4F-1

データベースマシン　GRACE

═ バケットのパイプライン処理機構 ═

喜連川　優　　田中　英彦　　元岡　達

（東京大学　　工学部）

1.　はじめに

GRACE［1］は、Hash とSort に基づく処理技法により処理負荷の重い関係代数演算を高速に処理する事が出来、Join の多用される環境に対しても高い性能が得られる点に特徴がある。今回はその中心機構であるバケットのパイプライン処理メカニズムについて報告する。

2.　バケットのパイプライン処理機構

GRACEではリレーションが複数のメモリモジュール上に並列格納されるが、その際、バケットが特定のメモリモジュールに対応するわけではない。図1に示される如く、あるバケットに属するタプルは複数のメモリモジュールに分散され、プロセッサはこれらをパイプライン的に処理してゆく。パイプのセグメントは各々のメモリモジュールに相当しプロセッサがこれらを順次通過してゆくことになる。当該バケットのタプルはモジュール毎に含まれる数が異なる為、又バケット自体の大きさが異なる為、パイプのセグメントタイムは動的に変化することとなり、パイプラインの擾乱が性能の低下を招くと予想される。

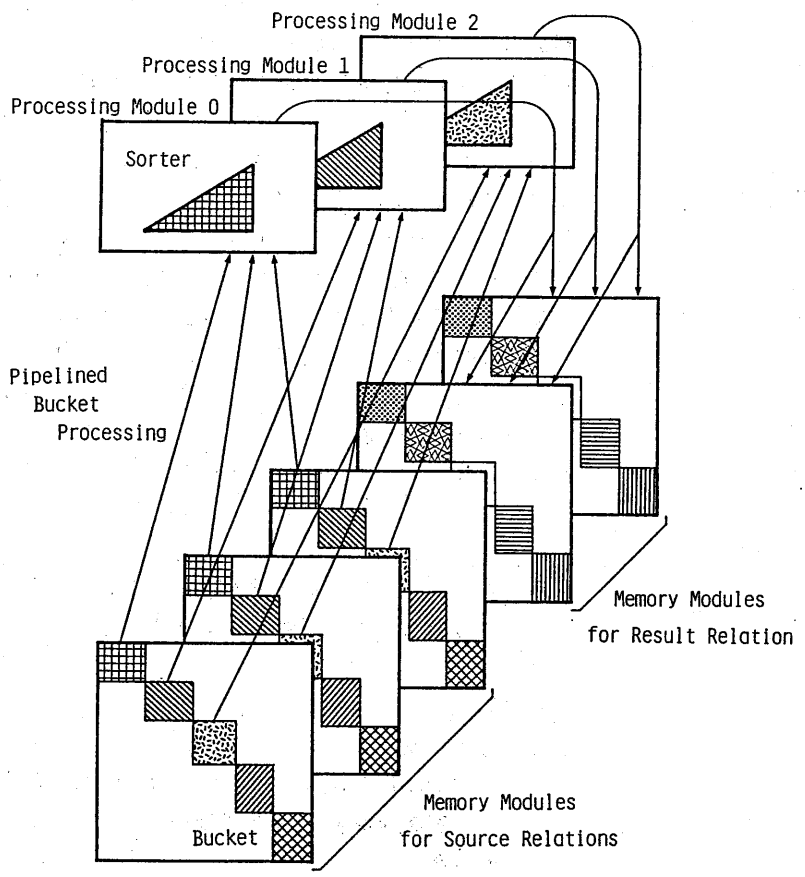


Figure 1　Conceptual View of Pipelined Bucket Processing

2.1　バケットサイズのゆらぎによるパイプラインの擾乱

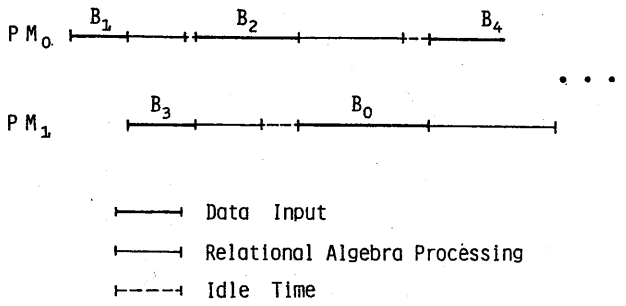
簡単の為、メモリモジュールが一台から構成される場合について考えると、全てのバケットの大きさが等しい場合には2台のプロセッサによって無駄のない完全なパイプラインが形成できるのに対し、一般の場合には、図2に示される如く、セグメントタイムのゆらぎにより、メモリ、プロセッサ両者に待ち時間が生ずる事となる。ここで、プロセッサはデータ流に沿った処理が可能なハードウェアソータを内蔵しており、1つのバケットの処理はその大きさを nとすると、2×n に比例した時間で処理可能としている。

2.2　スケジュールされたバケットのパイプライン処理

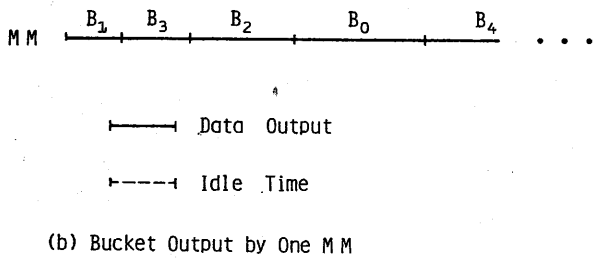
GRACEの処理対象はバケットシリアルなデータ流でありこの流れの速度によりシステムの性能が決定される為、メモリモジュールからの流れを出来る限り乱さずに処理をすることが望まれる。プロセッサの待ち時間よりも寧ろメモリのそれを最小化することが重要となる。図3は図2と同じバケット流に関し処理順序を変えた場合を示し、バケットを大きさの昇順に処理することによってメモリモジュールに於ける無駄時間をなくし効率よく処理できることがわかる。バケットサイズのゆらぎは性能に影響しない。

2.3　バケットの平坦化

複数台のメモリモジュールから構成される場合にも、全てのバケットがモ

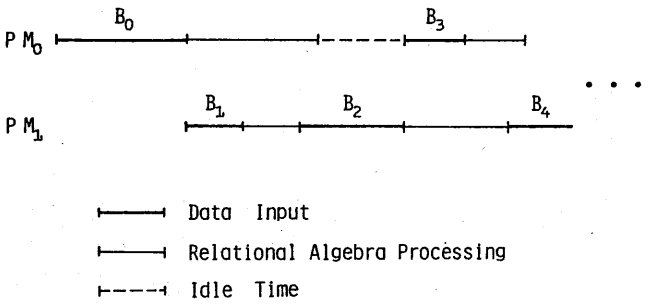


(a) Pipelined Bucket Processing by Two P M's

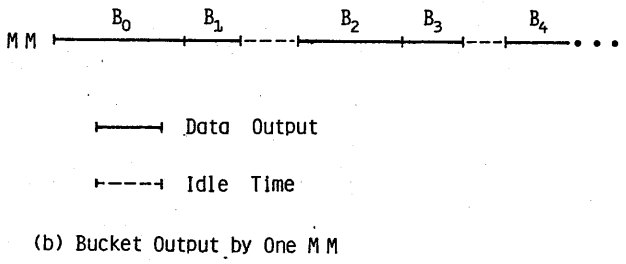


(b) Bucket Output by One M M

Figure 2. Bucket Processing Pipeline with One Memory Module
(Pipeline Disturbance Due to Bucket Size Fluctuation)



(a) Pipelined Bucket Processing by Two P M's



(b) Bucket Output by One M M

Figure 3. Scheduled Bucket Processing Pipeline
for The Data Stream in figure 2

ジュールにわたって均等に分配される時には先の方策が適用可能である。実際にはバケットのモジュール間にわたるゆらぎが生じる為、パイプラインは大きく乱れることになる［２］。ゆらぎを考慮に入れた最適バケット処理のスケジュールは困難であり、ここでは、バケットの平坦化処理を組み込むこととした。ＧＲＡＣＥではモジュール間結合網の１つとして多重チャネルリングバスの採用を考えているが、以下の如き分散化された手順によってリングバス上でバケットの平坦化を実現している。即ち、送信モジュールは生成タプルにハッシュを施しチャネルに送出し、受信モジュールは送られてくるタプルの中で最も適切なものを取り込む。ここで各モジュールはバケット毎に取得タプル数をテーブルに管理しているものとする。受信モジュールは送られてくるタプル群から（ＭＡＸ－Ｎ）／（Ｎ－ＭＩＮ）が最大となるタプルを取り込む。ここで、Ｎは当該モジュールの有する対象バケットのタプル数、ＭＡＸ（ＭＩＮ）はそのバケットのタプルを最も多く（少なく）含むモジュールのタプル数である。リングバスの為位置的優先順位が存在するがシミュレーションによって調べた結果、モジュール間のゆらぎはほとんどの場合高々３タプルにとどまることが判った。尚、実際のタプル取り込み動作と次タプルに対するＭＡＸ，ＭＩＮ設定処理は重畳化することが可能である。

2.4 バケットサイズの調整

ＧＲＡＣＥではバケットサイズのゆらぎが性能を大きく左右する事はなく、又メモリレベルでのオーバフローも回避されているが、プロセッサの利用効率向上並びにそのオーバフロー防止の点からバケットサイズはプロセッサ容量以下でかつそれに近いことが望ましい。ここではバケット数のより多いハッシュ関数を用いた後、生成されたバケットを統合する方式を採っている（Ｂin Ｐacking）。

3. おわりに

ＧＲＡＣＥのバケット処理機構についてまとめた。現在、より詳細な性能評価をすすめている。

参考文献

［１］喜連川他：
Ｒelational Ａlgebra Ｍachine ＧＲＡＣＥ
京都数理解析研究所講究録 1982年 6月
［２］坂井他：
ＧＲＡＣＥに於けるモジュール間結合網とその評価 本大会予稿集４Ｆ－２