

1. 序論

巨大なデータベースを検索する場合、1つのクエリの応答時間が非常に大きなものとなる場合がある。それは、SQL文の効率が悪い、あるいはインデックスの張り方が悪いなど、パフォーマンスチューニングの基本動作ができていない場合もあるが、そもそも「そんな巨大なクエリを1台のマシンでやらせることに無理がある」という場合もある。これに対する強力な解決策の1つがパラレルクエリである。

パラレルクエリとは、クエリを並列処理することにより、クエリの速度を向上させ、パフォーマンスを稼ぐ方法である。しかし、並列処理は均等に仕事を割り振らなければ、クエリの速度はうまく向上しないので、進捗よく状況のチェックや処理の分割をきちんとしなければならぬ。

本研究は、パラレルクエリを用いて、高負荷な処理のパフォーマンスを向上させていくことを目的とする。

2. 研究内容

本研究では、OS : Windows、Apache : XAMPP、Database : MySQL、テキストエディタ : TeraPad を使用して、データベースによるパラレルクエリを行なう。パラレルクエリの手順を以下に示す。

1. データの挿入

データ全体を2つに分け、各データをテーブル1及びテーブル2に挿入する。挿入後、2つのテーブルをマージしてテーブル0を作成する。このように、パラレルクエリを念頭に置きながら、分割した個々の部分を構築して、それをまとめ上げる。

2. 各テーブルへのクエリ処理

テーブル1とテーブル2それぞれに対してクエリを行なう。どこにクエリを投げるべきかの判断が複雑な場合もあるので、無駄にリソースを食いつぶさないため、必要十分な範囲にクエリを投げる。

3. 手分けして同時に処理

並列処理の主要な部分である。mysql コマンドをバックグラウンドで実行することによって、mysql コマンドを2つ同時に実行する。2つ同時に処理を行い、結果を作成した一時ファイルに追記し、プロセスの状態の管理を行なう。

4. 結果をまとめる

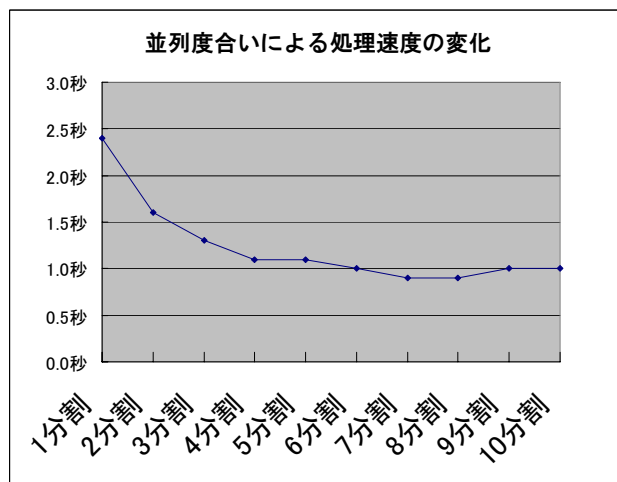
上記「3」のテーブル1とテーブル2の処理がすべて完了しているかをチェックし、一時ファイルの内容を合計し、結果を表示する。

以上の手順に沿って、膨大なレコードからなるテーブ

ルへのクエリを、「通常のクエリ」と「パラレルクエリ」の2通りを行い、それぞれの時間を比較する。

3. 結果・考察

「通常のクエリ」と「パラレルクエリ」との処理時間の比較の結果、1台のマシンでパラレルクエリを行うと、3割ほどパフォーマンスが改善した。この結果は、2分割のパラレルクエリだが、これを3分割・4分割としていくと、更にパフォーマンスが高まった。しかし、5分割・6分割のクエリでは、4分割と比べて、パフォーマンスはあまり向上しなかった。



個々のステップを並行処理にすることで速度を高めることができるが、単独マシンでの計算速度には限界がある。出現頻度の高い値 (例えば、ある値の平均値など) をあらかじめ計算しておく、または計算するたびキャッシュしておくなどにより、シリアル処理の計算ステップ数を減らしていくしかない。

4. 結論

1台のマシンで、どこまでパフォーマンスが上がるのかを調べていくとあるところで頭打ちになることがわかった。環境によって異なるので、今回と違うパターンもあると思うが、1台のマシンでのクエリでは、並列度「4」が1つの限界であるということがわかった。本当のパラレルクエリでは、複数のマシンで分散して、処理するのが普通なので、このパラレルクエリでは、高負荷になると安定しない。なので、実用に耐えることはできないので、まだまだ改良すべき点は多々あるだろう。

参考文献

- 1) Software Developer's Think IT :
(<http://www.thinkit.co.jp/>)