

Using SAS® to Get More for Less

Faron Kincheloe, Baylor University

ABSTRACT

Especially in this current financial climate, many of us are being asked to do more with less. For several years, the Office of Institutional Research and Testing (IRT) at Baylor University has been using SAS® software to increase the efficiency of the office and of the University as a whole. Reports that were once prepared manually have been automated. Data quality processes have been implemented in order to reduce the number of duplicate mailings. Predictive modeling is used to focus recruiting efforts on those prospective students most likely to respond. A web-based portal has been created to provide self-service report generation for many administrators across campus. Along with this, a number of data processing functions have been centralized, eliminating the need for additional programming skills and software support. This presentation discusses these improvements in more detail and provides examples of the end results.

INTRODUCTION

The Office of Institutional Research and Testing at Baylor University has always been the primary administrative user of SAS® software on campus. While it may be typical for the department of Information Technology (IT) to be the primary distributor and supporter of software, the use of SAS® at Baylor has evolved from the Office of Institutional Research and Testing. It started with the use of SAS® Foundation on a mainframe platform to accomplish the university's mandated and ad hoc reporting requirements. Demand for information was growing and proficiency in SAS® programming was almost non-existent outside of IRT. A series of programs were written for the Internal Audit Department. The programs were deployed through SAS® AF so the internal auditors would not be required to interact with SAS® program code. This required separate installations of SAS® on each auditor's desktop computer.

IRT had been archiving data since the mid-1980s. From that perspective, a data warehouse existed but there was no mechanism for users from other departments to access the data. In 1999, a data mart was created at the request of the Dean of the Graduate School who was seeking the ability to get data quickly enough to use it to make informed decisions. A University data warehouse soon emerged as similar data marts were created for other areas of the university. The SAS® Public Sector Portal and Data Warehouse Administrator provided the infrastructure for this data warehouse.

The Office of Admissions Services was attempting to use targeted marketing to optimize its communication with prospective students. The primary method of managing this process was a software solution provided by an outside consulting firm. The limitations of the process and the technology in place at the time caused the data to be between two weeks and three months old, making it difficult to use with confidence. In 2003, Admissions Services moved their student information system off of the mainframe to a new software solution on a new platform. The retirement of the mainframe systems forced a move of the SAS® software to a Windows® environment and to a version that included the newly released Output Delivery System (ODS). ODS facilitated the automation of many reports that previously required data to be manually copied from SAS® output to other document formats. This allowed the targeted marketing process to be integrated into the University data warehouse. The automated process refreshes data every night so that it is now no more than 24 hours old at any time. The Office of Admissions Services became the largest consumers of the data warehouse.

In the early 2000s, Baylor obtained the SAS® HR Vision solution to track payroll expenditures and provide other personnel related information. The HR Vision software was heavily modified to accommodate the University organizational structure and academic calendar.

By 2006 the number of groups using the data warehouse had outgrown the security structure provided by the Public Sector Portal. The SAS Institute had dropped support of this product. Some components of the HR Vision solution had become an integral part of doing business at Baylor but, overall, HR Vision had not been widely adopted. The decision was made to migrate the data warehouse to the relatively new SAS® BI Platform. It was economically unfeasible to repeat the modifications made to HR Vision in its new version. Therefore, only the essential elements were replicated within the BI Platform and HR Vision was dropped from Baylor's list of licensed software.

The following is a list of the currently licensed software products:

- Education Analytical Suite
- Platform Suite for SAS®
- SAS® Enterprise Data Integration Server (includes Data Quality Server)
- SAS® Enterprise Miner
- SAS® Enterprise BI Server

The BI Platform ended up touching almost all aspects of Baylor's use of SAS®. This current configuration of SAS® and the efficiencies it provides the basis for the remainder of this paper.

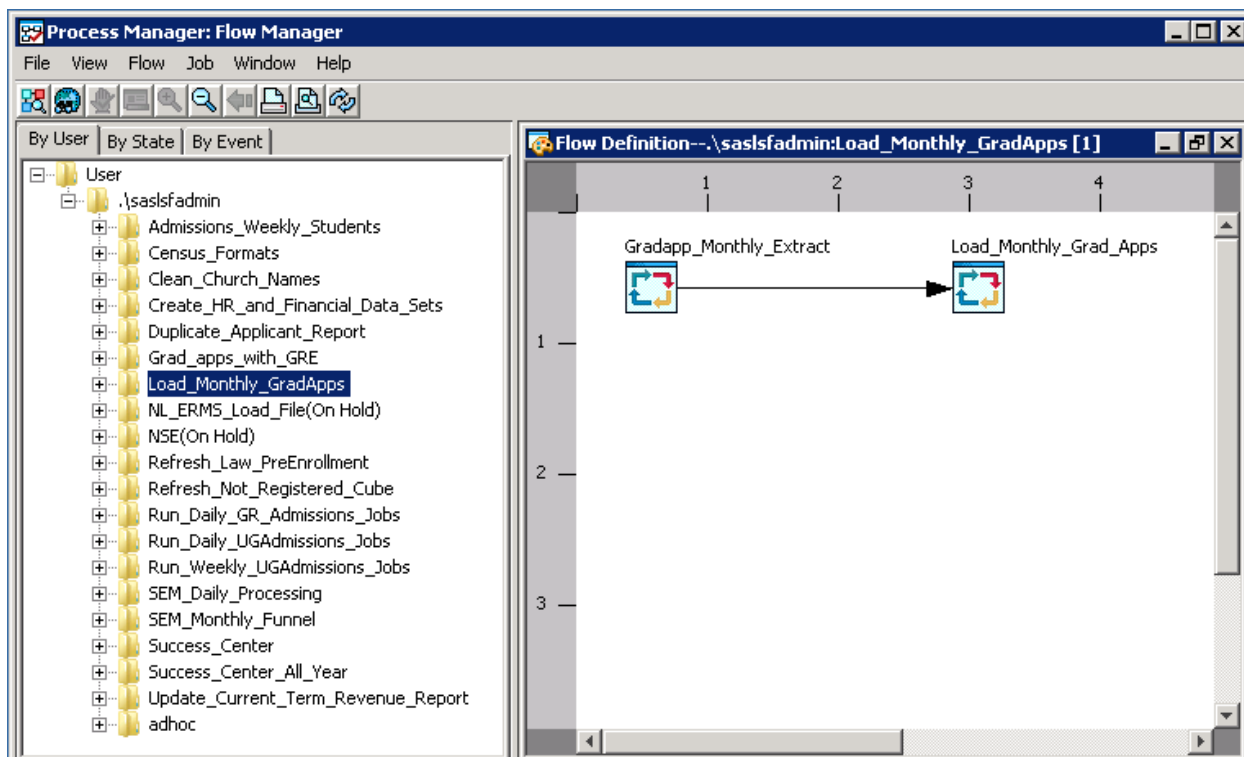
AUTOMATION OF REPORTS AND PROGRAMS

Scheduling jobs to run in batch mode is a part of everyday life in the mainframe environment. However, before ODS, the production of reports was still a highly manual process that required copying SAS® output to the desired document format. ODS can write output directly to a number of document types including HTML, Excel®, PDF, and RTF. In many cases, this eliminates the need for user intervention to produce the final report. By using a specially crafted filename statement, SAS® has the capability of writing ODS files directly to the content server used by the Portal. Once these capabilities are incorporated into a scheduled job, BI content can be updated as often as desired without any user intervention. This freed up analysts to do other things and allowed the jobs to be run during off-peak times when system performance would not be an issue.

Even though Windows® provides a scheduling service, the use of scheduled jobs is not as prevalent as in the mainframe environment. Baylor used the Windows® task scheduler to automate the extraction of data and the creation of reports as much as possible. However, there were some inherent weaknesses in this process. There was no good process for notification of failed jobs. Additional code was included in scheduled jobs to notify administrators when there was an error. Sometimes this was unreliable or would mask other problems. Administrators were still required to routinely monitor the logs to ensure that all jobs ran successfully. There was also no good way to establish dependencies among scheduled jobs. If a second job was dependent upon the first, the second job would be scheduled to run significantly later than the first in order to allow enough time for the first job to complete. If the first job ran longer than anticipated or aborted with an error, the second job might append old data to an archive table creating duplicates. This would create extra work to remove the duplicates from the table, correct the error and rerun the process.

SAS has partnered with Platform Computing to make a version of the Process Manager software available for use with BI installations. This software integrates seamlessly with the SAS® Management Console and Data Integration Studio to schedule both load processes and user-written programs. The Process Manager includes built-in notification of job status. It also allows for the creation of complex job dependencies where one job can be configured to run only upon the successful completion of one or more other jobs. Other triggers are available, such as file modification, so that a job could be launched when a particular data set is updated. Multiple jobs are placed in a Flow Definition and the flow is then scheduled to run at the desired day and time. Status notification can be configured as desired. At Baylor, the first flow sends an email to the administrator when the flow starts to run. This is confirmation that the service is active. Subsequent flows only send a message if one of the jobs in the flow encounters an error. It is rare for the scheduler service to crash, but when this happens, no messages are sent giving the appearance that everything is running successfully. The final flow of the day also sends a message upon successful completion as an assurance that the scheduling service is still active at the end of the day. If the administrator is checking email, there is no need to manually monitor the scheduler.

Display 1 shows the list of scheduled processes and an example of the dependency of two scheduled jobs within a Flow in the Process Manager.



Display 1. Process Manager Interface

DATA QUALITY PROCESSES

The SAS® Data Quality Server brings the data cleansing power of DataFlux® software into the data step. This is extremely useful because it allows data cleansing to be integrated into scheduled jobs that are already working with the data that needs to be cleaned. Data Quality Server has been bundled as part of the Data Integration Server package. Baylor University uses Data Quality Server primarily to standardize organization names and to identify duplicate records among prospective students.

Since Baylor is a private university with a religious background, churches are a fertile ground for recruiting new students. However, church names are entered into the prospect database as free form text. As a result, these data are unusable for reporting purposes without some sort of standardization. For example, a large local church, known to be home to many Baylor students did not appear in the top 25 in a report listing the number of students by church. In fact, this church appeared as number 30, 37 and 57 in the list. Further investigation revealed that 56 students had listed this church 17 different ways in the database. The DQSCHHEME procedure was used to create a standardization scheme for church names. A program runs every night using DQSCHHEME to apply the scheme to church names entered into the database. This converts the irregular entries into standardized entries that can be used for reporting. After standardization, the church mentioned above turned out to be fourth in the report of number of students by church.

As a private university, Baylor depends almost totally upon tuition and fund raising for its financial existence. Consequently, it is very important for Baylor to maintain a good relationship with prospective students, alumni, donors, and churches. Duplicate records create a number of problems in maintaining good relations. Duplicates often go undetected due to errors in the name or address. No one likes to receive multiple copies of the same piece of mail. This not only makes the University appear unprofessional, it is also costly. The most obvious cost is the expense of printing and mailing the second copy of the communication. There are also indirect costs related to lost opportunity for those who did not receive the communication. It is cost prohibitive to mail to every prospective student. Therefore, a specific number of targeted prospects receive a communication based on the amount budgeted for a specific campaign. Every duplicate represents a prospect that did not receive the communication. At some point, a number of missed prospects translate into a student who does not enroll which, in turn translates into lost revenue.

Another serious problem occurs when duplicate records belong to a person who is applying for admissions. In this case, there is the potential for the application to be entered for one record and important documents such as transcripts and test scores entered into the duplicate record. If this happens, an applicant who completed all the steps

for admission would not be considered for acceptance because neither record would appear to be complete in the records system. The DQMATCH function in Data Quality Server is used to identify these duplicate records so the database can be cleaned to avoid the aforementioned problems.

PREDICTIVE MODELING – TARGETED COMMUNICATION

It is neither feasible nor cost effective to maintain equal levels of communication with every prospective student. SAS® Enterprise Miner is used to create predictive models at various stages of the matriculation process. It starts with prospective students who have not yet initiated contact with Baylor. Each prospective student is assigned a score indicating probability of enrollment based on available information. Depending on the budget, a certain number of those students with the highest probability scores are introduced into the communication stream. After the prospective student begins corresponding with Baylor, a second model is produced to determine if additional communication, such as a telephone call from a current Baylor student, is warranted. It would be cost and time prohibitive to attempt to call everyone, so the model scores are used to prioritize the calls. Additional models are created to predict the subsequent enrollment of those who have applied for admission to Baylor. These models are used to prioritize further communication and to help predict the size of the incoming class. After a student has enrolled at Baylor a retention model is created to help identify those students who may be at risk of leaving Baylor. The retention model score is used to help determine the need for and shape early intervention programs.

CENTRALIZED PROCESSING

As mentioned in the introduction, there were several Internal Auditors who had individual PC SAS® licenses in order to run the reports that were deployed using SAS® AF. There are also a few users in the Cashier's Office and the Financial Aid Office who need to run SAS® programs but do not have the need or proficiency to do frequent programming. The Office of Institutional Research and Testing also received requests to run certain report programs on a frequent but unpredictable basis. The Stored Process Server, included in the BI Platform, provided a means of converting the SAS® programs to stored processes that could be accessed through a web interface on demand. The SAS® AF application used by the internal auditors was converted to a series of stored processes. This allowed us to reduce the number of workstation licenses that we were maintaining. The use of stored processes also eliminated the need for several staff members to learn to use the SAS® programming interface to open and submit programs and allowed them to focus on the core competencies of their respective jobs. It also provided the capability of running a job remotely while at home or traveling. Another benefit was less time spent by support staff maintaining the individual installations.

SELF-SERVICE, WEB BASED PORTAL

Early in the development of the original data warehouse it became apparent that we did not have the resources to support client software. Therefore, we adapted a paradigm that our data warehouse and, later, the BI platform must be accessible without any additional software on the user's computer. The SAS® Information Delivery Portal and Web Report Studio have enabled us to stay true to this paradigm. They have become the medium through which all the efficiencies previously discussed are delivered. BI content is grouped according to the needs of groups of users. When a user logs onto the portal, she is presented with a unique set of tabs based on the groups of which she is a member.

The content on these tabs is in a wide variety of formats based on the type of data, the frequency with which the data changes, and the needs of the user. Formats include:

- Drillable cubes that allow the user to see both detailed and summarized data sliced and or filtered across a number of dimensions.
- Interactive lists, some of which include model scores allowing users to sort and filter according to their priorities.
- Stored processes which provide parameterized reports on demand.
- Static reports in HTML, Excel® or PDF format, many of which are automatically updated on a daily basis.
- Interactive Excel® documents.
- Links to other web resources and a limited amount of content management as requested by the user groups.

New content is being added to the portal every week as users request additional information and seek to make the portal their "one stop shop".

CONCLUSION

There was an expectation that the data warehouse would make IRT's job easier, reducing the number of ad hoc requests as users were able to access information on their own. While users have been able to produce many of

their own reports, requests have not diminished. As one question is answered a tougher question arises. The organization of data in the warehouse and the BI Platform exposes data issues. Even though this causes extra work for cleanup and maintenance, the result is better data for analysis purposes. The anticipated spare time never materialized but Baylor has been able to do more with less through automation, increased data quality, predictive modeling, centralized processing and self-service, web-based reporting.

CONTACT INFORMATION

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

Faron Kincheloe
Baylor University
One Bear Place #97032
Waco, TX 76798
(254) 710-8835
Faron_Kincheloe@baylor.edu

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.

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