Using Solid State Drives As a Mid-Tier Cache In Enterprise Database OLTP Applications

September 17, 2010

Badriddine Khessib, Microsoft

TPCTC 2010
Singapore

Agenda

- SSD by the numbers
- SSD caching approaches
- Cost model
- Benchmark
- Results
- Conclusion

SSD by the numbers

Storage Type	Size (GB)	Price ($)	Perf	$/GB	$/Perf	Watts	W/GB
DRAM	4	143	1000000	35.75	0.000143	6	1.5
SSD (SLC)	120	1244	10000	10.37	0.1244	2	0.017
SSD (MLC)	160	480	10000	3.00	0.048	2	0.013
SAS(15K)	300	216	200	0.72	1.08	14	0.047
SAS(10K)	300	186	150	0.62	1.24	8	0.027
SAS(7.2K)	2000	293	100	0.15	2.93	5	0.003
SATA(7.2K)	2000	293	100	0.15	2.93	5	0.003

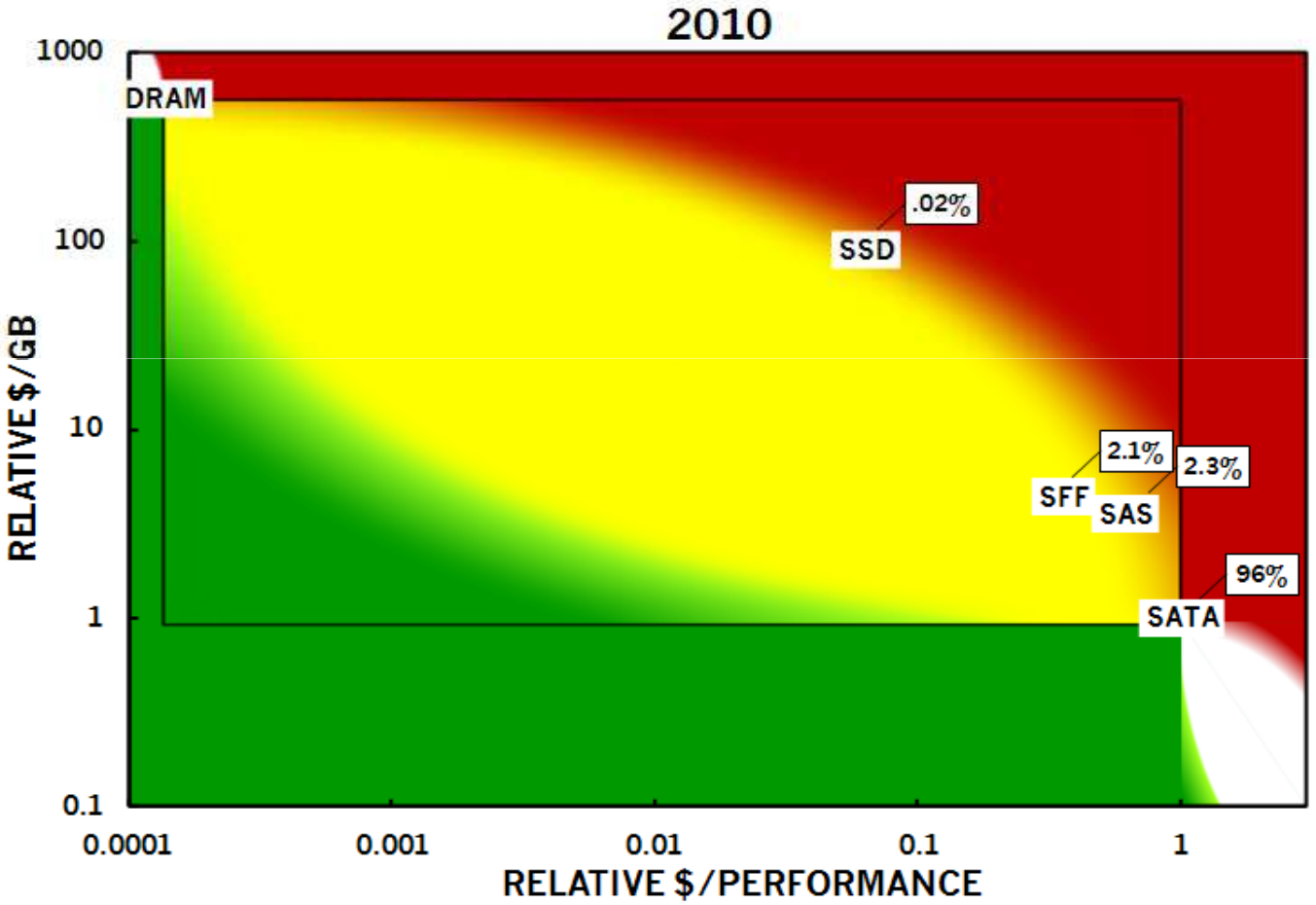
SSD by the numbers: Cost

- SSD price: $11/GB
  - Cheaper than DRAM: $36/GB
  - More expensive than SATA: $0.15/GB
- SSD Perf Cost: $.13
  - Cheaper than SATA: $2.93
  - More Expensive than RAM: $0.000143

SSD by the numbers: Power

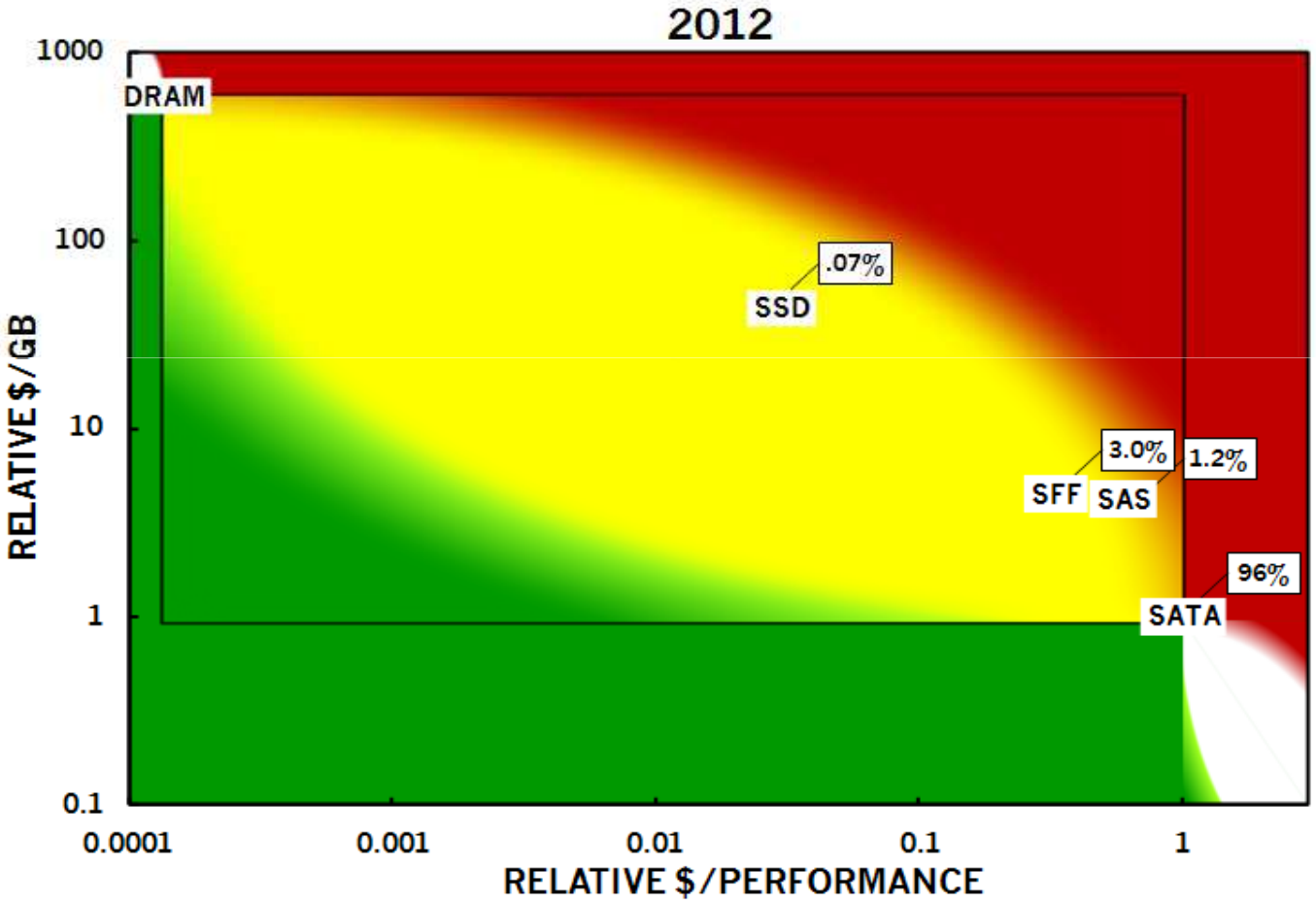
- SSD Power Consumption: 0.013 Watt/GB
  - DRAM more power hungry: 1.5 Watt/GB
  - SATA more power efficient: 0.003 Watt/GB

Storage Technologies by the $



* Hetzler, Steven R. The storage chasm: Implications for the future of HDD and solid state storage. *http://www.idema.org. [Online]* December 2008.

Storage Technologies by the $



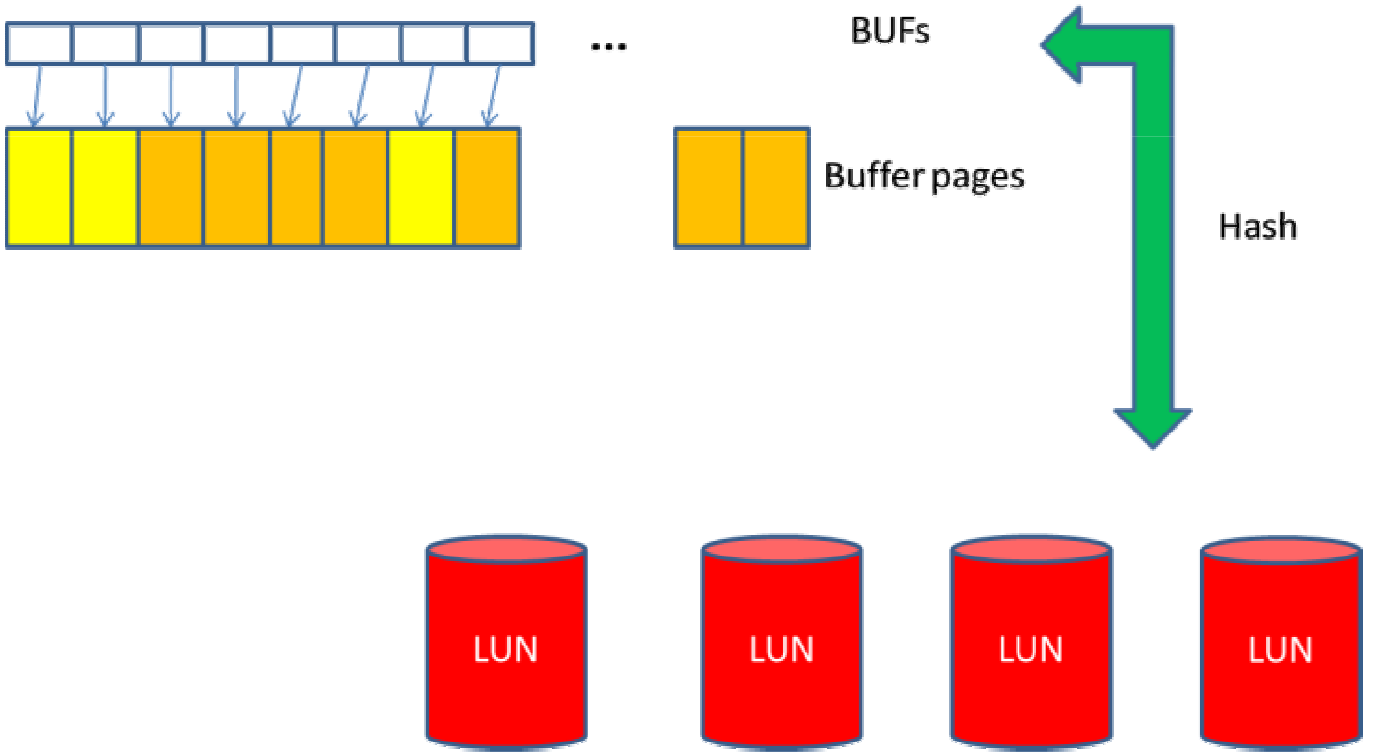
* Hetzler, Steven R. The storage chasm: Implications for the future of HDD and solid state storage. *http://www.idema.org. [Online]* December 2008.

Recap

- Price will continue to drop but …
  - SSD/HDD will remain high
- Endurance/Reliability getting better
  - 4/5 year Warranty
  - Full path data integrity being implemented
  - Battery backup/Super Caps (DBMS) will be available
- SSD Watt/GB much lower than memory but higher than HDD
- SSD caching needs to be explored as an alternative to complete HDD replacement

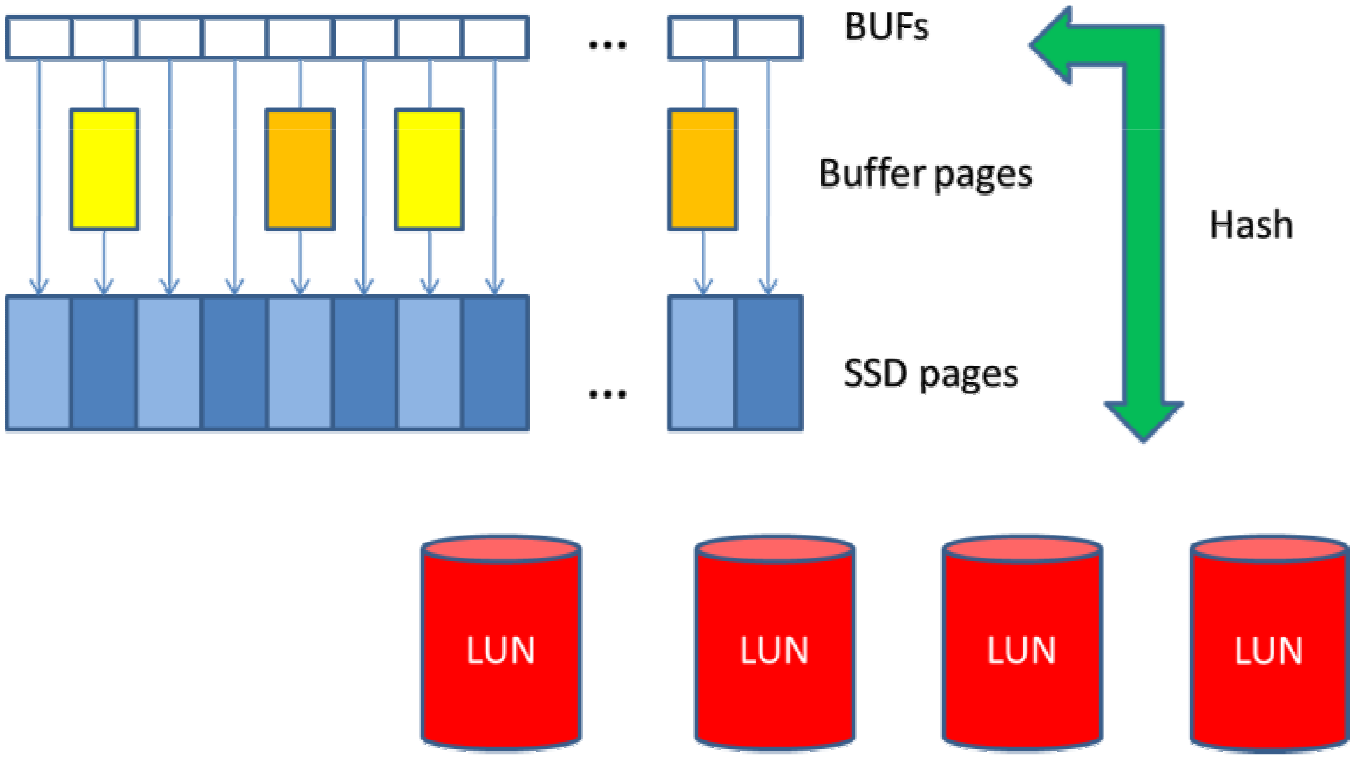
SSD Caching: Software

- Typical Working of DBMS memory manager.



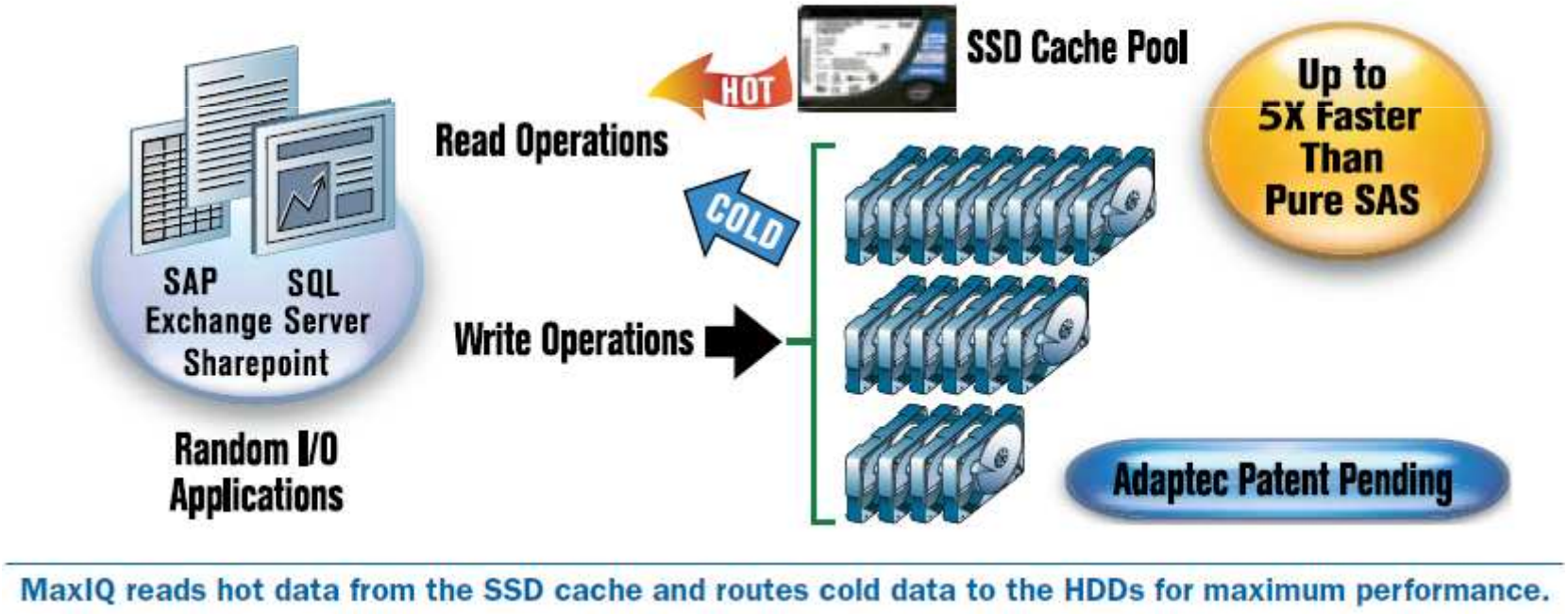
SSD Caching: Software (cont.)

- A modified DBMS memory manager that uses SSD as cache



SSD Caching: Hardware

RAID controllers from PMC/Adaptec & LSI



Microsoft IP licensed and productized by Adaptec

The Cost Model

- HDD: IO is expensive
  - $Cost_{HDD} = IOPS * \$/IOPS_{HDD} + Power_{HDD} * \$/Watt$
- SSD: GB is expensive
  - $Cost_{SSD} = GB * \$/$
- For SSD to be viable:
  - $Cost_{HDD} > Cost_{SSD}$
  - $IOD * \$/IOPS_{HDD} + PD_{\Delta} * \$/Watt > \$/GB_{SSD}$

The Cost Model (cont.)

- $IOD * \$/IOPS_{HDD} + PD_{\Delta} * \$/Watt > \$/GB_{SSD}$
  - IOD: IOPS/GB, workload dependent
  - $\$/IOPS_{HDD}$: $1.24
  - PD_{Δ}: 0.01 Watt/GB
  - \$/Watt: $10
  - $\$/GB_{SSD}$: $10.37

- Solve for IOD:
  - IOD > 8.28

Benchmark

- Need a benchmark that generates different IO densities
- Standard TPC-E does not that
- a modified TPCE:
  - GB/user
    - Change 300 day Trade Orders to 30 days
    - Change scale factor from 500 to 2000

Benchmark (cont.)

- a modified TPCE:
  - Transaction Mix:
    - Market watch has a weight of 0
    - Trade lookup frame 4 has 100% weight
    - Trade update “Read rows” set to 2
    - Trade update “update rows” set to 1
    - No data maintenance transactions

Setup (System)

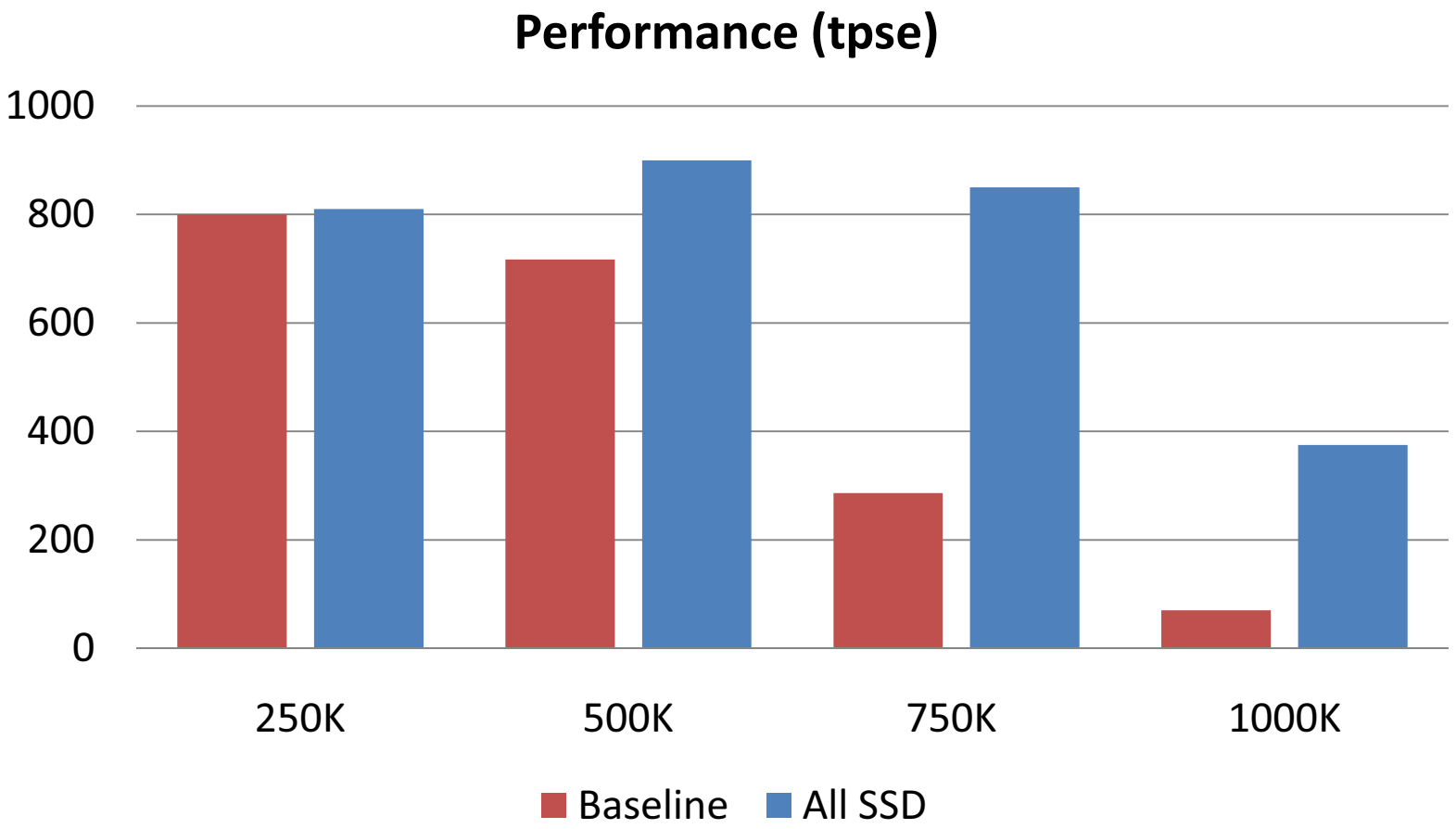
System configuration	
Platform	HP ProLiant DL380 2U Model
CPU	2 x Intel Xeon L5520 2.26GHz, 8MB L3 Cache, 60W
Memory	12 x 4GB PC3-10600R (48GB total)
HDDs	20 x 146GB 10K SAS 2.5” , 5 x 300GB 10K SAS 2.5”
Controller	PMC-Sierra’s Adaptec hardware accelerator
Power Supply	2 x 750W Gold N+1 Redundant Power
Network	Single Dual Port Embedded Intel NIC

Setup (database)

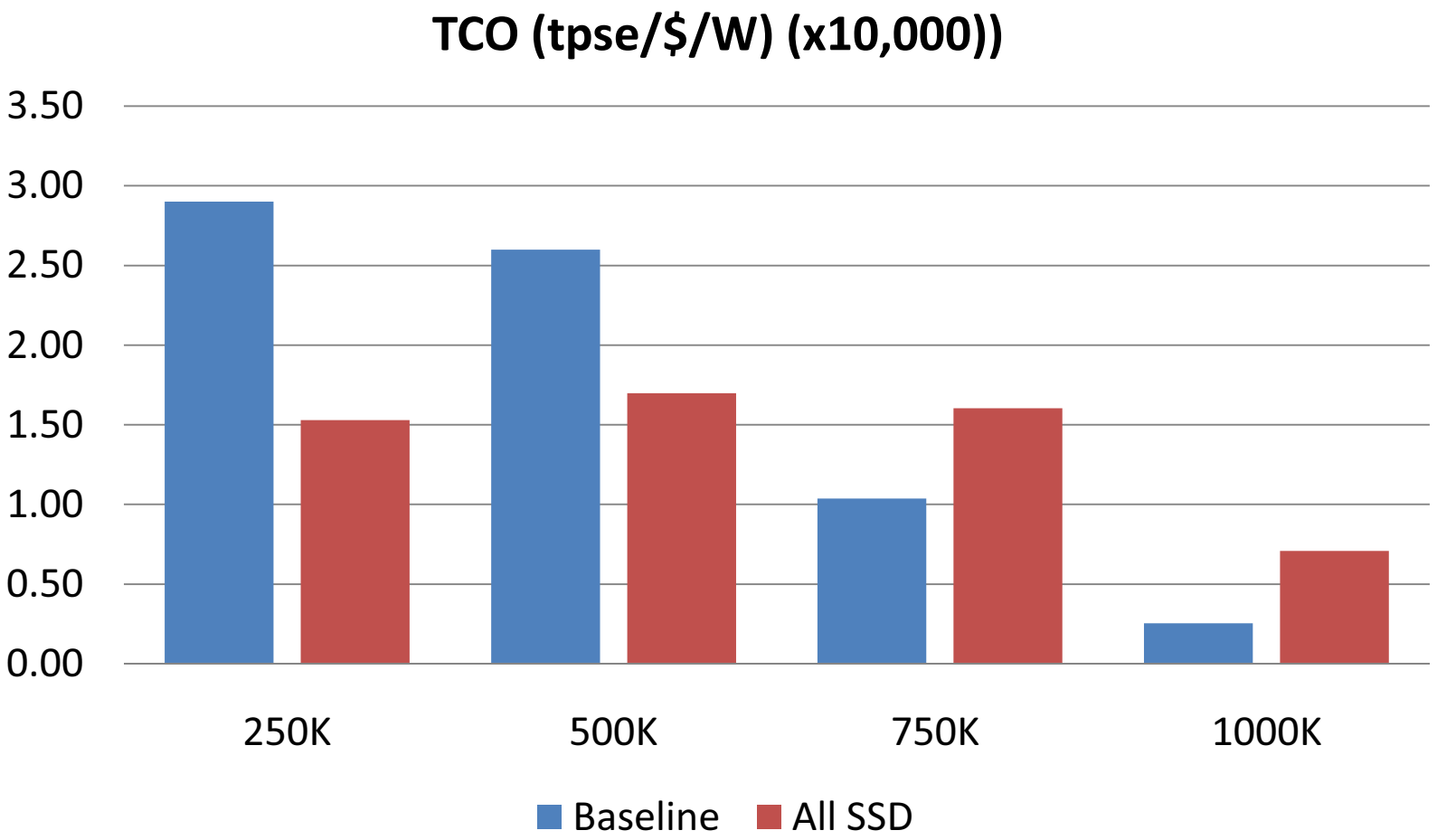
- a modified TPCE:
  - 1 Million customer
  - 600GB database
  - We simulate: 100K, 250K, 500K, 750K, 1M
    - Different IO density

Customer count	250K	500K	750K	1000K
IO density	2.27	7.31	16.30	25.74

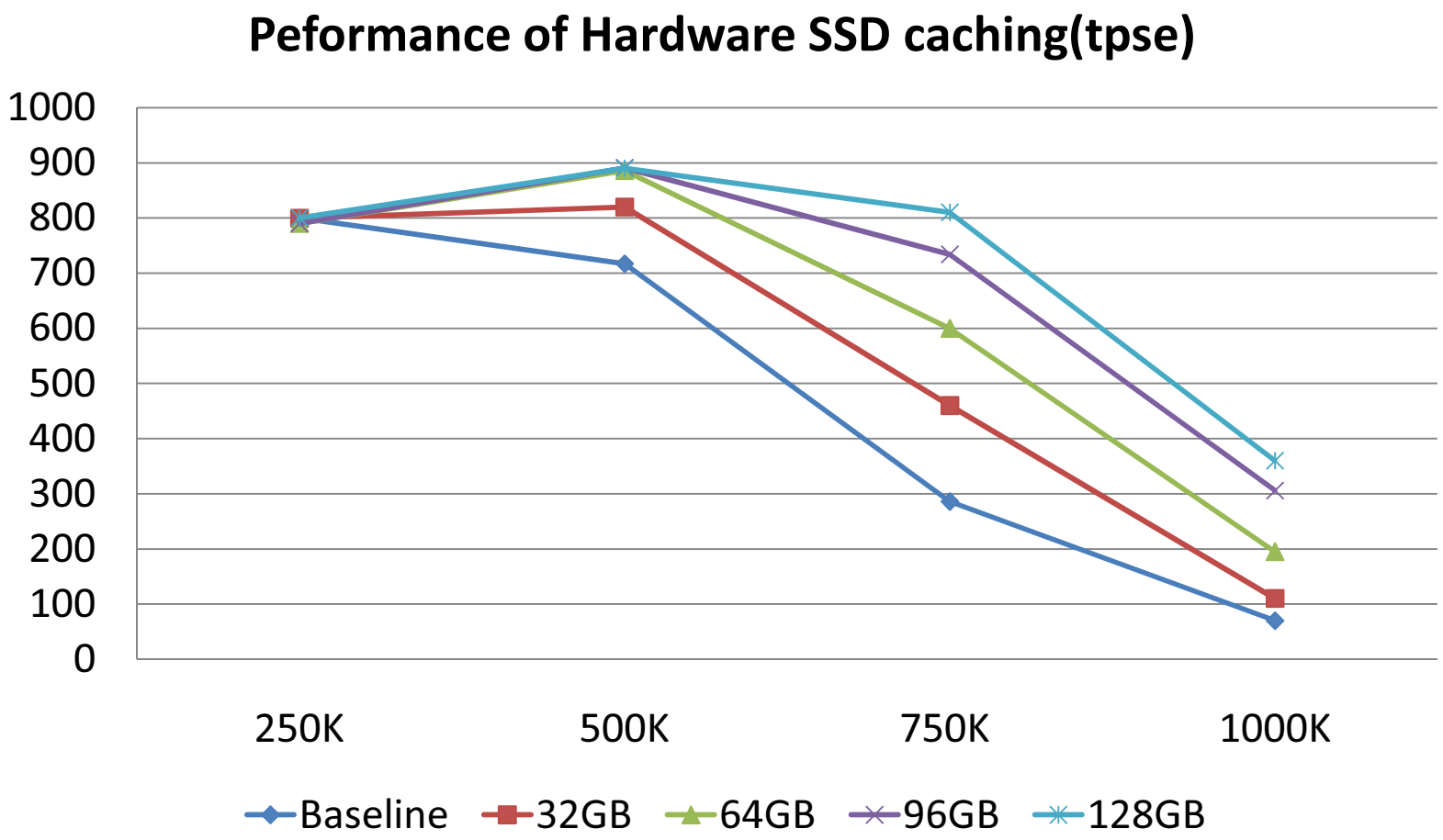
Results: HDD vs. All SSD



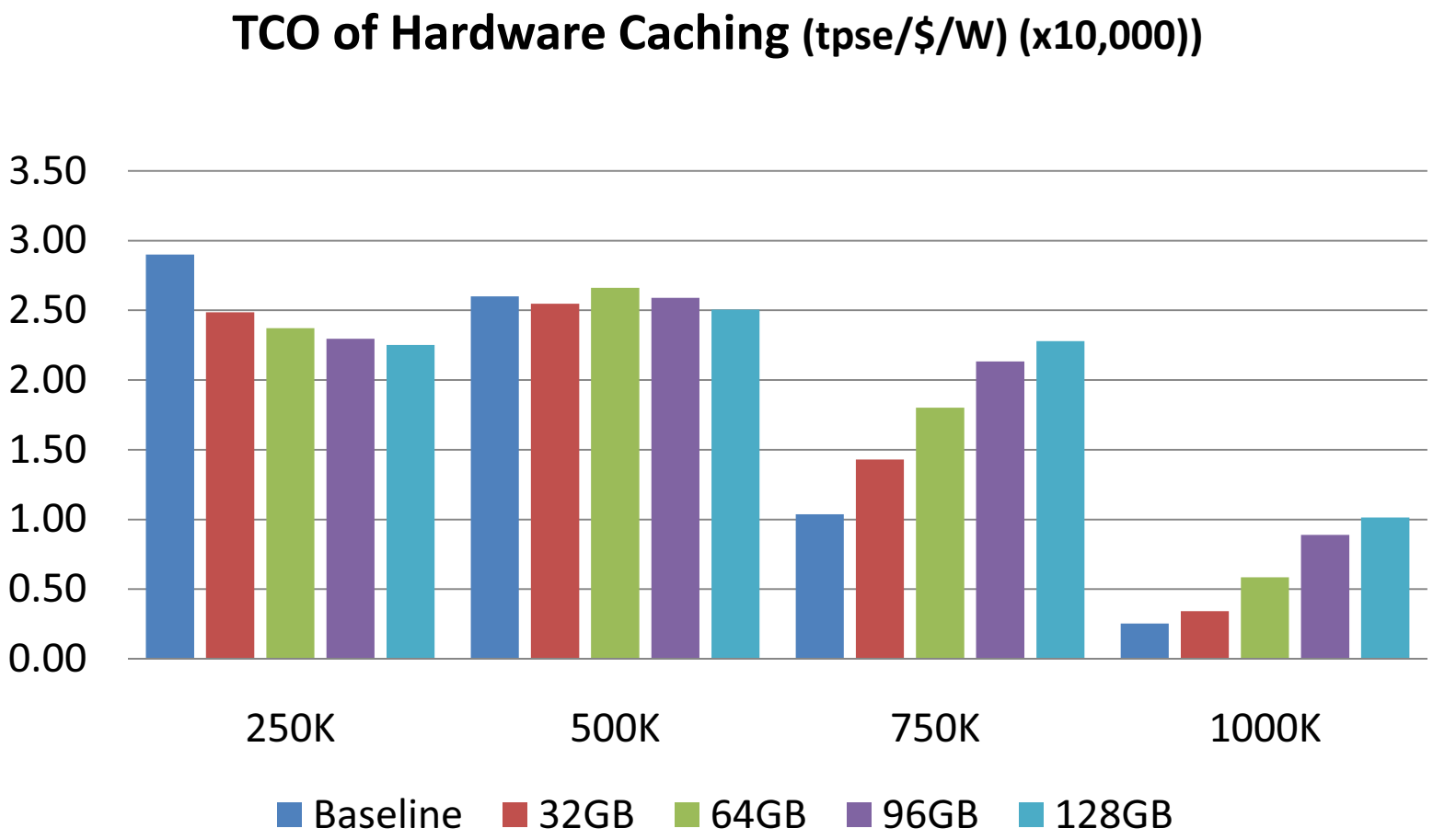
Results: HDD vs. All SSD (cont.)



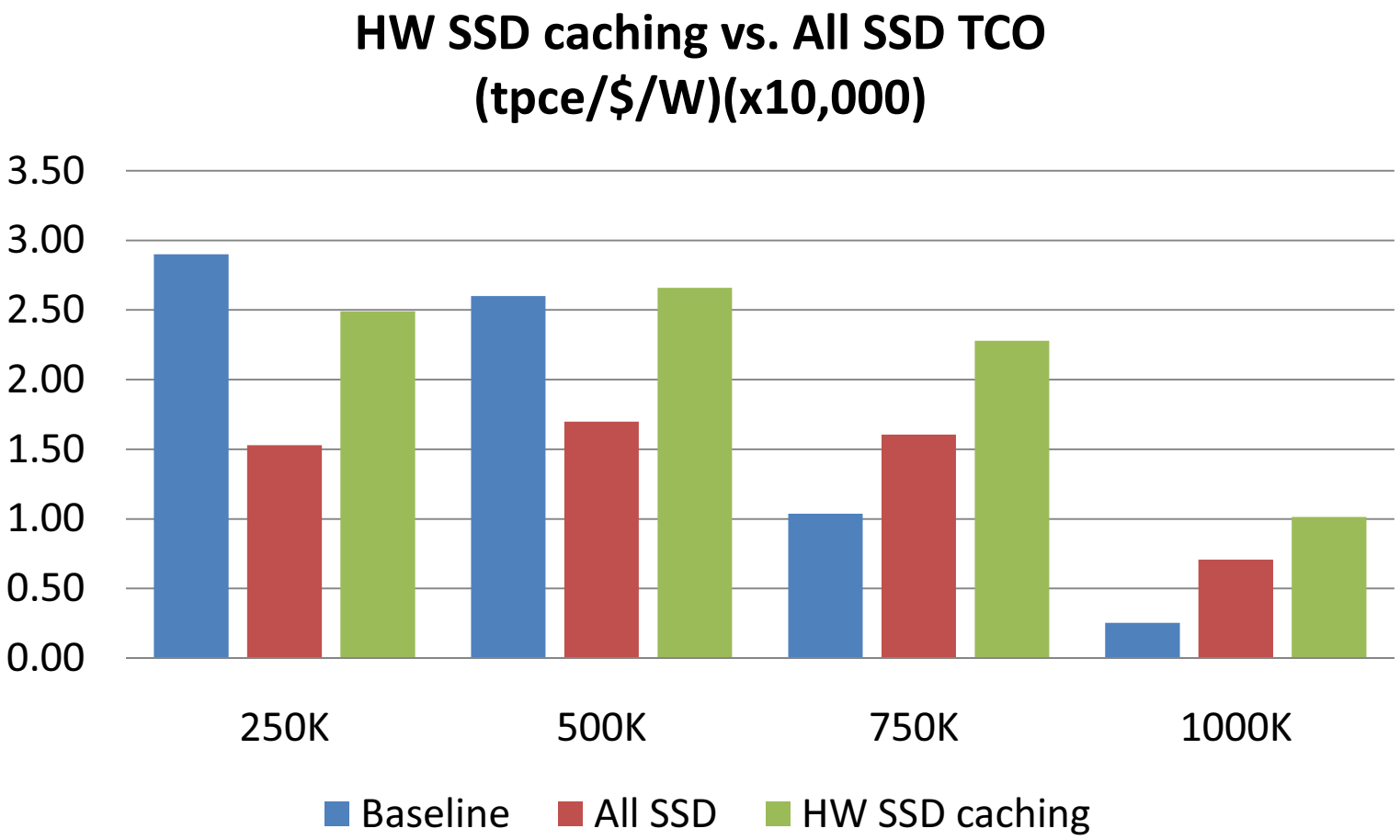
Results: Hardware SSD caching



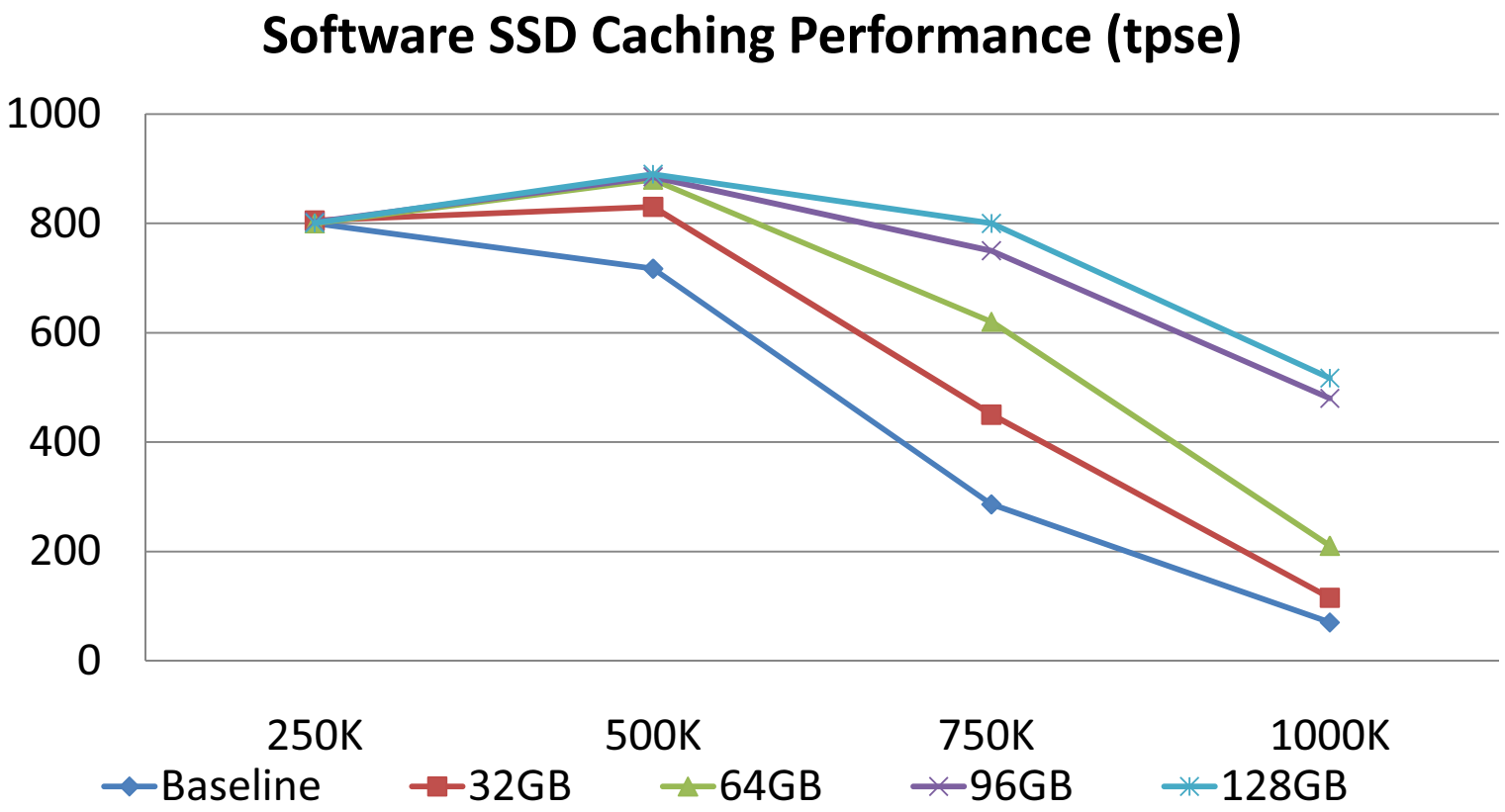
Results: HW SSD caching (cont.)



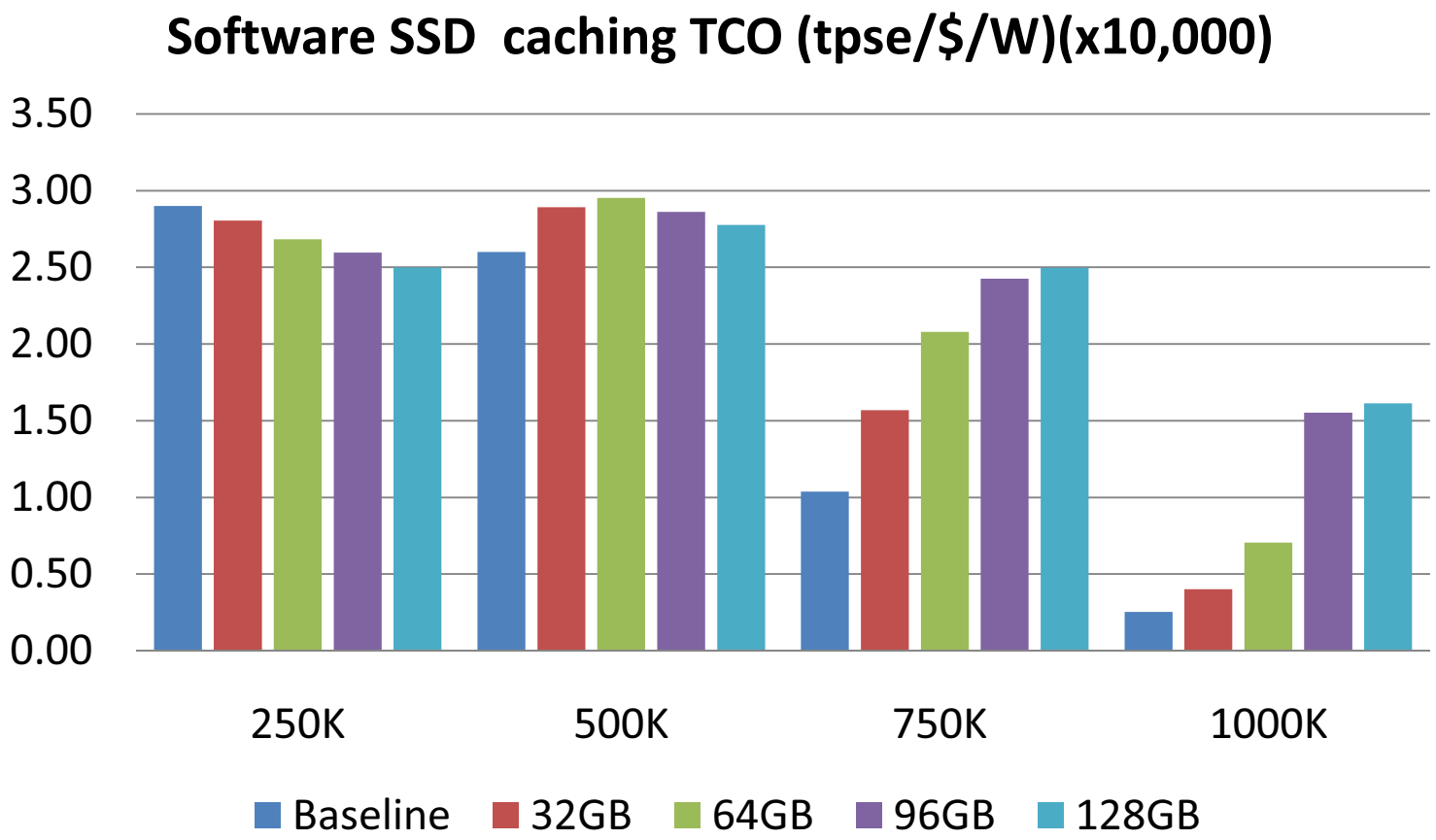
Results: HW SSD caching (cont.)



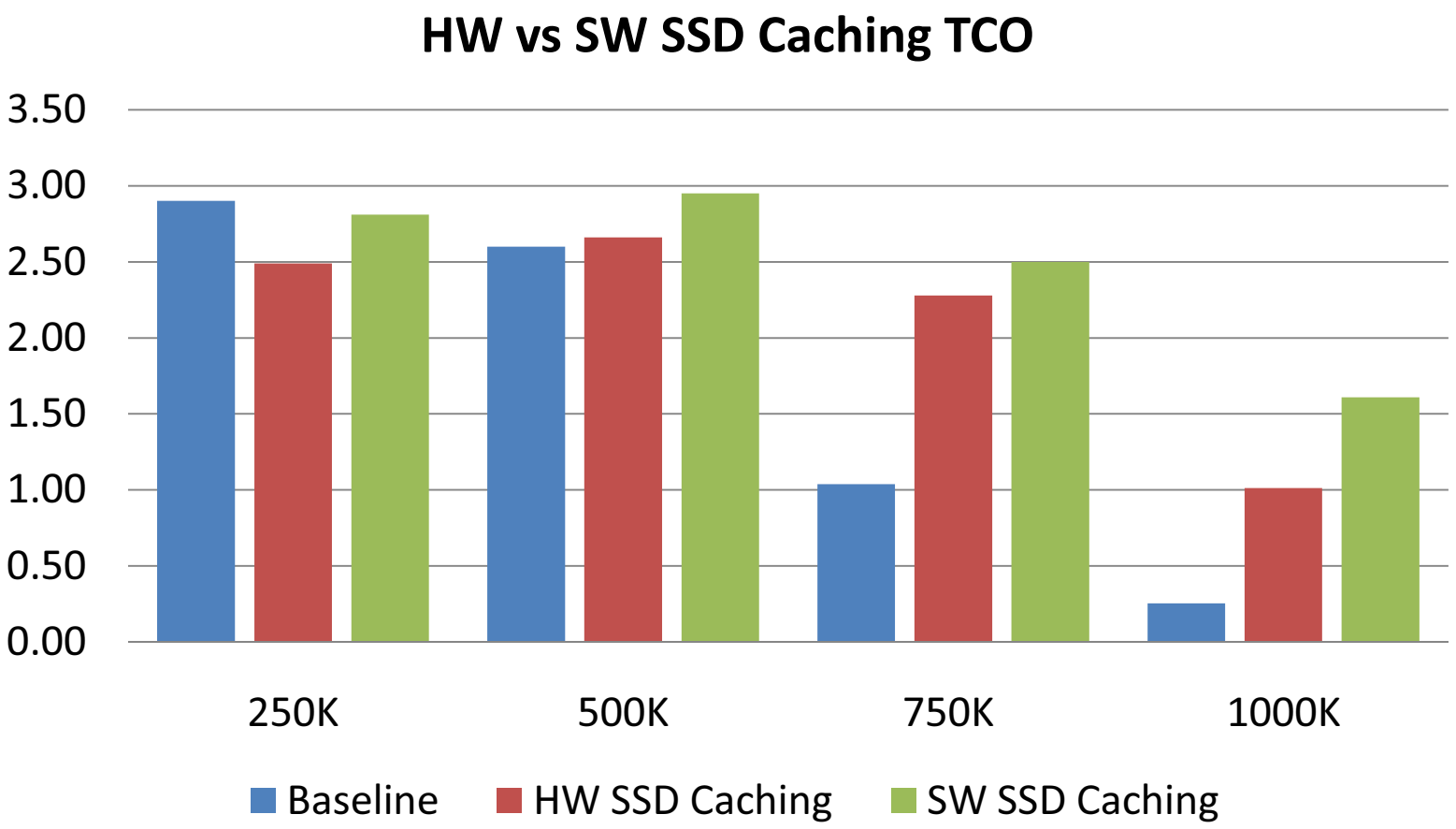
Results: SW SSD caching



Results: SW SSD caching (cont.)



Results: SW SSD caching (cont.)



Conclusion

- SSD technology is maturing
- Will continue to be expensive
- SSD caching is the best approach and could be used as a substitute for the Total SSD solution
- Hardware Solutions are available immediately
- Software solutions to follow