

UNIX での高信頼ファイルシステムの実現

4 F - 7

猪飼 誉夫, 鶴 薫, 加賀屋 芳美, 渡辺 秀次

三菱電機 (株) 情報電子研究所

1 はじめに

リアルタイム性、高信頼性を要求される分野で使用される工業用計算機等でも、市場流通 S/W の利用、S/W 資産の有効活用の点から、標準化、オープン性が求められている。

一般に、これらの標準化、オープン性といった要求に応える OS として UNIX がある。しかし、UNIX のファイルシステムでは、更新データを扱うにあたり、以下のような問題があげられている。

- 高速化のためにキャッシュを使用しているのに、更新データの確定の保証がない。
- ファイルシステムとして排他制御機能が弱い。
- 索引順編成など、高度なファイル編成機能がない。

これらの UNIX ファイルシステムが抱えていた問題点に対応するファイルシステムとして、高性能ファイルシステム (MELQISAM) を UNIX を実行環境とした工業用計算機「MELCOM 350-60/AR」上に開発した。本稿では、この高性能ファイルシステム (MELQISAM) の機能と実現方法を述べる。

2 特長

高性能ファイルシステム (MELQISAM : MELcom Queued and Indexed Sequential Access Method) の特長を以下に示す。

1. ファイルの二重書き、トランザクション制御機能等によりデータに対する信頼性を高めている。
2. 高速アクセス用には主メモリ、大容量データ用にはディスク、他系とのデータ共有のためには共有メモリ、共有ディスクが使用できる。
3. ファイル編成として、テーブル構造の直接編成、キュー構造の巡回編成、索引を持つ索引順編成が使用できる。
4. アプリケーション・プログラムからのデータの取り扱い、記憶媒体やデータ構造に依存しない統一のインターフェースとしている。

3 実現方法

MELQISAM の S/W 構成を図 1 に示す。

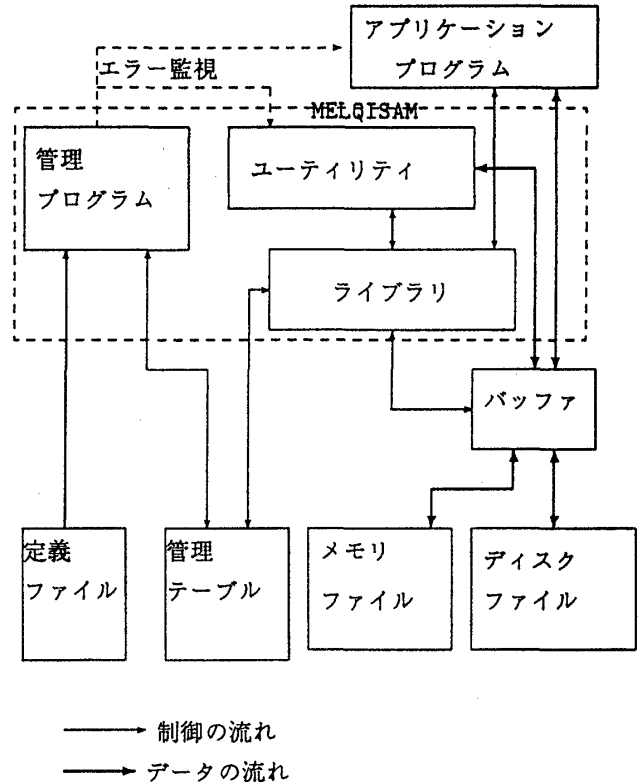


図 1: MELQISAM の S/W 構造

MELQISAM は、ライブラリとユーティリティと管理プログラムから構成され、主メモリ上のバッファ、管理テーブルを用い、各種のメモリ、ディスク上のファイル領域を操作している。ファイルシステム立ち上げ時に、管理プログラムは起動され、定義ファイルに従い管理テーブルを確保し、MELQISAM 使用プログラムのエラー監視を行う。

以下に、最初に述べた UNIX ファイルシステムの問題に対する MELQISAM の実現方法を述べる。

3.1 ディスク領域管理

UNIX ファイルシステムは、最初に述べたように、更新データの確定の保証がとりにくい。このため、本ファイルシステムでは、raw I/O により実現している。ディスクの領域管理には、区分編成ファイル構造を採用し、連続領域をファイル領域として使用している。これにより、ファイルシステムによる I/O エラーの管理、制御が可能となり、連続領域の使用で、I/O の高速化を図っている。

3.2 排他制御

UNIX 機能のセマフォ、fcntl システムコールを排他制御機能実現のために使用した場合、下記のような問題が発生する。

- プログラム障害等で対象リソースを解放せずにプログラムが終了すると、以降そのリソースが使用できなくなる。
- プログラム障害発生により無条件にリソースが解放されると、ファイルの情報に矛盾が発生する。

これらの問題に対し、MELQISAM では以下の方式で実現することとした。

排他制御の基本は、各排他要求リソース対応の待ち行列管理で、プログラムがリソース解放せずに終了した場合に備えて、管理プログラムがプログラム監視を行うこととした。管理プログラムはリソースを確保しているプロセスの存在を監視しており、プロセスが存在しない場合、対応するリソースに付随するファイルの修復と、リソースの解放処理を行う。

3.3 高速性と信頼性

本ファイルシステムでは、高速なデータ処理用として、主メモリ上にファイルデータを置くことが可能である。しかし、主メモリのファイルは、システム障害によるメモリ破壊など信頼性に問題がある。これに対し、図2のように、ファイルデータをディスクに退避/復旧するチェックポイント・セーブ機能をサポートし、信頼性の確保を可能としている。

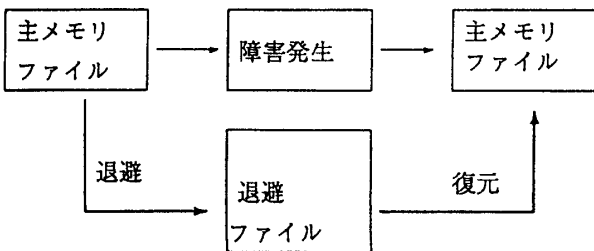


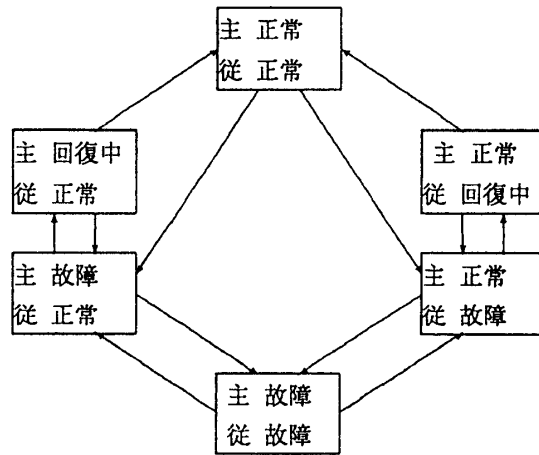
図 2: チェックポイント・セーブ機能

3.4 多重系システム構築のための機能

工業用計算機では高信頼システムを構築するため複数の計算機で共有可能なディスクを用いた冗長な多重系システムを構築することが要求され、ファイルシステムも多重系構成をサポートする必要がある。

また、アプリケーション・プログラムからは、他系の状態に応じ、ファイル状態の制御や二重書き時のファイル復旧の手順等を行なえることが要求されている。

MELQISAM は、共用ディスク、共用メモリをサポートし、ファイル状態の制御 (主側/従側ファイルの切り離し、正常化) や二重書き時のファイル復旧の手順 (二重書き開始、ファイルのコピー、正常化) 等をアプリケーション・プログラムから制御できるように、図3のようなファイルの各ステータス間の遷移を行えるインターフェースを提供している。



—— 応用プログラムから制御可能な遷移

図 3: ファイル状態の遷移

4 おわりに

以上、従来の UNIX ファイルシステムでは実現が困難であった高信頼なファイルシステムとして、UNIX 上で動作するファイルシステム (MELQISAM) について、機能とその実現方法等を述べてきた。現在、MELQISAM は工業用計算機システム MELCOM 350-60/AR 上のファイルシステムとして、使用されている。