#snsarticle http://snseurope.info/n/myzp

NV-DIMMs: The next critical technology for your data center storage strategy

As the world experiences an ever increasing numbers of devices connected to the internet with 15 Billion devices expected to be connected by 2015. Data centers serving these devices have either been creaking under the pressure or have adopted new technologies to help increase efficiency & performance.
By Adrian Proctor, VP, Marketing, Viking Technology.

Solid State Disks (SSD) have been driving part of this technology upgrade, delivering far increased I/O performance and efficiency when compared to HDDs. However, these SSDs derive their fundamental hardware structures and software characteristics from the much slower hard disk drive (HDD), including IO bus interfaces and file system schemes. Because of this evolutionary history, flash based SSDs have not yet attained their maximum potential value in the system / data center architecture.

Due to the incessant drive to lower cost, NAND flash vendors and SSDs are sacrificing endurance – SSD vendors and flash controller developers have to scramble to design better wear leveling and error correction algorithms at the introduction of every new NAND process geometry shrink.

Similar to HDDs, where increases in performance were sacrificed for cost and capacity; NAND flash is sacrificing reliability for cost. Table 1 illustrates the extremely low SSD endurance that would be expected when using the drive for anything other than reads. (Note: this SSD illustration assumes MLC NAND flash).

A new device category, called Non Volatile (NV) DIMMs can now be integrated into standard Intel x86 based servers to act as the write cache for any storage hierarchy. These DRAM based NV-DIMMs, have infinite write endurance and deliver a huge performance boost to the system.

Also utilizing NV-DIMMs allows the flash SSDs to be used primarily for read operations which means that the SSDs will have a far longer service life. The net result is that data centers gain significantly better ROI on SSD adoption when integrated and used in conjunction with NV-DIMMs.

Introducing ArxCis-NV™, a non volatile DIMM (NV-DIMM)

Viking Technology, a pioneer in Non-Volatile memory, DRAM and SSD technologies, takes a different approach to the integration of SSDs into data centers. Flash based SSDs have the potential to radically accelerate application performance, but at the same time, users must be fully cognisant of the major shortcomings such as - endurance, bottlenecking and high availability. Viking's ArxCis-NV, a Non-Volatile DIMM, uses a trusted paradigm, access to DRAM in main memory, to provide the fastest possible I/O performance, practically infinite write endurance (the Achilles heel of flash) and all this delivered in a solution that provides increased levels of data security and high availability.

A Non-Volatile DIMM, is a module that can be integrated into the main memory of an industry standard compute platform (i.e. server), perform workloads at DRAM speeds, yet be persistent & provide data retention in the event of a power failure or system crash. Viking's NV-DIMM, is a hybrid memory subsystem that combines the speed and endurance of DRAM, together with the non-volatile data retention properties of NAND flash. This marriage of DRAM and NAND technology delivers a high speed and low latency "non-volatile & persistent" memory module. Designed from the ground up to support unlimited write activity, it performs at fast DDR3 speeds (12,000 MB per second) that could wear out many SSDs in days or even hours. This solution can be viewed as the first commercially viable "Storage Class Memory" for the latest Intel based x86 servers.

When integrated synergistically with SSDs, NV-DIMMs allow applications to use much faster techniques for protecting data. For example, a Memcache scratchpad database typically becomes persistent by conversion to MemcacheDB, with frequent update writes to HDD or SSD storage. Now, with NV-DIMMs the application has two notably better options. Either, the MemcacheDB instance can store data in a persistent NV-DIMM RAMDisk, or, with some application tweaks, Memcache can run with the database stored in the NV-DIMM and use direct memory operations for updates. Both approaches offer incredible performance increases when compared to just integrating SSDs.

To better understand the benefits of NV-DIMMs, let's look at the issues with current Flash/SSD approaches and how NV-DIMMs solve them:

SSD's – performance problems with the legacy HDD model for SSDs

When Flash technology arrived on the scene, its characteristics of block-mode write operations and its speed were similar to HDD technology. Using the SATA drive interface and a block model allowed existing software structures to be used, and in effect SSDs became fast HDDs.

Flash speeds have improved as a result of parallel access techniques and better SSD controller chip design. As a result, the industry is now facing a set of performance bottlenecks. These include:

- File systems designed for millisecond access
- As many as 5 time consuming address translations from the application to the media
- High interrupt volume from I/O due to the much greater rate of IOPS
- Interface contention in SAS/SATA and PCIe

Additionally, transfer block sizes have expanded from a typical 1024 byte block to 4096 bytes, compounding the performance stress from random IO. Applications further strain the bottlenecks by an increased emphasis on scratchpad files for key data, which typically are updated very frequently with a few words at a time. These operations involve a multiple access read-modify-write operation.

SSD latency is currently around or above 50 microseconds today depending on the interface and is mostly driven by NAND flash device limitations. This limits IO operations to wait from when the IO command is started by the application, to disconnecting and going to sleep, to when the IO finally completes. This was a good model for HDD, but dramatically increases the amount of overhead with the high IOPS rate of SSD devices due to frequently swapping state. On top of this, most IO follows an indirect, buffered model, so the CPU expends additional precious cycles copying data between buffers and the application space.

NV-DIMMs are dramatically faster than any SSD

It is clear that the rest of the infrastructure must do better in order to use all of the rapidly growing CPU horsepower that we are paying for.

Block Mode ArxCis-NV™ integration

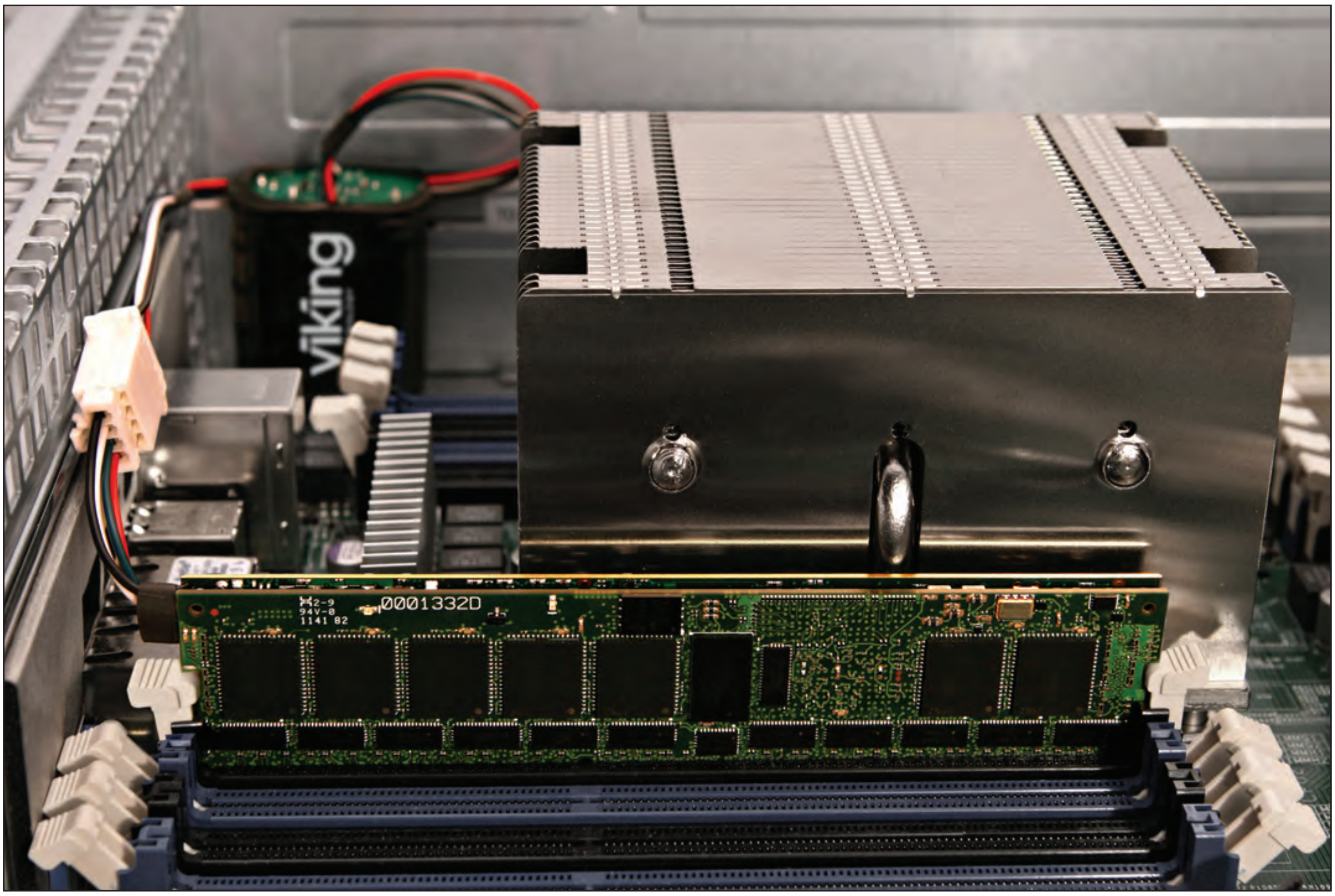
Using a Persistent RAMDisk driver, a very fast RAMDisk can be built in an NV-DIMM. This operates like a standard RAMDisk, except that data will be safe in the event of a power or system fail.

Since the ArxCis-NV™ operates as in-system DDR3 DRAM, it is blindingly fast in operation, compared even to the very best PCIe SSD solutions. But, there's more! The NV-DIMM supports single byte or word operations, meaning that applications no longer have to do tedious read-modify-write operations, nor do they have to make long transfers for minimal updates. The result is dramatic performance. Even in block mode, the standard C/C++ write operation #write(file ID, offset, transfer length) can be implemented with a transfer length to media of as small as 1 byte, as opposed to the 1024 or 4096 bytes required by all SSDs.

Another approach that further impacts the application (in a good way) is to switch to direct IO. This can be done in a (Persistent)

	SSD 6Gb SATA	PCIe SSD	ArxCis (Block Mode)	ArxCis (NVRAM Mode)	ArxCis Wins
Write Bandwidth (Seq 128K)	260 MB/s	1,300 MB/s	2,000 MB/s	4,000 MB/s	YES
Write IOPS (Ran 4K)	50K	105K	2 million	10 million	YES
WR IO Latency (microseconds)	100	15	0.5	0.025	YES
Read Bandwidth (Seq 128K)	360 MB/s	1400 MB/s	2,000 MB/s	4,000 MB/s	YES
Read IOPS (Ran 4K)	60K	140K	2 million	10 million	YES
RD IO Latency (Microseconds)	100	47	0.5	0.025	YES
Drive Size	512GB	640GB	8GB per DIMM	8GB per DIMM	
Memory Type	MLC	MLC	DRAM	DRAM	
Endurance - Random 4KB block	(75 % duty-cycle)				
99% Read - 1% Write	4.1 Years	6.1 Years	10's of years	10's of years	YES
95% Read - 5% Write	0.8 Years	< 1.2 Years	10's of years	10's of years	YES
90% Read - 10% Write	0.4 Years	0.6 Years	10's of years	10's of years	YES

Table 1



RAMDIsk software module. The copying of data between buffers and application memory is no longer needed, and the app runs faster.

Putting this in context for applications, let's compare the MemcacheDB application with SSD and with NV-DIMMs. Assume that a transaction ID is updated for each transaction. When using an SSD, that would require a read-modify-write operation on the block containing the ID field. This severely limits transaction rate, since one drive is being addressed for all the ID activity and it has other work to do.

When utilizing NV-DIMMs, we can do a single 8 byte transfer through a very low overhead Persistent RAMDisk to achieve the update. The operation occurs at DRAM speeds (12GBytes/second), and the result is orders of magnitude faster than SSD or PCIe solutions.

Flash Wears Out !

NAND flash is fallible, it wears out. The continual industry pressure to deliver higher capacity and more economical flash means that the endurance (number of times a block can be written to) and data retention is sacrificed. At the raw flash level over the past few years we have seen NAND degrade from 100,000 to 3,000 program/erase (P/E) cycles.

The latest SSDs attempt to counteract this NAND flash endurance issue in a number of ways:

- **Wear-leveling:** spreading write operations evenly across all the available flash
- **Over-provisioning:** by reducing the usable capacity, it permits wear to be spread over many more cells
- **Advanced ECC:** error correction allows devices with more bits that have gone bad to still be used

Even with these fixes, commercially available SSD and PCIe card life expectancy under typical database workloads can be less than a year, and is very dependent upon the percentage of reads to writes (see Table 1.).

ArxCis NVDIMM solves the SSD endurance problem

ArxCis-NV™ resolves the SSD endurance problem in a simple but compelling way. Instead of complex firmware and physics, NV-DIMMs reduce the number of write cycles in a given period. NV-DIMMs simply remove the need to write to the SSD.

The result is cost saving from using less storage for wear protection (over-provisioning), no operational cycles spent on wear balancing and scrubbing, no "write cliff" and no need for the sysadmin to monitor usage, but most of all the SSDs in the system will remain in service life for many more years than those without NV-DIMMs.

Conclusion

NV-DIMMs will have a major impact on storage caching/tiering and how applications are tuned. The basic limitations of flash endurance and block mode IO have prevented the full potential of flash SSDs being realized. ArxCis-NV™ overcomes this with a use paradigm that is easily understood, delivering dramatic improvements in performance and ROI.