A Case Study: Performance Evaluation of a DRAM-Based Solid State Disk

Hitoshi Oi

The University of Aizu

November 2, 2007

Japan-China Joint Workshop on Frontier of Computer Science and Technology (FCST) 2007, Wuhan, China

Outline

- Introduction: Processor and I/O speed gap
- DRAM-based Solid State Disks
- Evaluation Methodology and Environment
- Evaluation Results
- Conclusions and Future Work

Introduction and Background

- Speed gap problem is more serious between processors and I/O devices than processors and main memory
- Only few percent of CPU time is utilized for modern processors running I/O intensive workload.
- In on-line transaction processing (OLTP) workload, this phenomena is more stressed since its write-intensive nature makes simply caching disk blocks less effective.

DRAM-based Solid State Disks

- DRAM-based Solid State Disks (or Drives, SSDs) are storage devices build upon the same technology as the main memory of computers.
- Their access speed is high, but they are volatile and high in bit-per-dollar price. However, reduced price of DRAM chips made DRAM-based SSDs feasible as commercial products (e.g. FujiXerox, Texas Memory Systems, Solid Data Systems).
- In the commercial product SSDs, reliability is enhanced by various mechanisms such as battery backing up. Unlike main memory, SSDs are stored in the separate cabinet from the server. Therefore, its data should be safe from the software crash of the operating system and applications.

Evaluation Methodology

- We run OSDL DBT-2, an open-source implementation of the TPC-C benchmark, on the system with SSDs and HDDs.
- DBT-2 is an OLTP workload and most of its disk accesses are writes. (the read/write ratio is around 1:5). Therefore, caching disk blocks is not very effective.
- We run DBT-2 for varying W (primary scale factor) and investigate scalability, I/O behavior and CPU utilization, and find how much improvement over HDDs can be expected with a DRAM-based SSD.

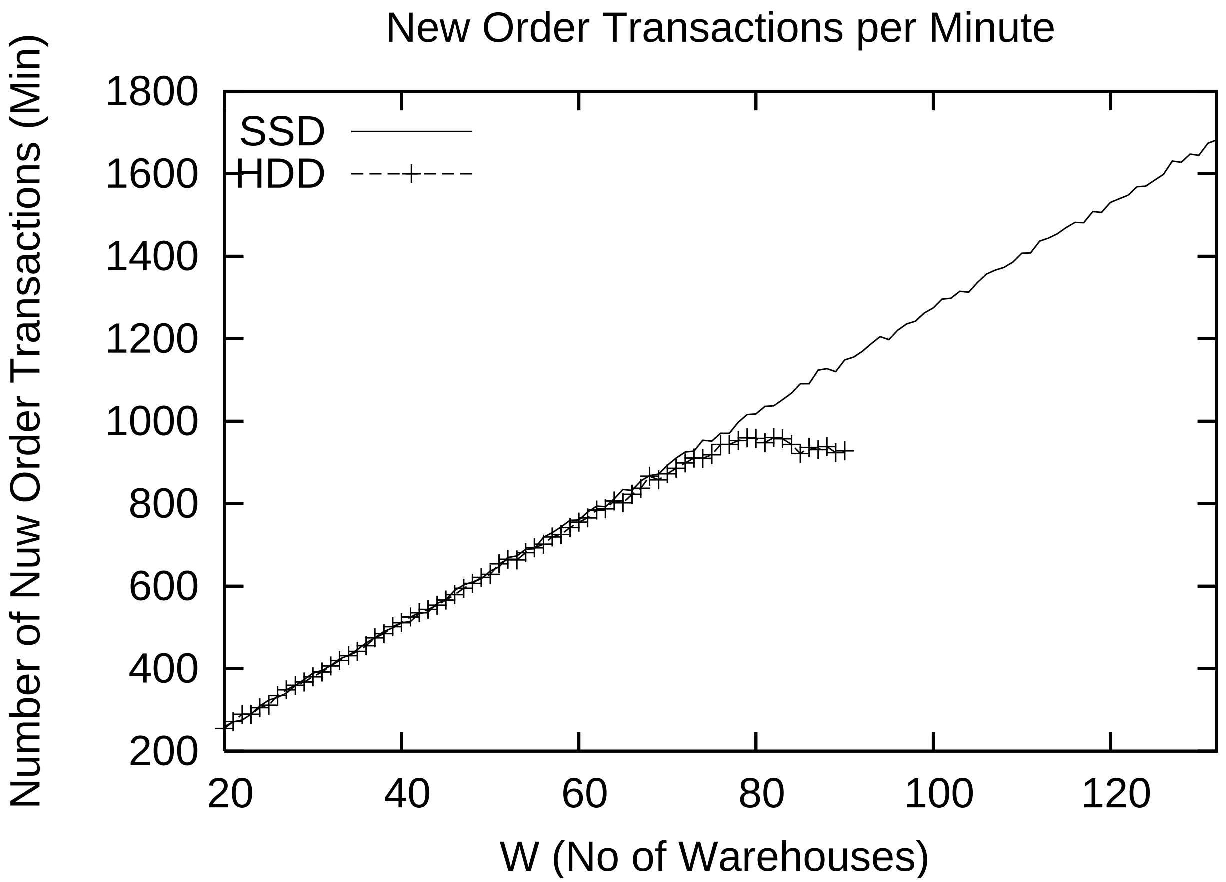
Evaluation Environment

Software	
Operating System	CentOS 4.4 (Linux)
Kernel	2.6.9
Benchmark	OSDL DBT-2 (V 0.40)
DBMS	PostgreSQL (V 8.2.4)
Hardware	
CPU	Xeon 3GHz $\times$ 2
Memory	2GB
HDDs	SCSI 15Krpm $\times$ 4
SSD	GigaExpress (16GB)

Evaluation Results: Scalability

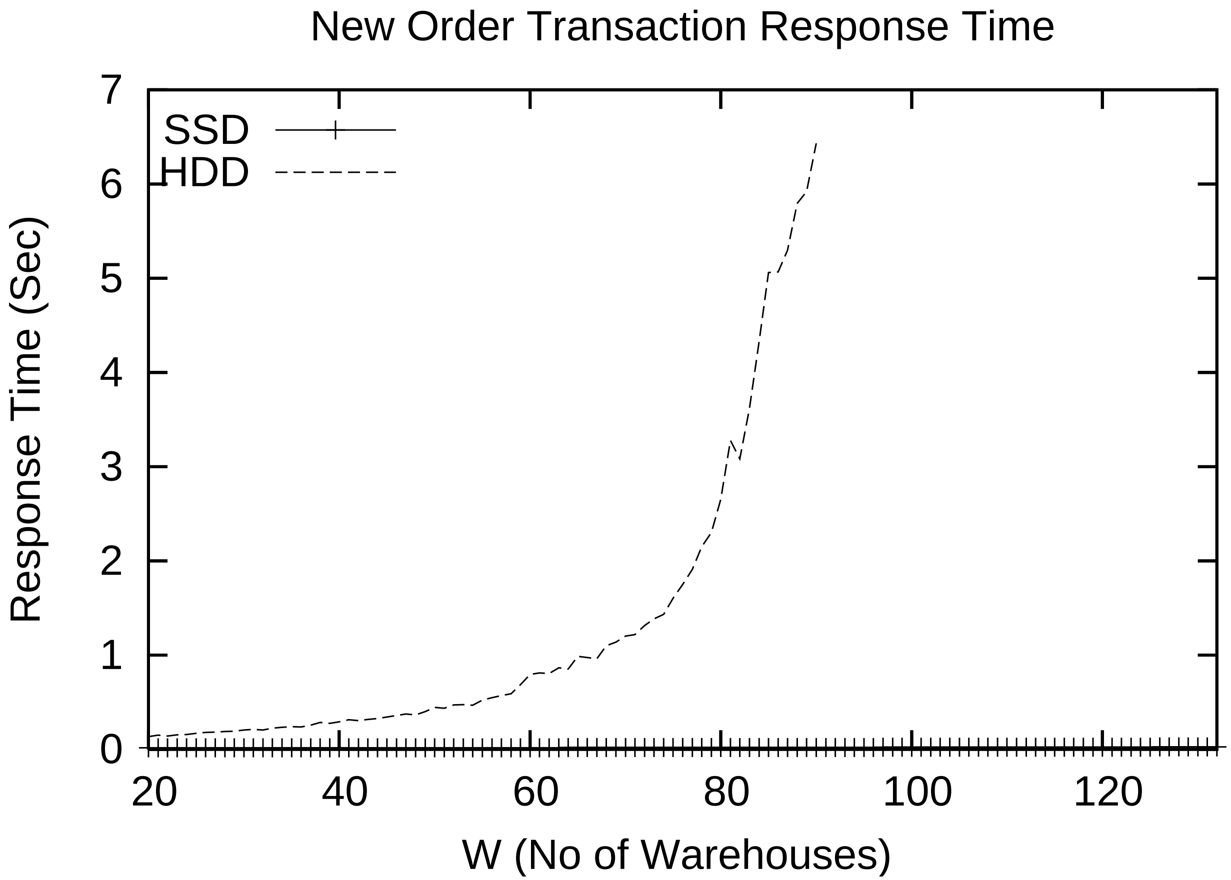
- New Order Transaction per Minute (NOTPM) is the primary performance metrics of TPC-C/DBT-2.
- In HDDs, NOTPM saturates at around W = 80, while in SSD, NOTPM grows W = 132 and beyond ($W = 132$ is the capacity limit of SSD).
- Response Time for New Order transaction must be 5 seconds or shorter. In HDDs, it starts increasing rapidly around W = 75. In SSD, it is around 0.02 second and increase is negligible.

Number of New Order Transactions Per Minute



New Order Transaction Response Time

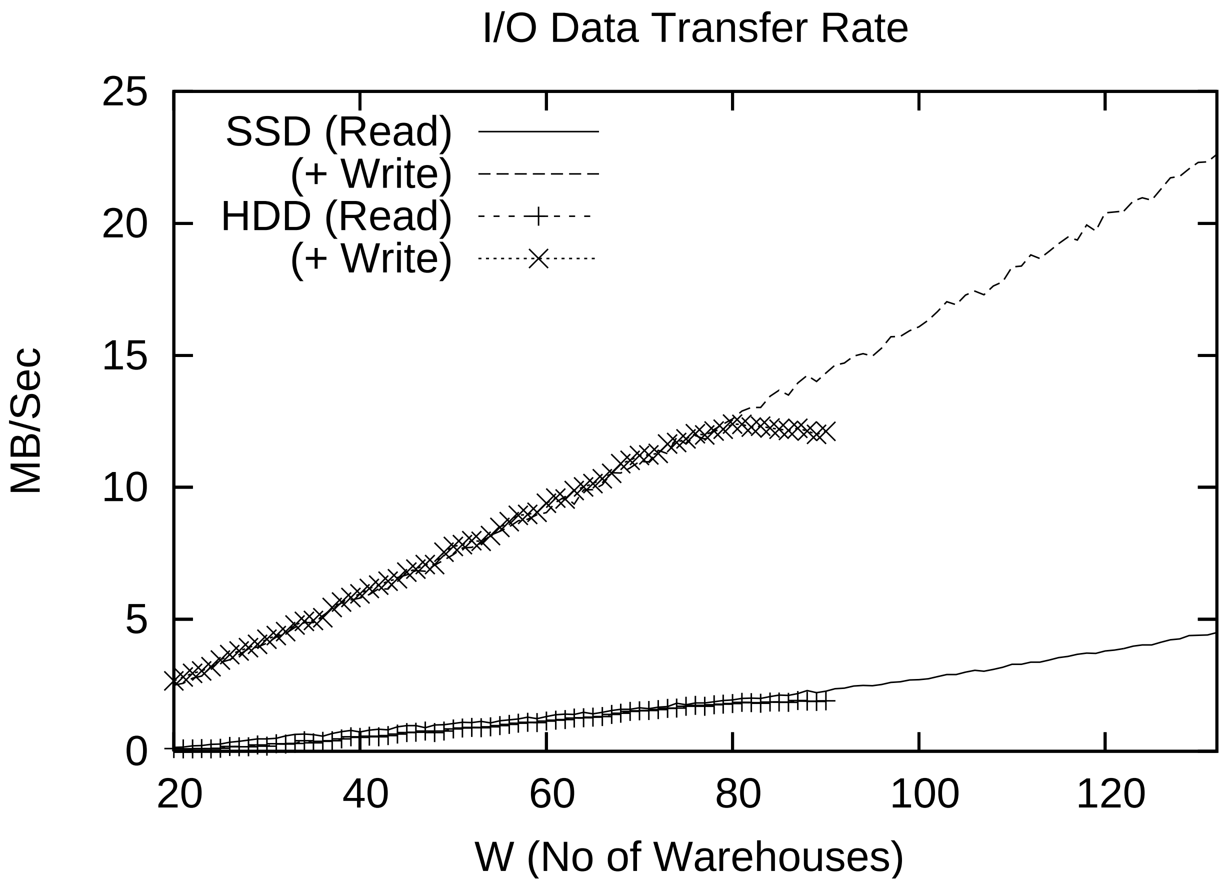
At 90th Percentile



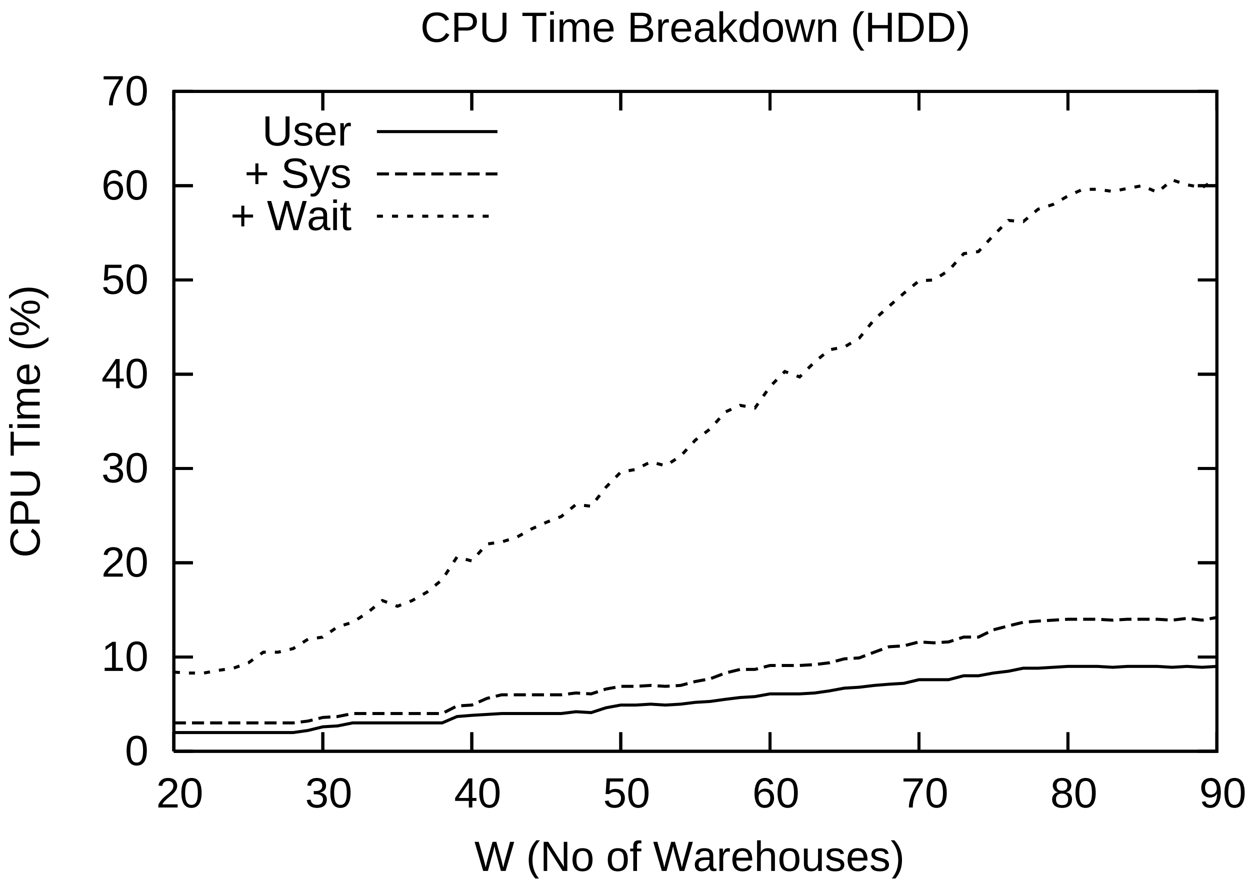
Evaluation Results: I/O and CPU Utilization

- Most I/O are for write accesses: read is 20% of less. In HDDs, I/O rate is saturated at around W = 80. In SSD, I/O rate is not saturated event at its capacity limit of W = 132.
- In HDDs, up to 45% of CPU time is wasted for I/O wait and only up to 14% is used for useful tasks (i.e. CPU time in user + system modes).
- In SSD, I/O wait is 2% or lower and 21% of CPU time is used for useful task.

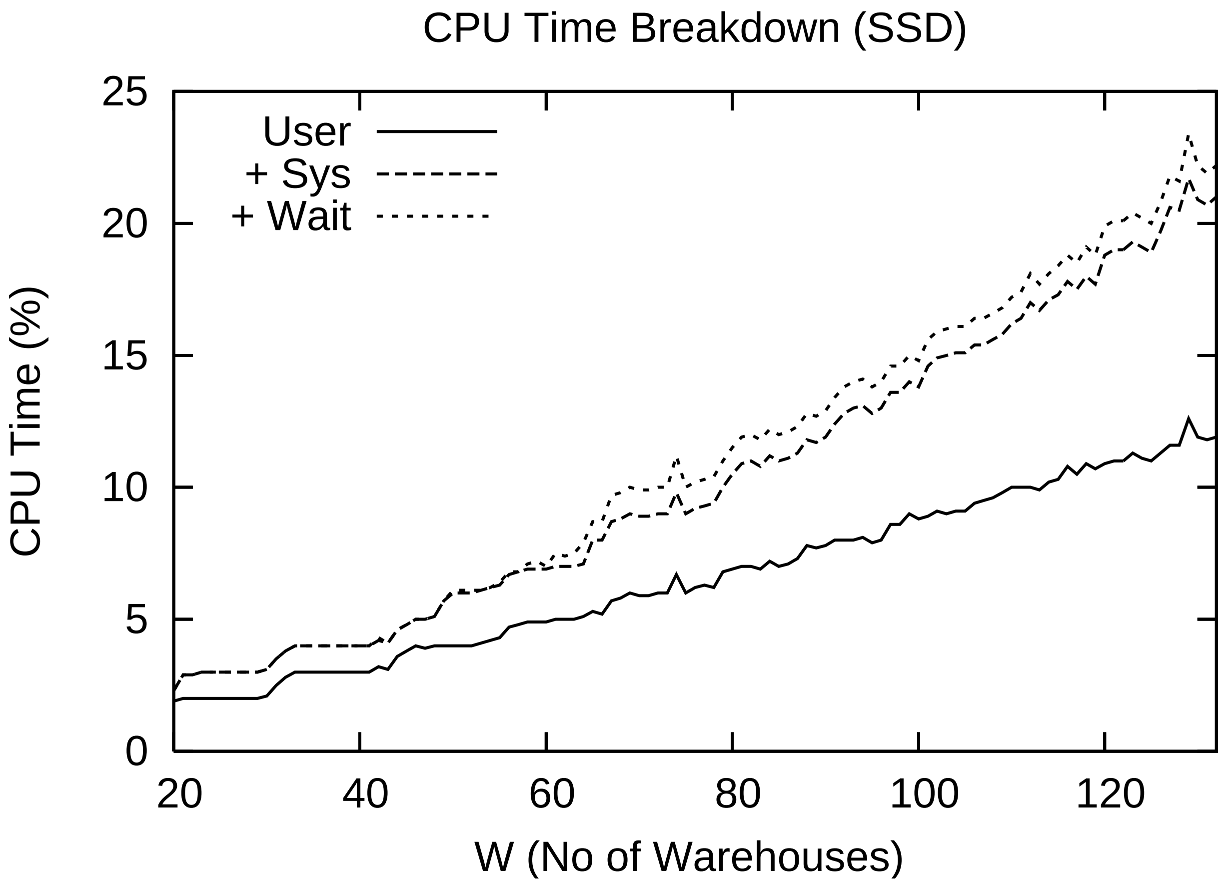
I/O Data Transfer Rates



CPU Time Breakdown (HDD)



CPU Time Breakdown (SSD)



Conclusions and Future Work

- DRAM-based SSD is lower in latency and higher in bandwidth. The scale factor for DBT-2 was increased by 70% or more (limited by the capacity of SSD).
- More than 40% of CPU time was wasted for I/O wait for HDDs but it was 2% or less for SSD.
- The per-bit price of SSDs is still much higher than that of HDDs (about two orders of magnitude). Selective storage space management (e.g. storing only 'hot' file in SSD) needed.
- The high throughput of SSDs looks attractive for virtualized systems. Enhancement of SSDs toward virtualization (file management, security, etc) is another topic for further investigation.

Thank You for Attention

Please send comments/questions to `hitoshi@u-aizu.ac.jp`