

スーパーデータベースコンピュータ SDC2 のソフトウェアシミュレータ

1F-3

北村 学 田村 孝之 中村 稔 喜連川 優 高木 幹雄
 東京大学 生産技術研究所

1 はじめに

スーパーデータベースコンピュータ SDC2 は、超高性能な、SQL による問い合わせ処理を目的として現在開発中の現在我々が開発中の高並列関係データベースサーバである。SDC2 は密結合の処理モジュール複数台を、データネットワークで疎結合した構成を採る。

実機でモジュールの並列度向上には限界があることから、さまざまな条件での性能測定を目的としたソフトウェアシミュレータを開発した。シミュレータの構成、ならびにその動作について報告する。

2 SDC2 の概要

2.1 ハードウェアアーキテクチャ

SDC2 は各モジュール内に最大でプロセッサ 7 台と磁気ディスク装置 4 台を密に結合した構成をとり、当該モジュールを高機能オメガネットワークにより疎に結合するハイブリッドアーキテクチャを採る。このオメガネットワークはモジュールの負荷を自律的に分散するバケット平坦化機能を有する。

2.2 SDC2 におけるデータの流れ

SDC2 ではデータメモリを固定長のページに分割し、モジュール内のデータ交換は全てこのページを受け渡すことで行う。またフリーページをモジュール内で一元管理して、ページが必要な時は必ずフリーページプールから取り出すようにして、データの受渡しを単純かつ統一的に扱えるようにしている。

また、データネットワーク間のデータのやりとりはタプルを単位として行い、ページ↔タプルの変換はネットワークインタフェース上で行う。

3 SDC2 のソフトウェアシミュレータ

SDC2 のソフトウェアシミュレータの実現には、

- モジュール内部のデータの流れ
- ネットワークインタフェースにおけるタプル単位のデータ転送

Software Simulator of the Super Database Computer (SDC2)
 M.Kitamura, T.Tamura, M.Nakamura, M.Kitsuregawa,
 M.Takagi

Institute of Industrial Science, University of Tokyo
 7-22-1 Roppongi, Minato, Tokyo, 106, Japan

- ネットワークの平坦化機能

を、並列に実行する必要がある。

また、シミュレータには、モジュール台数の増減、パラメータの変化などさまざまな条件が実現できるとともに、プログラムの書き換えが容易であることも要求される。そのため、プログラムはモジュール化されるとともに、拡張可能な構成でなければならない。

そこで今回は並列プログラムをシミュレートするライブラリソフトである YACSIM を利用してシミュレータプログラムを作成した。

3.1 モジュール内構成

図 1 にシミュレータの構成を示す。シミュレータで

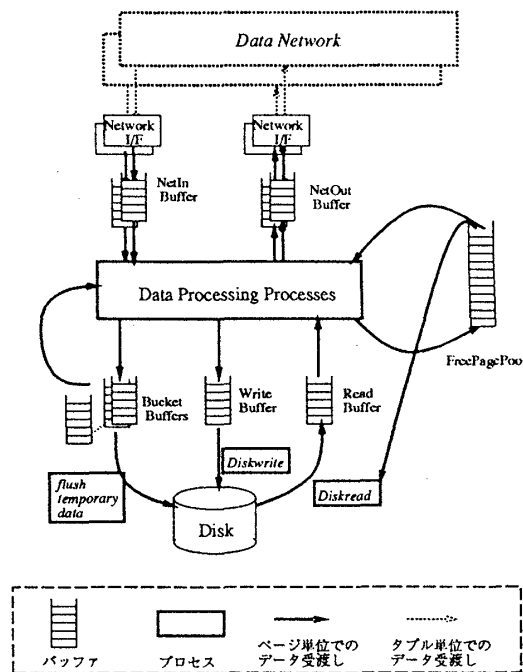


図 1: SDC2 シミュレータの内部構成 (1 モジュール内)

取り扱われるデータにはタプルとページがある。ページはタプルのリストと機能 ID を持った構造体で、多くの場合、データの受渡しにはページが用いられる。全てのページは、最初は Free Page Pool に存在する。タプルはキー値、タプル長、タプル ID、リレーション ID、タプルの宛先を持つ構造体で、ディスクから読み出されてページに入れられる。

ページ、タプルを処理するルーチンをプロセスと呼ぶ。プロセスは

1. ディスクの読み書き
2. ページの受渡し、ページ内タプルの関係演算処理
3. ネットワークでのデータ転送

を行う3種類のプロセスに分類される。

3.2 バッファ

シミュレータ上には、1モジュール内に5種類のFIFOバッファが用意される。ただし、データネットワークは2面あるために、NetIn/NetOut Bufferは2つずつ、またBucket Bufferは、分割バケットの数だけ用意される。

各バッファの大きさはいずれも可変で、フリーページプールにあるページを取り出して利用していく。

プロセスは、これらのバッファ間でのページのやりとりをする存在とみなせる。プロセスは、あるバッファから1ページを取り出して処理を行い、必要に応じてFree Page Poolからページを取り出して使用する。処理が終わると、新しくできたページは他のバッファに移される。また、タプルがなくなり不要になった空きページは、Free Page Poolに返却される。

3.3 ネットワーク

ネットワークインタフェースには送信側・受信側に1つずつプロセスが存在している。

ネットワークは、タプルに記された宛先にタプルを送信する宛先指定用のネットワークと、一旦全てのタプルを各モジュールに均等に分散する平坦化モードネットワークの2種の機能があり、それぞれ1面ずつ用意されている。

このシミュレータでは、平坦化モードはデータ転送の途中ではブロックが起きないという仮定を置いてシミュレートを行っている。

3.4 シミュレータの入力パラメータ

主なパラメータとして、

- モジュールの台数
- 1モジュールの保有するページ数
- ディスクのアクセス速度
- 入力リレーション

があり、これらを変化させて測定を行う。

3.5 出力

測定要素として、処理の経過時間とその時のメモリの使用状況が重要になる。このシミュレータでは処理の初めから終わりまでの各バッファの使用状況を保存して、その変化を測定することができる。

例として、

- モジュール数：8台
- タプル数(1モジュール)：12500×10000件
- 選択率：100%
- ディスクアクセス速度：4MB/sec
- モジュールクロック周波数：25MHz
- ネットワーククロック周波数：12.5MHz

の条件のもとで結合演算を行った時のバッファ使用状況のグラフを図2に示す。なお、この使用状況は実機での

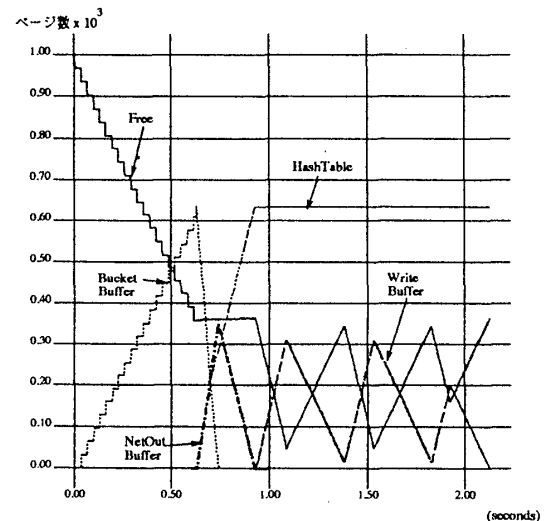


図2: 結合演算時のバッファ使用状況の変化

測定と比較し、ほとんど一致を確認することができた。

4 まとめ

以上のように、SDC2のシミュレーション環境を作成した。今後、このシミュレータを用いて、関係演算処理のモジュール台数の増減や、モジュール・ネットワーク・ディスクの速度の比に対する処理時間等の測定を行っていく予定である。

参考文献

- [1] 中村 田村 喜連川 高木: “スーパーデータベースコンピュータ第二版(SDC2)におけるデータ流制御の評価”, アドバンストデータベースシンポジウム, 1993
- [2] J.Robert Jump: “YACSIM Reference Manual”, Rice University Electrical & Computer Engineering Department, Mar. 1993.