Oracle Enterprise Data Quality - 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. EDQ Product Roadmap
2. EDQ Investment Themes
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
- EDQ Product Data
  - 5.6.2

**2013**
- Address Verification 13
  - Point-level Geocoding
- EDQ Product Data Extensions 11g R1
  - Endeca Integration
  - StatSim Matching

**2014**
- Address Verification 14
  - PowerSearch
  - CASS, SERP, AMAS Certification
- EDQ in Oracle Sales Cloud R9
  - Duplicate Prevention
  - Address Verification and Search

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

**2016**
- Address Verification Future
  - Extended data provider support
  - Improved search
  - Reference Data extension support

**2017**
- 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)

**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

**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
EDQ 12.2.1 – Next Generation Matching

Rich per tenant configuration: mappings, scoring, custom attributes

Request (+ tuning overrides)

EDQ Next Generation Cloud Matching Services
- 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’
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. EDQ Product Roadmap
2. EDQ Investment Themes
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
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
Oracle Fusion Data Quality Matching Cloud Service

- Matching for CRM
  - Dedicated matching service for Fusion CRM and other Fusion CX applications

Oracle Fusion Data Quality Address Verification Service

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

Oracle Data as a Service for B2B CRM

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

Oracle Data Enrichment Cloud Service

- Data Prep / Enrichment
  - 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

Oracle GoldenGate (Movement)
Oracle Data Integrator (Transformation)
Data Service Integrator (Federation)

Enterprise Data Quality
(Profile, Cleanse, Match and De-duplicate)

GoldenGate Veridata
(Online Data Verification)

Enterprise Metadata Management & Business Glossary
(Business Glossary, Data Lineage, Impact Analysis and Data Provenance)