

In the house

Teradata and SAS advance in-database analytics. *by Arlene Zaima*

Since the announcement of their partnership, SAS and Teradata have made great progress on an aggressive joint research and development (R&D) road-map. They have successfully delivered SAS Scoring Accelerator for Teradata and made enhancements to SAS business intelligence (BI) and SAS foundation technologies.

SAS also released a series of patches, code-named Supertanker, to help optimize customer deployments. These solutions, which function inside the Teradata Database, shorten the time it takes to analyze data. To support customers during product integration and in ongoing product enhancements, the partners established the SAS and Teradata Center of Excellence (COE).

SAS Scoring Accelerator for Teradata

Customers using SAS and Teradata products are leveraging Teradata Analytic Data Set Generator along with PROC SQL to build analytical training tables in the Teradata Database for modeling with SAS. SAS modelers then use SAS Enterprise Miner to develop models and Scoring Accelerator to publish the initial production model to the Teradata Database as a native scoring function.

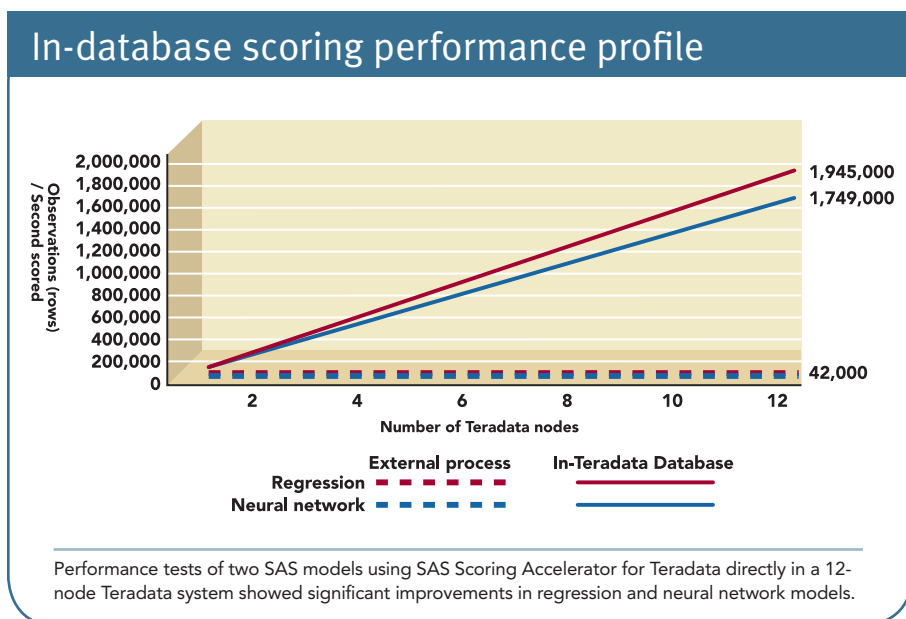
An alternative to scoring models using SAS/Access to Teradata, Scoring Accelerator combines the statistical transformation and modeling methods available in SAS Enterprise Miner with the scalability and processing speed offered through the

Teradata system. The combined Teradata and SAS products provide an integrated analytical environment for the massively parallel deployment of SAS analytics in the Teradata Database.

Previously, to achieve in-database scoring, some analysts manually translated SAS scoring code to SQL. Now Scoring Accelerator simplifies deployment of SAS Enterprise Miner models in the Teradata Database by making them automatically functional with any SQL program. The new solution has three components:

- > **SAS format library** is a file that is deployed once to the Teradata system in a post-installation step.
- > **Score Export Node** exports models created by SAS Enterprise Miner 5.3 into a set of files that are used as input to the publishing macro.
- > **SAS Publishing Agent** publishes the scoring function into the Teradata Database. SAS 9.2 translates the exported files into a SAS function definition and registers the function inside the Teradata Database for use in any SQL expression.

The SAS models are converted into functions that are deployed and executed on every available node within the Teradata system; therefore, the number of nodes and the current workload determine the batch scoring speed. Performance tests have been run on a number of different model types, including regression, decision trees and neural networks, and the performance enhancements have been consistent. “We



Teradata and SAS four-step best-practice model development process

When exploring the answer to a clearly defined business problem, much of the data preparation can be performed up front, where it is built once and used by the entire analytic community. This can be accomplished through joint SAS and Teradata solutions by following these four steps in the analytical development process:

> Step 1: Explore, integrate, prepare. Use Teradata Analytic Data Set Generator for easy data exploration and preparation directly in the data warehouse to avoid unnecessary data movement. Once a set of data elements is identified for modeling, build an analytical data set in the warehouse where data and tables are integrated, and transformations, aggregations and new predictive variables are derived.

> Step 2: Discover, transform and build. Extract the analytical data set into SAS Enterprise Miner in the model development phase using SAS/Access to Teradata. Data mining is highly iterative, meaning one analysis can lead to another. This could include exploring the data further to discover relationships, creating new optimized variable transformations,

filtering extreme values and selecting a subset of candidate variables for final modeling. The model comparison tool of SAS Enterprise Miner is used to select a production model that performs well on holdout data.

> Step 3: Translate and deploy. After the model development phase is complete, the production model is ready to be deployed. SAS Scoring Accelerator for Teradata translates the Enterprise Miner model and registers it as a scoring function in the Teradata system.

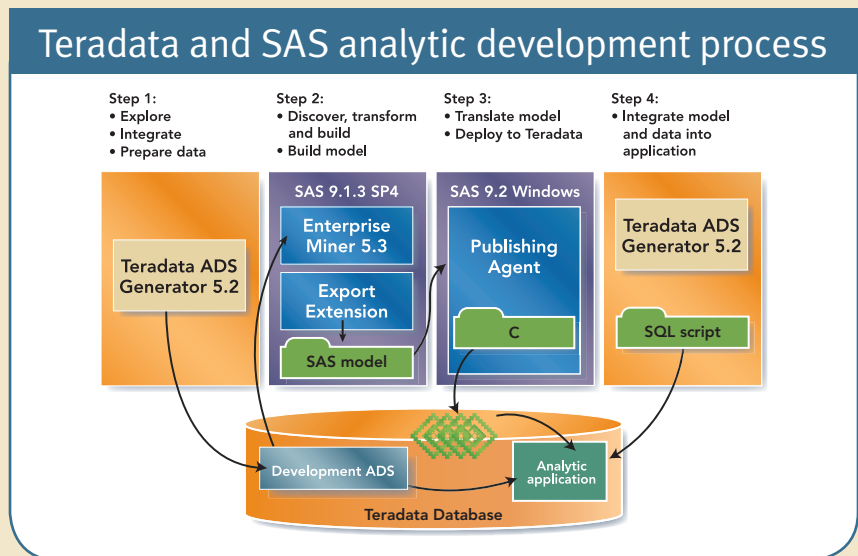
> Step 4: Integrate. Teradata Analytic Data Set Generator can integrate these scoring functions into a SQL program, eliminating the need for any SQL programming by the user. Its features allow users to select and parameterize the scoring functions with the production data, creating a complete and integrated application.

—David Shamlin,

Senior R&D Director, SAS

—Wayne Thompson,

Senior Product Manager, SAS



have seen performance gains of more than 10 times for some scenarios tested in our joint lab environment,” says David Shamlin, senior R&D director at SAS.

To validate the performance improvements, two SAS models were tested directly in a 12-node Teradata system using Scoring Accelerator. As shown in the figure (page 1), the regression model processing jumped

from 42,000 rows to 1,945,000 rows per second, and the neural network model increased from 42,000 rows to nearly 1,750,000 rows.

COE and Supertanker enhance performance

Through best practices and workshops, the COE empowers SAS and Teradata

customers to make the most of their data warehouse infrastructure by enhancing system performance and technology utilization. The COE works with customers to enable more of their SAS-generated BI queries to pass directly to the Teradata system.

SAS also developed the software patch Supertanker, which is instrumental in helping customers optimize their SAS

Teradata and SAS analytic environment

Analytical development server:

SAS 9.1.3 (BASE SAS, SAS/STAT, SAS/ACCESS to Teradata)
SAS Enterprise Miner 5.3

Windows client:

SAS 9.2 for Windows (BASE SAS and SAS/ACCESS to Teradata)
SAS Scoring Accelerator for Teradata
Teradata Analytic Data Set Generator 5.2

Database/operating system:

Teradata Database 12.0 on Linux

9.1.3 environments. Supertanker is a series of enhancements to SAS SQL processing technology that allows more queries to pass seamlessly from SAS BI software to the database as Teradata-specific SQL. SAS users can apply Supertanker to their environments via a set of SAS 9.1.3 SP4 hot fixes and turn on the enhancements using a SAS macro variable. (Usage information is provided in the hot fix documentation.)

Working with the COE, companies can benefit even more by accelerating SAS Enterprise Business Intelligence processes. SAS information maps, which are optimized using Teradata views with

Supertanker, provide a business metadata layer that describes physical data structures. At one company, by automatically generating appropriate query code based on the user's selection, information maps reduced processing time on a specific project from 40 minutes to less than one minute.

Enhancements were also made to SAS foundation technology, a format publishing agent that converts data values used throughout SAS applications from one type to another. The publishing agent works with Supertanker to allow SAS formats to run inside the Teradata Database. Performance benefits can be seen in SAS BI reports, data integration workflows and a number of its solutions. The roadmap for SAS foundation technology focuses on key data preparation functions that are part of Base SAS and SAS/STAT.

SAS collaborated with its customers to discover ways to leverage SAS components with Teradata system capabilities and push more sophisticated SQL to the Teradata Database. Scoring Accelerator currently supports SAS Enterprise Miner; however, research plans exist to pursue the ability to support SAS Data Step (SAS 4GL) by translating the code into functions that run in

the Teradata Database. This allows users and SAS to push a richer set of ad hoc functions into the database.

Future development of SAS analytic integrations will support a broader class of statistical transformations, sampling functions and models. SAS and Teradata are working with customers to prioritize these requirements. SAS Model Manager 2.2, an analytic model management and deployment environment, will also be fully integrated with Scoring Accelerator to further streamline the scoring of SAS models in the Teradata Database.

The SAS and Teradata partnership will continue to deliver best practices and new product integrations that allow customers to maximize their SAS and Teradata environments. "Our key emphasis moving forward is to push modeling in the database," says Wayne Thompson, senior product manager at SAS. "We've made good progress with some of our regression and variable selection models and are working on other descriptive and predictive modeling techniques." T

Arlene Zaima, a strategic intelligence program manager at Teradata, has more than 10 years of experience in advanced analytics.

SAS and Teradata Center of Excellence

The SAS and Teradata Center of Excellence (COE) is a dedicated team of solution architects and technical consultants. These SAS and Teradata experts share their experience and knowledge in data warehousing, advanced analytics, business intelligence (BI), IT infrastructure, data integration and industry-specific domain knowledge with their clients. Opportunities through the COE include:

> SAS and Teradata Optimization Services portfolio.

Designed to assist customers in leveraging their SAS and Teradata technologies more effectively, these services focus on providing joint customers immediate financial benefits and performance improvements, reduced data movement, improved analyst efficiency and accelerated time-to-value by streamlining the analytical processes.

> SAS and Teradata Architecture Assessment Workshop.

This popular program focuses on assessing and recommending the right blend of products, platforms and

releases to maximize efficiency of an organization's Teradata and SAS environments.

> **Optimization Workshop.** Programmers and analysts are provided a framework of best practices and given hands-on examples for effectively using the SAS and Teradata technologies in an integrated, optimal manner. Key concepts addressed are application efficiency, doing the right processing on the right platform and minimizing data transfer.

—Helen Fowler, COE director, Teradata