

データモデル概要

全社規模のデータモデリングと
ユーザーサービスの両立

Debbie Smith
Data Warehouse
Consultant
Teradata Global
Sales Support

データモデル概要

目次

概要	2
はじめに	3
E. F. Codd博士の12の 法則に立ち戻る	3
EDWとは何か?	3
データ・インフラストラクチャ	3
データモデリング理論	4
論理モデリングから 物理モデリングへ	7
物理モデル	7
情報配信ツールへの影響	10
抽出、変換およびローディング (ETL) ツールへの影響	10
代理キー (Surrogate Keys)	11
変化するディメンション	12
Teradataの方向性	13
データウェアハウジングを 高度化する効果	14
まとめ	14
補足資料A - Codd博士の 12の法則	16
補足資料B - 比較	17
補足資料C - 用語	17
補足資料D - 巻末注	19
補足資料E - 参考文献	19

概要

データウェアハウスのためのデータモデル選定が、大きな争点となることがよくあります。データモデルを選定する場合、ビジネス部門のユーザーが簡単にアクセスでき、満足できるレスポンスタイムを実現するセルフサービス環境を提供したいと考えることでしょう。しかしながらこの場合、レスポンスタイムとは新しいアプリケーションのコンセプトを確立し、提供できるまでの期間も意味合いとして含まれなければなりません。ビジネス部門のユーザーにとって使い易さを実現し、同時に更新、拡張、可用性および管理上の観点から、現在だけでなく将来のニーズも満たすためにはどんなデータモデルを採用すべきでしょうか? 本資料は、一般的なデータモデリングを概説し、データモデリングに関するTeradataの位置付けをご紹介します。

データモデル概要

はじめに

データウェアハウスの構築を検討する場合、データモデルの選定が大きな争点となることがよくあります。データウェアハウスを構築する場合、どのように構築されるでしょうか？ 企業全体の現状および将来に合わせて構築しようとするでしょうか？ それとも、現在のビジネス部門のユーザーニーズを満たすような方法で構築するでしょうか？ 本資料は、企業が長期的に活用することができ、しかもビュー・レイヤーを利用してビジネス部門の各ユーザーのニーズや要件を完全に満たせるセントラル・エンタープライズ・データウェアハウスを構築し、正規化されたデータモデルのメリットと、ディメンショナルモデルが持つメリットの両方を活かす考え方を検討していきます。

E. F. Codd博士の12の法則に立ち戻る

1985年にCodd博士は、リレーショナル・データベースシステム設計のガイドラインとなった12の法則を提唱しました（補足資料A参照）。¹ Codd博士はリレーショナルモデルの創設者としてその功績が認められており、このガイドラインにそってテーブル（表）の中で情報を構成する手法を提示しています。そしてこのガイドラインに基づくとき、ヌル値は適切に特定され、さらに高度な挿入、更新および削除が行われます。特にガイドラインの6、8および9に注目してください。

>ガイドライン6；

ビュー更新ルール - 論理ビューは、すべてのデータ操作種類をサポートする。

>ガイドライン8；

物理データの独立性 - ユーザーは、データの物理的な保持形態とは無関係である。

>ガイドライン9；

論理データの独立性 - ユーザーは、物理データ構造が変化しても影響を受けない。

これら3つのガイドラインは、特にIT部門の作業がユーザーに影響を与えないようにし、物理データ構造から独立した、論理的で、ユーザーフレンドリーなデータ構造を構築するための基盤となり、そのための方法論となります。この基盤により、ビジネス部門の各ユーザーニーズを満たしつつ、ETL、更新頻度およびデータ管理要件のニーズにも容易に対応できる物理データモデルを維持し、結果的にユーザーフレンドリーの観点を伴ったエンタープライズ・データウェアハウスを実現することが可能となります。

EDWとは何か？

本資料は、エンタープライズ・データウェアハウス（EDW）のコンセプトに基づいて作成されています。多くの方がデータウェアハウスという用語を用いて同じ様なコンセプトを表わしていますが、その意味が多少異なる場合があります。データウェアハウスとは、組織の意思決定を支援する環境であるというのが一般的な見解です。現在は、これに企業（エンタープライズ）という言葉を付け加え、この環境が企業全体を反映するようになると考えられています。更に話を進める前に、Teradataがエンタープライズ・データ

ウェアハウスと謳う場合に、どんな環境を指しているのか検討してみましょう。当社は、EDWとはビジネス（企業）データを集中的に統合して保存し、その企業の共通見解として確立された方法でアクセスできる領域であると定義しています。データを集中化することによる価値は、データを一度格納して管理すれば、様々な異なる方法、タイミング、方式、そして理由に基づいてアクセスできることによる相乗効果であると当社は確信しており、このことは当社のお客様の事例で立証されています。ビジネスデータを収集し統合することにより、ビジネス部門のユーザーはデータを360度、あらゆる角度から見渡すことが可能となるのです。勿論、ビジネスデータを蓄積するのは継続的なプロセスであり、常に進化しています。その目的の1つは、ビジネスデータを追加し充実させる方法論となる長期的な戦略を確立することです。お客様が成功を積み重ねるに従って、インフラストラクチャも拡大し、社内外のユーザーのあらゆるデータに対するあらゆる（新たな問い合わせや今後のアプリケーションニーズも含む）問い合わせに対しても、追加的なリソースの介在なしにいつでも回答できるようになります。この相乗効果により、企業はビジネス情報を活用してマーケットの変化に迅速に適応し、時にはマーケットに変化をもたらすことが可能となります。

データ・インフラストラクチャ

データモデルは、データウェアハウスの中核を形成します。データモデルを定義するときには下した決定により、

データモデル概要

データ・インフラストラクチャが決まります。このデータ・インフラストラクチャは、新たなアプリケーションのパフォーマンスや導入期間に影響を及ぼし、またマーケットの変化に対する適応能力、ビジネス部門のユーザーの情報検索能力、データの鮮度、そして長期に亘るデータウェアハウスの最適化にも影響を及ぼすことになります。そして一般に、データウェアハウスのためのデータモデリングを検討する場合、正規化理論とディメンショナル理論が中心となります。

TeradataデータベースはRDBMSとして、明確に定義されたどんなモデルでも選択することが可能です。正規化モデルを実行するお客様が多数いらっしゃいます。また、スノーフレイクモデルやスタースキーマを利用しているお客様もいらっしゃいます。更に、これらのモデルのいずれか/全ての変形、または派生モデルを利用しているお客様も多数いらっしゃいます。従ってTeradataは、データウェアハウジングに関して豊富な経験を有するベンダーとして、採用されたデータモデルが最大の柔軟性を発揮し、現在のニーズだけでなく将来のニーズについても企業の期待に応えられるようにすることを最優先しています。これは非常に重要なことです。

データウェアハウジングが進化し、企業のビジネスのやり方に大きな責任と積極的な役割を担うようになっている現在、この益々重要な役割をサポートするためデータモデルに対する要求や期待が急速に変化しています。データ

ウェアハウスは、企業の変化をより迅速に反映することを求められています。また、データウェアハウスは現在のビジネス手法を分析し、企業のビジネスのやり方を変えています。このようなデータウェアハウスに対する要件の増大が、データモデリングの選択に影響を与えています。EDWの目標は、企業のニーズに対応することです。このため、組織における各要員の特定のアプリケーションニーズに対応する限定的なソリューションではなく、エンタープライズ・ソリューションを実現するデータモデルが必要となります。

EDWの役割は、全社的観点からビジネスを行えるようにすることです。そのためには、データモデリングに関して厳しい判断が求められます。データモデルは、企業のニーズを反映すべきか、それとも個人のニーズを反映すべきか？個人もしくは企業のニーズが変化したとき、データモデルにはどのような影響があるか？1つのことを優先するとき何が犠牲となるのか？認識された犠牲を最小限に抑えられた場合にはどのような効果があるのだろうか？これらの問題が検討されなければなりません。これまでの実績は、Teradataデータベースが物理データの独立性、論理データの独立性、そしてビュー更新に関するCodd博士のガイドラインを完全に活かし、企業の現在だけでなく将来のニーズをも満たせる柔軟性を備えたデータモデルを作成できることを示しています。そして同時に、ビューを利用することによって個人のニーズに対応できるビューモデルを構築します。これは、正規化モデルを活用すると同時にディメンショナルモデリング技法

を利用したビューで正規化モデルを補完するという一石二鳥の効果をもたらす方法です。データベース・テクノロジーやデータウェアハウス設計に関して35年以上の経験を有し、データウェアハウス業界の権威として認められているBill Inmonの講演や著作により、データウェアハウスのための優れたモデルとして正規化モデルが普及し、スタースキーマモデルは必要な場合に限りデータマートに使用されるようになってきています。Inmonは、“エンタープライズ・データウェアハウスがなければ、効果的な意思決定支援環境を構築することはできません。一般に、エンタープライズ・データウェアハウスの設計は正規化されています。従来のエンティティ・リレーションシップ手法でデータを構築することによってデータ構造を正規化する手法は、エンタープライズ・データウェアハウスの利用要件とうまくマッチしています”と述べています。²

データモデリング理論

エンティティ・リレーションシップ・モデリング

物理データモデルを設計し構築する前に、論理データモデルを構築するプロセスがあります。このプロセスには、ビジネス部門と幅広い検討を行うことが含まれます。この検討で、ビジネス要件を特定し、特定したビジネス要件を満たすために必要なデータ・エンティティおよびエレメントを確定し、データ・エンティティ間の関係を把握します。そして、理解したことをダイヤグラム化してビジネスの流れを示します。これを、エンティティ・リレー

データモデル概要

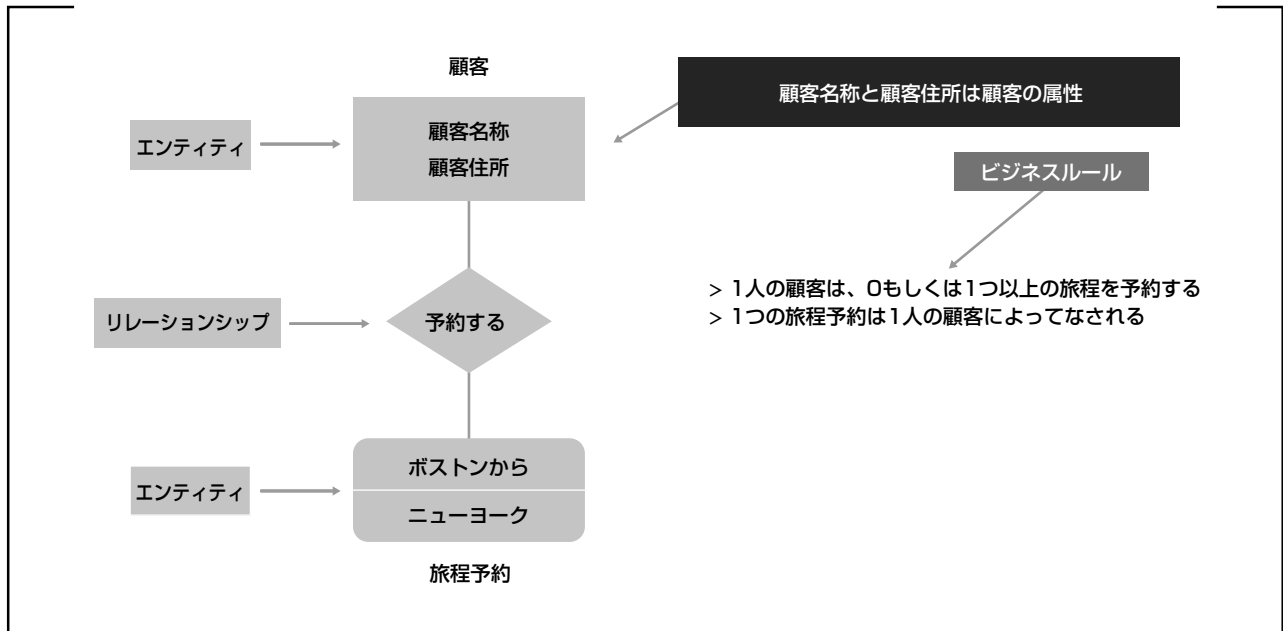


図1. ERダイアグラムの例

シオンシップ (ER) モデルと呼んでい
ます。ERモデルまたはダイアグラムは、
ビジネスのデータ・エンティティおよ
びこれらエンティティ間の関係を表わ
します。この時点では、モデルを利用
してサポートする具体的な機能や問い
合わせは含まれていません。これはビ

ジネスを論理的に表現したものであり、
後で物理データモデルやスキーマを開
発するときに利用します。ERダイアグ
ラムをマッピングするのに利用できる
ツールがいくつかあります。最も良く
知られているツールとして、ERwin®
(現在は、AllFusion™ ERwin Data

Modeler) があります。ERモデリング
は、正規化ルールに従います。これら
は、ビジネス関係の本質を確実に把握
することを意図したルールです。ERダ
イヤグラムのコンポーネント (エン
ティティおよび関係) の例を図1に示し
ます。

ディメンショナルモデリング

ディメンショナルモデリングは業務部
門のユーザーグループまたはビジネス
業務のためにデータを編成する別の論
理設計手法で、企業内のディメンショ
ンおよびレベルを把握し、ビジネスに
関する事実や指標を析出します。ディ
メンショナルモデルは、エンティティ
間における多対多の関係を排除する追
加ルールを適用し、多対1の関係のみを
考慮します。ディメンショナルモデル
は、複数のエンティティを1つの新しい
エンティティにまとめます。この新し
いエンティティは、ビジネスで発生す

るエンティティや関係を直接反映して
いませんが、ビジネス部門のユーザー
の対象となるグループにとって重要な
データポイントや、指標を保存する便
宜を図るために構築されます。ディメ
ンショナルモデリングの目的は、ビジ
ネス部門のユーザーがデータ・ナビ
ゲーションを簡単に行い、レポートに
迅速にアクセスできるプレゼンテー
ション層を提供することにあります。

ディメンショナルモデルは、時間、商
品、地域等をディメンション (軸) と

するキューブ (立方体) として考えら
れ、キューブとして表現されることが
よくあります。ビジネスの測定は、こ
れらのディメンションが交わる点にお
いて実施されます。ディメンショナル
モデルをキューブ (スタースキーマと
呼ぶことがある) として視覚化するこ
と (図2参照) により、そのデータ構造
をスライス・アンド・ダイス (切り口
を変更するためにキューブの一部を切
り取ったり、キューブを転がしたりす
ること) ができることが容易にイメージ
できるようになります。レポートイン

データモデル概要

グ目的でこれらのキューブを作成するときには、どんな問い合わせが行われるかを理解した上で設計する必要があります。各キューブは、特定のビジネス・アプリケーションに迅速且つ簡単にアクセスできるように設計されます。

ディメンショナルモデルは、非正規化によりディメンションをフラット化します。ビジネスルールを利用して、追加制約事項を検証し確定します（図2参照）。例えば、多くの企業は暦年度とは異なる会計年度を設けています。また、暦年度や会計年度とも異なる期間ディメンション（広告等）も設けて

いる場合もあります。広告週が会計週と一致しない場合、ビジネスルールを使用して広告週を判断し会計週と突き合わせ、より詳細なレベルまでドリルダウンできるようにします。実際には、業務ルールやプロセスをモデルに組み込みます。これを言い換えれば、モデルがサポートすべき問い合わせやアウトプットについてより深く理解する必要があります。ディメンショナルモデルはERモデルから作成することができますが、ERモデルはディメンショナルモデルから作成することはできません。複数のエンティティが1つにまとめられると、元のエンティティを

析出するのが不可能とは言えないまでも難しくなります。ディメンショナルモデルにおいて、中心となるファクトテーブルは単一の“結晶体”となり、最初のファクトテーブルの設計に固定化してしまことになります。新しいデータソースを組み入れるには、元のスタースキーマ設計をやり直し、別のファクトテーブル、またはデータマートを作成しなければなりません。従って、データを全社的な見方で把握するためには、これらの異なるスターやデータマートへのドリルパスが、全て構築されなければなりません。

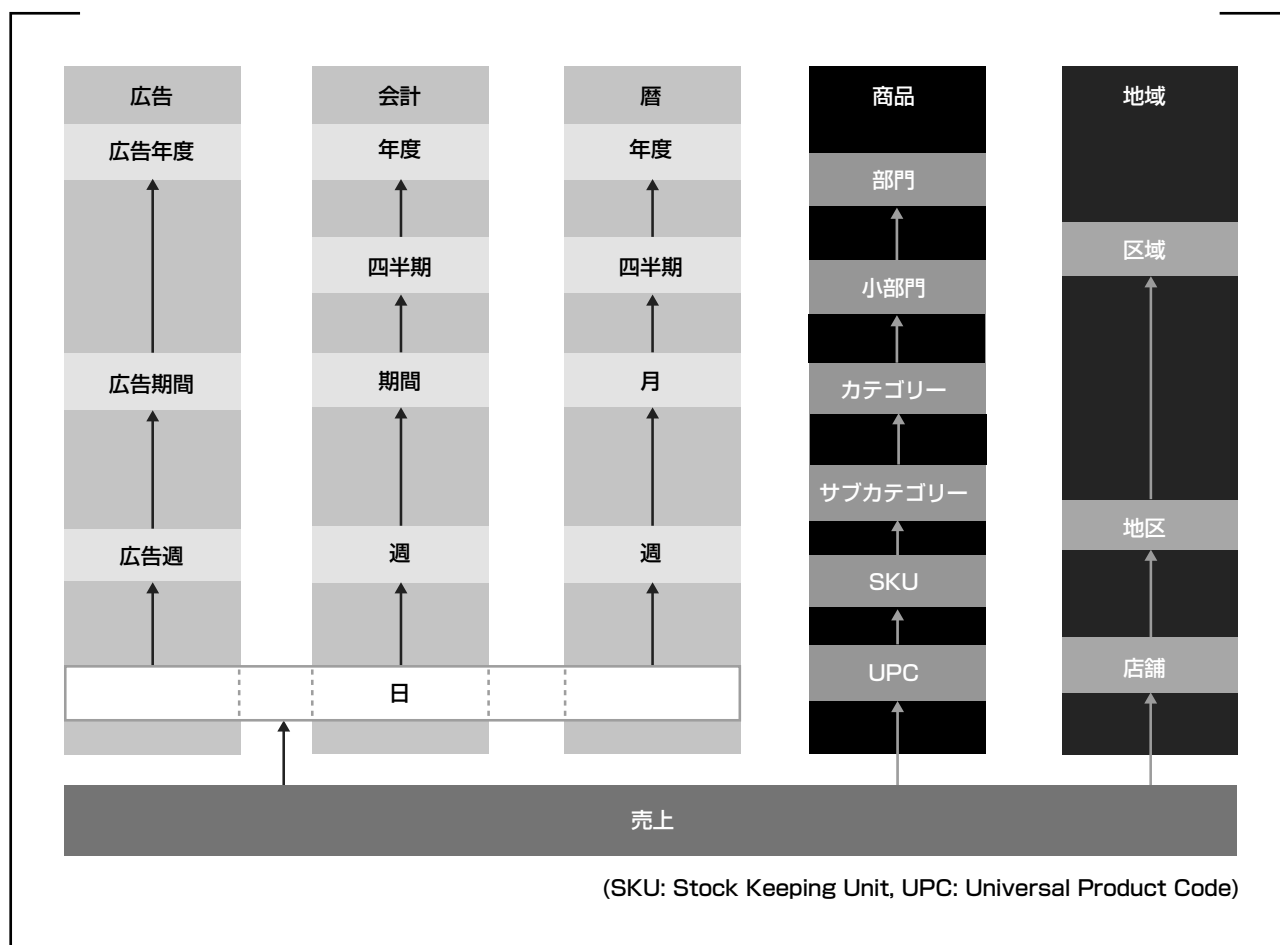


図2. キューブ（スタースキーマ）

データモデル概要

論理モデリングから物理モデリングへ

何故、論理モデルとは異なる物理モデルを作成するようになったのでしょうか？物理モデルは論理データモデルとは異なり、特定のデータベース・テクノロジー下において最高の総合パフォーマンスを実現できることがよくあります。それぞれのデータベース・テクノロジーは同じではなく、大きく異なります。Neil Radenは、Intelligent Magazineに寄稿した記事の中で、“ご存知の通り、データウェアハウスに利用される主要なスキーマは、スタースキーマか“正規化”スキーマです。正規化スキーマは定義が曖昧な用語ですので簡単に説明できませんが、一般にディメンションベースではない第3正規形（3NF、またはそれ以上高度に正規化された）スキーマと似ています。これらの3NF設計は問い合わせや分析をサポートせず、ステージングエリア、即ち、一連のスタースキーマやオンライン分析処理（OLAP）用キューブに代表される、分析担当者が問い合わせを直接実行するデータ構造の上位層に

位置づけられるデータ・リポジトリとしての役割を果たすことが唯一の目的です。但し、Teradataを導入した場合は異なります。Teradataは、超並列アーキテクチャおよびデータベース・最適化独自の機能によって3NFスキーマに対する分析SQLを適切なパフォーマンスで処理することが可能です”と語っています。⁴

これは論理モデルから物理モデルへの変換が、必ずしもビジネス要件に起因するものではなく、利用しているテクノロジーを一層活用し、問い合わせのスピードを実現するためである場合が多いことを意味しています。この論理モデルから物理モデルへの変換については、意見が様々に分かれ、違いが生じ、そのため当然のことながらデータモデリングに関して混乱が生じる原因となっています。片側にはフラット化された非正規化モデルが存在し、その対極に正規化モデルが位置づけられます。勿論、実際にはほとんどのお客様

が両極端の中間に位置するモデル、またはこれら2つを組み合わせたデータモデルを開発しています。これはデータウェアハウスが進化し、ビジネス要求を理解するようになった結果といえます。

Steve Hobermanは彼の著書であるData Modelers Workbenchにおいて、この傾向を正規化の“興隆”と呼んで取りあげ、非正規形の生き残りガイドを示しています。⁵ Hobermanの原則は、論理モデルが完全に（最大第5正規形まで）正規化されたとき、ビジネス業務をあたかも摩天楼から見下ろすように把握することができるかと述べています。従って、非正規形生き残りガイドは、正規化された論理モデルを物理モデルに変換する手法のガイドラインとなります。（この生き残りガイドの根底には、非正規化に関する幾つかの原則を提示し、非正規化を最小限度に抑え、ビジネスの現在および将来のニーズについて妥協しないようにする理念があります。）

物理モデル

正規化された物理モデル

モデリングの領域を推し進めていくとき（図3参照）、データモデルは左に行くほど正規化の度合いが高くなります。正規化モデルは、属性をプライマリーキーに関連付けするという、非常に厳格なルールに基づいて、データをテーブルに組み入れていきます。正規化の基本的ルールの1つは、モデル内のデータ冗長を除去し、各テーブル内の属性を機能的にプライマリーキーに依存させておくことです。このルールに従っ

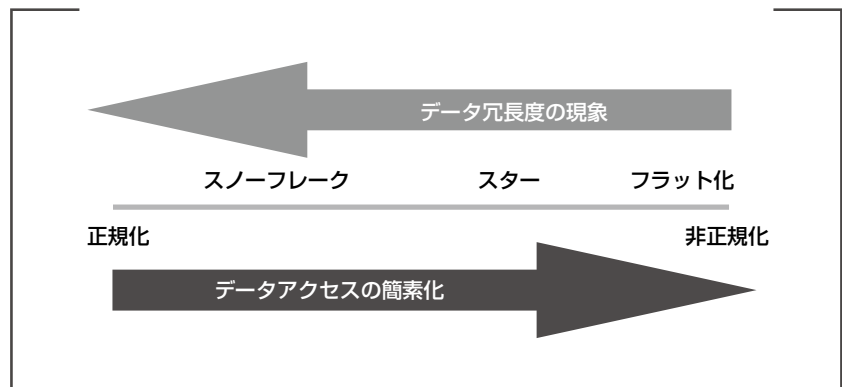


図3. モデリング領域

データモデル概要

てモデル内のデータの重複を除去していくと、テーブルの数とデータモデルの複雑性が増加していくこととなります。データウェアハウスに投入されるべき新たなデータサブジェクトが特定されたときには、この正規化ルールに従ってモデルが拡張され、新たなデータサブジェクトが組み込まれます。(データウェアハウスモデル担当者は、様々なデータウェアハウスをモデル化する手法についての議論に巻き込まれることが良くありますが、トランザクション・データベースの世界では正規化モデルが最適な選択であるという点で意見が一致しています。データウェアハウスが、最新のデータを利用したトランザクションベースの意思決定に進化していくのに伴って、正規化された物理設計の必要性は更に明確になっていきます。) 正規化モデルは、最も詳細なレベルのデータを収集し、蓄積する手法であり、同一のトランザクションから、幾つかのテーブルへの複数の更新を行う必要性を排除します。これは、ある特定のOLTPタイプのトランザクションにとって優れたモデルとなります。

多くのお客様は既に、正規化モデルはデータベースの構造を変えることなく、新たなビジネスの問合せや、それまで分からなかった問い合わせへの回答を提供できるということを体感しています。リレーションシップおよびエンティティが物理的に表示され、また企業にとっても最高の柔軟性を実現できるからです。正規化モデルはデータの重複を排除し、従って重複したデータを維持する煩雑さをも排除します。一

般に、正規化モデルには多くのテーブルがありますが、正規化の度合いが高くなると、データが重複していないのでストレージスペースを多く節約することが可能となります。一方でサマリーテーブル等の非正規化構造が、データウェアハウス内に構築された正規化構造の下に存在することも想定され、また実際に存在しています。サマリーテーブル構造を構築し管理するコストに見合うビジネス価値が生まれるときには、サマリーテーブルも正規化データモデルを補強する有効な方法です。

非正規化モデル

モデリング領域を右側に進んでいくと(図3参照)、非正規化の度合いが高まります。フラット化モデルは本質的に表計算ソフトのスプレッドシートと似通っています。非正規化データモデルは、ビジネス部門のユーザーが基盤となっているデータモデルを理解しなくても、データに迅速且つ簡単にアクセスできるようにする必要性から発達しました。ビジネス部門のユーザーはSQLやデータモデルのロジックを理解したり、テーブルをジョインする方法を理解したりすることが仕事ではないため、情報に簡単にアクセスする必要がありました。そしてこれらビジネス部門のユーザーは、スプレッドシートのフォーマットを理解しており、実際問題としてレポート形式での情報を要求することがよくありました。従って接続された一連のテーブルではなく、スプレッドシートに似た2次元形式でデータが提示されていた際に、ビジネス部門のユーザーはデータモデルの複雑性による影響を受けることはありませんでした。

データモデルは特定のビジネス部門の要員に対応するために作成されていたため、新たなユーザーグループ(パワーユーザーと呼ばれることがよくある)が組織内に生まれ、ビジネス部門のユーザーのためにレポートをコード化し作成するようになりました。しかしながら、これらのパワーユーザーは次第に多忙になっていき、レポート要求が数週間放置されるという事態も起こるようになりました。次にビジネス部門のユーザーは、直接的に情報にアクセスし、瞬時に結果を取得したいと考えるようになりました。これに対応する形で、非正規化モデルは単純で簡単なデータ・ナビゲーションを実現しました。通常、データは1つのディメンションに従って集約されません。要求された問い合わせのタイプが十分に理解され、これらの問い合わせに答えられるようモデルが構築されています。これらの予め定義されたモデルは、既知の問い合わせに対しての結果セットを迅速に返すことが可能です。新たな問い合わせが認識された場合には、既存データとの重複も許容した形で拡張され、それらがビジネス部門のユーザーをナビゲートし、新たな問い合わせセットに対して、迅速に結果セットを提供できるよう構築されます。

一般に、スタースキーマモデルはデータマートまたはデータウェアハウスで使用されると考えられており、特定のビジネスニーズ、レポートまたはアプリケーションに対応するように構築されています。各スタースキーマは個々のデータマートとして構築されます。ビジネスで新しいレポートが必要にな

データモデル概要

ると、別のスタースキーマまたはデータマートが構築されます。必要とされるディメンションが既に構築されているものとはほぼ同じである場合、新たなデータマートはその複製を保持することになります。これらのデータマートは、それぞれ個人のニーズに合わせています。これらのデータマートのグループがデータウェアハウスを構成すると考えがちですが、これらのデータマートは複製したデータを異なる方法で保存する、もしくは異なるデータベースやハードウェア上に保存するため別々の構造になっています。スタースキーマモデルは、データを一度保存しさえすれば、繰り返し利用可能な状態にすることを想定していません。その目的は、エンドユーザーが利用しやすいようにすることです。ビジネスグループに新たなニーズが発生すれば、通常新しいスタースキーマが構築され要求に応えることとなります。

例えば、ビジネス部門のユーザーが商品の月次売上を知りたい場合、作成されたレポートは通常、商品、1月度の売上、2月度の売上、3月度の売上等の項目を備えていることでしょう。一方で正規化された売上テーブルは、商品、日付および売上欄で構成されるはずで、そしてこの正規化されたテーブルは、最高の柔軟性を備えています。例えば、特定商品の週間売上は？ 特定会計月の売上は？ 月曜日に売れた商品は？ 週末に売れた商品は？ 等の問い合わせに答えることが可能です。しかしながらユーザーは、また時にはデータベースソフトウェアでさえも、テーブル構造をレポートレイアウトに変換するこ

商品	日付	売上
100012	01102001	10.00
300012	02122001	3.00
200012	01152001	2.50
100012	03042001	15.00

商品	1月売上	2月売上	3月売上	4月売上
100012	345.00	450.00	326.50	245.90
200012	456.60	376.50	210.00	390.00
300012	254.00	112.00	310.00	295.00
400012	510.00	610.00	590.00	545.00

図4. 非正規化されたテーブルの例

とが困難な場合がよくあります。従ってこのようなディメンショナルモデルの場合、データベース管理者はテーブルを単に非正規化されたレポート形式で保存することになります（図4参照）。適合性を維持するため、スタースキーマは限定されたディメンションで構築されます。Ralph Kimballは、データは変換を目的にステージングエリアに置かれると説明しています。そしてこれらのデータは、データウェアハウス・バスアーキテクチャ（同氏による命名）を利用して、そのデータを必要とするすべての異なるデータマートに伝播され、すべてのディメンションがステージングエリアで確立されたものと一致するようにします。このバスアーキテクチャは、使用されているすべての異なるデータマート及びサーバーを接続する作業です。Kimballは、“ステージングプロセスをサポートするために正規化データベースを作成しても構いませんが、これが最終目標ではありません。正規化構造は、ユーザーからの問い合わせがアクセスできないもので無ければなりません。なぜなら理解度やパフォーマンスに悪影響を与えるからです。問い合わせや表示サービスをサ

ポートするデータベースは、データウェアハウスにおける一部のプレゼンテーション層として考えなければなりません。通常、正規化されたデータベースはプレゼンテーション層から除外され、一方でプレゼンテーション層は、厳密なディメンション構造であるべきです”と語っています。⁶

データウェアハウスにはスタースキーマが最適であるというRalph Kimballの考えとは対照的に、Bill Inmonは、正規化モデル中心に構成されるのがEDWであり、スタースキーマモデルは必要な場合に限りデータマートに適用されるべきであると確信しています。Inmonは、“要するに、データウェアハウス設計のベースとして単純にディメンショナルモデルを構築することは、複数のスタージョインを検討する場合暗闇の中を手探りで進むようなものです。1つのスタージョインだけを検討したのでは、スタージョインに関して問題があることが明らかになりません。複数のスタージョインを検討すると、ディメンショナルモデリングの制約が明らかになります。これは、ディメンショナルモデリングはデータウェ

データモデル概要

アハウジングのデータベース設計技法として有効でないことを意味するのでしょうか? 決してそうではありません。ディメンショナルモデリングとスタージョインは、データマートの目的と完全に合致します。事実、私が明日データマートを設計しなければならないとすると、他の方法を利用することは考えられません。しかし、基盤データとなると、別の話です。基盤データ-データウェアハウス-は、ディメンショナルモデリングとは全く異なる取扱いが必要です。データウェアハウスは、スタージョインも含めたすべての意思決定支援活動に適切な基盤であり、非常に詳細で、かつ柔軟なデータを必要とします。データウェアハウスの理想的な構造は、正規化データです。正規化データは、従来から存在するいずれの方法にでも加工や絞込みが可能なのです”と語っています。⁷

また、Inmonは“ファクトテーブルは、外部キーとのリレーションシップという方法でディメンションテーブルに接続されます。ディメンションテーブルのキー群は、ファクトテーブルに存在します。ファクトテーブルに含められるリレーションシップは多数存在することでしょう。更にファクトテーブルは、必要な場合キー以外の他のデータも含んでいます。つまり、ディメンションテーブルは、キー以外のデータもファクトテーブルと共有する場合があります。ファクトテーブルとディメンションテーブルが接続されると、スターを形成します。このため“スタージョイン”と呼ばれています。ファクトテーブルは、非常に非正規化されとも効率的にアクセスできる構造であ

るという点で効果があります。ファクトテーブルが作成されると、効率的に分析することができます。しかし、一長一短があります。スタージョインは、非常に効率的であるが故に柔軟性に欠けています。スター内のデータを、本来そのスターを構築した際の目的以外の方法でアクセスしたい場合、スタージョインは非常に硬直的なものです”とも語っています。⁸

情報配信ツールへの影響

非正規化モデル、特にスタースキーマはビジネス部門のユーザーが直接データにアクセスできるという便利さにより広く受け入れられてきました。詳しいビジネス要件を収集したのち、開発担当者はリスクを軽減するために、またビジネス部門のユーザーが必要としているものを提供するために、最も簡単に導入できるのは非正規化モデルであると判断したことがよくありました。データウェアハウスは週または月1回しか更新またはリフレッシュされなかったため、更新の複雑さと関連する問題は緩和され、結果的に単純なデータモデルを維持することが可能でした。ビジネス部門のユーザーは情報を検索するために単純なSQLを理解し、コード化することを求められ、IT部門はデータウェアハウスの利用をビジネス部門のユーザーに任せ、データウェアハウスを最新の状態に保つことだけに責任を負っていました。データモデルで対応できない新しいビジネス上の問い合わせが発生し、スタースキーマ間のジョインが複雑になると、ビジネス部門のユーザーが簡単なインターフェースを介して必要な情報を検索するのに役立つ情報配信ツール

が開発されました。

これらの情報配信ツールの多くはスターモデルを前提に生まれましたが、高度化し、技術的に進化していくにつれて、より多くの正規化モデルも認識し、利用できるようになりました。今日では複雑なモデル上で、ビジネス部門のユーザーを簡単にナビゲートできるようになってきています。

抽出、変換およびローディング(ETL)ツールへの影響

情報配信ツールは、ビジネス部門のユーザーがデータウェアハウスから情報を検索するのに役立つように開発されましたが、ETL (Extract Transform & Loading) ツールは、トランザクションシステムからデータを抽出することを支援するために開発されました。ETLツールはトランザクションシステムからデータを抽出することを意図しており、これらのデータは正規形であるため、ETLツールは正規化されたデータモデルを前提に利用した場合に一層効果的であると言えます。

データウェアハウスが非正規化モデルで構築されていると、ほとんどの場合においてETLのためにデータを別のエリアに移すという追加ステップが必要になります。ETLツールは正規化モデルに効果的ですが、これらのデータを非正規化モデルに移すためには、加工や変換が必要になります。この加工や変換を取り扱うために推奨されている方法は、データ・ステージングエリアを設けることです。このステージングエリアは、抽出したデータをそこに置

データモデル概要

いて必要な形式に加工または変換した上でロードするのに使用します。またステージングエリアは、データ、特にディメンションを別のスタースキーマ・データマートに複製できるプラットフォームともなります。

正規化モデルへロードする場合、データをモデリングのために変換する必要が非常に少なくなり、実際にはロードプロセス中にその必要性が生じますのでステージングエリアが必要なくなります。正規化モデルでは、ステージングエリアをトランザクションシステムに存在するデータ品質の問題を解決するために利用するのが一般的です。例えば、キャラクターフィールドが同じルールセットに従うようにします。住所に、“street”が含まれている場合、“street”のすべてのフォーマットが同じになるように変換されます。例えば、STもSt.もstreetとなります。

代理キー (Surrogate Keys)

一般に、論理キーは各行について一意であり、行が有効な間は変更されない値のデータフィールドで構成されています。このキーは、物理的に導入される自然キーです。条件キーとインデックス（索引）が相互互換に使用されることがよくありますが、同じではありません。インデックスは、パフォーマンスを最適化するためのメカニズムです。物理的な導入方法は、採用されているデータベース管理システムにより異なります。

代理キーは人工的なキー、即ち生成されたキーで、自然キーの代わりにプラ

イマリーキーとして利用されます。代理キーを使用することは、絶対条件ではなく任意です。通常、これらのキーは数値で、行を一意に特定するためにランダムに作成されます。代理キーの使用は、データモデルの柔軟性やデータの独立性を向上させるために広まっています。ビジネスにおいて、変化は日常茶飯事の出来事です。企業は吸収・合併することもあれば、新しい商品やチャネルの追加、組織の再編成、新しいマーケットへの進出等が適時行われています。代理キーについて2つの議論があります。1つは、ビジネスが変化すれば、値も変わり、自然キーのフォーマットも変わる可能性があるという考えです。自然キーがプライマリーキーとして使用されている場合、そのキーの値またはフォーマットを変更することは煩わしい作業であり、履歴情報やデータ間の関係が失われる場合があります。代理キーがプライマリーキーとして使用されている場合、自然キーの値を変更することはテーブル内のカラムの値を変更するのと同程度の負荷であり、最小限度の影響で済みます。もう1つの議論は、自然キーが大規模、または複数のカラムから構成されている場合、代理キーの方が小規模でありモデル全体に一般適用できるため、必要なスペースが少なく済むということです。アクセスをしやすくするため、ジョイン処理や情報検索に利用する代理キーへとリードするためのプロセスが開発されます。

トランザクション環境から、ときには異なるデータをも統合するデータウェアハウスでは、代理キーは統合のため

の有効な手段となることがあります。データウェアハウスはサブジェクト指向であり、トランザクション環境のアプリケーションや業務指向とは異なります。同じサブジェクト（例えば、顧客）を対象とした複数のシステムを統合し再編成する場合、業務環境で発生しないデータを統合する必要が生じる場合があります。代理キーは、この統合されたサブジェクトの各インスタンス（各顧客データ）を一意に識別し、ビジネスが変化しても将来重複させる心配をしなくて済みます。これにより、時代の試練に耐えうる安定したデータモデルが実現します。しかしながら一方で、優れたモデリング手法に代わるものとして代理キーを利用しないよう気をつけなければなりません。

代理キーの使用を決定する前に、検討すべきことがあります。代理キーを作成するために、一意性と一貫性を確実に実現するプロセスを導入しなければなりません。行を更新したり、新たな行を挿入したりするには、まず代理キーを作成し、更に新しく作成されたキーが以前に使用されたことがないことを確認するルックアップ・プロセスを確立しなければなりません。ユーザーは代理キーを知りませんので、ユーザーが知っている値を代理値とリンクするプロセスを確立し情報を検索できるようにしなければなりません。一般に、これには追加的なセカンダリーインデックスを必要とします。代理キーの使用は各テーブルに一つのカラムおよび一意識別可能なインデックスの両方、更には複数のセカンダリーインデックスをも追加することになり

データモデル概要

がちで、全く新しいテーブルを追加する場合もあります。また、代理キーの値は外部キーとしてすべての従属テーブルに伝播しなければなりません。

代理キーはスタースキーマにおいてよく利用されますが、特定のデータモデルにだけに使用されるわけではありません。代理キーを使用するか否かの決定は、データモデル設計の検討事項の1つとして考えるべきです。この決定は、データモデル設計担当者が考える、その企業にとってベストと想定される方法に基づいて判断されるべきです。自然キーが変化する場合が存在するか？すべての行を一意にする必要があるか？

セカンダリーインデックスを追加することにより影響があるか？等が検討されなければなりません。

例えば、注文テーブルの自然キーが注文番号であり、ビジネス部門のユーザーが“ABC”社より受注した商品“X”の数量を知りたい場合、注文番号が分からない場合があります。しかし、商品番号及び会社名は分かることでしょう。代理キーがプライマリーキーとして使用されている場合、恐らく注文番号、商品および会社名がそれぞれセカンダリーインデックスとなっているでしょう。また、注文番号の自然キーがプライマリーキーとして使用されている場合、恐らく商品

および発注企業がそれぞれセカンダリーインデックスとなっているでしょう。代理キーの保守を行うという問題以外に、データモデル設計担当者は代理キーまたは自然キーをどんなケースにおいて使用するのが良いかを判断しなければなりません。例えば、注文番号をプライマリーキーとして使用することが有意義な場合がありますが、顧客テーブルの場合は顧客番号として代理キーを使用するほうが合理的かもしれません。代理キーを使用する決定はケースバイケースで検討すべきで、代理キーの使用を標準とすべきではありません。

変化するディメンション

変化するディメンションは、緩やかに変化するディメンション（SCD: Slowly Changing Dimension）とも呼ばれることがあります。時間という概念はときにデータウェアハウスに大きな影響を与えます。時間は、多くの変化を生み出す傾向にあります。顧客が転居する

と住所は変更となり、また名前が変わる人もいます。地域の人口統計が変化し、人口数も変わります。商品がこれまでのカテゴリーから別のカテゴリーに移動することもあるでしょう。これらが変化するディメンションなるものです。時間はすべてのデータモデルに

インパクトを与え、ディメンションは時と共に変化します。しかし、変化するディメンションは、モデルをディメンションに基づいて構築する場合のみ問題となります。一般に、SCDに対応する方法が3つあります（図5参照）。>タイプ1は、ディメンションの値が変

SCDタイプ1		SCDタイプ2			SCDタイプ3		
SKU	カテゴリー	SKU	カテゴリー	更新	SKU	カテゴリー	旧カテゴリー
1001	01	1001	03	20021029	1001	03	01
1002	02	1001	01	20010101	1001	01	15
1003	03	1002	02	20010101	1001	15	20
		1003	01	20020915	1002	02	02
		1003	03	20010101	1003	01	03
					1003	03	12
					1003	12	05
					1003	05	03

図5. SCDへの対応方法

データモデル概要

化したときに、これを置き換え、更新します。これは、常に最新のディメンションを提供し、履歴データの収集はできません。データと古いディメンションの値の関連が分からなくなってしまうからです。

>タイプ2は、ディメンションが変化すると、新しい行をテーブルに挿入します。これは、変化するディメンションの最新および履歴情報を提供し、通常どの行が最新の情報かを明確にする有効日を備えています。

>タイプ3は、変化したディメンションを古いカラムに移してからディメンションを置き換え更新します。この方法は、最新とその直前の変化に対応します。

通常、正規化モデルはタイプ2として開発され、変化した行がテーブルに挿入されるに伴い履歴が蓄積されます。ビジネス・アプリケーションが最新情報のみ必要とする場合、ビューを利用して最新の値のみをアプリケーションに提示することができます。履歴を必要とするまたは特定期間の値を特定できる必要があるアプリケーションは、すべての履歴データにアクセスします。

正規化モデルがタイプ2のSCDに対応するのに対して、ディメンショナルモデルの構築では変化するディメンションに対応するという課題が増大します。この課題には、変化するディメンションをどのように反映させるかという問題が含まれます。どのタイプがディメンションに対するビジネスニーズを満たすでしょうか？ すべてのビジネスニーズを満たすでしょうか？ どの方法が、ディメンショナルモデルにおける

すべてのデータマートのニーズを満たすでしょうか？

Teradataの方向性

Teradata Warehouseに最適なデータモデルの選択を行うことは、2つのステップからなるプロセスです。まず、Teradataデータベースが何に役立つかということとTeradataの専門要員が提唱することを区別して理解しなければなりません。Teradataデータベースはデータモデルに左右されることがなく、明確に定義されたデータモデルであればどんなモデルとでもその性能を発揮できる技術的機能とパワーを備えています。Teradataは問い合わせ処理の各ステップを並列化できますので、意思決定支援のために大量のデータを処理する際の技術的問題、そしてその複雑性（複数テーブルのジョイン、特定テクノロジー向けに要求される非正規化モデルのためのジョイン内ジョインを含む）を払拭することが可能です。Teradataのエンジニアリング部門では、継続的にオプティマイザのレビューとパフォーマンス改善エリアの識別をしており、データモデルの正規化、非正規化されたスタースキーマの類を問わずパフォーマンスの最適化を進めています。1990年代初期においては、スタースキーマを対象とした大規模テーブル/小規模テーブル・ジョイン等のジョイン効率向上手法を追求していきましました。Teradataデータベース V2R3で導入されたジョインインデックスおよびTeradataデータベース V2R4で導入されたアグリゲート・ジョインインデックスは高度なインデックス作成機能であり、Teradataデータベースユー

ザーが正規化モデルを利用することも、またデータ処理上メリットがある場合には、これらの機能を利用してスタースキーマを作成することも可能としています。Teradataデータベース V2R5は、IDカラムを導入することにより代理キーの作成を簡単に行えるようにしています。オプティマイザは、データモデルに関係なく学習およびインテリジェンス機能を発揮します。

多年にわたる現場の経験から、Teradataの専門要員はアトミックデータの正規化モデルが最高の柔軟性を実現し、従って、長期的に最大の効果を生み出すことを理解しています。上手に統合されたデータウェアハウスは、ビジネス業務を遂行する上での強力なメカニズムとなります。企業がデータウェアハウスを利用し、生産プロセス、発注プロセスまたは顧客の購買傾向のパターンを把握するようになると、それはビジネスプロセスの変化を促します。企業は、データウェアハウスがこれらビジネス上の変化への適応能力拡大を支援すると共に、それらの変化をデータウェアハウスに反映させることを期待しています。ビジネスの変化によりカラム（ディメンション）をファクトテーブルの1つに追加する必要が生じたとき、または商品のあるカテゴリーから別のカテゴリーに移したとき、企業が期待するのは、データウェアハウスがこれらの変化をビジネスの時間軸に従って反映させることなのです。ビジネス変化を反映させるデータモデルの変更に時間がかかり、ビジネスの時間軸に遅延を発生させた場合、データウェアハウスはビジネス上の成功を

データモデル概要

阻害するものになってしまいます。数多くのお客様との実績から、正規化モデルは（例えば）新商品を発売する上で最高の柔軟性と容易性を実現し、迅速にビジネス上の変化を反映させることが可能であることが繰り返して証明されています。Teradataの専門要員が正規化モデルから始めることを提案することが多いのはこれが理由であり、これによって新しいデータエレメントを柔軟且つ簡単に追加できるのです。しかしながら、これはデータウェアハウスが厳密な意味で第3正規形のすべてのルールに従うことを意味するものではありません。ビジネス要件によっては、サマリーテーブルが正規化モデルを補強するのに役立つこともよくあります。Teradataは、長期的に発生するすべての変化に対応し、ビジネスをサポートできるモデルの提供を目指しています。

Teradataデータベースが他のすべての競合製品とは異なる重要な要素の1つとして、管理の容易性、拡張の容易性に優れているために、ロード戦略を簡素化し、企業データを完全に統合できる正規化データモデルを活用できる点が挙げられます。これにより、全社規模を想定したデータウェアハウスの構築が容易であると同時に、個人のニーズに対しては、基盤データ構造をベースにビューを適用することが可能です。このビューによって、IT部門およびデータウェアハウスは個人のニーズに容易に対応することが可能となります。

ビューは基盤データ構造上に存在する

透明なレイヤーであり、ビジネス部門のユーザーによるアクセスを容易にするプレゼンテーション層を作成する手法です。Teradataデータベース上でビューを保持する際にはスペースを必要とせず、ほとんどどのようなプレゼンテーション層に対しても、効率的な基盤データモデルからの変換を可能としています。このため、ビューを利用する際の手間は、これらのビューを単に管理するのみで済むことになります。

データウェアハウジングを高度化する効果

データウェアハウスがアクティブ・データウェアハウジング環境に進化していくと、頻繁で継続的なデータ更新が必要となり、レポート主導からイベント主導の処理となり、マーケットの変化に一層迅速に適応する必要が求められるようになります。このとき、データモデルの位置づけは益々重要になっていきます。同じデータに対して、短い戦術的な問い合わせと長い戦略的な問い合わせが実行される混合ワークロードは、非正規形が内包する潜在的な課題を増大させることになります。アクティブ・データウェアハウスへの進化に伴い、既知の考慮条件（問い合わせの種類、問い合わせをする主体、想定したデータモデル等）の多くは、再検討が必要となります。この進化に対応するため、Teradataの競合他社によるテクノロジーも、正規化したモデルを実現できる機能や特徴を追加し始めてきています。情報配信のツールも、この進化に対応しつつあります。データウェアハウスがアクティブな環境に進化するに伴い、新鮮なデータに対し

ていつでもアクセスできるようにしなければならないため、非正規化モデルを維持するために費やせる時間はどんどん少なくなっていきます。

データウェアハウスのために選択するデータモデルは、それぞれのメリットを明確に理解した上で選択すべきです。データウェアハウジング初期の時代は、非正規化モデルへ依存する傾向にありました。モデル選択決定のほとんどは、ビジネス全体ではなく特定のビジネスニーズに対応することを目的に、または既存テクノロジー上の制約を考慮してなされました。これら初期の時代には、Teradataの専門要員が正規化モデルを提案するとお客様は意外に思われました。データウェアハウジングが進化してアクティブになり、また企業の変化に伴って発生するニーズや要件が増加していくにつれ、データウェアハウスに正規化モデルを利用するという考えはそれほど急進的ではなくなってきました。事実、有力アナリストやデータウェアハウス専門家は正規化モデルの価値を理解し、非正規化モデルに批判的になってきています。

まとめ

データウェアハウスの目的は、企業がよりの確かなビジネス決定を下し、その決定に基づいてタイムリーな行動をもたらすことができるようなツールを提供することにあります。データ・インフラストラクチャの強靭性こそが、データウェアハウスの長期的な成功を決定付け、企業が情報を活用してビジネスで成功を収めることを可能とします。Teradataデータベースはその強靭

データモデル概要

性により、市場における最も優れた組み合わせをサポートすることが可能です。データベース管理者は自身が管理しやすい正規化データモデルを作成することができ、このデータモデルは全社のデータをCodd、DateおよびInmonの教えに従って容易に、かつ完全に統合することが可能です。一方でユーザーにとって利用しやすいディメンショナルモデリングのプレゼンテーション層であるビューを利用することにより、個人のニーズへの対応も可能となります。Coddのガイドラインを活かしつつ、Kimballの教えに従ってビジネス部門のユーザーが簡単にアクセスできるように、データウェアハウスとのインターフェースを構築することが可能です。これは、データの集中的な蓄積、冗長データの排除、容易なデータ管理をもたらし、既存のデータ構造のまま将来のビジネスニーズに対応し、市場変化に適応する時間の短縮をもたらすことにつながります。

あるTeradataのお客様で、ビジネス部門のユーザーがIT部門へ新しいアプリケーションの作成を要求した例をご紹

介します。アプリケーション担当グループは、データを見つけるだけでも6ヵ月かかり、その後でアプリケーション開発に要する期間を判断しなければならないと伝えました。ビジネス部門のユーザーは、それでは問題をタイムリーに解決できないとあきらめました。時間はビジネスの成否に直接影響を与えるファクターです。しかし、データウェアハウス担当グループは、それらのデータがデータウェアハウス内に正規形で存在することを見つけたため、ビジネスニーズの解決に必要なプレゼンテーション層を2時間で作成することができました。そしてこの能力は、企業収益に大きな貢献をもたらしたのです。

データウェアハウスのデータモデルは、ビジネス要件を反映し、またビジネスそのものの実現要素であるべきです。特定のデータモデルタイプに囚われるとビジネスニーズに対応できないことでしょう。時間が企業に変化をもたらすため、データウェアハウスにはこれらの変化に対応できる柔軟性が必要となります。そしてそのためには正規化モデルが正しい出発点となります。

筆者紹介

Debbie Smithは、TeradataのGlobal Sales Support部門に所属するデータウェアハウス・コンサルタントです。ある小売業において14年にわたり、Teradataシステム上でのアプリケーション開発/サポート、データベース管理、ビジネス部門のパワーユーザー、ビジネス部門のユーザーのための情報配信戦略の開発/導入/管理を担当していました。Teradataに招聘されてからは、データウェアハウスを検討しているお客様、新規にデータウェアハウスを導入したお客様、または長期にわたって活用しているお客様のために幅広く活躍し、効果的な戦術的および戦略的データウェアハウジング・プロジェクトの遂行に貢献しています。

データモデル概要

補足資料 A—Codd博士の12の法則

法則 1: 情報

すべてのデータは、テーブル形式でユーザーに表現されなければならない。

法則 2: アクセスの保証

すべてのデータは、アクセス上の曖昧さが排除されなければならない。これは、テーブル名、プライマリーキーおよびカラム名を組み合わせることで実現できる。

法則 3: スル値の体系的な取扱い

フィールドは空のままの状態でも可とできなければならない。このためには、空のストリングまたは値がゼロである数値とは異なるスル値をサポートする必要がある。勿論、これはプライマリーキーには適用できない。更に、ほとんどのデータベース導入においては、特定のテーブルカラムでスル値を禁止する、非スル値フィールド制約というコンセプトがサポートされる。

法則 4: リレーショナルモデルに基づいたダイナミック・オンラインカタログ

リレーショナル・データベースは、データへのアクセスに利用するのと同じツールでその構造にアクセスできるようにしなければならない。通常、これは構造定義を特別なシステムテーブルに保存することにより実現できる。

法則 5: 包括的なデータサブ言語

データベースは、データ定義、データ操作、データ整合性およびデータベーストランザクション管理機能を含む明確に定義された言語を少なくとも1つサポートしなければならない。ほとんどの商用

リレーショナル・データベースは、サポートする包括的な言語として標準SQL (Structural Query Language: 構造化問い合わせ言語) を利用している。

法則 6: ビューの更新

データは、ビューという異なる論理的な組み合わせでユーザーに表示可能である。各ビューは、テーブルへの直接的なアクセスに利用できるすべての種類のデータ操作をサポートすべきである。実際には、論理ビューに対して更新や削除アクセスを行うことは難しく、現在のデータベースでは完全にサポートされていない。

法則 7: 高度な挿入、更新および削除

リレーショナル・データベース内のデータは、複数の行または複数のテーブルからのデータで構成されたセット単位で操作することができる。つまり、単一テーブル内の単一行ではなく操作可能なすべてのデータセットについて挿入、更新および削除操作をサポートすべきである。

法則 8: 物理データの独立性

ユーザーは、情報の保管方法やデータベース検索の物理的方法とは無関係である。基盤となるアーキテクチャ (ハードウェア、ディスクストレージ) をユーザーのアクセス方法に影響を与えることなく変更できる。

法則 9: 論理データの独立性

データベースの論理構造 (テーブル構造) が変化しても、ユーザーへのデータ表示方法は変更すべきではない。この法則に従うことは特に難しい。ほとんどのデータベースは、ユーザーへのデータ表示方

法と基盤テーブルの実際の構造とが強く結び付いている。

法則 10: 整合の独立性

データベース言語 (SQL等) は、データベースの整合性を維持するためのユーザー入力制約をサポートすべきである。この法則は、ほとんどの主要ベンダーが完全には導入していない。すべてのデータベースは、最低限以下の2つの制約を、SQLによって遵守する。

プライマリーキーのどのコンポーネントもスル値を持つことはできない (法則の3を参照)。

外部キーが1つのテーブルに定義されている場合、そのいずれの値も別のテーブルにおいてプライマリーキーとして存在しなければならない。

法則 11: 分散の独立性

ユーザーは、データベースが分散しているか否か (データベースの各部分が複数の場所にまたがって存在しているか否か) を全く気にしないでいられるようにすべきである。様々な理由により、この法則を導入するのが難しい。

法則 12: データ破壊からの独立性

データベース構造を修正する方法は、複数行を操作するデータベース言語 (SQL等) を利用する以外に存在してはならない。現在ほとんどのデータベースは、データ構造をある程度直接操作できる管理ツールをサポートしている。

(出典: ITWorld.com, 'Codd's 12 Rules - Data Management Strategies' 2001年5月7日)

データモデル概要

補足資料 B-比較

	正規化モデル	非正規化（スター）モデル
アクティブ・データウェアハウジング	アクティブ・データウェアハウジングの推進をサポートする。	アクティブ・データウェアハウジングへの対応上課題がある。
問い合わせ	正規化モデルは柔軟性に富み、どんな問い合わせにも答えられる。	非正規化モデルの単純さは、変更無しで回答可能な問い合わせに限界をもたらす。
複雑な分析	正規化モデルは、複雑な分析をサポートする。	非正規化モデルは意図的に単純化しているため、複雑な分析はサポートできない。
データの重複	正規化のルールはデータの一元性をもたらす。	異なる種類の問い合わせに答えるため、データが何度も複製される。
データの粒度	正規化モデルは、最も詳細なレベルのデータを容易に保存することが可能である。	ナビゲーションを簡単に行えるようにするため、非正規化モデルは通常1つのディメンションにそって集約される。
データ・ナビゲーション	正規化モデルの方が、ナビゲートするのは難しい。ビジネス部門のユーザーは、モデル、必要なジョイン、更に場合によってはデータエレメント間の関係を理解するのに苦労する。	非正規化モデルは、特にビジネス部門のユーザーのために簡単なナビゲーションを目的として設計されている。
柔軟性	データモデルの正規化の度合いが高くなると、柔軟性が高まる。	非正規化モデルは単純さを意図しており、特定の要件を対象に設計されている。要件が変化すれば、モデルを変更する必要がある。
メンテナンス	正規化モデルの更新は、1度更新すればその情報は1つの場所にしか存在しないため、容易である。標準的なトランザクション環境は正規化モデルである。データウェアハウス全体で正規化モデルを継続することにより、トランザクションプロセスと緊密にリンクし反映させることができる。	非正規化モデルの更新は、事前処理、集約を必要とし時間を要する。

補足資料 C-用語

属性（アトリビュート）

エンティティに從属するデータもしくはエンティティを記述するデータ。

キューブ

ディメンショナルモデルがどのように構成されているか視覚化する手法。また、指標やディメンションをOLAPツールが前提とするフォーマット（多くの場合、ベンダー固有のフォーマット）で保存するディメンショナルモデルの設計手法。1つのキューブは、1つのファクトテーブルとその構成ディメンションを表わすことがある。

（論理的な）データマート

データは1カ所に保存されるが、ビューを利用し、目的に基づいて基盤データの一部または派生データをエンドユーザーに表示する。データは複製されない。

（物理的な）データマート

エンドユーザーによる特定の問い合わせや要件に対応するために集約された、または事前処理されたデータ・セグメント。

データウェアハウス

サブジェクトエリア毎に構成され、全

社のかつ複数のエンドユーザーの要件や問い合わせに対応するため保存された、統合的な事実の投影。

ディメンション

ファクトを測定する企業の‘基準線’（期間、商品、地域）。

ディメンショナルモデル

アクセスの簡便さおよびレスポンスタイムの短縮を図るために事前定義された、一連の問い合わせに基づいてデータをモデル化する設計手法。

データモデル概要

ドリルアクロス

異なるディメンション間を移動するエンドユーザーのリクエスト。

ドリルアップ/ドリルダウン

1つのディメンションにそって移動するエンドユーザーのリクエスト。

エンティティ (主体)

他のすべてのオブジェクトから一意に識別可能なオブジェクト。

エンティティ・リレーションシップ・モデル

重複することなく企業のエンティティ (またはオブジェクト) および企業内のエンティティ間の関係を示す手法。

ETL

データを抽出し、変換し、そしてロードするツールグループを指す用語。抽出、変換およびロードの一連の変更プロセスを意味する場合もある。

ファクト

測定値または値 (量的数量)

フラット化モデル

ビジネスレポートが必要とする全てのカラムヘダーのデータ要素を含む行で構成される。レポートを作成するにあたって何も計算をする必要がない形式。

外部キー

他のエンティティのプライマリーキーを反映するテーブル内のカラム。エンティティ間の関係を決定する。

階層 (ヒエラルキー)

複数のエンティティを階級付けし順番

に並べたもの、もしくはそれらを行う処理。

インテリジェントキー

ビジネス要件、もしくはビジネス上で利用を反映している値で構成されているキー。

自然キー

各行の一意性を確立する値のデータフィールド。この値は行が有効である間変更されない。

正規化モデル

データ冗長性を排除するように設計されたエンティティのモデル。'形式 (Form-2NF、3NF、4NF等)' が高度であればあるほど、冗長性は低くなる。

OLTP

オンライントランザクション処理。数千のトランザクションを逐次的、機械的に繰り返し処理する一連の命令で構成されており、事前定義されたアクセスパスを介し、最終的に基盤となるデータストアを更新する。OLTPトランザクションは、データ取得に重点を置くトランザクションシステムが利用する。通常、OLTPトランザクションは1秒未満のレスポンスタイムを要求する厳しいサービスレベル合意 (SLA : Service Level Agreement) を伴う。OLTPはデータの更新、もしくは挿入に集中する。

プライマリーキー

テーブル内における行の、一意な識別子。

スキーマ

本資料では、エンティティ間の関係を確立し表示する構造的なフレームワークを指す。

サービスレベル合意 (SLA)

あるプロセスのパフォーマンスに関するIT部門とビジネス部門間の合意。可用性をベースにしてデータを特定の時間内にロードすることや、検索プロセスをベースとして特定の問い合わせや問い合わせ種類の処理時間を義務付けることが合意内容となる。

緩やかに変化するディメンション (SCD)

時間経過に伴うディメンションの変化もしくはこれらの特性をもつディメンション (例、顧客の住所)。

スノーフレークスキーマ

複数のスタースキーマで構成されており、ジョインできるディメンションテーブルは複製しないスキーマ。

スタースキーマ

複数のディメンションテーブルとそれに囲まれたファクトテーブルで構成され、ディメンションテーブルとファクトテーブルのジョインにより利用されることを想定したスキーマ。

代理キー

自然データキーの代理として利用される人工的なキー (例、顧客ID)。

データモデル概要

www.Teradata-j.com

補足資料 D-巻末注

1. Dr. E. F. Codd, Codd's 12 Rules – Data Management Strategies, ITWorld.com, 05-07-2001.
2. William Inmon, Enterprise Data Warehouse, <http://www.billinmon.com/library/articles/introedw.asp>.
3. Ralph Kimball and Margy Ross, The Data Warehouse Toolkit, Second Edition, Wiley Computer Publishing, New York, 2002, pp 11 – 12.
4. Neil Raden, edited by Ralph Kimball, Intelligent Enterprise, Data Warehouse Designer, 'Real Time : Get Real, Part II', www.intelligententerprise.com, June30, 2003.
5. Steve Hoberman, Data Modeler's Workbench, Tools and Techniques

- for Analysis and Design, John Wiley & Sons, New York, 1996.
6. Ralph Kimball and Margy Ross, The Data Warehouse Toolkit, Second Edition, Wiley Computer Publishing, 2002, page 9.
 7. William Inmon, The Problem with Dimensional Modeling, <http://www.billinmon.com/library/articles/artdimmd.asp>.
 8. William Inmon, Star Joins, <http://www.billinmon.com/library/articles/starjoin.asp>.

- & Sons, New York, 1996.
- Daniel L .Moody, Department of Information Systems, University of Melbourne, Kortink, Mark, From Enterprise Models to Dimensional Models : A Methodology for Data Warehouse and Data Mart Design, Simsion Bowles & Associates, 2000.
- William Inmon/Peter Meers, The Dilemma of Change : Managing Changes over Time in the Data Warehouse/ DSS Environment, White paper on <http://www.billinmon.com/library/whiteprs/Dilemma of Change.pdf>, March 2001.

補足資料 E-参考文献

Ralph Kimball, The Data Warehouse Toolkit, Practical Techniques for Building Dimensional Data Warehouses, Solutions from the Expert .., John Wiley

本資料および本資料に含まれている情報は、Teradata Corporationの占有資産です。次の条件で本資料の閲覧、複写、印刷および配布を許可します：本資料は商用目的以外の情報取得目的に限り使用することができ、“現状のまま”提供します。本資料または本資料の一部を複写した場合、本資料に記載してある著作権表示および他のすべての制約条件を含めなければなりません。

注：本資料で解説した製品、プロセスまたはテクノロジーはTeradataが留保している他の知的財産権の対象となる場合があり、本資料によりそのライセンスを許諾するものではありません。米政府による利用、複製または開示は、DFARS 252.227-7013 (c) (1) (ii) およびFAR 52.227-19に定められた制約に従うものとします。

TeradataはTeradata Corporationの登録商標です。本文中の社名、製品名は各社の商標または登録商標です。本資料に掲載されている情報は予告なしに変更されることがあります。本資料で記述した全ての機能、実行内容が全世界で販売されているとは限りません。最新情報についてはTeradata担当者にお尋ねください。

(c)2007 Teradata Corporation All Rights Reserved.

TERADATA
Raising Intelligence

日本テラデータ株式会社

〒104-0033 東京都中央区新川1-21-2 (茅場町タワー)
<http://www.teradata-j.com/>

TDMK-5009 (0712)