

Rei Takahashi
Solution Planning Center
Sale Div.
Jpnrat@jpn.sas.com
The Power to Know

Copyright © 2000, SAS Institute Inc. All rights reserved.



Issue & Solution

Issue

- 散在するデータソース、複数のテーブルから目的のテーブルを作成する
 - ✓ データソースはどこだったか？
 - ✓ どのテーブルを利用したのか？
 - ✓ いつ更新をかけたのか？
- WarehousingはいつもScrap and Build。培ったナレッジが生きていない
- ルーティンワークやメタデータのドキュメント作成が手作業

Solution

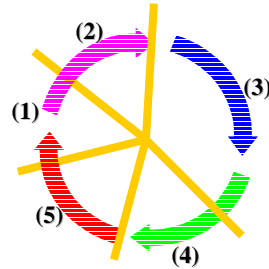


SAS/Warehouse Administratorによってメタデータ・レポジトリ・マネージメントを一元的に行ない、上記の問題点を全て解決

The Power to Know

DWH構築プロセス&サイクル

- (1) 分析する内容の選定
- (2) どのように分析をするのか吟味
- (3) データ抽出→データ変換→データ加工、再編成
- (4) データを公開して分析
- (5) 公開データに対してのフィードバックを得て(1)へ戻り繰り返す



(1)～(5)にかかる時間を短縮すれば早くデータウェアハウスを構築でき、また**工程(3)**が最も時間がかかると言われている。



(3)を**短縮**することのできる**ツール**が存在すれば**可能**となる

The Power to Know.

DWH構築プロセス&サイクル

SAS/Warehouse Administrator



DWH構築及び運用における
全行程をカバーする強力なツール

The Power to Know.

Why SAS Solutions ?

SAS \$10億ドルのソフトウェアSaleの 60% がDWH関連

Datamation 'The Product of the Year' Award 2年連続受賞

DATAMATION
PRODUCT
OF THE YEAR 1997
Datamation
Product
of the Year
1997
IS MANAGERS' CHOICE

1. **SAS**
2. Cognos
3. RedBrickSystems
4. Information Builders

DM Review DataWarehouse100 3年連続受賞

Data 100 Warehouse

1. IBM
2. **SAS**
3. Oracle
4. Platinum
5. NCR

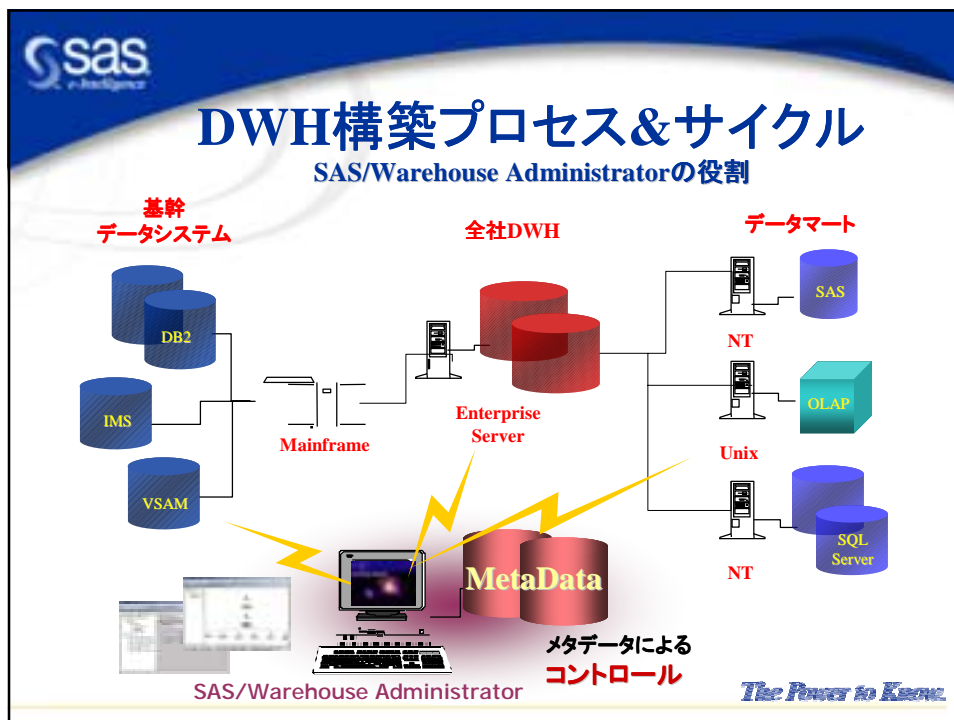
AWARD 100
WINNER
1999

1. **SAS**
2. IBM
3. Microsoft
4. Oracle
5. Cognos

DM REVIEW 100
2000
AWARD
WINNER

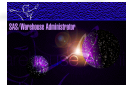
1. **SAS**
2. NCR
3. Oracle
4. ComputerAssociates
5. Cognos

The Power to Know.



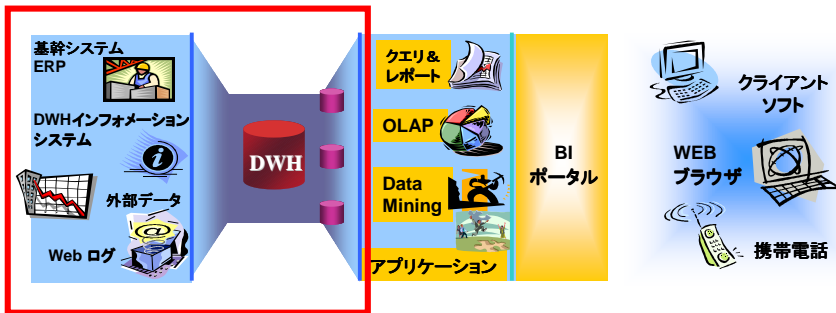
DWH構築プロセス&サイクル

SAS/Warehouse Administratorの役割



※DWH構築だけではなく、
データマート作成もOK

SAS/Warehouse Administrator



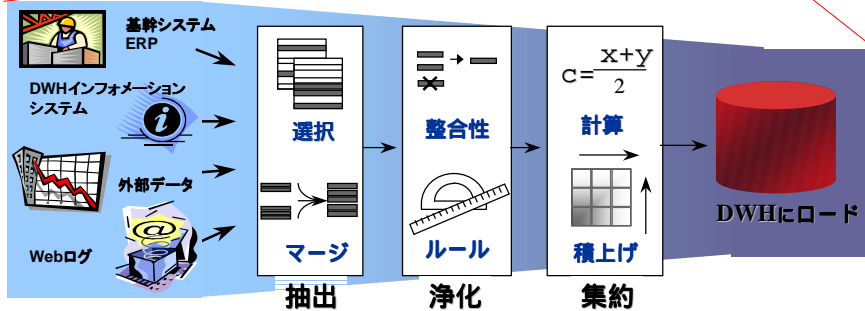
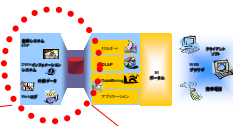
The Power to Know.

DWH構築の3つのコンポーネント

(1) 管理層

基幹システムや外部データソースからDWHへデータを運ぶ

- データソースへのアクセス
- データの選択・マージ
- DWHに適した形式に変換してロード



The Power to Know.

DWH構築の3つのコンポーネント

(2) 編成層

完全なDWHメタデータ管理

メタデータによってデータを整理して蓄積

システム側面でのベネフィット

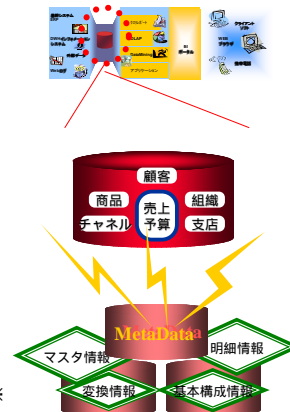
- 最適かつ効果的なデータ構造の判断材料に
- 中央集約型にするか分散型にするかなど戦略的な決定
- 将来の必要データの変化・多様化に的確に対応できる

業務的/ビジネスユーザーへのベネフィット

- 欲しいデータの取得がナビゲーションによって迅速で効率的に
 - 取得するデータの確実性を実感→データから生まれるアイデアに自信！
- DWH利用が促進され、組織内で一貫/共有化したデータ供給源となる

経営側面でのベネフィット

- 正しいデータの正しいタイミングでの取得による意思決定の精度向上
 - データ取得作業が容易になるため、システム要員/教育の軽減
 - データを利用する社員の作業効率の向上
- DWHへの投資効果の引き上げ！企業利益を1%引き上げるとい調査結果も！※



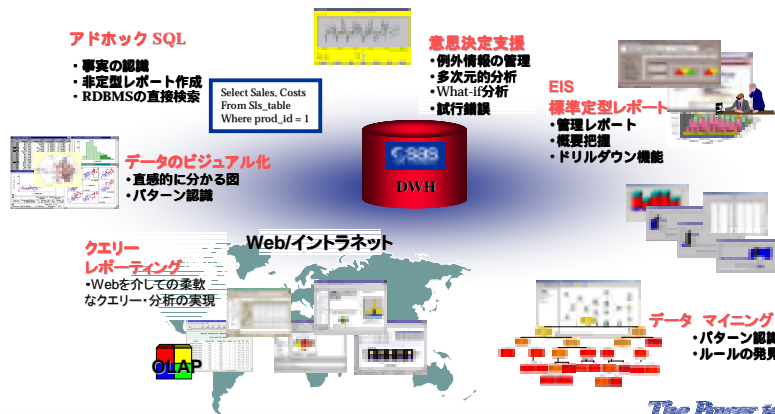
※source:
Marco, David(2000) Meta Data RIO 3-part series. DM Review, September, October, November

The Power to Know.

DWH構築の3つのコンポーネント

(3) 利用層

ユーザーや業務に適したアクセス手段を提供



The Power to Know.



This slide features the SAS logo and 'e-Intelligence' tagline in the top left. The background is a blue gradient with a faint image of a person's face. The title '完全なメタデータ管理' (Complete Metadata Management) is centered. The slide is divided into three main sections: 'メタデータとは?' (What is Metadata?), 'メタデータの役割' (Role of Metadata), and a diagram illustrating metadata management.

メタデータとは?

- データに関わるデータ→
データの構成、内容、キー、索引等

メタデータは膨大なデータウェアハウスの中から、
目的のデータの発見を容易にしてくれる

メタデータの役割

- データウェアハウス上にあるデータソースは**どこから来たのか?**
- 業務系システム→データウェアハウスに至るまでの**変形**は?
- データウェアハウスの**内容の記述**(権限、フィールド情報等)
- タイムスタンプやローディングスケジュールの為の**バージョン付け**

The diagram illustrates the flow of metadata. At the top, a red cylinder labeled '顧客' (Customer) contains '商品' (Product), '売上' (Sales), '組織' (Organization), 'チャネル' (Channel), '予算' (Budget), and '支店' (Branch). Below it, a red cylinder labeled 'MetaData' is connected by yellow lightning bolts to four green cylinders: 'マスタ情報' (Master Information), '明細情報' (Detail Information), '変換情報' (Transformation Information), and '基本構成情報' (Basic Configuration Information).

The Power to Know.

SAS/Warehouse Administratorでのメタデータ管理

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にします。(メタデータ
HTML書き出しアドイン利用)

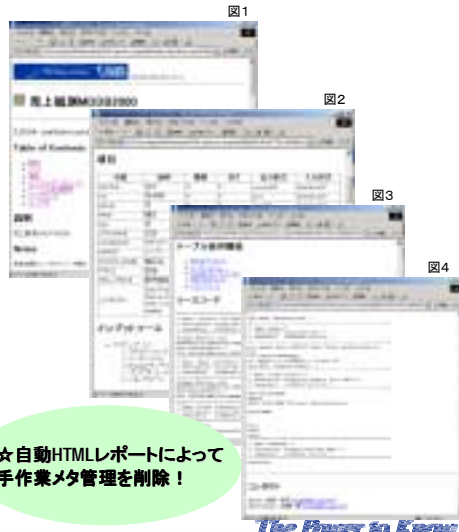
このレポートでは、メタデータ参照の対象のオブジェクトに関する
①説明②ノート③項目定義(名前、説明、種類、長さ、出力形
式、入力形式)④インプットデータソース⑤テーブル依存関係⑥
ソースコード⑦コンタクト(所有者/管理者の名前、連絡先)の情
報が得られます。(図1、2 参照)

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
されてなくとも、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/管理者のメールアドレスのリンクが表示
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を利用できます。(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でのメタデータ管理

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/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。

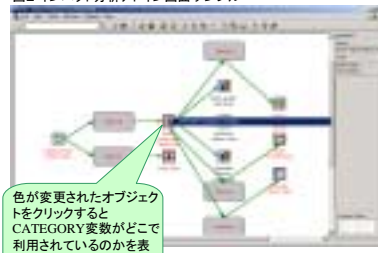
これらの機能は、SAS/Warehouse Administratorの完璧なメタデータ管理によって成り立つもので、DWHのメンテナンスにおいて
は、欠かせない機能になっています。

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.



各種DB対応の柔軟性

MEA—マルチ・エンジン・アーキテクチャー

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・アーキテクチャー)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	◆ACCESS99 ◆BMDP ◆DB2 ◆FAMECHLI ◆Infomix	◆ODBC ◆OLEDB ◆ORACLE ◆OSIRIS ◆MS SQL Server	◆Sybase ◆SPSS ◆TERADATA ◆XML ◆XPORT ※(※transport format file)	◆CA-OpenIngres


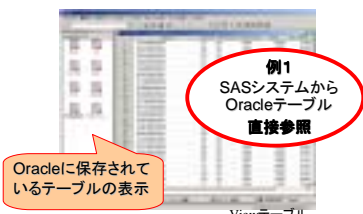



図1 DB登録ウィンドウ



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



データ/処理ホスト選択の柔軟性

MVA—マルチ・ベンダー・アーキテクチャー

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 処理ホスト登録ウィンドウ参照) このスケーラビリティはSASのMVA(マルチ・ベンダー・アーキテクチャー)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を行なう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ィンドウ参照)

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。

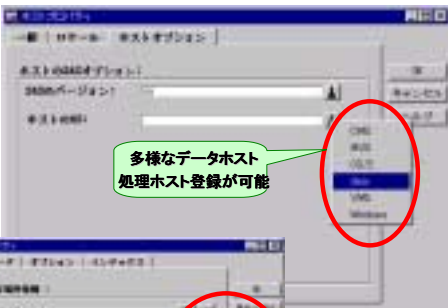


図1 処理ホスト登録ウィンドウ

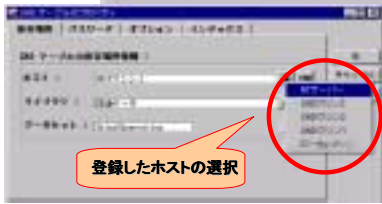


図2 データ保存場所ホスト選択ウィンドウ

The Power to Know.

スケジューリング機能

SAS/Warehouse Administratorによるジョブスケジューリングは、UnixホストではCron、NTホストではAT Commandを利用します。その他のスケジューラーに対応するためには、Nullスケジューラーを指定します。(ex. MVS スケジューラ CA7にも対応)

適切なコーディングを行なうと、処理の終了/エラー発生時の電子メール通知、マルチスレッドによる並列処理などの機能を利用することができます。UNIXホストでは、UNIXジョブスケジューラードインを利用してイベントドライブのスケジューリングを行なうことができます。

(Shell Scriptの利用)



スケジューラ選択ウィンドウ



スケジューリング ウィンドウ

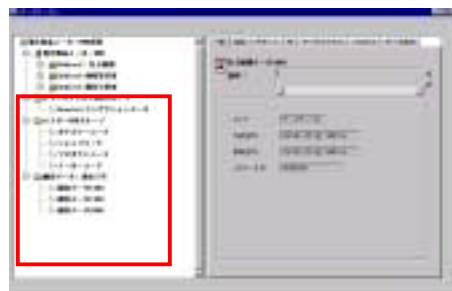
The Power to Know.

Extract


SASならではのデータソースバラエティ

SAS/Warehouse Administratorでは、登録できるデータソースの数に制限がありません。MVA, MEAを最大限に利用できるデータソース登録機能です。

Erwin, Powerdesigner, Designer2000などデータ・モデリングツールで作成したスキーマやデータモデルを読み込むことも可能です。



The Power to Know.

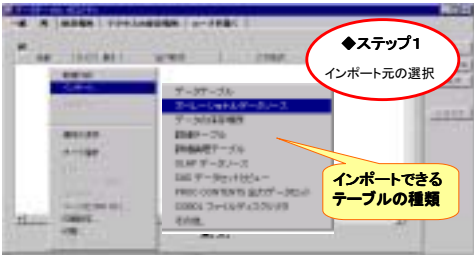


Extract

ターゲットテーブル(作成したいテーブル)の簡単定義づけ

SAS/Warehouse Administratorでは、ターゲットテーブルの項目定義が非常に簡単に行なえます。項目を一つ一つ定義していくことも可能ですが、他テーブルから項目名や定義の一括インポートすることも可能なので、項目定義にかかる手間と時間削減につながります。
(他テーブルから項目定義の一括インポート例参照)

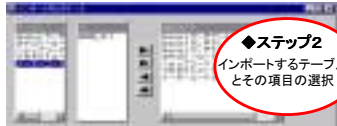
<例>他テーブルからの項目定義の一括インポート



◆ステップ1
インポート元の選択


インポートできる
テーブルの種類

ターゲットテーブルの項目定義ウィンドウ



◆ステップ2
インポートするテーブル
とその項目の選択

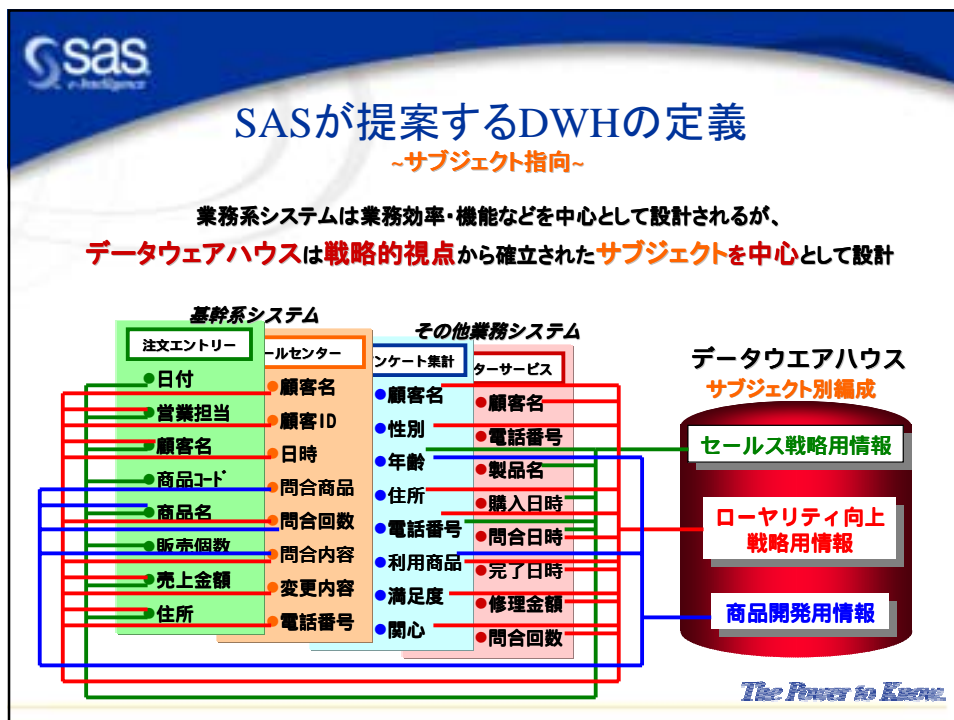
インポート項目選択ウィンドウ



◆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

The Power to Know.

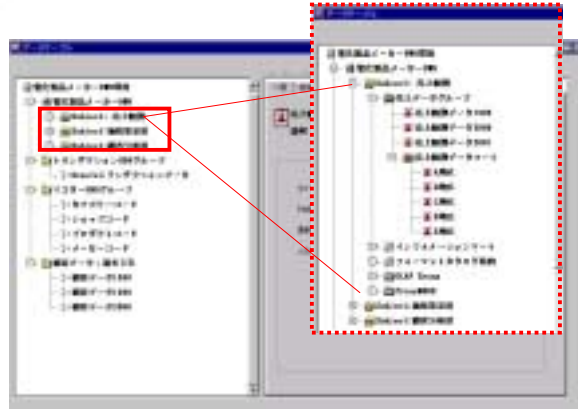


Transformation

サブジェクト指向

実際に利用される目的をもったウェアハウジングを行なうためには、やはりサブジェクトを設定してからETLをすすめていくべきです。

そこでSAS/Warehouse Administratorでは、メタデータの持ち方もサブジェクト指向です。それぞれのサブジェクト毎にデータやMDDBを作成していき、DWHを構築します。



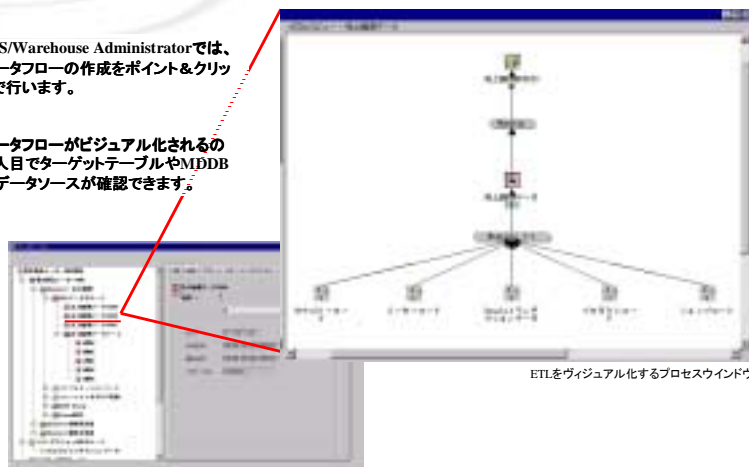
The Power to Know.

Transformation

データフローダイアグラム～データフローのビジュアル化

SAS/Warehouse Administratorでは、データフローの作成をポイント&クリックで行います。

データフローがビジュアル化されるので目でターゲットテーブルやMDDBのデータソースを確認できます。



ETLをビジュアル化するプロセスウィンドウ

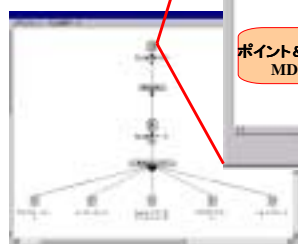
The Power to Know.

Transformation

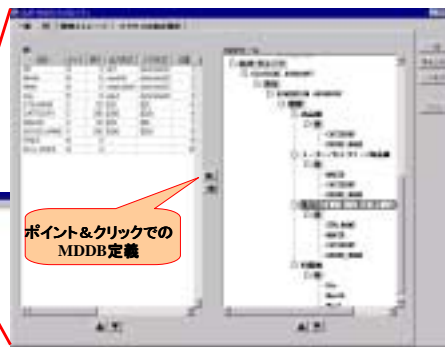
OLAP用MDDB作成/修正

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。



ETLをビジュアル化するプロセスウィンドウ



MDDB定義ウィンドウ

ポイント&クリックでの
MDDB定義

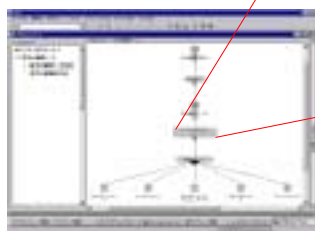
The Power to Know.

Transformation

ユーザー指定コード

SAS/Warehouse Administratorでは、GUIで行えない処理は、ユーザー指定コードをします。

この機能を利用するとSASシステムの柔軟なデータ加工機能を100%利用することができます。




ETLをビジュアル化するプロセスウィンドウ



SASコードウィンドウ

The Power to Know.



Load

SAS DWH/ RDB DWH

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)



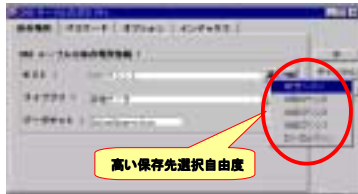




図1 保存形式指定ウィンドウ

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

The Power to Know.



SAS/Warehouse Administratorでのデータ利用

定型レポートのHTML生成

SAS/Warehouse Administratorでは、DWHのデータを利用してHTML定型レポート配信を可能にします。

レポート作成には、SAS言語コーディングが必要ですが、その分業的なレポート生成をお約束いたします。(図1 SASコードサンプル参照)

レポートは、HTMLをベースに、Java、ActiveX、Gifなどを利用してグラフィカルに作成することが可能です。作成したレポートを、Webサーバーに保存し、社内イントラにリンク付けする(図2、3 サンプルグラフ参照)ことや、HTMLレポートをEメールで配信する(図4 Eメール配信グラフサンプル参照)などの広範囲なレポート生成手段を提供します。





図1 SASコードサンプル

図2、3 サンプルグラフ

図4 Eメール配信グラフサンプル

The Power to Know.



SAS/Warehouse Administratorの優位性

ユーザーのニーズにあったアドインをフリーで提供

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
- インバクト分析アドイン
- 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
- リバースインバクト分析
- 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

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！

2001年5月現在

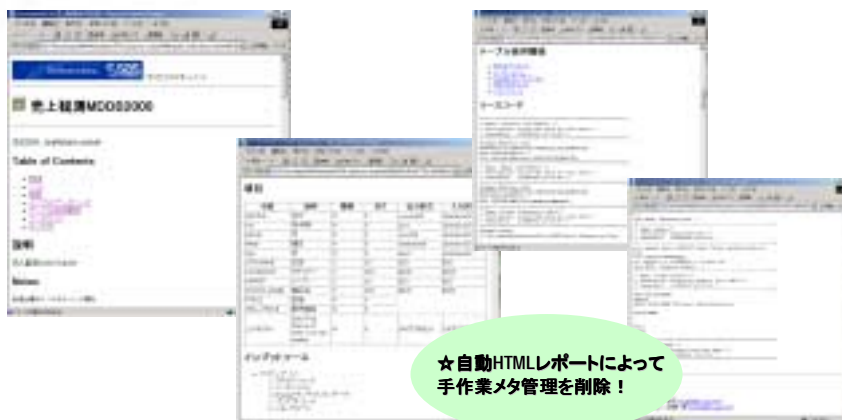
The Power to Know.



SAS/Warehouse Administratorの優位性

ユーザーのニーズにあったアドインをフリーで提供

メタデータHTML書き出しアドイン サンプル画面



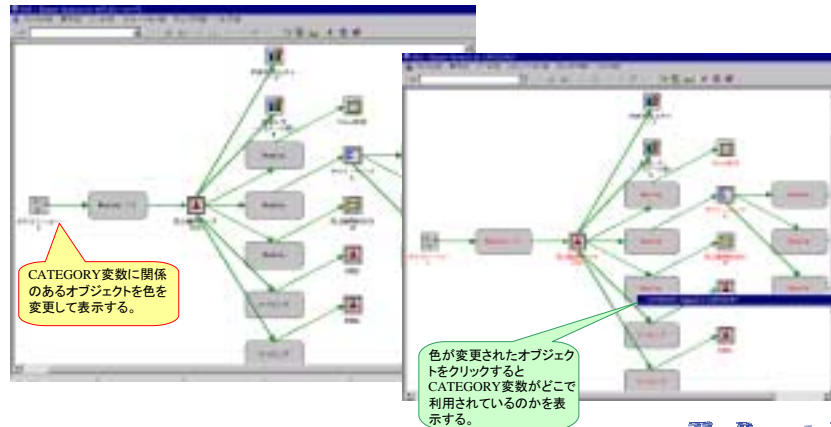
★自動HTMLレポートによって
手作業メタ管理を削除！

The Power to Know.

SAS/Warehouse Administratorの優位性

ユーザーのニーズにあったアドインをフリーで提供

インパクト分析 サンプル画面



The Power to Know.

SAS/Warehouse Administratorの優位性

ユーザーのニーズにあったアドインをフリーで提供

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。

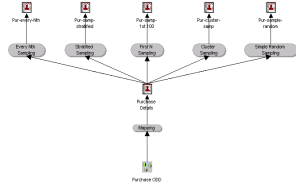


The Power to Know.

SAS/Warehouse Administratorの優位性

ユーザーのニーズにあったアドインをフリーで提供

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/Warehouses Administratorデータストアから多様なタイプのサンプルを生成、ブラウザ、分析機能を提供。項目の役割(model role)や測定レベル(measurement level)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す。

The Power to Know.