SAS® SOFTWARE AND IBM'S DB2:
AN END USER COMPUTING STRATEGY

Thomas Miron, Systems Seminar Consultants

ABSTRACT

End user computing can benefit from strategic planning. A useful approach is to break the process down into 3 components: end user data analysis, corporate data management, and the channeling of data from the corporate database to the user. The role of SAS software and the potential of SAS MEA (Multi Engine Architecture) are discussed in this context.

INTRODUCTION

Many data management schemes begin by regarding data as a valuable corporate resource, as indeed it is. The analogy between "hard" assets, like physical plant, and data is a useful. Like other assets data requires repair and maintenance, security, and, perhaps, periodic overhaul.

The asset model can also be applied to end user computing. The value added to corporate data via end user computing should be considered a valuable corporate resource. Like other assets, corporate data and user computing require strategic planning if they are to make cost-effective contributions.

Strategic planning can be beneficial to almost any organization. When it comes to managing the complex, fast paced, interactions between people and information that typify most data processing environments, and end user computing in particular, strategic planning is an absolute necessity for success.

In the absence of planning, the pace of technical change will foster informal patterns of interaction between users and data that tend to grow along the path of least resistance from the point of view of the end user. For example, an informal data access path may involve a Marketing Research analyst using a PC spreadsheet on data downloaded from an Accounting Department extract of "live" data in the corporate database. Such a pathway evolves because the analyst happens to know how to use the spreadsheet and download software, and happens to know someone in Accounting who knows about the Accounting Department extract. These informal data paths evolve out of real needs on the part of users of corporate data, but do not necessarily represent the most efficient methods or reflect longer term goals.

Much effort has been devoted to getting a handle on these informal networks, beginning with the original, mainframe based, Information Centers, through the rapid ascendancy of PCs, where the Information Center became a one stop support-consult-training department, to the vogue of dispersing IC functions throughout the user community. The situation is still evolving.

Of current interest is the relationship between the end user and the corporate database, inspired, in part, by relational database technology, a prime example being IBM's DB2. This paper is an attempt to layout one strategy for bridging the gap between the store of corporate data that may exist in a system like DB2; and the end user, who converts that data into a thing of tangible value.

A formal strategy can preclude the development of informal networks built with incomplete or inaccurate information about the corporate database, and with little regard to future development. In most cases, informal networks do not truly serve the user as well as a more structured connection between database and data analyst.

RECOGNITION OF RESPONSIBILITIES

The case for managing end user computing is easier to state than implement in fact. Almost by definition end user computing consists of "unorganized" efforts on the part of individuals or small groups attacking ad hoc problems with tools chosen more for familiarity than effectiveness. Attempts to rationalize this effort with respect to data processing methodology are often met with resistance. And indeed, most end users do not see themselves as data processing
professionals, bound by data processing tenets; but as data analysts within their own discipline, bound by the responsibility for good analysis.

Recognition of these differing roles and responsibilities is the first step in the development of an effective end user strategy. It will be more effective to offer users the tools needed to carry out their direct responsibilities and relieve them from data management responsibilities that more properly belong with the data processing specialist, than to attempt to enforce a data management scheme on them. A realistic end user computing strategy involves recognition of these differing missions and the provision of tools to carry them out. SAS and DB2 are such tools: DB2 from the point of view of the data manager and SAS software, the end user.

THIRD STRATEGIC ELEMENT

The end user and the data manager may look to different near term objectives but they must work together constructively to meet longer term, organizational objectives. The best data management means nothing if the data cannot be utilized, and the best analytic minds are "running on empty" with no information to work with. Tools such as SAS and DB2 are essential but a third component is required: a channel, or "gateway", must exist between corporate data and the users of that data. SQL and, when fully implemented, SAS MEA (Multi Engine Architecture) can act as software components of this gateway. Segregation of user computing into the three components: data management, analysis tools, and data gateway, with appropriate resources behind each, allows for optimization of the end user computing process.

SAS - FOR THE USER

There is a tremendous variety of end user oriented data analysis tools available. SAS is certainly one of the most widely used, and for good reason. One of the strong features of SAS software is the ability to access almost any data format, from ASCII files to VSAM. This makes SAS particularly well suited for the end user environment where data tends to grow organically and analysis may require the joining of information from diverse sources.

SAS also offers the combination of a procedural language and the non-procedural PROCs. This accommodates a broad spectrum of users from novice to "power user". It would be very difficult to out grow SAS.

If SAS does have a short coming it is the lack of a graphical or full screen interface for application development. The Display Manager System provides a full screen development environment for building systems from SAS statements, but does not afford a means of building applications directly. This makes for a rather steep learning curve that can discourage novice users.

In the IBM world, SAS can play an important role as a complement to QMF.

DB2 - FOR THE DATA MANAGER

For database management, DB2 and other relational systems are emerging as standards. Most of these systems provide an array of data management services including:

- database loading and unloading
- security/authorization
- backup and recovery
- maintenance procedures
- data integrity checks.

These services address many of the concerns of the data manager, and are representative of the sort of functions that end users should not have to worry about.

The relational data structure itself tends to be intuitive. Most users are familiar with tables of information. This does not mean that relational databases cannot be very complex and intimidating. Also, issues like normalization vs. performance, which are concerns of the database administrator, may not jibe with the concerns of users who probably know little about relational theory, but know what they want to accomplish.
DATA GATEWAY

It is the third strategic element, the data gateway, that must facilitate the flow from the well managed database to the user. The data gateway serves the function of translating information from a form optimized for data management to a form optimized for data analysis. SQL is an important part of the gateway. The SAS SQL interfaces, SAS/SQL-DS or SAS/DB2, allow users to include SQL statements in their SAS code to extract directly from tables into SAS datasets.

Another gateway component within SQL is the use of views. Views are combinations and subsets of any number of tables that look to the user like a single table. Views can insulate users from the complexity of the underlying relational database and provide security functions for the database administrator as well. They can be designed to contain just the information needed for analysis, greatly simplifying or eliminating the need for user coded SQL statements.

The beauty of views is that users can "design" data formats that suit their purposes while the database administrator maintains the underlying data in a format geared to effective data management. As with most good things, there is a cost involved with views: someone must maintain them and there can be a performance penalty involved.

The data gateway must also include user support personnel who are familiar with data management and user analysis. These people provide the critical communications and expert technical support required for the data gateway.

MULTI ENGINE ARCHITECTURE

An announced component of SAS Version 6 is MEA. When fully implemented MEA will have tremendous potential as a software gateway between the end user and the corporate database. MEA will allow users to deal with DB2 tables or views directly, on a row by row basis rather than extracting data at the table level into a SAS dataset. Further, dissimilar data sources may be accessed concurrently without regard to their underlying structure. This means data from many sources can be used as if they were SAS datasets.

Avoiding an extract step has important data management implications. Direct access to "live" data avoids redundancy. The information accessed via MEA will be the data, the current data, the data as managed by the database administrator. Extract data often becomes the responsibility of the extractor...the user assumes the role of data manager. Avoiding extract maintenance is important for an efficient end user computing strategy.

As with views, direct access is not without its drawbacks. Direct access via MEA means that there is the potential for an increase in the number of transactions against the corporate database. An extract may be run once a week, on Sunday afternoon, when there is little other demand on the system. The information may be stale but the extract would have little system impact. Direct access to a database during prime shift will present new problems for data managers.

The important point about MEA is that, when implemented, it will broaden the data gateway channel between corporate data and the end user. The more efficient, the more "broadband", this channel is, the easier it will be for users to optimize data analysis and for database administrators to optimize data management.

CONCLUSIONS

End user computing should be viewed as an organizational resource and managed as such. This implies the development of an overall strategy. One such strategy is to break the system down into three components: user processing and analysis, data management, and the data gateway. The data gateway is the channel between the corporate database and the data analyst. The gateway is important because the more efficient it is the less effort is diverted from database management and data analysis.

SAS software and a relational database are powerful tools for data analysis and management. The use of views and SQL within SAS can serve as a gateway component. The implementation
of SAS MEA, with direct access to data from many sources, including DB2, will broaden the data gateway, allowing for more effective data management and data analysis.

If you have comments or would like more information please contact:

Tom Miron
Systems Seminar Consultants
6014 Gateway Green
Madison, WI 53716
(608)255-3531, (608)222-7081

SAS software, SAS/DB2, and SAS/SQL-DS are registered trademarks of SAS Institute Inc., Cary, NC, USA.