Special Report

PCIe Solid-State Storage: What To Know Before Getting Started

Before you make any investments in PCI Express (PCIe) solid-state storage, take a few minutes to explore this Special Report, which highlights the key requirements you need to consider when deploying the technology.

Does PCIe SSD Live Up to The Hype?

Dennis Martin, independent SSD expert and founder of Demartek, discusses PCIe SSD in this Expert Answer.

PCIe SSD seems to be all the rage right now. Is that the best way to use solid-state or is it just a lot of hype?

It is certainly a good way to do it. With PCIe SSDs, you have a lot more bandwidth available to you, but the disadvantage is you are located in the server, so there are some limitations there as far as number of slots and how well you can share, which is a little bit different than in a SAN or that sort of architecture. Certainly, for bandwidth and low latency, it is a great thing to do.

There is also SCSI Express, which is SCSI, but on a PCIe bus. You are starting to see a lot of activity move toward PCIe solid-state in some fashion.

Is PCIe SSD Right for You? Deploying PCI Express SSD Devices

Learn about the PCI Express (PCIe) solid-state drive (SSD) market, how and where to deploy PCIe SSD in your environment, and challenges with the technology.

Solid-state storage based on NAND flash memory chips has drawn much attention in the last year, from traditional enterprise data storage vendors to less familiar names and newcomers. The message from storage vendors has been to deploy solid-state storage alongside disks in their arrays, with specialized software handling the migration of data to and from this high-performance tier. But newcomers and system vendors offer an alternative approach: solid-state storage deployed as a PCI Express card within the server itself. The PCIe approach eliminates the storage network entirely in certain situations. This tech tip focuses on the reality of the PCIe SSD market and where these devices should be deployed in today's systems.

Contrasting PCIe SSD and networked storage

Enterprise storage has slowly evolved from internal disks to direct-attached (DAS) RAID to networked arrays (SAN and NAS). Each step has maintained backward-compatibility with what went before, allowing applications to be deployed on SAN or NAS without major changes from DAS.

The technologies employed for block storage -- Fibre Channel (FC) and iSCSI -- rely on the SCSI protocol and drivers just like internal disk drives do, but with different results. Modern storage systems might use Ethernet adapters and switches, and can communicate with devices that are distant in terms of geography. Most storage systems are also virtualized, disguising their complex arrangement of caches and mobility. All of this effort goes to balance flexibility and performance, but it places an upper limit on storage performance.

PCI-based storage is entirely different. Rather than masquerading flash or DRAM memory as a SCSI-connected hard disk drive, PCI Express SSD products often use specialized drivers to communicate using direct memory access (DMA) over the PCI bus. This is game-changing in terms of I/O latency, enabling random read and write performance that's orders of magnitude faster than the quickest storage array. Although throughput is also improved thanks to the bandwidth of the PCI Express bus and memory, the expense of solid-state chips limits the amount of capacity that can be deployed.

Where to deploy PCIe SSD today

Enterprise systems architects face a wide variety of challenges, with each application or component placing unique demands on data storage subsystems. Some require massive storage capacity while others must constantly move vast amounts of data. Neither of these is appropriate for PCIe SSD at this point because of the high per-GB cost of SSD and the limited connectivity of the PCI Express bus.

Instead, architects should consider deploying PCI Express SSD in servers that demand extremely low storage latency or applications generating massive amounts of random read and write operations. The expense and difficulty of integrating these devices requires a careful examination of the various servers that make up critical applications. Consider investing in an application performance monitoring (APM) software suite to characterize application bottlenecks and identify the optimum locations for these cards.

It's too simple to say that databases are appropriate for PCIe SSDs because the performance profile of database-driven applications varies greatly. This is one product that requires a deeper knowledge of applications, so a sit-down with database and application managers is in order. Consider non-traditional applications as well: PCIe SSDs have found success in web applications and creative workstations, not just database servers.

PCI Express SSD implementation challenges, considerations

As PCI Express devices, these SSDs require an empty slot inside the server as well as an outage window for installation and maintenance. This can be problematic for mission-critical applications, but most will have some opportunity for installation.

Blade server users face special challenges when it comes to PCIe SSDs. Dedicated mezzanine SSDs exist for Hewlett-Packard (HP) Co.'s c-Class blade chassis, but it's more difficult to install them in other blade servers. Many vendors sell PCI Express expansion chassis, and companies like Aprius Inc. and Xsigo Systems Inc. enable these to be shared; however, these impact the performance of a PCIe SSD to an extent.

These devices are also expensive, though perhaps not when compared to a high-performance enterprise storage infrastructure. Because they're PCI Express devices and require special operating system-specific drivers, a PCIe SSD can't be easily shared with other servers. Such a card will be of great benefit to the server it's installed in and those that rely on its I/O processing abilities. But this investment can't be spread among a group of servers, and any excess capacity will go unused.

The future of PCIe SSD

PCIe SSD is an entirely new category of storage device, delivering unprecedented random I/O performance right inside critical servers. The rapid growth of sales at companies like Fusion-io, LSI Corp., Texas Memory Systems and others indicates that there are many buyers looking for this kind of performance.

These devices should be deployed as point solutions to specific performance demands. Use application performance monitoring software to determine if you have an I/O bottleneck and consult with database and application managers to decide whether a PCIe SSD is appropriate for their needs. As you can see, both the devices themselves and their use case are entirely new for enterprise storage managers.

PCIe SSD Pros and Cons, Use Cases, and VAR Recommendations

Learn the pros and cons of PCIe SSD, how NAND flash works, typical use cases for PCIe SSD, as well as the vendors and products in this space. Plus, get recommendations for VARs interested in selling PCIe-based SSD to customers.

PCI Express, or PCIe, is a computer motherboard interconnect that's used to expand the performance, functionality or connectivity of a server. HBAs, NICs, graphics accelerator cards and RAM expansion modules are common examples. Solid-state storage devices in the form of NAND flash are a newer application for the PCIe interconnect. This article will detail PCIe SSD: PCIe cards that contain on-board SSDs and provide *internal* storage to a server as opposed to *external* SSD arrays that are connected through a PCIe interface card.

Flash-based SSD is seeing increasing use because disk storage can't keep up with CPU performance; hard drive rotational speed hasn't increased in almost a decade. Even though some performance improvement has come from increased data density (more data per rotation), the fact remains that hard disk-based storage can't get data to and from CPUs fast enough to take advantage of continually improving processors. Memory (DRAM) is too expensive to provide the capacity needed, so flash storage has been tapped to fill this gap. With performance that's a lot closer to DRAM than HDDs but a cost per gigabyte that's in between the two, flash offers a viable alternative to spinning disk storage.

PCIe provides an efficient platform for integrating this flash into the compute environment, putting it onto the server motherboard and moving it closer (physically and logically) to the server compute engine. Compared with external devices -- such as traditional disk arrays with SSDs instead of hard drives, dedicated SSD arrays or caching appliances -- flash on the PCIe card storage can bypass the array controllers, interface protocol and cabling that direct-attached external devices have. They also eliminate the switches and networking connections of SAN-attached arrays.

How NAND flash works

As a data recording medium, NAND flash has a number of characteristics that require special processes. These processes represent overhead to be carried out by either a flash controller that's resident on the PCIe card or by the server CPU. Generally, they're associated with flash write operations but also handle reliability processes like error correction and replacement of bad flash blocks.

When data is written to flash, old data must be erased ahead of each write to make room for the new data. Also, this flash erase must occur in blocks, not at the byte level as with HDDs. This means that when a section of data is deleted, the entire block it's located on must be erased and the bytes *not* marked for deletion must be copied to another location.

NAND flash, unlike magnetic disk substrate, has a finite life span, meaning it can only be written to and erased a certain number of times. In response to this condition, flash controllers spread data writes across all available blocks so that the entire chip ages at a more even rate. Called "wear leveling," this process maximizes the life of the NAND substrate and, consequently, the SSD that contains it.

PCIe-based flash pros and cons

Flash on a PCIe card puts the SSD closest to the CPU memory complex within the server and affords a lower potential latency than external SSD devices. PCIe is bidirectional, meaning reads and writes can occur simultaneously, and it eliminates the protocol translation of SAS, SATA or FC storage. All this adds up to faster I/O than external devices, which use these protocols and often have to share CPU with RAID functions and storage features as well. Obviously, network-attached flash storage systems put the additional network latency into the equation.

Since it's a captive, dedicated storage device, a PCIe SSD card can be the easiest to install and use. External systems, especially those connected to a storage network and shared with other hosts, are typically more complex to implement. These larger-capacity storage devices represent a bigger investment and often involve strategies like storage tiering and caching to improve SSD utilization.

Being an internal, bus-connected card, PCIe SSD devices cause a reboot when installed or replaced, making expansion or servicing more disruptive. Also, since they consume a PCIe slot, they may reduce the number of other PCIe devices, like network adapters (or additional SSD cards) that can be used. As dedicated storage, PCIe SSD can't be shared with other servers, something that could help cost-justify this premium-priced capacity.

PCIe SSD use cases and products

PCIe SSD is ideal for I/O-intensive applications, like Web 2.0, social networking, OLTP, active databases, etc. It can also provide quick performance improvements for localized application problems that require a large, fast buffer area, like video editing, financial modeling, simulations and data acquisition. This direct-attached storage (DAS) implementation may also appeal to first-time SSD users, compared with the potential complexity of an external or network-attached array. This form factor is also ideal for OEMs to include SSD with servers as an upgrade option.

Texas Memory Systems puts the flash controller on the PCIe card, so it doesn't impact the server CPU or memory resources to handle the write overhead mentioned earlier. According to Levi Norman, director of marketing, "NAND flash returns errors; you just have to manage it. Texas Memory Systems designs PCIe cards as a system, with onboard RAM and CPU, giving it the intelligence to conduct more efficient I/O and improve performance, especially for write-heavy applications."

Mitch Crane, senior director of technical marketing at Fusion-io, said, "Everyone else is trying to make flash look like a disk drive; we're making it look like memory." He said Fusion-io's products use physical memory addresses and DMA to treat flash more like system memory than storage. Fusion-io products don't do flash overhead functions on the card, which Crane said increases complexity and chances for failure.

LSI's PCIe SSD solution takes a little different approach. Its card includes six SSD drive modules, as well as a SAS/SATA bridge that acts an interface between PCIe and these modules. The flash controller functions are handled by the SSD modules themselves. JB Baker, the company's principal product manager, said, "The use of SAS controllers in the solution architecture gives LSI the flexibility in future designs to consider including ports to external SAS/SATA devices, which could enable users to expand and scale the system."

VAR recommendations

Although PCIe-based SSD products are newer and represent a small portion of the total SSD market currently, they're certainly a hot topic. According to Gartner, market share for PCIe-based SSD was 6% of all solid-state shipments in 2009 and is predicted to be four times that by 2013, as all major SSD vendors come out with products. From a performance perspective, PCIe SSD represents a fast and efficient way to apply flash performance to a server IOPS problem. While performance is the key driver for SSDs, actual mileage may vary depending upon the PCIe solution chosen and the environment. But performance is a discussion integrators are used to having with customers. Another one is implementation.

As practitioners are finding out, getting the kinds of utilization they need to see from their SSD investment is easier said than done. Adding SSDs to existing arrays or putting in dedicated SSD appliances can be a complex and expensive solution. As a smaller, localized implementation of SSD, a PCIe card can be easier to get approved, easier to put in and be faster to show the performance improvements expected.

For VARs, this shouldn't be the only SSD product on the line card. Like other technologies, integrators should handle a range of products and represent multiple implementation options. Whether it's localized application acceleration, dedicated high-performance storage or the simplest way to get an IOPS problem resolved, PCIe SSDs provide a strong set of options for VARs.

Pumped Up Performance: PCIe SSDs

PCIe SSDs offer the best flash performance available; and now it can be shared among multiple servers, too.

PCIe SSDs offer the best flash performance currently available in your data center computing environment; and now they can be shared among multiple servers as well.

As solid-state storage increasingly becomes a storage system alternative, it's becoming universally recognized as the storage performance option. But its performance and near-zero latency can expose weaknesses in the rest of a storage infrastructure, including its backbone, the storage network. How and where solid-state is implemented is a significantly more important decision than it was with any hard drive-based storage systems. Given the cost and expected performance benefits, avoiding the storage infrastructure altogether by placing solid-state directly in line with the server CPU via the PCI Express (PCIe) bus may be the best option.

This direct connection to the server CPU has led to the rapid adoption of PCIe-based solid-state drives (SSDs). PCIe SSDs benefit from bypassing the storage network altogether, creating almost no latency when retrieving data via a PCIe data path.

Most, but not all, PCIe SSDs also ignore the entire storage protocol stack. This eliminates further latency and allows PCIe SSD vendors to provide custom drivers that are specifically designed for flash-based storage rather than hard disk storage.

All these factors have combined to boost adoption of PCIe-based SSDs and to a substantial expansion in the number of vendors offering PCIe SSD products. In fact, an entire ecosystem of hardware and software has evolved that can sometimes be overwhelming.

PCIe SSD differentiators

Most PCIe offerings have only a few basic components in common. First, they all use NAND flash chips to store data and, second, they put that flash on a PCIe card that's primarily designed to be installed in a server. After that, the features and capabilities of the cards vary greatly. Depending on particular environments, the differences can have a significant impact on performance and the return on a PCIe solid-state storage investment.

During the selection process, it's important to focus on solving the specific performance problem, not on selecting the fastest available card on the market. Paying extra for performance that can't be used is a waste of your IT budget.

Server-side SSD isn't just PCIe

When NAND flash first started gaining popularity, storage protocols and storage interconnect speeds were the performance bottlenecks. This made PCI Express (PCIe) immediately attractive. But in the past year, SATA, SAS, Fibre Channel and Ethernet have all improved performance. While the near-zero latency and direct CPU access of PCIe remains a distinct advantage, the latency gap has definitely narrowed.

If the most extreme level of performance isn't needed, you may benefit just as much from SAS- or SATA-based solid-state drives designed to populate existing hard disk drive bays. In addition to offering excellent performance, many servers allow these drives to be hot swapped for more highly reliable configurations.

Storage protocol compliance

A key differentiator is whether the PCIe SSD card is storage protocol compliant. In other words, when the card is installed in a PCIe slot in the server, will it be recognized as a storage device. A non-compliant card would need an additional driver installed in the operating system (OS) or hypervisor. There are pros and cons to both methods.

If the card isn't storage protocol compliant -- also called "native PCIe SSD" -- it should yield better overall performance because I/O to the card doesn't have to traverse a storage I/O stack designed for hard drive-based systems. It also means a custom driver must be shipped with the PCIe SSD for specific OSes and hypervisors. This driver knows how to directly access the card and should reduce overall protocol latency. How much of a benefit the reduction amounts to is highly dependent on the physical server and the application running on it. Also, most PCIe SSD vendors don't support every OS and hypervisor, so confirming compatibility is critical.

Some PCIe SSD vendors also provide an API set that users can leverage so their applications can have direct access to the PCIe SSD. This allows the application to not only bypass the overhead of the storage protocol stack but also of the OS itself. Of course, this is only valuable if it's possible to access the application source code.

Most native PCIe SSDs can't be booted from directly since the driver needs to load first. This means another boot device must be present in the server. Ironically, many users complement a PCIe SSD with a standard drive form-factor SSD. Having to buy two flash devices to complete the task at hand makes storage protocol-compliant PCIe SSDs that much more attractive.

PCIe SSD cards that are storage compliant look like actual storage devices to the server and OS. Typically, no drivers are needed to access the card and systems can boot from these cards. As a result, these PCIe SSD cards can be used almost universally.

Flash management processing

Flash memory is a unique storage medium that needs special handling. A NAND flash device is made up of multiple cells, each of which can store one or two bits of data. New data is written to cells by first erasing old data from the cell in complete blocks; data isn't overwritten at the byte level as it is with hard disk drives. This erase-then-write process is called the program/erase cycle. NAND flash cells have a limited number of program/erase cycles they can sustain before reliability diminishes. To ensure cells wear out evenly, flash vendors leverage a technique called wear leveling to make sure data is distributed evenly across all cells.

To maintain performance, many flash vendors will do the erase part of the cycle in advance. This technique, called garbage collection, scans for old data on an ongoing basis and "pre-erases" it. This allows for better performance when write traffic is high because the writes don't have to wait for the data to be erased first.

Flash vendors also add their own flash management features. For example, some vendors will add flash intelligence that can reduce flash substrate degradation by "softening writes" when I/O traffic isn't high in an effort to extend flash life. Others have added better data integrity and data protection routines.

All these flash management steps require processing power. Almost every other form of flash storage (all-flash arrays, flash appliances and SSDs) includes this processing power within the storage device. For PCIe SSDs, it's a divided camp: some PCIe vendors use their direct access to the host CPU and offload flash management processing to the host, while others have included onboard processing.

Leveraging server CPU and server memory resources to assist with flash management allows PCIe SSD vendors to shorten their development cycles. It may also reduce costs since they don't have to create their own silicon or use field-programmable gate arrays (FPGAs) to handle the processing. These designs do all their work in software, which means the speed and availability of server resources will directly impact overall flash performance.

Typically, there are plenty of server CPU resources to go around, but leveraging the host CPU may lead to unpredictable performance under high load conditions. Often, when the server CPU resources are being pushed to their limits, storage I/O traffic is also the greatest. The wrong combination could lead to a momentary, but unexpected, drop in performance.

PCIe vendors that have built their own hardware-based flash management are quick to point to this unpredictable consumption of host resources as a key problem with host-based flash management.

Sharing PCIe SSDs

One of the shortcomings of a PCI Express solid-state drive (PCIe SSD) is that it's exclusive to the server it's installed in. But for many environments, like server virtualization, sharing is required to provide high availability, redundancy, scalability and so forth. There are other more traditional methods of sharing PCIe SSD beyond the emerging flash-only, SAN-less architectures.

The first sharing method is for traditional storage vendors to incorporate PCIe SSD into their storage systems or infrastructures, such as arrays that leverage standard hard drives in conjunction with PCIe SSD. Hot data can take advantage of the faster PCIe SSD storage while cooler data is placed on traditional hard disks.

Some vendors have developed converged architectures that include host processing and storage across several nodes to create an all-in-one server/desktop virtualization offering. The local virtual machine (VM) images are stored in the node hosting each VM. The second copy is spread across the remaining nodes for redundancy and to enable VM migration.

Finally, a number of vendors have created fibre-attached appliances designed to house multiple PCIe SSD cards. Most of these appliances allow sharing of any installed PCIe card; servers can connect to the appliance via a PCIe network, InfiniBand or even Ethernet.

PCIe SSD software

A final PCIe solid-state drive consideration is the software that the vendor provides with the card or as an option. For native PCIe vendors, the device driver itself must be part of this software set. In addition to ensuring that it supports all the operating systems the card will be used with, you must ensure that the driver is compatible with other drivers in your server's boot stack.

Both native and storage protocol-compliant PCIe SSD vendors provide software that enhances the overall use of the board. In some cases the included software can analyze data access on the server and help determine which data would benefit the most by being placed on the SSD.

The PCIe SSD market has been partly responsible for the emergence of new vendors that have developed caching software so that static placement of data can be replaced with dynamic use based on data access activity. This makes the use of PCIe SSDs much simpler and eliminates the need to constantly analyze the environment for the most SSD-appropriate data. Because of the natural fit between PCIe SSDs and caching software, many PCIe SSD manufacturers have acquired a caching software vendor and now bundle the caching app with their hardware.

An interesting wrinkle to the caching market is the emergence of PCIe SSDs with the caching function built into the hardware. This provides the benefit of caching without the need to load additional software or to use server resources to perform the cache analysis.

Some PCIe vendors now provide the ability to mirror flash cards between servers via a high-speed 10 Gigabit Ethernet or InfiniBand network. This can eliminate the need for a SAN altogether. These software applications leverage the fault tolerance capabilities of products like VMware vSphere Fault Tolerance to provide application and data availability in the event of a server failure. This is a trend that should continue to grow in popularity since it provides the local access performance of PCIe SSDs while leveraging the shared resources and redundancy of a traditional SAN.

Summing up PCIe SSDs

All PCIe SSDs aren't created equal. The way the boards are designed and implemented can make a difference for both performance and durability. Also, the emerging software components of the solid-state drive ecosystem (like caching) can significantly impact how fully the PCIe SSD investment is exploited. Prospective PCIe SSD users should look for a board that most simply meets their performance needs at the best price, with some room for performance growth. Performance demands may require advanced features, but often a basic storage-compliant PCIe SSD will yield the best overall value and easiest implementation.

TechTarget

Free resources for technology professionals

TechTarget publishes targeted technology media that address your need for information and resources for researching products, developing strategy and making cost-effective purchase decisions. Our network of technology-specific Web sites gives you access to industry experts, independent content and analysis and the Web's largest library of vendor-provided white papers, webcasts, podcasts, videos, virtual trade shows, research reports and more —drawing on the rich R&D resources of technology providers to address market trends, challenges and solutions. Our live events and virtual seminars give you access to vendor neutral, expert commentary and advice on the issues and challenges you face daily. Our social community IT Knowledge Exchange allows you to share real world information in real time with peers and experts.

What makes TechTarget unique?

TechTarget is squarely focused on the enterprise IT space. Our team of editors and network of industry experts provide the richest, most relevant content to IT professionals and management. We leverage the immediacy of the Web, the networking and face-to-face opportunities of events and virtual events, and the ability to interact with peers—all to create compelling and actionable information for enterprise IT professionals across all industries and markets.

Related TechTarget Websites

- SearchDataBackup
- SearchCloudStorage
- SearchDisasterRecovery
- SearchVirtualStorage
- SearchSMBStorage
- SearchStorage