

# 多数台ディスク環境における データ配置制御を利用したストレージ省電力化手法の提案

理学専攻・情報科学コース 飯村 奈穂

## 1 はじめに

近年、情報爆発に伴いデータセンタの大規模化が進み、データセンタの管理運用コストは見越せないものとなっている。特に電力コストの削減は急務であり、一定の消費電力割合を占めるストレージの省電力化に注目が集まっている [1]。本研究ではデータの効率的な管理という点からデータセンタの省電力化を考え、アプリケーションの SLA (Service Level Agreement) を考慮したストレージ省電力化を行うことを研究目的とする。本論文では、業界標準のデータベースベンチマークである TPC-H を用い、データベース処理実行時におけるストレージ省電力化に向け、多数台ディスク環境でデータ配置制御手法の提案と評価を行った。

## 2 関連研究

これまでの研究で、我々はデータ配置制御によるストレージ省電力化が有効であるということを示した [2]。しかし、[2] の環境では評価に使用している HDD の台数や容量は少なく、大規模なデータを想定しているとは言い難い。そのため、データの配置方法も単純なものに限られる。そこで本稿ではより大規模な環境を想定し、提案手法の評価を行う。

ストレージにおける省電力を検討する場合、ハードディスクを SSD (Solid State Drive) に変える方法も考えられ、SSD を利用した様々な省電力手法が検討されてきた [3]。ハードディスクを SSD に置き換えられるかどうかはコスト次第であると言えるが、将来的にはそのような方向へ進むものと予想される。本研究のアプローチは、ハードディスクが SSD に置き換わってもそのまま適用できるものである。さらに、SSD はアクティブ状態とスタンバイ状態の切換えに、ハードディスクのような電力消費を要しないので、SSD の時代になったら、より有望なアプローチであると言える。

## 3 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) × 2、主記憶が 8GB、HDD が Seagate Barracuda 3TB × 11(うち OS 用として 1 台)、OS が CentOS 5.10 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。

## 4 提案手法の評価

本節では、TPC-H 実行時の I/O 発行間隔を制御するという目的のもと、TPC-H の各表と索引のデータに対する I/O の状況を調査し、データの配置に手を加えることでさらなる実行時省電力化を目指す。また、TPC-H のデータの中でも約半分のデータ量を占める LINEITEM 表および索引を 10 個に分割して、評価を行う。

### 4.1 入出力状況の調査

本研究におけるストレージ省電力化のための提案手法である、データ配置制御の評価を行うために、TPC-H 実行時の各データに対する入出力状況を調査する。DB の規模を示す SF (Scale Factor) は 10 とし、1 秒毎に pdbufs (DBMS 付属のバッファ情報表示ツール) を用いて入出力状況を取得する。調査の結果、LINEITEM を除いて、実行期間中に入出力が見られたデータは 12 個、入出力が見られなかったデータは 9 個であった。

LINEITEM 表および索引は、ハッシュ関数を用いてそれぞれ 10 個に分割する。分割した LINEITEM 表の入出力頻度を調査した結果を図 1 に示す。L<sub>xx</sub> の xx が分割した表の番号を示している。図 1 より、分割した表はそれぞれ分割した番号順に読まれていることがわかる。このことから、分割した表と索引は、できるだけ番号が近いものを同じディスクに配置することで、ディスクに対してより長い I/O 発行間隔を得られるということがわかる。

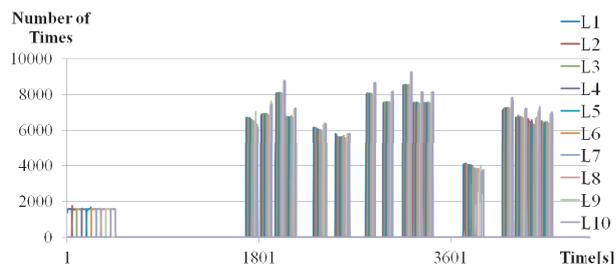


図 1: 分割表の I/O 頻度

### 4.2 分割表の配置方法

分割した LINEITEM 表と索引、およびその他の表と索引をディスク 3 台に配置する。ここで、分割した表と索引に関して、2 種類の配置方法を用意する。(1) 分割した表と索引をラウンドロビンに配置する方法、(2) 分割した番号の近いものをできるだけ同じディスクに配置する方法、の 2 種類である。その他のデータはディスク 3 台のデータ量が均等になるように配置するものとする。

#### 4.2.1 2 種類の配置方法での比較

2 種類の配置方法において、ディスクをスタンバイ状態へ移行するタイムアウトを設定した場合とそうでない場合の TPC-H 実行時における消費電力量の削減率と実行時間について比較を行った。消費電力量の削減率は、配置方法 (1) が約 15% であるのに対し、配置方法 (2) では約 33% であった。また、実行時間の遅延率は配置方法 (1) が約 22% であるのに対し、配置方法 (2) では約 8% であった。以上から、より長い I/O 発行間隔が得られる (2) の方が多くの消費電力を削減することができる。また、(2) の方がディスクをスタンバイ状態へ移行する際に生じる起動オー

バヘッドも少ないため、遅延率も小さいことがわかる。これらのことから、今回使用した TPC-H では、分割した表番号を万遍なくラウンドロビンに配置するよりも、表番号に近いものをできるだけ同じディスクに配置した方が、より効果的なストレージ省電力化が可能であるということがわかる。

#### 4.3 ディスク 10 台を用いた提案手法の評価

4.1, 4.2 節の結果に基づき、ディスク 10 台を利用して提案手法の評価を行う。ディスクの消費電力量と実行時間の関係を調べるために、使用するディスク台数および提案手法に基づき I/O 有のデータのみを配置した Hot 状態のディスク台数を変化させて評価を行う。今回用意したディスク環境は使用するディスク台数が 10 台、5 台、2 台の 3 種類で、さらにそれぞれ「ナイーブな手法」と「提案手法」の 2 種類を用意する。ナイーブな手法とは、提案手法であるデータ配置制御の対比となる単純なデータ配置手法である。つまり、提案手法はデータに対する I/O に基づいてデータを配置するのに対し、ナイーブな手法は、I/O の頻度に関わらず、データがおおむね均等になるように配置する。今回我々が用意したデータ配置は以下である。

##### 1. ディスク 10 台環境

- ナイーブな手法 (Hot10 台): データの利用頻度に関わらず、データ量が均等になるようにディスク 10 台に配置
- 提案手法 (Hot9 台): 分割した LINEITEM 表を HDD1 から HDD3 に配置し、I/O 有の残りのデータをスキーマ毎に 1 台のディスクに配置し (HDD4 ~ HDD9), HDD10 には I/O 無のデータを配置

##### 2. ディスク 5 台環境

- ナイーブな手法 (Hot5 台): データの利用頻度に関わらず、データ量が均等になるようにディスク 5 台に配置、残りのディスク 5 台は使用しない
- 提案手法 (Hot4 台): 分割した LINEITEM 表を HDD1 から HDD3 に配置し、I/O 有の残りのデータを HDD4 に配置し、I/O 無のデータを HDD5 に配置、残りのディスク 5 台は使用しない

##### 3. ディスク 2 台環境

- ナイーブな手法 (Hot2 台): データの利用頻度に関わらず、データ量が均等になるようにディスク 2 台に配置、残りのディスク 8 台は使用しない
- 提案手法 (Hot1 台): I/O 有のデータを HDD1 に配置、I/O 無のデータを HDD2 に配置、残りのディスク 8 台は使用しない

提案手法のうち、Hot9 台、Hot4 台において、分割した LINEITEM 表は 4.2 節の結果に基づいて、できるだけ近い番号の分割データを同じディスクに配置する。ここでは、L1, L2, L3, L4 を HDD1 に配置し、L5, L6, L7 を HDD2 に、L8, L9, L10 を HDD3 に配置する。ディスクの電力状態に関して、ナイーブな手法は Hot なディスク全てを Active/Idle 状態、Hot9 台と Hot4 台

では 10 台全てに Standby 状態へ移行するためのタイムアウト (5 秒) を設定する。これを省電力状態適用と呼ぶ。Hot1 台では HDD1 を Idle/Active 状態、HDD2 から HDD10 の 9 台に省電力状態を適用する。Hot4 台で HDD4 以外に省電力状態を適用したものを 'Hot4 台' と呼ぶ。

各配置方法における TPC-H 実行時の消費電力量と実行時間の相関について考察する。本評価での TPC-H の SF は 30 とする。各配置方法における、ナイーブな手法の Hot10 台と比較した実行時消費電力量の消費率と実行時間の遅延率の相関を図 2 に示す。図中の点線はパレート最適な曲線を示している。提案手法を利用した配置方法 (Hot9 台, Hot4 台', Hot4 台, Hot1 台) では、それぞれ同じディスク台数を用いたナイーブな手法に比べ、よりパレート最適に近づいていることがわかる。以上のことから、本測定における本研究の提案手法を利用した配置方法はパレート最適に近い配置方法であると言える。

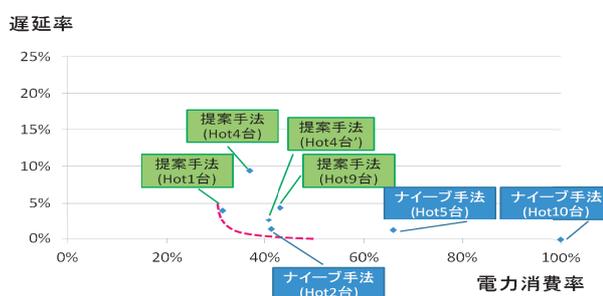


図 2: 消費電力量と実行時間の相関

## 5 まとめと今後の課題

本研究ではアプリケーションの SLA を考慮したストレージ省電力手法の提案を研究目的とし、本稿では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、提案手法であるデータ配置制御の評価を行い、本研究の提案手法がストレージ省電力化に有効であることを示した。さらに、本提案手法を利用することにより、ナイーブな手法と比較して、パレート最適に近いストレージ省電力化が可能であることを示した。今後はデータの規模をさらに大きくして評価を行い、本手法の更なる有効性を示したい。

### 参考文献

- [1] GIPC Survey and Estimation Committee Report FY2009 (Summary): <http://www.greenit-pc.jp/activity/reporting/100707/index.html>
- [2] 飯村奈穂, 西川記史, 中野美由紀, 小口正人: システム性能とストレージ省電力を考慮した TPC-H 実行時におけるデータ配置法の評価, DICOMO シンポジウム 2014, p.1818-1825, 8C-3, 2014 年 7 月
- [3] Jian Ouyang, Shiding Lin, Zhenyu Hou, Peng Wang, Yong Wang, and Guangyu Sun. Active SSD design for energy-efficiency improvement of web-scale data analysis, IEEE International Symposium on Low Power Electronics and Design (ISLPED 2013), pp.286-291, September 2013.