EXTREME PERFORMANCE

ORACLE®

ORACLE®

HP Oracle Exadata Product Overview

HP Oracle Exadata Storage Server

- Optimized Storage Product for the Oracle Database
- Extreme I/O and SQL Processing performance for data warehousing
- Combination of hardware and software

Hardware by

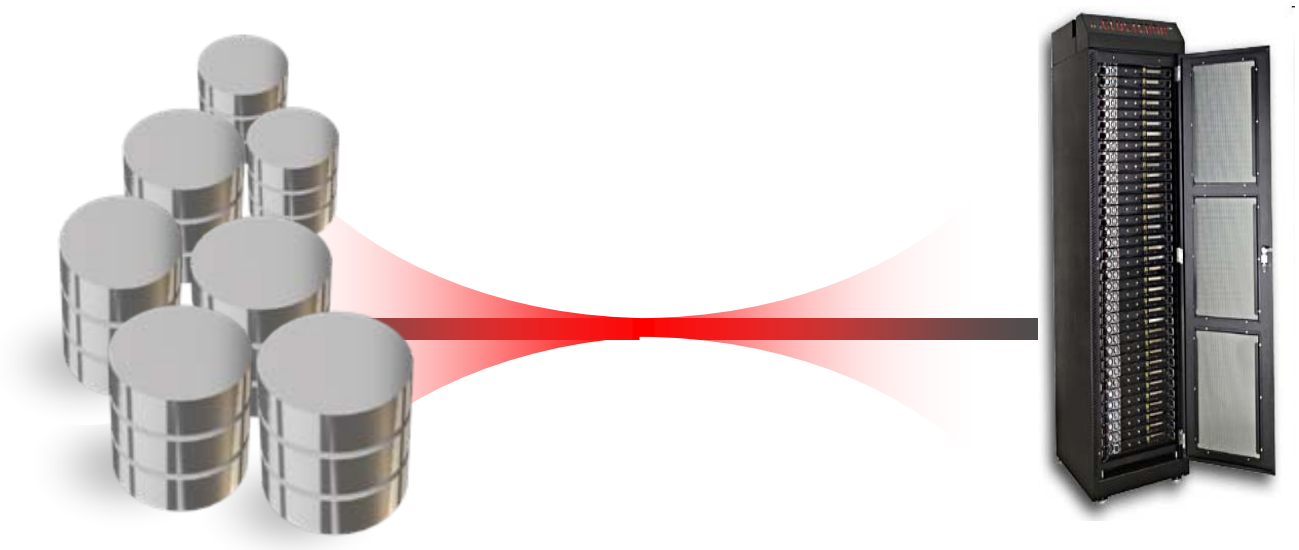
Software by

Exadata Benefits

- Extreme Performance
  - **10X to 100X** speedup for data warehousing
- Database Aware Storage
  - Smart Scans
- Massively Parallel Architecture
  - Dynamically Scalable
  - Unlimited Linear Scaling of Data Bandwidth
  - Transaction/Job level Quality of Service
- Mission Critical Availability and Protection
  - Disaster recovery, backup, point-in-time recovery, data validation, encryption

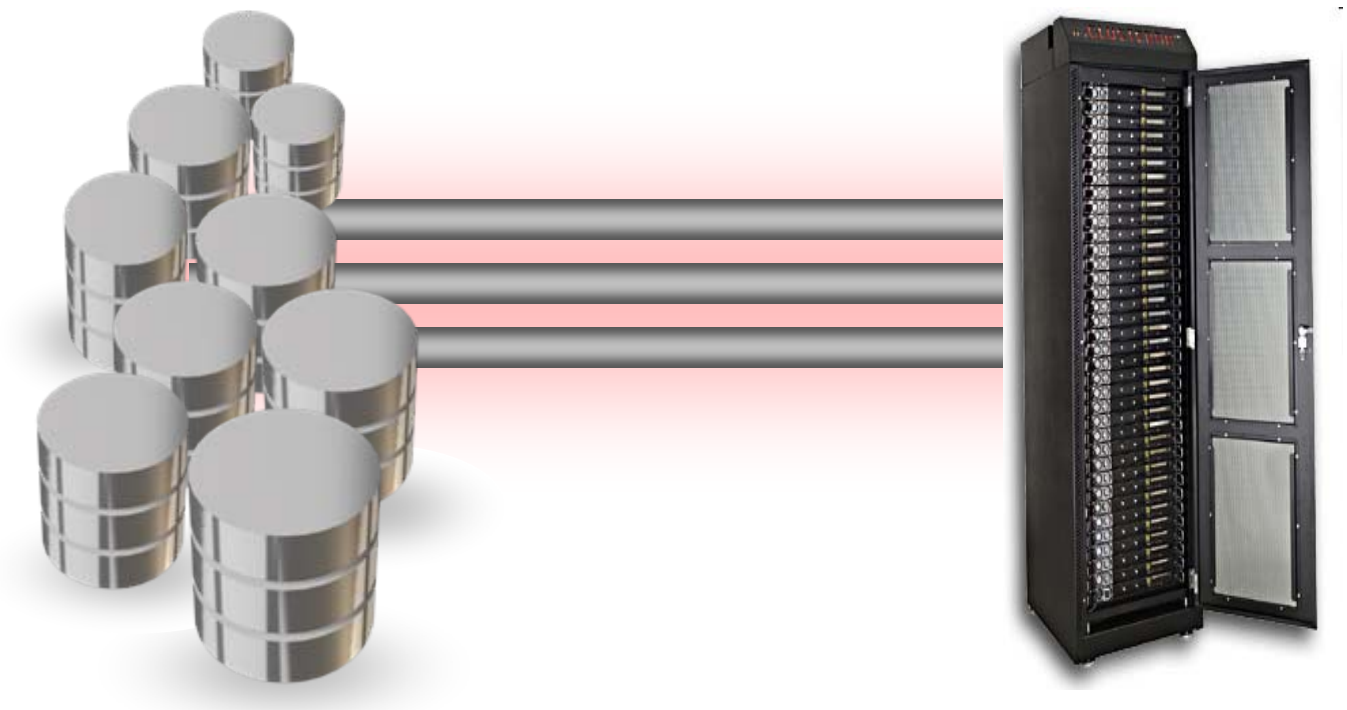
The Performance Challenge

Storage Data Bandwidth Bottleneck



- Current warehouse deployments often have bottlenecks limiting the movement of data from disks to servers
  - Storage Array internal bottlenecks on processors and Fibre Channel Loops
  - 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

Solutions To Data Bandwidth Bottleneck



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

Exadata – A New Architecture

Breaks Data Bandwidth Bottleneck

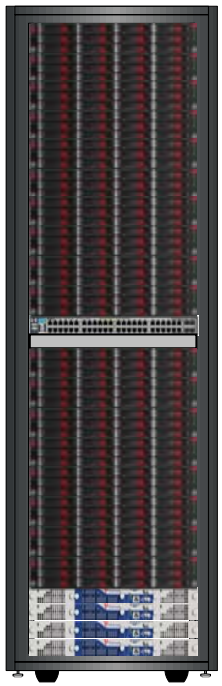
- Exadata Ships Less Data Through Pipes
  - Query processing is moved into storage to dramatically reduce data sent to servers while offloading server CPUs
- Exadata has More Pipes
  - Modular storage “cell” building blocks organized into Massively Parallel Grid
  - Bandwidth scales with capacity
- Exadata has Bigger Pipes
  - InfiniBand interconnect transfers data 5x faster than Fibre Channel

Exadata Moves a Lot Less Data a Lot Faster

HP Exadata Storage Server Hardware

Exadata Storage Server

Racked Exadata Storage Servers



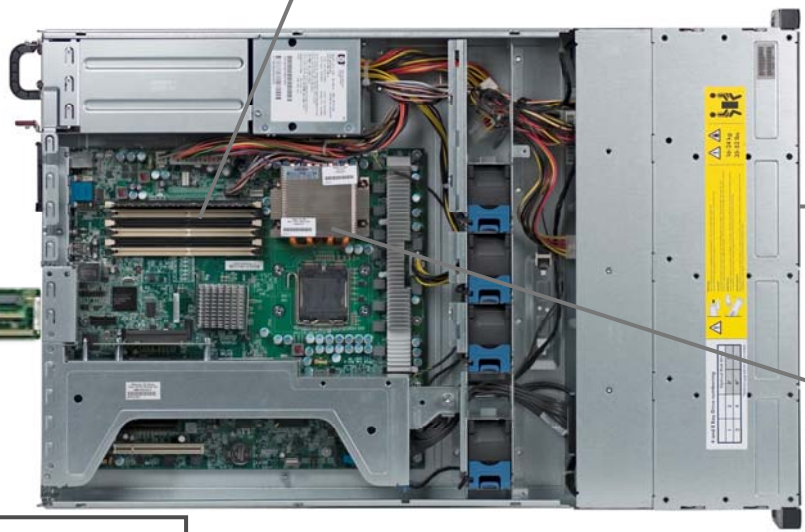
- Building block of massively parallel Exadata Storage Grid
  - Up to 1GB/sec data bandwidth per cell
- HP DL180 G5
  - 2 Intel quad-core processors
  - 8GB RAM
  - Dual-port 4X DDR InfiniBand card
  - 12 SAS or SATA disks
- Software pre-installed
  - Oracle Exadata Storage Server Software
  - Oracle Enterprise Linux
  - HP Management Software
- Hardware Warranty
  - 3 YR Parts/3 YR Labor/3 YR On-site
  - 24X7, 4 Hour response

HP Exadata Storage Server Hardware Details

Redundant 110/220V Power Supplies

8 GB DRAM

P400 Smart Array Disk Controller card
- 512M battery backed cache



12 x 3.5” Disk Drives

2 Intel Xeon Quad-core Processors

InfiniBand DDR dual port card

LO100c – Management Card

Included Software:

- Oracle Exadata Storage Server Software
- Oracle Enterprise Linux
- HP Management Software

SAS or SATA Disks in Exadata Servers

- Choice of either
  - 450 GB 15,000 RPM SAS disks
  - 1 TB 7,200 RPM SATA disks
- Choose SAS Based Servers for High Performance

SAS Advantages	SAS	SATA	Advantage
Throughput (MB/s)	1,000	750	1.33X
Average Seek Time (ms)	3.6	7.4	2.05X
Disk level read errors (per year)	6.3	63	10.00X
Years to disk failure	15.2	11.4	1.33X

- Choose SATA Based Servers for High Capacity

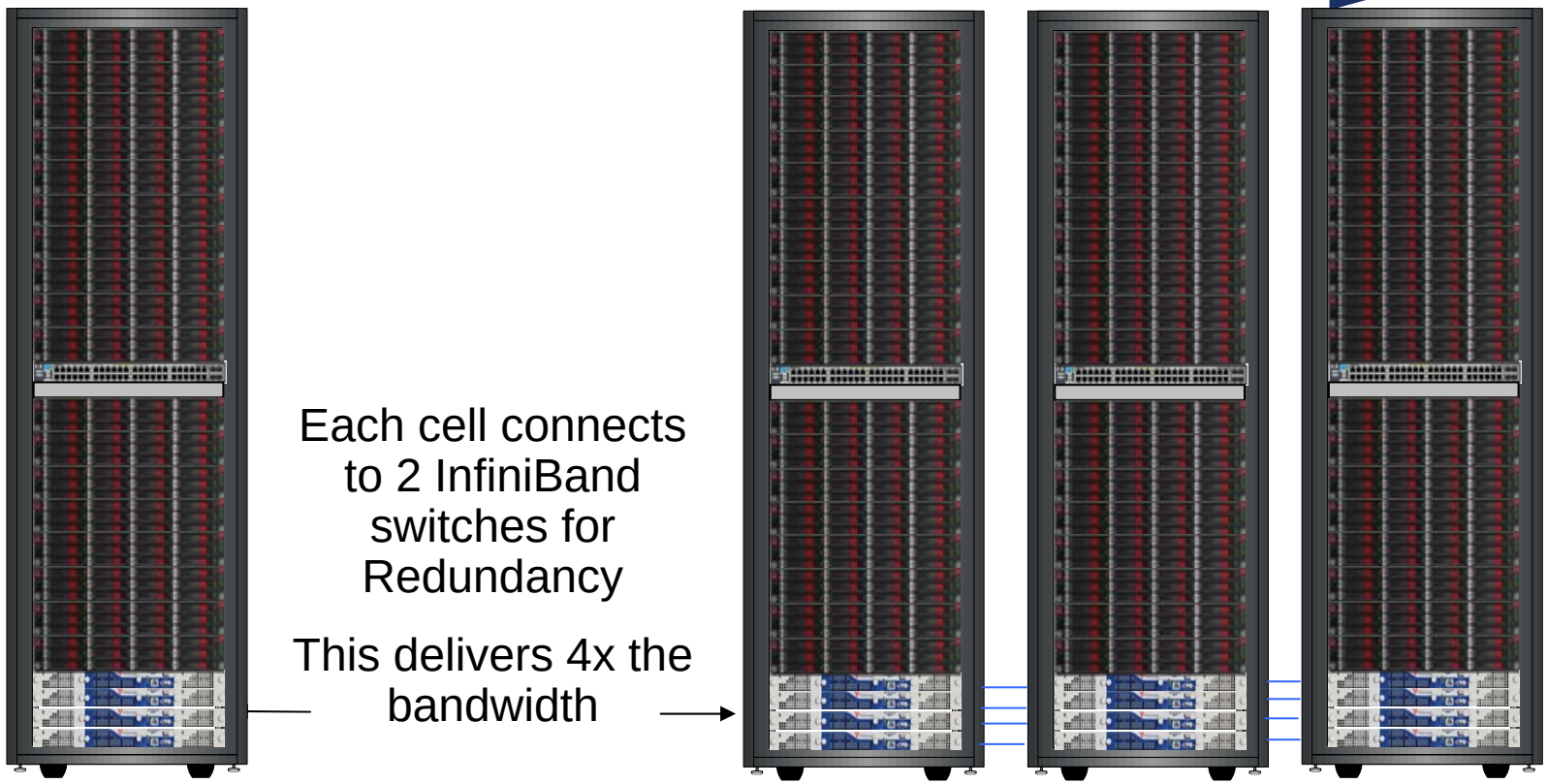
SATA Advantages	SAS	SATA	Advantage
Capacity (TB)	5.4	12	2.22X

ORACLE

Scalable

Add racks to scale further

Scale to 18 cells in one rack



Each cell connects to 2 InfiniBand switches for Redundancy

This delivers 4x the bandwidth

SAS raw capacity per rack: 97TB

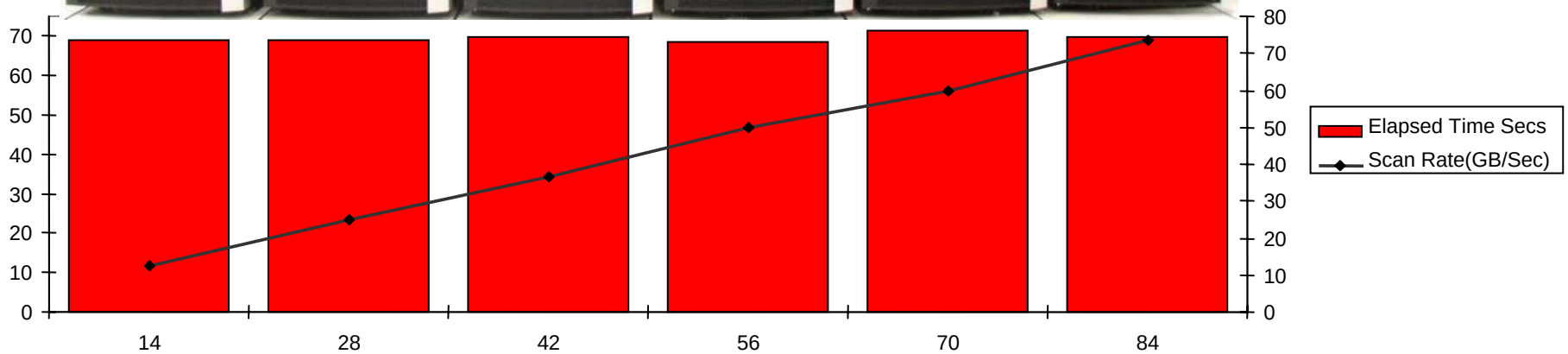
SATA raw capacity per rack: 216TB

Peak throughput per rack : >18GB/s

InfiniBand links across racks for full connectivity

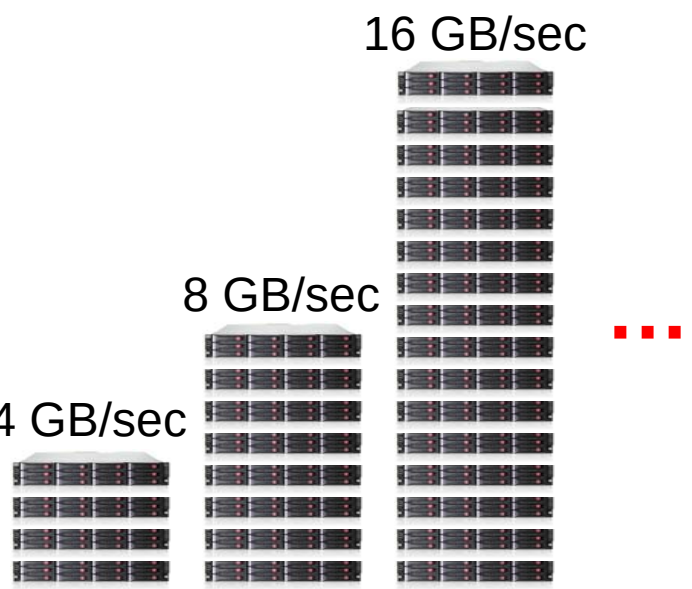
Extreme Performance in Action

Scan TB of User Data In 3.5 sec.



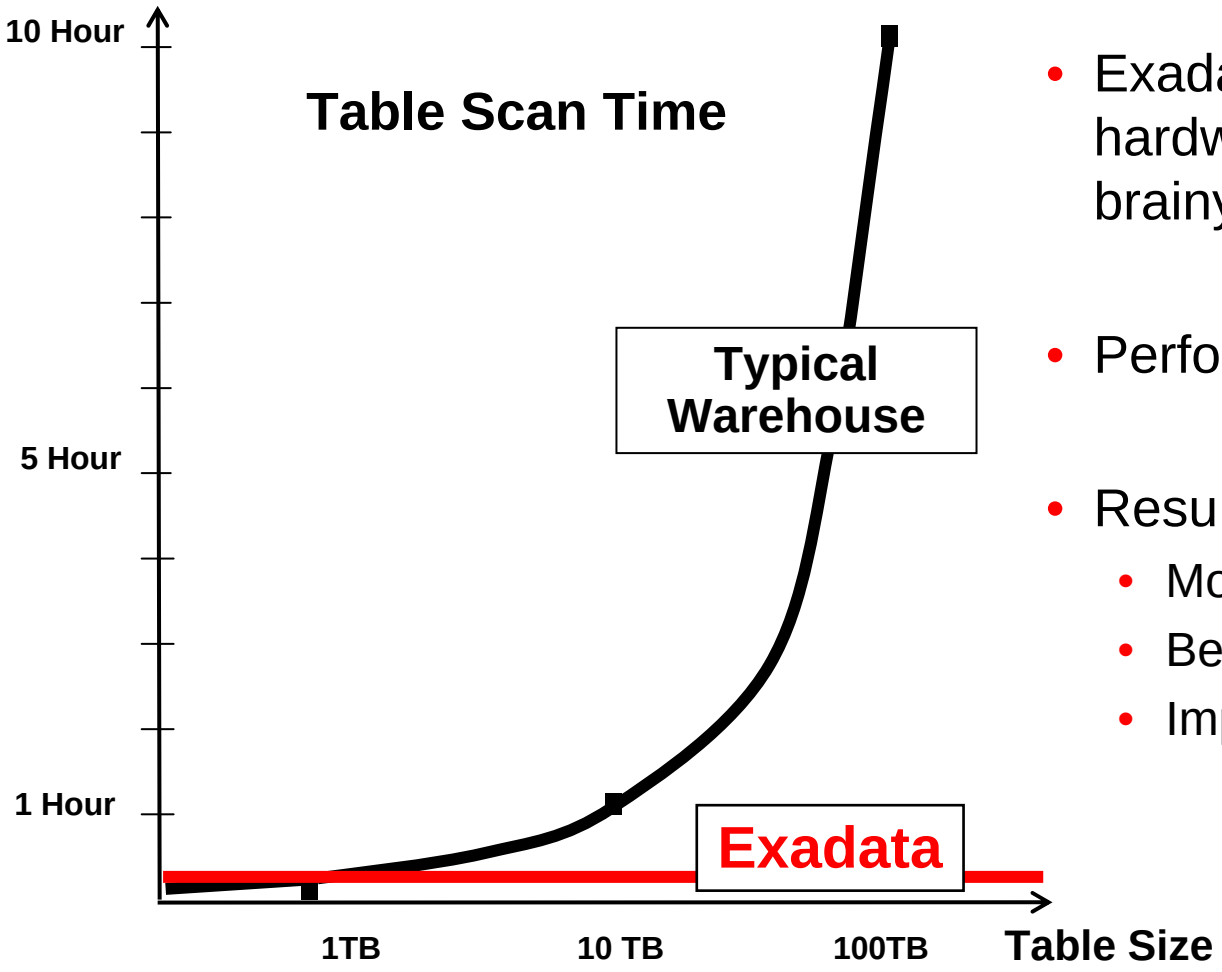
Massively Parallel Storage Grid

- Exadata Storage servers are organized into a massively parallel storage grid
- **Scalable**
  - Scales to hundreds of storage servers
  - Data automatically distributed across storage servers by ASM
    - Transparently redistributed when storage servers are added or removed
  - Data bandwidth scales linearly with capacity
- **Available**
  - Data is mirrored across storage servers
  - Failure of disk or storage server transparently tolerated
- **Simple**
  - Works transparently - no application changes



Exadata bandwidth scales linearly with capacity

Exadata Performance Scales

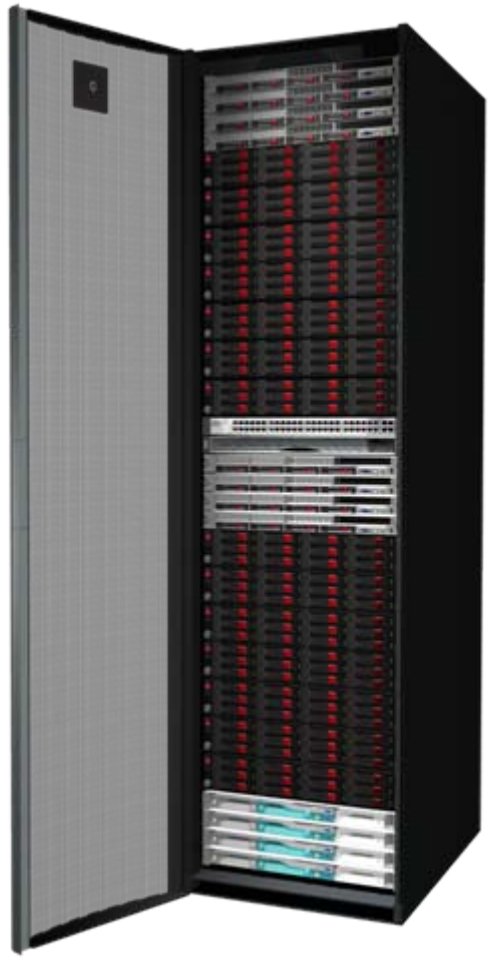


- Exadata delivers brawny hardware for use by Oracle's brainy software
- Performance scales with size
- Result
  - More business insight
  - Better decisions
  - Improved competitiveness

HP Oracle Database Machine

Pre-Configured High Performance Data Warehouse

- 8 DL360 Oracle Database servers
  - 2 quad-core Intel Xeon, 32GB RAM
  - Oracle Enterprise Linux
  - Oracle RAC
- 14 Exadata Storage Cells (SAS or SATA)
  - Up to 21 TB uncompressed user data (SAS)
  - Up to 46 TB uncompressed user data (SATA)
- 4 InfiniBand switches
- 1 Gigabit Ethernet switch
- Keyboard, Video, Mouse (KVM) hardware
- Hardware Warranty
  - 3 YR Parts/3 YR Labor/3 YR On-site
  - 24X7, 4 Hour response time



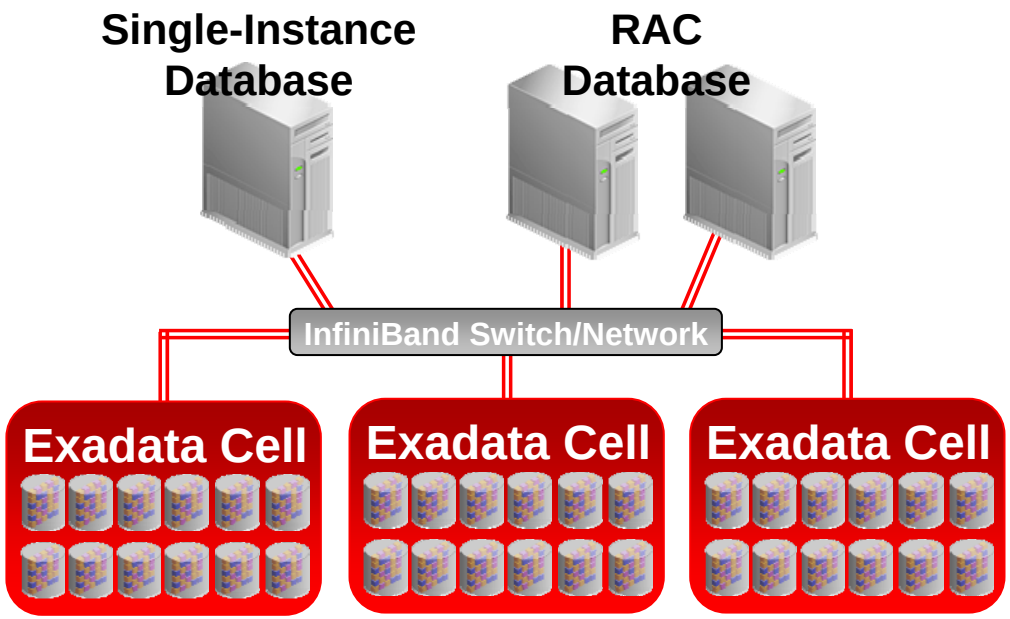
Add more racks for unlimited scalability

ORACLE

Exadata Product Capacity

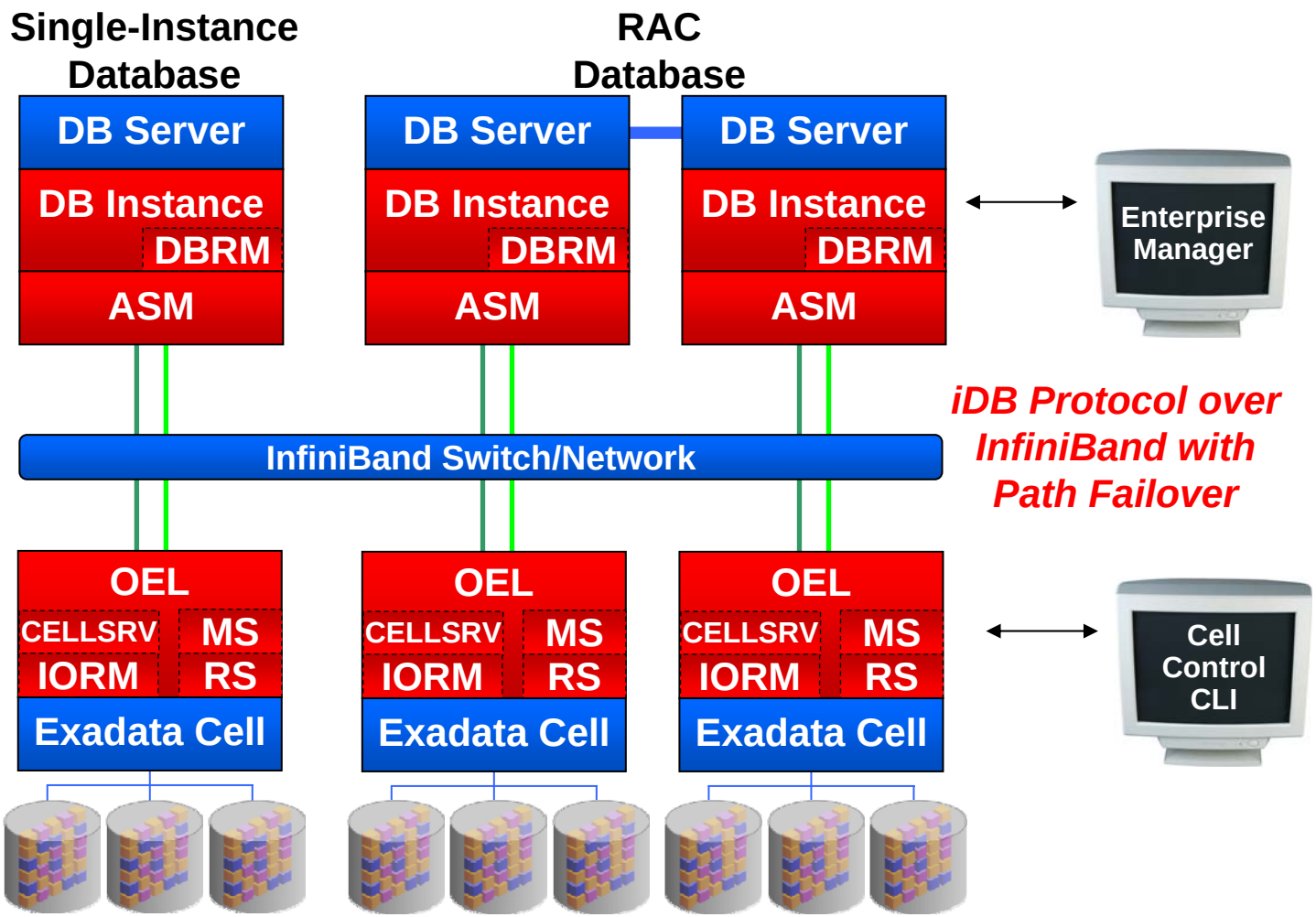
	Raw Storage	User Data	Data Bandwidth
HP Exadata Storage Server Hardware SAS	5.4 TB	1.5 TB	1 GB/s
HP Exadata Storage Server Hardware SATA	12 TB	3.3 TB	0.75 GB/s
HP Oracle Database Machine Hardware SAS	75 TB	21 TB	14 GB/s
HP Oracle Database Machine Hardware SAT	168 TB	46 TB	10.5 GB/s

Exadata Configuration



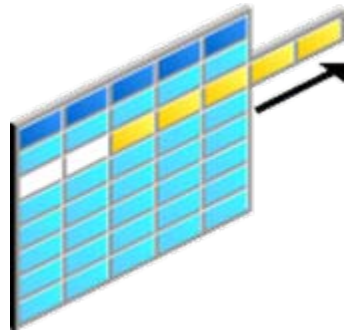
- Each Exadata Cell is a self-contained server which houses disk storage and runs the Exadata software
- Databases are deployed across multiple Exadata Cells
- Database enhanced to work in cooperation with Exadata intelligent storage
- No practical limit to number of Cells that can be in the grid

Exadata Architecture



Smart Scan Offload Processing

- Exadata Storage Servers implement smart scans to greatly reduce the data that needs to be processed by database hosts
  - Offload predicate evaluation
  - Only return relevant rows and columns to host
  - Join filtering



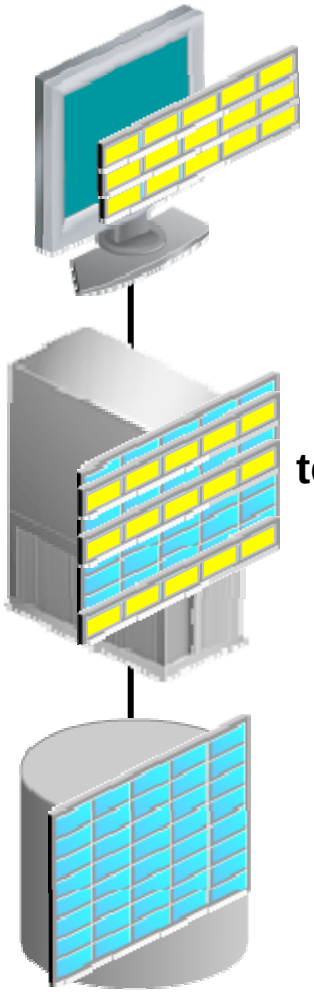
- Data reduction is usually very large
  - 10x data reduction is common
- Completely transparent
  - Even if a cell or disk fails during a query
- Smart Scan Example:
  - Telco wants to identify customers that spend more than $200 on a single phone call
  - The information about these premium customers occupies 2MB in a 1 terabyte table

Traditional Scan Processing

❶
```
SELECT
customer_name
  FROM calls
WHERE amount >
200;
```

❷
Table Extents Identified

❸
I/Os Issued



❹
I/Os Executed: 1 terabyte of data returned to hosts

❺
DB Host reduces terabyte of data to 1000 customer names that are returned to client

❻
Rows Returned

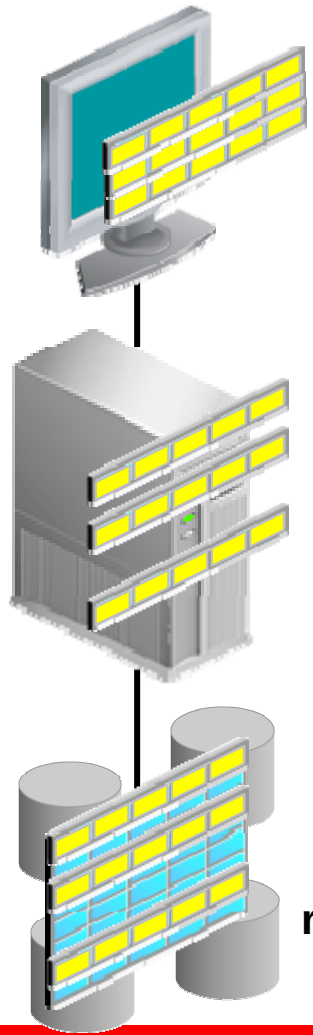
- With traditional storage, all database intelligence resides in the database hosts
- Very large percentage of data returned from storage is discarded by database servers
- Discarded data consumes valuable resources, and impacts the performance of other workloads

Exadata Smart Scan Processing

❶
```
SELECT
customer_name
  FROM calls
WHERE amount >
200;
```

❷ **Smart Scan Constructed And Sent To Cells**

❸ **Smart Scan identifies rows and columns within terabyte table that match request**



❹ **2MB of data returned to server**

❺ **Consolidated Result Set Built From All Cells**

❻ **Rows Returned**

- Only the relevant columns
  - customer_name

  and required rows
  - where amount>200

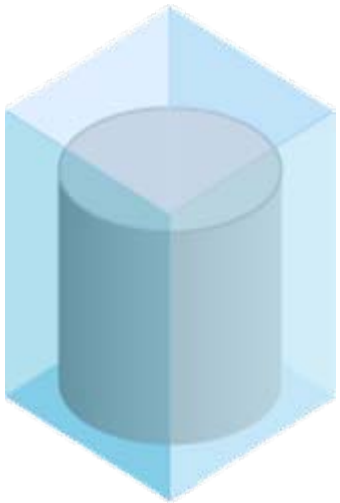
  are are returned to hosts
- CPU consumed by predicate evaluation is offloaded
- Moving scan processing off the database host frees host CPU cycles and eliminates massive amounts of unproductive messaging
  - Returns the needle, not the entire hay stack

Additional Smart Scan Functionality

- Join filtering
  - Star join filtering is performed within Exadata storage cells
  - Dimension table predicates are transformed into filters that are applied to scan of fact table
- Backups
  - I/O for incremental backups is much more efficient since only changed blocks are returned
- Create Tablespace (file creation)
  - Formatting of tablespace extents eliminates the I/O associated with the creation and writing of tablespace blocks

Smart Scan Transparency

- Smart scans are transparent to the application
  - No application or SQL changes required
  - Returned data is fully consistent and transactional
  - If a cell dies during a smart scan, the uncompleted portions of the smart scan are transparently routed to another cell

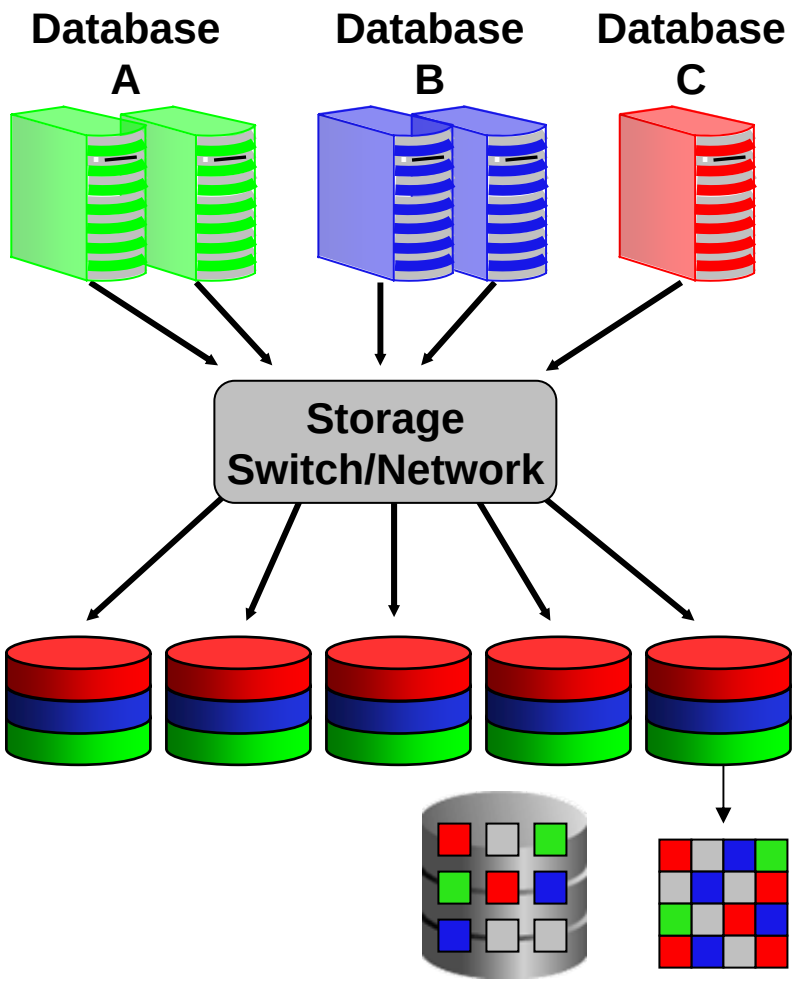


- Smart Scans correctly handle complex cases including
  - Uncommitted data and locked rows
  - Chained rows
  - Compressed tables
  - National Language Processing
  - Date arithmetic
  - Regular expression searches
  - Partitioned tables

High Throughput, Reduced Overhead, No Complex Tuning

Exadata Storage Grid

I/O Resource Management

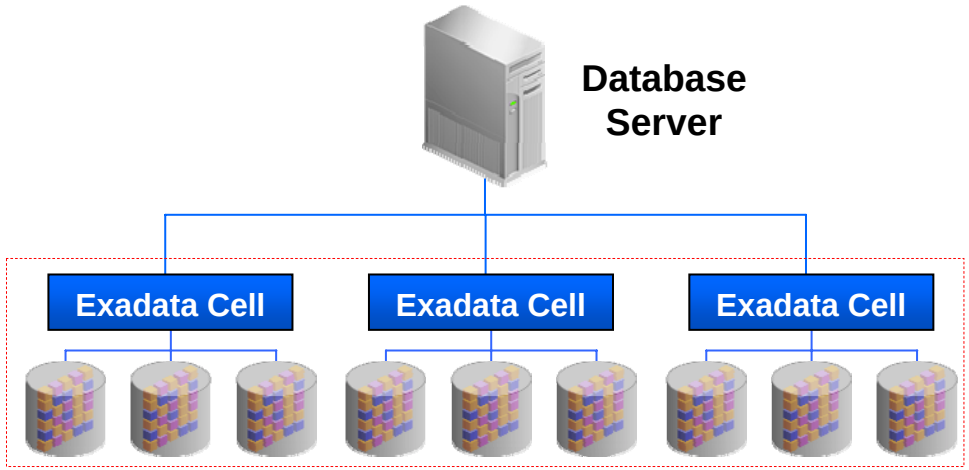


- With traditional storage,creating a managing shared storage is hampered by the inability to balance the work between users on the same database or on multiple databases sharing the storage subsystem
  - Hardware isolation is the approach to ensure separation
- Exadata I/O resource management ensures user defined SLAs are met
  - Coordination and prioritization between different groups/classes of work within a database and between databases

Exadata I/O Resource Management

DW and Mixed Workload Environments

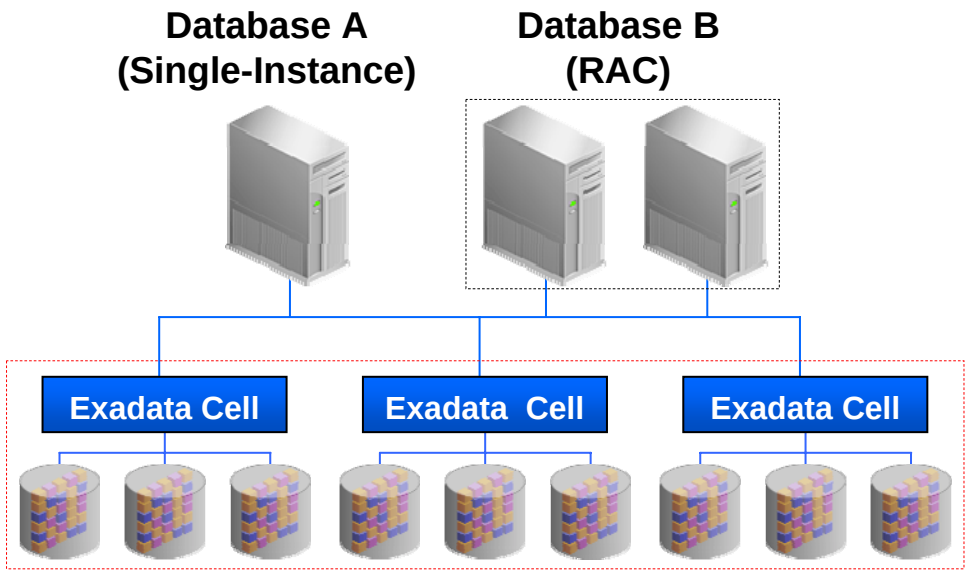
- Ensure different users and tasks within a database are allocated the correct relative amount of I/O resources
- For example:
  - Interactive: 50% of I/O resources
  - Reporting: 30% of I/O resources
  - ETL: 20% of I/O resources



Exadata I/O Resource Management

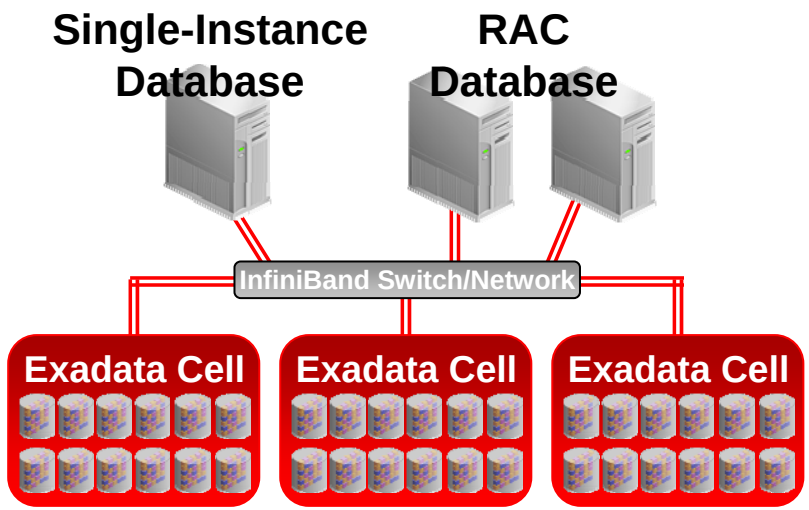
Multi-Database OLTP Environment

- Ensure different databases are allocated the correct relative amount of I/O bandwidth
  - Database A: 33% I/O of resources
  - Database B: 67% I/O of resources
- Ensure different users and tasks within a database are allocated the correct relative amount of I/O bandwidth
  - Database A:
    - Reporting: 60% of I/O resources
    - ETL: 40% of I/O resources
  - Database B:
    - Interactive: 30% of I/O resources
    - Batch: 70% of I/O resources



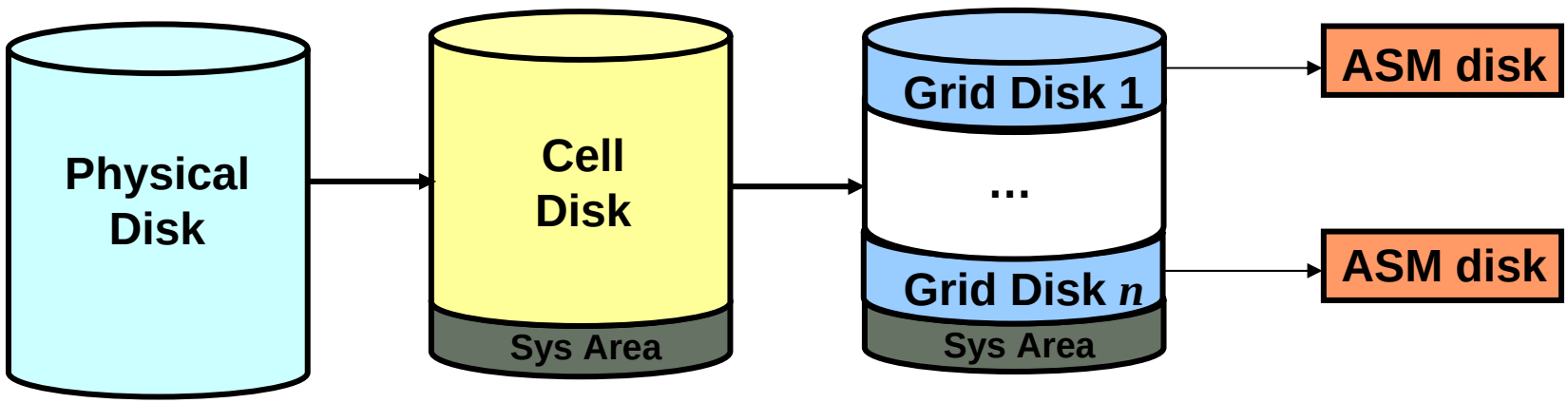
Exadata Scale-Out Storage Grid

- Dynamic virtualized storage resources using Automatic Storage Management (ASM)
  - Simple and non-intrusive resource allocation, and reallocation, enabling true enterprise grid storage
  - Database work spread across storage resources for optimal performance
- Powerful storage allocation options and management
  - Flexible configuration for performance and availability



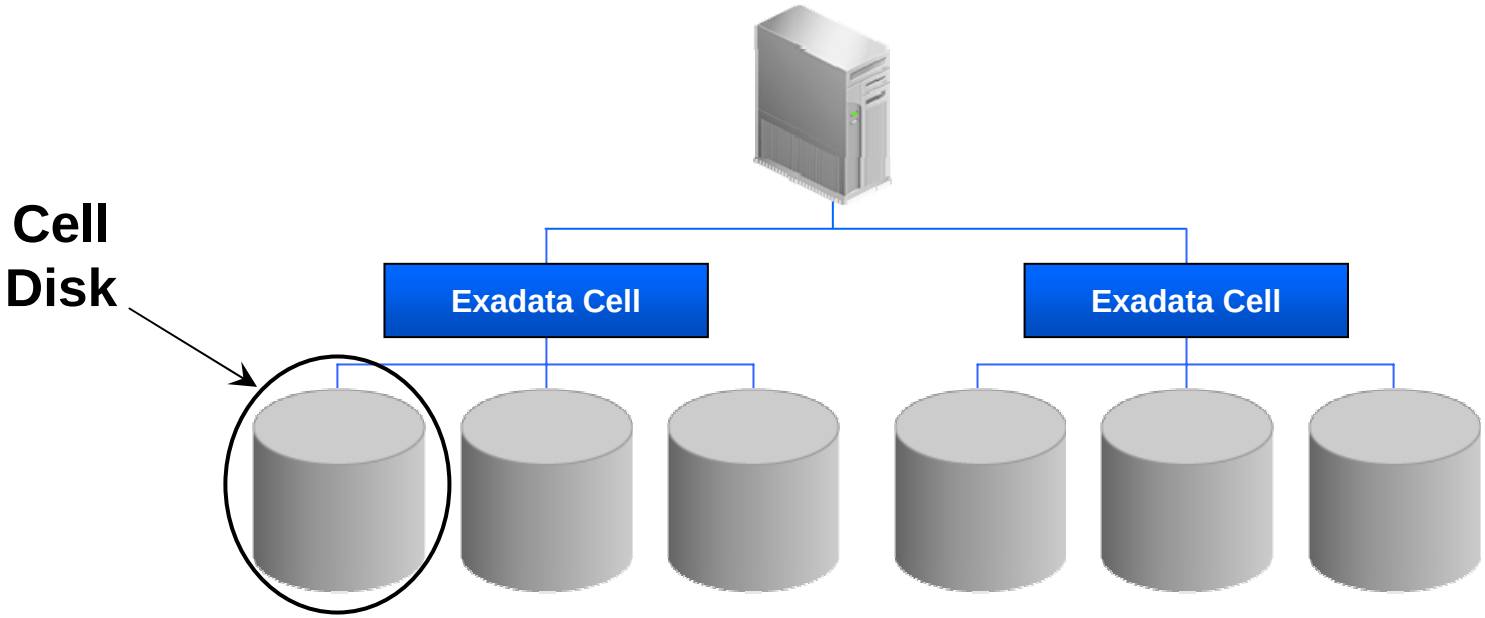
Exadata Storage Layout

- Physical disks map to a Cell Disks
- Cell Disks partitioned into one or multiple Grid Disks
- ASM diskgroups created from Grid Disks
- Transparent above the ASM layer



Exadata Storage Layout Example

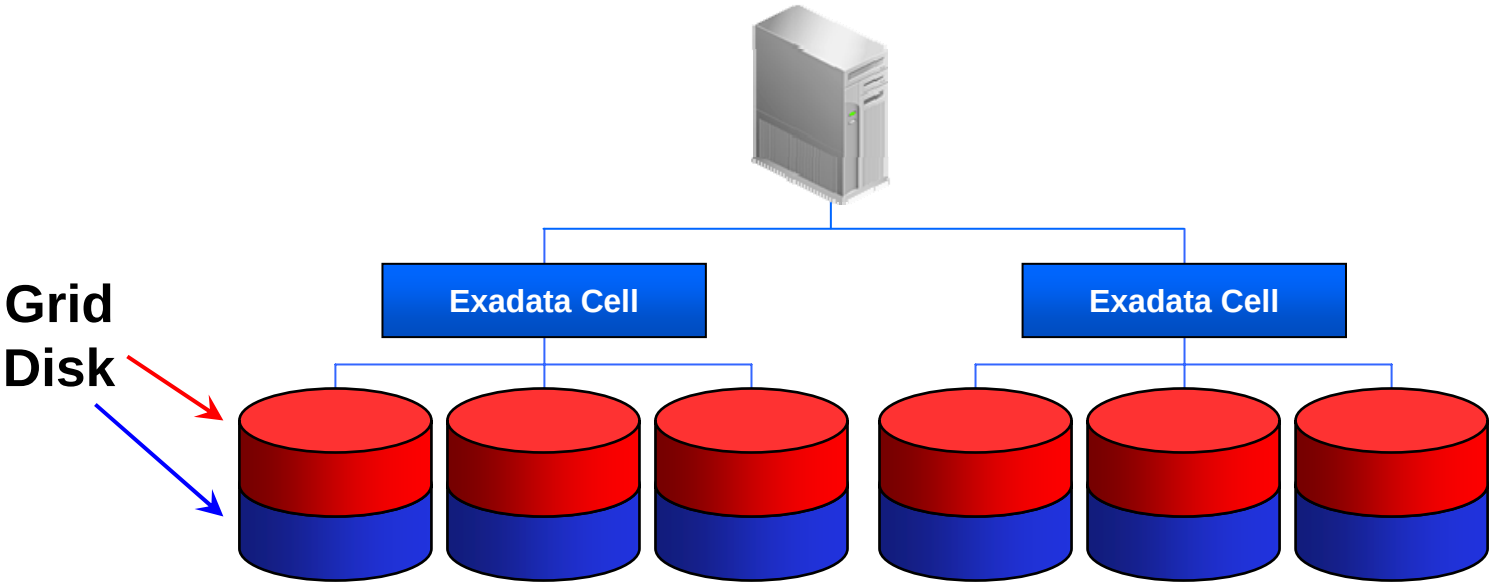
Cell Disks



- Cell Disk is the entity that represents a physical disk residing within a Exadata Storage Cell
  - Automatically discovered and activated

Exadata Storage Layout Example

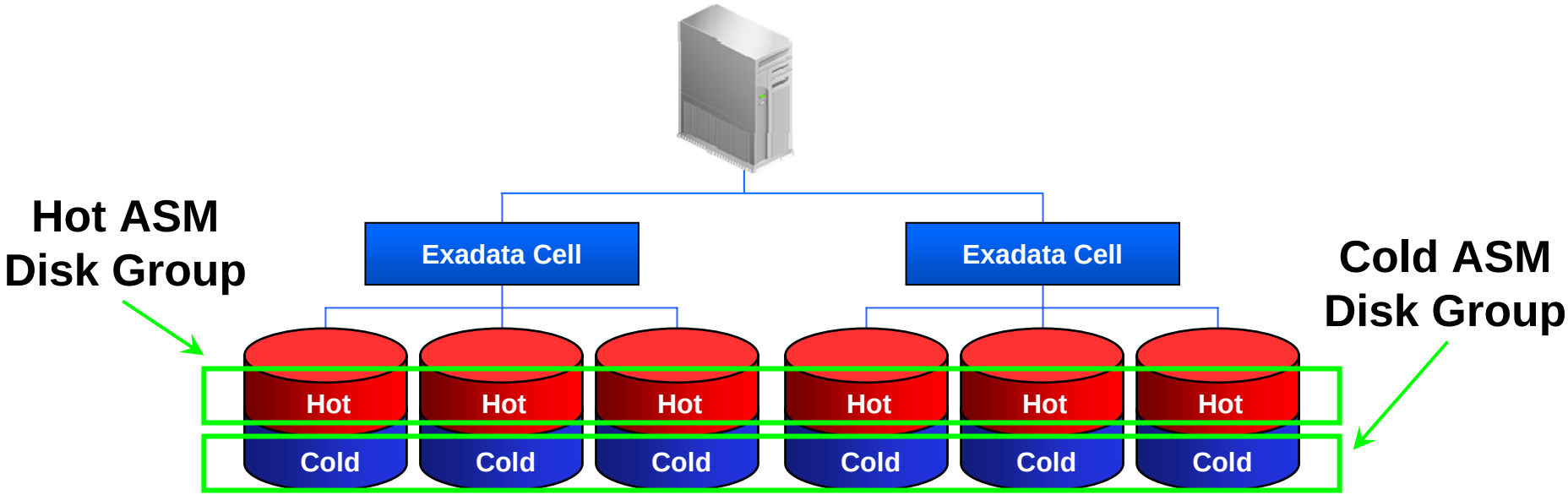
Grid Disks



- Cell Disks are logically partitioned into Grid Disks
  - Grid Disk is the entity allocated to ASM as an ASM disk
  - Minimum of one Grid Disk per Cell Disk
  - Can be used to allocate “hot”, “warm” and “cold” regions of a Cell Disk or to separate databases sharing Exadata Cells

Exadata Storage Layout Example

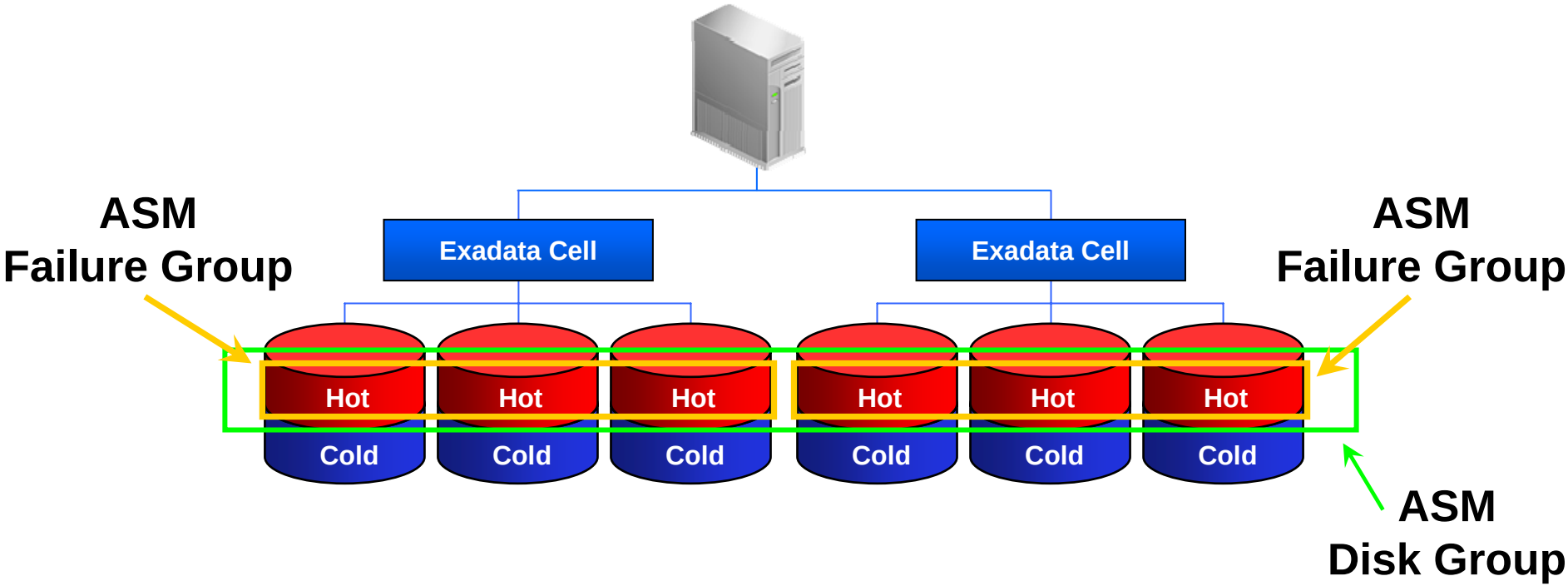
ASM Disk Groups and Mirroring



- Two ASM disk groups defined
  - One for the active, or “hot” portion, of the database and a second for the “cold” or inactive portion
- ASM striping evenly distributes I/O across the disk group
- ASM mirroring is used protect against disk failures
  - Optional for one or both disk groups

Exadata Storage Layout Example

ASM Mirroring and Failure Groups



- ASM mirroring is used protect against disk failures
- ASM failure groups are used to protect against cell failures

Exadata Storage Management & Administration

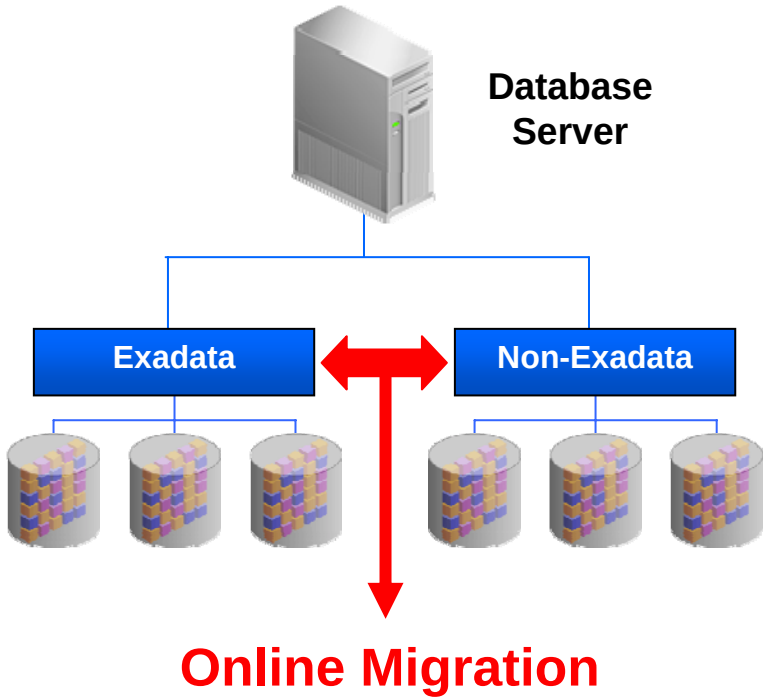
- Enterprise Manager
  - Manage & administer Database and ASM
- Exadata Storage Plug-in
  - Enterprise Manager Grid Control Plug-in to monitor & manage Exadata Storage Cells
- Comprehensive CLI
  - Local Exadata Storage cell management
  - Distributed shell utility to execute CLI across multiple cells
- Lights-out 100
  - Remote management and administration of hardware

Data Protection Solutions

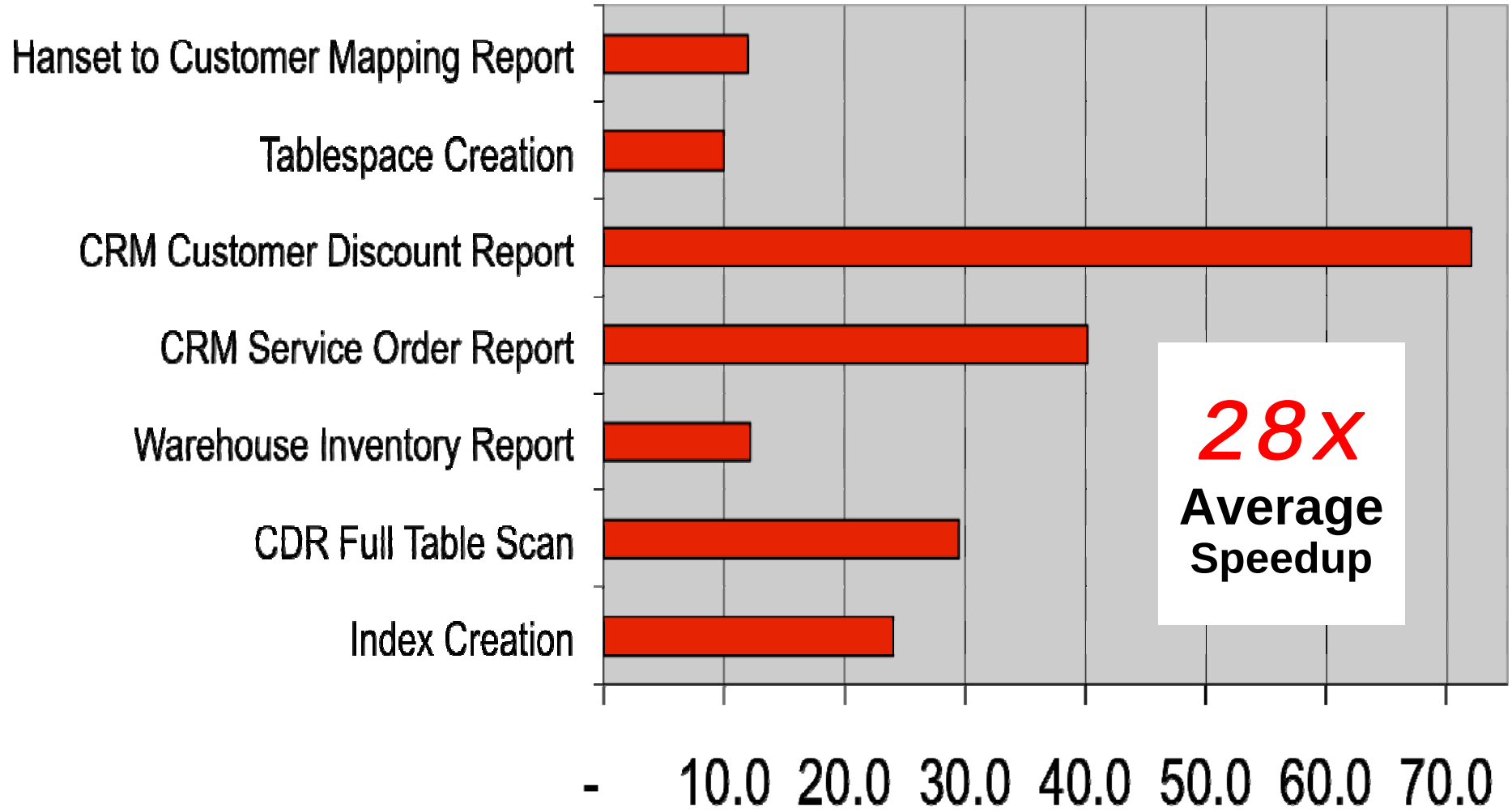
- All single points of failure eliminated by the Exadata Storage architecture
- Hardware Assisted Resilient Data (HARD) built in to Exadata Storage
  - Prevent data corruption before it happens
- Data Guard provides disaster protection and data corruption protection
  - Automatically maintained second copy of database
- Flashback provides human error protection
  - Snapshot-like capabilities to rewind database to before error
- Recovery Manager (RMAN) provide backup to disk
  - Archiving and corruption protection
  - Can be used with Oracle Secure Backup (OSB) or third party tape backup software
- These work just as they do for traditional non-Exadata storage
  - Users and database administrator use familiar tools

Exadata Co-Existence and Migration

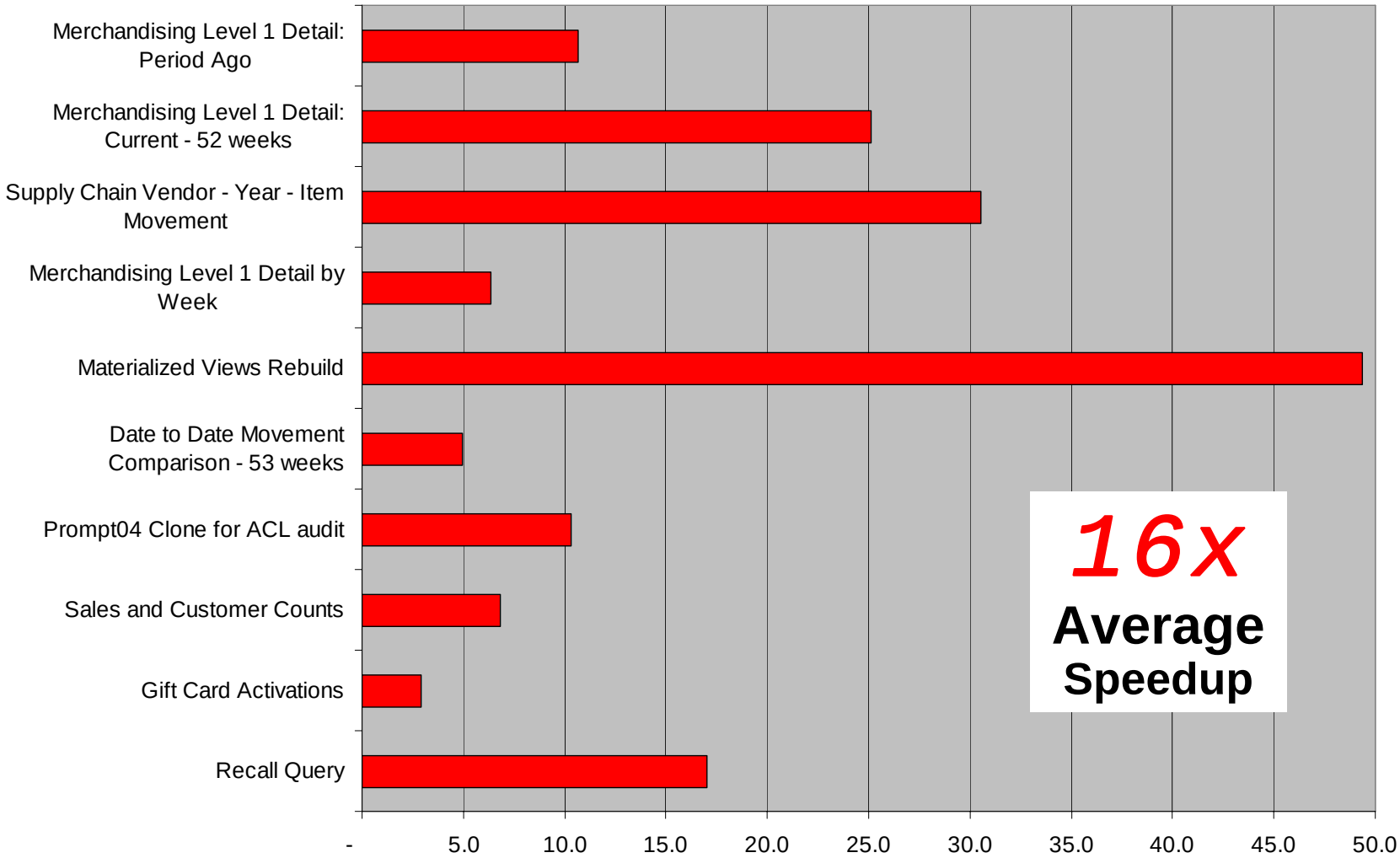
- Databases can be concurrently deployed on Exadata and traditional storage
  - Tablespaces can exist on Exadata storage, traditional storage, or a combination of the two, and is transparent to database applications
  - SQL offload processing requires all pieces of a tablespace reside on Exadata
- Online migration if currently using ASM and ASM redundancy
- Migration can be done using RMAN or Data Guard



Telco Exadata Speedup – 10X to 72X



Retailer Exadata Speedup – 3x to 50x



Exadata Benefits

- Extreme Performance
  - **10X to 100X** speedup for data warehousing
- Database Aware Storage
  - Smart Scans
- Massively Parallel Architecture
  - Dynamically Scalable to hundreds of cells
  - Linear Scaling of Data Bandwidth
  - Transaction/Job level Quality of Service
- Mission Critical Availability and Protection
  - Disaster recovery, backup, point-in-time recovery, data validation, encryption

Resources

- **Oracle.com:**
  http://www.oracle.com/exadata
- **Oracle Exadata Technology Portal on OTN:**
  http://www.oracle.com/technology/products/bi/db/exadata
- **Oracle Exadata white papers:**
  http://www.oracle.com/technology/products/bi/db/exadata/pdf/exadata-technical-whitepaper.pdf

  http://www.oracle.com/technology/products/bi/db/exadata/pdf/migration-to-exadata-whitepaper.pdf