Paper 410-2011

# Implementing a Data Mining Solution with SAS® and Teradata

Angela Watson, Overstock.com, Salt Lake City, UT

## ABSTRACT

This paper discusses the challenges that were overcome, and the benefits that were achieved by implementing SAS® Analytics with Teradata at Overstock.com. Follow the steps from "Decision to Implement" to "Production Analytics," as the Overstock.com team and SAS analytical consultants worked side by side. Experience business transformation, from initial model development and knowledge transfer to business process change, that resulted in significant performance and productivity improvements.

## INTRODUCTION

Overstock.com, Inc. (short cut: O.CO) is a Savings Engine offering brand-name merchandise at discount prices. The company offers its customers an opportunity to shop for bargains conveniently, while offering its suppliers an alternative inventory distribution channel.CRM (customer relationship management) analytics and database marketing methodologies are used by the Overstock.com team to drive strategic and operational decisions that are beneficial to the customer and the company.

As the scope of the analytical questions to be explored expanded and the size of the application of the team's findings grew, we quickly recognized the need for a robust, scalable and fast data mining solution. This paper shares the team's experience implementing SAS® Enterprise Miner and SAS® Scoring Accelerator for Teradata and offers recommendations for what to prepare for and keys to a successful implementation.

## PREPARING FOR INSTALLATION

In addition to the typical installation instructions for SAS® Enterprise Miner, following is a list of other documents that helped us prepare for installation:

1. Hardware sizing questionnaire – This was filled out between our Analytics and IT teams to determine the operating system of the server(s), the volume and frequency of the data we would be modeling, and the environments we planned to build out (e.g., production, development, test). After receiving the questionnaire, the SAS® team provided hardware recommendations that we could evaluate.
2. Reference architecture – This was an example hardware and software configuration that we used as a starting point to determine the appropriate architecture for our needs. We chose to build out production, development and test environments. The development environment is used as a sandbox for model development, the test environment is used for testing models and patches, and the production environment is for model deployment and scoring.
3. SAS® Accelerator Publishing Agent for Teradata – Aside from the configuration and installation of the software on the servers, there was also some work to be done by the Overstock.com Teradata DBAs to ensure that we had the right client packages on the Teradata side that matched the SAS® Enterprise Miner installation (i.e. 64-bit SAS® to 64-bit Teradata client packages).

Be sure to put some thought into the immediate and future workload of the hardware to ensure the expected level of performance from the system for a reasonable amount of time.

## INSTALLATION

When it comes time to install the software, there is the option of a remote install with a consultant assisting virtually or an on-site install. The Overstock.com IT team chose to do a virtual installation, working side-by-side with a SAS® resource for installation, configuration and testing.

The most challenging part of installing the software was making sure we had the right documentation and instructions for installing SAS® Scoring Accelerator for Teradata. In addition to documentation on the publishing agent mentioned above, we also found that we needed to apply a hot fix for the SAS® Formats Library for Teradata in order to successfully complete the installation.

### OVERCOMING THE UDF QUESTION

As many others have presented and discussed, there are significant performance and productivity gains realized when leveraging SAS® Scoring Accelerator for Teradata. Performance run times and model development are faster and analysts are free to invest more time in analyzing business questions rather than managing data volume, transformation and the mechanics of scoring.

The export utility of SAS® Enterprise Miner and SAS® Scoring Accelerator for Teradata exports the model scoring logic along with metadata about the required import and export variables. The publishing client then translates and publishes the model code into C source code for creating the scoring function in the Teradata database. The Teradata User Defined Functions (UDFs) created by the publishing client are then available to be called in Teradata SQL code.

The Teradata DBAs at Overstock.com had some concerns about the visibility they had into the UDFs created by SAS® Enterprise Miner and whether or not the UDFs would require extensive testing before deploying them into the production database. One differentiator of the UDFs created by the publishing agent is that there is no user involved in coding the UDF and the research and development groups at both SAS® and Teradata have rigorously tested the publishing client and resulting C code.

At Overstock.com, we employed the recommended sandbox methodology for creating an instance of the database that is specifically for SAS® Enterprise Miner model development that analysts have increased privileges to and can use for model development and testing the scoring UDFs. This, in addition to deploying the UDFs in development and test environments on both the SAS® server and Teradata database side have addressed some of these concerns. We expect to see even more success here the more times we develop and deploy models.

## BEYOND INSTALLATION

As part of our comprehensive implementation of SAS® Enterprise Miner and SAS® Scoring Accelerator for Teradata, we chose to scope some initial modeling projects and engage with SAS® analytical consultants to assist our CRM team in delivering a model deployed in production. This allowed for a quick-win project to train the team on best practices and begin delivering results right away.

### WHAT TO TACKLE FIRST?

One of the first modeling projects we tackled was to develop a response model that would allow us to decide which coupons to offer to which customers based past coupon use behavior and sales metrics. The goals of this project were to:

- Develop a response model for use in one of Overstock.com's email marketing campaigns
- Build a SAS® Enterprise Miner project (diagram) capturing the model and the export node to initiate the scoring UDF
- Transfer knowledge and expertise from the analytical consultant to the Overstock.com CRM team through side-by-side model building and analysis
- Calculate the expected lift from deploying the model based on a validation sample
- Successfully complete customer scoring using SAS® Scoring Accelerator for Teradata
- Evaluate the results of the model after deployment compared to the expected lift
- Enable the CRM team to use and extend experience from this project for future model building and deployment

As with all modeling projects, some time was spend on building a data set containing various customer behavior and sales variables that we could explore during model development. The analytical consultant worked closely with Overstock.com's CRM team as the data set was built, providing input into which variables would be interesting to explore in addition to some of the variables we had used in other modeling projects.

### KNOWLEDGE TRANSFER AND RESULTS

Although the Overstock.com CRM team had many skilled and experienced CRM analysts, there was value in combining the team's knowledge of Overstock.com customer behavior and the experience the SAS® analytical consultants had in building similar response models. The time spent in side-by-side model development allowed for close collaboration that resulted in a robust response model that addressed the specific business problem at hand.

The model was initially tested on a small sample of customers to validate the expected lift calculated during model building. After a successful test of the model compared to a control group, the model was deployed to production and is still generating the expected lift.

## ADDITIONAL TRAINING

After the success of the initial model development, Overstock.com also found it valuable to subsequently complete two on-site training courses:

1. Applied Analytics Using SAS® Enterprise Miner
2. Advanced Predictive Modeling Using SAS® Enterprise Miner

## CONCLUSION

When preparing to implement any large-scale software solution, it is important to have a good understanding of the business problems that the solution will address. This understanding will aid in identifying the right solution and provide a framework for configuring the software in the right way.

It is also extremely helpful to include the interested IT groups within the company from the beginning of exploring the solution. This allows the IT group to understand the business problems a successful implementation will solve and the impact of solving those problems to allow for appropriate prioritization of projects.

It is also strongly recommended to engage with analytical consultants in a manageable initial project that quickly delivers ROI and provides real-world experience for the teams that will be using the solution to build on.

Overstock.com continues to have success using SAS® Enterprise Miner and SAS® Scoring Accelerator for Teradata.

## ACKNOWLEDGMENTS

Congratulations to the Overstock.com and SAS® teams on a successful implementation. Thanks to members of the Overstock.com and SAS® teams that helped research and review this paper, especially Bryan Archibald, Fritz Van de Kamp, Phil Broadbent, Paul Longhurst, Micah Steadman, Russell Sedgwick, Matt Cranney, Peter Maughan, Isabel Braulio and Robert Stevens.

## RECOMMENDED READING

Please see below for recommended reading.

- Explanation of development, test and production environments: http://www2.sas.com/proceedings/forum2007/204-2007.pdf
- Forum posts from the SAS® and Teradata Product Advisory Council on in-database scoring best practices: http://support.sas.com/forums/thread.jspa?messageID=23746&

## CONTACT INFORMATION

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

Name: Angela Watson
Enterprise: Overstock.com
Address: 6350 S 3000 E
City, State ZIP: Salt Lake City, UT 84121
Work Phone: 801-947-3593
Fax:
E-mail: awatson@overstock.com
Web: http://www.overstock.com/