アプリケーション協調型大規模ストレージ省電力システムの開発と DSS を用いた評価

西川　記史†,　中野　美由紀†,　喜連川　優†
†東京大学　生産技術研究所

1. はじめに

デジタルデータの急増に伴い，データセンタの運用コストは増加の一途を辿っている．特にストレージの電力コストの成長率は他のコストを圧倒しており[1]，データセンタの運用コストの削減のためにはストレージの電力コストの低減が最重要課題となっている．

我々は，これまでアプリケーション実行時ストレージ省電力化にアプリケーションの I/O 挙動特性を活用する研究を行ってきた[2]．本論文では，開発した手法を実際のストレージ上に省電力管理システムとして実装する．さらに TPC-H を稼働し，省電力化しない場合のクエリ応答時間を保ちつつストレージ消費電力が大幅に削減できることを示し，提案手法が商用大規模ストレージ上で，有効に動作することを確認する．

2. アプリケーション協調型大規模ストレージ省電力方式

我々が提案しているアプリケーション協調型大規模ストレージ省電力手法[2]について述べる．本手法の特長は，i) アプリケーション実行時のストレージ省電力，ii) アプリケーションレベルにおける入出力発行間隔の長さや read/write 入出力の頻度等のモニタリング結果に基づくアプリケーションレベルでの入出力挙動のパターン化，及び iii) アプリケーションレベルの入出力挙動のパターンに基づく，適切なストレージ省電力手法の選択及び適用，である．

2.1 省電力ストレージモデル

我々が提案するアプリケーション協調型大規模ストレージ省電力を実現する省電力ストレージモデルを図 1 に示す．

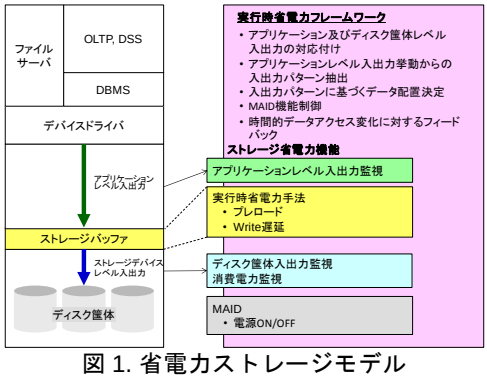

図 1. 省電力ストレージモデル

アプリケーションは，バッファに対してアプリケーションレベルの入出力を行う．そして，バッファを用いて先読み，書き込み遅延など省電力に有効な手法をデータごとに適用，その後ストレージへの入出力を行う．本モデルでは，アプリケーションレベル，ストレージレベル入出力の監視を行うと共に，消費電力の推移を計測し，アプリケーション入出力挙動の経時変化に追随するようデータ配置，入出力パターンなどの見直しを行う．

2.2 論理入出力パターン

ストレージの省電力にアプリケーションの入出力挙動特性を利用するために，論理入出力パターンという概念を導入する．論理入出力パターンとは，アプリケーションの入出力挙動をストレージ省電力手法を適用できるように分類・パターン化したものであり，実行時に省電力機能を適切に選択するために使用する指標である．

入出力パターンは次の 4 種類である．第一は，モニタリング期間中にアプリケーションから入出力が発行されなかったことを識別するための入出力パターン(P0)である．本パターンに該当するデータを識別することにより，容易に電源 OFF などのストレージ省電力機能を適用できる可能性が増加する．第二はストレージキャッシュを用いることで read 入出力間隔を延伸できる可能性があるデータを識別するためのパターン(P1)である．第三は同じくストレージキャッシュを用いるが，read ではなく write 入出力間隔を延伸できる可能性があるデータを識別するためのパターン(P2)である．最後は入出力の間隔が短く，ストレージ省電力機能を適用することができないデータを識別するための入出力パターン(P3)である．

論理入出力パターンに従ってストレージ省電力を行うことで，実行時に個々のアプリケーションレベルの入出力挙動ごとに適切な省電力化が図られ，アプリケーション実行時にもストレージ省電力を達成することが可能になると考えられる．

2.3 論理入出力パターンを用いたストレージ省電力

アプリケーション協調型大規模ストレージ省電力方式は，前述の論理入出力パターンを用いてデータ配置制御，及びキャッシュを利用した入出力発行制御を行う．

(1) データ配置制御

ストレージデバイスに省電力機能を適用するためには，ストレージデバイスに対する入出力間隔が，省電力機能を適用できる長さ以上が必要となる．．このために，省電力化が期待できない P3 型のデータは同一のディスク筐体に配置し，残りのデータの入出力発行間隔を省電力機能を適用できる程度に長くすることを試みる．

(2) キャッシュを利用した入出力発行制御

提案手法はキャッシュ上にデータを保持することでプレロード及び write 遅延による入出力間隔の延伸を行う．

プレロード: データがアプリケーションから read される前にキャッシュにロードする．これにより read 入出力間隔を伸ばす．

Development and Evaluation of Application Corroborative Large Storage Energy Saving System using DSS Application.
Norifumi NISHIKAWA†, Miyuki NAKANO†, and Masaru, KITSUREGAWA†
† The University of Tokyo Institute of Industrial Science

Write 遅延: データに対する更新を一時的にキャッシュに蓄積し，まとめてストレージデバイスに書き出すことで write 間隔を伸ばす.

3. ストレージ省電力管理機構の実装

アプリケーション協調型大規模ストレージ省電力管理機構とその実装を図 2 に示す．モニタリング機構は，論理入出力統計，及びストレージ性能・消費電力を収集・管理する．省電力管理機能は，データ要件管理，データ階層・ストレージ階層構築，ファイル配置計算，及びファイル配置変更機能を持つ．実行時省電力機能は，ユーザアプリケーション実行時にアプリケーションに動的にリンクされるストレージ省電力ライブラリ，仮想ファイルツリーファイルキャッシュ，及びストレージ電源制御から構成される．

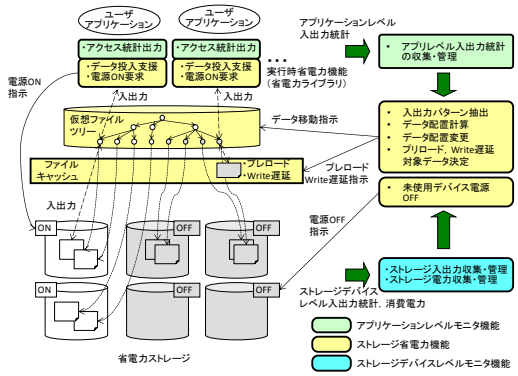

図 2. ストレージ省電力管理機構の実装

4. ストレージ省電力管理機構の評価

ストレージ省電力管理機構を実装したストレージ上で商用 DBMS を用いて TPC-H を実行し，ストレージの消費電力とクエリの応答時間を計測した.

4.1 評価環境

評価に用いたサーバのプロセッサは Intel Xeon X5670 2.93GHz (合計 24 コア)，主記憶は 48GB である．サーバの OS は Red Hat Enterprise Linux 5.4 (64 ビット版)，ファイルシステムは EXT2 である．ストレージは(株)日立製作所製の Hitachi Adaptive Modular Storage 2500(AMS2500)を用いた．AMS2500 は 15 台 7200 回転の SATA HDD を 15 台搭載したディスク筐体(13D+2P RAID6 構成)を 11 台，及び RAID コントローラ 1 台を搭載している．RAID コントローラのキャッシュ容量は 2GB，RAID 構成前のディスク筐体の容量は 11.25TB である．サーバとストレージは 4 本の 2Gbit Fibre Channel 1 本で接続されている.

DSS プログラムとして， DSS の代表的ベンチマークである TPC-H ベンチマーク[3]を用いて計測を行った．DB サイズは約 1.2TB(Scale Factor 300)，DBMS バッファサイズは 40GB とした．ログ及び作業表をディスク筐体 1 台に，表と索引を残りのディスク筐体 10 台にキーレンジ分割機能を用いて分散配置した．上記環境において，単一スレッドにて TPC-H のクエリ 1 から 22 までを順次実行し，ストレージの消費電力とクエリの応答時間を計測した.

4.2 評価結果

省電力ストレージ管理機構を利用しない場合(省電力制御なし) と省電力ストレージ管理機構を利用した場合(提案手法) のストレージ消費電力の平均値及びクエリ Q2, Q7 の計測結果をそれぞれ図 3 及び 4 にそれぞれ示す.

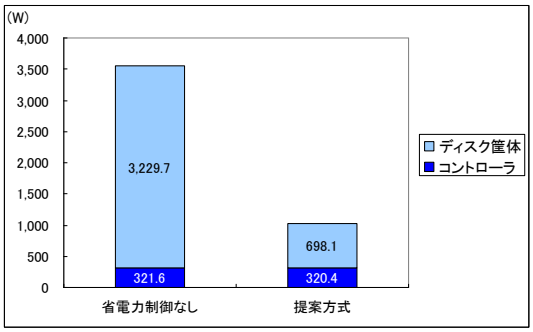

図 3. ストレージの平均消費電力

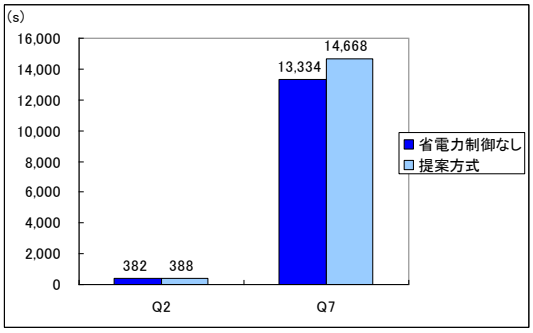

図 4. クエリ応答時間(Q2, Q7)

図から分かるように，提案手法はディスク筐体の平均消費電力を 3229.7 W から 698.1W に約 79%削減できた．これは，入出力が常時行われるデータを 2 台のディスク筐体に集めたことによる効果である．またストレージ応答時間は Q2 はほぼ同等，Q7 が約 9.1%倍増加した．Q7 の応答時間が増加したのは，クエリ実行中のディスク筐体の起動待ち(約 120 秒)，及び入出力が 2 台のディスク筐体の集約されたことによる入出力応答時間の増加のためである．これらの結果は，提案方式が大規模なストレージ上でも動作することを示している.

5. まとめ

本論文では，アプリケーションの入出力挙動を用いることによりストレージの省電力の機会を増加させる，アプリケーションと協調した新たなストレージ省電力方式を提案した．提案手法を実装し，TPC-H を用いて評価し，提案手法が大規模なストレージの消費電力を大きく削減できることを確認した.

参考文献

[1] S W Worth. Green Storage - The Big Picture. In Storage Networking World Spring 2010 Conference, 2010.
[2] Norifumi Nishikawa, Miyuki Nakano and Masaru Kitsuregawa, Energy Efficient Storage Management Cooperated with Large Data Intensive Applications, 28th IEEE International Conference on Data Engineering (IEEE ICDE 2012), 2012 (To Be Appeared).
[3] TPC BENCHMARK™H (Decision Support) Standard Specification Revision 2.14.3, http://www.tpc.org/tpch/spec/tpch2.14.3.pdf