Oracle Enterprise Data Quality - Overview and Roadmap

Mike Matthews
Director, Product Management

Martin Boyd
Senior Director, Product Strategy
Safe Harbor Statement

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.
Agenda

1. Why Worry about Data Quality?
2. EDQ Product Roadmap
3. EDQ Investment Themes
4. Q & A and For More Information
Why worry about Data Quality?

Garbage -in Garbage -out
Quality of Data Impacts Everything

Quality = Fitness for purpose
Complete – Valid – Consistent – Timely – Accurate

Poor data quality is the primary reason for 40% of all business initiatives failing to achieve their targeted benefits” Gartner

“Only 30% of BI/DW implementations fully succeed. The top two reasons for failure? Budget constraints and data quality.” Gartner

“More than 50 percent of data warehouse projects will have limited acceptance, or will be outright failures, as a result of a lack of attention to data quality issues” Gartner

“Through 2016, 25% of organizations using consumer data will face reputation damage due to inadequate understanding of information trust issues” Gartner

“It does not really matter how good your management sponsorship or your business-driven motivation is. If you do not have the data, or the data does not have sufficient quality, any BI implementation will fail.” Kimball

Copyright © 2014 Oracle and/or its affiliates. All rights reserved. |
Oracle Enterprise Data Quality

Common Access/UI

Govern
Monitor effectiveness & resolve problems

Match
Identify & merge duplicates

Standardize
Drive conformance to standards

Profile
Quickly understand data content

Enterprise DQ Platform

Market-leading usability for all types of data

Unparalleled time-to-value

High performance engine

Out-of-the-box global knowledge-base

Foundation for governance program
Common Data Quality Use Cases

**System Consolidation/Migration**
- Enforce new system standards on legacy data

**Business Intelligence Enablement**
- Enforce BI standards on disparate data

**MDM Enablement**
- Verify, standardize, match and merge data from disparate sources

**Application Enablement**
- Clean-up and govern application data for a single application (CRM, HR, PLM, Retail search, etc.)

**Compliance**
- Drive consistent data and processes to meet regulatory requirements (watchlist screening, anti-money laundering, tax compliance, etc.)

**Enterprise DQ Services**
- Common DQ services to clean-up and govern application data across multiple applications
Common Data Quality Use Cases

System Consolidation/Migration
• Enforce new system standards on legacy data

Business Intelligence Enablement
• Enforce BI standards on disparate data

MDM Enablement
• Verify, standardize, match and merge data from disparate sources

Enterprise DQ Services
• Common DQ services to clean-up and govern application data across multiple applications

Assuring quality of data before migration is essential to a smooth go-live (insurance policy against delays and operational problems)

Compliance
• Drive consistent data and processes to meet regulatory requirements (watchlist screening, anti-money laundering, tax compliance, etc.)
Managing Migration

US-Based Manufacturer
Using Oracle Enterprise Data Quality to migrate 14 legacy ERP Systems to Oracle EBS, also integration of customer data from acquisitions
Common Data Quality Use Cases

System Consolidation/Migration
- Enforce new system standards on legacy data

Application Enablement
- Clean-up and govern application data for a single application (CRM, HR, PLM, Retail search, etc.)

Compliance
- Drive consistent data and processes to meet regulatory requirements (watchlist screening, anti-money laundering, tax compliance, etc.)

Enterprise DQ Services
- Common DQ services to clean-up and govern application data across multiple applications

Even after initial data preparation, ongoing data governance is required
Application Enablement

Leading Independent Mobile Services Company
Oracle Enterprise Data Quality validates, standardizes and de-duplicates customer data for Siebel CRM
Common Data Quality Use Cases

System Consolidation/Migration
- Enforce new system standards on legacy data

Business Intelligence Enablement
- Enforce BI standards on disparate data

MDM Enablement
- Verify, standardize, match and merge data from disparate sources

Business Intelligence is perhaps the most obvious example of garbage-in, garbage-out

Compliance
- Drive consistent data and processes to meet regulatory requirements (watchlist screening, anti-money laundering, tax compliance, etc.)

Enterprise DQ Services
- Common DQ services to clean-up and govern application data across multiple applications

Oracle OpenWorld 2014
National Court System
Oracle Enterprise Data Quality parses and standardizes court decisions for compilation into a national database.
Common Data Quality Use Cases

System Consolidation/Migration
- Enforce new system standards on legacy data

Compliance
- Drive consistent data and processes to meet regulatory requirements (watchlist screening, anti-money laundering, tax compliance, etc.)

Enterprise DQ Services
- Common DQ services to clean-up and govern application data across multiple applications

Application Enablement
- Clean-up and govern application data for a single application (CRM, HR, PLM, Retail search, etc.)

MDM Enablement
- Verify, standardize, match and merge data from disparate sources

Sometimes best practices are mandated by law and have large penalties for non-compliance (FATCA, OFAC, etc.)
Watchlist Screening

3 of 4 top US and UK Banks
Screen hundreds of millions of customers against Government and other watch lists to ensure they are not managing money for terrorists
Common Data Quality Use Cases

System Consolidation/Migration
- Enforce new system standards on legacy data

Business Intelligence Enablement
- Enforce BI standards on disparate data

MDM Enablement
- Verify, standardize, match and merge data from disparate sources

Even where hub implementation is a core objective, DQ often comes first
- To prepare the way
- Lower implementation risk
- Deliver tactical ROI

Enterprise DQ Services
- Common DQ services to clean-up and govern application data across multiple applications

Compliance
- Drive consistent data and processes to meet regulatory requirements (watchlist screening, anti-money laundering, tax compliance, etc.)
Loyalty Program

Premier Roaster and Retailer of Specialty Coffee
Oracle Enterprise Data Quality plus Oracle Customer Hub used to manage ground-breaking loyalty program
Common Data Quality Use Cases

Enforcing common DQ standards across the enterprise is a simple, tactical fix to systemic problems and a good foundation for enterprise governance.

**Application Enablement**
- Clean-up and govern application data for a single application (CRM, HR, PLM, Retail search, etc.)

**Compliance**
- Drive consistent data and processes to meet regulatory requirements (watchlist screening, anti-money laundering, tax compliance, etc.)

**MDM Enablement**
- Verify, standardize, match and merge data from disparate sources

**Enterprise DQ Services**
- Common DQ services to clean-up and govern application data across multiple applications
Enterprise DQ Services

US-Based Manufacturing Company
Using Oracle Enterprise Data Quality to standardize data validations as all entry and transfer points to deliver centralized control in a distributed environment
Agenda

1. Why Worry about Data Quality?
2. EDQ Product Roadmap
3. EDQ Investment Themes
4. Q & A and For More Information
**EDQ Product History & Roadmap**

**EDQ 9.0**
- Customer Data Services Pack
- Product Data Processor
- New Address Verification (Loqate) Connector
- Dynamic Configuration & Data Overrides
- Operations UI (Server Console)
- Oracle Data Integrator integration

**EDQ 11.1.1.7**
- Fusion Middleware Integration
- Case Management expansion
- Reference Published Processors
- UI Localization to 9 Languages
- New Job Manager
- Real-time performance improvements

**EDQ 12.1.3**
- Oracle Access Manager integration
- Reference Data Generation
- Improved processor toolkit

**EDQ 12.2.1**
- WebLogic Clustering Support
- Improved Match Flexibility
- Big Data Connectivity
- REST APIs
- Extended Event Auditing
- Autotune matching services

**Future**
- Integrated Data Governance Center
- DQ Services to and in Cloud Applications
- Integrated Profiling
- Enterprise Manager integration
- Accessibility

---

**2012**
- Address Verification 12
  - Transliteration
  - Performance Improvements

**2013**
- Address Verification 13
  - Point-level Geocoding

**2014**
- Address Verification 14
  - PowerSearch
  - CASS, SERP, AMAS Certification

**2015**
- Address Verification 15
  - Improved Typo Handling
  - Data Pack Improvements
  - Geocoding Accuracy Improvements

**2016**
- Address Verification Future
  - Extended data provider support
  - Standalone Parsing

**2017**
- EDQ in Oracle Sales Cloud R9
  - Duplicate Prevention
  - Address Verification and Search

**Future**
- EDQ in Oracle Sales Cloud Future
  - DQ Health Checks
  - Integrated Profiling
Current: EDQ 12.1.3
Released June 2014

Performance Improvements
Improved performance of Director UI when working with large databases

Deeper Integration with the Oracle Stack
12c Certification
  Simplified Installation on WebLogic Server
  Oracle Access Manager Integration
  Recertification of ODI tool integration

Automated Rule Generation
Ability to generate rules by writing Reference Data directly from processes

Improved Processor Toolkit
Choose whether or not to package Reference Data with Published Processors
WebLogic Clustering Support
Simplified multi-server provisioning and operation

Next Generation Matching
Multi-tenant cloud matching services
Combine score-based and rule-based approaches
Rich match explanation output for users and Data Stewards
Auto-tune match services to data

Big Data Connectivity
Pull data from Hadoop (Hive)

REST Web Services
Dynamic Web Services support both REST and SOAP
REST Configuration Web Services v1 (get Projects, Jobs)
EDQ 12.2.1 – WebLogic Clustering

Operational Benefits
- Simplified operation of a multi-server system with one configuration
- Batch jobs routed to least busy server (extensible algorithm)

Elastic scale-out
- EDQ real-time jobs automatically start on a new server when added to the cluster
- New server is immediately available for batch jobs

Repository DB resilience
- Single, shared Config and Results schemas for all servers
- SQL Retries to tolerate RAC node failure in real-time processes
- High Availability of EDQ Case Management

Cluster
- Admin Server
- Managed Server 1
- Managed Server 2
- Managed Server 3

Database (RAC/non-RAC)

EDQ UIs
EDQ 12.2.1 – Next Generation Matching

Rich per tenant configuration: mappings, scoring, custom attributes

**Request (+ tuning overrides)**

- Person
- Organization
- Address

**Response data:**

- Which identifiers matched
- High-level reason (‘Exact’ / ‘Fuzzy’ / ‘No Data’)
- Detailed reason (‘Name Standardized’, ‘Email Typos’)
- Scores (‘Name 90’, ‘Address 80’, ‘Overall 86’)
EDQ 12.2.1 – Next Generation Matching

- Fewer match rules needed
- Easier to add identifiers and ways of matching
- Uses full breadth and extensibility of EDQ matching library; still easy to ‘tell Match what to do’

Define rules and scoring for each identifier

Choose how you want to score across identifiers (optional)

Richer rules define output decision
EDQ 12.2.1 – Other New Features

- REST APIs:
  - For dynamically created DQ web services
  - For external applications (ODI, Sales Cloud) to read metadata
- Extended Audit Framework
  - Configurable logging for any UI action (extended to Case Management, Case Management Admin)
  - Meets stringent FS requirements to counter fraud
- Hadoop HIVE Connector
  - Connector to pull in data from Hadoop via HIVE
- New UI
  - Alta UI improves usability of EDQ Dashboard on tablets
Agenda

1. Why worry about Data Quality?
2. EDQ Product Roadmap
3. EDQ Investment Themes
4. Q & A and For More Information
Data Quality Evolution and Disruptive Forces

Apps/Solutions
- Integrated and Accelerated DQ
  Faster, more flexible deployment, with more pre-built integrations

Data Governance
- Data Governance Foundation
  Use DQ as part of foundation for broad “Data Governance” process, tooling and organizational alignment

Cloud
- Cloud Services
  1. Matching service
  2. Address Verification service
  3. Enrichment service
  4. Integration platform services

Big Data
- Big Data Reservoir
  Use DQ techniques to identify, extract, enrich, merge data from Hadoop repositories

Faster, more flexible deployment, with more pre-built integrations

Use DQ as part of foundation for broad “Data Governance” process, tooling and organizational alignment

1. Matching service
2. Address Verification service
3. Enrichment service
4. Integration platform services

Use DQ techniques to identify, extract, enrich, merge data from Hadoop repositories
Integrated and Accelerated DQ

Integrated Data Profiling
- Generate and run a data profiling job in EDQ by calling APIs
- Link to new EDQ Dashboard UI to browse results
- Under consideration to build into a number of Oracle products

Integrated Data Services
- Further investment in out-of-the-box DQ validation services
- Apply DQ services to measure the quality of governed data
- Attach services according to Business Glossary
Data Governance Footprint

Data Governance Capabilities for Data Stewards & Stakeholders

- Data Flow Explorer
- Quality KPIs
- Issue Management
- Exception Review
- Metadata Management
- Business Glossary

Sources

Data Flows

Targets
Data Quality Powered Cloud Services

Matching for CRM
- Oracle Fusion Data Quality Matching Cloud Service
  - Dedicated matching service for Fusion CRM and other Fusion CX applications

Address Verification
- Oracle Fusion Data Quality Address Verification Service
  - Address verification and geocoding utility based on EDQ Address Verification. Can be called by any cloud or on-prem application

Matching for B2B Enrichment
- Oracle Data as a Service for B2B CRM
  - Matches data about common business entities such as companies & contacts from external sources such as Dun & Bradstreet

Data Prep / Enrichment
- Oracle Data Enrichment Cloud Service
  - Use DQ techniques to identify, extract, enrich, merge data from Hadoop repositories
Big Data still requires Quality & Governance

**Real-Time Data Movement**
- Low impact capture, stage in Hadoop
- Continuous data availability

**Data Transformation**
- Bulk data movement
- Pushdown data processing

**Data Federation**
- Query Hadoop SQL via JDBC

**Data Quality & Verification**
- Fix quality at the source
- Verify data consistency

**Metadata Management**
- Lineage and Impact Analysis
- Business Glossary Semantics
Agenda

1. Why Worry about Data Quality?
2. EDQ Product Roadmap
3. EDQ Investment Themes
4. Q & A and For More Information
Questions and Answers
<table>
<thead>
<tr>
<th>Time</th>
<th>Session Title</th>
</tr>
</thead>
<tbody>
<tr>
<td>MON</td>
<td>11:45AM – CON7776 Data Quality Maturity Journey: Building Toward Strong Enterprise Data Quality</td>
</tr>
<tr>
<td>TUE</td>
<td>10:45AM – CON7780 Oracle Enterprise Data Quality: Product Overview and Roadmap</td>
</tr>
<tr>
<td>WED</td>
<td>2:00PM – CON7775 The Essential Core of Data Governance with Oracle Enterprise Data Quality</td>
</tr>
<tr>
<td>WED</td>
<td>3:30PM – CON7934 Tapping into the Big Data Reserve with All Data</td>
</tr>
<tr>
<td>WED</td>
<td>4:45PM – CON7922 Tame Big Data with Oracle Data Integration</td>
</tr>
<tr>
<td>THU</td>
<td>12:00PM CON7931 Solving Big Data’s Big Problem with Data Preparation &amp; Enrichment in the Cloud</td>
</tr>
</tbody>
</table>
Resources
Oracle Data Integrator
Oracle GoldenGate
Oracle Enterprise Data Quality
Oracle Enterprise Metadata Management
Oracle Data Services Integrator

ORACLE FUSION MIDDLEWARE: CELEBRATE THIS YEAR'S MOST INNOVATIVE CUSTOMER SOLUTIONS

2014 Oracle Excellence Award Ceremony for Fusion Middleware Innovation

Tuesday, September 30, 2014 5:00-5:45pm
YBCA Theater (next to Moscone North)
Session ID: CON7029