Optimizing Benchmark Configurations for Energy Efficiency

Meikel Poess
Oracle Corporation, 500 Oracle Pkwy, Redwood Shores, CA-94065, +1-650-633-8012
meikel.poess@oracle.com

Raghunath Nambiar
Cisco Systems, Inc., 3800 Zanker Road, San Jose, CA 95134, USA, +1- 408-527-3052
rnambiar@cisco.com

Kushagra Vaid
Microsoft, One Microsoft Way, Redmond, WA 98052, +1-425-722-4038
kvaid@microsoft.com

ABSTRACT

Historically compute server performance has been the most important pillar in the evaluation of datacenter efficiency, which can be measured using a variety of industry standard benchmarks. With the introduction of industry standard servers, price-performance became the second pillar in the 'efficiency equation'. Today with an increased awareness in the industry for power optimized designs and corporate initiatives to reduce carbon emissions, data center efficiency needs to incorporate yet another key element in this equation: energy efficiency. Initial models based on 'name-plate' power consumption have been used to estimate energy efficiency [3][6][8] while recently industry standard consortia like SPEC, TPC and SPC have started amalgating new energy metrics with their traditional performance metrics. TPC-Energy, enables the measuring and reporting of energy efficiency for transaction processing systems and decision support systems [17]. In this paper we analyze TPC-C benchmark configurations that may achieve leadership results in TPC-Energy using existing, more energy efficient technologies, such as solid states drives for storage subsystems, low power processors and high density DRAM in back end server and middle tier systems. Even though the study is based on TPC-C configurations these configuration optimizations are applicable to other benchmarks and production systems alike. We envision that the energy efficiency metrics and related optimizations to claim benchmark leadership will accelerate development and qualifications of energy efficient component and solutions.

Categories and Subject Descriptors

B.8.2 [Hardware]: Performance and Reliability – Performance Analysis and Design Aids

General Terms

Measurement, Performance, Experimentation, Standardization

Keywords

Performance Evaluation, Industry Standard Benchmarks, Server Energy Efficiency

1. INTRODUCTION

Robust and meaningful computer system benchmarks have been crucial to the advancement of the computing industry. Without them, assessing relative performance between competing vendor architectures is virtually impossible. Verifiable benchmarks have existed since corporate IT environments were first confronted with a choice between purchasing one system over another, and have been driven by their desire to compare system performance on a fair basis ever since. Benchmarks have also proven very useful to system vendors. They use benchmarks to demonstrate performance competitiveness for their existing products, to improve and monitor performance of product lines, which are still under development, and to improve their products through competition.

Until recently absolute performance and price-performance have been the only two pillars of modern system evaluation. This is reflected in the main metrics in TPC benchmark publications, where achieving the highest performance number and/or the best price per-performance number have been the most important goals, which is a direct reflection of the industry demands for high transaction rate and low cost. The TPC-C trends are in line with industry demands, often 2-3 years ahead of adapting newer technologies, for example, clusters, multi-cores and multi-tiered cache hierarchies, just to name a few.

The unprecedented growth in the reliance on computers to run the world's industries and governments, as well as the Internet, has led to an explosion in server installations, both in size and in number and the amount of energy required to operate and to cool them. Energy consumption has been increasing exponentially over recent years – a trend that will continue to accelerate into the future. This is evidenced by the Environmental Protection Agency's report [20] in which data center energy consumption within the U.S. is projected to surpass 100 billion kWh by 2011, with an annual electricity cost of $7.4 billion.

The requirement to reduce energy costs and usage, while still satisfying the mounting demand for additional computing resources, has become the greatest challenge for many IT organizations today. Data center growth is constrained by hard limits on energy consumption due to facility constraints, limitations of the power grid and/or policy decisions. Public awareness of data center energy consumption and its impact on the environment has influenced many companies to place a higher priority on choosing "greener" technologies in "doing their part" to protect the environment.

Consequently, today's corporate IT environment demands the inclusion of energy efficiency in standard benchmarks as the third pillar in system evaluation. Especially large datacenters whose power consumption is becoming their key cost factor are very interested in purchasing energy efficient systems. To address this shift in priorities of IT purchasers, the Transaction Processing Performance Council (TPC), Standard Performance Evaluation Corporation (SPEC) and the Storage Performance Council (SPC) have developed new energy specifications to enhance their widely used benchmark standards [17]. Initial observation of the introduction of the energy metrics has been very positive. System vendors and industry experts have welcomed such initiatives. For example the SPECPower benchmark has gained significant momentum already and vendors are heavily investing in accelerating the development of energy efficient components. Similarly, with three published TPC-C benchmark results, the TPC-Energy benchmark is showing signs of industry acceptance.

We anticipate that the introduction of the energy metric into the TPC benchmark will have profound impacts on future systems architecture. Considering energy efficiency consciousness in data centers, we believe that some of the very first steps to produce leadership in energy efficient benchmarks will be to optimize the top energy consuming components, namely the storage subsystem, processors and memory. In the past years, the industry observed that performance, reliability and features of industry standard servers improved significantly while price-performance dropped at a rate of 1.28 times per year [7].

In [6] and [8] we analyzed historic TPC-C and TPC-H results and developed an energy consumption model for both TPC-C and TPC-H, which allowed us to estimate the power consumption of any TPC-C and TPC-H result. In this paper we enhance the power consumption model for TPC-C to analyze different approaches by which system vendors can achieve high performance and low energy consumption while keeping performance invariant. As a result we are able to quantify the energy savings of these approaches.

The remainder of this paper is organized as follows. In Section 2 we briefly review typical TPC-C configurations and their energy consumption. Section 3 reviews the power estimation model for TPC-C systems, which was originally presented in [8]. Section 4 develops modifications to [8] so that power savings of alternative low power components can be computed and applies the alternative components to 4 published TPC-C results to quantify their power savings. Section 5 summarizes our results.

2. OPTIMIZING TPC-C FOR ENERGY EFFICIENCY

The proposed TPC-Energy specification augments the existing TPC benchmarks by allowing for optional publications of energy metrics alongside their performance results. The energy metric is represented as the ratio of the Energy (typically measured in Watts-seconds) consumed by all subsystems of the benchmarked configuration– this includes servers, storage, clients, network switches – to the work completed (typically measured as number of Transactions) over the benchmark interval. After moving the time element to the denominator, the TPC-Energy metric is plainly represented as Watts/Performance. TPC-C energy efficiency is measured as electricity consumed [W] per work unit finished, as measure in [Watts/KtpmC]. For more information on the new TPC-Energy specification see [18].

Energy efficient systems will result in a lower value for the TPC-Energy metric, since the power consumption [W] value is in the numerator. In this paper, we will explore several techniques for component selection for achieving energy savings and a correspondingly higher score on the energy metric. The studies conducted in [8] shows that the storage subsystems are the predominant energy consumers of TPC-C benchmark systems. A typical TPC-C system is designed in 3 tiers as follows:

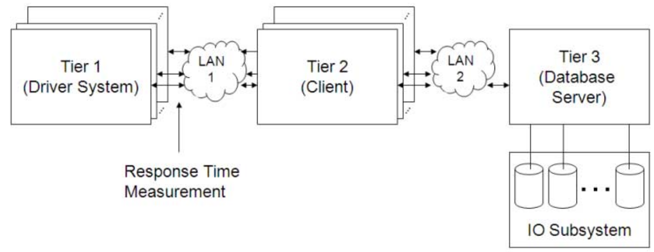

Figure 1: Typical TPC-C System Setup (conceptual)

The Driver System provides Remote Terminal Emulator (RTE) functionality, used to emulate the target terminal population and their emulated users during the benchmark run. The RTE is not part of the system under test. The clients run the TPC-C application and commercially available transaction monitor systems, such as Tuxedo or Com+. The Database Server runs the database management systems (DBMS) such as DB2, Oracle, SQLServer and Sybase. The tiers are connected through a local area network (LAN). The transaction response time is measured on the driver system: the start time is when the transaction is generated by the Driver System and the end time is when the commit is received by the Driver System.

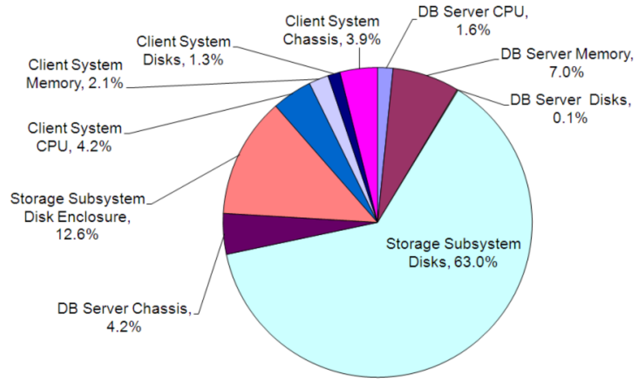

Figure 2: Power consumption of major parts used in TPC-C benchmarks. Source: PVLDB 1(2): Page 1229-1240 (2008)

Figure 2 shows the power consumption of each component of a typical TPC-C system as a percentage of the overall power consumption. 75.6 percent of all power is consumed by the storage subsystem (disks and disk enclosures). 12.9 percent is consumed by the database server (CPU, Memory, OS, Disks and Chassis), and 11.5 percent is consumed by the client systems (CPU, Memory, OS Disks and Chassis). Within the storage subsystem the largest power consumers are the disk drives (63 percent of the total power), followed by the disk enclosures (12.6 percent of the total power). 7 percent of the power is consumed by the database server's memory and 1.6 percent is consumed by the database server's CPUs. The power consumption of the internal disks for the operating system in the database server is insignificant. The clients' CPUs consume 4.2 percent; 2.1 percent is consumed by memory; 1.3 percent is consumed by the disks; and 3.9 percent is consumed by the chassis.

Recent developments in energy efficient technologies allow hardware vendors to reduce power consumption while still delivering ground braking performance at affordable cost. In a recent TPC-H publication HP has demonstrated that solid state drives can be deployed successfully in TPC benchmarks [19]. Our thesis is that in order to claim leadership in energy efficiency for TPC-C, hardware vendors will focus on energy efficient technologies in the following order: 1) Storage subsystem disks; 2) Database server memory; 3) Low power CPUs.

Replacing storage subsystem disks with solid state drives seems to be the most obvious choice because they contributed up to 63% of total power. Database server memory contributes to about 7% of the overall power consumption of TPC-C systems. Replacing regular memory DIMMs with low voltage memory can have a significant effect. Lastly client and server CPUs contribute to 5.8% of overall power consumption. Replacing regular CPUs with equivalent lower voltage CPUs can also have a positive impact on the overall power consumption while still guaranteeing high performance.

3. ENERGY ESTIMATION MODEL

To estimate the impact of various energy savings techniques on TPC-C benchmarks, we enhance the power consumption model proposed in [8] such that the most energy consuming components can be substituted with energy efficient alternatives. The model proposed in [8] assumes that the peak power consumption of an entire system during the measurement interval can be derived from the aggregate of the individual nameplate power consumptions. There are two key conditions that need to be met in order to apply this power consumption model. Only workloads that utilize a system one hundred percent for the entire duration of the measurement interval are suitable. The second requirement is system balance. Depending on the application and system, an optimal component ratio has to be maintained to keep all components (CPU, disks, controllers etc.) utilized during the measurement interval. If a system does not have the optimal ratio between these components, the power consumption model will not produce accurate estimates. This is because it assumes that all components are used during the duration of the measurement. TPC-C benchmark publications fulfill both requirements. The TPC-C benchmark is constructed such that the performance numbers are obtained during the steady state of the system, during which all components are fully utilized. Secondly, the typical business objective of a TPC-C benchmark publication is to demonstrate performance and price-performance. Hence all TPC-C publications maintain optimal component ratios. No vendor can afford to over-configure one part of the system because all parts that are used in a benchmark need to be disclosed and priced. And price-performance is widely being used by system vendors to showcase their advantages over those of their competitors. For instance, if a vendor over-configures a database server with 50% more CPUs, those CPUs need to be priced, and, since the number of CPUs is disclosed, the result will be used by competitors to show that they can achieve the same performance with fewer CPUs. Lastly, some database vendors tie their pricing model to the number of CPUs, while some tie it to the number of disks. This inconsistency makes it even more unattractive to publish unbalanced TPC-C performance results. All enhancements we make in Section 4 will need to take these two key conditions into consideration.

In the remainder of this section we briefly explain the original power consumption model, proposed in [8]. The modifications needed to accommodate other technologies, such as solid state drives, low voltage memory DIMMs and low voltage CPUs, are explained in their respective subsections of Section 4. Our energy estimation model includes Tier 2 and Tier 3 systems (see Figure 1). In TPC terms, these systems are referred to as the System Under Test (SUT). The database server (Tier 3) is typically comprised of one or more compute systems and a storage subsystem, usually comprising of one or more RAID devices (Redundant Arrays of Independent Disks). We refer to the 'container' of the RAID devices as disk enclosures. Our power estimation model includes the following components of Tier 2 and Tier 3: CPU, memory, disks, server chassis, and disks enclosures. We differentiate between database server and client CPUs since the client CPUs are typically less powerful than the database server CPUs. This is because a TPC-C system is usually sized around the database server, that is, the number of clients and their CPU choice is a function of how fast the database can drive those systems. We also differentiate between internal and external disks for the same reason. The following table summarizes the components that are considered in each Tier:

Tier	Component Description	Number of Components	Power Consumption per single component[W]
1	CPUs per client	C_{CC}	$P_C \in [55,165]$ see [1][4]
	Memory DIMMs per client	C_{CM}	P_M
	Internal disks per client	C_{CD}	$P_{DI} \in [7.2,19]$ see [9][10]
	Number of clients	C_{CL}	P_C calculated in Equation 1
2	CPUs per server	C_{SC}	$P_C \in [55,165]$ see [1][4]
	Memory DIMMs per server	C_{SM}	P_M
	Internal disks per server	C_{SD}	$P_{DI} \in [7.2,19]$ see [9][10]
	External enclosures	C_E	P_{ST}, calculated in Equation 4
	External disks per enclosure	C_{DE}	$P_{DE} \in [7.2,19]$ see [9][10]
	Number of servers	C_{SV}	P_{SV}, calculated in Equation 2

Table 1: Components of the Energy Estimation Model per Tier

We determine the peak power consumption for each of the components listed in Table 1. We obtain the peak power consumption of CPUs from their manufacturer's specification [1][4]. The peak power consumption for CPUs is depicted as Thermal Design Power (TDP). We approximate the power consumption of main memory by assuming 4.5 watts per Gigabyte memory DIMM[1] as done in other publications [3][5]. Finally the

[1] DIMM: Dual In-Line Memory Module

peak power consumption levels of the disk drives are obtained from the manufacturers' web sites [9][10].

Recent studies [2][11] suggest that the power consumption of a server chassis and its infrastructure (fan, power supply etc.) can be expressed as a percentage of the nameplate power consumption of its main components. In addition to the above, TPC-C servers, both for Tier 2 and Tier 3, contain CPU, main memory, and internal disks. Since the server chassis and its infrastructure (fan, power supply, and so forth) are sized according to its components (CPU, memory), we include its power consumption as 30 percent of the power consumption of its components plus a fixed overhead of 100 watts.

$$P_{CL} = C_{CL} * (C_{CC} * P_C + C_{CM} * P_M + C_{CD} * P_{DI}) * 1.3 + 100 \quad (1)$$

$$P_{SV} = C_{SV} * (C_{SC} * P_C + C_{SM} * P_M + C_{SD} * P_{DI}) * 1.3 + 100 \quad (2)$$

Similar to the server chassis case, we approximate the power consumption of the disks enclosures with 20 percent power overhead of the aggregate power consumption of all external disks.

$$P_{ST} = C_E * C_{DE} * P_{DE} * 1.2 \quad (3)$$

Hence, the total power consumption of the entire system (P_S) can be estimated as the sum of the power consumption of the clients, server and storage subsystem.

$$P_S = P_{CL} + P_{SV} + P_{ST} \quad (4)$$

Using the measurement methodology presented above, Table 3 shows the power consumption of the client systems, database server, and the storage subsystem of systems A, B, and C, both estimated by the power model and measured with power meters. The paper, "Power Provisioning for a Warehouse-sized Computer" [3] refers to this as the difference between the nameplate value and actual peak power. For each of the three layers, Table 2 shows the difference in percent between the estimate and the measurement. The power model over-estimates the system's power consumption. The difference varies between 10 and 25 percent. Overall the difference between the three systems is between 14 and 17 percent. The paper [3], which applied a different workload on smaller systems, shows that the difference between the nameplate model and actual peak power consumption is 30 percent. The 15 percent difference in the power consumption estimation with our power model and the actual measurements is smaller than the difference found in the paper [3]. The difference between the modeled number of a server (251W) and its measured power consumption referenced in the paper [3] is about 40 percent. In order to calibrate our power model to the TPC-C workload, we abate our power model number by 15 percent.

Tier	Power Consumption [W]								
	System A 2 CPUs, id=107111201[14]			**System B** 2 CPUs id=108010701 [16]			**System C** 4 CPUs, id=107090502 [13]		
	Power Model	**Measure-ment**	**Diff [%]**	**Power Model**	**Measure-ment**	**Diff [%]**	**Power Model**	**Measure-ment**	**Diff [%]**
Storage	7728	6720	13	6973	6240	11	11631	9600	17
Clients	1086	845	22	1813	1352	25	2369	2028	14
DB-Server	796	705	11	618	510	17	909	820	10
Total	**9610**	**8270**	**14**	**9404**	**8102**	**14**	**14909**	**12448**	**17**

Table 2: Comparison Power Consumption Model and Power Measurements of Three TPC-C Systems (Source: [8])

4. LOW POWER ALTERNATIVES

In this section we model the impact of using three low power alternative components to the power consumption of the four recent TPC-C results with TPC benchmark identifiers: 109012001, 109022301, 109052101 and 109052101. We chose these results because they were published within seven months by different hardware and software vendors on x86 processors for which we had power estimates available from the Intel websites. We first estimate the power consumption of the published result to establish a baseline by using the power model, developed in [8], and then build hypothetical systems substituting original components with energy efficient alternatives. The hypothetical systems are built such that they deliver the same performance as the original. This requires calculating how many low power components are needed to substitute their high power counterparts. In our particular case we will calculate how many solid state drives are needed to substitute the rotational disks, how many low voltage memory DIMMs are needed to replace the regular memory DIMMs how many low voltage CPUs are needed to replace the regular CPUs of the above systems.

4.1 Baseline Power Estimates

The result with the TPC benchmark id 109012001 was published on a HP ProLiant DL580G5 system with 4 Intel X7460 2.67 GHz processors and 256 GB of main memory, running the Oracle database 11g Standard Edition on Oracle Enterprise Linux . It was first published on January 1^{6th} 2009. This configuration was optimized to achieve the top 4-processor performance result.

The result with the TPC benchmark id 109022301 features a Dell PowerEdge 2900. This system supports up to 2 processors, but was configured with a single Intel Xeon X5440 2.83 GHz processor and 32 GB of main memory, running Oracle Database 11g Standard Edition on Microsoft Windows Server 2003 Standard Ed. It was first published on February 2nd 2009. This configuration was optimized for price-performance leadership.

The Result with the TPC benchmark id 108091501 was published on an IBM System x3850 M2 with 4 Intel X7460 2.67 GHz processors and 256 GB of main memory, running Microsoft SQL Server 2005 Enterprise x64 Edition on Microsoft Windows Server 2003 Enterprise x64 Enterprise R2. Configuration wise it is comparable to the server used in the first result (TPC benchmark id 109012001), It was first published September 15th 2008. Like

ID 109012001, this configuration also was configured to achieve the top 4-processor performance result.

The Result with the TPC benchmark id 109052101 features an HP ProLiant ML350 G6, which supports up to 2 processor, but this benchmark configuration was only configured with a single Intel E5520 2.27 GHz processor and 72 GB of main memory, running the Oracle Database 11g Standard Edition on Oracle Enterprise Linux. This result was first published on May 21st 2009. Like Result 109033001 this configuration was also optimized to achieve price-performance leadership. Compared to Result 109033001 it used a newer generation processor, 2.25 times the main memory, 2 times the number of hard drives and 3 times the number of client systems and, consequently, achieved 2.25 times the performance at about twice the total system cost, reducing the price-performance by 10%.

Table 3 highlights the hardware and performance characteristics of these four results and Table 4 calculates the power consumption of these systems using the power consumption model developed in [8]. For each of the results we list the number of components in each system tier (see column labeled C), how much power they consume (see column labeled P_C) and finally how much power is consumed by all components (see column P_T). Power consumption for the client systems, the database server and the storage subsystem are printed separately. If we sort the results by energy consumption [W], we get the following ranking: 3,1,4,2. That is the third result has the largest power consumption with 16209.1 Watts followed by the first result with 12512 Watts. The fourth result is on second to last energy consumer at 2864.7 Watts while the second result uses the least power at 1432.5 Watts. Total power consumption is not a fair comparison between systems because it does not take performance into consideration. Hence, we will use power per tpmC as our performance comparison. Using power per tpmC we get the same ranking. The third and first results are still the highest energy consumers with 0.0274 and 0.0223 Watts per tpmC, while the least power per tpmC consumer is the third and the fourth results with 0.0145 and 0.0161 Watts per tpmC respectively.

TPC Result ID	**109012001 (1) [12]**	**109022301 (2) [12]**	**108091501 (3) [12]**	**109052101 (4) [12]**
CPU	4x Intel X7460 2.67 GHz	1x Intel Xeon X5440 2.83 GHz	4x Intel X7460 2.67 GHz	1x Intel E5520 2.27 GHz
Memory	32 x 8GB PC2-5300	8 x 4GB 667 FB	32 x 8GB PC2-5300	9 x 8GB PC3-8500R
Disks	1002 x 36GB 10K 50 x 146GB 10K	25x72GB 15K 75x36GB 15K,6x146GB 15K	1344 x 73GB 15K 4G FC 16 x 500GB 7.2K	200 x 36GB 15K 6 x 300GB 10K
Performance [tpmC]	639,253	104,492	631,766	232,002
Price/Performance [$/tpmC]	0.97	0.6	2.58	0.54
Power/Performance [W/tpmC]	0.0223[2]	0.0161[2]	0.0274[2]	0.0145[2]

Table 3: Configuration details of sample results

Component	**TPC Result 1 109012001 [12]**			**TPC Result 2 109022301 [12]**			**TPC Result 3 108091501 [12]**			**TPC Result 4 109052101 [12]**		
	C[3]	**P_C[4][W]**	**P_T[5][W]**	**C[2]**	**P_C[3][W]**	**P_T[4][W]**	**C[2]**	**P_C[3][W]**	**P_T[4][W]**	**C[2]**	**P_C[3][W]**	**P_T[4][W]**
CPUs per client	1	80	80	1	65	65	2	80	160	1	80	80
DIMMs per client	2	2.3	4.7	2	2.7	5.4	4	1.8	7.2	2	2.3	4.6
Internal disks per client	1	9.2	9.2	1	11.4	11. 4	1	9.2	9.2	1	19	19
Client chassis	1	128.2	128.2	1	137.4	137.4	1	125.1	125.1	1	126.4	126.4
Total Client Power[6]	**8**	**188.7**	**1509.6**	**1**	**290.8**	**290.8**	**12**	**228.1**	**2736.6**	**3**	**234.7**	**704.2**
CPUs per server	4	130	520	1	130	130	4	95	380	1	80	80
Server Memory DIMMs	32	5.5	176	8	5.5	44	32	5.6	179.2	18	5.5	100.8
Internal disks per server	2	7.2	14.4	6	19	114	1	9.2	9.2	1	9.2	9.2
Server chassis	1	187.1	187.1	1	186.4	186.4	1	270.5	270.5	1	157	157
Total Server Power[5]	**1**		**870**	**1**		**403.2**	**1**		**713.1**	**1**		**295**
External enclosures	47	41.1	1930	4	38.4	153.6	84	30.3	2545.9	10	38.5	385.3
External disks	1050	9.1	9650	96	8	768	1360	10.3	12730	206	11.8	1926
Total Storage Power[5]			**10133**			**806.4**			**13366**			**2022.7**
Total Power			**12512**			**1432.5**			**16209.1**			**2864.7**

Table 4: Baseline power consumption for sample results

[2] 4 digit precision because of calculation: Estimated Power [W] divided by Performance [tpmC]
[3] Number of components
[4] Power consumption per component
[5] Total Power consumption
[6] Reduced by 15% according to power estimation model [8]

4.2 Low Power Alternative 1: Solid State Drives

We believe that solid state drives (SSD) are the future of high performance storage subsystems. There are many SSD options in the industry, but many of them are not fully enterprise ready yet. The most enterprise ready SSDs are SATA SSDs, which started to demonstrated fast and reliable performance at large scales in industry standard benchmark publications [19]. In our study we use Samsung SATA 2 SSDs (SLC). It uses 3 Watts of power during peak usage. Our tests show that its IO performance is equivalent to that of 12 traditional SAS drives.

Hence, we substitute SAS hard drives in our power model with SATA SSDs in the following way: Let's assume the storage subsystem in a benchmark publication that achieved x tmpC used e_{HDD} enclosures with de_{HDD} number of SAS drives per enclosure. That is, the benchmark used a total of $d_{HDD} = e_{HDD} * de_{HDD}$ number of disks. The number of SSD is bound by two factors, IOPS performance and overall capacity to hold the TPC-C database. First we calculate how many SSD are required to achieve the same performance as SAS drives and then we calculate how many SSDs are needed to hold the required TPC-C capacity. Finally, we take the maximum of the two numbers.

Since each SSD drive is performance wise equivalent to 12 SAS drives, the number of SSD drives can be computed by dividing the number of SAS drives by 12 and taking the ceiling:

$$d'_{SSD} = \left\lceil \frac{d_{HDD}}{12} \right\rceil \quad (6)$$

Due to the TPC-C scaling model the database size is tight to the performance of the database [tpmC]. Simplified, the higher the performance of a system is the larger must the database be. This equates to about 20 MByte per TpmC. Hence, capacity wise the number of SSD needed can be computed by multiplying the tpmC achieved by a factor of 20 and dividing the number by the capacity of the SSD, which is 72 Gbyte. The final number is computed as the ceiling of the result:

$$d''_{SSD} = \left\lceil \frac{x\,[TpmC]*20\left[\frac{MByte}{TpmC}\right]}{1024*72[MB]} \right\rceil = \left\lceil \frac{x*20}{1024*72} \right\rceil \quad (7)$$

Since the system needs to be sized for both d'_{SSD} *and* d''_{SSD}, we calculate the minimum number of SSD needed as the maximum of the two:

$$d_{SSD} = \max\,(d'_{SSD}, d''_{SSD}) = \max\left(\left\lceil \frac{d_{HDD}}{12} \right\rceil, \left\lceil \frac{x*20}{1024*72} \right\rceil\right) \quad (8)$$

Similar to hard drives, SATA SSDs are situated in enclosures. Performance of enclosures is limited by the performance of their controllers and network connections to the database server. Although TPC-C results publish their number of enclosures, they do not disclose the performance of them. Hence, in our modified model with SSDs we assume we use the same type of enclosures as the published hard drive configuration, while adjusting the number of SATA SSDs per enclosure accordingly. Hence, we compute the number of SSDs per enclosure by dividing the number of enclosures used in the hard drive configuration by 12 and taking the floor:

$$de_{SSD} = \left\lfloor \frac{de_{HDD}}{12} \right\rfloor \quad (9)$$

This approach of calculating the number of SSDs and the number of enclosures is conservative since it might overestimate the number of SSDs and the number of enclosures. For instance, if a system used one enclosure with 13 disks, we calculate the number of SSDs to 2 and the number of enclosures to 2.

Using Equation 8 to calculate the number of solid state drives and Equation 9 to calculate the number of solid state drive enclosures, we compute the power consumption for each system using solid state drives. Table 5 is an excerpt of Table 4, i.e. it shows those columns that are relevant to hard drives. In the last three columns it shows how many solid state drives and enclosures are need, their component power consumption and their total power consumption.

	Hard Drives			**Solid State Drives**		
Result 1	**C**	**P_C[W]**	**P_T[W]**	**C**	**P_C[W]**	**P_{ST}[W]**
Enclosures	47	41.1	1930	174	1.2	104.4
Disks/SSD	1050	9.1	9650	174	3	443.7
Total			**10133**			**548.1**
Result 2	**C**	**P_C[W]**	**P_T[W]**	**C**	**P_C[W]**	**P_{ST}[W]**
Enclosures	4	38.4	153.6	15	1.2	18
Disks/SSD	96	8	768	29	3	74
Total			**806.4**			**92**
Result 3	**C**	**P_C[W]**	**P_T[W]**	**C**	**P_C[W]**	**P_{ST}[W]**
Enclosures	84	30.3	2545.9	186	0.6	111.6
Disks/SSD	1360	10.3	12730	186	3	474.3
Total			**13366**			**585.9**
Result 4	**C**	**P_C[W]**	**P_T[W]**	**C**	**P_C[W]**	**P_{ST}[W]**
Enclosures	10	38.5	385.3	63	1.2	37.8
Disks/SSD	206	11.8	1926	63	3	160.65
Total			**2864.7**			**198.5**

Table 5: Power consumption with SSD

Since one solid state drive can substitute twelve traditional hard disks, the number of solid state drives is drastically reduced. For instance, Result 3 uses 1360 disk drives and only 186 SSDs. This reduces the storage power consumption from 13366 W to 585.9 W. To illustrate the power savings with SSDs, the bar chart in Figure 3 compares the estimated power consumptions per tpmC of the published configurations with our modified configurations that would use solid state drives for their storage subsystem. The chart shows that SSDs reduce power consumption of TPC-C systems between 53 and 79.2 percent. Compared to the power consumption of its baseline (0.0237 W/tpmC) the power consumption of the solid state drive solution of the third result is 79.2 percent lower (0.0049 W/tpmC). The power reduction for the first result is similarly high at 78.2 percent, while the power reductions are lower for the second and fourth results (53% and 64.4%), but overall considerably high.

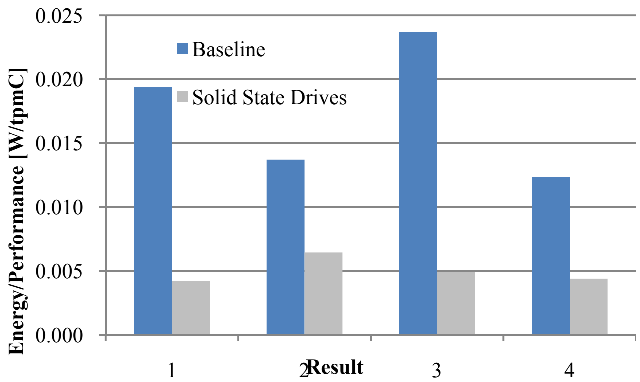

Figure 3: Effect on power consumption of solid state drives

4.3 Low Power Alternative 2: Lower Voltage CPUs

CPU manufacturers typically offer low voltage variants of the standard product lineup as an energy efficient option for customers. These CPUs are rated at about 25-30% lower power dissipation than the regular parts. Reducing the voltage on a product line has some disadvantages though, since the frequency may have to be lowered to counteract the effect of slower circuit speeds at the lower voltage setting. In some cases, the low voltage variants may have a smaller cache size to keep power dissipation under check. The tradeoffs made for the low voltage product offerings typically depend on variations across manufactured lots which is a function of process technology and micro-architectural design artifacts. For these reasons, in some cases it is entirely possible that a low voltage variant is available with no performance compromises over the regular voltage parts. Another point to note is that because the supply for low voltage parts is limited, typically the pricing on these parts is slightly higher compared to the standard product line, and hence the cost-benefit analysis is required for the power savings versus price.

Since there is no performance impact, substitution of low voltage CPU variants in our power model is very straight forward. The power consumption of all CPUs in the client (2 Tier) and database server (Tier 3) are modified to reflect the power consumption of their low voltage counterparts. The following Table 6 shows the power consumption for both types of CPU used in the TPC-C benchmarks analyzed in this paper.

Intel Xeon CPU	Power Consumption [W]		Used in Result
	Original	Low Voltage	
X7460 - 2.67 GHz	130	65	1
E5440 - 2.83 GHz	80	50	1,3
QC 5440 - 2.83 GHz	80	50	2
QC E5205-1.86 GHz	65	50	2
X5570 2.93GHz	95	60	3
E5520 2.27 GHz	80	60	4
E5405 2.00 GHz	80	50	4

Table 6: Power consumption: regular and low power CPUs

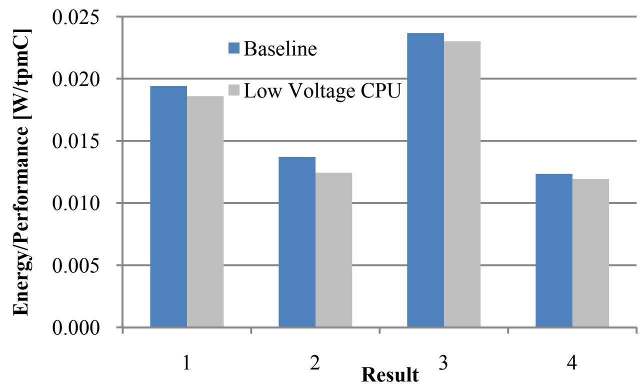

Figure 4: Effect on power consumption low voltage CPUs

Figure 4 shows the power savings with low voltage CPUs for all four TPC-C results compared to the baseline. The power savings is less dominant as in the previous case because there are far less components that are substituted. Result 2 has the highest power savings at 9.3% while Result 3 has the least power savings of 2.8%. The reason for the large range in power savings and with low voltage CPUs is due to different number and type of processors deployed in these benchmarks. Result 3 uses one processor in the database server and two processors in a single client.

When tuning a system for performance one always catches the next bottleneck. If we apply this to tuning for power-performance, we would consider the hard disks as the first power-performance bottleneck. Hence, it is more interesting to evaluate the power savings with low voltage CPUs once we have substituted the hard disks with SSDs. Figure 5 shows the effect on power consumption using low voltage CPUs after all traditional disks have been substituted with SSDs. Now the relative power saving is more dramatic at 17.4 percent for the first, 15.4 percent for the second, 13.9 percent for the third and 10.1 percent for the fourth result. Interestingly, in this calculation the result most profiting from low voltage CPUs is the first result, while in Figure 4 it was the second result. This is due to the size of the storage subsystem. With 96 hard disks the second result has by far the smallest storage subsystem. Hence, the power reduction achieved with low voltage CPUs has the largest affect in this result.

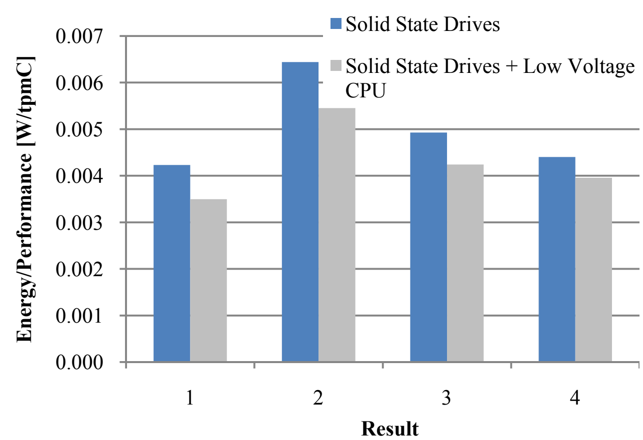

Figure 5: Effect on power consumption of low voltage CPUs after all traditional disks have been substituted with SSDs

4.4 Low Power Alternative 3: Higher density and lower voltage DRAM

With improvements in process technology, semiconductor manufacturers are able to pack more transistors on the same die by virtue of being able to shrink circuits. For memory technology, this means that DRAM devices manufactured on leading process technologies have higher transistor density (typically double) on the same die size compared to older generation. This has two important effects – by doubling the DRAM size, fewer DRAMs are needed per memory DIMM, and fewer DIMMs are needed for a given amount of system memory capacity - reducing the power per GB for memory. However, newer technology DRAMs are initially pricey because of supply limitations so it may not be possible to take advantage of this right away, and in any case, a careful tradeoff needs to be done for the price vs. power tradeoff.

Another way to achieve power savings with memory is to opt for lower voltage (LV) DRAMs. The LV variants for DDR3 run at 1.35V versus the standard 1.5V offering lower power for the same memory capacity. Here too, an analysis of price vs. power is necessary since LV DRAMs are typically slightly higher in cost compared to standard DRAMs. There is no significant performance tradeoff for either higher density DRAMs or low voltage DRAMs, since the memory DIMM has to meet the appropriate standards (JEDEC DDR2, DDR3 etc)[5]. Table 7 shows the difference in power consumption between regular and low voltage DRAM, which is about 25%.

As with the low voltage CPUs substitution of higher density DRAM variants in our power model is very straight forward. The power consumption of all regular DRAMS in the client (2 Tier) and database server (Tier 3) are modified to reflect the power consumption of their higher density counterparts. The following table shows the power consumption for both types of DRAM used in the TPC-C benchmarks analyzed in this paper.

Type of Memory DIMM	DRAM Power Consumption [W]	
	Regular	Low voltage
8 GB DDR2 DIMM w/ 2Gb	5.5	4.125
8 GB DDR3 DIMM w/ 2Gb	5.6	4.2
4 GB DIMM w/ 1Gb	9.3	6.9
4 GB DIMM w/ 2Gb	5.4	4.1

Table 7: Power consumption: regular and high density DRAM

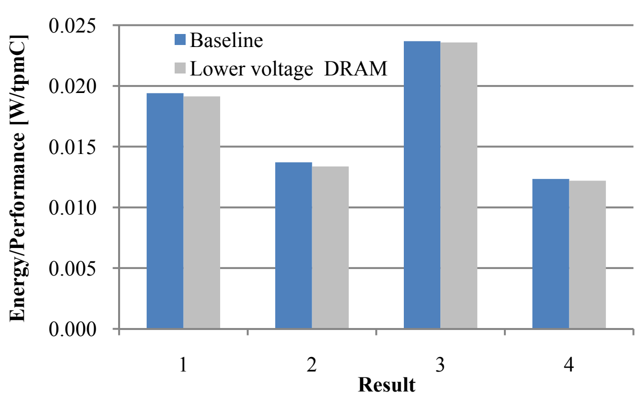

Figure 6: Effect on power consumption of high density and low voltage DRAM

Figure 6 shows the power savings in each of the four results. The power savings range between 0.4 and 2.5%. The second result gained the most from low voltage RAM, namely 2.5 percent. The first and fourth result had similar power savings of 1.4 and 1.2 percent respectively. The third result benefits the least with only 0.4 percent. On the first sight this is surprising because our previous study shows that memory in the database server and clients attribute to about 9.1% of the overall power consumption. The low power savings with low voltage DRAM in our sample systems is due to them using less memory compared to the main body of TPC-C results. Result 1 used 256 GB, result 2 used 32 GB, result 3 used 144 GB and result 4 used 72GB.

As in the low voltage CPU case we evaluate the benefits of low voltage DRAM after both solid state drives and low voltage CPUs are used. Although still low the power savings with low voltage DRAM is at about 2 percent. Results 1 and 2 have power savings of 1.8 percent, Result 3 has a power saving of 1.4 percent and Result 4 has a power saving of 2.6 percent.

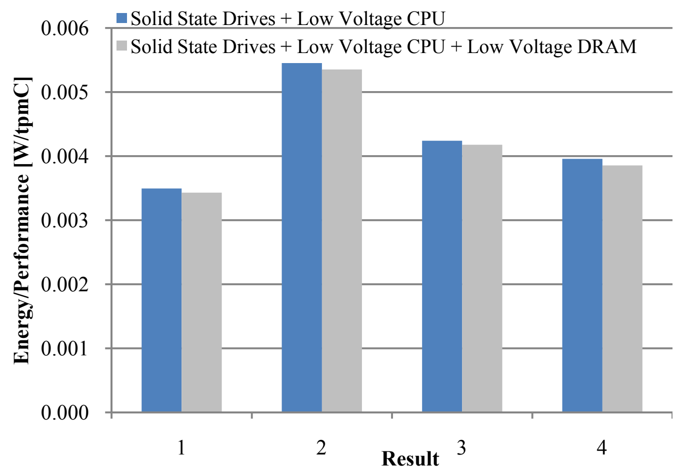

Figure 7: Effect on power consumption of low voltage DRAM after all traditional disks have been substituted with SSDs and low volt CPUs have been installed

5. SUMMARY

Energy efficiency benchmarks are expected to be quickly adapted by the industry by publishing benchmarks showcasing their energy savings techniques. The analysis in this model shows how systems might be deployed in benchmark publications if vendors first target energy saving technologies in the areas of storage subsystem, CPU and memory. The result show that if SSDs are used for the storage subsystem, regular CPUs are exchanged with low power CPUs and regular DRAM is substituted with high density DRAM power consumption per tpmC of TPC-C results can drop by as much as 82%. As a result the distribution of the 'power hawks' as identified in Figure 2 will change as depicted in Figure 8. It shows that server and client CPU resemble still a very high percentage of power consumption, namely 30 percent. The power consumption of the storage subsystem is also still very high 26 percent. Interestingly, the chassis used in the client and server machines are taking a larger percentage. Prior to the power savings techniques described in this paper they consumed 8.2

percent of the entire power (see Figure 2). Now they consume 29 percent. Memory power consumption in the client and server is relatively low at 10 percent. The client and server disks are almost negligible at 7 percent.

This indicates that once vendors have changed their system to use solid state drives for the storage subsystems and lower voltage CPUs and DRAMs they will need to explore other techniques to reduce power in the storage, for instance variable clock speed of processors and more power efficient fans or alternative cooling techniques.

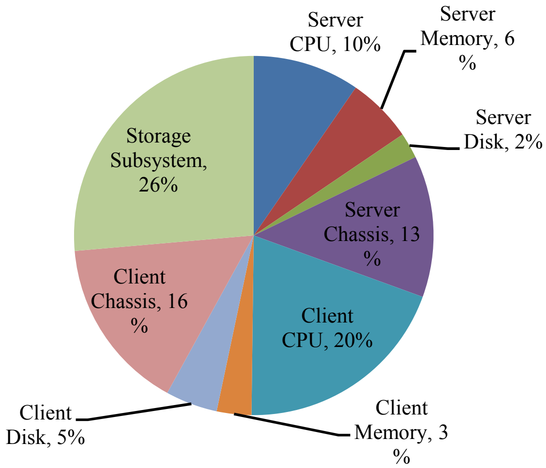

Figure 8: Average power consumption of major parts used in TPC-C after power consumption improvements

6. ACKNOWLEDGEMENTS

The authors thank Satinder Sethi and Ray Glasstone for their support in writing this paper and Travis Hindley and Roy Zeighami for their comments and feedback.

7. REFERENCES

[1] AMD Opteron family of processors: http://www.amd.com

[2] Dimitris Economous, Suzanne Rivoire, Christos Kozyrakis, and Parthasarathy Ranganathan: Full-system Power Analysis and Modeling for Server Environments, Workshop on Modeling, Benchmarking, and Simulation (MoBS), June 2006

[3] Fan, Xiaobo, Wolf-Dietrich Weber, and Luiz André Barroso. 2007. Power Provisioning for a Warehouse-sized Computer. Proceedings of the 34th International Symposium on Computer Architecture in San Diego, CA. Association for Computing Machinery, ISCA '07.

[4] Intel family of processors: http://www.intel.com

[5] Micron Technology, Inc. Calculating Memory System Power for DDR. http://download.micron.com/pdf/technotes/ddr/TN4603.pdf

[6] Poess M., Nambiar R. O.: A Power Consumption Analysis of Decision Support Systems. WOSP/SIPEW 2010: 147-152

[7] Poess M., Nambiar R. O.: Tuning Servers, Storage and Databases for Energy Efficient Data Warehouses. ICDE2010.

[8] Poess, M, Nambiar, R.O., Energy cost, the key challenge of today's data centers: a power consumption analysis of TPC-C results. PVLDB 1(2): 1229-1240 (2008)

[9] Specifications of disk drive from Hitachi: http://www.hitachigst.com/tech/techlib.nsf/techdocs/

[10] Specifications of disk drives form Seagate http://www.seagate.com/staticfiles/support/disc/manuals/, http://www.seagate.com/docs/pdf/datasheet/

[11] Suzanne Rivoire, Mehul Shah, Parthasarathy Ranganathan, Christos Kozyrakis, Justin Meza: Modeling and Metrology Challenges for Enterprise Power Management, IEEE Computer, December 2007

[12] TPC results with IDs: 109012001, 109033001, 108091501,109052101http://www.tpc.org/tpcc/results/tpcc_last_ten_results.asp

[13] TPC-C Benchmark Result 107090502: http://www.tpc.org/tpcc/results/tpcc_result_detail.asp?id=107090502

[14] TPC-C Benchmark Result 107111201: http://www.tpc.org/tpcc/results/tpcc_result_detail.asp?id=107111201

[15] TPC-C Benchmark Revision 5.9: http://www.tpc.org/tpcc/default.asp

[16] TPC-C Result 108010701: http://www.tpc.org/tpcc/results/ tpcc_result_detail.asp?id= 108010701

[17] Poess M., Nambiar R. O., Vaid K., Stephens, J.M., Huppler, K. Haines, E.: Energy Benchmarks: A Detailed Analysis. E-Energy 2010

[18] TPC-Energy Specification 1.2.0: http://www.tpc.org/tpc_energy/spec/TPC-Energy_Specification_1.2.0.pdf

[19] TPC-H result 09100801 http://www.tpc.org/results/individual_results/HP/HP_DL380G6_100GB_090409_2P_2.93GHzQC_ES.pdf

[20] EPA Report "EPA Report to Congress on Server and Data Center Energy Efficiency" http://www.energystar.gov/ia/partners/prod_development/downloads/EPA_Datacenter_Report_Congress_Final1.pdf