RAID型ファイルシステムVAFS/HRのファイル管理方式

6U－4

城田浩二　山下洋史*　高橋英男*　畠山敦*　裏谷郁夫*　山田秀則　高良亜紀子
日立コンピュータエンジニアリング（株）　*（株）日立製作所

1．はじめに

近年，ネットワークの普及に伴い複数のワークステーション（WS）間でファイルの共有化が進んでいる。共有化されたファイルには複数のユーザがアクセスするため，ファイルアクセスの高速化とファイルデータの高信頼化が要求される。筆者等は既に，複数のディスク装置にファイルを分割して格納する“バーチャルアレイ・ファイルシステム（VAFS）”を開発し，ファイルアクセスの高速化を実現している[1][2][3]。今回，更にパリティ・データを付与してディスク装置の故障に対するデータ保証を行うRAID型ファイルシステムVAFS/HR（Higih Reliability）を提案し，ファイルシステムの高性能化，高信頼化に取り組む。本稿では，VAFS/HR上でのパリティ・グループの管理とパリティ・データの格納を含むファイル管理方式について報告する。

2．VAFS/HRのファイル管理方式

2.1　ファイル管理の課題

ファイル管理方式やファイル格納方式を検討するにあたって，RAIDアーキテクチャを組み込む場合には，性能面，特にファイル書き込み性能の面から以下の二つの課題がある。

（1）断片化ファイルの書き込み性能の向上

Unix File System (UFS)では，ファイルをディスク装置に格納するときには物理的に連続したブロックを割り当てようとする。しかし，ファイルの書き込みや削除を繰り返していくと，物理的に連続したブロックに格納できず断片化されたファイルが徐々に増えてくる。断片化されたファイルの書き込み時にRAIDアーキテクチャに従ってVAFS/HRを構築する各ディスク装置の同一番地の物理ブロックでパリティ・データを生成しようとすると，パリティデータ生成のための読み出しが多発する。例えば，図1の場合では2ブロックの書き込みを行うと，パリティ生成のために4ブロックの読み出しが発生する。パリティ・データ2ブロックの書き込みを加えると計8ブロックのディスクI/Oが発生する。これは，断片化されていないファイルの場合の2倍のI/O数である。すなわち，断片化ファイルのファイル書き込み時のI/O数を減らし，ファイル書き込み性能を向上させる必要がある。

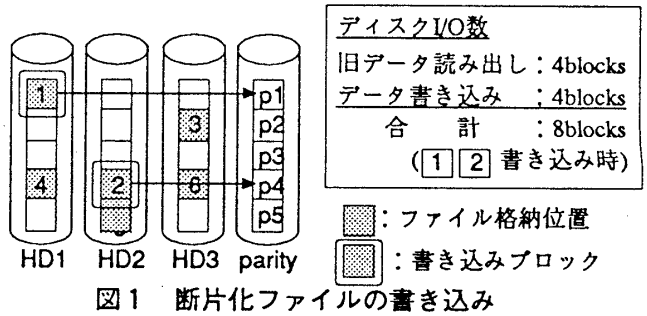

図1　断片化ファイルの書き込み

（2）パリティ・データ書き込み性能の向上

パリティ・データを格納するディスク装置を固定にしたRAID4の構成では，ファイル書き込み時にパリティ・データを格納したディスク装置にI/O負荷が集中しやすい。特に，(1)の断片化されたファイルへの書き込みや，断片化されていないファイルでもランダムに書き込みが行われるような場合に，負荷の集中は起きやすい。パリティデータ格納ディスク装置に対して負荷が集中しボトルネックが発生すると，複数のディスク装置への並列アクセスができなくなる。そこで，パリティデータ書き込み時でも複数のディスク装置への並列アクセスを可能にし，書き込み性能の向上させる必要がある。

上述した問題を解決するために，表1に示すアプローチを採用した。

The file management scheme of VAFS/HR - a software RAID file system
Koji Shirota, Hirofumi Yamashita*, Hideo Takahashi*, Atsushi Hatakeyama*, Ikuo Uratani*, Hidenori Yamada and Akiko Kora
Hitachi Computer Engineering Co.,Ltd.
*Hitachi, Ltd.

表1　ファイル管理の課題とアプローチ

No.	課　題	アプローチ
1	断片化ファイルの書き込み性能の向上	ファイル内パリティグループ管理方式
2	パリティデータ書き込み性能の向上	ファイル間パリティ分散配置方式

2.2　ファイル内パリティグループ管理方式

ファイル内パリティグループ管理方式は，パリティグループをファイルの論理ブロックごとに生成し管理する方式である。本方式により，RAIDアーキテクチャをそのまま適用した場合に比べて断片化されたファイルの書き込み性能が向上する。例えば，図1のような書き込みでも，パリティ生成のための読み出しが発生せず，断片化されていないファイルの場合と同数のI/O数に抑さえることができる。

本方式を実現するために，図2に示すようにファイルを格納する。すなわち，ファイル・データは分割しサブファイルとしてHD1～HDn-1に格納する。パリティ・データは各サブファイルの同一論理ブロック番号のブロック間で生成し，パリティファイルとしてHDn に格納する。これにより，各サブファイルとパリティファイルの同一論理ブロック番号のブロックがパリティグループとして管理される。

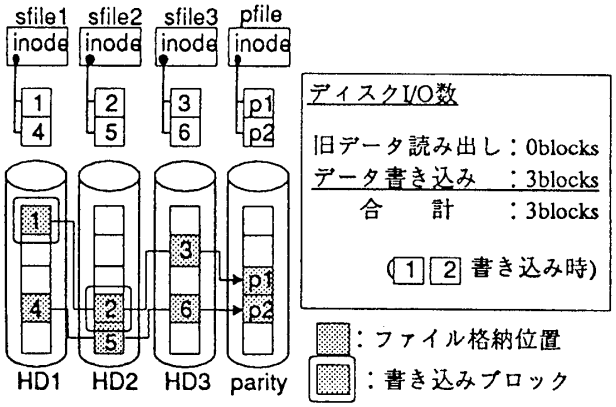

図2　ファイル内パリティグループ管理方式

2.3　ファイル間パリティ分散配置方式

ファイル間パリティ分散配置方式は，ファイルごとにパリティデータ格納ディスク装置を変更していく方式である。本方式により，パリティデータ書き込みの際に複数のディスク装置に負荷を分散でき，

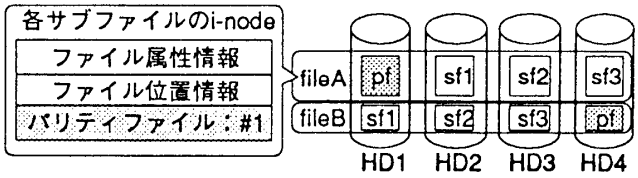

図3　ファイル間パリティ分散配置方式

パリティデータ書き込み性能が向上できる。

本方式を実現するために，各サブファイルのファイル情報を管理するi-nodeと呼ばれる領域にVAFS/HR専用の領域を設けて，そこにパリティファイルが格納されるディスク装置番号を登録する。各サブファイルにパリティファイルが格納されるディスク装置番号を登録するのは，ディスク装置の故障に対応できるようにするためである。すなわち，どれか一つのサブファイルが壊れていても，他のサブファイルのi-node情報からパリティファイルが格納されているディスク装置を特定できるようにしている。

パリティファイルを格納するディスク装置を決定する方法は，次の通りである。新規にファイルを作成するときにディスク装置の空き容量を調べ，最も空き容量が大きいディスク装置をストライピング開始ディスク装置とする。そして，ストライピング開始ディスク装置の一つ前の番号のディスク装置をパリティファイル格納ディスク装置と決定する。

3.　おわりに

VAFS/HRでは，ファイル内パリティグループ管理方式とファイル間パリティ分散方式を開発したことにより，従来のVAFSと遜色ない性能を保った状態で，高信頼なファイル管理を行えるようになった。

参考文献

[1]秋沢他5，「バーチャルアレイ・ファイルシステム(vafs)の基本構想」，情報処理学会第45回全国大会講演論文集4-62，(平4-10)

[2]秋沢他6，「ストライプド高速UNIXファイルシステムの開発」，情報処理学会システムソフトウェアとオペレーティングシステム研究会61-2，(平5-8)

[3]鬼頭他6，「高速UNIXファイルシステムの構想」他4件，情報処理学会第47回全国大会講演論文集，7B-1~5，(平5-10)

注)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。