

データベースの正規化による処理速度の検証

情報システム工学科 3年 74番 汲田章良

担当教員 笠原禎也助教授

使用した SQL の内容は次の 6 つである。

- ・ 地方名が中部地方のレコードを取得。
- ・ 地方名が中部地方で、月が 8 月のレコードを取得。
- ・ 地方名が中部地方で品目ごとの年別最高値を計算、取得する。
- ・ 地方名が中部地方で品目ごとの年別最安値を計算、取得する。
- ・ 地方名が中部地方で品目ごとの月別物価平均を計算、取得する。
- ・ 地方名が中部地方で品目ごとの年別物価平均を計算、取得する。

1. 目的

DB に関してよく正規化をすると効率がよくなるという話を聞くが、効率がよくなるならば、検索などに要する所要時間を短縮することができると思う。そのことを検証するために、この課題の目的として、正規化が行った DB と行っていない DB の検索の所要時間を計測し、正規化によって効率をよくすることができるのかを調べることにする。

2. 概要

Personal Computer (PC) に MySQL をインストール後に動作を確認する。次に、DB の規模を決定し DB を作成する。作成した DB を、正規化の有無で分け保存する。その後、複数の SQL を作成し、スクリプトを組み、その検索に要する所要時間を計測する。正規化の有無だけでなく、DB の使用目的を限定した場合に、完全な正規化を行わず、最初のテーブルに必要な属性を残した状態での検索に要する所要時間の計測も行う。正規化だけではなく、レコード数による DB の検索に要する所要時間も計測する。計測は 2 台の PC を用意し、VineLinux と MySQL を使用して、サーバとクライアントに分け LAN 経由でデータのやり取りを行う。

3. 研究内容

PC のスペックは、CPU が Celeron533MHz、メモリは 128MB である。

正規化とは、データの冗長性を少なくして、関連性の強いデータ項目をまとめて、DB の効率化を図るもので、主な正規形に第 1～第 5 正規形があり、本研究では第 1～第 3 正規形までを使用した。

DB のデータは、総務省統計局の小売物価統計調査の調査結果を使用した。このデータは、平成 12 年 1 月から平成 17 年 10 月まで HP 上で公開されており、その中の一部を使用した。DB の規模決定のために、テストとして 2000 件のデータを DB 化し、動作速度を計測したが、十分な時間が取れなかった。そこで、調整をし、288260 件の DB を作成した。

DB を作成後、正規化を行った。これは完全に正規化を行った DB と検索を行うときに“地方名”を条件に入れることとして、その検索に使用する項目は 1 つのテーブルに残して正規化を行った DB を作成した。

4. 結果

表 1：所要時間の結果

正規化			
SQL ファイル	最小値[s]	最大値[s]	平均[s]
price_reg1	11.69	12.56	11.91
price_reg2	30.05	30.28	30.11
price_reg3	39.58	39.77	39.66
price_reg2_local	10.43	11.19	10.65
price_reg3_local	10.46	11.02	10.63

(local は地方名等を 1 つのテーブルに残した DB)

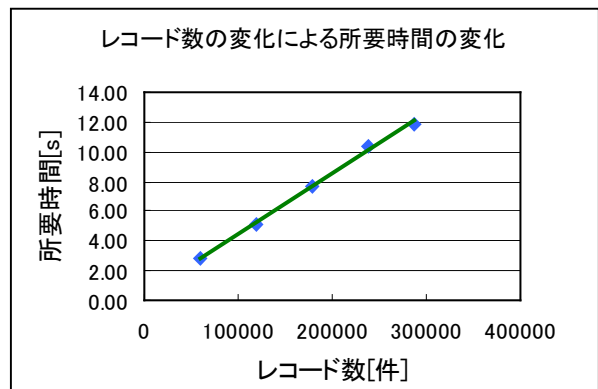


図 1：レコード数の変化による所要時間の変化

表 1、図 1 が結果であるが、所要時間は検索に 1 つのテーブルのみを使用した DB が一番速く、所要時間はレコード件数に比例していることが分かる。

5. 考察

正規化を行った DB の速度が遅くなったのには、検索方法が複数のテーブルを結合後に検索をかける SQL を使用したからと考えられる。そのため、サブクエリ等を用いて、テーブルを結合しない方法を取れば、結果も変わった可能性もある。レコード数に比例したグラフになったのは、レコードが増えてもレコードの処理のみが増えるだけで他の処理が加わることがないからだと考えられる。