



ビジネス インテリジェンスにおける データ モデリングの利点

2008 年 12 月

目次

概要	1
セクション 1	2
はじめに	2
セクション 2	2
BI 用のデータ モデリングが必要な理由	2
セクション 3	4
情報の意味を理解する	4
セクション 4	7
レポート作成を支援する	7
セクション 5	8
まとめ	8

Copyright © 2008 CA. All rights reserved. 本書に記載された全ての製品名、サービス名、商号およびロゴはそれぞれ各社の商標またはサービスマークです。本書は情報提供のみを目的としています。準拠法により認められる限り、CA は本書を現状有姿のまま提供し、商品性、特定の使用目的に対する適合性、第三者の権利に対する不侵害についての黙示の保証を含むいかなる保証もしません。本書の使用が直接または間接に起因し、逸失利益、業務の中断、営業権の喪失、業務情報の損失等いかなる損害が発生しても CA は責任を負いません。CA がかかる損害について明示に通告されていた場合も同様とします。

概要

課題

ビジネス インテリジェンス(BI)は、今や多くの組織で欠かせないものとなっています。膨大なデータが蓄積される中で課題となるのは、データの意味を理解して、ビジネスに役立つ情報を見つけ出すことです。データベース開発者がビジネスの要件を理解し、物理データベース システムへ実装するには、データモデルが重要なコミュニケーション ツールとなります。そのためには、次のような課題があります。

- 重要なビジネス用語の意味を理解する
- レポート作成の要件を満たし、ユーザーが正確な情報を使用して柔軟なクエリーを作成できるようにする

成果

データベース開発者がビジネスの要件を物理データベース設計に正しく反映すると、実装されたシステムを通じて、組織が保有する大量のデータから必要な情報を見つけ出すことができます。それによって、次のような成果が期待できます。

- 詳細情報から要約レベルに至るまで、任意のレベルで組織の業績に関する正確なレポートを作成できる
- 過去の実績から将来の出来事を正確に予測し、ビジネス ユーザーは組織の戦略を決定するために必要な情報を得ることができる

利点

BI システムのデータ モデリングを通じて、データ処理に関する多くの課題を解決できます。BI システムの論理および物理モデリングを活用すると、ビジネスに関する正確な情報をビジネス ユーザーに提供できます。主な利点は次のとおりです。

- ソース システムを詳しく理解することができ、BI システムの開発期間が短縮される
- BI システムから、より正確な結果を取得できる
- ビジネス ユーザーと開発者の双方が、利用可能な情報を分かりやすい形式で確認できる

セクション 1

はじめに

ビジネス インテリジェンス (BI) は、その名前が示すとおり、ビジネスに役立つ情報を取得するための手法です。BI システムからの出力は、企業が最高レベルの組織戦略を構築するために使用されます。BI は大きな利益をもたらす可能性があります、一方でデータの解釈に誤りがあると大きなリスクにつながります。そのようなリスクを避けるには、使用するデータについて詳しく理解しておくことが不可欠です。その点で、データ モデルは、ビジネス ユーザーと IT 技術者とのコミュニケーションを効果的に行うために有用です。BI システムのために構築されるデータベースには、正しい情報がビジネスで利用できる形式で格納されている必要があります。

ビジネス ユーザーは通常、BI システムの出力をレポートやデジタル ダッシュボードの形式で確認しますが、これらのシステムはオンライン分析処理 (OLAP) キューブによって実現されています。OLAP キューブには、あらかじめクエリーの結果が集計されており、本来は数時間かかるような処理であっても、わずか数秒で結果を生成できます。

このホワイトペーパーでは、さまざまな BI コンポーネントの出力を文書化、説明、および明確化する上で、データ モデリングが果たす役割について説明します。各コンポーネントは構造が異なるため、データ モデルの作成者はコンポーネントごとに別々のアプローチをとる必要があります。しかし、モデリング ツールを注意深く使用すれば、情報の誤りを避けるだけでなく、システム的设计を十分に理解できるという利点も得られます。

注意: 本書では、分かりやすさを重視した単純なモデルを使用しています。実際のモデルには、通常、さらに多くのエンティティと属性が含まれています。

セクション 2

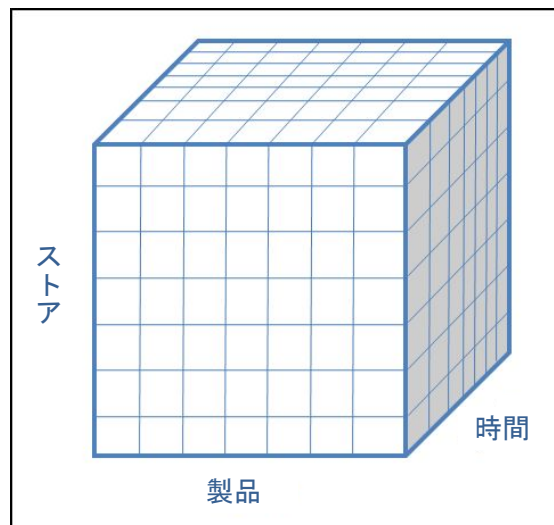
BI用のデータ モデリングが必要な理由

例として、食料雑貨の国際企業を考えてみましょう。毎週月曜日の朝に、取引チームはピボット テーブルを使用して、価格と数量からなる総売上高を確認します。売上高は製品グループ、個々の製品、地域、および店舗のレベルまで細分化されています。このピボット テーブルを作成するには、データベースでクエリーを実行して数百万ものレコードを集計する必要があります。情報を異なる基準 (月次ではなく週次など) で集計する必要が生じるたびに、非常に負荷の高いクエリーを再実行することになります。OLAP はこのような問題に対処するためのものです。

OLAP ストレージは、リレーショナル システムとはまったく異なります。OLAP を視覚化するもっとも簡単な方法は、キューブ (立方体) のような多次元構造として表すことです。キューブの軸はディメンション (次元) を表し、各ディメンション メンバーの交わる部分が集計結果となります。

図 1 に、3 次元のキューブを示します。

図 1: OLAP キューブ



製品、ストア、および時間のディメンションから構成された OLAP キューブ

実際には、より多くのディメンションが使用されることがありますが、図式化するのが困難になるため、OLAPの構造はディメンション数に関わらず常にキューブとして表されます。各製品は、「製品」軸のカラム(スライス)に対応します。同様に、各ストアは「ストア」軸のスライス、各日付は「時間」軸のスライスに対応します。製品、ストア、および日付の値を指定すると、セルの座標が決まります。このセルには、指定したディメンションのメンバーに対応するファクト テーブルが存在し、それぞれのメジャーによる集計値が格納されています。

ディメンションは通常、フラットな構造ではなく階層構造になります。たとえば、時間ディメンションには日、週、月、四半期、および年のレベルを含めることができ、すべてのレベルで値を集計できます。さらに、時間ディメンションは単一の階層構造にはなりません。週レベルは月レベルにぴったり収まらないため、このディメンションには複数の階層構造が必要です。

OLAP の設計は複雑な作業なので、処理対象のデータについて十分に理解しておく必要があります。そのために論理モデルを作成すると、データ ウェアハウスの物理モデル構造よりも作業に役立ちます。物理モデルは、データ型やインデックスなど、データベースのメタデータを正確に記述します。一方、論理モデルは物理モデルの上位にある抽象レイヤーであり、他のユーザーにとって分かりやすい構造です。

モデルに含まれるエンティティに詳細情報を追加することも重要です。通常、エンティティには単純な名前が付けられることが多いため、そのエンティティについての疑問が出てきます。たとえば「システム内で時間の構造はどのようになっているのか?」、「ストアとは何か? 実際の店舗が必要なのか、あるいはオンラインの Web サイトでもいいのか?」、「製品は買ったものなのか売ったものか?」などです。実際に役立つ BI 出力を作成するには、扱うデータについて完全に理解しておく必要があります。

システムを正しくモデル化することにより、わずか数秒で必要としている結果を得られるような OLAP キューブを生成することができます。

セクション 3

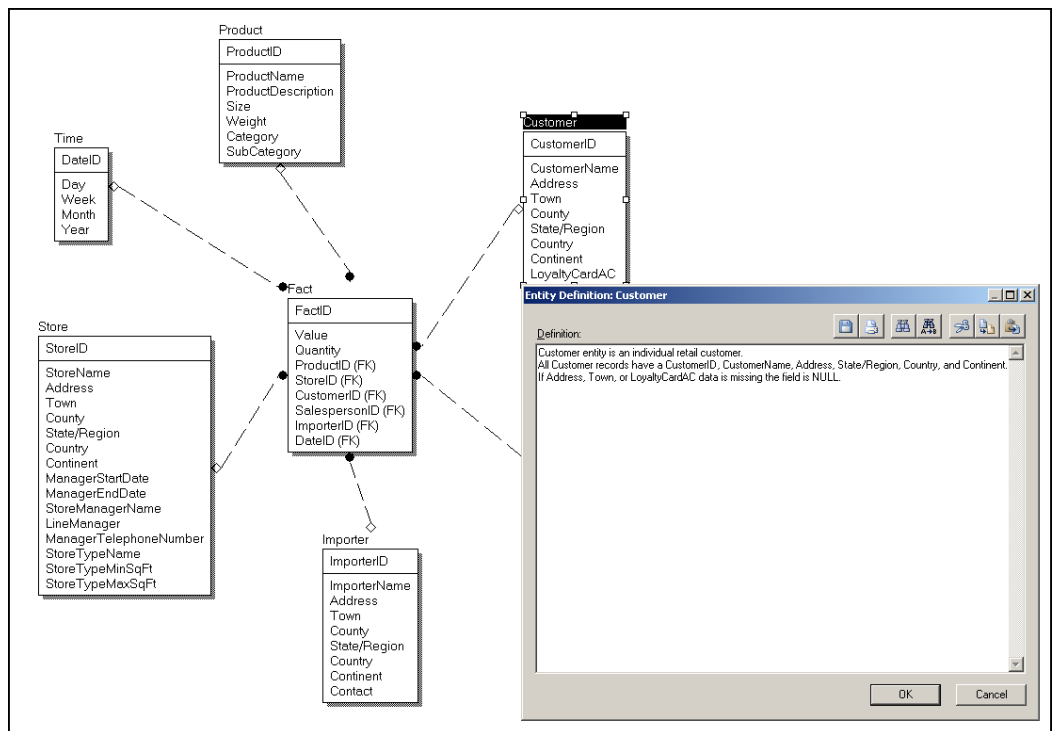
情報の意味を理解する

再び、例として食料雑貨の国際企業を考えてみます。顧客には、会社、および実店舗とインターネットで購入する個人がいます。これらの顧客タイプは一律に「顧客」として記述されていますが、それぞれ異なるソース システムで処理されています。このような状況から、ビジネス ユーザーが使用するピボット テーブルに、無効な顧客データが追加されるという問題が BI システムに発生しました。

この問題を解決するには、論理モデルに詳細情報を記述するべきであり、設計時に次のような事柄を考慮すべきです。たとえば、顧客とは何か？ という問題です。顧客は個人なのか、それとも会社なのか？ 顧客が会社であれば、会社に属する個人の情報も記録する必要があるのか？ 製品とサプライヤなどのディメンション間のリレーションシップはあるのか？ これらの疑問をあらかじめ解消しておけば、作成する論理モデルに、OLAP キューブの各要素の関係および機能を記述できます。食料雑貨の国際企業の例では、ビジネス ユーザーはピボット テーブルを利用できますが、ピボット テーブルが作成される仕組みを理解する必要はありません。求められるのは、ビジネス ユーザーとデータ モデル作成者との間で、顧客、ストア、製品、月、または販売単位などの定義が同一であることです。

図 2 に示す論理モデルには、OLAP キューブの作成時に必須となる詳細レベルが追加されています。

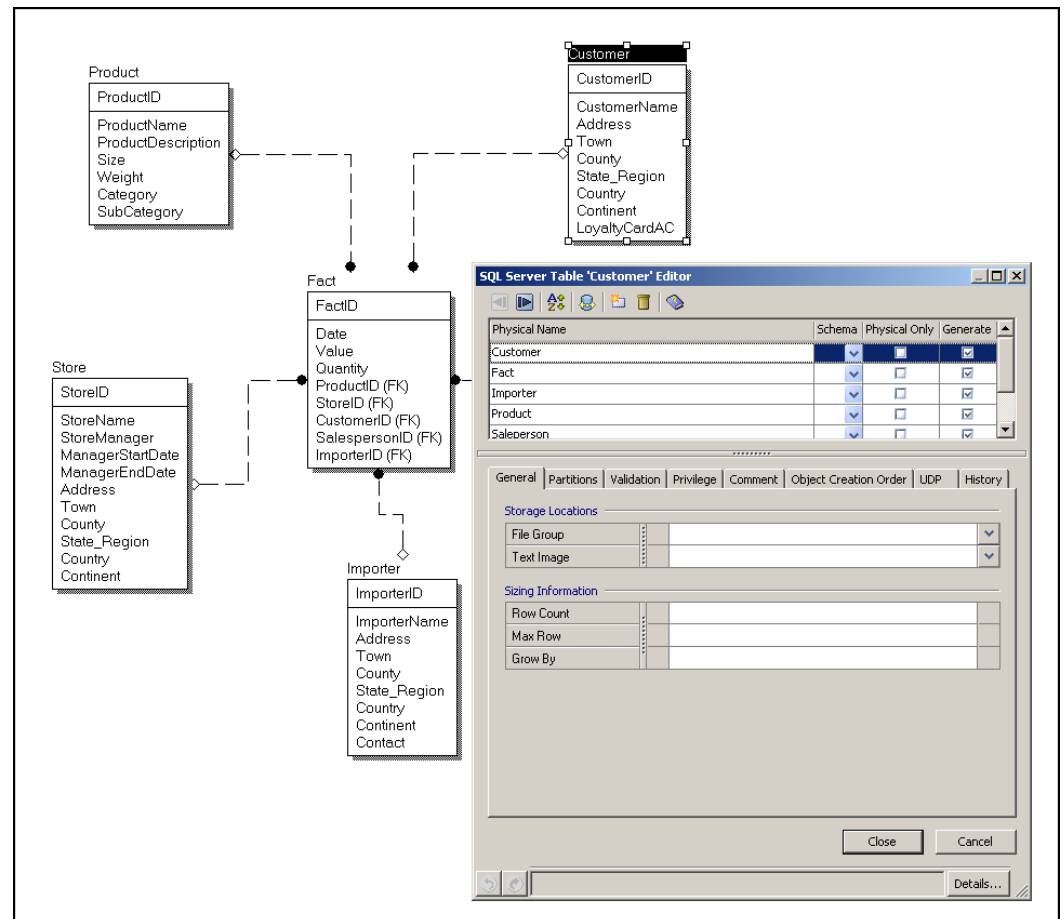
図 2: エンティティ定義



エンティティ定義が追加された論理モデル

さらに、ビジネス レポートの作成に役立つモデルも作成する必要があります。それというのも、先に説明したとおり、ディメンションは通常、フラットな構造ではなく階層構造だからです。集計はディメンションのすべてのレベルで実行されるため、ディメンションを標準的なトランザクション テーブルとして扱うことはできません。同様に、ファクト テーブルも標準的なトランザクション テーブルとして扱うことはできません。ファクト テーブルのメジャーとなる属性はどれかを知り、また、各メジャーが加算的(additive)、準加算的(semi-additive)、または非加算的(non-additive)であるのかを理解する必要があります。図 3 に示すのは、典型的な非多次元モデルにおける顧客テーブルのプロパティです。

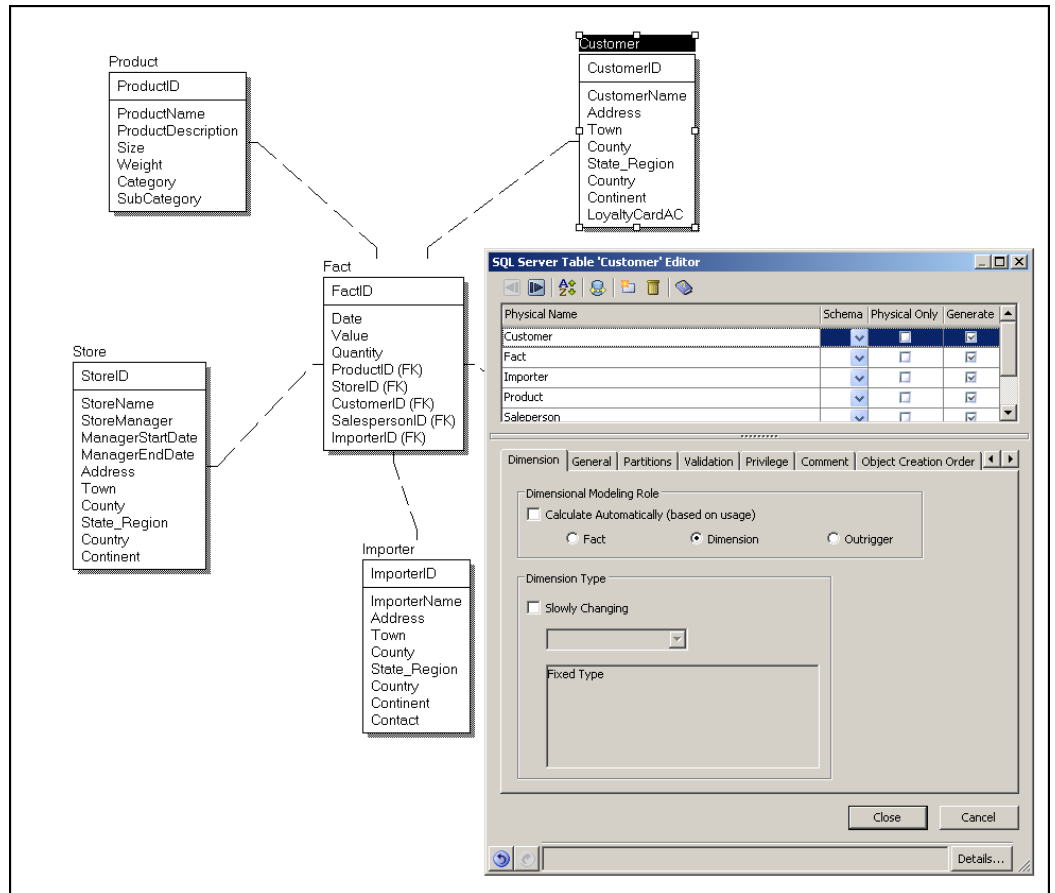
図 3: 非多次元モデル



ファクト テーブルの属性が表示された非多次元モデル

図 4 も同じテーブルを表していますが、多次元モデリングが使用されており、テーブルのモデリング ロールをディメンション テーブルとして定義できます。

図 4: 多次元モデル



ファクト テーブルの属性が表示された多次元モデル

OLAP システムから出力された構造は、ソースとなるデータ ウェアハウスとは大きく異なっているため、OLAP データを使用する開発者にとって、データ ウェアハウスの物理設計はほとんど役に立ちません。そのため、OLAP 構造へマッピングされる論理モデルを作成することが不可欠です。

OLAP キューブの結果を使用する開発者は、データ ウェアハウスの物理モデルや論理モデルを理解できません。その構造が大きく異なっているためです。そこで、OLAP キューブの結果を表す論理モデルを作成する必要があります。たとえば、データ ウェアハウスには売上および売上日の値が含まれていますが、週次売上の概念はありません。しかし、OLAP データを使用する開発者にとっては、そのような情報が重要になります。図 5 に示すのは、OLAP 出力を表す論理モデルの例です。このホワイト ペーパーの他の図と同じシステムに基づいています。

図 5 から分かるのは、OLAP 出力の論理モデルは、データ ウェアハウスの論理モデルや物理モデルとは大きく異なるということです。さらに、OLAP キューブの構造を理解するには論理データ モデリングが不可欠であることが分かります。

正しいモデルを作成して詳細情報を記述すれば、会社は情報を誤って解釈することなく、正確で有益な情報を得ることができます。はじめにビジネス側がレポートの要件を定義し、その要件に基づいてデータ モデリングのチームが各エンティティを定義して、ビジネス ユーザーがそれを確認するという手順を踏むことで、情報が誤って解釈されることはなくなります。データ モデリングのチームは、ビジネスの要件を満たし、ビジネスの基本構造が反映された論理モデルを作成することができます。このようなプロセスによって、曖昧さや混乱を避けることができ、ユーザーに正確な情報を届けることができます。

セクション 4

レポート作成を支援する

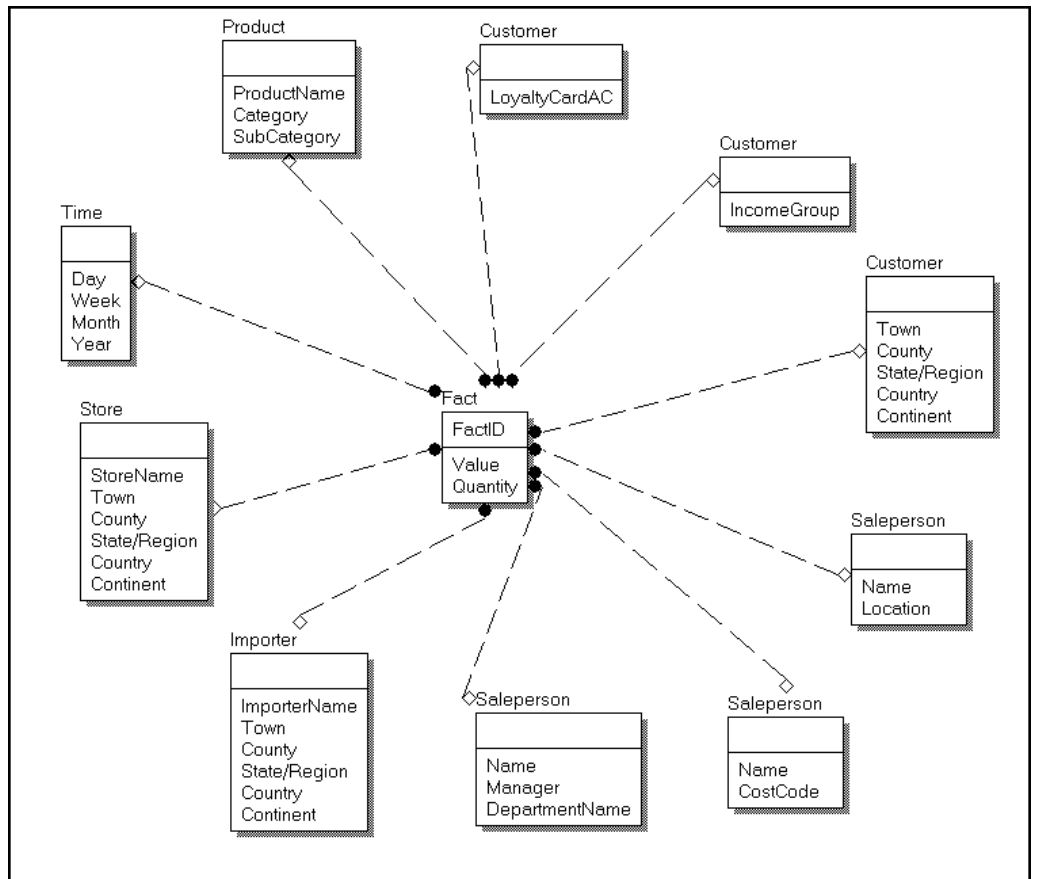
先に説明したとおり、例として挙げた食料雑貨の国際企業では、ビジネス ユーザーはピボット テーブルを利用してビジネスの業績を分析できます。このニーズには応えることができていても、一部のビジネス ユーザーは、他にどのような情報が利用できるかを知りたいと思っています。彼らには企業のエンタープライズ データの論理モデルが提供されたものの、モデルの構造がどのようにピボット テーブルやその他のレポートと関連しているかを理解することができません。そこで、モデルをピボット テーブルを作成するアプリケーション開発者に渡しましたが、開発者が必要とするモデルではありませんでした。

OLAP の論理モデルだけでなく、レポート入出力の論理モデルを作成することもまた重要です。レポートで使用されるデータは、OLAP やデータ マイニング クエリーの出力であるため、その論理モデルのエンティティは、リレーショナル ストレージ設計とは異なります。前のページで、OLAP システムやデータ マイニングの出力を表す論理モデル作成について説明しましたが、レポート作成システムでも、この論理モデルのデータが使用されます。

たとえば、月間売上高はデータ ウェアハウスには保存されていません。その属性は、データ ウェアハウスの論理モデルにも物理モデルにも存在しません。

図 5 に OLAP 出力の論理モデルを示します。このホワイト ペーパーの他の図と同じシステムに基づいていますが、ディメンションは各階層に分割されて表示されています。たとえば、顧客ディメンション (Customer) には、ポイントカードのアカウント (LoyaltyCardAC) を表す階層があります。ポイントカードは家族で共有されるためです。所得層 (IncomeGroup) や居住地域 (Town、County...) を表す階層もありますが、個々の顧客についての記録はありません。OLAP キューブでは、そのレベルの詳細情報が要求されることはほとんどないためです。他にも、この論理モデルでは、OLAP キューブでは使用されない多くの属性が削除されていることが分かります。

図 5: OLAP 出力モデル



OLAP 出力モデル

レポートを設計する際には、図 5 に示すような論理モデルが役立ちます。レポートで使用できる情報が一目で分かるためです。また、アプリケーション開発者がピボット テーブルを作成したり、ビジネス ユーザーがピボット テーブルで利用できる情報を確認したい場合にも、このモデルの構造が役立ちます。

セクション 5

まとめ

ビジネス インテリジェンスには非常に大きな利点がありますが、同時に大きなリスクを伴います。ビジネス ユーザーは、レポートに記載された情報を事実として受け取り、それらの事実は経営上の意思決定を左右します。そのため、情報が不正確な場合は誤った決定がなされる可能性があります。BI システムを注意深くモデル化すれば、この問題を避けることができます。さらに、開発者とビジネス ユーザーの双方にとって、情報へのアクセスが容易になるという利点もあります。データ モデリングは、情報の正確さと可用性を高め、開発時間を短縮することができ、BI 開発において重要な役割を果たします。

