
ORACLE®

ORACLE

# Announcing Exadata X4

## Overview of Changes from X3


Last updated – Dec 11, 2013

# Exadata Database Machine

- **The ultimate platform for all database workloads**
  - OLTP, Warehousing, Database as a Service
- **Most advanced <u>hardware</u>**
  - Fully scale-out servers and intelligent storage with unified InfiniBand connectivity and PCI flash
- **Most advanced <u>software</u>**
  - Database optimized compute, storage, and networking algorithms dramatically improve performance and cost
- **Standardized, optimized, hardened end-to-end**

# Exadata X4 Compared with X3

## Much More Performance and Capacity – Same Price

### X4 Storage

| | |
|---|---|
| 2X Larger Physical Flash Memory | 44 TB of Flash Memory |
| Up to 4X Larger Logical Flash Memory | 88 TB using Flash Cache Compression |
| 77% More Flash IOs/sec on X4-2 | 2.66M Reads, 1.96M Writes from SQL |
| 33% Larger High Capacity Disks | 672 TB using 4TB Disks |
| 2X Larger High Performance Disks | 200 TB using 1.2 TB Disks |

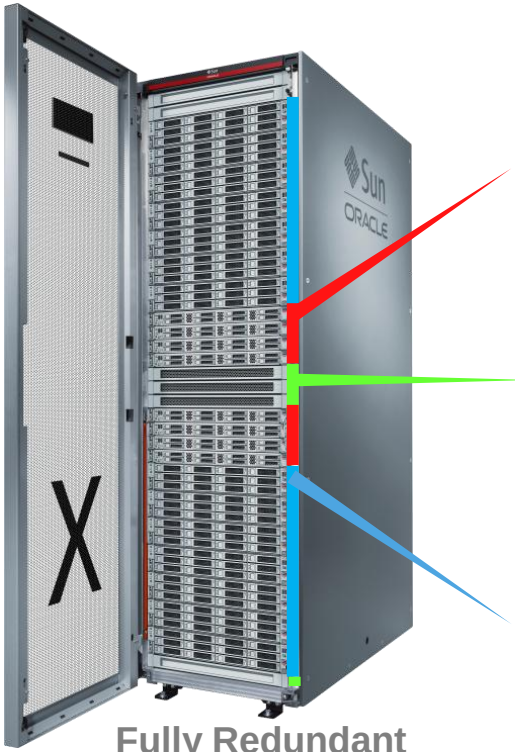
Per DB Machine Full Rack

### X4-2 Compute

| | |
|---|---|
| 50% More Database Cores | 192 Cores using 12-Core Xeon® CPUs |
| 2X Larger DB Server Local Storage | 2.4 TB per server using 600GB Disks |
| 2X Faster InfiniBand | InfiniBand PCI-3 Card. All Ports Active |

# Same Exadata Architecture

Complete | Optimized | Standardized | Hardened Database Platform



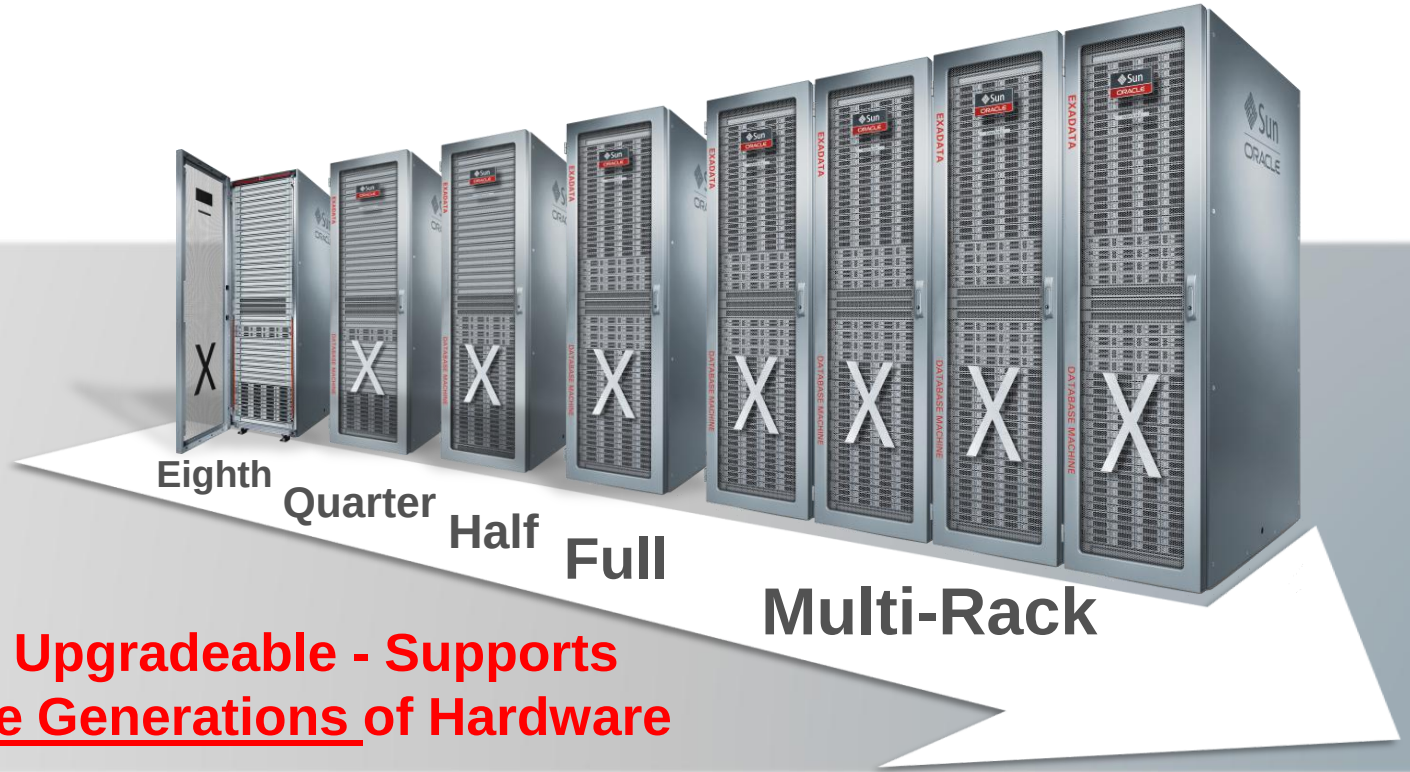

Fully Redundant

- **Standard Database Servers**
  - 8x 2-socket servers ➔ 192 cores, 2TB DRAM
    or
  - 2x 8-socket servers ➔ 160 cores, 4TB DRAM


- **Unified Ultra-Fast Network**
  - 40 Gb InfiniBand internal connectivity ➔ all ports active
  - 10 Gb or 1 Gb Ethernet data center connectivity

- **Scale-out Intelligent Storage Servers**
  - 14x 2-socket servers ➔ 168 *faster* cores in storage
  - 168 SAS disk drives ➔ 672 TB HC or 200 TB HP
  - 56 Flash PCI cards ➔ 44 TB Flash + compression

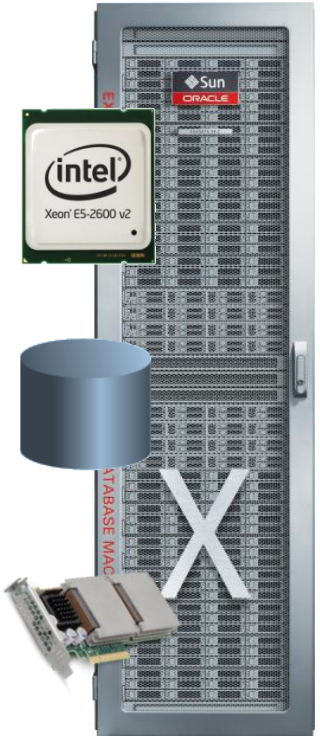
# Scalable from Eighth-Rack to Multi-Rack




**Field Upgradeable - Supports Multiple Generations of Hardware**

# Unique Software Optimizes Database Processing

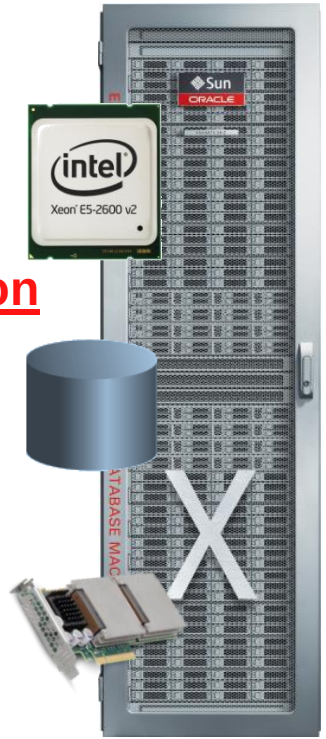
- **Query offload in storage**
  - Data intensive query operations offloaded to storage CPUs
  - 100 GB/sec SQL data throughput
  - Storage Index data skipping
- **Database storage compression**
  - Hybrid Columnar for 10x DB size reduction and faster analytics
- **Database optimized PCI Flash**
  - Smart caching of database data
  - 2.66 Million Database IOs/sec
  - Smart Flash log speeds transactions




- **Database optimized QoS**
  - End-to-end prioritization from application to DB and storage
- **Database optimized availability**
  - Fastest recovery of failed database, server, storage or switch
  - Fastest backup. Incremental offload
  - Exachk top-to-bottom validation of hardware, software, settings
- **Database optimized messaging**
  - SQL optimized InfiniBand protocol for high throughput low latency SQL

# Innovation Continues: Recent Enhancements

- **Query offload in storage**
  - Offload searches on LOBs (12*c*)
  - Offload joins for non-parallel queries (11.2.0.4)
- **Database optimized compression**
  - Hybrid Columnar enhanced for OLTP
  - and for Spatial and Text data (12*c*)
- **Database optimized PCI Flash**
  - Ultra high speed flash compression (X3 & X4) at multi-million IOs/sec
  - Automatic caching for table scans
  - Faster file initialization




- **Database optimized QoS**
  - Prioritization of CPU and IO by multitenant pluggable database (12*c*)
- **Database optimized availability**
  - Prioritize recovery of critical DB files (11.2.0.4)
- **Database optimized messaging**
  - End-to-End prioritization of critical database messages (11.2.0.4), including log writes and RAC

# Exadata X4 is the Fifth Generation DB Machine


DATA WAREHOUSING

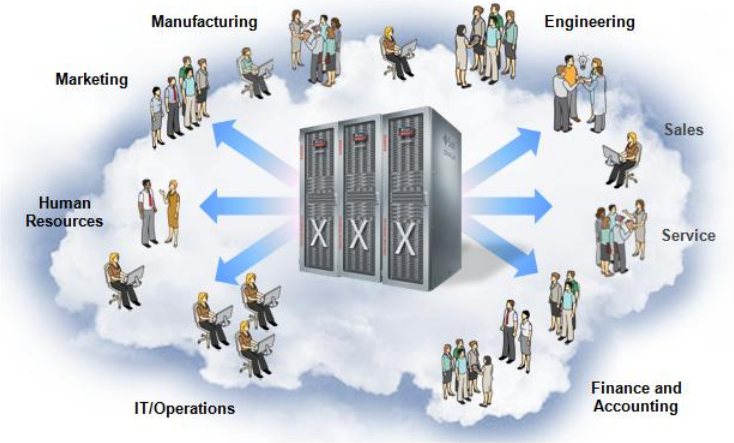
OLTP

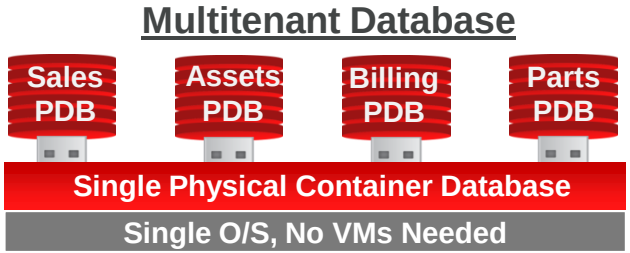
DATABASE CONSOLIDATION

FLASH CENTRIC

**DATABASE AS A SERVICE**

# Comprehensive Database as a Service Platform



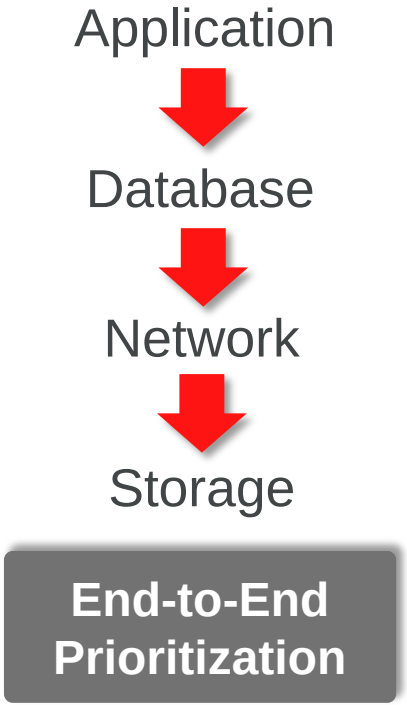




- **Scale-Out Platform optimized for Database**
  - Scale to any size
- **Deploy 100s of databases**
  - Using separate databases or 12c multitenant databases
- **Supports complex & varying mix of workloads**
  - No Performance Bottlenecks
  - Performance Isolation – CPU, I/O, Network

# No Performance Bottlenecks for Consolidation


- Best way to ensure performance is to avoid bottlenecks
- Exadata has unique extreme performance for complex workloads that mix OLTP, DW, batch, reporting
  - Millions of I/Os per second, 100 GB/sec of throughput
  - Sub-millisecond response times
  - Highest bandwidth network
- Unique software optimizations that eliminate bottlenecks
  - e.g. Storage Offload, Smart Flash logging

# Unique Performance Isolation for Consolidation




- Database Resource Manager provides CPU resource management for normal and pluggable databases
  - Both CPU prioritization and limits on CPU usage
- Exadata uniquely provides I/O resource management by pluggable database, job, user, service, etc.
- Exadata uniquely provides database aware **network resource management**
  - Prioritizes critical DB messages through entire fabric

# Technical Details of Exadata X4 Hardware

# X4-2 Database Server

## New 12-core “IvyBridge” CPUs,  Faster InfiniBand Card, Larger Disks

| | |
|---|---|
| Processors | **2 Twelve-Core Intel® Xeon® E5-2697 v2 Processors (2.7GHz) - Turboboost** |
| Memory | 256 GB (16 x 16GB) – **Expandable to 512GB** (16 X 32GB) |
| Local Disks | 4 x **600GB** 10K RPM SAS Disks (Hot-Swappable) |
| Disk Controller | Disk Controller HBA with 512MB Cache **– Battery Online Replaceable** |
| Network | 2  x InfiniBand 4X QDR (40Gb/s) Ports (**PCIe 3.0**) – **Both Ports Active**<br><br>4  x  1GbE/10GbE Base-T Ethernet Ports<br>2  x 10GbE Ethernet SFP+ Ports |


Changes from X3 are in Red

# X4-2 Storage Server

## Larger Flash Cards, Flash Compression, Larger disks, Faster Processors

| | |
|---|---|
| Processors | 2 Six-Core Intel® Xeon® **E5-2630 v2 Processors (2.6 GHz) - Faster clock** |
| Memory | **96 GB** (4 x 8GB + 4 x 16GB) **- More memory needed to manage larger flash** |
| Disks | 12 x **1.2 TB 10K RPM** High Performance SAS (hot-swap) – **2.5" disk size**<br>OR<br>12 x **4 TB** 7.2K RPM High Capacity SAS (hot-swap) – 3.5" disk size |
| Flash | 4 x **800 GB** Sun Flash Accelerator **F80** PCIe Cards – **Hardware Compression** |
| Disk Controller | Disk Controller HBA with 512MB Cache **- Battery Online Replaceable** |
| Network | 2 x InfiniBand 4X QDR (40Gb/s) Ports (**PCIe 3.0**) – **Both Ports Active**<br><br>Embedded Gigabit Ethernet Ports for management connectivity |

Changes from X3 are in Red

# X3-8 Updates


- X3-8 database machines updated with X4-2 storage servers
  - Benefit from increased flash size, disk size, and faster processor in storage
- 8-socket database servers remain the same
  - New Xeon EX Ivybridge chips due in 2014
  - Local disks and InfiniBand card in 8-socket DB servers are also unchanged – no dual active InfiniBand ports
- Name continues to be X3-8

# Exadata Rack Level Changes

## Removal of Default Spine Switch, Change in Spares Kit


- InfiniBand spine switch no longer included in Database Machines
  - Spine switch is used to connect multiple racks together
    - At bottom of rack
  - Previously included in Full and Half Racks
  - Multi-rack connectivity now requires purchase of spine switch
- Storage Expansion Racks continue to ship with spine switches
- Changes in included spares kit:
  - All configurations now ship with 1 spare disk and 1 spare flash card

# X4-2 and X3-8 Infrastructure

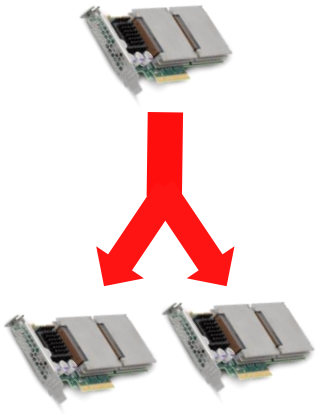

- Similar power, cooling, airflow as X3-2 and X3-8
  - Even though processing power greatly increased
- Same PDUs as the X3-2 and X3-8
- Same Cisco switch for management connectivity

# Technical Details of Exadata X4 Software

# 11.2.3.3.0

- Available on all - V1, V2, X2, X3
- Required on X4 systems

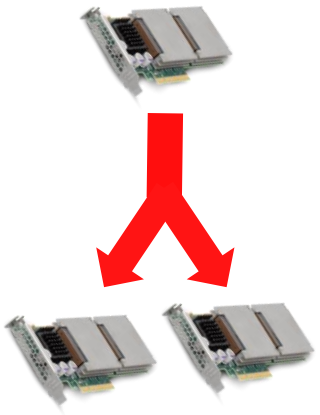
# Exadata Flash Cache Compression



- Exadata uses compression to expand smart flash cache
  - Data automatically compressed as it is written to flash cache
  - Automatically decompressed when it is read out of flash cache
  - Up to 2X more data fits in smart flash cache, so flash hit rates will improve and performance will improve for large data sets
- Flash cache compress/decompress implemented in hardware
  - Performance is same as uncompressed – millions of I/Os per second
    - ZERO performance overhead
  - Supported on X3 or X4 storage servers (requires F40 or F80 cards)

# Exadata Flash Cache Compression

- As always, compression benefits vary based on data

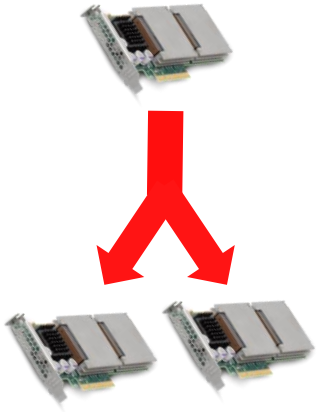


| Data Type | Typical Compression |
|---|---|
| Uncompressed Tables | 1.3X to 5X |
| OLTP Compressed Tables | 1.2X to 2X |
| Indexes | 1.3X to 5X |
| Oracle E-biz uncompressed DB | 3x to 5x |
| HCC Compressed Tables or Compressed LOBs | Minimal |

Many OLTP Databases will see 2x Flash Increase

- X4 flash cache compression expands capacity to 88TB (raw) per rack
  - Up to 4X more than X3 (depending on compressibility of data)

# Flash Cache Compression Commands



- Trivial to implement, no management
- Enable using simple cell command
  - **`alter cell flashCacheCompress=TRUE`**
  - **`On X3 machines also run:`**
    - **`alter cell flashCacheCompX3Support= TRUE`**
  - Requires **Advanced Compression Option** on all databases that access compressed flash cache (therefore not enabled by default)
    - Most Exadata customers already have Advanced Compression
- Amount of data cached in Exadata Smart Flash Cache grows and shrinks dynamically and automatically based on data compressibility
- Monitor Flash Cache Compression using cell metric FC_BY_USED
  - Reported flash cache size will increase to up to double physical flash size

# Exadata Smart Flash Table Caching

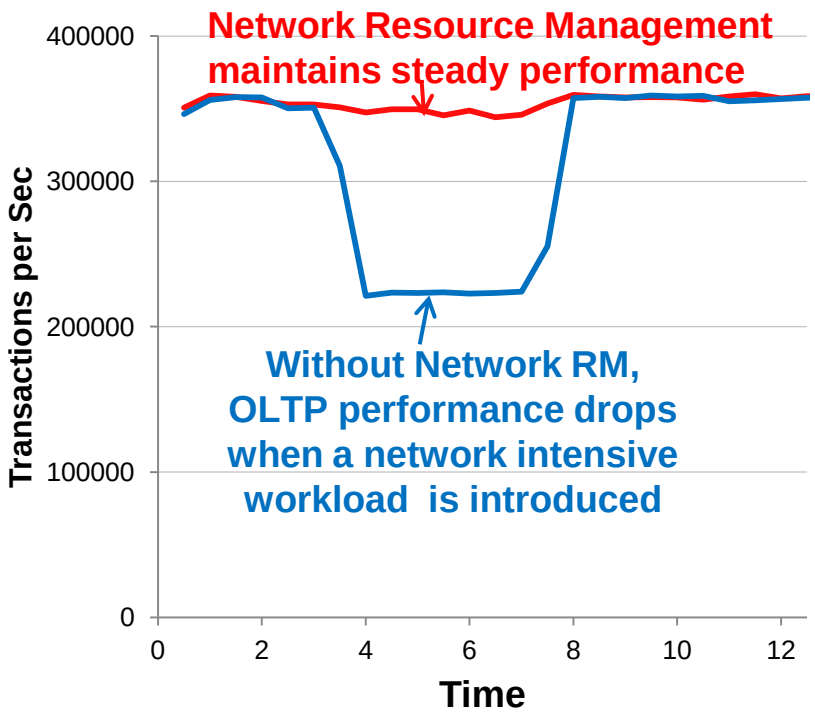

- Smarter flash caching for large table scans
  - Exadata software understands database table and partition scans and automatically caches then when it makes sense
  - Avoids thrashing flash cache when tables are too big or scanned infrequently or scanned by maintenance jobs
  - If scanned table is larger than flash, then subset of table is cached
  - No need to manually “KEEP” tables that are only scanned

# Exadata Effective Flash Size Often 10X Larger

- Due to Exadata Smart Flash Caching, the effective size of flash is much bigger than the physical size
  - Active data is automatically loaded into flash, inactive data is kept on disk
  - Caching is dynamic & fine grained, mirrors of data usually kept on disk
  - Flash cache hit rates are often above 95% or even 98% in real-world databases even when total database size is 10X larger than flash size
  - Get flash performance for databases many times bigger than physical flash

- Compression further expands the effective flash size
  - Both database compression (Basic, Advanced Row Compression, Hybrid Columnar) and Smart Flash Cache Compression
- Most databases run entirely in flash, even DBs much bigger than flash

# Exadata Network Resource Management

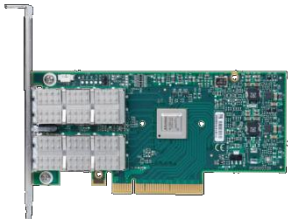



DB Version 11.2.0.4 or 12c, Switch 2.1.3-4

- Exadata Network Resource Management uniquely prioritizes critical database messages through the entire fabric
  - From database to InfiniBand card through InfiniBand switches to storage
  - Latency sensitive messages prioritized over batch, reporting, and backup messages
  - Log file writes have highest priority to ensure low latency transactions
- Combines with Exadata CPU and IO Resource management to ensure safe consolidation of workloads and databases
- Completely automatic & transparent

# InfiniBand Active-Active ports

## Double the InfiniBand Bandwidth



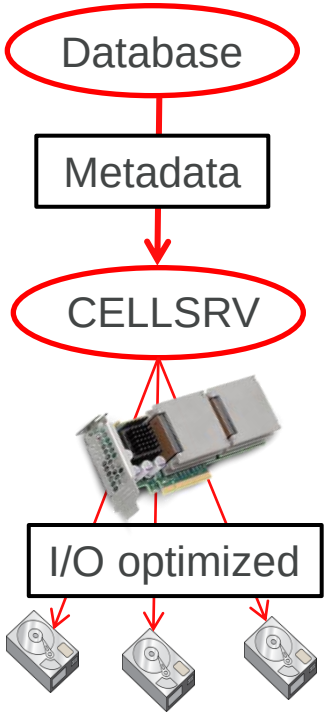
- Dual-ported InfiniBand PCIe-3.0 cards run in active-active mode
  - Double the bandwidth
- Requires 2 private IP addresses per InfiniBand card as opposed to the current 1
  - When port fails, IP assigned to that port will automatically fail over to the other port
- Don’t enable when inter-connecting with X3 or older systems
  - Older systems don’t have newer InfiniBand cards
- If connecting external servers using InfiniBand (e.g. Exalogic) then see 888828.1

# Enable Xeon Turbo-Boost

- Automatically allows processor to run faster than base frequency if operating below power and temperature limits
- Enabled on both DB Servers and Storage Servers on X4 systems
- Clock rate of DB Servers increases to maximum of 3.5 Ghz when few threads are active
  - Up to 28% improvement in processor throughput
- Clock rate often increased to 3.0 Ghz for heavy workloads
  - 11% boost
- Enabled by default

# Exadata Smart File Initialization

- Database file creation sped up by an order of magnitude
  - Create tablespace, file extensions, autoextend show benefit
- Combine the benefits of previous Smart Initialization and Writeback Flash Cache
  - Write file creation meta-data to writeback flash cache
  - Write I/Os to disk deferred, or not performed if data loaded
- Happens automatically, no tuning needed

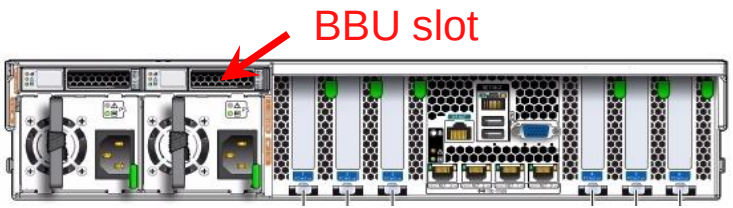


# Availability and Robustness Enhancements


- Disks and Flash
  - Improved algorithms for slow or flaky disk/flash detection and confinement
  - Faster recovery for bad sectors
  - Automatic hard disk scrubbing for bad sectors with 11.2.3.3 and Grid Infrastructure 11.2.0.4
  - Rebalance avoids reading disks that are in predictive failure (“sick” disks) with 11.2.0.4
- ILOM
  - Automatic reset of hung ILOM

# Online Replacement of Disk Controller Battery

- BBU (battery) removed from HBA, relocated to HDD slot
  - All X4 systems. X3 systems shipped after May 2013. Retrofit for previous X3.
- Software enables online replacement – flushes cache, sets write-through
  - See documentation for complete set of commands
- Much simpler, faster, and fully online battery replacement




Storage Server Rear View


Database Server Front View

# InfiniBand Manageability Enhancements


- New InfiniBand switch firmware – 2.1.3-4
  - Applied online by patchmgr
- Automatically disable InfiniBand network links showing poor performance
  - Can happen when cable not properly connected or bent too much
- Sub-second recovery from InfiniBand switch failure on multi-rack configs
- Automatic Service Request (ASR) for InfiniBand switches

# Management Enhancements


- Alert email reminder
  - Storage periodically sends reminder alert if broken HW is not replaced or other alerts are outstanding
- DBNodeUpdate (introduced in Spring 2013)
  - Much simpler OS and firmware upgrades on Database Nodes
  - Automatically adjusts OS settings
- Patchmgr emails progress alerts
  - No need to continuously watch system during patching
- Patchmgr Plugins to deliver known issue coverage
- Exachk continuously enhanced
- More robust patching

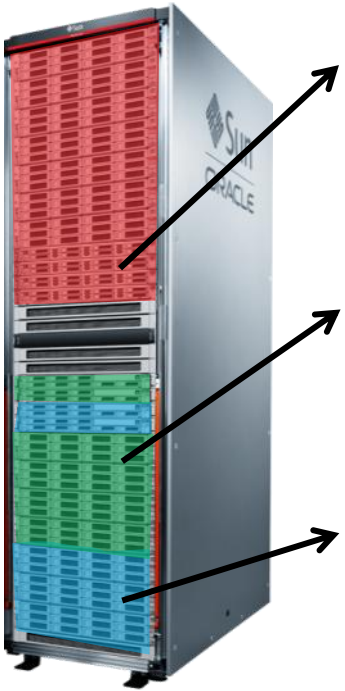
# Operating System Updates in 11.2.3.3.0


- Oracle Linux distribution updated to 5.9
  - Including all recent security updates


- Linux kernel updated to UEK2 kernel (2.6.39-400)
  - Same kernel on database servers (2-socket and 8-socket) and the storage servers
- Oracle Solaris updated to S11 Update 1 SRU 9

# Seamless Upgrades and Expansions

## Upgrade Example



**X4-2**
Half to Full Upgrade in 2014

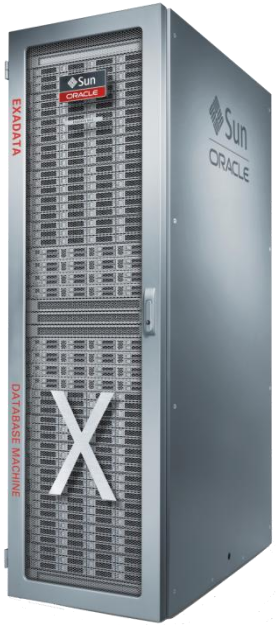
**X3-2**
Qtr to Half Upgrade in 2013

**X2-2**
Initial Quarter Rack deployed in 2011

- A single Database Machine configuration can have servers and storage from different generations
  - V2, X2, X3, X4
- Databases and Clusters can span across multiple hardware generations
- New software runs on older hardware

# Exadata X4 Change Summary

## Much More Performance and Capacity – Same Price




- Exadata X4 provides large performance and capacity increases
  - Large increase in Flash Capacity
  - Large increases in Disk Capacity
  - Large increases in processor throughput on X4-2
  - Large increases in InfiniBand throughput on X4-2
- Exadata X4 software (11.2.3.3.0) provides
  - Automatic Flash compression on X3 and X4 systems
  - Improved Flash caching
  - Improved support for consolidation and Database as a Service
  - Many management and robustness improvements

# Hardware and Software

ORACLE

# Engineered to Work Together


ORACLE®