




## SAS/Warehouse Administrator as ETL Tool

Rei Takahashi  
Solution Planning Center  
SAS Institute Japan Ltd.  
Jpnrat@jpn.sas.com



## SAS/Warehouse Administrator 優位性

- 完全なメタデータ管理
- コードの修正が容易
- ログ書き出し、ジョブの再スタート、途中スタート
- データとSQLのマルチパス
- データ/処理ホスト選択の柔軟性
- 各種DB対応の柔軟性
- Extractエンジン
  - データ重複の最小限化
  - 簡単で柔軟な項目設定
  - SAS/ACCESSソフトウェアの利用
- Transformationエンジン
  - 効率的なマッピング
  - 単数/複数テーブルに有効なWhere文
  - 簡単縦横変換と～毎の集計
- Loadエンジン
  - 柔軟なロード先の選択と各種DBバルクロード対応
- 多様なアドインツールでユーザーのニーズにこたえる



## SAS/Warehouse Administratorの優位性

### 完全なメタデータ管理

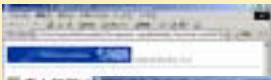



SAS/Warehouse Administratorでは、保存されたメタデータを利用してHTMLレポートを自動生成します。(メタデータHTML書き出しアドイン利用)

このレポートでは、メタデータ参照の対象のオブジェクトに関する①説明②ノート③項目定義(名前、説明、種類、長さ、出力形式、入力形式)④インプットデータソース⑤テーブル依存関係⑥ソースコード⑦コンタクト(所有者/管理者の名前、連絡先)の情報が得られます。(図1、2 参照)

テーブル依存関係では、依存しているテーブルのメタデータへのリンクが自動的にはられるので、オブジェクトのデータソースのメタデータにも簡単にたどり着くことが可能です。(図3 参照)

ソースコードは、SAS/Warehouse Administratorがインストールされてなくても、そのままコピー&ペーストでSASシステム上で利用できるようになっているので、柔軟にシステム資源を無駄なく利用できます。(図3、4 参照)

コンタクトでは、所有者/管理者の名前と連絡先が表示されるので、オブジェクトに関して問い合わせたい場合は、表示されたEメールアドレスや電話番号が利用できます。(図4 参照)

☆自動HTMLレポートによって手作業メタ管理を削除！

The Power to Know



## SAS/Warehouse Administratorの優位性

### 完全なメタデータ管理

SAS/Warehouse Administratorでは、保存されたメタデータを応用して、影響分析を行なうことができます。(インパクト分析アドイン或いはリバースインパクト分析アドイン利用)

インパクト分析アドインでは、データソースに変更を加えた場合、そのデータソースを利用するオブジェクトにどのように影響を与えるのかをビジュアルで表現します。(図1、2 インパクト分析アドイン画面サンプル 参照)

リバースインパクト分析アドインでは、作成したオブジェクトの特定の項目は、その元となったデータソースのどの項目と関連しているのかをビジュアルで表現します。

これらの機能は、SAS/Warehouse Administratorの完璧なメタデータ管理によって成り立つもので、**DWHのメンテナンスにおいては、欠かせない機能**になっています。

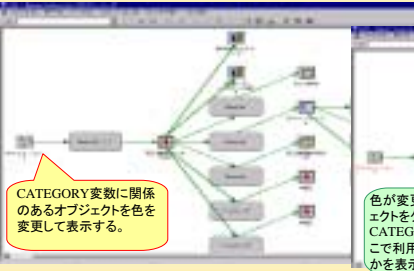





図1 インパクト分析アドイン画面サンプル

CATEGORY変数に関係のあるオブジェクトを色を変更して表示する。

図2 インパクト分析アドイン画面サンプル

色が変更されたオブジェクトをクリックするとCATEGORY変数がどこで利用されているのかを表示する。

The Power to Know



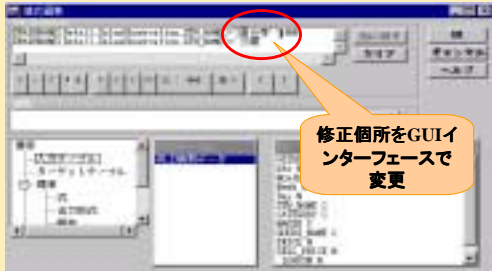
## SAS/Warehouse Administratorの優位性

### コードの修正が容易

• ジョブが処理中エラーになった場合、ログが残るので、エラー個所の修正が容易！修正方法は二つ；

- ①エラー個所に関連するGUIインターフェースから修正
- ②コード書き出し後に、直接修正

**①GUIで修正する場合**

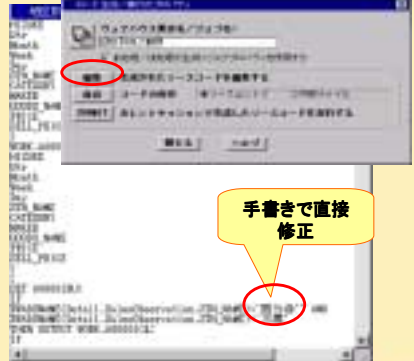


修正箇所をGUIインターフェースで変更

項目マッピングにおいて項目の内容定義(関数・数式定義など)ウィンドウ

**②直接で修正する場合**


コードは、GUIで設定したものは自動生成されるが、手書き編集が可能



手書きで直接修正

自動生成コード編集ウィンドウ

*The Power to Know*



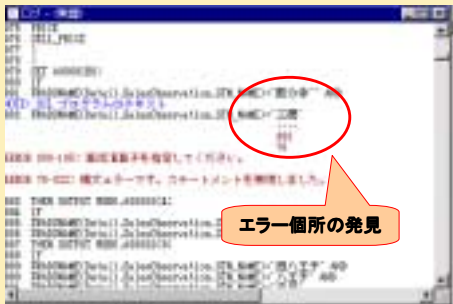
## SAS/Warehouse Administratorの優位性

### ログ書き出し、ジョブの再スタート、途中スタート

• かき出されたログを参照することによってエラー部分を発見

• 修正後は、中間ファイルが利用できるため、エラーになった場所から再スタートできる。

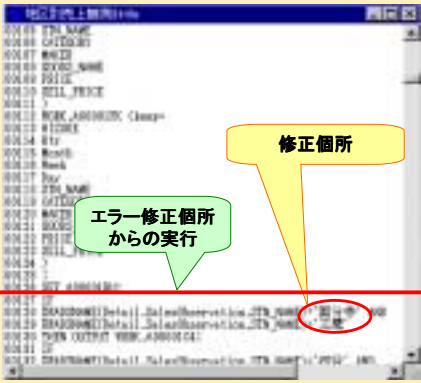
**エラーログの参照**



エラー個所の発見

ログウィンドウ

**対応個所の修正後、修正個所からスタート**



修正箇所

エラー修正箇所からの実行

自動生成コード編集ウィンドウ

*The Power to Know*

**SAS**  
in Intelligence

## SAS/Warehouse Administratorの優位性

### データとSQLのマルチパス

SQLで一度に複数テーブルを作成できるか？→できない  
**SAS言語だと一度のジョブで複数テーブルの作成が可能！**

SAS/Warehouse Administratorによって生成されるコードはSAS言語。SAS言語はデータ加工に非常に柔軟です。

SAS言語の処理は基本的にシングルスレッド。(マルチスレッド化にはSAS/CONNECTソフトウェアまたはSPDSソフトウェアのライセンスが必要です。)。一見ディスアドバンテージのように聞こえますが、他社ETLツールがマルチスレッド対応の主な理由は、SQL言語によってETLを行うため、左イメージのような複数のターゲットテーブルを作成する場合に、アウトプットの数に応じて何度もデータソースにバススルーする必要があり、それにかかる時間をマルチスレッドによって短縮するため。SAS言語では一度に複数のテーブルが作成できるので、バススルーも一度、よってシングルスレッドでも十分対応できます。しかもマルチスレッドで処理を行いたい場合は、コーディングを工夫することで可能です！

SQLやC言語でETLを行うツールは、必ずデータソースへのバススルーがアウトプットの数だけ発生しますが、SASはSAS独自の4GLであるため、1回のバススルーで複数のアウトプット生成が可能！

ETLをビジュアル化するプロセスウインドウ

**L: 複数テーブルのロード**  
**T: マッピング、データ加工**  
**E: 複数データソースからのエクストラクト**

The Power to Know

**SAS**  
in Intelligence

## SAS/Warehouse Administratorの優位性

### データ/処理ホスト選択の柔軟性

SASシステムは、MVSからWindowsまで多様なOSに対応しています。(図1 処理ホスト登録ウインドウ参照) このスケラペリティはSASのMVA(マルチベンダー・アーキテクチャ)の考え方にのっとるものです。

SAS/Warehouse Administratorでは、一度登録すると、どのホスト上でも自由にデータ保存や処理の実行をすることが可能になります。(図2 データ保存場所ホスト選択ウインドウ参照)

SAS/Warehouse Administratorはインストールされたホスト上で、ETL処理フローを作成し、処理はそれぞれのサーバーに実行命令を出すという形をとります。


多様なデータホスト  
処理ホスト登録が可能

登録したホストの選択

図1 処理ホスト登録ウインドウ

図2 データ保存場所ホスト選択ウインドウ

The Power to Know



## SAS/Warehouse Administratorの優位性

### 各種DB対応の柔軟性

SASシステムは、複数のDBに対してネイティブアクセスエンジンを提供しています。このスケーラビリティはSASのMEA(マルチ・エンジン・アーキテクチャ)の考え方にのっとるものです。

SAS/Warehouse Administratorでは、アクセスするDBを事前に登録しておきます。(図1 DB登録ウィンドウ参照)

Extractのフェーズにおいては、各種DBへSASシステムからテーブル参照が可能なので、テーブル選択の際にもSASのインターフェースだけで対応できます。(例1 SASシステムからOracleテーブル直接参照)

Loadのフェーズにおいては、各種DBのバルクロード機能のSASからの利用ができます。

SASシステムから直接参照できるデータの種類の種類	ACCESS99	ODBC	Sybase	CA-OpenIngres
	•BMDP	•OLEDB	•SPSS	
	•DB2	•ORACLE	•TERADATA	
	•FAMECHLI	•OSIRIS	•XML	
	•Infomix	•MS SQL Server	•XPORT ※(transport format file)	


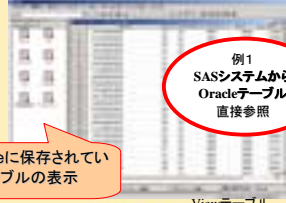


図1 DB登録ウィンドウ



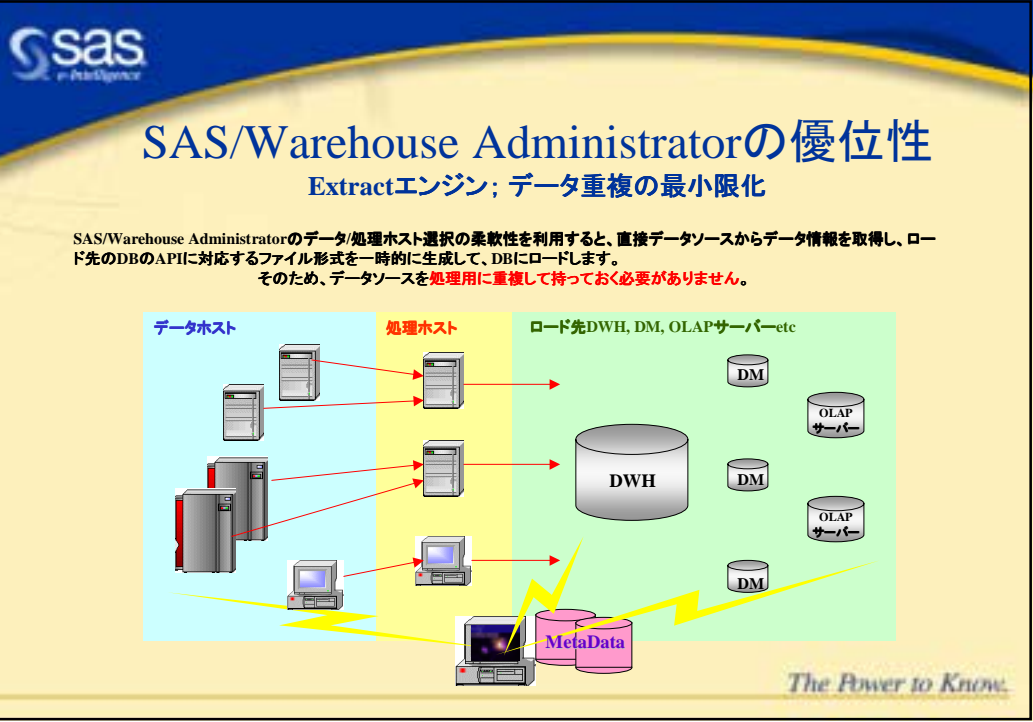
例1  
SASシステムから  
Oracleテーブル  
直接参照

Oracleに保存されている  
テーブルの表示

Viewテーブル

*The Power to Know*

※他社DBからのデータ取込のネイティブエンジンとして、Access to XXXのライセンスが別途必要です





**SAS**  
e-Intelligence

## SAS/Warehouse Administratorの優位性

Extractエンジン; 簡単で柔軟な項目設定

ターゲットテーブルの項目の追加・削除が非常に柔軟  
 ①項目の新しい定義や  
 ②他テーブルから項目名や定義の一括インポート  
 が可能なため、項目定義にかかる手間と時間削減に!

①項目の新しい定義 <例>値引率の項目の追加

◆ステップ1  
新規作成指定

◆ステップ2  
新規項目の定義

ターゲットテーブルの項目定義ウィンドウ

The Power to Know

**SAS**  
e-Intelligence

## SAS/Warehouse Administratorの優位性

Extractエンジン; 簡単で柔軟な項目設定

②他テーブルから項目名や定義の一括インポート

<例>項目定義が利用できる他テーブルからのインポート

◆ステップ1  
インポート元の選択

◆ステップ2  
インポートするテーブルとその項目の選択

◆インポート完了!

ターゲットテーブルの項目定義ウィンドウ

ターゲットテーブルの項目定義ウィンドウ

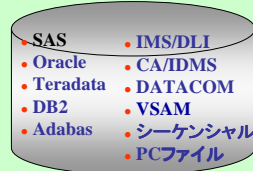
The Power to Know

# SAS/Warehouse Administratorの優位性

Extractエンジン; SAS/ACCESSソフトウェアの利用

SAS/ACCESSソフトウェアの多様なネイティブ・アクセス・エンジンで実現するデータアクセス

アクセスシナリオ1: メインフレーム MVS, VM/CMS

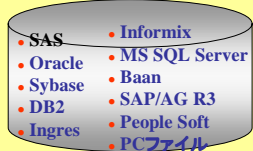


ネイティブ・アクセス・エンジン



メインフレーム MVS, VM/CMS,  
Hitachi, Fujitsu

アクセスシナリオ2: HP/SUN/AIX



ネイティブ・アクセス・エンジン



HP/SUN/AIX

V8.2 ステータス  
2001年5月現在

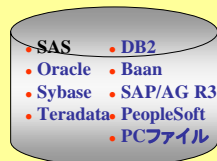
to Know

# SAS/Warehouse Administratorの優位性

Extractエンジン; SAS/ACCESSソフトウェアの利用

SAS/ACCESSソフトウェアの多様なネイティブ・アクセス・エンジンで実現するデータアクセス

アクセスシナリオ3: Windows9X, NT/NTS, OS2



ネイティブ・アクセス・エンジン

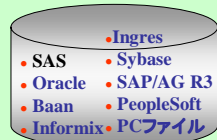
ODBC

OLE/DB



Windows 9X, Windows2000  
NT/NTS, OS2

アクセスシナリオ4: Compaq Tru64



ネイティブ・アクセス・エンジン

ODBC



Compaq Tru64

V8.2ステータス  
2001年5月現在

The Power to Know

**SAS** Intelligence

## SAS/Warehouse Administratorの優位性

### Extractエンジン; SAS/ACCESSソフトウェアの利用

SAS/ACCESSソフトウェアの多様なネイティブ・アクセス・エンジンで実現するデータアクセス

**アクセスシナリオ5; Intel ABI**

**アクセスシナリオ6; Vax/VMS**

V8.2 ステータス  
2001年5月現在

*The Power to Know*

**SAS** Intelligence

## SAS/Warehouse Administratorの優位性

### Transformationエンジン;効率的なマッピング

マッピング作業が効率的

- ①データソースとターゲットテーブルの項目名が同一で項目値をそのまま利用する場合は1-1マッピング機能を利用
- ②データソースとターゲットテーブルの項目名が異なる場合は、GUIを利用

①データソースとターゲットテーブルの項目名が同一で項目値をそのまま利用する場合の1-1マッピング機能利用例

◆ステップ1  
1-1マッピングの選択

マッピングウィンドウ; マッピングタブ; マッピング前

◆ステップ2  
データソースの選択とマッピングの実行

1-1 マッピングウィンドウ

①で対応できないものは、②で

◆1-1マッピング完了

マッピングウィンドウ; マッピングタブ; 1-1マッピング後

*The Power to Know*



**SAS**  
in Intelligence

## SAS/Warehouse Administratorの優位性

### Transformationエンジン;効率的なマッピング

②データソースとターゲットテーブルの項目名が異なる場合のGUI利用例

今回利用している関数はDATEPART関数。データソースの日付項目は年月日時分秒まで持っているが、年月日までの値だけを取ってきたい場合に利用される。SASから呼び出せる値の付け関数は160以上にのぼる。

◆ステップ1  
派生マッピングの選択

マッピングウィンドウ;マッピングタブ;GUIマッピング前

◆ステップ2  
利用する関数、数式などを設定

データを加工するための関数をポイント&クリック選択

◆派生マッピング完了

マッピングウィンドウ;マッピングタブ;GUIマッピング後

項目マッピングにおいて項目の内容定義(関数・数式定義など)ウィンドウ

ユーザーが作成したシンボル、数式、出力形式(フォーマット)※、関数、マクロなどを格納して、このインターフェースから恒久的に利用することも可能

※フォーマットとはSAS独自のデータの書き出し形式のこと。例えば、002のような数値データにWORDSEVフォーマットを使うとsix hundred ninety-twoのような文字列に変換して書き出すことが可能。あらかじめSASシステムで提供されている数多くのフォーマットに加え、ユーザーが特定のフォーマットを作成・利用することができる。しかし、フォーマットを使っても元データに変更はないため、うまくフォーマットを利用するとデータの容量を削減することが可能だ。例えば、あるテーブルのなかで商品コードですべてもってて、コード001なら電子レンジといった形でフォーマットを作成すると元データは10バイトで済む。(電子レンジでは10バイト必要)

The Power to Know

**SAS**  
in Intelligence

## SAS/Warehouse Administratorの優位性

### Transformationエンジン;単数/複数テーブルに有効なWhere文

データソースのデータにWhere文をあてる機能は、  
**単数/複数ターゲットテーブルに有効**  
データソースにwhere文を当てる例

◆ステップ1  
ターゲットテーブルの選択

◆ステップ2  
Where文の対象になる項目を選択して値を入力

図1 プロセスウィンドウ

マッピングウィンドウ;出力データタブ

値式編集ウィンドウ

図1のように、1つのマッピングオブジェクトから複数のテーブルを作成する場合、その複数テーブルの全てに当てはめるWhere文と、それぞれのテーブルに異なるWhere文の両方を設定可能

The Power to Know

**SAS/Warehouse Administratorの優位性**  
Transformationエンジン; 簡単縦横変換

SASのデータ加工エンジンは正に柔軟性に富んでいます。  
行列縦横変換もSAS言語を利用すれば自由自在

トランザクションテーブル(下記)があって、施設コードごとに商品の売上を見たい場合、行列の縦横変換が必要

縦横変換サンプルコード

データビューウィンドウ

施設ごとソート

一つ一つの商品が項目に!

コード編集ウィンドウ

ソート  
対象テーブルの指定  
施設コードでソート  
縦横変換  
対象テーブルの指定

ソートしたテーブルを別テーブルとして保存

縦横変換したテーブルを別テーブルとして保存

各種パラメーターの設定  
分析変数: 金額  
~毎: 施設コード  
縦横変換項目: 商品名

特定の項目の値ごとに、行列縦横変換がわずか10行程度のコードで可能に!

The Power to Know

**SAS/Warehouse Administratorの優位性**  
Transformationエンジン; 簡単MDDB定義/修正

MDDBの作成、保存もGUIで行います。


図1の項目をポイント&クリックでキューブ定義するだけなので、データソースに変更が加わった場合の修正が非常に簡単に行なえます。

ポイント&クリックでのMDDB定義

MDDB定義ウィンドウ

ETLをビジュアル化するプロセスウィンドウ

The Power to Know



## SAS/Warehouse Administratorの優位性

### Loadエンジン;柔軟なロード先の選択と各種DBバルクロード対応

SAS/Warehouse Administratorでは、ロード先がSASだけでなく、もちろん各種DBMSに対応しています。(図1 保存形式指定ウィンドウ参照)

ロード方式には更新と追加があります。(図2 ロード方式指定ウィンドウ参照)  
すでにあるテーブルの一部の変更(アップロード)は、SAS言語によるコーディングで対応します。

ロード時間はSASDWHへの高速ローディングのほか、各種DBのバルクロード機能を用い、高速ローディングを実現します。

ロード先に関しても、登録したホストを選択して保存します。  
(図3 保存先指定ウィンドウ参照)




図1 保存形式指定ウィンドウ




図2 ロード方式指定ウィンドウ

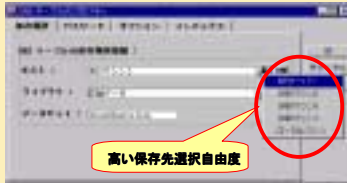


図3 保存先指定ウィンドウ

2001年5月現在

ロード対応DB

- DB2
- Informix
- MS SQL Server
- Oracle
- OracleRDB
- SAS
- SYBASE
- Teradata
- ODBC
- OLE DB

*The Power to Know*



## SAS/Warehouse Administratorの優位性

### 多様なアドインツールでユーザーのニーズにこたえる

**現在入手できるアドイン**

- Oracleロードアドイン
- Teradataロードアドイン
- Excel ロードアドイン
- 変更データ取得デルタロードアドイン
- カンマ区切りファイルロードアドイン
- カーディナリティ生成アドイン
- 制約条件設定アドイン
- DB2アドイン
- フォーマット生成アドイン
- HOLAPラッキングアドイン
- インパクト分析アドイン
- DDL(データ定義SQL)インポートアドイン
- ジョブ依存関係レポートアドイン
- ロードウィザードアドイン

- テーブルの縦横変換アドイン
- 情報配信管理アドイン
- フラットファイル読み込みアドイン
- データモデル登録/ブラウズアドイン
- メタデータナビゲーター
- リバースインパクト分析
- データステップ(SAS言語)マッピングアドイン
- データステップ(SAS言語)マッピングアドイン、ジョインとマージ用
- 複数ジョブのソースコードの書き出しアドイン;
- テーブル登録アドイン
- 標準項目管理アドイン
- 代理キー生成アドイン
- 2テーブルSQLジョイン
- UNIXジョブスケジューラー
- Enterprise Miner アドイン

2001年5月現在

ユーザーによる  
アドイン開発のために  
APIを開放しています!

提供されているアドインは  
フリー!

*The Power to Know*

**SAS**  
in Intelligence

## SAS/Warehouse Administratorの優位性

多様なアドインツールでユーザーのニーズにこたえる

メタデータHTML書き出しアドイン サンプル画面

☆自動HTMLレポートによって  
手作業メタ管理を削除！

*The Power to Know.*

**SAS**  
in Intelligence

## SAS/Warehouse Administratorの優位性

多様なアドインツールでユーザーのニーズにこたえる

インパクト分析 サンプル画面

CATEGORY変数に関係  
のあるオブジェクトを色を  
変更して表示する。

色が変わったオブジェ  
クトをクリックすると  
CATEGORY変数がどこで  
利用されているのかを  
表示する。

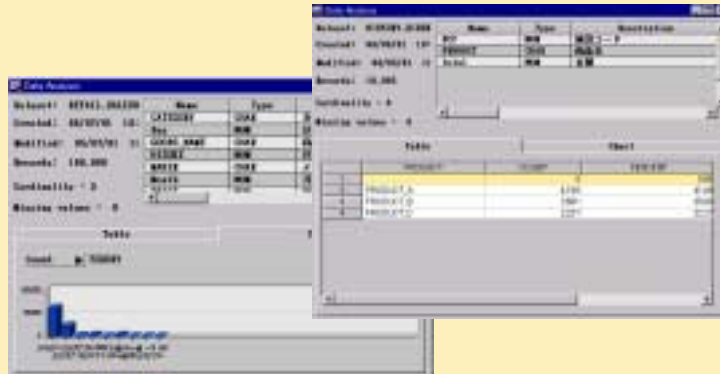
*The Power to Know.*

## SAS/Warehouse Administratorの優位性

### 多様なアドインツールでユーザーのニーズにこたえる

#### データ分析アドイン サンプル画面

テーブルのカーディナリティや欠損値、テーブル名、生成日、変更日、レコード数などの情報をテーブルとグラフで提供。

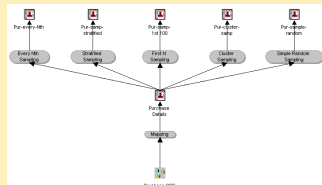


The Power to Know

## SAS/Warehouse Administratorの優位性

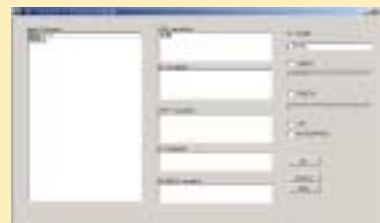
### 多様なアドインツールでユーザーのニーズにこたえる

#### Enterprise Minerアドイン サンプル画面



Enterprise Minerアドイン(以下EMアドイン)はSAS/Warehouse Administratorデータストアから多様なタイプのサンプルを生成、ブラウズ、分析機能を提供。モデルの役割や測定水準など特定のメタデータ定義を行うこともできます。ウェアハウスの管理者から見た、EMアドインを利用することのメリットは、ウェアハウスのデータストアから分析用のサンプルデータや最大値・最小値・欠損値の割合などの統計値メタデータの生成が可能になるということです。これらのサンプリングや統計値の算出は、EM上で利用するだけではなく、データストアのコンテンツやデータの質の分析にも利用できます。分析担当者としてのメリットは、EM上では分析対象データが同じであってもプロジェクトの数だけ入力データの登録が必要だが、EMアドインを利用すると、一度で行うことが可能になります。

#### テーブルの縦横変換アドイン サンプル画面



行列の縦横変換のニーズはかなり存在します。SASシステムでは、実際には10行ほどのコーディングをユーザーに行なっていたのですが、アドインをインストールさせていただいて、ユーザーにできるだけプログラムレスな環境を提供することも可能で

The Power to Know