

データベース設計品質の評価方法について

5G-6

村田達彦 川下満
NTT情報通信研究所

1. はじめに

データベース（DB）の設計の良否を判断するとき、これまで客観的な評価基準がないため、設計されたDBの性能や拡張の容易さ等の様々な要因について漠然とした判断がなされていた。

筆者らは、DB設計方法の効果を客観的に判断することを目的に、DB設計の品質の評価方法を検討している。本報告ではDB設計の品質を決定する項目と、各項目について定量的に判断するための評価尺度を提案する。

2. 前提条件と方針

(1) 以下の前提条件をおく。

- ①リレーショナルDBの設計を対象とする。
- ②集中型DBの設計を対象とする。
- ③DBシステムの一連の設計の内、DBの論理設計の結果を対象とする。即ちDBの論理的な構造である、テーブル、カラム、インデックスの設計結果を対象とする（図1参照）。

(2) 以下の方針とする。

- ①DB設計方法の効果の評価方法確立を第一の目的とするが、既に設計されたDBの品質評価への適用も考慮する。
- ②詳細な比較評価よりも、様々な観点からの大局的な評価を可能とする。
- ③経験者の感覚と一致する結果が得られる、定量的な評価尺度とする。
- ④DB設計方法に依存しない評価方法とする。

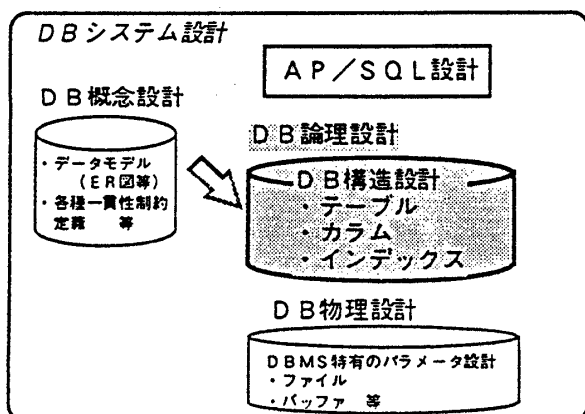


図1 DBシステム設計と本評価の対象 (●部分)

3. DB設計の品質項目

DB設計の品質を考えると、性能や拡張の容易さ等の様々な要因がある。そこで品質項目を決めるに当たり、ソフトウェアの品質項目[1][2]を参考にし、これをDB設計の場合に当てはめ、表1のように定義した。

表1 DB設計の品質項目

項番	品質項目	DB設計の品質項目の定義	重要度
1	正確性 (Correctness)	要求通りの業務を実現可能なDB構成であること	①
2	頑文性 (Robustness)	DB内のデータが異常な状態にあっても制約付きで使用できること	
3	拡張性 (Extendibility)	要求の変更に対しDB/APの追加、変更が容易であること	③
4	再利用性 (Reusability)	他のDB設計者が利用し易い(データを扱い易い)構造であること	④
5	互換性 (Compatibility)	他のDBを組み合わせて容易に業務を構築できること	
6	効率性 (Efficiency)	アクセス性能が良いこと	②
7	携帯性 (Portability)	種々のDBMSのDBとして構築できること。特定のDBMSに依存する構造になっていないこと	
8	実証性 (Verifiability)	誤ったデータの検出が容易にできること。正しいデータであることの検証手段が示されていること	
9	健全性 (Integrity)	一貫性を損なう等の不当な更新から保護されていること	
10	使い易さ (Ease of use)	APが作り易いこと(中味が理解し易いこと)	

表1のうち、以下の理由により正確性、効率性、拡張性、再利用性、の4項目をDB設計品質項目の代表として抽出した。

- 正確性：正確性を満たしていることが他の項目の評価の前提であり、基本的な項目である。
- 効率性（性能）：運用に供するにあたっての最も重要な項目である。
- 拡張性：性能を追求すると歯止めがなくなる。その歯止めとなる項目でもあり、性能とのトレードオフの関係にある。
- 再利用性：従来軽視されていたが、貴重なDB資産を有効活用する観点から重要である。

以降では、上記4項目に関し、具体的にDB設計の結果のどこを測定して数値化し評価するかという、定量的な評価尺度について述べる。

なお、残りの6項目については次の理由により優先度を下げた。

- ・頑文性、実証性、健全性：DB構造の問題よりも

APあるいはDBMSに依存する。

- ・互換性：前提条件の②より、今回は対象外とする。
- ・携帯性：前提条件の③より、本評価の対象自体がDBMSに依存しない論理設計の結果である。
- ・使い易さ：DB構造の設計を評価する観点からは再利用性に帰着される。

4. DB設計の品質項目の評価尺度

4.1 正確性

DB設計方法の良さを判断するとき、どの程度正確に設計できるかは重要な評価項目である。一方、既に運用中のDBの場合には、要求どおりの業務を既に実現済みであり、正確性は満たしていると考えられる。

正確性に関しては、筆者らはDB構造の設計結果の良さを評価することを目的としていることから、以下の尺度を採用することとした。

「与えられた業務を実現する上で、設計されたDB構造上の間違い箇所の割合」

具体的には、業務を正確に実現する上で、設計されたDBに対して修正が必要な以下に示す箇所の、全体に占める割合を尺度とする。

- ①不足しているカラム
- ②主キーの間違い
- ③参照キーの抜け
- ④カラムの属性（型・桁・精度）間違い
- ⑤ユニークでないカラムに付けられたユニーク・インデックス

4.2 効率性（性能）

以下の2つのケースについて評価尺度を検討した。

<ケース1>：性能実測環境があるケース

実測環境がある場合には、様々な観点からの性能測定が可能である。ここでは一般に代表的と考えられる以下の3つを尺度とし、実測値で評価する。

- (1) 各々の業務APを単独走行させたときの平均応答時間
- (2) 全APを多重走行させたときの平均応答時間
- (3) 全APを多重走行させたときのスループット

<ケース2>：性能実測環境がないケース

この場合、性能に大きな影響を与える以下の2つを評価尺度とする。

- (1) 検索条件で指定するカラムのうち、何らかのインデックスが定義されているカラムの割合
- (2) 性能向上が期待できるDB構造の改善箇所
具体的には、カラムのコピー、テーブルの統合／分割等により正規形を崩した箇所の数。^(*)

(*)：全体数が不明確なため、割合での表現は不可

4.3 拡張性

本来は、想定する仕様変更時にDB/APの修正にかかる稼働を測定すべきであるが、網羅的な変更仕様の設定が困難等の測定上の問題があり、次の尺度を採用する。

「DB構造の容易な変更を阻害する箇所」

具体的には、以下のような拡張性の阻害要因となる箇所の数を測定する。^(*)

- ①正規化されたDB構造に対して、テーブル統合や分割、繰り返し等により正規形を崩した箇所
- ②導出カラムや余分なカラム、カラムの統合／分割等を行っている箇所

4.4 再利用性

他のDB設計者が利用しやすい、という中には次の2つがある。①他のDB設計者がそのDBを理解しやすいこと（この意味で使い易さと等価）。

②他のDB設計者がそのDBのデータを扱い易いこと。以上の観点から次の2つの尺度を採用する。

- (1) データ標準化の要件を満たしていること[3]

具体的にはテーブルとカラムの名称の付与方法が正しいこと、同じ対象を表すカラムは同じ名称、同じ属性が定義されていることを採点する。^(*)

- (2) DBの中味の理解に有効な説明ドキュメントが充実していること。具体的には作成されたドキュメントの種類とその記述内容の充実具合を、以下の4つの観点から採点する。^(*)

- ①スキーマ構造の説明
- ②データの意味の説明（現実世界との対応関係）
- ③データの値の説明（値域、コード体系等）
- ④業務の説明（トラヒック、データの量／成長度）

5. おわりに

以上により4つの品質項目を評価することによって、DB設計の良否を客観的にかつ総合的に判断できるようになる。現在筆者らは本評価方法に基づいて、DB設計方法の効果を測定し本評価方法の有効性を検証中である。将来は既に運用中のDBについても適用できるよう、本評価方法を拡張していく予定である。

参考文献

- [1] Deutsch, M.S. and Willis, R.R.: "Software Quality Engineering: A Total Technical and Management Approach", Prentice-Hall, pp.46~50, 1988
- [2] Bertrand MEYER (二木厚吉監訳): "オブジェクト指向入門" アスキー出版局, p3~p9, 1990
- [3] 関根他: "体系的なDB構築のための用語辞書を用いたデータ標準化手法" 情報論文誌, vol.34, No.3, pp457~467, 1993