Vol.2015-DPS-163 No.18
Vol.2015-MBL-75 No.18
2015/5/28

応用の動作情報を用いた仮想化環境におけるストレージ電力の低減

谷貝　俊輔†　　山口　実靖†‡

データセンターなどで多数の計算機が稼働しており、その消費電力が大きな問題のひとつとなっている．ストレージは，電力消費の大きなの要因のひとつであり，その消費電力の低減は大きな課題のひとつといえる．本稿では，クラウドコンピューティングなどで使われる仮想化環境に着目し，アプリケーションの動作情報を用いた，ストレージ省電力手法の
仮想計算機での適応について考察する．具体的には仮想化環境におけるデータ保存位置の改善や，RAM Disk を使用したストレージアクセス回数の低減を行い，少ない性能劣化での大きなストレージ消費電力の低減手法を提案し，性能評価によりその有効性を示す．

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り今後 10 年でデジタル情報量は約 44 倍になるといわれている[1]. これに伴い，サーバの消費電力の増加が問題となり，データセンターのエネルギー消費量は 2050 年には 2010 年度の日本の発電電力量の約 3 倍になると予測されている [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レイアウトを変更することで HDD の消費電力削減する方法がある[2,3].

本研究では上記手法を仮想化環境に適用し，その有効性の調査をおこなう．具体的には代表的な仮想計算機システムである Xen を用いて，仮想計算機上に mysqlDB を立ち上げ TPC-C 実行時の各テーブルのアクセス量を調査し，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の程度とトランザクション性能の評価を行う．そして上記調査結果に基づくデータ配置手法の提案，RAM ディスクを用いた省電力手法の提案，性能評価によるこれらの手法の有効性の検証を行なう．

2. 応用情報を用いたストレージ省電力

前西川らは,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 データに対して 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っている．

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

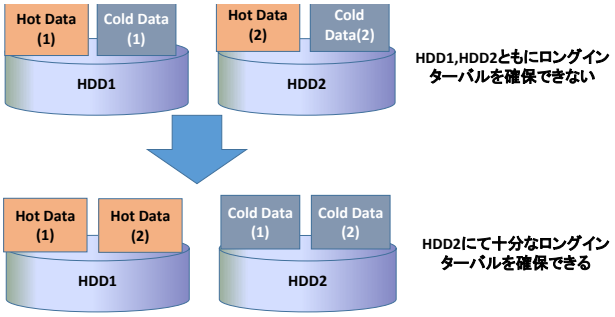

図 1　応用情報を用いたストレージ省電力

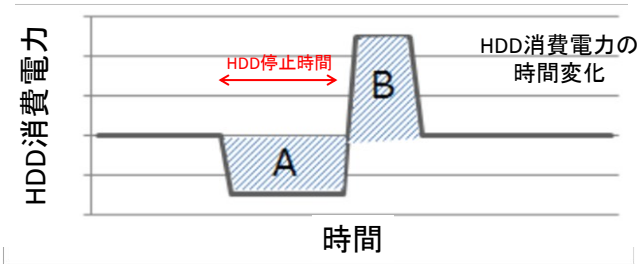

図 2　停止と再起動時の電力の変化

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 Break Even Time 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 Break Even Time 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

3.1 ファイル再配置によるストレージ停止時間の確保

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．図 3 に提案手法のが

†1 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
†2 工学院大学工学部情報通信工学科
Department of information and Communications Engineering, Kogakuin University.

Vol.2015-DPS-163 No.18
Vol.2015-MBL-75 No.18
2015/5/28

いようを示す．通常，各 VM のファイルを特定の HDD に集中して配置すると，考えられるが，本手法では VM と HDD を関連付けず，各 HDD に複数の VM のファイルを格納する．具体的には次のように格納する．

対象の計算機に搭載されている HDD の数を *n* とすると，これらを停止時間を確保し省電力化を試みる 1 台と，停止を試みず稼働させ続ける *n-1* 台に分ける．

次に，アプリケーション実行時の DB テーブルファイルごとのアクセス頻度をもとに，各 HDD に DB テーブルファイルを割り当てる．DB テーブルファイルごとのアクセス頻度は，予備実験を行い事前に調査しておく．本稿ではアプリケーションとしては TPC-C を想定し，予備実験では 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停止用の 1 個の HDD 上に集中して配置, 残りのファイルを残りの HDD に分散配置する．

停止用 HDD に格納する DB テーブルファイルは，以下の様に決定する．DB テーブルファイルをアクセス頻度が少ない順に並べ, アクセス頻度が少ない順に停止用 HDD に格納していき，全格納 DB テーブルファイルの合計アクセス頻度が 1/(*S**BreakEvenTime)を超えない範囲で可能な限り多くのファイルを停止用 HDD に格納する．S はチューニングパラメータであり，“アクセス間隔拡大率”と呼ぶ．アクセス間隔拡大率 S が大きいほど停止用 HDD のアクセス間隔を積極的に拡大することを意味する.「合計アクセス頻度が 1/(*S**BreakEvenTime)を超えない」は, 換言すると「平均アクセス間隔が *S**BreakEvenTime を下回らない」という意味であり，停止用 HDD に BreakEvenTime を超えるアクセス間隔を確保するためには *S* を 1 より大きく設定する必要がある．残りの HDD(停止用 HDD 以外の *n*-1 個の HDD)へのファイルの配置は以下の様に定める．残りのファイル(停止用 HDD に格納したファイル以外のファイル)を，アクセス頻度が多い順に並べ，アクセス頻度が多い順に残りの HDD に格納していく．その際，各 HDD のアクセス頻度の合計が均等にできるだけ近い様に格納する．具体的には，各ファイルの格納先を決定する時点で，各 HDD の格納済みファイルの合計アクセス頻度を求め，最も合計アクセス頻度の少ない HDD に対象ファイルを格納する. この様に，非停止用 HDD 群へのアクセス頻度が均等に近いように分散配置することにより、性能の劣化を抑えることができると期待できる．

3.2 RAM ディスクによるストレージアクセスの回避

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）が可能であるため，遅延書き込みの制限時間やバッファリング量の拡大によりアクセス間隔の拡大を実現できる.また，西川らの手法[1][2]においては,RDBMS の先読み機能に割り当るメモリ量の拡大により読み込みファイルをメモリ内に保持させる量を拡大している．本実験で使用する RDBMS の実装には同等の機能が無いため，以下の手法により DB テーブルファイルのメモリへの積極的格納を実現し, HDD へのアクセスの量の削減を行う．

物理メモリの一部を RAM Disk として活用し，アクセス数が少なくかつ読み込みアクセスのみのファイルは RAM Disk 上に配置する．これにより,停止用 HDD のアクセス間隔をより大きくできると期待できる．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．

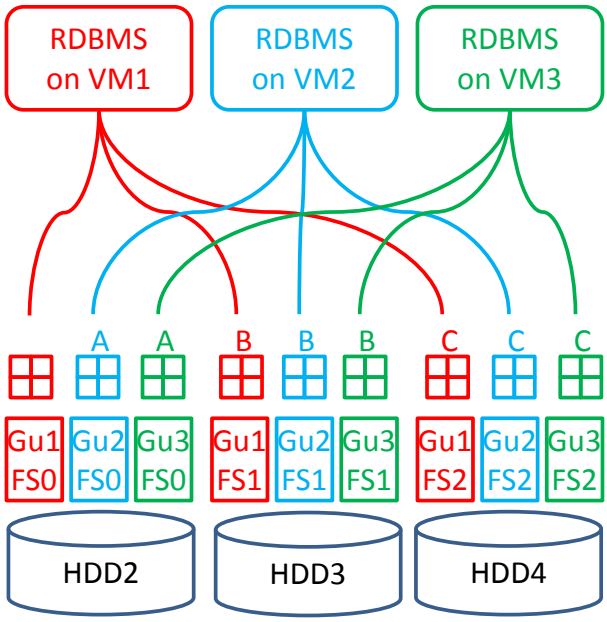

図 3　配置方法

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,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．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. DB テーブルファイル格納用 HDD が 3 台であり，非停止用 HDD が 2 台(HDD2,HDD3)，停止用 HDD が 1 台(HDD4)である．また本実験の計測時間は 1200 秒とし，実験は VM を 2 台使用した実験と,3 台使用した実験を行った．アクセス間隔拡大率 *S* は 2 とし，平均アクセス間隔が 2*BreakEvenTime 以上となる様にした．

配置手法としては，以下の 3 個の配置を用意し性能と消費電力を比較した．

Vol.2015-DPS-163 No.18
Vol.2015-MBL-75 No.18
2015/5/28

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 4 がアクセス頻度の観察結果である．図 4 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/(2*BreakEvenTime)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 BreakEvenTime は 15 秒である． 残りのファイルは HDD2 と HDD3 に配置するが,図 4 により得られるアクセス頻度の合計が同等近く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,各 HDD の合計アクセス頻度が低い方の 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 A”と呼ぶ．この手法は，3.1 節に記述した提案手法が適用され,3.2 節に記述した提案手法が適用されていない状態である.

三つ目の配置手法は２つ目の配置手法(提案配置 A)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アクセスのみのファイルを HDD4 ではなく RAM Disk に配置する．この配置手法を“提案配置 B”と呼ぶ．この手法は，3.1 節に記述した提案手法と，3.2 節に記述した提案手法の両方が適用されている状態である.

また本実験では，消費電力を以下の方法により計算により求め，各手法における HDD へのアクセスを Linux OS の SCSI 層で監視し, このアクセス情報と HDD 停止設定(HDD を停止させるまでのアイドル時間)より HDD の状態(稼働中または停止中)を推定する. そして事前に, 稼働時の HDD の消費電力と，停止時の HDD の消費電力を求めておき，これから累計消費電力を推定する.

4.2 実験環境

表 1，表 2，表 3 に実験に使用した機器の仕様を示す. 本実験ではアクセス間隔の拡大のために, dirty_expire_centisecs を変更した．この値はキャッシュ上に存在しているページがキャッシュ内に存在できる時間を指しておりこの値をすぎても書き込みされていないデータがある場合，自動的に HDD に書き込まれる．この値を初期設定の 30 秒から 300 秒に拡大することで書込みを一括化している．また本実験で使用した HDD の Break Even Time は 15 秒である.

表 1　物理計算機仕様

HostOS	CentOS release 6.4(final)
Host Kernel	linux 2.6.32.57
仮想化システム	Xen version 4.1.2
HDD	Seagate Barracuda ST2000DM001-1CH164
	回転数:7200rpm
	キャッシュ:64MB
	台数4台
Memory	4096MB ×２
Filesystem	ext2
dirty_exoire_centicecs	300 秒

表 2　仮想計算機仕様

Guest OS	CentOS release 6.3 (Final)
Guest kernel	Linux vm02 2.6.32.57
Virtual CPU Core	1
Virtual Memory	2048[MB]
Vitual HDD	50GB+100GB×3
File System	ext2

表 3 Mysql DB　仕様

Mysql	innodb_buffer_pool_size	512[MB]
	innodb_log_file_size	128[MB]

実験で使用した mysql の設定の詳細は表 3 の通りである.

4.3 応用情報について

実験で用いる tpcc-mysql は 9 種類のテーブルファイルを使用する．また今回，実験ではサイズの異なる 3 種類の DB(SF:4, SF:16, SF:64) を使用する．表 4 に各 DB の各テーブルの要領を示す． VM 上で各 DB を 1 種類ずつ動作させ，ファイルアクセスをカーネル内で監視し，各テーブルファイルの書込み，読み込みのアクセス頻度を調査した．計測時間は 1200[sec]とする.

図 4 に調査した各 DB の各テーブルのアクセス頻度についての図を示す．テーブルの配置方法についてはこの図 4 のデータを用いて決定する.

図 4 の結果から Stock，Order_line，Customer などのファイルはアクセス間隔が小さくこれらを格納した省電力化はできないファイルだと考えられる．これに対して，Warehouse，District，item などのファイルはアクセス間隔が比較的長く省電力化が期待できるファイルだと考えられる.

またこの結果から item ファイルは書込みをせず, 読み込みだけを行なうファイルであることがわかり, RAM Disk に適していることがわかる.

表 4　DB 仕様

	データサイズ[KB]		
テーブル名	SF:4	SF:16	SF:64
stock	149,696	591,408	2,346,640
order_line	127,216	510,800	2,027,520
customer	82,096	327,168	1,309,696
history	13,872	50,784	200,496
orders	10,272	35,936	142,608
new_orders	9,744	9,744	15,888
item	1,552	4,624	9,744
district	16	48	96
warehouse	16	16	16
合計	394,480	1,530,528	6,052,704

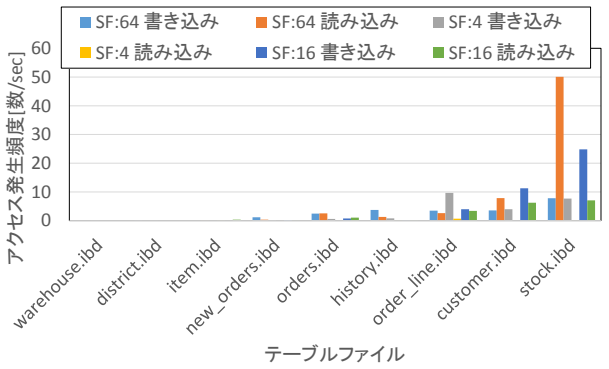

図 4　各テーブルの平均アクセス間隔

4.4 評価方法(仮想マシン 2 台使用時)

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,VM2 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前述した”サイズ均等”，”提案配置 A”，”提案配置 B”の 3 つを用意し，DB 性能，HDD4 のアクセス間隔，HDD4 の消費電力を測定し比較した.

表 5 に紹介した 3 つの配置方法による各 HDD への各テーブルファイルの配置の詳細を示す.

表 5　配置方法詳細

配置方法		HDD2	HDD3	HDD4	RAMディスク
サイズ均等	VM1	stock, district, warehouse	order_line, item, history	customer, new_orders, orders	
	VM2	stock, district, warehouse	order_line, item, history	customer, new_orders, orders	
提案配置A	VM1		stock, order_line, new_orders, orders, history	item, warehouse, district	
	VM2	stock	order_line, new_orders, orders, history	item, warehouse, district	
提案配置B	VM1		stock, order_line, new_orders, orders, history	warehouse, district	item
	VM2	stock	order_line, new_orders, orders, history	warehouse, district	item

4.5 測定結果(仮想マシン 2 台使用時)

前節で説明した配置により TPC-C を実行し，トランザクション性能と最大アクセス間隔とアクセス間隔ごとの発生頻度，そしてアクセス頻度の低いデータを集めた HDD4 の消費電力を測定した．VM を 2 台使用した測定結果を図 5, 図 6, 図 7，図 8 に示す．図 7 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 5 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，提案手法 A，B は HDD4 を積極的に使用しておらず，これが性能を低下させる原因となるが，アクセス頻度の均等化を行なっており，これが性能向上要因となったと考えられる．図 6，7 より，提案配置 A，B 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 8 より HDD4 が常時稼動しているサイズ均等と比べ提案配置では 7[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 2[w]以上の低減ができ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.

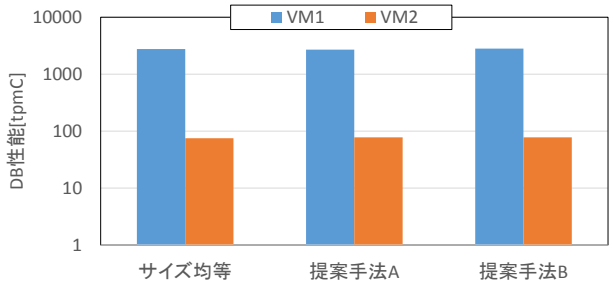

図 5　各配置方法のトランザクション性能

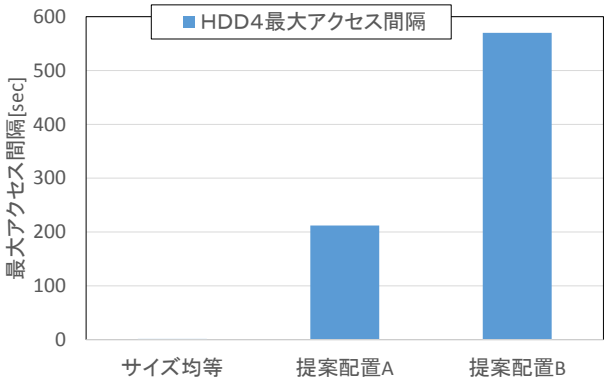

図 6　HDD4 の最大アクセス間隔

Vol.2015-DPS-163 No.18
Vol.2015-MBL-75 No.18
2015/5/28

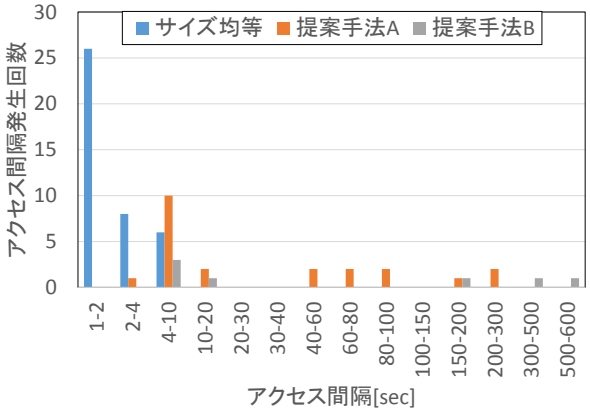

図 7　HDD4 のアクセス間隔頻度分布

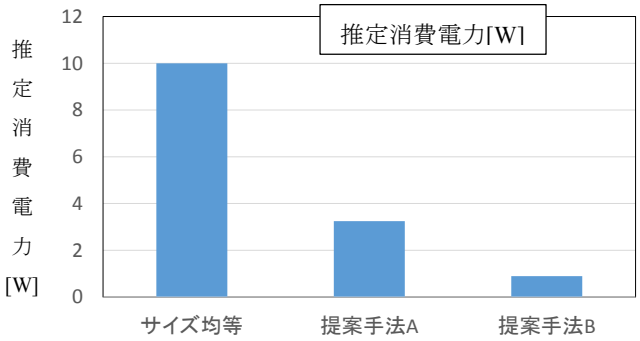

図 8　HDD4 の消費電力

4.6 評価方法(仮想マシン 3 台使用時)

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,全 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，VM2 には SF が 16 でサイズが 5.8[GB]を，VM3 では 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前述した”サイズ均等”,”提案配置 A”,”提案配置 B”の 3 通りを用意し DB 性能，HDD4 のアクセス間隔, HDD4 の消費電力を測定し比較を行った.

表 6 に 3 通りの配置方法の各 HDD への各テーブルファイルの配置の詳細を示す.

表 6　配置方法詳細

配置方法		HDD2	HDD3	HDD4	RAMディスク
サイズ均等	VM1	stock, district, warehouse	order_line, item, history	customer, new_orders, orders	
	VM2	stock, district, warehouse	order_line, item, history	customer, new_orders, orders	
	VM3	stock, district, warehouse	order_line, item, history	customer, new_orders, orders	
提案配置A	VM1	customer, orders, new_orders	stock, order_line, history	item, warehouse, district	
	VM2	stock	stock, customer, item	warehouse, district	
	VM3	stock, customer, orders	order_line, new_orders, item, history	item, warehouse, district	
提案配置B	VM1	customer, orders, new_orders	stock, order_line, history	warehouse, district	item
	VM2	stock	stock, customer, item	warehouse, district	
	VM3	stock, customer, orders	order_line, new_orders, item, history	warehouse, district	item

4.7 測定結果(仮想マシン 3 台使用時)

前節で説明した配置により TPC-C を実行し，トランザクション性能と最大アクセス間隔とアクセス間隔ごとの発生頻度，そしてアクセス頻度の低いデータを集めた HDD4 の消費電力を測定した. 測定結果を図 10，図 11，図 12，図 13 に示す. 図 12 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 9 より,提案手法の適用による性能の劣化がほぼなく，VM2 においては改善がみられることがわかる. 前述のとおり，性能向上が実現した理由は提案手法により HDD へのアクセス頻度の均等化が行われたことであると予想される.

図 10，12 より,提案配置 A，B 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 13 より HDD4 が常時稼動しているサイズ均等と比べ提案配置では 7.5[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 1.5[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 を用いることにより，アクセス間隔のさらなる拡大が実現できることが確認できた.

Vol.2015-DPS-163 No.18
Vol.2015-MBL-75 No.18
2015/5/28

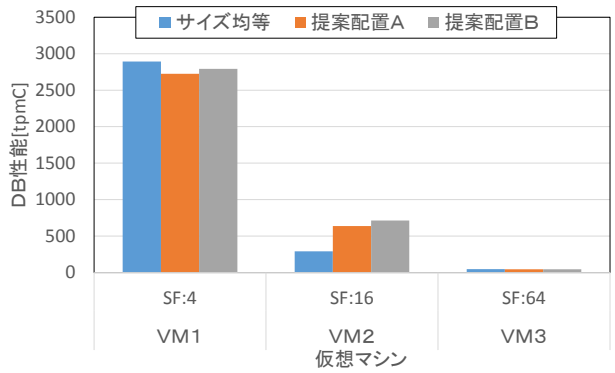

図 9　各配置方法のトランザクション性能

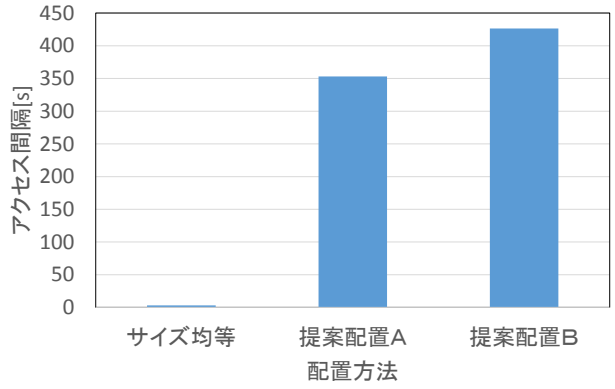

図 10　HDD4 の最大アクセス間隔

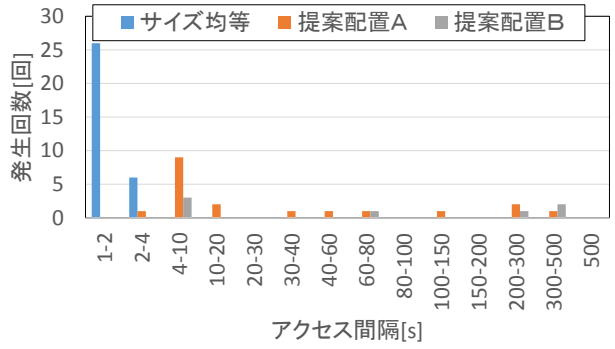

図 11　HDD4 のアクセス間隔頻分布

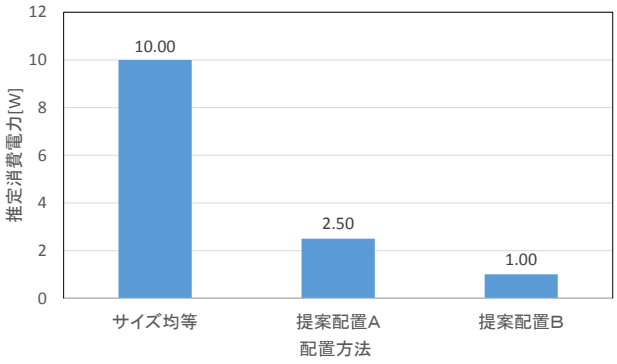

図 12　HDD4 の消費電力

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．またさらに RAM Disk を用いる手法を提案し評価によりその有効性を示した．

今後は，マイグレーションを利用したより細かなデータレイアウトの最適化，先読みや遅延書き込みの拡大などによる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．

謝辞

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

参考文献

[1] 飯村　菜穂, 西川　記史, 中野　美由紀, 小口正人　“データベース処理実行時における省電力化のためのストレージ制御手法の提案”(DICOMO2013 7/12　7C-1)

[2] Norifumi Nisikawa , Miyuki Nakano and Masaru Kitsuregawa,” Energy Efficient Storage Management Cooperated wuth Large Data Intensive Applications,” 28th IEEE International Conference on Data Engineering (IEEE ICDE 2012),

[3]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)

[4]若色　匠，山口　実靖　“仮想化環境における応答性能を考慮したストレージ稼働時間の低減” 情報処理学会 2013 年全国大会　1L-9

[5] Transaction Processing Performance Council (TPC) TPC BENCHMARKC Standard Specification　Revision 5.11　February 2010