

# 全社的な情報活用基盤としての DWH/BI 構築方法論

## Enterprise DWH / BI methodology in building

昨今、国内の企業を取り巻く環境は激変しており「変化に強く経営に有益な情報を迅速に引き出せる」全社的な情報活用基盤の構築が必要不可欠となっている。当社は、これまで国内の100社を超える幅広い業種においてBIの適用とDWH構築といった企業内に蓄積された情報の有効活用を促進するプロジェクトに参画してきた。

本論文では、国内の企業における現状を踏まえたうえで全社的な情報活用を如何に実現していけばよいのか技術的な手法や考察を交えながら導入方法論について述べる。

|       |                    |
|-------|--------------------|
| 桑島 義行 | Kuwajima Yoshiyuki |
| 星 芳彦  | Hoshi Yoshihiko    |
| 佐藤 徹  | Sato Toru          |
| 千葉 憲昭 | Chiba Noriaki      |

### 1. はじめに

昨今、我が国における企業を取り巻く環境は激変しており、米国におけるサブプライム問題を起因とする金融危機による株価の下落、世界的な不景気の煽りを受けた消費の低迷により収益の確保が困難になっている。また、業種の垣根を超えた異業種からの参入や BRICs 諸国の台頭による国際的な競争力の強化が求められており、生き残りのため、業務提携や M&A、事業の選択と集中が進められている。

これらの経済状況の下、企業は、環境の変化に柔軟、かつ、迅速に適応していく必要があり、企業活動を記録しているデータ群もあわせて整理・統合していく必要がある。また、商品のコモディティ化が進む市場の中で自社の優位性を見出し、コア・コンピタンスとして他社に対する優位性を確保するためには、情報の有効活用が必要であり、そのためにも「変化に強く経営に有益な情報を迅速に引き出せる」全社的な情報活用基盤の構築が必要不可欠となっている。

当社は、これまで国内の製造業・流通業・金融業・官公庁といった 100 社を超える幅広い業種において BI (Business Intelligence) の適用と DWH(Data Warehouse)構築といった企業内に蓄積された情報の有効活用を促進するプロジェクトに参画してきた。本論文では、国内企業における現状を踏まえたうえで全社的な情報活用を如何に実現していけばよいのか技術的な手法や考察を交えながら導入方法論について説明する。

### 2. 企業における典型的な情報システムの現状

我が国における多くの企業の情報システムは、企業活動における中核的なデータを記録する基幹系システムとそれを取り巻く業務系システム群で構成されている。これらの業務系システムはペーパーレス化や業務効率化を目的に個別最適化されており長年の利用の中で環境の変化にあわせ改修や再構築が繰り返されてきた。また、これらの業務系システムはマルチベンダー化や EUC(End User Computing)の促進で各部門や個人でのみ管理している情報システムも存在しているため、いわゆるサイロ化し、情報の共有や整合性確保が困難になってきている。これは、初めからシステム間連携やデータ共有を意識せず開発された経緯があるため当然の結果と考えられる。なお、これらの問題を解決すべく業務系システム群を統合した ERP(Enterprise Resource Planning)パッケージの導入を進める企業も多く現れたが、部分的な導入に留まったり、導入後の環境の変化に対応するためパッケージを補完するための周辺の業務システムを新たに構築したりと根本的な解決には至っていない。また、情報系システムの導入により情報を活用する取り組みもある。しかし、その多くは、ある業務・部門における蓄積であったり、業務システム間のデータ整合性を確保しないまま蓄積しているだけであるためデータの品質に問題があったり、複雑なデータ加工を行わなければならないかたたりと情報を有効に活用できていないのが現状である。具体的には、システム横断的な集計の精度が担保できない状態にある。

### 3. 全社的な情報活用基盤とは

#### 3.1 階層・部門毎の情報活用ニーズ

一般的に、企業における経営層とマネージャ層、各業務担当者では必要な情報とその活用ニーズは異なる。当然、情報の「見え方」に対する要求も異なる。

すなわち、経営層のニーズは、全社の状況把握である。業務分掌を考慮すれば、担当業務の状況把握が主要ニーズであり、ドリルダウンによる詳細把握は二義的ニーズにとどまると思われる。次に、マネージャ層を考えると、管理範囲の全体状況把握は当然のことながら、問題を含んだ箇所の特定のため、詳細状況を表示可能なドリルダウン機能は必須と思われる。そして、業務担当者においては、業務状況を具体的に調べ、個別にアクションプラン立案の材料となる詳細情報が必要となる。また、全ての層で業務実績の集計分析だけでなく、予算・実績対比、各種パフォーマンス指標 (KPI : Key Performance Indicator) のトレースを可能にする機能も望まれる。これらの実現には、当然、全社で整合性がとれた予算データ、および、KPI の設定と抽出ロジックの実装が必要である。図 1 に階層・部門毎の情報活用ニーズを示す。

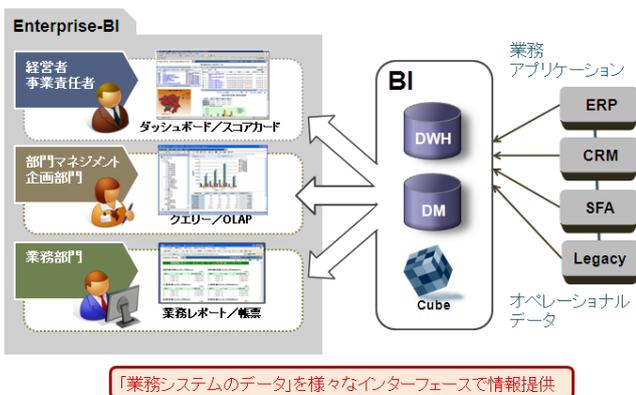


図 1 企業における階層・部門毎の情報活用ニーズ

全社的な情報活用を促進するためには、階層毎のニーズを的確に把握し、各階層に合わせた形での情報提供が必要となる。しかし、活用ニーズは異なっても各階層で見ている情報のソースは同じでなければならない (シングルソースデータの原則)。これは、経営層がマクロ的に捉えた企業レベルでの問題点や気づきを各マネージャや担当者にブレイクダウンして指示する際、共通の事実認識がなければ、効果的なアクションに繋がらないからである。

#### 3.2 全社的な情報活用基盤構築の前提となる要件

全社的な情報活用基盤にデータを供給する各処理系のシステム(ソースシステム)内で、クリアすべき事柄について考察が必要である。すなわち、事実を用いないで正しい情報活用は成り立たないからである。

##### (1) One Fact One Place の原則

事実を表現するデータが重複なく格納されているか。重複データが存在する場合、整合性等の問題が発生する。

##### (2) マスターデータの精度

マスターで管理するインスタンスがユニークに維持管理されていない場合、後の集計等が不正確になる。

##### (3) ファクトデータ(トランザクション)の粒度統一

たとえば、販売管理システムのトランザクションで、POS 明細とバルク販売データは別エンティティであるべき。

##### (4) 複数システムに存在するマスターの統合

同一インスタンスを名寄せする必要がある。MDM(Master Data Management)と呼ばれる統合をどこで行うのかは、複数の解決策が提案されている。

##### (5) 業務用語の統一

各システムでの用語(すなわちデータ)が同一基準で使用されているか。ビジネスリポジトリの確立が求められる。

#### 3.3 業務系システムとの関係および位置付

業務系システムの主要ミッションは基幹系であれ部門業務系であれ、ビジネスの遂行状況を反映した「業務トランザクション」の発生にともなう様々な業務処理(対外仕入販売, 在庫管理, 物流, 決済等)をサポートすることにある。したがって、ビジネスの「今」をデータとして保持することが主な目標となっている。

情報活用基盤は、これらのシステム群から、業務トランザクションを受け取り、さらに経時的に蓄積し、それらを活用可能な姿で保持することが基本的な要件として考えられる。この際、業務トランザクションの種類、蓄積期間、そして「粒度」が十分に考察されるべきである。粒度とは、そのトランザクションが単一の事実、事実の集約のどちらを表現しているのかを示す概念である。一般論でいえば、全てのトランザクションが発生単位であることが望ましいが、データ容量、アクセス性能の面から検討を要する事柄である。そして、マスターデータの

ありようについて考察してみる。それぞれの業務システムが部分最適を追求して作られた状況においては、商品、顧客、組織等のマスターが各々のシステムに存在していると思われる。このこと自体は、異常な事ではないが、情報活用基盤へのデータ供給を考えると、これらのマスターを統合するステップが必要になる。これを MDM と呼ぶ。以下に、MDM の技法について述べる。

(1) マザーマスター方式

情報活用基盤とは直接の関係なしに MDM を実施し全ての業務系システムのマスターにデータを供給するマザーマスターを構築する。加えて、統合リポジトリも確立すれば、トランザクションの生成から活用まで、一貫した意味づけでデータを扱うことが可能になる。

(2) MDM-HUB 方式

業務系システムから情報活用基盤へデータ流入経路 (ETL) 部分に各々の業務システムのマスターを集約統合する HUB を構築し、情報活用基盤内でのマスターバリューの統一を実現する。

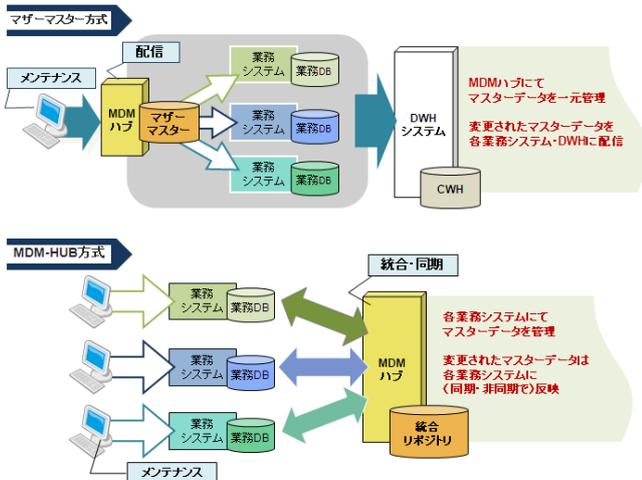


図 2 MDM の実現方式

MDM の主目的は以下のように示すことができる。

(1) データガバナンスの確立

各業務部門や業務システムが、部門の都合のみで用語・データを定義し、使用すると、会社全体の統一がくずれ、他部門へのデータ供給に齟齬が生ずる。

(2) データセキュリティ・プライバシーの確保

社会問題化している、情報漏洩を未然に防止するため、顧客・社員マスターの一元管理は重要課題である。

また、MDM の主な対象データは次の 2 エリアである。

(1) 商品マスター

場合によっては、製造業における基幹マスターである BOM(Bill of Materials)を含むデータを統合管理する。

(2) 顧客マスター

MDM-CDI(Customer Data Integration)として、近年注目を集めているテーマである。社外(顧客・取引先)、社内(社員・組織)を一元的な Party Model と呼ばれる「統合関係者マスター」に包含し、モデル化する技法である。

MDM は、場合によって、全社 DWH 構築プロジェクトと切り離して実施されるケースもある。また、抜本統合と疑似統合のパターン分けも有効な分類である。

(1) 抜本統合

マザーマスター方式に対応する概念で、全面再構築 (ERP 導入を含む) のタイミングで実施するのが妥当である。

(2) 疑似統合

抜本統合が不可能な場合、全社 DWH 構築の際、ハブ型 MDM として実現する。

最後に、情報活用基盤は、部門 Warehouse であれ全社 Warehouse であれ、ソースシステムのデータを抽出・変換し DWH として構築することになる。これを、そのまま、もしくはデータマート経由で情報活用する機能部分が BI と呼ばれることになる。

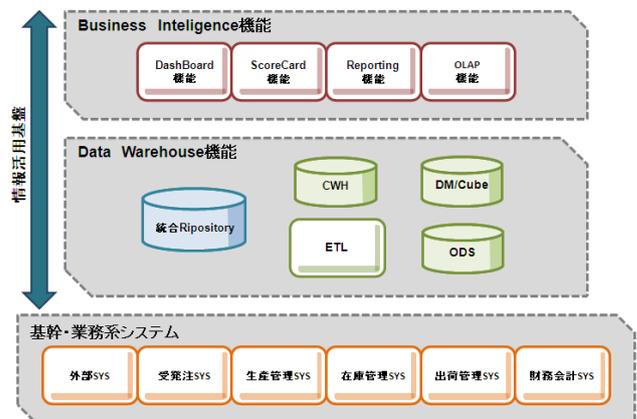


図 3 情報活用基盤

### 3.4 全社的な情報活用基盤に求められる要件

全社的な情報活用基盤に求められるのは、「正しい情報を必要な形で素早く出力できる」ことにある。「正しい」情報を出力するためには、全社に跨る個別最適化された情報システム間のデータ整合性と品質を確保した形で「繋げる」ことが必要となる。例えば、コールセンターのシステムとダイレクトマーケティングのシステムが個別で構築されている場合、顧客からコールセンターにダイレクトメールを送ってほしくないという電話があったにも関わらずダイレクトメールを送り続けてしまう、といった事態が起きる。この問題を解決するためには、顧客を軸に各システムで管理されているデータを繋げることが必要となる。そのためには、データを一元的に集中管理する情報系システムとしての DWH が必要となる。ただし、ここで言う DWH は広義の意味であり、狭義には CWH(Central Warehouse)とも呼ばれる。DWH 内のデータは一般的に RDBMS にてレコードとして管理されているため「必要な形で素早く出力」するためには、加工・編集が必要であり、目的別に最適化されたデータ群である DM(Data Mart)、目的に応じた形式でビューを提供する BI が必要となる。

なお、昨今、各情報システムの機能をサービスとして捉え、それらをバスで接続する SOA(Service Oriented Architecture)は、データ発生源である基幹系システム・業務システムの間でデータを繋げることを実現するための手法である。

全社的な情報基盤としての DWH/BI に求められる要件は以下の通りである。

- (1) データが一元管理され整合性と品質が確保されていること。  
基幹系や業務系といった各々のソースシステムから供給されるデータがオリジナルなのか複写なのか、加工後なのか不明確では DWH に取込むべきではないし、同様に品質、とりわけ、そのデータの帰属・性格を表すマスター項目の品質確保は重要である。
- (2) 外部変化(要求)に強く、長く使い続けられること。  
社内組織の再編による集計への影響、新商品の取扱いによる実績前年対比への影響、新事業分野の扱い等、ダイナミックな業務の遂行に追従可能でなければならない。
- (3) 増大する管理対象データに耐えられる容量と性能を併せ持つこと。

第一に、DWH を構成する RDBMS が十分なスケーラビリティを有する事、同様にハードウェア機材のスケーラビリティも重要で、場合によりブレードシステムの導入も有力な選択肢である。もちろん BI ツール自身がどの程度のスケーラビリティを有するかは、最重要の検討課題といえる。

- (4) セキュリティが確保され、証跡情報の取得が可能であること。  
社会問題化している情報漏洩の事例を見るにつけ、DWH/DM に対するセキュリティの確保は必須条件といえる。外部からのアクセス遮断は当然のこと、権限保有者であっても、アクセス対象データが特定可能な証跡情報(ログ)の取得は当然である。
- (5) 多様な情報活用ニーズに応じたビューが提供されていること。

具体的には、下記のように分類される。

- ① ダッシュボード機能  
売上げや利益、その達成度合いといった KPI やアラート情報を一画面に集約し、グラフ表示、チャート表示、ゲージチャート表示等を自在に出力可能な機能で、全体状況の把握に有効である。
- ② スコアカード機能  
あらかじめ設定した各種 KPI の達成度合いを信号機(青・黄・赤)といった形や、トレンドを矢印(↑・↓)で表現する機能で、ビジネスパフォーマンスの測定に有効である。
- ③ レポート機能  
クロス集計表と呼ばれる各種集計軸によるサマリ表や、その集計軸に含まれる明細の一覧を出力する機能で、定型フォーマットの出力だけでなく、対象データの絞り込みや集計のブレイクダウンのためドリルダウンを画面上で行うことも一般的である。
- ④ 分析機能  
OLAP 分析と呼ばれることが多く、ある集計軸に注目して掘り下げたり(ドリルダウン)、集計軸にそってデータを抽出したり(スライス)、集計の縦軸・横軸を自在に入れ替えたり(ダイス)する機能を備えている。
- ⑤ データマイニング  
大量の明細データを元に、相関関係やセグメント特性を発見する機能を有するソフトウェアで、専用ツールも各種販売されている。

### 3.5 全社的な情報活用基盤の構成

全社的な情報活用基盤は、基幹系システムや各業務系システムからデータを抽出し変換を行う ETL(Extract Transform Load)、ETL システムからデータを受け取り蓄積する ODS(Operational Data Store)、情報システム間の重複を省き整合性を確保し、経時的な蓄積を行った CWH、そして、目的別に最適化された DM の 3 層構造のデータ群と、セキュリティを確保した上で用途に応じた利用者へのビューを提供する BI で構成される。また、MDM 実施や DWH 運用で大きな戦力となるメタデータ管理も重要要件となる。

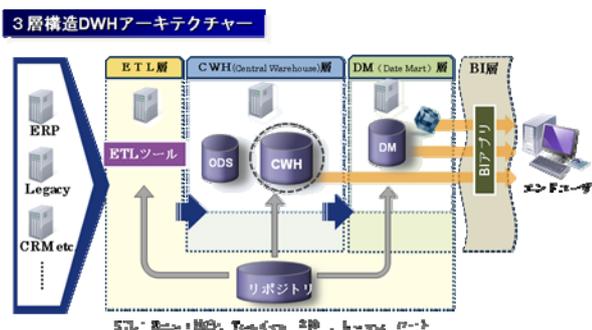


図 4 全社的な情報活用基盤の構成イメージ

### 4. 全社的な情報活用基盤の構築ステップ

情報活用基盤すなわち DWH の構築ステップに関しては、1990 年代後半、本格的には 2000 年になって研究成果が発表されるようになってきた。中でも中心的な役割を果たしたのが、TDWI(The Data Warehousing Institute)である。1995 年に設立された NPO 法人で毎年数回開催されるカンファレンスや教育プログラムによりビル・インモンやラルフ・キンボールの方法論やノウハウを発信している。

ラルフ・キンボールはボトムアップ方法論を提唱し、ユーザーニーズを直接実現する DM 群を考え、それらを集約して DWH を構築する方法論を提案した。それに対し、ビル・インモンの方法論はトップダウンアプローチであり、サブジェクトエリアの選定に始まり、ファクトデータの定義、時間軸を含むディメンジョンの考察へ進む、より全社レベルを意識した方法論を提示した。

ここでは、DWH の構築に際し、検討を要するポイントについての考察を、おおよその手順に沿って記述する。前記両者の方法論を取り入れたハイブリッドな方法論となっている。

#### (1) ニーズ収集

このステップは DWH 以外のシステム構築プロジェクトでも実施されるもので、ユーザーの要求をマクロ、ミクロを問わずに収集し文書化する事になる。具体的な方法はインタビューが中心になることが多い。DWH プロジェクトにおいては、何をしたいのかを中心にまとめる。マクロニーズはファクトの選定とディメンジョンの決定、ミクロニーズは DM の構成につながると考えられる。

#### (2) CWH の概念設計

要件書をもとに、下記のタスクを実行する。

##### ① ファクトの選定

全社の業務領域を見渡し、DWH にデータを収容すべきサブジェクトエリアを選定し、それぞれ 1 ないし 2 のファクトデータを粒度の考察を経て、決定する。ファクトテーブルとして実装される。ファクトには 3 種のタイプが存在する。

(ア)業務トランザクション

(イ)定期的(月末・期末)スナップショット

在庫データ、勘定科目元帳等

(ウ)累積型スナップショット

商品累積動態(受注・生産・出荷・請求・入金)

##### ② 蓄積期間の決定

業務要件を元にファクトデータの蓄積を行う期間を決定する。当然、蓄積期間が長ければ、ストレージ容量が増大し、パフォーマンスの維持に、高性能なプロセッサが必要になる。一般的には、5~7 年の蓄積を行うことが多い。

##### ③ ディメンジョンの決定

それぞれのファクトをどのような属性で集計・分析するのか、要件書から抽出し、ディメンジョンテーブルとして実装する。業務プロセスがどのようなディメンジョン単位に実施されているのかも重要な要素である。業務横断的にディメンジョンを発見・決定するために Bus Matrix 法<sup>5)</sup>がキンボールによって提案されている。

| Bus Matrix <sup>5)</sup> | ディメンジョン |    |     |      |        |     |      |
|--------------------------|---------|----|-----|------|--------|-----|------|
|                          | 日付      | 商品 | 仕入先 | 運送業者 | 物流センター | 販売店 | 活動区分 |
| 発注                       | X       | X  | X   |      | X      |     |      |
| 物流センター発注                 | X       | X  | X   | X    | X      |     |      |
| 物流センター在庫                 | X       | X  |     |      | X      |     |      |
| 販売店発注                    | X       | X  |     | X    | X      | X   |      |
| 販売店在庫                    | X       | X  |     |      |        | X   |      |
| 店舗販売                     | X       | X  |     |      |        | X   | X    |

図 5 Bus Matrix 法

次に、ディメンションテーブル間のリレーションを分析し、部分的な DM を作成する。ディメンションテーブルが階層構造になっていれば、結果的にスノーフレイクスキーマを構成することになり、OLAP 分析やレポート作成でのパフォーマンス阻害要因になることが考えられる。この場合、部分的に非正規化を実施し、できる限りスタースキーマ構造を実現すべきである。スタースキーマとは、ファクトを複数のディメンションが取り囲む構造で言う。基本的にディメンションはフラットにモデル化される。それに対し、スノーフレイクスキーマでは、ディメンション同士でリレーションシップが発生し、階層を構成するパターンである。当然ながらアクセス時の SQL 操作が複雑になる。

④ DWH の概念データモデル作成

ここまでの成果物を元に、DWH の概念データモデルとして纏める。

(3) 主要 DM の概念設計

ユーザニーズの中で、主としてレポート作成に対する要求と OLAP 分析の要件を元に DM を設計する。普通、DWH のファクトデータはレポート・OLAP 画面を直接作成できる並び・集計を持っていないので、利便性に重点をおいて設計する。結果として DM は冗長なデータということが出来る。DM に必要な項目に関しても、この段階でほぼ確定する。

(4) CWH/DM の外部設計

概念設計の成果物を元に、各テーブルの項目を具体化する。テーブル分割等の物理設計は、この段階では行わない。

① ファクトテーブルの設計

ファクトテーブルに収容する項目はファクトのキーを構成する項目群と数量項目、および少数の属性項目に分類される。必要と思われる項目を統合リポジトリに適合した型で定義する。事前に統合リポジトリが確立していなければ、ここでの定義を、リポジトリとして集積する。

② ディメンションテーブルの設計

ディメンションテーブルの役割は分類・集計軸の提供と意味づけにある。概念設計で提示されたテーブルに対しキーを含む各項目を定義する。統合リポジトリとの関係は、ファクトテーブルと同様に考える。

③ SCD(Slowly Changing Dimension)の考察

一般的に SCD の対応は、変更が発生した場合のデータの扱い方により「何もしない」と「対応する」の 2 種類に大別されるが、詳細は下記の 5 タイプに分類される。

(ア)タイプ 0

何もしない。したがって、変更後の属性のデータは新データとして存在し、変更前との紐付け機能はない。

(イ)タイプ 1

変更前データを変更後の属性で書き換える。履歴をとらないので、オリジナルは保存されない。

| 顧客KEY | 顧客名     | 所在地 |     |
|-------|---------|-----|-----|
| 001   | (株)日立〇〇 | 宮城  | old |
| 001   | (株)日立〇〇 | 神奈川 | new |

(ウ)タイプ 2

履歴データを同一テーブル内に作成し、バージョンやスタート・エンド日のような項目で識別する。集計時に注意が必要になる。

| 顧客KEY | 顧客番号 | 顧客名     | 所在地 | バージョン |     |
|-------|------|---------|-----|-------|-----|
| 001   | 001  | (株)日立〇〇 | 宮城  | 0     | old |
| 002   | 001  | (株)日立〇〇 | 神奈川 | 1     | new |

| 顧客KEY | 顧客番号 | 顧客名     | 所在地 | 開始日       | 終了日      |     |
|-------|------|---------|-----|-----------|----------|-----|
| 001   | 001  | (株)日立〇〇 | 宮城  | 2000/1/15 | 2003/3/3 | old |
| 002   | 001  | (株)日立〇〇 | 神奈川 | 2003/3/4  |          | new |

(エ)タイプ 3

変更データ項目を同一テーブル内に作成し、適用日等で管理する。事前に、予備項目を用意する必要があり、集計時の対応も必要である。

| 顧客KEY | 顧客名     | 前所在地 | 変更日      | 所在地 |
|-------|---------|------|----------|-----|
| 001   | (株)日立〇〇 | 宮城   | 2003/3/4 | 神奈川 |

(オ)タイプ 4

ヒストリテーブルを作成する。変更前データをヒストリテーブルに避難し、最新データを保持する。集計時の影響はない。

| 顧客   |         |     |
|------|---------|-----|
| 顧客番号 | 顧客名     | 所在地 |
| 001  | (株)日立〇〇 | 神奈川 |

| 顧客履歴 |         |     |          |
|------|---------|-----|----------|
| 顧客番号 | 顧客名     | 所在地 | 開始日      |
| 001  | (株)日立〇〇 | 宮城  | 2003/3/4 |

最近の研究によれば、上記の技法を組み合わせた方法も存在する。現実的には、上記 0~4 のタイプから選択し、設計に反映すべきである。ただし、運用開始後の対応法変更は避けるべきである。この結果を外部設計に反映する。

(5) ETL プロセスと ODS の検討

ODS で実装するテーブルは経時的な蓄積を目指すものではないので、基本的には CWH にロードするデータの一時テーブルと考える。考慮すべき項目は下記ようになる。

- ① ファクトテーブルのソースデータ  
ファクトテーブルデータのソースとなる業務系システムのトランザクションを抽出し取込む。そのサイクルは CWH の更新サイクルと同期するのが簡明である。
- ② ディメンションテーブルのソースデータ  
業務系システムにマスターデータがソースとなるが、基本的にはマスターデータに追加・変更があった都度、抽出し取込む。ただし、MDM を ODS の近傍で行う場合、さらに MDM 用のマザーマスターとの摺り合わせが必要である。
- ③ MDM-HUB 実装の検討  
複数の業務システムから取込んだマスターデータを統合する HUB 型マスターを実装する場合、ODS データの周辺に実装するのが適当と思われる。個別に作り込むか、専用ツールを利用するかの判断が要求される。
- ④ ETL での変換プロセスの検討  
収集されたデータを CWH のテーブルに変換するプロセスで考慮すべきポイントを列挙する。
  - (ア) コード値変換  
例えば、性別コード {1,2} を {M,F} に変換
  - (イ) 項目値のエンコーディング  
例えば、{Doctor} を {Dr} にコード化
  - (ウ) 導出データの生成  
数量と単価から金額を導出
  - (エ) フィルタリング  
レコード、項目を選択
  - (オ) 複数テーブルの結合  
分割されたトランザクションの紐付けやマスターのルックアップによるデータの

生成

- (カ) アグリゲーション  
合計値や平均値の生成
- (キ) コンカチネートカラムの分離  
複合要素を持つカラムを分離
- (ク) 整合性エラーの対応  
マスター突合でエラーになるデータの扱い項目値の変更または、データの削除

⑤ 中間テーブルの検討

最終的な CWH の項目が直接的に ETL の結果として生成できず、多段階のプロセスが必要となる場合、中間的なテーブルを実装し、ETL を構成することも考えられる。

⑥ ETL 実現方式の決定

以上の結果を元に ETL の変換プロセスの規模を想定し、ユーザープログラムを作成するか、ETL ツールを導入するか検討する。一般的に、ユーザープログラムの場合、開発・運用費用は、DWH 構築プロジェクト費用の 70% に達すると言われ最大の負荷になると考えられる。

(6) データ容量の見積

ここまでの成果を元に、ODS、CWH、DM の容量を積算し、必要なストレージを見積もる。その際、テンポラリーDB 領域を 3~4 割追加する必要がある。(BI 機能の種類によって、必要量は変動すると考えられるので、一般的な目安として提示する。)

(7) ツールの選定

この段階で、各種ツール(DBMS・ETL・MDM・BI)を選定する。ETL と MDM に関しては、プロジェクトの判断による。さらに、出力の高速化のため、インメモリ DB/BI 製品を検討することも有効と考えられる。実用レベルのインメモリ DB/BI 製品が発売され始めたので、現実的な選択肢と考えてよい。

(8) 内部設計以降

内部設計以降のステップに関しては、通常の業務システム構築手順と同様に考えて差し支えない。

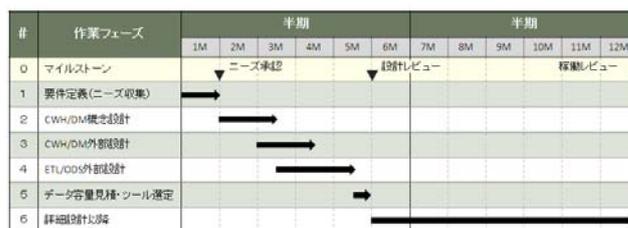


図 6 DWH/BI システムの構築ステップ(例)

## 5. 今後の課題

DWH/BI に対する技術的な要求は、ユーザーの適用効果が明瞭になるにつれ、ますます高度なものになっている。事実、SCD の対応技法は、黎明期の概念には含まれていない。ビル・インモンによれば、DWH のデータは非揮発性であることを求められ、変更されないことが前提になっていた。ただ、複数年度のデータ分析のニーズの中から必然的に対応を迫られた。また、MDM の概念にしても、主として、業務システムのデータ整合性の観点から研究され、業務横断的なデータを扱う DWH/BI にも波及してきたというのが経緯と思われる。

現在、最もホットな技術的課題はリアルタイム DWH/BI の実現である。ダッシュボード表示が一般的になると、リアルタイム性の要望が顕在化するの、必然と思われる。複数のベンダーが対応ツールを発売している。ただし、弊社導入事例も含めて真のリアルタイムを実現した事例はあまり報告されていない。いずれにしても、我が国では、部門 DWH・DM の実装がせいぜいで、全社 DWH/BI に取り組んでいる事例は非常に少ないのが現状である。

また、ユーザーに DWH の取組みを促す上で、「小さく始めて、大きく育てる」アプローチをとらなければ、予算の大きなプロジェクトの開始は難しい。となれば、部門 DWH の構築を突破口にするのが現実解の一つと考える。ただし、構築方法論は全社レベルでも適用可能な本格的なステップを踏むべきであろう。

今後、競争力の強化が求められる中で、全社情報活用基盤の構築ニーズは確実に増加すると思われる。その中で当社は、国内における数多くの DWH/BI 構築実績と本論文で纏めた実践的な方法論をもとに、全社情報活用基盤の「あるべき姿」とそのロードマップを示すとともにお客様のデータ整備状況や事業環境にあわせた構築ソリューションの提供を行っていく。

## 6. おわりに

DWH 構築方法論については、まだ決定的な手法が確立していない段階といえる。さらに我が国の現状では、プロジェクトの数が少ない状況なので、今後も米国での成果に期待せざるを得ない事情が、しばらくの間続くとと思われる。継続的に、論文、文献をウォッチする必要がある。また、理論的整合性よりも実用性に着目し、実践的な方法論をブラッシュアップしていくことが、ユーザにソリューションを提供する側の務めと考える。

## 参考文献

- 1) Bill Inmon, 「Building the Data Warehouse」, 2001
- 2) Ralph Kimball, 「The Data Warehouse Toolkit」, 1996/2/2
- 3) Ralph Kimball, 「The Data Warehouse Toolkit, 2<sup>nd</sup> Edition」, 2002/4/26
- 4) Ralph Kimball, 「The Data Warehouse ETL Toolkit」, 2004/10/1
- 5) Ralph Kimball, 「The Microsoft Data Warehouse Toolkit」, 2006/2/3
- 6) Alex Berson, 「Master Data Management and Customer Data Integration」, 2007/5/24
- 7) Silverston, 「The Data Model Resource Book」, 1997/3
- 8) Silverston, 「The Data Model Resource Book Vol2」, 2001/3/7
- 9) Silverston, 「The Data Model Resource Book Vol3」, 2009/1/9
- 10) Candace C. Fleming, Barbara von Halle  
「Handbook of Relational Database Design」, 星 芳彦  
訳, 1997/12/12



桑島 義行 2000 年入社  
金融 BI ソリューション G  
金融業への BI ソリューション提供  
kuwajima@hitachi-to.co.jp



星 芳彦  
(株)東北電子計算センター  
金融業を中心としたデータモデリング、  
コンサルティング  
yoshihiko.hoshi.01@hitachi-to.co.jp



佐藤 徹 1988 年入社  
BI ソリューション部  
金融業への BI ビジネス展開  
to-satou@hitachi-to.co.jp



千葉 憲昭 1986 年入社  
BI ソリューション部  
BI ビジネス展開  
n-chiba@hitachi-to.co.jp