

データベースアプリケーションの実行時省電力に関する評価

飯村 奈穂 (指導教員：小口 正人)

1 はじめに

近年のデジタル情報量の爆発的な増加により、ストレージの出荷台数も急増していることから、ストレージの管理運用コストは見逃げせないものとなっている。本研究ではデータの効率的な管理という点からクラウド上のデータベースの省電力化を考え、データベースベンチマークの業界標準である TPC-H の省電力化に向けて、実行時のシステム性能と消費電力量の解析を行い、ディスクの省電力状態を適用することにより、実行時省電力化が可能であることを示した。

2 基礎性能測定

2.1 測定環境

本研究では、サーバ PC として、CPU が AMD Athlon 64 FX-74 3GHz (4 cores) を 2 つ、主記憶が 8GB、HDD が Seagate Barracuda 1TB を 6 台、OS が CentOS 5.6 64 ビット版、DBMS は Hitachi HiRDB Single Server Version 9 を使用する。また、電力計は YOKOGAWA WT1600 Digital Power Meter を使用する。

2.2 ディスクアクセス時の性能測定

ディスクに対してシーケンシャルおよびランダムアクセスを行った際のディスクの消費電力とスループットを 1 秒毎に測定する。アクセスの方法は Read 処理、Write 処理の 2 種類。測定対象のディスクは 1 台とし、アクセス単位 (バッファサイズ) を 4, 8, 16KB と変化させる。データの読み書き量はシーケンシャルの場合は 4, 8, 16GB (アクセス単位*1M 回)、ランダムアクセスの場合は 4, 8, 16MB (アクセス単位*1K 回) とする。I/O の方法としては Direct I/O を用いる。

シーケンシャルアクセスの場合、バッファサイズが大きいほど一度に I/O を行うデータ量が増加するため、スループットはバッファサイズ 4KB と 16KB を比較すると 33.65MB/s から 77.77MB/s まで向上した。ランダムアクセスの場合もスループットについては同様である。また、ランダムアクセスの方が読み書き量が少ないにも関わらず、同じバッファサイズのシーケンシャルアクセスより消費電力がそれぞれ約 0.5W 程度大きかった。これは I/O を行う際のディスクアドレスが毎回異なるためである。これらのことから測定結果は妥当であると考えられる。

2.3 TPC-H 実行時のディスクの性能測定

データベースベンチマークである TPC-H を動作させた際の消費電力と I/O トレースの解析を行う。測定対象のディスクは 2 台、HDD1 には LINEITEM テーブルを、HDD2 にはその他テーブルを配置してある。DB の規模を決めるスケールファクタ (SF) を 1, 2, 3 と変化させ、それぞれの SF ごとに DB とクエリを用意し、DB はブロックデバイス上に直接配置して測定を行う。

図 1, 2 には SF=3 の時の HDD1, 2 の測定結果を示

す。SF=1, 2 の時も同様の傾向であった。上段は I/O トレースの取得結果を、下段は実行時の消費電力をクエリごとに色分けしたものを表している。I/O トレースと消費電力を比較すると、消費電力の上下とディスクアクセスの頻度が一致しており、ランダムアクセスを行っているクエリは応答時間が長く、クエリプランも複雑である。従って、この結果は妥当な振舞であると考えられる。

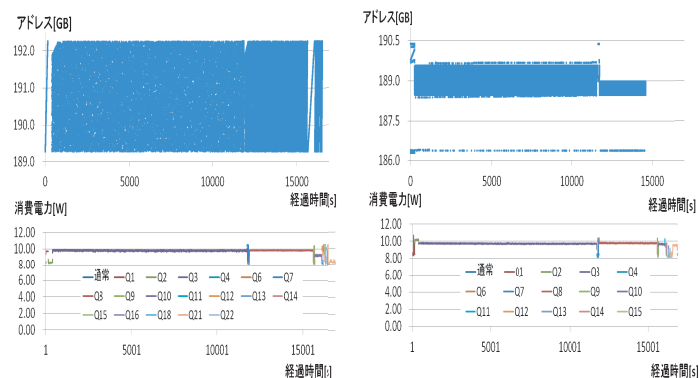


図 1: HDD1(SF=3)

図 2: HDD2(SF=3)

3 ディスクの消費電力特性

3.1 遷移状態と消費電力

本研究で使用したディスクの遷移状態は Standby1, Standby2, Idle, Active の 4 種類である。Idle/Active 状態から Standby1 状態に移行することを Spindown, Standby1, Standby2 状態から Idle/Active 状態に移行することをそれぞれ Spinup1, Spinup2 と呼ぶ [1]。

各状態におけるディスクの消費電力の測定を行った。測定対象のディスクは、先程の測定 2.2 と同様のディスク 2 台である。Standby1 時、Standby2 時、Idle 時、Active 時の最大消費電力と、Spindown, Spinup1, Spinup2 に必要なエネルギーを表 1 に示す。

表 1: ディスクの遷移状態における消費電力とエネルギー量

Disk	Standby1(W)	Standby2(W)	Idle(W)	Active(W)
HDD1	1.81	1.21	8.42	10.5
HDD2	1.92	1.24	8.43	10.8
Disk	Spindown(J)	Spinup1(J)	Spinup2(J)	
HDD1	16.31	159.03	184.41	
HDD2	13.77	181.39	180.05	

3.2 Break-Even Time

ディスクの Spindown 及び Spinup により消費されるエネルギーと、ディスクを Standby 状態に移行し、その状態を維持することにより削減できるエネルギーが等しくなる Standby 状態の持続時間を Break-Even Time と呼ぶ。これは Spindown に必要なエネルギーを E_d , Spinup2 に必要なエネルギーを E_{u2} , Standby1 状態の消費電力を P_{s1} , Standby2 状態の消費電力を P_{s2} ,

Idle 状態の消費電力を P_i , Spindown と Spinup2 に必要な時間をそれぞれ T_d, T_{u2} , Standby1, Standby2 状態の持続時間をそれぞれ T_{s1}, T_{s2} とすると, Break-Even Time T_{be} は,

$$T_{be} = (E_d + E_{u2} - P_{s2} * T_d - P_{s2} * T_{u2} + T_{s1} * (P_{s1} - P_{s2})) / (P_i - P_{s2})$$

により求めることができる [2].

HDD1, HDD2 では Break-Even Time はそれぞれ約 26 秒であった. これより Standby 状態を利用して省電力化を実行するためには, ディスクへの I/O 発行間隔が HDD1, HDD2 それぞれのディスクにおいて約 26 秒以上必要である. 図 3 は HDD1 において Idle 状態から Standby1 状態, Standby2 状態に移行した後, 再び Idle 状態に移行した時の消費電力の推移を示している.

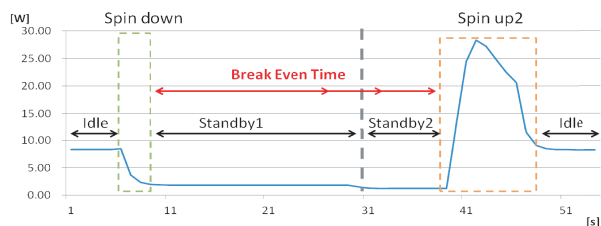


図 3: ディスクの状態遷移における消費電力と Break-Even Time(HDD1)

4 実行時省電力可能性

本節では, 前節をふまえて SF=3 時の TPC-H 実行時の I/O 発行間隔を調査し, Break-Even Time を考慮した際に, 実行時の消費電力量をどの程度削減することができるかを見積と実測により評価する.

4.1 I/O 発行間隔

ディスクに I/O が行われてから, 次の I/O が発行されるまでの時間を I/O 発行間隔と呼ぶ. 本研究では, TPC-H 実行時のディスク I/O の利用状況を取得・解析し, I/O 発行間隔を調査する. 調査対象のディスクは前節と同様のディスク 2 台とする. I/O 発行間隔が Break-Even Time 以上である回数が HDD1 では 4 回, HDD2 では 10 回であった. また, I/O 発行間隔が Break-Even Time 以下である回数は HDD1 では 8 回, HDD2 では 11 回であった.

4.2 削減可能エネルギー

ディスクのスタンバイ状態を利用する場合を省電力状態適用あり, 利用しない場合を省電力状態適用なしとする. さらに, 省電力状態適用ありの場合は, ディスクをスタンバイ状態に移行するまでのタイムアウトを設定すると仮定して, 実行時消費電力量の見積を行う【I/O 発行間隔 - タイムアウト時間】を, 省電力状態適用期間とする. この場合は Break-Even Time より短い I/O 発行間隔にも省電力状態を適用することになるため, I/O 発行間隔の長さによって見積式を選ぶ必要がある. 3.2 節で使用した項目に加えて, I/O 発行間隔を T_i , 設定するタイムアウトを T_t としたときの削減エネルギー E_s は,

省電力状態適用期間 < $T_d + T_{s1}$ の場合

$$E_s = (T_i - T_t) * P_i - E_d - E_{u1} + P_{s1} * (T_i - T_t - T_d)$$

I/O 発行間隔 - タイムアウト $\geq T_d + T_{s1}$ の場合

$$E_s = (T_i - T_t) * P_i - E_d - E_{u2} + P_{s1} * T_{s1} + P_{s2} * T_{s2}$$

により求めることができる. 見積式より, 最も消費電力量の削減率が大きいタイムアウトは HDD1 では 15 秒, HDD2 では 10 秒であった.

次に, 見積式から得られた最適なタイムアウトを各ディスクに設定し, TPC-H 実行時の消費電力量を測定する. 図 4 は省電力状態適用なし, 省電力状態適用ありの実測値, 省電力状態適用ありの見積値の TPC-H 実行時の消費電力量の比較を示している.

消費電力量削減率の見積値は, HDD1 では 2.1 %, HDD2 では 1.6 % であるのに対し, 実測値は HDD1 では 1.6 %, HDD2 では 1.3 % であり, 誤差は 0.3 ~ 0.5 % に収まった. TPC-H 実行時の I/O 発行間隔の長さや回数, ディスクの消費電力は毎回若干異なるため, 見積値と実測値の誤差は許容範囲であると考えられる. これらのことから, ディスクの省電力状態と I/O 発行間隔を利用した TPC-H の省電力化は可能であり, 本研究で提案した見積式は妥当であると考えられる.

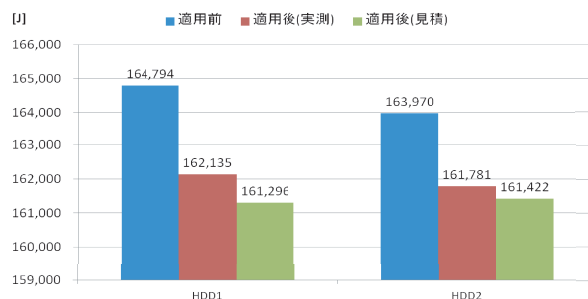


図 4: 省電力状態適用時の TPC-H 実行時消費エネルギー比較

5 まとめと今後の課題

本研究ではデータの効率的管理という点からクラウド上のデータベースの省電力化を考え, データベースベンチマークである TPC-H の実行時省電力化に向けて, 実行時のシステム性能と消費電力量の解析を行った. また, 実行時の I/O 発行間隔とディスクの省電力状態を利用することによって TPC-H の省電力化が可能であることと, 提案した見積式が妥当であることを実測により示した. 今後はデータ配置等による実行時のディスク I/O 発行間隔制御など, 更なる省電力化に取り組んでいきたい.

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), pp.126-137, 2012.04
- [2] Y.H. Lu, G.D. Micheli: Comparing System-Level Power Management Policies, IEEE Design & Test of Computers, 2010
- [3] 飯村奈穂, 西川記史, 中野美由紀, 小口正人: データインテンシブアプリケーション実行時のストレージ省電力に関する一検討, DEIM2013, 2013年3月発表予定