



## CATCHING UP TO MOORE'S LAW:

### EMC VFCache Speeds I/O Performance in the Server, Protects Data in the Array

Historically, in the Moore's Law-driven quest for ever faster application performance, disk drives have seriously lagged server chip sets for processing speed. Flash drive technology is helping narrow that gap, supercharging performance for both servers and back-end storage, but independently, on opposite ends of the switch. So while you now have the ability to enjoy faster application performance using Flash on your servers, shouldn't you worry that the data you manage out there is at risk? After all, there's a lot of high-performing storage infrastructure on the other side of the wall that gives you the snaps, backups, recovery, and availability assurances you need to sleep at night, and it probably has Flash too.

Seems server Flash and storage Flash ought to be friends. Shouldn't somebody introduce them?

“It feels like a return to the old days. But it isn’t.”

## INTRODUCTION

EMC® VFCache™ is a PCIe Flash card that caches hot data for fast performance on the server but reaches simultaneously into tiered storage for the full high availability protection of the array. A server running VFCache processes more transactions with less waiting time for users. Response times are cut in half, throughput increases up to 300 percent, and data remains as safe as the storage infrastructure can make it.

Something about VFCache, though, says “back to the future.” After all, a few decades ago EMC pioneered external storage when servers couldn’t handle the volume of data that applications were generating. Big boxes full of disks provided a place to store and protect information, so servers performed better and everybody was happy.

Well, as time went on, not everyone was entirely happy: Conforming to Moore’s Law (chip processing speed doubles roughly every two years), CPU performance has improved 100-fold per decade since the late 1990s. Speed of spinning disk drives has been essentially flat over that period. Storage operating systems could do only so much to keep up, say by pre-fetching likely-to-be-accessed data and placing it on disk in a place where it could be accessed a millisecond faster. The latency introduced by data flying back and forth from the server to the array through the network switch remains a necessary evil in the existing model.

Still, storage intelligence to manage and protect data, for most customers, far outweighed a millisecond here and there lagging behind nirvana speed for applications. Storage networking allowed servers and storage to be broken apart, setting servers and storage on separate evolutionary paths riding different cost curves.

Then Flash drives arrived in the enterprise, along with the irresistible temptation to bring all that storage speed back to the server where the computing was taking place and say goodbye to all that latency. It sort of feels like a return to the old days.

But it isn’t.

## FIRST IN FLASH

In 2008, EMC was first to use Flash-based solid-state drives (SSDs) in enterprise storage and has shipped more Flash capacity than anyone in the industry. Half or more of all EMC Symmetrix® VMAX™ high-end storage systems and EMC VNX® unified storage systems ordered now incorporate Flash capacity. EMC is making Flash technology more affordable as it uses a new generation of high-capacity, low-cost MLC (multi-level cell) SSDs for enterprise applications, in addition to SLC (single-level cell) technology it currently offers.

But when EMC shipped its first enterprise Flash drives in 2008, the first and last word was performance. The first customers to be called on were the ones who needed ultra-high performance, who seemingly always said that the next-generation platform didn’t have high enough IOPS (input/output operations per second) or low enough response times. EMC could tell those customers that enterprise Flash drives solve a specific problem for a specific application. It wasn’t the best price per gigabyte, but it wasn’t a capacity problem. It was a performance problem for the neediest users. Then FAST came along—EMC’s pioneering Fully Automated Storage Tiering—and brought array Flash to a much wider audience.

EMC, of course, wasn’t the only vendor looking closely at what Flash could do for application performance, and if you looked at the situation conventionally you might wonder if there was a logical threat to EMC’s model. EMC has innovated for years to help customers change Direct Attached Storage (DAS) into Storage Area Network (SAN) storage, and for a couple of decades the overwhelming trend has been falling DAS and rising SAN.

Then Flash arose as a force that said maybe DAS Flash on the server could pull the fight back the other direction. Given network latencies, putting Flash on the other side of the network from the server carries a high access cost, so wouldn’t there be significant value from a compute perspective to putting it on the server side of the network?

“The key mechanism is integration with EMC’s FAST technology.”

Not so fast. The downfalls of DAS are the fact that it’s not shareable and it’s not highly available. Customers consider such attributes of SAN storage to be absolute requirements. What happens to a write sitting on that high cost/high-performance DAS Flash that’s not hot data anymore? What if it needs to live someplace that’s warm and cost-effective, and then needs to migrate to someplace that’s cooler and even more cost-effective? And what if it needs to be compliant with retention policies and available at many sites? What happens if the auditors show up and it needs to be brought back and scrutinized?

If you put it all back on the server, you either take yourself back to the old DAS problems or you have to come up with a solution to get around them, or else your application has to be responsible for it. So each application you deploy might use a different solution, so you have to manage multiple ways of solving those issues.

With Flash as a game-changer in terms of squeezing higher performance out of an I/O subsystem in the server, it seems like a shame to sacrifice the value that you get out of a shared SAN storage. So how do you get the best of both worlds?

## FLASH IN CACHE

Whichever side of the server-storage debate you sit on, from a customer perspective, it’s all about feeding the server—rapid access to important information. EMC engineers looked at the situation and saw that the solution was cache.

If you put Flash into the server by turning it into a cache, especially a write-through cache that passes writes to the array for protection while optimizing reads, then you preserve all the value of SAN storage and gain performance at the same time. All EMC arrays have cache on the front end. Why not simply slide some cache up the wire and have it sit in the server? Compared to DAS Flash vendors, who talk only about performance, the cache approach knits together server Flash and array Flash and has them working together. Isn’t this the most natural way you’d want to do this? Who else could do that but EMC?

So EMC launched an initiative called “Project Lightning” and developed what came to be known as VFCache. Here’s how it works:

- VFCache is a software driver that runs in the operating system on the host—Windows, Linux, or VMware®. Essentially, it intercepts the I/Os that the server thinks it’s sending out to the array, looks them over, and determines whether it can service them locally out of the Flash without having to send it back to the EMC array behind the server.
- It’s called “write-through cache” because writes behave differently in different types of caching algorithms. VFCache puts a copy of the data in the Flash device, in the cache itself, but also back onto the EMC storage behind it. Why? To make the server cache look like an extension of the array cache, which includes bringing all the value that the array brings to the table in terms of write protection, disaster recovery, and so on.
- The key mechanism is integration with EMC’s FAST technology, by which data is automatically allocated to the right kind of storage depending on how urgently it is expected to be needed by an application. The FAST policy engine that operates today makes choices all the time to move data between tiers on the array. What VFCache does is provide another aspect of this intelligence leveraging the server flash as a higher tier for this data.

The idea is to preserve all the enterprise value around protecting the data and creating a high-availability environment. To do that requires always putting the data on the EMC array. So VFCache writes the data both to the cache and to the EMC array, so if the server or the Flash card in the server blows up, the data is safe. It’s still on the array.

## COMMUNICATION

VFCache intelligence boils down to one word: Knowledge. How does one gather knowledge? Communication—integration with FAST. By extending automated tiering into the server, VFCache becomes a higher tier for the hottest data, drawing from the efficiencies FAST brings to data management. By recognizing and absorbing much of the read workload, VFCache takes load off the array, so the array can better handle write workloads. You will prolong the life of your array with this investment in servers, and you'll get higher performance from your data center. The result of all this cache-array communication is higher total performance in terms of aggregate IOPS, and much lower latency I/Os. It lets the data center continue with business as usual in managing storage, but gives a huge performance boost to all applications.

But, you might ask, is that where the intelligence ends? Isn't there cache on the array too? Isn't that redundant? Not necessarily—as VFCache intelligence evolves, the server and array caches will work even more closely together. Let's say you have a 300-gigabyte Flash up on a server and another 300-gigabyte cache down in an EMC VNX storage array. If you do nothing, the 300-gigabyte array cache would end up holding some of the same data that's in the 300-gigabyte cache up in the server, which is pretty inefficient. VFCache will soon be able to increase communication between the cache and the array to make it even more intelligent.

Possible intelligence integration points would include cache coordination. VFCache could tell the array what is in its cache. If VFCache decides to cache something, the array will be able to decide not to—reducing the number of copies in both the array cache and VFCache. Couple that with a similar communication, as with any cache, that when the 300-gigabyte cache is full and needs to make room for something new, it has to push out something old. So when VFCache tells the array what it's getting rid of, the array can drop it into its cache because the application might want it in the future. The combination of those two pieces of information—what VFCache is caching and what it's evicting—will help the two caches operate more efficiently than if they were running independently.

## SWEET SPOT

Who should use VFCache? Financial services. Retail. The information-centric, “time-is-money” people. A small boutique hedge fund for whom 1,000 or 10,000 more IOPS means they make \$10 million more in a day. You'll use it if...

- You're measured by how quickly you acknowledge an order and fulfill it.
- You want to give customers a two-second response instead of a five-second response so they won't click away from your website and buy something from somebody else.
- Your data set just continues to grow and the more business you do the slower things get.
- 80 percent of your IOPS are on 20 percent of your data.

When a fifth of your data generates 80 percent of your IOPS, you're in the VFCache sweet spot. And you can start to examine which parts of ERP or CRM or data warehousing or business intelligence or whatever have workloads like that would really benefit from VFCache. Maybe your transactional workload is getting high. The more you can get out of a single instance of a database the better off you are, but scaling out a database is no small feat. So it's a simpler solution to scale up your server by adding VFCache and increasing throughput by as much as five fold.

In a VMware environment where you're getting higher utilization from the processors by putting lots of small independent loads on it, you can drive more I/Os out of the system than would have been driven out of it otherwise. The VFCache software runs inside the virtual machine itself, managed by the hypervisor, so each virtual machine on a physical server would appear to have VFCache, but each would use only a portion of the PCIe card that's plugged into the physical server.

“When a fifth of your data generates 80 percent of your IOPS, you’re in the VFCache sweet spot.”

And regardless of the environment, because reads are captured by VFCache, applications will accelerate and users will see that response time is much improved. So they’re more productive. They can get more done because the system is more responsive. The IT department sees that the CRM system that used to be able to support only a hundred users can now support a couple hundred or maybe three or four hundred users simultaneously.

## THE FUTURE

VFCache is more than an intelligent card for a server, more than just an improvement to dumb disk. It is clearly the start of something. Servers are becoming powerful enough that some big questions have to be asked: Do storage systems start to become more like application services as well as storage services, or does a server, which is now just an application home, include storage services? Where in the vicinity of the middle will two ends of the spectrum meet?

In the near term, VFCache will become more and more tightly integrated with storage systems. Getting distributed storage management and coherency together, and achieving all the reliability that systems provide today is much harder than just providing a compute service that can go up and down at any time. That gives the advantage to the storage side. Then again, compute services are commodity priced while storage services have been premium priced, so we’re back to driving the extremes back toward each other.

One way or another, VFCache comes to market as an interesting part of a solution that dovetails into today’s storage systems. Phase one of its lifecycle is about a kind of creative tension between what the market can use today and what will fit in a new ecosystem and architecture. Phase two will be the next generation of that new vision coming to fruition, not just an updated VFCache with better speeds and feeds.

So VFCache is part one of a product family, but it’s also the first salvo in a new generation or architecture. VFCache asks the Big Question: How do we redraw the boundaries between memory and storage?

*This paper was written by EMC’s Terry Brown, assisted by EMC Flash Business Unit Senior Director Barry Ader and VFCache Engineers Dan Cobb, Bob Beauchamp, and Roy Clark.*

EMC<sup>2</sup>, EMC, the EMC logo, Symmetrix, VFCache, VMAX, and VNX are trademarks or registered trademarks of EMC Corporation in the United States and other countries. VMware is a registered trademark or trademark of VMware, Inc. in the United States and/or other jurisdictions. All other trademarks used herein are the property of their respective owners. © Copyright 2012 EMC Corporation. All rights reserved. Published in the USA. 02/12 EMC Perspective H9605