

While people are often a corporation's true intellectual property, data is what feeds the people, enabling employees to see where the company stands and where it will go. Quick access to quality data helps employees excel. Delays in obtaining quality data could result in information bottlenecks and loss of competitive advantage. According to Ferdinandi, "the purpose of a data warehouse is to store clear, concise, and consistent data in a format that can be easily and quickly accessed by a business". What defines a data warehouse from other forms of data storage is the ability to use it specifically for analytical analysis. Any business that requires complex analysis of a significant amount of data to remain competitive needs a data warehouse, (Ferdinandi 8).

Chen, Nemati, and Barko recount DW history, detailing the evolution of solutions to the problems with business processes and applications of the 1980s. The early 1990s were spent developing relational database management system (RDBMS) technology, the structured query language (SQL), and improved visual presentation of information. The improvements of the 1990s allowed users to view historical in addition to current data and more cohesively integrate data to look at the entire corporate picture. Online analytical processing (OLAP) tools developed in the mid 1990s from corporate in-house system integration and product improvements as a bridge between the warehouse storage and presentation layer, (Chen 14).

Today, DW is centered on three knowledge creation activities: sense making, knowledge making, and decision-making, (Nemati and Barko 4). Sense making is the ability of corporate employees to interpret and understand information about the environment and events happening both inside and outside the organization. The sense making process is what empowers employees to make sense of what is going on in the

external environment. Knowledge making is the ability to create new knowledge by combining the expertise of the company employees with historical and current data to learn and innovate. The learning and innovation can often lead to enhancing current products, improve organizational processes, and even create new products to meet new and changing demands by customers. Decision-making is the ability to process and analyze information and knowledge in order to select and implement the appropriate course of action.

Data warehousing, hereinafter DW, is often paired with database management systems (DBMS) and artificial intelligence (AI), (Chen 3). The DBMS model depends on the dimensional model, as opposed to an entity-relationship model. The dimensional model has a primary table which contains numeric and additive measurements of the business, (Chen 14).

Ferdinandi defines a data warehouse as a pool of data organized in a format that enables users to interpret data and convert it to information to gain knowledge from this interpretation.

### **Types and Applications of Data Warehouses**

As we have seen, the application of a data warehouse is virtually unlimited and can be used in every area of business, organizations, and government agencies. How data warehouses are applied will depend on the focus of the enterprise. Those organizations that focus on customer relationship management will utilize their capabilities to assist them in making better and more precise decisions about their customers. Data warehouses are primarily used in transaction environments to record historical data. They can also be used in such departments as human resources to keep lengthy historical

records for as long as 20 years or more. These data warehouses will be smaller than sales and order warehouses because very few transactions occur.

Data warehouses come in a variety of types that can be categorized according to processing methods. There are traditional data warehouses that hold all of the historical transactional data in its entirety for an organization and then there are more specialized data warehouses that are used for specific applications. There are also data marts and mini marts. Data marts are smaller versions of a data warehouse and will normally focus on a particular subject or department. Some are subsets of larger data warehouses and are referred to as dependent data marts. The overall goal of a data mart is to meet the specific needs of a group of knowledge workers. Mini marts are a small slice of a data warehouse that is very focused and used by a small number of users. It will normally contain much less data than a data mart but will provide substantial granularity for the targeted group of knowledge workers.

There are 3 major types of data warehouse processing methods including ROLAP (Relational Online Analytical Processing), MOLAP (Multidimensional OLAP), and HOLAP (Hybrid OLAP). Before providing detail on each type a definition of OLAP may be needed. OLAP is used to provide a means of quickly providing answers to complex database queries. Relational databases, if normalized properly, function well for operational databases. However, for more complex multidimensional queries it is significantly slower. OLAP (Online Analytical Processing) takes a snapshot of a relational database and puts it into dimensional data for analytical processing. It is estimated that for complex queries OLAP can provide an answer in around 0.1% of the time for the same query on a relational database.

ROLAP is used in a data warehouse that is stored in a relational database rather than in a multidimensional database. In this type of environment the processing of the data may take place either within the database system, a mid-tier server, or the client. In a two-tiered system the client sends the request as a SQL query to the database and receives the response from the database system. In a three-tiered system the client sends a multi-dimensional analysis request and the ROLAP engine converts the request into SQL which is then submitted to the database system for processing and then sends the data back to the client in reverse. The ROLAP engine is used to connect the client to the database engine by converting traditional multi-dimensional analysis requests to a relational database and from a relational database to the client in multi-dimensional form. ROLAP was found to be more useful in environments like marketing and sales due to its dynamic access capabilities.

The next type of system is MOLAP which is a multidimensional system. It is sometimes just referred to as OLAP since it is the traditional form. A key part of MOLAP is aggregation and it works well with file structures such as multidimensional arrays where the data has uniform distribution. It does not work well with data that has skewed distributions. In a multidimensional file system the size of the “page” regions vary depending upon the density of the data which can slow down aggregation considerably since more dense regions will need to be read multiple times. Multidimensional databases are used to ask questions that are more difficult to answer using SQL or slower. This system provides the ability to rapidly process the data in the database so that the answers to specified questions can be answered more quickly.

MOLAP was determined to be more useful in financial applications due to its high performance and analytical power.

Lastly, there is the Hybrid OLAP system. When ROLAP and MOLAP first hit the market MIS personnel had to determine which method of analytical processing would work best for their environment. Since both ROLAP and MOLAP support multidimensional views of data and provide the ability to get specific information from these queries. There are significant differences between the two and when a company built its decision support system it had to live with the speed of the MOLAP or ROLAP's dynamic access to highly detailed and specific data. HOLAP provides a system that combines the best of both MOLAP and ROLAP and hybrid of the two systems. The system purports to combine the dynamic access to relational data with the more sophisticated analytical capabilities and instant access to aggregated data. However, these new systems were varied in structure even though they are some combination of the two preceding technologies. For a system to truly meet the requirements of a HOLAP system it must provide for the following: Dimensions can be dynamically updated, multidimensional view based on RDBMS metadata, fast access at all levels of aggregation, easy aggregate maintenance, and compact aggregate storage. If a system does not meet these criteria then it is not a true fully developed HOLAP system.

Works Cited

Adelman, Sid, et al. Impossible Data Warehouse Situations: Solutions from the Experts.

Boston: Addison-Wesley, 2003.

Chen, Zhengxin. Intelligent Data Warehousing: From Data Preparation to Data Mining.

Boca Raton: CRC Press, 2002.

“Data mart.” Webopedia. 18 Nov. 2004

<[http://www.webopedia.com/TERM/D/data\\_mart.html](http://www.webopedia.com/TERM/D/data_mart.html)>.

Drozdenko, Ronald G., and Perry D. Drake. Optimal Database Marketing. London: Sage

Publications, 2002.

Ferdinandi, Patricia L. Data Warehousing Advice for Managers. New York: AMA

Association, 1999.

“Hybrid Online Analytical Processing.” DMReview. 18 Nov. 2004

<<http://www.dmreview.com/>>.

Inmon, Bill. “Information Management: Charting the Course: Different Data Warehouse

Types.” DM Review Magazine, June 2000.

Mento, Barbara, and Brendan Rapple. “Data Mining and Data Warehousing.” SPEC Kit

274 July 2003.

“Multidimensional OLAP.” Association for Computing Machinery. 18 Nov. 2004

<<http://portal.acm.org/>>.

Nemati, Hamid R., and Christopher D. Barko. Organizational Data Mining: Leveraging

Enterprise Data Resources for Optimal Performance. Hershey: Idea Group

Publishing, 2004.

“OLAP.” Wikipedia. 18 Nov. 2004 <<http://www.wikipedia.com/>>.

Padilla, Ramon. "Beginning and planning your data-warehousing project." TechRepublic.

25 Feb 2004 <<http://techrepublic.com.com/5100-6329-5155894.html>>.

"ROLAP Definition." Techtarget. 18 Nov. 2004

<<http://www.searchdatabase.techtarget.com/>>.