

審査の結果の要旨

氏名 中野 美由紀

本論文は「大規模並列関係データベース処理における高速化技法に関する研究」と題し、大規模データを扱う関係データベースシステムの並列処理に関し、共有メモリ、分散メモリ、分散共有メモリ等の異なる並列計算機環境に適合した新しい処理方式の提案を行うと共に、試作機の開発、実機上での評価を行い、その有効性について論じており、8章から構成される。

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、「並列関係データベース処理と従来の研究」と題し、並列関係データベースシステムに関し、現時点までに提案されている並列処理化および高性能化手法の特徴をまとめ、並列データベースシステムの実例を紹介すると共に、従来手法における問題点を明らかにしている。

第3章は、「機能ディスクシステム」と題し、2次記憶システムへの関係データベース処理機能の有機的融合の試みについて論じている。即ち、関係データベース処理の性能ボトルネックとなるストレージ入出力に着目し、専用の高効率入出力ドライバ、ディスクからのデータ流に追従する動的クラスタリングならびにフィルタリングハードウェア機構、当該機構を活用するデータベース処理系などを新たに提案し、その有効性を示すべく試作機を開発し、1988年に当該機上にて性能評価を行い、従来関係データベースシステムと比較し、数十倍の性能向上が得られることを示している。

第4章では、「共有メモリ計算機における並列ハッシュ結合演算処理」と題し、共有メモリ計算機環境における大規模関係データベース処理における並列処理方式について論じている。即ち、入出力処理を司るプロセスとメモリ上のデータ処理を司るプロセスを明確に区分、管理することにより、プロセッサおよびストレージの台数に比例して効率よく性能の向上を図る手法を提案している。商用共有メモリ計算機(Sequent S81)上に提案方式を実装し、1990年代初期に、高い並列処理効果が得られることを明らかにしている。

第5章では、「分散メモリ計算機における並列多重結合演算処理の最適化技法」と題し、分散メモリ計算機における主要な資源である、ネットワーク転送コスト、CPU処理コスト、2次記憶入出力コストを用い、いずれかがボトルネックにならないように、資源消費が均衡する最適実行プランの生成手法について論じている。分散メモリ計算機 IBM SP2 をモデルにシミュレータを生成し、提案手法により生成される実行プランが従来手法と比較して、優れたプランを導出可能であることを明らかにしている。

第6章では、「分散共有メモリ計算機におけるデータベース処理に適合したバッファ管理方式」と題し、分散化された共有メモリのアクセスタイムの特性を考慮した並列結合演算処理方式を提案

している。即ち、一般に意思決定などにおける問合せ処理ではデータ局所性が殆ど無いと見做されているのに対し、ハッシュ分割されたクラスタデータ並びにハッシュテーブルのアクセスにおいては、高いデータ参照局所性が存在することに着目し、局所性を利用する新しい手法を提案すると共に、当該手法により大きく性能向上可能であることを明らかにしている。加えて、商用分散共有メモリ計算機が登場した 90 年代後半に他に先駆けて実装し、その有効性を示している。

第 7 章は、「GN ハッシュ結合演算処理」と題し、ハッシュ関数に基づいた関係データベース並列処理技法と旧来のネストループ処理方式の処理コストを解析的に推定・比較し、実行時にコストの小さい方式を選択する新しい方式 (GN ハッシュ) を提案している。ハッシュ結合方式に比べ、当該手法はとりわけデータの偏りが大きな場合に性能劣化が少なく有効であることを明らかにしている。

第 8 章「結論」では、本論文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、大規模データを扱う関係データベースシステムの並列処理技法に関し、高度なストレージ入出力制御手法を導入した機能ディスクシステムを開発し、加えて、1980 年代後半から 2000 年までに大きく進展した並列計算機システム技術に着目し、共有メモリ、分散メモリ、分散共有メモリなる代表的な並列計算機に適合したデータバッファ管理手法、並びに、関係演算並列化手法について新しい提案を行うと共に、実機上への実装および詳細なシミュレーションを通じて提案手法により大幅な性能向上が得られることを他に先駆けて明らかにしており、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）の学位論文として合格と認められる。