: What are the basics of data warehousing?

by Daniel J. Power

Editor, DSSResources.COM

For at least one hundred years, warehousing has referred to the short- or long-term storage of items in a specially designed facility. Originally this general definition described the storage of inventory and other physical items. In the early 1990s, Bill Inmon advocated creating specialized data warehouses for decision support applications. The term data warehousing refers to the process of creating and maintaining a data warehouse.

A data warehouse is a database designed to support decision making in organizations. It is updated in batches or in real time, and it is structured for rapid online queries and for providing managerial summaries. Data warehouses contain large amounts of historical data. According to Inmon, a data warehouse is a subject-oriented, integrated, time-variant, nonvolatile collection of data in support of management's decision-making process. Ralph Kimball defines a data warehouse as "a copy of transaction data specifically structured for query and analysis (p. 310)."

The data warehousing process has changed in the past 10 years. Builders have fewer concerns related to data storage capacity and processing speed, but the task of creating a data warehouse remains difficult.

In 1997, Anahory and Murray wrote about data warehousing in the real world. They positioned their book as "a practical guide for building decision support systems." They define a data warehouse "in its simplest perception ... as no more than a collection of the key pieces of information used to manage and direct the business for the most profitable outcome (p. 4)." More technically, they define a data warehouse as the data and the "processes involved in getting that data from source to table, and in getting the data from table to analysts (p. 4)." Let's review the process they prescribed.

The process for delivering an enterprise data warehouse is "a variant of the joint application development approach .. the entire delivery process is staged in order to minimize risk (p. 9)."

First, understand the business case for investment. Identify the projected business benefits from using the data warehouse.

Second, experiment with the concept of data analysis and learn about the value of a data warehouse.

Third, specify the business requirements.

Fourth, develop an overall system architecture.

Fifth, quickly load some data to produce an initial production deliverable that satisfies the "most pressing business requirement for data analysis (p. 12)."

Sixth, finish loading required historical data into the data warehouse.

Seventh, "configure an ad hoc query tool to operate against the data warehouse (p. 13)."

Eighth, automate operational data management processes like extracting and loading new data, backing up data, and generating data aggregations.

Ninth, if there are additional business requirements, extend the scope of the data warehouse.

Tenth, monitor business requirements. During the life of a data warehouse "business requirements will constantly change (p. 14)."

Some of the potential benefits of putting data into a data warehouse include:

- Improving turnaround time for data access and reporting;

- Standardizing data across the organization so there will be one view of the "truth;"

- Merging data from various source systems to create a more comprehensive information source;

- Lowering costs to create and distribute information and reports;

- Sharing data and allowing others to access and analyze the data;

- Encouraging and improving fact-based decision making.

The major limitations associated with data warehousing are related to user expectations, lack of data and poor data quality. Building a data warehouse creates some unrealistic expectations that need to be managed. A data warehouse doesn't meet all decision support needs. If needed data is not currently collected, transaction systems need to be altered to collect the data. If data quality is a problem, the problem should be corrected in the source system before the data warehouse is built. Software can provide only limited support for cleaning and transforming data. Missing and inaccurate data cannot be "fixed" using software. Historical data can be collected manually, coded and "fixed," but at some point source systems need to provide quality data that can be loaded into the data warehouse without manual clerical intervention.

Data warehousing tasks and deliverables have changed in terms of the technical tools used, the risks and concerns and the time needed to complete some tasks, but the above process is still a good starting point for planning an enterprise data warehouse for an organization.

According to Westerman (2001), "To understand what is needed for your data warehouse, you have to speak with the business people. This is not an option; it is a requirement (p. 61)."

The easiest way to get started with data warehousing is to analyze some existing transaction processing systems and see what type of historical trends and comparisons might be interesting to examine to support decision making. See if there is a "real" user need for integrating the data. If there is, then IS/IT staff can develop a data model for a new schema, load it with some current data and start creating a decision support data store using a database management system (DBMS). Find some software for query and reporting and build a decision support interface that's easy to use.

: What are the basics of data warehousing?

Although the initial data warehouse/data-driven DSS may seem to meet only limited needs, it is a "first step." Start small and build more sophisticated systems based upon experience and successes.

References

Anahory, S. and D. Murray. Data Warehousing in the Real World: A practical guide for building decision support systems.

Inmon, W. H. Building the Data Warehouse.

Kimball, R. The Data Warehouse Toolkit: Practical Techniques for Building Dimensional Data Warehouses.

Power, D., "What are the advantages and disadvantages of Data Warehouses?" DSS News, Vol. 1, No. 7, July 31, 2000.

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

Westerman, P. Data Warehousing: Using the Wal-Mart Model.

About Dan Power

Daniel J. "Dan" Power is a Professor of Information Systems and Management at the College of Business Administration at the University of Northern Iowa and the editor of DSSResources.com, the Web-based knowledge repository about computerized systems that support decision making; the editor of PlanningSkills.com; and the editor of DSS News, a bi-weekly e-newsletter. Dr. Power's research interests include the design and development of decision support systems and how these systems impact individual and organizational decision behavior.

: What are the basics of data warehousing?

ORIGINALLY PUBLISHED MARCH 18, 2010 at http://www.b-eye-network.com/view/12929

Author: Daniel Power
Last update: 2014-04-24 08:15