

Teradata Analytic Data Set Generator

Deliver analytic solutions faster and under budget:

- > Accelerate your analytic environment with in-database technology and automation.
- > Improve performance of your models with Teradata.
- > Build better models with access to centralized, enhanced analytic data.
- > Deliver a comprehensive model with data refresh for real-time scoring.

A Recipe for Success

The wrong ingredients can ruin a meal, just as incomplete, inaccurate, or irrelevant data can cause your analytic model to fail. For example, if your customer value model mistakenly assigns a low score to profitable customers because of missing transactions, you could lose high-value customers. That's why you must begin with the most important phase of the data mining process: data preparation.

Typically, data in a relational data warehouse aren't suitable for input directly into data mining algorithms. New predictive variables need to be created using formulas, aggregations, and/or expansions on specific values of a dimensioning variable. Joining tables and/or de-normalizing or flattening relational tables may also be necessary. And, statistical transformations are often required because many algorithms require data in a specific format or range. These are some of the tasks required to create an analytic data set needed to build your predictive model.

Because data preparation is an iterative and data intensive process, data miners can spend weeks, if not months, exploring and preprocessing data to create analytic data for analysis. Unfortunately, most data mining tools don't provide the functions to efficiently support exploration and preprocessing of large data volumes – a primary pain point for data miners. And with the fast paced changes in technology and marketing campaigns combined with pressure from your competitors, you don't have the luxury of waiting to prepare data for analytic modeling. Yet your models may be useless unless you give proper attention to your data.

Faster, Smarter Analysis

But now, the Teradata® Analytic Data Set (ADS) Generator offers an effective tool for data preparation – an in-database technique to prepare data for data mining. Its software helps you move the necessary functions and processing directly into your database reducing data movement and redundancy. Not only does Teradata ADS

Generator speed up the analysis by leveraging the parallel database engine to perform analytic functions, it also provides features to facilitate and automate portions of data preparation tasks. With a point-and-click interface, you can analyze data in multiple tables quickly and efficiently, with results returned in easy-to-read tables and graphics. What does this mean? The data exploration and preprocessing steps that once took weeks now can be done in days.



Build Better Models Faster

Teradata ADS Generator is a data processing tool, not a data mining technique. It speeds the data mining process by moving repetitive, data intensive, data preparation functions directly in the Teradata Database. As a result, the Teradata ADS Generator delivers an optimal analytic data set for analysis using a variety of analytic tools.

Teradata Analytic Data Set Generator

The Teradata ADS Generator complements all other analytic tools. By assisting in data preparation, you can explore your data warehouse, merge, join, and sort tables, create new variables, transform data in the database, and export a sample of the analytic data to your favorite analytic tool(s) for further analysis.

As data mining tool vendors adopt in-database scoring with SQL and UDFs (User Defined Functions), the Teradata ADS Generator was enhanced to be UDF aware and enable SQL import. These two functions not only automate deployment of analytic models and access to analytic functions represented as UDFs or SQL, but also link the models to the detailed data in-database. This allows analysts to create a single script that converts detailed data to analytic insights enabling real-time scoring.

By building an ADS in your Teradata system, you can:

- > Build a comprehensive ADS that can be shared to build multiple models. Build once, use many.
- > Analyze data faster and more efficiently. Joins and aggregation that took hours can be done in seconds.
- > Minimize cost and errors attributed to data movement.
- > Focus analysts on business analytics, not the mechanics of data movement.

- > Create a collaborative environment by sharing analytic data, results, and metadata throughout your enterprise.
- > Automate the deployment of data and models.

Teradata ADS features that optimize analytic environments include:

Data Explorer with drill down analysis – automates profiling of multiple variables across multiple tables within a single analysis. Built-in intelligence determines which analysis to perform based on the data type, number of unique values, and optional thresholds set by the user. Results of analysis for each variable are displayed in a gallery of thumbnail graphs that can be expanded for detailed analysis. In addition, with a simple mouse click, the drill down feature retrieves records represented by the graph or a segment of the graph for detailed analysis. So not only can you determine which variables contain suspicious data, but you can examine the data, too.

Variable Creation – allows analyst to create new predictive variables by generating optimal SQL scripts for their Teradata system. Creating new variables is simplified through a visual SQL generator where you can select data elements and apply a group of functions and UDFs with a mouse click. In addition, this feature retrieves values of a particular variable and inputs the values into conditional statements, automating repetitive tasks. With this feature, anyone with some SQL knowledge can generate optimal SQL

queries comparable to those done by a Teradata Master

Variable Transformation – most data mining tools require data to be transformed into a specific format. For example, continuous variables may be binned or categorical values may be converted to numeric values. This feature navigates users through the transformation process with an array of techniques including bin coding, design coding, recoding, rescaling, null replacement, statistical transformation, user defined calculations, and date/time transformations.

ADS Build and Refresh capabilities – the ADS is typically a very large, flat table with hundreds or thousands of variables built by creating, aggregating, and combining data from multiple tables. The Build ADS function is designed to create the ADS by combining variables from multiple tables into a single table or view. All the analysis and results are linked together into one project to automate the deployment process. Refresh ADS automates the rebuilding process when moving into a production environment.

PMML Consumer – Teradata ADS Generator supports an industry modeling standard called Predictive Model Markup Language (PMML) that allows applications or tools to share analytic models. Through PMML, models developed with partner tools can be imported into Teradata ADS Generator for in-database scoring, linkage to the ADS development, and integration into Teradata Model Manager.

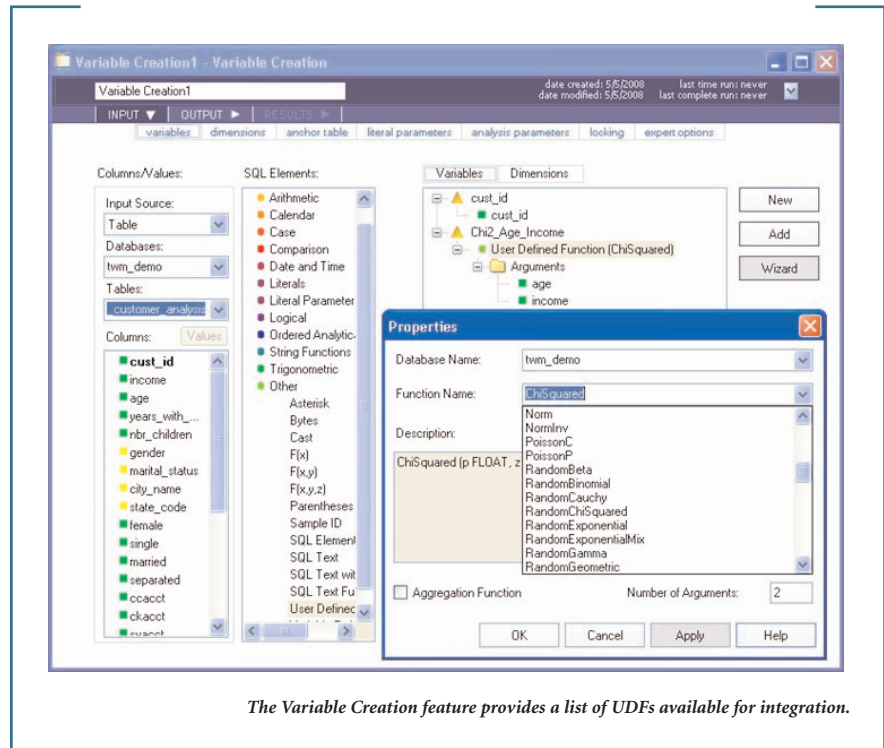
Teradata Analytic Data Set Generator

Teradata.com

ADS and Model Publisher – Introduced with Teradata Warehouse Miner 5.0, this feature allows users to link Teradata Warehouse Miner SQL models and partner PMML or UDF models to an ADS, automating deployment. Typically all the variables in the ADS used to develop the models are not required for the production model. The publishing feature will recreate the ADS with only the variables required for the production level model. This automates the deployment process and also eliminates the extra baggage not required for production.

Teradata Model Manager – This is a new product that allows the enterprise users to refresh analytic data and run models to meet their business requirements. Model manager provides a list of analytic data script and models that they can refresh and run against a particular segment of their data or the entire set. This interface provides the flexibility to generate scores as needed, but also security and controls offered through the Teradata Database.

Batch Interface – This feature opens Teradata ADS Generator functionality to partner products and user applications by exposing functions in a batch interface. The user creates an XML file to invoke any



The Variable Creation feature provides a list of UDFs available for integration.

pre-built ADS project or creates new analysis for data exploration and refreshing your analytic data set.

The Teradata Difference

Teradata Corporation has 25 years of experience building data warehouses where data quality is a critical core competency. This experience and our in-depth industry knowledge help us provide the

right solution for you. And Teradata technology provides efficient, scalable solutions for analyzing data in your Teradata Database.

For More Information

To maximize your data mining capabilities and streamline your data preparation steps, contact your Teradata representative or visit Teradata.com.

Teradata is a registered trademark of Teradata Corporation and/or its affiliates in the U.S. and worldwide. Windows is a registered trademark of Microsoft Corporation. Teradata continually improves products as new technologies and components become available. Teradata, therefore, reserves the right to change specifications without prior notice. All features, functions, and operations described herein may not be marketed in all parts of the world. Consult your Teradata representative or Teradata.com for more information.

Copyright © 2004-2008 by Teradata Corporation All Rights Reserved. Produced in U.S.A.