

Paper 086-2013

Time Is Precious, So Are Your Models: SAS® Provides Solutions to Streamline Deployment

Jonathan Wexler, Wayne Thompson, and Kristen Aponte, SAS Institute Inc.

ABSTRACT

Organizations spend a significant amount of time, often too much, operationalizing models. The more time you can spend on analytics, and the less time on deployment headaches, the better chance you have to address core business challenges. This paper shows you how SAS has accelerated data mining model deployment throughout the analytical life cycle, by providing key integration points across SAS® solutions. Whether you built your models using SAS® High-Performance Analytics Server, SAS® Enterprise Miner™, or SAS/STAT®, this paper shows you how to automate the management, publishing, and scoring of models by using SAS® Data Integration Studio, SAS® Model Manager, and SAS® Scoring Accelerator. Immediate benefits include reduced data movement, increased productivity of analytic and IT teams, and faster time to results.

INTRODUCTION

Big data has become *the* hot topic for analytical practitioners today. You hardly pick up a publication or research report without reading about big data. Data are exploding at a faster and faster pace. Whether you are analyzing candidates for home equity loans, generating offers within mobile applications, or printing next-best offers on store receipts, analytical models are at the core of many solutions. Analytical teams are being asked to develop hundreds, if not thousands, of statistical models that take advantage of all these data. Organizations that can wrap up and operationalize their analytical solutions experience significant benefits in terms of return on investment (ROI), enhanced productivity, and increased accuracy.

Whether you are deploying your models to databases such as Aster, DB2, Greenplum, Netezza, Oracle, or Teradata, to a web service, or to a real-time engine, models must be incorporated into production systems. This paper presents some easy steps that you can take to operationalize the models you have already built by using traditional SAS® tools such as SAS/STAT or SAS Enterprise Miner, while freeing up time to build significantly more models.

MODELING LIFE CYCLE

The process by which analytical models are built and deployed varies greatly by organization. Often, many users, departments, and tools are combined in order to build and deploy models into production. Whether your team follows an agile methodology or a waterfall approach to modeling, several steps are involved in the process. IT departments are often called on to support the entire process, which is taxing on resources and technology. Statisticians build models, but they need ways to deliver those models to IT that do not introduce errors. Business analysts summarize the data and might hand off a significantly wider data set to the statistician than what was originally selected. Data sets can grow from 50 input variables to more than 1,000 after data are transformed, binned, and modified. IT managers want to run and maintain processes that are centralized, automated, and stable. Business managers need to make informed decisions that are based on metrics such as sales, churn, and ROI. Operational interaction between these teams is vital to the success of model deployment.

Understanding the modeling life cycle helps you appreciate the time that is spent converting models from idea to production. The following figure displays the SAS modeling life cycle and the personas involved in it.

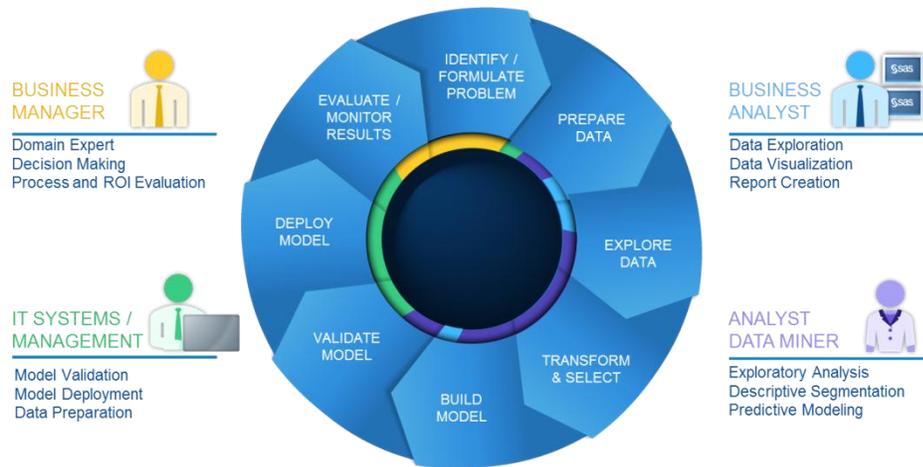


Figure 1. SAS Modeling Life Cycle

MODELING OPERATIONS BUSINESS IMPACT

It is widely reported that analysts spend 70% of their time gathering and transforming data from its rawest form to analytical base tables. Data extraction, transformation, and loading (ETL) can be quite complex and can involve gathering data from multiple data sources in a variety of formats that have different business rules. Given the permutations of business lines, products, and channels that companies face, hundreds of models might need to be built and scored monthly. In the health care industry, keeping the time from model to market as short as possible is key in preventing fraudulent claims. In the effort to minimize bankruptcy and foreclosure in the financial industry, just getting models out a few days earlier can save companies millions of dollars.

Assuming that 70% of analytical time will not be reduced, you need to optimize the remaining 30%. Deployment to production is arguably as challenging as building the models themselves. Translating hundreds of models into SQL is a daunting task even for the most experienced programmers. Models can take months to convert and automate into production systems. The cost in time and resources of manual conversion forces companies to live with their models in production longer than their accepted shelf life.

Don't fret! There are ways to break down those technological barriers. Read on to see how SAS has addressed the production challenge.

USING SAS TO SIMPLIFY MODELING OPERATIONS

Figure 2 outlines the hand-offs that occur in a standard modeling scenario. Depending on the number of departmental hand-offs, many tools are involved. Often, hand-offs occur manually across disparate systems and tools; these manual hand-offs require manual code conversion. SAS provides tools that shorten the time from data capture through model deployment. This paper describes how SAS automates the modeling process by removing manual barriers, enabling you to use your analytic talent more effectively.

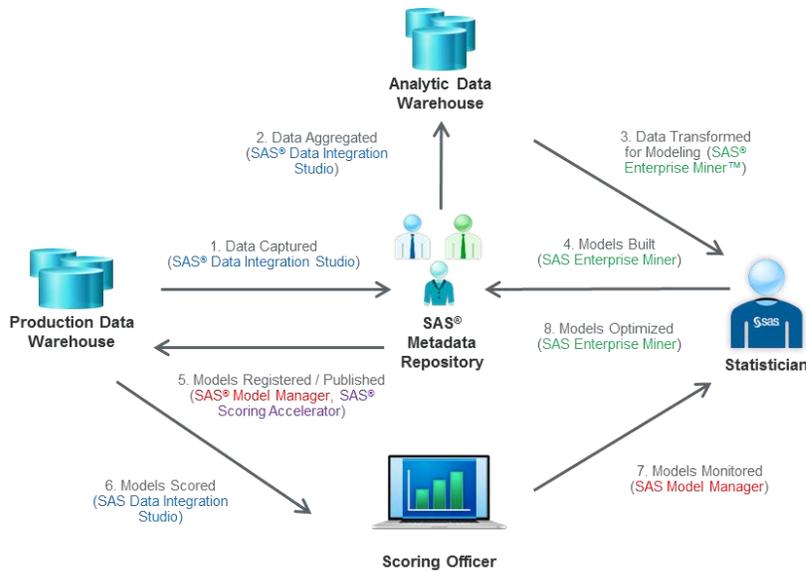


Figure 2. SAS Modeling Process Flow

Notice that the center of the modeling process in Figure 2 is the SAS® Metadata Repository, which enables users to centrally register data sources, programs, and models. Registered items are then available for consumption by SAS Enterprise Miner, SAS Model Manager, and SAS Data Integration Studio. You are not required to register items into the SAS Metadata Repository, but registering them greatly reduces redundant copies of programs and data sets, thus reducing the chance for errors.

IDENTIFYING OR FORMULATING THE PROBLEM

Models often start with one or more questions or business problems. An executive or business manager responsible for a particular line of business might read a sales report that depicts declining sales and ask the following questions: “What factors affect churn within the customer base? What are the next-best products marketers can offer to recent purchasers? Are there segments of customers within the customer base that behave differently?”

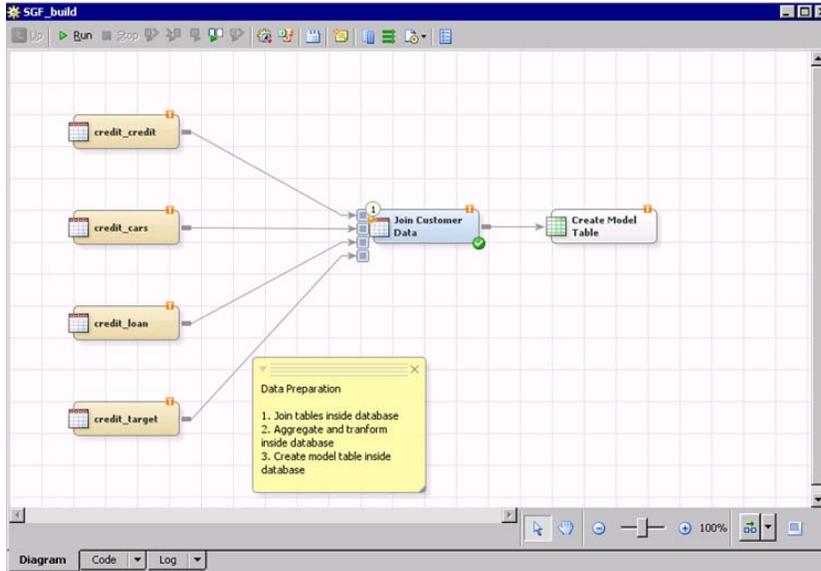
DATA PREPARATION USING SAS DATA INTEGRATION STUDIO

After the questions have been asked, not only are business analysts tasked with addressing the business problems at hand, but they are also charged with addressing the subsequent issues. In order to tackle the problem, data must be collected. Gathering data often takes the most time within the entire process. Data can reside in upstream transactional systems, multiple databases, text files, and so on. Database architects spend a significant amount of time running queries and joining tables, all the while competing for system resources with other users. Data preparation has become more of an art as big data has grown. The role of data scientist was created to address the need for analytical data mining.

The first step in modeling is to create the modeling data set. SAS Data Integration Studio 4.5 enables you to access your data from a graphical user interface. SAS Data Integration Studio enables you to query, transform, and merge data from multiple systems. Simply register your data sets and tables in the SAS Metadata Repository, and you can build your modeling data set with a few clicks. SAS Data Integration Studio provides analytical transformations, summary statistics, and custom transformations. If your source tables reside in a database, you can push down the query to the database, reducing data movement and speeding up processing. If you have SAS code from another process, you can also reference that code from within this job.

SAS Data Integration Studio provides the framework to automate your data set generation. After you have built and deployed your model, you can add the model-scoring process to this job. (Model scoring is described later.)

Display 1 represents a sample SAS Data Integration Studio job. Data sources are joined to create analytical base tables, which are ready for hand-off to the statistician.



Display 1. SAS Data Integration Studio Job

MODEL BUILDING USING THE SAS® PREDICTIVE ANALYTICS SUITE

A key step in the life cycle is to build the model based on the representative training data source that is collected from the data management stage. Building the model involves the application of exploratory statistical and visualization techniques, variable transformations, outlier filtering, data replacement, segmentation, clustering, predictive modeling algorithms, and model comparison, as examples.

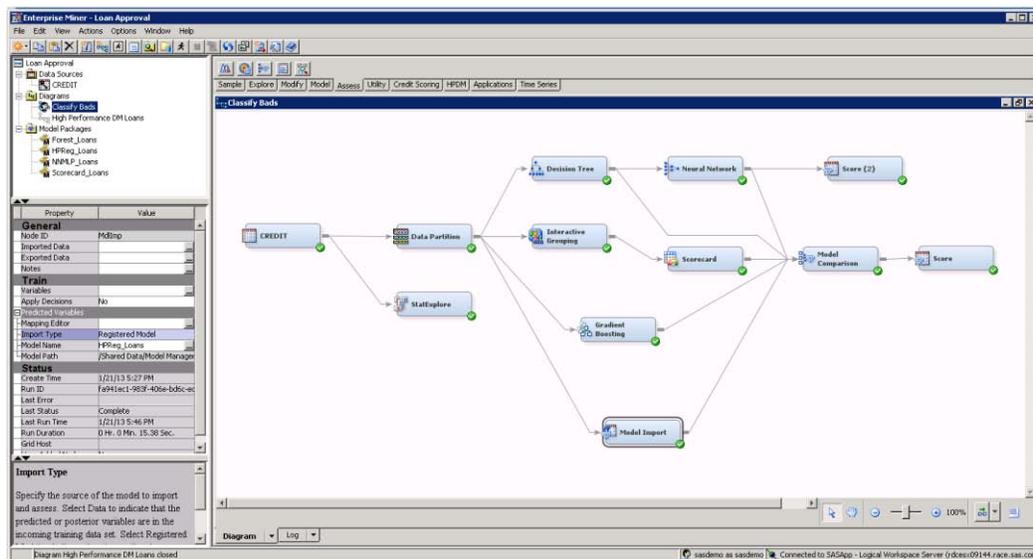
SAS provides a broad set of integrated tools that support the analytics life cycle process in a collaborative environment so that the analytics modeling group, business managers, and IT department can work together more efficiently. SAS Enterprise Miner 12.1 is a good choice for building models that must be integrated into scoring systems. You can often build models faster by using the self-documenting SAS Enterprise Miner process flow, which can also be reapplied to updated data and problems. You can also integrate your customized code and automatically capture the complete scoring logic, including preliminary statistical transformations for direct integration into scoring systems. Customers who license SAS High-Performance Analytics Server can also use massively parallel nodes within SAS Enterprise Miner to analyze more data faster within Teradata, Greenplum, and Hadoop appliances.

You can use the Data Source Wizard of SAS Enterprise Miner to import the model training data source that was created and registered to the SAS Metadata Repository by the analytics data architect in the previous section. Importing the data source helps ensure tight, integrated hand-off at the metadata layer and provides a consistent training source for collaborative workgroup model building. You can further automate the registration of standardized model training tables by using the Enterprise Miner %EMDS macro to register tables in batch for routine repetitive modeling scenarios.



Display 2. Importing the Registered Training Data from SAS Metadata Repository Folders

Display 3 represents a standard SAS Enterprise Miner process flow diagram. You can use SAS Enterprise Miner to build a model, create and register a model package to the SAS Metadata Repository, or save the model as a SAS package file.



Display 3. Build and Share Models Using SAS Enterprise Miner

After you have developed one or more suitable models, you can create a model package that can be shared with fellow data scientists, business managers, and IT staff. A best practice is to create the model package file for a specific model from a terminal Score node. A secondary pass of the score code is made to ensure that only the required input variables are defined in the scoring code. SAS calls this *optimized score code*. You can also register a model package to a SAS Metadata Repository folder for subsequent integration into SAS Model Manager, SAS Data Integration Studio, and other SAS products. Models can also be imported into a SAS Enterprise Miner process flow by using the Model Import node. Customers often define a development folder for registering models that need to be independently validated. You can save a SAS Package (SPK) file, which is a compressed container that includes the model score code and fit statistics. Some customers prefer to use an SPK file to promote models from a development machine to a test or production machine that cannot be physically connected.

Thousands of customers use SAS/STAT linear modeling procedures to build scoring models. The following SAS/STAT 12.1 procedures support a CODE statement that enables you to write a SAS DATA step to a file or catalog to score new data: GENMOD, GLIMMIX, GLM, GLMSELECT, LOGISTIC, MIXED, PLM, and REG. The SAS High-Performance Analytics modeling procedures (except for the HPFOREST procedure) also include a CODE statement, which enables you to save SAS DATA step score code to a file. The STORE statement in the following example generates an item store file that contains score code, information about the model target, inputs and outputs, class levels, fit statistics, and more:

```
proc logistic data=mydat.credit;
  class bad veh_make purpose;
  model bad (event='1') = rev_util bureau_score num_borrowers purch_price msrp
                        down_pyf loan_term ltv tot_income veh_mileage veh_make
                        purpose / selection=forward;

  freq Freq;
  code file="d:\models\logistic.sas";
  store work.logisticStore;
run;
```

The AACOMP macros that are distributed with SAS/STAT 12.1, SAS Enterprise Miner 12.1, and SAS Model Manager 12.1 include the AA_MODEL_REGISTER macro, which registers SAS/STAT linear models or a SAS High-Performance Analytics model to a folder in the SAS Metadata Repository. You can optionally create an SPK file for importing into SAS Enterprise Miner for integrated model comparison or into SAS Model Manager for validation and promotion. Registration of SAS/STAT linear models requires a DATA step plus a SAS item store file.

```
%aamodel;
%aa_model_register(
```

```

MODELNAME= CreditLogistic,
MODELDESC= Credit Class,
REGISTER= N,
MRPATH= %NRBQUOTE(/Shared Data/Model Manager/Stat Models/),
SPK = Y,
SPKFOLDER = D:\models,
ITEMSTORE = work.logisticstore,
DEBUG = Y
);

```

Data scientists and model validation specialists can import SAS models and predictive model markup language (PMML) models into SAS Model Manager. PMML is an XML markup language for exchanging predictive models between modeling systems and scoring platforms. SAS both produces and consumes PMML. The PSCORE procedure, which is distributed with Base SAS®, generates a SAS DATA step and SAS DS2 code for PMML 4.0 models. The following example shows how to use PROC PSCORE:

```

PROC PSCORE PMML FILE = "<full-pathname-of-PMML-file>"
DS FILE = "<full-pathname-of-DS-output-file>"

```

The SAS Model Manager project includes standardized training and test data along with several model comparison reports for both interval and classification targets (dependent variables). You can then select models for a champion and one or more challengers for further validation and promotion.

The screenshot displays the SAS Model Manager application window. On the left, the 'Repository' pane shows a 'Project Tree' with a folder named 'Credit Risk' containing several sub-items, including 'PROC_Logistic_Loans' which is highlighted. Below the tree is an 'Annotations - PROC_Logistic_Loans' section with a 'History' table containing one entry.

On the right, the 'PROC_Logistic_Loans' properties table is shown. The table has columns for 'Properties', 'Model Input', 'Model Output', 'SAS Code', 'Score Tasks', and 'Model Mapping'. The 'Properties' column is expanded to show various attributes.

Properties	Model Input	Model Output	SAS Code	Score Tasks	Model Mapping
General Properties					
Name			PROC_Logistic_Loans		
Description					
Owner			sasdemo		
Date Created			Jan 29, 2013		
Date Modified			Jan 29, 2013		
System Properties					
Specific Properties					
Default Scoring Task Input Table			Loans.CR_SCORE		
Default Scoring Task Output Table			Loans.CR_SCOREOUT		
Default Performance Table					
Default Train Table			Loans.CREDIT		
Expiration Date					
Model Label					
Subject					
Algorithm			PROC LOGISTIC		
Function			Classification		
Modeler			sasrwt		
Tool					
Tool Version					
Score Code Type			DATA step		
Template			STATSPKClassification		
Copied From					
Target Variable			bad		
Target Event Value			1		

At the bottom right of the window, the user is identified as 'SAS Demo User' and the session ID is 'RDCE5X02062.race.sas.com'.

Display 4. Candidate Models Registered in SAS Model Manager

MODEL VALIDATION AND PROMOTION

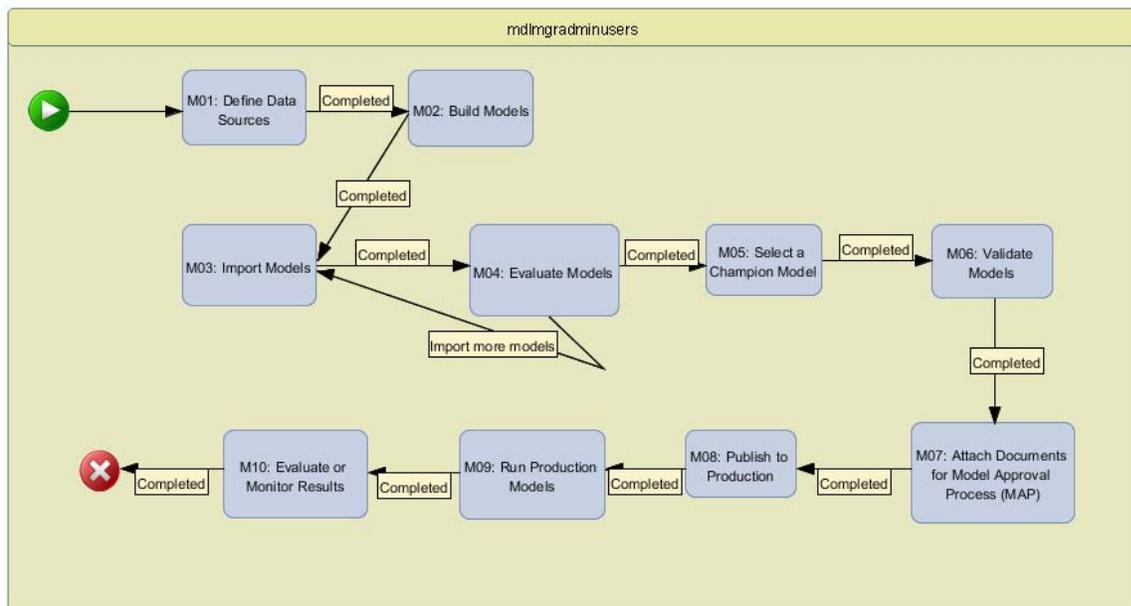
SAS Model Manager enables you to validate the performance of models and then promote them to a production environment. The model comparison, validation, and summary reports are tools that you can use to evaluate champion and challenger models before you promote them to production. The model comparison reports are analytical tools that project managers, statisticians, and analysts can use to assess the structure, performance, and resilience of candidate models. The model validation reports use statistical measures to validate the stability, performance, and calibration of Basel II risk models and parameters. The training summary data set report creates frequency and distribution charts that summarize the training table variables. The reports present information about a number of attributes that can affect model performance. Together, the reports provide qualified information that can serve as the analytical basis for choosing and monitoring a champion model.

SAS Model Manager also provides a comprehensive publishing environment for model delivery that supports sharing life-cycle and performance data. SAS Model Manager can publish classification, prediction, and segmentation (cluster) models to a database when the models use SAS DATA step score code. Application software such as SAS Data Integration Studio or SAS® Enterprise Guide® enables you to access models through the SAS Metadata Server and to submit on-demand and batch scoring jobs.

Using SAS® Workflow

SAS Workflow can be integrated with several of the model management tasks that are normally performed in the SAS Model Manager client. The SAS Model Manager Workflow Console is an interface to SAS Workflow that you can use to track the progress of models in a project's version. You can use SAS Workflow Studio to create process definitions and make them available for use in SAS Model Manager. You can configure workflow process definitions to use model management components with the workflow activities. When you activate the workflow process definition for use, the model management components are available through the object view for the associated activity in the SAS Model Manager Workflow Console. Workflow process definitions contain the set of activities, participants, policies, statuses, and data objects that constitute a business task. Activities can contain properties and comments so that you can share information with other users or make notes. The status that you select when you complete an activity determines the next activity in the workflow process.

A workflow forms the basis for model-management tasks through workflow activities, from data collection and model building to production conversion and model performance monitoring. Teams use a workflow to hand off tasks in a serial manner across multiple teams and personas.



Display 5. Model Management Workflow Diagram in SAS Workflow Studio

From the SAS Model Manager client application, you can launch the Workflow Console to create a new workflow or view the workflow for a version, manage all workflows, and view your workflow inbox to work with activities.

The screenshot displays the SAS Model Manager Workflow Console interface. The main window is titled "SGF 2013 Workflow 1-1". It features a left-hand navigation pane with "Workflow Definitions", "Workflows", and "Activities". The central area shows a list of activities with their names, states, and claimers. Below this is a process flow diagram for the "mdingradminusers" workflow, showing a sequence of activities from "M01: Define Data Sources" to "M10: Evaluate or Monitor Results". The right-hand pane contains "Properties" and "Participants" sections.

Activity Name	Activity State	Claimed By
M01: Define Data Sources	Completed	
M02: Build Models	Completed	
M03: Import Models	Completed	
M04: Evaluate Models	Completed	
M05: Select a Champion Model	Completed	
M06: Validate Models	Completed	
M07: Attach Documents for Model Approval Process (MAP)	Completed	
M08: Publish to Production	Completed	
M09: Run Production Models	Completed	
M10: Evaluate or Monitor Results	Completed	

Property	Value
Available statuses	Completed
Milestone ID	M1: Define Data Sources

Name	Workflow Role
mdingradminusers	Business administrator
mdingradusers	Potential owner

Display 6. SAS Model Manager Workflow Console—List of Workflow Activities and Process Flow Diagram

When a model management component is associated with an activity, you can perform one or more of the following tasks:

- import models
- view or evaluate models
- create and view reports to validate models
- set champion and challenger models
- add, view, or delete attachments
- publish models
- view performance results

For more information, see the SAS Model Manager product documentation page available at <http://support.sas.com/documentation/onlinedoc/modelmgr/>.

Model Validation

The **Reports Viewer** component enables you to create and view reports so that you can validate models before you promote them to a production environment. You can create reports in PDF, HTML, RTF, and Excel. Some reports are limited to PDF and RTF. With the release of SAS Model Manager 12.3, you can create and view reports when the **Reports Viewer** component is associated with an activity. Reports can be viewed only in the Workflow Console for SAS Model Manager 12.1. They must be created in the client application.



Display 7. Reports Viewer Component in SAS Model Manager Workflow Console

Display 8 shows an example of a delta report that compares the input, output, and target variable attributes for each of the variables that are used to score the champion model (Scorecard_Loans) and the challenger model (NNMLP_Loans). Delta reports display the differences in the variables of competing candidate models. The report output is a table that groups the variables by the variable name. For each variable, the report lists the attribute value for each model and whether the attribute value is the same or different from the other attribute values.

SAS Model Manager
Project: Auto Loan Default Version: Q4
Number of Differences: 91
User: sasdmo

Role	Name	Variable Attribute	NNMLP_Loans	Scorecard_Loans	Difference
	LTV	Label			
	LTV	Format			
	LTV	Level	INTERVAL	INTERVAL	
	P_BAD0	Type	N	N	
	P_BAD0	Length	8	8	
	P_BAD0	Label	Predicted: bad=0	Predicted: bad=0	
	P_BAD0	Format			
	P_BAD0	Level	INTERVAL	INTERVAL	
	P_BAD1	Type	N	N	
	P_BAD1	Length	8	8	
	P_BAD1	Label	Predicted: bad=1	Predicted: bad=1	
	P_BAD1	Format			
	P_BAD1	Level	INTERVAL	INTERVAL	
	Q_BAD0	Type	N		X
	Q_BAD0	Length	8		X
	Q_BAD0	Label	Unadjusted P: bad=0		X
	Q_BAD0	Format			
	Q_BAD0	Level	INTERVAL		X
	Q_BAD1	Type	N		X
	Q_BAD1	Length	8		X
	Q_BAD1	Label	Unadjusted P: bad=1		X
	Q_BAD1	Format			
	Q_BAD1	Level	INTERVAL		X
	REV_UTIL	Type		N	X
	REV_UTIL	Length		8	X
	REV_UTIL	Label			
	REV_UTIL	Format			
	REV_UTIL	Level		INTERVAL	X
	SCORECARD_BIN	Type		N	X
	SCORECARD_BIN	Length		8	X
	SCORECARD_BIN	Label			
	SCORECARD_BIN	Format			
	SCORECARD_BIN	Level		INTERVAL	X
	SCORECARD_POINTS	Type		N	X
	SCORECARD_POINTS	Length		8	X
	SCORECARD_POINTS	Label			
	SCORECARD_POINTS	Format			
	SCORECARD_POINTS	Level		INTERVAL	X
	SCR_AGE_OLDEST_TR	Type		N	X
	SCR_AGE_OLDEST_TR	Length		8	X
	SCR_AGE_OLDEST_TR	Label			
	SCR_AGE_OLDEST_TR	Format			
	SCR_AGE_OLDEST_TR	Level		INTERVAL	X
	SCR_BUREAU_SCORE	Type		N	X

Version location: /MMRoot/Credit Risk/Auto Loan Default/Q4

Display 8: HTML Output for the Delta Report

The **Attach Documents** component enables you to add, view, or delete documents that are part of the model approval process. The attachments are stored in the Documents folder, which is located within a project or version in the model repository.



Display 9. Attach Documents Component in SAS Model Manager Workflow Console

Model Promotion

The **Publish Models** component enables you to publish the project's champion model and challenger models that are associated with the DATA step score code type to a configured database or to the SAS Metadata Repository. SAS Model Manager uses the SAS Scoring Accelerator and SAS/ACCESS[®] interface to the database to publish models to the database. SAS Scoring Accelerator takes the models from SAS Model Manager and translates them into scoring files that can be deployed inside the database. After the scoring files are published to the database, they take advantage of SAS[®] Embedded Process to run the scoring model.

Publish destination: Teradata Publish

Publish Options

Publish method: SAS Embedded Process Scoring function

Select the models to publish, and specify a publish name for each model.

Select	Model Name	Role	Version	Model Type	Publish Name	Date Published
<input type="checkbox"/>	Scorecard_Loans	✓ Champion	Q4	Classification	Scorecard_Loans_Auto_Lo	Jan 30, 2013 02:16 PM
<input type="checkbox"/>	NNMLP_Loans	▣ Challenger	Q4	Classification	NNMLP_Loans_Auto_Loan	Jan 30, 2013 02:15 PM

Replace scoring files that have the same publish name

Specify an identifier to add to the database target table for each model:

Auto Loan Default

Database Settings

Database server:

Database:

User ID: Password:

[More options...](#)

Claim Status: Completed Update and Close Close

Display 10. Publish Models Component in SAS Model Manager Workflow Console

For more information about the SAS Scoring Accelerator, see the SAS[®] In-Database Technology page available at <http://support.sas.com/documentation/onlinedoc/indbtech/index.html>.

MODEL DEPLOYMENT

Regardless of the complexity and type of model selected, deploying models to production is just as important as the science behind it. Models are algorithms that need to be translated into an environment that they are not inherently designed for. Databases are SQL-based; SAS models are DATA step-based. Translating models requires not only translating the equation but ensuring that all the data preparation and transformations execute in the database. The SAS Scoring Accelerator enables you to automatically translate and publish SAS Enterprise Miner and selected SAS/STAT linear models into database-specific code that is easily incorporated into production processes.

Score Using SAS (%INCLUDE)

This example runs a traditional SAS DATA step that is generated by SAS Enterprise Miner and by the models that are generated by the GENMOD, GLIMMIX, GLM, GLMSELECT, LOGISTIC, MIXED, PLM, and REG procedures in SAS/STAT:

```
data server.credit_input_sas;
  set server.credit_input;
  %include 'D:\score_directory\score.sas';
run;
```

Score Using the SAS Scoring Accelerator Executed inside the SAS Program Editor

This example runs the scoring code that is published in the Model Promotion stage. The scoring code is executed against an input data set and executes entirely inside the database. Modeling flows that are built in SAS Enterprise Miner are fully executed. Unlike models that are built using SAS/STAT, the entire transformation and scoring process is executed by using the SAS Scoring Accelerator. This is a batch process that can be automated and invoked from within SAS, or executed directly inside the database by using stored procedures or their equivalent.

```

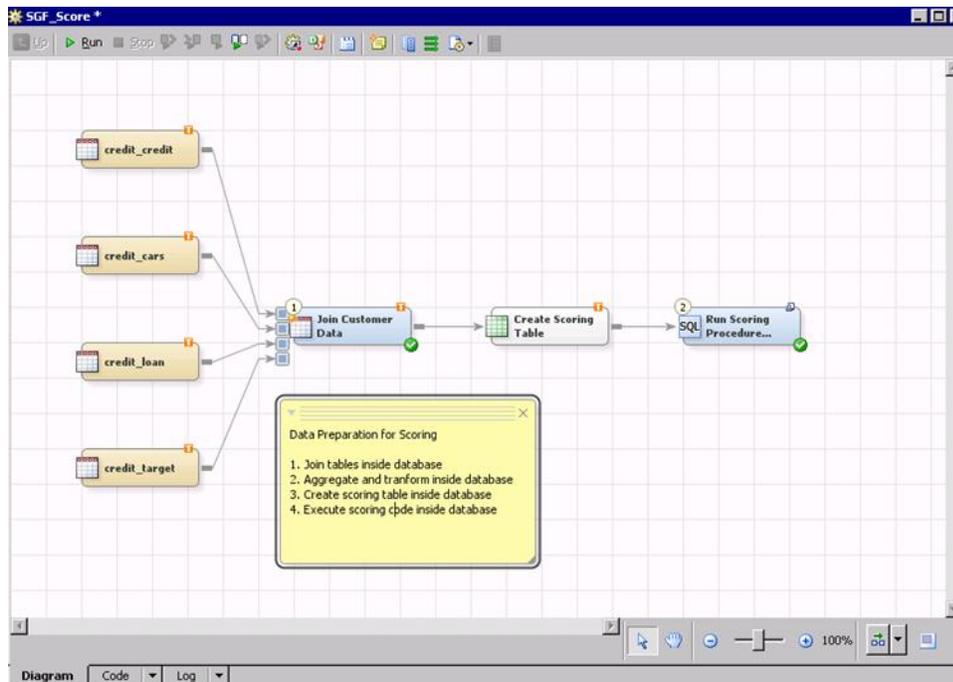
%let indconn=%str(user=xxx password=xxx server=xxx database=jowexl);

proc sql noerrorstop;
  connect to teradata (&indconn mode=teradata);
  execute(drop table jowexl.credit output)by teradata;
  execute(call SAS_SYSFNLIB.SAS_SCORE_EP(
    'MODELTABLE= jowexl."sas_model_table"',
    'MODELNAME=Scorecard_Loans_Default',
    'INQUERY=select "app_id","AGE_OLDEST_TR","BUREAU_SCORE","LTV","REV_UTIL"
      ,"TOT_DEROG","TOT_REV_LINE","TOT_TR"
    from jowexl.credit_input',
    'OUTTABLE=jowexl.credit_output',
    'OUTKEY=app_id','OPTIONS=VOLATILE=NO;' );
  ) by teradata;
  disconnect from teradata;
quit;

```

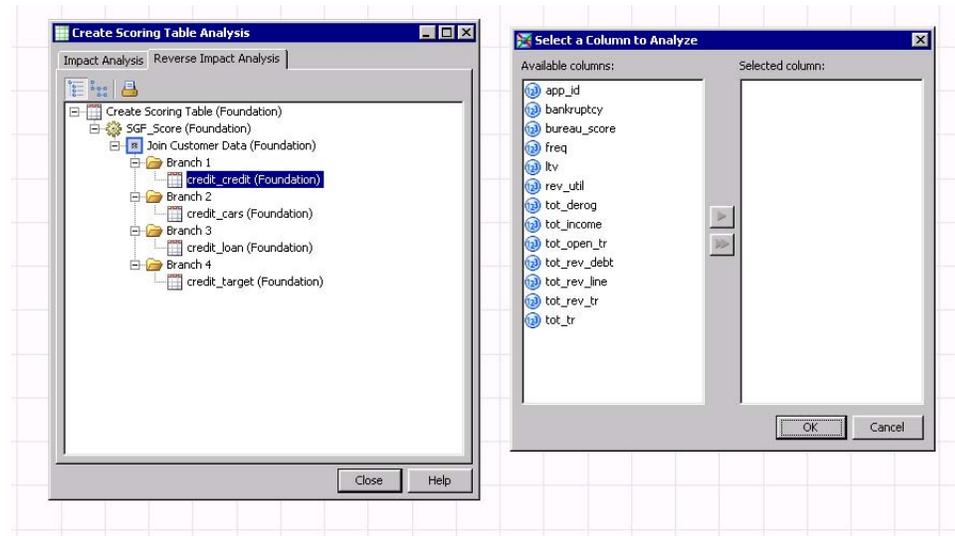
Automate Scoring Process Using SAS Data Integration Studio

This section describes how to create and schedule a job that executes all necessary scoring transformations. It uses the stored procedure that runs the published scoring model.



Display 11. Models Scored Using SAS Data Integration Studio

As you deploy to production, you might have to track data lineage. Using Impact Analysis in SAS Data Integration Studio enables you to trace the effect that changing a table or column has on upstream or downstream tables, columns, jobs, and transformations. This is especially useful for supporting compliance requirements and audits.



Display 12. Analyzing Data Lineage Using SAS Data Integration Studio

Deploy Models to Other Channels and Tools

After models are published to the SAS Metadata Repository, they are accessible by tools and services. Ensuring that your model is registered to SAS Metadata Repository allows it to be referenced within the SAS framework. For example, SAS[®] Real-Time Decision Manager accesses models that are registered in the SAS Metadata Repository to make real-time decisions. SAS Data Integration Studio enables you to convert a job to a stored process, which can be called from web services.

EVALUATING AND MONITORING RESULTS

It is absolutely critical that you monitor the impact your models are making on everyday decisions. SAS Model Manager provides several monitoring metrics for the continuous evaluation of models:

- model stability of inputs (characteristics) and score distributions
- model separation: Gini, ROC, K-S, lift, and so on
- MSE and residual diagnostics for prediction models
- probability of default and loss given default for risk models

You can generate model degradation profiles at any desired frequency. You can also define alert threshold conditions for any of the monitoring statistics with e-mail event notifications. SAS Model Manager also supports in-memory high-performance monitoring on Teradata and Greenplum appliances. You can set up alerts that help trigger an action to evaluate and recalibrate the model. You can also evaluate champion and challengers concurrently to help determine when perhaps a new challenger should be promoted as the champion. You can also create dashboard reports to evaluate the health of the entire model portfolio.



Display 13. Monitor Model Degradation across User-Defined Time Intervals

CONCLUSION

No matter who is involved in the process, the following factors can have a direct effect on the amount of time that you spend on each step of the process:

- **Data movement:** Data might be proliferated across environments, depending on the number of source systems involved. Tracking lineage becomes extremely important for conversion to production.
- **System resources:** Competing for system resources and workspace is challenging. This competition has given rise to the creation of the data scientist. Analyzing data requires staff who are not only logical thinkers, but also tech-savvy programmers.
- **Resource availability:** Deploying models requires many hand-offs, which in turn require resources that are ready to provide support. The more steps that are automated, the less the dependence on resource availability.

SAS' OPERATIONALIZING VISION

SAS is committed to delivering improved data management, analytical, and model management tools for streamlining decision making. SAS continues to invest heavily in massively parallel in-memory analytics to help ensure that no problem is too large or difficult for you to at least attempt to address. Key goals for SAS Model Manager include adding increased automation to support enormous analytical model factories and to support a wider class of analytical models.

SAS continues to focus on in-database model deployment. In 2013, SAS[®] Scoring Accelerator for Hadoop is scheduled to be released. As more and more models are developed to use data that reside in Hadoop, SAS Scoring Accelerator for Hadoop enables users to publish their models for deployment inside Hadoop, removing the requirement to hire Java MapReduce programmers. Code can automatically be converted and run inside the Hadoop framework.

SAS is also committed to delivering a truly integrated enterprise decision management platform in 2013: SAS[®] Enterprise Decision Management. This platform provides the business rule-writing interfaces that line-of-business users need; the scoring and strategy building that business leaders want to have created; the integration and execution environments that enable you to deploy the same logic in batch, in-database, and in real time; and the integrated monitoring that provides application or strategy, system, and economic views of what is happening in a particular environment or business process.

SAS is in the process of developing the decision management vision and platform that cover all aspects of the decision management life cycle, from capturing transactions and events, through applying business rules and scoring models, to ultimately completing the feedback loop that brings the *decided* data back to the data scientist for the next decision cycle.

SAS Enterprise Decision Management consists of the following integrated components:

- interactive data management and quality tools with impact analysis and lineage
- business rule design and management for convenient rule authoring and testing interfaces
- model management and validation for unified model management and monitoring
- decision services integration for combining rules and predictive analytics with control flow
- workflow for the control and policy management that are needed to control the flow of strategies
- multi-target deployment options, including batch, in-database, web-service, and real-time options
- monitoring and refreshing for decision service management, model retraining, and enabling decision services to challenge champions

REFERENCES

- Chu, Robert, David Duling, and Wayne Thompson. 2007. "Best Practices for Managing Predictive Models in a Production Environment." *Proceedings of the SAS Global Forum 2007 Conference*. Cary, NC: SAS Institute Inc. Available at <http://www2.sas.com/proceedings/forum2007/076-2007.pdf>.
- SAS Institute Inc. 2012. "SAS® Enterprise Data Integration Server Fact Sheet." Cary, NC: SAS Institute Inc. Available at <http://www.sas.com/resources/factsheet/sas-enterprise-data-integration-server-factsheet.pdf>.
- Tobias, R. and W. Cai. 2010. "Introducing PROC PLM and Postfitting Analysis for Very General Linear Models in SAS/STAT 9.22." *Proceedings of the SAS Global Forum 2010 Conference*. Cary NC: SAS Institute Inc.

ACKNOWLEDGMENTS

The authors express sincere gratitude to the SAS Model Manager and SAS Scoring Accelerator developers, testers, and also to our customers.

RECOMMENDED READING

- Taylor, James. May 2010. *Operational Analytics: Putting Analytics to Work in Operational Systems*. BeyeNETWORK. Available at http://www.decisionmanagementsolutions.com/attachments/155_BeyeNetworkOperationalAnalyticsResearchReport.pdf.
- *SAS 9.3 In-Database Products: User's Guide*. Available at <http://support.sas.com/documentation/onlinedoc/indbtech/index.html>.
- *SAS Model Manager 12.1: User's Guide*. Available at <http://support.sas.com/documentation/onlinedoc/modelmgr/>.
- *SAS High-Performance Analytics Server 12.1: User's Guide*. Available at http://support.sas.com/documentation/onlinedoc/securedoc/index_hpa.html.
- *SAS Data Integration Studio 4.5: User's Guide*. Available at <http://support.sas.com/documentation/onlinedoc/etls/index.html>.

CONTACT INFORMATION

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

Jonathan Wexler
SAS Institute Inc.
100 SAS Campus Drive
Cary, NC 27513
E-mail: jonathan.wexler@sas.com

Wayne Thompson
SAS Institute Inc.
100 SAS Campus Drive
Cary, NC 27513
E-mail: wayne.thompson@sas.com

Kristen Aponte
SAS Institute Inc.
100 SAS Campus Drive
Cary, NC 27513
E-mail: kristen.aponte@sas.com

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

Other brand and product names are trademarks of their respective companies.