

氏 名 目 木 信 太 郎
学位(専攻分野) 博 士 (工 学)
学 位 記 番 号 論 工 博 第 3438 号
学位授与の日付 平 成 11 年 3 月 23 日
学位授与の要件 学 位 規 則 第 4 条 第 2 項 該 当
学 位 論 文 題 目 Studies on Main Memory Database Systems for Vector Processors
(ベクトル計算機に対する主記憶データベースの研究)

(主査)

論文調査委員 教授 上 林 彌 彦 教授 富 田 眞 治 教授 湯 淺 太 一

論 文 内 容 の 要 旨

データベースの応用分野の拡大にともない、より高速なデータベースシステムに対する要求はますます大きくなりつつある。これまで高速なデータベースシステムを実現する方法として、並列データベースと主記憶データベースが利用されることが多く、これらは大量生産されているLSIを使っているの、経済的に実現することができる。一方、大規模数値計算に対する要求に応えるため、ベクトル計算に基づくスーパーコンピュータが開発され、実用に供されている。ベクトル計算機は強力なパイプライン演算装置と大容量の主記憶を備えており、データベースシステムのハードウェアプラットフォームとして有望であるが、その性質を利用したデータベースシステムの研究はほとんどなされていなかった。

本論文はこのような背景に基づき、関係データベース演算をベクトル計算機上で高速に処理する手法に関して研究を行なった成果をまとめたものであり、全体で9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意義を述べるとともに、本論文で提案するベクトル計算機上の主記憶データベースとこれまでの研究との違い、および論文構成を示している。

第2章では準備として、本研究で議論の対象としている関係データベース演算を定義し、ベクトル計算機の概要についてそのハードウェアアーキテクチャとベクトルプログラミングの2つの点から述べている。

第3章では、関係データベース演算に限らず、データベースシステムにおいて広くデータ処理に使われるソーティングについて、ベクトル計算機上での性能評価をおこなっている。データ数、キー長、キー値の分布といったサンプルデータの性質を変化させて網羅的なベンチマークを行なうことによって、これらの性質がソーティングアルゴリズムの性能に与える影響について明らかにしている。

第4章では、選択、射影、結合の3つの関係データベース演算について、そのベクトル化アルゴリズムを提案している。選択演算はベクトル計算機の持つベクトル収集命令で実行することができ、射影演算と結合演算はハッシュ法に基づいたアルゴリズムで実行することができることが示されている。ハッシュ法を用いる場合にはハッシュ値の衝突の取り扱いが問題となるが、一旦衝突を無視してハッシュテーブルに書き込みを行ない、上書きされたかどうかを後から調べることによって、これを解決している。

第5章では、関係データベース演算のアルゴリズム以外に、プロトタイプシステムを作成する上で必要となる、関係を主記憶上に割り当てる方法、質問処理の過程の中間結果の記憶方式、ハッシュ表の大きさ、文字列データの取り扱いといった

いくつかの点について考察を加えている。ハッシュ表の大きさに関しては、計算機を用いた実験を行ない、提案した評価式がこれと一致することを確認している。

第6章では、提案したベクトル化アルゴリズムをプロトタイプシステム上に実現し、性能評価を行なった結果について述べている。組数10万のウィスコンシンベンチマークにおいて、実行した4つの質問すべてが0.1秒以内に処理され、スカラーアルゴリズムと比較しても10倍以上の加速率が達成できたことが示されている。

第7章では、索引構造について考察を行なっている。AVL木、T木といったこれまでに提案された索引構造について述べた後、索引を用いた質問処理はベクトル処理に不向きなことを示している。

第8章では、関連する研究について述べている。本研究は、関係データベース演算をベクトル処理するという側面と、ベクトル計算機を数値計算以外に応用するという2つの側面があるが、そのそれぞれについて関連する2つの研究、データベースプロセッサIDPと記号処理ベクトル化のためのFOL法について述べている。

第9章は結論であり、本研究で得られた結果を総括している。

論文審査の結果の要旨

ベクトル計算機は大規模数値計算のための強力なパイプライン演算器と大容量の主記憶を備えており、高速なデータベースシステムを実現するためのハードウェアプラットフォームとしても有望であるにもかかわらず、ベクトル計算機の性能を十分に引き出すためのパイプラインアルゴリズムが開発されていなかった。本論文は、ベクトル計算機に適したハッシュ法に基づく関係データベース演算アルゴリズムを提案しており、得られた主な成果は次の通りである。

1. ベクトル化ソーティングアルゴリズムの網羅的なベンチマークを行ない、ソートするデータの性質がアルゴリズムの効率に与える影響を明らかにした。データ数、キー長、キー値の分布といった性質に応じて高速なソーティングアルゴリズムが異なることを示した。
2. 大容量の主記憶が利用できる場合には、質問処理のアルゴリズムとしてハッシュ法が有効であることが知られているが、ハッシュ値の衝突の処理は逐次的に行なう必要があるため、そのままでは正しくベクトル処理することができない。ハッシュ表に書き込みを行なう際に、一旦ハッシュ値の衝突を無視して書き込みを行ない、上書きされたものに対して後で例外処理を行なうことで、効率良くベクトル処理を行なう方法を提案した。
3. 上記の手法を用いる場合に、ハッシュ値の衝突を無視して処理を行なうことによるオーバーヘッドが問題となる。そこで、ハッシュ値の衝突によって上書きされるものの数を評価する式を導き、オーバーヘッドの定量的な評価を行なった。さらに、計算機を用いた実験を行ない、得られた結果が評価式と一致することを示した。
4. 選択演算や、最大値・最小値を求め全集約演算は逐次的に処理を行なう必要があり、ハッシュ法でもうまく処理することができないが、ベクトル計算機が備えているベクトル収集命令や、最大値・最小値を求める専用の命令を利用することによって10倍以上の加速率で処理できることを示した。

以上要するに本論文は、ベクトル計算機の持つ強力な演算能力と大容量主記憶を生かし、特殊なハードウェアを導入しないでソフトウェア的な方法だけで、関係データベースにおける質問処理を効率良く行なう手法を示したものであり、その成果は学術上、實際上寄与するところが少なくない。よって、本論文は博士（工学）の学位論文として価値あるものと認める。また、平成11年2月4日、論文内容とそれに関連した項目について諮問を行なった結果、合格と認めた。