

統合的データモデリング支援環境“Darwin”(1)

4E-4

ージェネラルコンセプトー

加藤 雅樹 郡司 賢 武田 和彦 丹羽 隆



NTTデータ通信株式会社 技術開発本部

1 はじめに

DOA(Data Oriented Approach:データ中心アプローチ)によるシステム開発が叫ばれ始めて久しいが、DOAの中心的な作業といわれるデータモデリングを行う上で、作業にとって真に有効とされるツール環境の整備がまだ不完全な状態であるという現実は否めない。

本稿では、データモデリングの作業を多角的に分析し、データモデリングに携わる要員のニーズに合ったツール環境のコンセプト及びツール個々の機能の概要について説明する。

2 コンセプト

本ツール環境のコンセプトを決定するにあたり、まず、当社において、本格的にDOAによるシステム開発に取り組んだ某金融システムでの事例を踏まえて、開発の指針とした。ここでの事例を分析する上で、以下の2つの側面に着目した。

2.1 作業的側面

企業情報の分析からデータベーススキーマの実装までの一連のデータモデリング作業の内容を分析すると、大きく論理データモデリングと物理データベース設計の2つのフェーズに分けられる。

●論理データモデリング

論理データモデリングのフェーズは、分析対象領域が有するデータ構造を明らかにし、特定のハードウェア・ソフトウェアに依存しない論理的なデータスキーマを導出するためのステップである。論理データモデリングのねらいは、企業内の情報システムの根拠となるところの企業活動を抽象化して把握しようとするもので、情報システム全体のフレーム・ワークを論じたり、長期的なシステム構想やその長期戦略を定義する拠り所となるべき論理スキーマを導出するものである。また、正規化及び標準化等の作業を繰り返し行い、最終的に論理的に矛盾のない、冗長性のないスキーマを導き出すのが目的である。

●物理データベース設計

論理データモデリングのフェーズまでで、データベースの論理的な構成は決定されるわけであるが、実際のコンピュータ・システムにおける効率の良いデータアクセスを達成するためには、内部的にデータをどのように格納するかを詳細に検討する必要がある。これを決定するのが物理データベース設計のフェーズの作業である。

DBMSにおいては、利用パターンに応じた効率の良いデータアクセスを達成するため、ファイル編成の選択やインデックスなどの設定が可能となっているのが普通で

ある。従って、物理データベース設計のフェーズにおいては、DBMSによって提供されるオプションと、システムの中での各トランザクションのデータベースアクセスパターンを考慮して、最適なデータベース構成(DDL構文)を導き出すのが目的となる。

以上の作業的側面を分析した結果、本ツール環境のコンセプトとして、データモデリングの作業におけるこれら2つのフェーズを総合的に支援することを目的とし、企業情報の分析という最上流のフェーズから、ターゲットDBMSのDDL生成といった下流フェーズまでを一貫したツール環境下でサポートすることとした。

2.2 人的側面

また、DOAによるデータモデリングを実現するためのリソース、特に、DOAを実践する人的資源に着目した。前述の事例を分析した結果、DOAを実践する要員として以下に示す3者が存在することを確認した。

●データ分析者(Data Analyst)

論理データモデリングを担当するスタッフである。

通常、1つのデータモデルをサブジェクトエリアという単位で分割し、サブジェクトエリア単位には数人のデータ分析者の共同作業によって行われている。

●データベース設計者(Database Designer)

物理データベース設計を担当するスタッフである。

通常、ターゲットDBMS単位に数人のデータベース設計者によって構成されている。

●データ管理者(Data Administrator)

データ分析者、データベース設計者が行ったモデル情報を一元的に管理・検査を行うと共に、用語・ドメイン辞書等の保守を行うスタッフである。

通常、1プロジェクトに1人ないし2人程度でプロジェクト全体のデータ資源の管理・統制を行う。

以上の人的側面の分析結果から、本ツール環境のコンセプトとして、データ分析者、データベース設計者、データ管理者の3者の視点からの統合的なデータモデリングの支援を実現することを目的とし、複数要員での分散協調作業ができるようにLAN等を用いたグループウェアにも対応することとした。

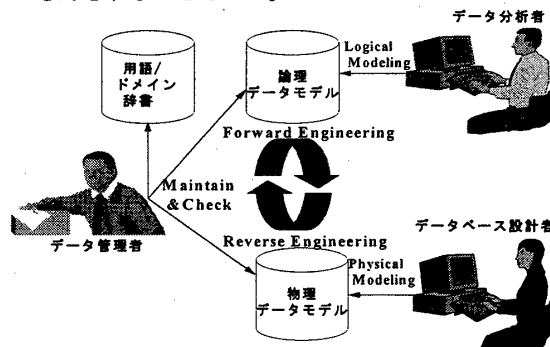


図1 データモデリングの作業形態

Integrated Data-Modeling Environment “Darwin”(1)

～General Concept

Masaki Kato, Masaru Gunji, Kazuhiko Takeda, Takashi Niwa
Research and Development Headquarters,
NTT DATA COMMUNICATIONS SYSTEMS Corp.

3 アーキテクチャ

以上のコンセプトを受け、本ツール環境は、データ分析者、データベース設計者、データ管理者の3者の作業を支援するために3つのコンポーネントと1つのリポジトリから構成される。

それぞれのツールはクライアント上で動作し、LAN環境を用いてサーバ上のリポジトリを共通的に利用する。

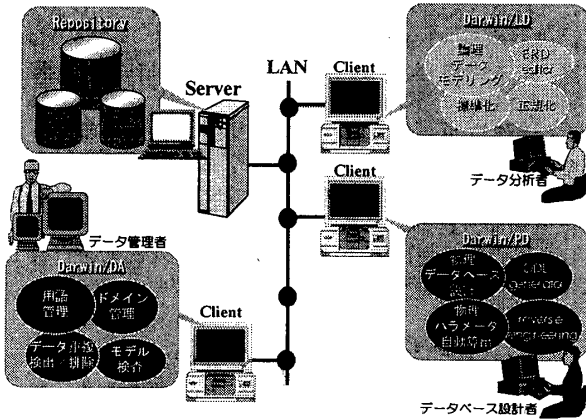


図2 ツール環境アーキテクチャ

3.1 論理データモデリング支援ツール“Darwin/LD”

“Darwin/LD”は、データ分析者が行うべき作業を支援する。

本ツールでは、こうした論理データモデリングを行うためのノーテーションとして、一般によく用いられているERダイアグラムを採用した。

GUIの特性を生かしたERダイアグラムドローイング機能により可視的にモデリングの作業が行えると共に、一覧表形式のインターフェイスも実装し設計情報の一括入力も可能とした。

また、Durellの提唱した名称標準化規則を拡張し^{[1][2]}、アトリビュート、エンティティ、ドメインといったオブジェクトに対して、用語辞書を用いて自動的に名称標準化を行い、データ重複の徹底的な排除を行う機能を有する。

更に、ERモデルの最適化を行うために、第1～第3正規化の支援、またリレーションシップの最適設計のための支援、論理データモデルの妥当性チェック機能等も有する。

3.2 物理データベース設計支援ツール“Darwin/PD”

“Darwin/PD”は、データベース設計者の行うべき作業を支援する。

本ツールでは、論理データモデルをフォワードし、ターゲットDBMSに特化した設計を行う。データベースのスキーマを表現するためのノーテーションとして、GUIを利用したDBダイアグラムというものを独自に考案した。また、Table、Index等の単純な構文のみならず最近のRDBMSに装備されているView、Trigger、Function、Package等複雑な構文にも対応した全DDL文を自動生成する。また、各種オブジェクトの格納容量等の物理的なパラメータの推奨値も自動計算できる。更に、既存のデータベース資源の再利用という観点からDDL文を構文解析しリバースエンジニアリングする機能も有する。

3.3 データ管理支援ツール“Darwin/DA”

“Darwin/DA”では、データ管理者の行うべき作業を支援する。

プロジェクトで共通で利用される用語、ドメインの定義及び保守を行うと共に、モデル全体の検査が実施できるようになっている。その他にも、モデル、サブジェクトエリア、ユーザ等の定義も行う。

3.4 リポジトリ

本ツール環境では、複数設計者による並行作業も可能にする為、LAN環境を用いたクライアント/サーバ形式のツール連動を行っている。

サーバ上にリポジトリを実装し、複数クライアントから同時に1つのデータモデルの構築が可能であり、設計情報の一元管理を行っている。論理データモデルから物理データモデルへのフォワードエンジニアリングは、このリポジトリを介して行われる。

クライアント単位には、サブジェクトエリアという概念で1つのデータモデルを細分化したビューで見せ、複数要員による協調作業もサポートしている。

リポジトリ上には、設計情報のメタモデルの他にも、名称標準化を行う際に必要な用語やドメイン等のプロジェクトで共通に利用されるオブジェクトを辞書として格納している。

以下に、本ツール環境のリポジトリ上に実装されているのメタモデルの概略を示す。

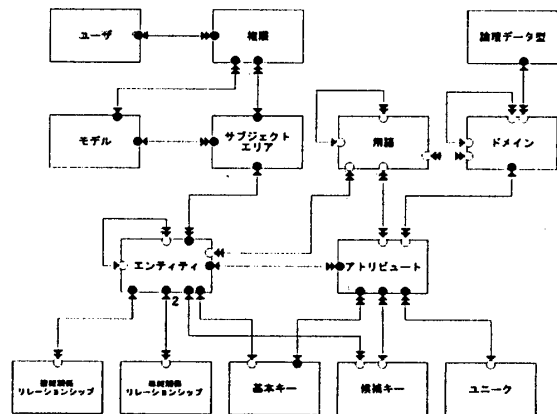


図3 メタモデル

4 おわりに

本稿では、統合的データモデリング支援環境“Darwin”のコンセプトと機能概要について報告した。

今後は、本ツール環境を実プロジェクトに適用、評価し、更にブラッシュアップしていくと共に、CORBA、OODBMS等を利用したリポジトリを実装し、より高度なグループウェア環境に対応していく予定である。

参考文献

[1]Durell.R.W：データ資源管理,日経マグロウヒル,1987
 [2]加藤：Durell命名規則の拡張によるデータ項目名称付与の自動化,情報処理学会第49回全国大会,3V-07,1994