

データベースの動作情報を用いた
仮想化環境における HDD のアクセス低減に関する一考察
Increasing HDD Access Interval Considering Database Runtime Behavior

若色 匠[†] 山口 実靖[‡]
Wakairo Takumi Yamaguchi Saneyasu

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、サーバ計算機の消費電力の増加が問題となっている。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 の消費電力を削減する手法[1]がある。

本研究では、計算機システムの Xen によって提供される仮想化環境において HDD へのアクセスについて調査を行う。そして、アプリケーションの動作情報を用いたデータレイアウト変更手法の仮想化環境への適用について考察する。

2. Web サーバへのランダムアクセス

2.1 基本調査

WEB サーバが稼働している VM を 6 台を 1 台の物理計算機上で起動し、他の物理計算機からランダムなアクセス間隔でランダム選択のファイルへの Web アクセスを発行し、HDD へのアクセス間隔と HDD 上のアドレス(ブロックアドレス)を調査した。HDD は 3 台(HDD1, HDD2, HDD3)使用し、各 HDD 上に VM を 2 台ずつ配置した。Web アクセス間隔は HDD1 上の VM から順に、平均が 256 秒, 128 秒, 64 秒, 32 秒, 16 秒, 8 秒の指数分布とした。ファイルサイズは 1MB であり、ファイルは 8192 ある。ファイル選択は平均 1500 の指数分布により行った。ホスト計算機には 1GB のメモリを割り当て、各 VM には 512MB のメモリを割り当てた。CPU は Intel Celeron CPU G1101 2.27GHz, ゲストとホストの OS は Linux 2.6.32.57, HDD は WD5000AZRX-0, 500GB を用いた。

測定結果を図 1 の"基本調査"に示す。アクセスの少ない HDD1 のアクセス間隔が最も長く、約 4 分であることが分かる。HDD1 においては HDD の停止による省電力が可能であると考えられる。

2.2 ブロック移動

前節の結果より HDD1 へのアクセスの間隔が長いことが分かった。そこで、HDD1 上のアクセスの多い(1 時間で 5 回以上)ブロックを HDD2 に移動し、HDD1 のアクセス間隔のさらなる拡大を図った。

測定結果を図 1 の"ブロック移動"に示す。ブロック移動により HDD1 へのアクセス間隔の拡大が達成できたことが

[†] 工学院大学大学院工学研究科電気・電子工学専攻
Electrical Engineering and Electronics, Kogakuin
University Graduate School

[‡] 工学院大学工学部情報通信工学科
Department of information and Communications
Engineering, Kogakuin University

分かる。HDD2 へのアクセス間隔が短縮されてしまっているが、このような短いアクセス間隔では HDD の停止による省電力の実現ができないため省電力の側面からは弊害がないと考えられる。

2.3 不均等メモリ割当

平均 HDD アクセス間隔の長い HDD1 上の VM2 台に多くのメモリを割り当てることにより、これら VM におけるキャッシュヒット率の向上を図り、HDD1 の平均アクセス間隔のさらなる拡大を図った。メモリ割り当て量を HDD1 上の両 VM に対して 1024MB, 他の VM に対して 256MB として測定を行った。

測定結果を図 1 の"メモリ割当"に示す。前節同様に HDD1 のアクセス間隔の拡大が実現されたことが分かる。

2.4 両手法

2.2 節, 2.3 節の両方を適用して測定した結果を図 1 の"両方"に示す。両手法の適用により、HDD1 のアクセス間隔のさらなる拡大が実現され、これによりさらなる省電力が可能になると考えられる。

また、基本調査、ブロック移動、不均等メモリ割当、両手法における Web アクセスの応答時間を図 2 に示す。HDD2 はブロック移動においてアクセスの発生するブロックを移動され負荷が増加しているが、HDD2 に大きな応答時間の劣化はなく、ブロック移動の性能への影響も小さいと考えられる。不均等メモリ割当により、応答時間の劣化が確認できるが、その程度は小さいと考えられる。

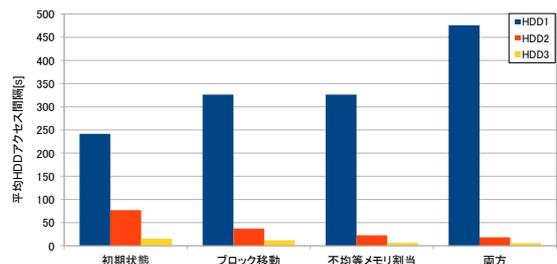


図 1 ブロック移動およびメモリ不均等割当による平均 HDD アクセス間隔

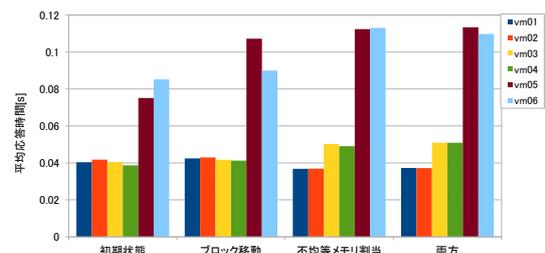


図 2 ブロック移動およびメモリ不均等割当による平均応答時間

3. データベース環境への適用

本章では、前章の手法のデータベース管理システムが稼働する環境への適用について考察する。

前章と同一の物理計算機上に VM を 3 台稼働させ(各 HDD 上に 1VM)、各 VM 上で MySQL を用いて TPC-C ベンチマーク (tpcc-mysql) を実行した。TPC-C における warehouse 数は 160 とし、データベースサイズは約 16GB である。物理計算機や仮想計算機の仕様は前章と同一である。

TPC-C 実行時の各 HDD への平均アクセス間隔を図 2 に、最大アクセス間隔を図 3 に示す。図より、TPC-C では各 HDD のアクセス間隔は非常に短く、この状態では HDD の停止時間を得ることが困難であることが分かる。

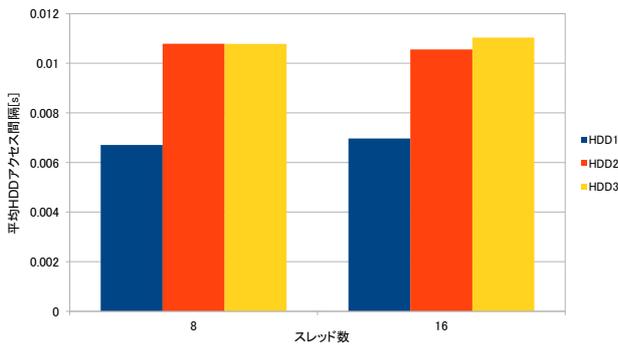


図2 tpcc-mysql 実行時の平均アクセス間隔

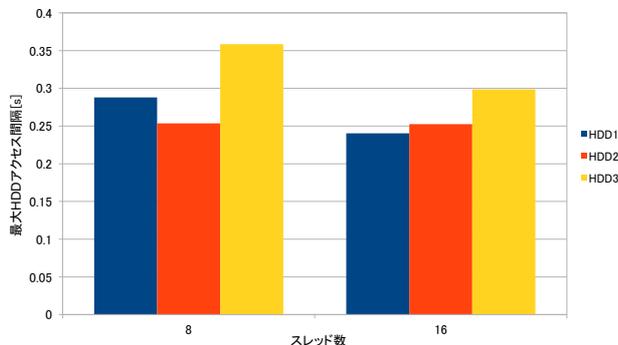


図3 tpcc-mysql 実行時の最大アクセス間隔

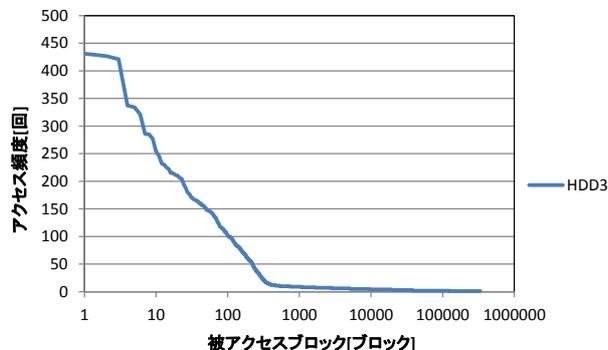


図4 8スレッドにおけるブロックのアクセス頻度分布

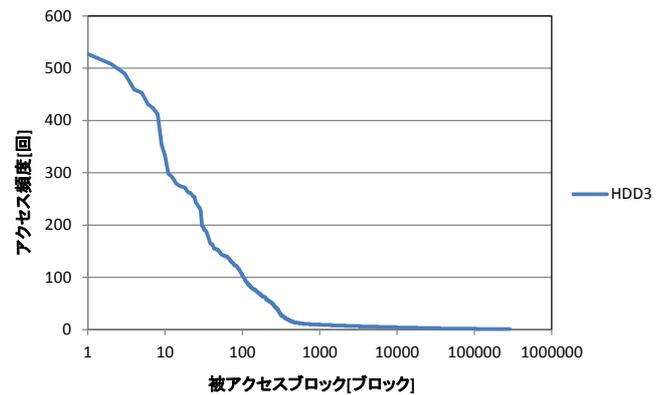


図5 16スレッドにおけるブロックのアクセス頻度分布

次に、TPC-C 実行時の各ファイルシステムブロック (4KB) のアクセス頻度を図 4、図 5 に示す。図 4 がスレッド数 8 における頻度分布、図 5 がスレッド数 16 における頻度分布である。HDD は 3 個搭載されているが、最もアクセスの少ない HDD の頻度を求めた。図の横軸は、アクセスがあったブロックをアクセス頻度順に並び替えたときのブロックの通し番号である。縦軸は、アクセス頻度である。図より、ブロックごとのアクセス頻度に極端な偏りがあり、アクセス頻度の高いブロックを重点的に移動することにより、特定の VM や特定の HDD へのアクセス頻度を極端に削減することが可能であると考えられる。たとえば図 4 においては、アクセスの多い上位 0.08% のブロックを移動すれば、全アクセスの 50% を移動することが可能となっている。

また、TPC-C では、テーブル毎にアクセス頻度が明確に異なるのが分かっており [1]、これを考慮することにより効果的にブロックの移動が可能になると考えられる。

4. まとめ

本稿では、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 の消費電力を削減する手法の仮想化環境への適用に着目し、Web サーバにおける適用結果を示し、その効果を示した。そして、データベース管理システムが動作する仮想化環境への適用について考察を行った。今後は、データ管理システムが動作する環境に当該手法を適用し、その効果の検証を行っていく予定である。

謝辞

本研究は JSPS 科研費 22700039, 24300034 の助成を受けたものである

参考文献

- [1] 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),