Enterprise Data Quality Dashboards and Alerts: Holistic Data Quality

Jay Zaidi – Fannie Mae

© Fannie Mae
About the Presenter

Jay Zaidi is the Enterprise Data Quality Program Lead at Fannie Mae, with over 15 years Data Management experience. He is responsible for the Vision, Strategy, Program Management, Solution Architecture, Best Practice Development, Training and Business Requirements for Fannie Mae’s Enterprise Data Quality Program. Jay blogs at Dataversity, TDAN and DataQualityPro and has authored 25 articles on data management (Twitter handle @jayzaidi).
What you will learn

- Business drivers and importance of data quality
- Mandate for Governance and Holistic Data Quality
- Our data quality maturity journey
- The Holistic Data Quality (HDQ) concept
- Why stronger controls result in trusted data?
- Business intelligence for data quality
- Project Methodology, Challenges Faced and Strategy Used
- Sample Enterprise Data Quality Dashboard
- Tangible business benefits achieved
### Business Drivers

<table>
<thead>
<tr>
<th>Question</th>
<th>Answer</th>
</tr>
</thead>
<tbody>
<tr>
<td>What did the Global Financial Crisis in 2008 cost the world economy?</td>
<td>Between $60 Trillion and $200 Trillion</td>
</tr>
<tr>
<td>What were the root causes?</td>
<td>“Data Gaps” and “Systemic Risk”</td>
</tr>
<tr>
<td>What was the reaction?</td>
<td>Regulations &amp; Panic - Dodd-Frank (BASEL &amp; Solvency)</td>
</tr>
</tbody>
</table>
Our Mandate

- **Measure** - Provide transparency into the quality of business critical data across the information supply chain
- **Automate** - Enable straight through processing of business transactions
- **Self Serve** - Reduce dependence of the business and operations teams on technology counterparts
- **Standardize** – Define a framework for consistent definition and measurement of data quality across the enterprise
- **Educate** - Develop design patterns, best practice documentation, eLearning modules and in-house training
- **Consult** - Consult with all Lines-of-Business to advise, mentor and support their staff to remediate data quality issues
- **Collaborate** - Collaborate with enterprise architects to enforce data management policies
The Data Quality Maturity Journey

**STEP ONE**
FOUNDATION & FRAMEWORK
- DQ Use Cases
- Solution Architecture
- Industry Tool Selection
- Consistent DQ Definitions
- DQ Dimensions

**STEP TWO**
CONSTRUCTING THE RAILROAD
- Tool Deployment
- Reporting Capabilities
- Training
- Communication

**STEP THREE**
EXECUTION
- Change Management
- Awareness
- Proactive DQ Controls
- DQ Continuous Improvement
- DQ Services

#EDW12

edw2012.dataversity.net
“Holistic Data Quality” monitoring and controls result in trusted data.
“Holistic Data Quality (HDQ) is a term that I coined to highlight a paradigm shift in data quality management. Quality should not be evaluated or managed in vertical business silos alone, but in a holistic integrated (cross-silo) approach, based on the HDQ Framework. The HDQ Framework provides a mechanism for the consistent definition of data quality requirements, supports proactive quality measurements and enables continuous improvement of the quality of business critical data.”
Replace Paper Reports With Business Intelligence

Operational Incidents
Audit Findings
Data Quality Issues Report
Regulatory Compliance Issues
Weekly Data Management Status Reports

Data quality metrics at summary and detail level, historical trending, comparison across departments and proactive alerts.
Do Not Boil The Ocean

General population of data elements* 10,000 to 20,000

Critical data for a line of business* (“LOB Critical”) 2,000 to 3,000

Critical data for the enterprise* (“Enterprise Critical”) 400 to 500

Initial Focus should be on “Enterprise Critical” data

Data quality monitoring and remediation should focus on enterprise critical data.

#EDW12

edw2012.dataversity.net
Business Intelligence For Enterprise Data Quality

- Data Quality COTS product
- Business intelligence tool (COTS)
- Data quality data mart (custom)
- Data quality issue management system
- Extract Transform and Load (ETL) product
- Enterprise Service Bus (SOA and Data Quality Services)
Project Methodology

1. Identify business critical data and trusted sources (TS)
2. Define data quality dimensions to be measured for each data element
3. Define data quality rules for each dimension
4. Code and test the rules in Data Quality tool against TS
5. Migrate execution results to the mart
6. Vend results to end-users via rich operational dashboards
7. Generate alerts when exceptions exceed pre-defined thresholds
8. Triage exceptions, conduct root-cause analysis and remediate data issues
Key Themes That Emerged

1. A holistic view of data and it’s quality is needed to solve the BIG BUSINESS problems (e.g. Systemic Risk, Governance, Certification etc.)

2. Business intelligence for DQ enables quick identification of outliers and patterns. Also, provides users with ability to slice and dice data on any dimension – to answer complex questions

3. Forces consistent definition and measurement of quality related KPI’s and Operational metrics (e.g. Governance, Quality, Master Data, Metadata etc.)

4. Facilitates dialog between business, operations and technology – to identify root causes and remediate issues

5. Combine proactive (data in motion) and reactive (data at rest) data quality analysis metrics

6. Instrumented to gain a view into all trusted data sources and systems of record
Challenges Faced

1. **Access to production data stores** and files, due to performance overhead and tight SLA’s

2. **Hard to get dedicated time with subject matter experts** to elicit business requirements

3. **Waterfall development** extended time-to-market. Recommend Agile development of EDQ Platform and Business Intelligence components

4. **De-couple EDQ Platform components** (DQ Tool and Results) from the Business Intelligence components (DQ Mart and BI tool) architecturally

5. **Canned BI interface** and exception reports limited users’ ability to conduct ad-hoc analysis and slice and dice data.

6. Determining the **root cause of data-related issues is hard** and time consuming, if data lineage isn’t available

7. **Many moving parts** and multiple teams/skills required to build and deploy the solution
Execution Strategy

1. Architect and design the solution at the Global level (Enterprise) and implement locally (in phases)

2. Deploy the data quality platform using a federated model – enterprise team owns and manages infrastructure and enterprise monitoring of enterprise critical data. Lines-of-business develop local data quality solutions and monitoring capabilities

3. Architect DQ Mart and BI Solution, based on questions related to data that need to be answered and operational metrics that must be captured

4. Provide self-service BI capability to end-users. Reduces custom BI development and removes dependency of users on technical teams

5. Deploy in phases. Focus on most critical data first or areas that have major pain points due to data related issues

6. Develop standards, policies and best practices for data elements and data quality rule development
Enterprise-wide Data Quality

Quality by Line of Business

Data Quality Maturity

Core Data Breakdown

Trending of Data Quality

Health Indicators

Overall Health

Quality Rating for Each Core Data Element
Enterprise Data Quality Dashboard (Commercial Business View)

OVERALL HEALTH

CRITICAL DATA BREAKDOWN

RELEASE 1

RELEASE 2

HEALTH INDICATORS

TRENDING OF DATA QUALITY – MASTER DATA

BORROWER DATA

LOAN DATA

DATA STORE TREND

QUALITY RATING FOR EACH LOB DATA ELEMENT

DATA QUALITY OPERATIONAL METRICS
Execution Related Benefits

1. **Transparency** - All teams have instant access to summary and detailed views of data quality – no more paging through hard copy reports

2. **Business Intelligence** - Data quality patterns and anomalies in graphical format

3. **Automated alerts** - Color coded outliers and automated alerts based on configurable thresholds

4. **Better Controls** - Complete transparency into data quality rules – facilitates version control and audits

5. **Rule Re-use** - Re-use and rule sharing – huge cost saving and eliminates redundancy

6. **Compliance** - Mechanism to enforce service level agreements related to data quality

7. **Accountability** - Exceptions tracked via aging reports and data patterns. Can tie departmental quality goals and enforce accountability
Summary

• A Holistic Data Quality (HDQ) program is a strategic initiative that requires C-level sponsorship and support
• Effective data management provides order out of chaos
• Focus must be on “Enterprise Critical” data initially. Do not try to boil the ocean
• The solution architecture’s core components are the DQ tool, a data quality Data Mart and a Business Intelligence tool
• Proactive monitoring and measurement of data quality, coupled with an alerting mechanism, significantly reduces risk from unknowns
• Implementing “Holistic Data Quality” provides transparency into data quality issues and helps identify systemic issues
• Every firm needs this capability to measure and monitor the quality of its business critical data, Governance metrics, Master Data, and Metadata
Questions?