

: How can organizations keep data-driven DSS and data warehouses current with changes in source systems?

by Daniel J. Power

Editor, DSSResources.COM

In my Ask Dan column of October 22, 2006 (Vol. 7, No. 22), I answered some of the challenging questions that Aileen MacKay had sent me. In this Ask Dan!, I'll focus only on her system change questions. She asks "What is the process companies are using to keep their Data Warehouse up-to-date? Once the source system is changed, what is the process for moving these changes into the Data Warehouse and DSS? Once changes are made to the source system, how long does it usually take to make the changes in the data warehouse and DSS? How long should this process take if it is run well? Do you have examples of state-of-the-art-processes that work really well where managers are willing to share their process from the source system to DW to DSS? Specific to the health care insurance industry or medicaid? Thank you."

In most data-driven decision support environments, multiple internal and external computerized systems provide source data that must be extracted, transformed and then loaded (ETL) into a data mart/warehouse (DSS data store). Routinely data is moved from source systems to a decision support data store. This process can be and usually is automated and routinized. But as businesses and organizations change processes, products and services the data models for operational systems change and because of changes in these source systems, the design of data warehouses and data marts used by data-driven DSS must also change. A design challenge occurs whenever it becomes necessary to make any changes to the data model in the source system. The data model defines what is stored in the source system, how it is stored, and how items in the database are related to each other. This source system data model impacts the data model in the decision support data store.

Let's quickly look at some of the changes that can occur in operational system data models. Products or services may be added, product codes may change, fields may be added or removed, data types may change, and data values may be altered. The process of data transformation can deal with some of these changes, like simple changes in product identifiers. As more significant changes occur in source systems more changes must be made in the DSS data store. New fields or tables may be needed. In some cases the DSS data store may

: How can organizations keep data-driven DSS and data warehouses current with changes in source systems?

need to be transformed, manipulated or even radically redesigned.

Waymire and Sawtell (2006) note a data transformation can be one of a number of "operations applied to your data before it is stored in the destination. This can include the calculation of new values, the concatenation of values, breaking of values like a name field into a first name and last name field, as well as any type of data cleaning you want to enforce." In general, data transformation involves processes for selecting, targeting, converting, and mapping data so that it may be used in a DSS data store. Data may be transformed using predefined rules or specific formulas or techniques.

When change is contemplated in an operational source system, analysts also need to examine the ramifications for systems using the data from the operational system. Part of the analysis and design should involve mapping the data elements between the data models. IDEALLY implementing changes in a source system should coincide with implementing changes in the DSS data store and the data-driven DSS. A specific data-driven DSS may need to be changed to display new data or query the revised data model.

Synchronizing changes between operational and external source systems and DSS data stores insures the quality of the data for decision support. Examining the mapping between the original and the new data model helps insure the DSS data is consistent, timely, well understood, and satisfies the needs of managers. The danger is that the two models will not be consistent and that users will be dissatisfied with the validity of the DSS data and the information derived from that data.

In general, we must deal with both simple and complex transformations. In some instances we can consolidate data. For example, if we have two similar products and we eliminate one and create a new shopkeeping unit (SKU) for the other, we may want to consolidate the historical sales data for the two products. A more difficult transformation is required if we further differentiate a product or service. For example, we may have one academic major in a department and we replace it with two more differentiated majors with new codes. In this case major #23 may become majors #230 and #231. In this situation, we may allocate historical enrollment data between

: How can organizations keep data-driven DSS and data warehouses current with changes in source systems?

the two new majors. We may use an arbitrary allocation (fixed %) or a more complex rule-based allocation. The easiest transformation involves an addition or a replacement. For example, adding a new SKU as a data value or substituting a new SKU for an old SKU. Time lags in adjusting the operational data model and the DSS data store can create many unnecessary problems for managers and IT staff in reconciling results from diverse systems.

Ralph Kimball identifies some "graceful modifications to existing fact and dimension tables". These include: new dimensional attributes, new types of measured facts, new dimensions, a dimension becomes specific or granular, an explicit hierarchy is added relating two dimensions, new dimensions are added from a new data source. Sadly many changes that result from changes in source systems are complex, messy and even awkward to make. In some situations, it seems easier to "start over" and not try to bring historical data forward to the revised data model and resulting DSS data store. This "blank slate" approach should be a last resort decision.

Part of the administration of a DW/DSS data store should involve governance of the data model and those staff members who are maintaining the integrity of the decision support system should be monitoring potential changes in source systems. Also, a process for change management should be established as part of the ongoing decision support/BI administration.

Data-driven DSS fail when source data is missing, inconsistent, or erroneous. Current, accurate data models are essential. Without them, data-driven decisions cannot be made with any confidence. At TDWI (tdwi.org), a research article on data transformation (1999) suggests both the magnitude of the problem and one technology solution: "as we embrace a greater number of sources and targets, we centralize all data transformation for the entire enterprise so that, effectively, the tool becomes the cornerstone of the corporate decision-making apparatus. This means a standard data transformation function, like data tracing, can be applied across the entire enterprise. This means aggregations and data analysis apply to all data, not just what is stored in a recognizable database. In all cases, data transformation has one goal: to guarantee accurate, clean information throughout the enterprise. To prevent garbage in, garbage out."

: How can organizations keep data-driven DSS and data warehouses current with changes in source systems?

The data warehouse/data-driven DSS team must continually monitor what data should go into the data store and where those particular pieces of information can be found. Some of the data will be internal to an organization and hence change is under management control. Source data extraction, selection, and transformation processes must be continually evaluated. According to a data warehousing introduction in the Minnesota State Archives, "Source data analysis and the efficient and accurate movement of source data into the warehouse environment are critical to the success of a data warehouse project. Good metadata is essential to the effective operation of a data warehouse and it is used in data acquisition/collection, data transformation, and data access. Acquisition metadata maps the translation of information from the operational system to the analytical system. This includes an extract history describing data origins, updates, algorithms used to summarize data, and frequency of extractions from operational systems. Transformation metadata includes a history of data transformations, changes in names, and other physical characteristics." Part of any change process is updating, revising and expanding the metadata for the DSS data store. The metadata helps users understand how the change in a source system impacted the historical record that is stored in the DSS data store.

Let's finish Aileen's questions. How long should this process take if it is run well? In general, updating the DSS data store should be less time consuming than making changes in source systems. The length of the change process is a function of the magnitude and complexity of the change in the source system. At some point the change may become so great that the historical record is no longer relevant to the "new" data model. Do you have examples of state-of-the-art-processes that work really well? NO, perhaps some readers or vendors will provide a case study for DSSResources.com. For additional information, I suggest reading some of the other Ask Dan questions on data warehouses and data-driven DSS and then check and search the following web sites:

1. Business Intelligence Network
<http://www.b-eye-network.com/home/>

B-eye-network delivers industry-based content hosted by domain experts and industry leaders. The Business Intelligence Network

Page 4/8

(c) 2022 Daniel J. Power, Power Enterprises <power@dssresources.com>

URL: <http://dssresources.com/faq/index.php?action=artikel&cat=&id=132&artlang=en>

: How can organizations keep data-driven DSS and data warehouses current with changes in source systems?

includes horizontal technology coverage from the most respected thought leaders in Business Intelligence, Business Performance Management, Data Warehousing and Data Quality.

2. Data Warehousing Information Center <http://www.dwinfocenter.org/>

This site aims to help visitors learn about data warehousing and decision support systems by providing links to vendors of tools for end-user data retrieval and analysis, tools for building and maintaining system infrastructure, and tools directed at specific industries and functions.

3. The Data Warehousing Institute <http://www.tdwi.org>

The Data Warehousing Institute (TDWI) is a member-based organization whose goal is to educate decision-makers and information professionals on data warehousing strategies and technologies. This site serves as a clearinghouse for case studies, white papers, and data warehousing events and conferences worldwide. Membership includes a subscription to the quarterly Business Intelligence Journal.

4. DM Review <http://www.dmreview.com/>

The DM Review web site offers content from DM Review magazine as well as online-only articles, resource lists, discussion forums, issue archives, etc. DM Review focuses on data warehousing and related topics such as metadata management.

5. Inmon Associates, Inc. <http://www.inmongif.com>

: How can organizations keep data-driven DSS and data warehouses current with changes in source systems?

This is the official web site produced by Bill Inmon, one of the leaders in data warehousing. This web site focuses on Inmon's framework for government information systems, the Government Information Factory (GIF).

6. Ralph Kimball Associates
<http://www.rkimball.com/>

Ralph Kimball is a leading proponent of the dimensional approach to designing large data warehouses. He teaches data warehousing design skills and helps selected clients with specific data warehouse designs.

As always your comments, suggestions, examples, etc. are appreciated.

References

Kimball, R., "Design Tip #29 Graceful Modifications To Existing Fact and Dimension Tables," 2001,
URL <http://www.rkimball.com/html/designtipsPDF/DesignTips2001/KimballDT29Graceful.pdf> .

Minnesota State Archives, "Data Warehouse Description," URL
<http://www.mnhs.org/preserve/records/dwintro.html>.

Power, D., Does data modeling differ for a Data-Driven DSS? DSS News, Vol. 2, No. 17, August 12, 2001.

Power, D., Is a Data Warehouse a DSS? What is a star schema? How does a snowflake schema differ from a star schema? DSS News, Vol. 3, No. 4,
Page 6/8

: How can organizations keep data-driven DSS and data warehouses current with changes in source systems?

February 17, 2002.

Power, D., What is ETL software and how is it related to DSS? DSS News, Vol. 3, No. 16, August 4, 2002.

Power, D., What do I need to know about Data Warehousing/OLAP? DSS News, Vol. 4, No. 5, March 02, 2003.

Power, D.J., "What is the process for designing decision support software?" DSS News, Vol. 7, No. 17, August 13, 2006.

Power, D.J., "What staffing is needed for DW/BI/DSS?" DSS News, Vol. 7, No. 22, October 22, 2006.

Schwartz, J., "Complex Data Transformation," Line 56, Tuesday, March 14, 2006, URL <http://www.line56.com/articles/default.asp?ArticleID=7412&pg=3&topicID=4> .

TDWI, "Beyond the Data Mart: A New Look at Data Transformation," What Works: Volume 8, November 1999, URL <http://www.tdwi.org/research/display.aspx?ID=5385> .

Waymire, Richard and Rick Sawtell, Sams Teach Yourself Microsoft SQL Server 7 in 21 Days, 2006, URL <http://www.sampublishing.com> .

Citation: Power, D., "How can organizations keep data-driven DSS and data warehouses current with changes in source systems?" DSS News, Vol. 7, No. 26, December 17, 2006.

: How can organizations keep data-driven DSS and data warehouses current with changes in source systems?

Author: Daniel Power

Last update: 2007-03-02 12:02