

Paper 087-2013

## SAS and Teradata In-Database Model Development Using the Business Example of Churn Modeling

Karl Krycha, Teradata; Jonathan Wexler, SAS Institute Inc.

### ABSTRACT

This paper provides an overview of SAS® in-database capabilities available on Teradata and uses a typical business application to illustrate how those capabilities can be used to greatly improve the end-to-end analytical modeling process. We will focus primarily on several benefits of using SAS Analytics Accelerator for Teradata. This product reduces modeling execution time from hours to seconds, allowing users to produce models faster. In addition, it can greatly reduce data duplication and data movement by performing key SAS analytic computations within the Teradata database.

### BACKGROUND

Traditionally, analysts extract data from Teradata and create SAS analytical datasets on separate hardware. They may explore the data first and then build models using SAS clients.

Although this method offered advantages, it has some inherent business challenges:

- Decisions on how to define the modeling population or how to sample data are typically taken in the beginning of the modeling process. Alternative modeling strategies are considered during this initial modeling step. This means the extracted population or sample might not be the sample the analyst wants for modeling and re-doing the sample requires additional time.
- Typically, the target event is a collection of the most recent observations. Any extraction of data brings latency into the modeling data, so sampling of modeling data should consider most recent observations as much as possible. This is another case where extracted data may not be the best sample for the model.

The modeler needs to cope with technical challenges, as well:

- Offloading the data essentially means downloading a large set of data through a typically thin network pipe which can cause some wait time before the actual model building process can start.
- Offloaded data will be run through the modeling process on infrastructure that may have significantly less computing power than the Teradata database. This may result in performance issues when executing the required analytical SAS procedures used in a particular model build cycle. This may result in insufficient analytics and/or oversimplified model settings (e.g., heavily sampling).
- When data is extracted from Teradata, it is taken from a highly secure and tightly governed database environment and becomes vulnerable for security breaches and data theft.
- Data duplication is inherent with this process. Data may be spread out in the organization with no version controls.

### SOLUTION

SAS Analytics Accelerator for Teradata, a component of SAS' in-database analytics product portfolio for Teradata, addresses the issues described above, by improving the integration between the SAS System and Teradata, and by moving SAS' analytic computations capabilities to where the data is stored already in the Teradata database.

In the following sections we will walk through a sample modeling scenario to demonstrate the use of the particular SAS procedures that are available as in-database versions through the SAS Analytics Accelerator for Teradata. (Other useful products will be mentioned as well.) The resulting advantages are manifold. From a modeler's perspective, the major benefits are the increase of performance and access to more detailed data with bigger historical depth.

In the following benchmark example you can see how in database integration can significantly reduce the preparation efforts and create more time for actual modeling. In Display 1, the Input SAS Table Size was 5,750,299 observations and 201 variables. The SAS processing time before using Analytic Accelerator for SAS PROC SORT plus SAS PROC SCORE was 20 min + 58 min = 1 hour 18 min = 4,739 seconds. Using Analytic Accelerator the same process requires only 15.87 seconds.

To improve this process the in-database PROC SCORE by SAS Analytics Accelerator for Teradata was used. As a result the original SAS Execution Time of 1hour and 19 minutes went down to only 16 seconds on an 8 Node Teradata 2580. This means using SAS Analytic Accelerator could get the data processing done in 1/296<sup>th</sup> the time, i.e. 4,739 seconds / 16 seconds = 296 Times Faster.

```

333 proc score
334 data=td.score01
335 score=td.PARAM_USCC_90_02
336 out=td.score02(dbcreate_table_opts='primary index(unique_id)')
337 type=parms
338 ;
339 by modell;
340 var &var;
341 run;

NOTE: SQL will be used for in-database computation for the data set OUT=TD.score02.DATA.
NOTE: SQL reports that a total of 5.7503e+06 rows were inserted into the data set OUT=TD.score02.DATA.
NOTE: PROCEDURE SCORE used (Total process time):
      real time          15.87 seconds
      cpu time           0.03 seconds

```

**Display 1. Performance improvements of Analytics Accelerator, Teradata 2580 8 Node (288 AMPs)**

## PRODUCT OVERVIEW

SAS Analytics Accelerator for Teradata enables in-database analytics. It improves data governance on predictive analytics projects and produces faster, better results. An overview to this product can be found at

<http://www.sas.com/software/high-performance-analytics/in-database-processing/accelerator-teradata.html#section=3>

SAS Analytics Accelerator for Teradata delivers a set of SAS' core analytical procedures from the SAS/STAT and from SAS/ETS module as in-database capable procedures.

- SAS/STAT<sup>®</sup> in-database versions of the following procedures:
  - CORR (correlation).
  - CANCORR (canonical correlation).
  - FACTOR (factor).
  - PRINCOMP (principal components).
  - REG (regression analysis, including stepwise regression).
  - SCORE (scoring of linear models).
  - VARCLUS (group variables into clusters).
- SAS/ETS<sup>®</sup> in-database version of the following procedure:
  - TIMESERIES (analyzes time-stamped transactional data and aggregates the data into a time series format for trending and seasonal analysis).

Additionally, EM Sampling macro generates custom SQL - %EMIDBSAM, and the EM binning macro calculations - %EM\_INDBIN. Consequently, the EM data source wizard can create sample using in-database processing.

- SAS<sup>®</sup> Enterprise Miner<sup>™</sup> in-database versions of the following macros and procedures:
  - Sampling macro (simple random, stratified and oversampling).
  - Binning macro (quantile binning with Weights of Evidence calculation).
  - DMDB (data mining database model effects summarization).
  - DMINE (variable selection).
  - DMREG (linear and logistic regression analysis, including support for stepwise regression).

SQL generation options specify the type of in-database computing (see SAS Products & Solutions / in-database processing):

```
libname teralib teradata
  server=tera5500
  user=edutest pw=edutest1
  database=dsb_result;

                                /* Options that trigger in-database process */
options SQLGENERATION=DBMS MSGLEVEL=I sastrace=',,,'d'
sastraceloc=saslog nostsuffix;

proc freq data=teralib.ads_3;
  tables churn_flag*age;
  where subscriber_id=120444;
Run;
```

- NONE (specifies that no in-database computation be performed).
- DBMS (specifies the SAS procedure to be used for in-database processing when possible):
- Uses conventional SAS processing when the specific procedure statement and options do not support in-database processing.
- ALL (specifies that in-database computation be performed whenever possible).

These options can be specified as either a LIBNAME statement option or as a system option in an OPTIONS statement.

It is important to note that SAS provides seven Base SAS in-database procedures (FREQ, RANK, REPORT, SORT, SUMMARY, MEANS and TABULATE) through the use of the SAS/ACCESS interface for Teradata software that can be used together during with the procedures listed in the table above the development of an analytical model. These Base SAS in-database procedures, with the exception of the CORR procedure, assume SAS/ACCESS to TD in place but do not require the presence of SAS Analytics Accelerator for Teradata.

Cf. Table 1 for an overview of available in-database enabled SAS PROCs).

Product	Module	In-Database enabled SAS PROCs
SAS/ACCESS for Teradata	Base SAS	FREQ, MEANS, RANK, SUMMARY, REPORT, SORT, and TABULATE
SAS Analytics Accelerator for Teradata	SAS/STAT	CORR, CANCELL, FACTOR, PRINCOMP, REG, SCORE and VARCLUS
	SAS/ETS	TIMESERIES
	SAS EM	DMDB, DMINE, DMREG

**Table 1. In-Database enabled SAS PROCs**

The following table depicts capabilities and supporting technologies of the SAS and Teradata integration products, including the most recent release of the SAS Analytics Accelerator for Teradata (Cf. Display .)

SAS PROC	SAS Language Elements	Teradata Capability Direct Support	Analytical Process
<b>SAS/ACCESS Interface to Teradata 9.3 (M2)</b>			
<u>Base SAS:</u> FREQ, RANK, REPORT, SORT, SUMMARY, MEANS, TABULATE	<ul style="list-style-type: none"> <li>SAS Formats</li> <li>SQL Pass-Through</li> </ul>	<ul style="list-style-type: none"> <li>Fastload, Mload, TPT API, T pump, FastExport</li> <li>Query Banding / Query Interruption</li> <li>Temporal Data Types</li> <li>Temporary Tables</li> </ul>	Data Exploration
<b>SAS Analytics Accelerator for Teradata 2.1</b>			
<u>SAS/STAT:</u> CORR, CANCELL, FACTOR, PRINCOMP, REG, SCORE, VARCLUS <u>SAS/ETS:</u> TIMESERIES		<ul style="list-style-type: none"> <li>Vendor-Defined Functions</li> </ul>	Data Exploration Model Development
<u>SAS Enterprise Miner:</u> DMDB, DMINE, DMREG	<ul style="list-style-type: none"> <li>%EM_INDBSAM</li> <li>%EM_INDBIN</li> </ul>		

**Display 4. SAS and Teradata analytics integration navigator**

## A SIMPLIFIED BUSINESS USE CASE

Based on a specific use case from the telecommunications industry, we would like to demonstrate the capabilities and best practice implementation of in-database analytics using the SAS Analytics Accelerator for Teradata to build a churn prediction model. The case of churn prediction is a rather standard application but is used only as a vehicle to demonstrate the basic approach of how to work in the integrated SAS and Teradata analytic environment.

In the telecommunications sector, the identification of potential defectors is a very important task. Every telecommunication provider uses churn prediction for segment early warning for customer retention marketing. Segment early warning would be an application of churn analysis in campaign management. It basically is a combination of customer churn segmentation with customer value segmentation. The resulting sub-segments (e.g., Highest Value customer with Neutral Churn risk between 30%-70%) are used for differentiated retention offers to a telcos' customer base and help to reduce overall costs of customer retention.

At the same time churn prediction heavily relies on granular customer usage data and therefore the amount of data for any type of analysis is very large by nature.

The challenge in churn modeling is to minimize modeling latency to enable timely reactions to market changes, to optimize automated scoring processes in order to have churn scores available on a daily or weekly basis, and at the same time, to take advantage of granular and large amounts usage data in the form of call detail records (CDR) or signaling information. Moreover a series of additional customer behavior data is available for inclusion in the Analytic Data Set. It can include Premium Content (Downloads, Purchases, Music, etc...) and KB usage, just to mention a few.

This challenge directly translates to the requirement of having a well and tightly integrated data warehousing and analytics environment to minimize any type of latency between the environments, which historically have been only loosely integrated.

## THE BEST PRACTICE APPROACH

The end-to-end process to develop an analytical model to identify characteristics in the customer base that can predict a customer churning starts with a business specification. In the business specification, the analysts work closely with business representatives, for example from organizational groups like segment management or a retention office, to formulate the analytical "problem" at hand.

Based on a solid business specification, the analyst commences the model building work, typically starting with gathering and preparing the data (statistical data preparation) for his/her needs before actually developing and validating the model. Once the analyst settles on a model that he/she deems most sufficient to predict customer churn, this model will have to be deployed to an automated and governed IT process for the recurring scoring of new input data.

The following paragraphs will explain how the in-database enabled SAS Analytics Accelerator for Teradata procedures can be applied in the various phases of this model building process to predict customer churn.

## CREATE AN ANALYTICAL DATA SET (ADS)

The most essential element of information is the customer level analytical data set. A single data record in this data set is also referred to as the Customer Analytic Record (CAR.). A series of data preparation steps are usually required to create this record-level ADS; data needed includes contract information, customer contact information, recharging for pre-paid, and other behavioral details including service usage. Behavioral information can be found in CDR and rated CDR data. Such usage information, gathered and transformed on a record-level for each customer entity, typically includes the number of calls or minutes of use, ratio of week-days versus weekend calls or similar metrics. The granular CDR data on call level is typically aggregated up to a weekly or monthly granularity before used in model development.

We recommend the use of the Teradata Data Set Builder for SAS (DSB) for these tasks of the model building process. The behavioral data aggregated from CDR's is combined with contract, recharging, and customer contact data. Everything is available in full history and detail on the Teradata warehouse. The aggregation and combination steps are executed directly in Teradata in a minimum of time typically resulting in ADS for all active lines comprising several hundreds of attributes ready to be used in churn modeling. The combination of data is typically implemented as joins and union on Teradata. The SAS PROC SORT, when enabled for in database processing, will heavily profit of the acceleration and would be translated to select ... order by. You can see Display 1 earlier for the resulting improvements in performance.

One Teradata best practice is to 'historize' the resulting ADS. Historization means ensuring that all relevant changes of customer behavior are tracked and recorded and that proper timestamps are assigned to these changes. Generally taken, historization is a prerequisite to achieving the goal of establishing the data warehouse as a corporate memory. The concepts for historization on the ADS level implies that all behavior is tracked over a certain time. In the modeling process, the historized ADS is translated into ABT (Analytic Base Tables) to be used for the actual modeling.

### Source Data Exploration

Prior to building the ADS, the analytic user would work against tables in a customer database, either a standalone data model or a domain within an enterprise data model inside the Teradata database, for all customers in the active population. Based on the understanding of the source data, the analytic user would then move to the next step, building the ADS as input for developing the analytic model.

With SAS/Access to TD the performance of explorative analysis of source data improves significantly. Assume that you would like to collect a statistics of minutes of use by weekday. PROC FREQ will come up with the required information and will push down the heavy load of summarizations and aggregations to Teradata without having to change anything in your preferred SAS code.

In the following code example the effect of enabling in-database processing is becomes clear when looking into the command log:

Source code:

```
options SQLGENERATION=DBMUST ;

PROC MEANS DATA=teralib.ads_3;
  VAR minutes;
RUN;
```

Command Log.

```

TERADATA_5: Executed: on connection 3
  select COUNT(*) as "ZSQL1", COUNT(*) as "ZSQL2",
         COUNT(TXT_1."minutes") as "ZSQL3",
         MIN(TXT_1."minutes") as "ZSQL4",
         MAX(TXT_1."minutes") as "ZSQL5",
         SUM(TXT_1."minutes") as "ZSQL6",
         COALESCE(VAR_SAMP(TXT_1."minutes")*(COUNT(TXT_1."minutes")-1),0) as
"ZSQL7" from "SAS_WORKSHOP"."ads_3" TXT_1
TERADATA: trget - rows to fetch: 1
TERADATA: trforc: COMMIT WORK
NOTE: There were 1 observations read from the data set TERALIB.ads_3.
TERADATA: trforc: COMMIT WORK
NOTE: PROCEDURE MEANS used (Total process time):
      real time          1.18 seconds
      cpu time           0.15 seconds

```

The same applies to PROC RANK, PROC REPORT, PROC SORT, PROC SUMMARY, PROC MEANS, and PROC TABULATE (cf. Display , SAS/ACCESS Interface to Teradata). All these procedures are typically used to get a first overview of how your customers behave in general and with a simple 'by' step you can contrast churner v. non-churners already in this early step of the analysis.

### Building the ADS In-Database

After having explored the source data for mediated call data, billing data, and recharging data Teradata Data Set Builder for SAS (DSB) can further be used to transform this source data to the CAR at subscription level granularity, with calling, information about a customer recharging his phone, and many more attributes for model development; there can be multiple subscriptions for a single customer. For churn prediction modeling, active subscription CARs as well as churned subscription CARs are captured in the ADS for the churn prediction modeling activities.

NOTE: Of course there is also the option to work with a graphical interface to the Teradata database. The DSB statistics and exploration capabilities can be used for the Initial data exploration and to support the construction of data preparation processes building the ADS directly in Teradata.

DSB offers a handy and simple to use tool to find out all available Teradata functions for data exploration (cf. DSB ADS and DSB analysis – statistical analysis).

After the modeling process the deployment of a churn model will run only against active subscriptions and will exclude inactive subscriptions. Besides only requiring a subset of predictors this is another best practice reasons for having separate DWH processes in place: One for the Modeling ADS v. a separate one for the Scoring ADS.

Typically, a CAR contains approximately 100 to 300 attributes for each subscription, derived from a semantic analytic data layer in the Teradata database that could have several thousands of enterprise-wide modeling attributes implemented for a single organization.

SUMMARY: It is good practice to create an ADS for the full modeling population (e.g., all active subscriptions) on Teradata with DSB, then do initial explorations and transformations using DSB on the full population, create the semi-final ADS for the full population, and access it with SAS to create the (sampled) ABT in SAS.

## MODEL DEVELOPMENT

There are several data mining methodologies on the market which are, like SAS' own SEMMA methodology, recommended sources for more process-specific details. In this section we will focus on demonstrating how the SAS Analytics Accelerator in-database procedures can be applied in the model development step of such a process. Please follow the link to find SAS documentation and material on SAS Analytics Accelerator.

<http://support.sas.com/documentation/onlinedoc/analyticsaccel/teradata.pdf>

These SAS procedures and SAS macros from the available spectrum of in-database capabilities (see the analytic navigator in Display ) will be used in the following example of churn prediction modeling: FREQ, MEANS, SUMMARY, PRINCOMP, REG, DMREG, DMINE, %EMIDBSAM, %EM\_INDBIN and SCORE.

### Model Data Exploration

Model data exploration can either be done using point-and-click interfaces in SAS Enterprise Guide or SAS Enterprise Miner or manually coding SAS statements from within these client tools, or typically by a combination of all three.

By using the in-database SAS procedures SUMMARY, FREQ and MEANS the analyst will further explore the ADS at this stage to understand the behavior of attributes across subscriptions or groups of customers within the population to identify the expected behavior but also unexpected patterns in subgroups; especially by using the FREQ analysis with stratified analysis by computing statistics across as well as within churner versus non-churner groups of subscriptions. SUMMARY, MEANS and FREQ all compute descriptive statistics for variables across all CAR observations or within groups of CAR observations in the ADS.

In addition, when working in SAS Enterprise Miner, the macro %EM\_INDBIN performs in-database quantile-based binning of all interval variables in the ADS residing in the Teradata database.

### Model Data Preparation

In the model data preparation step, the analyst will use the customer or related subscription behavior derived from the data in the ADS to refine existing attributes or by generating additional attributes into the ADS using statistical derivations. A good example is principal component analysis (PCA) techniques using the in-database PRINCOMP procedure. Depending on research objectives, a similar technique would be a factor analysis, also available as in-database FACTOR procedure.

The PRINCOMP procedure computes a principal component transformation of the SSCP matrix (sum-of-squares and cross-product matrix) into orthogonal components. The first component accounts for the maximum amount of variation; the second accounts for the next largest amount of variation; and so on. This procedure is typically used in exploratory data analysis and visualization. Principal component analysis reduces the dimensionality of data for predictive modeling by replacing the variables in the original data with fewer principal component terms for the final model. For the in-database version of PRINCOMP, the SSCP matrix is computed in-database and is downloaded to the SAS system for the remaining steps of the procedure processing.

Sample code for the in database enabled principal component analysis:

```
PROC PRINCOMP data = teralib.ads_3;
RUN;
```

The results of in-database processing is presented in Display 2. The essential information can be found in the NOTE on the bottom saying that 'The SQL Query for in-database computation of the SSCP matrix was built, submitted, and ran successfully'.

```
NOTE:
NOTE: The SQL query for in-database computation of the SSCP matrix was built, submitted, and ran successfully.
NOTE: SQL will be used for in-database computation for the data set DATA=TD_RESUL.ADS_3.DATA.
NOTE: The PROCEDURE CORR printed page 1.

37
38      /* -----
39      End of task code.
40      ----- */
41      RUN; QUIT;
```

### Display 2. Log of PRINCOMP

A typical application in churn prediction modeling would be to reduce a group of usage behavior attributes (e.g., three months, 6 months, 12 months average number of calls) with factor analysis. As a result we can use a reduced number of attributes for modeling which has a number of advantages (statistical and practical ones) over using a larger number of attributes.

Alternatively the modeler can use the in-database capable DMINE procedure to run through this step in a more automated fashion. Many data mining databases have hundreds of potential model inputs. The DMINE procedure, a component of SAS Enterprise Miner, allows quick identification of input variables that are useful for predicting the target variables based on a linear models framework. The procedure facilitates ordinary least squares or logistic regression methods.

NOTE: With the SAS Analytics Accelerator deployed, it is possible and also recommended to implement model data preparation steps and resulting variable derivations directly on the full ADS inside Teradata, without previous sampling or data offloading.

### Model Data Sampling

After completion of the steps above, the analyst will create data samples to be used for the model development. This is different from sampling data earlier in the process, which is not required as stated above. Even if we could apply

the analytical procedures, when developing the model, on the full set of data, it is recommended and common practice to create statistical samples.

We use statistically valid samples, either to support the modeling process itself (for example oversampling of the target variable, if required) or to create validation data sets. The sampling will be applied directly to the Teradata in-db ADS resulting in one or several sampled data sets (ADS's or ABT's, only according to the personal preferences of an individual modeler.) Nevertheless it is recommended to use a Data Lab approach inside Teradata for better security and improved data governance reasons.

A Teradata Data Lab is a solution that provides SAS users immediate access to required data for explorations and examinations. A Teradata® Data Lab enables agile analytics by simplifying the provisioning and management of analytic workspace within the production data warehouse. By allocating that space, it provides data lab users easy access to critical production information without moving or replicating data ([www.Teradata.com](http://www.Teradata.com).)

In our case of churn prediction modeling, any of these smaller ADS tables is a sample of customers from the original ADS table. The samples can be created either using the sampling functionality of Teradata (using SAMPLE function) or using the DSB sample node.

If the sampling is done only once, for a single model, the analyst could also use the in-database SAS macro %EMIDBSAM, part of the SAS Analytics Accelerator software in support of SAS Enterprise Miner. The macro %EMIDBSAM performs in-database sampling for a data table that resides in a relational database. The sample rows are output to a SAS data set. The sampling methods include simple random sampling, top-n sampling, and stratified sampling.

### **Model Construction**

Regression is a very robust method and we see it very often used for building churn prediction models. To be used by regression analysis, a certain number of churned lines are combined with a number of lines still active in order to prepare modeling.

The analyst can use the in-database procedures DMREG or REG. The DMREG procedure enables to fit both linear and logistic regression models. Linear regression attempts to predict the value of a continuous target as a linear function of one or more independent inputs. Logistic regression attempts to predict the probability that a categorical (binary, ordinal, or nominal) target will acquire the event of interest as a function of one or more independent inputs. In our example of churn prediction modeling the Logistic regression will be used in case when a categorical target attribute is used (note that this is the predominant case). The procedure supports forward, backward and stepwise selection methods. It also enables to immediately score a data set or generate SAS DATA step code to score a data set later on.

The REG procedure computes a model in which the dependent churn target variable is modeled as a linear equation of multiple independent variables by a least squares function. We might use REG, e.g., when we decide to define the churn target attribute as percentage decrease in usage. The REG procedure has numerous model-fitting options, which include many model selection variations and hypothesis tests. Regression analysis is often used to identify a subset of independent variables that have unique and significant relationships with the dependent variable. In this case, the REG procedure is used for its exploratory analysis, in addition to its extensive model fitting and reporting functions.

As a result of this core model development step the analyst defines the churn prediction model, to be deployed for regular scoring (on a set frequency) of newly incoming data inside the Teradata database.

### **Model Deployment**

Optimized in-database scoring is required in the deployment step of the churn prediction modeling. Deployment of prediction models in the customer retention business process requires recurrent scoring of the active subscriber base with the selected churn model. Model Development and Model Deployment are separate processes and specific tools are available from SAS side to support the development of optimized DWH processes around either of it.

We usually refer to the model construction process or Model Building to be the SAS Analytic Accelerator process (AA). Deployment into Production is a different task and SAS Scoring Accelerator (SA) for Teradata is available to specifically support the deployment process.

Nevertheless deployment with AA is possible to some extents. For DMREG, i.e., when using SAS EM with Scoring Accelerator we always have the option to publish churn models to create Teradata User Defined Functions. Alternatively we can create a scoring SAS embedded process (EP) for recurrent churn scoring.

Without going into details on either of these options, we do wish to highlight in-database scoring using PROC SCORE. It is a way to deploy in-database scoring in cases where we want to score, for example, PROC PRINCOMP or PROC FACTOR which actually would happen outside Analytics Accelerator.

So, in case we used PROC REG for the churn prediction (cf. previous section on modeling) the SCORE procedure could be used to apply the regression coefficients estimated using PROC REG to a scoring data set and compute the churn score for active subscriptions (generically called scores.) For example, you can use the SCORE procedure to produce churn scores per subscription in the form of a table in Teradata that contains predicted churn risk and/or residuals from a model that is estimated using the REG procedure in-database. The PROC SCORE procedure dynamically generates SQL code for the given model. PROC SCORE then submits the SQL code to the database, which produces a Teradata table without having to extract any rows of data into the SAS System.

## **FUTURE DIRECTION**

SAS is committed to providing solutions which streamline the modeling process in tandem with traditional databases. Over the coming months, new products will be released. SAS In-Database Code Accelerator for Teradata enables customers to run SAS DS2 code inside Teradata. Data preparation and manipulation will speed up and take advantage of the massively parallel processing power of Teradata. Data Quality Accelerator for Teradata allows users to execute SAS data quality functions inside Teradata.

## **SUMMARY**

This paper gives an overview of how to work with SAS Scoring Accelerator and SAS Analytic Accelerator in an integrated SAS Teradata analytical environment. The central portion of it is briefly describing how the various SAS procedures may be used in data prep, statistical analysis, modeling, and deployment. The integration preserves all existing SAS and Teradata standards.

The analytic navigator and summary tables are intended to find a quick overview of what is recommended to use for performing modeling within Teradata, including options coming with SAS Analytics Accelerator. Following the outlined path the analysts will reduce data movements and model latency.

## **ACKNOWLEDGMENTS**

With valuable review notes and recommendations by James Stewart (Teradata), Mike Rote (Teradata), and Nicolas Adamek (Teradata).

## **CONTACT INFORMATION**

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

Name: Karl Krycha  
Enterprise: Teradata  
Address: Storchengasse 1  
City, State ZIP: 1150 Vienna, Austria  
Work Phone: +43 (0) 664 32 40 195  
E-mail: karl.krycha@teradata.com

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.