

Providing Access to a Data Warehouse using SAS/Intrnet™ and the SAS® MDDB Report Viewer Phil Rhodes, Baylor University, Waco, TX

ABSTRACT

Providing timely and accurate information is the primary function of the Office of Institutional Research and Testing at Baylor University. With the advent of the Web, putting that information at fingertips of administrators and decision-makers is easier than ever before. Using The SAS System®, a web-based Enterprise Information System was developed to allow University decision-makers to evaluate various programs and academic departments over time. This system serves as the access point into the University data warehouse, and includes graphical and tabular reports with drilldown and reachthrough capability into individual student records. The only requirement on the client side is a web browser, with Java required to view the interactive graphics. The presentation will include a discussion of the requirements and implementation phases of the systems, as well as a demonstration.

INTRODUCTION

The Office of Institutional Research and Testing (IRT) at Baylor University is charged with providing timely and accurate information for decision support to the University administration. This information ranges from student enrollment and retention to faculty workload and compensation, as well as data for the University's new Five Year Budget Process. A decision was made to build a University-wide data warehouse in order to consolidate this information and to provide a single access point to the data. The responsibility for implementing the project was assigned to IRT, due to the knowledge IRT had of the necessary data. Using The SAS System, a web-based Enterprise Information System (EIS) was developed to allow University decision-makers to access data contained in the data warehouse.

THE DATA WAREHOUSE

Academic institutions often have an advantage over many businesses when it comes to building a data warehouse. Many times these institutions have years of data extracts, frozen on the twelfth class day of each semester. These frozen datasets are used for official reporting purposes. They can also form the core of a new data warehouse.

At Baylor University, IRT maintains frozen data for each term dating back to 1986, stored online. The data, which was initially retrieved, cleansed, and stored using the SAS® System, formed the initial core of the data warehouse. It was decided to include data back to the summer of 1996. This date corresponds to a major upgrade in the University's Student Information System, and avoided a great deal of data transformation from old system codes to the new system codes.

Baylor University's formal data warehousing project began early in the Spring of 1999. The staff of IRT began to research data warehousing methods and best practices, as well as contacting possible users and supporters of the data warehouse. Some data modeling was begun, beginning with the student data that was the data our office knew best

At the same time, a committee was formed to select software for the data warehousing project. The committee included members

from Information Technology, the School of Business, the Graduate School, University Development, and Institutional Research. RFPs were sent out to 10 different data warehousing companies, including SAS, Oracle, IBM, Cognos, and several others from the Data Management Review Top 100.

When the project bids were received, the committee met and The SAS System was chosen to use to implement Baylor's data warehouse. There were several reasons for this:

- **Completeness:** SAS was the only true end-to-end data warehousing solution, from data extraction to data cleansing, to storage, to presentation to end users.
- **Experience:** The staff of IRT already had SAS programming skills as well as a large library of SAS programs that could be modified and used for data warehousing tasks. Additionally, all frozen data and a large amount of information about the University was stored in SAS datasets and formats.
- **Cost:** When all costs (including training and consulting fees) were considered, none of the other bids came close to SAS. This was primarily due to the fact that minimal training would be necessary, as well as the fact that the university already licensed many of the needed modules of SAS.

Since a large amount of data was already stored in SAS datasets, converting them to another database format would not have been efficient. Instead, they were left in SAS format and the SAS/Warehouse Administrator® was used to manage the initial data extraction and load of the warehouse from the frozen datasets, especially the large amount of metadata generated. The Administrator was also used to add additional data to the warehouse.

Once the software was chosen and installed, the the Graduate School was chosen as the first subject area in the warehouse. This choice was made for two reasons: first, the Dean of the Graduate School was new to the position, and was requesting a large number of ad hoc reports from IRT. Second, the President of the University had recently made the decision to strengthen the University's graduate programs in an effort to move up the Carnegie classification scale. IRT staff members began meeting with the Dean to determine what type of questions he needed answered, as well as to find out what information could be provided now.

The data warehouse will eventually contain a wide range of data of very different types. This data will include individual student records, with information on courses, standardized test scores, and grade point averages, as well as demographic information and financial aid data. At the time of this writing, the majority of the data in the data warehouse was the student data, which was the easiest to obtain. Data marts have been created that focus on undergraduate data, the Graduate School, prospective student and admissions data, and the Law School. Human resource data will also be added, including information on faculty publishing and grantsmanship. Finally, departmental financial data will be included to allow administrators to track resource use and allocation.

ACCESSING THE DATA WAREHOUSE

REQUIREMENTS

In planning for providing access to the data warehouse, three main features were required:

- Flexibility
- Ease of Use
- Ease of installation and maintenance

The EIS had to be flexible to accommodate a wide variety of users. On the academic side of the University, these users range from administrative assistants in individual academic programs to the academic Deans and the University Provost. On the administrative side, users include research analysts in IRT, budget and financial analysts, and the University Vice Presidents and the President. All of these users have needs for different types of information as well as differing levels of access to information. The EIS had to be able to accommodate a wide variety of user types, giving each user access to the information they need to do their job.

This wide variety of user types with a wide range of computer skills also mandated that the EIS be easy to use. Since most users were already familiar with graphical user interfaces (in the form of the Microsoft Windows® and the Macintosh® operating systems), providing a point and click environment for most tasks was added to the requirements. Online help and data dictionaries were made available from all sections of the system. Additionally, most users would not be using the system every day. The system had to be intuitive enough for a user to come back after several weeks (or months) and still be able to get needed information in a short amount of time.

Finally, the system had to be easy to install and maintain. The project began with minimal assistance from the Information Technology Center (ITC). Without ITC support, it was not possible for the small staff of IRT to install software on the computer of each user. Again, due to staffing constraints, the system had to be easy to maintain and extend. The University computing environment included mostly Macintosh computers on the academic side, along with a mix of Macintosh and Windows computers on the administrative side, and this limited the choice of software. This desire for ease of maintenance led to an additional requirement that little or no software be installed on the client side.

After evaluating several different options, the decision was made to deliver access to the data warehouse over the University intranet, and that the SAS/Intrnet™ product would be used. This package met all three of the requirements listed above, and allowed the Institutional Research staff to make use of existing SAS programs and knowledge.

THE INITIAL PROTOTYPE

Taking advantage of existing SAS programs, a prototype application was developed to demonstrate the capabilities of the system. This initial version of the EIS used the SAS Application Dispatcher component of the SAS/Intrnet package, along with the Web Formatting Tools, to dynamically generate web pages. The prototype provided access to student demographic and enrollment data only.

The initial user request to enter the EIS was sent to the Application Dispatcher, which used a SAS program to dynamically generate an HTML form based on the user's access privileges. This form reflected the actual contents of the datasets. After the user chose the academic programs and statistics to display, submitting the form requested the Application Dispatcher to run another program to dynamically generate the

required report and return it to the user's web browser using the Web Formatting Tools.

This initial interface was demonstrated for the Graduate Dean and, while possibly adequate for some users, was not sufficient to meet the needs of the Dean and his staff. They had additional requirements:

- Drilldown capability, through both the time and organization dimensions, all the way into the detail data
- Dynamic reports that could be reformatted and displayed with different statistics and/or variables
- Export capability into Microsoft Excel for additional analysis

A SECOND PROTOTYPE USING webEIS™

It would have been difficult to meet these requirements using only the Web Formatting Tools and the Application Dispatcher, due to the large amount of SAS programming required. After some research, a second prototype was developed using the SAS webEIS™ package. This product from SAS is used to develop Java™ applets with the functionality of many of the standard SAS/EIS® objects.

Using webEIS, several Java applets were developed and placed on the IRT departmental web server. These applets allowed the user to investigate student data interactively, including drilling down through various dimensions, examining the detail data, and exporting into Excel. This second prototype was then demonstrated for the Dean and his staff.

Unfortunately, the second prototype was not without its problems. While it met the additional requirements of the Graduate Dean, it failed to meet two of the three main requirements for the EIS. First, it was not easy to use. The variable selection process was non-intuitive and somewhat complex. Second, the use of Java required the user to download and install a specific version of the Java Plugin from Sun Microsystems® - any version besides the one specified by SAS failed to work. Finally, the Java applets were non-functional on Macintosh computers due to the lack of a browser support for the correct Java Development Kit.

THE FINAL CHOICE

After evaluating the rest of the SAS/Intrnet package and other web enablement options, a decision was made to use a combination of custom Application Dispatcher programs and the SAS MDDB Report Viewer. This combination provided maximum flexibility and the ability to rapidly generate new data marts as information was added to the data warehouse. Since the MDDB Report Viewer was implemented as a SAS/AF® object, it was also possible to customize it relatively easily.

To aid in ease of use, the MDDB Report Viewer was modified to include some additional features. These included titles indicating the source of the data being viewed, as well as additional buttons on the report page that let the users view online help and the data dictionary as they were generating reports. Further customizations were made using Cascading Style Sheets and the style names provided with the Report Viewer.

THE DATA WAREHOUSE INTERFACE

Once the tools for the Enterprise Information System were decided on, a web-based EIS was set up as the access point to the data warehouse. This page contained links to public information about the University, as well as basic information about the contents and accessibility of the data warehouse. After

entering a valid user ID and password, the user is redirected to the appropriate data mart, depending on their level of access.

STRUCTURE

The EIS is structured around different data marts, mirroring the different subjects focused on in the data warehouse:

Main Page (<http://bearhaus.baylor.edu/>)

- Documentation
- Public Information
- Data Marts
 - Graduate School
 - Standard Reports
 - Interactive Reports
 - Special Reports
 - Undergraduate Data
 - Standard Reports
 - Interactive Reports
 - Law School
 - Standard Reports
 - Interactive Reports
 - Admission Services
 - Standard Reports
 - Interactive Reports
 - Special reports

The standard report page in each data mart links to reports produced on a regular basis by IRT that are related to the appropriate subject. These include student profiles as well as other information. Also included here are any ad hoc reports produced by IRT about the subject area.

The interactive report pages allow students to choose a data source (applicant data or degree data, for example), and then explore that data using the OLAP capabilities of the SAS MDDB Report Viewer. After choosing a data source, the user then uses the Report Layout page to design a report, selecting variables, statistics, and the format of the report (tabular or graphical).

Interactive Reports

- Data Source Selection
- Report Layout
 - Help & Instructions
 - Data Dictionary
- Report
 - Help & Instructions
 - Data Dictionary

Special reports are responses to requests for information by users. They generally are not as flexible or far-reaching as the Interactive Reports, and tend to focus on a narrower portion of one data source or involve small segments of multiple data sources. They are usually custom-written SAS programs run through the Application Dispatcher. For example, one of the special reports in the Admission Services data mart allows the director of recruiting to analyze recruiters performance by region in seconds at the click of a button, a process that used to take up to four weeks!

SECURITY

Access to the data is currently controlled via a combination of Active Server Page scripting and SAS coding. Each user is assigned a User ID and password, as well as a security level and access level. The security level governs what data mart the user

sees when they log on, and the access level determines what level of detail data the user may access (all users in a particular data mart may see the summary-level information). This data is coded in a non-persistent cookie sent to the user's web browser at logon time. Each time the user accesses a static or a dynamic page, this cookie is retrieved and checked to make certain that the user is allowed to access to the page or program requested.

Additionally, all of the dynamic pages use the POST method in the HTML form. This prevents the full URL from appearing in the browser address bar and prevents the user from tampering with the parameters sent to the Application Dispatcher.

FUTURE PLANS

Two large issues are on the horizon with respect to the data warehouse interface and its implementation.

First and foremost, the number of data marts and data sources will continue to expand. The next main data source is likely to be Human Resource data, and work has already begun on modeling the data and identifying potential questions. Another future area is University Development, including alumni information, to identify potential new donors to the University. This will most likely include data mining using the SAS Enterprise Miner™ product, and will probably involve custom-written SAS programs.

The second issue is the conversion from SAS 6.12 to SAS 8. This will be a gradual conversion, with new programs using the version 8 Application Dispatcher, and old programs converted as needed. The MDDB Report Viewer will require the largest amount of work, in order to include all of the customizations made to the version 6 Viewer. One program that will be converted quickly is the login program, in order to use the SAS Integration Technologies to tie into the University's BearID system and eliminate yet another user ID and password for users to remember.

BENEFITS

Several important benefits came out of the initial phases of the data warehousing process. First and foremost was greater access to information for decision making. This information was made available directly to administrators at their desks, and was quickly available to answer questions. This took some of the burden of ad hoc reporting off of the analysts of IRT and placed it in the hands of those who needed the information. Additionally, the turnaround time on the ad hoc reports was reduced from days or weeks to minutes or seconds.

A second benefit to the process was a greater emphasis on data cleanliness. As administrators began to access the data directly, they began to communicate the importance of clean data and error checking to the staff members responsible for data entry. It also led to a higher awareness of how 'dirty' some of the data was, even some of the archived data that had been cleansed already.

Finally, the process of developing the data warehouse and eliciting input from users led to a greater understanding of the needs of the those users. This should help the Office of Institutional Research and Testing continue to meet the information needs of University decision-makers, and to do so in a more effective manner.

CONCLUSION

In summary, providing access to a data warehouse through the

web enabled the Office of Institutional Research and Testing to provide information more effectively to the administrative decision-makers of the University. By giving access to information directly to the end users, a significant savings in time was realized, both for the staff of IRT and for the administrators. This allowed decisions to be made more quickly and more effectively.

CONTACT INFORMATION

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

Phil Rhodes
Institutional Research and Testing
Baylor University
PO Box 97032
Waco, TX 76798
Work Phone: (254) 710-2061
Fax: (254) 710-2062
Email: Phillip_Rhodes@baylor.edu
Web: <http://bearhaus.baylor.edu>