

情報系データベースシステムの構築手法

7K-6

佐藤 正 河村 浩

(株) 東芝 東京システムセンター

1. はじめに

近年、運用中の基幹業務システムに蓄積されたデータを有効活用するため、検索を目的としたデータベースに整理して、企業戦略の一端として利用しようとする気運がたかまっている。

本稿では、そうした流れを受けて新たに情報系システムを構築した事例に基づき、基幹業務からのデータ移行に関する手法と、実際にデータ移行に携わった立場からそこで実際に発生した問題点について考察する。

2. 概要

ここで紹介する事例は東京、大阪、福岡、名古屋に本社組織を持つ企業の情報システムである。それぞれの本社において販売業務、総務、経理の処理系システムが稼動し、一部の本社において出版、広告業務のシステムが稼動している。こうした一連の基幹業務システムが稼動する環境は、ハードウェア、基本ソフト、ファイルシステム等が互いに異なるうえ、本社間における業務上のコード体系も独立したコード体系となっている。

情報系システムは、このようにいくつものシステムに分散して、長年にわたって蓄積されてきたデータを1箇所に集め統合し、企業戦略や業務の意志決定の手段として多目的なデータ検索をすることにねらいがある。

3. 開発手法

開発事例においては情報系システム以外に基幹システムを将来ダウンサイジングするためのデータモデル作成も目的としていた。このため、主にボトムアップ手法によりデータを分析した

結果、異義同音語、導出項目の切り捨て、統合を経てデータ項目数は2.3%に整理された。さらに正規化を行うことで得られたERDはCASEツール(ORACLE* Designer2000)によりRDBの物理表へマッピングした。

データモデル化の手順

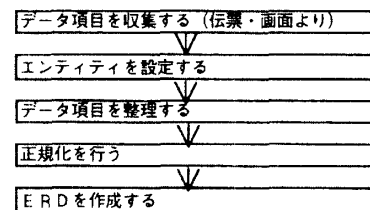


図1

4. データ移行手法

データを基幹業務システムから移行する規約、手段については1本化した。基幹業務システムは情報系システムと比較してデータの構造やコード体系が異なるためデータ構造の変更およびコード変換が必要になる。このため、変換パラメータの指示に従いデータ変換するツールを開発した。代表的な変換機能を図2に示す。

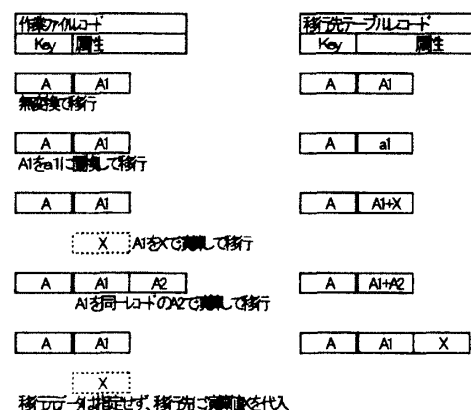


図2

The building technique of the information database system
Tadashi Sato, Hiroshi Kawamura
TOSHIBA Corporation.

(*)ORACLEはOracle Corporationの登録商標です。

5. データ移行時の問題点

基幹業務システムから、情報系システムへのデータ移行は基幹業務システムの十分な調査もとに実施した。机上の調査では汲み取れなかった問題は移行検証作業により発見し除去した。この作業では、物理マッピングされた表に対して一通りデータを移行して整合性を検証する。以下では基幹業務システムのデータ調査から移行検証時において発生した主な問題点についてまとめた。

ここで、データ移行を表現するモデルを

(a) 基幹業務を表現する概念データモデル

(b) 基幹業務システム上のデータモデル

(c) 情報系システム上のデータモデル

の3つの集合で考え、 $(x) = \{(a), (b)\}$ の要素（データレコード）を(c)へマッピングする行為 f を

$$f(x) \rightarrow (c)$$

と定義すると、今回の移行は、 f の集合 A の部分集合 $B = \{f \in A \mid f(b) \rightarrow (c)\}$ となる。

データ移行の概念

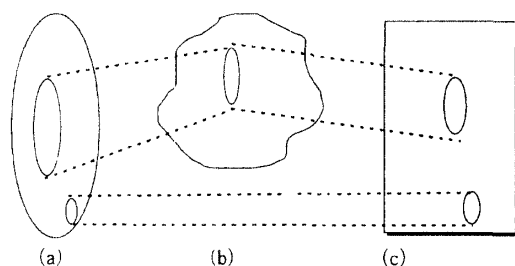


図 3

5-1. 基幹業務システムにデータが存在しない

データ分析に使用した帳票が実はワープロにより人がおこした帳票だった場合がある。すなわち、 f は A の要素であるが B の要素ではない。そのような場合は情報系に取り込む有効なデータは存在せず、移行は成立しない。

5-2. 元データと移行先データがN:1

$f \in B$ によってマップされる要素が N:1 の関係になる場合を指す。

情報系システムの1レコードが複数の基幹業

務システムデータで構成される場合、競合を考慮して処理する必要がある。またこのような場合には重複更新となっていないかどうか、データのメッシュの統一、コードの統一について確認をする必要がある。

5-3. データのクリーン度の問題

$f(a) \rightarrow (b)$ で行われるデータチェックのレベルは基幹業務システムの要求にあわせたものであり、 $g \in B$ で行われるデータチェックでエラーとなり移行が成立しないことがある。例えば、日付型の項目に対する'19960231'の登録はエラーとなる。また、基幹業務システムの処理の都合で数値が入るべき項目に空白がセットされていることもある。さらに多くの場合は、 $g \in B$ においてもチェックされていないため、別の手段による内容の確認が必要となる。

5-4. 基幹システムの処理の都合による問題

$f \in B$ によってマップされる要素に論理的に重複するものが含まれるケースがある。例えば、テストデータや過去に洗い替えたはずのデータが残っていた場合である。情報系システムでは上書きの扱いとなりデータの精度に問題が生じる。

5-5. 変換パラメータで表現できないケース

変換パラメータは $f \in B$ における変換の機能全ては表現できない。例えば、姓名などのように1項目で表現されるデータを姓と名に分離して移行する機能は組み込まなかった。

6. まとめ

検証作業を経てまずはデータがRDBに入ることになるが、データを利用者に開示する際に注意を要するのがデータの質、セキュリティの問題である。現在、移行対象となる基幹データがどのように運用されていたかについての調査や、データセキュリティに関する仕組みの実装を進めている最中である。さらに今後は利用者に対するサポート体制、データの精度の維持、監視機能の検討をすすめて行く必要がある。