

# Teradata Warehouse Miner



Make the most of your data warehousing capabilities with analytic technology that:

- > Reduces the analytic modeling development cycle, allowing faster iterations to refine your model and increase analytic intelligence.
- > Simplifies data profiling and creation of an analytic data set with built-in intelligence, complementing any data mining tool.
- > Allows you to quickly and easily integrate models into business applications.
- > Accelerates model scoring allowing you to analyze and score data in your warehouse efficiently.

## Take the Next Step

You've successfully implemented a data warehouse. Now here's the real challenge: maximizing the value and knowledge locked inside your data. But how do you turn data into predictive insight?

If you're like other advanced businesses, you turn to data exploration and data mining to discover meaningful new data

patterns and trends. Data exploration analyzes the data to identify relationships and anomalies that affect business decisions. Data mining sifts through huge amounts of detailed data stored in your data warehouse using pattern recognition technologies and mathematical modeling techniques. What are the biggest benefits of data exploration and data mining? They discover hidden data patterns that aren't obvious to humans or OLAP tools – patterns that can lead to actionable business decisions to drive greater profits.

But many tools developed to analyze data were not created with today's large data volumes in mind. And they require you to extract larger amounts of data from your data warehouse and move it to a specialized tool that implements data mining. So, as data volumes expand, how do you keep pace with scalability issues that may limit the potential of your warehouse? You can start with Teradata® Warehouse Miner.

## Keeping Pace with Your Business

Teradata Warehouse Miner provides an array of data profiling and mining functions ranging from data exploration and transformation to analytic model development and deployment that are performed directly in Teradata Database. While most data mining solutions require analysts to extract data samples to build and run analytic models, the Teradata solution allows you to analyze detailed data without data movement, streamlining the data mining process.

In addition to the performance and scalability advantages of in-database mining, Teradata Warehouse Miner brings a new level of efficiency to your analytic development process by automating and simplifying the creation of an analytic data set. Typically, the data in your warehouse are not suitable for data mining algorithms. First, new predictive variables and transformations must take place to condition your data. Creating this analytic data set (ADS) from which to build an analytic model can be the most intensive part of the data mining process. Teradata Warehouse Miner's ADS capabilities facilitate variable transformations and variable creations by capturing best practices from our Advanced Analytic Center of Expertise and building this capability into the software. What used to take months can now be accomplished in days with Teradata Warehouse Miner.

## Optimizing Analytics with Teradata

Teradata Warehouse Miner can also optimize your existing analytic environment by integrating the efficiency of in-database mining with your current analytic tool. This integrated environment delivers a flexible, optimized environment where you can perform data intensive steps, such as data preparation, directly in Teradata Database. This analytic data can then be exported into any data mining tool for experimentation and modeling. If the model is represented in PMML (Predictive Model Markup Language), the industry standard for analytic models,

# Teradata Warehouse Miner

<p><b>Descriptive Statistics</b></p> <p><b>Frequency</b> Computes frequency of column values or multi-column combined values.</p> <p><b>Histogram</b> Determines the distribution of a numeric column(s) giving count with optional overlay and statistics. Additional options for removing spikes and sub-binning.</p> <p><b>Overlap</b> Count overlapping column values in combination of tables to determine cross-table data integrity.</p> <p><b>Scatter Plot</b> Graphically displays data based on the relationship among two or three variables.</p> <p><b>Statistical Analysis</b> Computes minimum, maximum, and mean values, standard deviation, skewness, kurtosis, and many more univariate statistics.</p> <p><b>Values Analysis</b> Counts different values for a given variable(s), such as null, unique, and negative values.</p> <p><b>Data Explorer</b> Explores variables across multiple tables with built-in intelligence to determine the best analysis to run on data elements based on type and results from preliminary analysis.</p> <p><b>Correlation Analysis</b> Calculates the Pearson-Product Moment Correlations of any number of numeric variables.</p>	<p><b>Kolmogorov-Smirnov</b> Kolmogorov, Smirnov, K-S, Lilliefors, and Shapiro-Wilk tests</p> <p><b>Parametric</b> Two Sample T-Tests and One-Way/N-Way ANOVA</p> <p><b>Data Visualization</b></p> <p><b>Descriptive Statistics</b> Data Explorer Graphics, Proportional/Absolute Values Box Plots, Box and Whiskers Plots, Histograms, Frequency Bar Charts, and Scatter Plots. Drill down on records represented by a graph or segments within the graph from descriptive statistics analyses.</p> <p><b>Linear Regression</b> Coefficients and T-Statistics Box Plots</p> <p><b>Logistic Regression</b> Coefficients, T-Statistics, Wald Statistic, Log Odds Ratio, Partial R Box Plots and Lift Chart</p> <p><b>Factor Analysis</b> Factor Patterns and Screen Plot</p> <p><b>Decision Tree</b> Tree Browser and Lift Chart</p> <p><b>Cluster Analysis</b> Size and Distance and Similarity charts</p>
<p><b>Transformation and Data Reduction Functions</b></p> <p><b>Variable Creation</b> Simplifies variable creation through a visual SQL generator where you can select data elements and apply a group of functions with a mouse click. Any valid Teradata SQL keyword can be generated, including Aggregate functions, Ordered analytical functions, Arithmetic operators and functions, Trigonometric and Hyperbolic functions, CASE expressions, Comparison operators, Logical predicates, NULL operators, Calendar fields, String functions, Type conversion, and SQL predicates and User Defined Functions.</p> <p><b>Variable Transformation</b> Performs multiple data transformation functions on variables within a single pass, streamlining the entire transformation process. Transformations supported include Bin Coding, Design Coding, Recoding, Rescaling, Statistical Transformations, User-Defined Derive, NULL Value Replacement, and Date/Time Transformations.</p> <p><b>SQL Node</b> Attach any SQL code to the data mining project.</p> <p><b>Data Reduction Matrices</b> Calculates correlation, covariance, Sums of Squares and Cross Products (SSCP), and Corrected SSCP values of pair-wise combination for each variable.</p> <p><b>Build ADS</b> Joins together intermediate results tables or views into a development analytic data set.</p> <p><b>Refresh ADS</b> Generates a subset of SQL from the development ADS to create the production ADS.</p>	<p><b>Organization and Partitioning Functions</b></p> <p><b>Denorm</b> Create new denormalized table by removing index column(s).</p> <p><b>Join</b> Join tables or views into a combined results table.</p> <p><b>Sample</b> Select sample(s) from a table by size or fraction with options to stratify.</p> <p><b>Partition</b> Select partition(s) from a table using a hash key.</p> <p><b>Merge</b> Appends rows from tables or views by using UNION and/or MINUS.</p> <p><b>Analytic Algorithms</b></p> <p><b>Linear Regression</b> Estimate a numeric data value based upon linear combinations of other variables.</p> <p><b>Logistic Regression</b> Predict a binary discrete data value based upon multivariate analysis of other variables.</p> <p><b>Factor Analysis</b> Combine variables into factors using a variety of multivariate techniques.</p> <p><b>Decision Trees</b> Predict a discrete or continuous data value using rules-based techniques.</p> <p><b>Clustering</b> Search for groups of like characteristics using distance-based techniques.</p> <p><b>Association Rules/Sequence Analysis</b> Describes the relationship among items in a group or the sequence of items within a group over time.</p> <p><b>Neural Network Add-in</b> Predict a discrete or continuous data value using non-linear techniques against other variables; powered by <i>STATISTICA</i>.</p>
<p><b>Statistical Tests</b></p> <p><b>Binomial</b> Binomial and Sign tests</p> <p><b>Chi Square</b> Chi Square and Median tests</p> <p><b>Rank</b> Mann-Whitney, Kruskal-Wallis, Wilcoxon, and Friedman tests</p>	<p><b>Model Management and Scoring</b></p> <p><b>Model Manager</b> Easy to use, web-based interface for analytic data refresh and on-going model scoring and management.</p> <p><b>SQL-Based Models</b> Publish any SQL code including models for easy integration into business applications.</p> <p><b>PMML Consumer</b> Import third-party PMML models for execution in-database.</p> <p><b>Analysis References/Publishing</b> Link the output of one analysis as input to another enabling complete analytic data refresh and model automation.</p>

# Teradata Warehouse Miner

Teradata Warehouse Miner can import and execute the model in the Teradata Database for optimal performance.

## Deploying Models into Your Business Process

The model development and deployment processes are typically done in different environments. During the model development process, you want your analysts to have access to all the data required to build an accurate, robust model. You need a flexible environment with tool choices that allow analysts to experiment with new ideas.

The production environment is very different and requires controls, efficiency, performance, and scalability. Production models need to be efficient to process large volumes of complex data. In addition, your production analytic data must be fresh to ensure your models run against current data.

Teradata Warehouse Miner automates this previously manual process by recoding the analytic data and models for you. The ADS Refresh feature trims down the development ADS that typically includes hundreds or thousands of potential predictive variables to only the variables required for your model.

Teradata Database does this by linking

UDF, SQL or PMML models to a development ADS, then it determines which variables are needed for the model and creates a new production ADS for your model. All the complexity is hidden, so any user with appropriate permissions can refresh the data and score the model to meet their business needs.

In-database technology from Teradata Corporation allows you to improve advanced analytics by:

- > Eliminating data movement – there's no need for multiple iterations of data extraction for sampling, exploration, model development, or scoring.
  - > Minimizing errors – eliminates potential human errors from sampling, and eliminates data movement across platforms, tools, and databases.
  - > Simplifying data management – provides a single source of data, eliminating inconsistencies and data redundancy.
  - > Leveraging parallelism – leverage the high performance and scalability of Teradata Database.
  - > Building accurate models – enables analysis of large volumes of detailed data.
  - > Speeding development to deployment cycles – provides quicker access to quality information and enables you to act faster on issues, such as fraud detection and customer attrition.
- > Sharing models and data – analytic data and models represented in UDFs, SQL, and PMML are shared across all enterprise users.
  - > Exporting models – SQL representing models can be exported and integrated into business applications for real-time use.
  - > Open analytic environment – complements any data mining tool by performing data intensive steps in-database, while offering the flexibility of using any analytic tool.
  - > Automate integration of advanced analytics into the business process.

## Why Choose Teradata?

Because we have a 25-year history of providing analytical solutions that solve real-world problems. Because we've given companies like yours the knowledge and technology to drive strategic and operational decisions and to create close, personal customer relationships. And because Teradata has built-in industry knowledge, consulting expertise, global support services, and world-leading hardware technology – a combination unmatched in the industry.

## For More Information

To learn more about Teradata Warehouse Miner, contact your Teradata representative or visit [Teradata.com](http://Teradata.com).

Teradata is a registered trademark of Teradata Corporation and/or its affiliates in the U.S. and worldwide. Microsoft and Windows are registered trademarks 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](http://Teradata.com) for more information.

Copyright © 2007-2008 by Teradata Corporation All Rights Reserved. Produced in U.S.A.