

Paper 222-31

# The Value of Real Time Scoring Technology using SAS®

Young Chun KIM, IMC Korea Ltd. Co., Seoul, Korea

## ABSTRACT

Imagine having built a Data Mining model to detect fraudulent credit card transactions. If this model is applied once a day by the analyst to determine the nature of transaction, this implies the loss of preventing fraudulent transactions at the moment it occurred due to the fact that the model could not be applied in “real-time”, in another words, at the moment of transaction. To overcome this loss, the key technology required is to be able to systemize the determination process of normal against fraud at the moment it requests for authorization. SAS technology enables to build this real-time detection system.

The biggest challenge of Call-Centers is to balance the distribution of calls at peak time. Being able to direct the customer call to the most appropriate Call-Center agent based on the customer needs will be the basic principle of distribution. For example, the customer forgot to pay his/her loan interests, so would like to ask how much he needs to pay at the time of the call, or would like to ask the processing status of his/her loan application. Instead of the agents responding individually, if IVR could take over these kinds of requests automatically and let the agents allocate more time to cross selling and marketing, no one would doubt the effectiveness of operating a Call-Center. Of course, the prerequisite here would be to be able to predict the purpose of the call in order to provide the most suitable service, which is feasible by using Data Mining techniques. In terms of technology, the challenge is to systemize the result of Data Mining for it to be applied to every call coming in.

Data Mining is commonly used in batch. However more and more business is requiring the application in real-time to meet the needs of customers at the moment it occurs. Common examples were mentioned above – credit card transactions and call-center strategic routing. This paper will present how to systemize a real-time Scoring Engine for Data Mining.

A system for real-time scoring consists of 4 components.

### 1) ANALYTICAL DATA MART

This is a data set in order to build a model. It includes a target variable, input variables and derived variables to increase the accuracy of the model. Summarized tables are also included for exploration. During real-time scoring, look up tables are required to calculate the scores. We name this look up table a “Reference Operation Table”.

### 2) MODEL MANAGEMENT

This covers the full process of building a model - metadata about the model such as variables used, options used in SAS Enterprise Miner, C code generated using the C\*Score Node.

### 3) REAL-TIME SCORING ENGINE

In the case of fraud detection, a real-time scoring engine is provided to determine authorization. This engine is composed of C or Java program that can sit on top of the existing authorization system and a reference table to refer during scoring.

In case of Call-Center, a Data Mining model to predict the purpose of the call needs to be processed using the real-time scoring engine.

The prerequisites of a real-time scoring engine are

- be able to calculate the score where a Base SAS environment does not exist
- the resource to calculate the score need not be a burden to the existing operation system
- must be easy to migrate and operate the model
- the accuracy of the model must be maintained as expected
- the input data must be the most recent one

### 4) POST REPORTING & MONITORING

Reports to understand the expected accuracy, the score distribution, lift values, etc are provided. In the case of Fraud Detection, a monitoring system for the fraud agents to view the fraud score and authorization status are provided as well.

## INTRODUCTION

Generally the generation of score is being built through batch-processing in the field of Data Mining.

If the score were required in real-time, there would be a requirement of a certain decision at a time an event (a

certain problem, transaction) occurred.

For example, take the case of credit card. When the customer purchases the product, the store requests the approval of Credit Card Company through VAN. And if the request of approval is normally done, the purchase of credit card is completed by the signature of the customer.

However, there would occur any requirement if the request of approval from the store is normal at the moment the approval is requested.

If real-time scoring is available, we could identify if the request of approval of credit card is normal through **the fraud score** when it occurred.

Another case is to allocate the concentration of calls from customers at a Call-Center. When a customer makes simple inquiries such as the current month billing, current arrears on interest and result of process for current paid, if the one stop process is taken in the IVR (Intelligent Voice Response) without the response of Call-Center's agents, this could bring relatively quick process and cost reduction.

Instead, the Call-Center's agents can generate more value added in focusing consultation to sell other products such as cross-sell, up-sell and difficult response.

When a customer calls, if the IVR send an appropriate voice message, the purpose of calling from customer should be predicted accurately at the same time. The purpose of calling could be identified by using **the score for the purpose of calling** by the types of calling.

Generally, the technology of Data Mining is used to extend the application in the area of batch-processing.

But when a certain event occurred, the decision-making should be done in real-time under the business environment recently. So the requirement of using the technology of Data Mining is increased.

The following will explain how to realize the real-time technology of Data Mining. A system for real-time scoring consists of 4 components: analytical data mart, real-time scoring engine, post reporting and monitoring.

### ANALYTICAL DATA MART

A Analytical Data Mart is the base of Data Mining models.

This Analytical Data Mart can be different depends on business conditions required in each industry.

The most important things in Analytical Data Mart are the variables and unit which determines the value of variables.

In most of the cases, the period of value applied in Analytical Data Mart is longer on a monthly or daily basis than the time to approve card transaction. This characteristic of period is very useful in batch-processing environment.

However, we need a different characteristic in real-time processing environment and this is called as 'Reference Operation Table'.

We realized a real-time processing environment by generating two representative tables. One is a '**base model table**' to use in the process of building the Data Mining models of batch-processing environment and the other is a '**reference operation table**' to use in the process of building the real-time score of real-time processing environment.

The figure 1 is the layout of is '**base model table**'.

Figure 1. Base Model Table Layout

Key	Input Variables	Target
	The Input variables consist of the related information which can be explained the Target.	

\*A Key is a set of IDs determined by the analyzing unit.

Input variable used as we generate the Data Mining model can build information through batch-process way by using historical data. However, input variables in the real-time scoring process have been used as 'Reference Variables' and 'Operation Variables'. 'Operation Variables' is updated through real-time on legacy environment. 'Reference Variables' is updated from BI system periodically. See the Picture 2 Reference Operation Table Layout.

The figure 2 is the layout of is '**reference operation table**'.

Figure 2. Reference Operation Table Layout

Key	Input Variables		Score
	<i>Reference Variables</i>	<i>Operation Variables</i>	

\*\*A Key is a set of IDs determined by transactions.  
A Score is the value corresponding to Tr by a Model.

## MODEL MANAGEMENT

The main function of Model Management is the process of making a model.

We used SAS Enterprise Miner as a model generator and used the '**base model table**' mentioned in Analytical Data Mart as input data for Input Data Source of Enterprise Miner.

After all, '**base model table**' has the necessary information to generate Data Mining models which makes possible real-time processing. In most of the cases for building Data Mining models, they use SAS Code to generate scores in Data Mining process using SAS Enterprise Miner.

But, we converted SAS Score code into C code and used it instead of SAS Score code on real-time scoring processing.

Or it can be used SAS Score code to implement real-time scoring system. But, here it is omitted explanation about this part. If you are interested, don't hesitate to contact me; I'm willing to share the information.

We used C\*Score Node offered in SAS Enterprise Miner to convert SAS Score Code into C Score Code.

In SAS 9 and Enterprise Miner 4.3 or higher C\*Score Node is changed to Score Converter.

C\*Score Node has been very useful to us in building a real-time Data Mining models.

If SAS Enterprise Miner had not offered the code converting function in C\*Score Node, we would have wasted a lot of time on applying the solution like real-time scoring process.

It only took you few minutes (seconds) to convert C Score Code from SAS Score Code.

The figure 3 is an example of Score Converter Node in SAS Enterprise Miner.

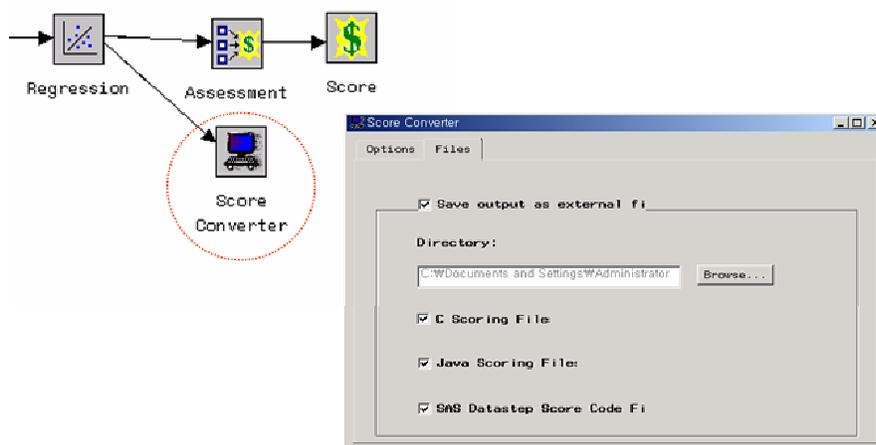


Figure 3. An Example of Score Converter Node in Enterprise Miner

## REAL-TIME SCORING ENGINE

To implement Real-Time Scoring Engine in the legacy environment, we should take some points into account.

In most cases, a scoring engine is applied in legacy environment without Base SAS. Therefore, scoring processes should be possible without SAS. If a process is added in legacy environment, we should think that there would be a burden the existing operation system. Next, real-time Data Mining models should migrate to and operate in the legacy environment. We applied real-time Data Mining model to Authorization system of Credit Card Company, and in case of Call-Center we did so to UNIX-based transaction processing system. Applying real-time scoring does not mean we can ignore the accuracy of the models. The core reason applying real-time scoring is that ROI (Return of Investment) of model application is highly realized. Lastly, the input data must be the most recent one. The figure 4 shows the components of Real-Time Scoring Engine in legacy environment.

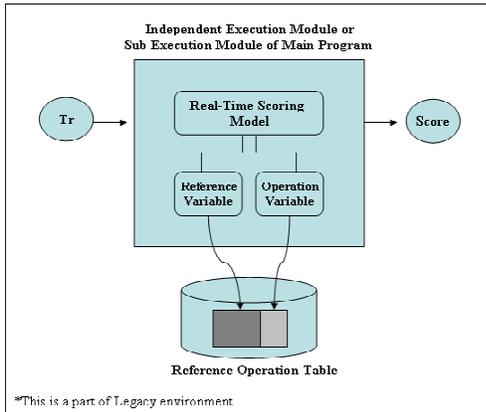


Figure 4. The Diagram of Real-Time Scoring Engine

Mostly, Data Mining Scoring Engine is composed of model formula and input variables. Especially when we use SAS System, model formula is SAS Code and the value of input variable is in SAS Dataset. But it is different in real-time processing environment. Model formula is not SAS code and the value of input variable may be saved in RDMS or in system memory. Our real-time Scoring Engine was composed of model formula, 'reference variable', and 'operation variable'. Model formula is the compiled executing program that converted SAS Score Code to ANSI C Code by utilizing C\*Score Node, and 'Operation variable' is the one that is saved in system memory. 'Reference variable' is the one that is done in RDMS.

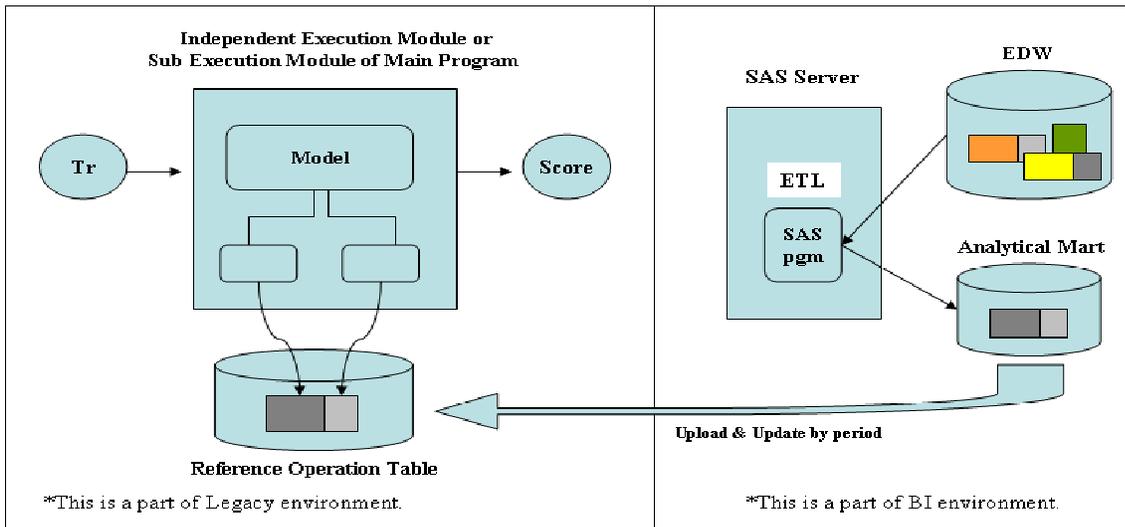


Figure 5. The Diagram of Legacy and BI system

**POST REPORTING & MONITORING**

Reports to understand the expected accuracy, the score distribution, lift values reports, the monitoring information, etc are provided.

We need information to know the status of scoring in real-time processing environment. In particular, the information needed for monitoring is very important to check to the real-time scoring status in the case of Fraud Detection in Finance Service. For example, the fraud detection agents of Credit Company can look at if the status of approval of credit card is normal or not by scoring point.

Also, this could be information to estimate the Score Cut-off Strategy. This is used to resolve the claims, which could occur from customers by rejecting the use of credit card even if the request of approval is normal.

As for Call-Center, we should know if the purpose of calling is expected correctly.

If this expectation were correct, customers wouldn't connect to Call-Center agents after hearing the voice message.

Most of the BI systems are applied in the environment of post-processing or batch-processing or to create reports after business hour. Above cases mentioned applied to BI system at the moment of decision-making.

We would like to apply our case to more new business territory as well as to the domain where our case is now being applied.

#### POINTS TO CONSIDER FROM THE VIEWPOINT OF SYSTEM IN THE FUTURE

- To appraise a lifetime of models being applied in legacy, and to automate the function of re-modeling process: we should apply the pattern to real-time scoring models as soon as possible because the pattern of fraudulent usage changes so quickly in the field of Credit Card.
- The sharing function of meta-data information
- Applying Java code and alerting the efficiency of models, etc

#### AN EXAMPLE OF REAL SYSTEM

The following is our image of the prevention system of fraudulent usage of credit card.

Our FDMS (Fraud Detection Management System) consists of Analytical Data Mart, predictive models (templates depends on sectors of fraudulent card usage), operation strategies, and modeling tool like SAS. See the Figure 6.

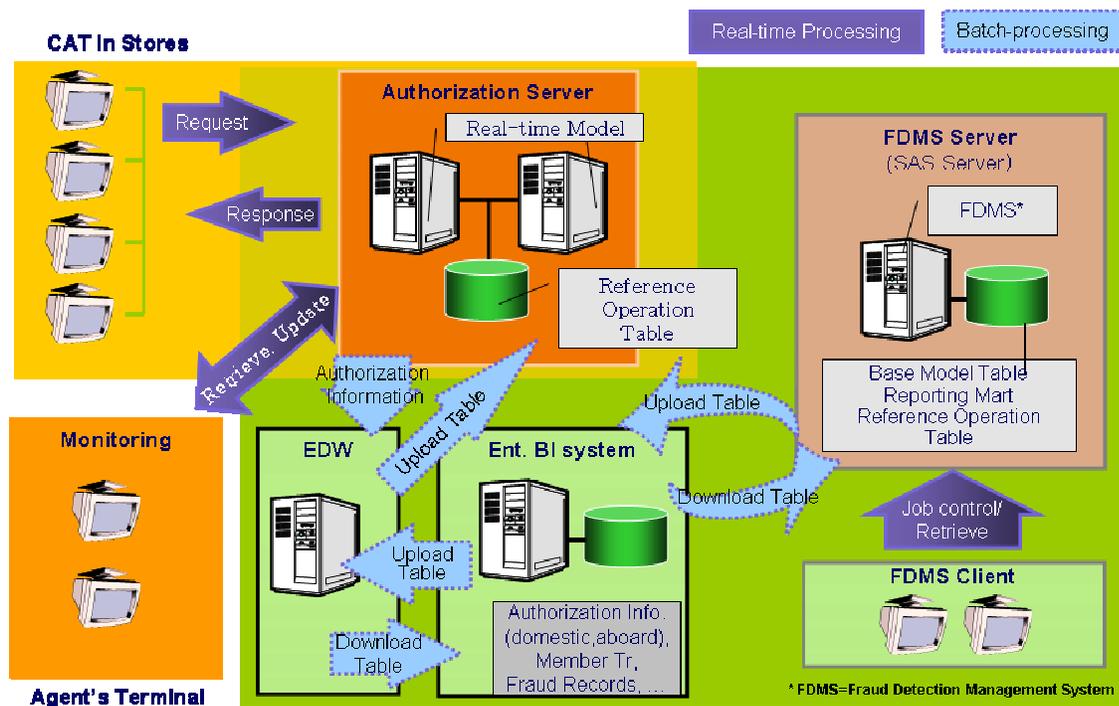


Figure 6. The System Diagram of Credit Card Company

Using SAS V8.2, Enterprise Miner 4.1, we have applied real-time Data Mining Model by converting SAS Score Code to C Score code to predict fraudulent usage of credit card in legacy environment. We also have applied real-time Data Mining Model using SAS V8.2, Enterprise Miner 4.1 to predict the client's purpose of calling in Call-Center in Finance Service.

If you need specific information, please contact me via the number or email below. I am willing to share that information with you.

## CONCLUSION

Typically when we think of Business Intelligence systems, we are more apt to think of post processing or batch-processing data and creating reports after working hours are closed. However cases mentioned above are examples where business intelligence solutions are applied in real-time at the moment decisions need to be made, domains which the market is demanding more and more and where SAS can fully play its role.

## REFERENCES

SAS Institute Inc. 2005. Enterprise Miner C and Java Score Code Basic. Cary, NC: SAS Institute Inc. Available [http://support.sas.com/documentation/onlinedoc/miner/cjscore\\_em5.pdf](http://support.sas.com/documentation/onlinedoc/miner/cjscore_em5.pdf)

## ACKNOWLEDGMENTS

I would like to thank my colleagues at INWAVE. And, particularly, I would like to thank So Rah Kim and Hye Chung Lee for their assistance and many ideas.

## CONTACT INFORMATION

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

Young Chun KIM  
IMC Korea Ltd. Co.  
#1601 Seocho World O/T Bldg.  
1355-3 Seocho-Dong, Seocho-Gu  
Seoul, 137-070  
Republic of Korea  
Work Phone: +82-2-3487-4191  
Fax: +82-2-3487-4192  
E-mail: [anether@paran.com](mailto:anether@paran.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.