



ORACLE®



Oracle Exadata V2: A Technical Overview

**Marshall Presser
David Partridge**



The Architecture of the Future

Massively Parallel Grid



**Best for Data
Warehousing**

Best for OLTP

Best for Consolidation



ORACLE



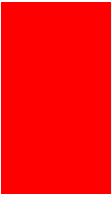
Sun & Oracle Partnership Advantage

A Legacy of Joint Market Leadership

- ✓ 20+ year relationship in sales & service
- ✓ Excellence in cooperative customer support
- ✓ Leading platform for Oracle Database
- ✓ Leading platform for Oracle Applications
- ✓ Leading UNIX platform for Oracle
- ✓ Top Java/J2EE partner



ORACLE®



The following is intended to outline our general 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 remain at the sole discretion of Oracle.



ORACLE®



Agenda

- Overview – Business Benefits of Exadata
- Exadata Based Product Offerings
- Exadata Architecture and Features
- Best Data Warehousing Machine
- Best OLTP Machine
- Best Consolidation Machine



ORACLE®



Business Value of Exadata

Extreme Performance

- Data Warehousing performance improvements of 10-100X
- OLTP performance improvements of 20X

Linear Scalability

- Performance scales linearly with increase in data volumes

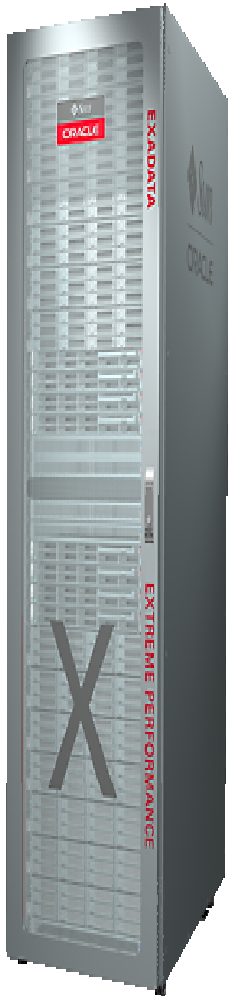
Enterprise Ready

- Get up and running quickly with a complete system
- Single Oracle Point Of Contact for all hardware and software support
- No changes to applications required



ORACLE®

Drastically Simplified Deployments



- Eliminates complexity
- Ready on day one
 - Pre-built, tested, standard, supportable configuration
 - **Runs existing applications unchanged**
- Extreme performance out-of-the-box

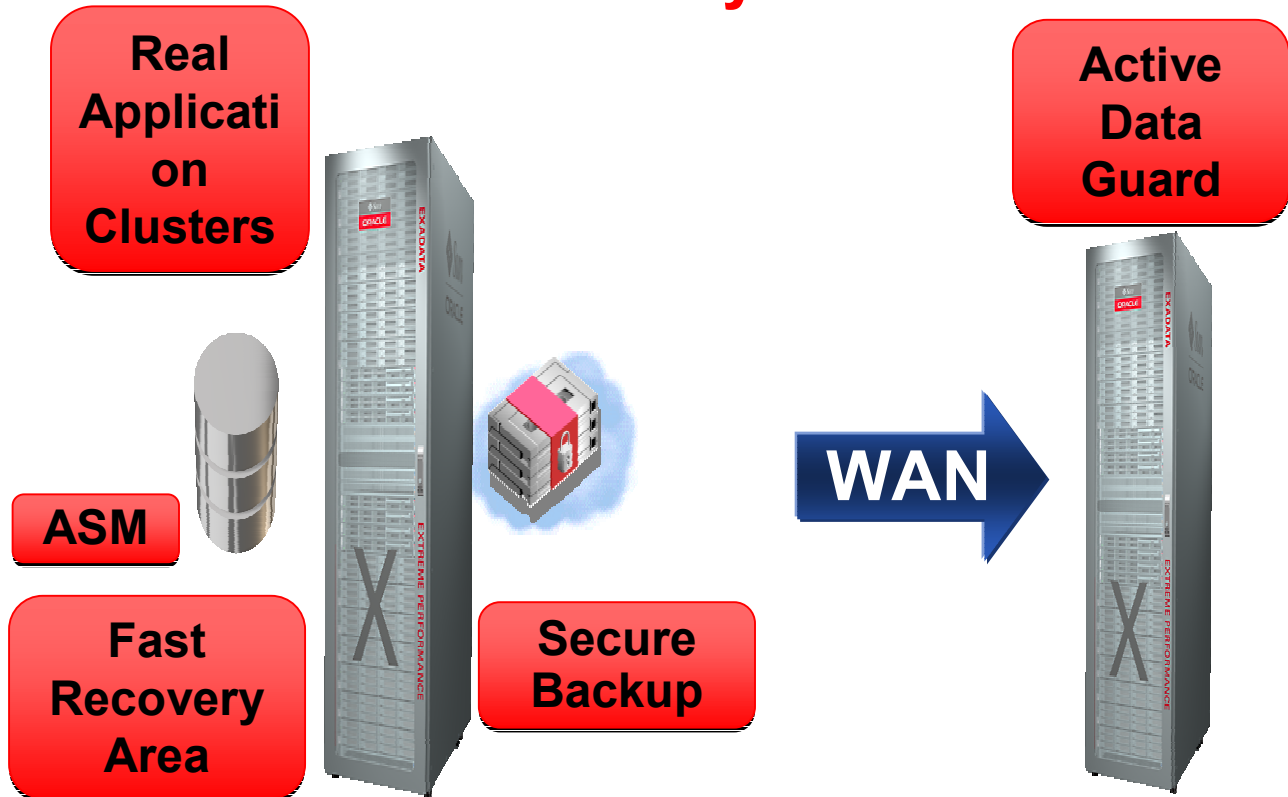
Months
to
Days



ORACLE

Complete, Open, Integrated Availability

Maximum Availability Architecture



- Protection from
 - Server Failures
 - Storage Failures
 - Network Failures
 - Site Failures
- Real-time remote standby open for queries
- Human error correction
 - Database, table, row, transaction level
- Online indexing and table redefinition
- Online patching and upgrades



Data Protection Solutions

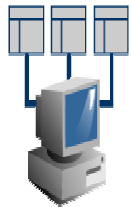
- All single points of failure eliminated by the Exadata Storage architecture
- Hardware Assisted Resilient Data (HARD) built in to Exadata Storage
 - Prevent data corruption before it happens
- Data Guard provides disaster protection and data corruption protection
 - Automatically maintains one or more copies of the database
- Flashback provides human error protection
 - Snapshot-like capabilities to rewind database to before error
- Recovery Manager (RMAN) provides backup to disk
 - Archiving and corruption protection
 - Compatible with Oracle Secure Backup (OSB) or third party tape backup
- Work the same as for traditional non-Exadata storage
 - Users and database administrator use familiar tools



ORACLE

Complete, Open, Integrated Security

Monitoring



Configurat
ion
Managem
ent



Audit
Vault



Total
Recall

Access Control



Databa
se
Vault

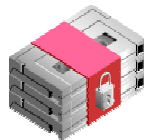
Secret

Label
Securit
y

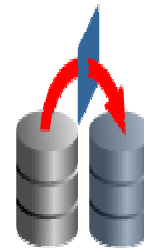
Encryption and Masking



Advanced
Security



Secur
e
Backu
p



Data
Maskin
g



ORACLE

Database Machine Success

Representative customers in all geographies and industries



ORACLE®

The Products

Exadata Storage Server and Database Machine

Exadata Storage Server

- Exadata Storage Server
 - Storage product optimized for Oracle Database
 - Extreme I/O and SQL processing performance
 - Combination of hardware and software
- Exadata Storage Server Software

ORACLE
EXADATA



Sun Oracle Database Machine

- Sun Oracle Database Machine
 - Pre-configured high performance
 - Balanced performance configuration
 - Straight-forward Oracle deployment
- Exadata Storage Server Software
- Oracle Database 11g Release 2



ORACLE

Sun Oracle Database Machine

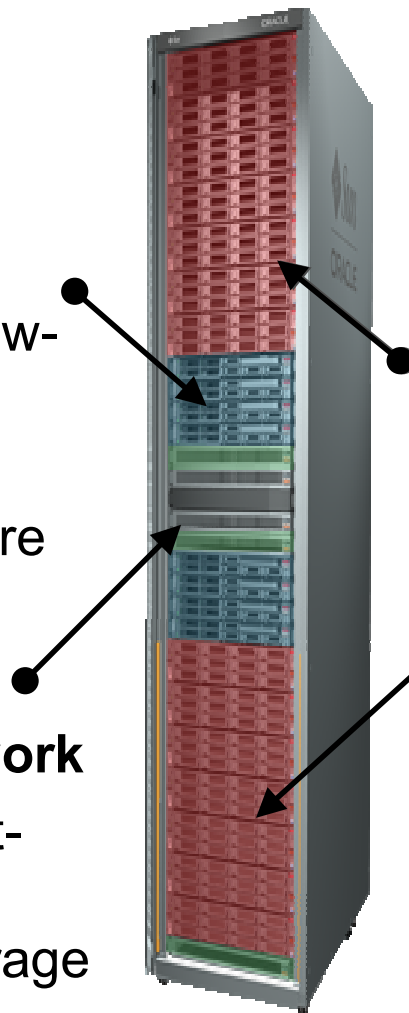
First and only complete grid architecture for all data management needs

RAC Database Server Grid

- 8 High-performance low-cost compute servers
- 2 Intel quad-core Xeons each

InfiniBand Network

- 40 Gb/sec fault-tolerant unified server and storage network



Exadata Storage Server Grid

- 14 High-performance low-cost storage servers
- 100 TB raw SAS disk storage or 336 TB raw SATA disk storage
- **5TB+ flash storage!**



ORACLE

Start Small and Grow



Quarter Rack

Half Rack

Full Rack



ORACLE

Scale Performance and Capacity



- Redundant and Fault Tolerant
 - Failure of any component is tolerated
 - Data is mirrored across storage servers
- Scalable
 - Scales to 8 rack database machine by just adding wires
 - More with external InfiniBand switches
 - Scales to hundreds of storage servers for multi-petabyte databases



ORACLE®

Sun Exadata Storage Server Hardware

Sun Exadata Storage Server Hardware



Hardware
by



Software
by

ORACLE®

- Building block of Exadata Storage Grid
 - Up to 1.5 GB/sec raw data bandwidth per cell
 - Up to 75,000 IOPS with Flash
- Sun Fire™ X4275 Server
 - 2 Quad-Core Intel® Xeon® E5540 Processors
 - 24GB RAM
 - Dual-port 4X QDR (40Gb/s) InfiniBa card
 - Disk Options
 - 12 x 600 GB SAS disks (7.2 TB total)
 - 12 x 2TB SATA disks (24 TB total)
 - 4 x 96 GB Sun Flash PCIe Cards (3 GB total)
- Software pre-installed
 - Oracle Exadata Storage Server Software
 - Oracle Enterprise Linux
 - Drivers, Utilities
- Single point of support from Oracle



ORACLE®

Sun Exadata Storage Server Hardware



Dual-redundant, hot-swappable power supplies

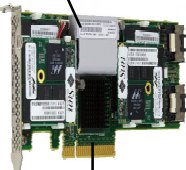
24 GB DRAM

ILOM

Disk Controller HBA with 512M battery backed cache

12 x 3.5" Disk Drives

2 Quad-Core Intel® Xeon® Processors



InfiniBand QDR (40Gb/s) dual port card

4 x 96GB Sun Flash PCIe Cards

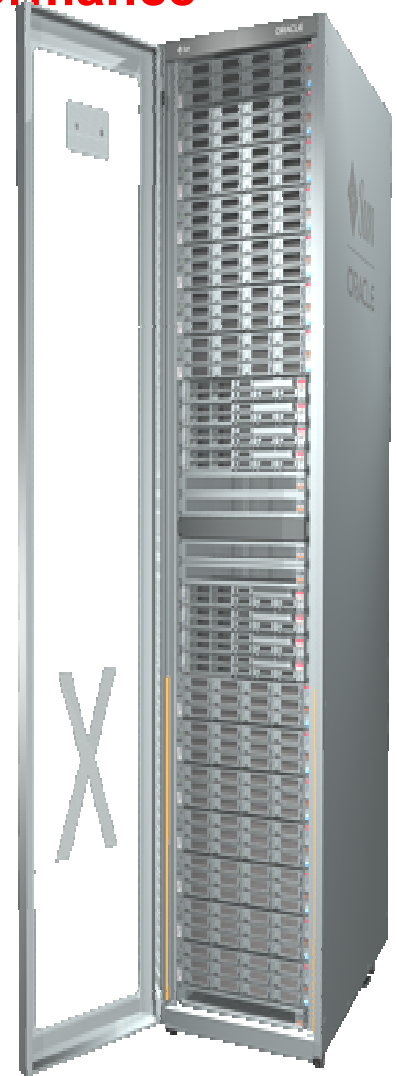


ORACLE

Sun Oracle Database Machine Full Rack

Pre-Configured for Extreme Performance

- 8 Sun Fire™ X4170 Oracle Database servers
- 14 Exadata Storage Servers (All SAS or all SATA)
- 3 Sun Datacenter InfiniBand Switch 36
 - 36-port Managed QDR (40Gb/s) switch
- 1 “Admin” Cisco Ethernet switch
- Keyboard, Video, Mouse (KVM) hardware
- Redundant Power Distributions Units (PDUs)
- Single Point of Support from Oracle
 - 3 year, 24 x 7, 4 Hr On-site response



Add more racks for additional scalability



ORACLE

Exadata Product Capacity

		Single Server	Quarter Rack	Half Rack	Full Rack
Raw Disk ¹	SA S	7.2 TB	21 TB	50 TB	100 TB
	SAT A	24 TB	72 TB	168 TB	336 TB
Raw Flash ¹		384 GB	1.1 TB	2.6 TB	5.3 TB
User Data ² (assuming no compression)	SA S	2 TB	6 TB	14 TB	28 TB
	SAT A	7 TB	21 TB	50 TB	100 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.

2 - User Data: Actual space for end-user data, computed after single mirroring (ASM normal redundancy) and after allowing space for database structures such as temp, logs, undo, and indexes. Actual user data capacity varies by application. User Data capacity calculated using 1 TB = 1024 * 1024 * 1024 * 1024 bytes.



ORACLE

Exadata Product Performance

		Single Server	Quarter Rack	Half Rack	Full Rack
Raw Disk Data Bandwidth ^{1,4}	SAS	1.5 GB/s	4.5 GB/s	10.5 GB/s	21 GB/s
	SATA	0.85 GB/s	2.5 GB/s	6 GB/s	12 GB/s
Raw Flash Data Bandwidth ^{1,4}		3.6 GB/s	11 GB/s	25 GB/s	50 GB/s
Max User Data Bandwidth ^{2,4} (10x compression & Flash)		36 GB/s	110 GB/s	250 GB/s	500 GB/s
Disk IOPS ^{3,4}	SAS	3,600	10,800	25,000	50,000
	SATA	1,440	4,300	10,000	20,000
Flash IOPS ^{3,4}		75,000	225,000	500,000	1,000,000
Data Load Rate ⁴		0.65 TB/hr	1 TB/hr	2.5 TB/hr	5 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.

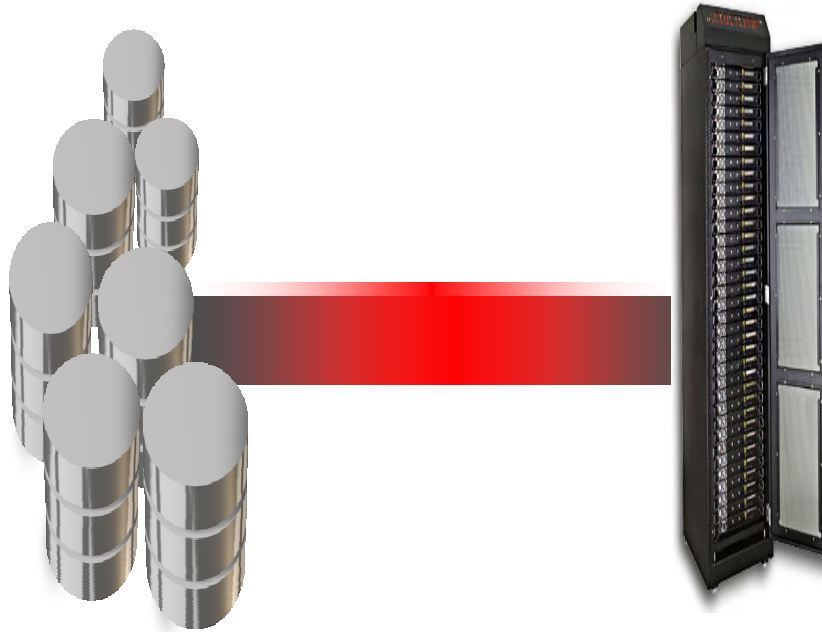
3 – IOPs – Based on IO requests of size 8K

4 - Actual performance will vary by application.



ORACLE

Storage Bottlenecks



- Today, database performance is limited by storage
 - Storage systems limit data bandwidth from storage to servers
 - Storage Array internal bottlenecks
 - SAN bottlenecks
 - Random I/O bottlenecks due to physical disk speeds
- Data bandwidth limits restrict data warehousing performance
- Random I/O bottlenecks limit OLTP performance



ORACLE®

Exadata Smart Storage

Solves Data Bandwidth and Random I/O Problems

- Massively parallel storage grid
 - High performance Exadata storage servers (cells)
 - Data bandwidth scales with data volume
- Offloads data intensive processing
 - Queries run in storage as data streams from disk, offloading database server CPUs
- Columnar compression reduces data volume 10x
 - Provides 10x lower cost, 10x higher performance
- Exadata Smart Flash Cache solves random I/O bottlenecks
 - Increase random I/Os by factor of 20X



ORACLE

Exadata Software Features

- Exadata Smart Scans
 - 10X or greater reduction in data sent to database servers
- Exadata Storage Indexes
 - Eliminates unnecessary I/Os to disk
- Hybrid Columnar Compression (HCC)
 - Increases effective storage capacity and increases user data scan bandwidths by a factor of 10X
- Exadata Smart Flash Cache
 - Breaks random I/O bottleneck by increasing IOPs by 20X
 - Doubles user data scan bandwidths
- I/O Resource Manager (IORM)
 - Enables storage grid by prioritizing I/Os to ensure predictable performance
- Inter-leaved Grid Disks
 - Enables storage grid that allows multiple applications to place frequently accessed data on faster portions of the disk



ORACLE



Exadata Storage Management & Administration

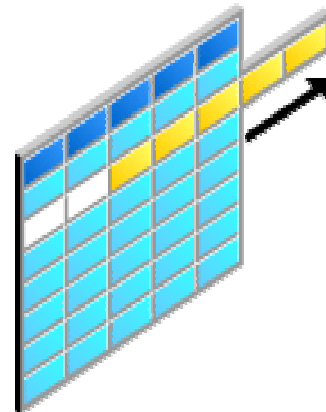
- Enterprise Manager
 - Manage and administer Database and ASM
- Exadata Storage Plug-in
 - Monitor and manage Exadata Storage Cells
- Comprehensive CLI
 - Local Exadata Storage cell management
 - Distributed shell utility to execute CLI across multiple cells
- Sun Embedded Integrated Lights Out Manager (ILOM)
 - Remote management and administration of hardware



ORACLE®

Exadata Smart Scan

- Exadata cells implement scan offload to greatly reduce the data sent to database servers
 - Row filtering based on “where” predicate
 - Column filtering
 - Join filtering
 - Incremental backup filtering
 - **Scans on encrypted data**
 - **Data Mining model scoring**

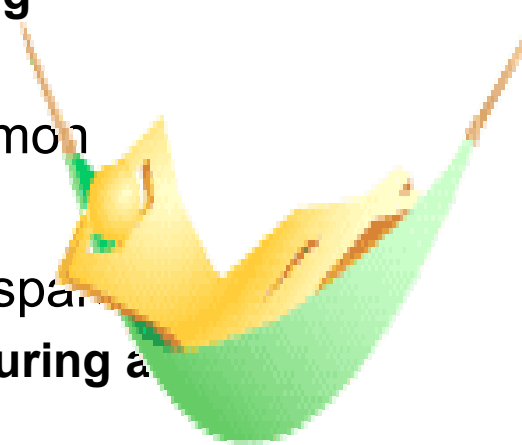


11.2

11.2

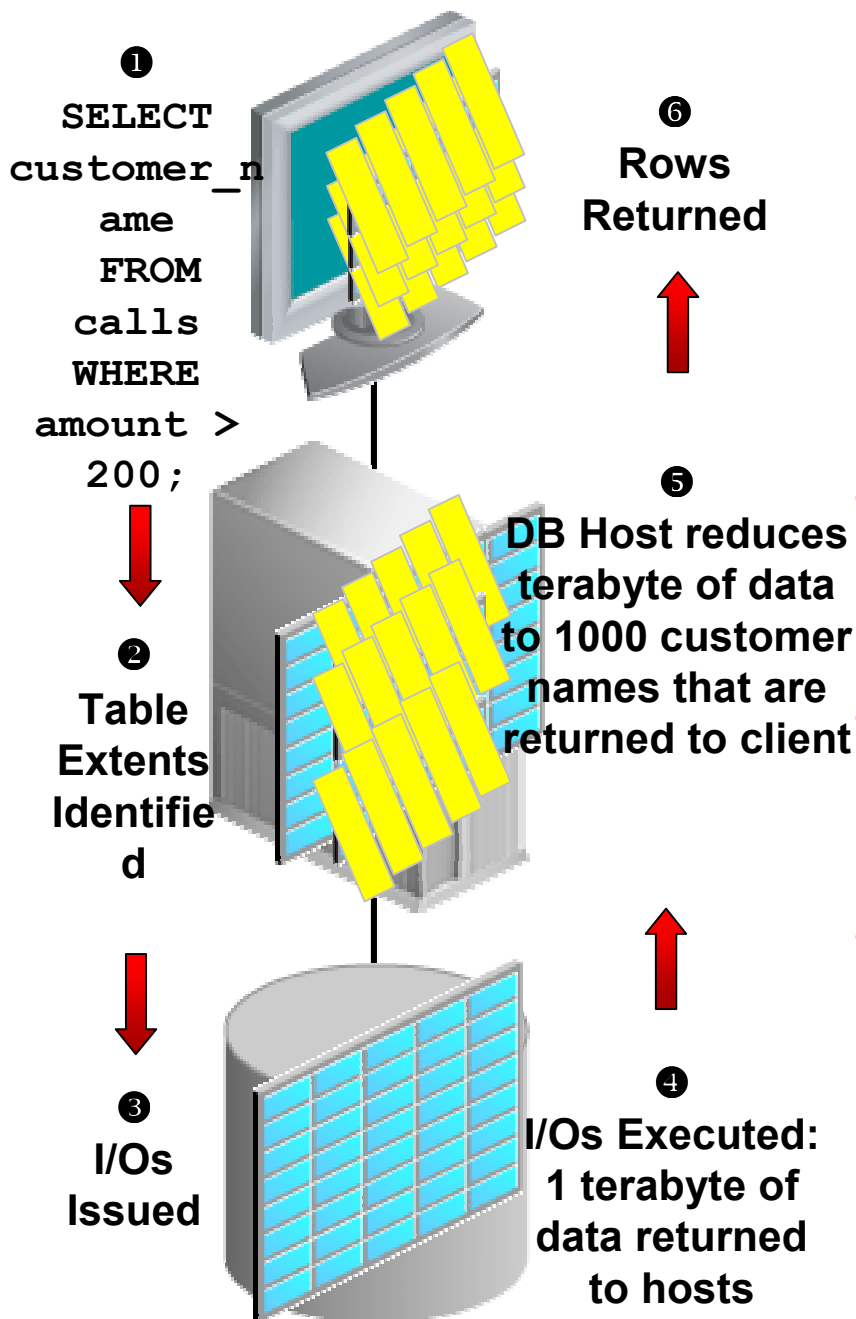
10x data reduction is common

- Completely application transparent
 - **Even if cell or disk fails during a query**



ORACLE

Traditional Scan Processing

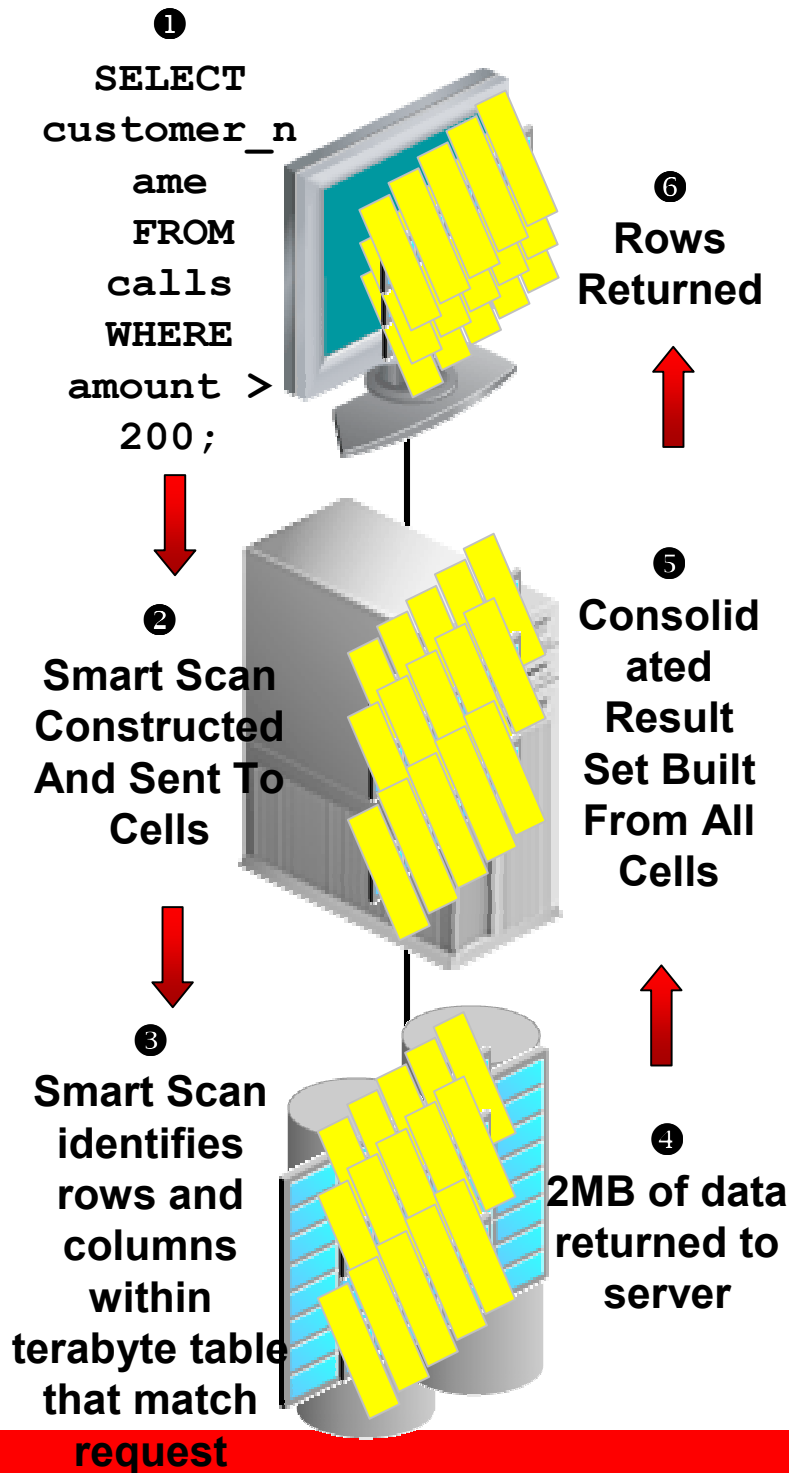


- Smart Scan Example:
 - Telco wants to identify customers that spend more than \$200 on a single phone call
 - The information about these premium customers occupies 2MB in a 1 terabyte table
- With traditional storage, all database intelligence resides in the database hosts
- Very large percentage of data returned from storage is discarded by database servers
- Discarded data consumes valuable resources, and impacts the performance of other workloads



ORACLE

Exadata Smart Scan Processing



- Only the relevant columns
 - customer_name and required rows
 - where amount > 200 are returned to hosts
- CPU consumed by predicate evaluation is offloaded to Exadata
- 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



ORACLE

Exadata Smart Scans

Offloaded Data Mining Scanning

11.2

- Data mining scoring executed in Exadata:

```
select cust_id
from customers
where region = 'US'
and prediction_probability(churnmod, 'Y'
using *) > 0.8;
```

Scoring
function
executed in
Exadata

- All data mining scoring functions offloaded to Exadata
- Up to 10x performance gains
- Reduced CPU utilization on Database Server



ORACLE

Exadata Storage Index

11.2

Transparent I/O Elimination with No Overhead

Table Index

A	B	C	D
	1		
	3		
	5		
	5		
	8		
	3		

Min B = 1
Max B = 5

Min B = 3
Max B = 8

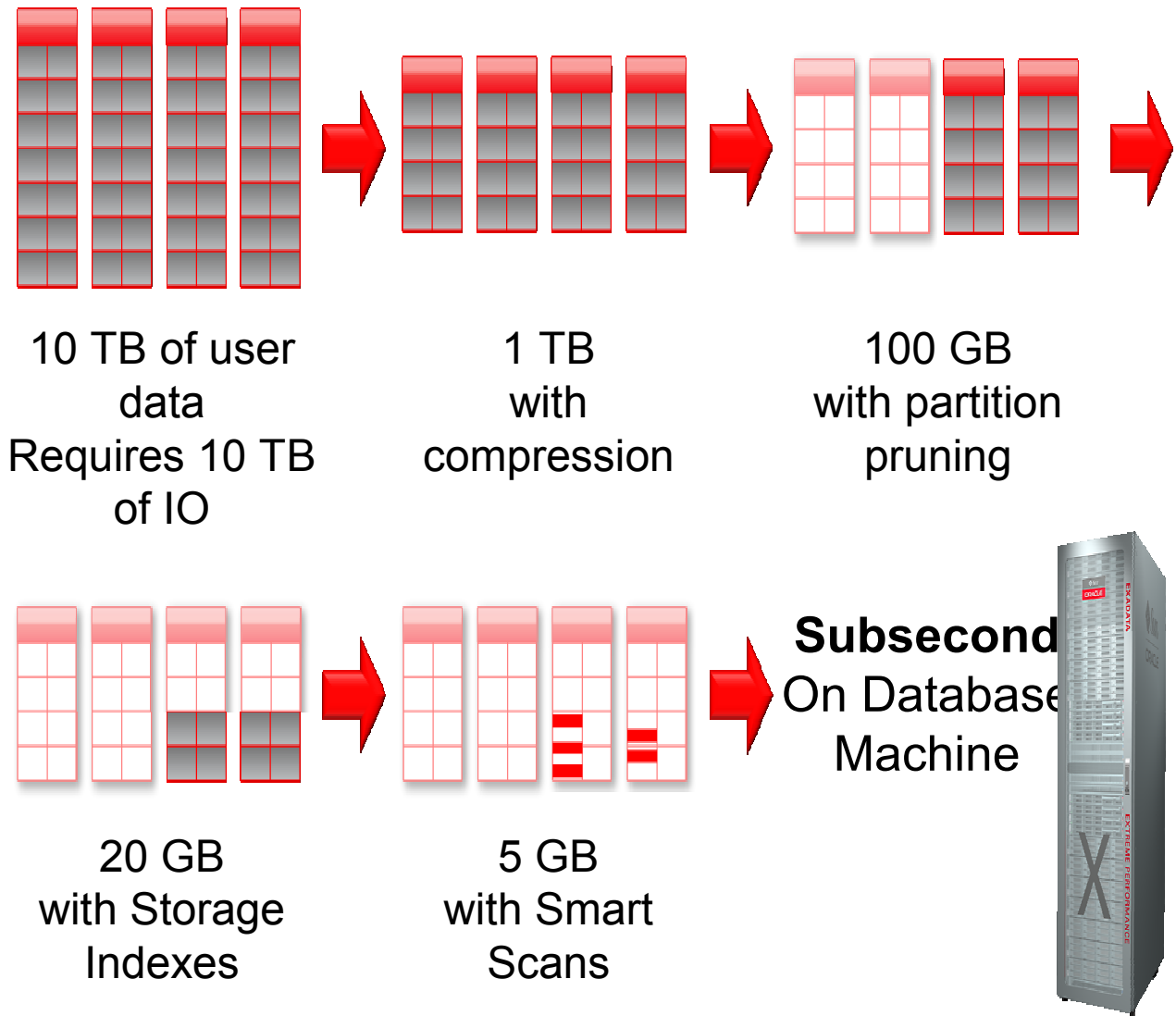
- Exadata Storage Indexes maintain summary information about table data in memory
 - Store MIN and MAX values of columns
 - Typically one index entry for every MB of disk
- Eliminates disk I/Os if MIN and MAX can never match “where” clause of a query
- Completely automatic and

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



ORACLE

Benefits Multiply



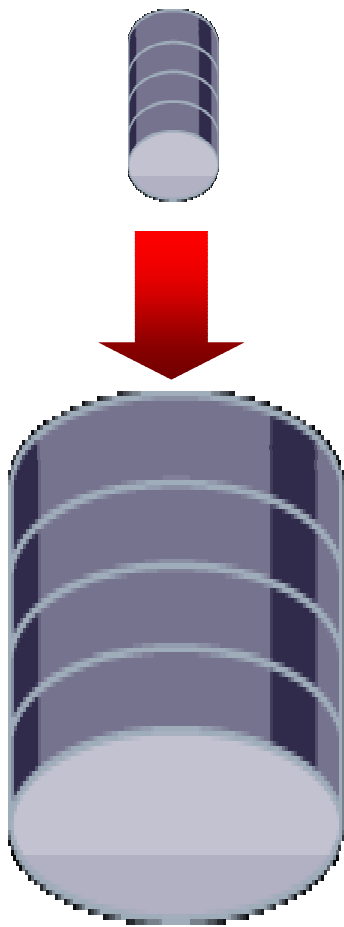
Data is 10x Smaller, Scans are 2000x faster



ORACLE

Data Growth Challenges

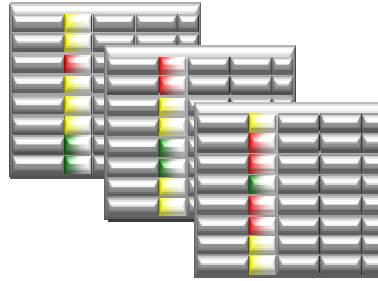
- Support exponentially growing amounts of data
 - Without hurting performance
 - Without growing cost



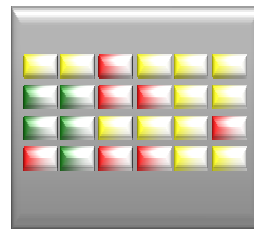
ORACLE®

Exadata Hybrid Columnar Compression

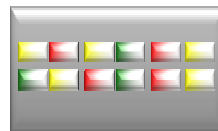
- Data is stored by column and then compressed
- **Query Mode** for data warehousing
 - Optimized for speed
 - 10X compression ratio is typical
 - Scans improve proportionally
- **Archival Mode** for infrequently accessed data
 - Optimized to reduce space
 - 15X compression is typical
 - Up to 50X for some data



11.2



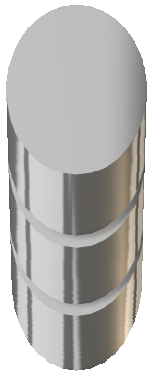
Up To
50X



ORACLE

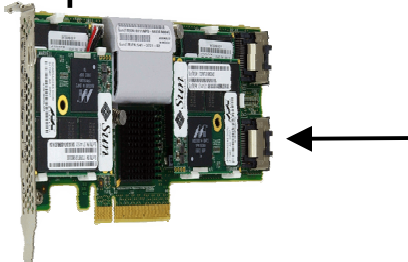
The Disk Random I/O Bottleneck

300 I/O per Sec



- Disk drives hold vast amounts of data
 - But are limited to about **300 I/Os per second**
- Flash technology holds much less data
 - But can run **tens of thousands of I/Os per second**

Tens of
Thousands of
I/O's per Second



- Ideal Solution
 - Keep most data on disk for low cost
 - **Transparently move hot data to flash**
 - Use **flash cards instead of flash disks** to avoid disk controller limitations
 - Flash cards in Exadata storage
 - High bandwidth, low latency interconnect



ORACLE

Exadata Smart Flash Cache 11.2

- Caches Hot Data Transparently in the 4 Flash Cards
- Use PCI Express based Flash Cards for greater throughput and IOPs and avoid disk controller limitations
- Smart Caching
 - Smarter than basic LRU algorithm
 - Knows when to skip caching objects to avoid polluting or flushing the cache
- Allows applications to explicitly optimize caching



4 x 96 GB Flash Cards



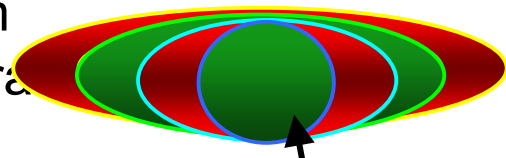
ORACLE

Interleaved Grid Disks

11.2

- Interleaved grid disks place frequently accessed data in all grid disks on higher performing outer tracks
- All applications benefit from higher performance outer tracks of disks

Grid Disk 1
Hot Data, Cold Data

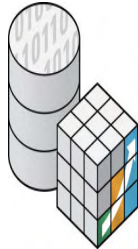


Grid Disk 2
Hot Data, Cold Data



ORACLE

Best Data Warehouse Machine



OLAP

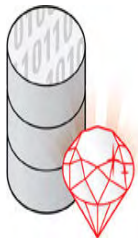
- Massively parallel high volume hardware to quickly process vast amounts of data

- **Exadata runs data intensive processing directly in storage**



ETL

- Most complete analytic capabilities
 - OLAP, Statistics, Spatial, Data Mining, Real-time transactional ETL, Efficient point queries



- Powerful warehouse specific optimizations
 - Flexible Partitioning, Bitmap Indexing, Join indexing, Materialized Views, Result Cache

Data Mining **Dramatic new warehousing capabilities**

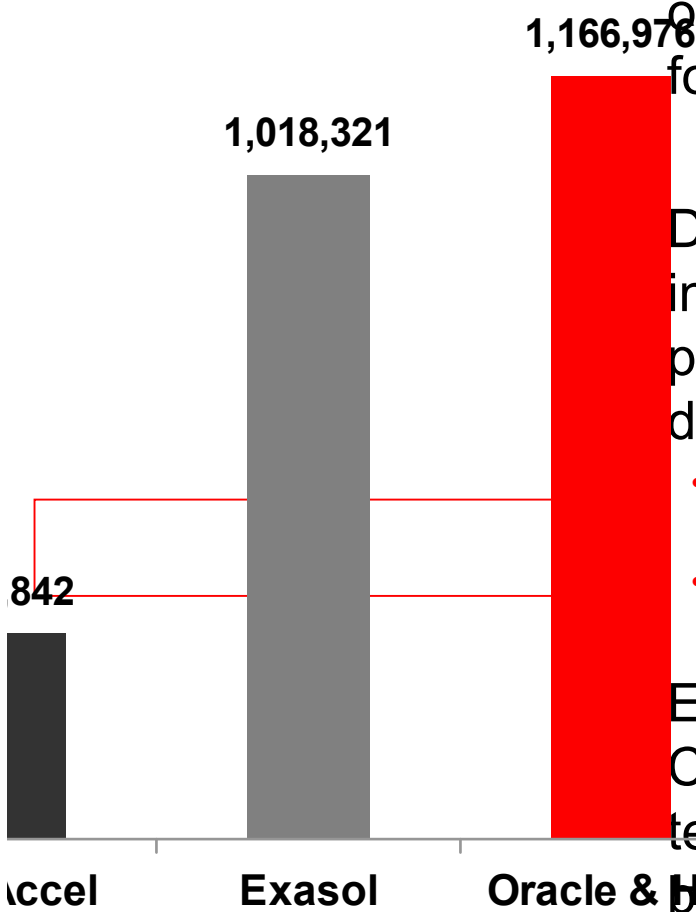


ORACLE

In-Memory Parallel Execution

New

QphH: 1 TB TPC-H



- A single database machine has over 400GB of memory usable for caching

Database release 11.2 introduces parallel query processing on memory cached data

- Harnesses memory capacity of entire database cluster for queries
- Foundation for world record 1TB TPC-H

Exadata Hybrid Columnar Compression enables multi-terabyte tables or partitions to be cached in memory

Exadata memory has 100x more bandwidth than Dis

Faster than specialized in-memory warehouse databases



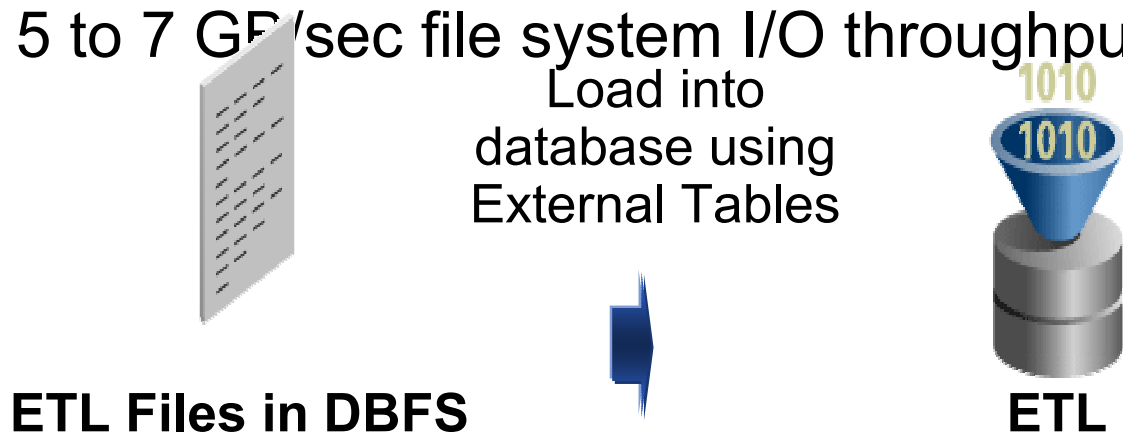
ORACLE

As of 9/14/09. Source: Transaction Processing Council, www.tpc.org
 Oracle on HP Bladepystem c-Class 128P RAC, 1,166,976 QphH@1000GB, \$5.42/QphH@1000GB, available 12/1/09.
 Exasol on PRIMERGY RX300 S4, 1,018,321 QphH@1000GB, \$1.18/QphH@1000GB, available 08/01/08.

DBFS - Scalable Shared File System

11.2

- Database Machine includes DBFS shared Linux file system
 - Shared storage for ETL staging, scripts, reports, other application files
- Files stored as SecureFile LOBs in database tables in Exadata
 - Protected like any DB data – mirroring, DataGuard, Flashback, etc.
- 5 to 7 GB/sec file system I/O throughput



More File Throughput than High-End NAS Filer



ORACLE

Best OLTP Machine



ORACLE

SIEBEL
CUSTOMER RELATIONSHIP
MANAGEMENT

ORACLE

PEOPLESOFT ENTERPRISE

ORACLE

E-BUSINESS SUITE

- Only Oracle runs real-world business applications “**on the Grid**”
- Unique fault-tolerant scale-out OLTP database
 - **RAC, Data Guard, Online Operations**
- Unique fault-tolerant scale-out storage suitable for OLTP
 - **ASM, Exadata**

- **Dramatic New OLTP Capabilities**



ORACLE

Exadata Flash

Solves the Random I/O Bottleneck

11.2



- Has **5+ TB** of flash storage
- Exadata Smart Cache caches hot data
- **Database Machine achieves:**
 - **20x more random I/Os**
 - Over 1 million per second
 - **2x faster sequential query I/O**
 - 50 GB/sec
 - **10x better I/O response time**
 - Sub-millisecond
 - **Greatly Reduced Cost**
 - 10x fewer disks for IOPS
 - Lower Power

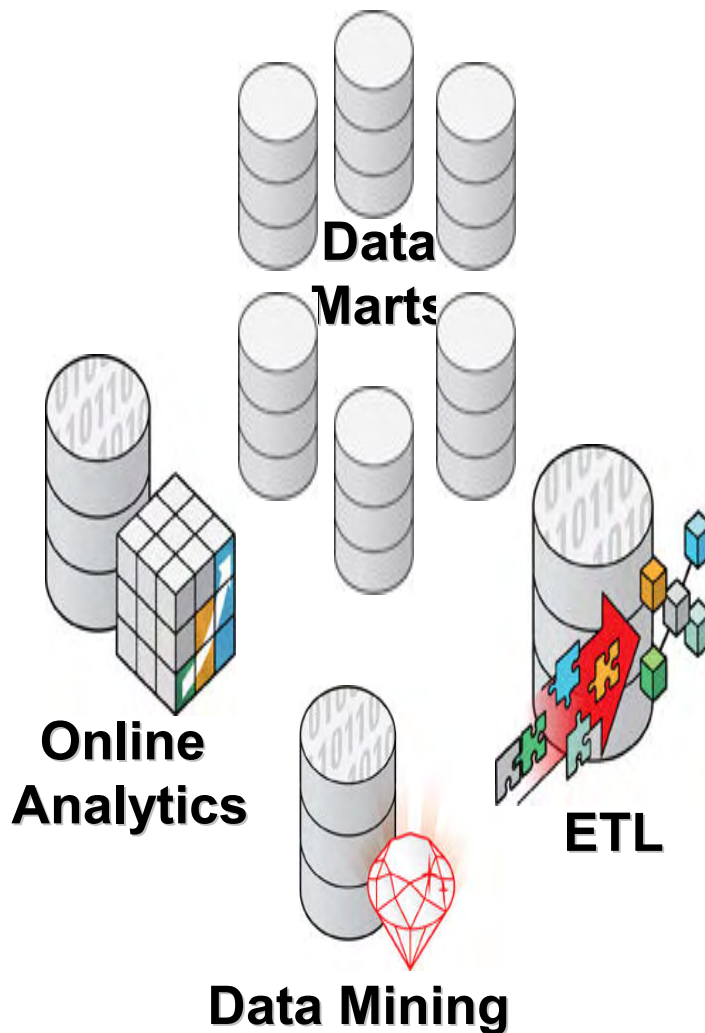
**Oracle is the First
Flash Optimized
Database**



ORACLE

Why Consolidate?

Biggest driver of ongoing cost: **multitudes of special-purpose systems**



ORACLE

Best Consolidation Machine



- Mixes different workloads in one system
 - Warehouse oriented **bulk data processing**
 - OLTP oriented **random updates**
 - Multimedia oriented **streaming files**
- Extreme performance for all workloads
- Predictable response times for all workloads

- Low cost platform for all applications
- Handles all data management needs
- Complete, Open, Integrated



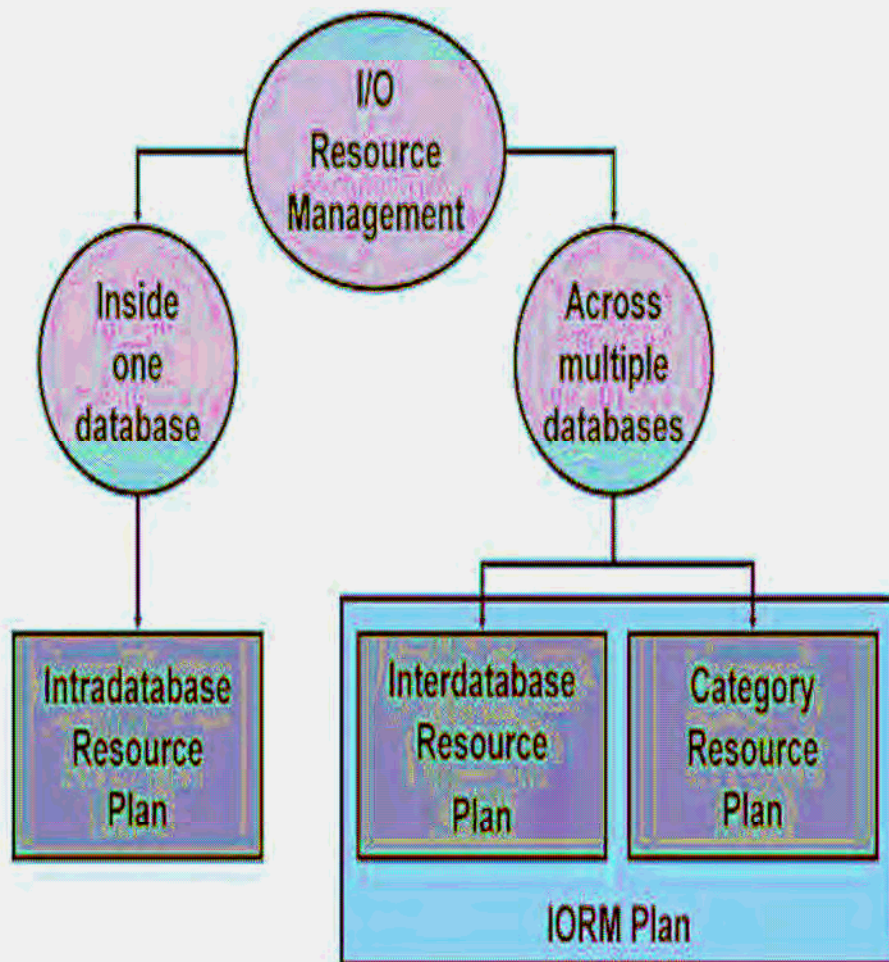
ORACLE

I/O Resource Management Overview

- Traditional benefits of shared storage:
 - Lower administration costs
 - More efficient use of storage
- Common challenge for shared storage:
 - Workloads interfere with each other. For example:
 - Large queries impact on each other
 - Data loads impact on warehouse queries
 - Batch workloads interfere with OLTP performance
- Exadata I/O Resource Management allows you to govern I/O resource usage among different:
 - User types
 - Applications
 - Workload types
 - Databases



I/O Resource Management Plans



Intradatabase Plan Example

```
BEGIN
  DBMS_RESOURCE_MANAGER.CREATE_SIMPLE_PLAN(SIMPLE_PLAN => 'my_simple_plan',
    CONSUMER_GROUP1 => 'high_priority', GROUP1_PERCENT => 80,
    CONSUMER_GROUP2 => 'low_priority', GROUP2_PERCENT => 20);
END;
/
```

```
ALTER SYSTEM SET RESOURCE_MANAGER_PLAN = 'my_simple_plan';
```

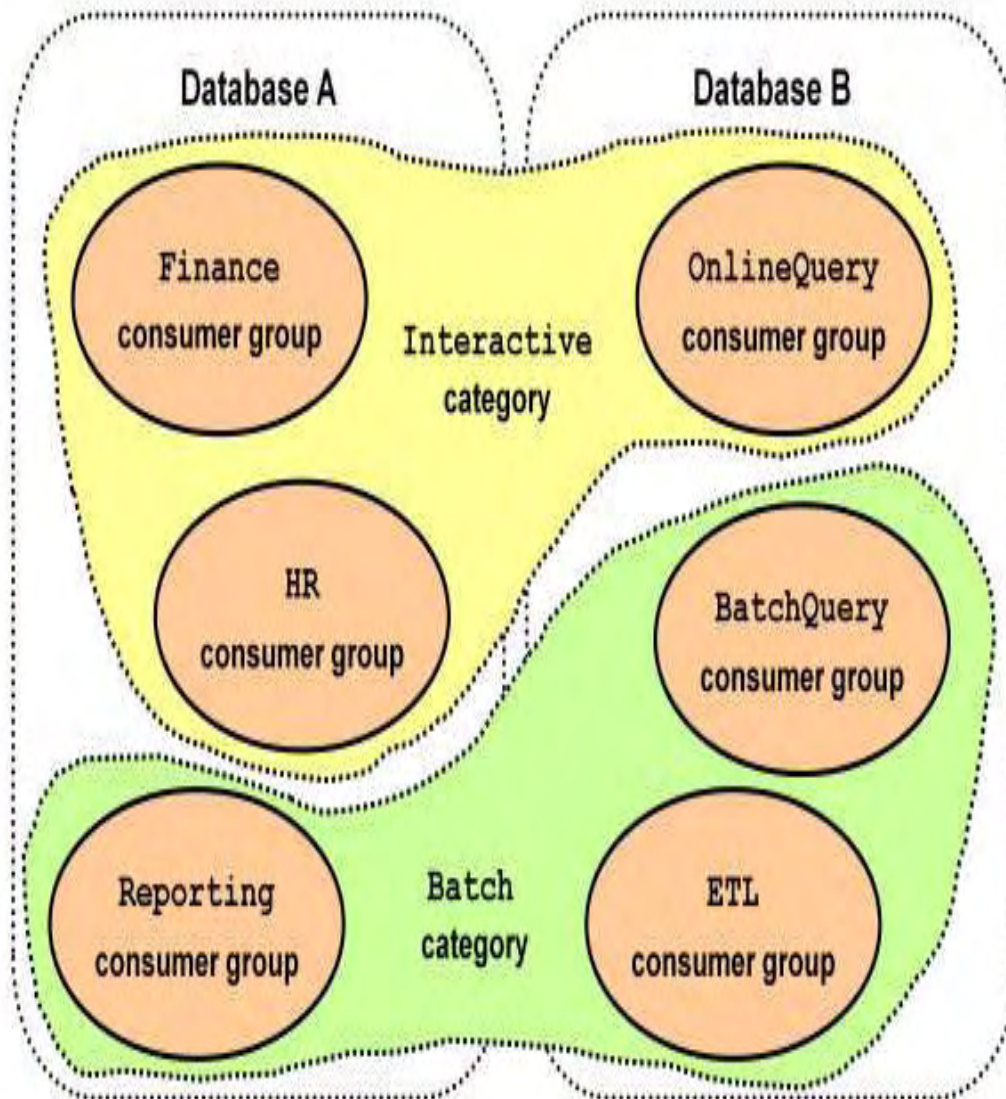
The plan is sent directly to the Exadata cells via iDB.

Consumer Group	Level 1	Level 2	Level 3
SYS_GROUP	100%		
HIGH_PRIORITY		80%	
LOW_PRIORITY		20%	
OTHER_GROUP			100%

Percentages are used for both CPU and I/O resources.

```
CellCLI> ALTER IORMPLAN ACTIVE
```

I/O Resource Management



Category Plan Example

DBA_RSRC_CONSUMER_GROUPS

CONSUMER_GROUP	CATEGORY
SYS_GROUP	ADMINISTRATIVE
BATCH_GROUP	BATCH
INTERACTIVE_GROUP	INTERACTIVE
ORAS	MAINTENANCE
OTHER_GROUPS	OTHER
DEFAULT_CONSUMER_GROUP	OTHER
LOW_GROUP	OTHER
AUTO_TASK_CONSUMER_GROUP	OTHER

dbms_resource_manager.create_category

Category	Level 1	Level 2	Level 3
Interactive	90%		
Batch		80%	
Maintenance			50%
Other			50%

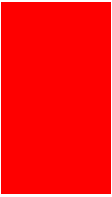
```
CellCLI> alter iorplan <-
>   dbplan=')' <-
>   catplan=( <-
>       (name=interactive, level=1, allocation=90), <-
>       (name=batch, level=2, allocation=80), <-
>       (name=maintenance, level=3, allocation=50), <-
>       (name=other, level=3, allocation=50) <-
>   )
CellCLI> alter iorplan active
```

Interdatabase Plan Example

```
CellCLI> alter iormplan  
>         dbplan=((name=sales_prod, level=1, allocation=80),  
>                 (name=finance_prod, level=1, allocation=20),  
>                 (name=sales_dev, level=2, allocation=100),  
>                 (name=sales_test, level=3, allocation=50),  
>                 (name=other, level=3, allocation=50)),  
>         catplan=''
```

```
CellCLI> alter iormplan active
```

Database	Level 1	Level 2	Level 3
sales_prod	80%		
finance_prod	20%		
sales_dev		100%	
sales_test			50%
other			50%



The Architecture of the Future

Massively Parallel Grid



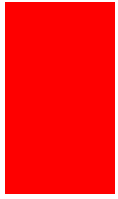
**Best for Data
Warehousing**

Best for OLTP

Best for Consolidation



ORACLE

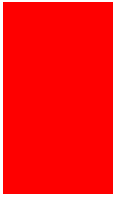


Demonstration

Exadata v2 – Management and Performance



ORACLE



ORA



ORACLE



For More Information

<http://search.oracle.com>

Oracle exadata



or

www.oracle.com/exadata



ORACLE®