

SAS® Scoring Accelerator 1.6 for DB2 under UNIX User's Guide



The correct bibliographic citation for this manual is as follows: SAS Institute Inc 2009. SAS® Scoring Accelerator 1.6 for DB2 under UNIX: User's Guide. Cary, NC: SAS Institute Inc.

SAS® Scoring Accelerator 1.6 for DB2 under UNIX: User's Guide

Copyright © 2009, SAS Institute Inc., Cary, NC, USA

All rights reserved. Produced in the United States of America.

For a hardcopy book: No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, or otherwise, without the prior written permission of the publisher, SAS Institute Inc.

For a Web download or e-book: Your use of this publication shall be governed by the terms established by the vendor at the time you acquire this publication.

U.S. Government Restricted Rights Notice: Use, duplication, or disclosure of this software and related documentation by the U.S. government is subject to the Agreement with SAS Institute and the restrictions set forth in FAR 52.227–19 Commercial Computer Software-Restricted Rights (June 1987).

SAS Institute Inc., SAS Campus Drive, Cary, North Carolina 27513.

1st electronic book, November 2009 2nd electronic book, December 2009 3rd electronic book, May 2010 4th electronic book, October 2010

SAS® Publishing provides a complete selection of books and electronic products to help customers use SAS software to its fullest potential. For more information about our e-books, e-learning products, CDs, and hard-copy books, visit the SAS Publishing Web site at support.sas.com/publishing or call 1-800-727-3228.

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 registered trademarks or trademarks of their respective companies.

Contents

	What's New in SAS Scoring Accelerator 1.6 for DB2 under UNIX	1
Chapter 1 • Intro	oduction to the SAS Scoring Accelerator for DB2 under UNIX	1
onaptor i ma	Overview of the SAS Scoring Accelerator for DB2 under UNIX	
	How the SAS Scoring Accelerator for DB2 under UNIX Works	
Chapter 2 • Dep	loyed Components for In-Database Processing	5
	Overview of Deployed Components for In-Database Processing	
Chapter 3 • Exp	orting the Scoring Model Files from SAS Enterprise Miner	7
	Overview of the Score Code Export Node	7
	Using the Score Code Export Node Compared with Registering	
	Models on the SAS Metadata Server	
	Using the Score Code Export Node	8
	Output Created by the Score Code Export Node	10
Chapter 4 • Pub	lishing the Scoring Model Files	19
-	Overview of the Publishing Process	19
	Running the %INDB2_PUBLISH_MODEL Macro	20
	Special Characters in Directory Names	26
	DB2 Permissions	27
Chapter 5 • Sco	ring Functions Inside the DB2 Database	29
•	Scoring Function Names	
	Using the Scoring Functions	30
Chapter 6 • SAS	S Scoring Accelerator and SAS Model Manager	33
	Using the SAS Scoring Accelerator with SAS Model Manager	33
	Index	35

What's New in SAS Scoring Accelerator 1.6 for DB2 under UNIX

Overview

The SAS Scoring Accelerator 1.6 for DB2 under UNIX has the following new features.

- availability of the Score Code Export node software in SAS Enterprise Miner
- the values for the %INDB2_PUBLISH_MODEL macro's MODE argument are now FENCED and UNFENCED

Score Code Export Node Software Availability

In the December 2009 release, the Score Code Export node software is available in SAS Enterprise Miner. In previous versions of the SAS Scoring Accelerator for DB2 under UNIX, the software was provided in a ZIP file.

MODE Argument Values

In the May 2010 release, the values for the %INDB2_PUBLISH_MODEL macro's MODE argument changed from PROTECTED and UNPROTECTED to FENCED and UNFENCED.

SAS Model Manager Support

In the November 2010 release, you can use the SAS Scoring Accelerator for DB2 under UNIX in conjunction with SAS Model Manager Version 2.3 to manage and deploy scoring models in DB2 under UNIX.

Chapter 1

Introduction to the SAS Scoring Accelerator for DB2 under UNIX

Overview of the SAS Scoring Accelerator for DB2 under UNIX
How the SAS Scoring Accelerator for DB2 under UNIX Works

Overview of the SAS Scoring Accelerator for DB2 under UNIX

When using conventional processing to access data inside a DB2 database, SAS Enterprise Miner asks the SAS/ACCESS engine for all rows of the table being processed. The SAS/ACCESS engine generates an SQL SELECT * statement that is passed to the DB2 database. That SELECT statement fetches all the rows in the table, and the SAS/ACCESS engine returns them to SAS Enterprise Miner. As the number of rows in the table grows over time, network latency grows because the amount of data that is fetched from the DB2 database to the SAS scoring process increases.

The SAS Scoring Accelerator for DB2 under UNIX embeds the robustness of SAS Enterprise Miner scoring models directly in the highly scalable DB2 database. By using the SAS In-Database technology and the SAS Scoring Accelerator for DB2 under UNIX, the scoring processing is done inside the database, and thus does not require the transfer of data.

The SAS Scoring Accelerator for DB2 under UNIX takes the models that are developed by SAS Enterprise Miner and translates them into scoring functions that can be deployed inside DB2. After the scoring functions are published, the functions extend the DB2 SQL language and can be used in SQL statements like other DB2 functions.

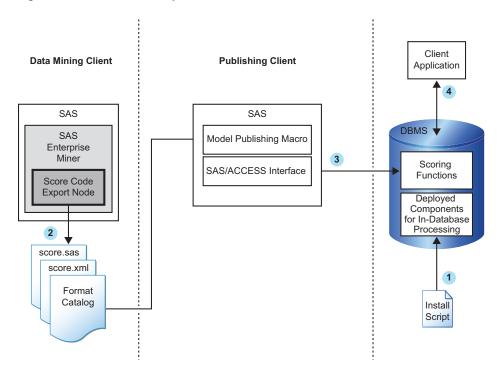
The SAS Scoring Accelerator for DB2 under UNIX consists of two components:

- the Score Code Export node in SAS Enterprise Miner. This extension exports the model scoring logic, including metadata about the required input and output variables, from SAS Enterprise Miner.
- the publishing client that includes the %INDB2_PUBLISH_MODEL macro. This macro translates the scoring model into .c and .h files for creating the scoring functions and generates a script of DB2 commands for registering the scoring functions. The publishing client then uses the SAS/ACCESS Interface to DB2 to publish the scoring functions to DB2.

How the SAS Scoring Accelerator for DB2 under UNIX Works

Using SAS Enterprise Miner, you can generate SAS DATA step code that contains scoring functions. The SAS Scoring Accelerator for DB2 under UNIX takes the scoring model code, the associated property file that contains model inputs and outputs, and a catalog of user-defined formats, and deploys, or publishes, them to the DB2 database. Inside the DB2 database, one or more scoring functions are created and registered for use in SQL queries. Figure 1.1 illustrates this process.

Figure 1.1 Process Flow Diagram



1 Install the components that are necessary for in-database processing in the DB2 database.

For more information, see Chapter 2, "Deployed Components for In-Database Processing," on page 5.

Note: This is a one-time installation process.

2 Use SAS Enterprise Miner to create a scoring model, and use the Score Code Export node to export files that are used to create the scoring functions to a score output directory.

For more information, see Chapter 3, "Exporting the Scoring Model Files from SAS Enterprise Miner," on page 7.

Start SAS 9.2 and run the SAS publishing macros.
For more information, see Chapter 4, "Publishing the Scoring Model Files," on page 19.

4 After the scoring functions are created, they are available to use in any SQL expression in the same way that DB2 built-in functions are used.

For more information, see Chapter 5, "Scoring Functions Inside the DB2 Database," on page 29.

4 Chapter 1 • Introduction to the SAS Scoring Accelerator for DB2 under UNIX

Chapter 2

Deployed Components for In-Database Processing

Overview of Deployed Components for In-Database Processing

The following components are deployed:

- the SAS 9.2 Formats Library for DB2. The SAS 9.2 Formats Library for DB2 contains many of the formats that are available in Base SAS and processes any formats that might be included in your scoring model.
- the binary file for the SAS_COMPILEUDF function. The %INDB2_PUBLISH_COMPILEUDF macro publishes the SAS_COMPILEUDF function in the SASLIB schema of a DB2 database. The SAS_COMPILEUDF function compiles the scoring model source files, links to the SAS 9.2 Formats Library for DB2, and then copies the new object files to a specified location.
- the binary file for the SAS_DELETEUDF function. The %INDB2_PUBLISH_DELETEUDF macro publishes the SAS_DELETEUDF function in the SASLIB schema of a DB2 database. The SAS_DELETEUDF function removes existing object files.

Components that are deployed to DB2 for in-database processing are contained in a self-extracting TAR file (acceldb2fmt*.sh) on the SAS Software Depot.

For more information about creating the SAS Software Depot, see your Software Order e-mail. For more information about installing and configuring these components, see the SAS In-Database Products: Administrator's Guide.

Chapter 3

Exporting the Scoring Model Files from SAS Enterprise Miner

Overview of the Score Code Export Node	/
Using the Score Code Export Node Compared with Registering Models on the SAS Metadata Server	8
Using the Score Code Export Node	8
Using the Score Code Export Node in a Process Flow Diagram	8
Score Code Export Node Properties	9
Output Created by the Score Code Export Node	10
Results Window	
Output Files	11
Output Variables	12
Fixed Variable Names	13
SAS Enterprise Miner Tools Production of Score Code	14

Overview of the Score Code Export Node

Users of SAS Enterprise Miner develop data mining models that use measured attributes to either characterize or predict the value of an event. These models are developed on historical data where an event has been measured or inferred. The models are then applied to new data for which the attributes are known, but the event has not yet occurred. For example, a model can be created based on a credit institution's records of payments that customers made and missed last year and then used to predict which customers will miss payments this year.

SAS Enterprise Miner creates SAS language score code for the purpose of scoring new data. Users run this code in production systems to make business decisions for each record of new data.

The Score Code Export node is an extension for SAS Enterprise Miner that exports files that are necessary for score code deployment. Extensions are programmable add-ins for the SAS Enterprise Miner environment.

The following icon is the Score Code Export node as it appears in a SAS Enterprise Miner process flow diagram.



The following files are exported by the Score Code Export node:

- the SAS scoring model program (score.sas).
- a properties file that contains a description of the variables that are used and created by the score code (score.xml).
- a format catalog, if the scoring program contains user-defined formats.
- an XML file containing descriptions of the final variables that are created by the score code. This file can be kept for decision-making processes.
- a ten-row sample of the scored data set showing typical cases of the input attributes, intermediate variables, and final output variables used to develop the score code. This data set can be used to test and debug new scoring processes.
- a ten-row sample table of the training data set showing the typical cases of the input attributes used to develop the score code.

For more information about the exported files, see "Output Files" on page 11. For more information about using SAS Enterprise Miner, see the SAS Enterprise Miner Help.

Using the Score Code Export Node Compared with Registering Models on the SAS Metadata Server

SAS Enterprise Miner can register models directly in the SAS Metadata Server. Models registered in the SAS Metadata Server are used by SAS Data Integration Studio, SAS Enterprise Guide, and SAS Model Manager for creating, managing, and monitoring production and analytical scoring processes.

The Score Code Export node exports score code created by SAS Enterprise Miner into a format that can be used by the SAS Scoring Accelerator for DB2 under UNIX. The exported files are stored in a directory, not the SAS Metadata Server.

The Score Code Export node does not replace the functionality of registering models in the SAS Metadata Server.

Using the Score Code Export Node

Using the Score Code Export Node in a Process Flow Diagram

The **Score Code Export node** icon is located on the **Utility** tab, as shown in Figure 3.1:

Figure 3.1 The Diagram Toolbar with the SAS Score Code Export Node Icon Highlighted



To use the Score Code Export node, you need a process flow diagram that contains nodes that produce score code and that flow to a Score node. The Score node aggregates the score code for the entire analysis path. The Score node must precede the Score Code Export node in the process flow diagram.

Figure 3.2 shows a valid data mining process for exporting score code:

Figure 3.2 Data Mining Process Flow Diagram



Requirement: The Score Code Export node exports score code that contains only one DATA step. For a list of SAS Enterprise Miner nodes that produce score code, see "SAS Enterprise Miner Tools Production of Score Code" on page 14.

After the process flow diagram is in place, set the properties for the Score node and the Score Code Export node:

- 1. Select the Score node. Ensure that the following properties are set to their default value of Yes:
 - **Use Output Fixed Names**
 - C Score
- 2. Select the Score Code Export node and set the properties. The **Output Directory** property specifies the directory to store the export files. The Name property specifies the folder that contains the output files created by the Score Code Export node. For information about the properties, see "Score Code Export Node Properties" on page

After the properties are set, you are ready to export the score code. Right-click the Score Code Export node and select Run. When SAS Enterprise Miner completes processing, the Run Status window opens to indicate that the run completed. Click the Results button to view the output variables and the listing output. For information about the output, see "Output Created by the Score Code Export Node" on page 10.

Score Code Export Node Properties

When the Score Code Export node is selected in the diagram workspace, the Properties panel displays all of the properties that the node uses and their associated values, as shown in Figure 3.3.

Property Value General CodeXpt2 Node ID Imported Data Exported Data Notes Train No. Rerun Output Directory e:\models simple_test Name Status Create Time 3/6/08 6:11 PM Run Id d44b7835-2b53-46f2-b Last Error Last Status Complete Last Run Time 3/6/08 6:29 PM 0 Hr. 0 Min. 5.48 Sec. Run Duration Grid Host

Figure 3.3 Properties Panel

The following Train properties are associated with the Score Code Export node:

- **Rerun** Use this property to force the node to run again. This property is useful if the macro variable controlling the target directory and folder name has changed.
- Output Directory Enter a fully qualified name for the location of an output directory to contain the score code files. If no directory is entered, a default directory named Score is created in the SAS Enterprise Miner project directory. You can change the value of the default directory by setting the &EM SCOREDIR=directory macro variable in the SAS Enterprise Miner project start-up code or server start-up code.
- Name Enter the name of the model that you are creating. The name is used to create a new subdirectory in the output directory that contains the exported score files. If no name is entered, a default name is generated as a combination of the &SYSUSERID automatic macro variable and an incremental index (for example, userID, userID 2, userID 3).

You can replace the &SYSUSERID automatic macro variable with a custom name by setting the &EM SCOREFOLDER=score-folder-name macro variable in the SAS Enterprise Miner project start-up code or server start-up code. An incremental index preceded by an underscore is added to *score-folder-name*.

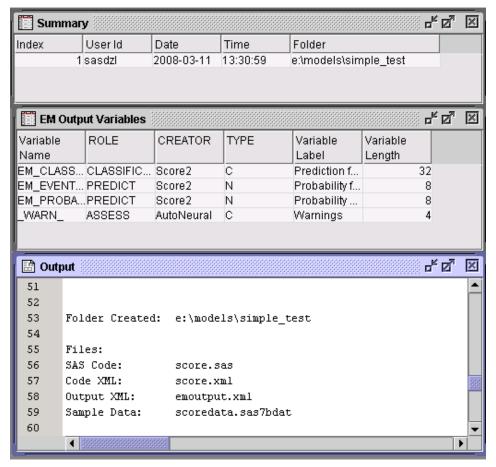
The General and Status properties for the Score Code Export node function just as they do for other nodes.

Output Created by the Score Code Export Node

Results Window

Using the values set in the Properties panel (Figure 3.3), the Score Code Export node creates the following output in the Results window:

Figure 3.4 Results Using Sample Properties



Output Files

The Score Code Export node writes the following output files, and a format catalog, if applicable, to the location specified by the Output Directory property. These files are used as input to the %INDB2_PUBLISH_MODEL macro that creates the scoring functions.

File or Folder	Description
score.sas	SAS language score code created by SAS Enterprise Miner. This code can be used directly in a SAS program. A sample program based on the properties shown in Figure 3.3 looks like this:
	<pre>data testout ; set simpletest.scoredata ; %include "c:\models\simpletest\score.sas"; run;</pre>
score.xml	A description of the variables that are used and created by the scoring code. XML files are created by a machine process for the use of machine processes. Do not edit the XML file.
	Restriction: The maximum number of input variables for a scoring function is 128.

File or Folder	Description
emoutput.xml	A description of the final variables that are created by the scoring code. This file can be kept for decision-making processes. These variables include the primary classification, prediction, probability, segment, profit, and loss variables created by a data mining process. The list does not include intermediate variables created by the analysis. For more information about these variables, see "Fixed Variable Names" on page 13.
	<i>Note:</i> The emoutput.xml file is not used by the %INDB2_PUBLISH_MODEL macro.
scoredata.sas7bdat	A ten-row sample of the scored data set showing typical cases of the input attributes, intermediate variables, and final output variables. Use this data set to test and debug new scoring processes.
	<i>Note:</i> The scoredata.sas7bdat file is not used by the %INDB2_PUBLISH_MODEL macro.
traindata.sas7bdat	A ten-row sample table of the training data set showing typical cases of the input attributes used to develop the score code.
	<i>Note:</i> The traindata.sas7bdat file is not used by the %INDB2_PUBLISH_MODEL macro.
Format catalog	If the training data contains SAS user-defined formats, the Score Code Export node creates a format catalog. The catalog contains the user-defined formats in the form of a look-up table. This file has an extension of .sas7bcat.

Output Variables

The score code produced by SAS Enterprise Miner creates both intermediate variables, such as imputed values of missing values, transformations, and encodings; and output variables, such as predicted value and probability. Any of these created variables can be used in a scoring process.

TIP The number of input parameters on a scoring function has a direct impact on performance. The more parameters there are, the more time it takes to score a row. A recommended best practice is to make sure that only variables that are involved in a model score evaluation are exported from SAS Enterprise Miner.

The most important output variables for the scoring process follow a naming convention using a prefix, as shown in the following table.

Role	Туре	Prefix	Key	Suffix	Example
Prediction	N	P_	Target variable name		P_amount
Probability	N	P_	Target variable name	Predicted event value	P_purchaseYES P_purchaseNO

Role	Туре	Prefix	Key	Suffix	Example
Classification	\$	I_	Target variable name		I_purchase
Expected Profit	N	EP_	Target variable name		EP_conversion
Expected Loss	N	EL_	Target variable name		EL_conversion
Return on Investment	N	ROI_	Target variable name		ROI_conversion
Decision	\$	D_	Target variable name		D_conversion
Decision Tree Leaf	N	_NODE_			_NODE_
Cluster number or SOM cell ID	N	_SEGMENT_			_SEGMENT_

Fixed Variable Names

The Score node of SAS Enterprise Miner maps the output variable names to fixed variable names. This mapping is appropriate in cases where there is only one prediction target or one classification target. In other cases, refer to the output variable names described in the previous table.

Using the fixed variable names enables scoring users to build processes that can be reused for different models without changing the code that processes the outputs. These fixed names are listed in the emoutput.xml file and are described in the following table. Most scoring processes return one or more of these variables.

Role	Туре	Fixed Name	Description
Prediction	N	EM_PREDICTION	The prediction value for an interval target.
Probability	N	EM_PROBABILITY	The probability of the predicted classification, which can be any one of the target variable values.

Role	Туре	Fixed Name	Description
Probability	N	EM_EVENTPROBABILITY	The probability of the target event. By default this is the first value in descending order. This is often the event of interest. The user can control the ordering in SAS Enterprise Miner.
Classification	\$	EM_CLASSIFICATION	The predicted target class value.
Expected Profit	N	EM_PROFIT	Based on the selected decision.
Expected Loss	N	EM_LOSS	Based on the selected decision.
Return on Investment	N	EM_ROI	Based on the selected decision.
Decision	\$	EM_DECISION	Optimal decision based on a function of probability, cost, and profit or loss weights.
Decision Tree Leaf, Cluster number, or SOM cell ID	N	EM_SEGMENT	Analytical customer segmentation.

SAS Enterprise Miner Tools Production of Score Code

The following table shows the types of score code created by each node in SAS Enterprise Miner. Users can develop their own nodes, known as extension nodes, which can create either SAS DATA step or SAS program score code. However, this code is not converted to PMML, C, or Java.

Node	SAS DATA Step	SAS Program	PMML	С	Java	DB2
Sample						
Input Data	*	*	*	*	*	*
Sample	*	*	*	*	*	*
Partition	*	*	*	*	*	*
Append	N	Y	N	N	N	N
Merge	N	Y	N	N	N	N
Time Series	N	Y	N	N	N	N

Node	SAS DATA Step	SAS Program	PMML	С	Java	DB2
Filter	Y When the user keeps the created filter variable.	*	N	Y	Y	Y
Explore						
Association	N	Y	Y	N	N	N
Cluster	Y	N	Y	Y	Y	Y
DMDB	*	*	*	*	*	*
Graph Explore	*	*	*	*	*	*
Market Basket	N	Y	N	N	N	N
Multiplot	*	*	*	*	*	*
Path	N	Y	Y	N	N	N
SOM	Y	N	N	Y	Y	Y
Stat Explore	*	*	*	*	*	*
Text Miner	N	Y	N	N	N	N
Variable Clustering	Y	N	N	Y	Y	Y
Variable Selection	Y	N	N	Y	Y	Y
Drop	*	*	*	*	*	*
Impute	Y	N	Y	Y	Y	Y
Interactive Binning	Y	N	N	Y	Y	Y
Replacement	Y	N	N	Y	Y	Y
Principle Components	Y	N	N	Y	Y	Y
Rules Builder	Y	N	N	Y	Y	Y
Transform Variables	Y	N	N	Y	Y	Y
Model	1	1		ı	<u> </u>	ı

Node	SAS DATA Step	SAS Program	PMML	С	Java	DB2
Autoneural	Y	N	Y	Y	Y	Y
Decision Tree	Y	N	Y	Y	Y	Y
Dmine Regression	Y	N	Y	Y	Y	Y
Dmine Neural	Y	N	N	Y	Y	Y
Ensemble	Y	N	N	Y	Y	Y
Gradient Boosting	Y	N	N	Y	Y	Y
MBR	N	Y	N	N	N	N
Model Import	*	*	*	aje.	*	*
Neural Network	Y	N	Y	Y	Y	Y
Partial Least Squares	Y	N	N	Y	Y	Y
Rule Induction	Y	N	N	Y	Y	Y
SVM — Linear Kernel	Y	N	Y	Y	Y	Y
SVM — Nonlinear Kernel	N	Y	N	N	N	N
Two Stage	Y	N	N	Y	Y	Y
Assess						
Cutoff	Y	N	N	Y	Y	Y
Decisions	Y	N	N	Y	Y	Y
Model Comparison	Y	N	N	Y	Y	Y
Score	Y	N	N	Y	Y	Y
Segment Profile	*	*	*	*	*	*
Utility				,		
Control Point	*	*	*	*	*	*

Node	SAS DATA Step	SAS Program	PMML	С	Java	DB2
Start Groups	Y	N	N	Y	Y	Y
End Groups	Y	N	N	Y	Y	Y
Metadata	*	*	*	*	*	*
Reporter	*	*	*	*	*	*
SAS Code The user can enter either SAS DATA step code or SAS program code	Y	Y	N	N	N	N
Credit Scoring						
Credit Exchange	*	*	*	*	*	*
Interactive Grouping	Y	N	N	Y	Y	Y
Scorecard	Y	N	N	Y	Y	Y
Reject Inference	Y	N	N	Y	Y	Y

Chapter 4

Publishing the Scoring Model Files

Overview of the Publishing Process	19
Running the %INDB2 PUBLISH MODEL Macro	20
%INDB2_PUBLISH_MODEL Macro Run Process	20
%INDB2_PUBLISH_MODEL Macro Syntax	22
Modes of Operation	
Model Publishing Macro Example	25
Special Characters in Directory Names	20
DB2 Permissions	27

Overview of the Publishing Process

The SAS publishing macros are used to publish the formats and the scoring functions in DB2.

The %INDB2_PUBLISH_MODEL macro creates the files that are needed to build the scoring functions and publishes the scoring functions with those files to a specified database in DB2. Only the EM_ output variables are published as DB2 scoring functions. For more information about the EM_ output variables, see "Fixed Variable Names" on page 13.

Note: SFTP is used to transfer the source files to the DB2 server during the publishing process. Certain software products that support SSH-2 or SFTP protocols must be installed before you can use the publishing macros. For more information, see Setting up SSH Client Software in UNIX and Windows Environments for use with the SFTP Access Method located at http://support.sas.com/techsup/technote/ts800.pdf.

The %INDB2_PUBLISH_MODEL macro uses some of the files that are created by the SAS Enterprise Miner Score Code Export node: the scoring model program (score.sas file), the properties file (score.xml file), and, if the training data includes SAS user-defined formats, a format catalog.

The %INDB2 PUBLISH MODEL macro performs the following tasks:

- takes the score.sas and score.xml files and produces the set of .c and .h files. These .c
 and .h files are necessary to build separate scoring functions for each of a fixed set of
 quantities that can be computed by the scoring model code.
- if a format catalog is available, processes the format catalog and creates an .h file with C structures, which are also necessary to build the scoring functions.

- produces a script of the DB2 commands that are used to register the scoring functions on the DB2 database.
- transfers the .c and .h files to DB2 using SFTP.
- calls the SAS_COMPILEUDF function to compile the source files into object files, links to the SAS 9.2 Formats Library for DB2, and copies the new object files to <code>db2path/sqllib/function/SAS</code>, where <code>db2path</code> is the path that was defined during installation. The object filename is <code>dbname_schemaname_modelname_segnum</code>, where <code>segnum</code> is a sequence number that increments each time the model is replaced or recreated. The object file is renamed to avoid library caching in DB2.
- calls the SAS DELETEUDF function to remove existing object files.
- uses the SAS/ACCESS Interface to DB2 to run the script to create the scoring functions with the object files.

The scoring functions are registered in DB2 with shared object files, which are loaded at run time. These functions are stored in a permanent location. The SAS object files and the SAS 9.2 Formats Library for DB2 are stored in the <code>db2path/sqllib/function/SAS</code> directory, where <code>db2path</code> is the path that was defined during installation. This directory is accessible to all database partitions.

DB2 caches the object files after they are loaded. Each time that the updated objects are used, the database must be stopped and restarted to clean up the cache, or the object files need to be renamed and the functions reregistered with the new object filenames. The SAS publishing process automatically handles the renaming to avoid stopping and restarting the database.

Note: You can publish scoring model files with the same model name in multiple databases and schemas. Because object files for the SAS scoring function are stored in the db2path/sqllib/function/SAS directory, the publishing macros use the database, schema, and model name as the object filename to avoid potential naming conflicts.

Running the %INDB2_PUBLISH_MODEL Macro

%INDB2 PUBLISH MODEL Macro Run Process

To run the %INDB2 PUBLISH MODEL macro, complete the following steps:

- 1. Create a scoring model using SAS Enterprise Miner.
- Use the SAS Enterprise Miner Score Code Export node to create a score output directory and populate the directory with the score.sas file, the score.xml file, and, if needed, the format catalog.
- 3. Start SAS 9.2 and submit the following commands in the Program Editor or Enhanced Editor:

%indb2pm;

%let indconn = server=yourserver user=youruserid password=yourpwd
 database=yourdb schema=yourschema serveruserid=yourserveruserid;

The %INDB2PM macro searches the autocall library for the indb2pm.sas file. The indb2pm.sas file contains all the macro definitions that are used in conjunction with the %INDB2 PUBLISH MODEL macro. The indb2pm.sas file should be in one of

the directories listed in the SASAUTOS= system option in your configuration file. If the indb2pm.sas file is not present, the %INDB2PM macro call (%INDB2PM; statement) issues the following message:

macro indb2pm not defined

The INDCONN macro variable is used to provide credentials to connect to DB2. You must specify server, user, password, and database. The schema name and the server user ID are optional. You must assign the INDCONN macro variable before the %INDB2 PUBLISH MODEL macro is invoked.

Here is the syntax for the value of the INDCONN macro variable for the %INDB2 PUBLISH MODEL macro:

SERVER=server USER=user PASSWORD=password DATABASE=database <SCHEMA=schemaname> <SERVERUSERID=serveruserid>

The SERVER argument supplies the name of the server. The name must be consistent with the way the host name was cached when PSFTP server was run from the command window. If the full server name was cached, you must use the full server name in the SERVER argument. If the short server name was cached, you must use the short server name. For example, if the long name, disk3295.unx.comp.com, is used when PSFTP was run, then server=disk3295.unx.comp.com must be specified. If the short name, disk3295, was used, then server=disk3295 must be specified. For more information about running the PSFTP command, see DB2 Installation and Configuration Steps in the SAS In-Database Products: Administrator's Guide.

The USER, PASSWORD, and DATABASE arguments are the same as those for the LIBNAME statement. For more information about the USER, PASSWORD, and DATABASE arguments, see the LIBNAME statement in the SAS/ACCESS Interface to DB2 Under UNIX and PC Hosts section of SAS/ACCESS for Relational Databases: Reference.

The SCHEMA argument supplies the schema name for the database. If you do not specify a value for the SCHEMA argument, the value of the USER argument is used as the schema name. The SERVERUSERID argument supplies the user ID for SAS SFTP and enables you to access the machine on which you have installed the DB2 database. If you do not specify a value for the SERVERUSERID argument, the value of the USER argument is used as the user ID for SAS SFTP.

- Note: The person who installed and configured the SSH software can provide you with the SERVERUSERID (SFTP user ID) and the private key that need to be added to the pageant.exe (Windows) or SSH agent (UNIX). Pageant must be running for the SFTP process to be successful.
- TIP The INDCONN macro variable is not passed as an argument to the %INDB2 PUBLISH MODEL macro. This information can be concealed in your SAS job. You might want to place it in an autoexec file and set the permissions on the file so that others cannot access the user ID and password.
- 4. Run the %INDB2 PUBLISH MODEL macro. For more information, see "%INDB2 PUBLISH_MODEL Macro Syntax" on page 22.

Messages are written to the SAS log that indicate the success or failure of the creation of the scoring functions.

%INDB2 PUBLISH MODEL Macro Syntax

%INDB2 PUBLISH MODEL

```
(DIR=input-directory-path, MODELNAME=name
 <, DATASTEP=score-program-filename>
 <, XML=xml-filename>
 <, DATABASE=database-name>
 <, FMTCAT=format-catalog-filename>
 <, ACTION=CREATE | REPLACE | DROP>
 <, MODE=FENCED | UNFENCED>
 <, INITIAL WAIT=wait-time>
 <, FTPTIMEOUT=timeout-time>
 <, OUTDIR=diagnostic-output-directory>
 );
```

Arguments

DIR=input-directory-path

specifies the directory where the scoring model program, the properties file, and the format catalog are located.

This is the directory that is created by the SAS Enterprise Miner Score Code Export node. This directory contains the score.sas file, the score.xml file, and, if userdefined formats were used, the format catalog.

Requirement: You must use a fully qualified pathname.

Interaction: If you do not use the default directory that is created by SAS Enterprise Miner, you must specify the DATASTEP=, XML=, and, if needed, FMTCAT= arguments.

See: "Special Characters in Directory Names" on page 26.

MODELNAME=name

specifies the name that is prepended to each output function to ensure that each scoring function name is unique on the DB2 database.

Restriction: The scoring function name is a combination of the model and output variable names. A scoring function name cannot exceed 128 characters. For more information, see "Scoring Function Names" on page 29.

Requirement: The model name must be a valid SAS name that is 10 characters or fewer. For more information about valid SAS names, see the topic on rules for words and names in SAS 9.2 Language Reference: Concepts.

Interaction: Only the EM output variables are published as DB2 scoring functions. For more information about the EM output variables, see "Fixed Variable Names" on page 13 and "Scoring Function Names" on page 29.

DATASTEP=score-program-filename

specifies the name of the scoring model program file that was created by using the SAS Enterprise Miner Score Code Export node.

Default: score.sas

Restriction: Only DATA step programs that are produced by the SAS Enterprise Miner Score Code Export node can be used.

Interaction: If you use the default score sas file that is created by the SAS Enterprise Miner Score Code Export node, you do not need to specify the DATASTEP= argument.

XML=xml-filename

specifies the name of the properties XML file that was created by the SAS Enterprise Miner Score Code Export node.

Default: score.xml

Restriction: Only XML files that are produced by the SAS Enterprise Miner Score Code Export node can be used.

Restriction: The maximum number of output variables is 128.

Interaction: If you use the default score.xml file that is created by the SAS Enterprise Miner Score Code Export node, you do not need to specify the XML= argument.

DATABASE=database-name

specifies the name of a DB2 database to which the scoring functions and formats are published.

Interaction: The database that is specified by the DATABASE argument takes precedence over the database that you specify in the INDCONN macro variable. For more information, see "%INDB2 PUBLISH MODEL Macro Run Process" on page 20.

FMTCAT=format-catalog-filename

specifies the name of the format catalog file that contains all user-defined formats that were created by the FORMAT procedure and that are referenced in the DATA step scoring model program.

Restriction: Only format catalog files that are produced by the SAS Enterprise Miner Score Code Export node can be used.

Interaction: If you use the default format catalog that is created by the SAS Enterprise Miner Score Code Export node, you do not need to specify the FMTCAT= argument.

Interaction: If you do not use the default catalog name (FORMATS) or the default library (WORK or LIBRARY) when you create user-defined formats, you must use the FMTSEARCH system option to specify the location of the format catalog. For more information, see PROC FORMAT in the Base SAS 9.2 Procedures Guide.

ACTION=CREATE | REPLACE | DROP

specifies one of the following actions that the macro performs:

CREATE creates a new function.

REPLACE overwrites the current function, if a function by the same name is

already registered.

DROP causes all functions for this model to be dropped from the DB2

database.

Default: CREATE

TIP If the function has been previously defined and you specify ACTION=CREATE, you will receive warning messages from DB2. If the function has been previously defined and you specify ACTION=REPLACE, no warnings are issued.

MODE=FENCED | UNFENCED

specifies whether the running code is isolated in a separate process in the DB2 database so that a program fault does not cause the database to stop.

Default: FENCED

See: "Modes of Operation" on page 25

TIP After the SAS scoring functions are validated in fenced mode, you can republish them in unfenced mode. You might see a performance advantage when you run in unfenced mode.

INITIAL WAIT=wait-time

specifies the initial wait time in seconds for SAS SFTP to parse the responses and complete the SFTP –batchfile process.

Default: 15 seconds

Interaction: The INITIAL_WAIT= argument works in conjunction with the FTPTIMEOUT= argument. Initially, SAS SFTP waits the amount of time specified by the INITIAL_WAIT= argument. If the SFTP—batchfile process is not complete after the initial wait time, retries occur until the wait time is equal to or greater than the time-out value specified by the FTPTIMEOUT= argument. All retries double the previous wait time. SAS SFTP fails after the time-out value is reached or exceeded, and an error message is written to the SAS log.

For example, assume you use the default values. The initial wait time is 15 seconds. The first retry waits for 30 seconds. The second retry waits for 60 seconds. The third retry waits for 120 seconds, which is the default time-out value. So the default initial wait time and time-out values enable 4 possible tries—the initial try, and 3 retries.

See: FTPTIMEOUT= argument

FTPTIMEOUT=time-out-value

specifies the time-out value in seconds if SAS SFTP fails to transfer the files.

Default: 120 seconds

Interaction: The FTPTIMEOUT= argument works in conjunction with the INITIAL_WAIT= argument. Initially, SAS SFTP waits the amount of time specified by the INITIAL_WAIT= argument. If the SFTP—batchfile process is not complete after the initial wait time, retries occur until the wait time is equal to or greater than the time-out value specified by the FTPTIMEOUT= argument. All retries double the previous wait time. SAS SFTP fails after the time-out value is reached or exceeded and an error message is written to the SAS log.

For example, assume you use the default values. The initial wait time is 15 seconds. The first retry waits for 30 seconds. The second retry waits for 60 seconds. The third retry waits for 120 seconds, which is the default time-out value. So the default initial wait time and time-out values enable 4 possible tries—the initial try, and 3 retries.

TIP Use this argument to control how long SAS SFTP waits to complete a file transfer before timing out. A time-out failure could indicate a network or key authentication problem.

See: INITIAL WAIT= argument

OUTDIR=diagnostic-output-directory

specifies a directory that contains diagnostic files.

Files that are produced include an event log that contains detailed information about the success or failure of the publishing process and sample SQL code (SampleSQL.txt). For more information about the SampleSQL.txt file, see "Scoring Function Names" on page 29.

TIP This argument is useful when testing your scoring models.

See: "Special Characters in Directory Names" on page 26

Modes of Operation

The %INDB2 PUBLISH MODEL macro has two modes of operation: fenced and unfenced. You specify the mode by setting the MODE= argument.

The default mode of operation is fenced. Fenced mode means that the scoring function that is published is isolated in a separate process in the DB2 database when it is invoked, and an error does not cause the database to stop. It is recommended that you publish the scoring functions in fenced mode during acceptance tests.

When the scoring function is ready for production, you can run the macro to publish the scoring function in unfenced mode. You could see a performance advantage if the scoring function is published in unfenced mode.

Model Publishing Macro Example

```
%indb2pm;
%let indconn = server=db2base user=user1 password=open1 database=mydb;
%indb2 publish model( dir=C:\SASIN\baseball1, modelname=baseball1);
```

The %INDB2 PUBLISH MODEL macro produces a text file of DB2 CREATE FUNCTION commands as shown in the following example.

Note: This example file is shown for illustrative purposes. The text file that is created by the %INDB2 PUBLISH MODEL macro cannot be viewed and is deleted after the macro is complete.

```
CREATE FUNCTION baseball1_EM_eventprobablility
"CR ATBAT" float,
"CR BB" float,
"CR HITS" float,
"CR_HOME" float,
"CR RBI" float,
"CR RUNS" float,
"DIVISION" varchar(31),
"LEAGUE" varchar(31),
"NO ASSTS" float,
"NO ATBAT" float,
"NO BB" float,
"NO ERROR" float,
"NO_HITS" float,
"NO HOME" float,
"NO_OUTS" float,
"NO RBI" float,
"NO RUNS" float,
"YR MAJOR" float
RETURNS varchar(33)
LANGUAGE C
NO SOL
PARAMETER STYLE SQL
DETERMINISTIC
FENCED THREADSAFE
NO EXTERNAL ACTION
ALLOW PARALLEL
```

```
NULL CALL

EXTERNAL NAME '/users/db2v9/sqllib/function/SAS/

dbname_username_baseball1.so!baseball1_em_ eventprobablility '
```

After the scoring functions are installed, they can be invoked in DB2 using SQL, as illustrated in the following example. Each output value is created as a separate function call in the select list.

```
select baseball1_EM_eventprobability
"CR_ATBAT",
"CR_BB",
"CR_HITS",
"CR HOME",
"CR_RBI",
"CR RUNS",
"DIVISION",
"LEAGUE",
"NO_ASSTS",
"NO ATBAT",
"NO_BB",
"NO_ERROR",
"NO_HITS",
"NO_HOME",
"NO OUTS"
) as homeRunProb from MLBDB2;
```

Special Characters in Directory Names

If the directory names that are used in the macros contain any of the following special characters, you must mask the characters by using the %STR macro quoting function. For more information, see the %STR function and macro string quoting topic in *SAS Macro Language: Reference*.

Character	How to Represent
blank ¹	%str()
2	%str()
;	%str(;)
,	%str(,)
=	%str(=)
+	%str(+)
-	%str(-)
>	%str(>)
<	%str(<)

Character	How to Represent
۸	%str(^)
	%str()
&	%str(&)
#	%str(#)
1	%str(/)
~	%str(~)
%	%str(%%)
•	%str(%')
"	%str(%")
(%str(%()
)	%str(%))
7	%str(¬)

¹Only leading blanks require the %STR function, but you should avoid using leading blanks in directory names.

Here are some examples of directory names with special characters:

Directory	Code representation
c:\temp\Sales(part1)	<pre>c:\temp\Sales%str(%()part1%str(%))</pre>
<pre>c:\temp\Drug "trial" X</pre>	<pre>c:\temp\Drug %str(%")trial(%str(%") X</pre>
c:\temp\Disc's 50% Y	<pre>c:\temp\Disc%str(%')s 50%str(%%) Y</pre>
c:\temp\Pay,Emp=Z	c:\temp\Pay%str(,)Emp%str(=)Z

DB2 Permissions

You must have DB2 user permissions to execute the SAS publishing macros to publish the scoring functions. Some of these permissions are as follows.

EXECUTE user permission for functions that were published by another user

²Asterisks (*) are allowed in UNIX directory names. Asterisks are not allowed in Windows directory names. In general, you should avoid using asterisks in directory names.

- READ user permission to read the SASUDF_COMPILER_PATH and SASUDF_DB2PATH global variables
- CREATE_EXTERNAL_ROUTINE user permission to the database to create functions
- CREATEIN user permission for the schema in which the scoring functions are published if a nondefault schema is used
- CREATE_NOT_FENCED_ROUTINE user permission to create UDFs that are not fenced

Permissions must be granted for each user that needs to publish a scoring function and for each database that the scoring model publishing uses. Without these permissions, publishing of the scoring functions will fail.

The person who can grant the permissions and the order in which permissions are granted is important. For complete information and examples, see the installation and configuration instructions in the SAS In-Database Products: Administrator's Guide.

Chapter 5

Scoring Functions Inside the DB2 Database

Scoring Function Names	. 2	29
Using the Scoring Functions	. 3	30

Scoring Function Names

The names of the scoring functions that are built in DB2 have the following format:

modelname_EM_outputvarname

modelname is the name that was specified in the MODELNAME argument of the %INDB2_PUBLISH_MODEL macro. modelname is always followed by _EM_ in the scoring function name. For more information about the MODELNAME argument, see "%INDB2_PUBLISH_MODEL Macro Syntax" on page 22.

outputvarname is derived from the names of the EM_ output variables in the score.xml file that is generated from the SAS Enterprise Miner Score Code Export node. For more information about the score.xml file, see "Fixed Variable Names" on page 13.

One scoring function is created for each EM_ output variable in the score.xml file. For example, if the scoring model DATA step program takes ten inputs and creates three new variables, then three scoring functions are defined, each with the name of an output variable. For example, if you set MODELNAME=credit in the %INDB2_PUBLISH_MODEL macro, and the EM_ output variables are "EM_PREDICTION", "EM_PROBABILITY", and "EM_DECISION", then the name of the scoring functions that are created would be "credit_EM_PREDICTION", "credit_EM_PROBABILITY", and "credit_EM_DECISION".

Note: A scoring function name cannot exceed 128 characters.

CAUTION:

When the scoring function is generated, the names are case-insensitive.

Consequently, if you have model names "Model01" and "model01", and you create two scoring functions, the second scoring function will overwrite the first scoring function.

Using the Scoring Functions

The scoring functions are available to use in any SQL expression in the same way that DB2 built-in functions are used. For an example, see "Model Publishing Macro Example" on page 25.

There are four ways to see the scoring functions that are created:

From DB2, you can log on to the database using the DB2 client tool (command line processor) and submit an SQL statement. The following example assumes that the model name that you used to create the scoring functions is mymodel and the DB2 installation instance is located in /users/db2v9. The first line of code executes a db2profile script which sets the DB2 environment variables so the DB2 command line processor (CLP) can execute.

```
>./users/db2v9/sqllib/db2profile
>db2
db2 => connect to database user username using password
db2 => select * from syscat.functions where funcame like '%MYMODEL%'
```

From SAS you can use SQL procedure code that produces output in the LST file.
The following example assumes that the model name that you used to create the scoring functions is mymodel.

```
proc sql noerrorstop;
  connect to db2 (user=username pw=password db=database);

select *
  from connection to db2
     (select * from syscat.functions where funcname like '%MYMODEL%');
  disconnect from db2;
quit;
```

You can also use the SASTRACE and SASTRACELOC system options to generate tracing information. For more information about these system options, see the SAS 9.2 Language Reference: Dictionary.

 You can look at the SampleSQL.txt file that is produced when the %INDB2_PUBLISH_MODEL macro is successfully run. This file can be found in the output directory (OUTDIR argument) that you specify in the macro.

The SampleSQL.txt file contains basic code that, with modifications, can be used to run your score code inside DB2.

For example, the SampleSQL.txt file refers to an ID column in **allmush1_intab** that is populated with a unique integer from 1 to *n*, with *n* being the number of rows in the table. The ID column uniquely identifies each row. You would replace the ID column with your own primary key column.

Note: The function and table names must be fully qualified if the function and table are not in the same schema.

The following example assumes that the model name that you used to create the scoring functions is **allmush1**.

```
drop table allmush1_outtab;
create table allmush1_outtab(
  id integer
```

```
, "EM_CLASSIFICATION" varchar(33)
, "EM EVENTPROBABILITY" float
,"EM_PROBABILITY" float
insert into allmush1_outtab(
,"EM_CLASSIFICATION"
, "EM_EVENTPROBABILITY"
,"EM_PROBABILITY"
)
select id,
allmush1_em_classification("BRUISES"
, "CAPCOLOR"
, "GILLCOLO"
, "GILLSIZE"
,"HABITAT"
, "ODOR"
, "POPULAT"
, "RINGNUMB"
, "RINGTYPE"
,"SPOREPC"
, "STALKCBR"
,"STALKROO"
, "STALKSAR"
, "STALKSHA"
, "VEILCOLO")
 as "EM_CLASSIFICATION",
allmush1_em_eventprobability("BRUISES"
, "CAPCOLOR"
, "GILLCOLO"
, "GILLSIZE"
,"HABITAT"
, "ODOR"
,"POPULAT"
, "RINGNUMB"
,"RINGTYPE"
, "SPOREPC"
, "STALKCBR"
, "STALKROO"
, "STALKSAR"
, "STALKSHA"
, "VEILCOLO")
 as "EM_EVENTPROBABILITY",
allmush1_em_probability("BRUISES"
, "CAPCOLOR"
, "GILLCOLO"
, "GILLSIZE"
,"HABITAT"
, "ODOR"
,"POPULAT"
, "RINGNUMB"
, "RINGTYPE"
, "SPOREPC"
,"STALKCBR"
, "STALKROO"
, "STALKSAR"
```

```
,"STALKSHA"
,"VEILCOLO")
   as "EM_PROBABILITY"
from allmush1_intab ;
```

• You can look at the SAS log. A message that indicates whether a scoring function is successfully or not successfully executed is printed to the SAS log.

Chapter 6

SAS Scoring Accelerator and SAS Model Manager

Using the SAS Scoring Accelerator with SAS Model Manager

You can use SAS Scoring Accelerator in conjunction with the SAS Model Manager to manage and deploy scoring models in DB2 under UNIX.

The **Publish Scoring Function** of SAS Model Manager enables you to publish classification and prediction model types to DB2. When you publish a DB2 scoring function for a project, SAS Model Manager exports the project's champion model to the SAS Metadata Repository and calls the SAS Scoring Accelerator to create the scoring functions. The scoring functions are deployed inside DB2 based on the project's champion model score code. The scoring function is then validated automatically against a default train table to ensure that the scoring results are correct. A scoring application (for example, a call center application that calls the SAS Model Manager Java Scoring API) can then execute the scoring functions in DB2. The scoring functions extend the DB2 SQL language and can be used on SQL statements like other DB2 functions.

For more information, see SAS Model Manager: User's Guide.

Index

Special Characters	M
%INDB2_PUBLISH_MODEL macro 1,	macro string quoting 26
19	macros
example 25	See also xisError - indexSee primary
modes of operation 25	entry
running 20	"%INDB2_PUBLISH_MODEL
syntax 22	macro" not found
	publishing macros 19
	special characters in directory names 26
C	masking 26
case sensitivity 29	model registration
,	on SAS Metadata Server, compared
	with Score Code Export node 8
D	1
data mining models 7	
DB2 database 1	N
scoring functions in 29	Name property 10
DB2 Formats Library 5	names
DB2 permissions 27	fixed variable names 13
deployed components	of scoring functions 29
for in-database processing 5	nodes
directory names	extension nodes 14
special characters in 26	score code created by SAS Enterprise
	Miner nodes 14
	user-defined 14
E	
EM_ output variables 19	
exporting scoring model files	0
from SAS Enterprise Miner 7	output
extension nodes 14	created by Score Code Export node 10
extension nodes 11	Output Directory property 10
	output files 11
F	output variables
fenced mode 25	EM 19
fixed variable names 13	Score Code Export node 12
Formats Library for DB2 5	Score code Export node 12
Tornats Elorary for DB2 3	
	Р
1	permissions, DB2 27
in-database processing	process flow diagrams
deployed components for 5	for SAS Scoring Accelerator 2
deproyed components for 3	101 DAD DOUTING ACCERTAINT 2

using Score Code Export node in 8	Score Code Export node 1, 7
properties 9	compared with registering models on
publishing client 1	SAS Metadata Server 8
publishing macros 19	files exported by 7
publishing process 19	fixed variable names 13
publishing scoring model files 19	output created by 10
running %INDB2_PUBLISH_MODEL	properties 9
macro 20	using in process flow diagrams 8
	scoring functions 1
	%INDB2_PUBLISH_MODEL macro
R	and 1
registering models	inside DB2 database 29
on SAS Metadata Server, compared	names of 29
with Score Code Export node 8	using 30
Rerun property 10	viewing 30
Results window 10	scoring model files
	exporting from SAS Enterprise Miner 7
	publishing 19
S	SFTP protocol 19
SAS_COMPILEUDF function 5	special characters
SAS_DELETEUDF function 5	in directory names 26
SAS Enterprise Miner	SSH-2 protocol 19
exporting scoring model files from 7	
score code created by each node 14	
SAS Metadata Server	U
registering models on, compared with	unfenced mode 25
Score Code Export node 8	user-defined nodes 14
SAS Model Manager 33	
SAS Scoring Accelerator for DB2 under	
UNIX 1	V
components 1	variables
how it works 2	EM_ output variables 19
process flow diagram 2	fixed variable names 13
score code	output variables 12
created by each node of SAS Enterprise	
Miner 14	