“Extending IN-Memory Database Processing to Shared Flash

Gurmeet Goindi
Master Product Manager
Safe Harbor Statement

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 remains at the sole discretion of Oracle.
Exadata Database Machine
Performance, Availability and Security

**Best Platform for Oracle Databases on-premises and in the Cloud**

Enabled by:

- Single-vendor accountability
- Exclusive focus on databases
- Deep h/w and s/w integration
- Revolutionary approach to storage
Exadata Achieves Memory Performance with Shared Flash

• Exadata X6 delivers **300GB/sec flash bandwidth** to any server
  – Approaches 800GB/sec aggregate **DRAM** bandwidth of DB servers

• Must move compute to data to achieve full flash potential
  – Requires owning full stack, can’t be solved in storage alone

• Fundamentally, storage arrays can share flash **capacity** but not flash **performance**
  – Even with next gen scale-out, PCIe networks, or NVMe over fabric
  – e.g. new EMC DSSD has 3-6 times lower throughput than Exadata X6

• **Shared storage with memory-level bandwidth** is a paradigm change in the industry
  – Get near DRAM throughput, with the capacity of shared flash
NVMe PCI-e Flash Disrupts the Storage Array Model
New improvements are causing 100X bottlenecks across shared storage stack

Latest PCIe Flash 5.4 GB/sec
SAN Link = 40Gb 5 GB/sec Less than 1 Flash card
Leading All Flash Array 24 GB/sec Less than 5 Flash card

All-Flash Storage Array IO Path: many steps, each adds latency and creates bottlenecks
Redesigning Scan Offload for Memory Throughput

- With Exadata Flash throughput approaching memory throughput, SQL bottleneck moves from I/O to CPU
- Exadata will automatically transform table data into In-Memory DB columnar formats in Exadata flash cache
  - Dual format architecture extended from DRAM to flash
- Enables fast vector processing for storage server queries
  - Smart Scan results sent to DB using In-Memory Columnar format to reduce DB CPU usage
- Uniquely optimizes next generation flash as memory
In-Memory Columnar Formats in DRAM (pre 12.2.1.1.0)
Super-Fast Scans from Memory, but All Queries Complete

In-Memory Columnar scans

Database Server

SGA
IMC
Up to 1.5 TB DRAM

Data not in DRAM

12.8 TB Flash

Storage Server
In-Memory Columnar Formats in Flash Cache (12.2.1.1.0)
3 - 4x Overall Analytics Performance Improvement

12.8 TB Flash x 3 = 38.4 TB (or more)
IMC (In-Memory Columnar) data

Storage Server

Hybrid Columnar Compressed Data
Smart Analytics: Join and Aggregation Smart Scan

- Extend In-Memory Aggregation technique into storage (vector joins and vector aggregation)

- Find Sales per country

  ```sql
  SELECT /*+ VECTOR_TRANSFORM */ country_id,
  sum(amount_sold) amount_sold
  FROM customers, sales
  WHERE customers.cust_id = sales.cust_id
  GROUP BY customers.country_id
  ORDER BY customers.country_id;
  ```

- Storage cells scanning sales fact table return tuples
  `{country_id, sum_amount_sold }`

- Join and Aggregation offloaded to the storage server
Smart Analytics: More Smart Scan Enhancements

- Smart Scan enhancements for XML and JSON
  - `JSON_EXISTS`, `JSON_VALUE`, `JSON_QUERY`, "IS JSON" and "IS NOT JSON"
  - XML: `XMLExists`, `XMLCast(XMLQuery())`

- Significant speedup in JSON analytic workloads
  
  ```sql
  select count(*)
  from pictures
  where json_value(photo, '$.tag')
  like '%Spain%';
  ```
Data Tiering

- Real Time Data Analysis
  - DBIM
- Hot (Frequently Accessed) Data
  - DBIM & Flash Cache
  - HCC on Exadata Storage
- Cooling Data
- Cold or Raw Data
  - ZFS or Big Data SQL
**Smart Analytics:** Smart Write Bursts and Temp IO in Flash Cache

- Write throughput of four flash cards has become greater than the write throughput of 12-disks
- When database write throughput exceeds throughput of disks, Smart Flash Cache intelligently caches writes
- When queries write a lot of temp IO, Smart Flash Cache intelligently caches temp IO
  - *Write to flash for temp spill reduces elapsed time*
  - *Reads from flash for temp reduces elapsed time further*
- Smart Flash Cache prioritizes OLTP data and does not remove hot OLTP lines from the cache
- Smart flash wear management for large writes
- Supports Database 11.2.0.4, 12.1.0.2 and 12.2.0.1

Accelerates Large Joins and Sorts and Large Data Loads
Exadata Cloud – Your Way

Oracle Public Cloud

Exadata Cloud Service

Same Product
Same Price
Oracle Managed
Oracle Owned

Cloud at Customer

Exadata Cloud Machine
Exadata Customer Case Studies
**Benefits**

<table>
<thead>
<tr>
<th>Faster Billing Processing</th>
<th>Maximum Availability</th>
<th>Reduced Operational Cost</th>
<th>Reduced Introduction Cost</th>
<th>Data Center Cost Savings</th>
</tr>
</thead>
<tbody>
<tr>
<td>10X speedup</td>
<td>Local &amp; Remote Standby</td>
<td>50%</td>
<td>25%</td>
<td>90% Space Reduction</td>
</tr>
<tr>
<td>3 million SQL/sec</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

"MoBills is a very important position as a mission-critical system to promote efforts toward the realization of "+d". Oracle Exadata is running very stable as a expected performance. We will continue to use the "Oracle Exadata" and we would like to establish a further advantage for our business."
- Shimamura, Manager, Information System Department, NTT docomo

**Business Objectives**
- Real-Time Billing Platform for 66 million customer
- Dramatically improve performance and availability
- Reduce cost and complexity

**Solution**
- Oracle Exadata : 30 racks
- Oracle MAA (RAC / Active Data Guard - Local & Remote Standby database)

**Pre-Exadata**
- Real-Time Billing Processing
  - High-end SMP Server
  - High-end Storage : 350 racks
- Storage Mirror Backup
- Storage Mirror Replication
- Oracle 9i Database Release 2

**Exadata MAA**
- 30 racks / Local & Remote Standby / RMAN backup
- Tokyo
  - Rating (primary) 5node RAC * 5 racks
  - Local Standby 5 racks
- Osaka
  - Billing (primary) 5node RAC * 5 racks
  - Remote Standby 5 racks

**Upgrade & Migration**
DCM Holdings: System Consolidation of 3 companies

**Benefits**

<table>
<thead>
<tr>
<th>Faster</th>
<th>Batch Processing</th>
<th>High Consolidation Ratio</th>
<th>Simplified Support</th>
</tr>
</thead>
<tbody>
<tr>
<td>Standardization</td>
<td>Multitenant Architecture</td>
<td>Oracle Full Stack</td>
<td></td>
</tr>
<tr>
<td>2X speedup</td>
<td>40% Off</td>
<td>6DBs Consolidation</td>
<td>Non Stop Support</td>
</tr>
</tbody>
</table>

Realized the Database consolidation and integration due to the high performance provided by Oracle Exadata. And, Oracle Database 12c Multitenant Architecture also achieved high consolidation ratio while maintaining the independence of each group companies. Platinum Service could improve the service level, Oracle Full-stack products could provide One-Stop Support.

**Business Objectives**

- $10 billion Sales, Faster M&A
- High Consolidation ratio and improve service level
- Reduce operational cost

**Solution**

- Oracle Database 12c Multitenant on Exadata
- Oracle Full Stack (Middleware and Server products)

Oracle Multitenant on Exadata

- Consolidation and Integration 3 group companies (Homac, Kahma and DAIKI) of system infrastructure
- Replaced from IBM p Servers
- Teradata Migration to Exadata
- Oracle Database 12c Multitenant
- Platinum Service
- Zero down time System Migration by using GoldenGate
Sprint: Call Data Record - Data Warehouse

<table>
<thead>
<tr>
<th>Benefits</th>
</tr>
</thead>
<tbody>
<tr>
<td>“We reduced the queries from 30 seconds down to sub-second response time. Quick information, quick queries give Customer Care the ability to do their job better and meet the customer’s needs.”</td>
</tr>
<tr>
<td>- Richard Ewald, Senior Technical Architect, Data Warehousing</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Faster Queries</th>
<th>Faster Reports</th>
<th>Storage Savings</th>
<th>Maximum Availability</th>
<th>Data Center Cost Savings</th>
</tr>
</thead>
<tbody>
<tr>
<td>&gt; 10x</td>
<td>24 X</td>
<td>6 x</td>
<td>No unplanned downtime</td>
<td>3:1 Consolidation</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Business Objectives</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Improve performance</td>
</tr>
<tr>
<td>• Improve sustainability</td>
</tr>
<tr>
<td>• Improve availability and maintainability</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Pre-Exadata</th>
<th>Production</th>
<th>Dev/Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>2012 X3-2</td>
<td>Exadata X3-2 Full Rack</td>
<td>Exadata X3-2 Half Rack</td>
</tr>
<tr>
<td>2014 X4-2</td>
<td>HCC: 950 TB to 150 TB</td>
<td></td>
</tr>
<tr>
<td>2015 X5-2</td>
<td>ZFS Storage</td>
<td>ZFS Storage Appliance (Backup)</td>
</tr>
<tr>
<td>4 x Sun Fire E6900, 1 x M9000</td>
<td>Exadata Storage Expansion</td>
<td>Exadata X3-2 Half Rack</td>
</tr>
<tr>
<td>Mixed Storage</td>
<td>Exadata Storage Expansion</td>
<td></td>
</tr>
<tr>
<td>Multiple backup systems</td>
<td>Exadata X5-2 Half Rack</td>
<td></td>
</tr>
<tr>
<td>90 Day CDR DW 1.15 PB</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Oracle DB 11gR2</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Solution</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Full Rack (Prod), Half Rack (Dev/Test); ZFS</td>
</tr>
<tr>
<td>• Storage Expansion</td>
</tr>
<tr>
<td>• Half Rack (Prod)</td>
</tr>
</tbody>
</table>

Copyright © 2017, Oracle and/or its affiliates. All rights reserved.
Pulte Group: Multitenant Consolidation

Benefits

“Exadata delivered tremendous improvements in productivity. Users no longer have to wait for data. Data sharing is now real time.”
- Brian Pawlik, IS Manager, Pulte Homes

<table>
<thead>
<tr>
<th>Business Impact</th>
<th>Faster Applications</th>
<th>Lower Admin &amp; Support Costs</th>
<th>Cost Savings</th>
</tr>
</thead>
<tbody>
<tr>
<td>40% Productivity</td>
<td>2x - 15x Faster</td>
<td>40% Reduction</td>
<td>40% CapEx</td>
</tr>
<tr>
<td>Monthly Close 2 Days Faster</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Business Objectives

- Scalability
- Supportability
- Sustainability

Solution

- quarter rack & eighth rack

Pre-Exadata

- IBM P7
- EMC storage arrays

Exadata Quarter Rack
Production / Standby / Test Dev / UAT

- Infor Lawson S3 ERP; Rebate Tracking
- Consolidate 35 DBs: 4 CDBs, 35 PDBs
- Production, Local Standby and QA
- Primary databases: > 5 TB

Active Data Guard
WAN @ 800 miles

Exadata Eighth Rack
Disaster Recovery
Exadata Advantages Increase Every Year

Dramatically Better Platform for All Database Workloads
Integrated Cloud
Applications & Platform Services