

# 高性能サーバ向けディスクアレイ「DF400」の開発（1）

4H-4

## ～アーキテクチャと高I/O処理技術～

松並 直人, 八木沢 育哉, 山本 政行, 大枝 高, 佐藤 雅彦\*

(株) 日立製作所 システム開発研究所、ストレージシステム事業部\*

### 1. はじめに

オープン系サーバの基幹系/情報系業務への適用が進展し、そのディスクサブシステムとして価格対性能比と信頼性に優れたRAID[1]型のディスクアレイが一般的に用いられている。

サーバの性能は、プロセッサの高速化に伴い2年につき3倍のペースで順調に向上し、同時にサーバとディスクサブシステムを接続するホストインタフェースの転送速度もFibre Channelの登場等により高速化し続けている。

一方、インターネットを活用した電子商取引等の新ビジネスの普及に伴い、サーバには24時間365日の連続運転が可能な高信頼性が求められている。

以上の背景より、サーバ向けのディスクアレイには、サーバやインタフェース速度に追随した高トランザクションI/O性能と高速データ転送性能、および24時間運転に対応する高信頼性が要求される。

本稿では、高速データ転送と高信頼化を実現するディスクアレイ「DF400」のコントローラアーキテクチャと、高I/O処理技術について報告する。

### 2. コントローラアーキテクチャ

#### (1) 高性能化の実現（高性能サーバ対応）

Fibre Channelの高速転送性能をフルに引出すため合計800MB/Sの転送帯域を備える専用データ転送LSIを搭載した。各ホストインタフェース用Fibre Channelコントローラは、データ転送LSIに独立バスで接続する。また、ディスクインタフェ

ース用SCSIコントローラは、2つずつ3本の独立バスで接続する。サーバからのライトデータはデータ転送LSIのDMA機能により、デュアルコントローラの両方のキャッシュに二重書きする。

#### (2) 高信頼化の実現（24時間無停止運転対応）

ホストインタフェースとして最大4本のFibre Channelをサポートし複数台構成のクラスタサーバに独立ポートで直結できるマルチポート機能と、障害時にクラスタサーバとDF400が連動して障害箇所を回避し正常部位に処理を引継ぐ連動フェイルオーバー機能を実現した。

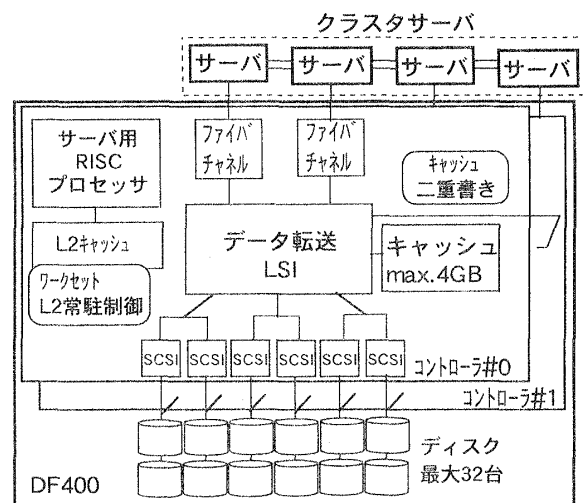


図1. DF400アーキテクチャ

### 3. 高I/O処理技術

#### (1) 目標の設定

トランザクションI/O処理性能の目標は、サーバの高性能化率(2年で3倍)に追随することと定め、従来機比3倍に設定した。

#### (2) I/O処理のボトルネック部位

トランザクションI/O処理においては、コントロ

Development of Disk Array "DF400" for High Performance Servers (1)  
 ~Architecture and High I/O Throughput Technique~  
 Naoto Matsunami, Ikuya Yagisawa, Masayuki Yamamoto, Takashi Oeda,  
 Masahiko Sato\*  
 Systems Development Laboratory, Data Storage & Retrieval Systems  
 Division\*, HITACHI, Ltd.

RAID: Redundant Arrays of Inexpensive Disks  
 SCSI: Small Computer System Interface  
 IOPS: Input / Output Per Second  
 RISC: Reduced Instruction Set Computer

一上の性能ボトルネック部位はプロセッサのみである。よって上記目標より、I/O処理当りのプロセッサ処理時間を従来比1/3に低減することが必要になる。

(3)処理時間短縮化の実現手法

プロセッサの処理時間を短縮するために、次に列挙する実現手法を検討した。

(a)プロセッサ方式

(i)組込システム専用のプロセッサに代え、サーバ等に用いる200MHz超の高速RISCプロセッサをディスクアレイに適用する。

(ii)サーバ用の高速L2キャッシュをディスクアレイに適用すると共に、制御プログラムのワークセットを常駐させる (L2常駐制御方式)。

(b)インタフェース制御方式 (負荷分散制御方式)

(i)ホストI/F、(ii)ディスクI/F各々のコントローラにハードウェア制御を主とするプロセッサ処理の一部分を分担させる。

(4)性能予測評価手法

プロセッサの処理内容は、(I)インタフェース (ハードウェア) 制御部、(II)非インタフェース制御部、に大別できる。上記の実現手法(a)(b)各々の効果を定量的に予測評価するため、プロセッサ処理時間Tの算出式を次のように定式化した。

$$T = \{s + (1-s)/k\} \times (T_s - T_{ol}) + T_g/k \dots (式1)$$

$T_s$ : 従来機のインタフェース制御部の処理時間  
 $T_g$ : 従来機非インタフェース制御部の処理時間  
 $T_{ol}$ : インタフェースコントローラへの一部処理のオフロード時間  
 $k$ : プロセッサの高速化倍率  
 $s$ : ハードウェア依存率 (プロセッサ高速化効果がない処理部分の割合)

(5)ケーススタディ

上記実現手法の効果を検証するため、表1、表2に示すバリエーションでケーススタディを行う。

表1. プロセッサ方式比較

方式	プロセッサ	ベンダ	(i)高速プロセッサ	(ii)L2常駐制御	高速化倍率*1	備考
A	P-a	X社	×(66MHz)	×	1.1倍	従来機プロセッサの高速版
B	P-b	Y社	○(233MHz)	×	1.8倍	
C	P-c	X社	○(200MHz)	○(3-1-1-1)*2	2.7倍	
D	P-d	Z社	○(200MHz)	○(2-1-1-1)*2	4.4倍	DF400採用方式

(注)\*1: 高速化倍率は従来機プロセッサ性能比。\*2: 数値は7GHz速度を表す。

表2. インタフェース制御方式比較

方式	①ホストI/F	②ディスクI/F	ハード依存率	Tol オフロード量*2	備考
1	×	×	61%	0	従来方式
2	△	○	43%	0.17	
3	○	△	33%	0.26	DF400採用方式
4	○	○	23%	0.28	

(注)\*1: ○印はインテリジェントメディアコントローラ使用、△印はノンインテリジェントメディアコントローラ使用。  
\*2: 従来機プロセッサ処理時間を1としたときの相対量。

(6)評価結果

各方式の効果を予測評価した結果を図2に示す。

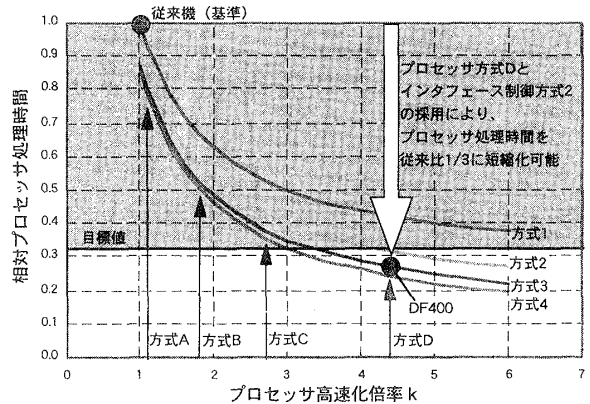


図2. ケーススタディ評価結果

4. 性能評価

図3に、DF400と従来機のI/O処理性能の実機測定に基づく比較結果を示す。上記の予測評価結果通り、従来比3倍の目標高I/O処理性能を実現できた。

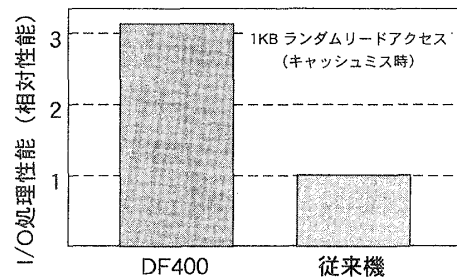


図3. I/O処理性能比較

5. おわりに

サーバ要求に対応した高性能、高信頼を実現するディスクアレイ「DF400」のアーキテクチャを開発した。また、高I/O処理技術として、高速RISCプロセッサとL2キャッシュ常駐制御を採用したプロセッサ方式、およびハードウェア制御をインタフェースコントローラに分担させる負荷分散制御方式を検討し、その効果を検証するための定量的な性能予測評価手法を考案した。これらの高I/O処理技術をDF400に適用することで、目標を上回る従来比3倍超の高I/O処理性能を実現した。

参考文献

[1] David A. Patterson, et al.: "A Case for Redundant Arrays of Inexpensive Disks (RAID)", Report no. UCB/CSD 87/391, Computer Science Division Department of Electrical Engineering and Computer Science, University of California, Berkeley, 1987.