

SQL処理特性把握のための  
ベンチマークモデルの一提案

4K-6

飯塚 哲也、 黒岩 淳一  
NTT情報システム本部

## 1. はじめに

NTTでは、業務システムのマルチベンダ環境への移行を推進中であり、今後多様なRDBMS (Relational DBMS) を利用することになる。RDBMSは同一の問い合わせ要求を達成するに当たり、選択条件の指定方法やDB構成の差異による性能への影響が大きい。性能面の問題から単純に業務のアプリケーションを移行することはできない。このため、各ベンダのRDBMSのSQL処理特性を把握し、業務システムの設計に個々の処理特性を反映し、性能面の問題を回避、解決する手段を確立することが重要な課題である。

性能、コストパフォーマンスを得る手段としてベンチマークモデルがあるが、TPC-A/B/C等の既存のベンチマークモデルでは、RDBMSのSQL処理特性を把握するには不十分である。そこで我々は、ベンチマークモデルによる実測結果からSQL処理特性を推測可能とすることを目的としたベンチマークモデルを作成し、その有効性を確認した。

本稿では、このようなベンチマークモデルの特徴、作成に際しての考え方、および効果の例を示す。

## 2. ベンチマークへの要求

ベンチマークに対する我々の要求として大きく以下の2つがある。

①コストパフォーマンスの良いDBMSを如何に選ぶか。

②DBMSの持つ性能を如何に有効に利用するか。

①は業務システムの開発においてDBMSを調達する際の要求である。②は性能問題を回避するためにAP、DBを設計する際の要求である。

①の要求に対する既存のベンチマークモデルで模擬している業務と実システムの業務との隔たりは大きい。ベンチマークモデルによる測定結果は、あくまでも目安として用いられる。マルチベンダ環境では事前に性能を有効に利用する方法を把握することは困難であるため、環境、条件の差異により必ずしも望んだ性能が出るとは限らない。そのため、業務システムの設計、運用の段階で性能的な問題が発生する恐れがある。そこで、我々は②の要求が非常に重要であると考えた。

## 3. 既存ベンチマークモデルの特徴

既存ベンチマークモデルの主目的は、前節で述べた①の要求を満たすことにある。その点で、それらの測定結果は単純に数値の比較を行うことができるため、コストパフォーマンスの良いDBMSを選ぶことが可能である。しかし、前述の②の要求を満たすには不十分である。

## 4. 本ベンチマークモデルの特徴と作成の考え方

## (1)特徴

RDBにおいて、AP (SQLステートメントの記述) 及びDBの設計次第で、同一の要求を実現するための処理性能が大きく変化することから、前述したように設計時にはRDBMSの性能を有効に利用することが重要である。そのため、本ベンチマークモデルでは、測定結果からSQL処理特性を把握し、性能上最適なDBMSへのアクセスを実現することにより前述の②の要求を満たすことを目的とする。この目的を達成するために本ベンチマークモデルは、SQL処理の性能に大きな影響を与えると思われる要因をパラメータとすることで、その要因が性能にどのような影響を与えるかを明らかにする。

## (2)作成の考え方

SQL処理の性能はAP及びDBの設計に多大な影響を受ける。つまり、SQLステートメントの記述及びDBの構成 (テーブルの構成、データ量等) に影響を受けるということである。そのため、SQL処理の特性を把握するためには、SQLステートメントの記述、DBの構成に関して、性能への影響が大きいと思われる要因を抽出し、それらの要因が性能にどのような影響を与えるかを調べればよい。

そこで、①SQLステートメントの記述、DBの構成に関して、SQL処理の性能に大きな影響を与えると思われる要因を抽出し、②その要因をパラメータとして、SQL処理の特性を把握可能なベンチマークモデルを作成した。①で抽出した要因を表1に示す。

## 5. ベンチマークモデルの例

本ベンチマークモデルの例を以下に示す。データ量に関する例としては、カラムの構成は同じでレコード件数の異なるテーブル、カラム数は同じでレコ

表1 性能に影響を与える要因

項目	要因	
DB構成	テーブルの構成	カラム数, 索引付与カラム値の並び, 索引構成カラム数, カラム値の並び, 索引付与カラム値の重複の有無
	データ量	レコード数, レコード長, ファイルサイズ
問い合わせ 検索系処理	選択	選択レコード数, 選択キーカラム値の並び, 選択条件カラムに索引が付与されているか否か
	射影	射影カラム数, 射影カラムの連続性
	ソート	ソートレコード数, ソート条件カラム値の並び, ソート条件カラムに索引が付与されているか否か, ソート条件カラム数
	集計	集計レコード数, 集計条件カラム値の並び, 集計条件カラムに索引が付与されているか否か
	結合	結合レコード数, 結合テーブル数, 結合条件カラムに索引が付与されているか否か, 結合条件カラム値の重複の有無
問い合わせ 更新系処理	追加	追加レコード数, 索引付与カラム値の重複の有無, コミットの契機
	変更	変更レコード数, 索引付与カラム変更の有無, 変更カラム数, コミットの契機
	削除	削除レコード数, コミットの契機

ード長の異なるテーブル等を定義した。また、テーブル構成に関する例としては、昇順の並びのカラム、ランダムな並びのカラム、単一のカラムから構成される索引、複数のカラムから構成される索引等を定義した。また、問い合わせの例を以下に2つ示す。

①レコード件数の異なるテーブルからレコードを1件選択する

この結果から、データ量の変化に対する特性を得る。

各問い合わせに対し以下の3パターンで測定する。

- A: 選択条件カラム(値の並びは昇順)に索引を付与
- B: 選択条件カラム以外のカラム(値の並びはランダム)に索引を付与
- C: 索引なし

②レコードの選択率を変え複数のレコードを選択し、ソートする(選択条件カラムとソート条件カラムは異なる)

この結果から、ソートレコード数の変化に対する特性を得る。

各問い合わせに対し以下の3パターンで測定する。

- A: ソート条件カラムに索引を付与
- B: 選択条件カラムに索引を付与
- C: 索引なし

## 6. 測定結果例と考察

### (1)測定結果例

5節で示した①、②の例を用いて、あるベンダのRDBMSを測定した結果を図1、図2に示す。

図1からは、以下の特性がわかる。

- ・選択条件カラム以外に索引が付与されているときより、索引をまったく付与していないときの方が性能が良い。つまり、索引を付与するカラムによ

っては、性能が悪くなる可能性があるということである。これから、選択処理では常に索引が付与されているカラムを利用して処理を行っていることがわかる。

図2からは以下の特性がわかる。

- ・選択条件カラムに索引が付与されているときより、ソート条件カラムに索引が付与されているときの方が性能が良い。これから、ソート処理ではソート条件カラムに索引が付与されているならその索引を利用して処理を行っていることがわかる。

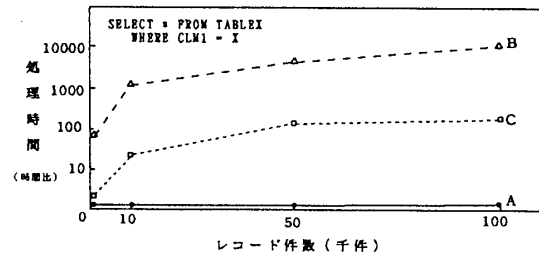


図1 例①の測定結果

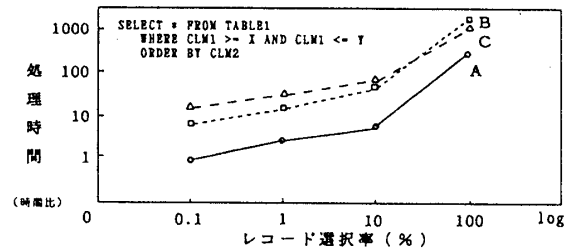


図2 例②の測定結果

### (2)考察

結果から、測定で用いたDBMSの特性を把握することができ、該DBMSを用いてAP及びDBを設計するときの注意点を得ることができた。1ベンダのDBMSの測定結果ではあるが、本ベンチマークモデルの有効性を確認することができた。

### 7. まとめと今後の課題

本稿で作成したベンチマークモデルを用いてRDBMSのSQL処理を測定することにより、その特性を把握できる見通しを得た。

今後の課題として以下を実施する予定である。

- ・複数のベンダのDBMSを測定することによるモデルの有効性の検証
- ・システム全体の処理特性を把握可能とするモデルへの拡張

### 参考文献

[1] Jim, G.: The Benchmark Handbook, Morgan Kaufmann, 1991  
 [2] Multivendor Integration Architecture, Version 1.1, 第2編 アプリケーション・プログラム・インタフェース仕様 第4部 データベース言語, TR550001-1, NTT, Apr., 1992