

OLTP を対象としたアプリケーション協調型 大規模ストレージ省電力制御方式の提案及び評価

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

サーバやストレージの集約によるデータセンタの高密度化に伴い、データセンタの消費電力は増加の一途を辿っている。中でも、データセンタが管理するデータ量の急増に伴うストレージの消費電力の増加は著しく、その増加率は年率約 13%と他の IT 機器の消費電力の増加率を圧倒している[1]。データセンタにおけるストレージの電力消費量は急激に高まるものと考えられ、その消費電力の削減はデータセンタにおいて喫緊に解決しなければならない重要な課題である。

また、データセンタで使用されるハイエンドストレージのアプリケーションに目を向けると、その約 6 割は DBMS 用途に出荷されており [2]、さらにその半数以上が Enterprise Resource Planning (ERP)や Customer Relationship Management (CRM)などの Online transaction processing (OLTP)向けに出荷されている。即ち、OLTP はデータセンタにおける主要なアプリケーションの一つである。

本論文では、OLTP 系アプリケーションが用いるストレージの消費電力手法の確立を目指し、OLTP の I/O 挙動特性を用いたアプリケーション協調型のストレージ省電力機構を提案する。OLTP の I/O 挙動特性と消費電力特性を実測に基づき明らかにすると共に、シミュレーションにより提案手法の評価を行う。

2. ストレージの消費電力特性

大規模ストレージ向けの省電力制御方式を開発するに当たり、我々は市販のストレージに電力計を取り付け、その消費電力特性を計測した。ストレージは 1 台のコントローラ筐体と 10 台のディスク筐体を持つ。コントローラ筐体は 2 つの I/O プロセッサ(各 2 cores)と 2GB のキャッシュメモリを有する。ディスク筐体はそれぞれ 15 台の 750GB, 7200 回転の SATA ディスクを有し、RAID6(13D+2P)構成を取る。また、本ストレージはディスク筐体内のディスクの電源のみを OFF にする Spin down, 及びディスク筐体全

体の電源を OFF にする Power off の 2 種類の省電力機能を有している。各筐体は 2 本の 200V 電源ケーブルを持つ。我々はこの電源ケーブルにクランプオンセンサを装着し、電力計測を行った。

負荷生成サーバとストレージは 4 本の 4Gbps ファイバチャネルで接続されている。サーバは 32 台のプロセッサ(各 2 cores)と 512GB のメモリを有し、OS は AIX5.3 64 ビット版、ファイルシステムは JFS2 である。また、ストレージのディスク筐体の容量はそれぞれ 11.25TB, 合計 112.5TB である。

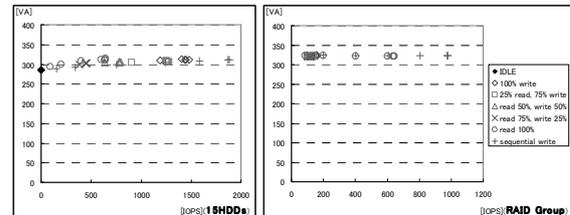


図 1. I/O 時のディスク筐体消費電力(左)とコントローラ筐体消費電力(右)

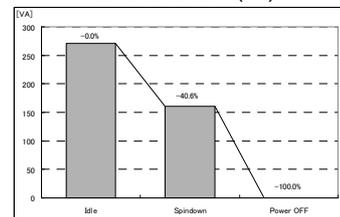


図 2. ストレージ省電力機能使用時の消費電力

ディスク筐体及びコントローラ筐体の消費電力の計測結果を図 1 に、省電力機能使用時のディスク筐体の消費電力を図 2 にそれぞれ示す。図 1(左)からわかるように、ディスク筐体の消費電力は I/O 数が増加するに従って増加している(アイドル時+10.6%)。一方コントローラ筐体の消費電力は I/O 数が増加してもほとんど変化していない。また、図 2 より Spin down 機能を用いた場合のディスク筐体の消費電力は-40.6%、Power Off 機能を用いた場合の消費電力は 0 であることが分かる。このことから、高い省電力効果を得るためにはディスク筐体の省電力機能を最大限活用しなければならないことが分かる。

3. OLTP 系 DBMS の I/O 挙動特性

OLTP 系アプリケーションの I/O 特性の活用方式を検討するために、我々は TPC-C ベンチマークの簡易実装であ

Propose and Evaluation of Application Corroborative Storage Power Saving Function for OLTP.

Norifumi NISHIKAWA[†], Miyuki NAKANO[†], and Masaru, KITSUREGAWA[†]

[†] University of Tokyo Institute of Industrial Science

る tpcc-mysql[3]を用い OLTP 系アプリケーションの I/O 特性の調査を行った. 図 3(上)に DBMS からストレージに対して発行された秒当り I/O 数を, 図 3(下)にストレージコントローラからディスクに発行された秒当り I/O 数を示す.

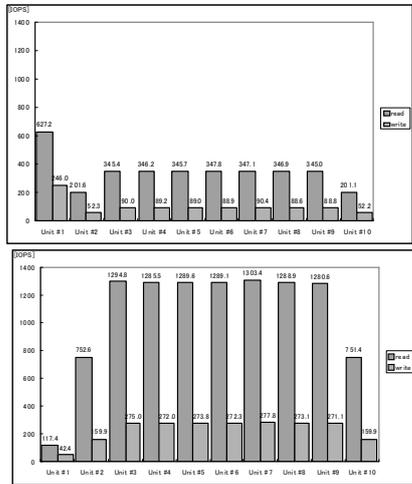


図 3. ディスク筐体毎の秒当り I/O 数
(上段: 対コントローラ, 下段: 対ディスク間)

ディスクに対する I/O 数はディスクへの write に伴う RAID のパリティ生成のために OLTP がストレージに発行した I/O の数倍に達しディスクが性能ボトルネックとなっていることが分かった. これは, write 数を削減することによりパリティ生成数を削減できればディスクの負荷を大きく下げることができ, この結果アクセス頻度が高いデータをより少数のディスク筐体に集め他のディスク筐体に省電力機能を適用するチャンスが生じることを示していると言える.

そこで我々は, ストレージキャッシュのサイズを大きくした場合にディスクへの I/O 数がどの程度削減できるかを調査した. この結果を図 4 に示す. 図 4 から分かるように, DB 容量の 5% のキャッシュを追加で用意することにより, write 数を 20% 低減できることが分かる.

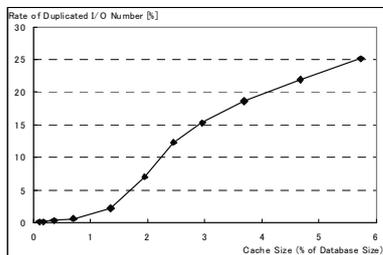


図 4. ストレージキャッシュ容量と write 数削減率

4. TPC-C の I/O 挙動特性を用いた省電力制御

TPC-C 向けのストレージの省電力化を行うため, 我々は TPC-C が持つ I/O の局所性に着目した省電力制御手法を提案する. その主たるアイデアは, i) 適切なサイズのストレージキャッシュの割当てによるディスク筐体に対する

write 数の削減, ii) i)により生成されたディスク筐体の I/O 余力を用いて高アクセス頻度のデータを少数のディスク筐体への集約と残りのディスク筐体への省電力機能の適用, iii) ディスクへの write の遅延によるアイドル時間の延伸, である.

5. 評価結果

ストレージキャッシュサイズを変化させた場合の消費電力とトランザクションスループットを図 5 上・下に示す. 図 5 から分かるように, 提案手法によりトランザクションスループットの低下をほぼゼロにしつつストレージの消費電力を最大 44.9%削減することが可能である.

消費電力が削減できた理由はアクセス頻度が高いデータを少数のディスクに集約できたためである(キャッシュ容量が DB の 20% の場合 4 ディスク筐体). また, write 遅延は DBMS の I/O とは非同期であり性能にはほとんど影響を与えていない.

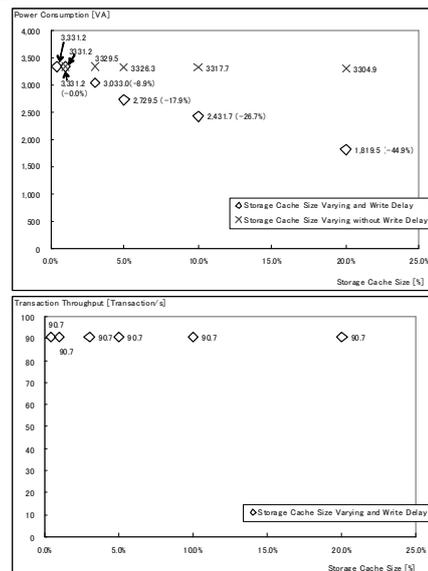


図 5. 消費電力とトランザクションスループット

5. まとめ

本論文では, OLTP 系 DBMS の I/O 挙動特性をもちいることにより, ストレージの大幅な省電力化が可能になることを示した,

参考文献

- [1] D. Reinsel, The Real Costs to Power and Cool All the World's External Storage, IDC #212714, 2008
- [2] Reinsel, D., White Paper Datacenter SSDs: Solid Footing for Growth, IDC #210290, 2008.
- [3] tpcc-mysql, <https://code.launchpad.net/~percona-dev/perconatools/tpcc-mysql>.