Strategic Use of the SAS® System in Banking

Laurie A. Rose, SAS Institute Inc., Cary, NC

ABSTRACT

As the banking industry continues its trend of record breaking mergers and acquisitions, the systems that provide these companies with critical business information become more complex. The focus has turned from product and branch performance to understanding and predicting behaviors of individual customers. This customer centric approach has created the need for business units within the bank to look at new levels of detail in the corporate data store. In turn, Information Technology (IT) groups must respond to the demand by capturing more detailed information from operational systems and make it available in an understandable format that supports this new focus.

The SAS system is used strategically throughout the banking industry for numerous applications from CPU performance in the data center to financial reporting and consolidation for executives. This paper will focus on the data warehousing and data mining initiatives that support the customer focused approach to doing business.

INTRODUCTION

Banks are implementing "Customer Relationship Management" (CRM) strategies that include segmenting and targeting customers for increased profitability. While continuing to use traditional methods to determine short-term product profitability, it is important to develop new models to determine a customer's lifetime value to the bank.

While on-line analytical processing (OLAP) has become a common practice for uncovering relationships in customer data, data mining techniques are being implemented more frequently to gain a deeper understanding of actual behavior. A business analyst can use OLAP to discover how many customers at a particular branch are using ATM machines for their checking account transactions. Data mining is used to determine the characteristics of the customer who uses the ATM and why the customer who meets that profile prefers ATM use over branch banking.

The SAS System provides the technology for both types of analysis. A multidimensional database (using SAS/MDDB™ software) provides an efficient means of storing aggregate data that can then be surfaced through the multidimensional viewer of SAS/EIS® software. Enterprise Miner™ Software is an extensive environment for data mining, providing an array of modeling and statistical techniques for identifying patterns and trends in customer data. SAS/Warehouse Administrator™ software gives the Information Technology (IT) group the ability to manage the data extracted from the operational sources and organize it by subject area relating to the activities of the customers.

CUSTOMER RELATIONSHIP MANAGEMENT

"CRM" is not just a buzzword that banks are trying to incorporate into their dialogues. Many banks have embraced CRM as a way of doing business that is affecting various aspects of the business, from the core systems of the enterprise to the way employees are compensated. CRM encompasses all aspects of the customer life cycle, from acquisition through cross-sell and retention strategies, to analyzing churn.

From a technology perspective, we can view CRM at three different levels of implementation.

Level 1:
At the first level, the company representative is linked to an interactive server that provides up-to-date information regarding the customer's status. This person is aware of any account balances, outstanding payments, and open complaints initiated by the customer.

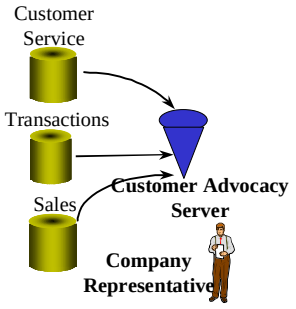


Figure 1

Level 2:
At the second level, we include a campaign management system and a customer data mart that includes historical information as well as current customer status. The company representative now knows how the customer has

responded to promotions in the past and is better equipped to offer a new product or service (cross-sell or up-sell). We also include OLAP tools to help analyze the success or failures of the various campaigns and give us a better understanding of success within certain segments.

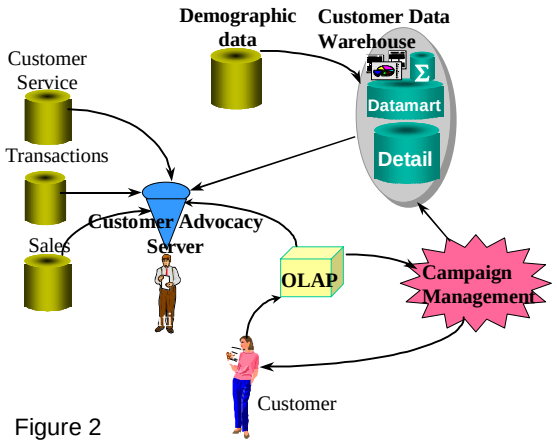


Figure 2

Level 3:
In the final stage, the data mining activity is added. We are able to better exploit the detail data that we have collected on our customer population, and marketers can more efficiently identify those customers most likely to respond to certain offers or identify the most profitable customers both short term and long term.

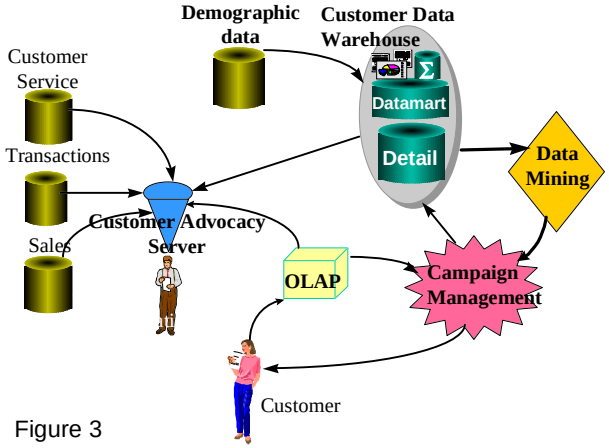


Figure 3

From the business side, we can identify some of the key solution areas where customer relationship management, in particular data mining, can expand our knowledge of the customer and impact the actions taken to increase the lifetime value of the customer. At level three of implementation, banks are better able to

- model actual behavior and predict future behavior
- assess the risk of the customer
- gain a stronger understanding of the customer's profile and how they are like other customers
- manage campaigns that are more accurately targeted to the right individuals
- ensure that profitable customers are retained.

These are just a few of the areas in which customer focused warehouses can surface information for use in data mining or business intelligence applications. It is critical however, to make sure the proper building blocks are in place when creating the warehouses that support the bank's CRM strategy.

BUILDING THE CUSTOMER WAREHOUSE

Banks are rich with data, but often that data is locked away in operational systems that are difficult to access and cumbersome to navigate. Key business drivers in the areas of marketing, risk management, and profitability analysis are pushing banks to develop data warehouses to translate untapped data about customers into valuable information to be used for CRM.

Data Validation and Metadata

An important step in building a warehouse that will be accepted and exploited by the users is data cleansing. The SAS® System programming language and the SAS/Warehouse Administrator™ provide the tools to assist in the cleansing and transformation of data from operational systems into properly aggregated systems for decision support. Data validation is critical in this process and holds the key to a successful data warehouse.

The quality of the data loaded into a customer warehouse affects many components of CRM: cross-selling efforts, risk assessment, targeted marketing campaigns, and retention strategies, to name a few. If the quality of the data is questionable, the business analysts doubt its validity and have a more difficult time justifying their models. With inaccurate data, the customer profiles are weakened, and in turn, the response

rate for retention or cross-selling campaigns is reduced.

The process of validating data is iterative. As warehouses are built, users are shifted from hitting the operational data tables to accessing cleansed information in the warehouse. Too often, the source files, both internal transaction files and often external demographic files, feeding the warehouse are inaccurate, and the data often contains redundancies or missing values. Once the inaccuracies are discovered, it is imperative to correct the errors at the source level. Since data in the warehouse is often updated on a weekly or monthly interval and summarized for those same intervals, there is no guarantee that "this month's data" is accurate if the errors have not been corrected from the inputs. Accurate data leads to improved marketing effectiveness, reduced IT intervention, increased use of the data warehouse, and better business decisions.

Another critical step in building the warehouse is the inclusion of *metadata.* Metadata can be defined as "data about data". A typical bank may create relational database tables that contain column names and descriptions of elements from the operational systems and call these tables "metadata". To make the customer warehouse more effective, the user must be able to discern how, when, and from what source the information is derived and any business rules associated with the data. For example, the definition of "profitability" may vary from one source to another, so it is critical to understand how profitability is derived. The business rules (how profitability is calculated) along with the technical information (its last update and whether it came from an accounting spreadsheet or a marketing system) needs to be created at the time the warehouse is populated. The SAS System can manage metadata from several environments, including SAS/EIS and the SAS/Warehouse Administrator, and even surface the metadata through an intranet browser application for easy access.

Data Mining Warehouses

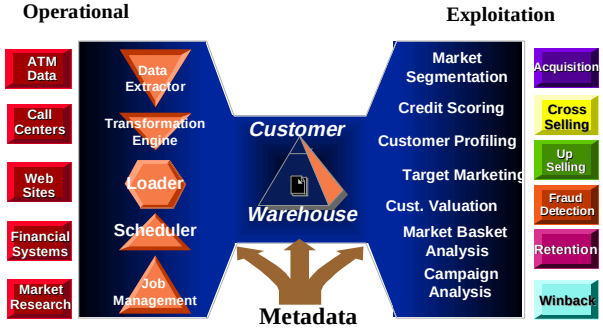
Data warehousing and OLAP tools have become mature technologies while data mining is still in its early application phases. Many banks have followed the trend of building a warehouse of relational tables, adding multidimensional data stores, then providing OLAP tools which have led analysts to gain exciting new insights about relationships between products, regions, and segments of customers never before realized. But that knowledge discovery can be greatly enhanced with additional data mining tools.

To enable data mining, the IT group may need to enhance the current systems or develop a new data warehouse or data mart for the data mining activity. The data preparation for data mining is fundamentally the same as other warehouse projects. Extraction from transactional systems and proper transformation of the data are still key activities that make the project succeed. The merging of files is more likely have the goal of adding specific new value to individual records of data that represent individual customers. Atypically, data mining data is *not* summarized. Summarization suggests that the dimensions around which the data are to be summarized are known. Data mining actually helps define which dimensions should be examined. Rather than summarize, the next important step for data mining is sampling (refer to SUGI 22 Proceedings, pp. 578-581, for a complete discussion of the SAS data mining methodology which explains *sampling*). The key point for IT is that the discipline of data warehousing is critical for data mining, but the difference will be that the end product will contain samples of detail rather than summaries.

Figure 4 illustrates the data mining warehouse environment the SAS System provides.

Figure 4

Warehouse Solution
for Data Mining and CRM



On the left-hand side, the data is extracted from the operational sources. It is cleansed and transformed into a customer focused warehouse that is augmented by business and technical metadata. On the right-hand side, the CRM strategies such as customer acquisition, cross-

selling, and retention are supported by business applications such as credit scoring, customer profiling, and target marketing.

Data Warehouse vs. Data Mart

Most financial institutions have adopted some type of data warehousing methodology, but debate still exists between the support of enterprisewide warehouses and departmental data marts. The obvious benefit to an enterprise or corporate system is that it offers "one version of the truth". The biggest pitfall is that it usually takes years to develop and deploy. With extensive cooperation between IT and the various business units who create it, the corporate warehouse provides an enriched environment for sharing critical information about the customers and the business across all entities of the bank. On the other hand, certain departments may need more detailed and historical data (especially those involved in data mining efforts) than is necessary at the corporate level. Due to the complexity and storage requirements, most organizations prefer to keep that data in a departmental data mart.

A successful strategy for a bank includes a combination of both data marts and an enterprise warehouse. The enterprise warehouse contains data that has been summarized by subject area. It most likely contains a 13-month interval of data which has been derived from the operational systems: internal transaction systems such as ATM, deposits, loans and credit cards; and external sources, such as demographic files. The summarized data is stored in relational files and in multidimensional cubes and is accessed primarily for management reporting, business queries, and OLAP applications. Departmental data marts for areas like Marketing, Risk Management, and Fraud Detection include information from the enterprise warehouse, but also include historical activity-based data. To achieve the highest level of value from these data marts, new information derived from data mining efforts or other analysis should be loaded into the enterprise warehouse when it is appropriate for other areas of the organization. For example, the Risk Management group may score a customer file and identify a high-risk profile for credit card users. Those scores need to be added at the customer level so the Marketing group can eliminate the customers meeting that profile in the next credit card campaign.

DATA MINING IN ACTION

Various solutions are available for data mining. On the low end, niche software packages exist that perform only certain types of analysis. An analyst may try to perform logistic regression or enlist a neural network application with PCs and spreadsheets. On the high end, organizations do none of the work themselves but outsource it to service bureaus. Financial institutions are clamoring for software that will drive data mining into the mainstream. In-house data mining solutions are well suited for banks that want to empower themselves over the long run. A common misconception is that data mining is a substitute for human intelligence. Those who are employing the techniques know that it takes a basic understanding of the data and the business. The business analyst within the bank can best understand the clusters, the product associations, the rules of the models, or the value of a correct prediction like acquisition cost or expected life-time profit.

Data Mining with Enterprise Miner™

The SAS System's Enterprise Miner solution offers a complete environment for data mining. From its ability to sample and modify data, to its automatic assessment of statistical models, it gives the quantitative and business analysts the most effective suite of tools for understanding data. Different analytical techniques work well for different problems, so the optimal solution for data mining is to use multiple techniques including traditional statistical methodologies. A breadth of data mining algorithms assures the best results, especially since each method has its own inherent advantages and disadvantages. Decision trees, logistic regression, discriminant analysis, clustering, neural networks, associations, and sequences are types of analysis typically used for data mining.

The three primary analysis tools in Enterprise Miner for statistical modeling follow:

- **Regression** for linear and logistic regression. The Regression tool performs linear or logistic regression analysis and saves the models and results for comparison and assessment.

- **Neural Networks** for nonlinear or linear modeling. Common applications of neural networks include credit risk assessment, market segmentation, and sales prediction. In

general, neural networks are especially useful for analytical problems that

- have large amounts of example data
- have relationships within the data that are not fully understood, with many potential models that could be specified
- require the iterative use of the data to detect patterns
- are solved by generating predictions of complicated phenomena rather than by generating explanations.

- **Decision trees** for CHAID or CART classification trees. Decision trees perform multi-way splitting on nominal, ordinal, and continuous data. A tree or subtree is evaluated using a utility function that incorporates the profit of an outcome for a decision alternative. The decision tree tool in Enterprise Miner automatically ranks input variables by the importance they contribute to the tree. This ranking may be used to select variables to use in subsequent modeling. In addition, dummy variables that represent important "interactions" of variables can be automatically generated to use in subsequent modeling.

The combination of these methods within a single environment allows the analyst to compare various models, chose the one that produces the best results, and put it into practice. The models are then tested and the results incorporated in subsequent models to gain the maximum competitive advantage.

CONCLUSION

Banks that are embracing customer relationship management strategies are determining how to take advantage of the mass amounts of customer data that exists in their operational systems. The data represents a huge asset of the business. When appropriately surfaced to the analysts and decision makers of the organization, it provides valuable insights about their customers. The implementation of customer focused data warehouses and data marts, containing both summarized data and detail data for data mining, provides the environment to better understand the relationships that exist with customers. SAS Institute's data mining, OLAP, and reporting software exploits the data and increases the bank's chance of achieving a higher level of profitability over the lifetime of their customer relationships.

REFERENCES

Brown, Mark (1997), "Leveraging the Power of One-to-One Marketing, The Data Mining Solution," DCI Data Warehouse Conference Presentation.

Kivel, G., McGinnis, P., and Wong, R. (1997), The Tower Group, "Financial MIS: New Technology Enables New Views."

McCall, Glen (February 1997), "Ensuring Accuracy," The DCI Data Warehouse Conference Proceedings, Volume 1, B20, 1-8.

Author Information: Laurie Rose, SAS Institute Inc., SAS Campus Drive, Cary, NC 27513, (919) 677-8000. Email: saslar@wnt.sas.com