

データ活用のための情報プラットフォームの実現 — 情報システムを支えるインフォメーション・マネジメント —

IT 技術の普及・浸透により日々さまざまなデータが大量に生み出されている中、企業にとっていかにデータを活用できるかがビジネス成功の鍵となっています。これまで多くの企業が基幹業務からのデータを元に情報系システムを構築してきました。しかしながら、業務変革、新規ビジネス参入、提携・合併などさまざまな経営環境の変化に伴い、データの範囲・種類も多様化・複雑化・大量化しています。一方、経営のニーズに柔軟かつ迅速に対応することがこれまで以上に求められており、そのためにはデータの発生から収集・蓄積を経て分析までつながるように情報システムが整備されていることが必要となります。Business Analytics and Optimization (BAO) では企業が保有するデータ、取り扱うデータを情報活用に結び付けていくための基盤、情報プラットフォームとして Enterprise Information Management (EIM) を定義し、関連するソリューションを提供しています。ここではデータ活用の情報システムの構成要素について EIM ソリューションの基本領域を通じて解説します。

① BAO ビジネスの基盤とは

BAO がターゲットとする市場を見たとき、ビジネスの大きな部分が EIM ソリューションの基本的な領域で占められています。ここでいう EIM ソリューションの基本

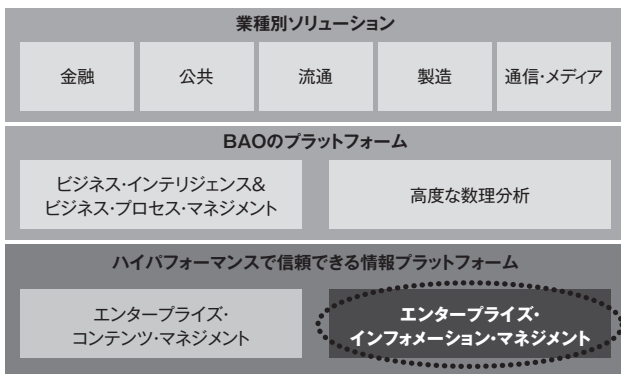


図1. BAOで提供されるビジネス領域

Achievement of an Information Platform for Data Utilization - Information Management Supporting Information Systems -

A large and varied amount of data is generated every day due to the spread of IT technology. In this situation, how well enterprises can use data has become the key to a successful business. In the past, many enterprises have build information systems based on data used in the line of business. However, in conjunction with various changes in management environment due things like business reform, entrance into new business, alliances and mergers, the range and type of data has also diversified, become complicated, and expanded. On the other hand, enterprises are expected to respond to the needs of management more flexibly and promptly than ever. Meeting these needs requires that information systems are in place that can handle data from its generation through to its collection, accumulation, and analysis. BAO defines the infrastructure or information platform that enables the utilization of the data held and handled by enterprises as Enterprise Information Management (EIM), and offers associated solutions. Here, we cover the components of the information systems used for the utilization of data in line with the basics of the EIM solution.

的な領域とは、図1の右下、点線で囲まれている部分で、Data Definition、Database Design、Data Integrity といった IT システムの基本となる領域から、データ・ウェアハウス構築やマスター・データ管理導入、データ・クレンジングの実現という領域を指しています。

Business Intelligence (BI) や Predictive Forecasting といった高度な情報活用を行うためにも、まずはそれらの分析で使用するための情報を整備することが重要であり、お客様が最初に取り組む領域となるのです。

② 適切なデータを適切なタイミングで

なぜ、EIM ソリューションの基本的な領域が重要なのでしょうか？ 例えば、BI ツールを使用して経営情報を分析することを考えてみましょう。

BI ツールで分析を行うためには、分析対象として整

合性の取れた「データ」を、適切なタイミングで、使用可能な状態にする必要があります。この状態にするためには、データソースからのデータの抽出、データの統合、加工、クレンジング、ターゲット・データベースへのデータの格納という作業を、データ・リポジトリの配置に応じて数サイクル回す必要があるのです。

また、最終的に届けられる「データ」の鮮度は非常に重要な要素です。リアルタイムで情報を分析できるのか、3日前のデータでしか分析できないのかでは、経営判断にも大きな影響を与える可能性があります。このため、データソースからターゲット・データベースに格納するまでのパフォーマンスを考慮した設計が重要であり、実装においてはパフォーマンス・チューニングのスキルが必要になるのです。

データソースからBIツールまでの情報の流れ（データ・フロー）において、EIMソリューションは「Data Integration」「Data Quality」「Data Architecture」「Master Data Management」の4つのカテゴリーに分類できます（図2）。ここでは、各カテゴリーの実現を支える技術領域の詳細について説明していきます。

③ Data Quality - データ・ガバナンス&データ品質管理

分析対象とするデータソースは、基幹システムのデータベースから、個々のユーザーが扱うExcel®表や紙の明細など、環境によってさまざまです。必要な項目が埋まっていなかったり、まったく異なるカテゴリーのデータが入っていたり、同じデータが違うキーで扱われていたり、ということが起こり得るのです。情報活用の第一歩として、

これらのデータを整理・洗浄して整合性が取れた状態にし、適切に扱える状態にする必要があります。

●**データ・クレンジング**：システムで想定する形にデータを整形することです。例えば、電話番号というデータ項目としては、“0312345678”や、“03(1234)5678”といった入力が行われることが考えられますが、これらのデータをすべて“03-1234-5678”に統一する、という作業を行うことを指します。複数のデータソースがあり、管理部門が異なる場合などは、まずデータ・クレンジングの実施を検討する必要があります。

●**名寄せ**：重複するデータを洗い出すという意味ではデータ・クレンジングと同じですが、名前や住所などのデータを標準化し、類似データを比較して、同一データかどうかを判断する、という作業のことを指します。名寄せをすべてシステム化することは難しく、自動名寄せと手動名寄せの組み合わせの仕組みを検討します。また、手動名寄せを実現するためには、それを実施するためのお客様組織の検討を並行して進める必要があります。

●**データ・マスキング**：データソースに格納されているデータについて、セキュリティの観点からそのままの形でデータ・ウェアハウスに格納したくない、という要件があるかもしれません。このような場合には、適切な形でデータを加工し、セキュリティを確保しつつ、分析対象のデータだけを抽出する、という作業が必要になります。単純なアルゴリズムではセキュリティが低くなってしまいますし、複雑な暗号化を実施しようとすると、大量データを処理する際にパフォーマンスが懸念される、というトレードオフが発生します。

●**データ・モデリング**：データの整合性を保つために、データを構造化し、組織化します。ビジネス要件に従って、ステージング・エリア、ODS（オペレーショナル・データ・ストア）、データ・ウェアハウス、データ・マートなどのデータ・リポジトリを定義し、実装します。特にデータ・ウェアハウスやデータ・マートではスター・スキーマで実装し、Queryが単純になるよう設計します。それぞれの表の特性に合わせて正規化のレベルを変更して、パフォーマンスを考慮した設計にします。

●**メタデータ管理**：メタデータを管理することで、開発時や保守時における、整合性の取れたデータの変更管理を容易にすることを目的とします。ただし、既存システムまで含めたメタデータ管理を実施するには、全社的に用語を統一したり、既存の管理ルールを変更したりする必要があり、容易ではないといわれています。

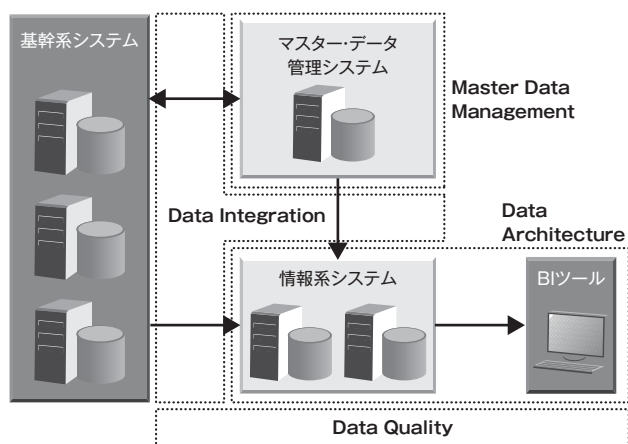


図2. EIMソリューション・カテゴリー

4 Data Integration - データの統合

基幹系システムからマスター・データ管理システム、情報系システムへの情報連携を実現するカテゴリーです。ビジネス分析の対象となる大量データをいかに高速に連携することができるか、開発コストをどれだけ下げることができるかが大きなポイントとなります。

●**データ連携**：データソースからの情報抽出については、プラットフォームへの依存度が低いこと、責任分解点を明確にできることから、ファイル転送によって実施されるケースが依然多いようです。ただし、環境によっては、フェデレーションやレプリケーションによるデータ連携手法によって、開発生産性の向上や運用保守コストの削減を実現することもできます。また、大容量データの高速度移動という観点では、ストレージ・コピー機能の採用も候補の1つとなり得ます。

●**ETL**：データベースからデータベースへ、ファイルを入力とした大容量データの取り込みなど、EIMソリューションにおいてETL(Extract/Transform/Load)を欠くことはできません。IBM InfoSphere™ DataStage®を使用することで、高い開発生産性、および高いETLパフォーマンス、そして、保守の容易性を実現することができますので、多くのプロジェクトで採用されるようになってきました。ただし、開発規模が小さい場合は、SQLやシェル・スクリプト/バッチ・プログラムで実装されるケースもあります。

●**データ統合**：マスター・データ管理システムや情報系システムへのデータ連携の中で、簡易的なデータ・クレンジングや名寄せについて、専用のツールを使うのではなく、ETLツールを使用してデータ統合を実施することも1つの方法です。一連のETLの流れの中で対応できれば、開発工数を抑え、保守を容易にできる可能性があります。

●**データ移行**：システム構築時において、大容量データの移行にかかる開発工数は無視できないほど大きくなる場合があります。これは、データ量が大きくなるほど、例外データの量・種類が大きくなりやすいため、個別に対応すると工数がどうしても大きくなってしまふからです。プロジェクト当初からきちんと開発工数を見積もっておくこと、また、ETLを使用するなどの工夫も検討すべきです。

●**データ・ウェアハウス設計・構築**：もし、分析対象のデータがデータ・ウェアハウスに存在しなければ、ODSやデータソースにさかのぼってデータを取得しなければならないため、データ統合やデータ連携の仕組みを再設計する必要が出てきます。これを避けるために、データ・ウェ

アハウスにおいては、まずは分析に必要なデータをすべて蓄積することを優先します。基本的には、分析対象のデータをファクト表として定義し、分析軸となる次元をディメンジョン表として定義するスター・スキーマの構造にします。

5 Data Architecture - 情報系システム全体のアーキテクチャー

情報系システム全体のアーキテクチャーを策定します。アーキテクチャーに基づき、実際のシステム構築・実装へつなげていきます。ステージング・エリア/ODS/データ・ウェアハウスなどのデータ・リポジトリの最適な配置や大容量データを扱うためのストレージ設計をリードします。特に、提供するデータの鮮度を考慮した、システム全体でのパフォーマンス・マネジメントが重要です。BIツールを使用して、EUC(End User Computing)でユーザーが自由にQueryを発行する場合、すべてのケースでパフォーマンス要件を満たすことは非常に難しいといわれています。これらの要件を整理し、必要なデータ・マートを適切に配置していきます。

6 Master Data Management - マスター・データ管理

これまで、ビジネス・ニーズ・業務ごとに個別に最適化されたさまざまなシステムが構築された結果、本来、一元管理されるべき顧客や商品が複数のコード体系で管理されていることが多くなっています。顧客の売上・履歴を

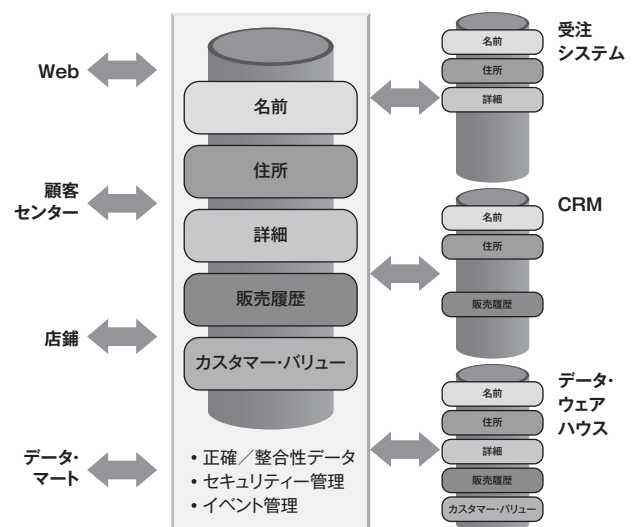


図3. マスター・データ管理

業務横断的に把握したい、企業グループ内を統一した商品コードで管理したい、といった全社的なニーズに対応するためには、マスター・データの整備・管理が不可欠となります。マスター・データ管理の重要なポイントの1つはデータ・モデルの構築です。EIM では顧客および商品のマスター・データ管理ツール (IBM InfoSphere Master Data Management Server および IBM InfoSphere Master Data Management Server for Product Information Management) と共にソリューションを提供しています。マスター・データ管理は全社的なデータ・ウェアハウスの基本構成要素であると同時に、業務・組織横断的な視点で顧客情報・商品情報を提供します。

マスター・データ管理へのアプローチのポイントには、「データ・モデル」「データ属性」「データ連携」の3つの観点があります。特に「データ属性」「データ連携」においては、現状の情報システムの課題を検討・解消してからマスター・データ管理を検討する必要があります。

●**データ・モデル**：顧客・商品の観点で、全社的な視点で情報を提供します。全社データ・モデルとして統合されたマスター・データ管理を実現することで、より正確で整合性のあるデータを提供できるようになります。また全社マスター・データ管理の実装については、結合型、登録型、共存型、ハブ型の4つの中から、既存システムとの依存性や全体的な開発工数・期間を照らし合わせ、適切なデータ・モデルの統合の仕方を選択します。

●**データ属性**：複数のコード体系を保持していることによって管理負荷が掛かっている場合や、レポート内容が不足している場合は、マスター・データ管理の前段階として、コード統一またはデータ・ウェアハウスの構築を検討します。

●**データ連携**：データ取り込みに時間がかかっている場合や、取り込みが月次バッチなどで即時性に課題がある場合は、まず、システム間インターフェースを検討することになります。マスター・データ管理の構築に当たっては、現状の情報システムの課題の把握・分析と、データ活用に向けた全社的な取り組みを合わせて計画して取り組むことが必要です。

7 まとめ

EIM ソリューションを構成する技術要素は多岐にわたりますが、重要なことは最適なコストで最適なソリューション

を組み合わせられるかどうか、という点です。情報活用の前段階の部分になりますので、それだけでは投資対効果が出にくく、経営判断として投資がしにくい部分になるかもしれません。ただ、情報活用のためには避けては通れない部分になりますし、この部分をいち早く実現したお客様こそが、情報活用による優位性を持つこととなります。そういう意味でも、将来計画を明確にし、お客様の進むべき道筋を示すことが合わせて必要になります。

単純な“情報活用”というキーワードで、フロントのBIツールだけが提供されているケースがあります。お客様自身、「BIを使いこなせていない」と認識されているケースのほとんどはこれに当たるかもしれません。

BAO がターゲットとする EIM ソリューションにおいて、個々の技術領域については、これまで基幹系システム構築でわれわれが培ってきた経験・技術が最大限に活用できる部分です。これら IBM の総合力を結集することによって、データソースからデータ・ウェアハウスを含めてトータルで本当に必要な情報を最適な形でお客様に提供することが可能になるのです。



日本アイ・ビー・エム株式会社
グローバル・ビジネス・サービス事業
コンサルティング・サービス
ビジネス・アナリティクス & オプティマイゼーション
エグゼクティブ IT アーキテクト

鈴木 至 Itaru Suzuki

【プロフィール】

2007年よりグローバル・ビジネス・サービス事業にてインフォメーション・マネージメント分野のデリバリー部門および、グローバル・コンピテンシーリーダーを担当。2009年よりBAO EIM 担当としてソリューション展開、コンピテンシー推進に従事。



日本アイ・ビー・エム株式会社
グローバル・ビジネス・サービス事業
コンサルティング・サービス
ビジネス・アナリティクス & オプティマイゼーション
アドバイザリー IT アーキテクト

大久保 将也 Masaya Ookubo

【プロフィール】

ソフトウェア事業でInformation Managementのテクニカル・セールスを経て、グローバル・ビジネス・サービス事業でITアーキテクトとしてメディア業界の開発プロジェクトに従事。特にInformation Architectureの観点で、システム全体の安定化やパフォーマンスの最適化、必要なデータの整合性を実現する設計を手掛ける。