

GPU やマルチコア CPU を用いた際のデータベース検索性能の評価

先端工学基礎課程 学籍番号：1024004 成見研究室 阿部昭志郎

1. はじめに

近年、GPU を多目的に利用する GPGPU 技術が様々な分野で活用されている。現在、注目されているビッグデータを取り扱う情報分析用大規模データベース (DWH) 等でも、膨大なデータを分析する為の高速な検索性能が求められており、データベースの検索性能向上に GPU を利用することが有用な試みの 1 つとなっている。本研究では、データベース検索処理に GPU やマルチコア CPU を使用した場合の性能向上傾向を評価し、有効性を考察した。

2. GPU

GPU (Graphics Processing Unit) とは描画処理の高速化を目的とした 1 チッププロセッサのことである。一般的な CPU 単体のコア数は 4~8 コア程度であるが、GPU のコア数は数百~数千個存在する。このため、GPU を用いて大量のデータを並列に処理することで、CPU 単体で処理する場合よりも高速に演算を行うことができる。

3. データベース

データベースとは、一般的に「コンピュータで、相互に関連するデータを整理・統合し、検索しやすくしたファイル。また、このようなファイルの共用を可能にするシステム。」のことを指すものである。共有データとしてのデータベースを管理し、データに対するアクセス要求に応えるソフトウェアを DBMS (データベースマネジメントシステム) という。管理するデータの表現形式 (データモデル) によりいくつかの種類に分類でき、代表的なものには リレーショナル型、階層型などがある。現在最も広く普及しているのはリレーショナル型 (RDBMS) である。

4. PG-strom

PG-strom[1]は、PostgreSQL の外部データラップとして動作するモジュールである (図 1)。大量データの連続スキャン処理を加速する為に、GPU を利用するように設計されている。外部データラップとは、PostgreSQL に備わった機能拡張の仕組みの 1 つであり、外部データソースに対して標準的な SQL クエリで通常の表と同様に検索することを可能とするものである。

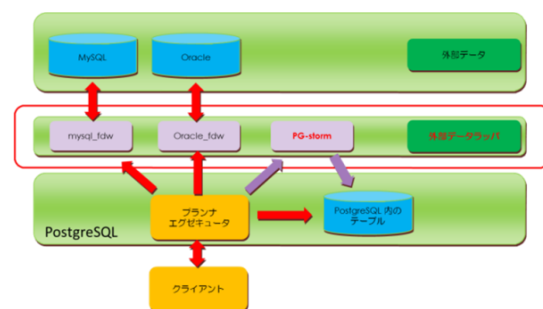


図 1 PG-strom の実装イメージ

5. Stado

Stado[2]は、PostgreSQL において、パラレルクエリを実現するミドルウェアである。PostgreSQL は、マルチコア CPU を実装したマシンを使用した場合でも、1 クエリの処理をマルチコア CPU で並列分散処理することができないが、Stado を使用することにより並列分散処理が可能となる。

Stado による並列分散処理には、パーティショニングと呼ばれる技術が使用されている。パーティショニングとは、図 2 に示すようにデータソースを複数のデータベースに分散化して格納し、クエリ処理を並列に実行することにより高速化させることが可能となる技術である。

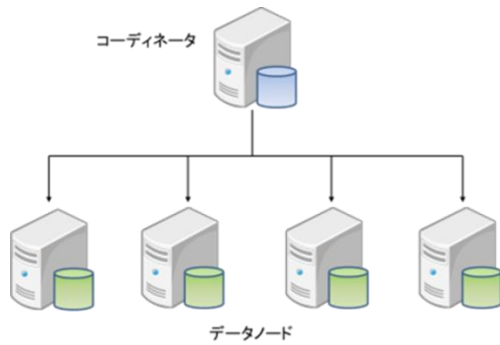


図2 データソースの分散

6. 性能評価

本研究では、DBMSに PostgreSQL、GPU アクセス用の外部モジュールに PG-Strom を利用。このシステム構成で、データベースのアクセラレータとして GPU を利用した場合の検索性能を観察する。また、マルチコア CPU を利用したデータベースの検索性能の観察には、パラレルクエリを実装するミドルウェア Stado を利用した。

6.1 GPU による検索性能

GPU による検索処理性能測定結果を図3に示す。データベース検索処理に GPU を利用した場合、約 30 倍程度の性能向上を観察した。但し、検索対象レコードが少ない条件では、CPU 単独処理の方が処理時間が少ない結果である。

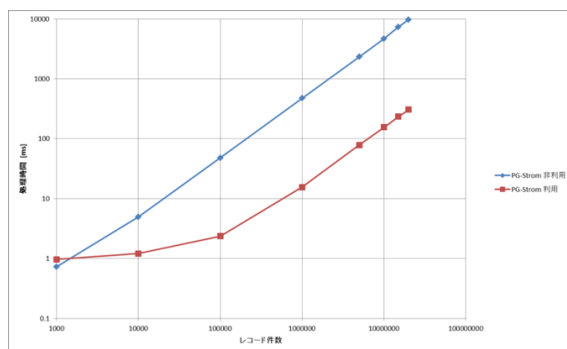


図3 GPU による検索性能

6.2 マルチコア CPU による検索性能

マルチコア CPU による検索処理性能測定結果を図

4に示す。データベース検索処理にマルチコア CPU を利用した場合、約 3~4 倍程度の性能向上を観察した。

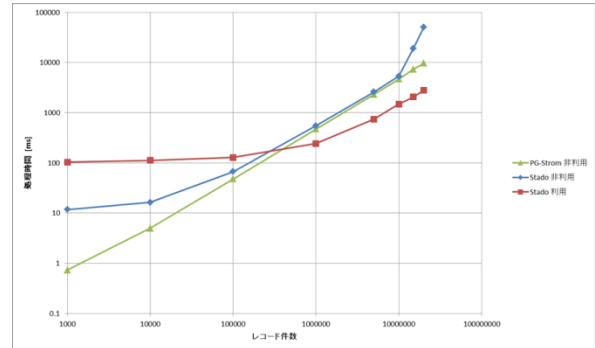


図4 マルチコア CPU による検索性能

7. まとめ

GPU およびマルチコア CPU による処理時間測定結果から、全体を通して検索対象量や検索条件量が多い条件においては、コア数の多い GPU に優位性があることが観察できる。一方で、検索対象量や検索条件量が少ない条件では、並列演算を実行する為の前段処理のオーバーヘッドにより、GPU やマルチコア CPU の優位性を発揮できない結果となることがいえる。

データベースの性能向上に GPU やマルチコア CPU を利用することは性能評価結果から有効性のあるものと考えられるが、現状の DBMS では、GPU やマルチコア CPU の有効性を発揮する為にデータベース自体の構成を変更しなければならない。今後、一般に利用される DBMS が、GPU やマルチコア CPU に対応することがデータベース分野で複数コアによる並列実行処理技術が利用される為に必要であると考ええる。

参考文献

[1] PG-Strom

<http://gpucscience.com/software/postgresql-gpu-pgstrom>

[2] Stado

<https://launchpad.net/stado>