

: *What is data warehousing?*

by Dan Power

Editor, DSSResources.com

For at least a 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 finished goods 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 remains conceptually 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.

: *What is data warehousing?*

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)."

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)."

: *What is data warehousing?*

References

Anahory, S. and D. Murray. Data Warehousing in the Real World: A practical guide for building decision support systems. Reading, MA: Addison-WQesley, 1997.

Inmon, W. H., "What is a Data Warehouse" at http://www.cait.wustl.edu/cait/papers/prism/vol1_no1/

Kimball, R. The Data Warehouse Toolkit: Practical Techniques for Building Dimensional Data Warehouses. 1996.

Power, D., "What are the advantages and disadvantages of Data Warehouses?" DSS News, Vol. 1, No. 7, July 31, 2000. <http://dssresources.com/faq/index.php?action=artikel&id=180>

Power, D., What do I need to know about Data Warehousing/OLAP? DSS News, Vol. 4, No. 5, March 02, 2003. <http://dssresources.com/faq/index.php?action=artikel&id=47>

Westerman, P. Data Warehousing: Using the Wal-Mart model. San Diego: Academic Press, 2001.

Author: Daniel Power

Last update: 2010-02-18 10:39