Craig S. Mullins

February 1998

Computing : News&Review

Data Warehouse & Data Mart Administration

By Craig S. Mullins

After many months of hard work you have finally succeeding in implementing a useful data warehouse. Your clients and your management is happy to have access to all of that data. But, believe it or not, the hardest part of your job has just begun.

Data warehouses provide value to the organization by providing flexible access to more accurate information in a more timely manner. Because the data warehouse enables quick reaction to rapidly changing business conditions organizations can gain a competitive advantage by using the warehouse. So the data warehouse becomes as critical in terms of administration and management as the production systems are. Therefore, it is important to make sure the data warehouse is operating efficiently and effectively every day of the week.

Keeping the data warehouse up and working over its useful life cycle is a significant challenge which needs to be addressed during the planning stages of the project. It is important to include administration and management in the planning phases or you risk creating an unstable and non-supportable data warehouse. When putting your implementation plan together, be sure to include these items:

Consistent Data Acquisition: As the data in operational systems changes, so must the data warehouse. Over time, fields will be eliminated, meanings will change, international growth occurs, sizes change, and more. The business reacts and adapts to respond to industry trends. You must plan to keep track of physical data changes, as well as changes to the semantics of the data. Regardless of the type of change you will need utilities and tools as well as processes to allow you to keep on top of these issues and respond appropriately.

Backup and Recovery: Backup and recovery needs special consideration within the context of the data warehouse. The data warehouse should have a backup and recovery strategy that will enable the organization to recover essential data in an emergency. Depending on the size of the data warehouse, you may choose not to do a backup, because you can refresh the data more efficiently. Review the cost/benefit of each warehouse and mart, keeping in mind how often the data is updated or refreshed and how long recovery will take to

implement. Additionally, disaster recovery requirements must not be overlooked. Organizations are becoming more dependent on the information that a data warehouse provides thereby raising the importance of the warehouse application. This means the warehouse must be treated like any other critical system in terms of disaster recovery planning.

Financial Chargeback: In most organizations, data warehouse projects are managed by multiple departments, each of which has its own financial goals. Data warehouse managers should ensure that they can charge back appropriate costs to business units and users so that they can meet financial reporting requirements. An integrated solution is required that monitors IT costs by providing critical chargeback services that track information resources used organization-wide.

Scalability: As a data warehouse becomes accepted in an organization, demand for its services grows. The need for new reports and summary tables increases, and the data warehouse can explode to several times its original size. Some industry surveys indicate that 60 to 70 percent of data warehouses are filled with duplicate or redundant data such as summary tables and indexes. This can more than doubles the size of the disk subsystem required to store the data. The more users on the system, the more simultaneous queries, and the more potential there is to frustrate users with delays in response time. It is important therefore to architect the system so that it will be able

to scale linearly with demand. Parallel processors, parallel databases, bit mapped indexes, data compression, and other techniques can been applied to these issues.

Performance: System performance is closely linked to scalability and can be viewed from three perspectives: extract performance (how smoothly data is updated and refined), data management (the quality, maintenance, and query performance), and server performance. The server on which data warehouses reside requires peak performance around the clock. However, performance may have a different definition for analytical warehouse access than for OLTP. A query may realistically execute for hours in the warehouse environment, but not so for OLTP. Organizations should seek an agent-based performance monitor that collects, analyzes, and stores thousands of performance measures, is configurable for multiple environments, and offers both a real-time and a historical perspective on viewing all critical metrics. In this manner organizations can implement an integrated database performance solution which is capable of monitoring and managing the performance of: relational databases in Windows NT, UNIX, and MVS environments, servers in distributed environments, the entire enterprise network, and distributed client/server transactions. Additionally, it is imperative to optimize the speed by which the data warehouse is loaded, unloaded, reorganized, and accessed. High speed database

utilities can be used to optimize the flow of data throughout the lifecycle of the data warehouse.

Synopsis

Astute managers will constantly be on the lookout for signs of problems in their warehouse implementations. The longer one waits to act upon a problem, the more it will cost to remedy that problem. Clearly, the best alternative, is not to allow the problem to occur, but that is not always possible. In business today, the only certainty is change. Each of these areas of change necessarily impacts the structure and operations of the data warehouse, which means the warehouse must be able to react to these changes in order to maintain its value to the organization and to leverage the substantial financial and resource investment.

To ensure stability in the face of change and growth enterprise technology that manages the applications, systems platforms, and data must be used to keep pace with demanding business requirements. Operating a data warehouse smoothly is as challenging as running any sophisticated OLTP system in a day to day operational environment. In the end it is people, tools, and methods and their interaction that will get it built and make it last.

From *Computing News and Review*, February 1998.