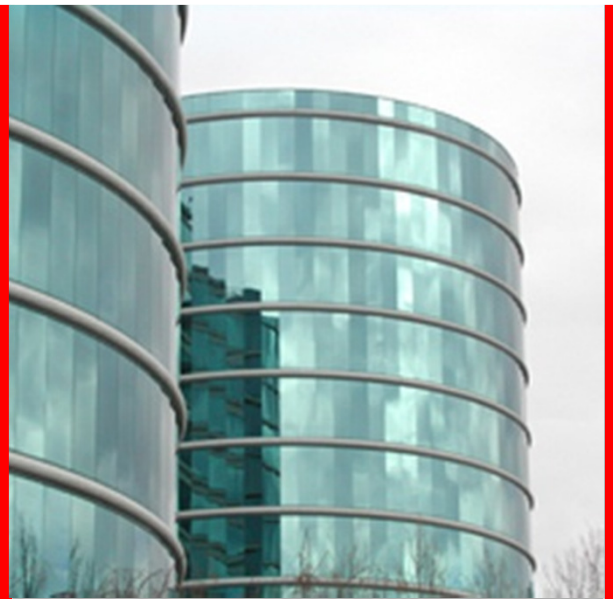


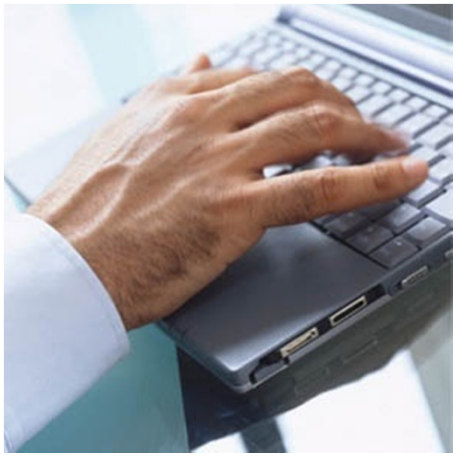
ORACLE®

Oracle Exadata Database Machine Overview

Srikant Gopalan
Senior Solution Specialist
Oracle Public Sector

The following is intended to outline our general capabilities and product direction. It is intended for information purposes only, and may not be incorporated into any contract. It is not a commitment to deliver any material, code, or functionality, and should not be relied upon in making purchasing decisions. The development, release, and timing of any features or functionality described for Oracle's products remains at the sole discretion of Oracle.

Agenda



- Exadata Hardware Architecture
- Exadata DB Machine Architecture Overview
- Exadata Storage Software Features
- Exadata Demos

Exadata Hardware Architecture

Scaleable Grid of industry standard servers for Compute and Storage

- Eliminates long-standing tradeoff between Scalability, Availability, Cost

Database Grid

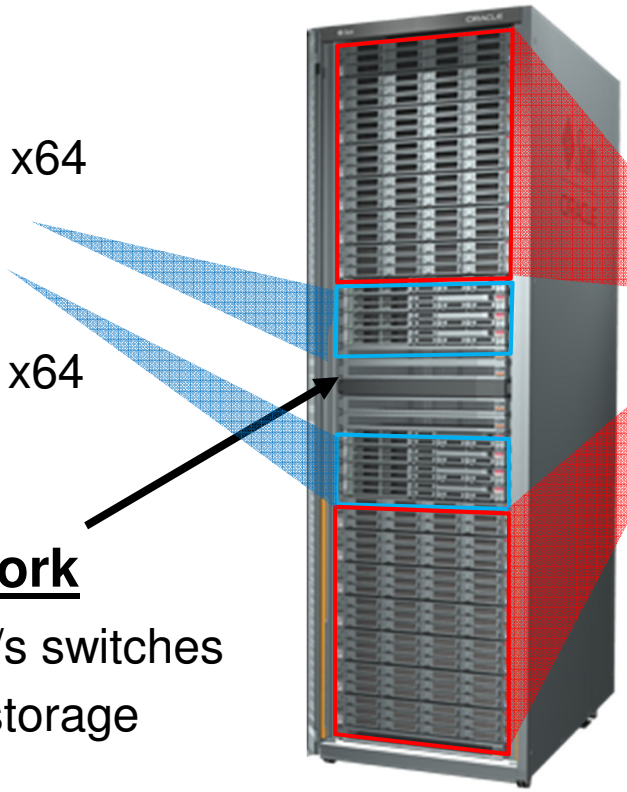
- 8 Dual-processor x64 database servers

OR

- 2 Eight-processor x64 database servers

InfiniBand Network

- Redundant 40Gb/s switches
- Unified server & storage network



Intelligent Storage Grid

- 14 High-performance low-cost storage servers

- 100 TB **High Performance** disk, or
  336 TB **High Capacity** disk
- **5.3 TB PCI Flash**
- Data mirrored across storage servers

Exadata Database Machine X2-2 Full Rack

Pre-Configured for Extreme Performance

- 8 x64 Dual-procesor Database Servers
  - 96 cores (12 per server)
  - 768 GB memory (96GB per server)

- 14 Exadata Storage Servers X2-2
  - 168 cores (12 per server)
  - High Performance Disks: 45 TB of usable capacity (full rack)

  OR
  - High Capacity Disks: 150 TB of usable capacity (full rack)

Add more racks for additional scalability

New - Exadata Database Machine X2-8 Full Rack

Extreme Performance for Consolidation, Large OLTP and DW

- 2 x64 Eight-processor Database servers
  - High Core, High Memory Database Servers
  - 128 CPU cores (64 per server)
  - 2 TB memory (1 TB per server)

- 14 Exadata Storage Servers X2-2
  - 168 cores (12 per server)
  - High Performance Disks: 45 TB of usable capacity (full rack)

  OR
  - High Capacity Disks: 150 TB of usable capacity (full rack)

Add more racks for additional scalability

Exadata Storage Server Building Block

- **Hardware by Sun**
- **Software by Oracle**

- High-performance storage server built from industry standard components
- 12 disks - 600 GB 15000 RPM High Performance SAS or 2TB 7200 RPM High Capacity SAS
- 2 Six-Core Intel Xeon Processors (L5640). 168 Storage Cores for the cluster
- Dual ported 40 Gb/sec InfiniBand
- 4 x 96 GB Flash Cards
- Intelligent Exadata Storage Server Software

Start Small and Grow

Field Upgradeable

Balanced Incremental Scaling for OLTP and DW

Scale to 8 Racks by Just Adding Cables

Full Bandwidth and Redundancy

InfiniBand Network

- Unified InfiniBand Network
  - Storage Network
  - RAC Interconnect
  - External Connectivity (optional)
- High Performance, Low Latency Network
  - 80 Gb/s bandwidth per link (40 Gb/s each direction)
  - SAN-like Efficiency (Zero copy, buffer reservation)
  - Simple manageability like IP network
- Protocols
  - Zero-copy Zero-loss Datagram Protocol (ZDP RDSv3)
    - Linux Open Source, Low CPU overhead (Transfer 3 GB/s with 2% CPU usage)
  - Internet Protocol over InfiniBand (IPoIB)
    - Looks like normal Ethernet to host software (tcp/ip, udp, http, ssh,…)

Database Server Operating System Choices

- Two Operating System Choices on the database servers
  - Oracle Linux
  - Solaris 11 Express (x86)
- Customers will choose their preferred Database Server OS at installation time
- Exadata Storage Servers will continue to be Oracle Linux

Exadata Product Capacity

		X2-8 Full Rack	X2-2 Full Rack	X2-2 Half Rack	X2-2 Quarter Rack
Raw Disk[1]	High Perf Disk	100 TB	100 TB	50 TB	21 TB
	High Cap Disk	336 TB	336 TB	168 TB	72 TB
Raw Flash[1]		5.3 TB	5.3 TB	2.6 TB	1.1 TB

1 – Raw capacity calculated using 1 GB = 1000 x 1000 x 1000 bytes and 1 TB = 1000 x 1000 x 1000 x 1000 bytes.

Exadata Product Performance

		X2-8 Full Rack	X2-2 Full Rack	X2-2 Half Rack	X2-2 Quarter Rack
Raw Disk Data Bandwidth[1,4]	High Perf Disk	25 GB/s	25 GB/s	12.5 GB/s	5.4 GB/s
	High Cap Disk	14 GB/s	14 GB/s	7 GB/s	3 GB/s
Raw Flash Data Bandwidth[1,4]		50 GB/s	50 GB/s	25 GB/s	11 GB/s
Disk IOPS[3,4]	High Perf Disk	50,000	50,000	25,000	10,800
	High Cap Disk	25,000	25,000	12,500	5,400
Flash IOPS[3,4]		1,000,000	1,500,000	500,000	225,000
Data Load Rate[4]		5 TB/hr	12 TB/hr	2.5 TB/hr	1 TB/hr

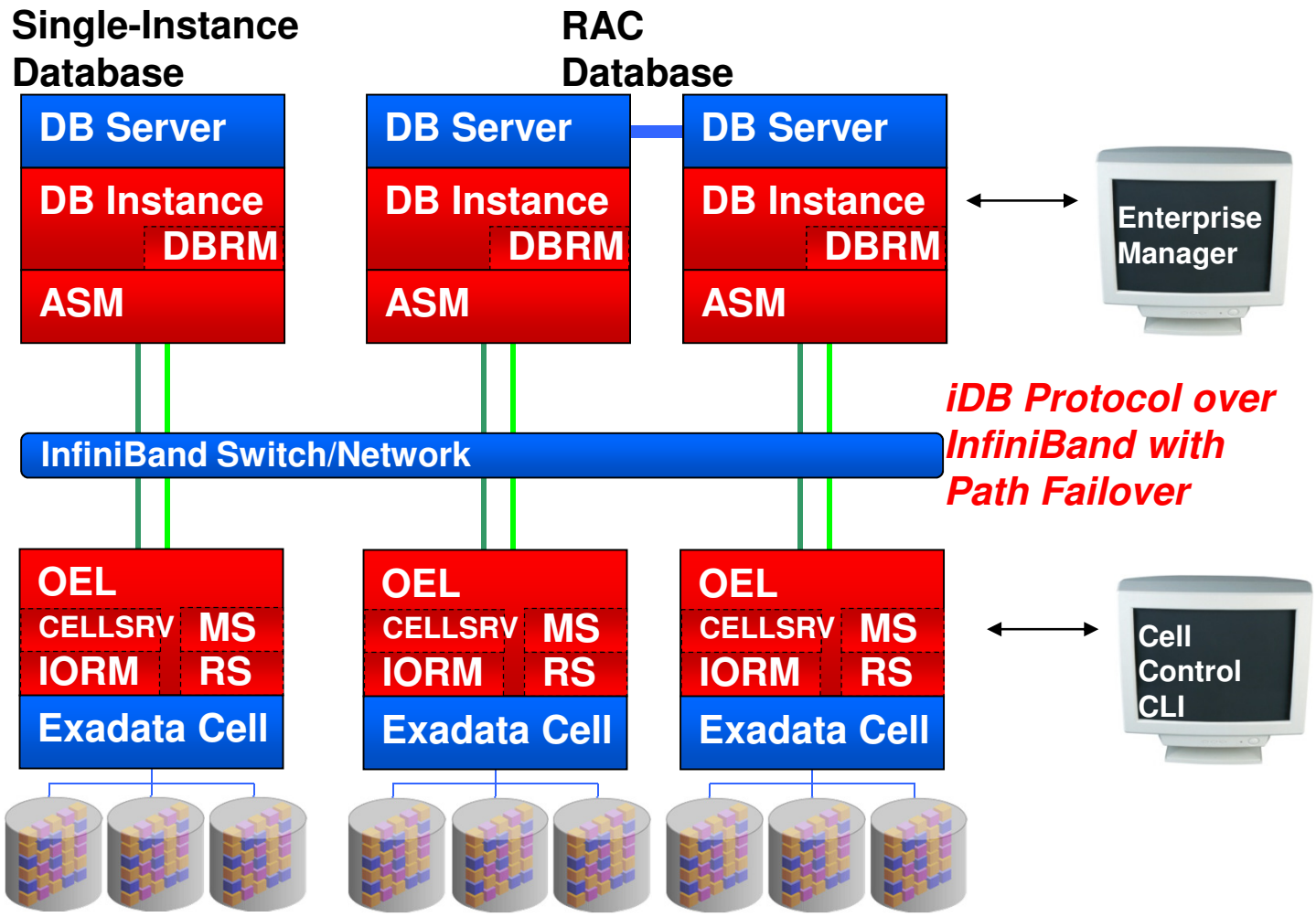
1 – Bandwidth is peak physical disk scan bandwidth, assuming no compression.

2 - Max User Data Bandwidth assumes scanned data is compressed by factor of 10 and is on Flash.

3 – IOPs – Based on IO requests of size 8K

4 - Actual performance will vary by application.

Exadata Architecture



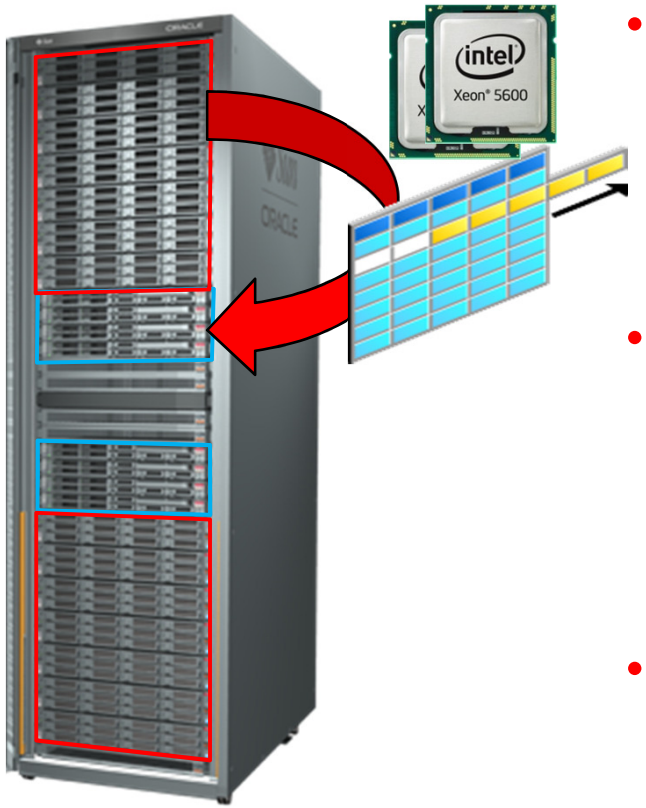
Exadata Cell Components

- Code running on the Exadata cells
  - cellserv
    - Primary component of cell
    - Provides advance sql offload capabilities - smartscan
  - ms – management server
    - Primary management interface
    - Facilitates cellcli commands and EM plug-in
    - Provides information for IORM
  - rs – restart server
    - Ensures ongoing operation of Exadata software
  - IORM - IO Resource Manager –
    - used in conjunction with DB Resource Manager
- iDB protocol
  - iscsi like I/O built using ZDP protocol
  - Runs in DB kernel
  - Eliminates unecessary copying of blocks

- Disk concepts:
  - Lun – physical disk in cell
  - Cell disk – lun presentation in Exadata
  - Grid disk – lun presentation for ASM

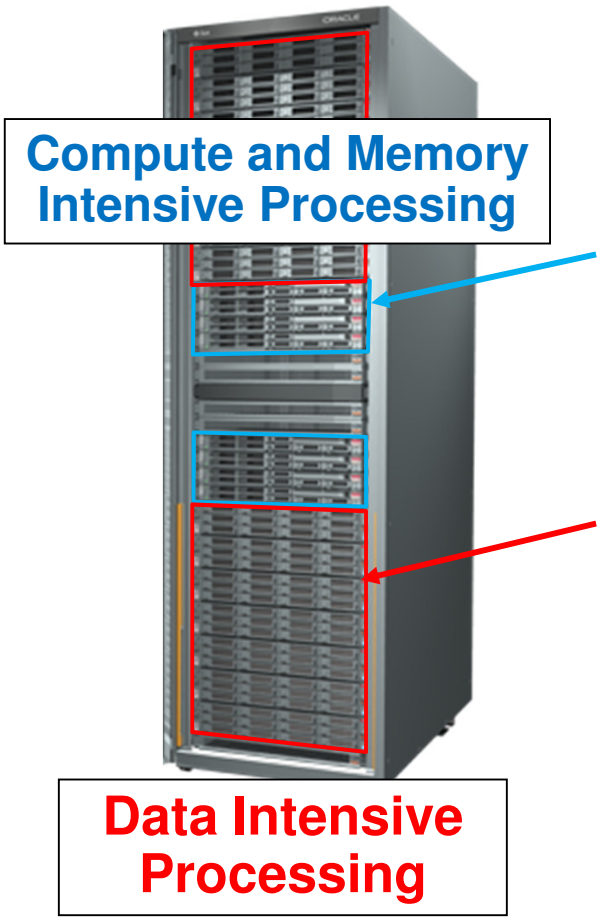
Exadata Intelligent Storage Grid

Most Scalable Data Processing



- **Data Intensive processing runs in Exadata Storage Grid**
  - Filter rows and columns as data streams from disks (168 Intel Cores)
- Example: How much product X sold last quarter
  - Exadata Storage Reads 10TB from disk
  - Exadata Storage Filters rows by Product & Date
  - Sends 100GB of matching data to DB Servers
- Scale-out storage parallelizes execution and removes bottlenecks

Exadata is Smart Storage



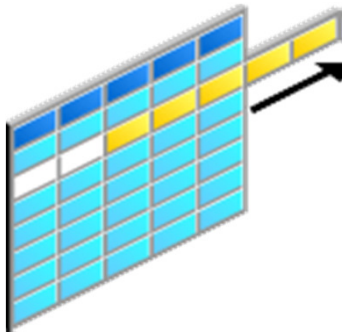
- Storage Server is smart storage, not a DB node
  - Storage remains an independent tier
- **Database Servers**
  - Perform complex database processing such as joins, aggregation, etc.
- **Exadata Storage Servers**
  - Search tables and indexes filtering out data that is not relevant to a query
  - Cells serve data to multiple databases **enabling OLTP and consolidation**
  - Simplicity, and robustness of storage appliance

Exadata Storage Software Unique Features

- Exadata Smart Scans
  - 10X or greater reduction in data sent to database servers
- Hybrid Columnar Compression
  - Efficient compression increases effective storage capacity and increases user data scan bandwidths by a factor of up to 10X
- Exadata Smart Flash Cache
  - Breaks random I/O bottleneck by increasing IOPs by up to 20X
  - Doubles user data scan bandwidths
- I/O Resource Manager (IORM)
  - Enables storage grid by prioritizing I/Os to ensure predictable performance

Exadata Smart Scans

- 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



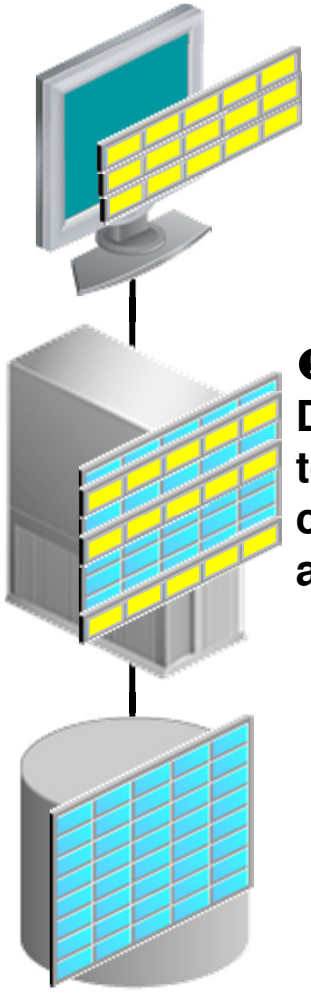
- Data reduction is usually very large
  - 10x data reduction is common
- Completely transparent
  - Even if a cell or disk fails during a query
- Smart Scan Example:
  - Telco wants to identify customers that spend more than $200 on a single phone call
  - The information about these premium customers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

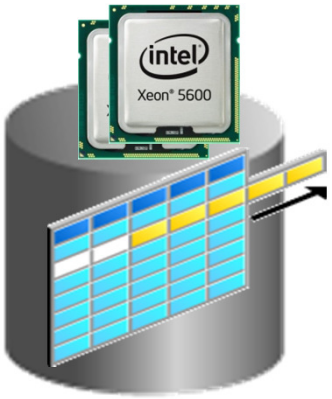
❻ **Rows Returned**

- 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

Exadata Intelligent Storage

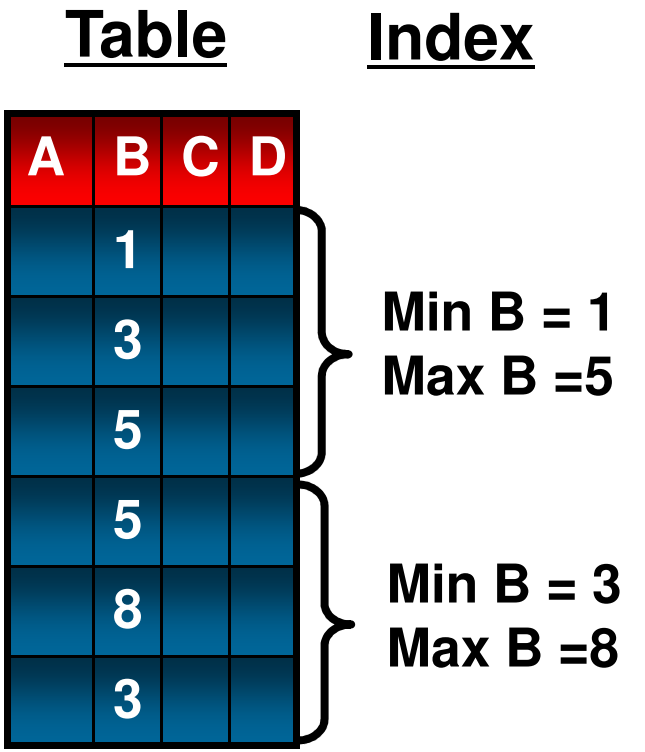


Exadata Intelligent Storage Grid

- Exadata storage servers also run more complex operations in storage
  - **Join filtering**
  - **Incremental backup filtering**
  - **I/O prioritization**
  - **Storage Indexing**
  - **Database level security**
  - **Offloaded scans on encrypted data**
  - **Data Mining Model Scoring**
- 10x reduction in data sent to DB servers is common

Exadata Storage Index

Transparent I/O Elimination with No Overhead

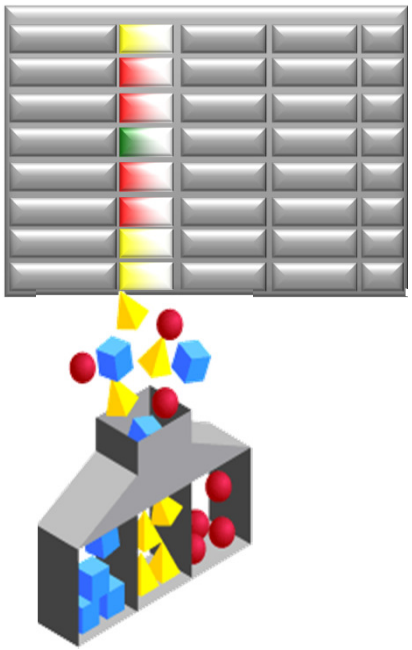


- Exadata Storage Indexes maintain summary information about table data in memory
  - Store MIN and MAX values of columns
  - Typically one index entry for every MB of disk
- Eliminates disk I/Os if MIN and MAX can never match “where” clause of a query
- Completely **automatic and transparent**

Select * from Table where B<2 - Only first set of rows can match

Exadata Hybrid Columnar Compression

Highest Capacity, Lowest Cost



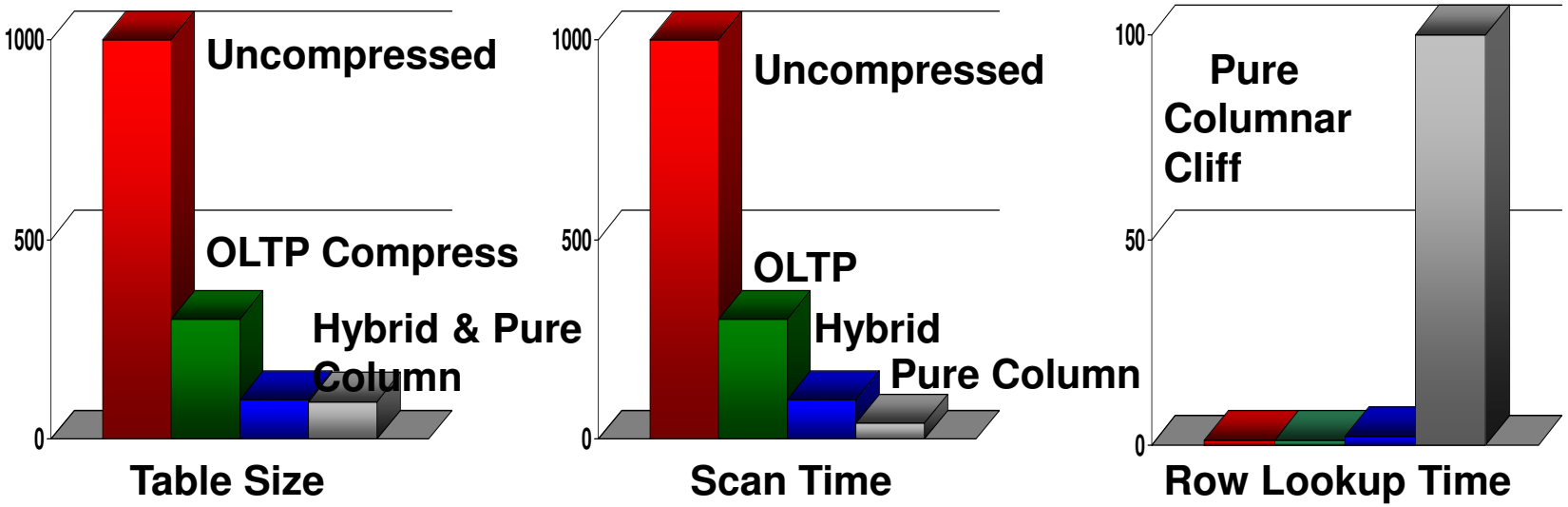
- Data is organized and compressed by column
  - Dramatically better compression
- Speed Optimized **Query Mode** for Data Warehousing
  - **10X compression typical**
  - **Runs faster because of Exadata offload!**
- Space Optimized **Archival Mode** for infrequently accessed data
  - **15X to 50X compression typical**

Faster and Simpler

Backup, DR, Caching, Reorg, Clone

Benefits Multiply

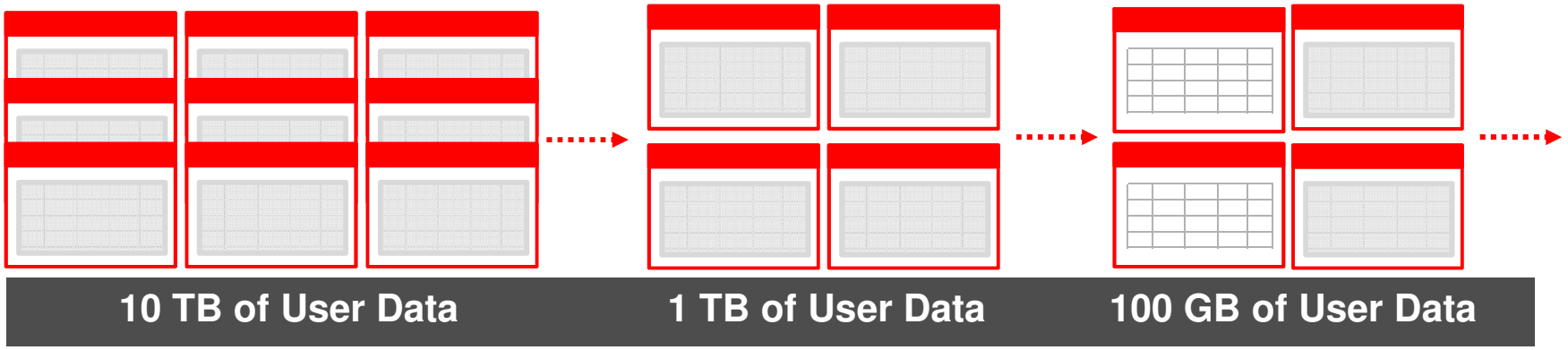
Hybrid Columnar Comparisons



- Hybrid Columnar Compression is a second generation columnar technology combining the best of row and column formats
  - Best compression – matching full columnar
  - Excellent scan time – 93% as good as full columnar
  - Good single row lookup – no full columnar "cliff"
- Row format remains best for workloads with updates or trickle feeds

Benefits Multiply

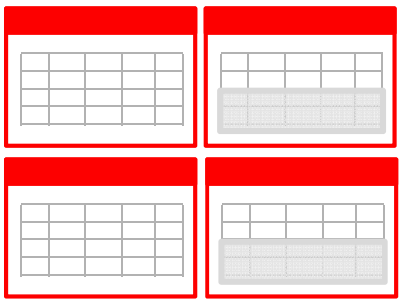
Converting Terabytes to Gigabytes

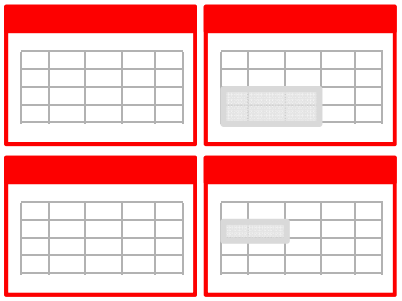


10 TB of User Data → **1 TB of User Data** → **100 GB of User Data**

With 10x Compression

With Partition Pruning





20 GB of User Data → **5 GB of User Data** → **Sub second "10 TB" Scan**

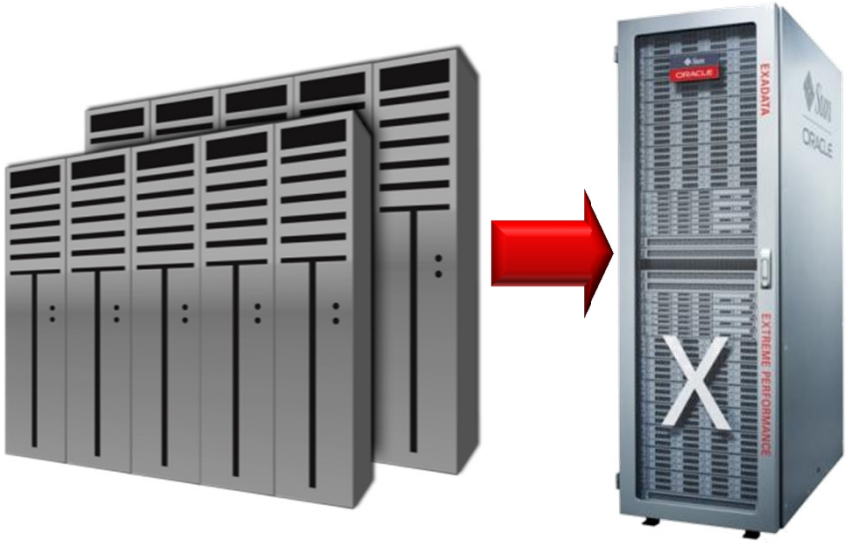
With Storage Indexes

With Smart Scan

No Indexes

Exadata Smart Flash Cache

Extreme Performance OLTP & DW

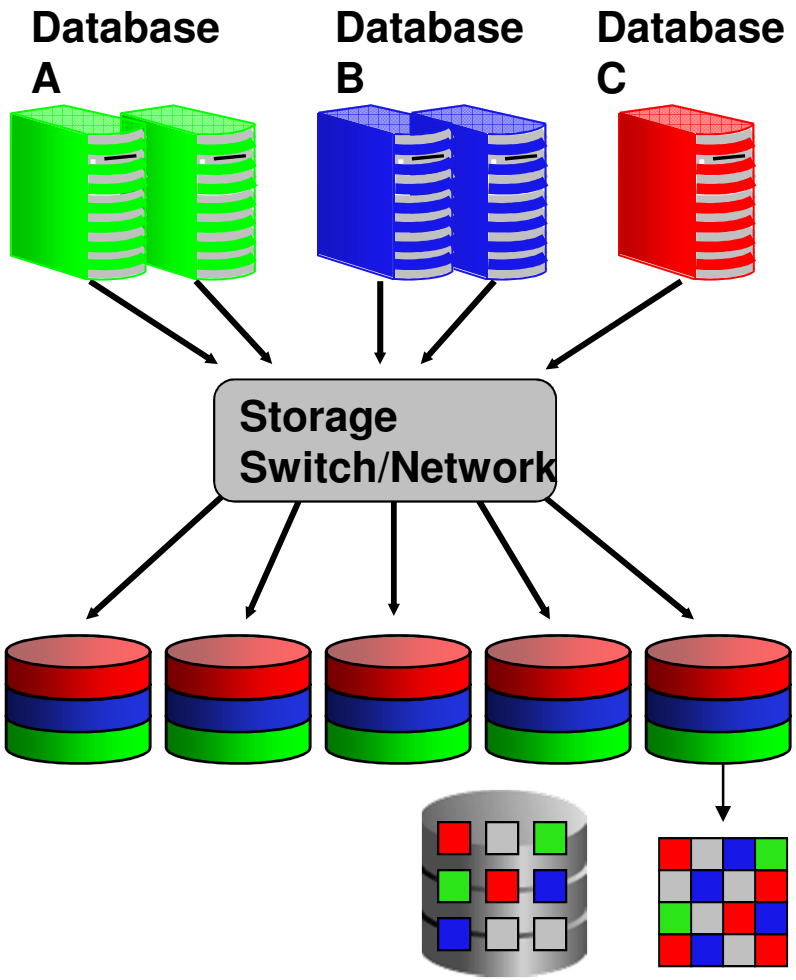


5X More I/Os than 1000 Disk Enterprise Storage Array

- Exadata has **5 TB** of flash
  - **56 Flash PCI cards avoid disk controller bottlenecks**
- **Intelligently manages flash**
  - Smart Flash Cache holds hot data
    - Avoids large scan wipe-outs of cache
  - **Gives speed of flash, cost of disk**
- Exadata flash cache achieves:
  - Over **1 million IO/sec from SQL** (8K)
  - Sub-millisecond response times

Exadata Storage Grid

I/O Resource Management

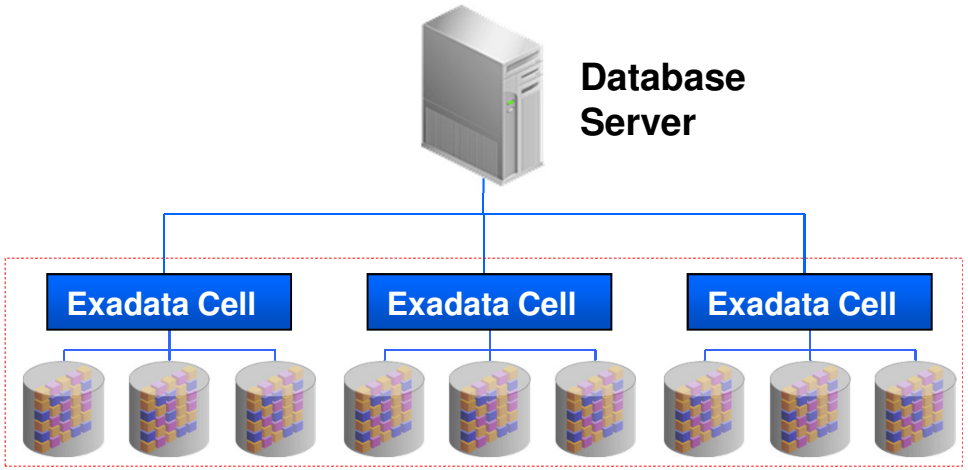


- With traditional storage,creating a managing shared storage is hampered by the inability to balance the work between users on the same database or on multiple databases sharing the storage subsystem
  - Hardware isolation is the approach to ensure separation
- Exadata I/O resource management ensures 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

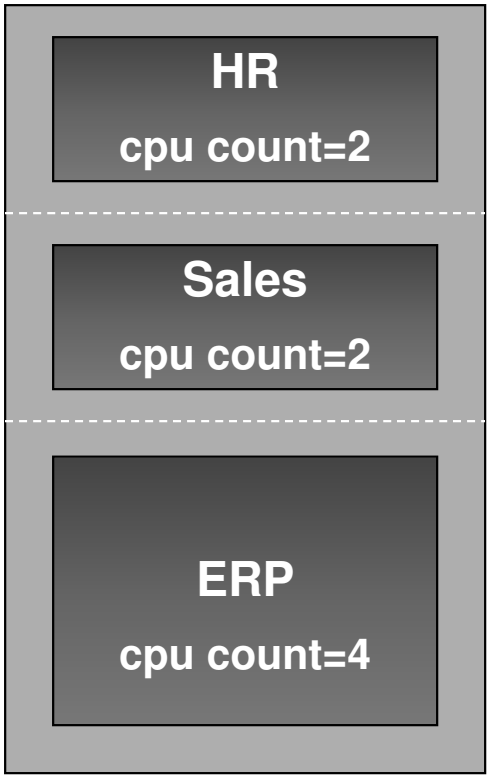
- Ensure different users and tasks within a database are allocated the correct relative amount of I/O resources
- For example:
  - Interactive: 50% of I/O resources
  - Reporting: 30% of I/O resources
  - ETL: 20% of I/O resources



Resource Isolation

Instance Caging

- Enables cpu core limits for instances on shared server
- Protects service levels by preventing runaway cpu consumption
- Can be adjusted dynamically, while databases are online.
  - Controlled by cpu_count parameter
  - Supports partitioning and overprovisioning cpu
- Works with Resource Manager



8 core server

First Secure Database Machine

- Moves decryption from software to hardware
  - Over 5x faster
- Near zero overhead for fully encrypted database
- Queries decrypt data at hundreds of Gigabytes/second


Q&A