Solid State Drive搭載オンライントランザクション処理サーバにおけるDynamic Voltage and Frequency Scalingを用いた省電力化の実験的検討

早水　悠登† 　合田　和生† 　中野美由紀† 　喜連川　優†

† 東京大学生産技術研究所　〒 153–8505 東京都目黒区駒場 4-6-1
E-mail: †{haya,kgoda,miyuki,kitsure}@tkl.iis.u-tokyo.ac.jp

あらまし　近年データセンタの消費電力が増加しており、なかでも主たるアプリケーションであるオンライントランザクション処理 (OLTP) の省電力化は重要な課題である。OLTP のように停止できないアプリケーションでは、低消費電力なハードウェア構成だけではなく、ソフトウェアによる省電力化が重要である。本研究では、省電力構成で多く用いられる Solid State Drive 搭載サーバを対象とし、OLTP 実行時における負荷の偏りを利用した DVFS による省電力化の効果について検討を行った。

キーワード　省電力化, オンライントランザクション処理, Solid State Drive, Dynamic Voltage and Frequency Scaling

An Experimental Study on Energy Saving for Online Transaction Processing Servers Equipped with Solid State Drives using Dynamic Voltage and Frequency Scaling

Yuto HAYAMIZU†, Kazuo GODA†, Miyuki NAKANO†, and Masaru KITSUREGAWA†

† Institute of Industrial Science, the University of Tokyo
4-6-1 Komaba, Meguro-ku, Tokyo 153–8505 Japan
E-mail: †{haya,kgoda,miyuki,kitsure}@tkl.iis.u-tokyo.ac.jp

1. はじめに

近年、情報技術分野においても低炭素化が求められており、IT 機器の消費電力削減が求められている。その一方で、Web サービスの大規模化、スマートグリッドや ITS など情報技術を活用したサービスの登場などにより人類が扱う情報量は急増しており、それらを支えるデータセンタの消費電力は年々増加している。米国環境保護庁が 2007 年に発表した調査 [1] によると、米国内におけるデータセンタの消費電力は 2000 年から 2006 年にかけて約 2 倍となり、2011 年にはさらにその 2 倍になることが予測されている。データセンタの消費電力増加は電力コストの増加を招くだけでなく、電力供給の制約により計算・記憶資源の新規導入や増強が阻害されるという問題を引き起す。そのため、データセンタの省電力化の必要性が高まっている。

EMERSON の報告 [2] によると、データセンタで消費される電力のうち 44%がサーバによって消費されており、データセンタの省電力化のためにはサーバの省電力化が効果的である。サーバの消費電力を削減することにより、配電機器や無停電電源装置（UPS）などの給電設備の消費電力もまた削減することができる。これによってデータセンタ内の総発熱量が低下するため、冷却設備の消費電力も削減することができる。つまり、サーバの省電力化はデータセンタ全体の省電力化を行う足がかりとなる。

データセンタのサーバの中でも、データ処理の中心的役割を果たしているのはデータベースサーバであり、データベースサーバの省電力化は重要な課題である。データベースサーバにおける主要なアプリケーションの一つに、オンライントランザクション処理 (OLTP) がある。OLTP は電子金融・商取引などの基盤となる技術であり、その処理性能が社会活動・経済活動のスピードに直結するため、高スループット・低遅延であることが求められる。また OLTP アプリケーションの停止は機会損失へと直結するために、常に稼動し続け利用可能な状態にあることが求められる。このような性能・可用性への厳しい要求のために、OLTP の省電力化は難しくこれまでに有効な省電力化手法はほとんど提案されていない。データベースサーバの省電力化において、OLTP の省電力化は避けて通ることができない課題である。

近年では、高いスループットを達成するためにデータベースサーバに数十 GB のメモリを搭載し、インメモリデータベースシステムを用いる構成が随所にみられる。我々はこのような構成の OLTP 向けデータベースサーバを対象とした省電力化手法の提案を行ってきた [3]。しかし、データがメモリに収まりきらない規模のデータベースは依然として多く存在する。そのような場合には、従来用いられてきたハードディスクに代わって Solid State Drive(SSD) が用いられる場面が増えている。SSD はハードディスクに比べてアクセス遅延が短くデータ転送速度が高いため、ハードディスクを SSD で置き換えることでより高いスループットを達成することができる。このような構成の OLTP サーバにおける省電力化の検討はこれまでに行われていない。SSD が普及し低価格化するにつれこのような構成がさらに多くの場面で用いられるようになることが予想され、その省電力化の重要性は高い。

本研究では、SSD を搭載した OLTP サーバにおける省電力化を目的として、OLTP アプリケーション性能と消費電力を実測できる環境を構築し、性能と消費電力の関係を調べる実験を行った。実験では、OLTP アプリケーションとして業界標準ベンチマーク TPC-C [4] を利用し、現在広く普及している省電力化技術である Dynamic Voltage and Frequency Scaling(DVFS) を用いてプロセッサの動作周波数を変動させ、TPC-C 性能・消費電力への影響を測定した。

2. オンライントランザクション処理と Solid State Drive

オンライントランザクション処理 (OLTP) とは、同時に多数のトランザクションを処理するデータベース処理のことを指す。OLTP は電子商取引や電子金融取引など現代社会に欠くことのできないサービスの基盤技術として用いられている。たとえば電子商取引であれば、商品を発注する、代金を支払うなどという行為の 1 つ 1 つがトランザクションに対応する。これらトランザクションの処理が滞ることはビジネスの機会損失に直結するため、OLTP アプリケーションには高スループットを達成することが求められる。

OLTP アプリケーションに求められる高スループットを実現するために、近年ではハードディスクを用いないインメモリデータベースシステムの採用が進んでいる。その代表例としては、東京証券取引所が 2010 年から稼動を始めた株式売買システム arrowhead [5] が記憶に新しい。しかし一方で、メモリに収まらない大規模なデータベースを扱う必要がある場面も依然として存在する。そのような場合に、ハードディスクの代替として使用されるようになっているのが Solid State Drive(SSD) である。SSD はハードディスクに比べてデータ転送速度が高くアクセス遅延が少ないために、ハードディスクと置き換えることで OLTP アプリケーションのスループットを高くすることができる。SSD を二次記憶として採用するデータベースシステムの需要が高まるにつれ、このような構成のデータベースサーバの省電力化の重要性も高まっている。

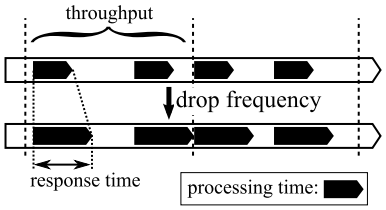

図 1　動作周波数を下げたときのスループット・応答時間の変化

3. Dynamic Voltage and Frequency Scaling を用いたオンライントランザクション処理の省電力化

OLTP アプリケーションには、常に稼動し続けることが求められるため、その省電力化においては実行時省電力化技術が必要である。アプリケーションの実行を止めることなく省電力化を行う手法として現在広く用いられているものにプロセッサの Dynamic Voltage and Frequency Scaling(DVFS) がある。DVFS とは、プロセッサの動作周波数・電圧を実行時の負荷状況などをもとにして動的に制御し、省電力化を行う技術の総称である。たとえばボリュームサーバで広く用いられている Intel Xeon プロセッサでは Intel SpeedStep Technology という形で DVFS 機能が提供されている。プロセッサの消費電力は動作周波数に比例し、動作電圧の 2 乗に比例するため、これらの値を小さくすることで消費電力を削減することができる。ただし、DVFS により動作周波数を下げるとプロセッサの命令スループットが低下するため、アプリケーション性能も低下する可能性がある。

DVFS が効果的に機能するのは、プロセッサの使用率がある程度低い場合である。例えば OLTP の場合には、動作周波数を低くすることで命令スループットが下がると応答時間 (トランザクション 1 つあたりの処理にかかる時間) は増加するが、スループット (単位時間あたりのトランザクション処理数) のへの影響は小さい (図 1)。OLTP アプリケーションの利用率は特定の短い時間帯にピークを迎え、それ以外の時間帯は低いというパターンが一般的であるため、プロセッサの使用率は 1 日のうち多くの時間は低い値であると考えられる。2007 年に Google が行った報告 [6] によると、Google のデータセンタにおける平均的なサーバのプロセッサ使用率は、ほとんどの時間において 10%〜50%ということであった。また、今回対象として考えている SSD を搭載したデータベースサーバにおいてはディスク IO によるプロセッサの待ち時間が発生する。このように、OLTP においては負荷の偏りや IO によってプロセッサのアイドル時間が生まれるため、スループット性能の低下を抑えつつ DVFS によって省電力化を行う余地が大きいと期待される。

DVFS を用いて SSD を搭載したデータベースサーバの省電力化を行うためには、動作周波数と OLTP アプリケーション性能・消費電力の関係を理解する必要がある。そこで本研究ではこれを明らかにするために、動作周波数を変化させた場合の OLTP アプリケーション性能と消費電力を実際に稼動するシステムにおいて測定する実験を行った。

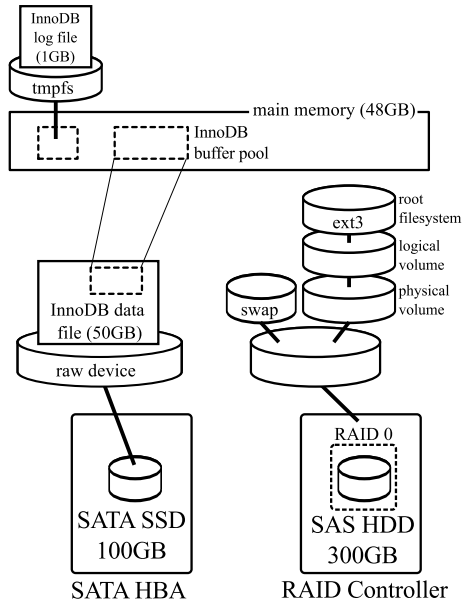

図 2　データベースサーバのストレージ構成

4. 実験環境

データベースサーバとして Dell PowerEdge R510 を 1 台用いた。このサーバには、Intel Xeon X5550 2.67GHz を 2 ソケット、メモリは 48GB(DDR3 8GB RDIMM × 6)、SSD は Samsung SS805 2.5”SATA 100GB SLC が 1 台、ハードディスクは TOSHIBA MBD2300RC 2.5”SAS 10,025RPM HDD が 1 台搭載されている。Intel Xeon X5550 では、DVFS 機能である Intel SpeedStep によって実行時の動作周波数を 1.60GHz から 2.66GHz まで 9 段階で設定することができる。ストレージの構成は図 2 に示すように、SSD の raw device をデータベース用の領域として利用し、HDD は swap 領域とルートファイルシステム用の領域として利用した。消費電力は、データロガー (Hioki 2332-20 Power Meter Module) とクランプ電流計を用いてストレージも含めたデータベースサーバ全体の消費電力を 1 秒間隔で測定した。OLTP アプリケーションのフロントエンドサーバとしては、Dell PowerEdge R900 を 1 台用いた。このサーバには、Intel Xeon X7460 2.66GHz が 4 ソケット、メモリは 128GB(DDR2 4GB FB-FIMM × 32) 搭載されている。フロントエンドサーバとデータベースサーバは 1Gb Ethernet で接続されている。

実験に用いたソフトウェア構成を以下に述べる。データベース管理システムには MySQL 5.1.41、ストレージエンジンとして InnoDB を利用し、Linux カーネル 2.6.18 上で実行した。図 2 に示すように SSD4 台で構成した RAID0 のデバイスを raw device として利用し、200GB の InnoDB データファイルを作成した。InnoDB のバッファプールは 5GB とした。また、ログファイルはサイズを 1GB とし、メモリをブロックデバイスとしてマウントする tmpfs 上に配置した。DVFS の制御には、Linux カーネルのモジュールである cpufreq を用いた。

オンライントランザクション処理のアプリケーションとして、業界標準ベンチマークである TPC-C [4] を用いた。TPC-C はフロントエンドサーバ上で実行され、データベースサーバへトランザクションリクエストを送信する。TPC-C のデータサイズの尺度である warehouse 数は 100(データサイズはおよそ 10GB) とした。また TPC-C のターミナル数は warehouse 数と同数の 100 とし、各 warehouse に一つずつターミナルを割り当てた。TPC-C のその他のパラメータは version 5.11 に準拠するよう設定を行った。

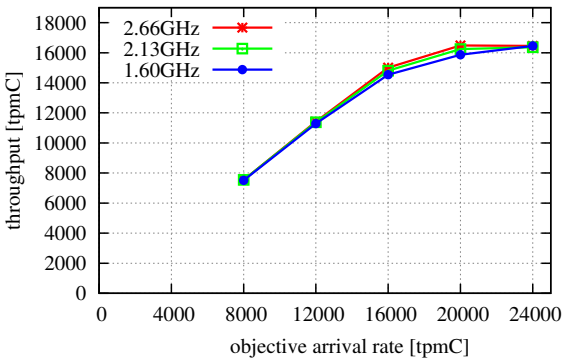

図 3　目標トランザクション到着率とスループットの関係

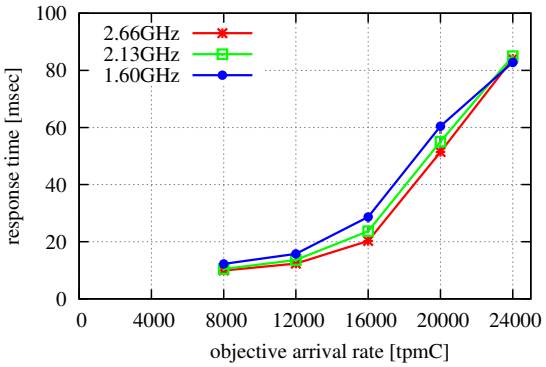

図 4　目標トランザクション到着率と応答時間の関係

5. プロセッサの動作周波数と TPC-C 性能・消費電力の関係

DVFS によりプロセッサの動作周波数を変えた場合に、それが TPC-C の性能、またシステム全体の消費電力に与える影響を調べるために実験を行った。データベースサーバにおけるプロセッサの動作周波数は、2.66GHz(最大)、1.60GHz(最小)、2.13GHz(最大と最小の中間値) の 3 段階を用いて、それぞれの動作周波数において TPC-C の目標トランザクション到着率(注1)を 8000、12000、16000、20000、24000 tpmC と変化させてスループット、応答時間、消費電力を測定した。それぞれの測定においては TPC-C を 120 分間実行し、実行開始 30 分後から 110 分後までの間の平均値を測定値とした。

実験結果を図 3、4、5 に示した。図 3 は目標トランザクション到着率とスループットの関係を表したグラフである。デー

(注1)：実験におけるトランザクション到着率は TPC-C の keying time および think time の値により調整を行った。このときの制御目標値をここで目標トランザクション到着率と呼ぶ。

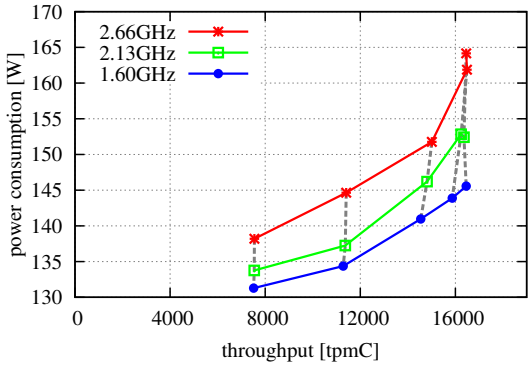

図 5　スループットと消費電力の関係

タベースサーバの処理性能が十分な場合には、スループットは目標トランザクション到着率とほぼ同じ値になるはずである。この図では、いずれの動作周波数においても目標到着率 16000tpmC まではほぼ同じスループットを達成できているが、目標到着率 20000tpmC 以降では到着率の増加に対してスループットの伸びは頭打ちとなった。このグラフにおけるスループットの最大値 16500tpmC が当該システムの最大スループット性能である。動作周波数の違いによる性能差に注目すると、目標到着率 16000、20000tpmC 以外では動作周波数にかかわらずほぼ同じスループットであった。また、目標到着率 16000、20000tpmC の場合においても動作周波数が最大の場合と最小の場合の性能差は 3%程度であった。

図 4 は目標到着率と応答時間の関係を示したグラフである。このグラフより、いずれの動作周波数においても目標到着率が大きくなるほど応答時間は長くなることが確認された。目標到着率が 24000tpmC の場合を除き、動作周波数が低くなるほど応答時間が長くなることもわかった。動作周波数を最大から最小へと下げることによる応答時間の増加量は目標到着率 20000tpmC のときに最も大きく 9.0msec であった。

図 5 はスループットと動作周波数の関係を表したグラフであり、点線で結ばれた点は同じ目標トランザクション到着率のもとで測定されたデータ点であることを表している。このグラフより、いずれの目標到着率においても動作周波数が低いほど消費電力が少ないことがわかる。また、目標到着率が大きいほど動作周波数を下げることによる消費電力の削減量も大きくなった。動作周波数を最大から最小へ下げることによる消費電力削減量は、目標到着率 8000、12000、16000、20000、24000tpmC について、それぞれ 6.9、10.3、10.8、18.0、18.6W であった。

以上の結果より、当該システムにおいては動作周波数を下げることにより消費電力を削減できることが確認された。その削減量は目標到着率が高いほど大きく、また動作周波数を最大から最小へと下げることによる性能への影響は、スループットが最大 3%の低下、応答時間が最大 9.0msec の増加であった。

6. 関 連 研 究

アーキテクチャレベルでの DVFS による省電力技術は、様々な制約条件、目的の下で提案が行われてきた。特にここ数年はマルチコアにおける DVFS の研究が主流である。Herbert らは [7] において動作周波数・電圧の制御単位である VFI(voltage/frequency island) の粒度選択のトレードオフについて検証した。また Herbert は [8] において、製造プロセスのばらつきを考慮にいれて DVFS の制御を行うことで電力効率を大幅に改善できることを示した。また DVFS 関連の研究では SPEC2000 などのベンチマークが用いられることがほとんどであったが、[7], [8] では Apache や TPC-C、TPC-H などを評価に用いて実際のサーバのワークロードでの効果を測定した。

データベースシステムの省電力化技術に関する研究では、データベース分野研究の方針を決める上で重要な役割を果たしている The Claremont Report on Database Research の 2005 年版 [9] にて、データベースエンジンの見直すべき項目として省電力化技術が指摘された。その他にもデータベースシステムの省電力化技術を指摘する文献は多い。Graefe は、データベースシステムのクエリ最適化の新たな指標としてエネルギー効率を導入するべきであると指摘した [10]。また同著においては、クエリオプティマイザやスケジューラなどの各モジュールが解決するべき課題が提示されている。Harizopoulos らはデータベースシステムの消費電力削減はハードウェアの省電力化技術だけでは不十分であり、データベースエンジンの改良が不可欠であることを指摘した [11]。しかし具体的な省電力化手法の提案は少ない。

データベースシステムの中でも、オンライン分析処理 (OLAP) に関しては省電力化技術への取り組みが比較的多く見られる。Henkel らは OLAP に適したシステム構成のシステムにおいて、各コンポーネントの詳細な消費電力の測定を行い、最適なハードウェア構成を調べた [12]。Poess らは業界標準ベンチマーク TPC-H のトップシステムの消費電力を見積り、電力効率の時系列変化を分析した [13]。また彼らは様々なハードウェア構成における DBMS の消費電力と性能の関係を分析した [14]。Dmitris らはハッシュ結合、ソートマージ結合などのデータベース処理の消費電力の分析を基に、エネルギー効率を最も高めるようなシステム設計の指針を提案した [15]。Lang らはクエリスケジューラを改善することでプロセッサの省電力モードを活用し、消費電力を削減する手法を提案した [16]。Xu らは消費電力を電力モデル化することによってクエリオプティマイザを構築し、データベースシステムの消費電力を削減する手法を提案した [17]。

オンライントランザクション処理に関する省電力化の取り組みとして、Poess らは TPC-C の歴代トップシステムの消費電力を見積り、消費電力あたりの性能の増加率が十分ではなく、現在の性能向上率が維持すると消費電力の絶対量は増え続けると予測した [18]。

7. お わ り に

本研究では、SSD を搭載した OLTP サーバにおける省電力化を目的として、DVFS により動作周波数を変化させた場合の省電力効果および性能への影響を測定した。TPC-C を用いた実験の結果、当該システムにおいては動作周波数を下げること

により消費電力を削減できることが確認された。その削減量は目標到着率が高いほど大きく最大で 18.6W であった。また動作周波数を最大から最小へと下げることによる性能への影響は、スループットが最大 3%の低下、応答時間が最大 9.0msec の増加であった。

文　献

[1] EPA. Epa report to congress on server and data center energy efficiency. Technical report, U.S. Environmental Protection Agency, 2007.

[2] EMERSON Network Power. Energy logic: Reducing data center energy consumption by creating savings that cascade across systems. White paper, Emerson Electric Co., 2009.

[3] Yuto Hayamizu, Kazuo Goda, Miyuki Nakano, and Masaru Kitsuregawa. Application-aware power saving for online transaction processing using dynamic voltage and frequency scaling in a multicore environment. In *Architecture of Computing Systems - ARCS 2011*, Lecture Notes in Computer Science. Springer, 02 2011. (to appear).

[4] Kim Shanley. Tpc releases new benchmark: Tpc-c. *SIGMETRICS Performance Evaluation Review*, 20(2):8–9, 1992.

[5] Tokyo Stock Exchange Group. Launch of “arrowhead”, the next-generation equity/cb trading system –the tokyo market enters the millisecond world with “arrowhead”–, Jan. 2010.

[6] Luiz André Barroso and Urs Hölzle. The case for energy-proportional computing. *Computer*, 40:33–37, December 2007.

[7] Sebastian Herbert and Diana Marculescu. Analysis of dynamic voltage/frequency scaling in chip-multiprocessors. In *ISLPED '07: Proceedings of the 2007 international symposium on Low power electronics and design*, pages 38–43, New York, NY, USA, 2007. ACM.

[8] S. Herbert and D. Marculescu. Variation-aware dynamic voltage/frequency scaling. In *High Performance Computer Architecture, 2009. HPCA 2009. IEEE 15th International Symposium on*, pages 301 –312, Feb. 2009.

[9] Rakesh Agrawal, Anastasia Ailamaki, Philip A. Bernstein, Eric A. Brewer, Michael J. Carey, Surajit Chaudhuri, AnHai Doan, Daniela Florescu, Michael J. Franklin, Hector Garcia-Molina, Johannes Gehrke, Le Gruenwald, Laura M. Haas, Alon Y. Halevy, Joseph M. Hellerstein, Yannis E. Ioannidis, Hank F. Korth, Donald Kossmann, Samuel Madden, Roger Magoulas, Beng Chin Ooi, Tim O'Reilly, Raghu Ramakrishnan, Sunita Sarawagi, Michael Stonebraker, Alexander S. Szalay, and Gerhard Weikum. The claremont report on database research. *SIGMOD Rec.*, 37(3):9–19, 2008.

[10] Goetz Graefe. Database servers tailored to improve energy efficiency. In *Proceedings of EDBT'08 Workshop on Software Engineering for Tailor-made Data Management*, pages 24–28, 2008.

[11] Stavros Harizopoulos, Mehul A. Shah, Justin Meza, and Parthasarathy Ranganathan. Energy efficiency: The new holy grail of data management systems research. In *CIDR 2009, Fourth Biennial Conference on Innovative Data Systems Research*, 2009.

[12] Justin Meza, Mehul A. Shah, Parthasarathy Ranganathan, Mike Fitzner, and Judson Veazey. Tracking the power in an enterprise decision support system. In *Proceedings of the 2009 International Symposium on Low Power Electronics and Design*, pages 261–266, 2009.

[13] Meikel Poess and Raghunath Othayoth Nambiar. A power consumption analysis of decision support systems. In *WOSP/SIPEW '10: Proceedings of the first joint WOSP/SIPEW international conference on Performance engineering*, pages 147–152, New York, NY, USA, 2010. ACM.

[14] Meikel Poess and Raghunath Othayoth Nambiar. Tuning servers, storage and database for energy efficient data warehouses. In *Proceedings of the 26th International Conference on Data Engineering, ICDE 2010*, pages 1006–1017, 2010.

[15] Dmitris Tsirogiannis, Stavros Harizopoulos, and Mehui A. Shar. Analyzing the energy efficiency of a database server. In *SIGMOD '10: Proceedings of the 36th SIGMOD international conference on Management of data*, New York, NY, USA, 2010. ACM.

[16] Willis Lang and Jignesh M. Patel. Towards eco-friendly database management systems. In *CIDR 2009, Fourth Biennial Conference on Innovative Data Systems Research*, 2009.

[17] Zichen Xu, Yi-Cheng Tu, and Xiaorui Wang. Exploring power-performance tradeoffs in database systems. In *Proceedings of the 26th International Conference on Data Engineering, ICDE 2010*, pages 485–496, 2010.

[18] Meikel Poess and Raghunath Othayoth Nambiar. Energy cost, the key challenge of today's data centers: a power consumption analysis of tpc-c results. *Proc. VLDB Endow.*, 1(2):1229–1240, 2008.