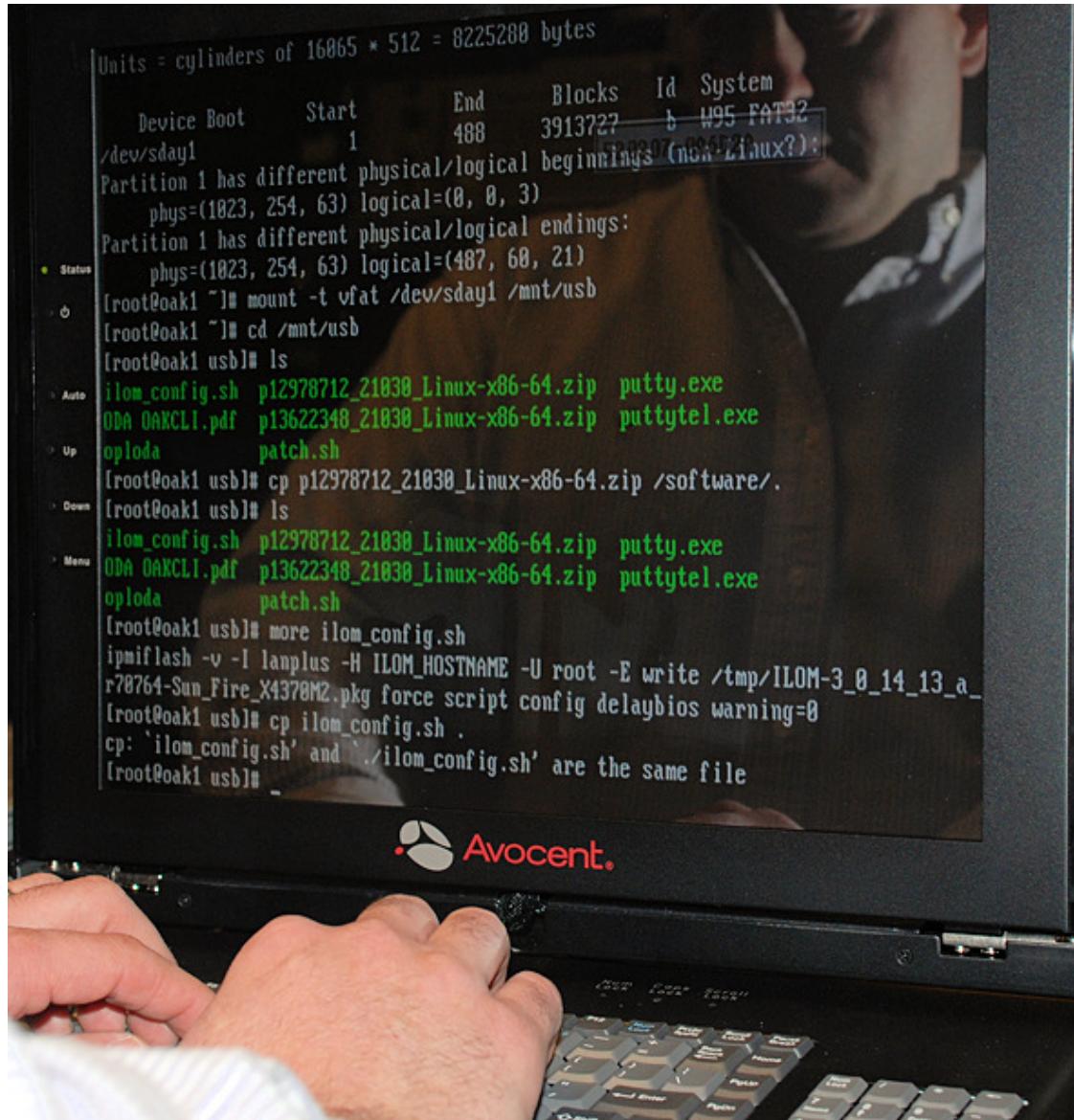
Step 2: Wait for OK Lights



Step 3: Log In as root



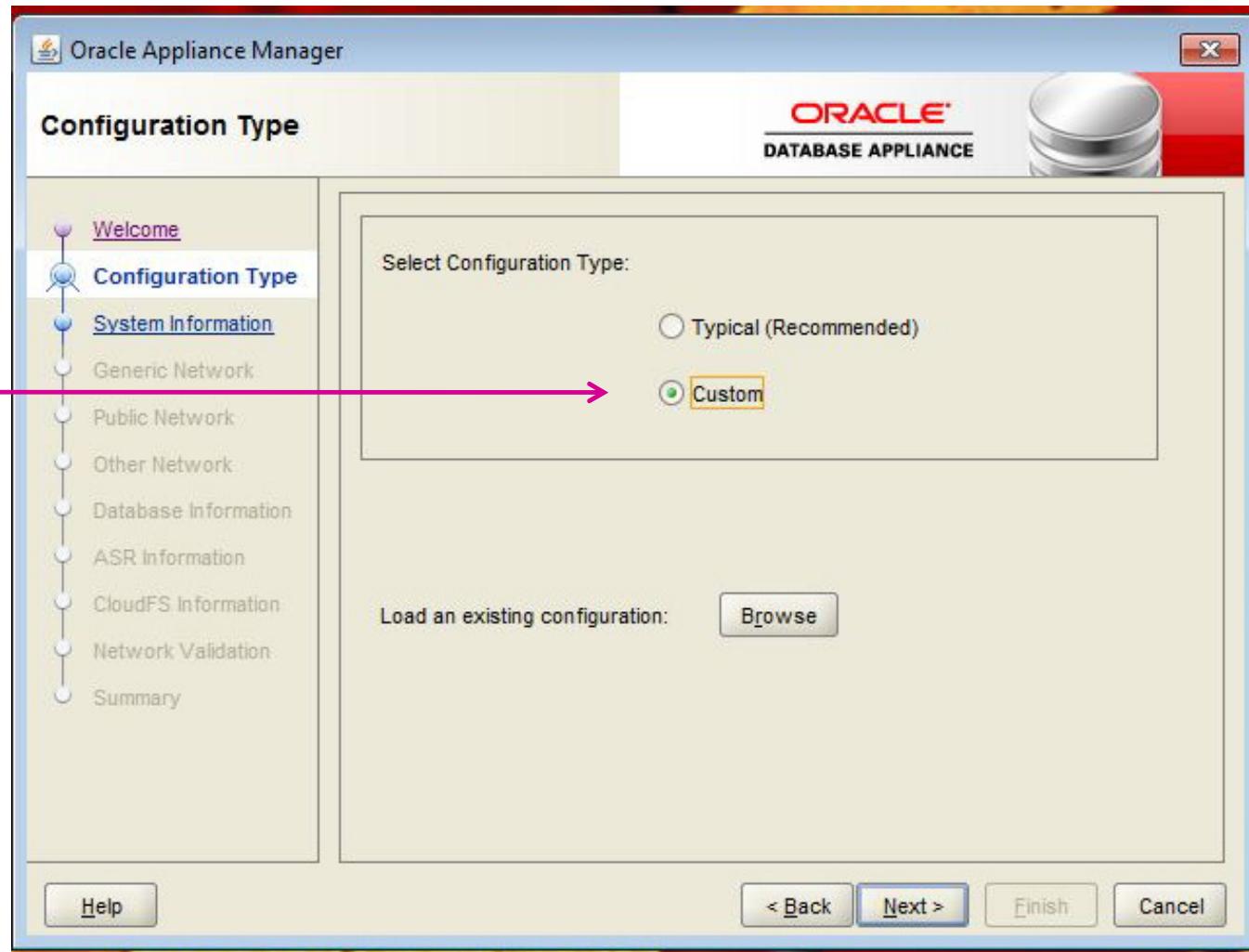
Step 5: Perform ILOM Configuration



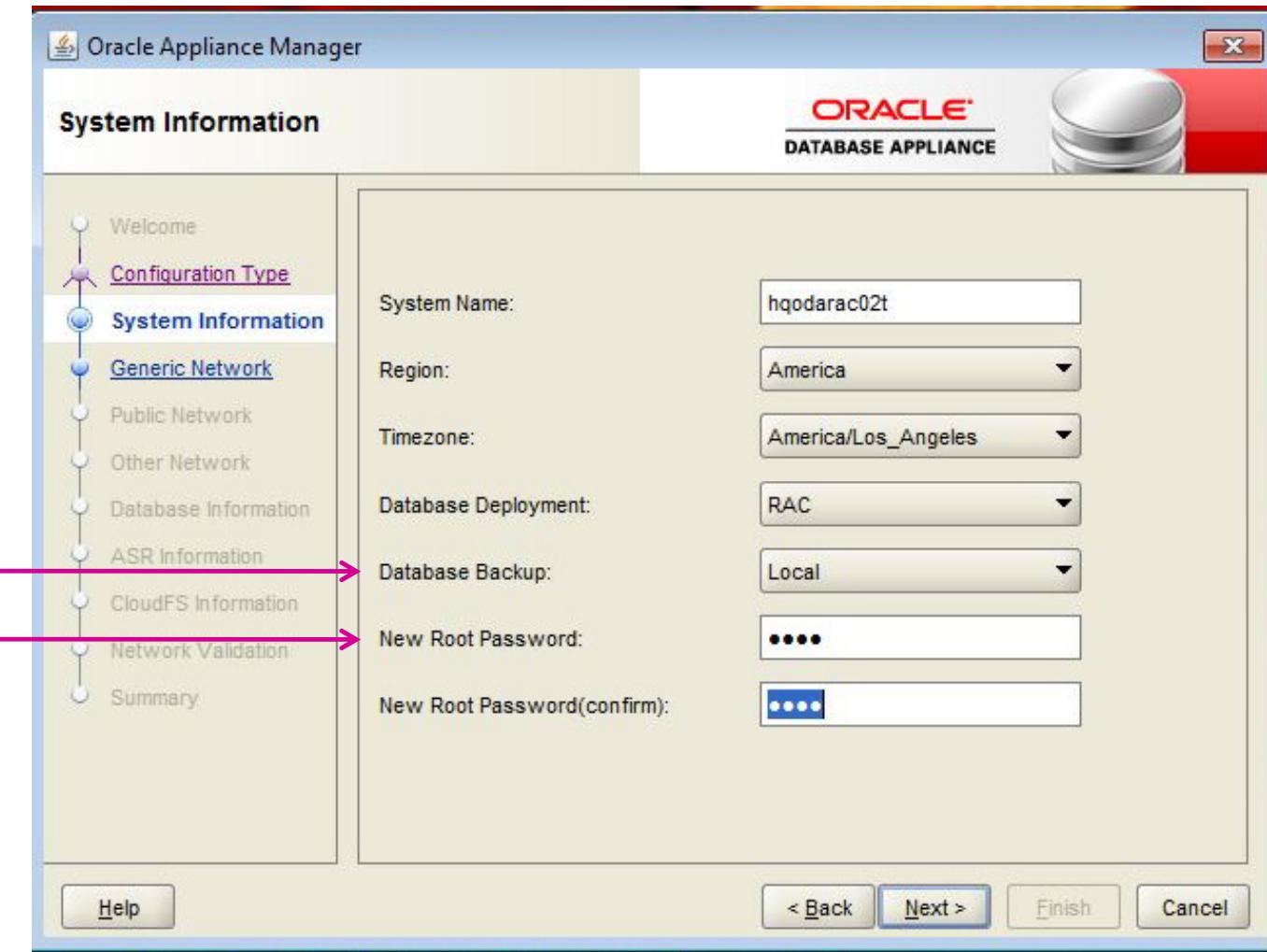
Install Screens: 1



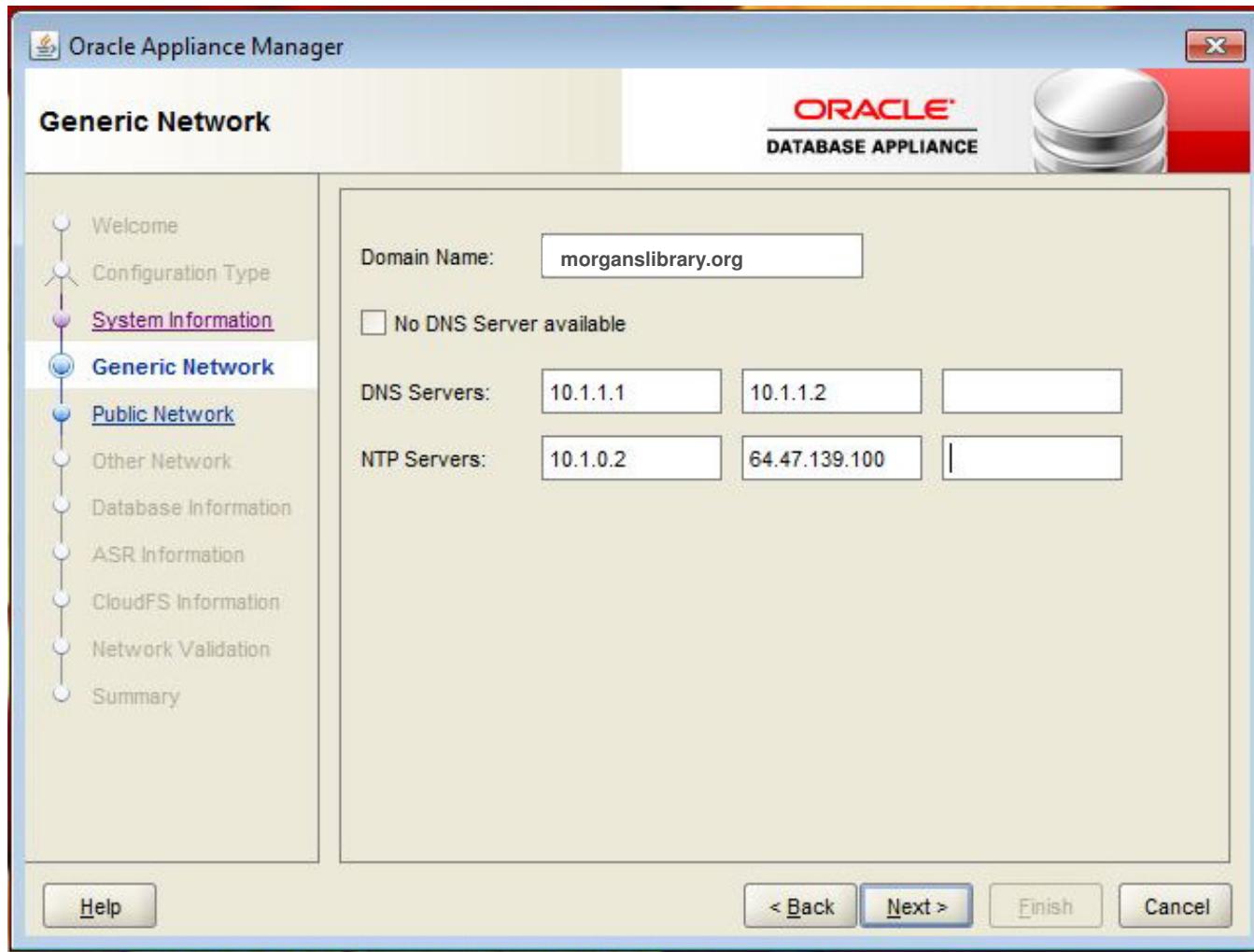
Install Screens: 2



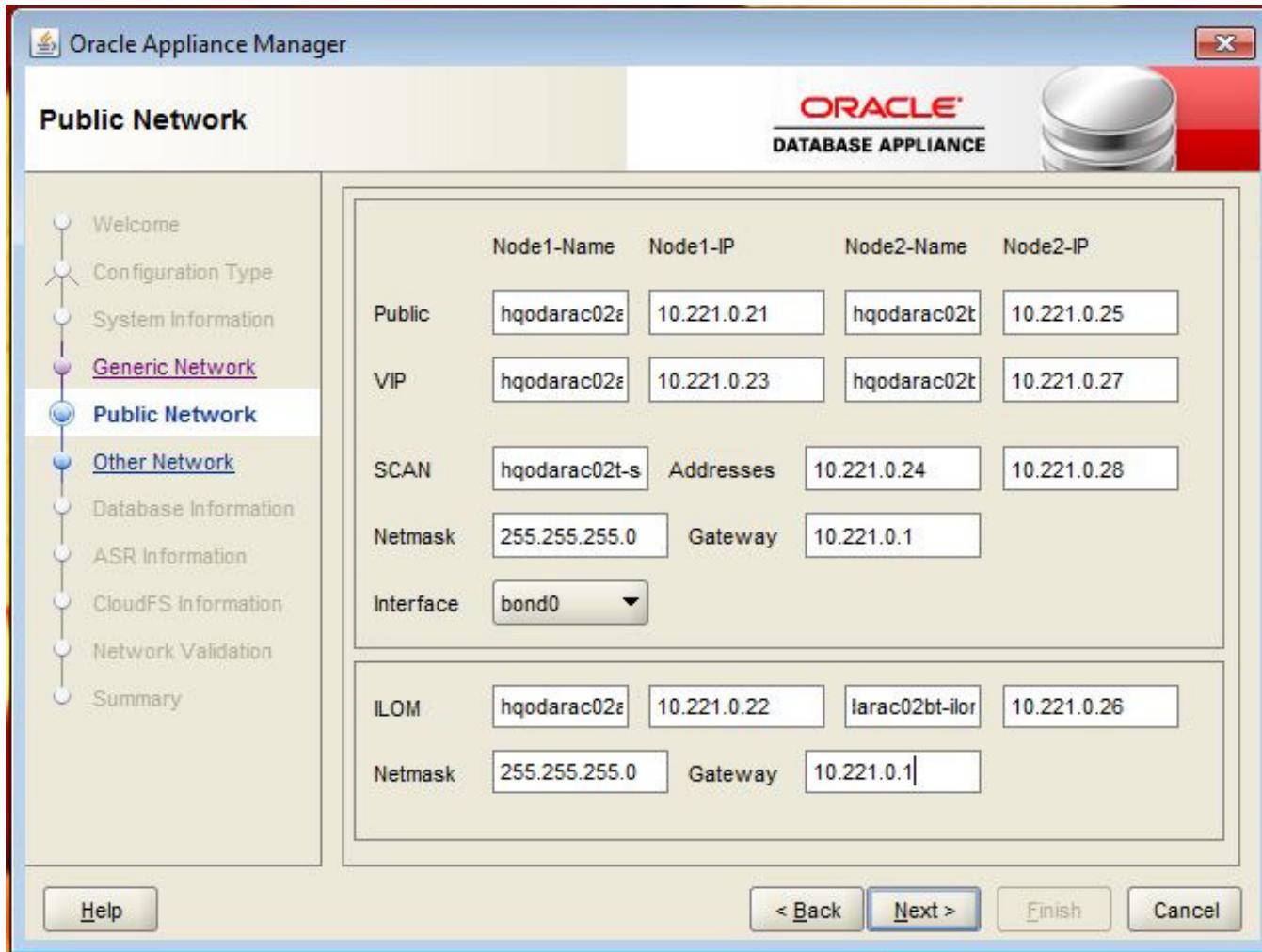
Install Screens: 3



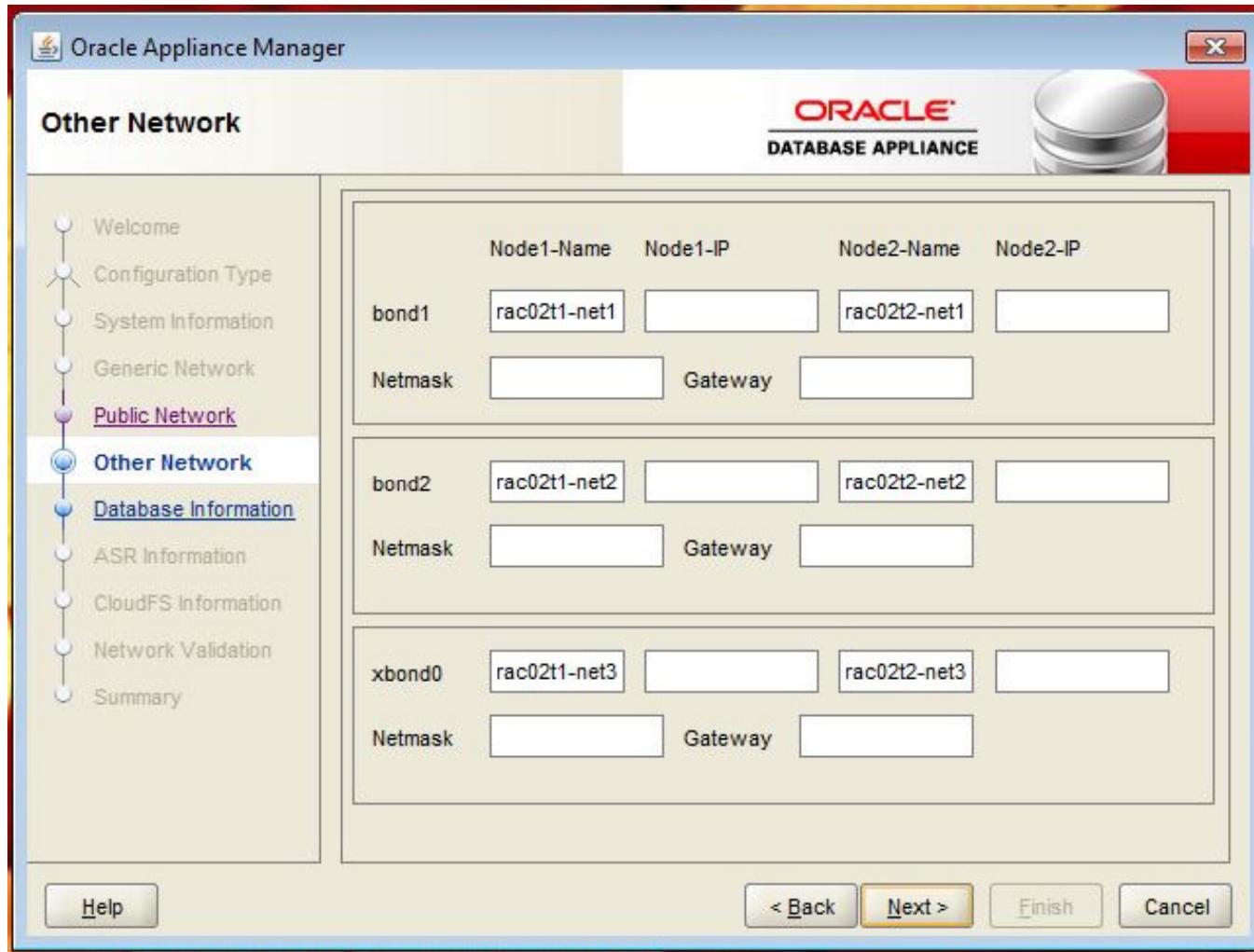
Install Screens: 4



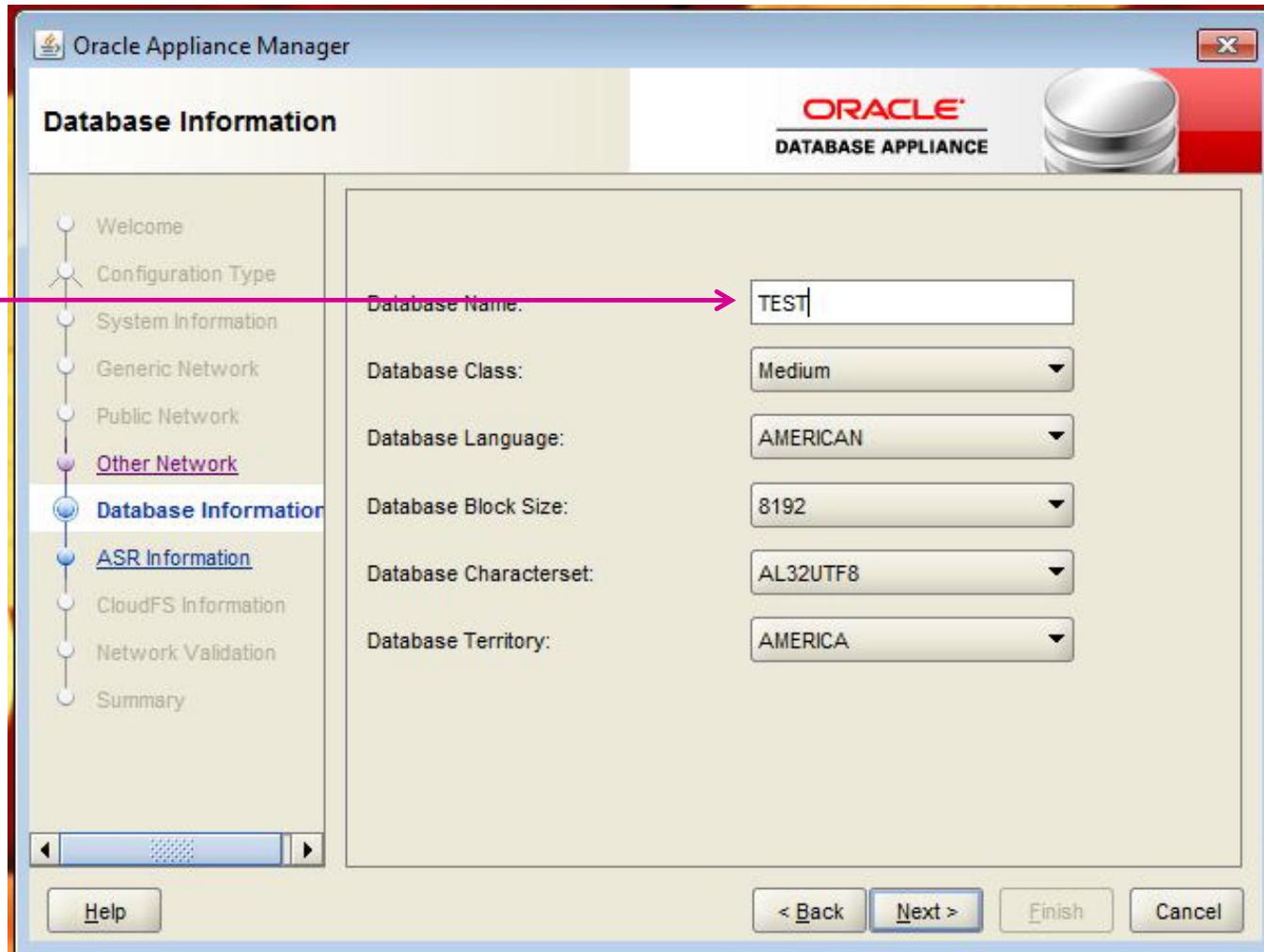
Install Screens: 5



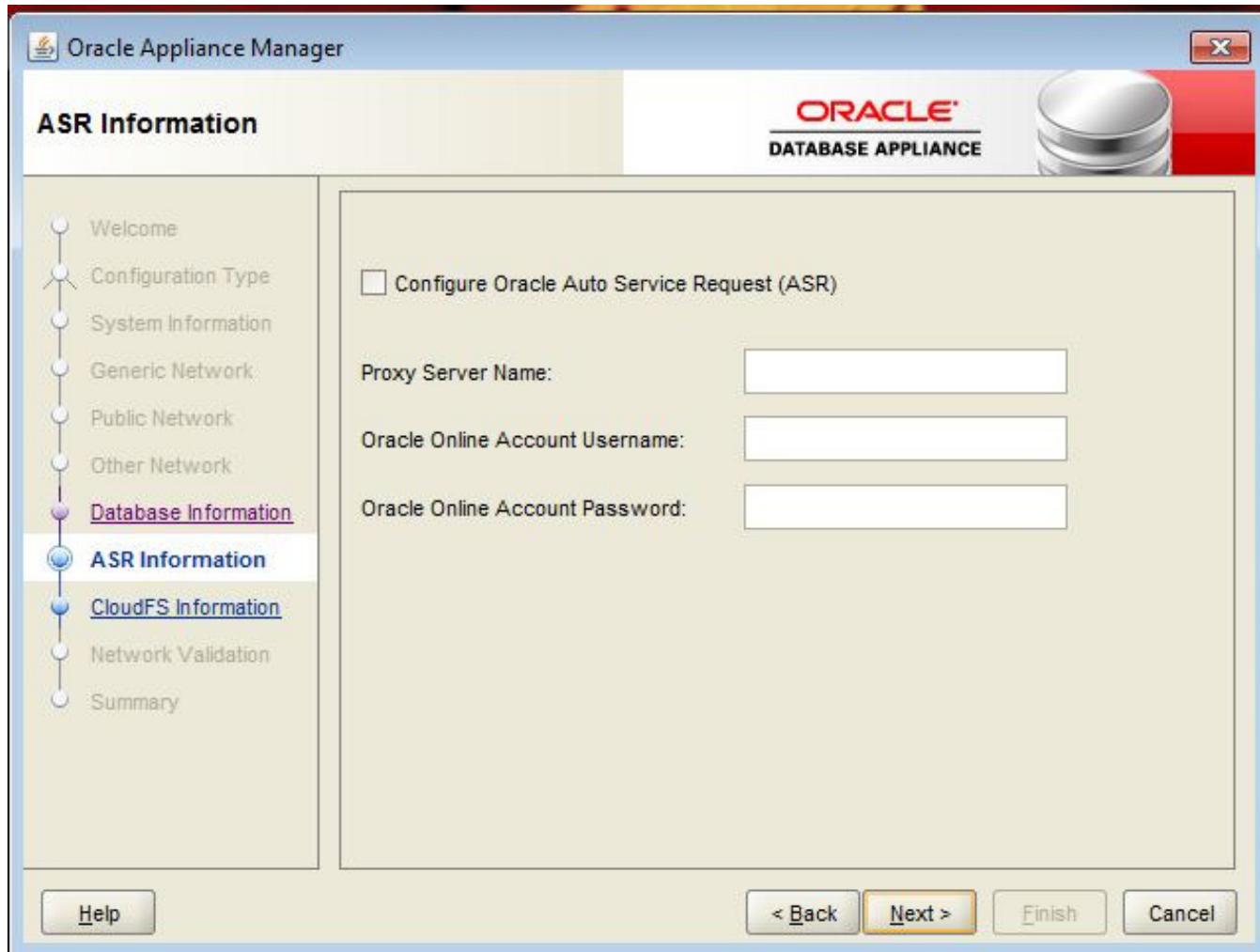
Install Screens: 6



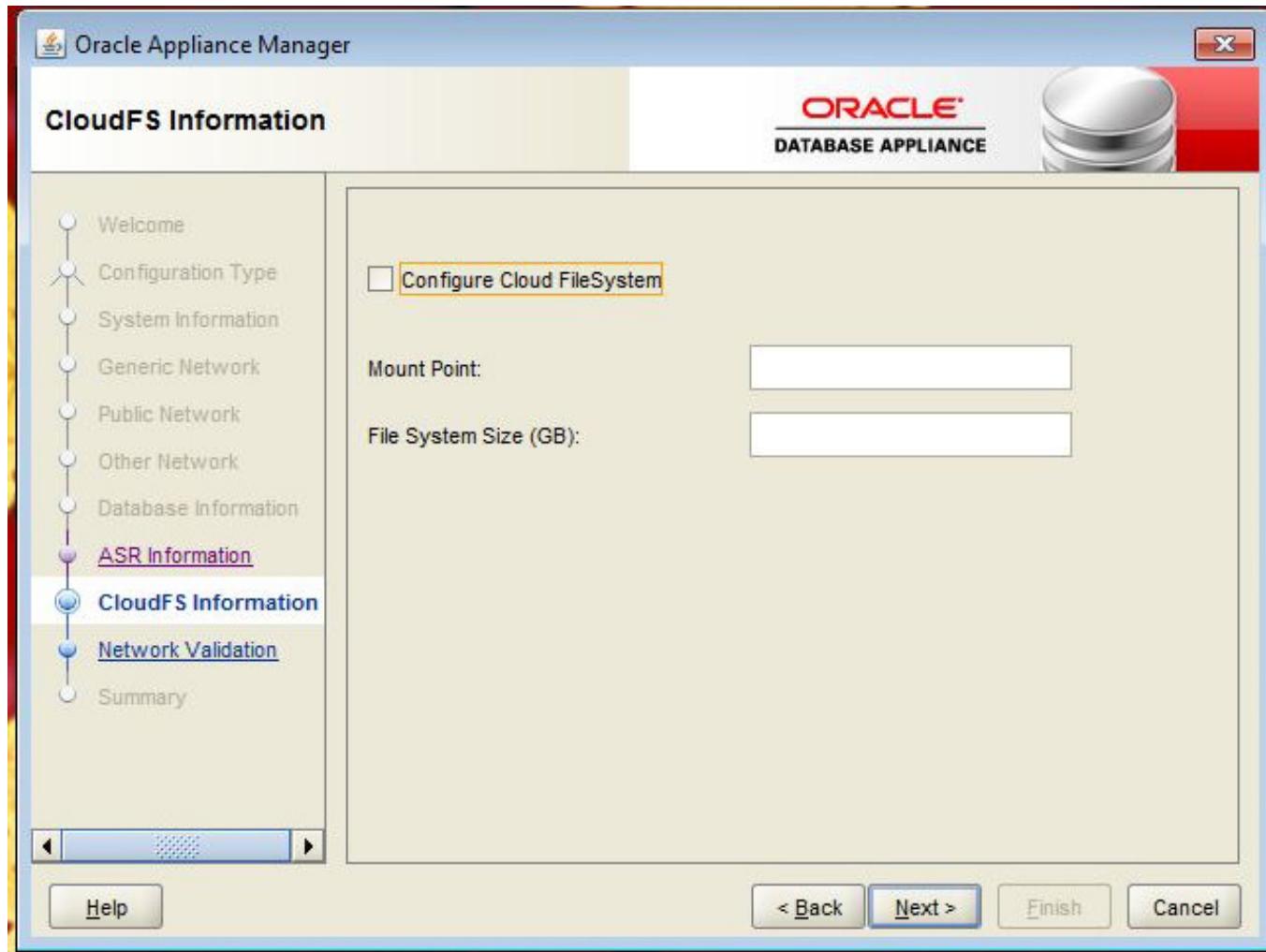
Install Screens: 7



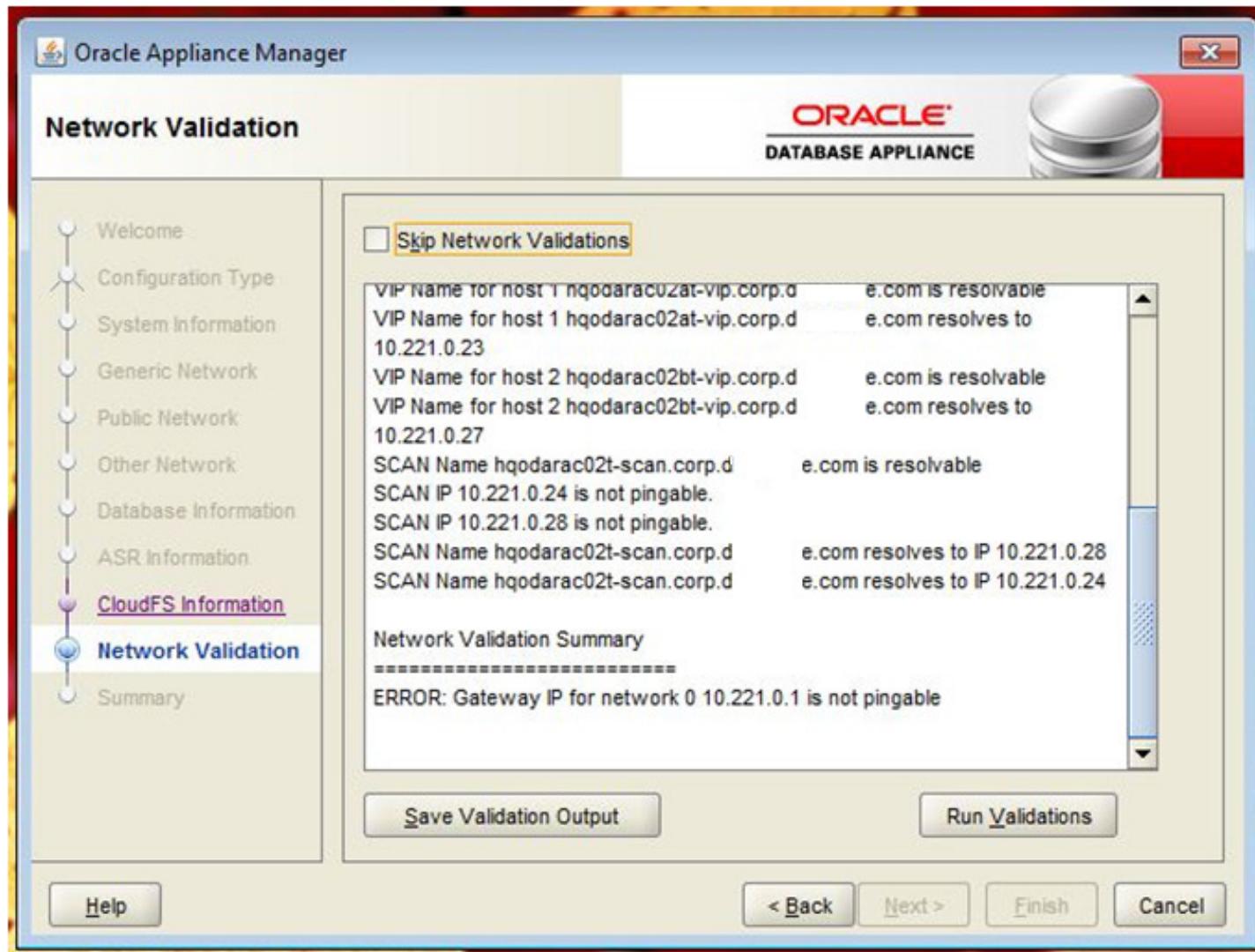
Install Screens: 8



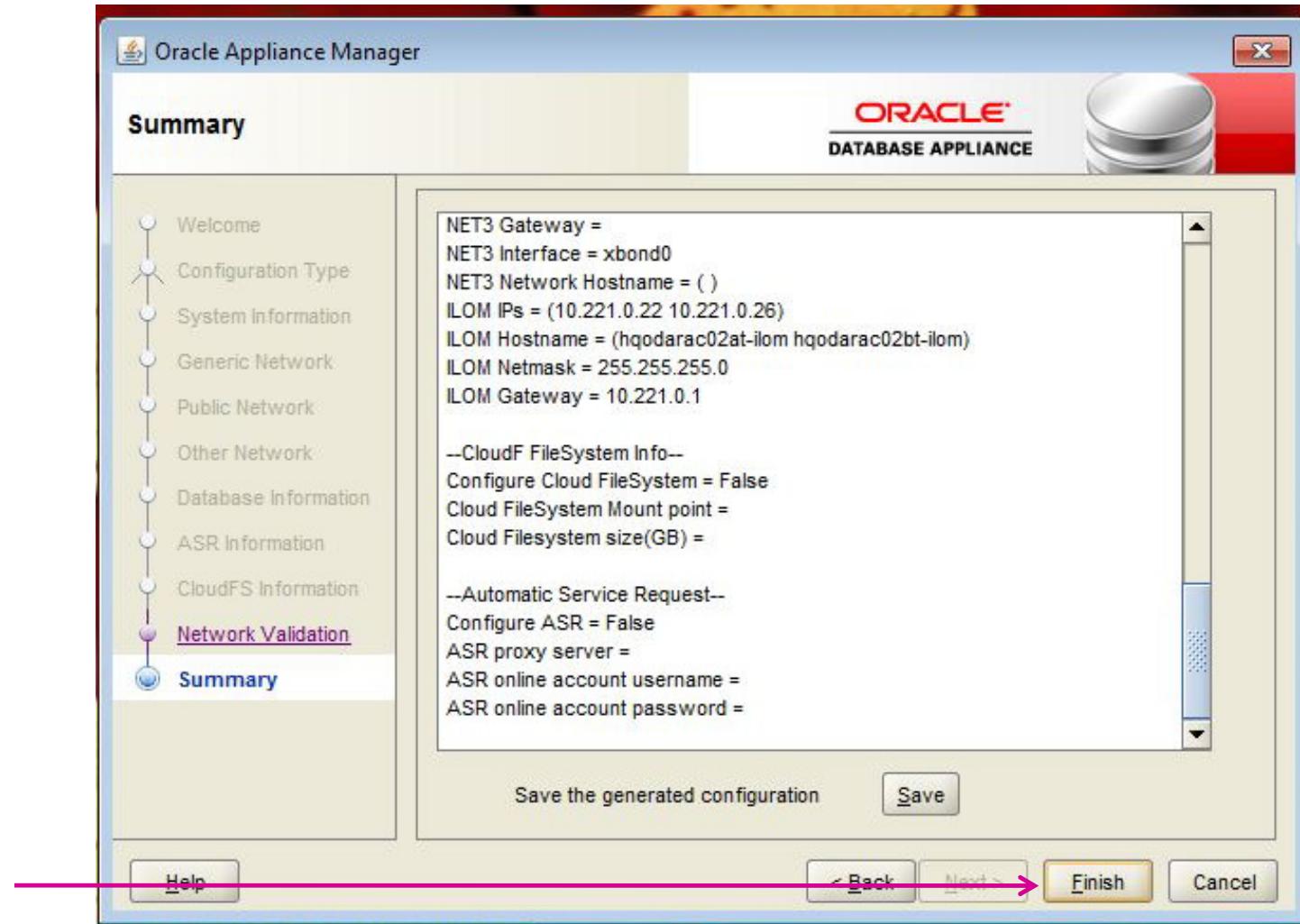
Install Screens: 9



Install Screens: 10



Install Screens: 11



Value Adds

Daniel Morgan | damorgan12c@gmail.com | morganslibrary.org

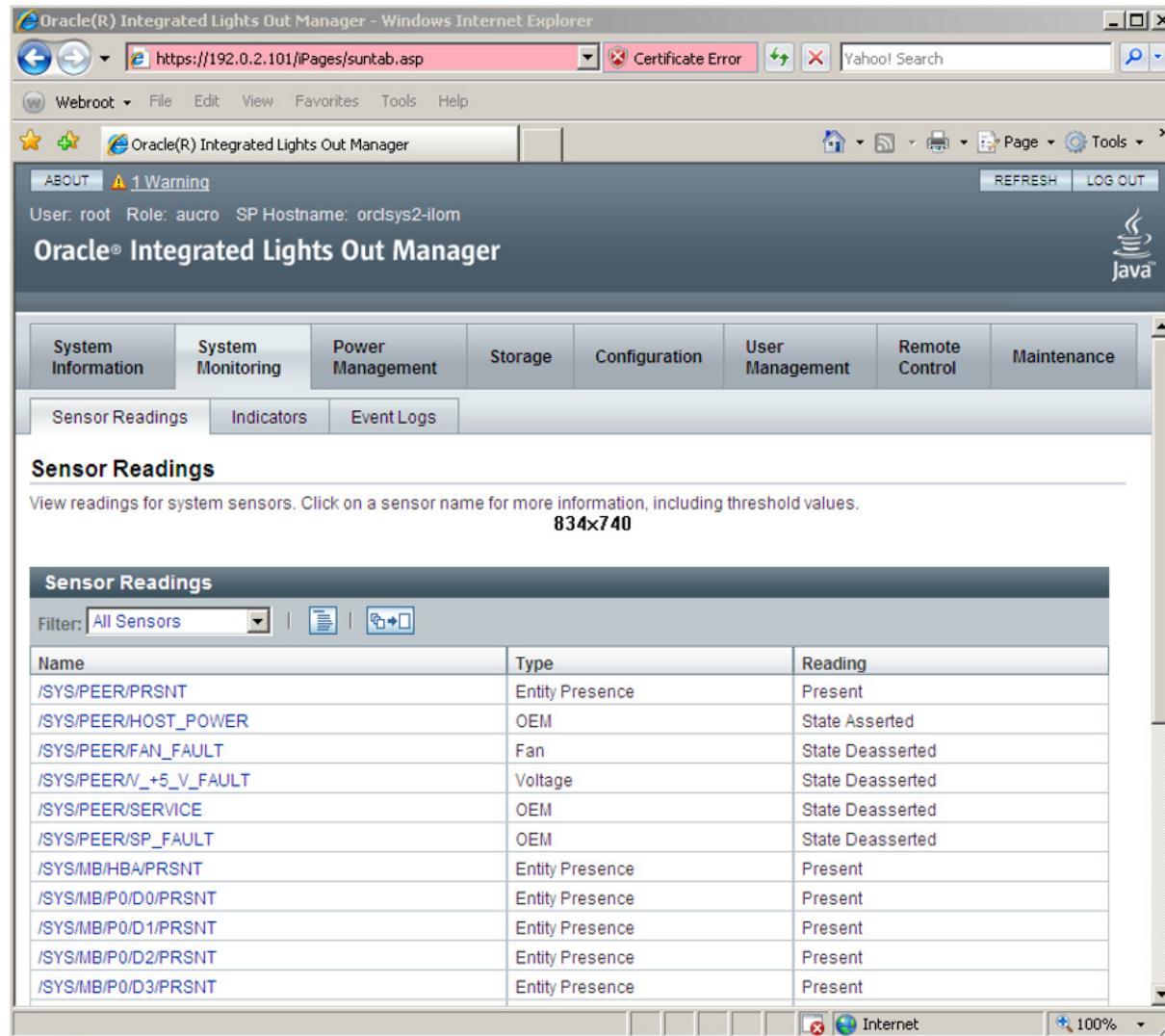
Integrating ODAs with ZFS to Create a Secure Scalable HA Environment

Presented: Rocky Mountain Oracle User Group - 13 January, 2013

ILOM: System Information: Overview

Oracle(R) Integrated Lights Out Manager - Windows Internet Explorer
https://192.0.2.101/iPages/suntab.asp Certificate Error Yahoo! Search
Webroot File Edit View Favorites Tools Help
Oracle(R) Integrated Lights Out Manager REFRESH LOG OUT
ABOUT 1 Warning User: root Role: aucro SP Hostname: orclsys2-ilom
Oracle® Integrated Lights Out Manager Java™
System Information System Monitoring Power Management Storage Configuration User Management Remote Control Maintenance
Overview Components Fault Management Identification Information Banner Messages Session Timeout Versions
System Overview
View system summary information. You may also change power state and view system status and fault information.
Product Name: SUN FIRE X4370 M2 SERVER
Part/Serial Number: 30102851+1+1 / 1146FMW00R
Host Power: **On** Change...
System Status: Normal View...
BIOS Version: 12010304
SP Hostname: orclsys2-ilom
Uptime: 0 days, 00:40:38
IP Address: 192.0.2.101
ILOM Version: v3.0.14.13.a r70764

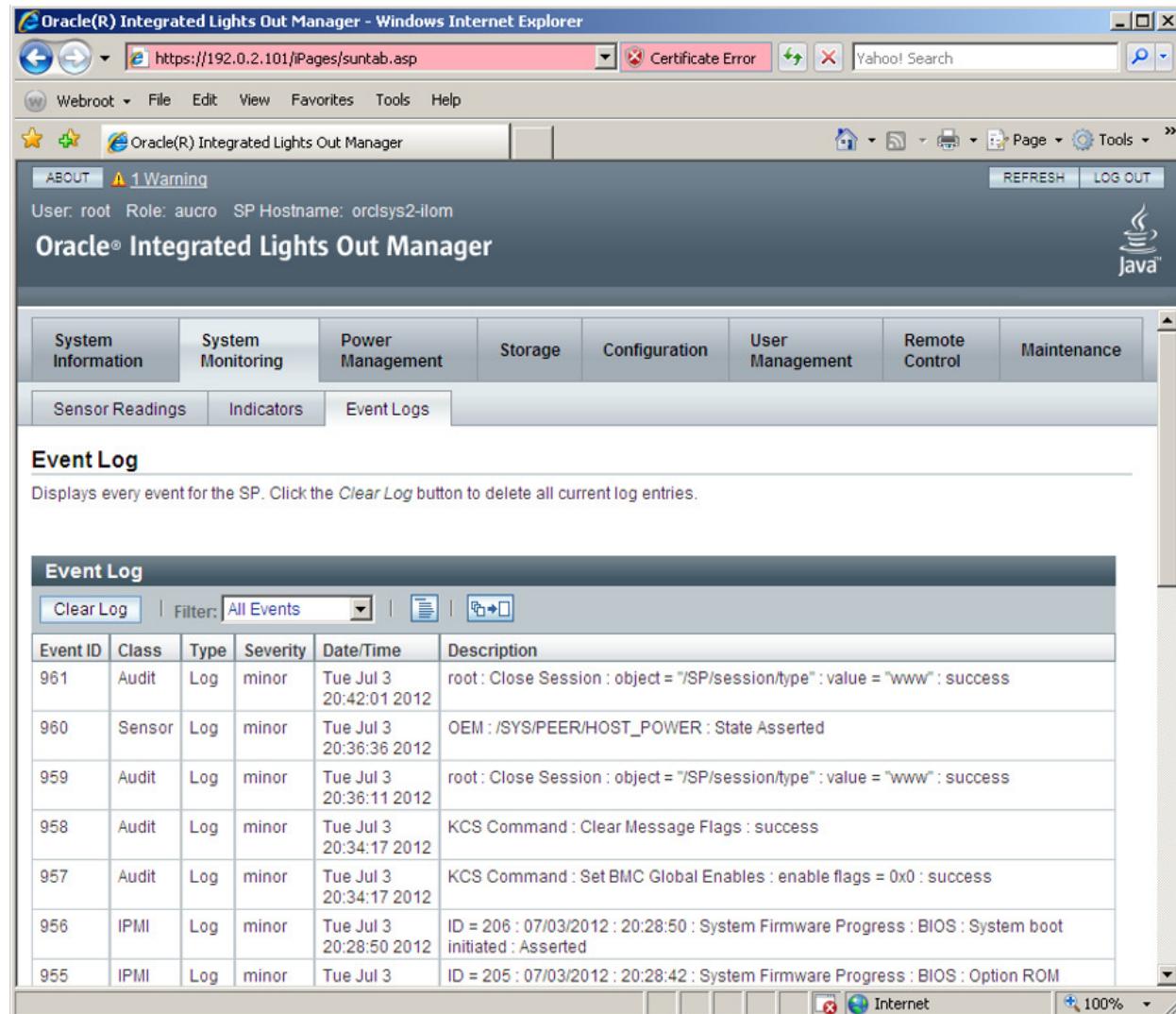
ILOM: System Monitoring: Sensor Readings



The screenshot shows the Oracle Integrated Lights Out Manager (ILOM) web interface. The browser title is "Oracle(R) Integrated Lights Out Manager - Windows Internet Explorer". The URL is "https://192.0.2.101/iPages/suntab.asp". A "Certificate Error" message is displayed in the address bar. The page header shows "User: root Role: auro SP Hostname: orclsys2-ilom". The main menu includes "ABOUT", "1 Warning", "REFRESH", and "LOG OUT". The navigation bar has tabs for "System Information", "System Monitoring", "Power Management", "Storage", "Configuration", "User Management", "Remote Control", and "Maintenance". The "System Monitoring" tab is selected. Below it, sub-tabs are "Sensor Readings", "Indicators", and "Event Logs", with "Sensor Readings" being the active tab. The main content area is titled "Sensor Readings" and displays a table of system sensor data. The table has columns for "Name", "Type", and "Reading". The data is as follows:

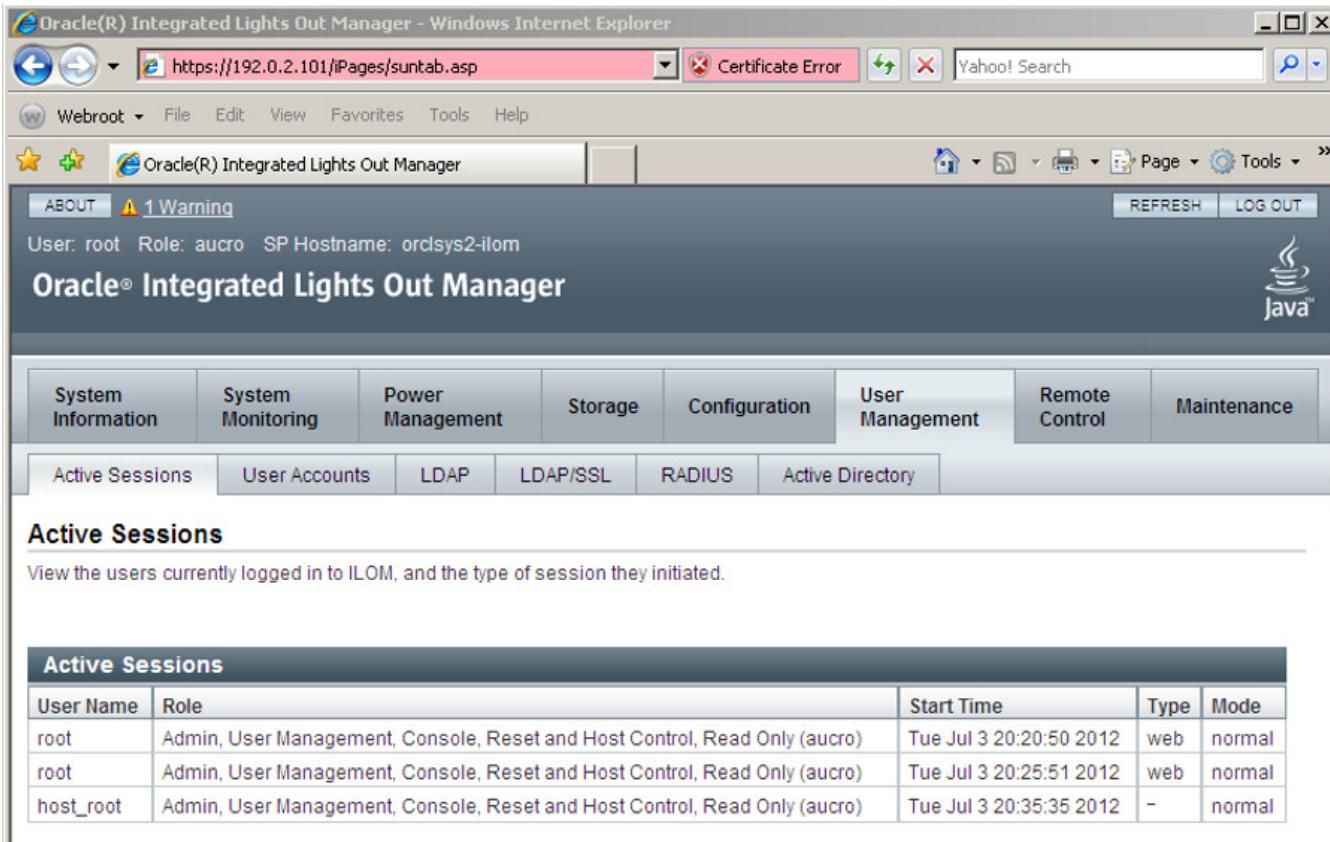
Name	Type	Reading
/SYS/PEER/PRSN	Entity Presence	Present
/SYS/PEER/HOST_POWER	OEM	State Asserted
/SYS/PEER/FAN_FAULT	Fan	State Deasserted
/SYS/PEER/V_+5_V_FAULT	Voltage	State Deasserted
/SYS/PEER/SERVICE	OEM	State Deasserted
/SYS/PEER/SP_FAULT	OEM	State Deasserted
/SYS/MB/HBA/PRSN	Entity Presence	Present
/SYS/MB/P0/D0/PRSN	Entity Presence	Present
/SYS/MB/P0/D1/PRSN	Entity Presence	Present
/SYS/MB/P0/D2/PRSN	Entity Presence	Present
/SYS/MB/P0/D3/PRSN	Entity Presence	Present

ILOM: System Monitoring: Event Logs



Event ID	Class	Type	Severity	Date/Time	Description
961	Audit	Log	minor	Tue Jul 3 20:42:01 2012	root: Close Session : object = "/SP/session/type" : value = "www" : success
960	Sensor	Log	minor	Tue Jul 3 20:36:36 2012	OEM : /SYS/PEER/HOST_POWER : State Asserted
959	Audit	Log	minor	Tue Jul 3 20:36:11 2012	root: Close Session : object = "/SP/session/type" : value = "www" : success
958	Audit	Log	minor	Tue Jul 3 20:34:17 2012	KCS Command : Clear Message Flags : success
957	Audit	Log	minor	Tue Jul 3 20:34:17 2012	KCS Command : Set BMC Global Enables : enable flags = 0x0 : success
956	IPMI	Log	minor	Tue Jul 3 20:28:50 2012	ID = 206 : 07/03/2012 : 20:28:50 : System Firmware Progress : BIOS : System boot initiated : Asserted
955	IPMI	Log	minor	Tue Jul 3	ID = 205 : 07/03/2012 : 20:28:42 : System Firmware Progress : BIOS : Option ROM

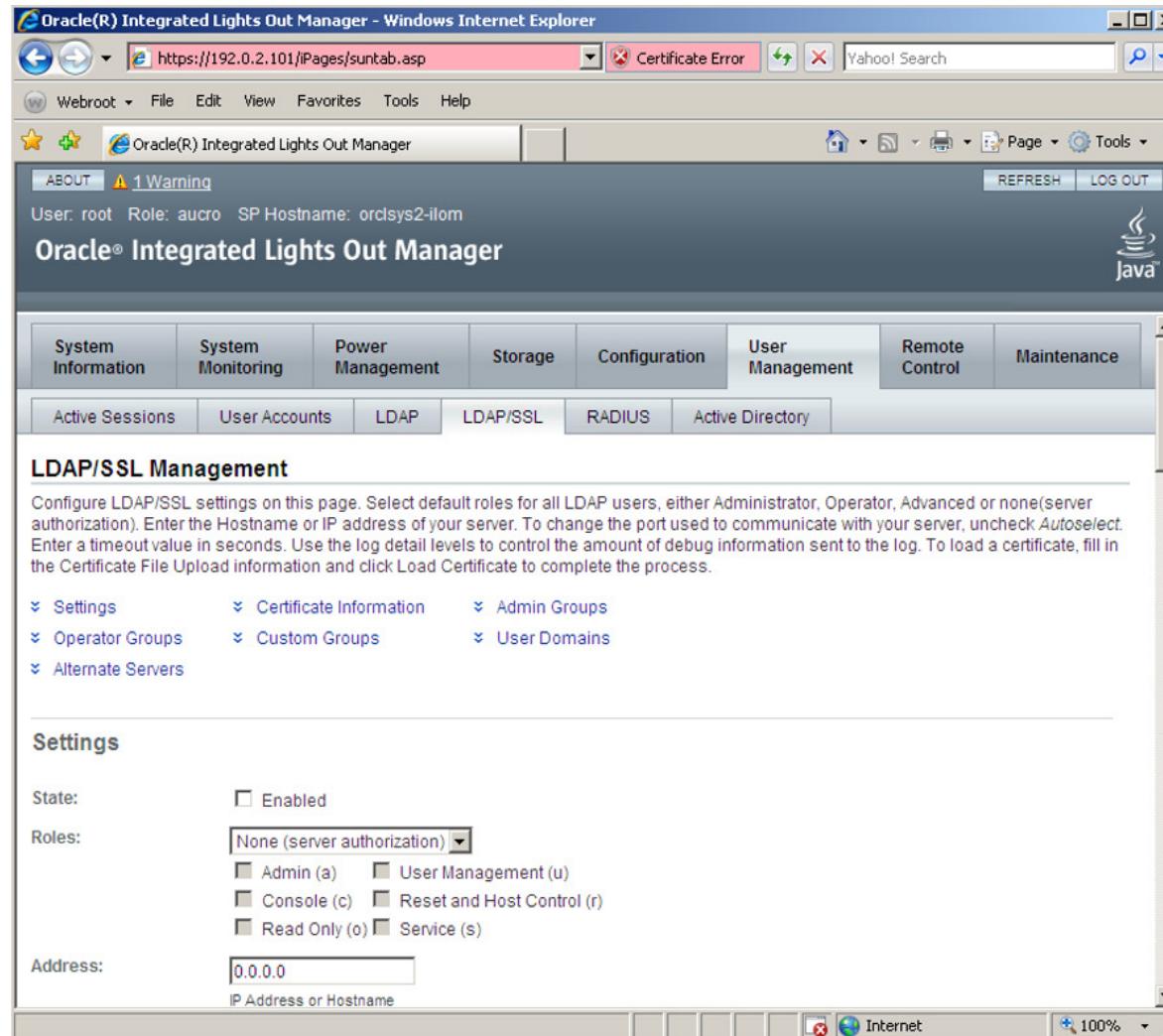
ILOM: User Management: Active Sessions



The screenshot shows a Windows Internet Explorer window displaying the Oracle Integrated Lights Out Manager (ILOM) interface. The URL in the address bar is <https://192.0.2.101/iPages/suntab.asp>. A 'Certificate Error' message is visible in the top right of the browser window. The ILOM interface has a dark blue header with the title 'Oracle® Integrated Lights Out Manager'. Below the header is a navigation menu with tabs: System Information, System Monitoring, Power Management, Storage, Configuration, User Management, Remote Control, and Maintenance. The 'User Management' tab is selected, and its sub-tabs are: Active Sessions, User Accounts, LDAP, LDAP/SSL, RADIUS, and Active Directory. The 'Active Sessions' sub-tab is selected. The main content area is titled 'Active Sessions' and contains the following text: 'View the users currently logged in to ILOM, and the type of session they initiated.' Below this text is a table titled 'Active Sessions' with the following data:

User Name	Role	Start Time	Type	Mode
root	Admin, User Management, Console, Reset and Host Control, Read Only (aucro)	Tue Jul 3 20:20:50 2012	web	normal
root	Admin, User Management, Console, Reset and Host Control, Read Only (aucro)	Tue Jul 3 20:25:51 2012	web	normal
host_root	Admin, User Management, Console, Reset and Host Control, Read Only (aucro)	Tue Jul 3 20:35:35 2012	-	normal

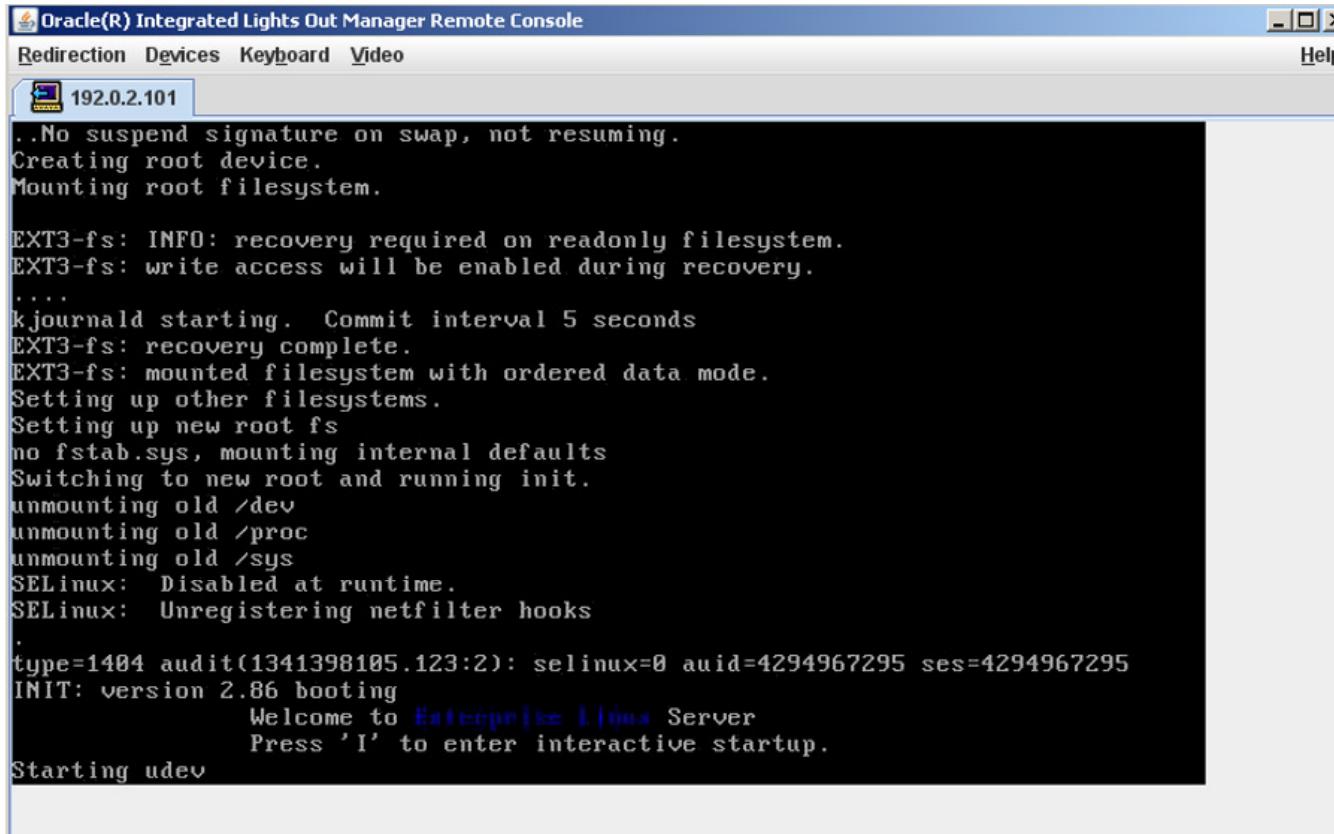
ILOM: User Management: LDAP / SSL



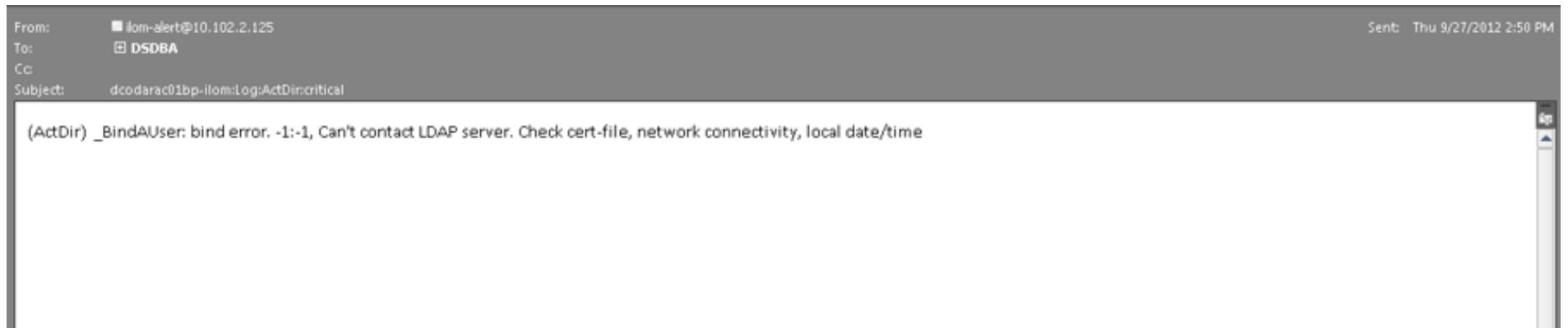
The screenshot shows the Oracle Integrated Lights Out Manager (ILOM) web interface. The URL in the browser is <https://192.0.2.101/iPages/suntab.asp>. The page title is "Oracle(R) Integrated Lights Out Manager". The top menu bar includes "Webroot", "File", "Edit", "View", "Favorites", "Tools", and "Help". The toolbar includes "Home", "Back", "Forward", "Stop", "Search", and "Tools". The main header displays "User: root Role: auco SP Hostname: orclsys2-ilom" and the "Oracle® Integrated Lights Out Manager" logo. The navigation menu at the top includes "ABOUT", "1 Warning", "REFRESH", and "LOG OUT". The main menu tabs are "System Information", "System Monitoring", "Power Management", "Storage", "Configuration", "User Management", "Remote Control", and "Maintenance". The "User Management" tab is selected. Below the tabs, sub-menu items are shown: "Active Sessions", "User Accounts", "LDAP", "LDAP/SSL", "RADIUS", and "Active Directory". The "LDAP/SSL Management" section is currently active. It contains a configuration form with the following fields:

- Settings**:
 - State: Enabled
 - Roles:
 - Admin (a) User Management (u)
 - Console (c) Reset and Host Control (r)
 - Read Only (o) Service (s)
- Address:** IP Address or Hostname

Remote Control: Host Control: Remote Console



ILOM Warning Message



Full Support for High Availability

- RAC
- Data Guard
- RMAN
- Streams
- OEM Cloud Control 12c
 - Diagnostic Pack
 - Tuning Pack
 - Data Masking Pack

One Button Solutions

- Fully scripted, one button, solutions, for
 - Build
 - Secure
 - Migration
 - Bare Metal Restore
 - Data Guard
 - GoldenGate
 - RMAN Backup

Support Center

The screenshot shows a Firefox browser window displaying the Oracle Support Center. The URL is https://support.html.oracle.com/epmos/faces/ui/km/SearchDocDisplay.jspx?_afrLoop=6463580628317000&type=DOCUMENT&id=1449552.2&displayInd=1. The page title is "Information Center: Oracle Database Appliance [ID 1449552.2]".

The left sidebar shows a search for "Search : ODA STIG" and a list of results:

- 1 - 2
- [Back to Results](#)
- [Place holder for STIG Implementation Script for Oracle Database Appliance](#)
- [Information Center: Oracle Database Appliance](#)

The main content area is titled "Information Center: Oracle Database Appliance [ID 1449552.2]" and was last refreshed on 27 Jun 2012. It contains the following sections:

- Information Centers**: Overview, Hot Topics, Resources, Sun System Handbook, Hardware Compatibility Lists.
- Refine Search By Task**: Use Product, Troubleshoot, Patching And Maintenance, Install And Configure, Upgrade, Optimize Performance.
- Alerts**: View the most up-to-date high impact and urgent issues for your product. One alert is listed: "ALERT - ODA (Oracle Database Appliance) Mandatory OAK Patch 2.1.0.3.1 [Document 1452085.1 Updated: 04/23/2012]".
- News & Announcements**: Read recently published news and announcements about your product. No results found.
- New Knowledge Documents**: Read recently published documents about your product. Two documents are listed: "Oracle Database Appliance FAQ [Document 1463638.1 Updated: 06/05/2012]" and "How to use a separate Linux machine as your RPM repository to download and apply RPM's [Document 1455719.1 Updated: 05/04/2012]".
- New Troubleshooting and Problem-Solution Documents**: Read recently published Troubleshooting and Problem-Solution documents about your product. Two documents are listed: "NEW ODA SETUP: FAILS WITH DOMAINNAME OF "EXAMPLE", NOT EXAMPLE.COM [Document 1455719.1 Updated: 05/04/2012]" and "DBUA (DataBase Upgrade Assistant) failing with "Cannot find the ORACLE_HOME for the database" on ODA (Oracle Database Appliance) [Document 1455719.1 Updated: 05/04/2012]".

STIG Download

★ Oracle Database Appliance DoD C&A STIG [ID 1456609.1]

To Bottom

Modified: Jul 18, 2012 Type: README Status: PUBLISHED Priority: 3

Comments (0)



In this Document

[Main Content](#)

[References](#)

Applies to:

Oracle Database Appliance
Generic Linux

Main Content

The Department of Defense(DoD) DISA Information Assurance Process includes Certification and Accreditation(C&A) including the Security Technical Implementation Guides(STIGs). These are guidelines and scripts that are run to advise on securing and locking down database, operating system, application servers, and other system components.

Currently, DoD customers are running various Oracle products that go through the DoD C&A process including the STIG process. General STIG Information is available at: - <http://iase.disa.mil/stigs/>

The Oracle Database Appliance(ODA) is a fully integrated system of software, servers, storage, and networking in a single box that delivers high-availability database services. Oracle engineered Oracle Database Appliance for simplicity. Accordingly, Oracle aims to provide a more simplified configuration and patching process.

STIG Script

- STIG Script Syntax
 - The script logs its actions in the "/opt/oracle/oak/log//hostname/stig/" directory
 - **-check** checks the system for guideline violations
 - **-force** re-runs the script even if there are no violations
 - **-fix** used to implement guideline recommendations
 - lock and unlock options can be used to enable or disable direct ssh logging as root. Direct ssh login as root is required for Patching and therefore before patching, the unlock needs to be executed.

Sample usage

```
# ./stig.sh -fix
```

STIG Script Output: Category 1

```
2012-06-28 01:18:12 : Running stig script version: '1.0'
2012-06-28 01:18:12 : Executing script : ./stig.sh -check
2012-06-28 01:18:12 : Checking for stig violations on system 'orclsys1'

2012-06-28 01:18:12 : List of Category-1 stig violation found by script
2012-06-28 01:18:12 : [STIG ID : LNX00140] : [CHECK] : Password for grub not enabled : FOUND
2012-06-28 01:18:12 : [STIG ID : GEN004640] : [CHECK] : sendmail decode command is not commented in /etc/aliases : FOUND
2012-06-28 01:18:12 : [STIG ID : LNX00320] : [CHECK] : Privilege account 'shutdown' is present : FOUND
2012-06-28 01:18:12 : [STIG ID : LNX00320] : [CHECK] : Privilege account 'halt' is present : FOUND
2012-06-28 01:18:12 : [STIG ID : LNX00580] : [CHECK] : Ctrl-Alt-Del combination to shutdown system is enabled : FOUND
2012-06-28 01:18:13 : [STIG ID : 2006-T-0013] : [CHECK] : RealVNC rpm is installed on system : FOUND
2012-06-28 01:18:13 : [STIG ID : LNX00040] : [CHECK] : Support for usb device found in kernel : FOUND

2012-06-28 01:18:13 : List of Category-2 stig violation found by script
2012-06-28 01:18:13 : [STIG ID : GEN000020] : [CHECK] : Single user mode boot is enabled without a password : FOUND
2012-06-28 01:18:13 : [STIG ID : GEN000340] : [CHECK] : Non privileged account oprofile found on system : FOUND
2012-06-28 01:18:13 : [STIG ID : GEN000340] : [CHECK] : Non privileged account avahi-autoipd found on system : FOUND
2012-06-28 01:18:13 : [STIG ID : GEN000460] : [CHECK] : pam_tally not used to lock account after 3 consecutive failed logins : FOUND
2012-06-28 01:18:13 : [STIG ID : GEN000800] : [CHECK] : remember not used in PAM configuration files : FOUND
2012-06-28 01:18:13 : [STIG ID : GEN000600] : [CHECK] : Force of at least one lower case character is not set for password : FOUND
2012-06-28 01:18:13 : [STIG ID : GEN000600] : [CHECK] : Force of at least one upper case character is not set for password : FOUND
2012-06-28 01:18:13 : [STIG ID : GEN000620] : [CHECK] : Force of at least one numeric character is not set for password : FOUND
2012-06-28 01:18:13 : [STIG ID : GEN000640] : [CHECK] : Force of at least one special character is not set for password : FOUND
2012-06-28 01:18:13 : [STIG ID : GEN000480] : [CHECK] : Login delay is not enabled in /etc/pam.d/system-auth : FOUND
2012-06-28 01:18:13 : [STIG ID : GEN000700] : [CHECK] : Maximum age for a password change is more than 60 days : FOUND
2012-06-28 01:18:13 : [STIG ID : GEN000540] : [CHECK] : Password can be changed more than once in 24 hours : FOUND
2012-06-28 01:18:13 : [STIG ID : GEN000580] : [CHECK] : Password length is less than 8 characters : FOUND
2012-06-28 01:18:13 : [STIG ID : GEN001120] : [CHECK] : Direct login as root is enabled from ssh : FOUND
2012-06-28 01:18:13 : [STIG ID : GEN002100] : [CHECK] : ekshell supported by the pam.rhost : FOUND
2012-06-28 01:18:13 : [STIG ID : GEN002960] : [CHECK] : Access to cron is not through cron.allow and cron.deny : FOUND
2012-06-28 01:18:13 : [STIG ID : GEN003080] : [CHECK] : Permission of file /etc/crontab is more permissive than octal 600 : FOUND
2012-06-28 01:18:13 : [STIG ID : GEN003200] : [CHECK] : Permission of file /etc/cron.deny is more permissive than octal 600 : FOUND
2012-06-28 01:18:13 : [STIG ID : GEN005400] : [CHECK] : Permission of file /etc/syslog.conf is more permissive than octal 640 : FOUND
2012-06-28 01:18:13 : [STIG ID : GEN000920] : [CHECK] : Permission of directory /root is more permissive than octal 700 : FOUND
2012-06-28 01:18:13 : [STIG ID : GEN003865] : [CHECK] : tcpdump rpm is installed on system : FOUND
2012-06-28 01:18:13 : [STIG ID : GEN004000] : [CHECK] : Permission of file /bin/traceroute is more permissive than octal 700 : FOUND
2012-06-28 01:18:13 : [STIG ID : LNX00340] : [CHECK] : Unnecessary account ftp found on system : FOUND

2012-06-28 01:18:35 : List of Category-3 stig violation found by script
2012-06-28 01:18:35 : [STIG ID : GEN004560] : [CHECK] : sendmail version is not hidden. : FOUND
```

However

However

- We MAY want to preserve the 6TB ASM disk for data
- We may want more storage for
 - FRA, Flashback DB files, RMAN files ...
 - Clone
 - Data Masking
 - Real Application Testing
 - Staging
 - Logs
 - And so on

Data Masking Pack

- Offers the ability to mask regulated or confidential data on test and development systems
- Mask format libraries
- Mask definitions
- Masking techniques
 - Condition-based masking
 - Compound masking
 - Deterministic masking
- Application masking templates import or export
- Mask format library import or export
- Masking script generation
- Clone and Mask workflow

Source: Linux Today: 2009

Daniel Morgan | damorgan12c@gmail.com | morganslibrary.org

Integrating ODAs with ZFS to Create a Secure Scalable HA Environment

Presented: Rocky Mountain Oracle User Group - 13 January, 2013

Choices

- ASM
 - Raw devices
- Clustered Storage
 - Which one? OCFS2, VxFS, ...
- Non-Clustered Storage
 - Non-blocking visibility on both nodes
 - dNFS, CIFS ...

ASM?

- Excellent decision for database storage
- Perhaps not optimal as a file system
 - ACFS?
- Requires raw disk to be presented to ODA
- Traditional HBA discussion

Clustered File System?

- Several CFS available for Linux
 - Need expertise
 - Wire it yourself
 - Tech concerns
 - File sizes
 - File counts
- Still traditional HBA discussion

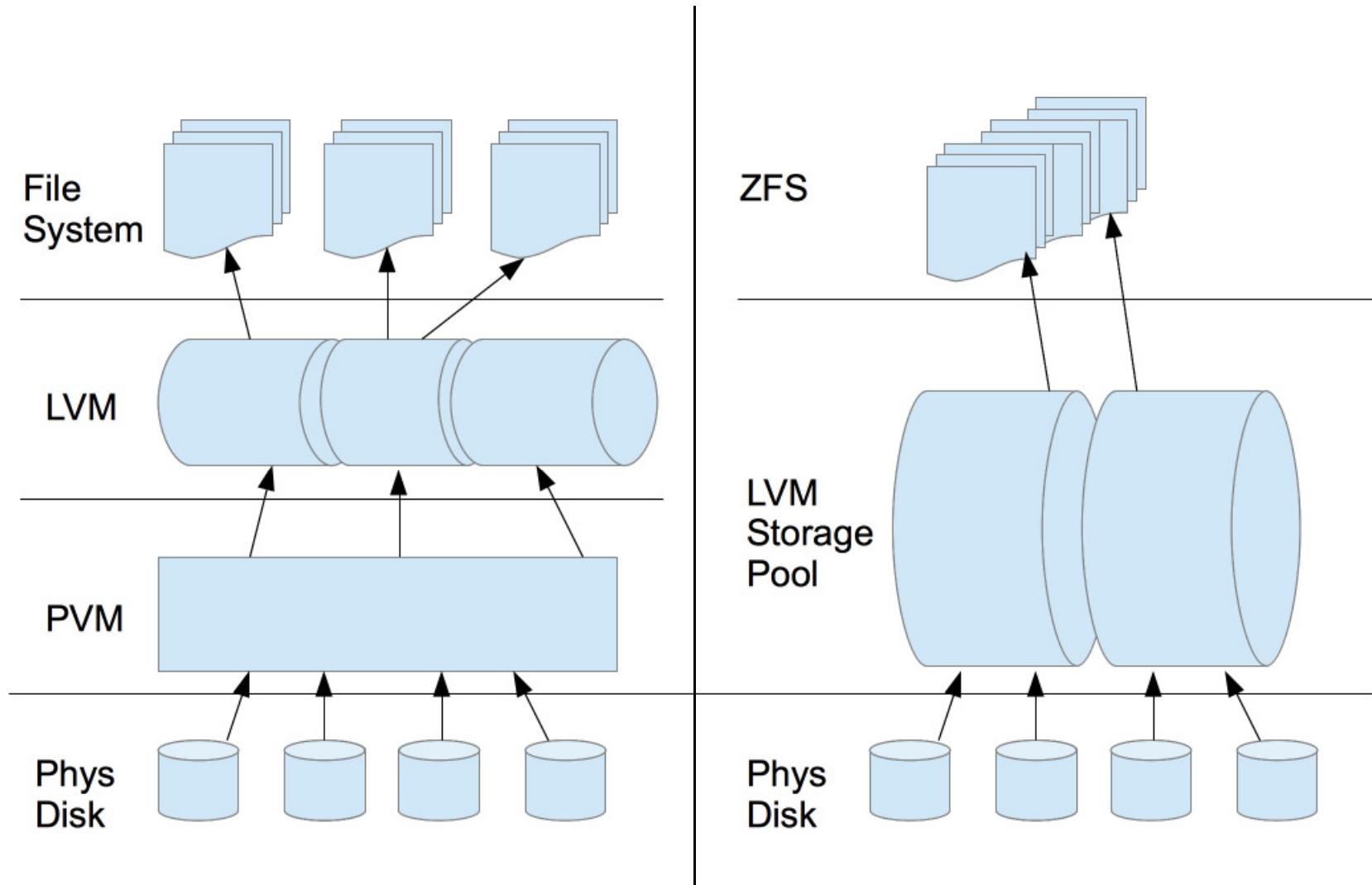
Non-clustered File System?

- Local File System
 - May be suitable for some applications,
 - But we have two separate hosts in ODA
 - Standard Linux-oriented
 - Still traditional HBA discussion
- [d]NFS
 - Vendor: NetApp, Oracle ZFS Appliance
 - OpenFiler?

Additional concern – silent corruption

- An undetected or uncorrectable error can occur on average once every 10-20 TB of data storage OR transfer
 - In modern systems that could mean a corruption in as little as 15 minutes
- ZFS was designed to combat this challenge
 - Checksum on all blocks
 - Copy on Write (preserve original block, not write in place)
 - Hot spares in pool
 - Auto-healing from ZFS mirror
 - Scrub instead of fsck
 - Monthly (or weekly for consumer disks)

Traditional File System stack vs ZFS



Quick Notes

- RAID
 - ZFS cannot fully protect the user's data when using a hardware RAID controller, as it is not able to perform the automatic self-healing unless it controls the redundancy of the disks and data.
 - Instead, ZFS provides its own RAID counterparts within the Storage Pool
- ZFS provides a hot-spare storage pool manager and a 128-bit, Copy on Write File System
- Capacity
 - Single file: 16 exabytes
 - Files in a pool: 264
 - Disks in a pool: 264
 - Pools in a system: 264

Where do you want to invest your time and treasure?

- Reinventing the wheel?
- Designing physical architecture?
- Applying one-off patches?
- Becoming Linux security experts?
- Writing shell scripts?

or would you rather be ...

- Managing your applications, users, and data?
- Optimizing your applications to maximize customer satisfaction?

ZFS Storage Appliance

Daniel Morgan | damorgan12c@gmail.com | morganslibrary.org

Integrating ODAs with ZFS to Create a Secure Scalable HA Environment

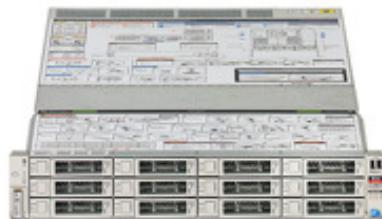
Presented: Rocky Mountain Oracle User Group - 13 January, 2013

ZFS Storage Appliance

- ZFS file system with advanced error detection and self-healing capabilities
- Integrated with Oracle Engineered Systems
- Both ZFS Deduplication and Compression or Hybrid Columnar Compression
- Hybrid Storage Pools
- Simultaneous multiprotocol support across multiple network interconnects, including GbE, 10 GbE, fibre channel and InfiniBand
- Integrated with OEM Grid Control
- Web-based storage management
- Integrated real-time storage analytics

What is a ZFS Appliance?

- Enterprise class Network Attached Storage (NAS)
- Choose the size that meets your needs
- Hybrid Columnar Compression (w/o an Exadata)
- Hybrid storage pools for DRAM and Flash caches
- DTrace storage analytics
- Use for
 - Backup and Restore
 - Cloning
 - Data Masking



ZFS Configurations

Sun ZFS Storage Appliance Configurations						
	Key Requirement	Maximum Storage Capacity	Space (Rack Units)	Write Optimized Flash	Read Optimized Flash	Cluster Option
Sun ZFS Storage 7120	Low-priced entry-level system with all software features	177 TB	2U/controller, 4U/disk shelf	73 GB	N	N
Sun ZFS Storage 7320	Entry-level cluster option for high availability	432 TB	1U/controller, 4U/disk shelf	Up to 1.2 TB	Up to 2 TB per controller	Y
Sun ZFS Storage 7420	Best price/performance	1.73 PB	3U/controller, 4U/disk shelf	Up to 7.0 TB	Up to 2 TB per controller	Y

ZFS Specifications

Sun ZFS Storage Appliance Specifications			
	Sun ZFS Storage 7120	Sun ZFS Storage 7320	Sun ZFS Storage 7420
Architecture			
Processor	1x 4-core 2.4 GHz Intel® Xeon® Processor	2x 4-core 2.4 GHz Intel® Xeon® Processor, per controller	4x 8-core 2.0 GHz or 10-core 2.4GHz Intel® Xeon® Processors per controller
Base Configurations			
Configuration options	<ul style="list-style-type: none">• 3.3 TB to 177 TB using either high-speed (15,000 RPM) or high-capacity (7,200 RPM) SAS-2 disks• Controller contains 11 HDDs and one SSD cache, supports up to two additional disk shelves with 24 disks each (300 GB, 600 GB, 2 TB, or 3 TB)	<ul style="list-style-type: none">• 6 TB to 432 TB using either high-speed (15,000 RPM) or high-capacity (7,200 RPM) SAS-2 disks• Supports up to six disk shelves with 20 or 24 disks each (300 GB, 600 GB, 2 TB, or 3 TB) and up to four optional write-optimized SSDs per shelf	<ul style="list-style-type: none">• 6 TB to 1.73 PB using either high-speed (15,000 RPM) or high-capacity (7,200 RPM) SAS-2 disks• Supports up to 24 disk shelves with 20 or 24 disks each (300 GB, 600 GB, 2 TB, or 3 TB) and up to four optional write-optimized SSDs per shelf

ZFS In The Data Center



Daniel Morgan | damorgan12c@gmail.com | morganslibrary.org

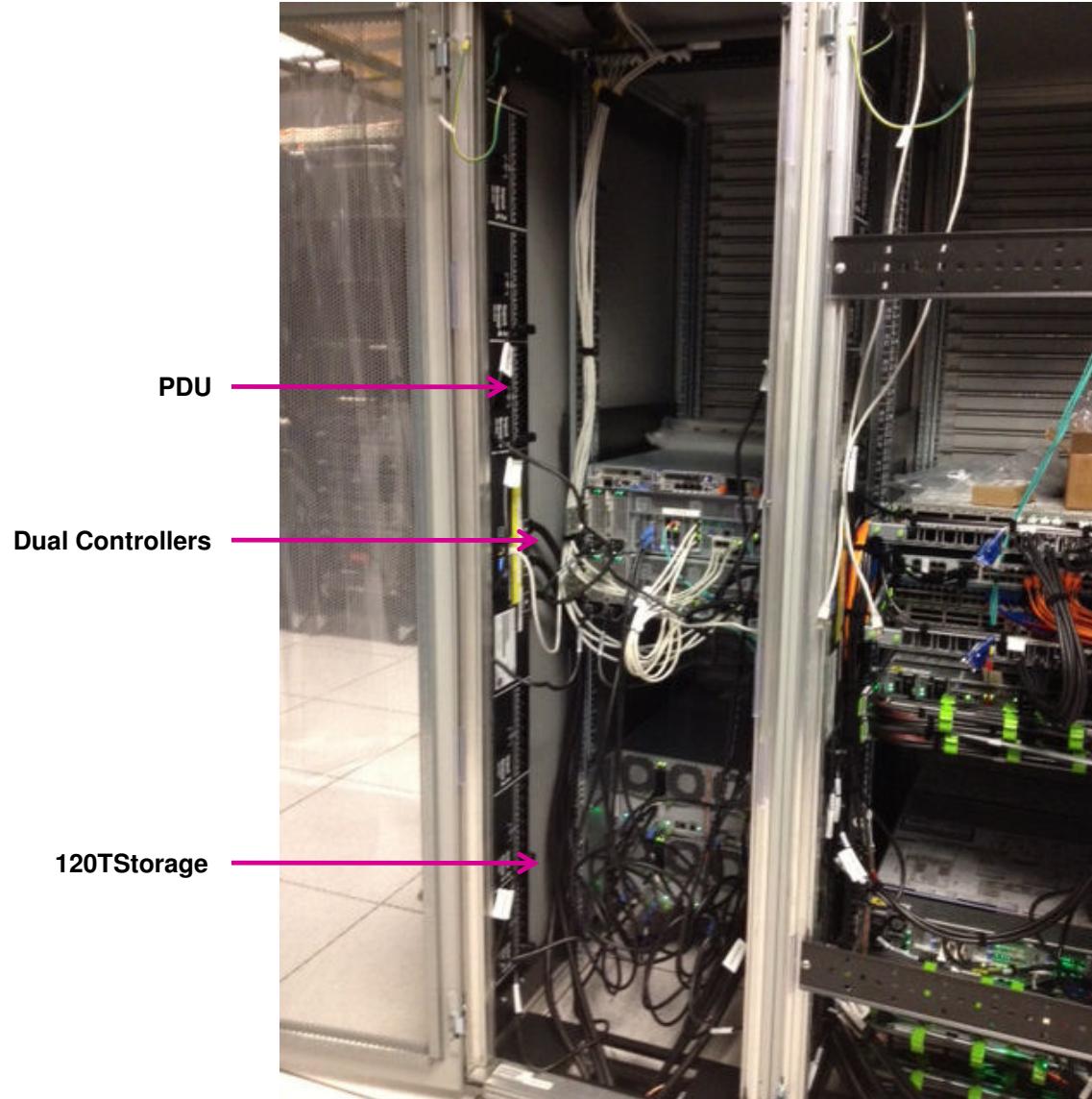
Integrating ODAs with ZFS to Create a Secure Scalable HA Environment

Presented: Rocky Mountain Oracle User Group - 13 January, 2013

ZFS 7420



ZFS Internals



ZFS BUI: Config Services

The screenshot shows the Sun ZFS Storage 7420 BUI interface. The top navigation bar includes the Sun Oracle logo, the text 'SUN ZFS STORAGE 7420', a warning message 'The cluster peer has rejoined the cluster.', and links for 'LOGOUT' and 'HELP'. Below the navigation is a main menu with tabs: Configuration (selected), Maintenance, Shares, Status, and Analytics. Under Configuration, there are sub-tabs: SERVICES, STORAGE, NETWORK, SAN, CLUSTER, USERS, PREFERENCES, and ALERTS. The 'Services' tab is selected, showing a list of services categorized into Data Services, Directory Services, System Settings, and Remote Access. Each service entry includes a status icon (green for online, grey for disabled), the last update time, and configuration icons for edit and delete.

Data Services			
● NFS	Online	2012-10-10 09:28:04	
● iSCSI	Disabled	2012-9-25 16:19:00	
● SMB	Disabled	2012-9-25 16:19:15	
● FTP	Disabled	2012-9-20 17:49:03	
● HTTP	Disabled	2012-9-20 17:49:03	
● NDMP	Disabled	2012-9-25 16:19:21	
● Remote Replication	Online	2012-9-20 17:49:50	
● Shadow Migration	Online	2012-9-20 17:49:50	
● SFTP	Online	2012-9-21 18:50:18	
● SRP	Disabled	2012-9-20 17:49:03	
● TFTP	Disabled	2012-9-20 17:49:54	
● Virus Scan	Disabled	2012-9-29 17:49:03	

Directory Services			
● NIS	Disabled	2012-9-25 16:12:36	
● LDAP	Online	2012-9-25 16:12:36	
● Active Directory	Online	2012-9-25 15:59:30	
● Identity Mapping	Online	2012-9-25 15:56:16	

System Settings			
● DNS	Online	2012-9-25 16:12:36	
● IPMP	Online	2012-9-20 17:49:51	
● NTP	Online	2012-9-24 14:23:46	
● Phone Home	Disabled	2012-9-20 17:49:50	
● Dynamic Routing	Online	2012-10-8 14:53:10	
● Service Tags	Online	2012-9-20 17:49:50	
● SMTP	Online	2012-9-25 16:12:36	
● SNMP	Online	2012-9-21 18:39:52	
● Syslog	Online	2012-9-21 18:20:47	
● System Identity	Online	2012-9-20 17:52:32	

Remote Access			
● SSH	Online	2012-9-20 17:52:33	

Daniel Morgan | damorgan12c@gmail.com | morganslibrary.org

Integrating ODAs with ZFS to Create a Secure Scalable HA Environment

Presented: Rocky Mountain Oracle User Group - 13 January, 2013

ZFS BUI: Config Storage

The screenshot shows the Sun ZFS Storage 7420 BUI interface. At the top, there is a message: "The cluster peer has rejoined the cluster." On the right, there are "LOGOUT", "HELP", and "Dismiss" buttons. Below the message, a confirmation message reads: "Confirm that all devices are present and minimally functional, and allocate them to a storage pool." There are "ABORT" and "COMMIT" buttons. The main area is titled "Verify and allocate devices" and includes a note: "Devices may be added on a per-device basis, however SATA devices in SAS-1 enclosures may be added in half- or whole-chassis units only. While affected devices may be allocated, they will not be available for use and cannot be added later without reconfiguring the pool; for best results, defer configuring storage until any problems can be repaired. Mixing devices of differing speeds within a storage pool is strongly discouraged." A table lists three devices: "c0zfs742001p" (Sun ZFS Storage 7420), "1235FMD003" (Sun Disk Shelf (SAS-2)), and "1235FMD002" (Sun Disk Shelf (SAS-2)). The "DATA" column for "1235FMD003" shows a dropdown menu with options from 0 to 18, with "5" selected. The "DATA" column for "1235FMD002" shows a dropdown menu with options from 0 to 18, with "18" selected. The "LOG" and "CACHE" columns are empty for all devices.

Daniel Morgan | damorgan12c@gmail.com | morganslibrary.org

Integrating ODAs with ZFS to Create a Secure Scalable HA Environment

Presented: Rocky Mountain Oracle User Group - 13 January, 2013

ZFS BUI: Config Storage

SUN ZFS STORAGE 7420 ⚠ The cluster peer has rejoined the cluster.

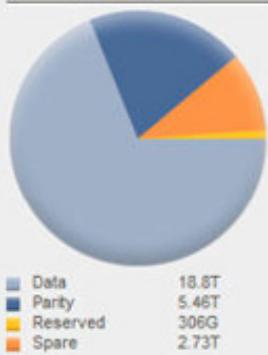
LOGOUT HELP Dismiss X

Confirm that all devices are present and minimally functional, and allocate them to a storage pool.

Choose Storage Profile Step 2 of 2

Configure available storage into a pool by defining its underlying redundancy profile. Carefully read the profile descriptions to understand how each balances the inherent trade-offs between availability, performance, and capacity, and select the profile that best fits your workload. If available, NSPF indicates no single point of failure, which affords certain profiles the ability for a pool to survive through loss of a single disk shelf.

Storage Breakdown



■ Data	18.8T
■ Parity	5.46T
■ Reserved	308G
■ Spare	2.73T

Data Profile

TYPE	NSPF	AVAILABILITY	PERFORMANCE	CAPACITY	SIZE
Double parity	No	<div style="width: 20%;"></div>	<div style="width: 20%;"></div>	<div style="width: 20%;"></div>	18.8T
Mirrored	Yes	<div style="width: 100%;"></div>	<div style="width: 100%;"></div>	<div style="width: 100%;"></div>	10.7T
Mirrored	No	<div style="width: 100%;"></div>	<div style="width: 100%;"></div>	<div style="width: 100%;"></div>	10.7T
Single parity, narrow stripes	No	<div style="width: 100%;"></div>	<div style="width: 100%;"></div>	<div style="width: 100%;"></div>	16.1T
Striped	No	<div style="width: 100%;"></div>	<div style="width: 100%;"></div>	<div style="width: 100%;"></div>	26.9T
Triple mirrored	Yes	<div style="width: 100%;"></div>	<div style="width: 100%;"></div>	<div style="width: 100%;"></div>	8.06T
Triple mirrored	No	<div style="width: 100%;"></div>	<div style="width: 100%;"></div>	<div style="width: 100%;"></div>	8.06T
Triple parity, wide stripes	No	<div style="width: 20%;"></div>	<div style="width: 20%;"></div>	<div style="width: 20%;"></div>	16.1T

Disk Breakdown

Data + Parity	9 disks
Spare	1 disks
Log	0 disks
Cache	0 disks

Data profile: Double parity

Each array stripe contains two parity disks, yielding high availability while increasing capacity over mirrored configurations. Double parity striping is recommended for workloads requiring little or no random access, such as backup/restore.

ZFS BUI: Config Network Config

The screenshot shows the Sun ZFS Storage 7420 BUI interface. The top navigation bar includes the Sun logo, Oracle branding, and links for LOGOUT and HELP. The main menu tabs are Configuration, Maintenance, Shares, Status, and Analytics, with Configuration being the active tab. Sub-tabs under Configuration are SERVICES, STORAGE, NETWORK, SAN, CLUSTER, USERS, PREFERENCES, and ALERTS, with NETWORK being the active sub-tab. The left sidebar is titled 'Network' and contains sections for 'Devices' (12 total), 'Datalinks' (4 total), and 'Interfaces' (4 total). The 'Devices' section lists network interfaces categorized by PCIe slot: BUILT-IN (igb0, igb1, igb2, igb3), PCIe 3 (ixgbe0, ixgbe1), PCIe 6 (ixgbe2, ixgbe3), PCIe 7 (ibp2, ibp3), and PCIe 2 (ibp0, ibp1). The 'Datalinks' section shows connections between interfaces: igb0 via igb0, igb1 via igb1, ixgbe0 via ixgbe0, and ixgbe2 via ixgbe2. The 'Interfaces' section lists network interfaces with their IPv4 addresses: head1 (net0, 192.168.40.248/22), head2 (net1, 192.168.40.249/22), private10gb (10.221.112.49/24), and private10gb2 (10.221.112.50/24). Below the interface list are 'Configuration', 'Addresses', and 'Routing' tabs, with 'Configuration' being the active tab. At the bottom of the page are 'REVERT' and 'APPLY' buttons.

ZFS BUI: Configuration Services LDAP

The screenshot shows the Sun ZFS Storage 7420 Configuration Services LDAP page. The top navigation bar includes links for Configuration, Maintenance, Shares, Status, and Analytics. The Configuration tab is selected, and the Services sub-tab is also selected. The LDAP service is currently active, indicated by a green dot next to its name. The page displays the configuration for the LDAP Directory Service, including options for SSL/TLS protection, base search DN (dc-unix, dc-m1ib, dc=com), search scope (Subtree (recursive)), authentication method (Simple (RFC 4513)), bind credential level (Anonymous), and schema definition. A table below lists a single server entry: perrito3.m1ib.com:389. The bottom of the page includes 'See Also' links for Help: LDAP and Wikipedia: LDAP, and a footer with the text 'Super-User@c0zfs742001p LOGOUT HELP'.

Super-User@c0zfs742001p LOGOUT HELP

Configuration Maintenance Shares Status Analytics

Services Storage Network SAN Cluster Users Preferences Alerts

Services LDAP

Back to Services 2012-10-11 12:30:32 Online REVERT APPLY

LDAP Directory Service
Recognize users and groups defined in an LDAP directory. Once the LDAP service is configured, go to Configuration/Users to give users permission to log into the administrative interface.

See Also
Help: LDAP
Wikipedia: LDAP

Protect LDAP traffic with SSL/TLS

Base search DN dc-unix, dc-m1ib, dc=com

Search scope One-level (non-recursive) Subtree (recursive)

Authentication method Simple (RFC 4513)

Bind credential level Anonymous Proxy

DN

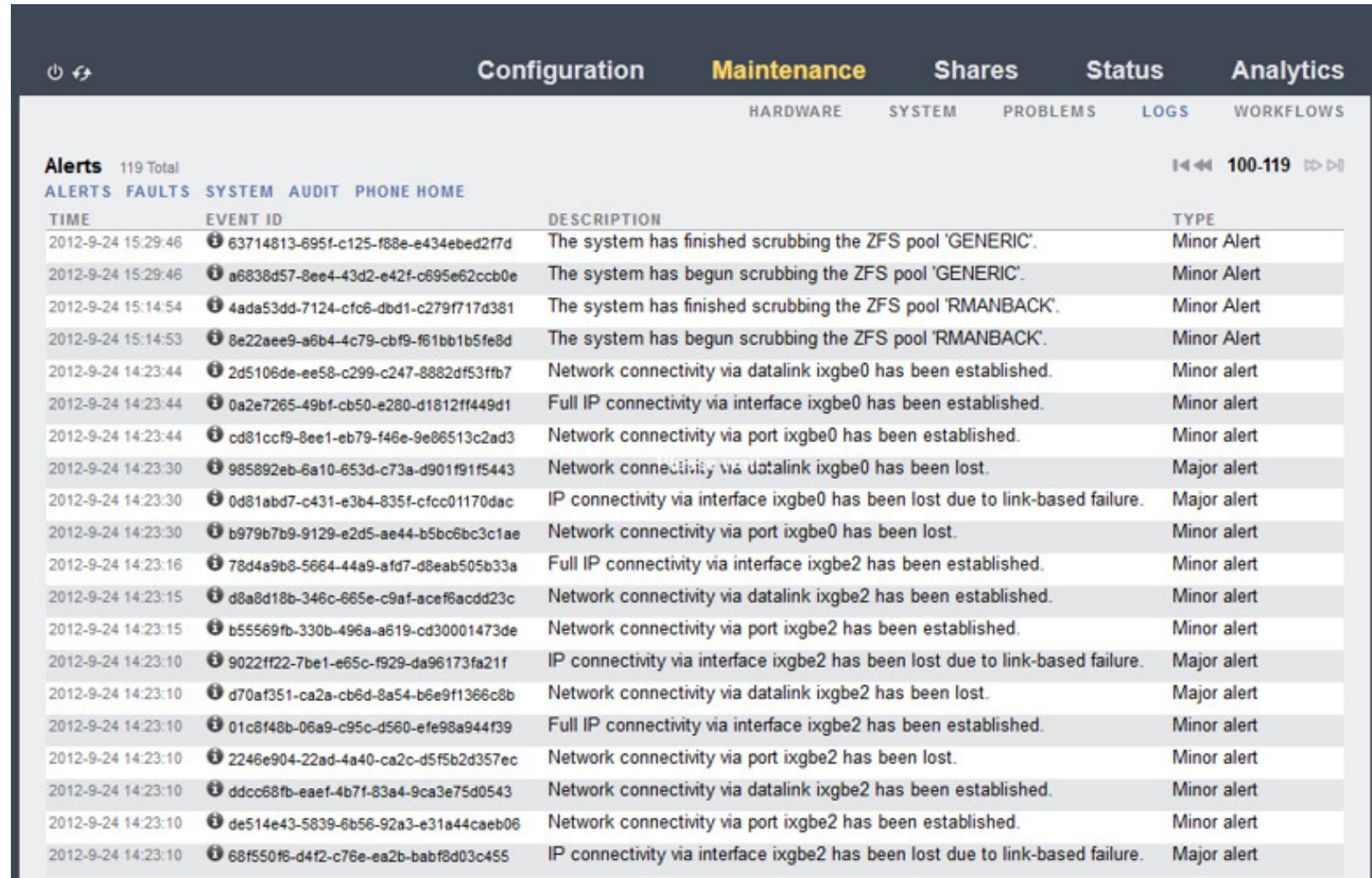
Password

Schema definition [Edit...](#)

Servers 1 Total

SERVER	CERTIFICATE	EXPIRES
perrito3.m1ib.com:389		

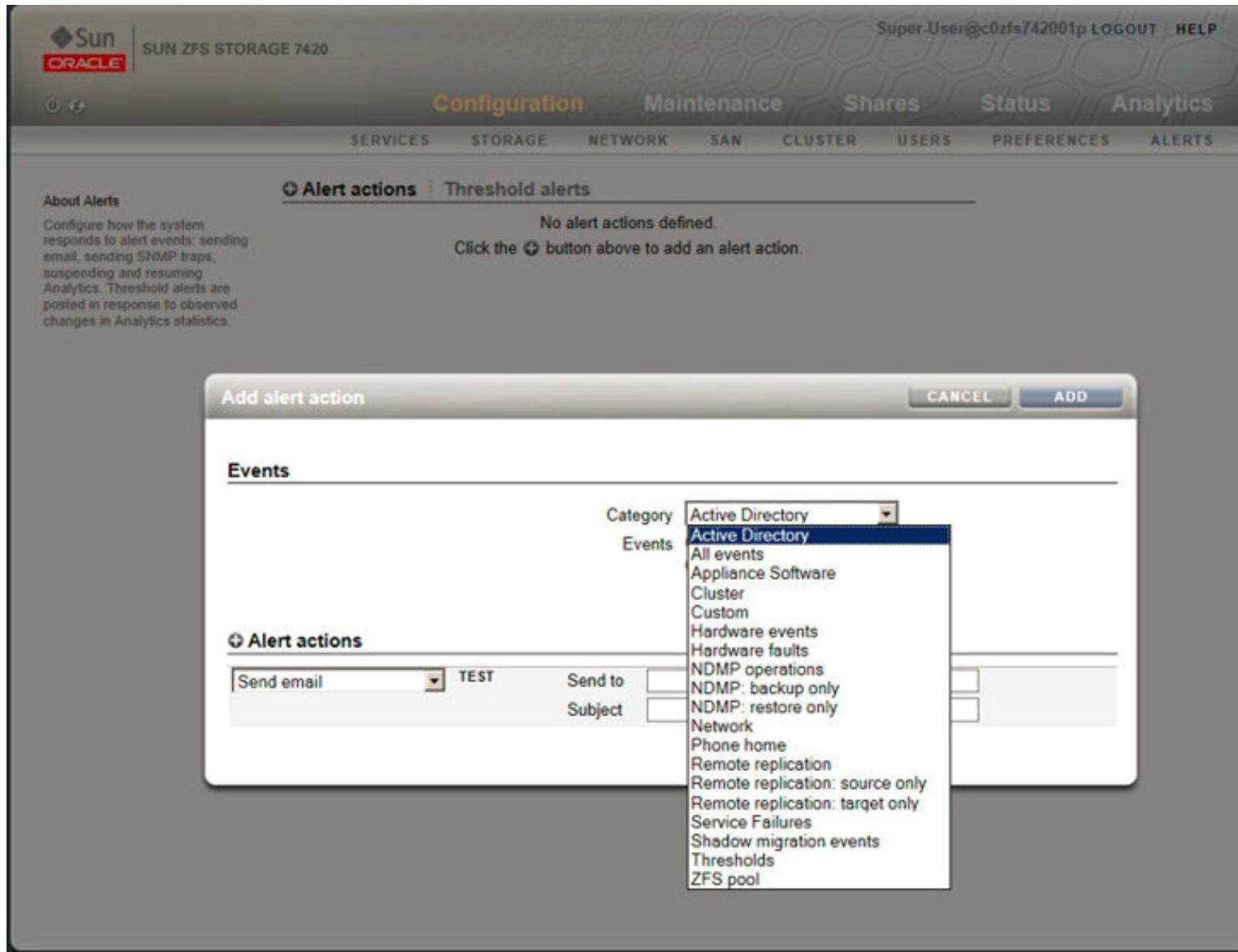
ZFS BUI: Maintenance Logs



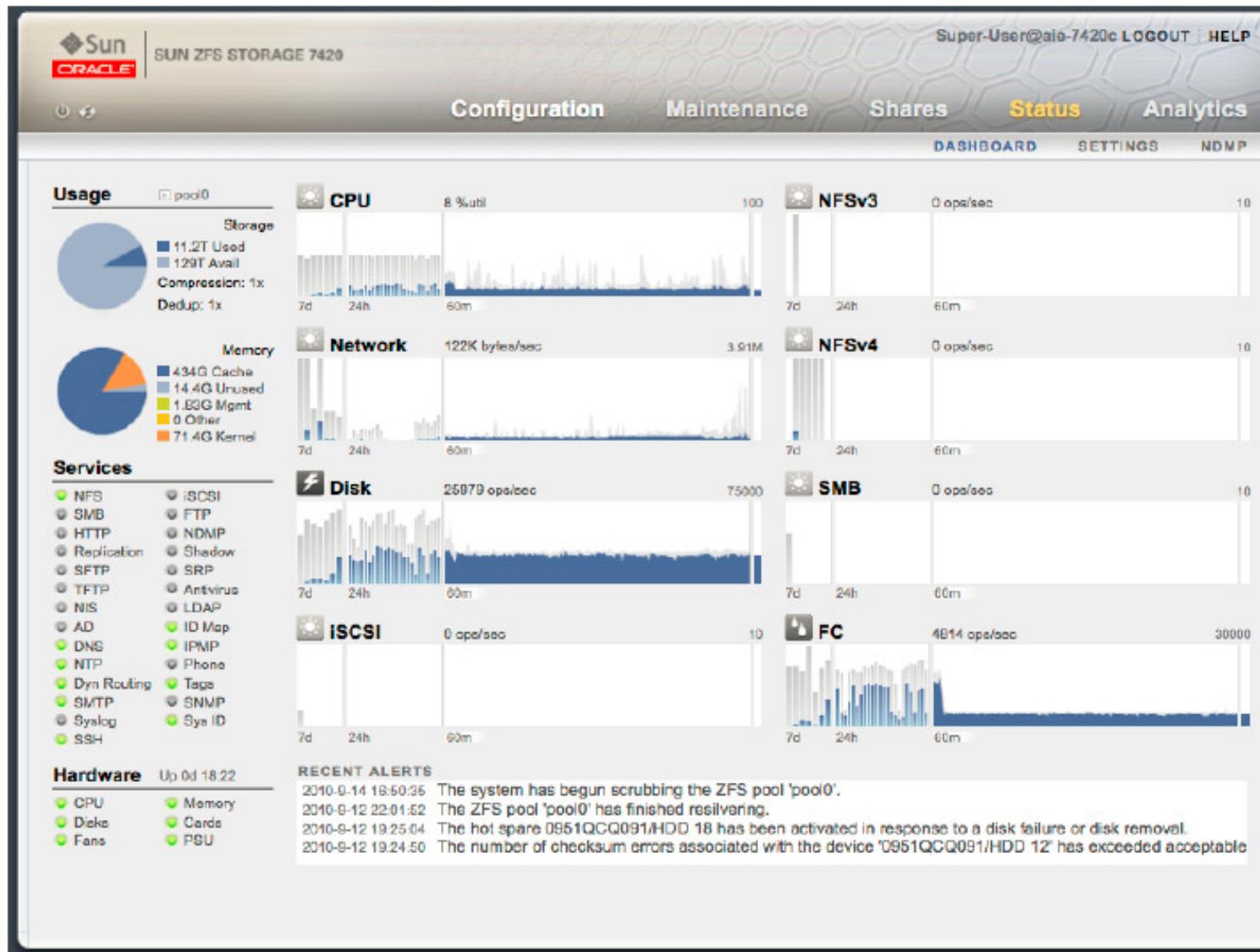
The screenshot shows the ZFS BUI Maintenance Logs page. At the top, there are navigation links for Configuration, Maintenance (which is highlighted in yellow), Shares, Status, and Analytics. Below these are sub-links for Hardware, System, Problems, Logs, and Workflows. A search bar with the placeholder '100-119' is located above the main table. The table has columns for TIME, EVENT ID, DESCRIPTION, and TYPE. The rows list various system events, mostly related to network connectivity and ZFS pool scrubbing, with most being minor alerts.

TIME	EVENT ID	DESCRIPTION	TYPE
2012-9-24 15:29:46	63714813-695f-c125-f88e-e434ebcd27d	The system has finished scrubbing the ZFS pool 'GENERIC'.	Minor Alert
2012-9-24 15:29:46	a6838d57-8ee4-43d2-e42f-c695e62ccb0e	The system has begun scrubbing the ZFS pool 'GENERIC'.	Minor Alert
2012-9-24 15:14:54	4ada53dd-7124-cfc6-dbd1-c279f717d381	The system has finished scrubbing the ZFS pool 'RMANBACK'.	Minor Alert
2012-9-24 15:14:53	8e22aee9-a6b4-4c79-cbf9-f61bb1b5fe8d	The system has begun scrubbing the ZFS pool 'RMANBACK'.	Minor Alert
2012-9-24 14:23:44	2d5106de-ee58-c299-c247-8882df53ff7	Network connectivity via datalink ixgbe0 has been established.	Minor alert
2012-9-24 14:23:44	0a2e7265-49bf-cb50-e280-d1812ff449d1	Full IP connectivity via interface ixgbe0 has been established.	Minor alert
2012-9-24 14:23:44	cd81ccf9-8ee1-eb79-f46e-9e86513c2ad3	Network connectivity via port ixgbe0 has been established.	Minor alert
2012-9-24 14:23:30	985892eb-6a10-653d-c73a-d901f91f5443	Network connectivity via datalink ixgbe0 has been lost.	Major alert
2012-9-24 14:23:30	0d81abd7-c431-e3b4-835f-cfcc01170dac	IP connectivity via interface ixgbe0 has been lost due to link-based failure.	Major alert
2012-9-24 14:23:30	b979b7b9-9129-e2d5-ae44-b5bc6bc3c1ae	Network connectivity via port ixgbe0 has been lost.	Minor alert
2012-9-24 14:23:16	78d4a9b8-5664-44a9-af07-d8eab505b33a	Full IP connectivity via interface ixgbe2 has been established.	Minor alert
2012-9-24 14:23:15	d8a0d18b-346c-665e-c9af-acef6acdd23c	Network connectivity via datalink ixgbe2 has been established.	Minor alert
2012-9-24 14:23:15	b55569fb-330b-496a-a619-cd30001473de	Network connectivity via port ixgbe2 has been established.	Minor alert
2012-9-24 14:23:10	9022ff22-7be1-e65c-f929-da96173fa21f	IP connectivity via interface ixgbe2 has been lost due to link-based failure.	Major alert
2012-9-24 14:23:10	d70af351-ca2a-cb6d-8a54-b6e9f1366c8b	Network connectivity via datalink ixgbe2 has been lost.	Major alert
2012-9-24 14:23:10	01c8f48b-06a9-c95c-d560-efef98a944f39	Full IP connectivity via interface ixgbe2 has been established.	Minor alert
2012-9-24 14:23:10	2246e904-22ad-4a40-ca2c-d5f5b2d357ec	Network connectivity via port ixgbe2 has been lost.	Minor alert
2012-9-24 14:23:10	ddcc68fb-eaef-4b7f-83a4-9ca3e75d0543	Network connectivity via datalink ixgbe2 has been established.	Minor alert
2012-9-24 14:23:10	de514e43-5839-6b56-92a3-e31a44caeb06	Network connectivity via port ixgbe2 has been established.	Minor alert
2012-9-24 14:23:10	68f550f6-d4f2-c76e-ea2b-babf8d03c455	IP connectivity via interface ixgbe2 has been lost due to link-based failure.	Major alert

ZFS BUI: Configuration Alerts



ZFS Storage Appliances



How Does This Change Our Jobs?

Job Title	Loses	Gains
Storage Admins	Time wasted monitoring competing loads on the storage appliance balancing competing need to read/write cache, and allocation of disk.	More efficient storage environment as it is all file system.
Network Admins	Pain and suffering	Time to devote to troubleshooting, security monitoring, and other value-added tasks.
System Admins	<ul style="list-style-type: none">Gives up appliance root passwordGives up 2:00am support calls	
Database Admins		Patching operating system, firmware, and database as a single unit with patches previously tested for compatibility

Your ODA is not a general purpose computer, will not be hosting files, applications, middleware, etc.

How Does This Change Our Jobs?

- Storage Admin
 - No longer required
- Network Admin
 - Only required for public network interface
- System Admin
 - Advise on configuration
 - Install backup agent (ie Networker)
 - Install security software (ie TripWire)
- DBA
 - Just like with ASM ... assumes broader responsibility for deployment and patching
 - Gives up large amounts of unproductive time debugging configurations

Executive Summary

- Never make a technology argument when a financial argument will suffice
- My ODA Challenge - Cost Compare an ODA to:
 - Two pizza box servers with comparable CPU, RAM, internal disk, HBA and NIC interfaces
 - A 12TB SAN
 - Dual Cache Fusion Interconnect switches
 - Dual SAN switches
 - Add the cost of all of the time necessary to identify the components (and internal meetings to discuss it)
 - Add the cost for admins to rack and stack, cable, install, configure, patch, and build a working 2 node RAC cluster
- ODAs are the least expensive solution: by far

Questions

**ERROR at line 1:
ORA-00028: your session has been killed**



Thank you