

SQL 処理特性把握のためのベンチマークモデルの効果

7W-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、基幹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、今後複数ベンダのRDBMS (Relational DBMS) を用いることになる。しかし、現在各々のRDBMSの性能特性を把握できていないため、業務目的に合致したベンダ選定や性能を意識したシステムの設計に不安がある。

既存のベンチマークモデルは、共通化した尺度、すなわち特定の条件によるRDBMSの性能、コストの比較を目的としているため、RDBMSのSQL処理特性を把握するには不適當である。そこで、我々は各ベンダのRDBMSのSQL処理特性を把握し、業務の目的に合致した調達や、業務システム設計に適用することを目的としてベンチマークモデルを考案し、[2]で報告した。本稿では、考案したベンチマークモデルに基づく複数RDBMSの実測による、本ベンチマークモデルの効果を報告する。

2. 本ベンチマークモデルの特徴

(1) 目的

既存のベンチマークモデルは、「コストパフォーマンスの良いDBMSを如何に選ぶか」という要求を満たすことを目的としている。しかし、これでは業務システムの目的に合致した調達や、設計、運用における性能的な問題の対処に利用することはできない。本ベンチマークモデルは、「DBMSの持つ性能を如何に有効に利用するか」という要求を満たすことを目的にし、実測した結果からDBMSの性能特性を把握可能とする。これにより、業務目的に合致した調達、各DBMSの性能特性を反映した業務システムの設計が行える。

(2) モデルの構成

以下のようにモデルを規定した。

- ・カラムの構成は同じでレコード件数が異なる複数のテーブルを規定
- ・SQL処理の基本的な選択、射影、ソート等の処理毎に性能に影響を与える要因をパラメータとし変化させるようにSQLステートメントを規定
- ・索引付与カラム値の格納順や索引構成カラム数を変化させるように、各SQLステートメントに対する実測時に各テーブルに付与する索引を規定

(3) 利用法

業務システムの開発のフェーズは大きく調達、設計、運用と分けられる。既存のベンチマークモデルでは、このうち調達時のコストパフォーマンスの良

いDBMSを選択する際に利用される。本ベンチマークモデルは先にも述べたように、より広い視点による調達、及び設計、運用時の性能問題の対処に利用することを考えている。

モデルの各SQLステートメントは、個々の基本処理の性能特性を実測可能なことから、各々の開発フェーズ、あるいは各々の業務システムにおける性能評価目的に合わせた選択組み合わせに対してのみ実測を行えばよい。表1に例を示す。調達時はII、IV、設計時はすべて、運用時はI、IV、Vに対して実測を行えばよい。

表1 SQLステートメントの例

項番	SQLステートメント	調達	設計	運用
I	SELECT * FROM xxx WHERE RANKEY1 = '000100'		○	○
II	SELECT * FROM TENKTBL WHERE SEQKEY1 >= '000000' AND SEQKEY1 <= 'xxxxxx'	○	○	
III	SELECT * FROM TENKTBL WHERE RANKEY1 >= '000000' AND RANKEY1 <= 'xxxxxx'		○	
IV	SELECT * FROM TENKTBL WHERE SEQKEY1 >= '000000' AND SEQKEY1 <= 'xxxxxx' ORDER BY SEQKEY1	○	○	○
V	SELECT * FROM TENKTBL WHERE SEQKEY1 >= '000000' AND SEQKEY1 <= 'xxxxxx' ORDER BY RANKEY1		○	○

3. 実測結果及び解析

本ベンチマークモデルの一部の以下のSQLステートメントで、2ベンダ(A社、B社)のRDBMSを実測した結果をグラフ1からグラフ5に示す。

I : 1レコード選択

選択条件カラムはランダムな並びでユニーク
SELECT * FROM xxx
WHERE RANKEY1 = '000100'

II : 複数レコード選択

選択条件カラムは昇順並びでユニーク
SELECT * FROM TENKTBL
WHERE SEQKEY1 >= '000000'
AND SEQKEY1 <= 'xxxxxx'

III : 複数レコード選択

選択条件カラムはランダムな並びでユニーク
SELECT * FROM TENKTBL
WHERE RANKEY1 >= '000000'
AND RANKEY1 <= 'xxxxxx'

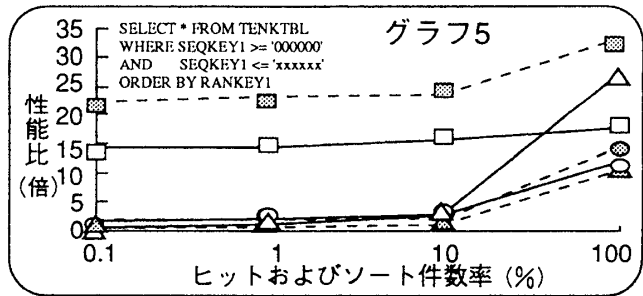
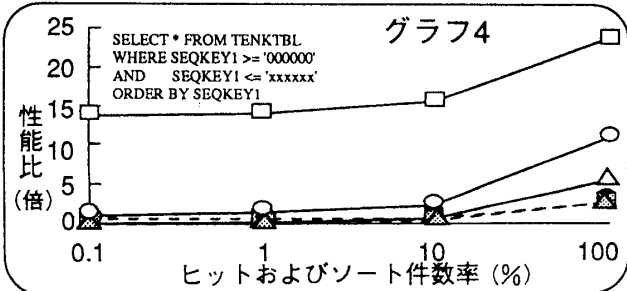
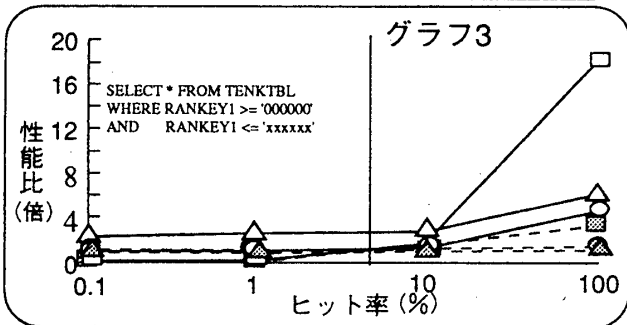
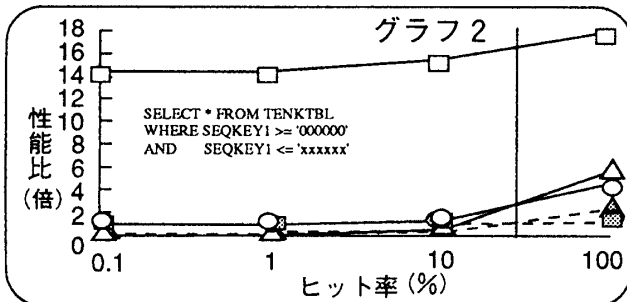
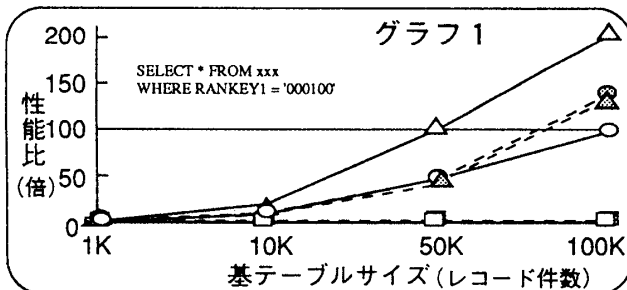
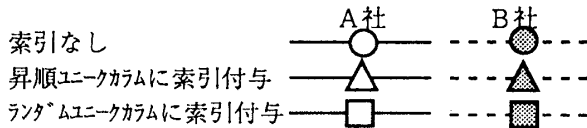
IV : 複数レコード選択、ソート

選択、ソート条件カラムは同一で昇順並びでユニーク
SELECT * FROM TENKTBL
WHERE SEQKEY1 >= '000000'
AND SEQKEY1 <= 'xxxxxx'
ORDER BY SEQKEY1

V: 複数レコード選択、ソート
 選択条件カラムは昇順並びでユニーク
 ソート条件カラムはランダムな並びでユニーク

```
SELECT * FROM TENKTBL
WHERE SEQKEY1 >= '000000'
AND SEQKEY1 <= 'xxxxxx'
ORDER BY RANKEY1
```

上記すべてのステートメント実測時に、アクセスするテーブルに索引を付与しないとき、昇順並びのユニークなカラム(SEQKEY1)に索引を付与したとき、ランダム並びのユニークなカラム(RANKEY1)に索引を付与したときの3つのパターンで実測した。各グラフの線の意味を以下で示す。



以降で上記グラフを解析した結果について述べる。
 グラフ1

A社: 選択条件カラム以外に付与されている索引は性能に悪影響を与える

B社: 選択条件カラムに索引が付与されていないときはアクセスするテーブルサイズ増に対する性能劣化の度合いが大きい (大規模時に要注意)

共通: 選択条件カラムへの索引付与は効果的でアクセスするテーブルサイズの影響はない

グラフ2, グラフ3

A社: 選択条件カラム以外に付与されている索引の性能への悪影響が大きく、ヒットレコード数増による性能劣化が比較的大きい

B社: 選択条件カラムに索引が付与されていないときはヒットレコード数の影響はない

共通: 効果的な索引もヒットレコード数が増すと逆効果となる (ランダム並びのカラムに索引を付与したとき要注意)

グラフ4, グラフ5

A社: ソート条件カラムと選択条件カラムが不一致のときヒット/ソートレコード数に応じて有効な索引が異なる (使い分けが必要)

B社: ソート条件カラムのみに索引を付与しているときヒット/ソートレコード数に関わらず性能が極めて悪化する

共通: ソート条件カラムと選択条件カラムが異なるときは同一の時に比べてヒット/ソートレコード数増に対する性能劣化の度合いが大きい

4. まとめと今後の展開

本稿で作成したベンチマークモデルを用いて複数のベンダのRDBMSのSQL処理を測定し、実測結果から性能の特性を得た。実測結果のグラフ及び解析からわかる様にベンダによる性能特性の違いが大きく、極めて重要な情報が得られる。

今後の展開として以下を実施する予定である。

- ・更に複数ベンダのRDBMSを測定し、各RDBMSの性能特性データを収集する。
- ・収集した性能特性データ、測定時の環境条件の分析により、各ベンダのRDBMSの最適利用法を整理する。

参考文献

[1] Jim, G: The Benchmark Handbook, Morgan Kaufmann, 1991
 [2] 飯塚哲也、黒岩淳一: SQL処理特性把握のためのベンチマークモデルの一提案、第47回情報処理学会全国大会、1993