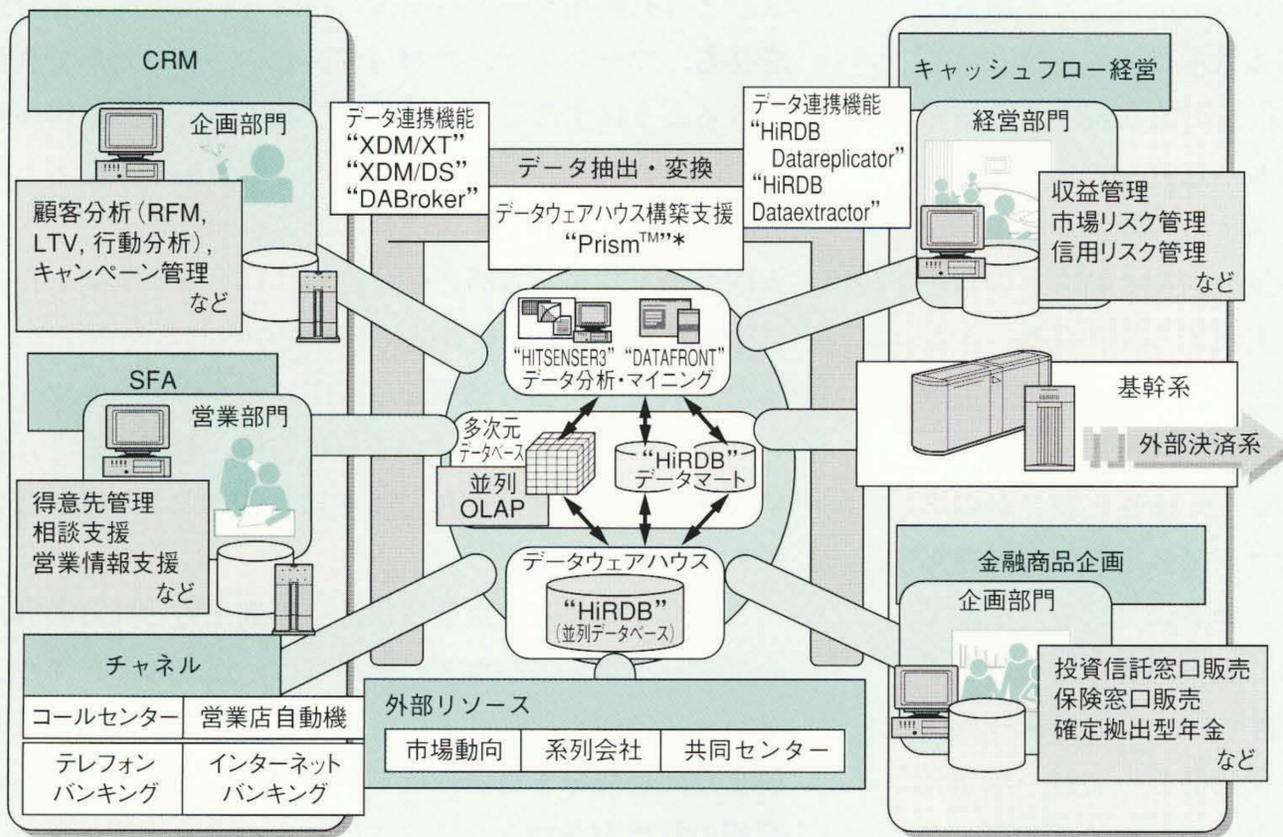


新しい情報システムのバックエンドを支える 大規模データウェアハウスの実現

Enterprise Data Warehouse Technologies for New Information Systems

亀城嘉人 Yoshito Kamegi 高橋政美 Masami Takahashi
加納直紀 Naoki Kanô 山中一朋 Kazutomo Yamanaka



注：略語説明ほか
CRM (Customer Relationship Management)
RFM (Recency, Frequency, Monetary Value)
LTV (Lifetime Value)
SFA (Sales Force Automation)
OLAP (Online Analytical Processing)
*Prismは、米国Prism Solutions, Inc.の米国での商標である。

顧客情報を活用するデータウェアハウス
大競争時代に対応できる企業の戦略的情報システムとして、データウェアハウスシステムの位置づけはますます重要になる。

急激な市場環境変化の下で、各企業は情報の活用の位置づけを強めており、企業内外にまたがる顧客や営業情報を統合的に収集し活用する「企業データウェアハウス」構築の気運が高まっている。この統合化された企業データウェアハウスは、広範囲なデータ収集とその活用により、精度の高い需要予測と的確な顧客セグメンテーション、ビジネスリスクの迅速な把握と顧客中心型の先手経営の実現を目指すものである。この構築のためには、(1) 多様なデータソースからの効率的データ収集と所在の管理、(2) 企業経営に活用できるデータ品質の維持、および(3) 収集したデータ全域にわたる詳細レベルでの横断的な分析が必要であり、これにこたえられる新しい情報活用技術が必要となる。

日立製作所は、このような企業をまたがる広範囲な情報活用を図るデータウェアハウスを「エクストラ データウェアハウス」と位置づけ、「日立データウェアハウス ソリューション」として、構築と分析ソリューションの提案を展開している。さらに、データの分析結果や営業情報などを統合的に蓄積し活用する、ドキュメントウェアハウスやナレッジマネジメントの分野にも発展できる基盤技術を提案している。

1 はじめに

今日の企業を取り巻く経営環境では、規制緩和に伴う金融ビッグバンなどによる大競争時代の中で、合併や買収、組織のリストラクチャリング(再構築)、スキルのある労働者の不足、流動的な労働力などが課題となっている。この中で、市場でのトップの座を得て、顧客満足のために質の高い商品とサービスを提供し、コストを削減して市場のトップの座を得るためには、最新の経営環境

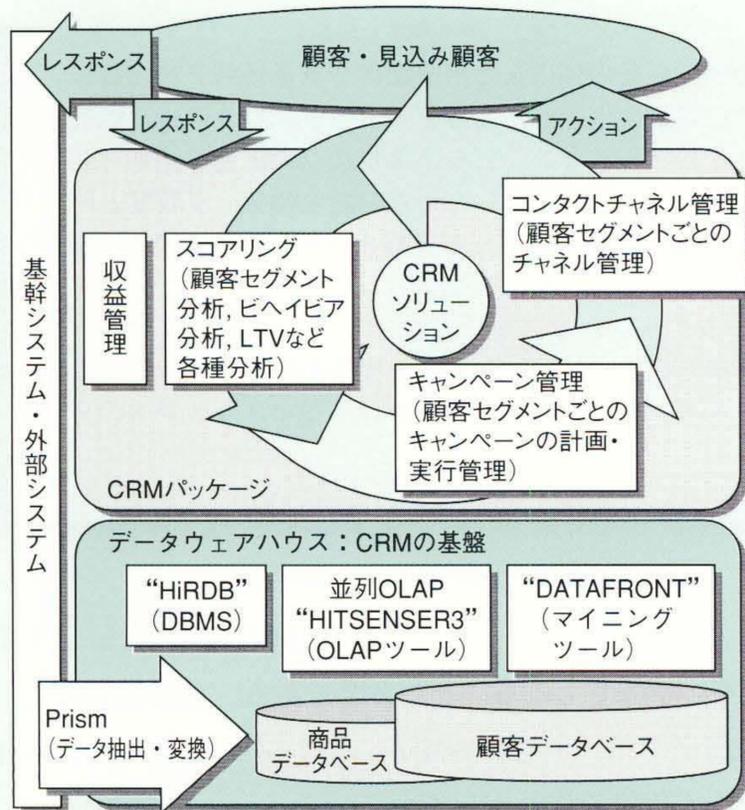
の情報をフィードバックしながら目標と戦略を策定するプロセスを作り上げることが必要である。金融機関でも、顧客満足を第一とし、みずからのコスト削減と高収益化のための仕組み作りが急がれている。

ここでは、顧客情報を中心として、フロントオフィスとバックオフィスのチャネルからの情報収集とフィードバックを主体とした企業情報システムを構築する必要性、このような情報システムを実現する情報技術、および今後の製品開発の方向性について述べる。

2 戦略的企業情報活動を作る 「エクストラ データウェアハウス」

金融機関ではすでに顧客情報を格納するデータベースを支店ごとに運用しているが、他支店で取り引きのあった顧客の属性や履歴などを確認する方法は十分整備されておらず、店舗単位での顧客情報管理に限界を感じはじめていられる。これは、従来の業務視点での情報システムでは実現できず、顧客視点に立ったチャネルでサービスを提供する、新しいシステムの整備が必要になってきている。つまり、顧客満足度向上と顧客との緊密なりレーションシップの構築を実現する経営戦略、すなわちCRM(Customer Relationship Management)の実践が求められている。

CRMは、取引内容や属性に応じて顧客を幾つかのタイプに分けて優良顧客に適切なサービスを提供し、さらに優良顧客以外の層に対しても、顧客に見合った手段で商品やサービスを提供しながら、優良顧客に近づける手法である。この手法では、「課題発見→問題提起→仮説生成→データ収集→仮説検証→行動計画→実践→課題発見」というマーケティングプロセスの継続的繰り返しが



注：略語説明

LTV (Lifetime Value)

DBMS (Database Management System)

図1 CRMの基盤となるデータウェアハウス

CRMでの繰り返し型マーケティングプロセスは、データウェアハウス技術で支えられる。

行われる。このため、顧客の取引履歴や契約内容、問い合わせやクレームの内容を顧客ごとに収集し、一元管理できる顧客データベースを中心とした「顧客中心型データウェアハウス」の構築が必要になる(図1参照)。

このような顧客データベースでは、「分析用データベース」と「行動用データベース」の両方が論理的に整合性を持ち、マーケティングサイクルにのっとった形で利用できるようにすることが重要になる。例えば、顧客属性(プロフィール)の重複や不整合などが発生し、顧客への二重アクセスやセールス機会のロスが起きたり、システム開発でもシステム間の受け渡しプログラム開発と保守に追われたりして、投資的にもむだが生じるからである。また、収集した顧客情報をOLAPツールやデータマイニングツールで分析しやすいように蓄積、管理すると、特に大量データが蓄積された場合は、OLAPツールなどの強みが出せる。

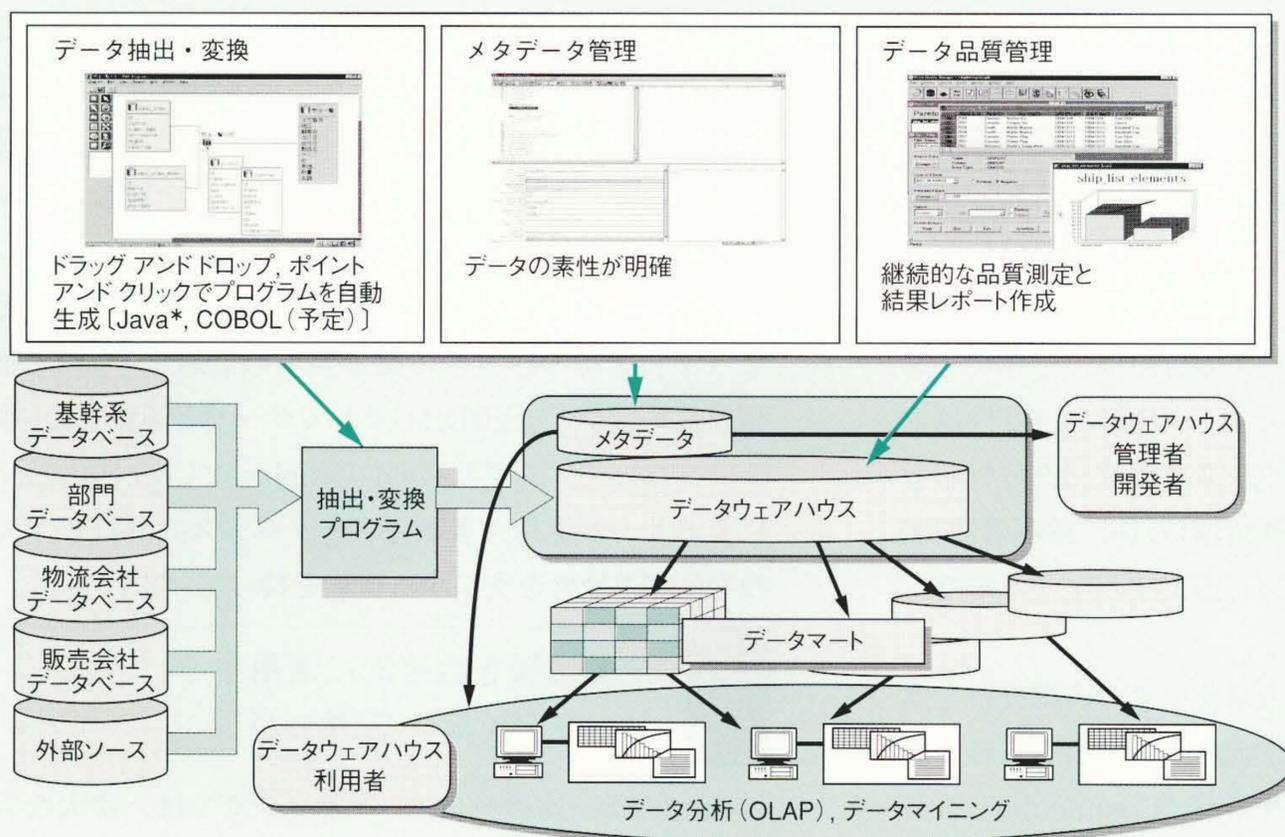
このような戦略的企業情報活動の推進力となる顧客中心型データウェアハウスの構築では、従来のデータウェアハウスに加え、これまで蓄積されていなかった全社規模または企業をまたがる形での情報や、企業外部からの情報の収集が必要となる。このような形態のデータウェアハウスを「エクストラ データウェアハウス」と呼び、ここに蓄積したデータを基にユーザーごとに的確な情報提供を行うことにより、顧客ひとりひとりに適切な商品をタイムリーに提案するなど、顧客満足を第一とした営業活動へのフィードバックができる。

エクストラ データウェアハウスの構築では、(1) 多岐にわたる情報収集の経路の構築と経路やデータの管理、(2) 増大するデータの収集を効率よく行えるツール、(3) 収集した広範囲なデータを多角的に分析に生かせる分析ツール、(4) 大規模データの蓄積や管理に加え、分析を支援する文書、その他の非定型データを一元管理できるデータベースがシステムのポイントとなる。

エクストラ データウェアハウスの構築に求められる技術要件について以下に述べる。

3 大規模データウェアハウス構築の実際

エクストラ データウェアハウスの構築では、収集するデータを会社活動全域や企業外部からの広範囲なものとするとともに、単に営業末端からの生データの蓄積だけでなく、それを顧客情報(プロフィール)として整理、統合することも必要である。この過程では、取り扱うデータや、要求の規模の拡大と多様化により、次のような課



注：* JavaおよびすべてのJava関連の商標およびロゴは、米国およびその他の国における米国Sun Microsystems, Inc.の商標または登録商標である。

図2 構築支援ツールによるデータウェアハウス構築の流れ

データウェアハウス構築支援ツール“Prism”では、データ抽出・変換プログラムの自動作成、メタデータ管理による高生産性、データ品質管理によるデータウェアハウスの品質の作り込みを提供し、スピーディなデータウェアハウス構築を支援する。

題が発生する。

- (1) 多様なデータ収集のために必要な多数のプログラム開発
→高生産性で低コストな開発手段が必要
- (2) 分析ニーズの変化に伴う頻繁なプログラム修正と、
収集したデータの整合性の確保、効率的な利用
→見通しのよいデータ利用を図る仕掛けが必要
- (3) 企業経営に直結するデータの品質の確保
→データの品質を測定できる仕掛けが必要

日立製作所は、広範囲なデータ収集を行うエクストラデータウェアハウス構築のために、データウェアハウス構築支援ツール“Prism”を提供している(図2参照)。Prismのような支援ツールを適用することで、データウェアハウス構築で全体の70%とも言われるデータ抽出、変換の仕掛け作りの生産性を高めることができ、データウェアハウス構築全体の開発効率を向上させ、短期開発とプロトタイピングや上流・下流の設計への工数注入が容易にできる。さらに、ツールによって作業を標準化できることから、開発プロセスの簡素化と保守性の向上、開発・保守要員の確保を容易にすることができる。

データウェアハウスを企業経営に生かすためには、それにこたえる情報量を確保し、情報の精度を磨く必要がある。この過程は、繰り返し型の開発プロセスとして実施することが有効で、これによって短期に効果を出しつつデータウェアハウスを成長させていくことができる。一般に、繰り返し型開発では、繰り返しのフェーズを3, 4か

月とするのが効率が良いとされている。このプロセスを効率よく推進するためにも、ツールの支援は有効である。

また、データウェアハウスの規模が大きくなると、構築、維持の過程で、データの所在を明らかにし、利用者の各階層でこの情報の利用が図れる情報や仕掛けの提供が望まれる。このようなデータの情報を示すデータを「メタデータ」と言い、対象データ範囲や利用規模の大きなデータウェアハウスでは、メタデータの整理と利用が構築と維持の大きなポイントとなる。

データウェアハウス構築のもう一つの重要なポイントとして、データ品質があげられる。データ品質では、収集したデータが分析に十分な品質と精度を持っているかどうか問われ、データウェアハウスの重要な課題として最近注目が集まっている。データの品質の低下は、信頼できない情報による意思決定と、それによる顧客サービスの低下を招き、信用されない収集データを修正するための多大なコストの発生など、企業経営に与える影響は無視できないものがある。データ品質の確保には、品質の測定と維持のプロセスを作り上げることが重要であり、ここでもツールの利用が有効である。

4 顧客全域の情報をセグメントする OLAPサーバ

現在、金融分野で構築が進められている顧客データベースでは、顧客をセグメントに分け、それぞれの顧客に

ついでに情報を分析し、適切なタイミングでそれぞれの顧客の関心事に対応した情報を提供することによって顧客満足度の向上を実現することが期待されている。このためには、支店営業担当ごとに集約されていた従来のデータのレベルでの分析では顧客特性を抽出するには精度が不十分であり、個人単位の取引履歴や属性の分析が求められる。

このような詳細レベルのデータ分析では、母体データの件数と分析軸が多く、それぞれの項目数は1万件以上も考えられる。多次元データベースでのデータ分析を考えると、対象顧客10万人、分析指標1万件、顧客属性300バイトとして、総データ量は、運用開始時点ですでに300 Gバイト以上が必要となる。

日立製作所は、このような大量データの詳細分析を高速に実行可能にする並列型多次元OLAP技術を世界に先駆けて実用化し、並列OLAPサーバ“Cosmicube”として提案している(図3参照)。

Cosmicubeは、スケーラビリティに優れた“Shared Nothing(シェアードナッシング)”型並列サーバ構造と合わせ、並列化された多次元データベースを分析軸の空間の対角面に沿って複数のディスクに配置し、多次元分析操作の負荷の均等化を図る、独自の「対角分割格納方式」により、チューニングなしでスケーラブルな応答性能が得られるという特徴を持つ。これにより、従来のOLAPツールでは実現できなかったテラバイト級のデー

タが収集でき、刻々と増え続ける顧客データにもスケーラブルに対応できる。また、扱える分析軸の数の上限が高くとれるので、多次元データベースの設計の自由度を高くできる。これにより、全社レベルにまたがって収集した、顧客データのような大量データの全域をセグメントするような分析に効果が期待できる。

またCosmicubeは、大量データのデータ集約の時間とデータ利用時の応答性の配分要求の最適化を図る「並列事前集約機能」、HiRDBパラレルサーバと連動した「並列データローディング」、GUI(Graphical User Interface)によるサーバ運用と多次元データベース設計など、大規模なOLAP分析を支援する機能を持っている。

5 企業情報を戦略的に活用する データマイニングサーバ

CRMやデータベースマーケティングでは、膨大なデータをさまざまな方向から分析してみる必要がある。特にマイクロマーケティングになると、分析する軸の数は、従来のマスマーケティングよりも飛躍的に拡大する。データマイニングは、さまざまな軸のデータを整理し、その中から有効な情報を取り出す技術である。OLAPツールでもデータの傾向を見て知識を獲得することができるが、軸の数があまりに多いと、何を見ればよいかわからなくなる。データマイニングはOLAPに必要となる有意な軸の候補を示唆してくれるので、データマイニングと

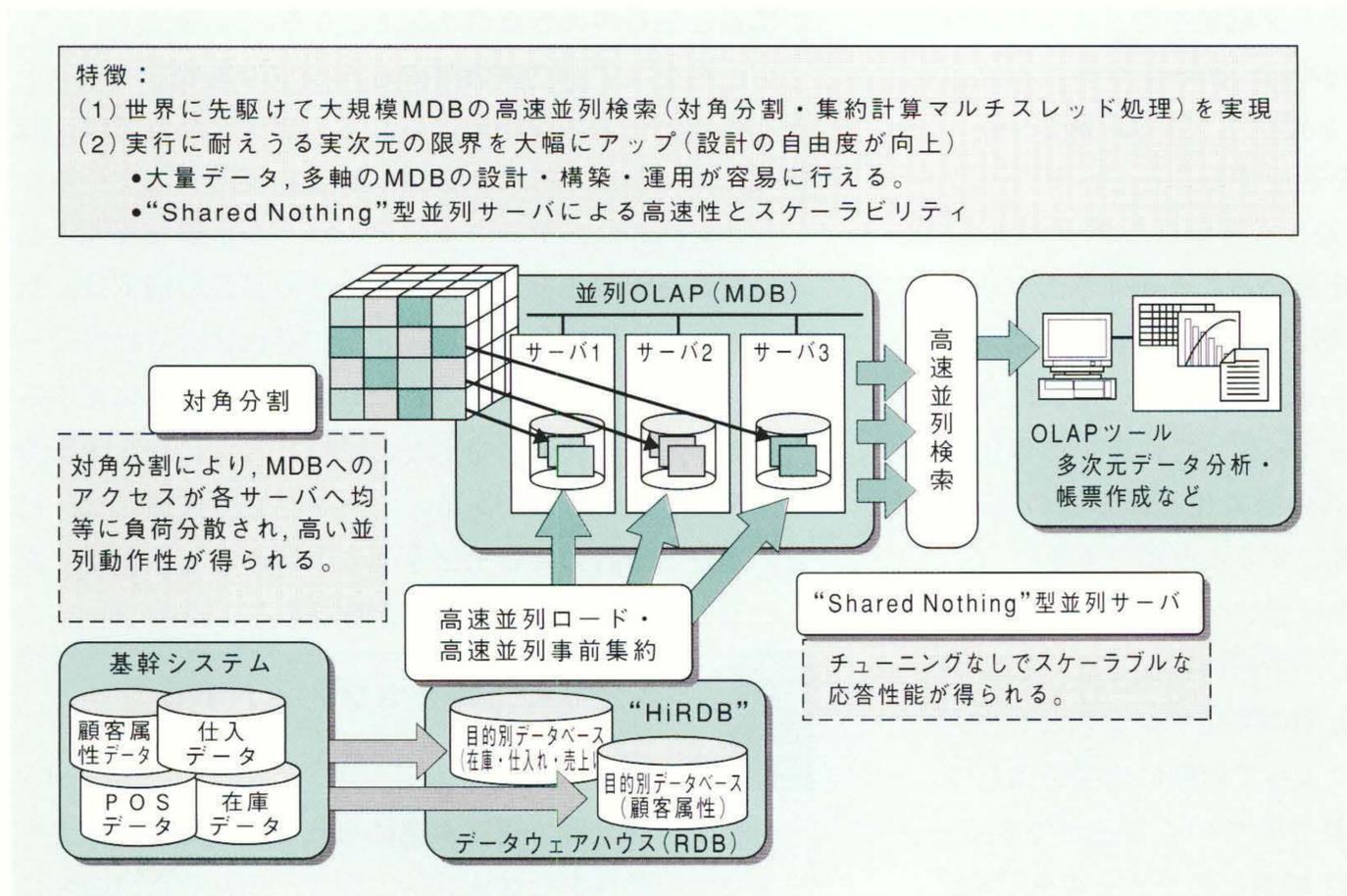


図3 並列OLAPサーバ“Cosmicube”の特徴と機能
 Cosmicubeは、独自の並列OLAP技術により、従来のOLAPツールではなしえなかった大量のデータの多次元分析を可能にする。

OLAPを組み合わせることによって適切な分析軸を選択することができ、「仮説—検証」のデータ分析プロセスを構成できる。

データベース マーケティング システムを実現するには、データマイニングをデータウェアハウスと密に連携し、データウェアハウスに蓄積された大量な情報(データ)から有用な知識を取り出し、“Plan-Do-See”の各場面で有効に活用しなければならない。日立製作所は、このような大規模データウェアハウスと連携し、100万件を超えるような大量データのマイニングを実現するデータマイニングサーバ“DATAFRONT/Server”を提案している(図4参照)。

DATAFRONT/Serverは、(1) データを分析してセグメントする「特徴ルール生成機能」と、(2) 蓄積したデータの傾向から新しいデータを予測するための「メモリベース推論機能」を持っている。

特徴ルール生成機能では、CHRIS(Characteristic Rule Induction by Subspace Search)というアルゴリズムを用い、複数の軸をカテゴリーで区切り、 n 次元空間のセグメントの中から結果に対する特徴的なセグメントを見つけ出す。例えば、「最近、預金金額の減少した顧客は、どういう顧客か。」を見つける場合、さまざまな要素の中から、結果として「年齢区分がXで、公共料金の引き落としのある顧客だ。」とか、「〇〇クレジットの引き落としが増えていて、電話料金の引き落としも増え

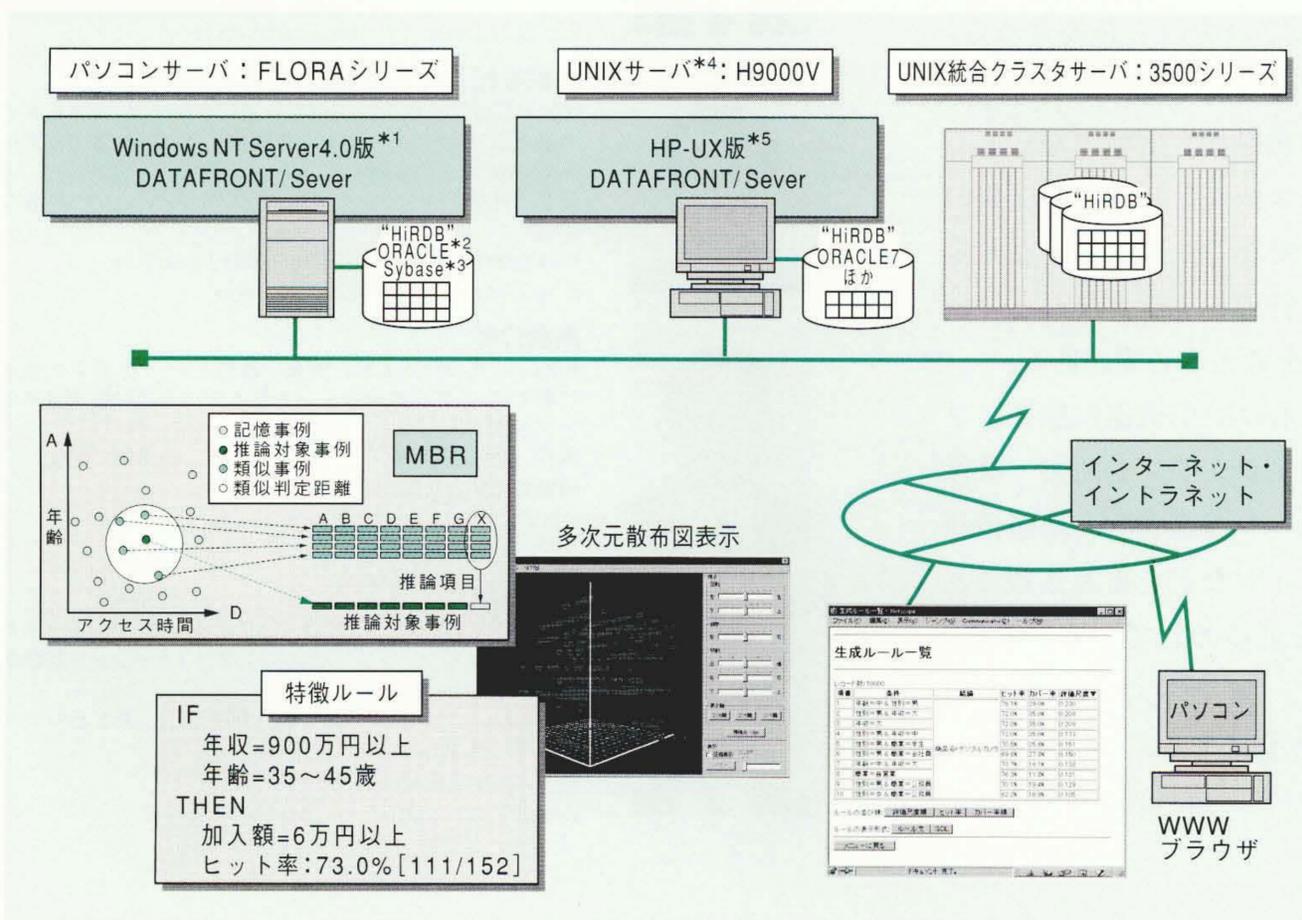
ている顧客だ。」などという複数のことが同時にわかる。

一方、MBR(Memory-Based Reasoning: メモリベース推論)では、大量の蓄積されたデータを知識のベースとして、予測したい顧客に近い顧客の情報を集めてきて、予測値を算出する。例えば、「ある生命保険の顧客Xさんは、過去のAさん、Bさん、Cさんに属性、契約金額が似ている。AさんとBさんは、その後、学資保険も追加契約をしてくれたから、Xさんも $a\%$ の確率で追加契約をしてくれるだろう。」ということがわかる。これにより、営業部門は、顧客を個別訪問したり、ダイレクトメールを発送するなどのアクションにつなげられる。この手法は、保険料率や解約率といったリスク管理、各種スコアリングなどに幅広く利用できる。

また、DATAFRONT/Serverでは、データをビジュアルに見るための多次元散布図表示、分散オブジェクトとしてのアプリケーションへの組み込み、データマイニングの結果のWebページとしての配信など、データ分析を有効に利用できるトータルな利用環境を提案している。

6 リレーショナルデータとマルチメディアデータを融合する、新しい情報ウェアハウス

顧客チャネルを利用して顧客データベースに蓄積した情報は、顧客満足度を上げるためにだけにしか利用できないわけではない。現在の市場の傾向を読み、企業経営をいっそう戦略的に導くための重要な生きたデータを入



- 注: 略語説明ほか
- WWW (World Wide Web)
 - *1 Windows NTは、米国およびその他の国における米国Microsoft Corp.の登録商標である。
 - *2 ORACLEは、米国Oracle Corporationの登録商標である。
 - *3 Sybase, Sybaseのロゴは、米国法人Sybase, Inc.の登録商標である。
 - *4 UNIXは、X/Open Company Limitedがライセンスしている米国ならびに他の国における登録商標である。
 - *5 HP-UXは、米国Hewlett-Packard Companyのオペレーティングシステムの名称である。

図4 データマイニングサーバ“DATAFRONT/Server”の機能 DATAFRONT/Serverでは、100万件を超える大規模な母体データに対して、スケーラブルな性能を発揮できるアーキテクチャとともに、多次元散布図やWebによる結果の配信など、トータルなマイニング環境を提供する。

手できるはずであり、このために、既存の業務系データベースと顧客チャネルを利用した顧客データベースの双方からうまくデータを抽出、変換できることが、今後のデータウェアハウスの利用にとっては重要となる。

日立製作所は、これまでも、従来のデータウェアハウスでの要件に対して、スケーラブル データベース“HiRDB”を提案しており、パラレルサーバによる大規模データベースへの対応、Prismの導入によるセントラルウェアハウスの構築、“HITSENER DataMart Server”によるマートサーバ構築や“HiRDB Datareplicator”によるレプリケーションによるデータの連携機能などにより、データウェアハウス向けRDBMS (Relational Database Management System)として機能を強化してきた。

今後の戦略的情報システムでは、基幹データに限らず企業に集まるすべてのデータを統合し、さまざまな角度から分析することによって市場や企業の状況を読むことが求められ、多様なデータを管理できるデータベースを基盤とすることが重要になる。

例えば、コールセンターや営業活動も、顧客や消費者の生の声を聞ける現場であるが、この生の声は、データベース上のデータとして整理するには向かない、非定型データである。この非定型データを蓄積し、マーケティングや顧客満足度向上の材料にできることが望まれる。また、営業日報や市場分析レポートなどの情報もナレッジ(知識)として企業活動の大きな資産であり、この蓄積と、営業などフロントオフィスでの活用も必要となる。さらに、エリアマーケティングなどで、地図情報を活用した分析や情報整理、さらに情報提供も重要になってくる。今後のデータ活用を図るシステムとしては、従来の数値データのハンドリングだけでなく、このような非定型データを扱えるドキュメント管理システムなどとの統合(ドキュメントウェアハウス)を考える必要がある。

HiRDB Version 5.0では、従来のRDB機能に加え、プラグイン機構により、全文検索プラグインや画像・映像ファイルを管理する“File Link”プラグイン、地図情報を管理する空間情報管理プラグインなどを組み込むことにより、非定型データの管理を可能としている。また、非定型データの整理には、複合文書管理基盤“Document Broker”との組合せにより、情報の整理、活用が強化できる。さらに、データの収集と提供経路で

もインターネットやイントラネット基盤が重視されており、HiRDBでも、“DABroker”や“Web Page Generator”などの組合せにより、Web環境からのアクセスを実現している。

7 おわりに

ここでは、これからの企業情報システムとしての「エクストラ データウェアハウス」と、その要求の背景および技術的要件について述べた。

今後の競争に対応するには、企業戦略として、顧客情報を直接とらえる「顧客中心型経営」の実践が急務となっている。「日立データウェアハウス ソリューション」では、今後も、顧客中心型データウェアハウスの構築ソリューションとその要件にこたえる技術をそろえ、積極的な提案によって幅広い活用を推進していく考えである。

参考文献

- 1) 前田：データベースマーケティングへの応用，オペレーションズ・リサーチ，1998年12月号

執筆者紹介



亀城嘉人

1983年日立製作所入社，情報・通信グループ ソフトウェア事業部 アプリケーション基盤本部 DB設計部 第1グループ 所属
現在，スケーラブルデータベース“HiRDB”の拡販とテクニカルサポート業務に従事
情報処理学会会員
E-mail：kamegi_y@soft.hitachi.co.jp



加納直紀

1984年日立製作所入社，情報・通信グループ ソフトウェア事業部 アプリケーション基盤本部 DB設計部 第7グループ 所属
現在，日立データウェアハウス ソリューションの推進に従事
情報処理学会会員，ACM SIGMOD Japan会員
E-mail：kano_n@soft.hitachi.co.jp



高橋政美

1972年日立製作所入社，情報・通信グループ ソフトウェア事業部 アプリケーション基盤本部 DB設計部 第4グループ 所属
現在，並列OLAPサーバ“Cosmicube”の開発に従事
情報処理学会会員
E-mail：takah_ma@gm.soft.hitachi.co.jp



山中一朋

1980年日立製作所入社，情報・通信グループ ソフトウェア事業部 アプリケーション基盤本部 アプリケーション基盤設計部 第4グループ 所属
現在，データマイニング，統計分析，OLAPなどのデータ分析関連製品の開発に従事
情報処理学会会員，人工知能学会会員
E-mail：yamanaka@soft.hitachi.co.jp