The Architecture of the Future
Massively Parallel Grid

Best for Data Warehousing
Best for OLTP
Best for Consolidation
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
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.
Agenda

• Overview – Business Benefits of Exadata
• Exadata Based Product Offerings
• Exadata Architecture and Features
• Best Data Warehousing Machine
• Best OLTP Machine
• Best Consolidation Machine
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
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
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
Complete, Open, Integrated Security

- Monitoring
- Configuration Management
- Access Control
- Encryption and Masking
- Advanced Security
- Secure Backup
- Audit Vault
- Total Recall
- Secret Security
- Label Security
- Data Masking
- Secure Backup
Database Machine Success

Representative customers in all geographies and industries

[Logos of various companies]
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

**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
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!
Start Small and Grow

Quarter Rack  Half Rack  Full Rack
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
Sun Exadata Storage Server

Hardware

- 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) InfiniBand 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
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
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
## Exadata Product Capacity

<table>
<thead>
<tr>
<th></th>
<th>Single Server</th>
<th>Quarter Rack</th>
<th>Half Rack</th>
<th>Full Rack</th>
</tr>
</thead>
<tbody>
<tr>
<td>Raw Disk$^1$</td>
<td>7.2 TB</td>
<td>21 TB</td>
<td>50 TB</td>
<td>100 TB</td>
</tr>
<tr>
<td>Raw Disk$^1$ (SAS)</td>
<td>7.2 TB</td>
<td>21 TB</td>
<td>50 TB</td>
<td>100 TB</td>
</tr>
<tr>
<td>Raw Disk$^1$ (SATA)</td>
<td>24 TB</td>
<td>72 TB</td>
<td>168 TB</td>
<td>336 TB</td>
</tr>
<tr>
<td>Raw Flash$^1$</td>
<td>384 GB</td>
<td>1.1 TB</td>
<td>2.6 TB</td>
<td>5.3 TB</td>
</tr>
<tr>
<td>User Data$^2$ (SAS)</td>
<td>2 TB</td>
<td>6 TB</td>
<td>14 TB</td>
<td>28 TB</td>
</tr>
<tr>
<td>User Data$^2$ (SATA)</td>
<td>7 TB</td>
<td>21 TB</td>
<td>50 TB</td>
<td>100 TB</td>
</tr>
</tbody>
</table>

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.
## Exadata Product Performance

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

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.
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
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
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
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
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**

**10x data reduction** is common

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

---

```
SELECT customer_name
FROM calls
WHERE amount > 200;
```
Exadata Smart Scans
Offloaded Data Mining Scanning

• Data mining scoring executed in Exadata:

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

• All data mining scoring functions offloaded to Exadata
• Up to 10x performance gains
• Reduced CPU utilization on Database Server
Exadata Storage Index
Transparent I/O Elimination with No Overhead

Table

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3</td>
<td>5</td>
<td>3</td>
</tr>
</tbody>
</table>

Index

- 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 match
Benefits Multiply

10 TB of user data
Requires 10 TB of IO

1 TB with compression

100 GB with partition pruning

20 GB with Storage Indexes

5 GB with Smart Scans

Subsecond
On Database Machine

Data is 10x Smaller, Scans are 2000x faster
Data Growth Challenges

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

Up To

- 50X
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

- 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
**Exadata Smart Flash Cache**

- 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
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.
Best Data Warehouse Machine

- Massively parallel high volume hardware to quickly process vast amounts of data
  - Exadata runs data intensive processing directly in storage

- 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
In-Memory Parallel Execution

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.

Faster than specialized in-memory warehouse databases.

---

Oracle on HP Bladesystem 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.

Copyright © 2009, Oracle Corporation and/or its affiliates.
DBFS - Scalable Shared File System

- 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
Best OLTP Machine

• 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
**Exadata Flash**

Solves the Random I/O Bottleneck

- 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
Why Consolidate?

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

- Inside one database
  - Intradatabase Resource Plan

- Across multiple databases
  - Interdatabase Resource Plan
  - Category Resource Plan

IORM Plan
Intradeskbase 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 /DB.  

<table>
<thead>
<tr>
<th>Consumer Group</th>
<th>Level 1</th>
<th>Level 2</th>
<th>Level 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>SYS_GROUP</td>
<td>100%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HIGH_PRIORITY</td>
<td></td>
<td>80%</td>
<td></td>
</tr>
<tr>
<td>LOW_PRIORITY</td>
<td></td>
<td>20%</td>
<td></td>
</tr>
<tr>
<td>OTHER_GROUP</td>
<td></td>
<td></td>
<td>100%</td>
</tr>
</tbody>
</table>

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

> ALTER FORMPLAN ACTIVE
I/O Resource Management
Category Plan Example

```sql
classcl> alter iormplan -
    > dbplane ' ' -
    > catplane( -
    > (name=interactive, level=1, allocation=90), -
    > (name=batch, level=2, allocation=80), -
    > (name=maintenance, level=3, allocation=50), -
    > (name=other, level=3, allocation=50) -
    )

classcl> alter iormplan active
```
### Interdatabase Plan Example

```sql
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
```

<table>
<thead>
<tr>
<th>Database</th>
<th>Level 1</th>
<th>Level 2</th>
<th>Level 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>sales_prod</td>
<td>80%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>finance_prod</td>
<td>20%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>sales_dev</td>
<td>100%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>sales_test</td>
<td></td>
<td>50%</td>
<td></td>
</tr>
<tr>
<td>other</td>
<td></td>
<td>50%</td>
<td></td>
</tr>
</tbody>
</table>
The Architecture of the Future

Massively Parallel Grid

Best for Data Warehousing
Best for OLTP
Best for Consolidation
Demonstration

Exadata v2 – Management and Performance
O & A
For More Information

http://search.oracle.com

Oracle exadata

or

www.oracle.com/exadata