High Performance Computing and Master Data Management

Running Oracle Siebel on Oracle Exadata and Oracle Exalogic case study
Executive Summary

Organizations today are faced with unique and oftentimes conflicting challenges to deliver more efficiency from their IT but at the same deliver improved business value from their applications. One of the most elegant ways to address this conundrum is to run these business applications on engineered systems platforms, providing a turnkey solution that offers exceptional price/performance and delivers optimal scalability and availability.

In this study, we take a specific look at the performance and scalability of Oracle MDM running with Oracle Siebel on both Oracle Exadata and Oracle Exalogic. The results produced impressive performance and scalability metrics, which highlight the advantages of Oracle’s engineered hardware and software systems. The test was performed running Oracle MDM on a half rack Exadata machine and a half rack Exalogic machine. All the components of Siebel along with the Web Servers and Gateway Server were installed on Exalogic. Highlights from the study include:

- Scalability performance for MDM on Exadata for user loads scaled up by 3X for complex queries, 2.3X higher throughput performance
- Throughput performance for MDM on Exadata was 11X higher than customer requirements for complex read transactions, and 7X higher than customer requirements for medium complexity read transactions
- Response times for MDM on Exadata were 10X faster for complex read transactions and 7X faster for medium complexity read transactions

Relative to other commercially available HW alternatives, Oracle Exadata and Exalogic machines operate at a lower total cost of ownership (TCO), require less space, and consume less energy. Comparing results for the same tests done on traditional commodity hardware with the same processing power shows that Exadata/Exalogic delivered 4X better throughput and 8X better response times.

Business Requirements for Master Data Management

Organizations today are faced with unique and oftentimes conflicting challenges – whereas it used to be sufficient to treat customers well, while delivering shareholders organizations, now however, managers have to do much more:

- **Delight Customers**: Know customers by capturing their preferences, transaction histories, and utilize this information to deliver truly ‘exceptional’ experiences, instantly during the course of the interaction itself.
- **Think Globally, Act Locally**: Operate internationally, to create a ‘reputation’ based on understanding local cultures, yet, manifesting processes, practices, products, and services in a manner that leverage global scale.
- **Be an Orchestrator**: Orchestrate the supply-chain to source, manufacture, fulfill, and service across countries all the while tracking ingredients/ components/ resources for optimal utilization, security, and synergy – especially in the face of supply chain disruptions caused by political or natural events.
- **Compete on Analytics**: Render analytical tools and systems that deliver ‘insight’ from ever-increasing volumes of data from internal (finance, operations, marketing, etc.) and external (social media like facebook and twitter, and 3rd-parties like AC Nielson, and other indices) sources.
- **Show Exceptional Agility**: Re-organize assets, businesses, products, and services to respond to market changes, competition, environmental threats, and governmental regulations, yet maintaining a history of previous structures to ensure compliant reporting.
• **Thrive in the Collaborative Enterprise:** Collaborate virtually, incorporating flexible team structures and investments that yield lasting savings and a strategic advantage.

While this is just a short-list of imperatives facing corporations and governmental institutions today, clearly, there are several key themes that emerge: From a strategy perspective: *speed and scale*, resulting in savings; and from a capability perspective: *data and information*, resulting in better decisions. Clearly, the underlying challenge that needs to be addressed is data and infrastructure complexity for both business and IT, such that infrastructure investments align with business priorities around enterprise data management and deliver tangible benefits.

This paper shows how Oracle can help reduce data and infrastructure complexity, and deliver speed and scale to help businesses succeed in a data intensive, fast-changing, and competitively dynamic environment.

**Mastering the Complexity**

**Case-in-Point #1:** Cross selling / up selling offers during customer interaction is easier said than done. Transactional data needs to be augmented with customer data, and during seasonal promotions and specials, the transaction volumes would be quite high, necessitating high performance HW that can support complex reads (from other systems/sources) and delivery in real time.

**Case-in-Point #2:** Border control and customs is a topic that results in heated debates, but when it comes to tracking movements of passengers, shipments, trucks, and people across borders, the underlying challenges of pulling against watch lists, ensuring compliant trade practices, and adequate documentation requirements take on a whole new meaning. Over the years, we see ever-growing integration between transactional and operational systems, very high data volumes, and real time verification requirements; not to mention the need for sub-second response times for queries to ensure streamlined flow of people, product, and policy.

**Case-in-Point #3:** Global synchronization of data (customer, product, supplier, asset, and financial) is an operational headache for not just finance and accounting teams. After all, it is the assets of the company (its customers, products, facilities, supplier relationships, and financials) when optimally organized, that result in lasting competitive advantage. While batch instead of real time updates may still work, the underlying data volumes are enormous, not to mention the complex reads required from multiple (if not 100s) of systems, all verified, cleansed, and presented for management reporting and analysis. Only the most reliable, scalable, and scan-intensive database machines will do the job when supported by robust data management and mastering practices.

![Data Complexity Factors](image1.png)

![Infrastructure Requirements](image2.png)

**Figure 1.** Data Complexity factors that drive Infrastructure Requirements
Oracle’s Technology Solution (MDM on Exadata/Exalogic) For Complex Infrastructure Requirements – A Test of Performance

In a recent performance test conducted by Oracle’s Master Data Management (MDM) – Oracle Customer Hub and the Exadata/Exalogic teams, the primary objectives were to:

1) Process and consolidate millions of customer records*

2) Deliver sub-second response time for all read operations

3) Meet high throughput requirements

Figure 2: Illustration of MDM on Exadata

Represented above are the middleware components, the web servers, application servers, gateway servers/load balancers, data quality servers all residing on the Oracle Exalogic, and the underlying database residing on Oracle Exadata.

Expectations Prior to the Test

i. Integrated system: Since any MDM system has many components including the Web Servers, Load Balancers, Application Servers, Data Quality Servers and some of them have their own dedicated database, the integrated system approach of having all the components installed on Exadata/Exalogic should reduce the IT operational cost significantly and reduce the complexity of the IT hardware landscape.

* In the retail, financial services, utilities, and communications verticals, typical customer data volumes can range from 100 million to 500 million records, just for one company.
ii. **High Scalability:** Because of the presence of multiple high performance compute nodes in the Exalogic rack, and use of parallelism, compression, disk striping, query offloading and massive flash cache in Exadata, the MDM deployment on Exalogic/Exadata will result in very high scalability.

iii. **Prebuilt Components:** Exalogic has preloaded Web logic components that are optimized and enhanced. Since any MDM system has to be integrated to the operational systems via the middleware, integration to the operational systems will become quite easy.

iv. **High Throughput:** WebLogic's networking, request handling and thread management mechanisms are enhanced in Exalogic which enable it to scale better on the high multi-core compute nodes that are connected to the fast InfiniBand fabric that ties all the compute nodes together. The net effect is that each server should be able to handle more client requests whilst also reducing the time taken to respond to each individual request.

v. **Reduced Response Time:** All the compute nodes within Exalogic, and Exadata and Exalogic, are all connected via the Infiniband I/O fabric and a native InfiniBand networking protocol called SDP is used to interact with the Oracle RAC database on Exadata. This should result in low latency for request-response times for calls between the application tier and the database. The performance gain should be most significant when large result-sets are transferred from the database. The desired net effect is applications are able to respond to client requests faster, leading to overall performance gain for enterprise MDM deployments.

Test Steps

The test was designed to mimic real-word data loads and transaction volumes. Outlined below are the steps that a typical MDM solution will require.

The Initial Batch Load consolidates initial customer data from all the operational systems to the MDM System. This is a one-time activity where:

a. Data from operational systems is extracted and loaded into Siebel UCM EIM (Enterprise Integration Manager) Tables via any SQL Loader tool such as ODI (Oracle Data Integrator).
b. Data from EIM Tables is loaded to UCM SDH (Source Data History) Tables.
c. UCM batch process is run in parallel workflow tasks to load the customer tables. The UCM batch process consolidates the data, de-duplicates the data and creates the golden customer records.
d. BI/DW picks up the customer data from the customer MDM system via ODI.
e. BI/DW picks up the transactional data from the operational systems via ODI.

After the initial load, daily operations are where MDM is used as the customer repository. All operational systems search the Customer Hub before creating a record.
For example, let’s take the Call Center as one of the operational systems. Please note that similar activities would happen for all the operational systems.

a. Customer calls the Call Center for Enquiry/Order/Support.

b. Using the match/fetch web services exposed by the Hub, the golden customer record is matched and fetched from MDM. Now the agent has complete information about customer rather than fragmented data.

c. If the customer is not found, a new customer record is created and synched to the Customer Hub. Usually the primary customer attributes are synched in real time.

d. Using the Insert/Upsert Web Service, a new customer record is published to all the subscribing systems via publish/subscribe mechanism.

e. BI/DW picks up the customer data from the customer MDM system via ODI.

Continuing with our daily operations example, there is generally the need to consolidate new/updated customer data from external / operational systems to the MDM system. This requires sending all the new/updated customer data to the Hub in a daily/weekly batch process. (Since this activity is not Performance intensive, this was kept out of scope of the current Performance benchmark, nevertheless, the typical steps involved are outlined below).

a. New/Updated customer data is extracted and loaded into Siebel UCM EIM (Enterprise Integration Manager) Tables via any messaging queue.

b. Data from EIM Tables is loaded to UCM SDH (Source Data History) Tables.

c. UCM batch process is run in parallel Workflow tasks to load the Customer tables. The UCM batch process consolidates the data, de-duplicates the data and creates the golden customer records.

d. New/Updated customer records are published to the subscribing systems.

e. BI/DW picks up the customer data from the customer MDM system via ODI.

Test Results

The test was done using a half rack Exadata machine and a half rack Exalogic machine. All the components of Siebel along with the Web Servers and Gateway Server were installed on Exalogic. Exadata was used as the database – 400 million conditioned customer records were loaded to the system via the Siebel EIM Process, with close to 150 attributes for each customer record, (name, profile, cross-reference data, addresses, etc.). Various real time transaction scenarios such as ‘Medium Complexity Read’, ‘Complex Read’, ‘Insert’ and ‘Update’ (refer to

Figure 3: Illustration of Customer Data Retrieval

Represented above is a situation where the call center agent retrieves customer information.
Figure 7 in the Appendix for descriptions) were tested and various performance metrics related to throughput, response times and CPU resource utilization were gathered.

Figure 4: Results of Performance – User Load & Speed Testing

- Scalability performance for MDM on Exadata for user loads scaled up by 3X for complex queries, translated into 2.3X higher throughput performance, with only a 20% increase in response times (see Figure 4 above).
- Throughput performance for MDM on Exadata was 11X higher than customer requirements for complex read transactions, and 7X higher than customer requirements for medium complexity read transactions.
- Response times for MDM on Exadata were 10X faster for complex read transactions, and 7X faster for medium complexity read transactions.

Clearly, the performance of a ½ rack Exadata machine produced impressive results against the initial hypotheses. A larger configuration would only yield faster response time and throughput figures. Furthermore, relative to other commercially available HW alternatives, Oracle Exadata and Exalogic machines operate at a lower total cost of ownership (TCO), require less space, and consume less energy.

Comparing results for the same tests done on traditional commodity hardware with the same processing power shows that Exadata/Exalogic delivered 4X better throughput and 8X better response times indicating the effectiveness of the “engineered” hardware-software solution.

**Conclusion**

The business context has changed drastically over the last couple of decades. Managers are faced with significant complexity and challenges, which when addressed correctly, can lead to significant savings through speed and scale. The notion of aligning business priorities with IT decisions and actions holds merit, regardless of whether the business manager is trying to operationalize strategy or whether an IT stakeholder is trying automate processes supporting the strategy.

<table>
<thead>
<tr>
<th>Business Needs</th>
<th>IT Initiatives</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Delight Customers</td>
<td>• Data integration and mastering (customer data)</td>
</tr>
<tr>
<td></td>
<td>• Manage, cleanse and enrich high volume customer data and interactions.</td>
</tr>
<tr>
<td></td>
<td>• Deploy architectures that provide sub-second response times.</td>
</tr>
<tr>
<td>• Think Globally, Act Locally</td>
<td>• Provide process automation &amp; consistency based on business rules-based workflows with fully cross-referenced business objects.</td>
</tr>
<tr>
<td></td>
<td>• Provide process integration based on SOA standards, seamlessly integrated with master data.</td>
</tr>
<tr>
<td></td>
<td>• Process, consolidate and synchronize master data across the enterprise.</td>
</tr>
</tbody>
</table>
| **Be an Orchestrator** | • Improve performance of company’s OLTP applications, via MDM aware applications.  
• Move from batch to real time integration, with master data cross-referenced in real time.  
• Data integration and mastering (product and supplier data)  
• Deploy application-neutral data models |
| **Compete on Analytics** | • Operationalize scan-intensive data warehouses with MDM.  
• Develop advanced analytic capability via BI systems augmented with MDM, and data sourced from many and varied sources/systems.  
• Deploy highly concurrent MDM-aware OLTP applications |
| **Show Exceptional Agility** | • Automate management of data hierarchies across financial, product, customer, supplier, and dimensions. (Data Relationship Management).  
• Rapidly on-board new application functionality.  
• Augment Enterprise Performance Management systems with MDM.  
• Provide Data Governance capabilities, and ensure auditability and compliance. |
| **Thrive in the Collaborative Enterprise** | • Consolidate multiple database instances and heterogeneous data models, simplify administrative tasks and reduce infrastructure costs.  
• Enable Web 2.0 capabilities  
• Implement SOA-based standards integrated with master data  
• Support comprehensive data security to control and monitor data access, update rights, and maintain change history. |

Figure 5: Translating Business Priorities Into IT Considerations

Given the inherent infrastructure complexity (multiple systems, multiple standards, a history of custom development projects, fragmented data sources, etc.) further compounded by business imperatives (need for high speed/throughput, response times, optimum HW utilization at lower TCO, etc.) render high performance computing and data management a necessity. IT and business managers, both, recognize the importance of accessing large volumes of data quickly, across systems – all the while leveraging data management best practices, IT prowess, and fast/scalable HW – each of which is a critical success factor in achieving business goals.

The Oracle test of MDM on Exadata shows the ability of the combination to deliver high performance computing capabilities in order to deliver speed, scale, and savings.
## Appendix

### ACRONYMS AND DESCRIPTIONS OF TERMS USED IN THIS DOCUMENT.

<table>
<thead>
<tr>
<th>Acronym</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>AIA</td>
<td>Application Integration Architecture</td>
</tr>
<tr>
<td>BI/DW</td>
<td>Business Intelligence / Data Warehouse</td>
</tr>
<tr>
<td>CRM</td>
<td>Customer Relationship Management</td>
</tr>
<tr>
<td>EIM</td>
<td>Enterprise Integration Manager</td>
</tr>
<tr>
<td>ERP</td>
<td>Enterprise Resource Planning</td>
</tr>
<tr>
<td>Exadata</td>
<td>Oracle Exadata Database Machine provides an optimal solution for all database workloads, ranging from scan-intensive data warehouse applications to highly concurrent OLTP applications. The Database Machine delivers extreme performance in a highly-available, highly-secure environment and is also well-suited for consolidating multiple databases onto a single grid. The Database Machine is delivered as a complete pre-optimized and pre-configured package of software, servers, and storage.</td>
</tr>
<tr>
<td>Exalogic</td>
<td>Oracle Exalogic is an engineered system that is highly optimized and pre-configured to run multi-tier application workloads. Most Oracle Applications are based on a multi-tier architecture and can therefore benefit from Exalogic.</td>
</tr>
<tr>
<td>MDM</td>
<td>Master Data Management</td>
</tr>
<tr>
<td>OLTP</td>
<td>On Line Transaction Processing</td>
</tr>
<tr>
<td>RAC</td>
<td>Real Application Clusters</td>
</tr>
<tr>
<td>RTD</td>
<td>Real-Time Decisions</td>
</tr>
<tr>
<td>SDP</td>
<td>Sockets Direct Protocol</td>
</tr>
<tr>
<td>SDH</td>
<td>Source Data History</td>
</tr>
<tr>
<td>SOA</td>
<td>Service Oriented Architecture</td>
</tr>
<tr>
<td>SQL</td>
<td>Structured Query Language</td>
</tr>
<tr>
<td>TCO</td>
<td>Total Cost of Ownership</td>
</tr>
<tr>
<td>UCM</td>
<td>Universal Customer Master</td>
</tr>
</tbody>
</table>
MDM on Exadata - Topology of the Test Environment

The above provided results clearly show the significant speed advantages of Exadata well beyond the throughput and response-time requirements.
The following are the results of medium query and complex query with increasing number of users.

<table>
<thead>
<tr>
<th>Operation</th>
<th># of Users</th>
<th>Throughput (tps)</th>
<th>Response Time (m sec)</th>
<th>Web CPU</th>
<th>App CPU</th>
<th>DB CPU</th>
</tr>
</thead>
<tbody>
<tr>
<td>Medium Query</td>
<td>100</td>
<td>4450</td>
<td>25</td>
<td>8%</td>
<td>25%</td>
<td>8%</td>
</tr>
<tr>
<td>Medium Query</td>
<td>200</td>
<td>8200</td>
<td>27</td>
<td>15%</td>
<td>355%</td>
<td>13%</td>
</tr>
<tr>
<td>Medium Query</td>
<td>300</td>
<td>10,450</td>
<td>29</td>
<td>22%</td>
<td>60%</td>
<td>16%</td>
</tr>
<tr>
<td>Complex Query</td>
<td>100</td>
<td>2397</td>
<td>40</td>
<td>6.5%</td>
<td>30%</td>
<td>10%</td>
</tr>
<tr>
<td>Complex Query</td>
<td>200</td>
<td>4696</td>
<td>41</td>
<td>13%</td>
<td>52%</td>
<td>18%</td>
</tr>
<tr>
<td>Complex Query</td>
<td>300</td>
<td>5657</td>
<td>49</td>
<td>15%</td>
<td>74%</td>
<td>21%</td>
</tr>
</tbody>
</table>

Figure 8: MDM on Exadata. Test results for Medium and Complex Query

Here the test shows the linear scalability of the application with step increases in number of users. As the number of users increase, we can see the throughput increasing linearly.