7D－5

超並列メインメモリデータベースシステムにおける永続データの取り扱い

今崎 憲児　小野 剛　牧之内 顕文
（九州大学 工学部 情報工学科）

1. はじめに

近年，プロセッサエレメント(PE)を数百台，数千台備えた分散メモリ型超並列計算機が商用化されている．このような計算機ではそれぞれのPEが数メガのメモリを持っているため，全体では数ギガのメモリを持つことになり，巨大なメモリ空間を持つ１台の計算機とみなすことができる．我々はこの点に着目し，現在データベースの分野で盛んに研究されているメインメモリデータベースシステム(MMDBS)を超並列計算機上で実現することを考えている．これにより，トランザクションの並列実行やクラッシュ・リカバリ処理の高速実行などが可能となる．本稿では，現在我々が開発しているオブジェクト並列プログラミング言語MAPPLEをデータベースプログラミング言語として用いる並列オブジェクト指向データベースシステムMAPPLE/DBの設計について述べる．特に永続オブジェクトの実現について言及する．

2. MAPPLE/DBの設計

2.1 メインメモリデータベースシステム(MMDBS)

MMDBS[GS92]は現在，DBの分野で盛んに研究されているトピックである．MAPPLE/DBはMMDBSの特徴をもつ．MMDBSはディスクを用いたDBに比べて次の様な利点が存在する．

- ダイレクトアクセスによるアクセス時間の短縮
- トランザクションの高速終了
- データがポインタで表されることによる領域の節約

しかし，DBのシステムがメモリに直接アクセスするため，信頼性に欠ける．そこで，クラッシュ時のためにディスク等にバックアップやログを残すような処理が必要となる．

Maniplating persistent data in massively parallel main memory database system
Kenji Imasaki, Ono Tsuyoshi, and Akifumi Makinouchi
Department of Computer Science and Communication Engineering,Kyushu University

2.2 MAPPLE/DBのアーキテクチャ

MAPPLE/DBでの構成要素は次のようになる．

- トランザクションマネージャ
  - PE間のメッセージの受渡し
  - トランザクションコミット・アボートの管理
- 永続・揮発ヒープ領域の管理

それぞれ構成要素はオブジェクトとして，PE上に常駐している．永続ヒープ領域の管理については，後の節で述べる．

3. MAPPLEでのヒープ領域の分類

MAPPLEでは，データベースのオブジェクトは各PE上のヒープ領域に割り当てられる．ヒープ領域には次の二種類がある．

- グローバルヒープ－他のPEから参照される可能性のあるオブジェクトを割り当てる．各オブジェクトはグローバルOID(PE+オブジェクト参照テーブル(ORT)エントリ番号)を持つ
- ローカルヒープ－そのPEからしか参照されないオブジェクトを割り当てる．通常のC++のヒープ領域がこれにあたる．

上の分類とは別に次の分類がある

- 永続ヒープ領域－複数のアプリケーションプログラムに渡ってメインメモリ上に存在する．
- 揮発ヒープ領域－単一のアプリケーションプログラムのみでメインメモリ上に存在する．

これらの分類の関係は図１のようにになる．

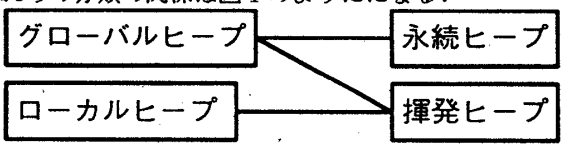

図１ ヒープ領域の両立する関係

MAPPLEでは永続ヒープに割り当てられたオブジェクトは永続化される．

4. MAPPLEでのオブジェクトの分類

MAPPLEではオブジェクトを集合オブジェクト(Set Object)とそうでない素オブジェクト(Atomic

Object) に分ける．集合オブジェクトはデータの集まり (collection) である．集合オブジェクトの要素を要素オブジェクト (Element Object) と呼ぶ．要素オブジェクトは素オブジェクトでも集合オブジェクトでもよい．

集合オブジェクトや素オブジェクトはある１つの PE 上に生成されるため，複数の PE をまたがることはできない．しかし，PE 上のオブジェクトのメソッドを並列に実行するためには，各 PE 上のオブジェクトをまとめて１つの集合と見る必要がある．MAPPLE ではこの問題に対し，“和集合”という概念を導入することで対処している．和集合オブジェクト (Union Object) は各 PE 上の同一の型の集合オブジェクトまたは素オブジェクトを管理するオブジェクトである．

これらのオブジェクトによる並列処理の実行の様子を図２に示す．並列実行は集合オブジェクトのメソッド及び要素オブジェクトのメソッドどちらでも行なえる (階層的な並列実行)．

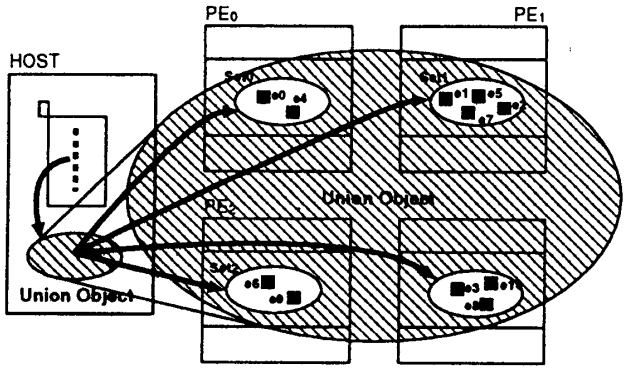

図２ オブジェクトの並列実行の様子

5. MAPPLE/DB の永続ヒープ領域管理

MAPPLE/DB では次に述べるような機能を提供し，オブジェクトの永続化を行なう．永続化については永続プログラミング言語 INADA[IN95] を参考にしている．

5.1 永続化のインターフェース

- オブジェクトに名前を登録し，別のアプリケーションプログラムではその名前で参照する．MAPPLE/DB では永続ヒープ管理オブジェクトに次のメソッドを用意している．
  - int save_obj(char* str,void* ptr) − *ptr* が指すオブジェクトを名前 *str* で登録する
  - void* load_obj(char* str) − 名前 *str* のオブジェクトを指すポインタを返す
- 検索した永続オブジェクトに対して処理を適用する．例えば，学生の中で２０才以上の永続データについて，あるアプリケーションプログラムで用いるという場合である．

5.2 永続ヒープ管理オブジェクトの機能

永続ヒープ (PH) および永続ヒープ管理オブジェクト (PHMO) は各 PE 上に常駐する．つまり，各 PE 上では，アプリケーションプログラムと PHMO が交互に動く．また HOST 上では前述の和集合オブジェクトの永続性を管理するためにマスターヒープ領域 (HPH) 及びマスター永続ヒープ管理オブジェクト (MPHMO) が存在する．PHMO および MPHMO は次のような機能を持つ．

- 定期的にダーティページをバックアップファイル上に書き出す (チェックポインティング)．PE 上に２次記憶が存在しない場合は，HOST 上にそれを送り，HOST はそのバックアップをとる．
- コミットしたトランザクションについて，ログ（アフターイメージのみ）をとり，２次記憶上にはき出す．

6. おわりに

本論文では，MAPPLE/DB の設計の概略について述べた後に，その上のデータベースプログラミング言語である MAPPLE について簡単に述べた．次に MAPPLE/DB 上での永続オブジェクトの実現方法について言及した．

現在は実際の並列計算機への実装は富士通社製 AP1000 上で予備評価[IF94] を行なったところである．そこで今後は本稿で述べた永続化を実装していき，この方法を評価していきたいと思う．

参考文献

[GS92] Hector Garcia-Molina, and Kenneth Salem:"Main Memory Database Systems:An Overview", IEEE Tansactions on Knowledge and Data Engineering, Vol.4, No.6, December 1992.

[IF94] 今崎，福見，牧之内："MAPPLE による超並列オブジェクト指向メインメモリデータベースの試み"，情報処理学会第５０回全国大会，4-79，1995.

[IN95] 牧之内研究室:" 永続プログラミング言語 INADA 言語仕様書，1995.