

パネル討論 ディスクアレイの現状と展望  
- RAIDレベル4、5の考察 -

中村俊一郎 三菱電機情報電子研究所

5. 2.5インチ、3.5インチといった小型ディスクが容量/性能面で急速に伸びてきており、これらを多数並べたディスクアレイが最近注目を集めている。又これには市販のLSIの技術的進歩が大きく係わっているという点も見逃せない。

ここではRAIDレベル4と5にしぼってその実現上の技術的問題点に対する考察を行なう。

1. 狭義のRAID

RAID (Redundant Array of Inexpensive Disks) は、単に従来方式のディスクをアレー状に並べただけのRAIDレベル0と呼ばれるものから、二重故障までを復元可能なRAIDレベル6と呼ばれるものまで広い意味で使われることがある。RAIDレベル1は所謂ミラーディスクのごとであり、ディスクの信頼性を向上させる目的で、既に多くの商用マシン上で使われている。RAIDレベル2と3についても、スーパーコンピュータ用とか、画像データのような大量のデータの高速転送用として以前から商用化されている。これに対しここ数年注目を集め初めてきた所謂狭義のRAIDとは、RAIDレベル4、5、6のごとである。この狭義のRAIDについては、ミラーディスクのように多くの冗長ディスク(100%)を持たなくてもディスクのエラー訂正が出来るという魅力を持っている反面、ディスクへのWriteに時間がかかるという欠点も持っており、その評価はまだ固まった段階にはない。ここ数年米国では10数社がこれらのRAID製品を発表しているが、国内メーカからはまだ本格的なRAID製品は出ていない。

2. RAIDレベル4、5

RAIDレベル2、3は複数のディスクを同期回転させて、データを並行に読み書きして高速転送を実現したものであるが、レベル3では1バイトを単位として各ディスクからデータ転送を行なう。RAIDレベル4は原理的にはRAIDレベル3のバイト単位を、ブロック(1~数セクター)単位にしただけの違いしかない。しかしながらブロックになると、単1ブロックのみの読み書きが可能になるためかなり変わった様相を呈してくる。ディスク

の同期回転とか複数同時アクセスは必要無くなり(即ち個々のディスクへの単独アクセスが可能となり)、レベル3では当たり前であったデータのストライピング(横並び)も必ずしも必要でなくなり、又パリティの生成方法等も変わってくる。RAIDレベル5はレベル4からパリティを全ディスクに分散させるという改良を加えたものであり、本質的にはレベル4と同じである。

3. WRITE性能

ディスクのエラー訂正のためにミラーディスクでは冗長ディスク量100%を必要とするが、RAIDレベル4、5では9~25%程度でよいため非常に魅力的である(前者は多重故障、後者は単一故障という違いは有るが)。その代わり図1に示されるようにRAIDレベル4、5ではディスクへのWRITE時には常に、①旧データREAD、②旧パリティREAD、③新データWRITE、④新パリティWRITEと従来型に比べ4倍のディスクアクセスが必要であり大きな性能低下を招く。これは換言すればディスクが故障したときのための保険料として、ミラーディスクは冗長ディスク量で払うという事が出来る。

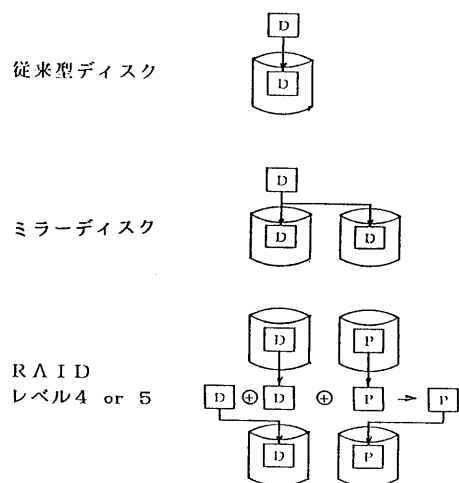


図1 ディスクのWRITE処理

#### 4. LSIの進歩

上記のようにRAIDレベル4、5ではディスクへのWRITE処理が遅いという大きな欠点を持っているが、この改善策としては以下のようなものが考えられる。

- (1)データディスクとパリティディスクを別のDKC(Disk Controller)に繋げ処理を並行して行なう。
  - (2)データとパリティをキャッシュに入れ、上記①旧データREAD、②旧パリティREADをディスクキャッシュから行なうようにする。
- さてここで注目すべきことは、Pacstor社のIntegralのような初期の製品を除き、現在市販されているRAID製品はほとんど例外なく上記(1)を実現していることである。即ち5~12個程度のDKCを配置しディスクへのREAD/WRITEを並行処理して性能向上を図っている。しかもこの5~12個のDKCが1枚のカードにコンパクトに収められているため手頃な価格で実現出来る。従来それ自体1つの装置であったDKCが、複数個まとめて1枚のカードに収まってしまうというようなLSI技術の進歩があって初めて、多数のDKCが必要とされるRAIDが商用化可能となってきたと言える。なおここでDKCとはディスクのデータ転送を独立して実行出来る単位のことを言っており、具体的にはSCSI I/O Processorを指す。(往年のDKCが備えていた機能の多くは、LSIの進歩により今やディスク装置自体の中に入っている。)

さてこれらの装置を眺めると「逆の面」が見えてくることに気付く。即ち1枚のカードに収まった5~12個のDKCが並行して処理を行なう訳であるから、RAIDのWRITE処理の遅さということとを差し引いても、ディスクシステム全体としてかなりの高性能を発揮するだろうということである。このことが現時点においてRAIDの大きな魅力となっていることは否めない。

#### 5. ストライピングと深さ

前章で述べた「ディスクシステムの性能」を考える時、それは以下の2点に集約される。

- (1)ディスク処理のスループット  
(トランザクション処理の高速化)
- (2)ディスク処理の応答時間  
(大量連続データREAD/WRITEの高速化)

さて、2章で述べたようにRAIDレベル4、5ではデータのストライピングは必ずしも必要ないが

、ストライピング無しの場合には、性能面では従来型のDKCが5~12個付いている場合に帰着でき、後はRAIDのWRITEによる性能低下を考慮すれば良い。

ストライピングをする場合には、RAIDレベル3型と従来型をミックスした使い方となるため、新たな技術的問題が生起する。ストライピングの効果は以下のようなものである。

- (1)データベースマシン等で使われる水平分割手法と同様、1つの論理的なファイルが自動的に複数のディスクにばらまかれるため、特定のディスクにアクセスが集中してネックになることを解消できる。
- (2)ディスク上の連続データを大量にREAD/WRITEする場合にはすべてのDKCが並行してデータ転送を行なうため高速転送が可能となる。一方次のようなデメリットも存在する。
  - ①個々のディスクのスケジューリングはRAID側が行わざるを得なくなり、それに伴いO/S側にも変更が必要となる。
  - ②上記の(2)使い方をした場合、連続データが大量でなかった場合に却って性能(スループット)の低下を招く。例えば連続データが数KByteで3個程度のディスクに渡るようなケースでは、1個のディスクに収まった場合に比べて明らかにスループットが落ちる。このことはストライピングの深さの問題に帰着され、これはユーザが選択可能なパラメータであるが重要な戦略的意味を持つことになる。

#### 6. まとめ

以上主としてRAIDレベル4と5の技術的問題点につき考察を行なった。今後益々システムが分散化し、その中心となるサーバ上のストレージシステムの重要性は増加し、この種の議論が活発になることを期待したい。

#### [参考文献]

- (1)D.A.Patterson et al. 'A Case for Redundant Arrays of Inexpensive Disks(RAID)' Report No. UCB/CSD 87/391, 1987
- (2)PRODUCT DESCRIPTION' RAID+SERIES MODEL RX' Array Technology Corporation, 1990