Managing data migration, Tier 1 to SSD Tier 0 July 1, 2009

When EMC became the first storage vendor to announce support for solid state drives over a year ago, the company didn't say what software they had that would migrate business-critical data into and out of the newest Tier 0 cache. It wasn't until April of this year, that EMC introduced its Fully Automated Storage Tiering Tool (FAST), which automatically moves data among storage tiers based on policies, predictive analysis and access patterns. The company's only problem – FAST won't be available until sometime in the second half of this year and full FAST implementation won't be available until 2010.

That issue – moving data into Tier 0 SSD storage – is one many vendors face. Some like Compellent with its Data Progression software say that movement is automatic based on frequency of access. Others such as Fusion-IO have yet to attack the problem.

Moving data manually – as is what happens now in most SSD implementations – is not necessarily something system administrators want to face – they don't want to deal with application downtime when moving data or even with the time-consuming data migration tasks. Determining which files should be in cache and how long they should reside there, and then moving them is an arduous process. Is data such as database redo logs cached in the SSD tier? Or should the most frequently accessed data reside there?

Some implementations such as Pillar Data's Adaptive Caching watch the flight of data across the storage network and allocate the busiest files for an application to SSD cache.

Nonetheless, when vendors introduce SSD drives into their portfolios, they need to take care of the movement of data. Without software to make the decision of where data resides and which data is placed on SSD, much of the benefit of SSD is compromised. *–Deni Connor, principal analyst*

Samples of SSD Management Software

	Product Name	Platforms	Availability
Compellent	Data Progression	Storage Center	Now
EMC	Fully Automated Storage Tiering (FAST)	Symmetrix, Clariion, Celerra	End of 2009
IBM	Smart Data Placement, Dat Facility Storage Manageme System	xSeries, Power, DS800	Now
Pillar Data	Adaptive Caching		Now

Over-provisioning: a winning strategy or a retreat?

Once, as a young systems programmer, I learned that 1KB was actually 1,024 bytes, and that 64KB was really 65,536 bytes. This is because memory is sized as multiples of 2. The difference between a decimal billion and a gigabyte (1024^3) is a substantial 73,741,824. And that's where the old fight between buyers of hard disk drives (HDDs) and manufacturers of HDDs began. The manufacturers started using that "extra" 73.74 million bytes for housekeeping, bad tracks and error correction, selling a gigabyte of data as 1,000,000,000 bytes. It took a while for IT buyers to figure it out and when they did, they were mad, leading to lawsuits, settlements and fine print when it comes to buying gigabyte and terabyte drives.

It is now well practiced that a 'gig' refers to the decimal amount of data storage, which is the net of the real capacity. So in the HDD world, there is a built-in amount of 'over provisioning,' equal to 7.37% per gigabyte.

When we begin talking about SSDs, the same old argument applies. How many of those 'extra' bytes are going into non-data capacity? It depends. Most manufacturers use the net amount of data storage as the size of the drive. We have seen some enterprise drives sized at 50GB (decimal), when we know that the real storage capacity is 68,719,476,736. That's a whopping 37% difference.

What is the justification for this type of 'unused' capacity? The answer is that it is being used for the 'wear-leveling' and endurance algorithm. Because flash memories are subject to wear each time a physical area is written, in order to achieve longer guaranteed use-life, manufacturer's algorithms will use the extra area for data storage after the original area has worn out. In addition to wear, there is a characteristic in flash memories called endurance. Endurance of flash, that is, the time that a memory will 'remember' is expected to be about ten years. But that will begin decreasing the more the area is written, like the lithium ion battery in your cell phone loses capacity after it has been recharged 500 times. So another reason for over-provisioning is to offset the endurance problem, which is where much of the rocket science in the development of SSDs is being developed.

Over-provisioning creates another problem in and of itself. The more devices there are, the more the chance of failure at the device level. No one said this would be easy.

Finally, there is the real factor that NAND flash memory, as a technology, is very mature and is reaching the end of its development cycle. This is good in some ways, bad in others. The performance is very predictable, the devices are relatively cheap and the science in the management is becoming increasingly well-practiced. That's the good news. The bad news is that the point of diminishing returns (in terms of Moore's law) will be reached fairly soon, so the dramatic cost reductions that we have enjoyed will begin to taper off.

As in all technology races, the winners will find solutions to the issues and the losers won't. The trade-offs between over-provisioning and other solutions (including slower performance and less use-life) will be critical factors in the battle.—*Jim Bagley, senior analyst*

Hitachi GST: An interview on the company's SSD strategy

SSG-NOW sat down with Brendan Collins, vice president of product marketing for Hitachi GST last month to talk about Hitachi's strategy for SSD. Following is a partial transcript of that interview.

What specifically did Hitachi announce about SSDs?
Last September Hitachi announced a joint development agreement with Intel to deliver SAS and Fibre Channel drives to the enterprise market. The idea was to extend Hitachi GST's portfolio beyond hard drives to include SSDs. In that announcement we said we would have early samples by the end of this year with full qualification samples at the beginning of 2010 and volume production shortly after that. We're still on schedule to make that happen.

What is the value proposition for Hitachi GST in SSDs?
We found in our quarterly product planning with OEMs that they had tested from 80 to 100 different flash SSDs. We've received varying degrees of feedback on how good or how bad those products work. The consistent piece of feedback was that a lot of flash vendors believed that the value is in the flash intellectual property, but to the people that are developing storage and storage systems and the things we take for granted like the host interface and how it operates, that is the area most flash vendors have problems with.

If you are a big OEM, this is exactly the discussion we have with them -- big OEMs have typically spent the last 10 to 15 years developing the hardware and software architectures, the file systems, the operating systems and the drivers and they are all pretty much designed around the ASIC and firmware architectures we use today in hard drives. So if we were to develop an SSD with the same ASIC and firmware architecture as hard drives that means that OEMs don't have to change anything in that software stack. Any change in the stack is a big risk for them.

If they take any other Serial ATA-based SSD off the shelf, they have been dramatically surprised by the varying degrees or reliability that they've seen. They've concluded that to protect their hardware and software at the system level, it is much easier for them to partner with a hard drive vendor that has that interface established and has the knowledge and the qualification expert team to use that class of product rather than something from a non-hard drive SSD vendor.

You decided to focus on SAS and Fibre Channel, not SATA. Why?
Think about how many vendors in the industry have that capability -- there is less than a handful. We decided to enter that market because there are only two or three that can do it. Of those you have the STEC's and the hard drive companies like Hitachi and Seagate. We have the host interface experience and stability the OEMs want. We can also come up with much more cost-efficient designs than those that are shipping today. When we launch our first SAS and Fibre Channel drives, we expect to have performance parameters that are best in class. They will be totally interoperable with everything else on the system.

If you look at where STEC plays with their Fibre Channel solution, it is at the highest end of the market -- in EMC Symmetrix'. They are sold into the financial services market where millions of stock transactions are

processed and where high response times are critical. When we launched the product we said that we need to have something that can be a direct replacement for Fibre Channel drives. In addition, when you think beyond that you have to have SAS in both 2.5-inch small form factor and 3.5 inch Fibre Channel format. We'll be able to cover the high end of the data center all the way down to the mid-range server space with 2.5-inch form factor drives.

We don't have access to flash or flash IP that Intel owns, so we couldn't even get into this market without them. Likewise Intel doesn't have the SAS or Fibre Channel expertise we have to qualify with these OEM customers. This is an appropriate win-win for both of us. In the first quarter the drives will be marketed as Hitachi drives exclusively.

Pliant Technology sees SSD future in high IOPS, SAS

SSG-NOW recently spoke with Greg Goelz, vice president of marketing for Pliant Technology, regarding his views about the SSD market and the trends which will drive SSD adoption. Pliant, who is developing the next generation of Enterprise Flash Drives (EFDs), recently completed a $15 million C-round of financing and will be announcing products sometime later this year.

"The argument that SSD's are expensive on a cost per gigabyte for a single drive misses the point," says Goelz, adding "when costs to achieve a given transaction rate or streaming output volume, SSD's can cost out at a fraction of HDD costs in both cost per gigabyte and cost per input-output operation (IOP)."

We walked through an analysis of a data warehousing requirement that achieves 640,000 transactions per minute for 18 terabytes characterized by the TPC-C benchmark. The

Tell us more about the SSDs Hitachi will be shipping ...

We are going to leverage Intel's 34-nanometer process for our SSDs. Our first-generation product is going to be a single-level cell (SLC-based) SSD. We believe that these products must be SLC-based to give you the performance and endurance you need. If you want to extend this beyond the high end of the enterprise, that will only happen when flash comes down the cost curve. And, that will only happen when you can begin to switch to MLC. It's the companies can come up with controller IP and solutions that can make an MLC product perform as reliably as an SLC-based solution at MLC-based prices. That will be a second or third-generation based solution.—*Deni Connor, principal analyst*

HDD requirement for this throughput is 1,000 HDD's "short-stroked" to 18GB each for speed. When adding the number of shelves, the total cost for the storage alone as $420,000. And that's just the beginning because that configuration takes 16,000 watts of power to operate and cool.

"Now let's compare that to a hybrid solution that uses enterprise-class SSDs with HDDs that are provisioned at 147GB each and four enterprise flash drives at 150GB each," Goelz explained. "The storage cost drops to $225,000 and operating power to 2,000 watts. The cost per GB is only 44% of the HDD solution and the cost per IOP is halved."

When asked about interfaces required to support enterprise SSD applications, Greg was particularly sanguine regarding the use of Serial ATA protocol (SATA).

"Why would you handicap a device that can read and write simultaneously with an interface that only goes one way at a time? If there is an area of

education that you can contribute with the newsletter, it would be regarding the advantages of Serial Attached SCSI [SAS] due to its bi-directional capability and overall throughput advantages."

Serial Attached SCSI (SAS) was developed as a low-cost method to approach the high throughput of Fibre Channel. The standard took off about four years ago when the ability to support Serial ATA devices was added to the protocol, allowing SAS controllers to use bi-directional devices (like SSDs) and still support the lower cost SATA drives. The protocol has advantages over Fibre Channel in its ability to multiplex up to four devices on a single "wide port." Pliant isn't the only enterprise SSD supplier working on SAS. (See the Hitachi article in this newsletter).

No discussion of SSD strategy is complete without a discussion of Single Level Cell (SLC) versus Multi Level Cell (MLC) technologies.

"We are of the opinion that MLC runs counter to the requirements of an enterprise drive," says Goelz. "Our customers are going to expect seven years of high performance, and we have error correction strategies to support it without going to large amounts of over-provisioning. That takes SLC."

Over-provisioning is a strategy which reserves varying amounts of device capacity for the wear-and-tear associated with re-writing flash memories. We discuss this strategy in detail in this newsletter.

When we asserted that MLC is very likely to become the standard for client side SSDs, Goelz agreed. "Client level cost/performance/IOPS are a good match for MLC on the side, particularly for mobile devices."

Recent industry reports have indicated that 20 and 30 nanometer fabrication processes for NAND flash memories will be required to jump-start the SSD market.

"We're perfectly happy at the 50 and 40 nm price points, and expect continued cost reductions in the 3x range per year without the new fabs coming online," says Goelz. "And that is without the type of dumping that was going on over a year ago."

When we probed regarding Pliant's launch cycle, Goelz says, "All I can say is that we have qualifications going on right now and we are about where I expected to be in this time frame." *–Jim Bagley, senior analyst*

DRAM combined with FLASH gives DDRDRIVE X1 a unique performance profile

SSG-NOW recently spoke with Christopher George, chief technical officer and founder of DDRdrive, a Palo Alto-based company with an interesting approach to achieving high input-output (IOPS) and high reliability in a solid state drive. Rather than using NAND flash as its primary storage, DDRdrive uses DRAM and a PCI Express interface to hit 300,000 IOPS for 512 byte random reads, 200,000 IOPS for random writes, all at low power consumption. In addition, random 4KB byte reads at 50,000 IOPS and 4KB random writes at 35,000 IOPS are supported, and sustained sequential reads at 215 MBPS and sustained writes at 155 MBS are achieved.

"We use the NAND flash strictly for backup of the DRAM," explained George, "and guarantee 100,000 backups during the life of the drive."

The use of the Single Level Cell (SLC) flash on a bit-for-bit backup basis eliminates the need for wear-leveling processes and speed-impeding write amplification to achieve an enterprise-class five - year operational guarantee. Backup can be

performed in less than one minute. DDRdrive's 4GB capacity drive is priced at about $1,500.

"It takes an FPGA solid-state storage accelerator to achieve the high IOPS," says George, "And, it has the advantage of being field-upgradeable as new features are supported."

DDRdrive has driver support for Microsoft Windows 2000 Client/Server, Microsoft Windows XP (32/64 bit), Microsoft Windows Vista (32/64 bit), Microsoft Windows 7 Beta (32/64 bit), Microsoft Windows Server 2003 (32/64 bit), Microsoft Windows Server 2008 (32/64 bit) and Microsoft Windows Server 2008 R2 Beta. "Device driver support has been a primary focus of our development efforts, with Linux drivers under development," says George.

The drive comes in a PCI Express full-Height, Half-Length form factor and is compatible with PCI Express v1.0a, v1.1, or v2.0 X1/X4/X8/X16 slots, double-wide clearance required. Because there is no need for a Serial ATA or other device controller, read-write performance is enhanced by direct-bus operation.

"Our customers to date have been both enterprise and super-users, and we have some OEMs evaluating the product," added Christopher.—*Jim Bagley, senior analyst*

Seanodes announces support for SSDs

Seanodes, a company that virtualizes direct attached and internal storage within Linux or VMware ESX servers, announced support for solid state drives. The company's Shared Internal Storage is now able to combine SSDs with other storage media to deliver 36,000 IOPs on random 4KB read tests, according to Seanodes claims. The Seanodes' SSD implementation will have the most benefits in blade servers shipped by Dell, HP and IBM.

RAID Inc. rolls out dual-controller SSD

RAID Inc. on Monday launched a 1U Fibre Channel or Serial Attached SCSI-based RAID controller. The Razor SSD complements a redundant controller version of the company's 1U Fibre Channel or SAS-base 2.5-inch SSD product, which was released earlier this year. Failover capabilities of the dual-controller prevent any single point of failure, making the controller a good fit for Tier 0 business-critical applications. According to company claims, a dual-controller implementation provided over 130,000 IOPs and 1.4GB/sec throughput. It is available in 32GB or 64GB capacities.

LSI 3ware SATA RAID Controllers support Intel SSDs

LSI this week rolled out a new controller firmware for its 3ware 9650SE SATA RAID controller cards that supports Intel's X25-E Extreme SATA Solid State Drives (SSDs). Intel's X25-E Extreme SSDs are single-level cell (SLC) NAND flash drives that are available with SAS or SATA interfaces. The 3Ware controllers are available now. Existing customer can upgrade to the new firmware for free at www.3ware.com/support/download.asp.

Links to SSD News

Intel's X25-E runs circles around rivals
http://news.idg.no/cw/art.cfm?id=2D3114E9-1A64-67EA-E49C67F5505FCE55

Solid state drive sales stalled
http://www.tgdaily.com/content/view/42953/135/

Can solid state drive maker STEC be stopped?
http://www.enterprisestorageforum.com/industrynews/article.php/3826166

Executive analysis: Traditional disk greener than SSD, on par for IOPS-per-dollar
http://searchstorage.techtarget.com.au/articles/33229-EXCLUSIVE-ANALYSIS-Traditional-disk-greener-than-SSD-on-par-for-IOPS-per-dollar

The bottom line on solid state drives
http://gcn.com/articles/2009/06/29/solid-state-vs-hard-drives-for-storage.aspx

How to Subscribe

The SSD Flash wants you. To subscribe, either fill out the form below, fill it out and scan it and email to info@ssg-now.com. We'll bill you. Or ...

Fill out the credit card information below, and you will be billed automatically each quarter.

The SSD Flash is $495 a quarter for three monthly newsletters, or $1,800 a year.

Name _____ Title _____

Signature _____ Date _____

Company Name: _____

Address _____

City _____ State _____ Zip Code _____

Visa __ Mastercard __Discover ___ Name as it appears on card _____

Account Number: _____ Expiration Date: _____ Card security code _____