
Sun
ORACLE

Sun Oracle Database Machine Hardware Overview



Agenda

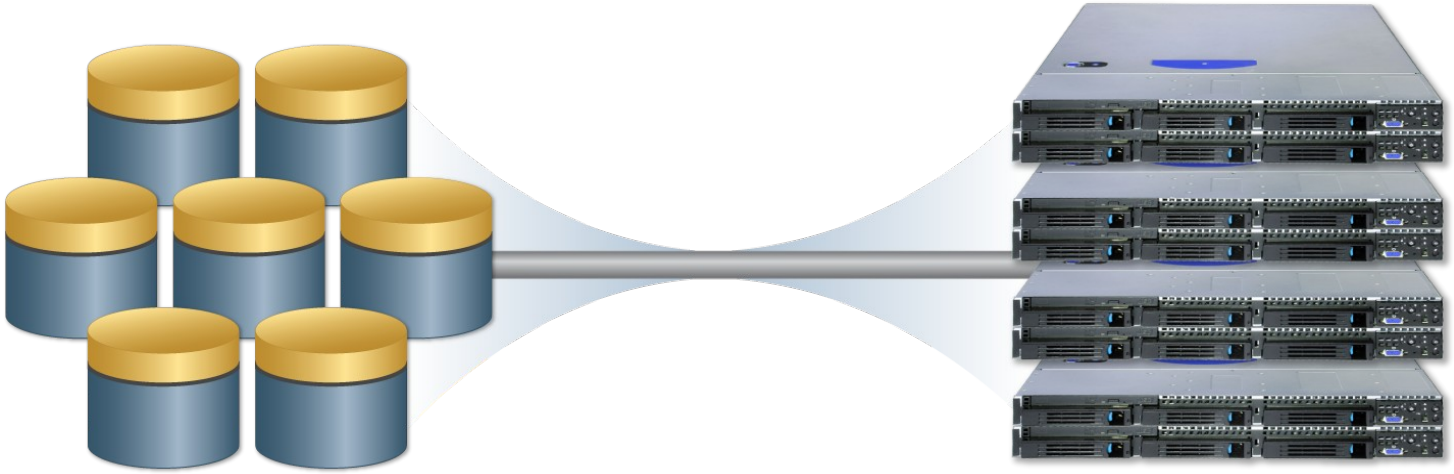
- Introduction
- Configurations
- Architectural Components
- Networks
- Resources



Sun Oracle Database Machine Exadata V2

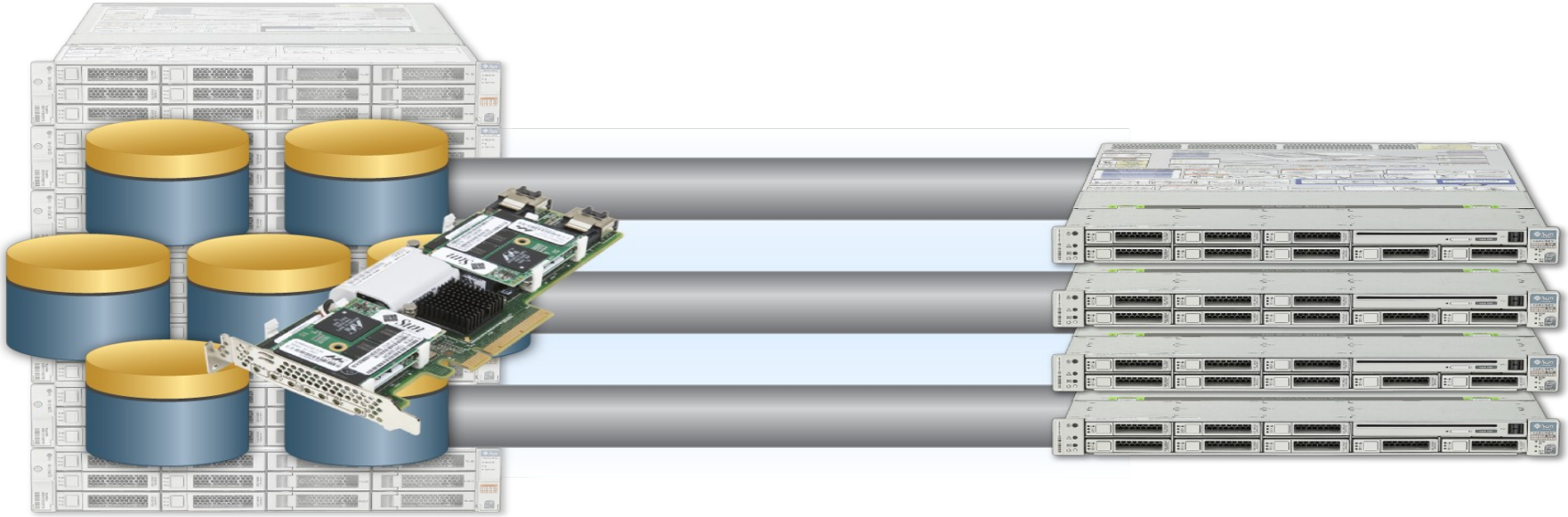
- A complete ***pre-optimized and pre-configured bundle of software, servers, and storage***, built for large-scale data warehousing and online transaction processing (OLTP)
- Hardware by Sun, Software by Oracle

Current Database Machine Issues



- Current database deployments often have bottlenecks limiting the movement of data from disks to servers
  - Storage Array internal bottlenecks on processors and Fibre Channel Architecture
  - Limited Fibre Channel host bus adapters in servers
  - Under configured and complex SANs
- Pipes between disks and servers are 10x to 100x too slow for data size

Sun Oracle Database Machine Solves the Bottleneck



- Adding **more pipes** – Massively parallel architecture
- Make the **pipes wider** – 5X faster than conventional storage
- Ship **less data** through the pipes – Process data in storage

Configurations

Configurations



Full Rack



Half Rack



Quarter Rack

Sun Oracle Database Machine Full Rack

- 8 Sun Fire X4170 Oracle Database servers
- 14 Exadata Storage Servers (Cells)
- 3 Sun Datacenter InfiniBand Switch 36
- 1 “Admin” Cisco Ethernet switch
- Keyboard, Video, Mouse (KVM) hardware
- Redundant Power Distribution Units (PDUs)



Sun Oracle Database Machine Half Rack

- 4 Sun Fire X4170 Oracle Database servers
- 7 Exadata Storage Servers (Cells)
- 2 Sun Datacenter InfiniBand Switch 36
- 1 “Admin” Cisco Ethernet switch
- Keyboard, Video, Mouse (KVM) hardware
- Redundant Power Distribution Units (PDUs)



Sun Oracle Database Machine Quarter Rack

- 2 Sun Fire X4170 Oracle Database servers
- 3 Exadata Storage Servers (Cells)
- 2 Sun Datacenter InfiniBand Switch 36
- 1 “Admin” Cisco Ethernet switch
- Keyboard, Video, Mouse (KVM) hardware
- Redundant Power Distribution Units (PDUs)



Sun Oracle Database Machine

Scalability: Start Small and Grow



Quarter Rack

Half Rack

Full Rack

- Upgrades can be performed with no scheduled downtime to the existing system

Scale to Multiple Racks



- Scale up to 8 Full Racks by using more InfiniBand cables
- Scale to more Full Racks by using external InfiniBand switches

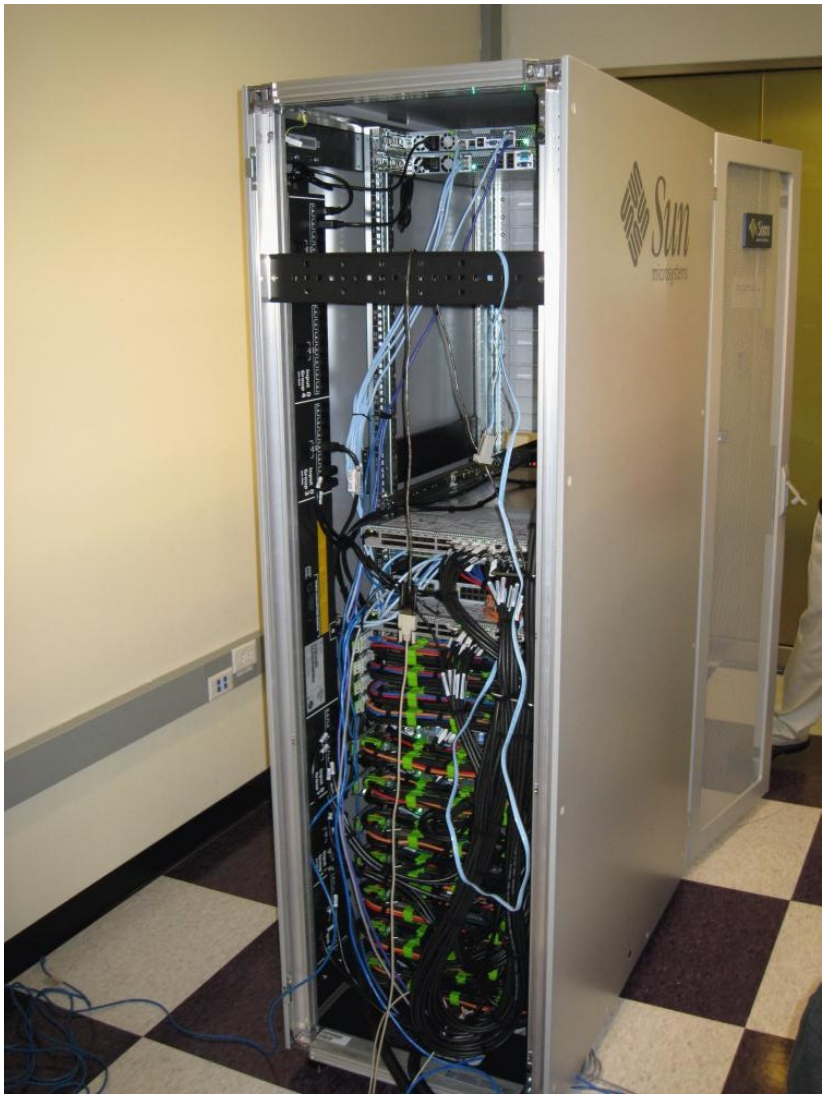
Sun Oracle Database Machine Configuration Big Rules

- No hardware changes allowed
  - No re-racking into customer supplied racks
  - Additional Storage Servers (Cells) should be installed into a separate adjacent rack
  - The only hardware modification that is allowed is to the Cisco 4948 Gigabit Ethernet switch
- No software changes allowed

Sun Oracle Database Machine (Front)



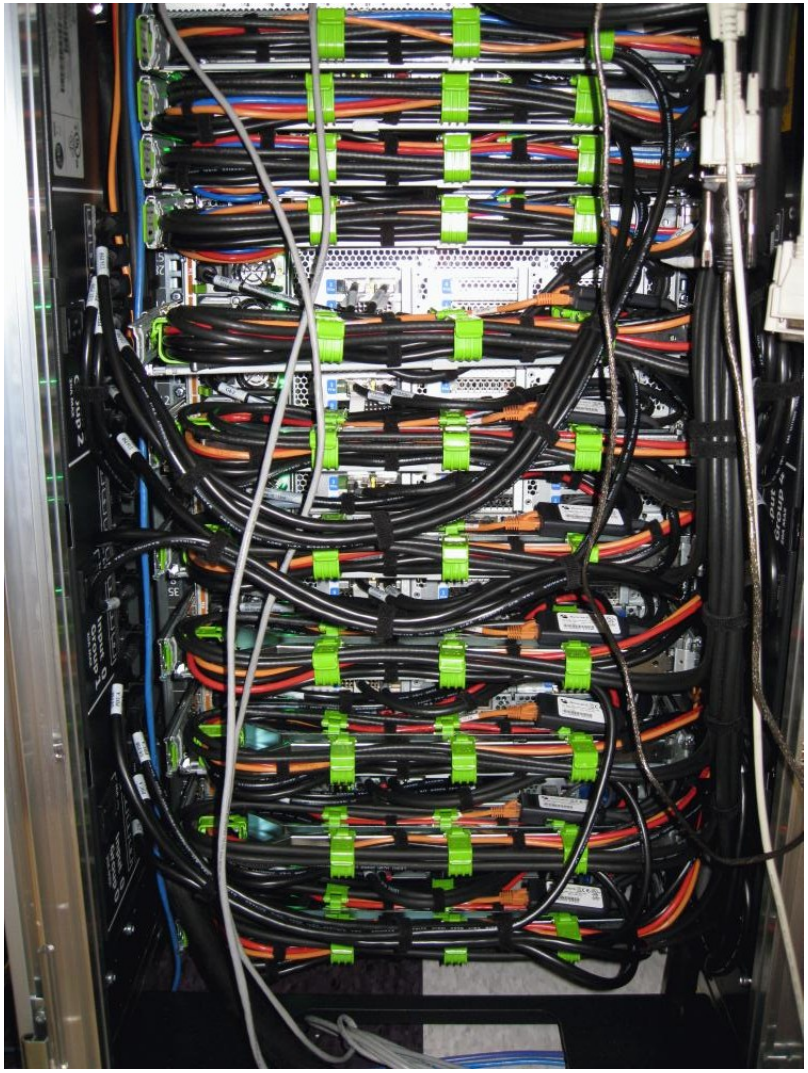
Sun Oracle Database Machine (Rear)



Sun Oracle Database Machine (Rear)



Sun Oracle Database Machine (Rear)



Architectural Components

Sun Rack II 1242E

- 42 RU spaces available
- 1200mm depth
- 600mm wide, 19-inch rails
- Enterprise Packaging suitable for shipping fully loaded
- Zero-RU PDUs



Database Server – SF X4170



Up to eight Database Servers per Exadata version 2 rack

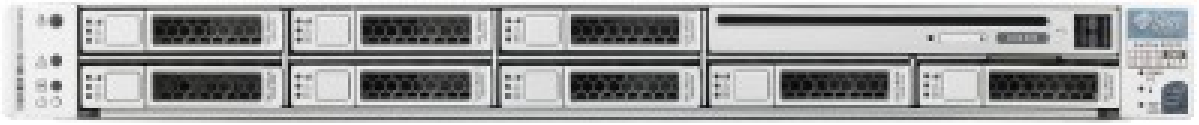
Oracle Enterprise Linux
Oracle Real Application Clusters

Maximum Reliability, Availability, and Serviceability

Ease of Deployment and Manageability

- **Compute**
  - CPU: 2x Intel Xeon E5540, 4C, 2.53 GHz
  - RAM: 72 GB (18x DDR3-1333 4GB ECC Registered DIMMs)
- **Internal Storage & I/O**
  - 4x 146GB 10K rpm SAS HDDs
  - 1x SAS2 RAID HBA w/ 512MB cache Installed in slot **PCIE-0**
  - 1x Sun QDR (40Gb) InfiniBand HCA Installed in slot **PCIE-2**
- **Availability**
  - Hot-swap disks
  - Hot-swap redundant 750W PSUs
  - Hot-swap redundant fans
- **Remote Management**

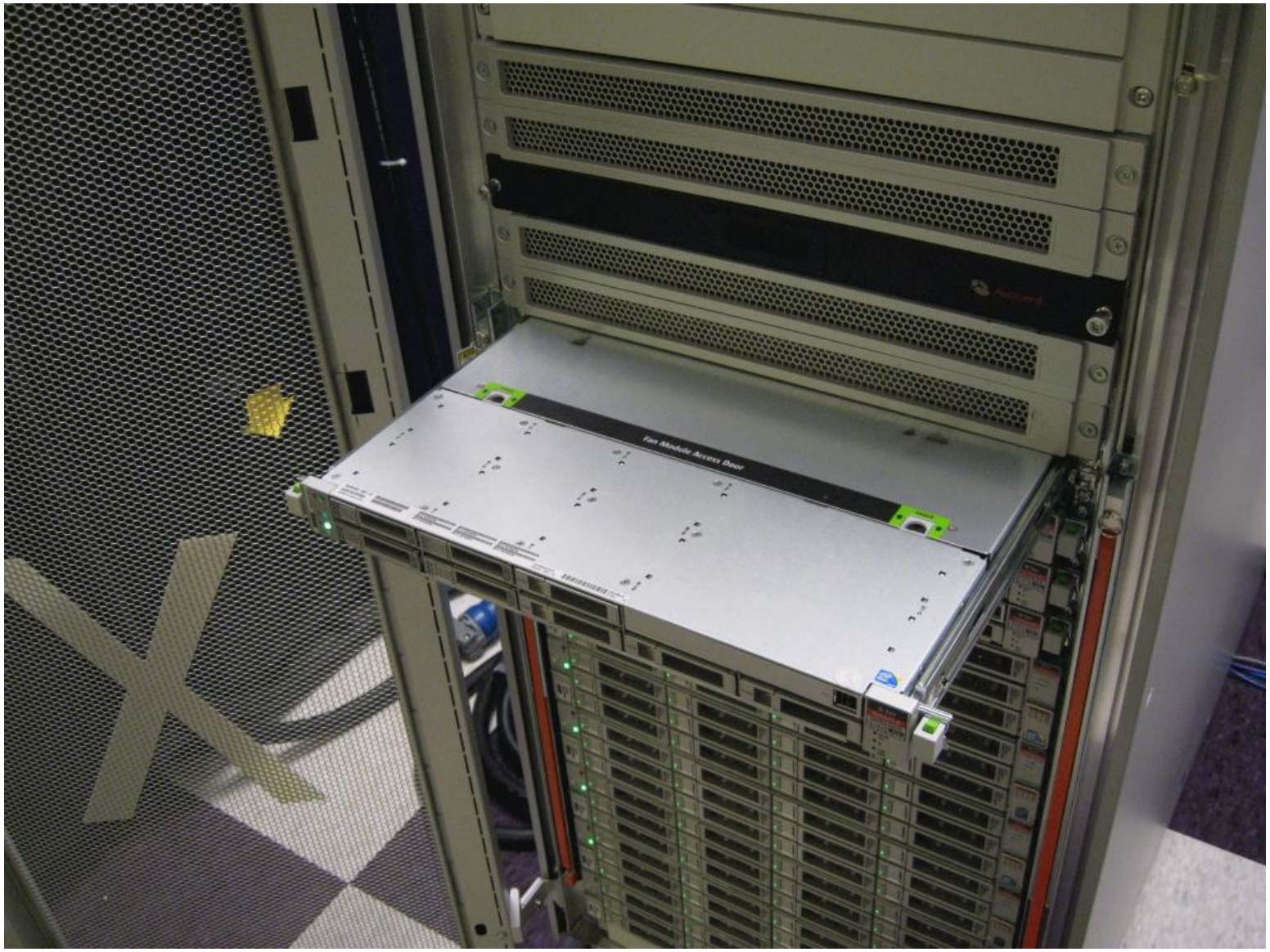
Database Server – SF X4170



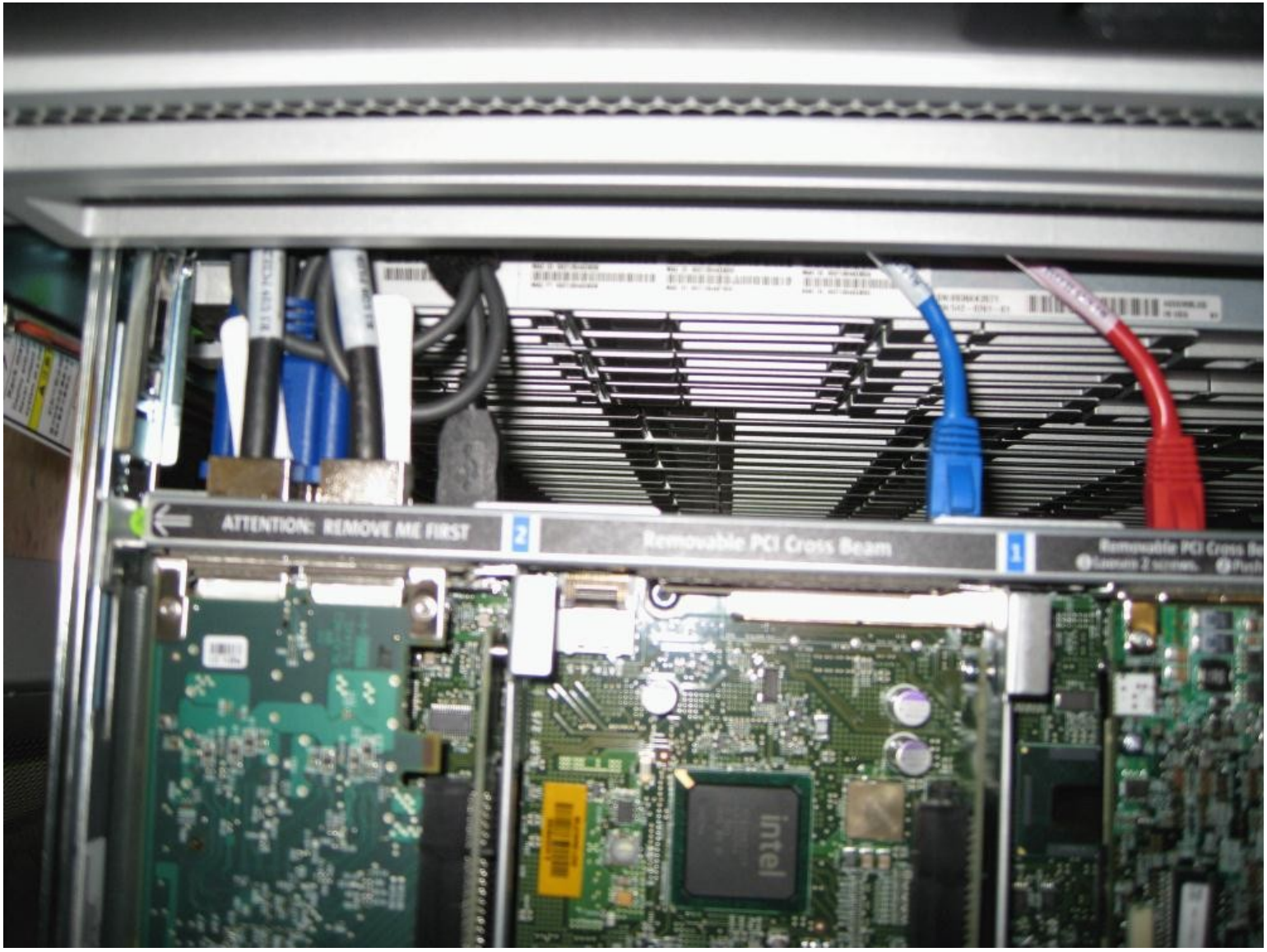
Sun SAS RAID PCIe 2.0
8 port HBA
Installed in slot PCIE-0

Sun Dual port QDR InfiniBand
PCIe HCA
Installed in slot PCIE-2

Database Server – SF X4170



Database Server – SF X4170



Sun Exadata Storage Server – SF X4275



Up to fourteen Storage Servers per Exadata version 2 rack

Oracle Enterprise Linux
Oracle Exadata Storage Server Software
Additional Drivers, Utilities

Building Block for Massively Parallel Storage Grid

- **Compute**
  - CPU: 2x Intel Xeon E5540, 4C, 2.53 GHz
  - RAM: 24 GB (6x DDR3-1333 4GB ECC Registered DIMMs)
- **Internal Storage & I/O**
  - 12x 600GB 15K rpm SAS HDDs or 2TB 7200 rpm SATA HDDs
  - 1x SAS2 RAID HBA w/ 512MB cache Installed in slot **PCIE-0**
  - 1x Sun QDR (40Gb) InfiniBand HCA Installed in slot **PCIE-3**
  - 4x Sun Smart Flash Accelerator F20 PCIe cards with 96GB Flash each
- **Availability**
  - Hot-swap disks
  - Hot-swap redundant 1050W PSUs
  - Hot-swap redundant fans
- **Remote Management**

Sun Exadata Storage Server – SF X4275



HDD Configuration
12x 600GB15K,SAS or
12x 2TB 7200K, SATA

SAS cable for
HBA is integrated

Sun SAS RAID PCIe 2.0
8 port HBA
Installed in slot PCIE-0

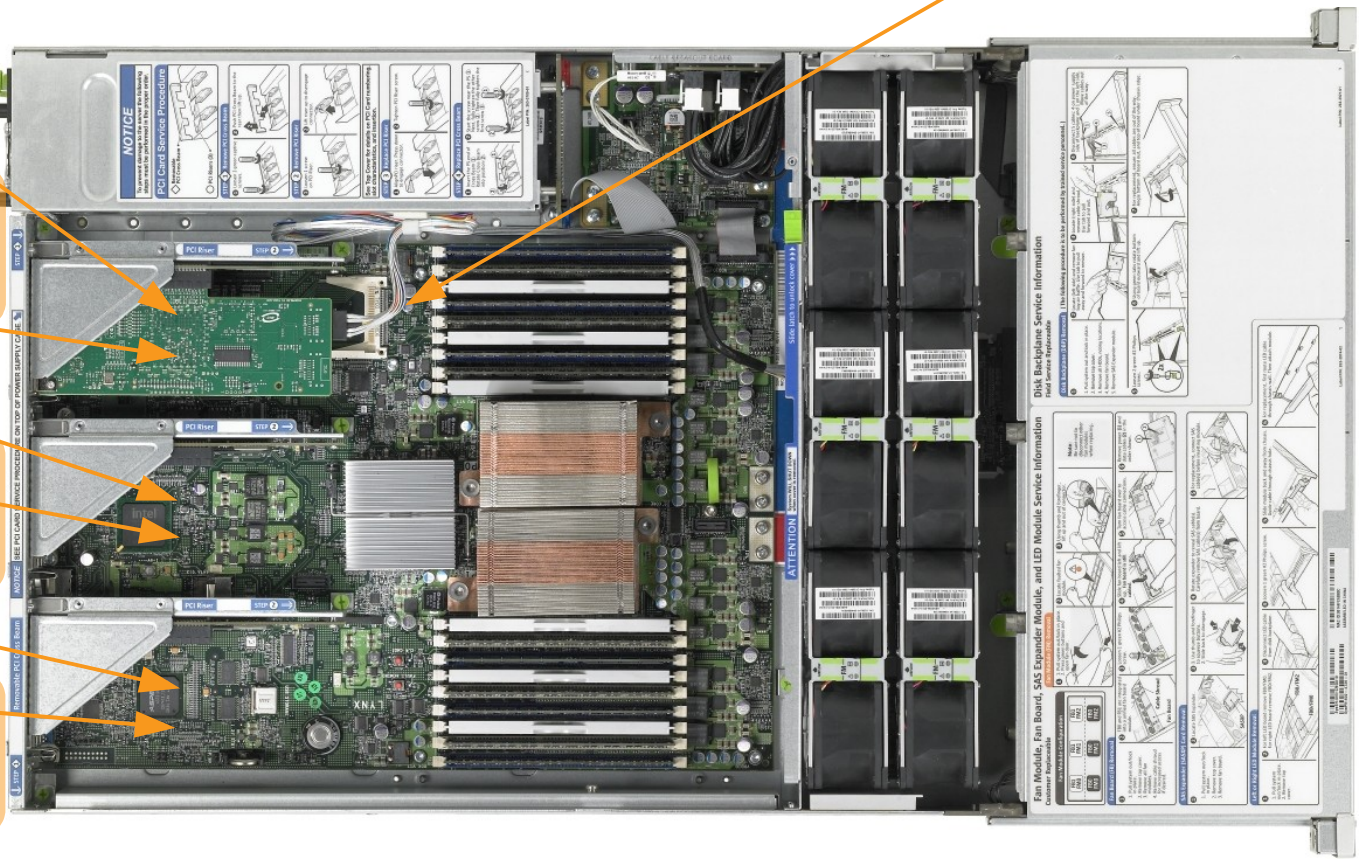
Sun Dual port QDR InfiniBand
PCIe HCA
Installed in slot PCIE-3

Sun FlashFire F20 PCIe card

Sun FlashFire F20 PCIe card

Sun FlashFire F20 PCIe card

Sun FlashFire F20 PCIe card

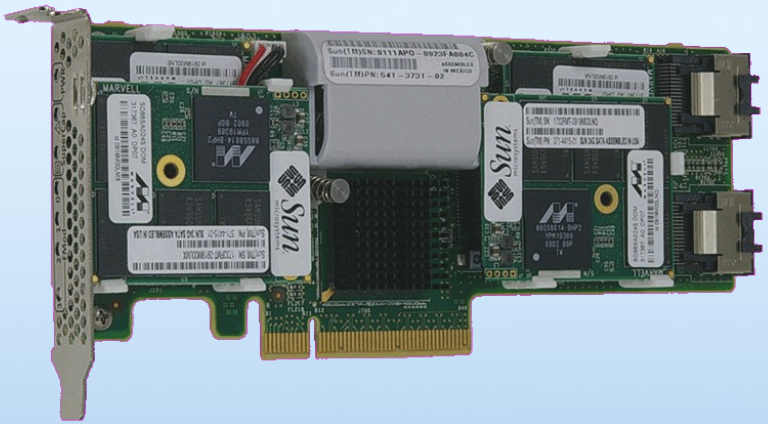


Sun Exadata Storage Server – SF X4275



Sun Flash Accelerator F20 PCIe Card

100k IOPS per PCIe slot



The “Secret Sauce” is how Oracle Exadata Storage Server Software utilizes and manages the cache!

Used as Smart Flash Cache Target

Capacity

- 96 GB of Flash in one 8-lane PCIe slot
- Attach up to 8 Internal SAS/SATA HDD

Performance

- 100k read IOPS (4K)
- 84k write IOPS (4K)
- < .4ms random read/write service times
- 1 GB/s sequential read (.5 GB/s write)

Reliability

- Advanced wear-leveling, write endurance
- Integrated Super CAP: no batteries!
- ENTERPRISE READY

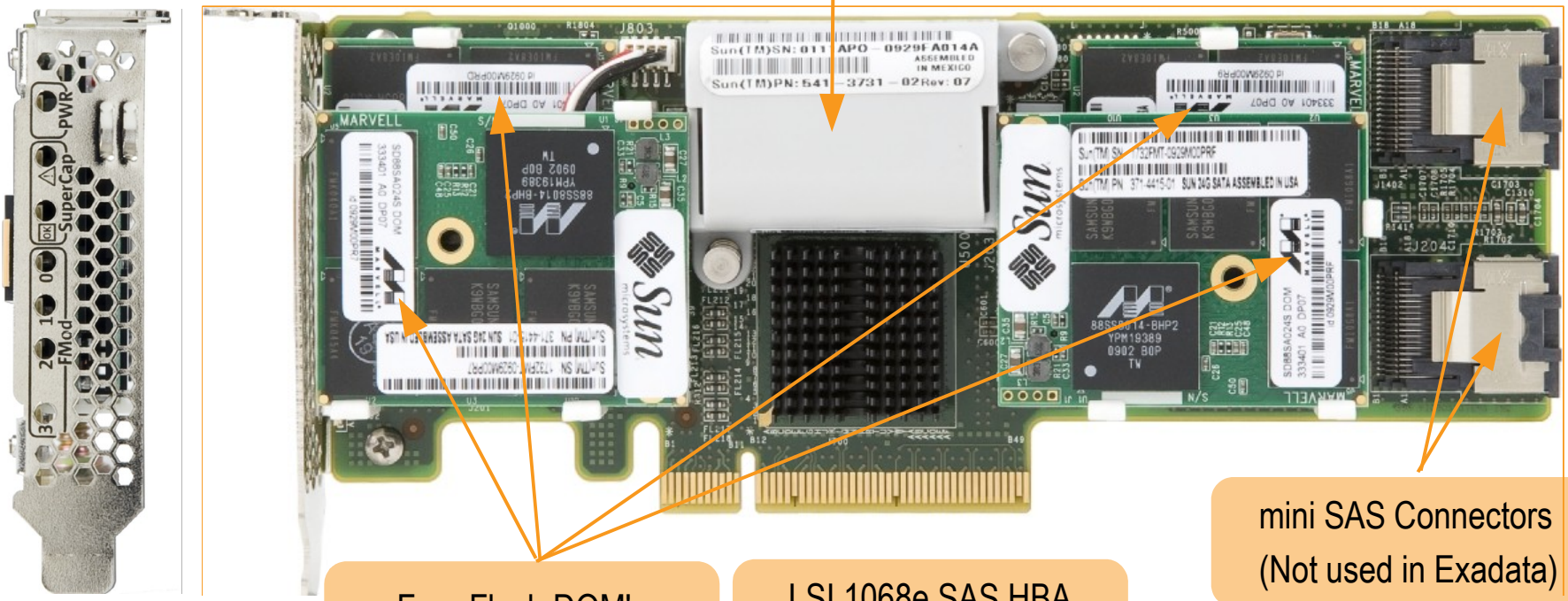
Connectivity

- 2 x 4-wide3Gb/s internal SAS ports
- SAS expander , HBA and Flash
- Standard HBA drivers

Sun Flash Accelerator F20 PCIe Card

Half Length, Low Profile form factor

Super Capacitor (ESM)

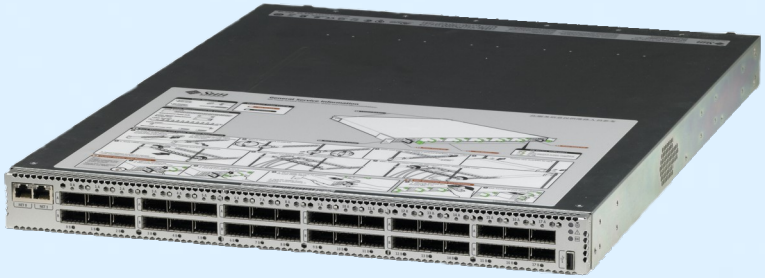


Four Flash DOM's (FDOMs)

LSI 1068e SAS HBA That uses 8 lane PCIe 1

mini SAS Connectors (Not used in Exadata)

Sun Datacenter InfiniBand Switch 36



Up to three IB switches per Exadata V2 rack

36 x 40Gbps (QDR) IB ports
Low Latency
Embedded OpenSM

Building Block for High-Bandwidth Low-Latency Non-Blocking IB Network Fabric

- **Density**
  - 36 QDR InfiniBand (IB) ports
  - 1RU form factor
- **I/O**
  - 36 physical standard QSFP ports
  - 40Gbps per port
  - 100ns latency
- **Availability**
  - Hot-swap redundant PSUs
  - Hot-swap redundant fans
  - Redundant network management ports

Ethernet Switch



Cisco Catalyst® 4948 10 Gigabit Ethernet Switch
48 ports of wire-speed 10/100/1000BASE-T
2 ports of wire-speed 10 Gigabit Ethernet
A wire-speed, low-latency,
Layer 2 to 4, 1-rack-unit (1RU)

- All Database Servers and Sun Exadata Storage Servers
  - A) Service Processors for Lights Out Management (ILOM)
  - B) Data Facing Servers
    - Provisioning
    - Administration
- Database Servers
  - Access point for customer network

Keyboard, Video, Monitor (KVM)

- Avocent ESC17
  1U Rackmount 17” LCD

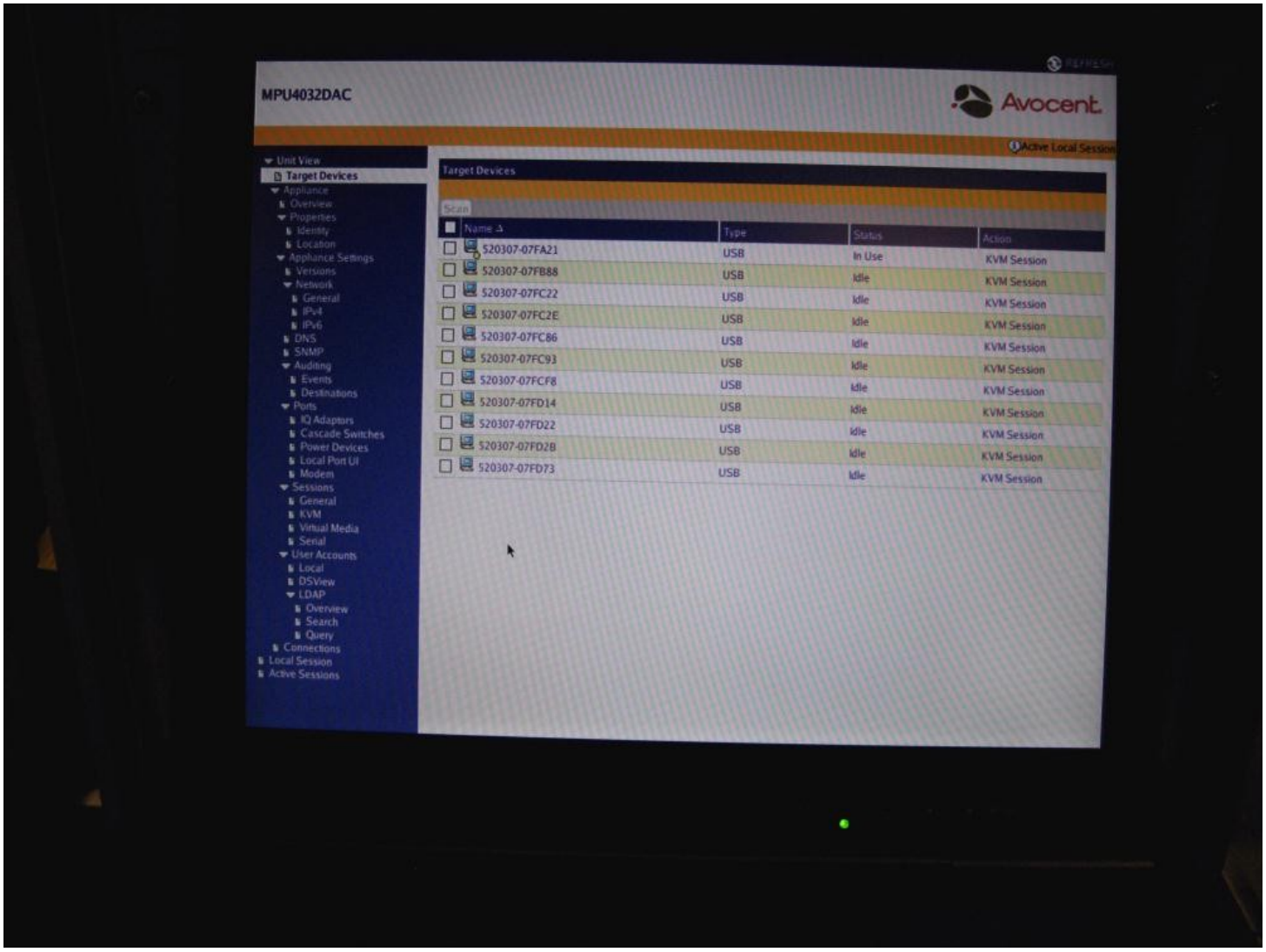


- Avocent MergePoint Unity 4032
  32-port KVM switch





Keyboard, Video, Monitor (KVM)



Networks

Networking Guidelines Networks

- Management/ILOM network
  - Administrative duties
  - Connects all servers, ILOM, and switches via Cisco switch
  - Single uplink from Cisco switch to customer network
- Client Access network
  - Application access (VIP & SCAN addresses)
  - Connects SF X4170 servers
- Private InfiniBand network
  - For storage network and RAC interconnect traffic, external connectivity (optional)
  - Connects all SF X4170 and SF X4275 servers
- Additional networks
  - Only for SF X4170 servers
  - Backup, Disaster Recovery, ...

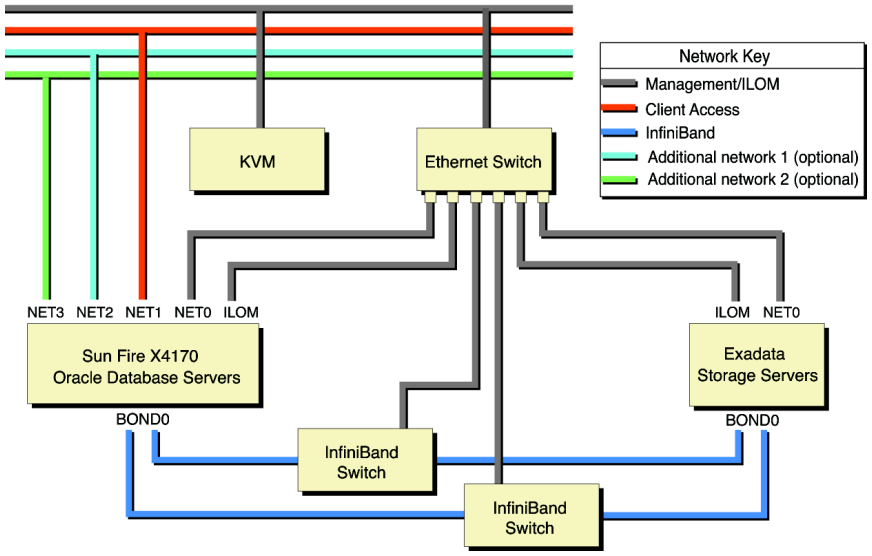
Sun Confidential: Internal Only

Networks

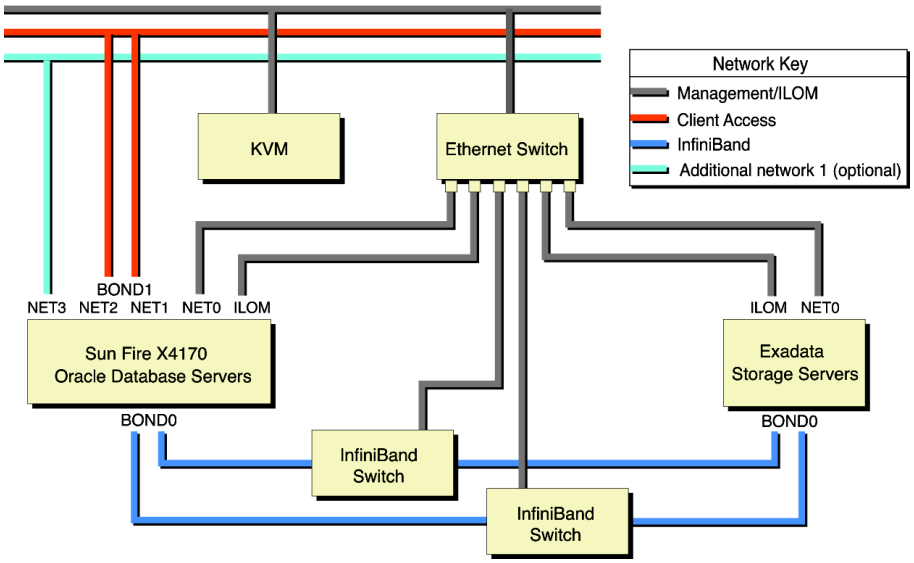


Network Topology

Non-Bonded Configuration



Bonded Configuration



Networking Guidelines
Ethernet Network Drops

- Management network
  - 1 Ethernet connection for Cisco switch to customer's management network
  - 1 Ethernet connection for KVM switch to customer's management network
- Client Access network
  - 1 Ethernet connection per SF X4170 server (NET1/NET2)

Type of Sun Oracle Database Machine	Bonded Network Configuration	Non-bonded Network Configuration
Sun Oracle Database Machine Full Rack	16	8
Sun Oracle Database Machine Half Rack	8	4
Sun Oracle Database Machine Quarter Rack	4	2

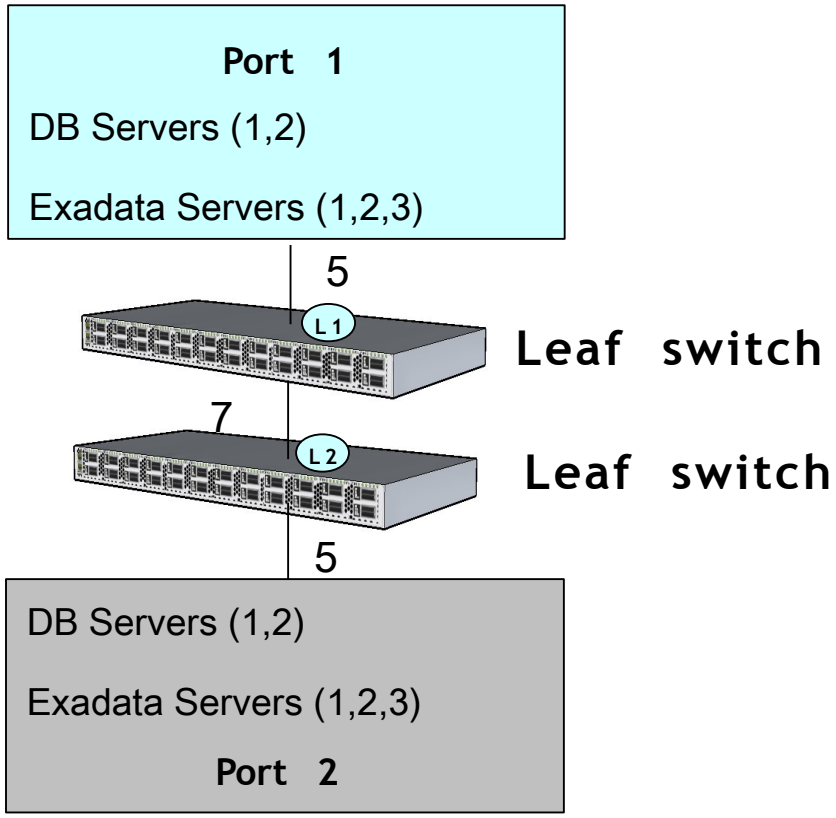
Networking Guidelines
Ethernet Network Drops

- Additional network NET2 (in non-bonded configuration)
  - 1 Ethernet connection per SF X4170 server
    - Full Rack: 8 connections
    - Half Rack: 4 connections
    - Quarter Rack: 2 connections
- Additional network NET3
  - 1 Ethernet connection per SF X4170 server
    - Full Rack: 8 connections
    - Half Rack: 4 connections
    - Quarter Rack: 2 connections

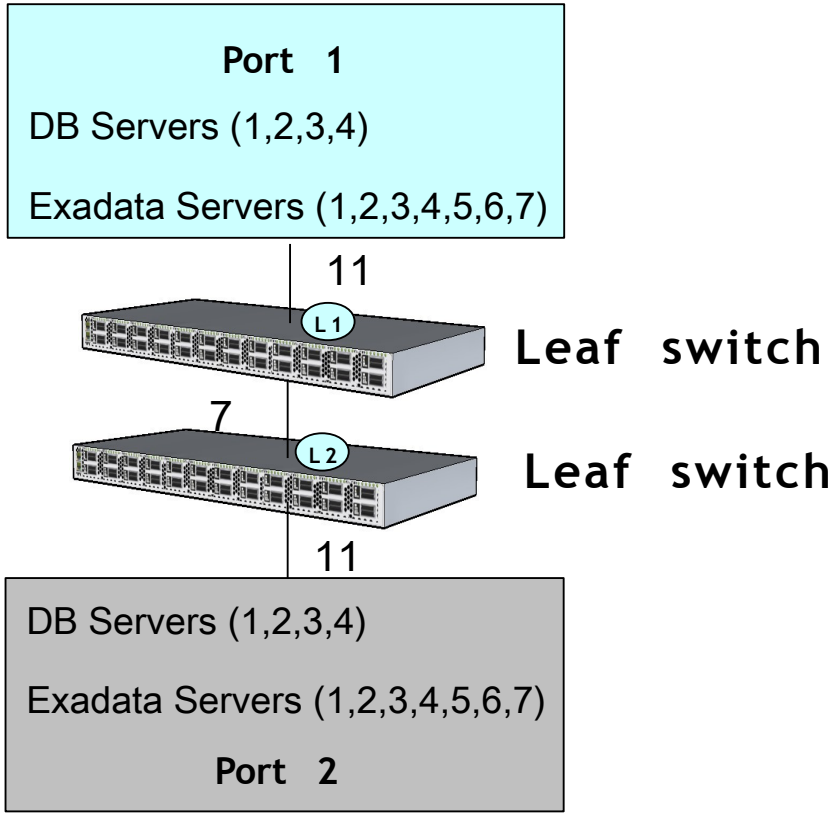
InfiniBand Network

- Sun Datacenter InfiniBand 36 Switches
  - 2x: Quarter & Half Rack
  - 3x: Full Rack
- Each server has 1 dual-port QDR (40Gb/s) HCA
- Active-Passive bonding – assign single IP address
  - Performance is limited by PCIe bus, so active-active not needed
- Connect one port from the HCA to one leaf switch and the other port to the second leaf switch for redundancy
- Connections pre-wired & default IP addresses pre-configured in the Factory

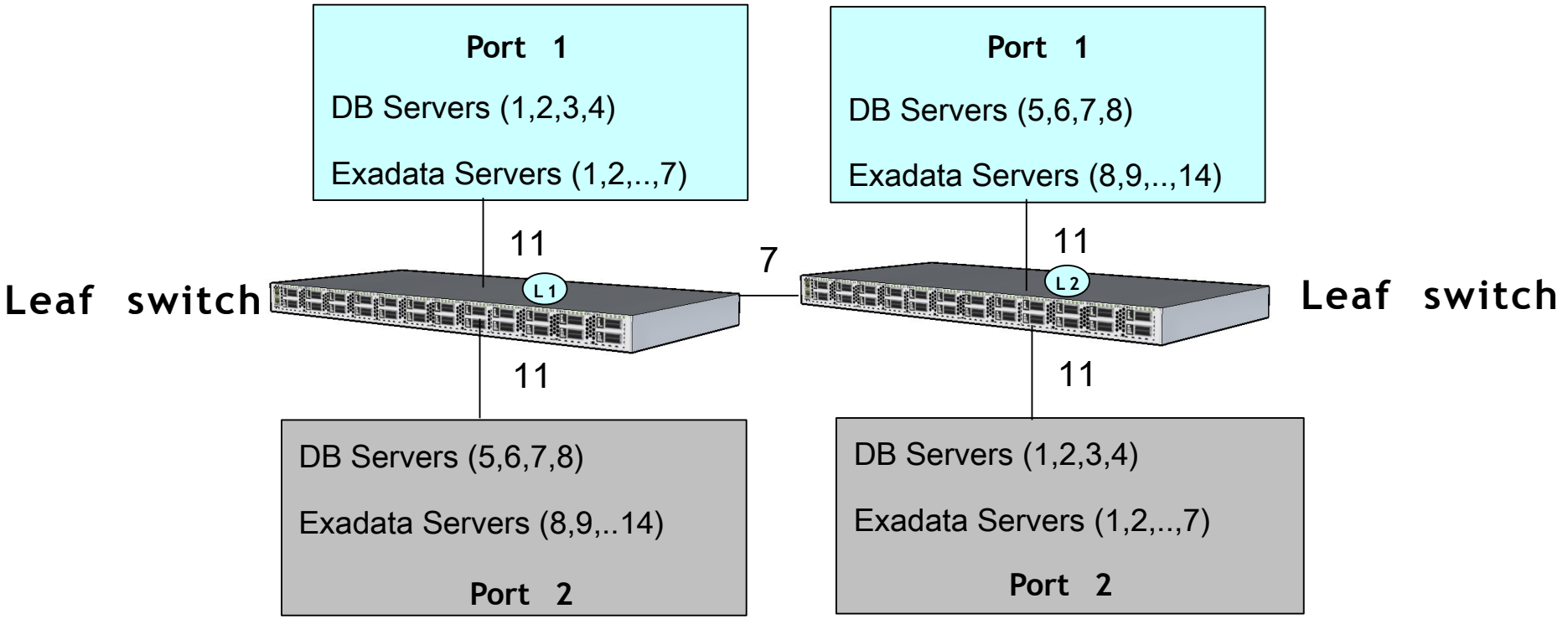
InfiniBand Network Topology Quarter Rack



InfiniBand Network Topology Half Rack

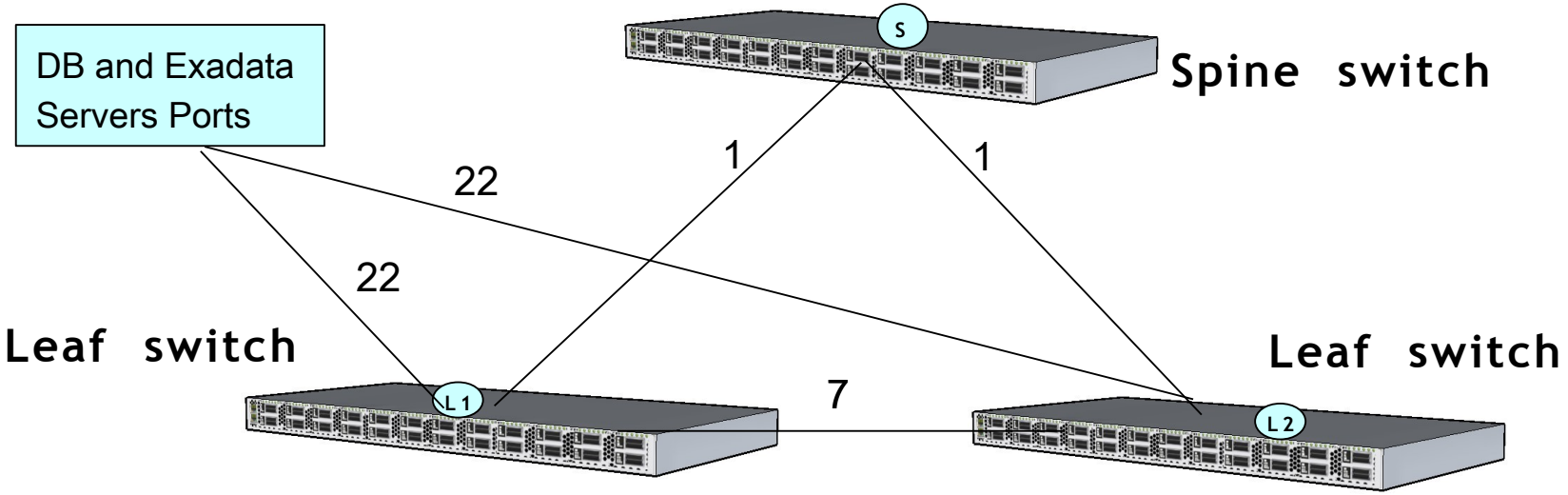


InfiniBand Network Topology Full Rack (1/2)



InfiniBand Network Topology Full Rack (2/2)

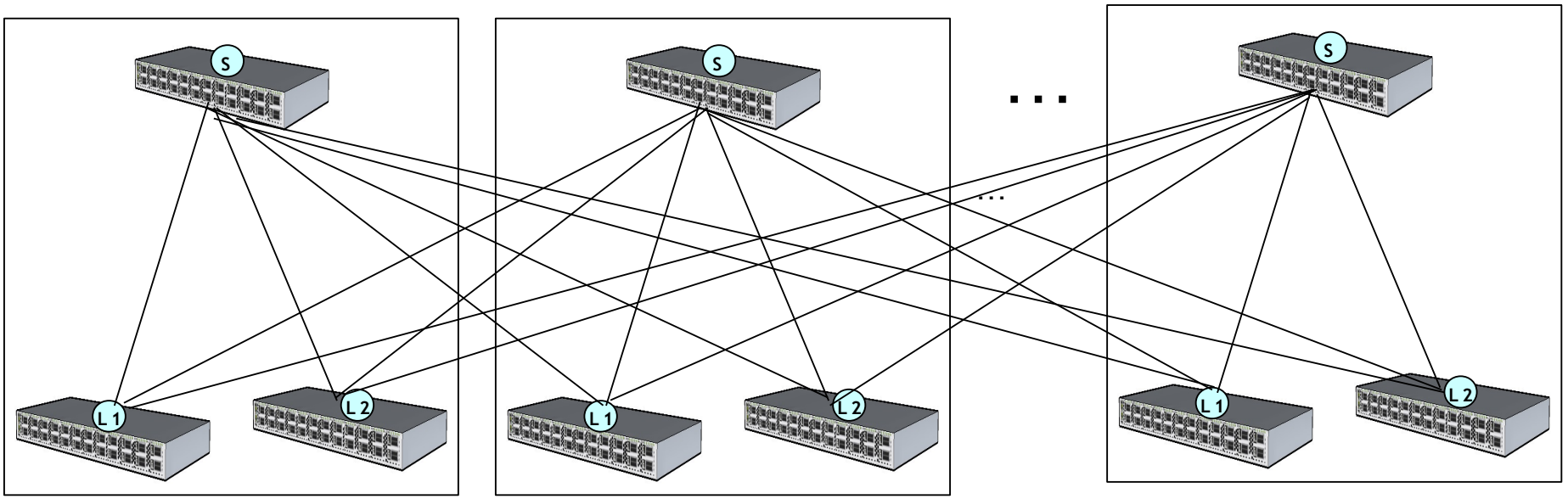
- Use 3rd switch (S) as spine switch
- Connect each leaf switch to spine switch (1 link wide)
- Interconnect leaf switches with each other (7 links wide)



InfiniBand Network Topology Multiple Full Rack

- Sun Oracle Database Machine Multi-Rack Cabling Guide 821-0854-10
- <= 3 Full Racks → Use spare IB cables
- > 3 Full Racks → Customer must purchase additional IB cables
  - Longest approved IB cable is 10 meters
- > 8 Full Racks → Customer must purchase larger external IB switches

InfiniBand Network Topology Multiple Full Rack



Resources

Training

- Sun Oracle Database Machine WZT-6404
  - How-To's
  - Resources & Links
  - Attachments: Student Guide, TOIs
- Sun Fire Nehalem-based Rack Mount Servers
  - Architecture and Features WZT-6508
  - Installation and Replacement WZT-6509
  - Configuration and Administration WZT-6510
  - Troubleshooting WZT-6511
- InfiniBand Fundamentals WZT-4989
- HPC InfiniBand Engineering Workshop, Dec09, Langen, GY
  - Slides | Audio Recordings

Resources

- Oracle
  - Exadata V2 homepage (internal | external)
  - FAQ (internal | external)
  - Documentation (internal)
    - Sun Oracle Database Machine - Owner's Guide
    - Oracle Database Machine – Configuration Worksheets
  - Exadata Forum Archives

Resources

- Sun
  - Sun System Handbook
  - EIS Checklist, EISdoc
  - Sun Installation Engineer (SIE) Checklist
  - PTS Site
  - Documentation (internal)
    - Sun Oracle Database Machine Site Planning Guide
    - Sun Oracle Database Machine Installation Guide
    - Sun Oracle Database Machine Multi Rack Cabling Guide
  - SunSpace/Onestop
  - FAQ: Technical | Business
  - Netadmin Mail Alias: exadata-support@sun.com

Sun Oracle Database Machine Hardware Overview

