



Enterprise Miner Version 4.0

Changes and Enhancements

Table of Contents

- General Information 1
- Upgrading Previous Version Enterprise Miner Projects 1
 - Converting Single User Enterprise Miner Projects from Version 3.0x to Version 4.0 1
 - Converting Single User Enterprise Miner Client/Sever Projects from Version 3.0x to Version 4.0. 1
 - Converting Single User Enterprise Miner Projects from Version 2.0x to Version 4.0 2
- Metadata Sample 2
- Tilting 3D Plots 3
- Summary of Node Updates 3
- New Enterprise Miner Nodes and Node Features 5
- Warehouse Administrator Add-ins 5
 - DM Tool Application Program Interface (API) 5
- Online and Web-based Documentation 6

General Information

Enterprise Miner Version 4.0 can now handle longer names for tables and variables. You can use strings that are up to 32 character long and that include upper- and lowercase letters when you name the variables and tables.

Enterprise Miner Version 4.0 runs under the following operating environments:

- for clients
 - Windows 2000, NT, 98, and 95.
- for servers
 - Windows NT Server
 - Solaris
 - OS/390
 - Compaq Tru64 Unix
 - HP/UX
 - ABI+ for Intel Architecture
 - AIX/6000.

It is now easier to install Enterprise Miner and SAS software together. A new installation wizard enables you to concurrently install Enterprise Miner Version 4.0 with Release 8.1 of SAS instead of requiring separate installations.

Upgrading Previous Version Enterprise Miner Projects

You can convert existing Enterprise Miner projects from previous versions. In fact, a project's underlying SAS data sets are automatically converted from Release 6.12 to Release 8.1 when the project is converted.

Note: We **strongly** recommend that you back up the earlier versions of your projects before you upgrade! After converting an Enterprise Miner Version 3.x project to Version 4.0, do NOT revisit that project using Release 6.12 of SAS software. Although it will appear that you can add and delete diagrams, the project file will become corrupted. If you want to use separate versions of Enterprise Miner (Version 3.x and Version 4.0) on the same project, you should make a copy of the project files to use with Enterprise Miner Version 3.x and make a different copy for conversion and for use in Enterprise Miner Version 4.0.

Converting Single User Enterprise Miner Projects from Version 3.0x to Version 4.0

To convert an Enterprise Miner project from Version 3.0x, select **File→Open** from the main menu to open the project in Version 4.0, select **Yes**, and click on **OK** in the Conversion to V8 Required window.

Converting Single User Enterprise Miner Client/Server Projects from Version 3.0x to Version 4.0

Warning: Before beginning the project conversion process, you should close all Enterprise Miner sessions and make sure that no other users are connected to the project that you plan to convert.

After you are sure that your Version 3.x project files and the files in the project's remote server directory are backed up, start Enterprise Miner Version 4.x / SAS Version 8.1 (or 8.2).

1. The first step in upgrading your Enterprise Miner Version 3.x project is to create a new server profile. The new server profile must be able to successfully connect to the remote server running SAS Version 8.1 or 8.2. To create a new server profile, use the main menu to select

Options → User Preferences

When the User Preferences window opens, go to the Server profile tab. Click the New profile button and fill out the required settings. *(If you need more information on creating a server profile, see the Enterprise Miner online reference Help.)*

Help→EM Reference→Getting Started with Enterprise Miner 4.0→Working with Projects→Creating a Client/Server Project→Defining a Server Profile

After you define your new server profile, save your changes and close the User Preferences window.

2. Next, you need to identify the Enterprise Miner Version 3.x project that you want to convert. From the main menu, select

File→Open

then navigate to the version 3.x Enterprise Miner project folder that you want to convert. Double-click on the project folder to open it.

3. After you open the project folder, you should see one or more diagram files (diagram files end with *.dmd* extensions.) Select a diagram file by double clicking on it. If the diagram file is from an Enterprise Miner Version 3.x project, and if you are running Enterprise Miner Version 4.x, a message box opens and informs you that the project must be converted to SAS Version 8 format. Select the **OK** button to begin the project conversion process.

4. When the conversion completes, a message box asks if you want to connect to the server. Click on the **No** button.
5. Now you need to modify your project settings.
 - a) Remember to make sure that no one is using the project you are upgrading.
 - b) Go to the Diagrams tab of the Project Navigator (the Project Navigator is just to the left of the Diagram Workspace).
 - c) Right-click on the project that you are converting, then select the **Explore** pop-up menu item. This opens a Windows file utility window for the directory containing your project. The window's title bar displays the system path to the directory.
 - d) Double-click on the **Emproj** folder to open it. The **Emproj** folder contains files ending in *bcat*, *bdat*, *bview*, and a folder named **Users**.
 - e) Delete all but the most recent user entry inside the Users folder. You should be left with one file, the current user entry in the folder. Do not delete the Users folder itself.

Caution: Enterprise Miner will halt the project conversion if the Users folder is empty or if the Users folder contains more than one user file.
 - f) Close the file utility window and return to the Diagrams tab of the Enterprise Miner Project Navigator.
 - g) Right-click on the project again, and select the **Properties** menu item. This opens the Properties window for the project you are upgrading.
 - h) Go to the Server tab. The Server tab contains the project settings used in the old Enterprise Miner Version 3.x project file. Make a note of the **Default remote data path** and also save any code that is currently in the **Start-up code** field. You will need this information later.
 - i) With the Properties window still open, from the main menu, select
 - j) Edit the **Default remote data path** and **Startup code** fields, using the information you copied from the old project properties in the previous step. The path and startup code should be the same as before.
 - k) If you used a **Signon script** or **Start-up code** for this project, be sure to click on those buttons to ensure that the new project points to SAS 8.1 or 8.2 on the server.
 - l) After making your changes, click the **OK** button to close the Properties window.
6. After Next, connect the project to your remote server. In the Diagrams tab of the Project Navigator, right-click on the project and then select **Connect to server** from the menu.
7. After If the connection is successful, a message box informs you that the default remote data path that you specified in the Properties window may require conversion to SAS Version 8, and recommends that you back up the files in this directory before proceeding. Be sure you have a back up of the files in that path, then click the **Yes** button to proceed with the conversion.
8. After The last step in converting your project is to re-run all process flow diagrams in your project using SAS Version 8.1 or 8.2/ Enterprise Miner version 4.x.

Converting Single User Enterprise Miner Projects from Version 2.0x to Version 4.0

To convert an Enterprise Miner project from Version 2.0x, select **File→Import version 2 Project** and use the Import Wizard within Version 4.0.

Metadata Sample

In order to determine the measurement level and variable roles in a data set, Enterprise Miner examines the metadata sample. Two parameters can be used to control processing time and memory requirements for this step:

- The number of variables to be processed at a time.
- The maximum number of distinct values to be retrieved before terminating the processing of a variable. This value is also used to determine which numeric variables are set to interval variables.

In Enterprise Miner Release 3.01 and earlier releases, variables were examined in groups of 100 variables with a maximum of 2,000 different levels. In Version 4.0, those defaults have been changed to 2,000 variables with a maximum of 128 levels. Moreover, you can control these parameters by using the following macro variables:

File→Load server profile settings

and choose the name of the new server profile that you created a few steps ago. This loads the new server profile and updates the server profile settings.

- **DM_METAVARS** specifies how many variables to be processed at a time.
- **DM_MAXLEVEL** indicates to process variables until the specified number of distinct levels is reached.

For example, to build metadata by using default settings for a data set with 400 variables now requires only one pass of the metadata sample instead of four passes in Enterprise Miner Release 3.01 and earlier releases. Numeric class variables now are assumed to have no more than 128 distinct levels, as opposed to the 2,000 levels used in earlier versions.

Note: You can change the default metadata determined by Enterprise Miner at any time by changing the configuration in the **Input Data Source** node or in the **Data Set Attributes** node.

Tilting 3D Plots

In Enterprise Miner Version 4.0, you can tilt the following three-dimensional plots around the horizontal axis interactively by using the **Tilt** icon on the tool box:

- Cluster pie chart in the Partition tab of the Clustering and SOM/Kohonen Results Browsers.
- Profile grid plot in the Profile tab of the Clustering and SOM/Kohonen Results Browsers.
- Tree ring plot in the Tree Results Browser.

In the Chart tab of the Distribution Explorer Results Browser, you can select **View→Perspective→3D Depth View** and use the tool icon to manipulate the three-dimensional plot.

Summary of Node Updates

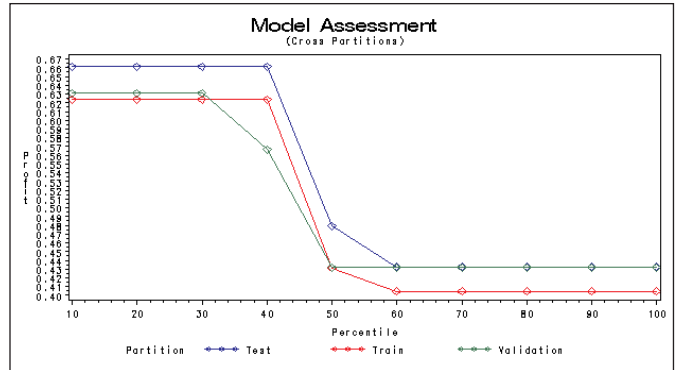


Assessment

In the HTML report generated by the **Reporter** node, the Model Assessment charts for all the partitions (test, training, and validation) are now displayed. The available charts include

- **%Response (cumulative or noncumulative)**
- **Profit or Loss (cumulative or noncumulative).**

Here is an example display of the noncumulative cross-profit chart of a **Tree** model:



Association [Context]

Enhancements to the **Association** node include the following:

- The **Minimum Transaction Frequency to Support Association** and the **Minimum Transaction Frequency to Support Sequence** can be specified as a percentage or a fractional percentage.
- The Rules tab of the Association Results Browser contains an additional column showing **Transaction Count** for each rule.
- Additional scatter controls are implemented in the Association Results Browser scatter plot. By default, the **confidence level** and the **support** for each rule determine the shape and color of the symbol in the plot. You can scroll the legend that displays statistics in categories. You can use the **Create Viewport Into Data** tool to reduce the scale of the axes on the plot.



Clustering

The **Clustering** node now exports the cluster statistics and the cluster seed data sets to the successor nodes.



Data Set Attributes

- In order for the **Data Set Attributes** node to run, the **Export** status of at least one variable must be set to **Yes**.
- You may now browse the data set in a table view. To see a tabular view of a data set's administrative information, select **File→Details** from the main menu when the **Data Set Attributes** node is open in the Diagram Workspace.



Ensemble

Enhancements to the **Ensemble** node include the following:

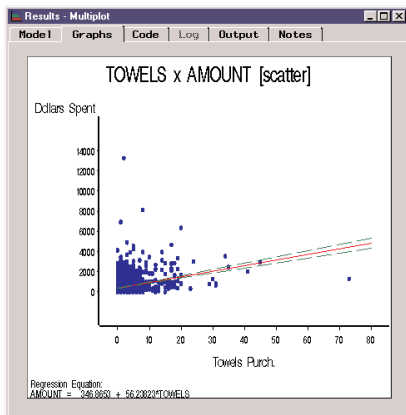
- The **Ensemble** node now computes all fit statistics for the ensemble model.
- The fit statistics, classification, and decision tables are printed in the output.
- For bagging, boosting, or stratification models, iteration (looping) counts for the fit statistics are now printed in the output.



Multiplot

Enhancements to the **Multiplot** node include the following:

- You can change the default configuration settings for the **Multiplot** node.
- The **target scatter plot** display can be toggled on and off. The new default setting does not automatically display the target scatter plot.
- If a target variable is defined, a **bivariate histogram** is created automatically. You can also specify whether to display the univariate histogram along with the bivariate histogram.
- You can change the variable roles that are exported to the successor nodes.
- You can now display an equation on the regression plot using an **interval input** on an **interval target variable**.
- Here is an example of an equation on the regression plot that uses an **interval input** on an **interval target variable**.



Neural Network

The Code tab of the Neural Network Results Browser contains information about training and score codes. The training code contains procedure-based code for training the model, and the score code contains the SAS DATA step for scoring.



Regression

- The **Regression** node automatically changes the variable role from **input** to **rejected** for variables that are not included in the model. This is true when model selection methods of **forward**, **backward**, or **stepwise** are applied.
- In the Estimates tab of the Regression Results Browser, you can now view parameter estimates in a plot or a table.
- The **Regression** node now exports the parameter estimates data set to the successor nodes.



Sampling

The **Sampling** node supports the **user defined stratified sampling** method, which allows you to retrieve information about levels and associated percentage from either the **metasample** or the **entire training data set**.

| Purchase (y/n) | Metadata Sample | Percentage | Percent |
|----------------|-----------------|--------------|---------|
| NO | 49.186164802 | 49.186164802 | |
| YES | 50.813835198 | 50.813835198 | |

Enhancements to the **SAS Code** node include the following:

- A tree view of macro variables with their current values is displayed.

| Macro Variable | Description | Current value | Data Set |
|---------------------|-------------------------|---------------|----------|
| EMDATAVIEW_IQM | Import: Train data... | | SAMPS |
| EMDATAVIEW_IQM | Import: Validation... | | SAMPS |
| EMDATAVIEW_IQM | Import: Test data... | | SAMPS |
| EMDATAVIEW_IQM | Import: Score data... | | SAMPS |
| EMDATAVIEW_IQM | Import: Raw data... | | SAMPS |
| PURCHASE AMOU... | All variables (not r... | | |
| PURCHASE INCOM... | Rejected variables | | |
| PURCHASE INCOM... | All inputs | | |
| INCOME HOMEVAL... | Interval inputs | | |
| PURCHASE MARIT... | Class inputs | | |
| INTITLE STATECOD... | Nominal inputs | | |

- In addition to the Exports tab, you can add the exported macro variables in the Macros tab.



Score !

With the implementation of long variable names, the variable names in the merge action are improved and become easier to follow.



SOM/Kohonen

Enhancements to the **SOM/Kohonen** node include the following:

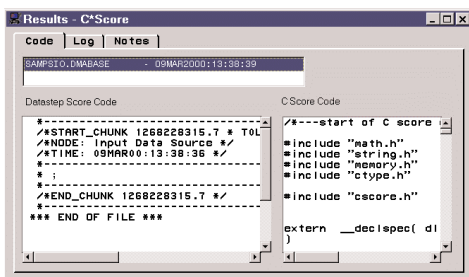
- The **SOM/Kohonen** node now uses a new procedure that provides enhanced speed.
- The **SOM/Kohonen** node now creates dummy variables that are incorporated into the score code.

In the HTML report generated by the **Reporter** node, the Tree Diagram is now available with zooming capabilities. If you do not view the Tree Diagram in the Reporter Results Browser before creating a report, the Tree Diagram in the HTML report will not support assessment coloring and is limited to the information displayed in each node. Otherwise, the Tree Diagram in the generated report will be identical to the one displayed in the Tree Results Browser.

New Enterprise Miner Nodes and Node Features

The **C*Score** node is released as a beta feature in Enterprise Miner Version 4.0, and it converts Enterprise Miner scoring code to C functions.

Here is an example display of the **C*Score** node Results Browser:



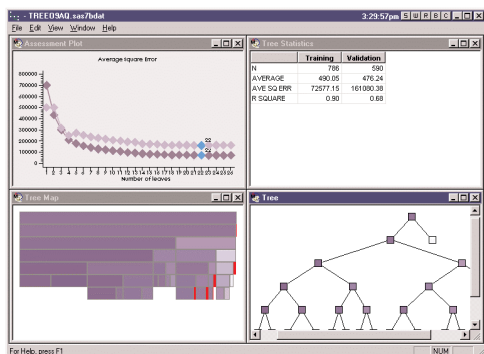
Tree

The **Tree** node in Enterprise Miner Version 4.0 supports additional graphic output displays. To invoke the new Tree View graphics, follow these steps:

1. Add the following code in the autoexec.sas file or submit it using the SAS Program Editor: `%let emv4tree=1;`
6. Click on the **Tree** node icon in the Diagram Workspace and select the **New view** pop-up menu item to launch the Tree Viewer.

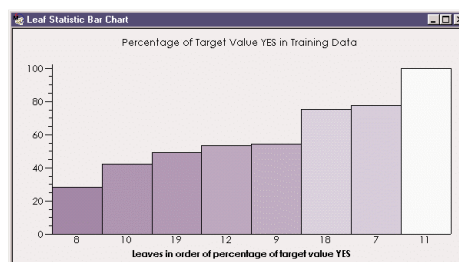


Tree



The new Tree View:

- Creates a thin-client viewer.
- Provides a competitive tree display that has improved printing ability.
- Improves the presentation quality of the analytical graphics.
- Includes a **Tree Map** that replaces the original Tree Ring graph. In the Tree Map, the width of the bars on each level of the map represents the number of observations that fall into that node.
- Includes a Leaf Statistic Bar Chart that provides another graphical display of the quality of the fitted model. This chart is available by choosing **View**→**Leaf Statistic Bar Chart** from the main menu of the new tree viewer.



Warehouse Administrator Add-ins

The following new sampling methods are available for you to create metadata samples before you import data into the Warehouse Administrator:

- Simple Random Sampling
- First N Sampling
- Every Nth Sampling
- Cluster Sampling
- Stratified Sampling.

DM Tool Application Program Interface (API)


Some data mining professionals have expressed interest in creating custom data mining tools to use within Enterprise Miner. The DM Tool class provides an SCL programming interface for you to create custom tools for the Enterprise Miner. You must license SAS/AF in order to use the DM Tool API.

DISCLAIMER: SAS Institute reserves the right to modify or eliminate the DM Tool class programming interface with each new release of SAS Enterprise Miner. The DM Tool class may not be supported in future releases of Enterprise Miner. SAS Institute does not provide end-user support for custom tools that are created by customers who use the DM Tool class.

Further documentation about DM Tool can be found on the World Wide Web at the following URL: www.sas.com. From the SAS main page, click on the **Demos and Downloads** link. From that page, click the **Business Solutions** link from **Downloads by Product Category**. The Business Solutions page contains a link to **Enterprise Miner Documentation and Examples (DM Tool)**.

Online and Web-based Documentation

Enterprise Miner Version 4.0 documentation is available in the following forms:

- Help from the Enterprise Miner main menu. Select **Help→EM Reference** and select your Help topic from the table of contents.
- Frame-specific Help. Click on the  **“What’s This?”** Help tool icon in the Enterprise Miner Tool Box above the Diagram Workspace; then click inside the window frame that you would like help on.
- Web-based tutorial. To view the tutorial, start Enterprise Miner and select **Help→Getting Started with Enterprise Miner Software** from the SAS menu bar. A link will lead you to the tutorial on the SAS Web site. The Getting Started tutorial guides you step-by-step through the process of creating a data mining project.
- Introduction to Enterprise Miner. Select **Help→EM Reference→Getting Started with Enterprise Miner**. This information is more detailed and thorough than the Web-based tutorial mentioned above.
- Introduction to the Cubic Clustering Criterion. Select **Help→EM Reference→Cubic Clustering Criterion**.



SAS World Headquarters
SAS Campus Drive
Cary, NC 27513 USA
Tel: (919) 677 8000
Fax: (919) 677 4444
U.S. & Canada sales:
(800) 727 0025

SAS Europe, Middle East & Africa
PO Box 10 53 40
Neuenheimer Landstr. 28-30
D-69043 Heidelberg, Germany
Tel: (49) 6221 4160
Fax: (49) 6221 474850

SAS Asia/Pacific & Latin America
SAS Campus Drive
Cary, NC 27513 USA
Tel: (919) 677 8000
Fax: (919) 677 8144

www.sas.com