

応用の動作情報を用いたデータ保存位置の最適化による

仮想化環境におけるストレージ電力の低減

Storage Power Saving Cooperated with Application in Virtualized Environment

谷貝 俊輔⁺ 山口 実靖⁺

Shunsuke Yagai Saneyasu Yamaguchi

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]。

我々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てデータ再配置を行なう手法を提案し、有効性を確認した[3]。本稿ではこれをさらに発展させ、HDD 上に巨大なイメージファイルを作成するという VM の特性を考慮した手法について考察を行う。

2. 応用情報を用いたストレージ省電力

2.1 物理計算機における省電力

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 データと呼ぶ。ある HDD から Hot データを取り除き、Cold データのみをその HDD に格納することで、その HDD のアクセス間隔を拡大させ省電力化を実現している。

ストレージは、停止させるとその間消費電力が低減し消費電力の削減が可能となる。しかし、再稼働(spin-up)時に通常以上の電力を消費する。よって、停止によって削減した電力量が再稼働により増加した電力を上回らないと消費電力の削減は実現できない。当該研究では削減電力が上回る停止時間をロングインターバルと呼び、配置制御によりロングインターバルを作り出している。

2.2 仮想環境への適用

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[3]では、アプリケーションとして TPC-C を想定し TPC-C 実行時の各 DB テーブルファイルへのアクセス要求数を調査する。アクセス要求数はカーネル内で監視する。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。これにより、要求の少ない HDD のアクセス間隔が大きくなると期待される。それ以外の 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。本手法はアクセス頻度情報を用いて各 HDD へのファイル配置(HDD 間のファイル配置)を決定しているが、各 HDD 内におけるファイル配置の考察は行っていない。

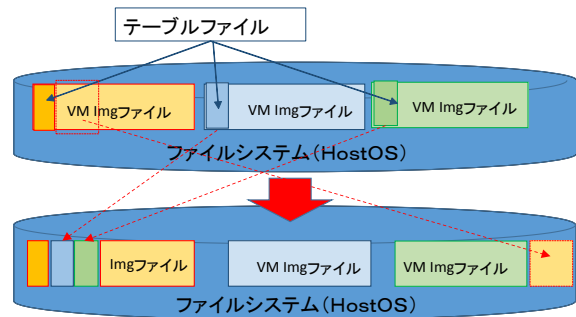


図1 データマイグレーションによるデータ再配置

3. 提案手法

複数の VM が稼働する環境では、単一の HDD 内に複数の VM イメージファイルが格納され、物理 HDD のヘッドが各 VM イメージファイル間の移動を繰り返す。一般に VM イメージファイルはサイズが大きく、この移動が性能劣化の大きな原因となると考えられる。本稿では、ファイルシステムの inode 情報のブロックアドレスを変更してファイルをブロック単位で再配置し、性能劣化をさらに低減する手法を提案する。具体的には、図1のようにイメージファイル内のテーブルファイルが存在するブロックを HDD の先頭の領域に集中して配置し、イメージファイル間のシーク距離の削減とそれによる VM の I/O 性能向上を図る。再配置前に HDD の先頭の領域に配置されていたブロックは、HDD の後方や移動により空く領域に配置する。

4. 性能評価

4.1 実験方法

仮想化システム Xen を用いて1台の物理計算機上に3台の VM(VM1, VM2, VM3)を起動させ、全 VM 上に MySQL を立ち上げた。そして全 VM 上に SF (スケールファクター) が 64 で DB サイズが約 6[GB]の TPC-C のテーブルを作成した。テーブルの配置方法としては後述の3通り(サイズ均等と、省電力配置、提案配置)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、他の3台(HDD2 から HDD4)には MySQL のテーブルファイルを格納した。各 HDD の VM イメージファイルの容量は 500[GB]である。また本実験の計測時間は 1200 秒とした。

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、本稿ではこれをナイーブな配置方法と考える。この配置手法を“サイズ均等”と呼ぶ。二つ目の配置方法は 2.2 節の省電力手法に基づきアクセス頻度を考慮し

て配置した配置方法である。図2がアクセス頻度の観察結果である。本配置方法では、アクセス頻度の合計が閾値の1/(ブレークイーブンタイム×2)以下になる様に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のブレークイーブンタイムは15秒である。残りのファイルはHDD2とHDD3に配置するが、図2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最も高いファイルを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。以下この配置手法を“省電力配置”と呼ぶ。次に省電力配置に加え、3章の提案手法によりHDD内再配置を行ない各VMのテーブルファイルをHDD内の近隣に配置する。今回は、VM1のイメージファイルの5GB目から15GB目のブロックをHDDの後方に移動し、その空いた領域にVM2とVM3のイメージファイルのテーブルを含むブロックを配置した。以下この配置手法を“提案配置”と呼ぶ。

以上の配置方法を用いて、DB性能を測定し、消費電量を推定した。消費電力の推定は、HDD4へのアクセスをLinux OSのSCSI層で取得し、それを元にHDDの状態(稼働中、停止中)を推定し、消費電力の推定値を算出した。

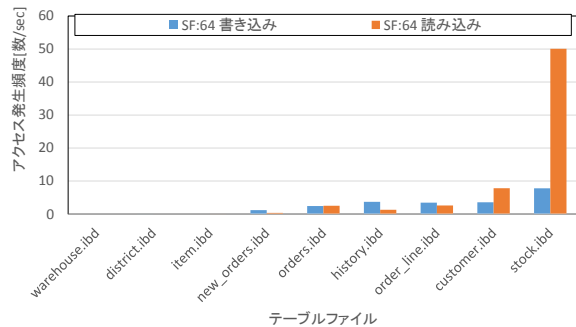


図2 各ファイルのアクセス頻度

4.2 評価結果

図3,4,5に各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の推定消費電力を示す。ただし、図4においてアクセス間隔0-1[sec]の発生回数は省いている。

図3より、サイズ均等と提案配置の性能差が小さい(10%以下)ことが確認できる。これは、HDD内再配置を行わない省電力配置より小さな性能劣化である。提案配置ではHDDのアクセス頻度が均等化されており、それに加えてHDD内配置最適化によりシーク距離の削減がなされたことによるものと予想される。図4より、提案配置により100秒を超える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。また図5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推定消費電力において9[w]の以上の低減をできていることを確認できる。以上より、提案配置により少ない性能劣化で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た。

5. まとめ

本稿では、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、それを仮想化環境に適用し、仮想化環境においてイメージファイルが使用されることを考慮した

HDD内再配置手法を提案した。評価の結果、提案手法による性能劣化の抑制が確認できた。

今後は4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。

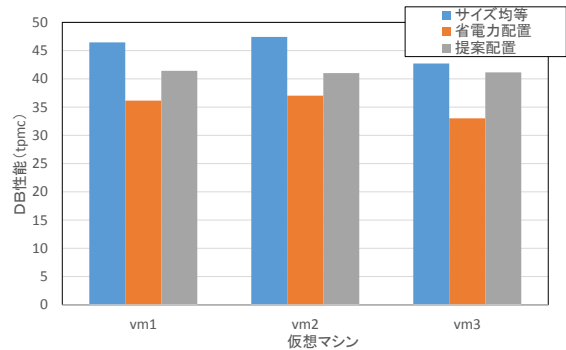


図3 トランザクション性能

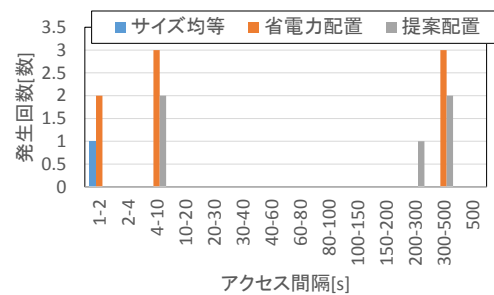


図4 HDD4のアクセス間隔頻度分布

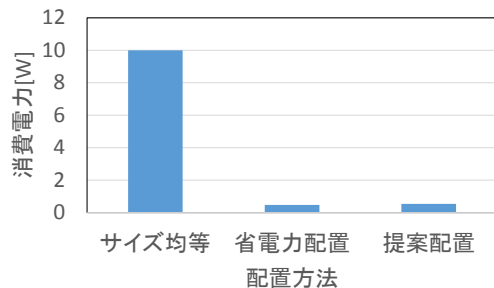


図5 HDD4の消費電力

謝辞

本研究はJSPS科研費25280022,26730040,15H02696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)
- [3] Shunsuke Yagai and Saneyasu Yamaguchi, "Energy Efficient Storage Management Cooperated with Data Intensive Applications in Virtual Machines", The 1st International Workshop on Future Technologies for Smart Information Systems (FTSIS 2014) in conjunction with the 33rd IEEE Symposium on Reliable Distributed Systems (SRDS 2014)(査読有)