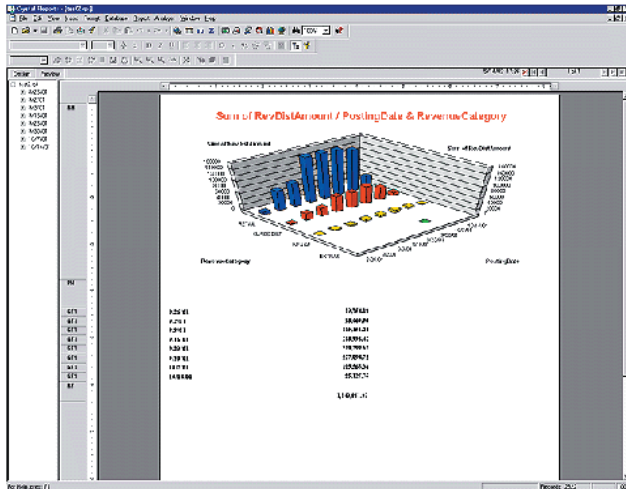




► Data Warehousing with AdMAX

By Richard J. Cichelli, SCS President



What should you expect from data warehousing? Is data warehousing simply a technology in search of a problem? This article presents a context for understanding these issues.

Enterprises are becoming more integrated and complex. Decision making is becoming more analytic. Senior management wants to “crunch the numbers.”

Can data warehousing help? Successful data warehousing projects have a management component and a technology component. Management perceives the need and defines the solution. Data warehousing technology is used to implement the solution. It may seem obvious, but an unused data warehouse is, at best, a foolish expenditure and, at worst, a disconcerting distraction. For a data warehouse to be useful requires analytic (i.e., numbers-oriented) people with a think-and-do attitude.

What will such analysts be looking for? Trends, synergies, exceptions. In short, where is the business going and what are its problems and opportunities.

Software Consulting Services, LLC
630 Selvaggio Drive, Suite 420
Nazareth, PA 18064
Sales: 1-800-568-8006
Fax: 610-746-7900
E-mail: sales@newspapersystems.com
www.newspapersystems.com

Our customers are newspapers. Unlike many businesses, newspapers are fortunate to design, produce and account for their products on computers. The most forward-looking newspapers think of their systems in terms of computer-integrated publishing. Computer-integrated publishing is the next step after linear workflow systems, based on serial processing via queues. Central repositories or databases characterize the technology of computer-integrated publishing.

Interestingly, integration and complexity form the yin and yang of enterprise newspaper systems. Unmanageable complexity characterizes both completely non-integrated and completely integrated newspaper systems. Balance seems to be required. In our highly integrated solutions, newsroom and advertising are cleanly interfaced while, within these systems, we have highly integrated subsystems.

How could data warehousing help newspapers? As the name implies, a data warehouse is a central repository of data. It may include data from many of a newspaper’s subsystems. It might even include data from multiple newspapers, such as those in a group. Reduced to its barest essentials, a data warehouse is a computer with data storage, network connectivity and COTS (Commercial, Off-the-Shelf) analytic software. A fast workstation with lots of disk space is likely to be sufficient. If you are planning to build a data warehouse, your investment in analysis and data gathering will likely far exceed your investment in the data warehouse’s hardware and software.

What kind of data analysis will you want to do? In general, management is trying to identify

1. supply and demand;
2. cause and effect; and
3. sensitivity to variability.

Think of trend-oriented questions like: “How do classified employment ad sales relate to the unemployment level month by month?” and “How do real estate ad sales compare to the federal bank discount rate over time?” Often questions will



involve three variables, such as region, time and amount of sales. Identifying successful promotions by comparing results over time or among regions are typical of management queries. Being able to predict the impact of price increases, web size changes, etc., are all issues of significant interest.

Since putting a copy of every data item of an organization into a data warehouse is a potential goal of naïve data warehousing technologists, failure to take small steps in data gathering and analysis leads to expensive and useless data gathering.

Given its effective use, the term “warehouse” is unfortunate. “Warehouse” makes me think of that last scene in the first Indiana Jones movie. Rows of ceiling-high shelves with stacks of boxes on each. And where did they put the Arc of the Covenant? If you just want to warehouse data, make a backup tape and stick it on a shelf. Doing useful analysis with enterprise data requires a keen sense of business objectives and a good knowledge of statistical methods.

A data warehouse allows you to separate operational (i.e., transactional) and analytic computing. Say the analytic task requires running a consolidated report, summarizing lots of data or searching for hard-to-find patterns. Such tasks might bring online operational systems to their knees. Assigning such analytic tasks to the data warehouse is a nearly ideal solution.

And what about the tools for analysis? Reporting programs, spreadsheets, statistical packages, text retrieval software, graphics and charting tools, etc., are designed for analytic work.

Here’s a problem: Say you have many subsystems from which you need data. Without a data warehouse, you’ll likely have to interface to many different technologies, perhaps a different one per subsystem. Sending the needed data to a data warehouse allows you to:

1. gather only the data you need,
2. place it in a common, easy-to-access format,
3. process the data without impacting the operational systems in prime time, and
4. keep secure data in operational systems and not make it generally accessible.

Analytic activities involve incremental refinement. Usually you don’t know the information you might

want until you start experimenting with the data. Data mining and data warehousing go hand in hand.

How do you turn a data warehouse into information? Usually, you want to do multi-variant analysis. This is a fancy term for wanting to see how one quantity varies against multiple, independent variables. You could plot ad sales against time and ad type, as shown in the chart above. You might compare who sold how much during several time periods. Clearly, this type of analysis would show you sales trends.

How do you build the data sets you need in a data warehouse? Most COTS software runs against relational databases. Unfortunately, many COTS packages can only link to one brand of database management system at a time. Once again, having a data warehouse helps solve this problem. All needed data can be extracted and loaded into the database on the data warehouse.

When you extract data for your data warehouse, you may only want or need a subset of your operational data. With SCS’s applications come built-in operational reports, general-purpose database reporting tools and general purpose application-specific reporting tools. The general-purpose database reporting tools allow users to define reports based on the information stored in data dictionaries. The application-specific reporting tools use information about how the data is used to support end-user report creation.

A component of SCS’s general-purpose reporting tools guides end-users and data warehouse specialists incrementally through the data extraction process. This usually involves denormalization of some database entities. For example, in advertising systems, ad data might be extracted in a way that ad insertion data has advertiser names and salesperson names rather than difficult-to-remember codes that uniquely identify them. Operational systems require this complexity. Following database normalization rules may be required in them, but not in a data analysis system.

In summary, data warehousing can be the kind of repository you can use for easy analysis of enterprise-wide data. Data warehouses can be made as flexible as you might need them.

For more information, visit the Data Warehousing Information Center at www.dwinfocenter.org.