HPC 向け高速・大容量ストレージの省電力化を図る階層ストレージアーキテクチャと階層管理方式の提案

赤池　洋俊† 藤本　和久‡ 岡田　尚也‡ 三浦　健司‡ 村岡　裕明‡

（株）日立製作所 システム開発研究所† 東北大学電気通信研究所‡

1. はじめに

近年、IT 機器の消費電力は無視できないほど増加しており[1]、大きな問題となっている。ストレージシステムはその中でも多くの電力を消費するシステムの一つである。特に HPC 分野では、計算機の性能向上に伴いデータ容量が著しく増加していることから、ストレージシステムの大規模化の要求が強く、今後さらなる消費電力の増加が予想される。

スーパーコンピュータと接続するストレージシステムには大量のデータを高速に入出力することを目的として高い性能が要求される。そのため、性能を維持しながら消費電力を削減するストレージアーキテクチャと、その管理方式が求められている。

2. 従来手法

データセンターでは、高性能なオンラインストレージ（以下、OL）と大容量のニアラインストレージ（以下、NL）の階層構成が用いられている。従来システムの概要を図 1 に示す。

従来手法では、①使用頻度の低いデータを OL から NL へデータ移行するデータ管理を行っている。データを使用頻度に応じて適切な場所に保存することで、ストレージシステムの容量を効率的に利用できる。また、②アクセスのない NL のディスクをスピンダウン制御することにより、省電力化を行っている

統計的にデータの使用頻度は、作成時からの経過時間が長いほど小さい。そこで、予め経過時間の閾値をユーザのポリシーとして設定しておき、閾値よりも経過時間の長いデータを使用頻度が小さいとみなしてデータ管理する。

閾値を短く設定した場合、OL 上に保存されるデータ容量は小さくなるため、OL のサイズを小さく設計することができる。これは結果として OL の電力を削減できることになる。しかし、OL のサイズ減少分のデータが NL に保存されるため、NL のデータアクセス頻度が高くなり、性能低下が発生してしまう。ポリシーの閾値を長く設定した場合はこの逆で、性能低下が発生しない一方で、OL のサイズを大きく設計する必要があるため電力が増加してしまう。

この様に従来方式では性能と電力の間にトレードオフの関係が存在している。性能を維持しつつ、省電力化することが課題となる。

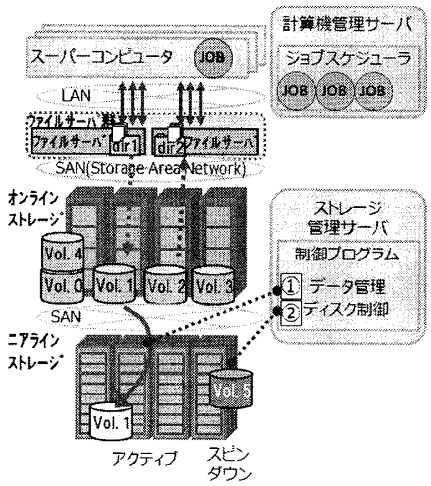

図 1．従来手法の概要

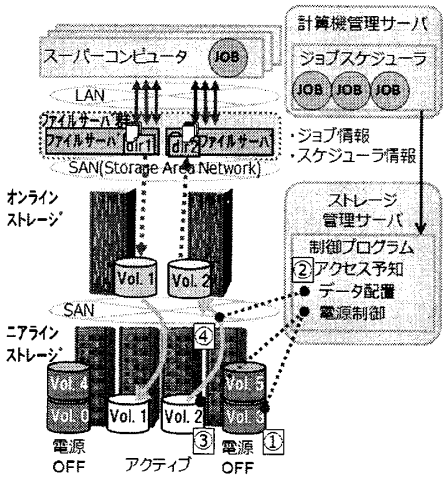

図 2．提案手法の概要

Tired Storage Architecture and Tiring Management Method for Saving Energy Consumption of High-Speed Mass Storage in HPC Systems

† Hirotoshi Akaike, Systems Development Laboratory, Hitachi, Ltd.

‡ Kazuhisa Fujimoto, Naoya Okada, Kenji Miura, Hiroaki Muraoka, RIEC, Tohoku University.

3. 提案手法

本研究では上記課題を解決し、性能を維持しながら消費電力を削減する手法を提案する。概要を図2に示す。従来手法と同様にOLとNLの2階層構成とする。違いはストレージの管理方式にあり、次の動作を行う。①通常は全データをNLに保存し、ディスクの電源をOFFする。②スーパーコンピュータからのアクセスを予知する。③アクセス前に予めディスクの電源をONにし、④データをOLへコピーする。ユーザがログインしてアクセスする時は、NLのディスクの電源をONに制御する。

この管理方式により、スーパーコンピュータがアクセスするデータのみをOL上に配置することができ、性能を維持しながらOLのサイズを減らすことで、省電力化が可能となる。そして、アクセスされないNL上のディスクを電源OFFにすることで、従来のスピンダウンよりも更なる電力を削減している。

なお、提案手法では、計算機管理サーバのスケジューラ情報やジョブ情報をヒントにして、対象ファイルと、そのアクセス時刻を予測し、データ配置と電源制御を計画する。アクセス予知の精度が高いほど、必要最小限のデータをOLに配置可能で、電力削減が可能となる。データ配置とそのタイミングが従来と逆方向に管理されている点に特徴がある。

4. 評価

システム全体の容量を1024TB、従来手法のOLサイズを60TB、提案手法のOLサイズを半分の30TBとした時で、消費電力の評価を行った。従来手法ではOLとNLにユーザデータを保存できるのに対して、提案手法ではNLのみになることに注意されたい。これは、OLを高速なデータアクセスを目的とした一時保存領域として用いているためである。OLの電力は、1日に24時間アクセスを受けると仮定して最悪値として評価する。なお、ユーザは1日に6時間システムへログインし、その中の3時間がディスクアクセスであると仮定する。

OLとNLのディスク筐体とコントローラの各状態の消費電力を見積もったところ、表1に示す結果を得た。提案手法は従来手法と比較して52.9%の電力削減が可能であることが分かった。

システム全体の容量が512TB, 256TBの場合で同様の評価を行った。結果を図3に示す。3つのどの容量でも、提案手法は省電力効果を示している。そして、システム容量が大きいほど、従来法と比較して電力量削減の割合が大きくなることが分かる。これは、コントローラで消費される電力の割合が小さくなり、その分提案手法による省電力化が寄与するためである。

システム構成要素		電力[w]	従来手法 時間[h]/1日	台数	電力量[kwh](1日分)	提案手法 時間[h]/1日	台数	電力量[kwh](1日分)
OL			容量60TB		**165.6**	容量30TB		**93.4**
ディスク筐体(2.1TB) HDD 300GB (7D+1P)	アクセス	215	24	29	149.6	24	15	77.4
	アイドル	205	0		0	0		0
コントローラ		668	24	1	16.0	24	1	16.0
NL			容量964TB		**469.0**	容量1024TB		**205.8**
ディスク筐体(8TB) HDD:1TB (6D+2P)	従来のデータ管理	232	8	121	224.6	—	128	—
	提案手法のデータ配置	232	—		—	3		89.1
	ユーザアクセス	232	0		0	3		89.1
	アイドル	202	0		0	0		0
	スタンバイ	112	16		216.8	—		—
	電源OFF	0	—		—	18		0
コントローラ		288	24	4	27.6	24	4	27.6
システム全体			容量1024TB		**634.6**[kwh]	容量1024TB		**299.2**[kwh]

表1. 従来手法と提案手法の消費電力

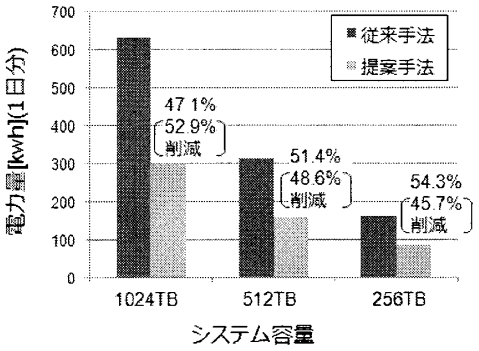

図3. システム容量と電力量の関係

5. まとめ

HPC向け高速・大容量ストレージの省電力化を図る階層ストレージアーキテクチャと階層管理方式を提案した。アクセス予知に基づくデータ配置と電源制御により、性能を維持しながら消費電力を削減できることを示した。今後は、さらなる省電力化を目指して、高い精度のアクセス予知手法を実現することが課題となる。

謝辞 本研究は、文部科学省の委託研究「高機能・超低消費電力スピンデバイス・ストレージ基盤技術の開発」の成果の一部である。

参考文献

[1] "Report to Congress on Server and Data Center Energy Efficiency Public Law 109-431", U.S. Environmental Protection Agency, ENERGY STAR Program, Aug. 2007.