

仮想化環境における VM ごとの I/O アクセス頻度を考慮した

テーブル配置によるストレージ省電力に関する一考察

谷貝 俊輔[†] 若色 匠[†] 山口 実靖[‡]

[†]工学院大学大学院工学研究科電気・電子工学専攻 [‡]工学院大学工学部情報通信工学科

1. はじめに

近年,情報技術が普及しデータセンター等において多数のサーバ計算機が稼働するようになった. これに伴い,サーバの消費電力の増加が問題となっている[1]. この問題に対する解決策の一つとして,アプリケーションの動作情報を用いてディスク上のデータレイアウトを変更し HDD の消費電力削減する方法がある[1][2].

本研究では上記手法を仮想化環境に適用し,仮想計算機 (VM)ごとのアクセス量を考慮したデータ再配置を行なう. 具体的には Xen を用いた仮想計算機上に MySQL を立ち上げTPC-C実行時,各テーブルのアクセス量を観察する. そしてアクセス量を考慮したテーブルの再配置を行い,HDD アクセス間隔の拡大および消費電力の低減の程度とトランザクション性能の評価を行い,評価によりその有効性を検証する.

2. 応用情報を用いたストレージ省電力

西川らは,データ(テーブル)へのアクセス頻度を考慮しディスクへのデータ配置を制御することにより,ディスクに省電力機能を適用できるだけの I/O 発行間隔を生成する手法を提案している[1][2]. 本手法ではアクセス数が多いデータを Hot データ,アクセス数が少ないデータを Cold データと呼び,図 1 の様にこの Cold データをひとつの HDD に集中させることで特定の HDD のアクセス間隔を拡大させ省電力化をはかっている. また Cold Data に対して DBMS の機能を用いて専用のバッファを割り当て,HDD アクセスを削減し HDD アクセス間隔の拡大を図る.

図 2 にストレージの停止時と再起動時の消費電力の変化について示す. ストレージ停止により削減できる電力量と

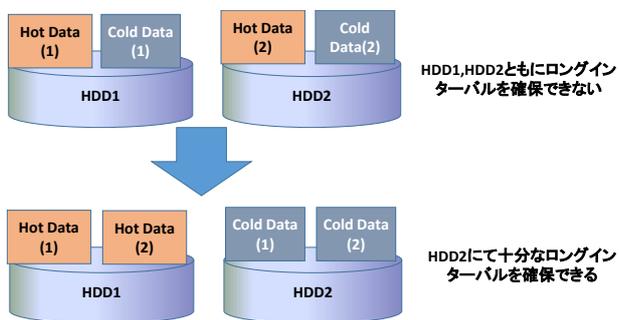


図 1 応用情報を用いたストレージ省電力

“Energy Efficient Storage Management Based on DB Table Layout Optimization Considering Table Access Frequency in Virtualized Environment” †Shunsuke Yagai, Takumi Wakairo, Electrical Engineering and Electronics, Kogakuin University Graduate School ‡Saneyasu Yamagushi, Department of information and Communications Engineering, Kougakuin University

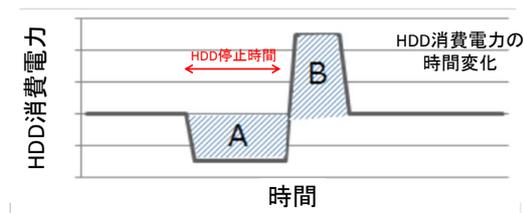


図 2 停止と再起動時の電力の変化

ストレージ再稼働により失われる電力量が等しくなる ($A=B$ となる) ストレージ停止時間をブレイクイーブンタイムと呼び,それより長くなる ($A>B$) ストレージ停止時間をロングインターバルと呼ぶ. 当該手法では配置制御によりロングインターバルを作り出すことで,省電力化を実現している. またブレイクイーブンタイムは HDD の実装に依存する.

3. 仮想環境における応用情報を用いたストレージ省電力

本章で,仮想化環境においてアプリケーションの動作情報を用いてデータ配置を制御し,特定の HDD におけるアクセス間隔を拡大する手法を提案する. アプリケーションとしては TPC-C を想定する.

本手法では,TPC-C 実行時の DB テーブルファイルへのアクセス要求数をカーネル内で監視し,各ファイルの書き込み,読み込み要求ごとのアクセス頻度を調査する. そしてアクセス要求の少ないテーブルファイルを特定の HDD 上に集中して配置する. 書き込み要求はキャッシュに溜め込み一括書き込みさせること (遅延書き込み) が可能なためアクセス西川らの手法[1][2]においては,DBMS の機能を用いて読み込みファイルをメモリ内に保持させているが本実験で使用している DBMS の実装には同等の機能が無い. そこで,アクセス数が少なくかつ読み込み要求のみのファイルは RAM Disk 上に配置する. これにより,要求の少ない HDD のアクセス間隔をより大きくできると予想される. これらのファイルは書込が行なわれないファイルであるため,揮発性メモリ上に保存してもデータの損失には繋がらないと考えられる. また別の HDD 群へのアクセス頻度は均等になるようアクセス頻度が高いファイルは別々の HDD に分散配置し,性能の劣化を抑える.

4. 性能評価

4.1. 評価方法

仮想化システム Xen を用いて 1 台の物理計算機上に 2 台の VM (VM1, VM2) を起動させ,両 VM 上に MySQL を立ち上げた. そして VM ごとにサイズが異なる TPC-C の表を作成した. VM 1 には SF (スケールファクター) が 4 でサイズ 0.38 [GB] の DB を, VM 2 には SF が 64 でサイズが 5.8 [GB]

のDBを作成した。表の配置方法としては後述の3通りを用意し比較した。HDDは4台(HDD1からHDD4)使用し、HDD1にはゲストOSのシステムファイルを格納し、ほかの3台のHDD(HDD2,HDD3,HDD4)にはMySQLのテーブルファイルを格納した。また本実験の計測時間は1200秒とする。

一つ目の配置方法では、HDD間でテーブルサイズの合計が均等になるよう配置する。この方法はアクセス頻度を考慮しておらず、本稿ではこれを標準的な配置方法として考える。この配置手法を“サイズ均等”と呼ぶ。

二つ目の配置方法はアクセス頻度を考慮した配置方法である。図3がアクセス頻度の観察結果である。図3より、テーブルのアクセス頻度は、DBのサイズ(SF)によって変化することがわかる。本配置方法ではアクセス頻度の合計が閾値の、 $1/(\text{ブレークイブタイム} \times 2)$ 以下になる様、アクセス頻度が低いものからHDD4に配置する。本実験で使用したHDDのブレークイブタイムは15秒である。残りのファイルはHDD2とHDD3に配置するが、図3により見積もられるアクセス頻度が同等になる様に配置を行なう。具体的には、まずアクセス頻度が高いファイルをHDD2に配置する。そして次にアクセス頻度が高いファイルをHDD3に配置する。それ以降のファイルは、HDDの合計アクセス頻度が低いHDDに配置する。これを繰り返してアクセス頻度が同等になる様にする。この配置手法を“提案配置”と呼ぶ。

三つ目の配置手法は二つ目の配置手法にRAMDiskを適応した配置方法であり、書き込みがなく読み込み要求のみのitemファイルをHDD4からRAM Diskに配置する。この配置手法を“提案配置RAM”と呼ぶ。

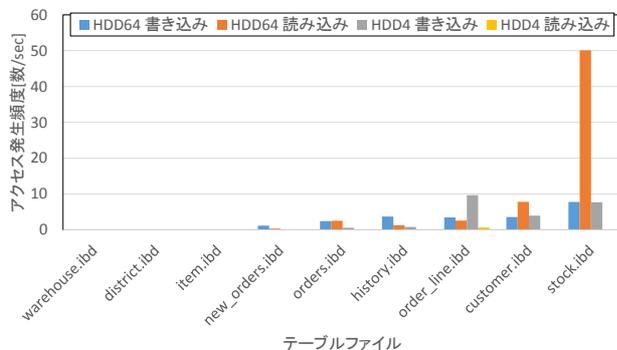


図3 各ファイルのアクセス頻度

4.2. 評価結果

図4,5,6に全配置方法におけるトランザクション性能、HDD4のアクセス間隔の発生頻度分布、HDD4の実験で消費した電力の合計を示す。図5において、アクセス間隔は0-1[sec]の発生回数は省いている。

図4より、サイズ均等配置、提案配置と提案配置RAMの性能はほぼ同等であることがわかる。性能がほぼ同等である理由は、サイズ均等配置ではHDDのアクセス頻度が均等化されていないためであると予想される。図5より、提案配置により100秒を超えるロングインターバルを確保することができ、ロングインターバルが複数回発生していることを確認できる。次に図6よりHDD4が常時稼働しているサイズ均等と比べ提案配置では8000[w]の消費電力を低減できており、RAMを用いた場合ではさらに2000[w]以上の低減を出来ていることを確認できる。以上より、提案配置により性能

劣化なくロングインターバルの確保、消費電力の低減が可能であることが確認でき、またRAM Diskを用いる手法の有用性が確認できた。

5. まとめ

本研究では、応用情報を利用したHDDのアクセス間隔の拡大手法に着目し、それをVMごとのデータサイズが異なる仮想化環境に適用し性能評価を行った。評価の結果、本手法による大幅なHDDアクセス間隔の拡大、消費電力の低減が確認できた。

今後は3台以上の仮想計算機上での実装、メモリの変動による性能の変化、アクセス間隔のさらなる拡大について考察していく予定である。

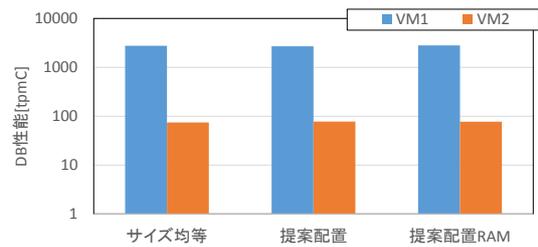


図4 トランザクション性能

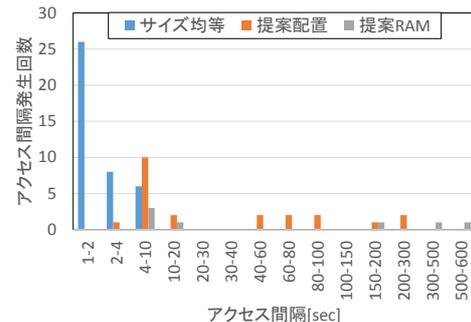


図5 HDD4のアクセス間隔頻度分布

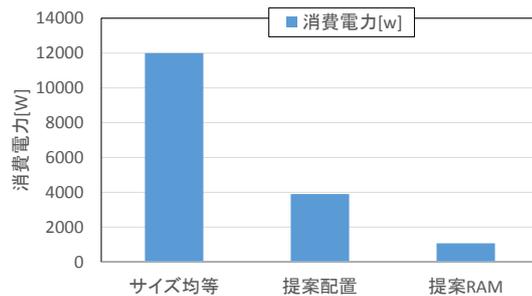


図6 HDD4の消費電力

謝辞

本研究はJSPS科研費24300034, 25280022, 26730040の助成を受けたものである。

参考文献

- [1] Norifumi Nisikawa, 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)
- [2] 西川 記史, 中野 美由紀, 喜連川 優 "アプリケーション処理のI/O挙動特性を利用したディスクの実行時省電力手法とその評価: オンライントランザクション処理における省電力効果" 電子情報通信学会論文誌, J95-D, 3, 1-13(2012.03)