バーチャルアレイ・ファイルシステム (vafs)の基本構想

4P−3

秋沢　充　山下洋史　加藤寛次　鈴木広義†　牧　敏行‡　山田秀則‡

(株)日立製作所 中央研究所　†(株)日立製作所 ソフトウェア開発本部
‡日立コンピュータエンジニアリング(株)

1．はじめに

プロセッサの高性能化によりワークステーション(WS)の性能が著しく向上したが，これに対応する入出力系の性能向上は十分とは言い難い。そのため，ファイルアクセス性能の向上を目的としてアレイディスク装置[1][2]やストライプド・ファイルシステム[3]の採用が試みられている。ストライプド・ファイルシステムは複数台のディスク装置にファイルを分割(ストライピング)して格納し，OSの並列アクセス制御によりファイル入出力を高速化するものである。

本稿では，ストライプド・ファイルシステムをベースとした"バーチャルアレイ・ファイルシステム(vafs)"と呼ぶ高速UNIXファイルシステムを提案するとともに，そのアクセス高速化方式とインタフェース仮想化方式について報告する。

2．ｖａｆｓの基本構想

vafsは1チャネルのSCSIバスに複数のハードディスク装置(hd)を接続して構成したバーチャルアレイディスク装置(vahd)にファイルを分割格納する。アクセスの際には一連の要求を非同期に次々発行し，多重アクセスによりSCSIバスの利用効率を高めて高速アクセスを実現する。

またvahdのストライピングを意識させず，あたかも1台のhdのようにアクセスできるアプリケーションプログラム・インタフェース(API)を提供する。これによりプログラムの書き換えなしにvafs管理のファイルがアクセス可能となる。

さらにアレイディスク装置(ad)もサポートし，段階的なファイルアクセスの高速化とファイル格納容量の拡張を可能とする。

このように多重アクセスによる高速な入出力の実現と，デバイス非依存なAPIを提供することがvafsの課題である。

以上の課題を解決するために，

(1)高速多重アクセス制御方式
(2)API仮想化方式

を開発した。以下これらについて述べる。

3．高速多重アクセス制御方式

高速多重アクセス制御方式とは，外部との高速アクセスが可能なhdの内部キャッシュメモリ(CM)とSCSIバスのディスコネクト/リコネクト機能を利用して，vahdを構成する複数台のhdへのアクセスを並列化するものである。

図1にvahdの構成と動作を，図2にファイル読み出しのタイムチャートをそれぞれ示す。

書き込みの場合，vafsはファイルを分割して各hdのCMに次々と書き込む。書き込み完了後は各hdとの接続をディスコネクトする。一方，書き込み完了前にhdはシークを開始し，シークと回転待ちの後にディスク媒体に分割ファイルを書き込む。この間vafsは他のhdへ次の分割ファイルの書き込みを行う。

読み出しの場合，vafsは各hdに読み出し命令を次々に発行する。発行後，各hdとの接続をディスコネクトする。一方，hdは命令を受け付け

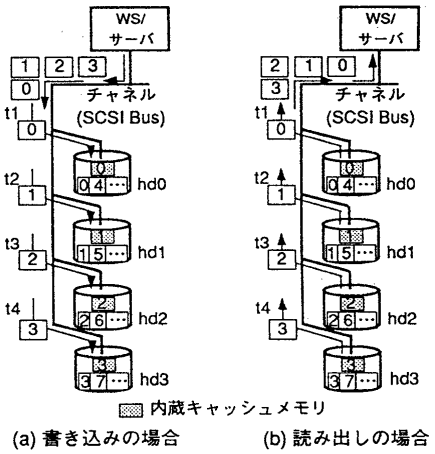

図1　vahdの構成と動作

An Architecture of Virtual Array File System (vafs)
Mitsuru AKIZAWA,Hirofumi YAMASHITA,Kanji KATO,Hiroyoshi SUZUKI†,
Toshiyuki MAKI‡,Hidenori YAMADA‡
Central Research Laboratory,Hitachi,Ltd. †Software Development Center,Hitachi,Ltd.
‡Hitachi Computer Engineering Co.,Ltd.

ると，シークと回転待ちの後にディスク媒体からデータをCMへ読み出してリコネクト要求を発行する。WSはこれを受け付けるとデータの読み込みを開始する。この間，他のhdもvafsが発行した命令により，ディスク媒体からCMへデータを読み出し，リコネクト要求をWSへ発行する。

このように各hdが独立にほぼ並列にアクセスされるため，高速なファイル入出力が実現できることになる。

4．ＡＰＩ仮想化方式

API仮想化方式はvahdを論理的に1台のhdのようにAPに見せるものである。このため，APIとなるvサブファイルシステムと分割したファイルを格納するsサブファイルシステムからvafsを構成し，vafsが両者の対応関係を管理する。

vafsはスペシャルファイル/dev/vaXをディレクトリへマウントして使用する。vafsをマウントした状態のディレクトリ構造を図３に示す。vafs内のファイルはvサブファイルシステム内にvファイルを持ち，各sサブファイルシステム内にsファイルを持つ。vファイルはデータの実体を持たず，これに対応する複数のsファイルにデータの実体を分割して持たせる。ディレクトリについても同様の構造をとる。

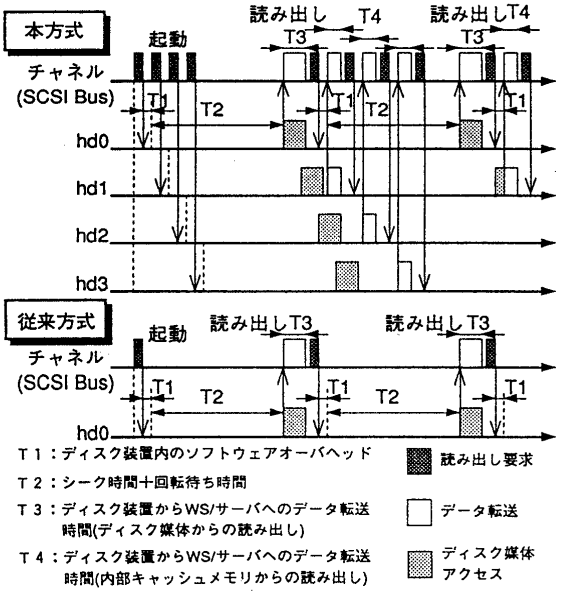

図２　ファイル読み出しのタイムチャート

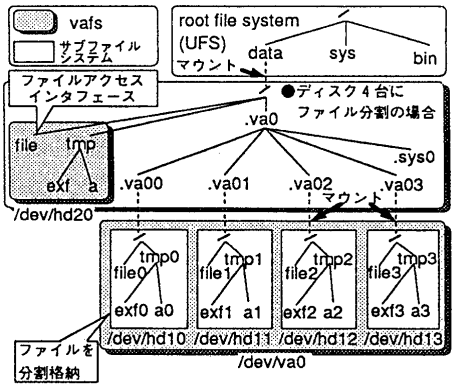

図3　API仮想化方式

ディレクトリの移動やファイルのパス名サーチの場合には，vファイルとvディレクトリを対象とする。vafs内のファイルやディレクトリにアクセスする場合は，vファイルとvディレクトリから対応するsファイルとsディレクトリのiノードを得る。sファイルのデータブロックの番号と格納デバイス番号をvファイルのiノードに書き込んでおき，これを参照してsファイルが格納されたhdをアクセスする。

このようにvafsをvサブファイルシステムとsサブファイルシステムから構成するAPI仮想化方式により，vahdを論理的に1台のhdのようにAPに見せることができる。

5．おわりに

高速UNIXファイルシステムとしてバーチャルアレイ・ファイルシステム(vafs)を提案し，基本方式について報告した。現在，本ファイルシステムはプロトタイピング中であり，今後は性能評価および動作解析を行う予定である。

参考文献

[1]D.A.Patterson, P.Chen, G.Gibson, and R.H.Katz,"Introduction to Redundant Arrays of Inexpensive Disks(RAID)",spring COMPCON '89, pp.112-117, Feb.1989

[2]J.Moad, "Relief for Slow Storage Systems",DATAMATION, pp.22-28, Sep.1990

[3]M.Loukids(砂原秀樹 監訳),「UNIXシステムチューニング」,アスキー,1991

*UNIXオペレーティングシステムはUNIX System Laboratories,Inc.が開発し，ライセンスしています。