



新データウェアハウジング環境 SAS ETL Studioの紹介

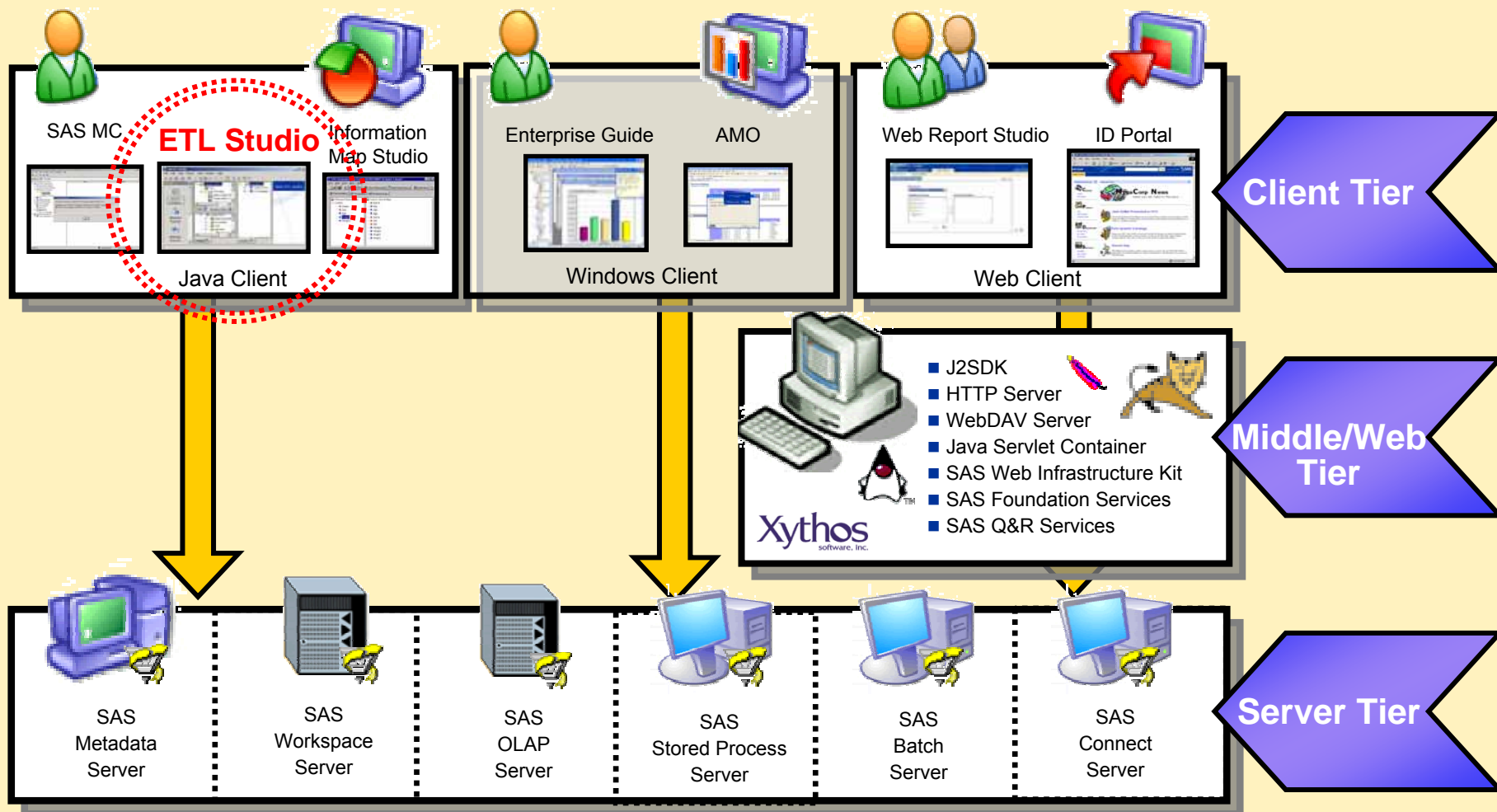
SAS Institute Japan株式会社 プロダクトマネジメント部

前田 幸一

SAS Institute Japan Ltd. Product Management Dept.

Koichi Maeda

SAS® 9 システムアーキテクチャ

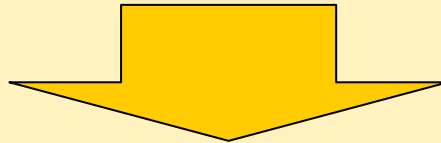


SAS ETL Studioの位置付け

SAS ETLツールの変遷

➡ SAS 8.2 :

SAS Warehouse Administrator 2.2



➡ SAS 9.1 :

SAS Warehouse Administrator 2.3

SAS ETL Studio 9.1

WA2.2のメンテナンス版

WAの後継ソフトウェア

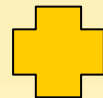
メタデータ・リポジトリの管理、SASコードの実行、ETLのデザインを閉じられた環境で一括で管理していたWAと比べると、使いやすさ、機能ともに大幅な向上が図られており、かつ、統合化された管理機能、分散化された処理機能を提供するSAS 9アーキテクチャに対応する、BIプラットフォームの中核的な役割を担うデータウェアハウス構築支援ツールです。

SAS ETL Studioの位置付け

Intelligence Value ChainにおけるETL^Q



- **E**xtract データの抽出
- **T**ransform データの加工 (変換)
- **L**oad DWHへのロード



- **Q**uality データの品質の向上

データウェアハウスを構築する上で、元のデータソースからデータを抽出 (Extract) し、ユーザーの利用目的に合わせた形に変換・統合 (Transform) し、ターゲットとなるデータウェアハウスやデータマートにロード (Load) する

SAS Data Quality Serverによって、上記 Transformのプロセスでいわゆる「名寄せ」処理を行う機能。ただし、現時点では日本語版は未提供 (将来のリリースで提供予定)

SAS ETL Studioは、企業に蓄積されたデータから競争優位を導くインテリジェンスを創出するための体系化されたプロセスである「Intelligence Value Chain」の「ETLQ」ステージで、ETLプロセスの構築・管理機能を提供する、フロントエンドツールとして位置付けられ、ETL - DWH/DM - BI - AIまでシームレスに連携します。

SAS ETL Studioとは

ポイント

➡ Javaベースのビジュアルなデザイン機能

ドラッグ&ドロップとウィザードなど、GUIによるETLプロセスの構築・管理

➡ 豊富な変換機能

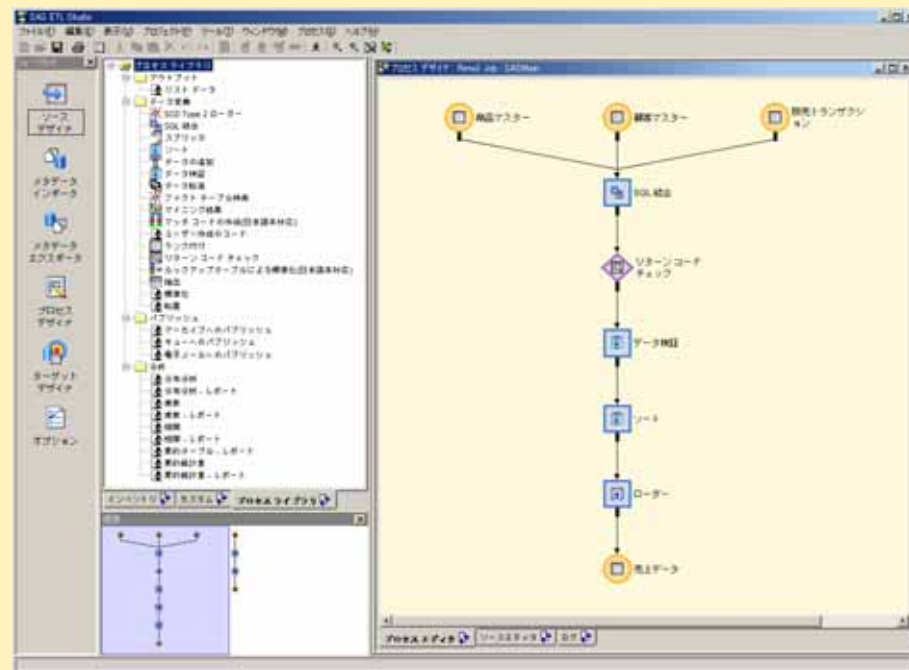
様々なテンプレートを提供することで、プログラムコーディングの手間を省力化し、ETLプロセスの構築とメンテナンスを効率化

➡ データアクセス機能とストレージ機能

あらゆるデータソースに素早く容易にアクセス、DWH/DMへのロードもスピーディーに

➡ 複数ユーザーによる開発環境

チェックイン/チェックアウト機能によるチームでのETLプロセスの構築とメンテナンス環境



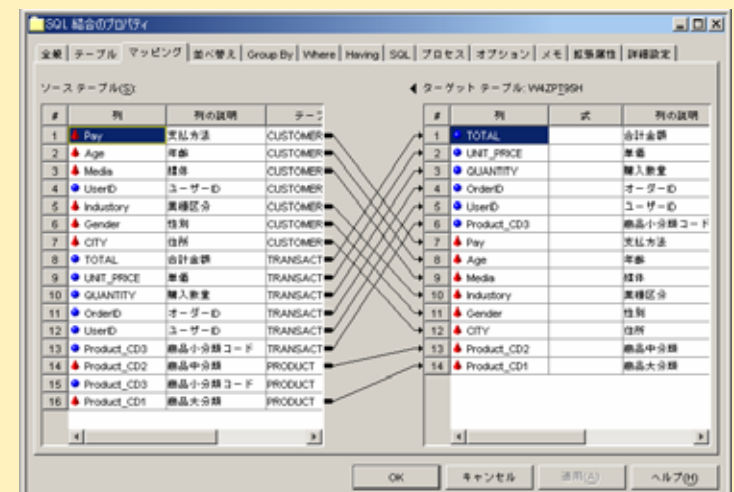
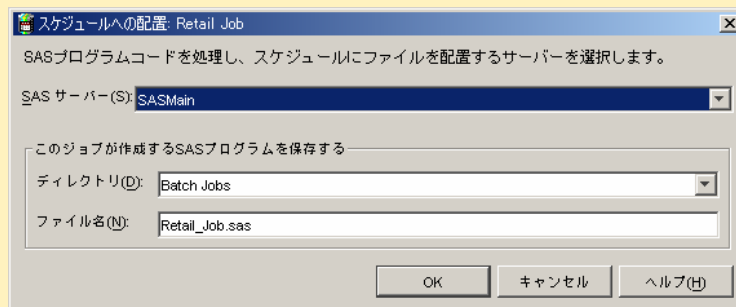
➡ データ検証機能

欠損値の自動変換や不完全レコードのフィルタリングなど、ETLに欠かせないデータクリーニング機能

Javaベースのビジュアルなデザイン機能

プロセスデザイナー (2)

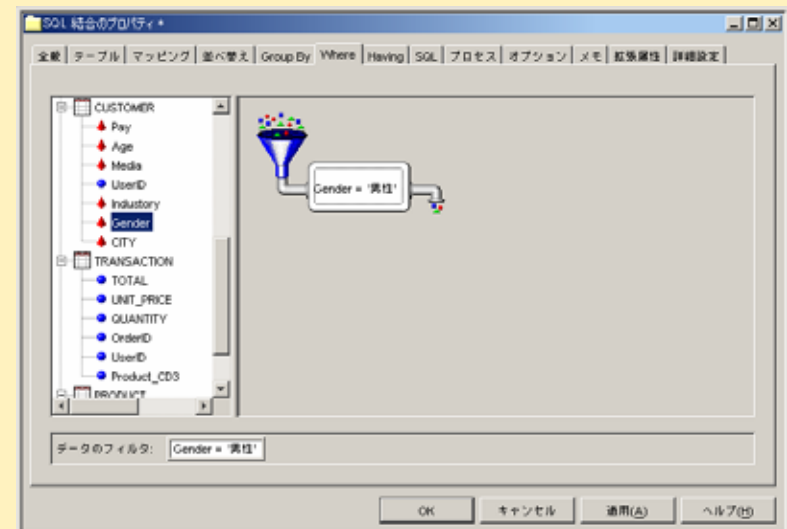
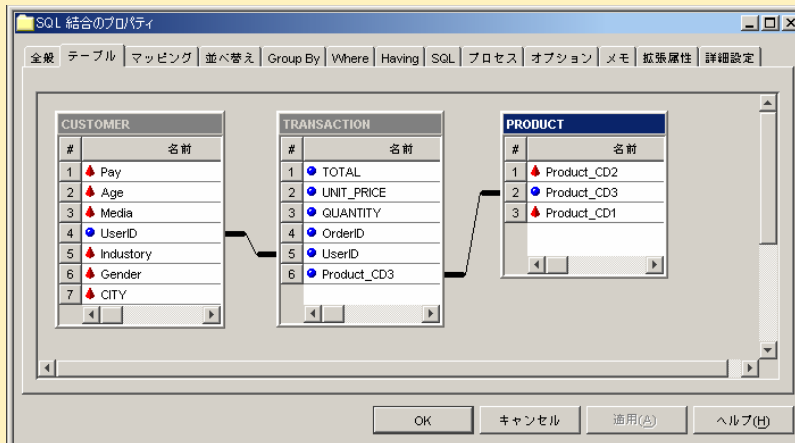
- ➡ 変換およびロードのステップにおいて自動的に1対1のマッピングを行い、変数マップの手間を省力化
- ➡ 開発時に対話的なテスト実行の機能を利用して、デバッグやジョブのテストを実行可能
- ➡ 統合されたスケジューラに、簡単にジョブを配置することが可能



Javaベースのビジュアルなデザイン機能

プロセスデザイナー (3)

- ➡ 複数ユーザー間でジョブを共有することにより、再作業を省き、サイクルタイムを削減
- ➡ テーブルの結合や抽出条件の設定もGUI操作で

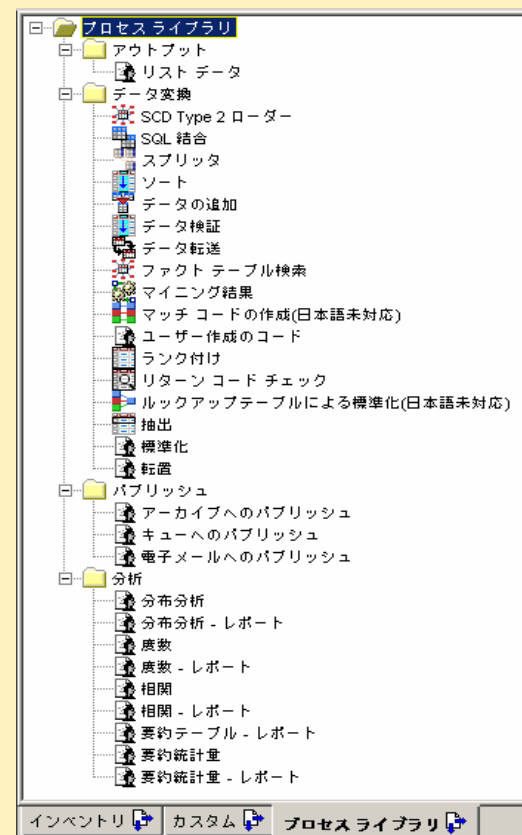


豊富な変換機能

テンプレート

➡ 「プロセスライブラリ」として豊富な変換テンプレートを提供、プログラミングの手間を省力化

- ☑ SQL結合: データの結合デザイン、自動結合
- ☑ ソート: SORTプロシジャのサポート
- ☑ スプリット: 複数データに分割するDATAステップの生成 (IF-THEN、ユーザー作成の選択基準)
- ☑ ランク付け: RANKプロシジャによるランク付け
- ☑ データ転送: サーバー間のデータ転送
- ☑ 標準化: STANDARDプロシジャのサポート
- ☑ 転置: データの縦横変換
- ☑ ユーザー作成のコード: 独自のSASコード挿入
- ☑ 相関: 相関分析による統計値算出機能
- ☑ 要約統計量: 要約統計量の算出機能 etc.



「プロセスライブラリ」ではSQLの制限を越えて、SASプロシジャ・言語がサポートされます。
Javaによる独自の変換テンプレートを作成することも可能です。

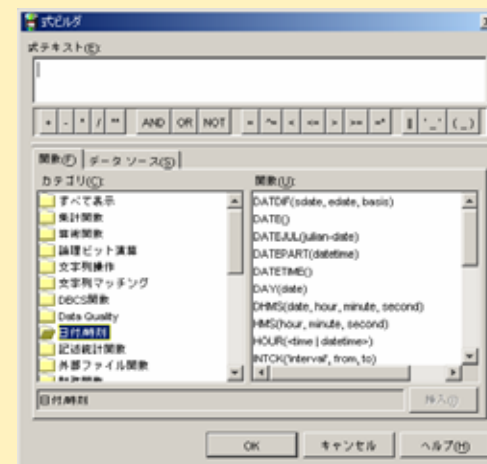
豊富な変換機能

SAS関数: 値の変換用関数

- ➡ SASが提供する豊富な関数群をGUIで利用可能
 - ☑ 文字列操作、算術関数、集計関数、記述統計関数、外部ファイル関数、日付/時刻 etc.

ウィザード

- ➡ SASプログラミングによるユーザー定義の変換テンプレートの操作
 - ☑ 変換ジェネレータウィザード: ユーザー定義の変換テンプレートの作成
 - ☑ 変換インポート/エクスポートウィザード: ユーザー定義の変換テンプレートの配布



変換テンプレートその他、ETLプロセス内で他のユーザが定義したSASコードを呼び出すことも可能です。

データアクセス機能

ウィザード

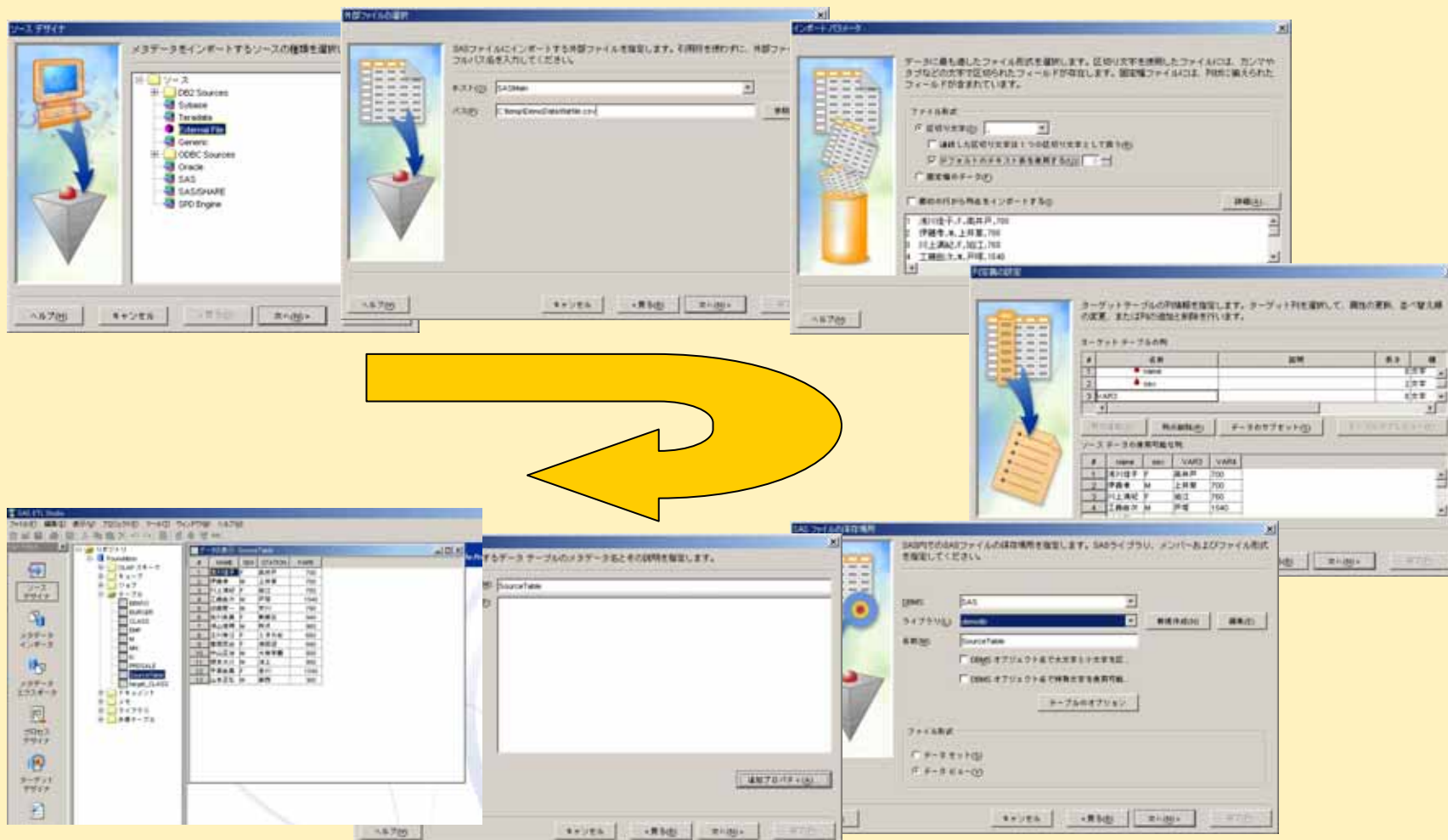
- ➡ ウィザードにより、SQLやSASプログラミングの知識、経験を問わず、ETLプロセスを構築可能
 - ☑ ソースデザイナウィザード: テキストファイルやRDBMSデータの読み込み
 - ☑ 新規ライブラリウィザード: ライブラリの割り当て定義
 - ☑ ターゲットデザイナウィザード: ロード先データの定義
 - ☑ キューブデザイナウィザード: OLAPキューブの定義

アクセス可能なデータソース

- ➡ DBMS, MS Excel, MS Access
 - ☑ Oracle, DB2, MS SQL Server, Teradata, MySQL, PC-Files, ODBC, OLE DB etc.
- ➡ テキストファイル
 - ☑ CSV, TAB区切り、ブランク区切り
- ➡ ERPシステム (SAS Data Surveyor)
 - ☑ SAP BW, SPA R3, Siebel, Oracle Applications, Peoplesoft

データアクセス機能

ウィザード(テキストファイルの例)

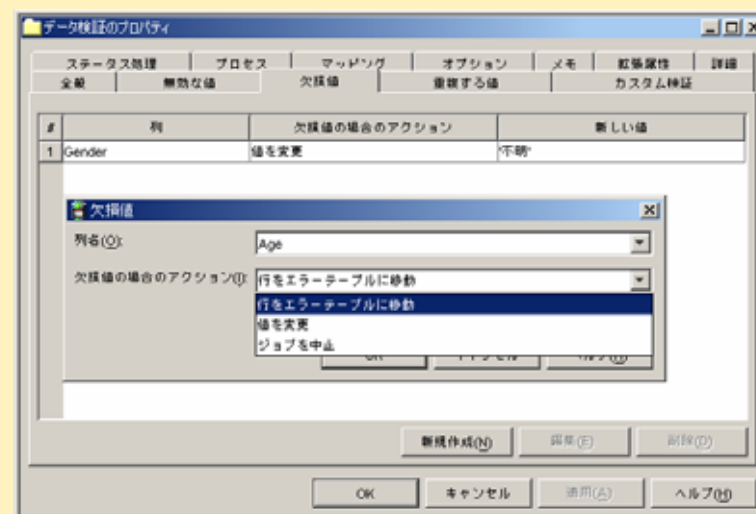


データ検証機能

欠損値の自動充填

➡ 3パターンの欠損値ハンドリングを提供

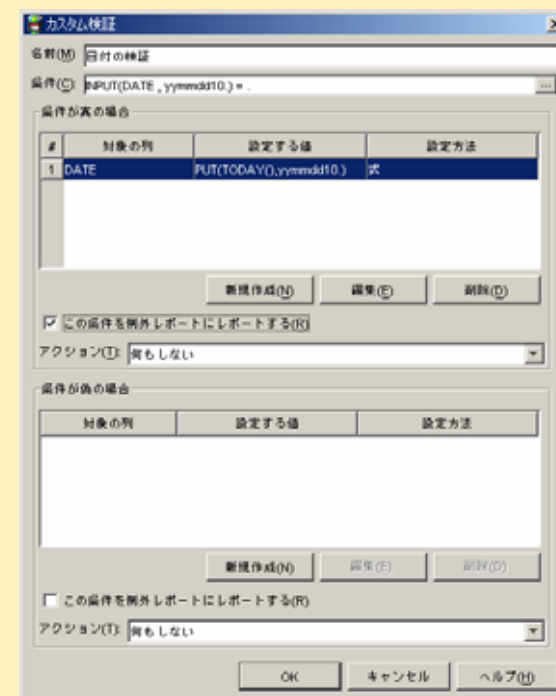
- ☒ 該当列の欠損値を、設定した値に自動変換
 - ➡ 定数を設定
 - ➡ 他の変数の値や、SAS関数を使用して得た/変換した値を設定
- ☒ 該当行をエラーテーブルに移動
 - ➡ ジョブの実行日時を付加して、欠損値を含む行をエラーテーブル
('データ検証ノード'の「プロパティ」>「オプション」タブで指定)に移動
- ☒ ジョブを中止



データ検証機能

不完全レコードのフィルタリング

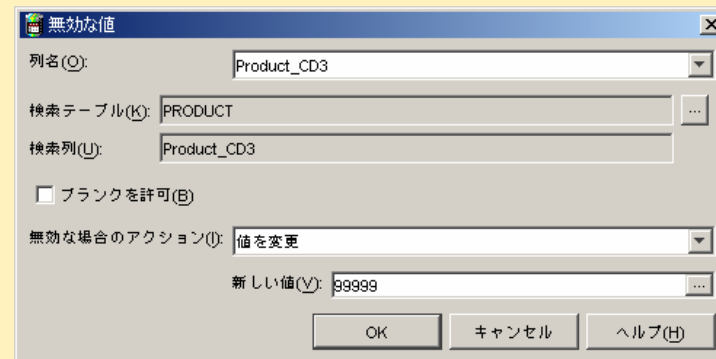
- ➡ 設定した条件の真/偽によるデータハンドリングを提供 (真/偽別にハンドリング可能)
 - ☑ 該当列の値を、設定した値に自動変換
 - ➡ 定数を設定
 - ➡ 他の変数の値や、SAS関数を使用して得た/変換した値を設定
 - ➡ 変換テーブル(参照テーブル)から変換値を取得して設定
 - ☑ 例外レポートへの出力
 - ☑ 該当行をエラーテーブルに移動
 - ➡ ジョブの実行日時を付加して、欠損値を含む行をエラーテーブル(「データ検証ノード」の「プロパティ」>「オプション」タブで指定)に移動
 - ☑ ジョブを中止



データ検証機能

検索レコード・列

- ➡ マスターデータを参照して値の突き合わせを行い、その結果による(存在しない場合の)データハンドリングを提供
 - ☒ 該当列の値を、設定した値に自動変換
 - ➡ 定数を設定
 - ➡ 他の変数の値や、SAS関数を使用して得た/変換した値を設定
 - ☒ 該当行をエラーテーブルに移動
 - ➡ ジョブの実行日時を付加して、欠損値を含む行をエラーテーブル(「データ検証ノード」の「プロパティ」>「オプション」タブで指定)に移動
 - ☒ ジョブを中止
 - ☒ ブランクの場合の許可設定可能



データ検証機能

ステータスハンドリング

➡ プロセスやジョブのリターンステータス(成功/警告/エラー)に応じて、それぞれの場合のアクションを設定可能

☒ 中止

☒ カスタム

➡ SASマクロコードで記述されたアクションを実行

☒ Email送信

➡ メッセージを電子メールで送信

☒ テキストファイルにエントリを送信

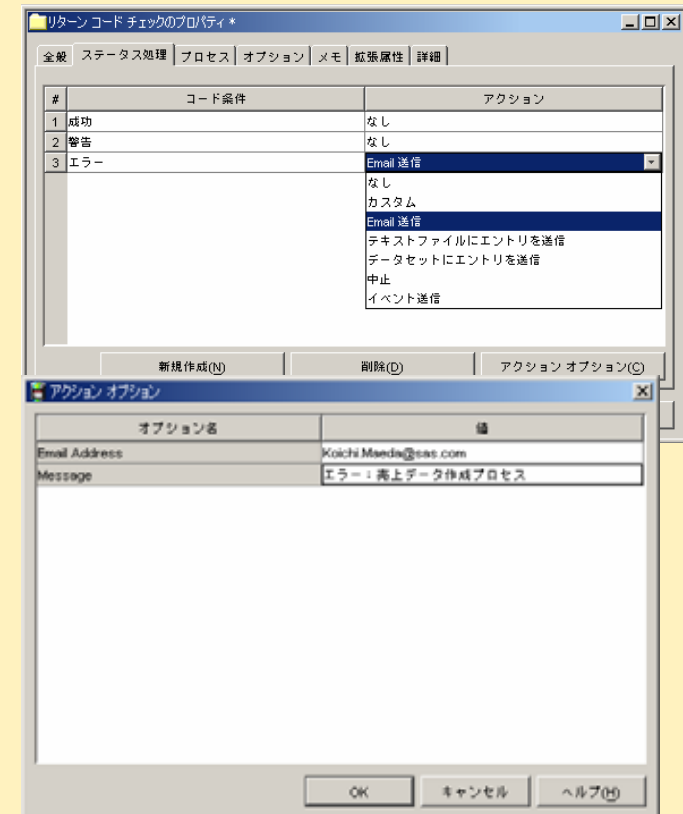
➡ メッセージをテキストファイルに保存

☒ データセットにエントリを送信

➡ メッセージをSASデータセットに保存

☒ イベント送信

➡ メッセージをイベントブローカーに送信



メタデータ管理機能

一元管理されたシステムメタデータ

- ➡ データライブラリ情報
- ➡ データ項目情報
- ➡ ユーザーアクセス権限 etc.

プロセスメタデータの主な機能

- ➡ ETLプロセスの開発環境情報
- ➡ 複数ユーザーからのアクセス制御: 変更管理
- ➡ インパクト分析 etc.

オープンメタデータ仕様

- ➡ Common Warehouse Metamodel (CWM) 規格に準拠し、同規格に準拠した他アプリケーションとのメタデータのインポート/エクスポートが可能
- ➡ Meta Integration Technology社のMeta Integration Model Bridge (MIMB) によるメタデータのインポート/エクスポートも可能



The Power to Know®

本論、またはSAS ETL Studioについてのご質問、お問い合わせは
プロダクトマネジメント部 前田幸一 <Koichi.Maeda@sas.com>
またはご担当のSAS営業までお願い申し上げます。