

## 履歴管理可能な企業内マスターデータのスキーマ設計方針

酒井理江<sup>†</sup> 水口正彦<sup>†</sup> 中川真一<sup>†</sup> 村山隆彦<sup>†</sup> 赤埴淳一<sup>†</sup><sup>†</sup>日本電信電話株式会社 NTT 情報流通プラットフォーム研究所

## 1. はじめに

企業内に存在する種々の情報資産を有効活用する事が求められている。特にマスターデータは重要性が高く、そのデータ管理のための製品も登場している。それらはデータ統合や配信をサポートするが、統合先のスキーマ設計は独自に行う必要があり、その指針も存在しない。

有効活用の観点では時系列の調査分析等での利用が増えつつあり、業務処理の最新データのみならず、履歴管理されたデータも効率的に活用することが不可欠である。マスターデータ管理製品では通常トランザクションログを履歴データの代わりに保存するだけであるため、ログからの履歴データのトレースは困難である。

本研究では履歴管理を可能とするマスターデータのスキーマ設計方針について、社員のマスターデータを具体例に挙げて検討した結果を報告する。

## 2. 企業内データの現状

## 2.1 企業内データ利用のニーズ

企業内には共有資産である多くのデータが存在するが、それぞれが異なるシステム上や組織毎の異なるフォーマットで蓄積されているため、データの統合した利用が困難である。これらのデータ群が既存の閉じた枠組みを超えて参照・有効活用できることが望ましい。

昨今では企業内サイトやデータを検索するためのエンタープライズサーチ[1]、企業内システムを連結しデータ連携を図る EAI (Enterprise Application Integration) [2]、企業内の知識共有が可能なようにするナレッジマネージメント[3]などの技術が注目され、多くの製品も登場している。また企業データの利用法として、例えば過去の購買履歴より今後の戦略を立てる等、蓄積された過去のデータを分析して意思決定に役立てるようなデータウェアハウス[4]も注目されている。

## 2.2 マスターデータ

企業内データには業務活動によって発生するオペレーショナルデータ (以下、OD) や、説明資料などの非構造データ等、数種のデータが存在するが[5]、本研究では企業内データの根幹となるマスターデータ (以下、MD) に着目する。MD とは業務で利用する基本データの事であり、社員を管理する“社員マスタ”や組織を管理する“組織マスタ”等が存在する。MD は OD と比較して更新頻度は低いものの、時間と共に変化する情報である。また他のデータとの関わりが多く、殆どの社内サービスで参照されるという特徴を持つ。

## 3. マスターデータの課題と解決法

## 3.1 マスターデータの課題

企業では例えば各組織での発表分野や件数の変遷を確

認し注力すべき分野を絞り込む等、今後の方針検討等に過去のデータ、つまり履歴データを利用する事が多い。

MD 管理製品も多く登場しているが[6]、これらの製品では、最新データのみを保持し時間情報や履歴データの管理を省くことにより、データのスリム化と処理の軽量化を図っている。一般的にトランザクションログを履歴データの代わりに保存しているが、このログから履歴データをトレースしたい場合はロールバック等の処理が必要となるため困難である。履歴データをサービスで用いるためには最新データと共にテーブルに保持する必要がある。

## 3.2 課題の解決法

最新データと履歴データを共にテーブル内で扱うためにはデータの有効期限を明確にする必要がある。有効期限を明示する単位としては (方法 1) テーブル単位、

(方法 2) レコード単位、(方法 3) フィールド単位が考えられる。具体的に (方法 1) は 1 日 1 回等のタイミングで新たなテーブルを作成、つまりその時点でのデータのスナップショットを残す方法、(方法 2) は各レコードに有効期限を示す“開始日”と“終了日”カラムを追加する方法、(方法 3) は属性 X の値を  $(X_1, \text{開始日}_1, \text{終了日}_1), (X_2, \text{開始日}_2, \text{終了日}_2), \dots$  と示すなど、フィールド内に有効期限を追加する方法である。

これらの方法を比較すると (方法 1) は実際の運用を考えると一定間隔でスナップショットを取る事となるため、任意の時間に戻せず課題の解決が出来ない。(方法 3) は履歴データの増加に伴い各フィールドの値の長さが膨大となり、指定期間の値を検索するのも困難となるため現実的でない。(方法 2) はカラム数が増えるが (方法 3) と比べ検索も容易であるため (方法 2) を採用する。

## 4. 履歴管理可能なスキーマ設計方針の検討

本章ではまずスキーマ設計案を比較する際の観点となるスキーマ設計が満たすべき条件を挙げ、どの企業でも必ず存在する社員マスタを例にとって 3.2 節の解決策を用いた具体的なスキーマ設計方針を検討する。

## 4.1 マスターデータの性質からくるスキーマ設計時の条件

(条件 1) 1 つの主キー (以下、PK) で他データと関係が取れる事。MD は他のデータとの関わりが多いため、関係が簡単に追えるようにするためである。

(条件 2) 重複箇所を排除する事。MD は利用範囲が大きいため、データ不整合が含まれると影響範囲も大きくなる。一般的に重複がデータ不整合の原因となる。

(条件 3) スキーマ変更要望に対して柔軟に追従できる事。MD は携わる組織やサービスが多いため、運用後の変更要望が多いと考えられるためである。

## 4.2 社員マスタの主キーの検討

条件 1 より、まず社員マスタの PK を検討する。通常社員マスタの PK には社員 ID を用いる事が多いが、3.2

A guideline for master data schema design including historical data

<sup>†</sup> SAKAI Rie, MINAGUCHI Masahiko, NAKAGAWA Shinichi, MURAYAMA Takahiko, AKAHANI Jun-ichi

NTT Information Sharing Platform Laboratories, NTT Corporation

節の解決策より異なる有効期限で同じ社員 ID を持つレコードが登録されるため、社員 ID は PK に成り得ない。社員 ID と“開始日”を組合せた ID の場合、同日に任命された兼務は同 ID となってしまふ。また社員 ID と組織 ID を組合せた ID の場合、出向後同じ組織に戻ると同 ID となってしまふ。

最小の組合せで作成できる PK を検討すると、社員 ID と組織 ID と“開始日”の組合せとなる。条件 1 より 3 カラムの複合 PK ではなく、3 カラムの値を組合せた新たな PK を作成する事とする。

### 4. 3 テーブルの分割法

条件 2 より、一般的に重複排除のためには正規化を行うが、正規化には時間軸の概念が含まれていないため、時間と属性でテーブルを分割する方法を図 1 に示す。

属性で分割 →		IDとそれ以外の属性		IDとそれ以外の各属性	
分割なし	①	②	③		
最新と履歴	最新テーブルと履歴テーブル間が主キーで関係が取れない				
スナップショット	任意の時間に戻せず、履歴管理が出来ない				

図 1. テーブル分割案

図 1 の時間で分割の“スナップショット”とは 3.2 節で挙げた（方法 1）と同等であるため、同じ理由で不採用とする。同じく図 1 の“最新と履歴”は最新データからなるテーブルと履歴データからなるテーブルに分割する方法である（図 2 参照）。

最新データテーブル

ID	社員ID	組織ID	部屋番号	...	starDate
112222220404	1122	2222	MH8F	...	2004-4-1
223311110004	2233	1111	MH9F	...	2000-4-1

履歴データテーブル

ID	社員ID	組織ID	部屋番号	...	starDate	endDate
112211110004	1122	1111	MH9F	...	2000-4-1	2004-3-31

図 2. 最新と履歴データテーブルに分割

この案では、例えば ID を形成する属性である組織 ID の変更時、新たな ID が振られるため最新と履歴データテーブル間で PK を用いた関係が取れない、つまり条件 1 が満たせないため、不採用とする。

図 1 の属性で分割の“ID とそれ以外の属性”と“ID とそれ以外の各属性”は PK を形成する属性群（以下、メイン属性）とそれ以外の属性（以下、サブ属性）を別テーブルに分割しており、前者ではサブ属性を 1 テーブルに、後者ではサブ属性毎にテーブルを分割している（図 3 参照）。

IDとそれ以外の属性						IDとそれ以外の各属性					
メイン属性テーブル						メイン属性テーブル					
ID	社員ID	組織ID	starDate	endDate		ID	社員ID	組織ID	starDate	endDate	
112211110004	1122	1111	2000-4-1	2004-3-31		112211110004	1122	1111	2000-4-1	2004-3-31	
112222220404	1122	2222	2004-4-1	NULL		112222220404	1122	2222	2004-4-1	NULL	
223311110004	2233	1111	2000-4-1	NULL		223311110004	2233	1111	2000-4-1	NULL	
サブ属性テーブル						部屋番号テーブル					
ID	部屋	電話番号	...	starDate	endDate	ID	部屋	starDate	endDate		
112211110004	MH9F	0422-59-1111	...	2000-4-1	2004-3-31	112211110004	MH9F	2000-4-1	2004-3-31		
112222220404	MH9F	0422-59-1111	...	2004-4-1	NULL	112222220404	MH9F	2004-4-1	NULL		
223311110004	MH9F	0422-59-2222	...	2000-4-1	NULL	223311110004	MH9F	2000-4-1	NULL		
電話番号テーブル						電話番号テーブル					
ID	電話番号	starDate	endDate			ID	電話番号	starDate	endDate		
112211110004	0422-59-1111	2000-4-1	NULL			112211110004	0422-59-1111	2000-4-1	NULL		
223311110004	0422-59-2222	2011-4-1	NULL			223311110004	0422-59-2222	2011-4-1	NULL		

図 3. 属性で分割

図 1 の案①は履歴データを含む全データを 1 テーブルで管理するものであるが、例えば部屋番号のみの変更では 4.2 節で定めた ID では PK と成り得ず、条件 1 より不採用とする。

案②と③ではサブ属性の追加時、案②ではサブ属性テーブルの変更が必要だが、案③では新たなテーブルを追加するだけで済むため、条件 3 より案③を採用する。

なお、サブ属性の中で PK と同じタイミングでしか変更されない属性は個別にテーブルを持つ必要がないため、メイン属性テーブルに含む事とする。また属性とテーブル名の対応が分かるよう対応表を追加する。本スキーマ設計指針を適用して作成した企業の社員マスタを図 4 に示す。

メイン属性テーブル					属性-テーブル対応表	
ID	社員ID	組織ID	starDate	endDate	属性名	テーブル名
112211110004	1122	1111	2000-4-1	2004-3-31	部屋番号	部屋番号テーブル
112222220404	1122	2222	2004-4-1	NULL	電話番号	電話番号テーブル
223311110004	2233	1111	2000-4-1	NULL		

  

部屋番号テーブル					電話番号テーブル			
ID	部屋	starDate	endDate		ID	電話番号	starDate	endDate
112211110004	MH9F	2000-4-1	2004-3-31		112211110004	0422-59-1111	2000-4-1	NULL
112222220404	MH9F	2004-4-1	NULL		112222220404	0422-59-2222	2011-4-1	NULL
223311110004	MH9F	2000-4-1	NULL					

図 4. スキーマ設計指針を適用した社員マスタ

## 5. 考察

本方針では全属性が PK に対し不変である場合はテーブルが 1 つとなるが、このようなケースは稀であるため本方針は有用であると考えます。

## 6. まとめ

本研究では企業内のマスターデータの履歴データに着目し、社員マスタを例に履歴管理が可能となるスキーマ設計方針を示した。

今後は他種のマスターデータも含めて本設計にて実装を行い、追加や検索処理を検証する必要がある。また、本設計が異業種他社のマスターデータにも適応可能か検討していきたい。

## 参考文献

- [1] Falk Brauer et al., Graph-Based Concept Identification and Disambiguation for Enterprise Search, WWW2010, 2010.
- [2] Athanasios Bouras, Panagiotis Gouvas, Gregoris Mentzas, ENIO: An Enterprise Application Integration Ontology, 18th International Workshop on Database and Expert Systems Applications, 2007.
- [3] Tiedong Chen, Ziyu Liu and Lei Huang, Research and Application of Enterprise Knowledge Management System Based on Ontology, International Federation for Information Processing, 2008.
- [4] 山崎慎一, データウェアハウスのモデリング, UNISYS Technology Review 第 68 号, 2001, [http://www.unisys.co.jp/tec\\_info/tr68/6815.pdf](http://www.unisys.co.jp/tec_info/tr68/6815.pdf)
- [5] Mario Gomez et al, The Art of Enterprise Information Architecture: A Systems-Based Approach for Unlocking Business Insight, IBM Press, 2010.
- [6] David Butler, Better Information through Master Data Management – MDM as a Foundation for BI, An Oracle White Paper, 2010, <http://www.oracle.com/us/products/applications/master-data-management/018877.pdf>