

PCサーバ向け「ハイブリッドRAID」の開発（1） ～アーキテクチャ～

2F-7

大枝 高, 兼田 泰典, 松並 直人, 荒川 敬史, 八木沢 育哉
(株)日立製作所 システム開発研究所

1 はじめに

PCサーバは、標準化されたアーキテクチャによる低価格を特徴とし、CPUの高性能化および高機能OSの登場により基幹業務への適用が進んでいる。この分野では、高信頼が要求されるためPCサーバでもディスクアレイの搭載が標準で、ディスク5台構成のRAID-5ディスクアレイを内蔵するのが一般的となってきた。

ところが、従来のRAID-5用ディスクアレイコントローラは5台のディスク装置と同程度以上の価格であり、RAID-1（ミラー構成ディスクアレイ）に比べ本来の価格メリットをだせていないし、PCサーバ本体の低価格化にも追いついていない。また、性能的にもRAID-5特有のライトペナルティによりRAID-1に劣っているという問題がある。

そこで、RAID-5をベースにディスクアレイコントローラ価格を従来の1/3、性能を2倍にする「ハイブリッドRAID」を考案した。その特徴はソフトウェア制御とハードウェア制御の最適配分、ライトペナルティ削減のためのキャッシュ方式とその効果をさらに高めるパリティ配置方式である。

本報告は、「ハイブリッドRAID」のアーキテクチャと試作機による性能評価について述べる。以降の報告でキャッシュ方式の詳細や本方式の課題であるホストCPUの負荷低減をハードウェアサポートにより実現する方式などを説明する。

2. ハイブリッドRAIDアーキテクチャ

(1) ソフトウェア制御の導入

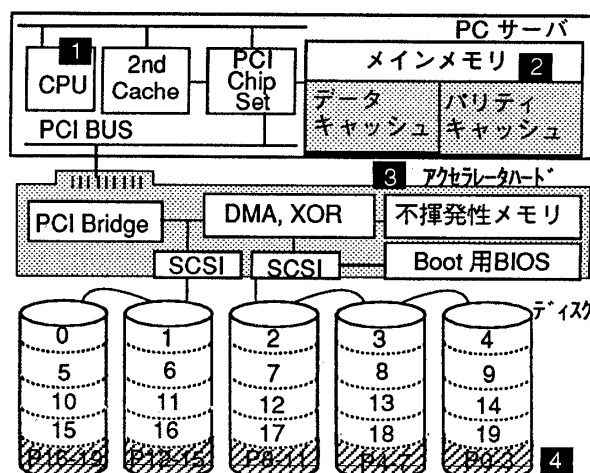
低価格化のためにアドレス変換とキャッシュ制御をソフトウェアで行う方式とし専用CPUを不要とした。高性能化が著しいホストCPUを使用するため高性能化が期待でき、かつ、ホストPCの性能向上によるディスクアレイコントローラの陳腐化が避けられる。

(2) デイレイドパリティ生成方式

RAID-5では、ライト時にパリティを生成し、ディスクに書き込むため通常のディスク装置に比べライト処理が遅くなる。このライト時の性能低下をライトペナルティと呼ぶ。

ライトペナルティの削減方式には大別して

- ・不揮発性キャッシュを利用する方式 [1]
- ・ダイナミックにアドレスマップを変化させ、複



項目	概要
1	ソフトウェア制御方式 上位CPUでRAID制御
2	デイレイドパリティ生成方式 パリティキャッシュを用いた遅延・まとめ処理
3	ハードウェアアクセラレータ データ転送, パリティ演算, 電源障害対策, Boot機能
4	分離パリティ方式 パリティ生成処理の高速化

図1 ハイブリッドRAIDの構成

Development of "Hybrid RAID" for PC Servers (1)
～Architecture～

Takashi Oeda, Yasunori Kaneda, Naoto Matsunami, Hiroshi Arakawa, and Ikuya Yagisawa

Hitachi, Systems Development Laboratory
1099 Ohzenji, Asao, Kawasaki, 215 Japan

数のライトをまとめて連続領域に書き込む方式 [2] の二種類の方式が知られている。不揮発性キャッシュを用いる方式はコストが高く、ダイナミックにアドレスマップを変える方式は実装とガーベジコレクションのオーバーヘッドが大きいという欠点がある。そこで、揮発性であるホストコンピュータの主記憶をキャッシュとして利用し、パリティ生成の遅延まとめ処理を行うダイレイドパリティ生成方式を考案した。電源障害対策のため、本方式ではデータをライトスルーとし、ダーティなパリティ領域のアドレスをアクセラレータボード上の不揮発性メモリに記憶する。不揮発性メモリはパリティそのものではなくアドレスだけでよいためコスト的に問題にならず、かつライト処理時間、およびライト処理時に占有するディスク台数をベストケースでは通常のディスクと同じ時間、台数まで削減できる。

(3) ハードウェアアクセラレータ

ハードウェア化した部分はソフトウェア制御では実現不可能な処理（不揮発性メモリ、ブートBIOS）およびホスト側CPUの効率を著しく落とす処理

（データ転送処理、パリティ演算処理、ディスク処理等）の二種類である。データ転送処理やパリティ演算処理をホスト側CPUで行うとCPUキャッシュのミス率を上げてしまう。また、ディスク処理をホストCPUで行うと割り込み回数が増大する。

(4) パリティ分離配置方式

図1のディスク領域で0から19の白抜きの部分がデータ、斜線部分がパリティである。通常のRAID-5ではパリティが対角線上にならぶ方式（flat-left-symmetric方式など[3]）を採る。パリティが対角線上に並ぶ方式はデータと対応するパリティが近い位置におかれるためデータ・パリティ同期更新を行うのには有利である。しかし、ダイレイドパリティ生成方式を採った場合、分離しておくことにより、パリティ生成のまとめ処理を行う場合ヘッドのシーク時間を短くできる効果がある。これによりダイレイドパリティ生成方式のキャッシュミス時でも20%程度の高速化効果が得られる。

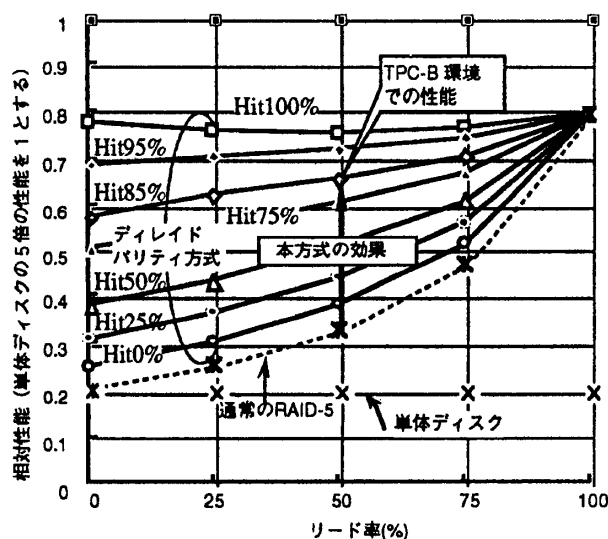


図2 試作機による性能評価結果

3. 性能評価結果

シミュレーションおよび試作機で性能を評価した。ディスク構成は5台（4D+1P）のRAID-5とした。リード50%、ライト時のダイレイドパリティキャッシュヒット率85%（TPC-Bベンチマークの標準的アクセスパターン）の時従来のRAID-5に比べ約2倍の性能を得た。試作機の性能はシミュレーションともよく一致し理論通りの性能が発揮できることが検証できた。

4. まとめ

PCサーバ用ディスクアレイを高コストパフォーマンス化するため、ハイブリッドRAIDアーキテクチャを考案した。ハード・ソフト制御の最適化およびダイレイドパリティ生成方式などの高速化方式により従来比約2倍の性能と大幅なハードウェア削減による低価格化を実現できた。

参考文献

- [1] Jai Menon and Jim Cortney. The Architecture of a Fault-Tolerant Cached RAID Controller. The 20th Annual International Symposium on COMPUTER ARCHITECTURE. May 16-19, 1993
- [2] Daniel Stodolsky, Garth Gibson, and Mark Holand. Parity Logging Overcoming the Small Write Problem in Redundant Disk Arrays. School of Computer Science and Dept of Electrical and Computer Engineering, CMU
- [3] E. Lee and R. Katz. Performance Consequences of Parity placement in Disk Arrays. Proceedings of ASPLOS-IV 1991, pp. 190-199.