Abstract

Every organization is faced with a set of common concerns when implementing a successful enterprise-wide data access strategy. This paper will define two distinct views of enterprise data -- the data provider and the information seeker. Data is a valuable resource for decision making only after it has been made accessible to seekers who need the data. There are two barriers which must be overcome to implement a successful enterprise-wide data access strategy -- diverse data sources and data distribution. This paper will also take a look at a company that has successfully implemented several decision support applications with the SAS® System, supplying users around the world with valuable information.

Introduction

Many obstacles stand in the way of accessing data in a large organization. Certainly one of the biggest problems is that data can come from a variety of different sources -- from proprietary system files and legacy applications to different database management systems. Data can also come from a variety of different hardware environments, raising a host of connectivity issues. Because most requests for data are ad hoc in nature and therefore unique, you can spend a lot of time repeating the same things over and over again. Often, accessing data means that some of the data has to be rekeyed, which can introduce errors. You must also work to ensure that you are looking at the most current data. And finally, there are a variety of methods you must learn because not all methods apply to all data types.

Types of Data Users

Before we explore the SAS System's approach to data access and integration, it is important to understand that there are different types of data users, the data providers and the information seekers, each with their own unique needs. The first perspective is that of data providers. Data providers are responsible for making sure that others have the information they need to do their jobs, either by processing information requests directly or by providing access to the underlying data. Typically, these are data processing professionals such as database administrators, systems analysts, applications developers, or those who manage these individuals. In serving others, the primary goal of data
providers is to make information accessible while preserving the integrity and security of data assets.

The second perspective is no less important. Information seekers are those who use the data provided to perform an essential part of their job function. They can include executives and other decision makers, business or research analysts, or administrative staff. Typically not technical professionals, information seekers are primarily business professionals interested in practical use of the data they receive.

**Information Seekers**

Anyone who has ever tried to fulfill a data seeker's ad hoc request for information knows the phrase, "I need it and I need it now." Easy and timely accessibility is the single most important requirement for information seekers. That is why information seekers are generally intolerant of the technical obstacles to information delivery. They simply cannot understand why their organization is so data rich but information poor. And it is hard to fault them. After all, getting the information you need to do your job should not be a job in itself. And enterprise data that cannot be accessed and used ceases to be an asset. So the first requirement of a successful data access strategy is the ability to get data into the hands of those who need the data quickly and efficiently.

A second requirement is that seekers need to have confidence in the data they are using. Making sure that the data is up-to-date and accurate is essential, especially when the data is being used to support mission-critical decisions.

Third, seekers are typically not data processing professionals and are not particularly motivated to become them. Therefore, a high degree of transparency in accessing data is important. Ease of use is a critical success factor.

And finally, many users want to be able to do things with the data once they get the data. They want reports, analyses, "what ifs", charts, graphs, and more. That means that data has to be dynamic in nature rather than static. It also means that the data should integrate well with other tools for information delivery.

**The Data Providers**

Data providers, on the other hand, have a different set of needs. As pointed out earlier, every organization has a diverse mix of data sources ranging from proprietary application files, database management systems, real-time data collection devices, legacy data types, and more. First, your enterprise data access strategy should enable access to any or all of these sources and let you combine them according to specific information needs. Second, your data access strategy should support multi-platform connectivity to enable you to get to data that does not reside locally. Third, because you are the one responsible for the preservation of data assets, your strategy should permit a high degree of control over security.

In addition, you want a strategy that gives you a consistent interface to data rather than having to learn methods for different types of data. At the same time, you want the flexibility to apply different methods depending on the specific need. And finally, you
want a solution that integrates with your existing technology, not one that requires a steep learning curve or one that requires you to retool the way you do business.

The SAS System has been designed and engineered with all these needs in mind. One of the SAS System's central strategies is to make enterprise data, regardless of source or structure, a generalized resource available to any application or anyone who needs it.

From an information seeker's perspective, this means that there is no need to learn specific database terminology or internal structures to access data and turn it into useful information. From a data provider's point of view, this means a consistent and reliable interface to all the diverse sources of data throughout the organization.

Multiple Engine Architecture

The central enabling technology in the SAS System's data access strategy is the software's Multiple Engine Architecture, or MEA.

In the SAS System, all data—regardless of type—is accessed through engines. Engines are simply internal groups of instructions that allow the SAS System to access, read, and write data stored in a variety of different data structures. Fundamentally, MEA provides the framework for translating read and write requests from the SAS System into the appropriate database management system or file structure calls. The engine initiates and monitors communication between itself and the database management system or file.

The SAS System currently provides more than 50 different engines for a wide variety of file types including:

* relational database management systems such as DB2® or ORACLE®
* hierarchical DBMSs such as IMS®
* data gateways and applications programming interfaces such as ODBC
* operating system files such as VSAM
* a variety of external file formats supported by other software products such as DIF and DBF files
* and SAS® data sets.

Development continues for many other data structures as well. A long-term goal is to provide engines for virtually any data structure required by our customers.

You have probably noticed that access to the internal file structure of the SAS System is also included in our list. That is because the SAS System accesses native data exactly the same way it accesses data external to the SAS System. In fact, the primary role of an engine is to represent all data as if it was native to the SAS System itself. That is what we mean when we describe enterprise data as a generalized resource, and it is particularly relevant as we explore how data are surfaced to a user or to an application.

Engines can surface data in two primary forms: as data views or as data files. Views contain no data, just information about the data, their attributes, and their location. Views can describe external files, DBMS structures, and native SAS data files.
Some advantages of data views are as follows:

* reduce data redundancy
* provide currency of data
* require little storage
* combine dissimilar data sources
* use of native SQL queries
* directly access data, even across platforms

Views can be:
* defined to legacy data files
* defined as subsets of larger structures
* defined as supersets of data with added calculated fields.

Data files are actually extracts of enterprise data structures, stored in SAS data sets. Data files contain information about the data, its attributes, and the physical data that comprises each record. Data files are row (observation) and column (variable) oriented and can have indexes associated with a key column or a group of key columns.

Some advantages of data files are as follows:

* avoid ill-timed or ill-framed queries
* are generally faster to access
* are good for multiple passes of data in same application.

With the ability to create data views and data files, the SAS System gives data providers an option when assessing the needs of an individual application. Data views provide access to data, even across platforms, which brings us to the second barrier: data distribution across multiple hardware and software platforms.

**MultiVendor Architecture®**

The framework and enabling technology for multiplatform data access is the SAS System's MultiVendor Architecture™, or MVA™. MVA was the result of a six year, $100 million investment in rewriting and re-engineering the SAS System for portability across the entire range of computing environments from the desktop to the data center. This layered implementation ensures that the applications you write will run the same regardless of the underlying hardware architecture.

**Client/Server**

From the data provider's perspective, the consistency we have achieved across operating environments means that the methods and techniques described earlier are the same whether you are accessing data from DB2 or IMS on the mainframe, ORACLE or
SYBASE® under UNIX®, or DBF files on the PC. You get the same consistent and reliable interface and you use the same techniques. Another important issue directly related to data distribution is connectivity between different hardware and software platforms. With connectivity comes one of the hottest terms in the industry today: client/server.

Connectivity services in the SAS System enable you to establish conversations with one or more remote SAS sessions from a local SAS session using industry standard communications protocols such as TCP/IP. Once established, these remote sessions can provide a variety of data services and compute services.

Remote data services can provide access to all the data you have defined to the system through views that have been created on these remote systems. Data can be downloaded to your local session for processing. Data generated locally can also be uploaded to data sources residing in the server environment, taking advantage of the SAS System's read/write access capabilities. Remote compute services are also provided, which distinguish the SAS System in the client/server environment.

The capabilities described for accessing diverse data sources on local and distributed platforms or environments offer data providers the control and security needed for planning a successful enterprise data access strategy. Information seekers are provided with a virtual window to an enterprise of data, allowing them quick and easy access to the information. And that is one of the essential goals behind the concept of the virtual information database.

The SAS System at Work at Eastman Chemical Company

Now that we have discussed how the SAS System can help you with accessing enterprise data, let us turn to Eastman Chemical Company and look at how they have used the SAS System to implement several enterprise information systems.

Eastman Chemical Company is a large manufacturer of chemicals, fibers, and plastics headquartered in Kingsport, Tennessee. Eastman Chemical Company employs roughly 18,500 people in manufacturing and sales locations around the world. The company is a recipient of the 1993 Malcolm Baldrige National Quality Award. Formerly a division of Eastman Kodak Company, Eastman Chemical Company became an independent company on January 1, 1994.

Like most large companies, Eastman has struggled with finding the best way to provide information to decision makers throughout the company. In the past, some progress was made as data–rather than reports–flowed to the seekers of information. Software tools such as the SAS System became important as the means for turning data into information. However, several barriers prevented this from being completely effective. For example, some people had neither the aptitude nor the desire to write their own data analysis programs. In other cases, data was not in an easily usable format. Data integrity and security also became important issues as data files were replicated in multiple places and in multiple formats.

In order to address some of these issues, a pilot project was begun with the intent to provide the Sales and Marketing organization with historical sales data. The data was to
be stored in SQL/DS®, which would provide a central repository for the data and levels of security which had been previously unavailable. Critical to the success of this project was the realization that the sales data taken from various on-line transaction processing systems needed to be redesigned before being downloaded to the SQL/DS data warehouse. Summarization and denormalization of the data were required to make the data usable in a decision support environment.

To further simplify access to the data, application views were created in the database to allow for a truly denormalized view of the data. These application views became the foundation of a decision support architecture. Upon this foundation, all levels of inquiring could be based. Experienced SAS software users could write programs using native SQL queries.

Marketing Decision Support System

The Marketing Decision Support System (MDSS) was designed and programmed by SAS Consulting® Services. This application enables Sales and Marketing personnel to construct queries against the sales history data. The application uses a completely data-driven approach and is virtually maintenance free. The best feature of the application is that it allows the users to manage the queries they construct. Users may save, copy, and share queries with each other. This application is being used by people around the world and enables them to quickly and easily access important information about sales of specific products to specific customers in specific market segments.

Sales and Earnings Analysis System

Once the sales history pilot project became a full-scale production application, other projects quickly followed. A Sales and Earnings Analysis System (SEAS) targeting the needs of business managers was next to go into production. It is a true decision support system which provides drill-down capabilities to business managers and financial analysts. Users of this system can construct an earnings statement using any combination of customer, product, geographic, or market information. Users can further drill down to get means for tracking costs at a lower level than ever before possible. Using SEAS, they are able to isolate areas of both high and low profitability, initiate action to control costs, and effectively manage costs for their business organizations.

The decision support system nature of SEAS enables users to navigate freely through the system and to drill down in any order desired. It also enables the user to maintain an application profile of various control limits which highlight out of control variance figures.

Estimated Comparative Earnings Statements

Yet another decision support system allows for the creation of Estimated Comparative Earnings Statements (ECES). It relies upon sales and earnings, factory cost, and sales
history data stored in the data warehouse. ECES enables a business manager to create a prediction, or estimate, of earnings for a potential sales situation based on the underlying cost of sales data. Once an estimate is created, the user can make use of what-if analysis features to change the selling price or the values of various cost elements to see the effect on earnings from operations. Or, the user can take a goal-seeking approach by changing the earnings from operations percentages to see the effect on the selling price. The time to prepare an estimate has been reduced from several days to a matter of minutes. The flexibility of the application results in better decisions which in turn increase earnings from operations.

Continuing Application Development Efforts

To date, the SQL/DS data warehouse contains sales and earnings, competitive price situation, customer complaint, customer profile, invoice line item, career interest, and employee data. A Human Resources Decision Support System (HRDSS) is currently under development for use by area personnel representatives throughout the company.

Eastman Chemical Company's Reasons for Choosing the SAS System

The SAS System was the logical tool to use for the implementation of these applications for several reasons. First, it was already available in both the MVS and CMS environments. Although other application development tools were available, the SAS System provided a much desired graphical user interface.

Another very important feature is the ability to utilize native SQL queries with the SAS System. Eastman Chemical Company has an enormous amount of data in SQL/DS tables so, for performance reasons, they must have the ability to submit native SQL statements to the database. This capability enables them to ensure that data are compressed as much as possible through subsetting and summarization before returning the selected data to the calling program.

The MultiVendor Architecture of the SAS System has been extremely important to Eastman Chemical Company. As they build more and more applications with the SAS System, they can rely on a library of common routines which can be shared by all applications. With very few exceptions, there is absolutely no environment-specific code in the common routines.

Eastman Chemical Company's strategic direction for decision support applications includes the use of the SAS System. Moving toward a more distributed computing environment, they will be able to rely on the SAS System to deliver applications on the platforms which make up this new environment.
Conclusion

The SAS System has been designed around making enterprise data, regardless of the structure or location, a generalized and available resource. High-level business decisions require access to corporate data in a timely manner. The key is starting with a successful enterprise-wide data access strategy and combining it with a software solution that together can solve your business needs. The SAS System is a proven solution for information delivery and decision support. SAS Institute will continue to provide software solutions that will help your business remain competitive with our continuing development efforts.

SAS and SAS Consulting are registered trademarks of SAS Institute Inc. MultiVendor Architecture and MVA are trademarks of SAS Institute Inc.

Other brand and product names are registered trademarks or trademarks of their respective companies.

Copyright © 1994  SAS Institute Inc. and Eastman Chemical Company