Essential Elements of Data Architecture

Data Quality – Indicator of Success

Celeste T. Harris, Affiliated Consultant, Metaview360™

Three Pillars of Data Management

Data Governance, Meta Data Management, and Data Quality are key components of an organization's Data Architecture. Together they represent the three points of stability for the discipline – like the three legs of a stool that allow it to stand and make it strong.

Data Quality Improvement Imperative

Corporations are more likely than ever to recognize the costs associated with poor data quality. The business-to-business relationships between vendors and customers require timely and accurate information for product and payment to flow smoothly. Errors in information provided to Federal agencies delay reimbursements for expenses associated with Medicare and special education services. In today's regulatory environment, inaccurate corporate reporting can result in fines or more severe legal charges against the officers of a company.

Data Quality assessment is no longer an ad-hoc activity. Methods have been developed to guide an organization from initial evaluation through long-term management of data quality to mediate the costs and risks of inaccurate data and information.

Initial Data Quality Assessment

The first step in a Data Quality Evaluation program is to examine existing sources of data. The organization needs an objective, measurable, and reportable analysis to begin the process. Anecdotal analysis of problems with the data is not enough.

A well-rounded evaluation team will include people who originate the data, people who develop and support the systems that process the data, and people who use the data. Some members of the team may be involved in Data Governance activities as Data Stewards.

The team begins by examining the domain value sets for each column of data. What percent of the occurrences of a value are incomplete? What percent are complete but invalid?

Examination of the structural integrity is the next step. Is each Account Number assigned to only one Customer? Does each Customer have only one Account Number? Is the State Code for a Customer found in the State Code Table?

The analysis of compliance to business rules is a more complex step and involves data across columns and tables.

- Are redundant values correct? Where Hours Worked, Pay Rate, and Pay Amount are stored, is Pay Amount calculated correctly?
- Are the relationships across columns correct? For example, in healthcare data is a row where the Procedure is "Infant Immunization" is the Patient's Age appropriate?
- Does the data comply with time sequencing rules? For example, should an order be packed before it is shipped?

Hundreds of business rules might exist and will be progressively uncovered as testing progresses. Definition and testing of these rules must be prioritized according to their positive (or negative!) impact to the business.

Assessment Goals

The evaluation team and business sponsors will define and prioritize enhancements to systems, processes, and training with the goal of improving the quality of the data. Once enhancements are in

24 DUGGAN ROAD
ACTON, MA 01720
978.790.3457 TEL
978.635.0457 FAX
WWW.METAVIEW360.COM

place, the subsequent improvement in data quality can be tracked. The *value* of the improvements can be quantified and reported.

Data Warehouse Initiatives

Data of dubious and varying quality will be sourced from multiple systems, sometimes from outside the company, and brought into a single representation of the "corporate truth" in a data warehouse. Early assessment of the capability of that source data to support warehouse requirements ensures that the data warehouse plan includes the tasks, time, and cost to deal with idiosyncrasies that are found in the source system data. The project team and its business sponsors are able make informed trade-off decisions about scope vs. cost and time. Removing nasty surprises can eliminate hidden costs.

Monitoring the Data in the Warehouse

The source data assessment provides a target level of completeness, validity, and compliance to business rules for the data in the warehouse. When the warehouse is loaded for the first time, the same analysis techniques can be used to measure how well the data meets those targets. Those measures become the baseline for monitoring the health of data in the warehouse. Trends in the quality metrics will tell the support team whether improvement efforts in the source systems are working. Trends may also detect changes in the source systems that are having a negative impact on the warehouse. Metrics may include:

- Structural integrity metrics, such as referential integrity or uniqueness of primary keys
- Analysis of "rows in" and "rows out" to verify success of the load

Meta data, such as data definitions, calculation, sources, timeliness, and data quality can now be provided to the users of the warehouse. A few key metrics, defined by the business end users, will provide at-a-glance understanding of the condition of the data. Metrics could include:

- Completeness and validity of key business elements, such as Sales Districts Reporting
- Compliance to key business rules, such as total exposure for a customer compared to credit limit

The actual value of a metric is not as relevant as its relationship to the last reading and to the baseline from the initial warehouse load. With a quick comparison, the business end user will know whether or not the data is at an acceptable quality level for decision making.

Critical Success Factors

As with all disciplines, the approach to Data Quality evaluation requires rigor. A methodology to guide assessment is critical.

Tools are available to perform data quality analysis tasks. They are useful when implemented in support of a defined methodology. Types include:

- Data Profiling Tools: use pattern matching and inference engines to uncover data irregularities
- Quality Analysis Tools: analysis of domain values and relational rules in data structures, ability to define and test business rules
- Name & Address/Householding Tools: specialized data profiling and analysis to consolidate and differentiate company as well as individual name and address data

The cost of a Data Quality evaluation program is really an investment in management of a critical asset of the corporation. Dollars spent on staffing, tools, and infrastructure yield a significant return on that investment.

About Metaview360

Metaview360 consults in the areas of data integration, data governance, data warehousing, meta data management, and information systems architecture. Metaview360 delivers solutions that positively affect your bottom line. Please visit us at www.metaview360.com

METAVIEW360
practical meta data solutions
24 DUGGAN ROAD
ACTON, MA 01720
978.790.3457 TEL
978.635.0457 FAX
WWW.METAVIEW360.COM