

A Case Study in Data Warehousing and Data Mining Using the SAS® System Mary K. Tucker, Royal & SunAlliance, Charlotte, NC

ABSTRACT

Since January 1999, Royal & SunAlliance Insurance Company's MIS (Management Information Systems) unit has been concentrating on the formal process of data warehousing after being in the business of informal data warehousing since the department's inception. As a unit, they possess an enormous amount of business and systems logic knowledge, regularly accessing, transforming, summarizing and reporting data from almost every enterprise data source.

A next logical step in their evolution as application developers has been to create a uniform way to access, store, manipulate and present data, metadata, business rules and process documentation, or in other words - Data Warehousing. Requiring a robust, departmental, client/server physical environment, an open architecture and logical design for integration into enterprise solutions and browser based end user interfaces, they chose the SAS System.

Though pleased with their data warehousing experiences to date, the most amazing part is that they have barely tapped into the true power of their new platform and methodologies for MIS applications development. Between the data they own and/or keep, the business knowledge they have and the new technology they now possess, their potential for positive impact on the corporation is phenomenal!

INTRODUCTION

As a management information systems department, our charge, simply stated, is to answer questions. We answer questions reactively when needed and proactively when possible. Being able to turn data into information, information into action and action into value is our reason for being. As technology changes, industry competition tightens, and our clients become increasingly computer savvy, our department must rapidly seek new technology and methods to meet the needs of our wide and varied list of clients.

Beyond the online mainframe systems, printouts, and electronic file attachments, we desire to provide internet based, intelligent, integrated systems that give our end users the best management information systems in the world. By using formal data warehousing practices, tools and methodologies, state of the art data extraction, transformation and summarization tools and thin client application deployment, we want to move beyond "data reporting" to "data mining."

According to the authors of Data Mining Techniques for Marketing, Sales and Customer Support, "to really achieve its promise, data mining needs to become an essential business process, incorporated into other processes, including marketing, sales, customer support, product design and inventory control. The 'virtuous cycle' incorporates data mining into the larger context of other business processes. It focuses on action based discovery and not the discovery mechanism itself." (Berry, 18)

To this end, MIS is developing a customized process to re-engineer existing MIS applications into a data warehousing environment where significant improvements and benefits for end users and the corporation can be realized. The process is founded in accepted data warehousing principles using an iterative rapid application development methodology, which is reusable across systems, functions and business solutions.

DATA WAREHOUSING

To successfully engage data mining in our processes, the first step is to know who our customers are. We are able to list them by name,

job title, function, and business unit, and communicate with them regularly.

Next we must be able to identify the appropriate business opportunities. In MIS, our priorities are based on business needs as articulated to us by our clients through ad hoc requests and project management meetings and processes. Constant communication, integration and feedback are required to ensure we are investing our resources in proper ways.

Once having identified our customer base and business cases, we must be able to transform data into useful information. Transforming and presenting data as information is our primary function in the corporation. We are constantly looking for new and improved ways to accomplish this directive. The latest evolution in efficiently transforming and presenting data is formal data warehousing practices with browser based front ends.

Source data is crucial to data quality and mining efforts. As each new on-line transactional system and data base platform is introduced the complexity of our tasks increases. "Using operational data presents many challenges to integrators and analysts such as bad data formats, confusing data fields, lack of functionality, legal ramifications, organizational factors, reluctance to change, and conflicting timelines (Berry, 25)." Also, the more disparate the input data sources, the more complicated the integration.

A clear definition of the business need is also required to ensure the accuracy of the end results. Defining a logical view of the data needed to supply the correct information, independent of source data restraints, is necessary. Here clients and analysts get the opportunity to discuss their business needs and solutions proactively.

Next, a mapping from the physical source data to the logical view is required and usually involves some compromises from the logical view due to physical data constraints. Then questions about presentation can begin to be answered. Who needs it? How often? In what format? What technology is available?

The first iteration of our SAS Data Warehousing solution accesses five operational systems existing on six data platforms. In addition to printed reports the users expect, the data warehouse is also accessible through MDDB OLAP technology over the intranet. Users can now ask and answer their own questions, enabling the creativity needed for successful data mining. With help from the SAS System, we are busily integrating additional data, accessing more data platforms and streamlining our processes.

PAYING ATTENTION TO DETAIL

MEASURING RESULTS

Real measurements are needed in a true data mining environment because feedback is necessary to continue the cycle of results improvement. (Berry, 28). Our first step has been to enable regular monitoring of traditional measurements for the business case by enabling users direct access to standard quarterly reports across business systems and platforms. Much more opportunity exists to expand and redefine what we measure and how we determine success.

METADATA

Formal recording of metadata is also crucial to data warehousing and data mining exercises. Metadata describes data in terms of entities, attributes and relationships that are meaningful on the business level. It describes product hierarchy, the customer

attributes, the relationships between the business and various partners, and other data attributes such as when and where the data is available and what applications use it. Metadata must be flexible to change since a data warehouse is not a static environment and must respond repeatedly to changes in the business and systems environment it was built to support. (Berry, 365)

“Metadata provides the key link between the business users and the data. It describes the data in business terms. A good metadata system gives users the ability to browse through the metadata on their desktops... making users more comfortable with the data warehouse and encouraging them to take full advantage of it (Berry, 365).”

SAS DATA WAREHOUSE ADMINISTRATOR

The SAS/Warehouse Administrator® software is a graphically based software product that allows implementers of a data warehouse to visually see how it is constructed. It is designed by storing metadata that pertains to individual elements and relationships between elements. It is also designed to contain metadata about the data extraction and transformation processes including who supports the data, how often it is refreshed and business rules that apply.

The SAS/Warehouse Administrator comes with an off-the-shelf web enabled browser call the “Metaspace explorer” which allows endusers to browse the metadata from the desktops.

In Strategic Data Warehousing Principles Using the SAS System, Peter Welbrock tells us:

“The best Data Warehouse designers have a detailed understanding of the reasons it is being developed. This is of great importance because a Data Warehouse is always being *developed*. This means that having control over the process is essential. The control is needed because of the constant changes that are being made. For this reason, tools like the SAS/Warehouse Administrator can be of invaluable help. It means that the designers can concentrate more on the vision of the Data Warehouse ... (Welbrock, 278).”

Incorporating the SAS/Warehouse Administrator is currently our biggest challenge and where our data warehousing efforts are most lacking. Capturing and storing business rules is another key to enabling data mining. As the highest level of abstraction, they describe why relationships exist and how they are applied. They are often difficult to capture and are found inside code and old memos. We have certain advantages because many, if not most, business rules we need already exist in our legacy systems and code. However, translating and re-engineering existing systems is easier said than done. We must maintain a balance between creating the new environment and sustaining the old one until all processes are fully supported in the new environment. As we continue to meet the data warehousing challenge and deliver successful new products we will increasingly reap the benefit from our efforts.

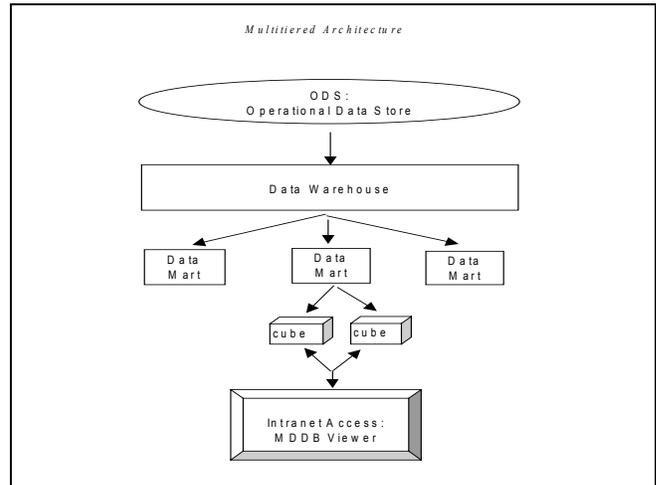
THE MULTI-TIERED ARCHITECTURE FOR A DATA WAREHOUSE

“The multi-tiered data warehousing approach recognizes that data and needs come in many different forms and provides a comprehensive solution for managing decision support (Berry, 369).”

Figure 15.4 in Strategic Data Warehousing Principles states the major components are:

Source Systems	where data comes from
Data Transport and Cleansing	moving and formatting data
Central Repository	main store for data warehouse

Metadata	what is available and where
Data Marts	fast, specialized access for end users and analysts
Operational Feedback	integrates decision support back to operating systems
End-users Warehouse	the reason for the Data in the first place



FEEDBACK

Giving our users the ability to provide operational feedback is key to the future success of our data warehousing and mining initiatives. We welcome suggestions on how to best present the end results our users seek, about how to streamline business processes and what future needs to anticipate. (Berry, 378).

We are able to demonstrate our commitment to designing our systems for our users by using an iterative design approach where new features and improvements occur on a regular basis, not more than ninety days apart. We also anticipate improvements in source data as our users begin to see the impact their decisions early in the process have on the results.

End users are, after all, the target audience for the information we provide. Their needs drive the development of applications, the architecture of the warehouse, the data it contains and the priorities for implementation. Many choose only to interface the data warehouse through printed reports as they always have. Even in this case they will experience improvements from the data warehouse as reports become more accurate, more timely and easier to access. (Berry, 380-81).

More importantly, business users who are willing to use the computers on their desks have the advantage of direct access to the data warehouse. (Berry, 381).

Through the off-the-shelf MDDB viewer that comes with SAS/IntrNet® software and custom applications under development using SAS AppDev Studio™ software, we plan to empower our business users in ways never before realized by providing on line analytical processing, report generation and drill down to detail.

RETURN ON INVESTMENT

What return on investment can be expected from a successful data warehouse? Initial returns include automation of existing processes, on line reporting, and re-engineering and streamlining existing applications. Our first success with our current data warehousing

initiative has been to re-engineer a process reducing the production cycle from four months to only two days. Many opportunities exist to realize other time and cost savings. Bigger returns include improved routine access to data, innovation and creativity for solving business problems and new ways of looking at and analyzing data.

CONSIDERATIONS FOR BUILDING A DATA WAREHOUSE TO SUPPORT DATA MINING

DESIGN

A data warehouse designed for data mining needs 1) a central repository that contains detailed data, 2) a hardware investment for the central repository that supports a variety of tools, and 3) regular use to measure the effectiveness of campaigns, especially those based on results from data mining. (Berry 386)

MIS is building an Operational Data Store in our HP-UNIX SAS environment that contains detail data from every operational data source needed to meet our users' business needs. We are implementing a second server to act as our 'deployment server' for all web based end-user applications. Consistent and timely information and upgrades help ensure our users will continue to use our data warehouse solutions.

ON LINE ANALYTICAL PROCESSING

Where Does OLAP fit in? OLAP, On Line Analytical Processing, provides powerful tools for reporting on data. OLAP is a presentation tool that enables manual knowledge discovery and is the latest advance in giving end users access to data, using client/server tools with advanced graphical interfaces for talking to powerful and efficient representations of the data, the "cube." The cube is ideally suited for queries that allow users to slice and dice the data anyway they see fit.

But OLAP tools alone will not contribute to the success of a data warehousing venture. End-user tools must provide the end-users with the answers they need to do their jobs in formats they can easily use and understand. Our phase II data warehouse deliverables include web-enabled standardized reports with links to OLAP tools when drill down to more detailed data is desired due to questions arising from standard results reporting. We are also evaluating SAS AppDev Studio to develop more customizable web based applications.

CONCLUSION

Only two years have passed since our newly ordered server and SAS products were delivered. Looking back, we have made tremendous progress, and when we take a step back and think of the successes we have had, they are wide and many. We are looking forward to the adventures to come and to the opportunity and ability to revolutionize MIS support in our organization.

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

REFERENCES

Berry, Michael J. A., & Linoff, Gordon (1997), *Data Mining Techniques For Marketing, Sales and Customer Support*. USA: John Wiley & Sons, Inc.

SAS Institute Inc. (1998), *Rapid Warehousing Methodology*. Cary, NC: SAS Institute Inc.

Taylor, David A. (1995), *Business Engineering With Object Technology*. USA: John Wiley & Sons.

Welbrock, Peter R. (1998). *Strategic Data Warehousing Principles*

Using SAS Software. Cary, NC: SAS Institute, Inc.

ACKNOWLEDGMENTS

I would like to acknowledge all Royal & SunAlliance employees and affiliates contributing to the implementation and success of my data warehousing initiatives. Without them, none of my recent achievements would have been possible.

I also acknowledge all SAS Institute, Inc. employees whom I know (and those I do not yet) who are dedicated to providing invaluable service, dedicated support and cutting edge products to help me meet my technology and business needs.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Mary K. Tucker
Royal & SunAlliance
9300 Arrowpoint Blvd
Charlotte, NC 28273
Work Phone: (704)522-3048
Fax: (704)522-3413
Email: Mary_Tucker@rsausa.com