

5H-4

並列ワークステーション上での
個人用データベースに関する研究

安藤誠 大森士郎 小沢英昭 安西祐一郎

慶応義塾大学 理工学部

1. はじめに

最近では、計算機の低価格化・高機能化により、オフィス・大学等にワークステーションが広く普及している。しかし、従来のものは、その利用度が高まるにつれて過負荷になりがちであり、ユーザからは「負荷の軽減」の要望が高まってきている。したがって、近い将来、オフィス・学校等ではシングルプロセッサのワークステーションに代わって、プロセッサ数が数個から数十個の並列ワークステーションが広まることが予想される。

さらに、オフィス内では、各労働者の個人所有情報(名刺・住所録等)をデータベース化しようとする傾向が見られる。情報のデータベース化には次のような利点があげられる。

- (1)情報の冗長性の低減
- (2)情報の無矛盾性・完全性の維持

しかし、以上の(1)(2)の機能を提供する代償として、以下の様な問題を抱えており、これらは従来のパソコンやワークステーションでは、大きな障害となっている。

- (i)処理負荷が大きい。
- (ii)情報の構造化によって、データ量が增大する。

本研究では、これらの問題を軽減させるために、並列ワークステーション上に個人利用データベースを構築する手法を提案する。

2. システムの特徴

前述の問題を軽減させるためには、次のような点が必要とされる。

- (1)従来のデータベースシステムを並列計算機に適した単位でモジュール化する手法。
 - 各モジュール単位にプロセッサを割り当てることにより、モジュールレベルでの処理の並列化が可能となる。
 - 各モジュールでの独立した事故復旧処理等、データ管理の簡素化が図れる。
 - 文字・数字データ専用のモジュールと、イメージ等のバイナリデータ専用のモジュールを設定することにより、両者のデータ処理の違いをモジュール内で吸収することが出来る。このことにより、ユーザは文字・数字データとバイナリデータの違いを意識することなく、データ処理が出来る。
- (2)2次記憶にアクセスする回数を減少させる手法。

本システムでは、これらの目標の実現に向けて、次のような方式を採用した。

2.1 概念・内部レベルのモジュール化

本システムは、関係型データベース(RDB)を基盤としている。一般にRDBは、図1に示すように3つの階層に分割できる。ここで、概念レベルにおける各表は「基底表」である。

本システムでは前述のモジュール化を、概念レベル以下において「基底表」を単位として行なった。図2における、ローカル概念レベルと内部レベルの組が、1つのモジュールである。この方式により、次のような利点を実現される。

- (1)基底表レベルでの処理の並列化が可能となる。これにより、複数の表にまたがる処理の場合、処理の並列化・負荷の軽減が図れる。
- (2)従来のデータベースでは、一部の表の事故がデータベース全体に影響する危険がある。しかし、本システムでは基底表単位でデータ管理が出来るので、あるモジュールでの事故は他のモジュールに影響を与えない。
- (3)文字・数字データを扱うモジュールと、バイナリデータを扱うモジュールをそれぞれ設定した。実際に内部レベルでは、両者は全く違う処理をしているのであるが、モジュールの

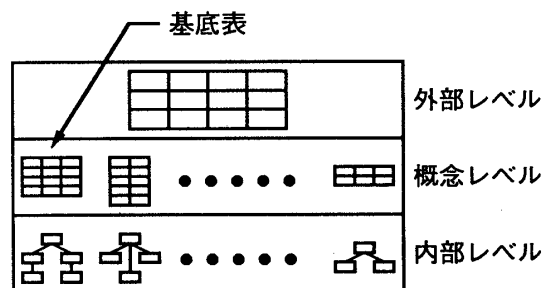


図1 RDBの3階層

A study of a personal database system on a parallel workstation.
Makoto ANDO, Shiro OHMORI, Hideaki OZAWA and Yuichiro ANZAI
Faculty of Science and Technology, Keio University

外側から見た場合には、両者の違いを意識する必要がなくなる。

2.2 キャッシュによるアクセスの高速化

本システムは、個人的に利用される情報のみを扱うことを前提としている。このことより、データアクセス頻度の分布には、その個人の特徴が現われるはずである。そこで、本システムではローカル概念レベルにソフトウェアによるキャッシュを採用し、アクセスの高速化を図った。

本システムでは、各表のタプルのうち特定のフィールドに関してキャッシングしている。この方式を採用した理由は、本システムが個人的な利用を対象としたものだからであり、「Aさんの住所」というような検索が比較的多い場合には高い効果が得られると予想される。

3.概念レベルの機能

本研究では、システムのモジュール化を行なうために概念レベルをグローバルな部分とローカルな部分に分けた。以下では、各概念レベルの持つ新しい機能について述べる。

3.1 グローバル概念レベル

本システムでは、従来のパソコンやワークステーション上のデータベースシステムと違い、各モジュールのプロセッサへのマッピングを行わなければならない。この処理は、このレベルで行なう。現段階では、プロセッサの個数による基底表の個数の制限をなくすため、1モジュールに1プロセスを対応させる方式を採用している。

一方で、このレベルには外部レベルから受け取った利用者からの問い合わせを、各基底表に対応したローカルな問い合わせ及び複数の基底表にま

たがる処理を含むグローバルな問い合わせに分解する機能がある。分解されたローカルな問い合わせは、それに対応するモジュールに渡される。また、各モジュールから処理された表を受け取った後、それらの表にグローバルな問い合わせに対応する関係演算を行なう。

3.2 ローカル概念レベル

このレベルでは、グローバル概念レベルから受け取った問い合わせを処理する。この処理は、一般のRDBの概念レベルと同じである。

4.おわりに

本システムは現在、SUNワークステーション上でシミュレーションを行ない、評価をとっている。今後の課題としては、次のようなことがあげられる。

- (1)並列計算機への実装をする。
- (2)現在のシステムでは、プロセッサ数よりもモジュールが多い場合の考慮がされていない。負荷分散のためには負荷を均等にするためのアルゴリズムが必要である。

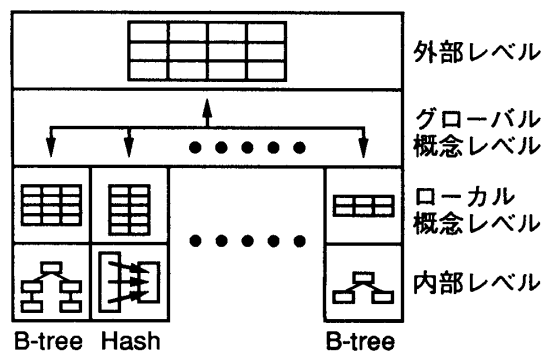


図2 本システムのプロット