高速UNIXファイルシステムの性能評価

7B－5

山下 洋史　秋沢 充*1 加藤 寛次　鬼頭 昭*2 牧 敏行*3 山田 秀則*3
(株)日立製作所 中央研究所　*1(株)日立製作所コンピュータ事業本部
*2(株)日立製作所 ソフトウェア開発本部　*3 日立コンピュータエンジニアリング(株)

1．はじめに

近年，UNIXワークステーション(WS)がCPU性能の向上に伴い急速に普及してきたが，WSシステムとしての性能を引き上げるためにはファイルアクセス性能の向上が不可欠とされている。筆者らは，ファイル・ストライピングを用いた高速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，WSシステムの高性能化に取り組んできた[1]。VAFSは，UNIXオペレーティングシステムに組み込んだファイルシステムであり，汎用外部I/OバスであるSCSIバスを介してWSに接続された複数のディスクを並列制御することによって，ファイルアクセスの高速化を実現する。本稿では，WS上に実装したVAFSの性能評価結果について報告する。また，この結果の分析を通して，VAFSの性能向上のための課題を明らかにする。

2．VAFSのファイルアクセス性能の評価

2.1　測定方法

ここでは，VAFSにおけるシーケンシャルアクセス性能とランダムアクセス性能を測定する。シーケンシャルアクセス性能からファイルアクセスの最大スループットが，ランダムアクセス性能から最小スループットが分かる。VAFSと比較するために標準UNIX File System(UFS)の性能も測定する。それぞれ，表1に示すアクセス方法で測定する。すなわち，読み出し時には，VAFSでは非同期システムコールのasread()を使用し，UFSでは従来の同期システムコールであるread()を使用して測定を行なう。同様に書き込み時には，VAFSでは非同期システムコールのaswrite()を使用し，UFSでは従来の同期システムコールであるwrite()を使用する。asread()やaswrite()では，

表1　アクセス方法

	読み出し時	書き込み時
VAFS	asread()	aswrite()/synchronousモード
UFS	read()	write()/synchronousモード

asread(), aswrite()：非同期システムコール関数

The Performance Evaluation of Performance Improved UNIX File System
Hirofumi YAMASHITA, Mitsuru AKIZAWA*1, Kanji KATO, Akira KITO*2, Toshiyuki MAKI*3, Hidenori YAMADA*3
Central Research Laboratory,Hitachi,Ltd.
*1 Computer Group,Hitachi,Ltd.
*2 Software Development Center,Hitachi,Ltd.
*3 Hitachi Computer Engineering Co.,Ltd.

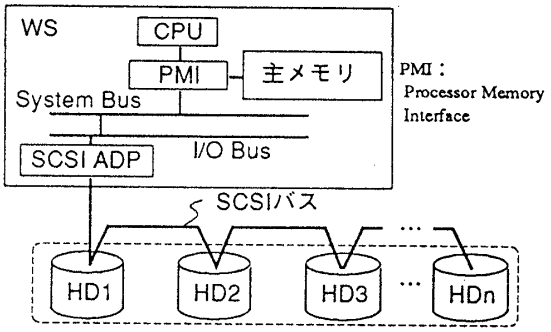

図1　性能測定のためのシステム構成

表2　測定機器の仕様

WS		ディスク装置		
名称	SCSIバス	サイズ	回転数	容量
3050RX	10MB/s	3.5inch	5400rpm	1.0GB

ディスク装置にアクセス要求を発行するとそのアクセスの終了を待たずに次の処理へ移行する。read()やwrite()では，ディスク装置へのアクセスが終わるのを待ってから次の処理へ移る。また，aswrite()とwrite()ではシステムバッファのキャッシングの影響を排除するために，ディスク装置までデータを書き込むsynchronousモードを用いる。

測定するシステムの構成を図1に示し，測定機器の仕様を表2に示す。ここでは，日立製WS3050RXを使用し，表3に示す条件でシーケンシャルアクセス性能とランダムアクセス性能を測定する。

表3　測定条件

	母体ファイル	総アクセスデータ量	1回当たりのアクセス量
シーケンシャルアクセス	8MB	8MB	8kB (1Block)
ランダムアクセス		2MB	

2.2　測定結果

3050RXにおけるVAFSのシーケンシャルアクセス性能およびランダムアクセス性能の測定結果を，それぞれ図2の(a)と(b)に示す。従来との比較対照のためUFSの性能も合わせて示す。

読み出し時には，ディスク装置3台で8.1MB/s (UFSの2.3倍)のシーケンシャル性能が，ディスク装置4台で1.8MB/s(UFSの3.0倍)のランダム性能が得られた。書き込み時には，ディスク装置4台で6.5MB/s (UFSの32.5倍)のシーケンシャル書き込み

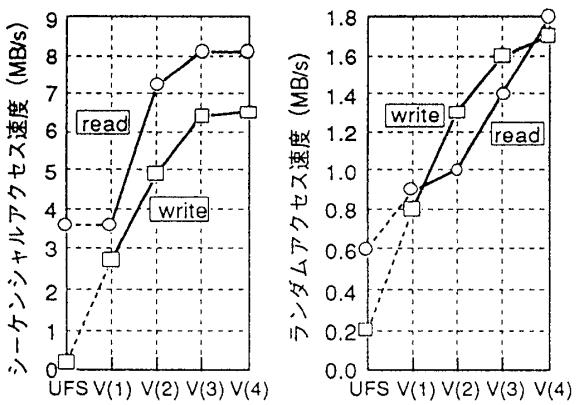

図２　測定結果

性能と，１.７MB/s(UFSの８.５倍)のランダム書き込み性能が得られた。シーケンシャルアクセスでは，ディスク装置４台以上で性能が飽和してしまう。

２.３　測定結果の考察

（１）シーケンシャル読み出し性能

ディスク装置４台以上で性能が飽和してしまうのは，SCSIバス転送速度がボトルネックとして作用するためである。このことは，ディスク装置３台接続時の性能がSCSIバス１０MB/sの約８０%(８.１MB/s)に達していることからも分かる。

（２）ランダム読み出し性能

ディスク装置１台構成のVAFSの性能がUFSに比べて約１.５倍高くなるのは，非同期入出力システムコールasread()の効果のためである。asread()を使用し先のアクセス要求の終了を待たずに次々とデバイスドライバに発行されたアクセス要求は，シークが最適に行なわれるようにデバイスドライバで並べ替えられるため，シーク時間が短縮され性能が向上する。ランダム読み出しの場合には，SCSIバスの性能に余裕があるため，ディスク装置台数に応じて性能が向上する。したがって，ディスク装置台数を５台以上に増やすことで，更に高い性能が期待できる。

（３）シーケンシャル書き込み性能

ディスク装置１台構成のVAFSの性能がUFS性能に対して１３.５倍向上するのは，VAFSで非同期システムコールaswrite()を用いているためである。すなわち，aswrite() を使用するとデバイスドライバに次々と発行されたアクセス要求は連続アクセスとなりブロックマージ機能によって一つにまとめられるため，シークと回転待ち時間が大幅に削減される。また，write()ではiノードの更新がシステムコールごとに行なわれるが，aswrite()ではaswait()が発行された時に一括して行なわれる。そのため，iノード更新のオーバヘッドが大幅に削減される。これら二つの理由により，UFSに対しディスク装置１台構成のVAFS性能が大幅に向上することになる。ディスク装置が３台を越えると性能が頭打ちになるのは，SCSIバスが飽和するためである。また，シーケンシャル読み出し性能が最高８.１MB/sであるのに対し書き込みでは最高６.５MB/sとなるのは，i-node更新によるオーバヘッドの影響である。

（４）ランダム書き込み性能

ディスク装置１台構成のVAFSの性能がUFSに比べて４倍高くなるのは，非同期システムコールaswrite()を用いているためである。aswrite()を用いると，ランダム読み出しの時と同様に，シークが最適化されるため，シーク時間が短縮される。また，シーケンシャル書き込み時と同様に，aswrite()では一括してiノードの更新が行なわれるため，iノード更新のオーバヘッドが削減される。これら二つの理由により，UFSに比べてディスク装置１台構成のVAFSの性能が大幅に高くなる。

２.４　VAFSの高性能化への課題

VAFSは理想的な状況では，ディスク装置台数に応じて性能が向上する。しかし，実際にはVAFSの性能は飽和曲線を描く。これは表４に示す要因がボトルネックとなるからである。したがって，VAFSを更に高性能化するためには，(1)Fast-Wide SCSIバスの採用によるSCSIバス転送速度の向上，(2)システムバスの性能向上によるDMA転送速度の向上が課題となる。またシステムバッファからユーザバッファへのデータコピー性能の向上も重要になると考えられる。

表４　ボトルネック要因

項番	ボトルネック要因	部位	処理内容
1	SCSIバス転送速度	SCSIバス	ディスク装置からシステムバッファへのデータ転送
2	DMAデータ転送速度	WS	

３．おわりに

VAFSをWS上に実装し，性能評価を行なった結果，ディスク装置台数に応じて性能が向上することが確認できた。これにより，VAFSにおける非同期I/O方式と，多重アクセス制御方式の有効性を検証することができた。同時に，VAFSを更に高速化するためには，SCSIバス転送速度やDMA転送速度の向上が必要であることが分かった。今後，これらの課題を踏まえ更なる性能向上を図っていく。

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」，情報処理学会全国大会講演論文集4-62,平成４年１０月

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。