Evolving To The Big Data Warehouse

Kevin Lancaster
Director, Engineered Systems, Oracle EMEA
Big Data In Action

Make Better Decisions Using Big Data
Oracle Integrated Solution Stack for All Data

ACQUIRE
- HDFS
- Oracle NoSQL Database
- Enterprise Applications

ORGANIZE
- Hadoop (MapReduce)
- Oracle Loader for Hadoop
- Oracle Data Integrator

ANALYZE
- Data Warehouse
- In-Database Analytics

DECIDE
- Analytic Applications
EXALYTICS
IN-MEMORY BI MACHINE
BIG DATA
APPLIANCE
Oracle NoSQL Database

- Key value pair database
- Dynamic data model
- Highly scalable, available
- Transparent load balancing
- Built using BerkeleyDB
Hadoop Architecture

- Management/Monitoring
- MapReduce
- Hadoop Distributed File System (HDFS)

Distributed file system with redundant storage
Map/Reduce programming paradigm
Highly scalable data processing
Cost-effective model for high volume, low density data
A Map/Reduce Pipeline
Oracle Big Data Connectors

- Oracle Data Integrator Application Adapter for Hadoop
- Oracle Loader for Hadoop
- Oracle Direct Connector for Hadoop Distributed File System
- Oracle R Connector for Hadoop
Oracle Data Integrator

Reduces Hadoop complexities through graphical tooling
Oracle Loader for Hadoop
EXADATA
DATABASE MACHINE
The World’s Best Platform for Data Warehousing

- Oracle **Database**: #1 for DW, for decades
- Oracle **Exadata**: #1 Platform for the #1 Database

**Oracle #1 for Data Warehousing**
Worldwide Data Warehouse Management Software Vendor Share

- Oracle 40.4%
- IBM 22.2%
- Microsoft 16.7%
- Teradata 12.0%

*Source: IQV Worldwide Data Warehouse Management Software Vendor Share, 2011 Report, Aug 2011*
Oracle Database 11g
The Best Database for Data Warehousing

<table>
<thead>
<tr>
<th>Feature</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Real Application Clusters</td>
<td>✔</td>
</tr>
<tr>
<td>Advanced Compression</td>
<td>✔</td>
</tr>
<tr>
<td>Partitioning</td>
<td>✔</td>
</tr>
<tr>
<td>Advance Security &amp; Firewall</td>
<td>✔</td>
</tr>
<tr>
<td>In-Database Advanced Analytics (Data Mining + R Enterprise)</td>
<td>✔</td>
</tr>
<tr>
<td>In-Database Multidimensional OLAP</td>
<td>✔</td>
</tr>
<tr>
<td>In-Database Semantics, Text Mining, Spatial, Statistics...</td>
<td>✔</td>
</tr>
</tbody>
</table>

- World record performance for fast access to information
- World-class security
- The most compete in-database analytics capability
Challenge: No Single Source of Truth
Expensive Data Warehouse Architecture
Consolidate Onto a Single Platform
Faster Performance, Single Source of Truth

Oracle Database 11g
Oracle Exadata Database Machine
Built-In Analytics
Secure, Scalable Platform for Advanced Analytics

- Complex and predictive analytics embedded into Oracle Database 11g
- Reduce cost of additional hardware, management resources
- Even more performance by eliminating data movement and duplication, specialist data structures
Partition for Performance

Partition Pruning

What was the total sales amount for May 20 and May 21 2010?

```
Select sum(sales_amount)
From SALES
Where sales_date between
  to_date('05/20/2011','MM/DD/YYYY')
And
  to_date('05/22/2011','MM/DD/YYYY');
```

- Performs operations only on relevant partitions
- Dramatically reduces amount of data retrieved from disk
- Improves query performance and optimizes resource utilization
Oracle Exadata Database Machine
For OLTP, Data Warehousing & Consolidated Workloads

- **Improve query performance by 10x**
  - Better insight into customer requirements
  - Expand revenue opportunities

- **Consolidate OLTP and analytic workloads**
  - Lower admin and maintenance costs
  - Reduce points of failure

- **Integrate analytics and data mining**
  - Complex and predictive analytics

- **Lower risk**
  - Streamline deployment
  - One support contact
Oracle Exadata Database Machine Family

Oracle Exadata Database Machine X2-2

Oracle Database Server Pool
- 8 2-processor Database Servers
  - 96 CPU Cores
  - 768 GB Memory
  - Oracle Linux or Solaris 11 Express

Exadata Storage Server Pool
- 14 Storage Servers
  - 5 TB Smart Flash Cache
  - 504 TB Disk Storage

Unified Server/Storage Network
- 40 Gb/sec Infiniband Links

Available in full, half, quarter racks
Oracle Exadata Database Machine Family

Oracle Exadata Database Machine X2-8

Oracle Database Server Pool
• 2 8-processor Database Servers
  – 160 CPU Cores
  – 4 TB Memory
  – Oracle Linux or Solaris 11 Express

Exadata Storage Server Pool
• 14 Storage Servers
  – 5 TB Smart Flash Cache
  – 504 TB Disk Storage

Unified Server/Storage Network
• 40 Gb/sec Infiniband Links

Full and multi-rack configuration
Exadata Smart Scan
Improve Query Performance by 10x or More

Select sum(sales) where salesdate= ‘22-Jan-2011’…

What Were Yesterday's Sales?

Return Sales for Jan 22 2011

- Data intensive processing runs in Exadata Storage Servers
- Rows and columns filtered as data streams from disks
- Complex operations also run in storage
- Parallelize query execution and removes bottlenecks
Exadata Smart Flash Cache
Extreme Performance for Random Access & Database OLAP Cubes

- Full rack has 5 TB of Smart Flash Cache
- Can process over 1.5 million IOs per second
- 50 GB/sec query throughput on uncompressed data
- 5x more I/Os than 1000 Disk Enterprise Storage Array
Exadata Hybrid Columnar Compression
Reduce Disk Space Requirements

- Uncompressed Data: 1.4x
- Data Warehouse Appliances: 2.5x
- OLTP Data: 3x
- DW Data: 10x
- Archive Data: 15x

Oracle
Oracle Exadata Momentum
Rapid Adoption in All Geographies and Industries
Oracle Exadata for Data Warehousing
### Challenge: User Requirements

#### Leading Data Warehouse Challenges

<table>
<thead>
<tr>
<th>Challenge</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Warehouse Performance</td>
<td>45%</td>
</tr>
<tr>
<td>Maintenance and Admin Costs</td>
<td>43%</td>
</tr>
<tr>
<td>Data Integration</td>
<td>35%</td>
</tr>
<tr>
<td>Maintaining/Updating Operational/Admin Skills</td>
<td>31%</td>
</tr>
<tr>
<td>Purchasing and Installation Costs</td>
<td>25%</td>
</tr>
<tr>
<td>Inability and Cost to Scale to Support Growing</td>
<td>16%</td>
</tr>
<tr>
<td>Can’t Support Real-Time Data Warehousing</td>
<td>14%</td>
</tr>
<tr>
<td>Inadequate High Availability</td>
<td>13%</td>
</tr>
<tr>
<td>Can’t Support Advanced Analytics</td>
<td>12%</td>
</tr>
<tr>
<td>Inadequate Security</td>
<td>11%</td>
</tr>
<tr>
<td>No Major Issues Encountered</td>
<td>11%</td>
</tr>
<tr>
<td>Can’t Support Mixed Workloads</td>
<td>9%</td>
</tr>
<tr>
<td>Don’t Know/Unsure</td>
<td>8%</td>
</tr>
<tr>
<td>Other</td>
<td>3%</td>
</tr>
</tbody>
</table>

Source: A New Dimension to Data Warehousing, 2011 IOUG Data Warehousing Survey
Hardware and Software

Engineered to Work Together