

審査の結果の要旨

氏 名 山 田 浩 之

本論文は「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」と題し、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すると共に、当該実行方式に基づく並列データ処理系の構成法を明らかにし、試作実装を用いた性能評価実験によりその有効性を示しており、加えて、当該実行方式を列指向データレイアウトに対して拡張し、同じく試作実装を用いた性能評価実験によりその有効性を論じており、8章から構成されている。

第1章は序論であり、本研究の背景および目的について概観し、本論文の構成を述べている。

第2章は「関連研究」と題し、並列データ処理の実行方式ならびに並列データ処理系の構成法に関する研究に関して、現時点までに提案されている手法の特徴を纏めると共に、これらの問題点を論じることにより本研究との関連を明らかにしている。

第3章は「並列データ処理のアウトオブオーダー型実行方式」と題し、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ている。Hadoopをはじめとする並列データ処理系においては、基本的にデータセット空間全体の走査が前提とされていたのに対して、選択的なデータ処理、即ち、データセット空間の一部のデータを対象とするデータ処理についてはこのような全体走査は必ずしも効率的でないことを示し、当該データ処理の高速化を目的として、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ている。当該方式によれば、並列データ処理系の各々の計算機は、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うことが可能となり、即ち、入出力を非同期化することにより、並列データ処理にかかる入出力スループットを大幅に向上し、顕著な高速化が達成可能であることを示している。また、並列データ処理のアウトオブオーダー型実行方式を考慮した実行コストの見積りに基づく実行計画最適化アルゴリズムを提案し、これによりコストベースの実行計画最適化が実現されることを明らかにしている。

第4章は「アウトオブオーダー型並列データ処理系の設計と実装」と題し、オープンソースの並列データ処理系であるHadoopを基礎として、第3章で提案された並列データ処理のアウトオブオーダー型実行方式を実現した新たな並列データ処理系であるHadooodeならびにその上で機能するSQL処理系であるHive++の設計と実装を示し、当該並列データ処理系によって、MapReduce型の並列データ処理ならびにSQL記述の問合せを、アウトオブオーダー型実行方式によって実行可能であることを明らかにしている。

第5章は「並列データ処理のアウトオブオーダー型実行方式の評価」と題し、第3章で提案された並列データ処理のアウトオブオーダー型実行方式の有効性を検証すべく、128台の計算機から構成されるクラスタマシン（合計2048プロセッサコア、3072磁気ディスクドライブ）を構築し、当該環境において第4章で提案されたHadooodeならびにHive++を使用して行った並列データ処理系の代表的なベンチマークである解析タスクベンチマーク並びに代表的なデータベースベンチマークTPC-Hを用いた性能評価結果を示し、従来型の並列データ処理系と比して、特に選択性のある並列データ処理において、著しい高速化が実現可能であることを実証している。また、同一のデータ処理に対して複数の実行プランによる実行性能を計測し、第3章で提案された最適化方式によるコスト見積りと比較することにより、当該方式によって適切な実行プランの導出が可能であることを明らかにしている。加えて、実データを用いた性能評価として、厚生労働省から提供された匿名化済の我が国医療保険レセプトデータを用いた実験結果を示し、同じく提案手法の高速性を明らかにしている。

第6章では「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」と題し、並列データ処理系において実装が進められている列指向データレイアウトに対して、第3章で提案された並列データ処理のアウトオブオーダー型実行方式を拡張した手法を提案している。即ち、列指向データレイアウトにおいては、複数の射影列からデータを取得して元のタプルを生成するタプル再構築なる処理が必要となることが多く、当該処理に掛かる入出力がデータ処理の実行時間に与える影響が著しいことに着目し、当該処理をアウトオブオーダー型実行方式によって実行することにより、データ処理を高速化可能であることを明らかにしている。

第7章では「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の評価」と題し、第6章で提案された列指向データレイアウトに対する並列データ処理のアウトオブオーダー型実行方式に基づいた並列データ処理系であるCoooddeの設計と実装を示している。加えて、当該試作実装を用いて、72ディスクドライブを具備する単一計算機ならびに128台の計算機から構成されるクラスタシステムにおいて、センサデータ解析を模擬した基礎的なベンチマークによる性能評価結果を示し、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の高い有効性を明らかにしている。

第8章は結論であり、本研究の成果と今後の課題について総括している。

以上、これを要するに、本論文は、並列データ処理の飛躍的な高速化を目的とし、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と共に、当該方式を列指向データレイアウトに拡張してその有効性を実証しており、情報理工学上貢献するところが少なくない。

よって本論文は博士（情報理工学）の学位請求論文として合格と認められる。