

# Data Warehousing and Knowledge Discovery Technology

## Data Utilization for Dollars and Sense

John Thomstatter, MKS Inc.

Rapid changes in information technology have dramatically increased our ability to generate, collect, and store data. The widespread use of corporate business software and data collection hardware, such as scanners and bar code equipment, continues to increase our ability to accumulate huge amounts of data. Increased capacity and reduced cost of storage devices further encourages the retention and storage of this data. Explosive growth in the collection and storage of data has led to the expansion of other facets of the information industry--Knowledge Discovery in Databases (KDD) and Data Warehousing. The intent of KDD, or Data Mining, another commonly used term, is to extract knowledge from historical data for strategic analysis and business planning.

Data Warehouse is the term used to describe "a copy of transaction data specifically structured for query and analysis." (Kimball, 1996) An article in the July issue of the DIGITAL Today newsletter estimates the Data Warehousing market at \$8B annually with a 47% annual growth rate. ("Targeting" 1997) This creates another opportunity for technical communicators to document the features, capabilities and use of the technology. Data Warehousing and KDD, like artificial intelligence (AI) technology, are not "plug and play" software solutions.

The elements of the knowledge discovery process include: the online transaction data which is restructured into a Data Warehouse format; a Decision Support Software (DSS) tool which is used to define the data access and rules for analysis, as well as extracting and interpolating the data; and a PC spreadsheet or similar software package which receives the DSS results and further manipulates or formats the data into a decision statement.

### Getting Started

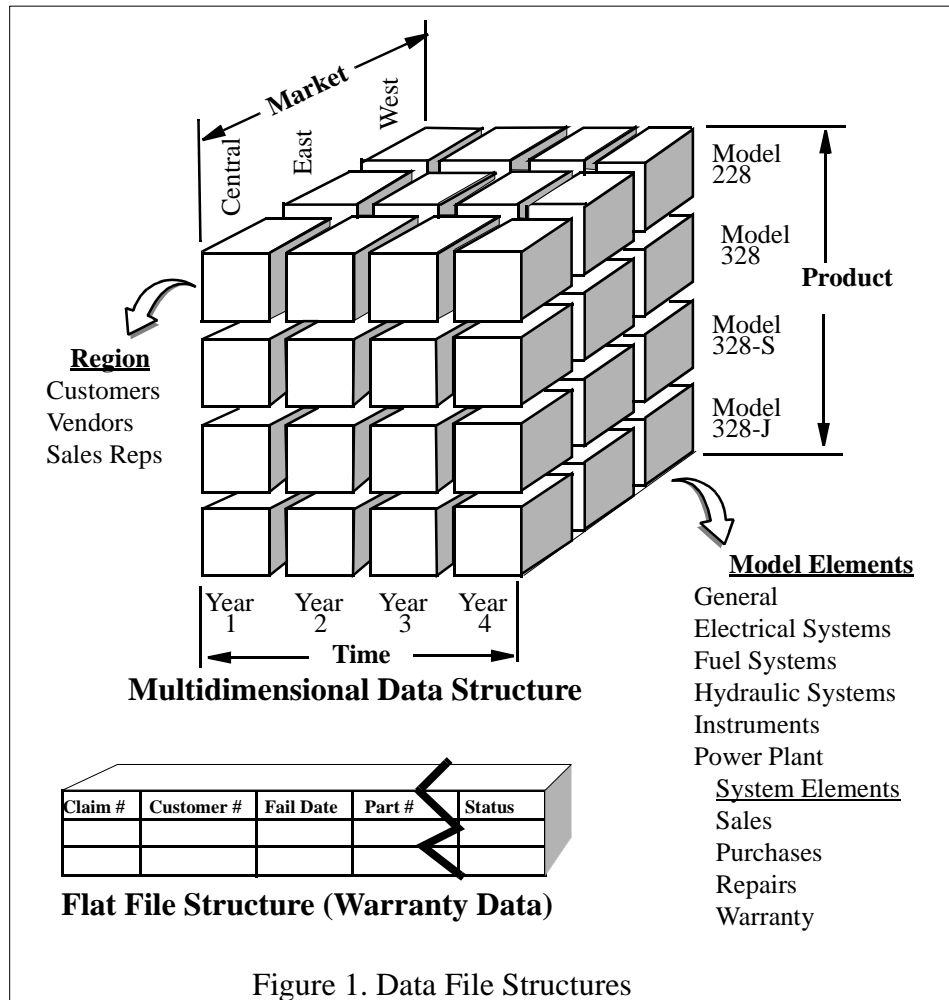
About a year ago the company I am working for released a Decision Support Software tool to analyze historical business data. My initial task was to gain an understanding of the new product, the related technologies, and the growing number of "helper acronyms." Product vendors continue to flood the arena with terms and catchy phrases in an attempt differentiate their product and capture a share of the market.

### Building a Data Warehouse

Creating a Data Warehouse involves more than just moving copies of your existing data files to a large central storage disk. Business transaction data are stored in flat files or relational databases. These files are updated by online transaction processing (OLTP) software as transactions take place during the business day. Records are individually accessed, fields are updated, and lists of records are displayed or printed. Data Warehouses are structured for analysis of historical data such as identifying business trends or marketing strategies. Data stored in a Data Warehouse is subject oriented as opposed to the transaction detail nature of OLTP databases. For example, in an OLTP order processing database, the level of detail on file would be down to the individual line items in an order. Users are interested in tracking a business transaction from the start to completion. The user of a Data Warehouse is more concerned with questions such as, "How many transactions were processed for a given customer and a specified time period?" The level of detail stored in your Data Warehouse is dependent on the type of analysis anticipated and is usually defined during the design phase of the warehouse implementation process.

One methodology for structuring a Data Warehouse is explained as a multidimensional format. Figure 1 shows a comparison of a flat file OLTP database for processing aircraft warranty claims and a typical multidimensional Data Warehouse for analyzing the aircraft model data. The information from several OLTP databases are combined and stored in the Data Warehouse. The multidimensional model stores data in conformance with product, market and time boundaries. (Starmen, 1993)

# Data Warehousing and Knowledge Discovery Technology



In this example we can evaluate aircraft data by model, system and component against time periods and market variables. How many component failures are occurring on aircraft with 3,000 to 5,000 flight hours and what systems (electrical, instruments, fuel, etc.) are affected? Which operating regions are experiencing the highest failure rates? The results of these queries could pose other questions. Which component manufacturers are experiencing the highest failure rates?

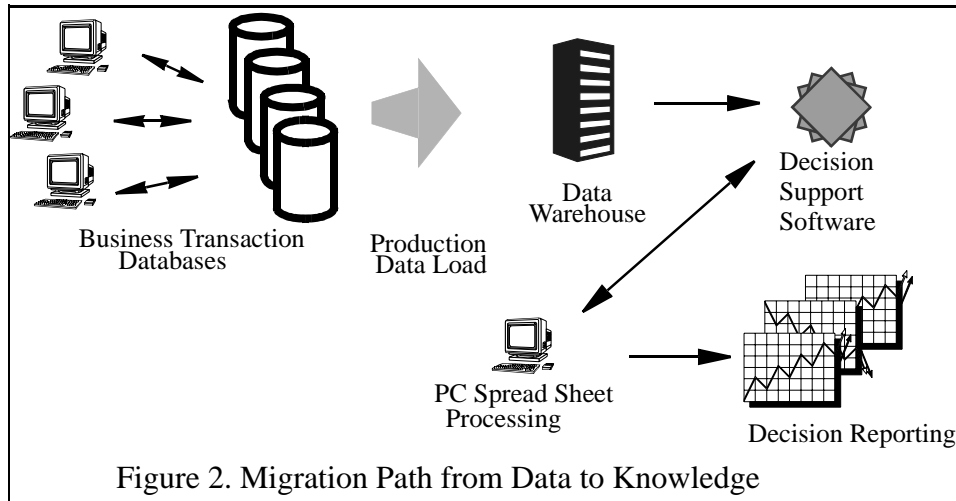
## Updating and Using the Warehouse Data

The information in Data Warehouses is "read access only." Data extracted from the OLTP databases is incorporated into Data Warehouses on a periodic basis by a production data load. The updating is performed during off hours at specified intervals to minimize interruptions to your user groups. The OLTP data is indexed and validated to ensure the new data is consistent with your existing data.

## Decision Support Software

The extracting of data from your warehouse is handled with a Decision Support Software tool using programming commands with financial and statistical algorithms. The DSS software will contain a definition of the Data Warehouse known as metadata or data about the data. The DSS query program uses the metadata to access, collect, and tally the values. The collected data can then be analyzed using the applicable algorithms. You may need to use a "drill down" (Kimball, 1996) process to examine the detail behind a calculation. The resulting values are then moved to your PC spreadsheet for further manipulation or formatting. The end result will be a combination of reports, graphs and or charts. Figure 2 shows how the process works.

## Data Warehousing and Knowledge Discovery Technology



### Solving a Problem

The implementation of a Data Warehouse program usually results from the desire to resolve a business problem. Several years ago an international aerospace company I was working for was faced with paying some stiff customs charges for suspected "bearing dumping." There was an urgent need to disprove the charges because of the significant cost and quantity of bearings used in the manufacture and maintenance of a helicopter. Using a primitive form of the Data Warehouse methodologies discussed in this article, we were able to extract historical international sales data, vendor information, pricing, and discount structures to disprove the dumping allegations. This ultimately resulted in the reduction of the customs fees and recovery of several million dollars being held in escrow. Reduction of these fees also impacted the selling price of the aircraft and spare parts. The deliverable items for this study consisted of several data extract files (on tape), copies of reports, record layouts, and programs used to analyze the data.

### Where to Begin

Today's Data Warehouse solutions require the integration of tools from more than one vendor. (Darling, 1996) If you are planning to tackle the Data Warehouse arena, brush up on the latest acronyms and be sure to keep track of your metadata. Be prepared to roll up your sleeves and "drill down" into the data layers to understand the results of your efforts. Add some hindsight, some foresight, a lot of patience, perseverance, and a dab of artificial intelligence, and you are on your way to Knowledge Discovery!

References:

- Kimball, R., *The Data Warehouse Tool Kit*. New York: John Wiley & Sons, 1996.
- "Targeting, Partners, Call Centers Are Keys to Success for Data Warehousing and ERP," *DIGITAL Today*, Business Partner Edition (July 1997).
- Starmen, Jeffrey P., "Structuring Databases for Analysis," *IEEE SPECTRUM* (October, 1993): 55-58.
- Darling, Charles B., "How to Integrate your Data Warehouse," *DATAMATION* (May 1996).

### Web site resources for additional information:

- International Data Warehousing Association: <http://www.idwa.org>
- The Data Warehousing Institute: <http://www.dw-institute.com>
- Data Warehouse Information Center: <http://pwp.starnetinc./larryg/index.html>
- Datamation: <http://www.datamation.com>

---

September, 1997 **STC - Philadelphia Metro Chapter News & Views**