



Dell, Inc.

Busting solid-state storage myths

Maximizing the impact of flash storage

An exploration of solid-state storage, a survey of the technology, its uses and applications.

Bob Ganley, Dell Inc.

March 2013

Contents

- Is this the start of something big? 1
- Innovations in solid state chip technology 1
- Endurance for Solid State Storage 2
- Is SSS reliable? Can I trust it with my data? 3
- Solid State Form Factors 3
 - Solid State Drives 3
 - Solid State Cards 4
- Architecture choices for solid-state storage 4
 - PCIe attached flash in the server 5
 - SSD in the Server 6
 - Flash as cache in shared storage 6
 - Flash Tier in shared storage 7
 - Flash in Hybrid storage 8
 - Flash shared storage in a dedicated high speed array 8
- Should SSS be a part of your IT architecture? 8
- Real world use cases 9
- Conclusion 10



Busting "solid-state" storage myths

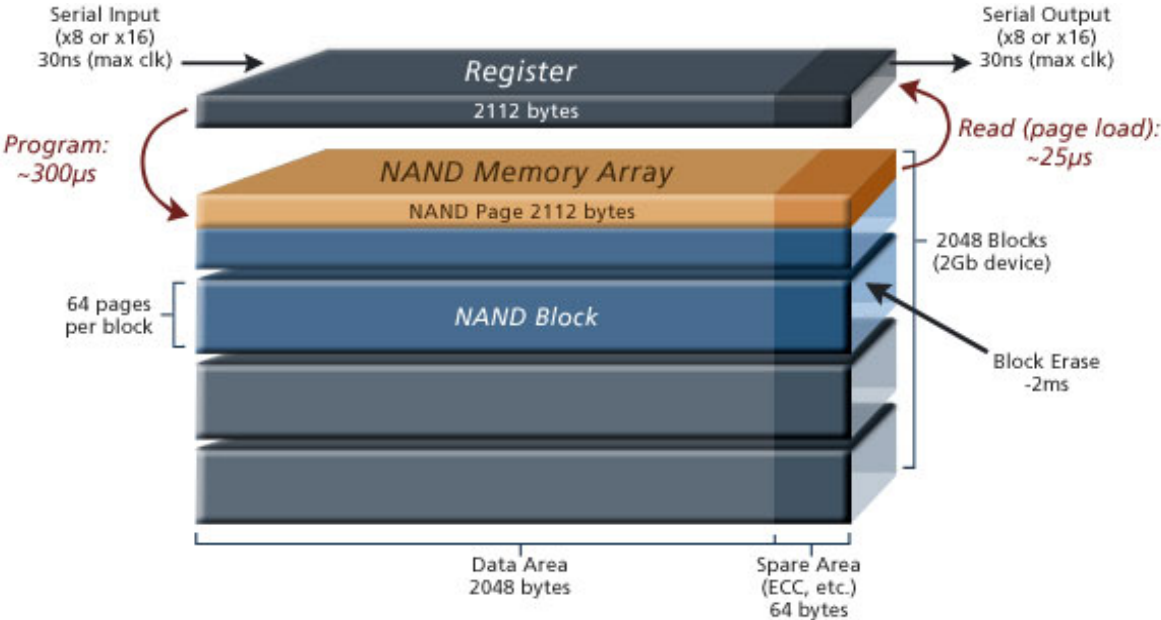
Solid-state storage (SSS) is receiving a lot of buzz today. Trade publications are regularly mentioning the topic. Startups involved in solid-state storage are announcing exciting innovations. Mainstream vendors are creating partnerships in the space and announcing innovations of their own. Why all the hype? What is the state of the technology? Is this just a flash-in-the-pan (pun intended...) or will most organizations be deploying some type of solid state solution soon? This paper summarizes the state of the technology and reviews some common applications for solid-state storage, including some examples of how organizations are using this technology today to make a positive difference in achieving their goals.

Is this the start of something big?

So where is the industry in the adoption of solid-state storage? Most analysts and thought leaders in the industry agree: we are right at the beginning of a major change in the market that will see massive adoption. Many organizations have small deployments today, but over the next few years, most "active" (regularly accessed) data will move to solid-state storage. Where are you on the adoption curve?

Innovations in solid state chip technology

Solid-state storage is the application of non-volatile memory (NVM) for long term data storage. NVM is memory that retains data even when it has no power. Innovation in the non-volatile memory market leading to increasing storage density is the main driver that is



1 2Gb SLC NAND flash device (image courtesy of Micron Technology, Inc.)



enabling the adoption of solid-state storage. The most significant increases in density have been due to increasing the amount of data that can be stored in a single cell. NAND flash is the most commonly used NVM for solid-state storage. The most common enterprise NAND flash technology is Single Level Cell or SLC Flash which stores a single bit in one cell. Multi-Level Cell - MLC Flash - allows more than one bit to be stored in a cell. A common application of MLC flash is the ubiquitous thumb drive seen in consumer applications. Increasing bit densities, combined with shrinking transistor sizes on the physical chips has resulted in dramatic increases in capacity for solid-state storage.

Higher density flash has led to more data being stored on the same amount of silicon. Increasing demand, innovations in manufacturing processes and efficiencies have led to dropping prices for NAND flash. Between 2010 and 2013 raw storage prices have dropped from around \$3 to around \$1 per GB. According to industry analysts, prices will drop an average of 38% per year moving forward, which equates to prices halving every two years. Of course raw flash costs are very different than \$/GB for useable storage which takes into account things like: costs of RAID parity, over-provisioning waste, and management software. While calculating total storage costs per GB for your data center, you also need to account for costs associated with data protection, snapshots, replication, backups, data center space and operational costs. So don't go running to your storage vendor asking for \$1/GB flash storage, and certainly don't use that number in any budget calculations!

Endurance for Solid State Storage

The challenge with the higher density chips is that they have less endurance. The chips wear out as they are used. This endurance is measured in write cycles. The standard for SLC endurance is 100,000 write cycles. Consumer grade MLC has a nominal endurance of 3,000 write cycles. A higher grade "Enterprise" MLC or eMLC has been developed that has an

Technology	Relative Attributes					
	Single-Package Density	Endurance	ECC Complexity	Performance	Price/GB	Interface Complexity
Single-Level Cell (SLC)	•	•••	•	•••	\$\$\$	••
Multi-Level Cell (MLC)	••	••	••	••	\$\$	•••
Triple-Level Cell (TLC)	•••	•	•••	•	\$	•••
Serial NAND (SLC)	•	•••	None ³	•	\$\$\$	•
Enterprise NAND – SLC ¹	•	••••	••••	••	\$\$\$\$	••••
Enterprise NAND – MLC ¹	••	••••	••••	•	\$\$\$	••••
e•MMC	•• ²	••	None ³	••	\$\$	•
MCPs – NAND with DRAM	•	•••	•	•••	\$\$\$	••
ONFI Modules – SLC	•••	•••	•	•••	\$\$\$	••
ONFI Modules – MLC	••••	•	••	••	\$\$	•••

• = Lower; •• = Medium; ••• = High; •••• = Highest

Notes: ¹Requires enhanced ECC and relaxed timings. ²Offered in a wide range of densities. ³ECC circuit and processing built in.

2 The many levels of NAND flash provide a wide range of solutions (courtesy of Micron Technology, Inc.)



endurance standard of 30,000 write cycles. What does this mean in terms of reliability for an SSD drive using SLC flash? If you had a 600GB drive that can last 100,000 write cycles, you'd have to write 450GB/hour 24 hours per day for 5 years to reach 100,000 write cycles for the drive. Furthermore, 100,000 write cycles is the minimum standard for SLC endurance that is guaranteed by chip foundries. In order to meet that minimum manufacturers produce chips that (on average) last a lot longer. Flash cells rated for a minimum life of 100,000 write cycles can actually last for millions of write cycles before they fail.

Is SSS reliable? Can I trust it with my data?

If a bit fails on a flash chip, your data storage does not fail. Errors on storage media have been an issue since the dawn of storage. Bit-level errors happen in traditional spinning media as well. Error correction codes (ECC) to manage bit errors have been in use for over 40 years in hard drive controllers and are very well-known. Controllers that manage flash storage use additional techniques that make these systems extremely reliable. There is an amazing amount of science and engineering that goes in to managing solid-state storage. Advanced techniques like wear-leveling, over-provisioning are combined with ECC to provide solid-state storage systems that easily last 5 years in production.

Solid State Form Factors

Solid-state storage is commonly packaged into at least 3 form factors:

- Solid State Drives (SSD) - This is flash storage packaged in a hard drive form factor such as 3.5" or 2.5" drives commonly found in servers and storage arrays.
- Solid State Card (SSC) – Flash storage packaged on a card such as a PCIe card for use in a standard interface.
- Solid State Module (SSM) – Flash storage packaged on a DIMM similar to how DRAM is packaged – this paper doesn't discuss this form factor.

Solid State Drives

Solid state drives can be used anywhere a spinning disk (HDD) can be used. The major benefit of SSD compared to HDD is direct access. HDD requires seek time, flash memory does not. Therefore SSD will provide higher transaction rates (IOPS) than HDD. The raw differences can be stunning: on the order of 100-200 IOPS for HDD vs. 5,000 -10,000 IOPS for SSD. As a rule of thumb: [a flash disk can provide 40x better IOPS](#) than a spinning disk. SSD and HDD use the same controller interface types (FC, SAS or NL-SAS or SATA) which provide the limit for bandwidth to a disk subsystem. Since sequential



3 Intel® SSD 520 (Courtesy of Intel Corporation)

workloads tend to be bandwidth limited, performance increases seen for SSD will be greater for random workloads than for sequential workloads. SSD read performance is better than write performance; A rule of thumb here might be that SSD read speeds are 2x SSD write speeds.

Another factor is cost. In terms of raw capacity, SSD costs more than HDD. With current (2013) technology, when you include controllers and other overhead a general rule of thumb might be that SSD are dollars per GB while HDDs are pennies per GB. Of course because of the dramatic performance differences, the cost for performance equation tilts in the other direction. SSD storage system price-per-performance is measured in pennies per IOPS while hard drive-based storage system IOPS costs are measured in dollars.

Power consumption may also be a consideration today for certain applications. SSD uses a lot less power than HDD: A single SSD consumes on the order of 1 watt while an HDD consumes on the order of 10 watts. When comparing HDD to SSD it is common to use IOPS/Watt as a way to compare the relative power impact for a given level of power consumption. SSDs power performance is on the order of 1000s of IOPS/Watt compared to HDD performance in the 10s of IOPS/Watt! As power density becomes more important in larger data centers, SSD technology becomes an important consideration.

Solid State Cards

PCIe attached NAND flash on a solid state card allows a high volume of data to be pulled closer to the processor when compared with DIMM-packaged DRAM. Compared to standard hard drive interfaces, this provides faster access to data because of the PCIe bus. This dramatically increases the throughput available when compared with SAS or SATA interfaces. Throughput for PCIe attached flash is measured in GB/S compared to throughput in the 100s of MB/S for SAS or SATA interfaces.

Architecture choices for solid-state storage

Solid state storage technology can appear in several different places in your information technology infrastructure:

- In the server PCIe attached
- In the server as an SSD
- In shared storage as a cache
- In shared storage as a tier
- In a hybrid storage system
- In shared storage as a dedicated high speed array

Each of these options will be discussed in the sections that follow.



PCIe attached flash in the server

A common use of PCIe attached flash is caching for disk Input/Output (I/O). As the operating system calls for disk I/O the data is read into the cache and kept there until overwritten. This is almost always used as read-only cache because write caching would risk data loss since some of those writes might not be written to the disk in the event of an interruption. This leaves the disk storage



4 Removable PCIe flash card in a Dell Poweredge R720 server

in an inconsistent state. Write-through cache requires waiting for acknowledgement from the back-end storage (whether DAS or shared) providing no performance advantage. An example of such an implementation is the [Dell PowerEdge PCIe Express Flash SSD](#). The flexibility of this type of solution is increasing with the introduction of PCIe capable expansion slots for removable media like the one shown here in a [Dell PowerEdge R720 12G server](#).

One drawback of this approach is that the acceleration available from this cache is available only to the workloads running on that server (either in a VM or physical) where the flash cache card is installed. This would prevent taking full advantage of some of the benefits of virtualization including dynamic load balancing and high availability through virtual workload mobility. Another drawback is the lack of integration with shared storage infrastructure. The data protection and management benefits of a shared storage environment are not extended to the data being stored in the PCIe cache. Traditional PCIe flash cache cards only benefit read-intensive workloads running on a dedicated server.

There are new products from some organizations to

Solid-State Storage Myths

Myth #1: SSS is unreliable

While it is true that NAND flash chips have a limited life-cycle, SSS systems on the market today are very reliable. You can trust your data to solid-state storage.

Myth #2: SSS is expensive

While it is true that in terms of \$/GB SSS costs more than spinning media consider this: Traditional solutions such as maximizing spindle count and disk short-stroking drive up costs for achieving high performance SLAs. A small amount of SSS can eliminate significant CAPEX costs resulting in an affordable solution.

Myth #3: SSS requires special skills

Some traditional SAN solutions still require complex manual configuration. However modern highly virtualized SAN designs automatically incorporate SSS as a tier. These solutions eliminate the OPEX associated with specialized expertise while providing all the performance benefits.

Myth #4: SSS is only for the large data center

The advent of hybrid storage solutions, scale-out architectures and automated tiering has made the benefits of SSS accessible to small and medium-sized organizations.

address these issues. One improvement is to provide write-consistent cache with data protection. This provides the ability to accelerate reads as well as writes while preventing data loss in the event the cache card fails. Dell has released [Dell Fluid Cache for DAS 1.0](#) application acceleration technology that combines the Dell PCIe SSD with Fluid Cache software to deliver read and write caching for direct-attached local disk. See this [Dell TechCenter blog](#) for some interesting data on accelerating OLTP database performance.

The next step in this technology will be to integrate this capability with shared storage. Dell has announced intentions to provide this capability in a project called [Fluid Cache for SAN](#) that is based on the technology obtained in Dell's acquisition of RNA Networks, Inc. An exciting thing about this project is that it provides high performance to workloads as they move through a cluster of servers along with data protection in the case of a server failure. Through the use of a high speed RDMA interface to connect multiple servers, data can be written to cache in multiple servers simultaneously. This provides two benefits: protection against data loss in the case of server failure and the ability to share the cache with other servers on the high-speed network. Brian Payne, Executive Director of Dell Server Marketing discusses this and the future of flash in the enterprise in [this video](#).

Additional capabilities planned for this project will extend the data protection and management benefits of a [Compellent storage area network \(SAN\)](#) to the data in the cache, essentially making server-attached flash a managed tier in the storage infrastructure. This will blur the lines between server memory and storage!

SSD in the Server

SSD in the server (or direct-attached storage (DAS)) is another option for flash deployment. This will provide performance benefits for that server, but also shares the drawbacks of dedicated PCIe attached flash when it comes to workload mobility. Furthermore given the higher \$/GB of SSD, it would be difficult to rationalize provisioning enough SSD on a server to meet all the capacity needs of DAS in a given server. The question then becomes "what data should I put on the SSD?" This creates a need for detailed workload storage analysis to determine the appropriate distribution of data between SSD and HDD on a server. In addition to requiring specialized expertise for the analysis it is likely that static data distribution will not meet the evolving needs of many workloads – especially in randomized I/O streams created by consolidated, virtualized environments. Of course there are other challenges presented by the need for data protection in these types of environments that are too long to detail here. Suffice it to say that this type of configuration is likely to be appropriate in highly specialized, dedicated installations.

Flash as cache in shared storage

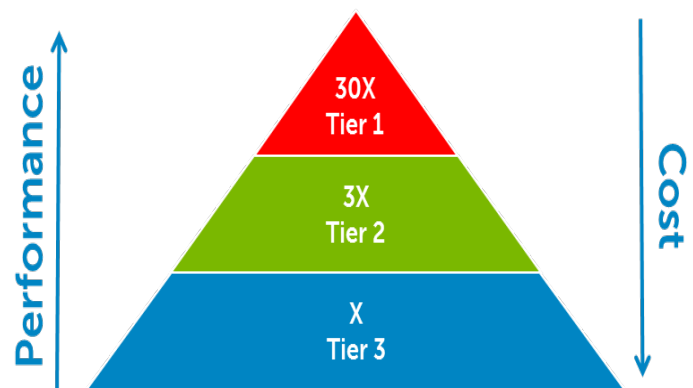
SAN/NAS controllers all use some sort of DRAM caching to accelerate performance. Using PCIe attached flash in a SAN/NAS controller, or dedicated SSDs in drive bays, can expand this caching capability to provide measurable storage performance gains. These configurations



typically provide read-cache only to provide on the order of 10x latency improvements for the right types of workloads (read-intensive). These solutions are expensive, depend on the “cachability” of the I/O stream and provide limited benefit to random I/O. Plus the benefits of this type of cache won’t persist across workload changes, reboots etc. because the cache has to “warm up”. When application “A” runs for a while, its hot data fully populates the cache (the cache warms up) and the performance of that application benefits from a high percentage of cache hits. If application “B” then runs for a while the cache becomes populated with the data from application “B”. When application “A” runs again its performance will lag until the cache warms up again. This behavior can lead to changes in application performance sometimes referred to the “Monday morning” problem.

Flash Tier in shared storage

Most modern shared storage products offer some level of tiering such as using multiple types of storage to allow optimization of both \$/GB and performance. Traditionally this has been done by installing drives with different speeds and interface technologies such as 15k and 10k SAS/FC and 7.2k SATA drives in a single unit. This can be augmented by including SSD drives as part of the mix to provide an additional storage tier.



5 Cost and performance multiples are estimates only

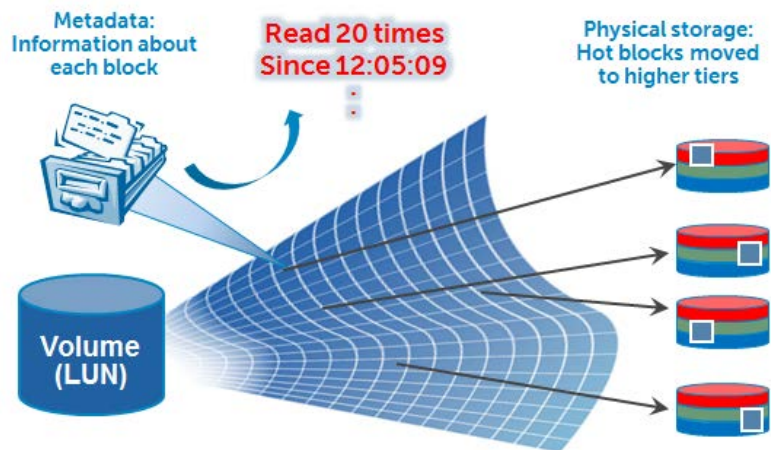
Legacy shared storage products may have the capability to house multiple drive types, including SSD drives, but they require administrators to specify which data or volumes (LUNs) are provisioned on each type of drive. The challenge with this type of LUN-based tiering is that it requires careful analysis and design to determine the “hot” portions of the workloads that would gain the most benefit from SSD storage. Those portions must be configured in a LUN that is then manually provisioned on a portion of the SSD storage. Careful up-front capacity planning, workload analysis and storage provisioning require specialized resources and precious time, often resulting in static configurations that may not remain optimal over time. Performance gains can be achieved, but at a steep cost in terms of expertise and time, reductions in IT agility and increasing administrative costs on top of the capital costs of the solid state drive.

More modern storage products allow “sub LUN” tiering and dynamic provisioning which provide more flexibility. The most effective solutions store historical access patterns as metadata about small blocks of data that comprise the volume (LUN). This metadata is then used to determine the portions of the workload with the highest latency or most frequent access. The system can then automatically move those sensitive portions of the workloads to the highest performing storage (including the SSD tier.) The analysis and movement of

data is done in a non-disruptive fashion so that existing workloads can be optimized while they are running and without manual intervention. This type of [automated tiering](#) requires a highly virtualized and flexible storage infrastructure.

One of the huge benefits of this type of architecture is that “a little goes a long way.” A relatively small amount of solid state drive capacity (as an overall percentage of a SAN) is all that is needed to solve common performance issues. The key to sub-

lun automated tiering is that it moves only the critical few blocks that are your performance bottleneck to your highest tier. This means that from a CAPEX perspective \$/GB storage capacity costs are optimized. Complex analysis of storage workloads is eliminated. Since the process is continuous there is no need for repeated changing of static configurations. The OPEX associated with manual tiering is significantly reduced as well with this type of solution.



6 Automated tiering using metadata to place data on the right storage

Flash in Hybrid storage

What is the concept of “hybrid” storage with flash? This is sometimes used to refer to either of the two scenarios discussed above: DAS or shared storage with a combination of SSD and HDD drives.

Flash shared storage in a dedicated high speed array

The dedicated flash storage array is a new category of shared storage being defined by startups and existing companies alike. This is a self-contained shared storage product that contains only flash-based storage. These products come in one of two options - using SSD drives internal to the storage, or using bus-based interface technologies between the shared storage controller and the flash memory. Both options eliminate the need for data to travel lower speed FS/SAS/SATA interfaces within the storage array for even more extreme performance.

Should SSS be a part of your IT architecture?

What place should solid-state storage assume in your architecture? And what are the best options for the average data center? After all, the vast majority of organizations do not have a team of dedicated storage architects. Most organizations are busy virtualizing their servers and are happy to have shared storage. They are taking advantage of things like workload mobility or snapshots or storage based backups. Few organizations have full-blown disaster recovery scenarios and storage-based replication. The question is, can Solid-state storage

provide benefits to them too or only for the rarefied few high-end data centers with specialized applications?

The fact is, solid-state storage is going main stream today. Adoption is starting to take off at all levels of the market. But it is not a panacea. The key to understanding where it makes the most sense is to categorize workload profiles. Highly transactional profiles – for example OLTP data bases - are a prime example of where SSD can help, see [this great article](#) which shows how automated tiering with SSD was able to improve SQL Server® TPC-TPS and concurrent end users by 170%! This may sound like it is only applicable to rare situations, but remember that common small to medium-sized transactional applications that have a database under them generally include systems like a web-based store or a payroll processing application. Another common workload that is transactional in nature is a virtual desktop infrastructure (VDI) implementation.

How is SSD being deployed to help with mainstream highly transactional applications? Though it would certainly be possible to put an SSD drive in a rack-mount server or a tower server, this will present a new set of challenges when it comes to deciding which data should be put on that particular drive. A much more common scenario for a practical application of SSD to accelerate transactional workloads is as part of a shared storage environment featuring automated data tiering.

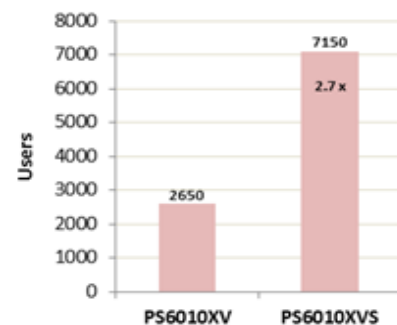
Real world use cases

Let me give a couple of specific examples: [Northwest Mississippi Community College](#) decided to implement a virtual desktop initiative, rolling out 48 virtualized desktops. They decided to use the Dell EqualLogic [PS6000XVS hybrid SAN](#) which contains both spinning media and NAND flash based solid state drives. The PS6000XVS SAN [intelligently tiers workloads](#) between the SSDs and the lower-cost 15K SAS drives. “The ability to distinguish between data that is in high demand versus less important data saves us the cost of an all-SSD SAN,” says Michael Lamar, network technician at Northwest Mississippi. The result has made a significant impact on users’ experience. “We cut login times from 74 seconds to 54 seconds, which is 26 percent less time users have to spend waiting for their work session to start,” says Lamar. “This was all based on moving to the hybrid SAN.”

Another example involves the databases which underpin many critical applications these days. [Nelnet, Inc.](#) provides loan processing outsourcing services and needs to maintain high performance for those applications. They decided to implement a [Dell Compellent SAN](#) with an SSD tier to optimize that performance. Compellent features intelligent automated tiering called [Data Progression](#). “We only allow our main reporting server, a Dell PowerEdge R710

“...we’re not paying to have all of that data sitting on SSD, and we still get the performance benefit.

Ryan Regnier
IT Manager
Nelnet, Inc.



7 SQL Server® performance benefits



server running Microsoft SQL Server® 2008, to access our two terabytes of Tier 1 solid-state drives [SSD]," says Ryan Regnier, IT Manager of Operational Engineering at Nelnet. "But because of Data Progression, most of that data is actually sitting on Tier 2, which is 15K SAS. We're not paying to have all of that data sitting on SSD, and we still get the performance benefit."

As these examples show, solid-state storage deserves a place in the IT tool kit for many organizations. The key to success depends on applying the technology to the right workloads, and leveraging systems that intelligently place the right portions of your workloads on solid state. The exciting thing is that "a little goes a long way", and with a storage system featuring automated tiering you can "set it and forget it."

Conclusion

What should your next steps be? Working with a vendor that has solutions that span the breadth of the Solid-state storage market is a good place to start. Dell provides a broad set of solutions that include solid-state storage in server, Storage Area Network and Network Attached Storage solutions including automated storage tiering that provides a simple way to leverage solid-state storage.

A good way to get more information is to use online resources like www.dellstorage.com. There you can find whitepapers, case studies and reference architectures to better understand the technology and deployment options. Another resource is www.delltechcenter.com. This is an online community where you can network with experts from inside and outside Dell as well as find other technical content for education. Dell provides services to help you understand the options, including a [Storage Assessment service](#) to provide analysis and recommendations on your specific situation. Finally Dell has online access to a tool called [Dell Performance Analysis Collection Kit \(DPACK\)](#) that can analyze your workloads to identify bottlenecks that may be.

Solid-state storage technology is quickly maturing. The basic building blocks for these solutions have been around for a long time, and recent innovations have pulled together those pieces in to integrated storage systems with specific benefits. Organizations are using these solutions to more effectively reach their goals today.

© 2013 Dell Inc. All rights reserved. Dell, the DELL logo, the DELL badge, PowerEdge, EqualLogic, and Compellent are trademarks of Dell Inc. Other trademarks and trade names may be used in this document to refer to either the entities claiming the marks and names or their products. Dell disclaims proprietary interest in the marks and names of others. This document is for informational purposes only. Dell reserves the right to make changes without further notice to any products herein. The content provided is as is and without express or implied warranties of any kind.

