

## 論文の内容の要旨

論文題目 並列データ処理系におけるアウトオブオーダー型実行方式に関する研究

氏名 山田 浩之

計算機システムを構成するハードウェア技術の潮流を見ると、プロセッサコアの動作周波数の向上は2008年からほぼ停滞し、また、磁気ディスクドライブのレイテンシの削減は年率5%以下に留まっており、当該傾向はメモリモジュールならびにネットワーク装置を構成するハードウェアにおいても見られる。即ち、これらの単一のハードウェアから構成される計算機システムにおいては、今後著しい性能向上は期待できないと考えられ、計算機システムの高性能化の実現には、複数のハードウェアを集約することが必須となりつつある。企業においてはハードウェアを高密度に集積した単一の大型システムが広く利用されている一方で、複数のコモディティサーバを高速なネットワークで接続したクラスタシステムが利用されるケースも少なからず見られ、昨今のビッグデータ解析と称される超大規模データの解析においては、価格性能比の点から、後者が進展しつつあることが伺える。クラスタシステム上で高速にデータ処理を行う並列データ処理系が産業界・学术界から次々と発表・提案されるに至っており、並列データ処理系の高速化はビッグデータ解析における一つの鍵になると言えるだろう。

本論文では、並列データ処理系の飛躍的な高速化を実現すべく、喜連川らが関係データベースシステムにおいて考案したアウトオブオーダー型実行方式を拡張し、並列データ処理のアウトオブオーダー型実行方式なる、並列データ処理の新しい実行方式を提案する。当該実行方式に基づく並列データ処理系では、クラスタシステムの各々の計算機において、並列データ処理の実行時にタスク分解を行い、分解されたタスクにおいて自らの二次記憶ならびにネットワークを介した他の計算機の二次記憶への入出力を行い、関連する演算を実行する。即ち、並列データ処理における入出力全体の非同期化を行う。データインテンシブな並列データ処理においては、入出力に性能が律速されることが多く、当該入出力を非同期化することにより、従来型の処理系に比して、飛躍的な高速化が期待される。

並列データ処理系においては、1980年代におけるハッシュクラスタリングアルゴリズムの開発以降、アルゴリズムの工夫によって基本的なファイルアクセスをシーケンシャルアクセスとすることにより、多数の磁気ディスクスピンドルを並列駆動

してスループットを高めるアプローチが広く採用されてきた。当時から磁気ディスクドライブのレイテンシの削減は停滞している一方で、記録密度の向上は指数関数的に増大していることから、今日においても同様のアイデアは広くみられる。オープンソースの並列データ処理系であるHadoopや、商用並列データ処理系であるNetezzaおよびVertica等もその一端をなしており、ファイルアクセスとしては全体走査を基本としている。当然のことながら、このような全体走査は必ずしも常に効率的であるわけではない。一般に、アクセスパスという観点では、ファイルの多くの部分のデータに対してデータ処理を行う場合、即ち、データ処理の選択性が低い場合は、ファイル全体を読み出す全体走査が有効である。反対に、ファイルのごく一部の部分のデータに対してデータ処理を行う場合、即ち、データ処理の選択性が高い場合においては、索引を始めとする構造化データを用いた選択的アクセスが有効である。選択的アクセスにおいては、商用の並列データ処理系等において実装されているものの、著者の知る限り、その有効性に関してはごく少数の文献で議論がされるに留まっており、特に解析的なデータ処理については、十分に活用されてきてはいなかったと見られる。しかしながら、ビッグデータ解析においては、並列データ処理系が抱えるデータセットの量は格段に増加する傾向にあり、当該傾向においては、ファイル全体を読み出す全体走査が可能な機会は限定的とならざるを得ず、また、エビデンスドリブンな意思決定支援を実現するためには、現状の所謂ビジネスインテリジェンスツールと比べて遙かに深度の高い解析が求められるようになると見られ、インタラクティブかつアドホックに対象空間をドリルダウンしながら発行されるデータ処理を機動的に実現することが求められると見られる。これまで並列データ処理系においては選択的アクセスが有効な領域は必ずしも顕著ではなかったが、当該領域は今後拡大する方向にあると言える。本論文で提案する並列データ処理のアウトオブオーダ型実行方式は、このようなデータセット空間の一部のデータを選択的にアクセスする場合に特にその効果を発揮し、当該アクセスの飛躍的な高速化を実現するものである。

本論文では、著者が試作を行ったHadoopをベースとするアウトオブオーダ型並列データ処理系Hadoopdeの構成法を示すとともに、128台の計算機からなるクラスタマシンにおいて人工データセット、ならびに実データを用いた当該試作による性能評価実験を示し、その有効性を明らかにする。

一方、並列データ処理系が保持するデータセットの特性を見ると、商品カタログ、RDF、センサデータを始めとし、多数の属性から構成されているものも少なくない。当該データに対するデータ処理においては、必ずしもすべての属性を必要とするとは限らず、むしろ、単純な集約問合せ等では、小数の属性を参照して解析を行うケースが少なくないことから、並列データ処理系においては、データを行形式ではなく列形式で二次記憶装置に格納する取り組みが見られ始めている。加えて、多数の

NULL値を持つ疎なデータセットにおいては、当該データを列形式に圧縮して保存することで、データの管理に必要な記憶空間を大幅に削減する取り組みも見られつつある。

このような列指向データレイアウトは、基本的には、予め想定する問合せに対してデータ配置の効率化を図るものであると考えられるため、必ずしもすべての問合せに対して有効に機能するわけではない。例えば、問合せの実行時に列形式のデータから行形式のデータ（タプル）を構築する必要がある場合、多数の属性への参照を必要とする問合せでは、当該処理において多くの入出力アクセスが発行される可能性があり、データ処理時間の大幅な増加が懸念される。列指向データレイアウトに対する問合せ処理においては、当該ケースを始めとし、入出力を非同期化し、入出力スループットを向上させることにより、問合せ処理の性能向上が期待できるケースが少なからず存在すると考えられる。

本論文では、前述の提案に加えて、列指向データレイアウトにおける問合せ処理の飛躍的な高速化を狙うべく、列指向データレイアウトに適応したアウトオブオーダー型実行方式を提案する。本論文では、著者が試作を行った列指向並列データ処理系の構成法を明らかにするとともに、72ドライブを具備する単一計算機、および128台の計算機からなるクラスタマシンにおいて当該試作機を用いて行った性能評価実験を示し、その有効性を明らかにする。

以上をまとめると、本論文は二つの主要なデータレイアウトにおける並列データ処理のアウトオブオーダー型実行方式を提案し、その有効性を明らかにするものである。即ち、本論文の前半においては、並列データ処理が対象とするデータセットが水平に分割（Horizontal Partitioning）され、複数の計算機に分配されたデータレイアウトにおけるアウトオブオーダー型実行方式を論じ、後半においては、データセットを垂直に分割（Vertical Partitioning）した列指向データレイアウトにおけるアウトオブオーダー型実行方式を論ずる。