

自動データ移行によるディスクアレイ保守容易化技術の開発

野中 裕介†

西本 哲†

永田 幸司‡

酒井 幸介‡

(株) 日立製作所 システム開発研究所†

(株) 日立製作所 RAID システム事業部‡

1. はじめに

データ量の増加・データ保護の強化へのニーズを背景として、中堅・中小企業においてもネットワークストレージの需要が増加している。中堅・中小企業向けストレージは、ハードウェア等の導入コストが低だけでなく、保守・管理にかかる運用コストもまた低いことが求められる。

保守・管理に、高度な知識・技術を持った専門技術者が必要であれば、それが、運用コストが高くなる要因の1つとして挙げられる。

本研究では、専門技術者による保守を不要とすることにより低い運用コストを実現するディスクアレイの新保守方式を検討し、実現に必要な自動データ移行技術の開発を行った。

2. 保守容易化のための要件分析

2.1 従来の保守方式

24 時間運用のネットワークストレージとして使用するディスクアレイは、高い可用性を実現するために、ハードウェアが冗長構成化されており、障害発生時には、運用を継続しながら部品交換を行う。部品交換の手順を以下に示す：

- (1) 障害部位の特定
- (2) 障害部品の取り外し
- (3) 交換用部品の取り付け

障害部位の特定は、LED 点灯や電子メール等、製品固有の方法によるディスクアレイからの障害部位通知を用いて行うため、この作業には製品固有の知識が必要である。また、障害部位の特定作業や部品交換作業において、取り外し対象を正常部品と誤ると、システム停止等を起こす可能性があるため、訓練された専門技術者により保守が行われることが望ましい。

2.2 新保守方式の要件定義

専門技術者による保守を不要化するため、部品単位で取り外しを行う手順を含まない保守方式とすることを、新保守方式の要件とした。

3. 新保守方式

3.1 アーキテクチャ

部品交換のない保守を実現するために、装置全体交換（以下、アレイ交換）アーキテクチャとした。表 1 に部位毎の新方式の保守手順と実現上の技術課題を示す。

ディスクアレイ内の実装台数が多く、かつ、稼働部位であるため障害発生率が他部位と比較して高いハードディスクの障害時には、空きスロットへハードディスクを追加し、RAID^[1]方式による修復コピーにより冗長性を回復する保守を行う。ハードディスク以外の障害についてはアレイ交換を行う。アレイ交換を実現するには、装置間のデータ移行を容易に実行できるようにすることが課題であった。

表 1 新保守方式

障害部位	保守手順	技術課題
ハードディスク	ハードディスクの追加*1	—
上記以外	アレイ交換	装置間データ移行の容易化

*1: スロットに空きが無いときはアレイ交換

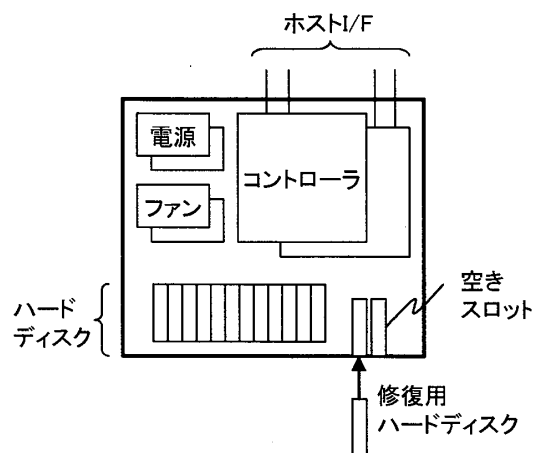


図 1 ハードウェアアーキテクチャ

Development of Auto Migration Function for Disk Array Easy Maintenance

†Systems Development Laboratory, Hitachi, Ltd.

‡Disk Array Systems Division, Hitachi, Ltd.

図 1に、新保守方式に対応するハードウェアアーキテクチャを示す。コンポーネントを冗長構成化しているため、障害発生からアレイ交換までの期間も運用の継続が可能である。また、ハードディスク追加用の空きスロットを持つ。

3. 2 自動データ移行機能

容易な装置間データ移行のためには、運用中に装置全体を複製する機能が必要である。さらに、新しいアレイへのパス切替時にもホスト I/O 処理を停止する必要が無いことが望まれる。

これらの要求を満たす自動データ移行機能を開発した。本機能は、

- ・ 構成情報・データの装置間コピー
- ・ ホストに透過的なパス切替処理

を一括実行する。処理フローの概要を図 2に示す。ホストと移行元アレイは IP ネットワークで接続され、iSCSI^[2]により I/O を行っている。

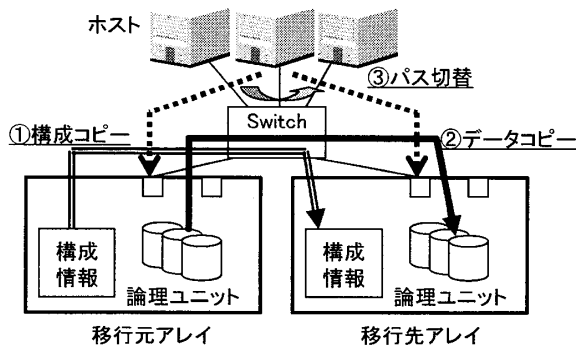


図 2 自動データ移行処理フロー

(1) 構成情報・データの装置間コピー

移行先アレイを IP ネットワークに接続し、自動データ移行の実行を指示すると、移行元アレイは、最初に、移行先アレイとのセッションを確立し、論理ユニットの構成等を含む構成情報を移行先アレイにコピーする(①)。

次に、論理ユニット内のユーザデータのコピーを開始する(②)。①、②の処理は、移行元アレイがバックグラウンドで実行し、この間ホスト I/O は継続可能である。コピー中のデータの更新は、両アレイに反映する。

(2) ホストに透過的なパス切り替え

両アレイの論理ユニット内のデータが一致した状態になった後、移行元アレイの IP アドレス・iSCSI Name 等のアレイ識別情報を、移行先アレイへ引継ぐことによってホストのアクセス先を変更する、パス切替処理を実行する(③)。

パス切替処理では、最初に、移行元アレイの IP アドレスを、一時的な IP アドレスに変更する。これによりセッションが切断されるが、移行元

アレイは、移行先アレイへの再接続を実行し、パス切替を移行先アレイへ指示する。次に、移行先アレイは、IP アドレスを、移行元アレイにパス切替開始前に設定されていた IP アドレスに変更し、Gratuitous ARP をブロードキャスト送信する。これを受信した各ホストは、ARP テーブルを更新し、iSCSI コマンドの発行先を、移行元アレイから移行先アレイに切り替える。

IP アドレスの引継ぎ後に、各ホストと移行先アレイが通信可能なことを保証するため、自動データ移行機能は、パス切替より前に、移行先アレイが移行元アレイと同一のサブネットに接続され、IP アドレスが引き継ぎ可能であることを、開始前の移行先アレイの IP アドレスから確認し、また、コピーの通信路として使用しないホスト I/F がネットワークに正しく接続されていることをテスト通信により確認する。これにより、移行先アレイの接続誤りによるホスト I/O の停止を防止する。

4. 評価

自動データ移行機能により、運用を継続しつつ容易にアレイ交換を実行することができる。表 2に示すように新保守方式は、部品の取り外しを含まないため、専門技術者以外による保守に、適した方式であると言える。

表 2 保守方式比較

	部品取り外し	部品追加	アレイ交換
難易度	高	低	中
従来方式	要	要	不要
新保守方式	不要	要	要

5. おわりに

部品の追加とアレイ交換で成る、ディスクアレイの新保守方式を提案した。また、運用を停止せずにアレイ交換を実行することができる自動データ移行機能を開発し、保守コストの低いディスクアレイ実現の見通しを得た。

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.
- [2] J. Satran, et al.: RFC 3720: Internet Small Computer Systems Interface (iSCSI), IETF, 2004.