Data Integration Design: Event Management
Capturing, Correcting & Preventing Data Defects

Is your organization spending large amounts of money, time and human resources implementing data integration processes? Are the customers of the new data structures satisfied with the information they are getting? Is your organization able to easily identify and correct the information defects that impact your customers?

Most Internet, ERP and data warehouse implementations experience significant information defects that negatively, and many times acutely, impact their intended customers. Efforts to detect and correct these defects frequently require maintaining or enhancing the data integration processes that move the data from its sources to the new data structure. However, defect correction in data integration processes is frequently a trial-and-error, time-consuming, and error-prone process. Some data integration processes are so complex and difficult to maintain that many organizations resort to "after the fact" data correction techniques, adding to the overall complexity of the environment.

However, there is an approach that can help: *Event Management*. This is an approach to manage and control data integration and similar processes; it provides information on data defects to make these processes reliable, maintainable and trustworthy. It makes defect detection and correction not only possible but highly effective. It enables the organization to manage and control the most significant information defects, their impact, frequency and location within the data integration process.

In order to understand the value proposition of Event Management, it is important to review its results and the response from business and IT executives and staff once it is in operation. To illustrate, here is a case study:

One of my clients integrated information about their operations in an Operational Data Store (ODS) to support operational analytics and decision making and to supply this information to its Enterprise Data Warehouse (EDW). This ODS was implemented several years before; however, the business areas experienced constant information quality issues such as production reports overstated with duplicate production or understated with missing production, "ghost producing units" (units that do not exist and that are not in the source systems) classified as "producing," and misclassified producing units (e.g., classified as "sold" or "plugged and abandoned" when in fact are "producing"). The myriad of issues caused confusion, rework, and more importantly, doubt. The IT and business areas conducted frequent, labor intensive data corrections on the ODS but data quality issues keep popping up. As a result, the business areas did not trust the ODS and it was largely underutilized.

After considering many alternatives, including the redesign and rewrite of all the ETL processes, or even abandoning the ODS, executive management, at the suggestion of our team, agreed to implement processes to monitor the information quality in the ODS using event management to shed some light into the problem.

Using the data quality tool already available at the client's site, the team implemented data quality monitoring processes to control the quality of critical information in the ODS; we called them "data controls." A data control process is an application that executes on a schedule and conducts one or more validations of the data in the ODS; for instance, a control may verify that the information is consistent with its source (e.g., detecting "ghost" records) or that it meets a critical business rule (e.g., are all the production details consistent with each other?).

Once the data controls were in production they provided a complete view of the information quality in the ODS: a perspective that informed and surprised business and IT personnel alike. During the initial execution of these controls the team obtained the following results:

Data Control Process	Controls	Events Logged
Completion	42	35,590
Completion Test	32	211,895
Monthly Production	20	85,888
Daily Production	45	1,022,030
Well	31	76,172
Producing Entity	3	24,115
TOTAL	173	1,455,690

The team first validated that the controls were working as expected. Then, they set out to work with the business and IT areas to determine the best action path for each set of defects. The first step was to apply the Pareto principle. Based on business impact and frequency of incidence, they identified the most critical defects to address: the identity issues or "ghost" properties.
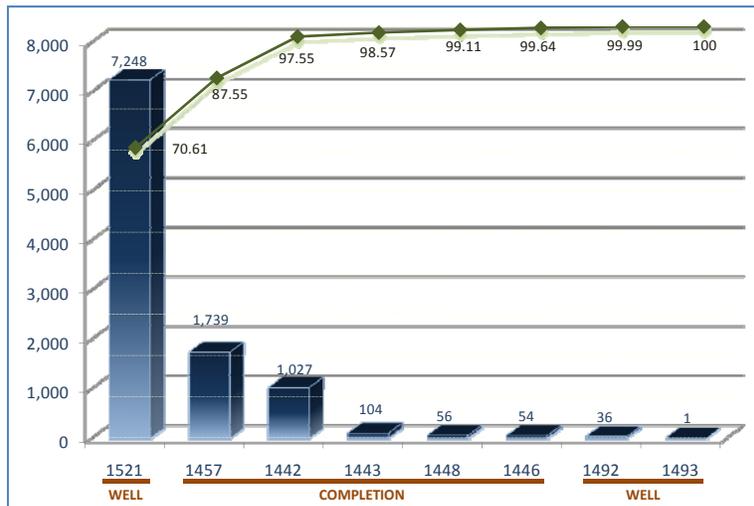

Figure 1 - Most Critical Information Defects

The list included five completion issue types and three well issue types. The issues were not new, what was new is that now they knew how many they were and the actual records that were defective in the ODS. The Pareto diagram allowed them to focus on the highest impact and work their way down the list. The business areas already had methods and procedures for correcting these conditions in the source systems and the ODS. They

knew by experience that correcting an issue immediately after it occurred takes a few hours; correcting the same issue days or months after the ghost record is created, once production and test records are posted to the incorrect property, could take days or even weeks!

Also new was the fact that as new issues occurred (due to an insert, update or delete into the ODS), the proper business and IT personnel were immediately informed. This is done via an Event Management notification function.
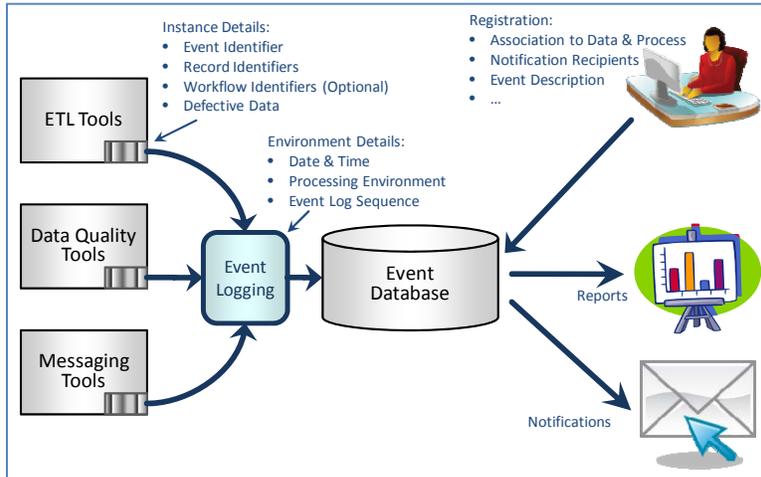

Figure 2 - Event Management Process

When an event is logged, the event management process issues an e-mail notification to each person registered for the event; regardless of the numbers and types of events, the addressee receives one e-mail with a hotlink to a report that then provides the details on all the logged events.

Figure 3 - Information Defects Notification

This early detection and notification process enabled the business areas to correct the "ghost" records as soon as they appeared, reducing their data correction efforts significantly.

However, the real value of implementing event management is to enable process improvement to prevent the occurrence of information quality defects. The event database contains a wealth of information on data defects, and because the data controls are continuously monitoring the data, they offer immediate feedback on the implementation of process improvements by showing a "before" and "after" picture of the data under control.

As an illustration, the team set out to identify and remove the major cause for event "1457: The Completion must NOT be present in the ODS based on its classification." They learned that the ETL was not handling properly the classification used to determine if a property should be in the ODS in the main source system. The source system had some values in the schema that were not accounted for in the ETL; also, the conversion of these values was not known by the business areas, creating even more confusion.

Once the ETL was corrected and fixed, the volume of invalid properties fell sharply from an average of 1,350 per day to an average of less than 100 per day. Additional analysis was required to identify the next cause; but the team decided to focus on the next high impact issue. The event management report for this event is as follows:
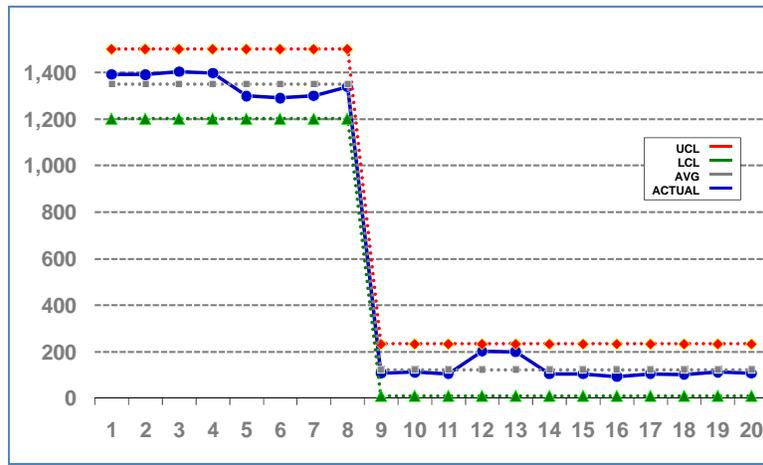

Figure 4 - Event "1457: The Completion must NOT be present in the ODS based on its classification"

The team has implemented a continuous process improvement around its ODS and all its operations information and continues to add new data controls where appropriate to monitor, correct and prevent information defects.

Conclusion

Managing the quality of information in a complex data integration environment is a challenging task; however, ***Event Management*** can enable the management and control of the quality of information replacing the classical trial-and-error, time-consuming, and error-prone process. In this case study, the business areas now trust, use, and depend on the information provided by the ODS. Furthermore, the assurance that when something goes wrong it will be noticed immediately and that the group will come together to address the problem has created a positive working atmosphere.

I hope that you find this article useful. What do you think? Let me know at andres.perez@irm-consulting.com.

Andres Perez is an Information Management Consultant, President of IRM Consulting, Ltd. Co., based in San Antonio, Texas, and the VP of Operations for DAMA International. Mr. Perez specializes in information resource management, information architecture and information quality management. He may be contacted via phone at: +1 210-413-1481 or via email at: andres.perez@irm-consulting.com.