Big Data – Are You Ready?

Thomas Kyte
http://asktom.oracle.com
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.
<table>
<thead>
<tr>
<th>Month</th>
<th>Event</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dec 2010</td>
<td>Oracle Exadata Database Machine X2-8 shipping</td>
</tr>
<tr>
<td>Feb 2011</td>
<td>Oracle Database Firewall</td>
</tr>
<tr>
<td>Apr 2011</td>
<td>Oracle Exadata Automatic Service Requests</td>
</tr>
<tr>
<td>May 2011</td>
<td>Oracle Exadata Solaris support</td>
</tr>
<tr>
<td>Jun 2011</td>
<td>Oracle Exadata certification of SAP</td>
</tr>
<tr>
<td>Jun 2011</td>
<td>Oracle Exadata Storage Expansion Racks</td>
</tr>
<tr>
<td>Sep 2011</td>
<td>Oracle Database 11.2.0.3 Patch Set released</td>
</tr>
<tr>
<td>Sep 2011</td>
<td>Oracle Database Appliance shipping</td>
</tr>
</tbody>
</table>
Big Data Buzz

“Why big data is a big deal”
InfoWorld – 9/1/11

“The challenge— and opportunity— of big data”
McKinsey Quarterly—5/11

“Ten reasons why Big Data will change the travel industry”
Tnooz -8/15/11

“Keeping Afloat in a Sea of 'Big Data”
ITBusinessEdge – 9/6/11

“Getting a Handle on Big Data with Hadoop”
Businessweek-9/7/11

“The promise of Big Data”
Intelligent Utility-8/28/11
### Big Data Use Cases

<table>
<thead>
<tr>
<th>Today’s Challenge</th>
<th>New Data</th>
<th>What’s Possible</th>
</tr>
</thead>
<tbody>
<tr>
<td>Healthcare</td>
<td>Remote patient monitoring</td>
<td>Preventive care, reduced hospitalization</td>
</tr>
<tr>
<td>Expensive office visits</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Manufacturing</td>
<td>Product sensors</td>
<td>Automated diagnosis, support</td>
</tr>
<tr>
<td>In-person support</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Location-Based Services</td>
<td>Real time location data</td>
<td>Geo-advertising, traffic, local search</td>
</tr>
<tr>
<td>Based on home zip code</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Public Sector</td>
<td>Citizen surveys</td>
<td>Tailored services, cost reductions</td>
</tr>
<tr>
<td>Standardized services</td>
<td>Social media</td>
<td>Sentiment analysis segmentation</td>
</tr>
<tr>
<td>Retail</td>
<td></td>
<td></td>
</tr>
<tr>
<td>One size fits all marketing</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
What Makes it Big Data?

- **Volume**
- **Velocity**
- **Variety**
- **Value**

Keywords:

- Social
- Blog
- Smart Meter

Binary Code:

```
101100101001
001001101010
101011100101
010100100101
```
## Why Is Big Data Important?

<table>
<thead>
<tr>
<th>Sector</th>
<th>Impact Description</th>
<th>Value/Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>US Health Care</td>
<td>Increase industry value per year by</td>
<td>$300 B</td>
</tr>
<tr>
<td>Manufacturing</td>
<td>Decrease dev., assembly costs by</td>
<td>−50%</td>
</tr>
<tr>
<td>Global Personal Location Data</td>
<td>Increase service provider revenue by</td>
<td>$100 B</td>
</tr>
<tr>
<td>Europe Public Sector Admin</td>
<td>Increase industry value per year by</td>
<td>€250 B</td>
</tr>
<tr>
<td>US Retail</td>
<td>Increase net margin by</td>
<td>60+%</td>
</tr>
</tbody>
</table>

Source: *McKinsey Global Institute: Big Data – The next frontier for innovation, competition and productivity (May 2011)*
Big Data in Action

Make Better Decisions Using Big Data
Big Data in Action

Acquire all available data
Acquiring Big Data Challenge

- Need to process high volume, low-density information
- Application will need to change frequently
- Must scale out to meet aggressive roll out plan
Oracle NoSQL Database

Key value pair database
Dynamic data model
Highly scalable, available
Transparent load balancing
Built using BerkeleyDB
Oracle NoSQL Database

- Key value pair database
- Dynamic data model
- Highly scalable, available
- Transparent load balancing
- Built using BerkeleyDB
Oracle NoSQL Database

Oracle NoSQL: Practically ACID

The serious part of Oracle NoSQL is a practical approximation of ACID compliance, the standard that SQL databases like to offer. ACID means "Atomic, Consistent, Isolated, Durable transactions," and there's a robust debate about just what this translates to in excruciating detail. Most NoSQL systems promise a different acronym, BASE, which stands for "Basically Available, Soft State, and Eventually Consistent." In other words, you'll probably get the right answer except when you don't.

Key value pair database
Dynamic data model
Highly scalable, available
Transparent load balancing
Built using BerkeleyDB
In all, Oracle NoSQL was a pleasure to try because it offered so many serious features developed by a company with a deep history of serious data management. There are dozens of small ways in which the tool is more thorough and sophisticated than the simpler NoSQL projects. You get a number of different options for increasing the durability in the face of a node crash or trading that durability for speed. The documentation is solid and written by working engineers with deep experience in storing data for enterprise customers.
Big Data in Action

Oracle NoSQL Database
Organize and distill big data using massive parallelism
Organizing Big Data Challenge

- Have existing Oracle data warehouse
- Also want to perform analysis on big data
- Can’t negatively impact data warehouse SLAs
Analysis Sandbox

Provides analysis workspace
Controlled access to resources and data
Doesn’t impact production system
Sandboxing with Oracle Enterprise Manager

- Simple to set up
- Efficient server utilization
- Secure and scalable
- Accountable via charge back
- Ideal for Oracle Exadata
Big Data in Action

Oracle NoSQL Database
Oracle Enterprise Manager
Organizing and Distilling Big Data Challenge

Must transform big data into something easily analyzed

Want to avoid writing lots of Hadoop code

Need to load data quickly into Oracle Data Warehouse
Hadoop Architecture

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

**Key Features:**

- 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 Data Integrator

Reduces Hadoop complexities through graphical tooling
Big Data in Action

Oracle NoSQL Database
Oracle Enterprise Manager
Oracle Data Integrator
Oracle Loader for Hadoop
Big Data in Action

Analyse **all** your data, at once
Analyzing Big Data Challenge

Require access to all data

Want to perform statistical analysis using R

Doing analysis on a laptop is slow and not secure
R Statistical Programming Language

Open source language and environment

Used for statistical computing and graphics

Strength in easily producing publication-quality graphs

Highly extensible
Why R Wasn’t Ready for the Enterprise

Small data models only are stored and run on user’s laptop
Oracle R Enterprise Approach

- Models run in-database
- Processes large data sets
- Uses the power of Oracle Database 11g and Exadata
- Same code, much faster
Big Data in Action

Oracle NoSQL Database
Oracle Enterprise Manager
Oracle Data Integrator
Oracle Loader for Hadoop
Oracle R Enterprise
Decide based on real-time big data
Making Decisions Based on Big Data Challenge

Big data has been transformed into actionable insight

Want to add new insights into BI dashboard

How do we quickly integrate R analytics into dashboard?
Dashboard Analytics

• Oracle Business Intelligence Enterprise Edition
  – Advanced dashboard visualization
  – Runs BI and EPM applications

• Integrating R Analytics
  – Embed R script’s web interface in BI dashboard
  – Graphics will stream to BI dashboard
Oracle Exalytics Hardware

Engineered for extreme analytics

- 40 Intel processor cores
- 1 Terabyte main memory
- 40 Gb InfiniBand connection to Oracle Exadata
Oracle Exalytics Software

• Oracle TimesTen In-Memory Database
  – Adaptive in-memory caching of analytics
  – In-memory columnar compression
  – Tightly integrated with Oracle Exadata
  – Enables speed-of-thought visualization

• Oracle Business Intelligence Foundation Suite
Oracle Integrated Solution Stack for Big 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
Oracle Big Data Appliance Hardware

• 18 Sun X4270 M2 Servers
  – 48 GB memory per node = 864 GB memory
  – 12 Intel cores per node = 216 cores
  – 24 TB storage per node = 432 TB storage
• 40 Gb p/sec InfiniBand
• 10 Gb p/sec Ethernet
Oracle Big Data Appliance Software

- Oracle Linux
- Java Hotspot VM
- Apache Hadoop Distribution
- R Distribution
- Oracle NoSQL Database
- Oracle Data Integrator for Hadoop
- Oracle Loader for Hadoop
The preceding 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.
Hardware and Software
Engineered to Work Together
Maximizing the Value of Enterprise Big Data

• Hardware and software for Big Data
• Integrates all enterprise data
  – Structured and unstructured
  – SQL and NoSQL
• Fastest time-to-market
• Single vendor support