NVDIMM - Fastest Tier in Your Storage Strategy

Contributor:

Viking Technology

January 2014

About the SNIA

The Storage Networking Industry Association (SNIA) is a not–for–profit global organization, made up of some 400 member companies spanning virtually the entire storage industry. SNIA's mission is to lead the storage industry worldwide in developing and promoting standards, technologies, and educational services to empower organizations in the management of information. To this end, the SNIA is uniquely committed to delivering standards, education, and services that will propel open storage networking solutions into the broader market. For additional information, visit the SNIA web site at http://www.snia.org or http://www.snia.org/international for a list of International Affiliates

About the SNIA Solid State Storage Initiative

The SNIA Solid State Storage Initiative (SSSI) fosters *the growth and success of the market for solid state storage in both enterprise and client environments*. Members of the SSSI work together to promote the development of technical standards and tools, educate the IT communities about solid state storage, perform market outreach that highlights the virtues of solid state storage, and collaborate with other industry associations on solid state storage technical work. SSSI member companies represent a variety of segments in the IT industry (see http://www.snia.org/forums/sssi/about/members). For more information on SNIA's Solid State Storage activities, visitwww.snia.org/forums/sssi and get involved in the conversation at http://twitter.com/SNIASolidState.

About Viking Technology

Viking Technology is recognized as a leader in NVDIMM technology. Supporting a broad range of memory solutions that bridge DRAM and SSD, Viking delivers solutions to OEMs in the enterprise, high-performance computing, industrial and the telecommunications markets. Viking Technology is a division of Sanmina Corporation (Nasdaq: SANM), a leading Electronics Manufacturing Services (EMS) provider. More information is available at http://www.vikingtechnology.com.

Table of Contents

Executive Summary

This paper reviews how Non-Volatile Memory can be utilized within current storage strategies to maximize performance and availability. Additionally, the paper will discuss how Non-Volatile Dual In-line Memory Modules (NVDIMMs) can be enabled as the fastest tier and the impact it has on your storage strategy.

What Is Non-Volatile DIMM?

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 (i.e. DDR3), yet be persistent and provide data retention in the event of a power failure or system crash. By understanding that flash based SSDs have the potential to radically accelerate application performance, but at the same time, being cognisant of some major shortcomings (endurance, performance and high availability) that present themselves when integrating SSDs into data centers. Non-Volatile DIMM, uses a trusted paradigm, DRAM access 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.

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 DDR3 speeds (12GBytes/second) and can sustain itself from host power failure or a system crash. This solution can be viewed as the first commercially viable “Storage Class Memory” for the latest Intel based x86 servers.

Figure 1. ArxCis-NV™ NVDIMM from Viking Technology

NVDIMM - Fastest Tier in Your Storage Strategy

NVDIMMs, integrated synergistically with SSDs, allow applications to use much faster techniques for protecting data. For example, a Memcache scratchpad database is usually converted to MemcacheDB in order to be persistent, with frequent update writes to HDD or SSD (i.e significantly slower than Memcache in DRAM). Now, with NVDIMMs the application has two higher performance options. Either, the MemcacheDB instance can store data in a persistent NVDIMM RAMDisk (Block addressable like an SSD/HDD), or Memcache can run with the database stored in the NVDIMM and use direct memory operations for updates (Byte addressable). Each approach offers incredible performance increases over simple SSD integration, and are discussed again later in this paper.

Below are comparison charts for review:

Performance Comparison - NVDIMM and SSD Technologies

Workloads	SSD (6Gb SATA)	PCIe SSD	NVDIMM (illustrated by ArxCis-NV) Block Mode	NVDIMM (illustrated by ArxCis-NV) NVRAM Mode
Write Bandwidth (Sequential 128K)	260 MB/s	1,200 MB/s	2,000 MB/s	4,000 MB/s
Read Bandwidth (Sequential 128K)	360 MB/s	1,400 MB/s	2,000 MB/s	4,000 MB/s
Write IOPS (Random 4K)	50K	105K	2 Million	10 Million
Read IOPS (Random 4K)	60K	140K	2 MIllion	10 Million
Write IO Latency (microseconds)	100	15	0.5	0.025
Read IO Latency (microseconds)	100	47	0.5	0.025
Memory Type	MLC	MLC	DRAM	DRAM
Drive size	512GB	640GB	85GB per DIMM	8GB per DIMM

Endurance Comparison - NVDIMM and SSD Technologies

	SSD (6Gb SATA)	PCIe SSD	NVDIMM (illustrated by ArxCis-NV) Block Mode	NVDIMM (illustrated by ArxCis-NV) NVRAM Mode
99% Read, 1% Write	4.1 Years	6.1 Years	10+ Years	10+ Years
95% Read - 5% Write	0.8 Years	<1.2 Years	10+ Years	10+ Years
90% Read - 10% Write	0.4 Years	0.6 Years	10+ Years	10+ Years

Applications

NVDIMMs can be used to speed up a number of Enterprise applications. There are a variety of usage models, with differing levels of complexity. This section is intended to sample some models that have considerable impact for industry. Individual vendors can provide more details on approaches in applications that are not described below. .

This paper addresses three groups of applications:

1) Little or no change to code will provide a major improvement
2) Higher yielding solutions involving code changes
3) Leading edge innovation with major impact, allowing completely new methods as well as fast performance.

Little or No Code Changes Providing Major Impact

Most sysadmins are familiar with RAMDisk as a way to speed up applications. Typically, data that is accessed often is stored in a RAMDisk (DRAM), but their use is limited to non-critical data due to the volatile nature of DRAM storage. A type of RAMDisk, usually labeled a “PRAMDisk” (P stands for Persistent), is able to safely retain the data held in DRAM should a power failure or system crash occur (thus being a Non-Volatile DIMM).

Once you have a PRAMDisk, it can be mounted to act as a drive and attached to any application as needed. Some examples of its use are:

1. **As the new “Fastest Tier in a storage hierarchy”.** The tiered storage manager (caching/ tiering software) handles loading the ‘hot’ data into the PRAMDisk, and returning it to slower SSD/PCIe/HDD storage tiers as necessary. Additionally, the NVDIMM has practically infinite write endurance, so will prolong the life of all other storage tiers (SSDs) that would otherwise be used for this write activity.

2. **As an alternative to SSD/HDD storage** for volatile files. This seems trivial, but many applications need a place to keep tally data, counters and other frequently updated information. Currently, persistence of this data requires writes to disk or to database storage, both of which effectively limit throughput to low transaction counts. A PRAMDisk is the ideal solution for speeding up this type of operation. With DRAM speed, these operations will no longer be a bottleneck.

3. **Boot image storage:** Most of us are unhappy with the time it takes to reboot a system. PRAMDisk addresses this issue. In its simplest form, creating a PRAMDisk allows you to store all of the app layer data, including current stored changes to files, in an “instant-on” space. When boot occurs, the kernel is loaded from another storage device, and finds the PRAMDisk ready and loaded on completing boot.

   A slightly more complex implementation involves loading a compatible PRAMDisk with the intermediate boot loader. This allows the kernel image to be kept in the PRAMDisk too, speeding up boot further. This approach is important in mission-critical environments such as financial, OLTP, and military and medical applications, where downtime must be minimized and high availability is critical.

More complex methods of improving boot are discussed further on in this article, including implementation of a persistent standby operation.

4. **To speed up a wide variety of image editing systems**, including medical imaging, surveillance scanning, movie rendering and editing CAD, etc. The essential benefit in all these cases is immediacy. Having to wait for file operations when processing an image is distracting, both directly and by breaking concentration, adds substantial amounts of expensive employee time to a task.

   Essentially, this effort requires scripting and working mount points. Take, for example, video editing, where a pool of technicians work around a central NAS file containing the entire work product (movie), and operate by 'borrowing' segments of the movie to their workstation disk. Replacing that disk with an NVDIMM/PRAMDisk will speed things up considerably thereby delivering increased efficiencies.

5. **To hold MemcacheDB files.** These small databases are being increasingly used to store vpointers and other volatile data, as well as metadata tables in advanced storage systems. The distributed replication model of MemcacheDB fits well with the PRAMDisk approach, and the performance speed up can be substantial.

Higher Yielding Solutions: More Sophisticated Implementations

By making some code changes to the installation it will open up new methods and further performance boosts. The spectrum of changes ranges from additional OS modules, creating word mode accesses to the NVDIMM, BIOS alterations that support backing up a complete system state. This section reviews those that an advanced user should be able to evaluate.

1. **Performance:** As an example, Viking Technology has identified a set of OS changes in LINUX that will substantially improve PRAMDisk performance. This involves some changes to the kernel build, adding new modules. The resulting performance boost is substantial in most applications.

2. **Time is Money**: Databases such as Oracle require considerable analysis to tune performance. A key fundamental is minimizing the time cost of an assured write operation. With distributed architectures, using a PRAMDisk is an alternative for the most frequently accessed files. With local writes and reads to a PRAMDisk such as Viking Technology's ArxCis-NV™ (in block mode) taking less than a microsecond, compared with the 50 microseconds or more to the fastest PCIe SSDs, it's clear the time cost is dramatically reduced.

3. **Performance**: Adding RAID/replication type protection to NVDIMM storage involves a spectrum of approaches. At the user-capability level, some operating environments or file systems support asymmetrical mirrors and preferred read operations. Here, a local NVDIMM copy is written, and a mirror copy to a remote storage device is also created. If the latter is also a system with NVDIMM storage, the time cost of the write is relatively small. However, on reading the data at a later time, the preferred read system goes to the local NVDIMM and gives by far the fastest access.

Leading Edge Innovation With Major Impact

These advanced techniques would typically involve OS or BIOS changes. Examples include:

1. **CPU Cost Savings**: Creating a word-mode access system to NVDIMM. Since NVDIMM behaves like DRAM, it is possible to update individual bytes or words. To do this, the application must contain variable constructs that reflect permanent storage, and be written to address those variables directly rather than through block I/O. The payback is substantial, allowing a lower cost CPU to reach the same performance levels as the latest high-powered CPUs.

2. **Recover from power-fail almost immediately**: BIOS changes, and simple hardware enhancements, will allow the implementation of “hibernate-in-place”, giving the host an ability to power down a system completely, but return to operation as though from standby. This would allow a system to recover from a power fail in the same way, like returning from a standby-like state.

3. **Fastest Boot**: In the same vein, BIOS changes will allow the system to boot directly to a PRAMDisk image.

4. **VM’s**: Having a single persistent image is a strong value proposition for Virtual Machine environments. Proprietary hypervisor code would need to be generated to handle the image, either as a PRAMDisk device, or more cost effectively, as a read-mostly, persistent shareable segment of DRAM. This is likely to be handled by the software vendors.

Summary

NVDIMMs as illustrated by Viking Technology's ArxCis-NV™ provide a means to speed up many areas of application and system operation. The items above are only a short list of examples, and we expect that users will find more and more innovative ways to use NVDIMMs to realize increased levels of performance, value and high availability.