Integrating the Oracle Database Appliance with the Sun ZFS Storage Appliance to Create an Ideal Database Environment

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

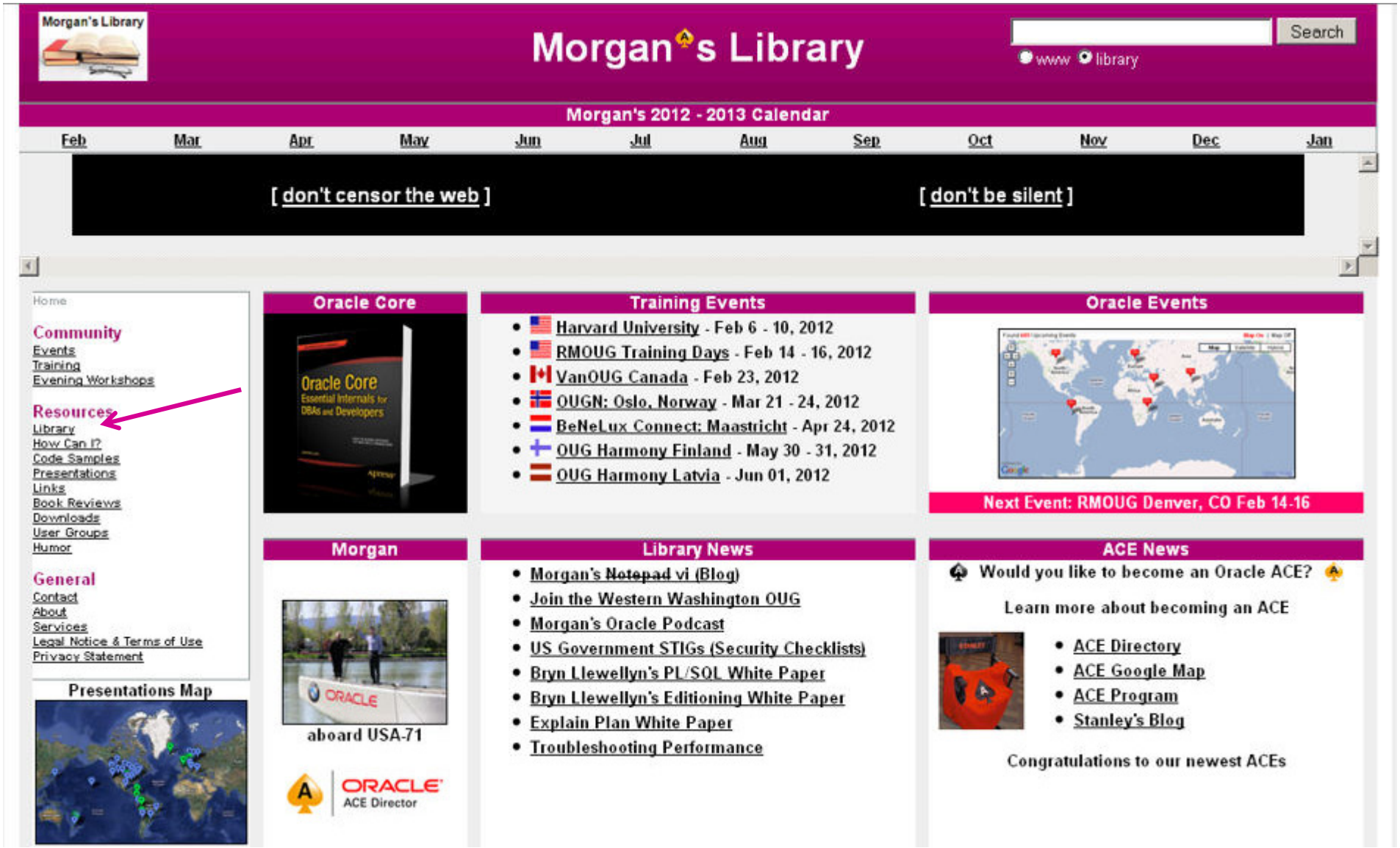
Daniel A. Morgan

- 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
- Executive Board: Vancouver/Victoria OUGs
- Upcoming Presentations
  - Apr 17-20: Oracle User Group Norway
  - May 15 Oracle User Group Azerbaijan
  - May 19 Bulgarian Oracle User Group

ORACLE RACSIG

International zSeries Oracle SIG

Morgan's Library: www.morganslibrary.org



Travels

LAD Tour: Machu Picchu Peru

Agenda

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

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

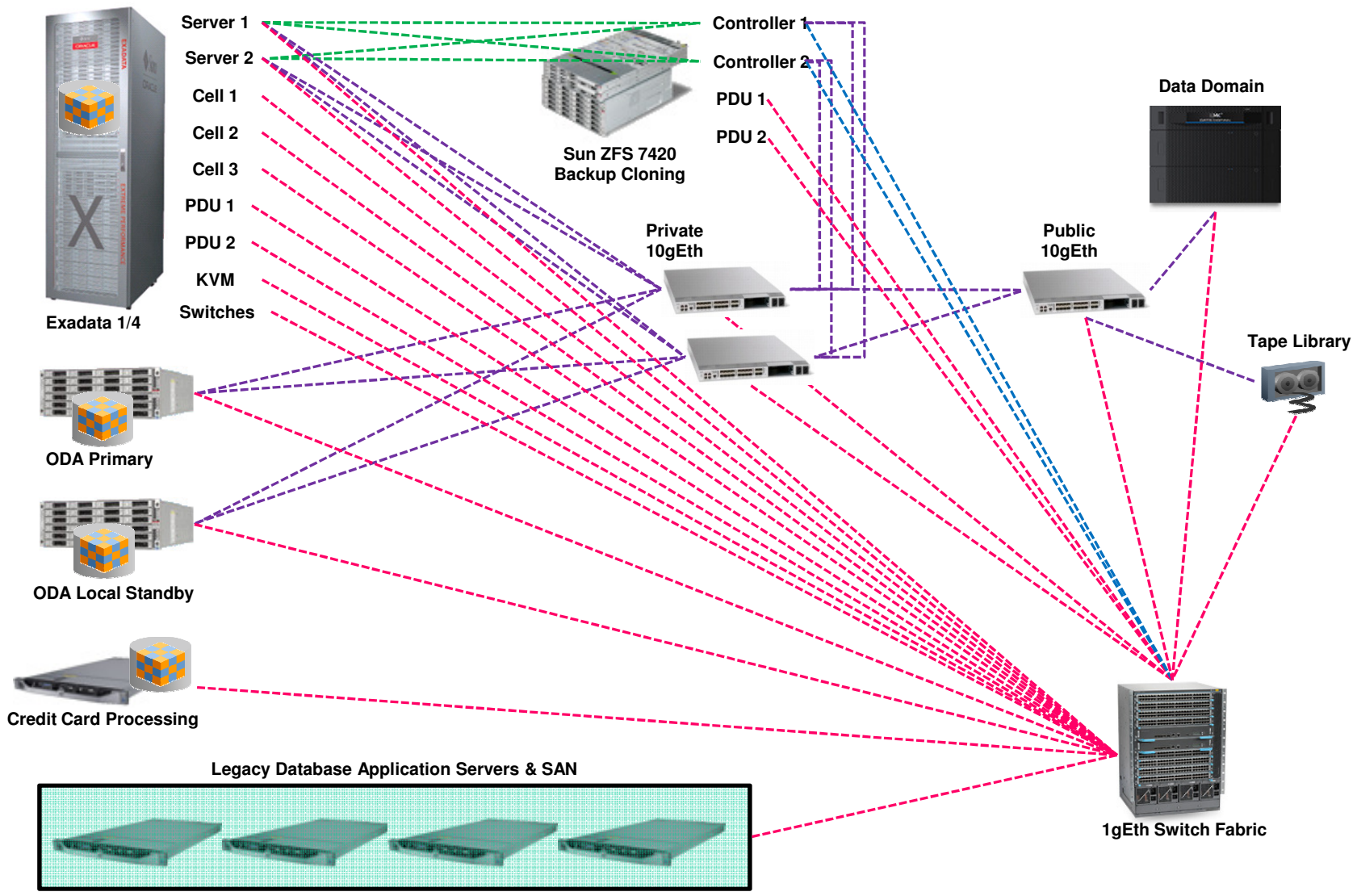
Questions we as IT professionals must answer

- 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

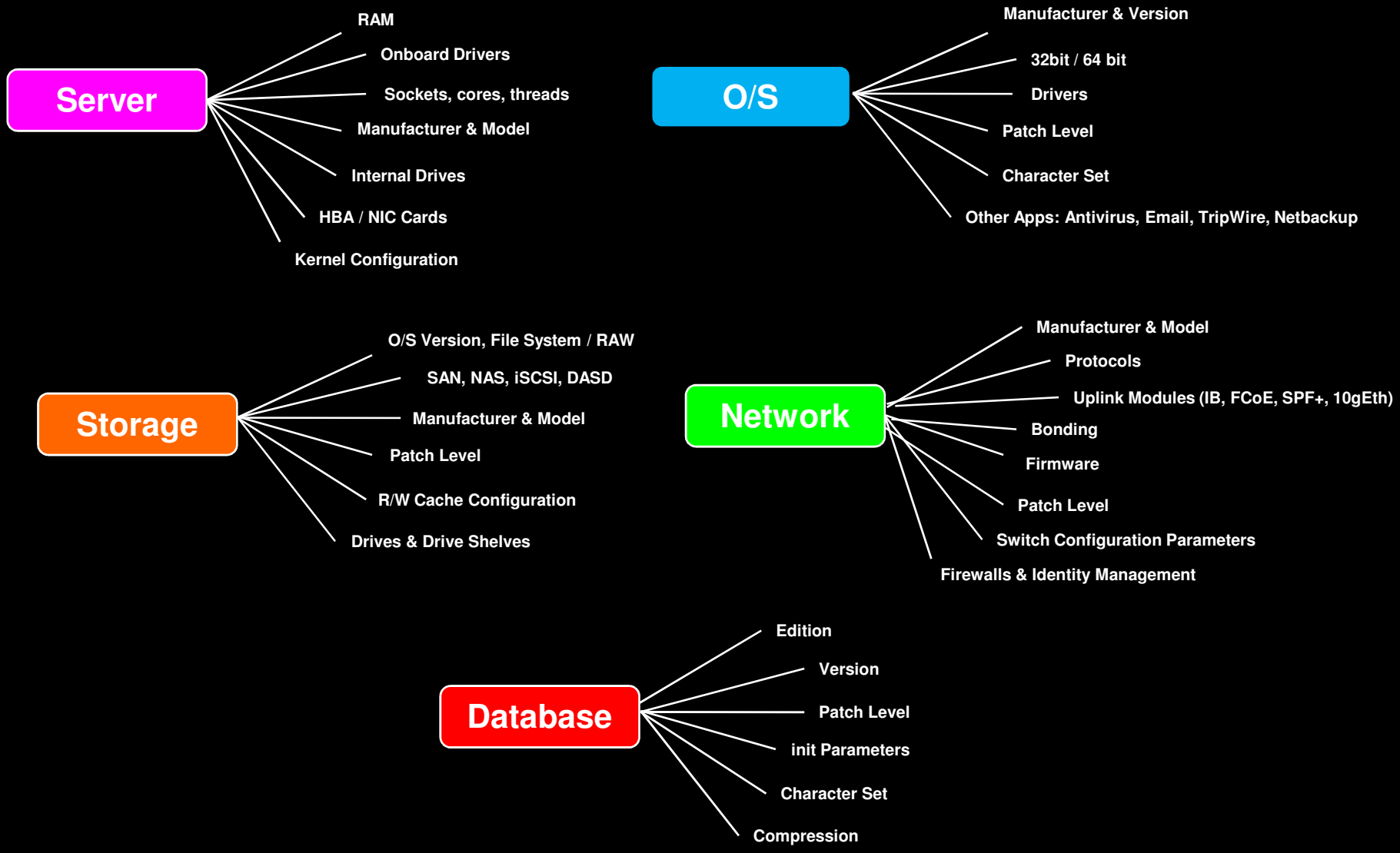
IT infrastructure meets a single point of failure

Puzzle Pieces

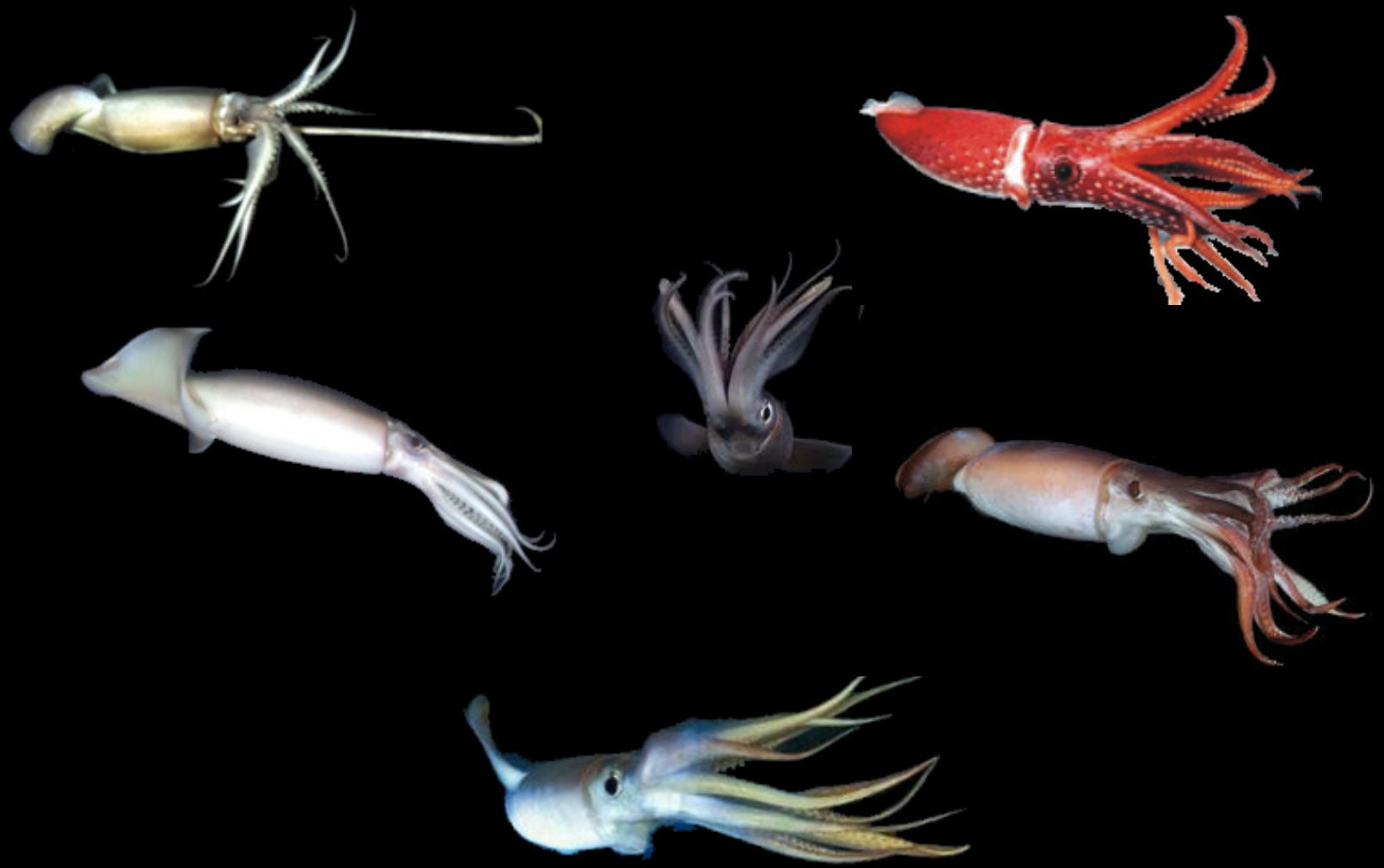


Each connection shown must be multiplexed and bonded

Static Puzzle Pieces



Animated Puzzle Pieces



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.

The Solution

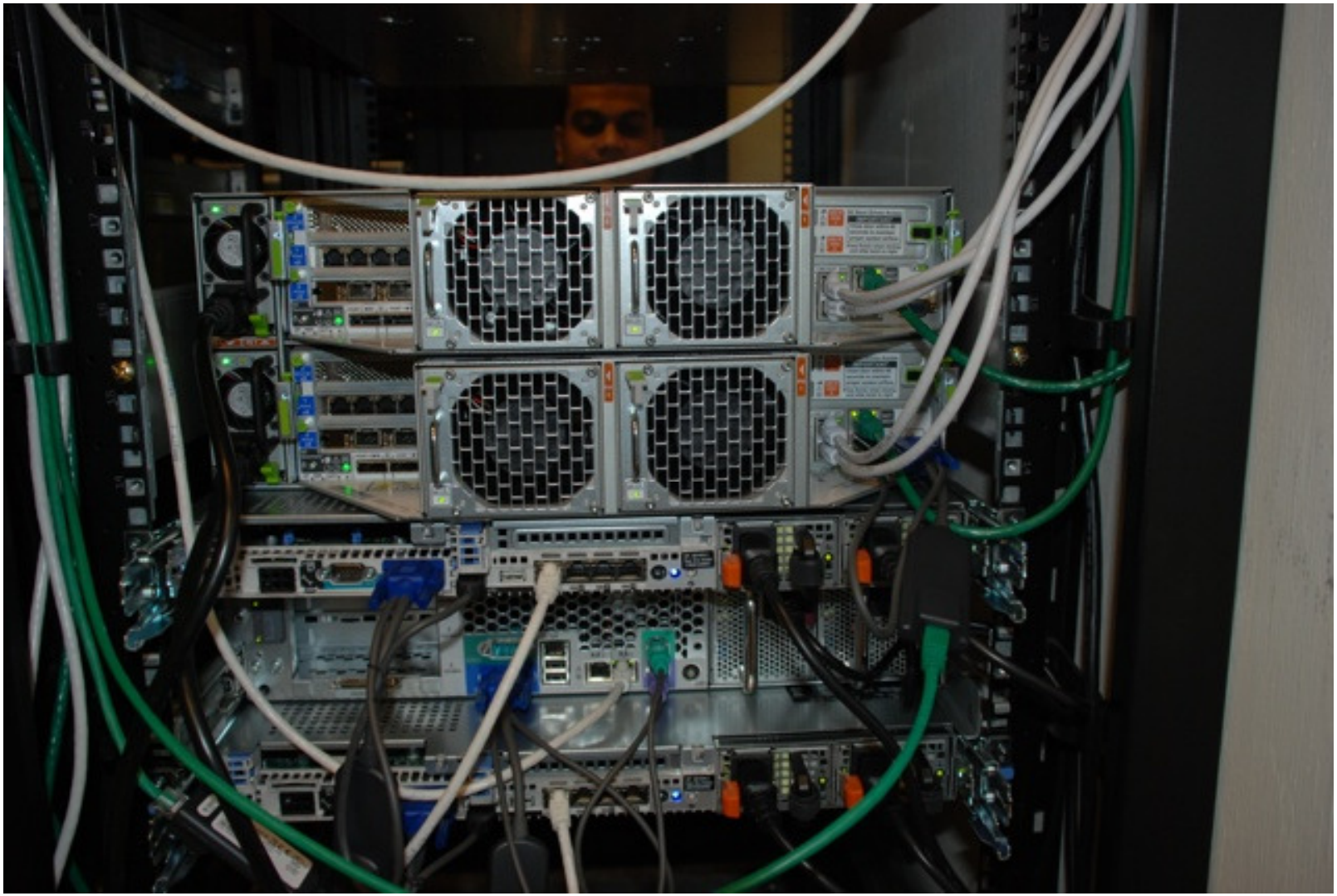
make different, and better, decisions

What is an ODA?

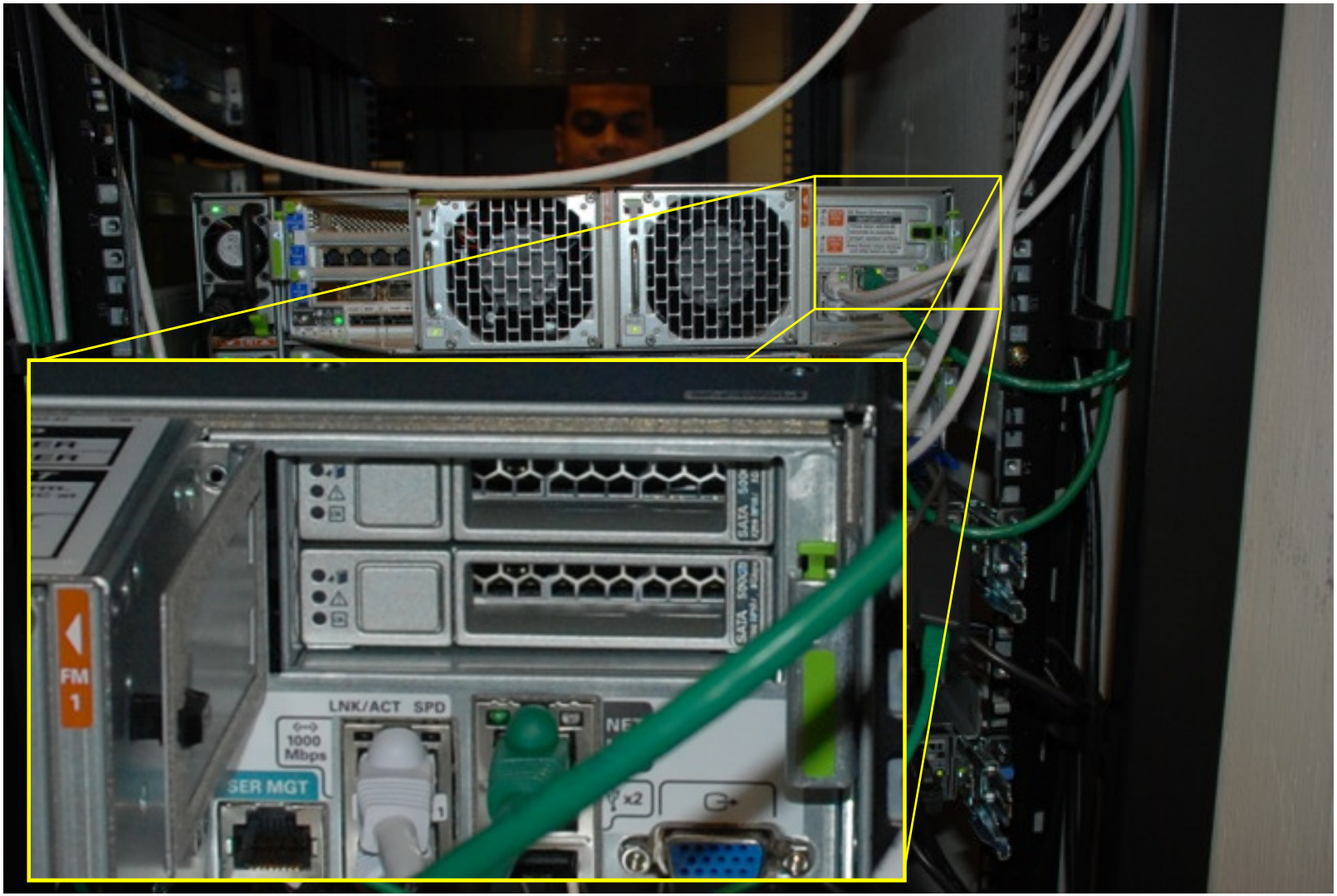
- An appliance
  - A single line on the invoice ... plus the power chords
  - But you get root and sys: The customer is in control
- Announced last year at OpenWorld
- Engineered two server RAC cluster in a 4U case
  - 24 CPU cores
  - 192 GB of RAM
  - 12 TB of direct attached storage with ASM mirroring
  - 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?

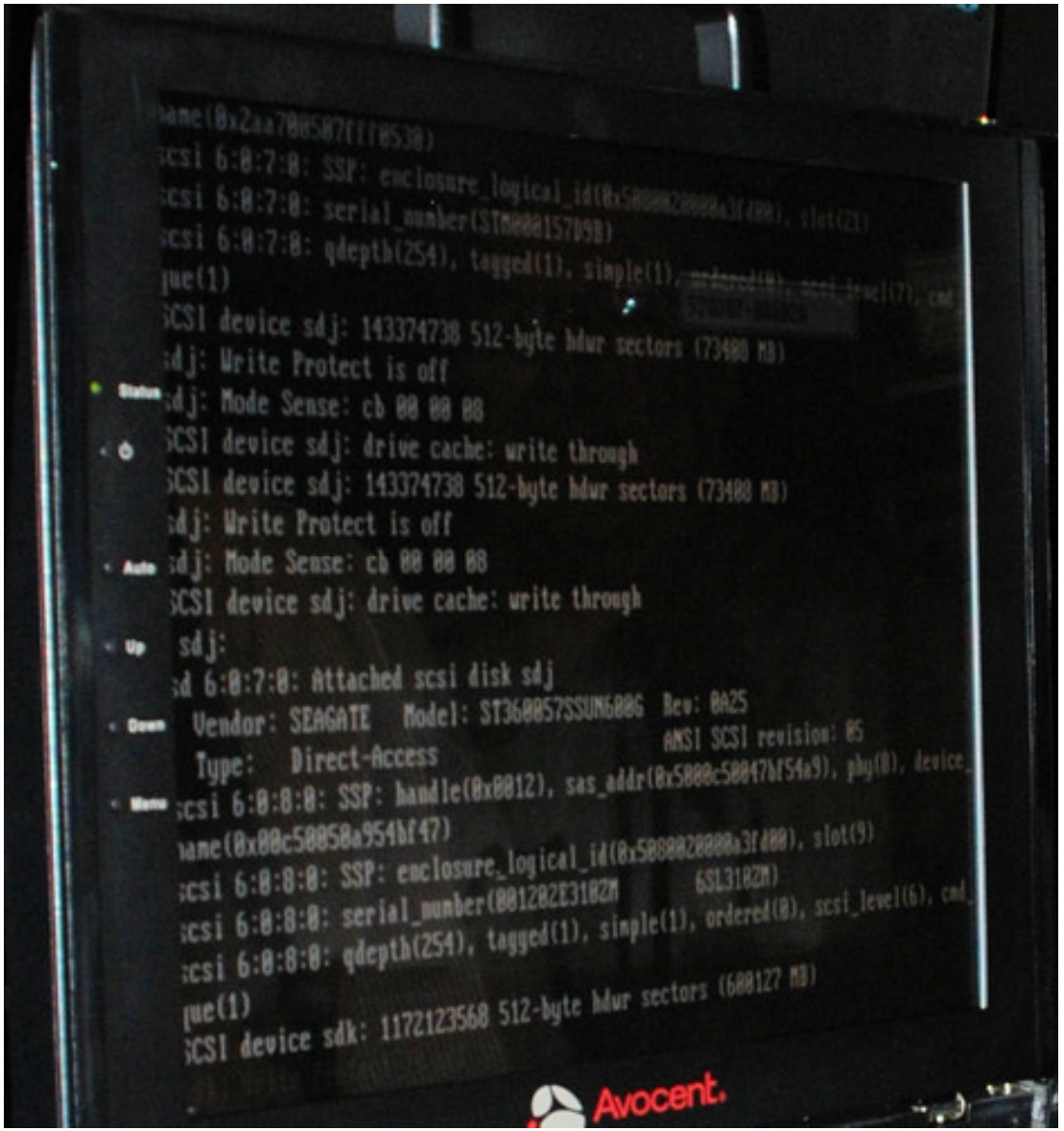
- Minimize complexity from rack-and-stack through database deployment
- Fewer resources required to deploy
  - UNIX System Admins: not required
  - Network Admins: not required
  - Storage Admins: not required
- Ease of maintenance and patching
  - One patch combines O/S, drivers, networking, infrastructure
  - 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

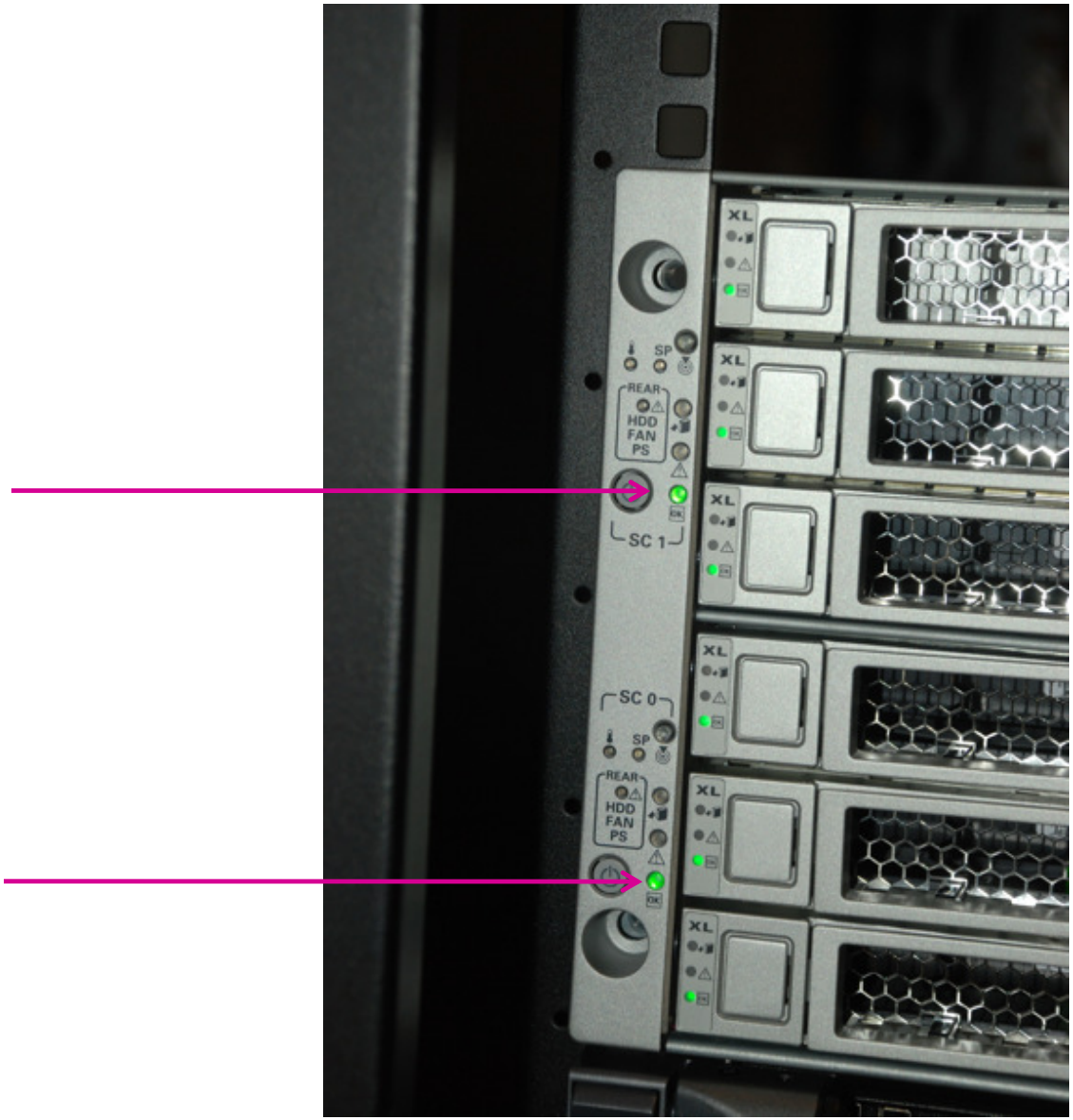
Discussion

Installation

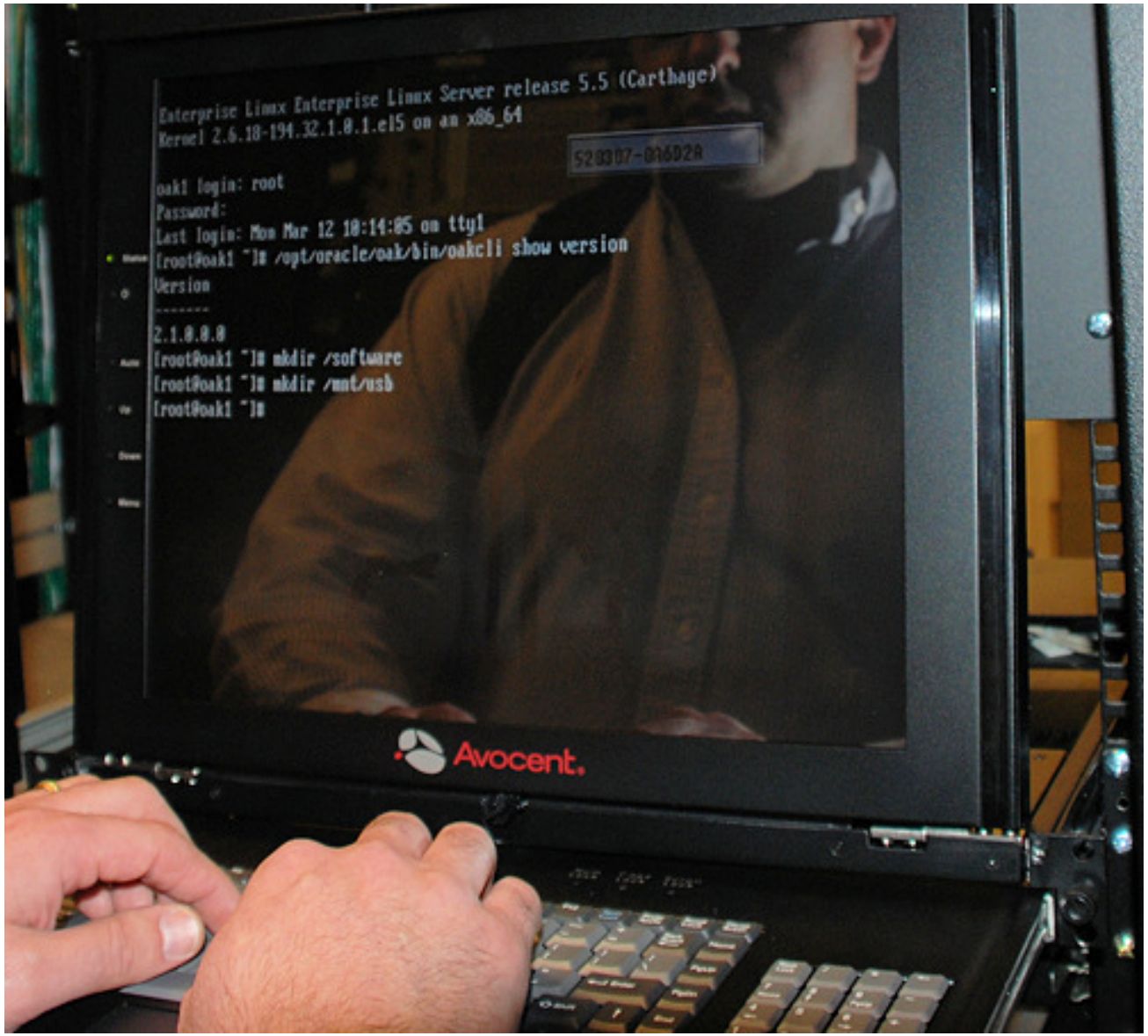
Step 1: Power On



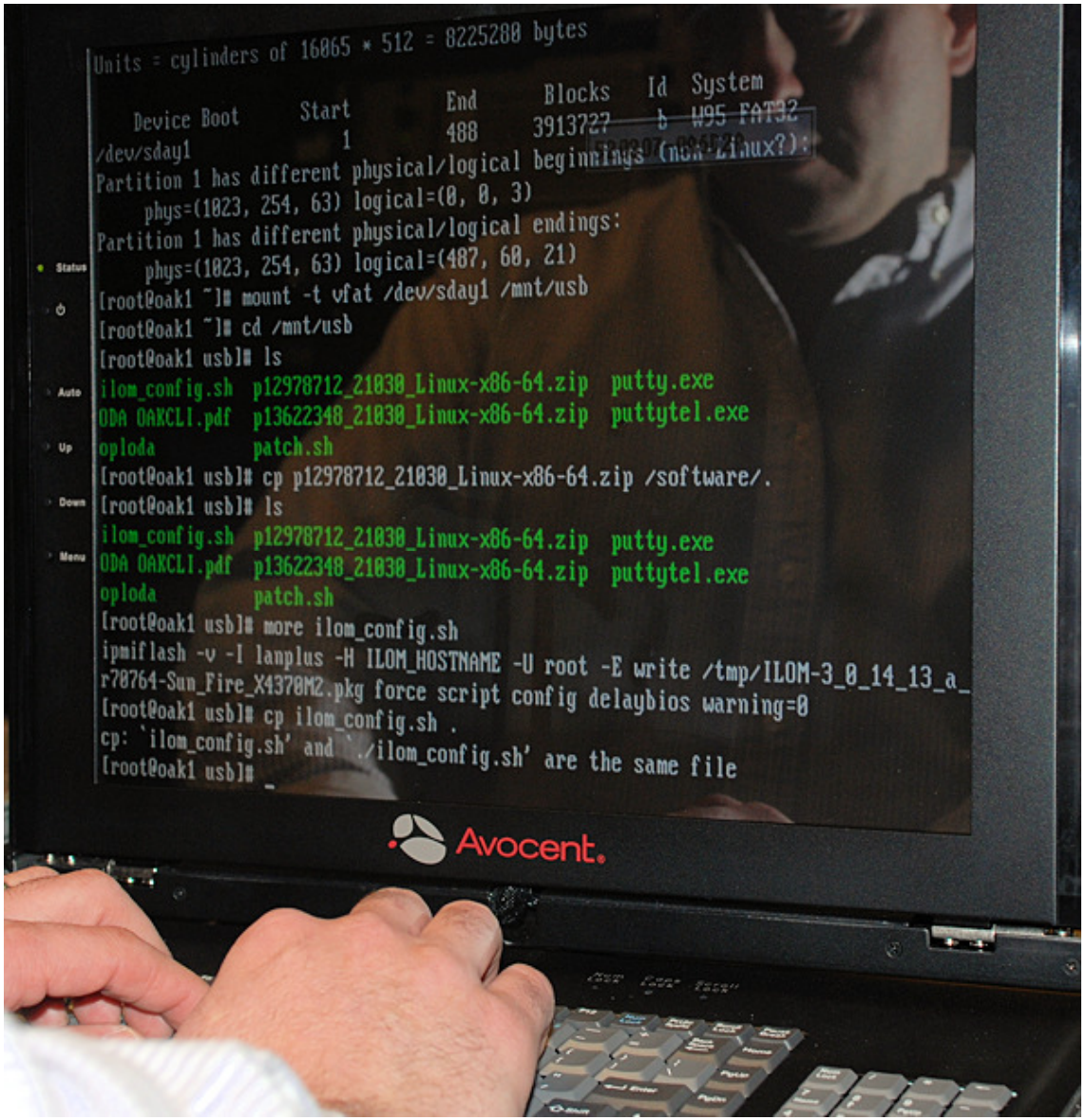
Step 2: Wait for OK Lights



Step 3: Log In as root

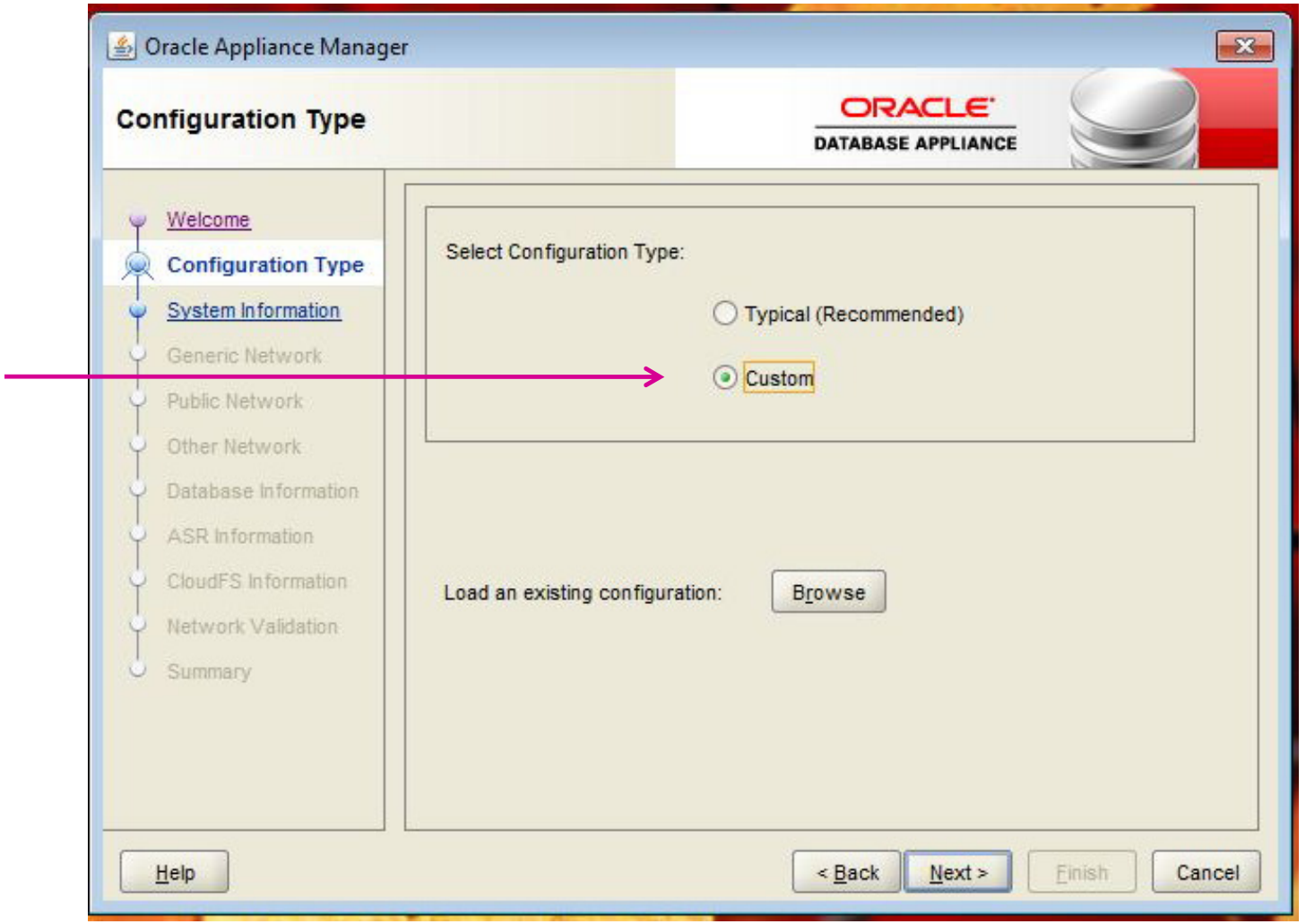


Step 5: Perform ILOM Configuration

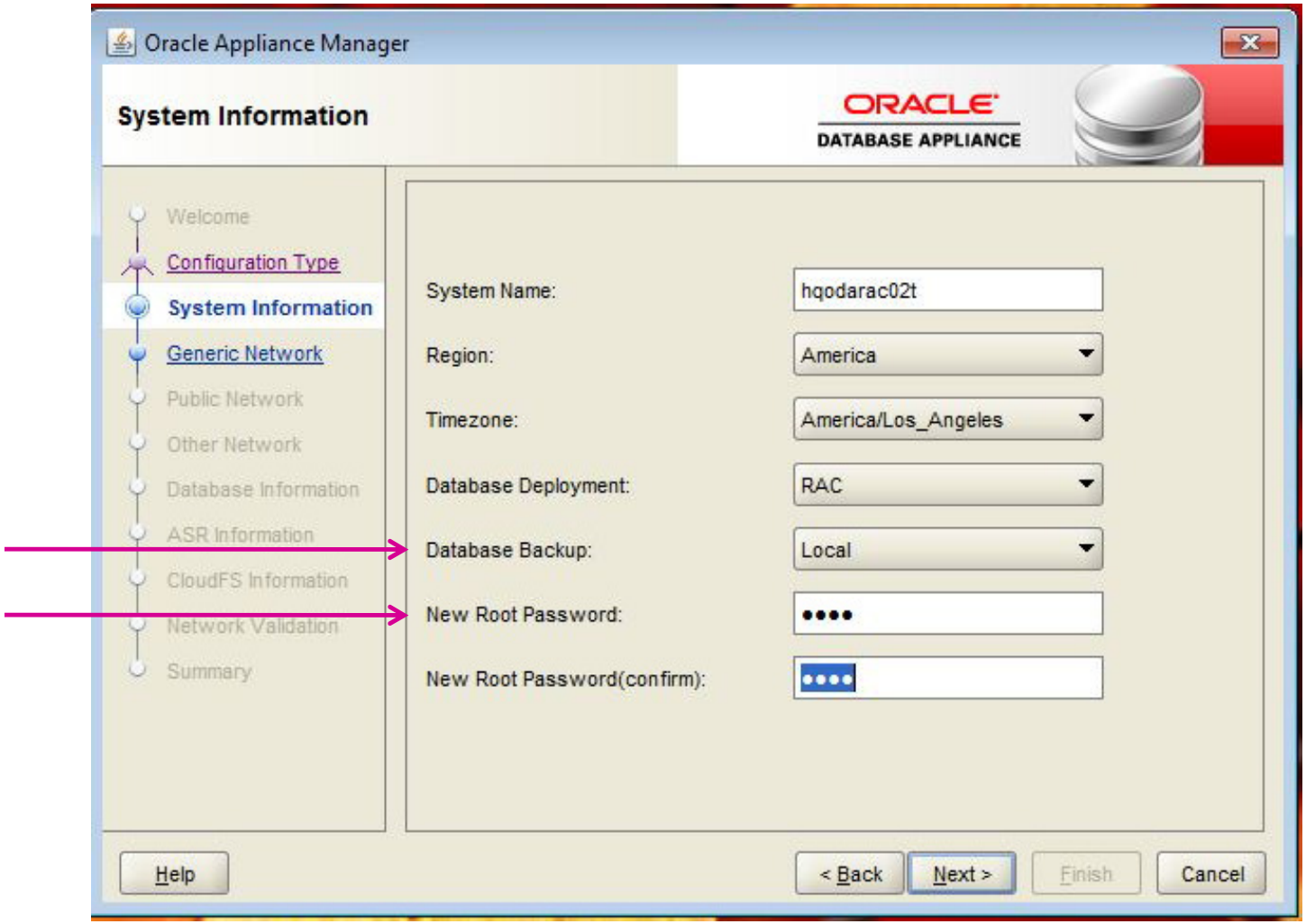


Install Screens: 1

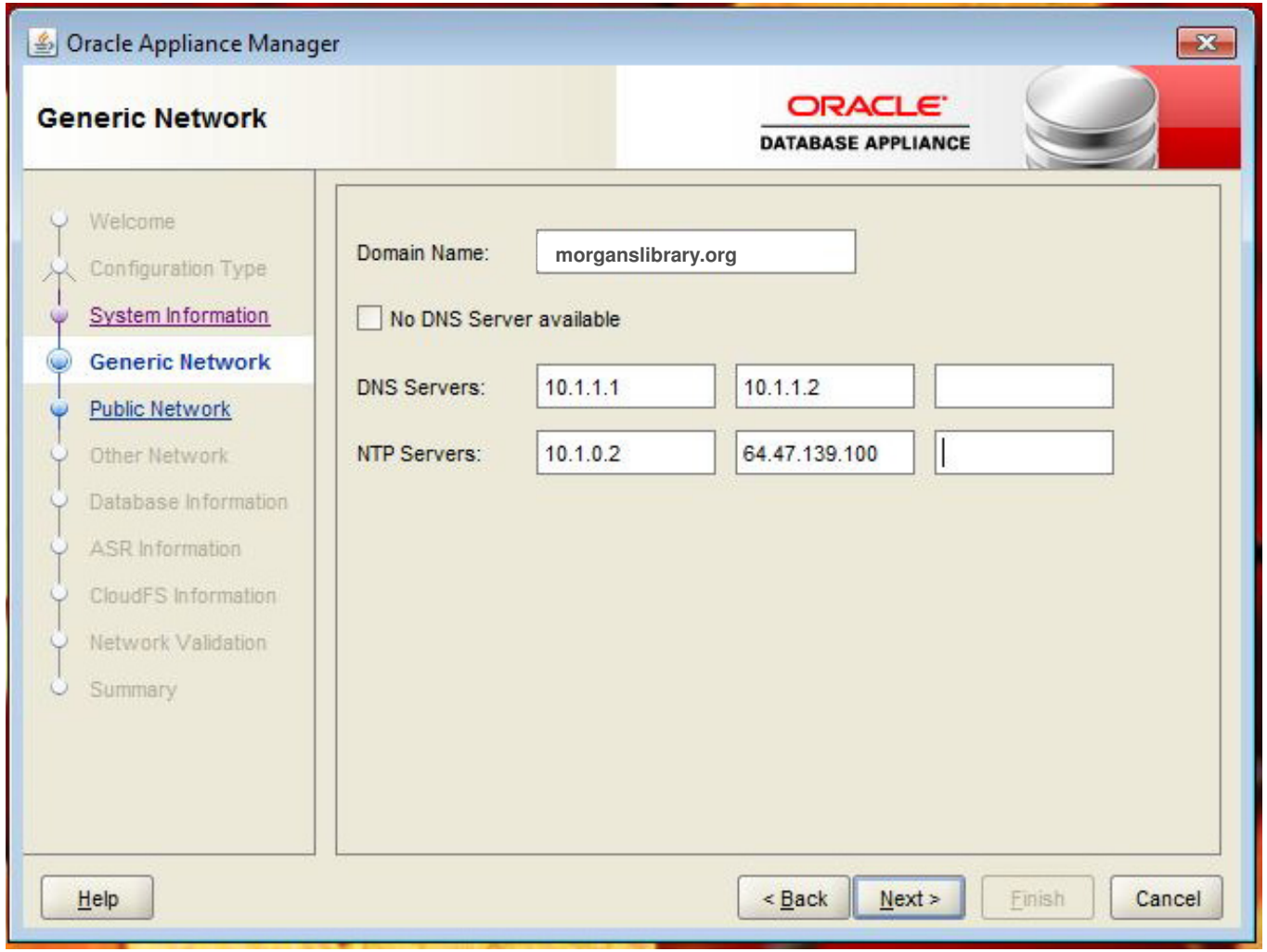
Install Screens: 2



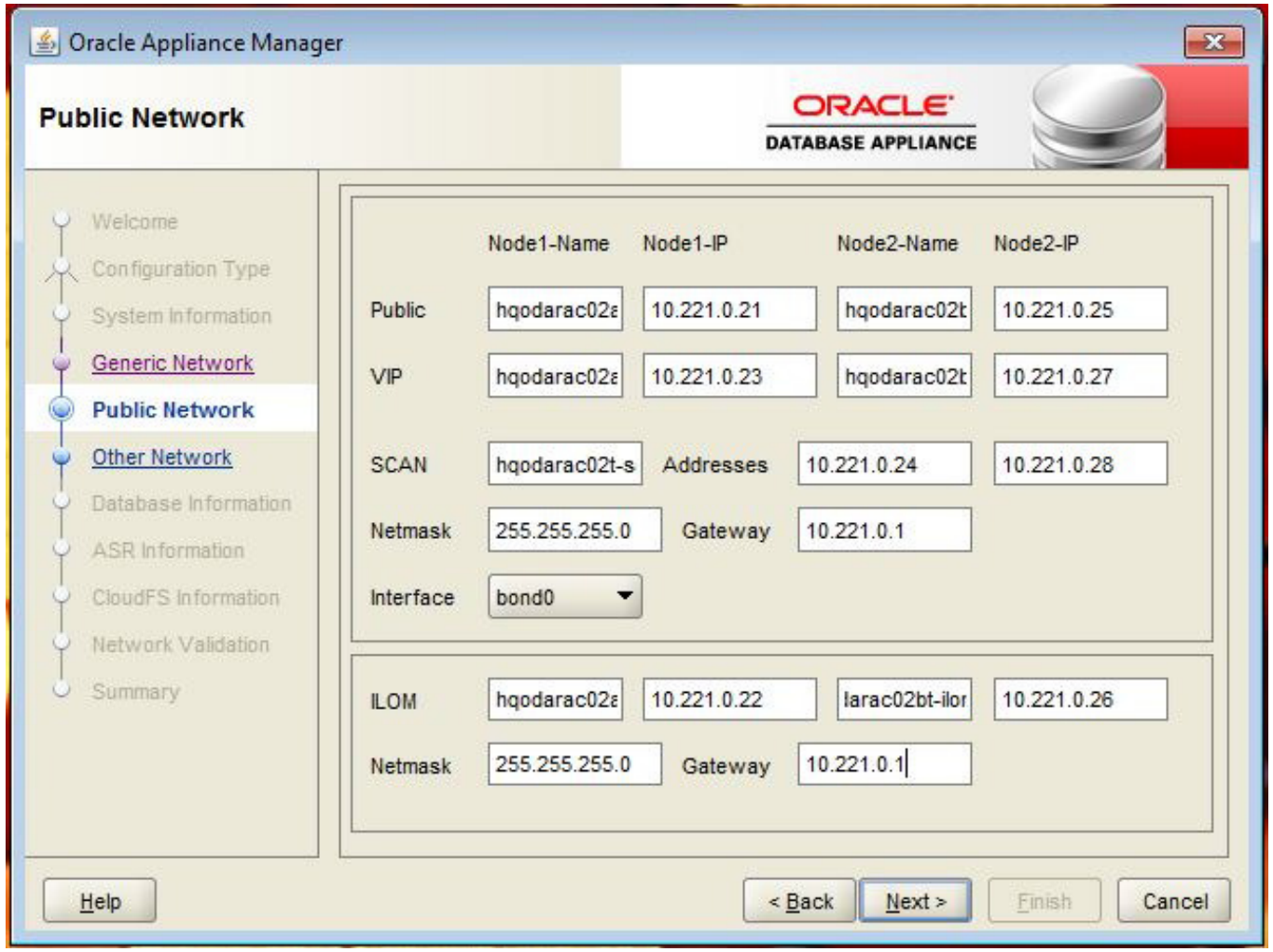
Install Screens: 3



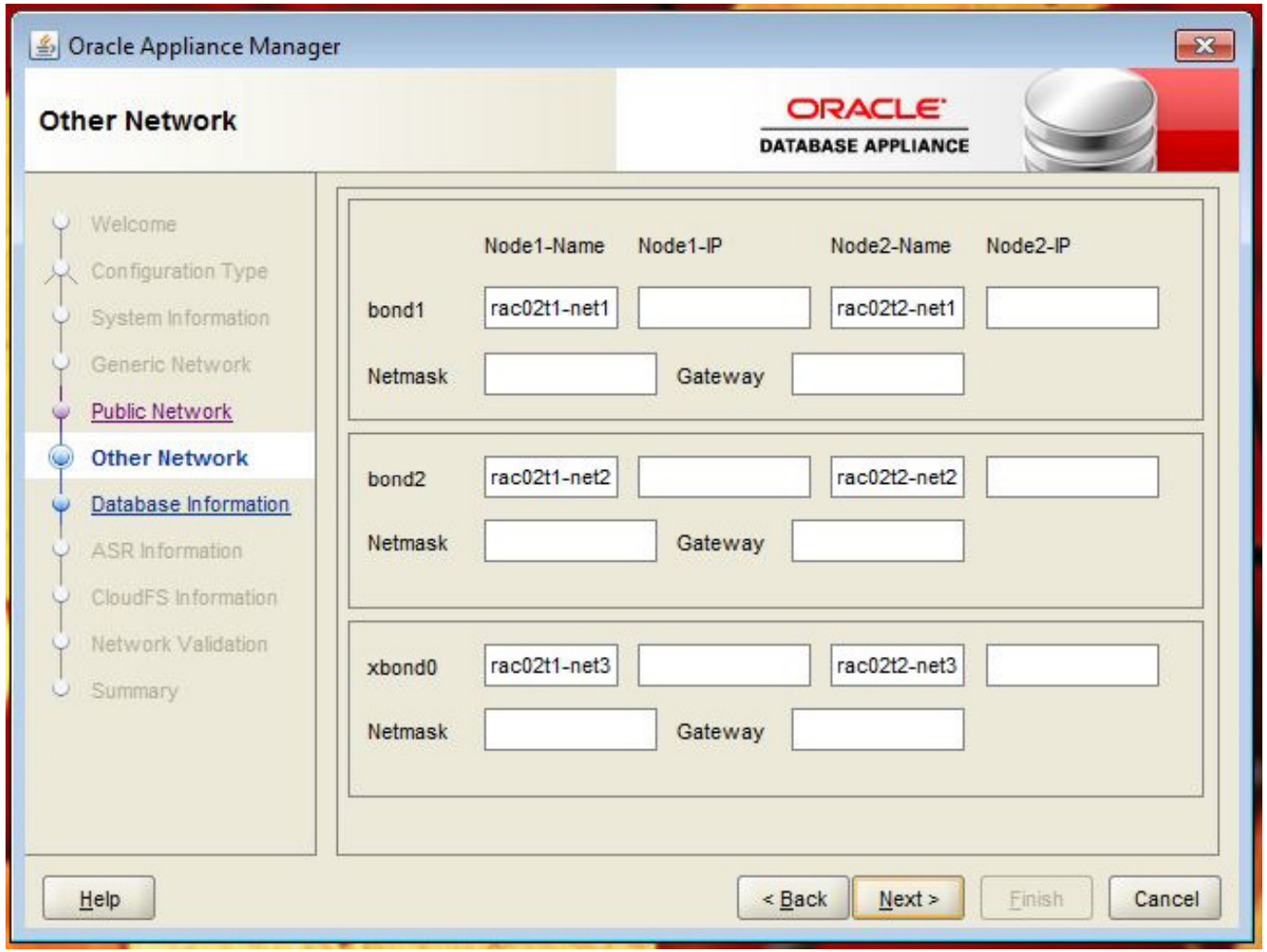
Install Screens: 4



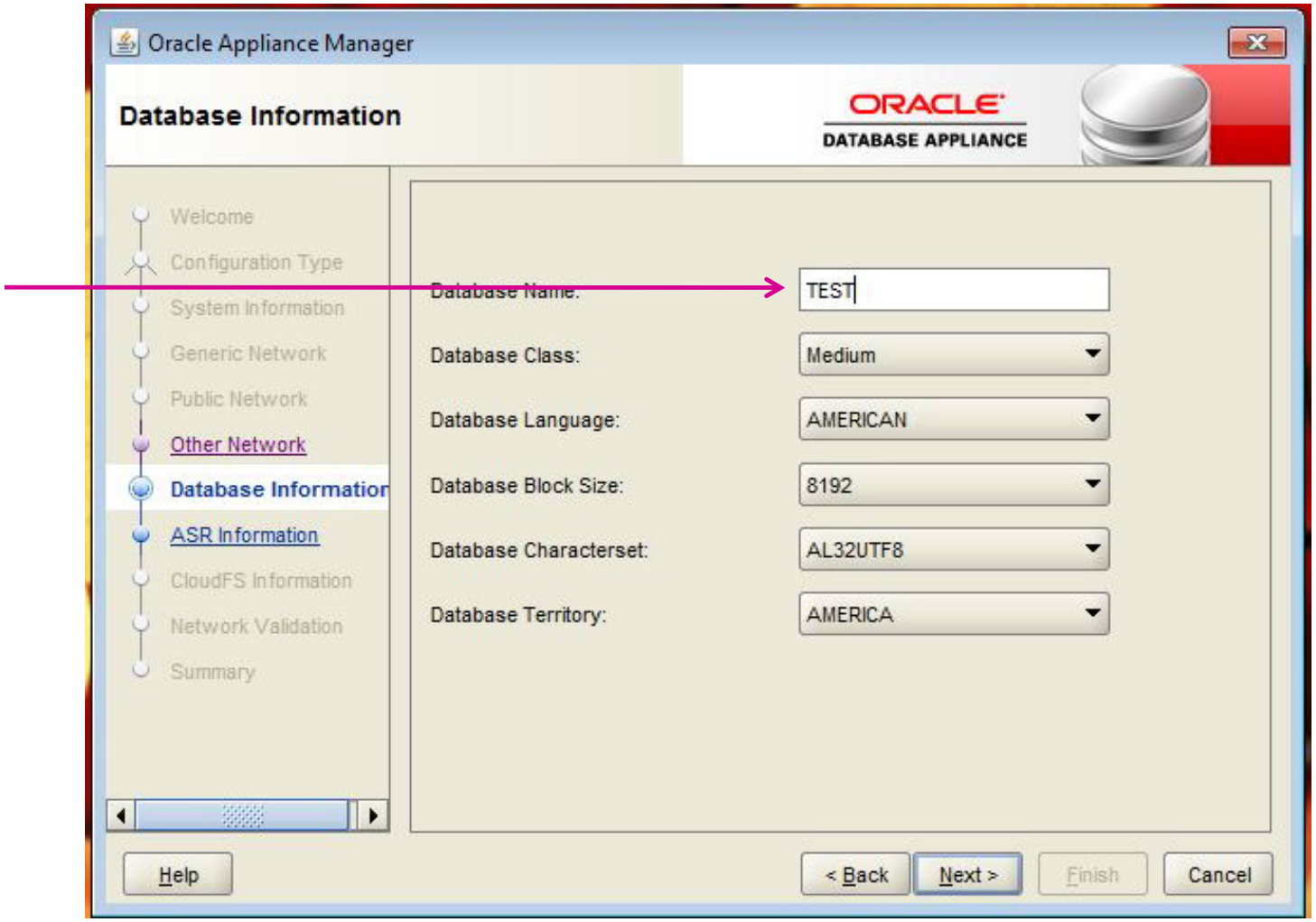
Install Screens: 5



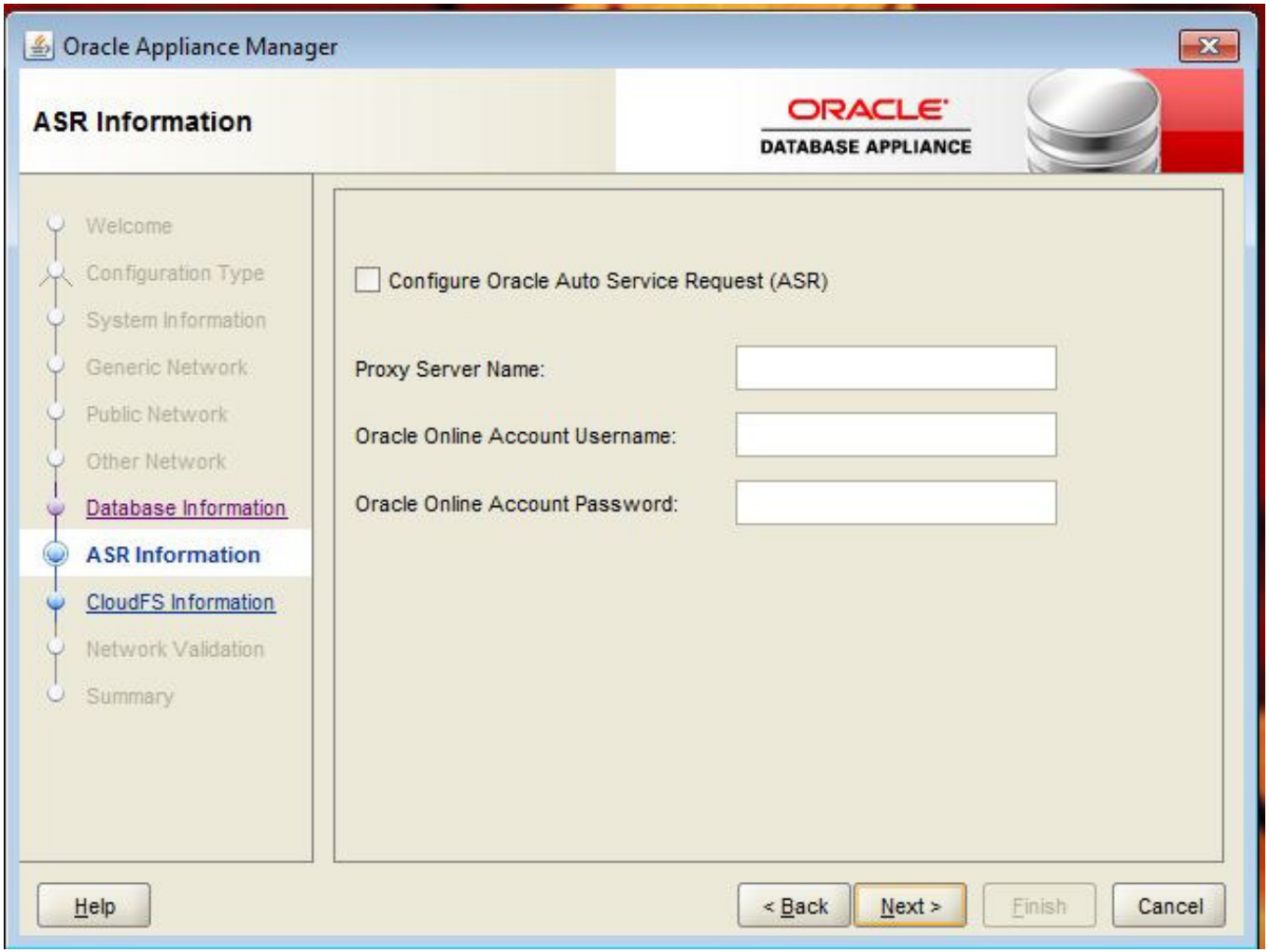
Install Screens: 6



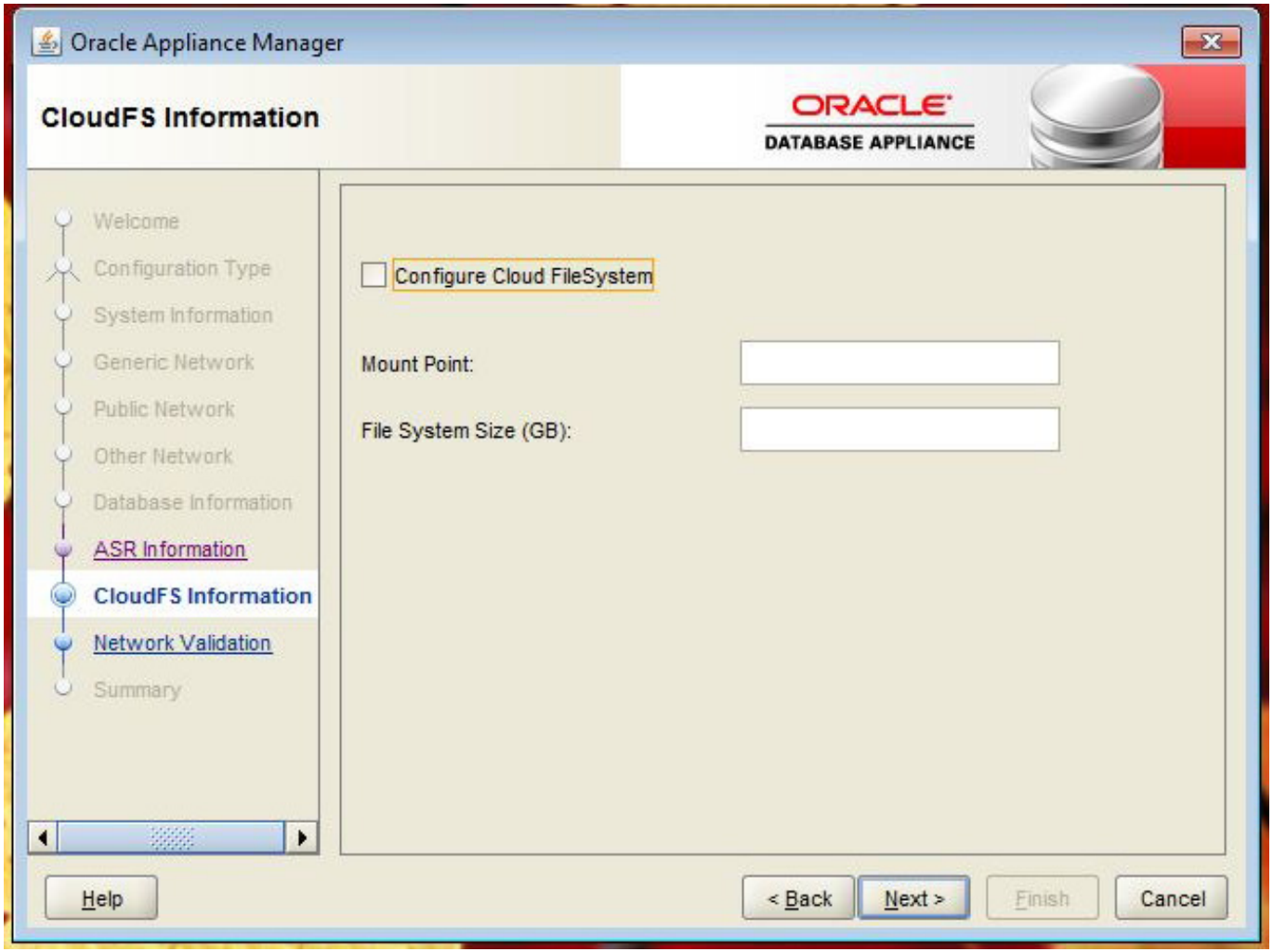
Install Screens: 7



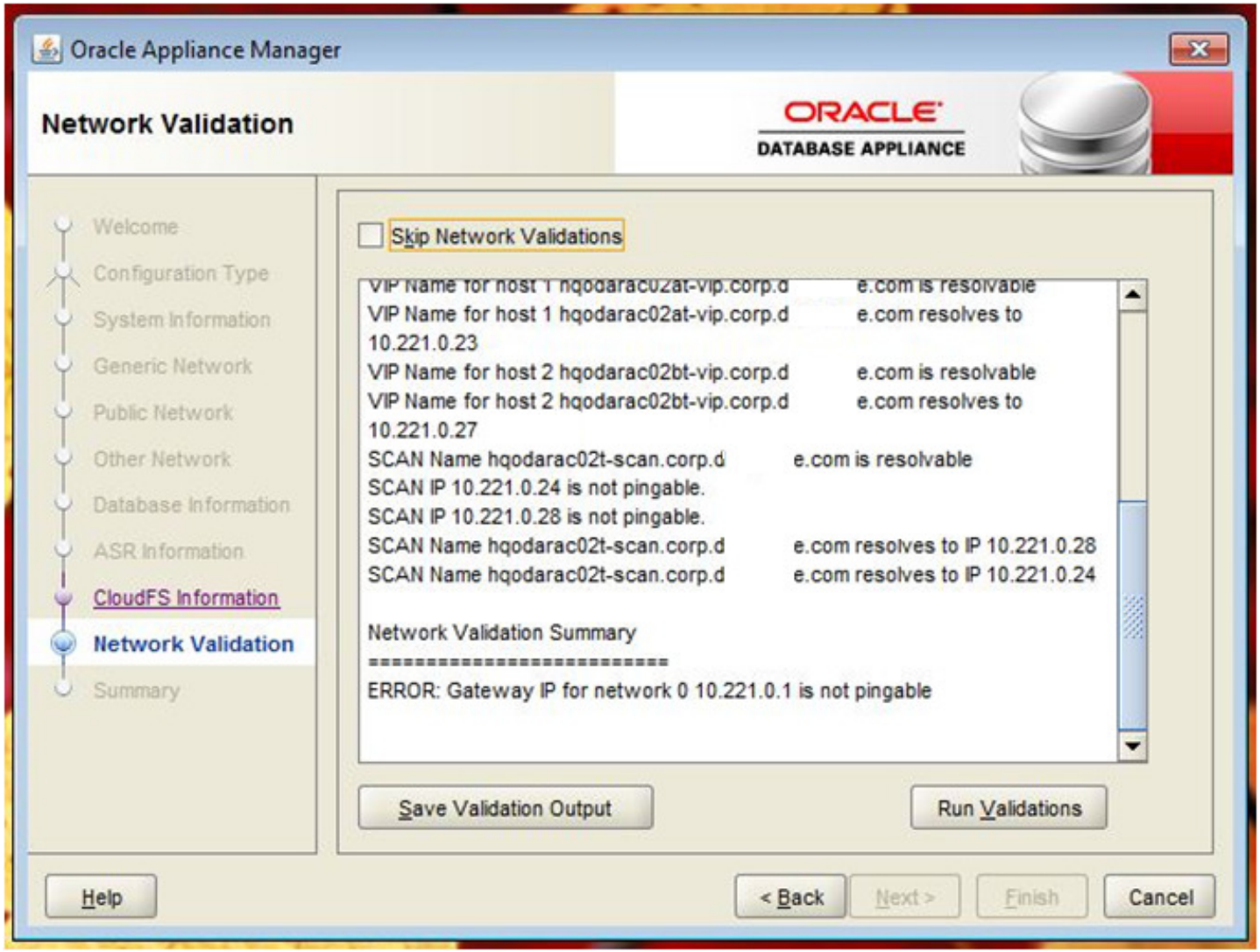
Install Screens: 8



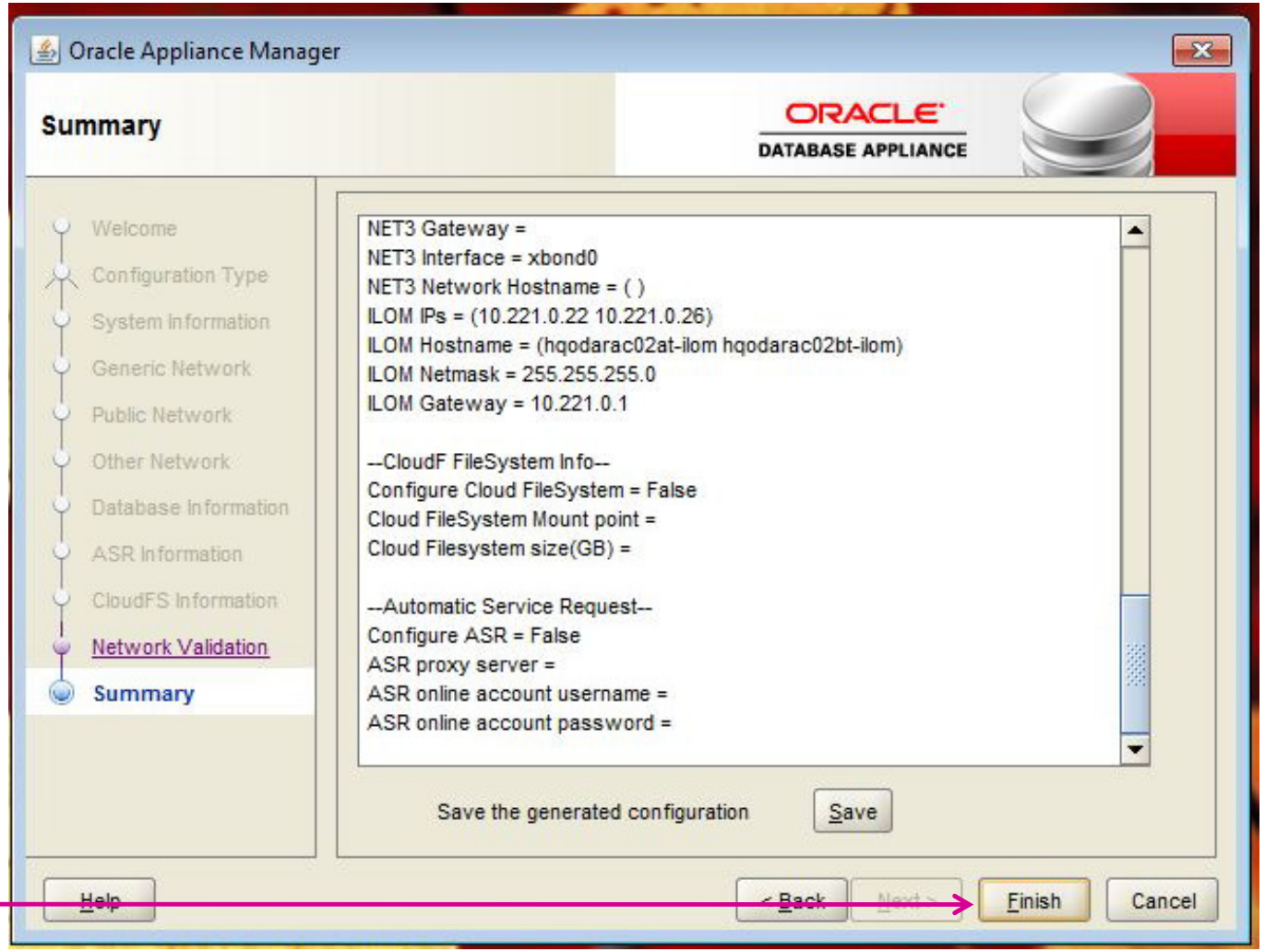
Install Screens: 9



Install Screens: 10

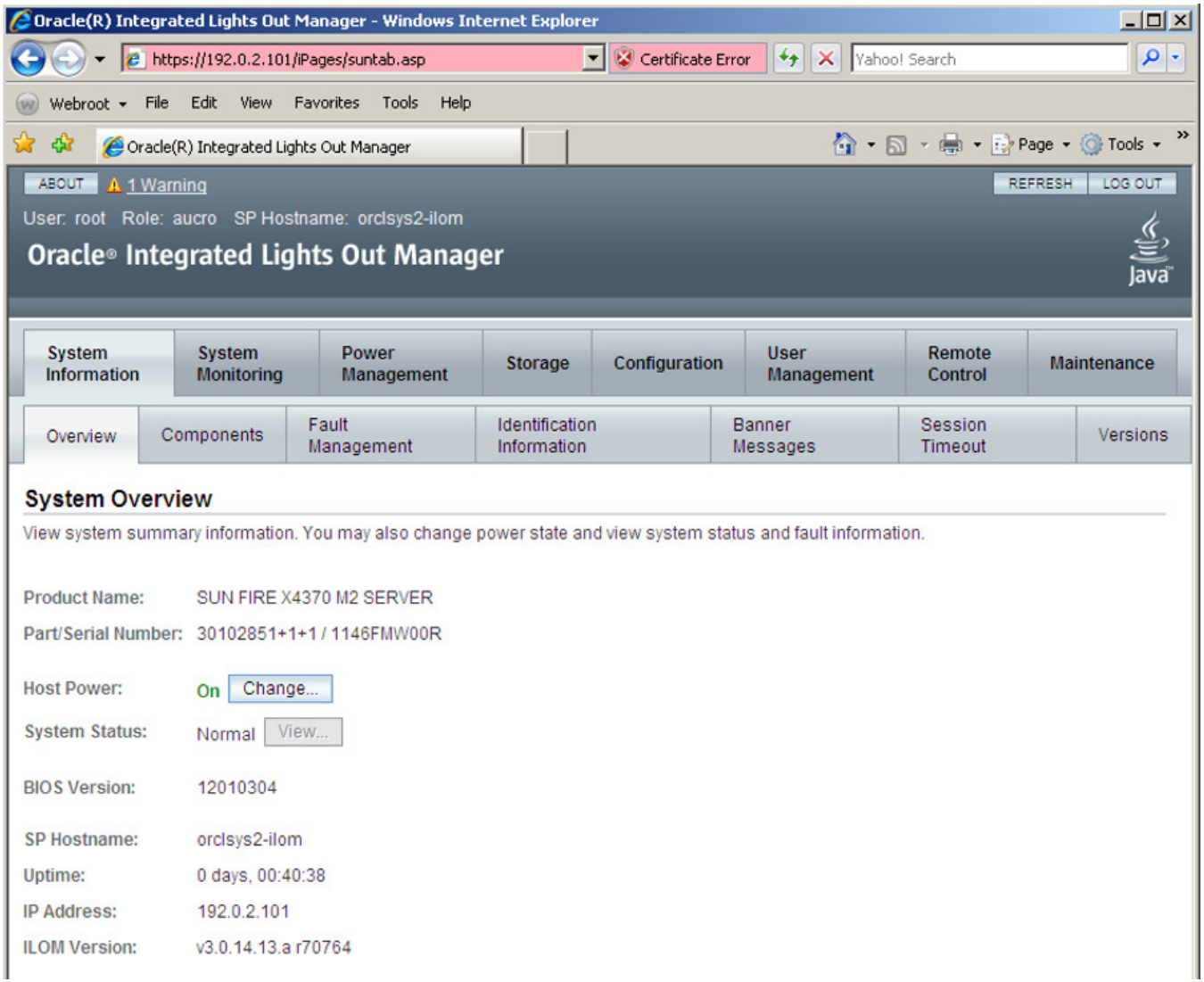


Install Screens: 11

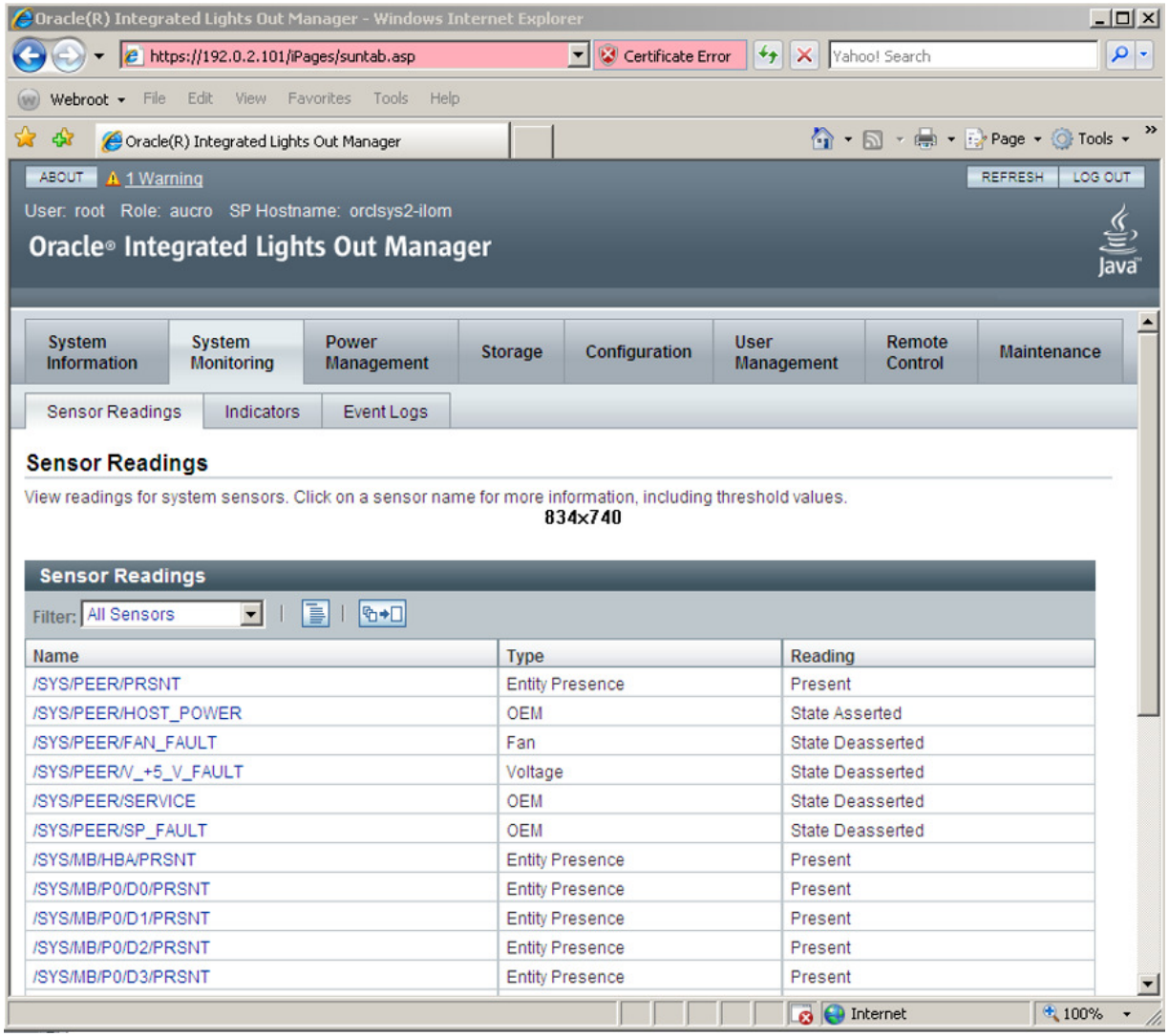


Value Adds

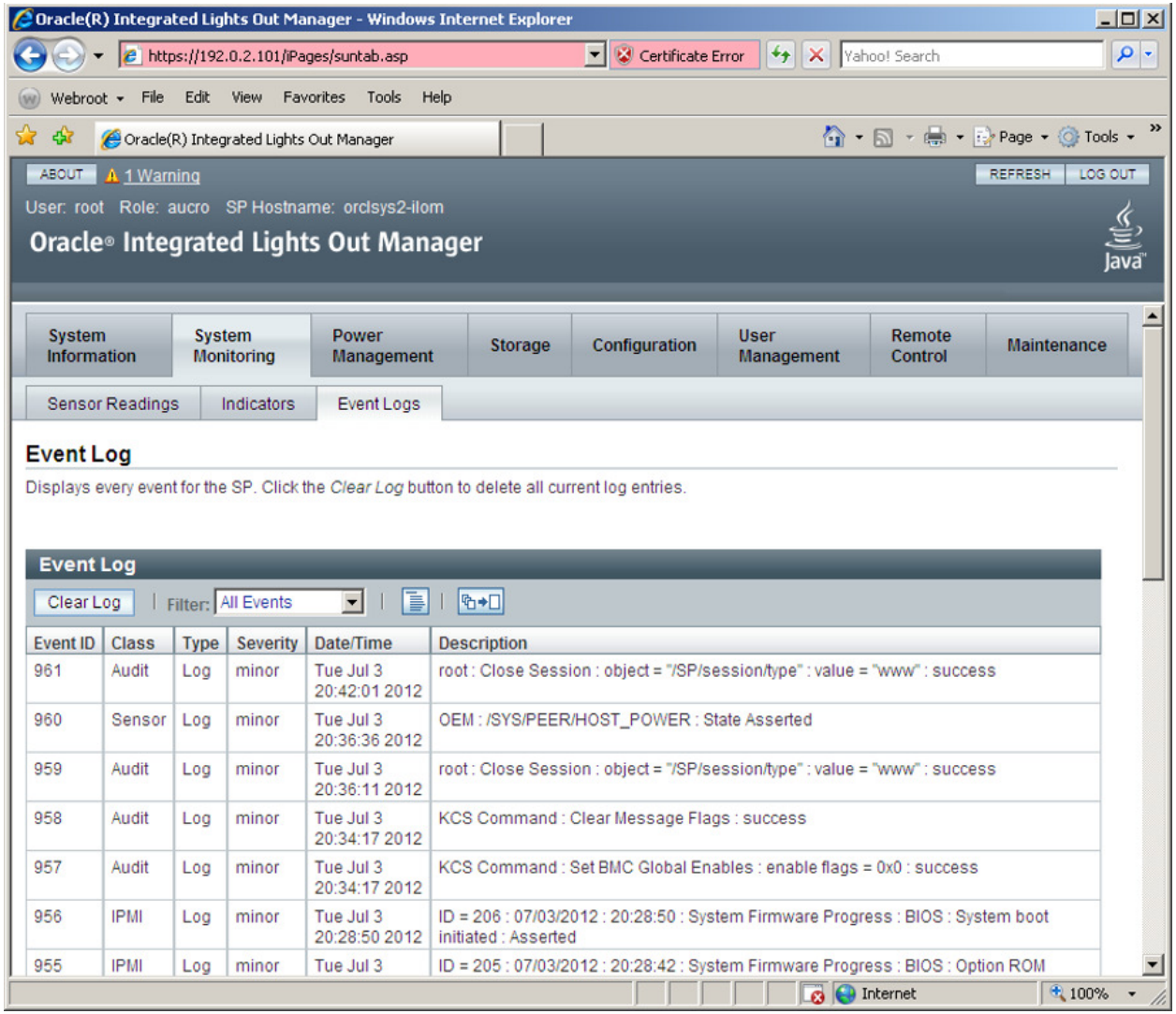
ILOM: System Information: Overview



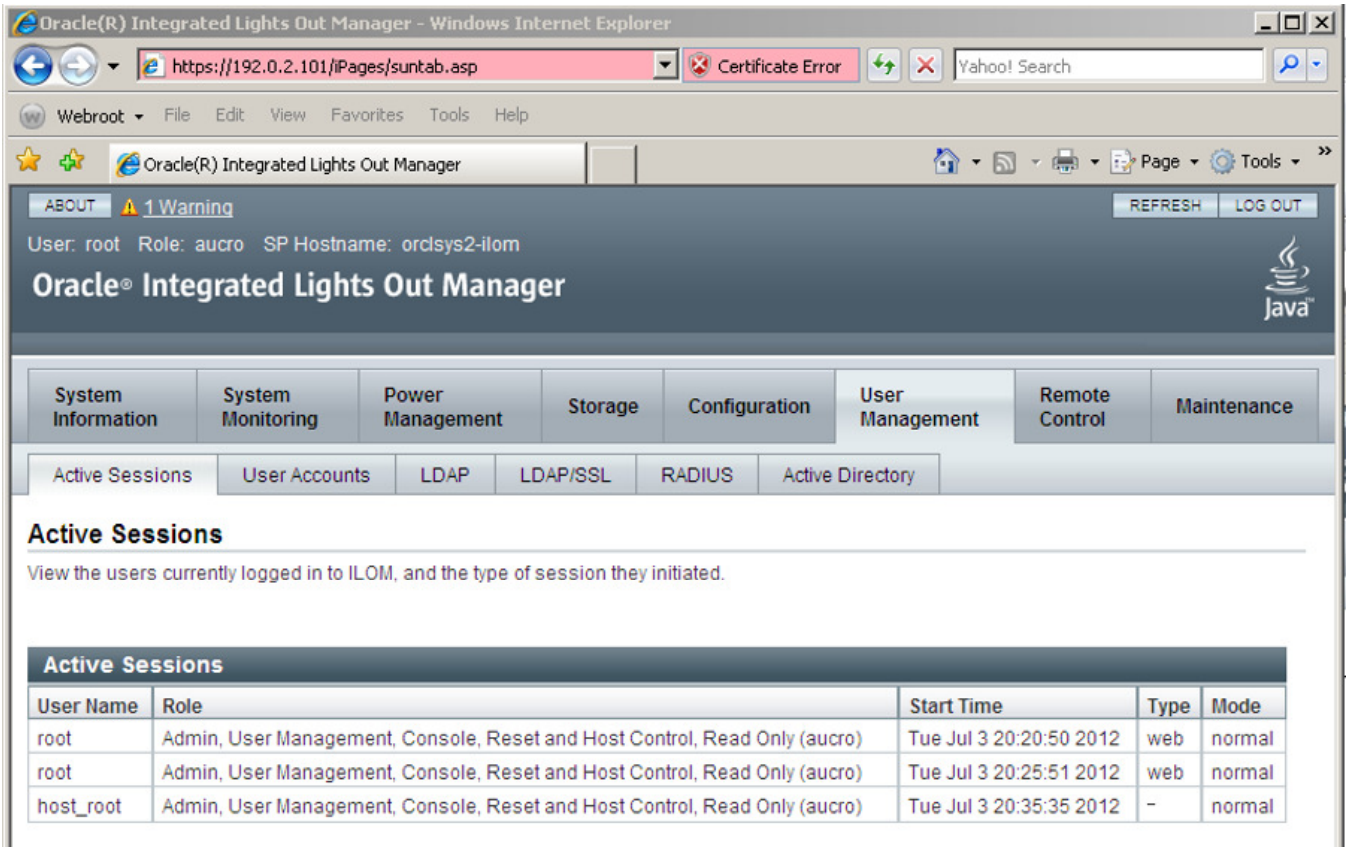
ILOM: System Monitoring: Sensor Readings



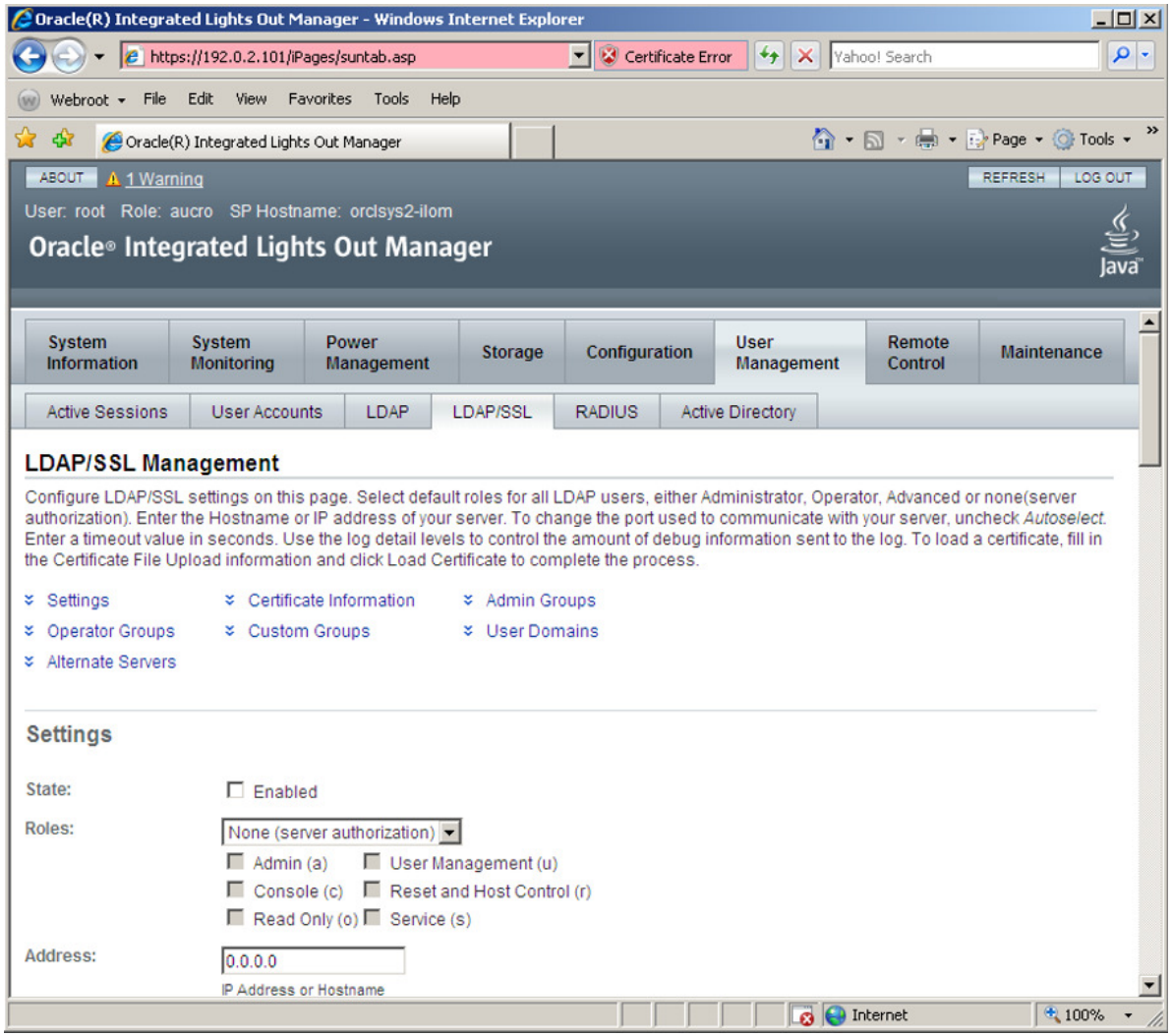
ILOM: System Monitoring: Event Logs



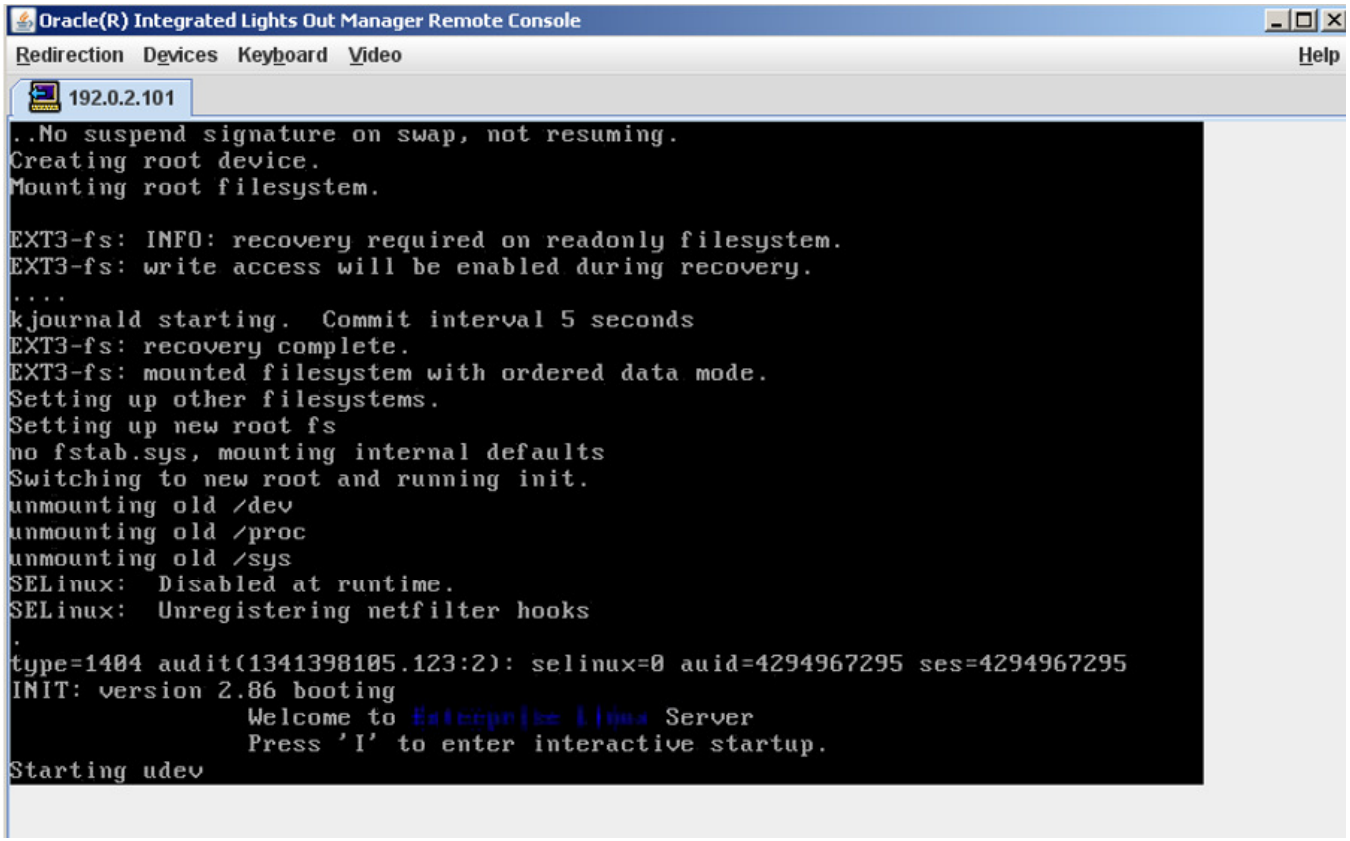
ILOM: User Management: Active Sessions



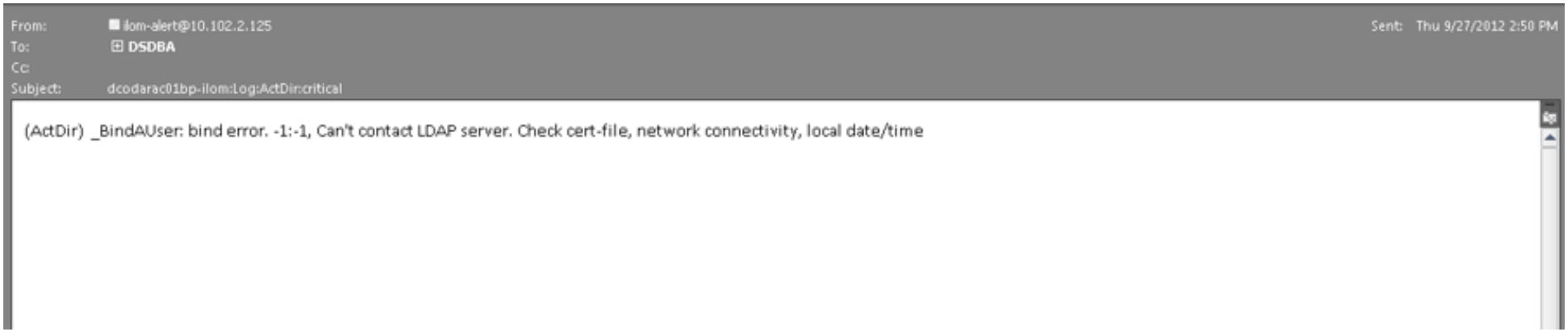
ILOM: User Management: LDAP / SSL



Remote Control: Host Control: Remote Console



ILOM Warning Message



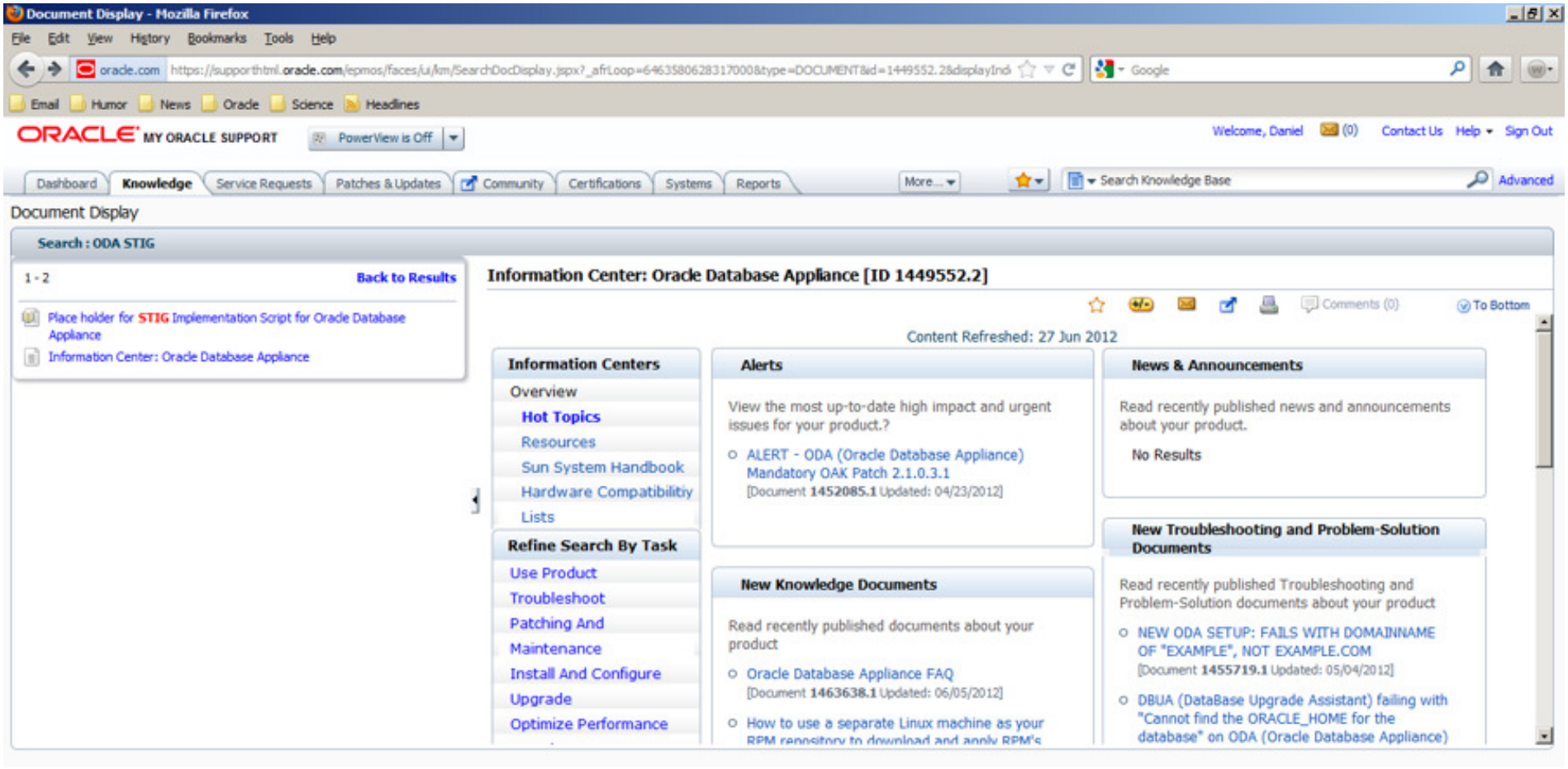
Full Support for High Availability

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

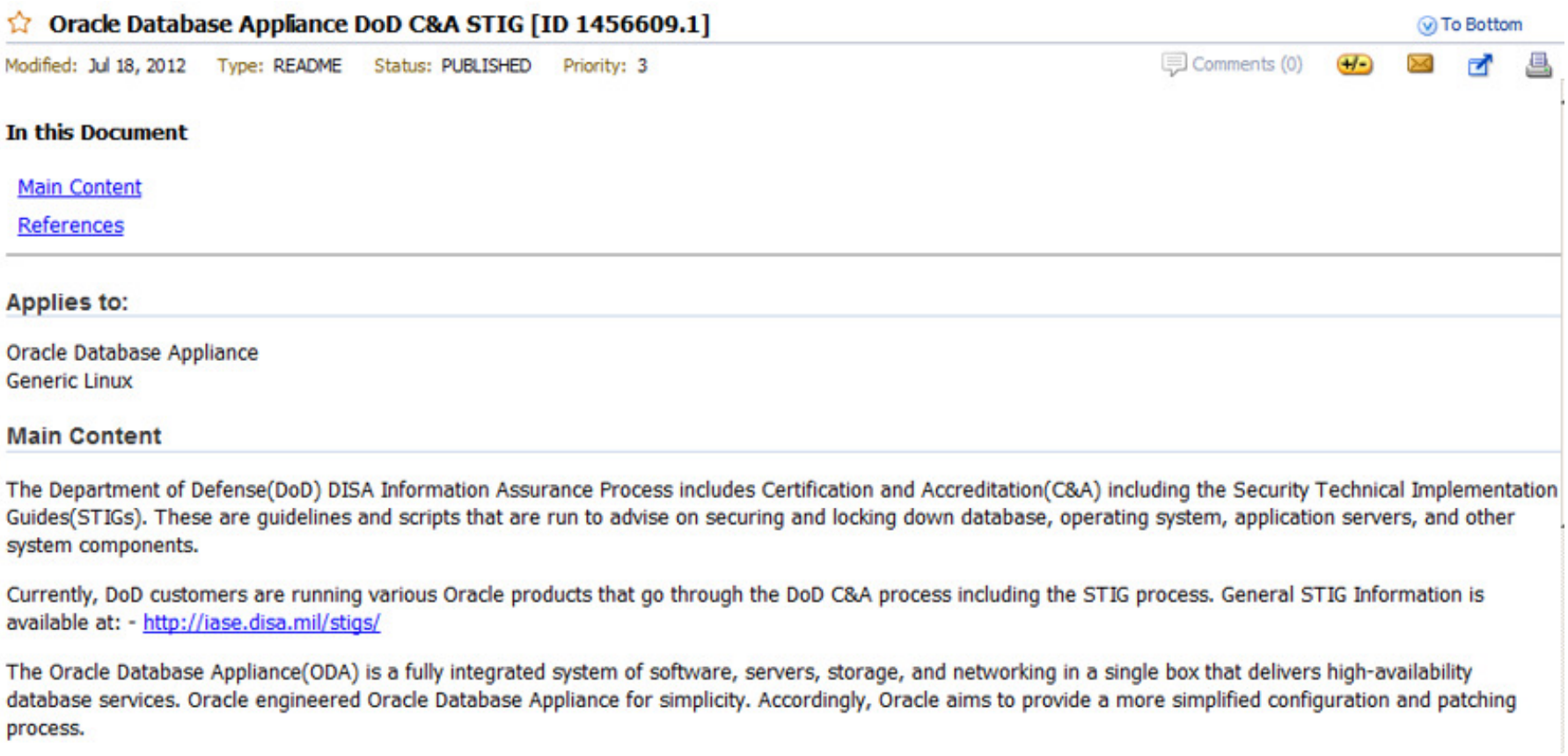
One Button Solutions

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

Support Center



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

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?

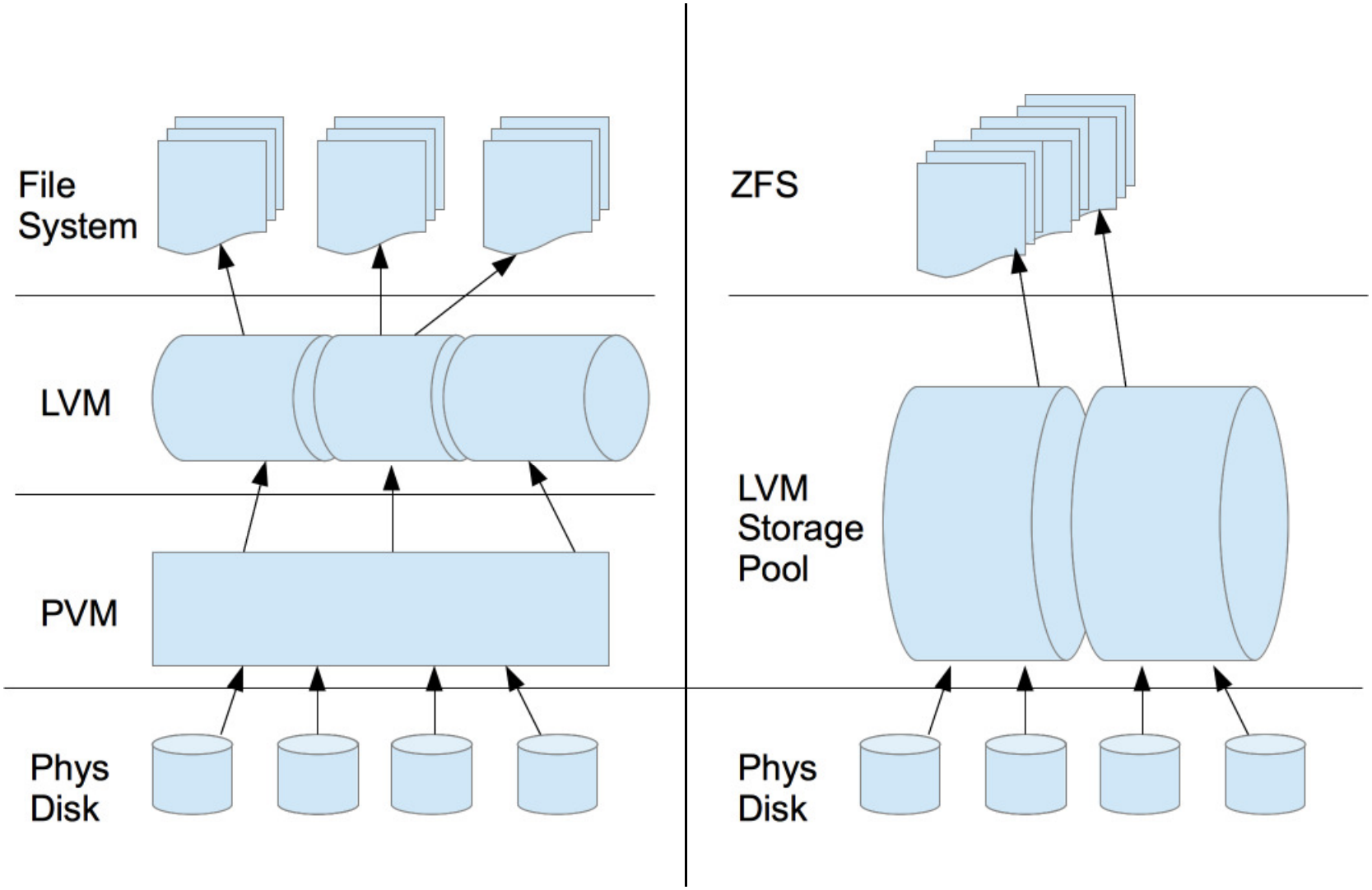
- Local File System
  - May be suitable for some applications,
    - But we have two separate hosts in ODA
  - Standard Linux-oriented
    - Still traditional HBA discussion

- [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 a 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 it's 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

CLARiiON CX4 480 Heat Map

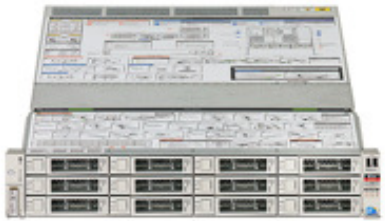
1. Usage over 50% guarantees an component failure will overwhelm remaining resources producing an outage
2. Metrics coded in yellow are warnings
3. Metrics code in red indicate the database has been forced into a wait state while resources are recovered

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

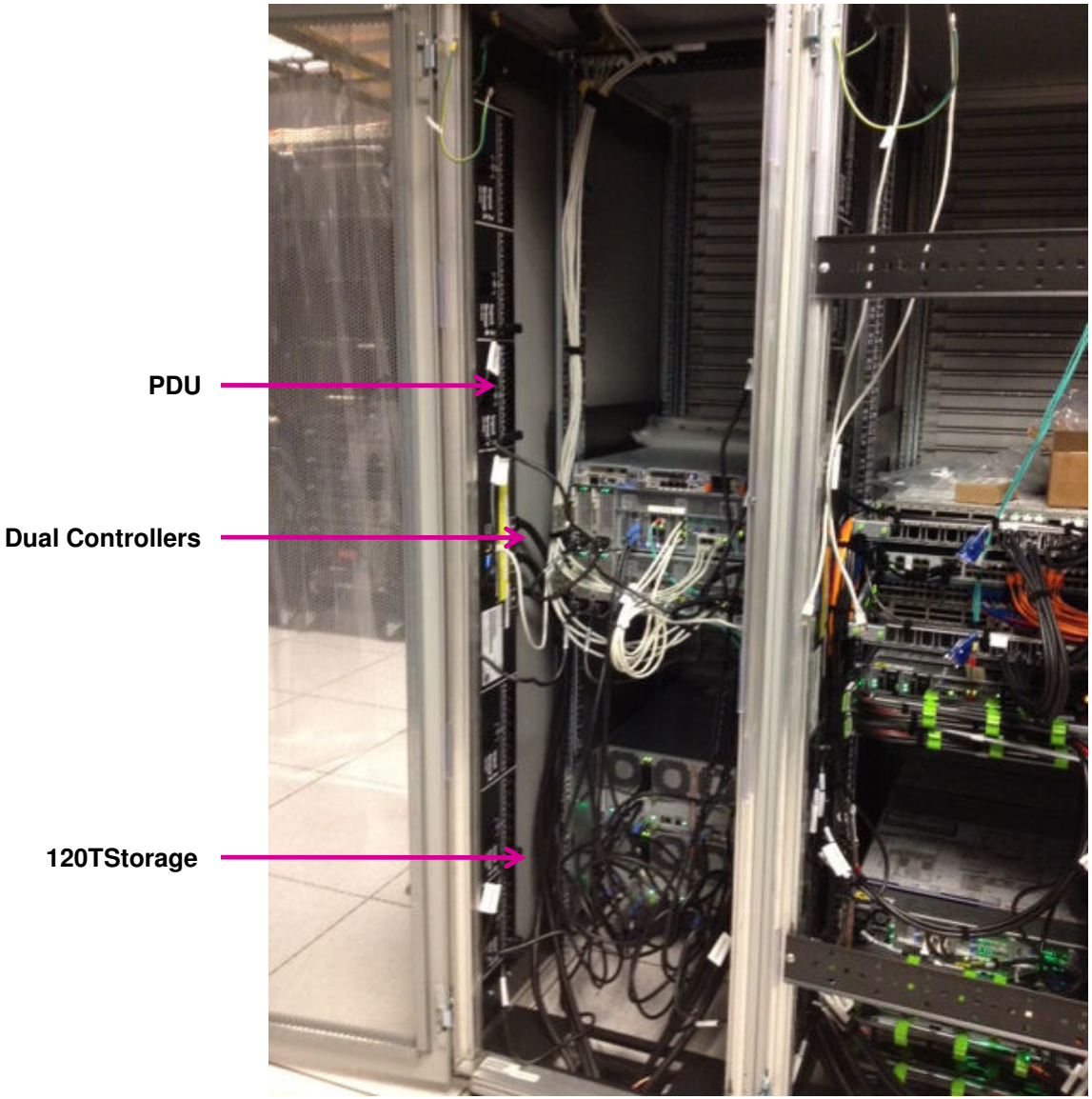
ZFS Specifications

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
Main memory	48 GB	Up to 144 GB per controller	Up to 1 TB per controller
Base Configurations			
Configuration options	• 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)	• 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	• 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

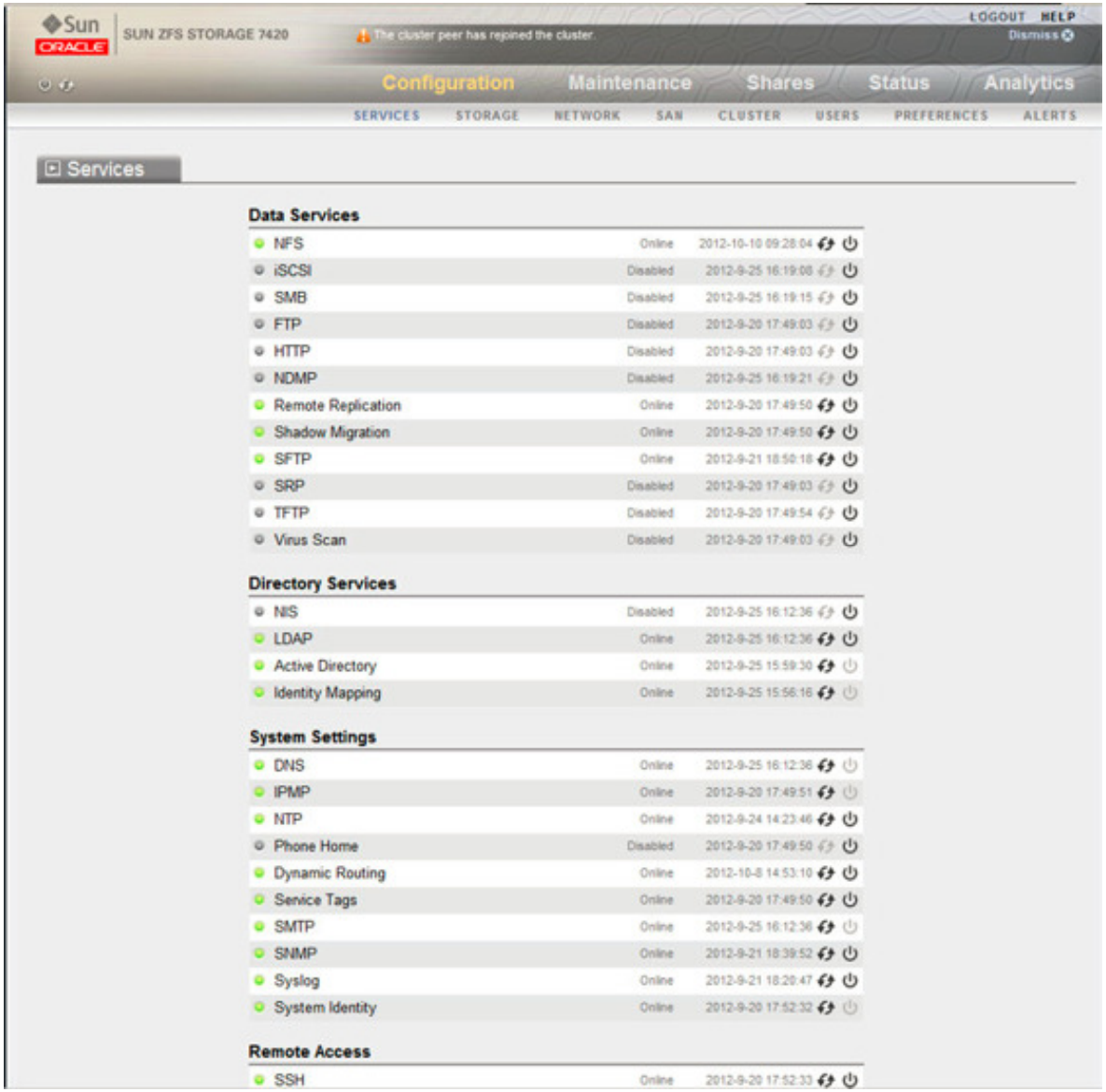
ZFS In The Data Center

ZFS 7420

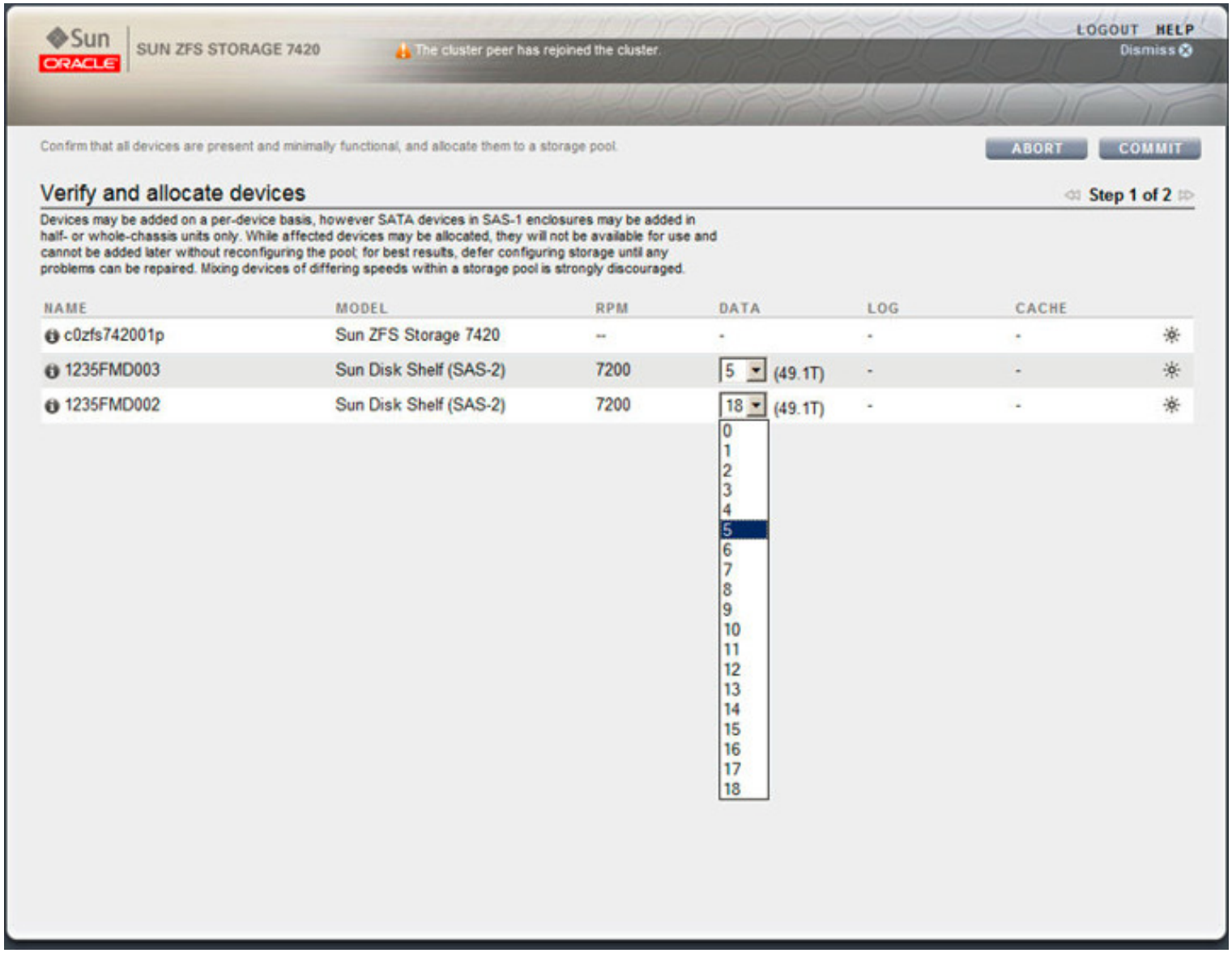
ZFS Internals



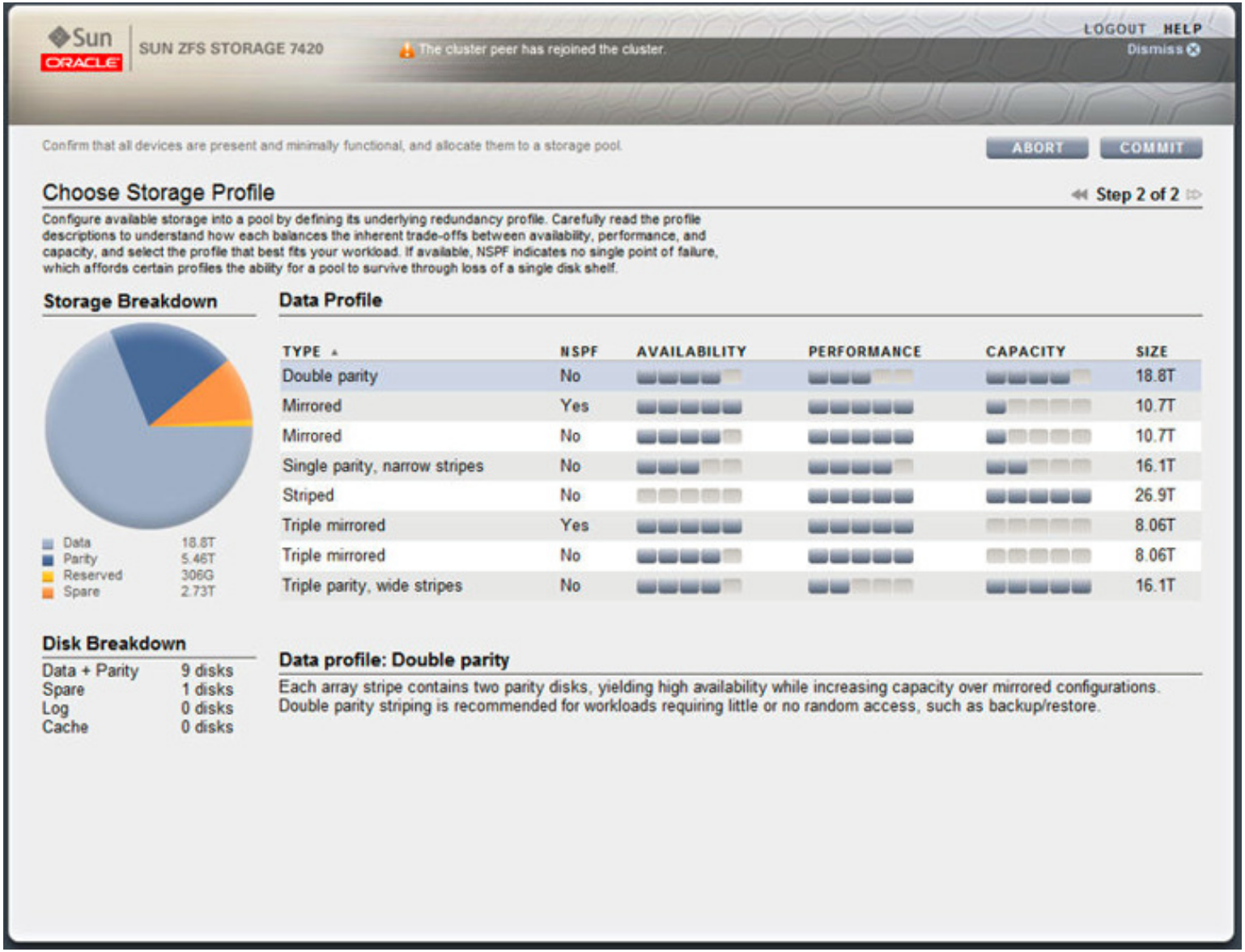
ZFS BUI: Config Services



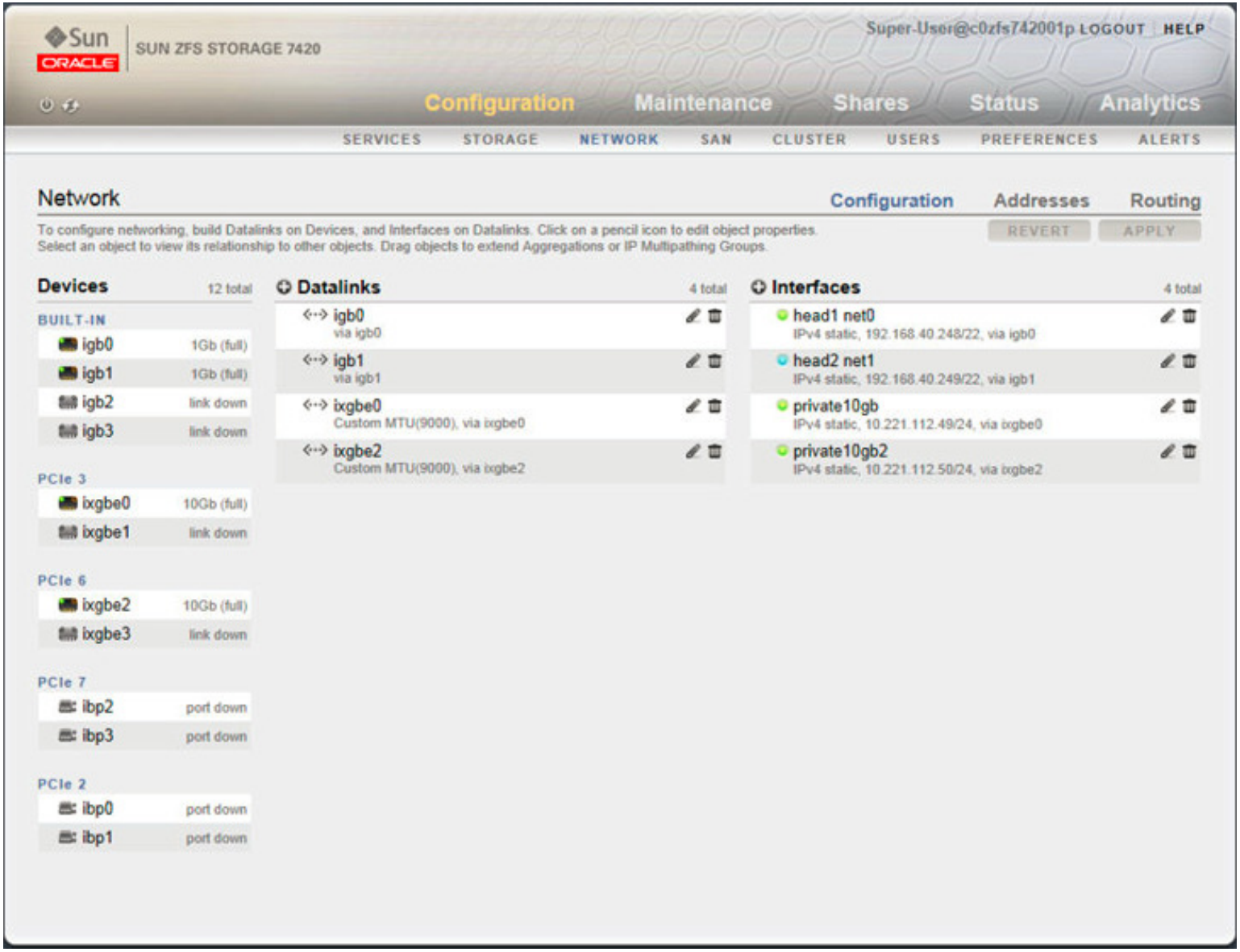
ZFS BUI: Config Storage



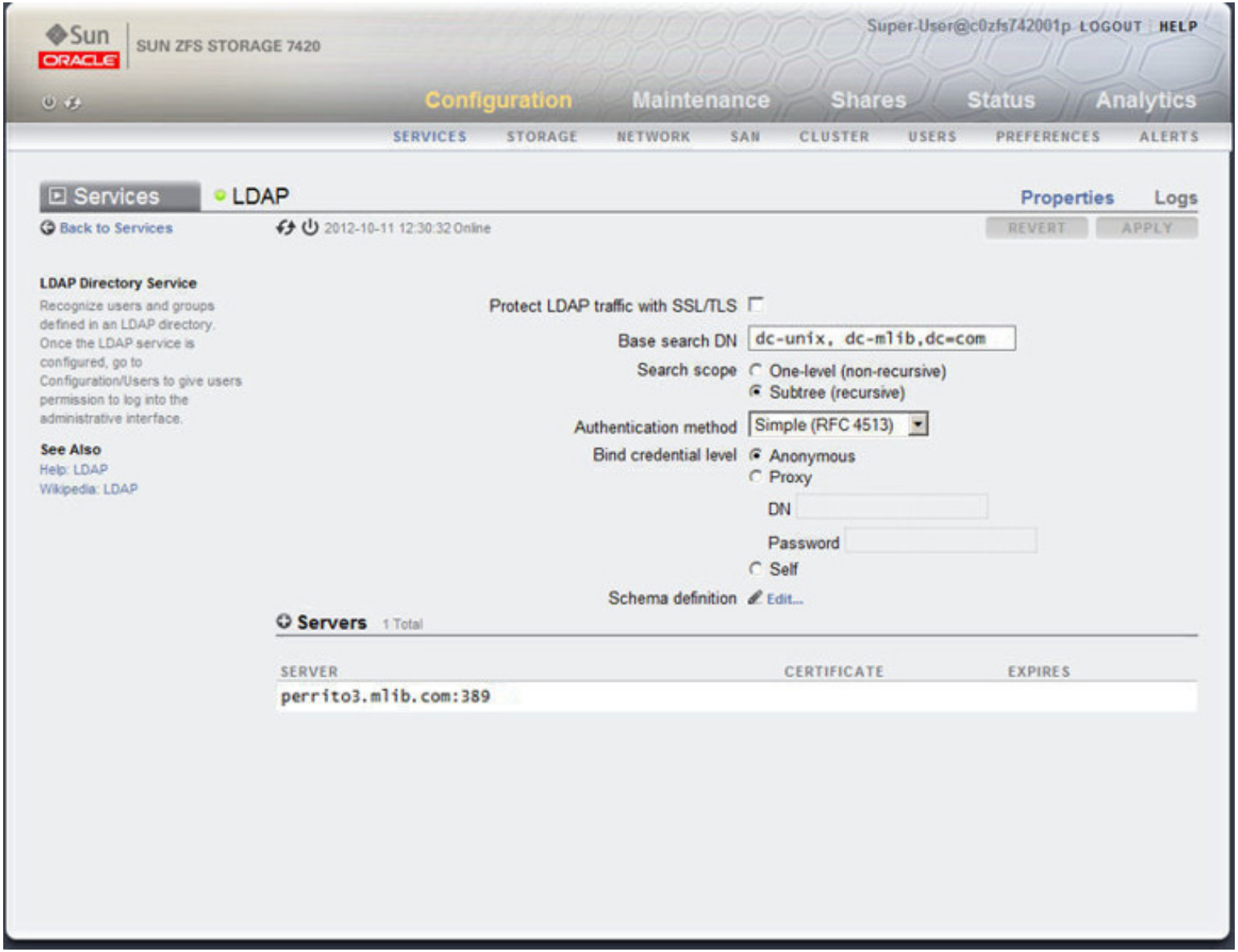
ZFS BUI: Config Storage



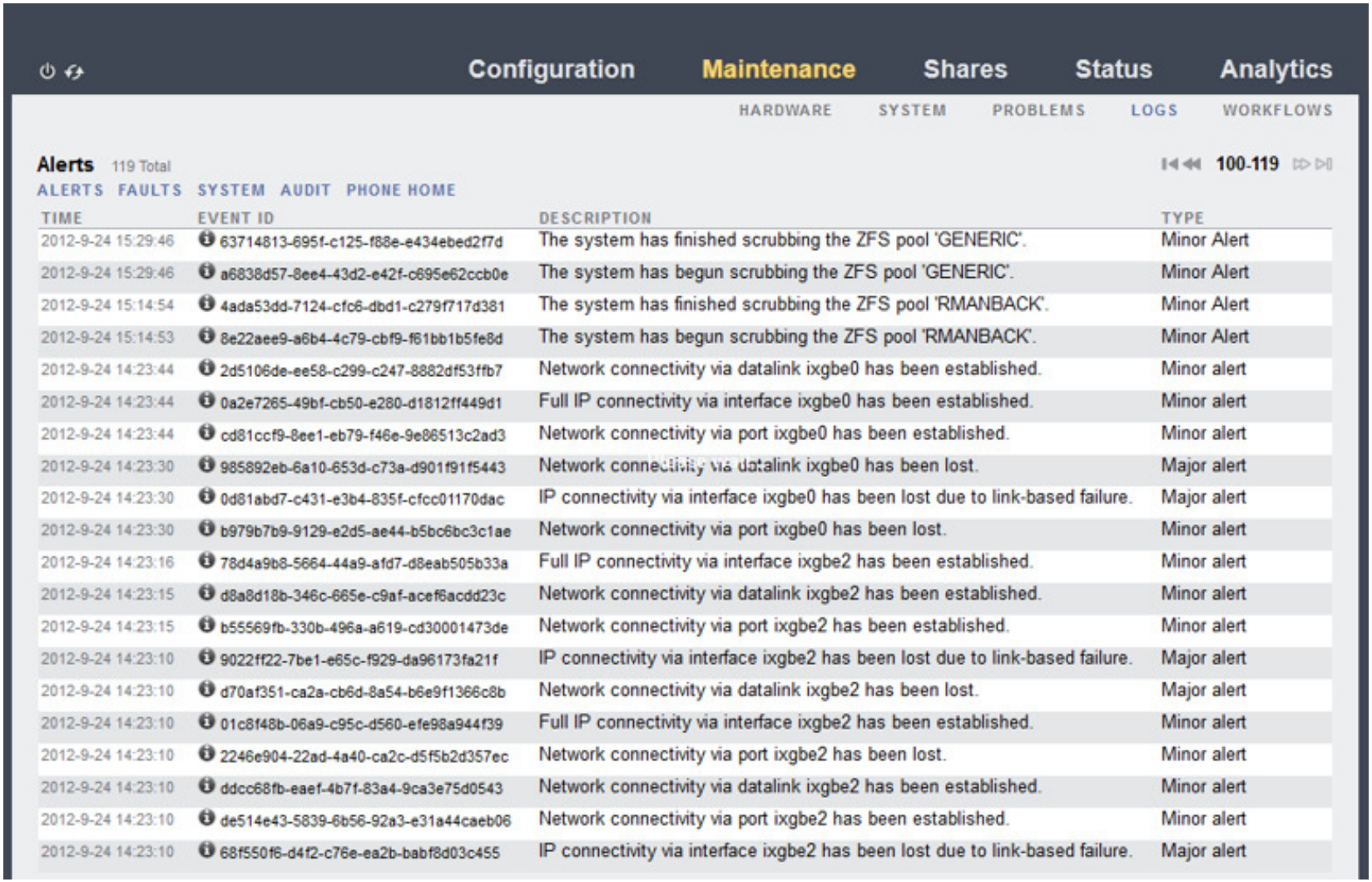
ZFS BUI: Config Network Config



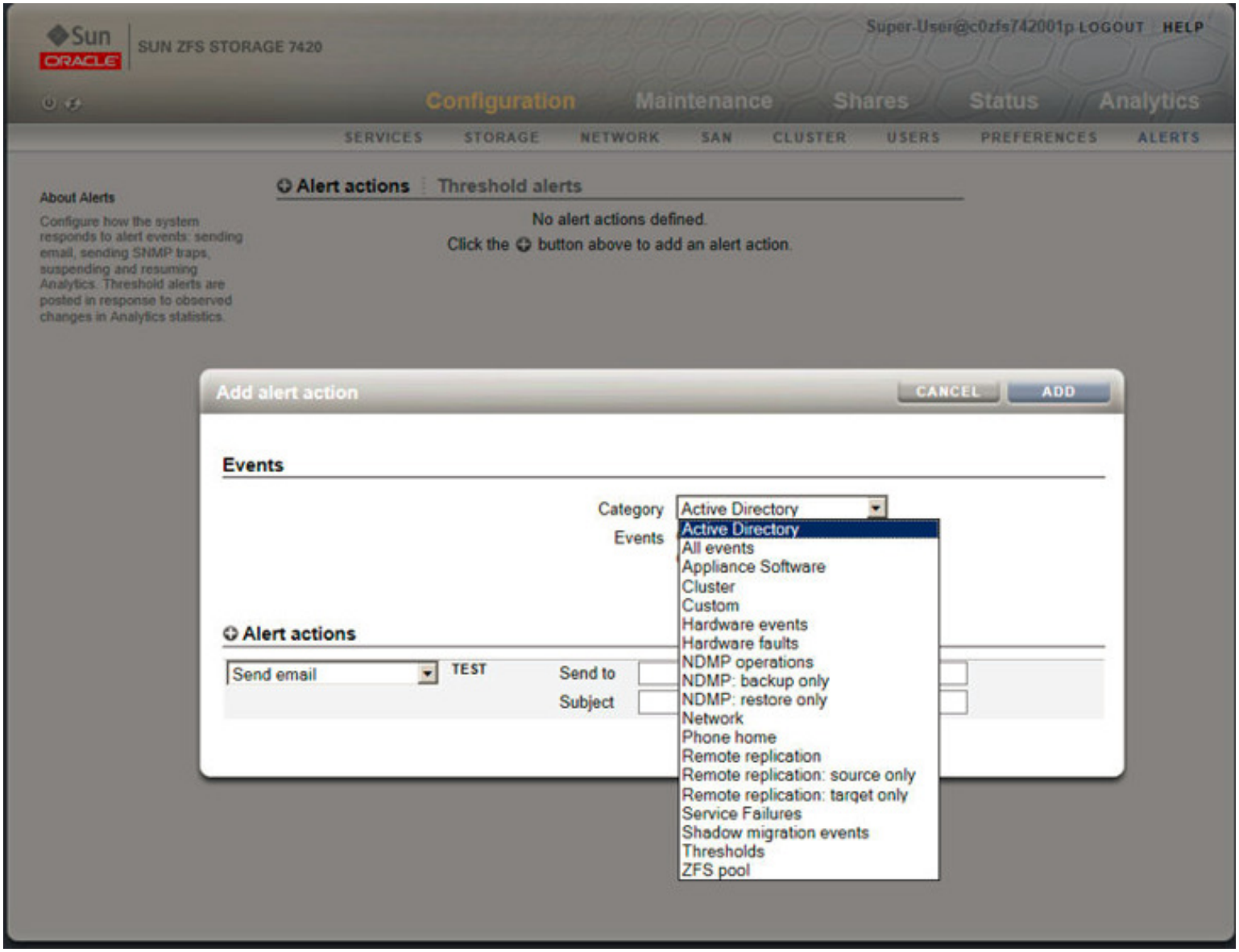
ZFS BUI: Configuration Services LDAP



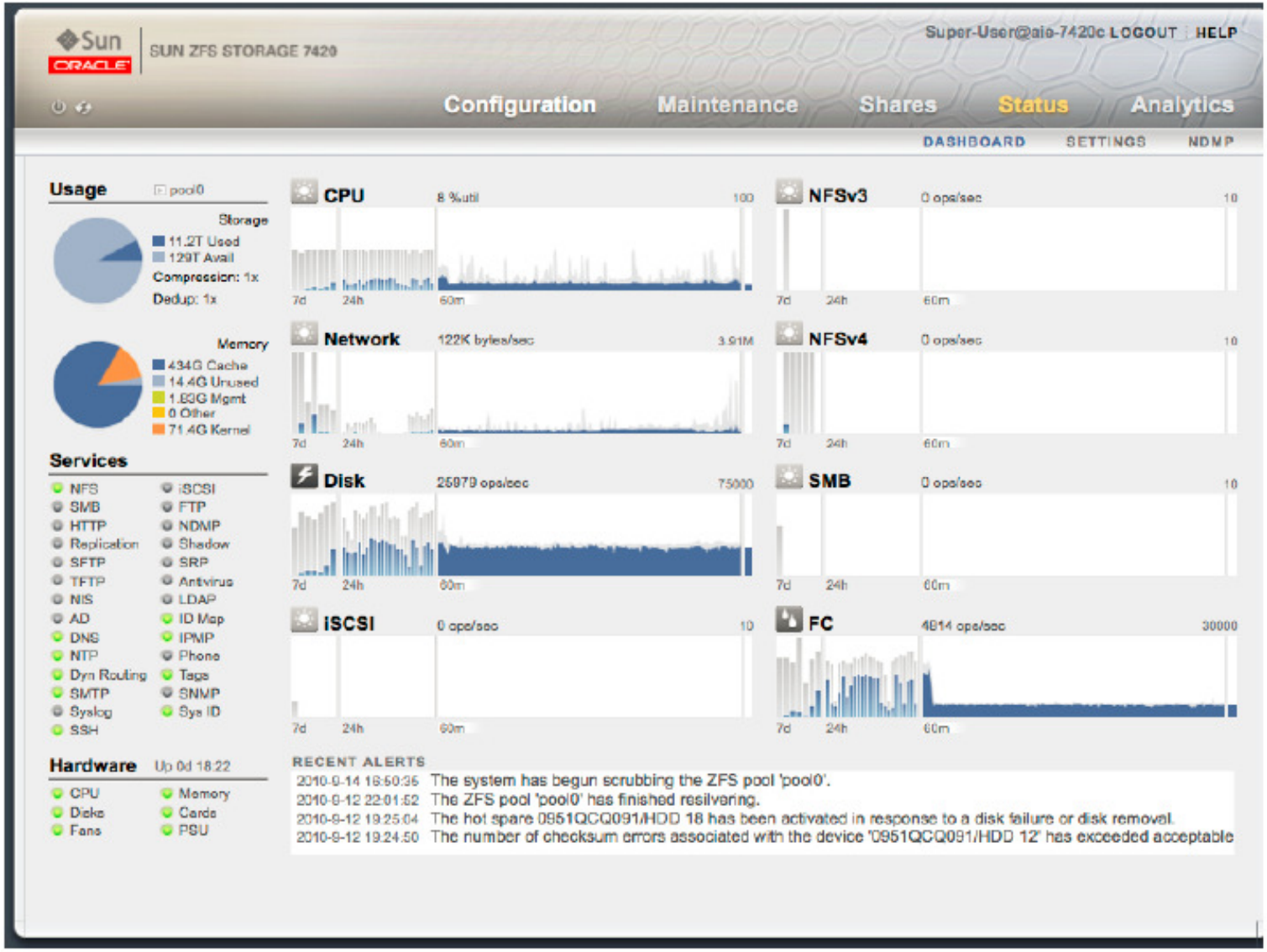
ZFS BUI: Maintenance Logs



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	▪Gives up appliance root password ▪Gives 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

Questions

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

Thank you