

7K-5

分散ファイルシステムにおける  
重複データの管理方式本田公男・田中英彦・元岡 達  
(東京大学工学部)

1.はじめに 分散データベースの一構成法として、まず網上で共通なアクセス法を提供するファイルシステムを作成し、その上に、そのアクセス法を利用してDBMSを構成してゆく方法が考えられる。一方、分散データベースにおいては、通信経費の低減、応答時間の短縮、信頼性の向上などの点から重複データを認めることが重要である。本発表では、研究用計算機網TECNETに実装された分散ファイルシステムの重複データ管理方式について報告する。この方式では、ユーザに重複データを意識させず、網上で共通なアクセス法を提供している。

2.重複データ管理のポイント 本システムでは重複データの管理にあたり、

- (1)ユーザからひとつにみえるファイル(論理ファイル)と、複数のファイル実体(物理ファイル)のマッピング情報をカタログに記録し、ユーザがJCLでファイルの物理的な重複を指定する必要をなくすこと。
- (2)ファイルにアクセスする際、アプリケーションがファイル実体の選択(マッピング)を行なう必要のないこと。を実現し、ユーザからデータの重複がみえないように努めた。

## 3.システムの構成(図1参照)

- (1)FMM(File Main Manager)
 

各サイトにひとつあるプロセスで、自サイト内のファイルを一括管理し、以下の機能を他のFMMと協力して実行する。

  - (a)ファイルの確保、解放(排他制御)
  - (b)ファイルの新規作成、消去の管理
  - (c)ファイルのOpen, Closeの管理
  - (d)ディレクトリの管理

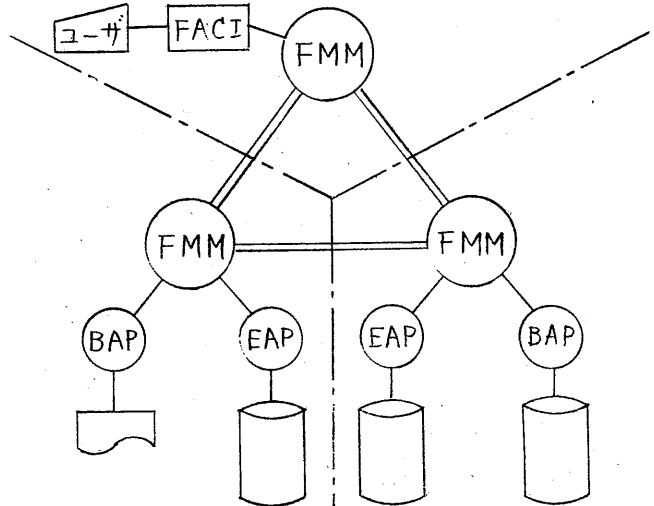


図1. システムの構成

- (2)BAP(Basic Access Process)
 

実際にファイル実体へアクセスするプロセスで、以下の機能がある。

    - (a)ファイルのOpen, Close
    - (b)ファイルの入出力(Read, Write)
    - (c)ファイルの新規作成, 消去
  - (3)EAP(Extended Access Process)
 

DBMSのために設けられたプロセスであり、データベースへの検索, 更新等のハイレベルなアクセス手法を提供し、通信量を低減する。
  - (4)FACI(File Access Command Interpreter)
 

各々のアクセスに対するマッピング情報に基づきマッピングを実行するリエントラントなルーチン群である。
- 4.ディレクトリの構成(図2参照)
- 以下のテーブルからなる、「三階層テーブル方式」と呼ぶ方式を採用した。
- (1)UFNT(User File Name Table)
 

ユーザがつけたファイル名とオーナーがつけたファイル名のマッピングを行なう表で、OFNTの検索に必要な情報を持つ。使用するファイルごとに1エントリある。

UFNTの1エントリ)      OFNTの1エントリ)

ユーザファイル名
OFNTの 存在するサイト
オーナー名
オーナーファイル名

オーナーファイル名
物理ファイルの 存在する全サイト
機密保護 情報

PFNTの1エントリ)

オーナーファイル名
物理ファイル 装置情報

UGTの1エントリ)

ユーザ名
UFNTへのポイント
OFNTへのポイント
PFNTへのポイント

図2. ディレクトリの構造

(2) OFNT (Owner File Name Table)

論理ファイルと物理ファイルのマッピングを行なう表で、複数のPFNTの検索に必要な情報を持つ。論理ファイルごとに1エントリある。

(3) PFNT (Physical File Name Table)

装置情報を記録してある表で、物理ファイルごとに1エントリある。

(4) UGT (User Group Table)

UFNT, OFNT, PFNTの検索の効率を考慮して設けられた表である。

5. プロトコルと管理手順      本システムが処理する網上で統一されたコマンド(プロトコル)とその発行手順を以下に述べる。

(1) ファイルアクセス

(a) GETFN コマンド

UFNTを検索し、OFNTの検索に必要な情報を得る。

(b) OBTAIN コマンド

OFNTを検索し、ファイルを論理ファイル単位に、デッドロックを回避するアルゴリズムにより、確保し、PFNTの検索に必要な情報を得る。

(c) OPEN コマンド

PFNTを検索し、装置情報を得た後、物理ファイルをOpenする。(c), (d), (e)は、重複データ更新の際には物理ファイルの存在する全サイトに対し

て発行されるばならないが、それらの処理はFACIが行なうのでアプリケーションは重複を意識しないでよい。

(d) READ, WRITE コマンド

物理ファイルと入出力を行なう。

(e) CLOSE コマンド

物理ファイルをCloseする。

(f) FREE コマンド

確保した論理ファイルを開放する。

(2) カタログ管理

(a) CREATE, DELETE コマンド

物理ファイルの作成、消去と、PFNTへの登録、削除を行なう。

(b) CATALOG, UNCATALOG コマンド

UFNTへの登録、削除を行なう。なおOFNTへの登録、削除はOBTAIN, FREE コマンドが兼任している。

(c) ADDCOPY, DLT COPY コマンド

物理ファイルを追加、削除したときに、OFNTの「物理ファイルの存在する全サイト」を更新する。

(d) PERMIT, INHIBIT コマンド

ファイルの共用を許可、禁止する。

6. おわりに      本システムでは、重複データによる通信経費の低減、応答時間の短縮を追求した。そのため、重複データの更新にあたって通信不可能なサイトへの更新要求を保持しておく機構、ディレクトリの多重化、などによる信頼性の向上については、実装するに至らなかった。しかし、本システムにより、分散データベースにおいて、重複データによる通信経費の低減、応答時間の短縮を容易に達成できる環境が整ったといえよう。

〔参考文献〕

- 堀口他：分散データベース用ファイル管理システム，第19回全国大会，pp. 957~958, 1978. 8.
- 本田他：分散データベース用ファイル管理システム，第15回データベース管理システム，第2回分散処理システム合同研究会，1979. 9.