Exadata Smart Scan
What is so smart about it?

Uwe Hesse
Senior Principal Instructor
Oracle Certified Master
http://uhesse.com
Exadata Quarter Rack – Hardware Viewpoint

- Quarter Rack
- Half Rack
- Full Rack
- Multiple Racks

Copyright © 2011, Oracle and/or its affiliates. All rights reserved.
X2-2 Database Servers (Sun Fire X4170 M2)

- 2 Six-Core Intel® Xeon® X5675 Processors
- Up to 144 GB DRAM (16 x 8GB)
- Dual-port 10GbE PCIe card
- Dual-port InfiniBand QDR (40 Gb/s) HCA
- 4 x 2.5” 300GB Disk Drives
- Disk Controller HBA with 512 MB BBWC
- Redundant Power Supplies, Fans
- ILOM

Installed Software:
- Oracle Database 11.2
- Oracle Linux or Solaris
Exadata Storage Server Hardware (Sun Fire X4270 M2)

- 2 Six-Core Intel® Xeon® L5640 Processors
- 24 GB DRAM (6 x 4GB)
- 12 x 3.5" Disk Drives (600GB SAS or 2TB SAS)
- 4 x 96GB Flash PCIe Cards
- Dual-port InfiniBand QDR (40Gb/s) HCA
- Disk Controller SAS2 HBA with 512MB BBWC
- Redundant Power Supplies, Fans
  - ILOM

Installed Software:
- Oracle Exadata Storage Server Software
- Oracle Linux
InfiniBand Network

- Unified InfiniBand Network
  - Storage Network
  - RAC Interconnect

- High Performance, Low Latency Network
  - 40 Gb/s bandwidth
    - Zero-copy Zero-loss Datagram Protocol (ZDP RDSv3)
      - SAN-like Efficiency (Zero copy, buffer reservation)
Exadata Quarter Rack – Database Viewpoint

Diagram showing the architecture with DB Servers, ASM disks, and failgroups connected through InfiniBand network and ASM disk groups with normal redundancy.
Database Viewpoint is unchanged

• Competitor: “Exadata is still Oracle!”
• Right – that is actually a good thing because:
  • No change is needed for the Database Design
  • No change is needed for the Application Code
  • Migration to Exadata is relatively easy therefore
  • DBA knowledge remains valid
Exadata in the Marketplace
Rapid adoption in all geographies and industries
Classic Database I/O and SQL Processing Model

```
SELECT customer_id
FROM orders
WHERE order_amount > 20000;
```

1. Extents identified
2. I/O issued
3. I/O executed: 10 GB returned
4. SQL processing: 2 MB returned
5. Rows returned

Copyright © 2011, Oracle and/or its affiliates. All rights reserved.
Exadata Smart Scan Idea

• “The fastest way of doing something is not to do it”
• Do not deliver the whole Data Volume up to the DB Layer!
• Instead, filter resp. project already on the Storage Layer!
• We want DB Intelligence built into the Storage Layer ...
Remove the Storage Network Bottleneck

**Without Smart Scan**
- Utilized DB Nodes
- Saturated Storage Network
- Throttled Storage

**With Smart Scan**
- Utilized DB Nodes
- Utilized Storage
Exadata Smart Scan Model

1. SELECT customer_id
   FROM orders
   WHERE order_amount>20000;

2. iDB command constructed and sent to Exadata cells

3. SQL processing in Exadata cells

4. 2 MB returned to server

5. Consolidated result set built from all Exadata cells

6. Rows returned
Exadata Smart Scan

- **Predicate filtering:**
  - Only the requested rows are returned to the database server rather than all the rows in a table.

- **Column filtering:**
  - Only the requested columns are returned to the database server rather than all the columns in a table.
  - For example:

  ```sql
  SQL> SELECT col1, col2 FROM t WHERE condition;
  ```

http://uhesse.com/2011/01/19/exadata-part-i-smart-scan/
Smart Scan Prerequisites

- Smart Scan is only possible for Full Table Scans or Full Index Scans.
- Smart Scan can only be used for Direct-Path Reads:
  - Direct-path reads are automatically used for parallel queries.
  - Direct-path reads may also be used for serial queries.
Exadata Smart Scan

• Join processing:
  – Star join processing is performed within Exadata Storage Server
• Scans on TDE encrypted data
• Scans on HCC compressed data
Exadata Hybrid Columnar Compression: Overview

Warehouse Compression
- Optimized for Speed
  - 10x average storage savings
  - 10x scan I/O reduction
  - Optimized for query performance

Archival Compression
- Optimized for Space
  - 15x average storage savings
    - Up to 50x on some data
  - Greater access overhead
  - For cold or historical data

Reduced Warehouse Size
Better Performance

Reclaim Disks
Keep Data Online

Can mix compression types by partition for Information Life Cycle Management
Exadata Hybrid Columnar Compression
Data Organization

Compression Unit

- A compression unit is a logical structure spanning multiple database blocks.
- Each row is self-contained within a compression unit.
- Data is organized by column during data load.
- Each column is compressed separately.
- Smart Scan is supported.

Exadata Storage Index Idea

• “The fastest way of doing something is not to do it”
• Even faster than offload scanning to the Storage Layer:
  • Exclude areas from scanning when we know the requested Data can’t be there

http://uhesse.com/2011/01/20/exadata-part-ii-storage-index/
Storage Index with Partitions: Example

<table>
<thead>
<tr>
<th>ORDER#</th>
<th>ORDER_DATE (Partition Key)</th>
<th>SHIP_DATE</th>
<th>ITEM</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2007</td>
<td>2007</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>2008</td>
<td>2008</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>2009</td>
<td>2009</td>
<td></td>
</tr>
</tbody>
</table>

- Queries on `SHIP_DATE` do not benefit from `ORDER_DATE` partitioning:
  - However, `SHIP_DATE` is highly correlated with `ORDER_DATE`.
- Storage index enhances performance for queries on `SHIP_DATE`:
  - Takes advantage of the ordering created by partitioning
Smart Scan related Statistics & Wait-Events

Important Statistics:
cell physical IO interconnect bytes
cell physical IO interconnect bytes returned by smart scan
cell physical IO bytes saved by storage index

Important Wait-Events:
cell smart table scan
cell smart index scan
cell single block physical read
cell multiblock physical read

http://uhesse.com/2011/07/06/important-statistics-wait-events-on-exadata/
Benefits Multiply

Multiple terabytes of user data normally require multiple terabytes of I/O.

Less with Exadata Hybrid Column Compression

Even less with partition pruning

Storage index skips worthless I/O.

Smart Scan filters the remaining Candidate Data and reduces the Data Volume sent to the Database Layer.

Results in real-time on Database Machine
The smartest thing about Smart Scan

If you ask me:
Smartest thing is that it works transparently together
with the same Oracle Database Code
as used on Non-Exadata platforms

Transparently except that it is 10 times faster

„Exadata is still Oracle!“ Love it :-}