Big Data
Mastering Big Data
and Your Analytic Architecture

Jonathan Beasley
Oracle Enterprise Architect
Agenda

- The Big Data Architecture Challenge
- Targeting an Agile and Adaptive Architecture
- Unified Information Management Architecture
- An Example for Building a Unified Analytics Solution
Big Data: Challenge to Value

Business Value

Challenges

➢ High Variety
➢ High Volume
➢ High Velocity

Today

➢ Deep Analytics
➢ High Agility
➢ Massive Scalability
➢ Real Time

Tomorrow
Agenda

- The Big Data Architecture Challenge
- Targeting an Agile and Adaptive Architecture
- Unified Information Management Architecture
- An Example for Building a Unified Analytics Solution
Is Your Analytic Information Agile and Adaptive?
Adaptive, ROI-Driven Information Cycle
Actionable, Real-Time, Integrated

Who are our customers?

Discovery

Learning

How do they respond to our brand, products, and services?

Actions to improve sales and intimacy

Application

Hypothesis

Test for patterns of customer need and improved sales response
Sentiment-Driven Information Cycle
Actionable, Real-Time, Integrated

What do our customers value?
- Discovery

What are their interests?
- Learning
  Who do they tell?

Actions to improve sales and intimacy
- Application

Test for new patterns for customer influence
- Hypothesis
The Agile and Adaptive Enterprise

Information Revelation

• Comes from everywhere
• Content is unpredictable
• Correlation adds value
• Analytics must be available to people and processes
Agenda

- The Big Data Architecture Challenge
- Targeting an Agile and Adaptive Architecture
- Unified Information Management Architecture
- An Example for Building a Unified Analytics Solution
Should Big Data be yet another Information Silo?
An Architect’s Approach to Enterprise Initiatives

Information Architecture Capability Model

Diverse Data Realms
- Master
- Transaction
- Reference
- Analytical
- Metadata
- Unstructured
- Big Data

Infrastructure
- Sharing & Delivery
- BI & Data Warehouse
- Integration
- Content Management
- Enterprise Data Model
- Master Data Mgmt
- Security
- Governance
Information Management Architecture

Source Data

Enterprise Data Warehouse

Data Integration

Information Access

Security and Metadata
Traditional Data Architecture

Source Data Layer
- Processes
- COTS/ERP
- External

Enterprise Data Warehouse

Staging Data Layer
- Strongly Typed Data
- Weakly Typed Data

Foundation Layer
- Enterprise Data with full history

Performance Layer
- Embedded Data Marts

Security and Metadata
- Data Quality

Data Integration
Add New Data Sources

Architecture Insights

Goals:
- Transform low density data to extract value
- Balance schema design with time to value
- Prevent Data Loss
- Provide Historic Data Retention

Value:
- Enrich data, enable correlation

How:
- Maximum parallelism, Map-Reduce,
  Targeted languages, Statistics

Key Decisions:
- Software Tools, Hardware, Storage, Network
- Repeatable Process - Productivity
- Standards and Support
Add Knowledge Discovery

Architecture Insights

• **Goals:** Enable business centric, rapid exploration
  Enable deep analytic capabilities

• **Value:** Assess value → find patterns

• **How:** Sampling, correlation, statistics, modeling

• **Key Decisions:**
  ➢ Minimize data movement
  ➢ Support individualized data discovery
  ➢ V-V-V requires enterprise scale, security
Knowledge Pools and Correlation

Architecture Insights

- **Goal:** Analytical Performance
- **Value:** Rapid modeling iterations
- **How:** Move analytics to the data
- **Key Decisions:**
  - Minimize data movement
  - Use optimizations: hardware & SQL
  - Consider “R” as portable statistical platform
Add Information Access

**Architecture Insights**

- **Goal:** Secure, timely data sharing (No silos!)
- **Value:** Widespread access
- **How:** Standardized services, federation
- **Key Decisions:**
  - Leverage existing skills, standards, and tools
  - Build services interfaces for easy sharing and operational integration
  - Plan architecture from access to ingest

**Source Data Layer**

- **Goal:** Secure, timely data sharing (No silos!)
- **Value:** Widespread access
- **How:** Standardized services, federation
- **Key Decisions:**
  - Leverage existing skills, standards, and tools
  - Build services interfaces for easy sharing and operational integration
  - Plan architecture from access to ingest
Translated into Oracle Product Architecture

Source Data Layer
- Processes
- COTS/ERP
- External
- Social/Text
- Sensors
- Streaming
- Security and Metadata

Enterprise Data Warehouse
- Staging Data Layer
  - Strongly Typed Data
  - Data Quality
  - Weakly Typed Data

Foundation Layer
- Enterprise Data with full history

Performance Layer
- Embedded Data Marts

Oracle Exadata

Oracle Exalytics

Information Access
- Performance Management
- Alerts, Dashboards, Reporting
- BI Abstraction & Query Federation
- Information Discovery
- Advanced Analysis & Data Science

Oracle Big Data Appliance

Oracle Big Data Connectors
Agenda

- The Big Data Architecture Challenge
- Targeting an Agile and Adaptive Architecture
- Unified Information Management Architecture
- An Example for Building a Unified Analytics Solution
An Example

Multi-Channel Retail Shopping Conversion

On-line and In-Store Customers

Business Goals

- Increase sales through all channels
- Reduce abandoned on-line carts
- Reduce out-of-stock promotional items
- Improve satisfaction & word-of-mouth

Challenges

- Linking structured and unstructured
- Capturing where items are placed
- Customer sentiment on social media
Goal: Make an Informed Recommendation

Structured Sources
- Customer Analytics
  - Customer History
- Product Analytics
  - Inventory Position

Unstructured Sources
- Sentiment & Influence
- Store Placement

Sensors

Real-Time Recommendations
- Analytics
- Cart
- Search
- Promotions
Using all Data Sources to Understand Customer Website Logs & Data
Discovering Valuable Data

Knowledge Discovery Engine

High Volume Distributed File System

Data Warehouse

Website Logs & Data

NoSQL DB

Structured

Unstructured

Semi-structured

Copyright © 2012, Oracle and/or its affiliates. All rights reserved.
Valuable Data Found – Now Store it Securely

Knowledge Discovery Engine

Data Warehouse

High Volume Distributed File System

Discoveries

MapReduce code separates valued data, then sent to via specialized adapters to Data Warehouse

Persistent Data Store for All Data of Value

Website Logs & Data

NoSQL DB
Deploy Widely Available Reports & Analytics

Knowledge Discovery Engine

Data Warehouse

BI Tools and Dashboards

Persistent Data Store for All Data of Value + In-DB Analytics

High Volume Distributed File System

Sensors

Website Logs & Data

NoSQL DB

MapReduce code separates valued data, then sent to via specialized adapters to Data Warehouse

Enterprise-class for reporting & analysis
Feed the Recommendation Engine

Knowledge Discovery Engine

High Volume Distributed File System

Data Warehouse

BI Tools and Dashboards

Real-Time Analytics and Recommendations

Website Logs & Data

NoSQL DB

Persistent Data Store for All Data of Value + In-DB Analytics

MapReduce code separates valued data, then sent to via specialized adapters to Data Warehouse

Update Website Recommendations

Sensors
Make Well-Tuned Real-Time Recommendations

Persistent Data Store for All Data of Value + In-DB Analytics

Knowledge Discovery Engine

Data Warehouse

BI Tools and Dashboards

High Volume Distributed File System

MapReduce code separates valued data, then sent to via specialized adapters to Data Warehouse

Real-Time Analytics and Recommendations

NoSQL DB

Website Logs & Data

Location & User Profile

Recommend

Sensors

Copyright © 2012, Oracle and/or its affiliates. All rights reserved.
Summary
Big Data Adds to Your Information Architecture

Challenge:
Exploiting Synergies

DECIDE
ACQUIRE
ANALYZE
ORGANIZE
Oracle Big Data Architecture Capabilities

Data
- Structured: Master & Reference, Transactions, Machine Generated
- Semi-structured: Files
- Unstructured: NoSQL, HDFS

Acquire
- DBMS (OLTP)
- ETL/ELT
- ChangeDC
- Real-Time
- Message-Based
- Hadoop (MapReduce)

Organize
- ODS
- Warehouse
- Streaming (EP Engine)
- In-Database Analytics

Analyze
- Reporting & Dashboards
- Alerting & Recommendations
- EPM, BI, Social Applications
- Text Analytics and Search
- Advanced Analytics
- Interactive Discovery

Decide

Specialized Hardware
- Big Data Cluster
- High Speed Network
- RDBMS Cluster
- In Memory Analytics

Management, Security, Governance

33 | Copyright © 2012, Oracle and/or its affiliates. All rights reserved.
Oracle Big Data Platform

**Oracle Big Data Appliance**
Optimized for Hadoop, R, and NoSQL Processing

- Cloudera & Hadoop
- Open Source R
- Oracle NoSQL
- Applications

**Oracle Big Data Connectors**

**Oracle Exadata**
“System of Record”
Optimized for DW/OLTP

- Oracle Big Data Connectors
- Oracle Advanced Analytics
- Data Warehouse
- Oracle Database
- In-DATABASE Analytics

**Oracle Exalytics**
Optimized for Analytics & In-Memory Workloads

- Oracle Enterprise Performance Management
- Oracle Business Intelligence Applications
- Oracle Business Intelligence Tools
- Oracle Endeca Information Discovery

Stream | Acquire | Organize | Analyze and Visualize
Hardware and Software
Engineered to Work Together