SSD limits in enterprise storage

Executive Summary

Solid state drives or SSDs have taken the storage market by storm. But while they are wonderful for notebooks and desktops they have serious performance problems in enterprise storage - despite their high cost.

In this brief you will learn – in non-technical language – why SSDs can't give you all the performance you pay for.

The SSD problem

SSDs use a form of semiconductor memory - NAND flash (or just flash) - that is very different from disk drives. Key differences:
1) Reading flash is much faster than disk drives, but writing takes much longer - 10-15x is common.
2) SSDs can only write data in large blocks - 256 thousand bytes or more - which means a lot of overhead to maintain a supply of fresh blocks.

One result is that many flash performance claims are based solely on reads. Another result is that the overhead required to maintain a supply of fresh blocks - for quick handling of writes - is ignored.

When a block has too much invalid data, due, say, to updates, good data has to be copied to another block along with other fresh data. Old blocks must be erased - a lengthy process - before re-use.

Workloads with many writes - or workloads with bursty write traffic - can use up all the free blocks. Then any read requests must wait until write requests complete to avoid retrieving stale or corrupt data. That means all traffic to and from the SSD must stop. Not a problem for notebook users, but for enterprise storage this fact means continual and uncontrollable slowdowns.

Two important metrics

Data storage is usually measured two ways: the number of reads and writes per second (known as IOPS or Input/Outputs Per Second); and, how long it takes for those reads and writes to complete (known as latency). Heroic IOPS numbers are now the norm in storage marketing, but now that all flash arrays offer more IOPS than most applications need, latency performance is more important.

The SSD performance effect

SSDs have dramatically improved IOPS. Small storage boxes are now commonly capable of 500,000 IOPS - a number that used to take a half dozen full racks of hot, failure prone, and expensive equipment.

Latency has not improved nearly as much, with average access times dropping from approximately 6

milliseconds (ms) to less than 1ms. According to TPC-C benchmarks - audited performance tests simulating common business database workloads - maximum access latency is often 20 times or more the average latency.

Predictable latency is important because slow accesses are bad for response times and system uptime. These "long-tail" latencies are typically a direct result of the use of SSDs in enterprise storage.

Mixed workload performance

In contrast, Violin Memory all flash arrays - that do not use SSDs - in recent independent performance tests that emulated real-world workloads - mixed block sizes, multiple workloads, varying read/write ratios - achieved ". . . latencies in the low hundreds of microseconds at the storage system while all four workloads were running."

SSDs and Big Data

Researchers at SanDisk Corporation, a major SSD producer, found that SSDs are not optimal for Big Data applications either. Big Data software, such as Hadoop, commonly use a technique called log-structured I/O. This means that new data is added to the end of the storage pool, rather than updating invalid data in place, much as a captain writes new data to the last page of the ship's log.

This is a problem, according to the researchers, because SSDs already use log-structured I/O. When an application does too, there are multiple negative impacts, including:

". . . write pressure to flash devices through randomization of workloads, unaligned segment sizes, and uncoordinated multi-log garbage collection. All of these effects can combine to negate the intended positive affects of using a log."(1)

Conclusion

For over 50 years computer system I/O was optimized for disk drives. When flash drives arrived, the easiest path to market was to make them interface to computers as if they were much faster disks.

However, that injected another layer of hardware and software between applications and their data storage, leading to undesirable performance impacts. These include unpredictable latency, extra software complexity, and sub-optimal performance in Big Data applications.

Systems that directly access flash, down to the chip level, can avoid these negative effects because nothing is hidden from them. Systems from Violin Memory embody this design philosophy and their performance validates the approach.

In 10 years few, if any, flash arrays will use SSDs. The question for today's buyer is: why buy SSD-based arrays today, when better designs are available?

Footnote:
(1)*Don't stack your Log on my Log* research paper by Jingpei Yang, Ned Plasson, Greg Gillis, Nisha Talagala, and Swaminathan Sundararaman,SanDisk Corporation. Presented at the INFLOW '14 conference. Accessed in December, 2015 at URL https://www.usenix.org/system/files/conference/inflow14/inflow14-yang.pdf

About the Author

Robin Harris is the president and chief analyst of research and advisory firm TechnoQWAN LLC, founded in 2005. He has over 30 years experience in the IT industry in product management, strategic planning, and emerging technologies in leading companies, including Digital Equipment Corporation (DEC) and Sun Microsystems. He earned degrees from Yale University and and Wharton School of the University of Pennsylvania, lives in northern Arizona, and may be reached at robin [at] StorageMojo.com.