



## WHITE PAPER

# Why AFA Architecture Matters as Enterprises Pursue Dense Mixed Workload Consolidation

Sponsored by: Violin Memory

Eric Burgener  
July 2015

## IDC OPINION

---

All flash arrays (AFAs) have proven themselves able to meet the requirements of high performance dedicated application environments over the last several years, and enterprise customers are now planning to move more workloads to these platforms over time. Enterprise storage arrays have always been used for dense mixed workload consolidation, and as the cost of flash storage drops customers are asking themselves when AFAs can effectively assume the primary storage array role, replacing entirely systems that use hard disk drive (HDD) technology. The answer to that question lies in how quickly AFAs can cost-effectively evolve to accommodate the other key performance, endurance, availability and reliability, scalability, data services and integration requirements that an array running a dense mix of primary applications must exhibit. For enterprises, IDC recommends the following:

- Grasp the significance of the AFA market that is evolving into a new phase where the ability to densely consolidate mixed workloads defines the competitive battleground going forward
- Understand the compelling business and financial justifications for flash deployment at scale, and evaluate the applicability of an "all flash datacenter" strategy to your environment
- Know how to evaluate the value that different AFA architectures bring to the table, and how the architectures impact these systems' ability to efficiently deliver enterprise-class performance, scalability and manageability at a lower cost than HDD-based systems

IDC believes that ultimately, AFAs will become the enterprise storage workhorses of record for all tiers of primary applications, as the total cost of ownership (TCO) benefits of flash deployment at scale are just too compelling relative to HDD-based platforms. However, AFAs will need to meet the expanded set of requirements driven by dense mixed workload consolidation before this vision can become a reality.

## IN THIS WHITE PAPER

---

This white paper discusses the current state of the AFA market and how IDC sees this evolving in the future as AFAs start to dominate primary enterprise storage environments. Today, this market is in the midst of a major shift to a new set of requirements that will challenge all participating vendors to successfully navigate the transition. IDC will identify this transition and the requirements enterprises should be looking for in AFA products, discussing the business and technical benefits of these features. Finally, the paper will close with a review of the Violin Memory Flash Storage Platform (FSP) in light of these requirements, providing IDC's view on the company's vision and its prospects for future success selling into primary enterprise storage environments as the market evolves.

## SITUATION OVERVIEW

---

The use of virtual infrastructure is not only the present, but it is the wave of the future for at least the next decade. In a recent IDC survey, 77.3% of respondents have a "virtual first" strategy where new servers are deployed on virtual infrastructure unless a specific exception argument can be made. Enterprises are still supporting legacy applications like databases, messaging and collaboration systems, and file-sharing environments, while at the same time being asked to support entirely new applications for mobile, social media, big data and analytics, and cloud computing. Consolidating these disparate workloads on virtual infrastructure generates a workload profile that cannot be cost-effectively handled by HDDs alone, and flash storage is becoming ubiquitous as a result. In the same survey referenced above, 90% of virtual desktop infrastructure (VDI) and 67% of virtual server infrastructure (VSI) environments have some form of flash storage deployed in production today.

Enterprises' interest in the use of flash as a persistent storage tier led to the emergence of the AFA market. IDC first began tracking this market in 2012, and pegs 2015 revenue in this space at \$2.238 billion, and expects it to grow at a 22.5% compound annual growth rate through 2019 to \$5.05 billion. Initial products in this space were most often deployed for dedicated application use – for example, a database that demanded extremely high performance no matter what the cost. AFAs met requirements for very high performance and enterprise-class data integrity, but lacked capacity scalability and many of the storage management features that the enterprise storage workhorses of record had been delivering for years: snapshots, clones, replication, encryption, thin provisioning and APIs that enabled integration into pre-existing datacenter workflows.

Satisfaction with the performance of AFAs was very high, and as flash costs started to come down, most customers became interested in moving other workloads to them. There was some early discussion about the possibilities for an all-flash datacenter for primary storage as early as 2013, but this would not become a viable future unless AFAs were able to more cost-effectively replace legacy enterprise-class storage arrays. To do that, they would have to deliver not only flash performance, but also be able to provide a highly scalable, highly available solution specifically targeted for dense mixed workload consolidation that provided all the other benefits of traditional enterprise storage solutions. And these systems would need to deliver flash-optimized performance and reliable storage efficiency technologies to bring the cost of effective flash capacity below that of HDDs.

Many AFA vendors foresaw this shift from dedicated application to mixed workload consolidation deployment, and began enhancing their systems to meet the new set of requirements in 2014. Today, there are AFAs that support flash performance, provide effective capacity in the hundreds of terabytes (TBs) to petabytes (PBs) range, include inline data reduction technologies like compression and deduplication, support many storage management features (also called data services), and have APIs that enable integration with many of the leading virtualization, monitoring, management, and data protection tools and platforms. As legacy primary storage arrays come up for technology refresh, IDC strongly recommends enterprises to evaluate AFA options that include such capabilities.

## Business Justifications for Flash Deployment at Scale

Flash deployed at scale brings much more to the table than just high performance, and many enterprises have found the impact of flash not only transformational to their IT infrastructures, but also to their business processes. Here are a number of examples illustrating these benefits:

- A financial services company had recently merged with a competitor, and planned over time to integrate the IT infrastructure of the acquired company into its own datacenter, which was already operating at near capacity from a floor space point of view. After looking at AFAs, the company realized it could cut its IT infrastructure floor space by more than 30% while continuing to run the existing primary application workload at the same or better performance. This freed up enough physical space to consolidate the acquired company's applications into the same datacenter over time, allowing the financial services firm to avoid having to expand to additional datacenter capacity outside of what it already owned.
- An online gaming company wanted to introduce a new very high performance, cloud-based gaming environment where an online gamer received a dedicated virtual machine (VM) that was provisioned on demand. The company was concerned that with an HDD-based storage back end it would take too long to do all this online and on demand with the customer waiting. With flash in its infrastructure, the company could take a credit card number online, verify and charge the card, and spin the new VM up within seconds, enabling the company to provide this never-before-offered service on demand.
- A government agency that builds ruggedized mobile datacenters for short-term secure projects significantly improved the reliability and the performance of its equipment by entirely replacing HDDs with flash. This allowed the agency to extend the longevity of its IT infrastructure, which it looks to share across multiple projects in order to drive much lower costs. Power, cooling, floor space, and weight requirements were also reduced on a per-project basis, driving evolutionary changes in the agency's mobile datacenter design that also lowered its costs.
- A large software development company moved to an all-flash solution for its DevOps environment to take better advantage of a high-performance and very scalable snapshot technology that was specifically architected for the AFA. By being able to use very space-efficient yet high-performance snapshots, the company was able to lower its storage capacity consumption, speed the time to set up and deploy new testing cycles, and increase the amount of parallelism the company could leverage in its QA and other testing regimens. This allowed the company to shorten the time to market for new releases as well as for updates to address support issues, making its developers more productive and improving customer satisfaction.

Flash not only increases the infrastructure density, allowing more applications to be run in the same or less physical space at higher levels of performance, but also opens up the ability to implement new processes, generate additional revenues, and pursue new market opportunities not before possible.

## Financial Justifications for Flash Deployment at Scale

Historically, the most common objection to considering flash storage was the cost of raw flash storage capacity. CFOs would look at the cost of the performance-intensive storage of the past (15K RPM HDDs) on a \$/GB basis, compare it to the \$/GB cost of a flash-based device, and conclude that few applications could justify that big of a cost differential. Today, the cost of a 15K RPM enterprise-class HDD is roughly \$.80/GB or so, and there are several AFA vendors that will sell their raw flash capacity at under \$5/GB (some are more expensive). The use of \$/GB is a familiar metric that has been in use for decades to evaluate HDD-based storage options, but the fact is that it is NOT a relevant metric when comparing HDD-based systems to flash-based systems that use SSDs or custom flash modules (CFMs). Note that

a CFM denotes a different packaging of NAND flash technology onto a printed circuit board (PCB) which has some different and positive implications in the areas of performance, scalability, density and cost that will be explained later in the paper. Here's why the traditional metric is not relevant:

- A 15K RPM HDD can deliver generally in the range of 200 input/output operations per second (IOPS), whereas the typical SSD can sustainably deliver anywhere from 10,000 to 20,000 IOPS in real world usage. To configure an HDD-based system to deliver 500,000 IOPS (a common requirement in today's virtual datacenters) would require 2500 HDDs, but it would require only 50 SSDs (assuming they deliver only 10,000 IOPS each). HDDs are cheaper on a \$/GB basis, but you would need to buy 50 times more of them to meet this same performance requirement. In general, enterprises that replace an HDD-based system with an AFA will buy 50% to 80% fewer devices, even after they ensure that they have enough SSDs to meet their capacity requirements.
- Disk drives (HDD or SSD) need energy for power and cooling, and they take up rack space. Flash-based systems consume significantly less of both since so few devices are needed to meet requirements. There will be a significant decrease in the energy and floor space consumption requirements associated with the purchase of an AFA assuming it is used to consolidate multiple workloads.
- Latencies on 15K RPM HDDs are generally in the 5 to 20 millisecond range, while SSD latencies are consistently sub-millisecond – generally at least 10x lower latencies. Because HDDs are so slow, today's high performance x86 CPUs spend most of their time waiting for the storage to respond, but with flash CPU utilization generally goes up significantly, often as much as 2x or more. As each CPU does more work, fewer CPUs are needed and if configurations are large, fewer servers are needed. When flash is deployed at scale (at least 80TB of raw capacity or so) to replace legacy HDD-based systems, enterprises generally need 5% to 30% fewer x86 servers. As a result, customers save on server expense.
- As interesting as needing fewer x86 servers is, that is not as compelling as the software license cost savings associated with needing fewer servers. Expensive software products like Oracle, Exchange, SQL Server and others generally cost more than the servers they run on; the savings of licensing these products on fewer cores and/or servers also results in significant savings.

IDC refers to these savings as the "secondary economic benefits of flash deployment at scale." When customers purchase a small AFA for a single application, these benefits do not kick in. But as the deployed flash capacity increases – as it will when customers use these arrays more and more for dense mixed workload consolidation – these cost savings come to dominate the TCO calculations. It is critical to understand that dense workload consolidation on AFAs is the economic enabler of the all-flash datacenter strategy.

But the economic benefits of flash deployment at scale do not stop there. With its extremely low latencies, flash enables the use of many storage efficiency technologies that cause unacceptable performance impacts on HDD-based systems. Historically, inline data reduction technologies like compression and deduplication were rarely used in primary storage environments because of these impacts. But flash latencies allow these technologies to be used while still consistently providing sub-millisecond latencies. With the mixed virtual workloads dominant in most enterprise environments today, customers are seeing high rates of data reducibility and average data reduction ratios across a mix of multiple workloads in the 4:1 to 6:1 range. Data reduction ratios are very dependent on the specific workload, so for more information see *Evolving Flash-Optimized Storage Architectures*, (IDC #256994, June 2015). But even if we assume only a 4:1 data reduction ratio and flash at \$5/GB, the effective cost of flash capacity *at acquisition* is \$1.25/GB. That is getting very close to 15K RPM HDDs at around

\$.80/GB – and note that this does not yet take into account the secondary economic benefits of flash deployment at scale.

To accurately evaluate the \$/GB acquisition costs of an AFA, the data reduction ratio should be taken into account, and the comparison should be around effective \$/GB, not raw \$/GB. Note, however, that effective \$/GB does not take into account data protection or provisioning overhead, and these do impose capacity overhead that decreases the usable capacity. Both overheads are generally a static percentage, though, that can be applied to a raw \$/GB number of either flash or HDDs to help get to usable and/or effective capacity, so it's important to take them into account.

There is one final issue to consider. Most AFA vendors bundle the data services software with the purchase price of the array, while most legacy arrays are still charging separate license fees for snapshot, clone, replication and other features. This can add up to tens of thousands of dollars of additional savings with AFAs on even midrange array configurations.

## Mixed Workload Support Requirements in the Enterprise

Enterprises have different application tiers, each of which have different performance requirements. Flash initially gained a foothold in the datacenter with tier 0 applications – those applications that needed additional performance regardless of cost. But as the effective \$/GB cost of flash has come down, it becomes cost-effective to host more primary workload classes on flash, even those that are not as performance-intensive. When the effective \$/GB cost of flash capacity approaches the effective \$/GB cost of HDDs, enterprises should consider moving most workloads to flash to take advantage of the secondary economic benefits of flash deployment at scale. To do so effectively, they need AFAs that offer the same features as the enterprise storage workhorses of the past in a flash-optimized platform. In this section, we'll review performance, endurance, availability, scalability, data services and integration requirements, explaining the nuances of evaluating those capabilities on AFAs.

### Performance

Server consolidation has been a great boon to increasing CPU utilization in datacenters, but it significantly increases storage performance requirements. In the past, physical servers running dedicated applications at relatively low rates of CPU utilization demanded generally no more than tens of thousands of IOPS from shared storage systems. Today, dense server consolidation on virtual infrastructure requires hundreds of thousands to millions of IOPS. And the I/O coming out of virtual hosts is significantly more random than I/O in the older environment, demanding lower latencies from storage to maintain performance in these environments. HDDs are notorious for generating low IOPS in random environments. Flash, however, with its extremely high throughput and very low latencies in both random and sequential environments, is a great match to deal with these requirements, and it is a major reason why flash has penetrated datacenters so quickly over the last few years.

Given this, performance is clearly a key purchase criteria. While AFAs in general can deliver hundreds of thousands to millions of IOPS and are capable of delivering sub millisecond latencies, if an array will be used for dense mixed workload consolidation then its ability to *consistently* deliver those latencies across its entire throughput range with varying workloads becomes important. Predictable performance is easy to achieve when AFA utilization rates are low, but as an AFA becomes more heavily loaded, certain design decisions will impact its ability to continue to provide consistent latencies. Free space management, often colloquially referred to as "garbage collection," must scale, delivering consistent read and write performance regardless of what else is happening in the system.

AFAs in the market today are designed around two distinct types of storage media: CFMs and SSDs. These two media types, although they are both flash-based, handle garbage collection very differently in ways which may significantly affect a system's ability to deliver consistently low latencies under heavy load. If possible, AFAs should be tested with real-world workloads (NOT the "hero" tests of old) across their entire throughput range. Technical audiences can refer to the *AFA Performance Testing Framework*, (IDC #251951, October 2014) for detailed information about why this is true and how to model today's typical virtual workloads. If an enterprise plans to deploy dense mixed workloads on an enterprise array, it is important to understand how that array will perform across its entire range against particular workloads. Hero tests that provide a throughput rating (in IOPS) based on 100% 4K random reads are misleading to IT managers attempting to understand how an AFA will perform in their environment against their mixed virtual workloads.

If you plan to use inline data reduction (and IDC recommends that you should unless there are business or regulatory proscriptions against it), then you will need to understand not only how many IOPS the AFA can handle on the front end (the host connection side) but also how many it can handle on the back end (the connection to the storage media). Block sizes vary with modern workloads, and can be as small as 4K or as large as 256K (or more). If the AFA uses a chunk size of 8K for deduplication, then a 32K I/O coming in the front end will actually generate 4 I/Os on the back end. On the back end, then, an AFA will need to support some multiple of front end IOPS to handle the additional load generated by deduplication, data protection (RAID or replication), and other required tasks that the system performs transparently. Ensure that a system can perform all of these types of inline operations while still meeting both read and write latency requirements at scale.

A related consideration is degraded mode performance after the failure of a node (if working with scale-out or clustered configurations), controller or disk (or CFM). A redundant controller design is a requirement for availability, but some are implemented in an active/active configuration while others are active/passive. The point is not that one is better than the other; the point is to understand the implications of each approach. Either case can be appropriate, as long as performance requirements are met and the customer understands the implications of the design and the impact on performance of a controller failure. Degraded mode performance may not be as big of an issue when an AFA is minimally utilized, but it is important to understand the impact of these types of failures on a heavily loaded system as well.

Finally, in dense mixed workload environments, there is always the possibility for the "noisy neighbor" problem to rear its ugly head. This is the concern that bursts of activity in one application can impact the performance of other applications running on the same array in unpredictable ways. More and more enterprises are managing to defined service-level objectives (SLOs) for their customers, and this is one of the factors that makes the performance consistency of AFAs very attractive.

## **Endurance**

Flash is a consumable resource. After a defined number of writes (referred to as program/erase cycles), flash will become read-only and ultimately inoperable. How an AFA manages writes and in particular the write amplification generated by garbage collection, flash management, and data protection determines the rate of flash media consumption and ultimately the usable life of the array. Most AFA vendors have refined how they handle writes to the point that, despite the fact that they are built using less expensive MLC NAND flash, many will replace any failed device for free for any systems under maintenance during a five year period.

Features to look for include a flash-optimized data protection implementation that provides the same as, or better protection than RAID 6 (support for two concurrent drive failures with no data loss) while minimizing writes through to flash. Systems that minimize writes by write coalescing, eliminating transient I/O, performing various data services (snapshots, clones, thin provisioning, etc.) in a very write-efficient manner, or inline data reduction all contribute to higher endurance. The efficiency of garbage collection can also contribute to flash endurance, and efficient wear-leveling ensures that write-driven wear is evenly distributed across all flash in the array. Enterprise storage life cycles are on the order of four-to-five years for most customers, and AFAs will need to meet or exceed that to be considered as viable primary storage platforms for dense mixed workload consolidation.

## **Availability**

Availability requirements have only increased in today's datacenters, with many Web-savvy customers expecting 24 x 7 operations on application services. Even with the many capabilities in virtual infrastructure to manage VM availability, a primary enterprise-class storage array must support at least five nines (99.999%) availability (roughly 5 minutes of downtime a year) if not better. This means that systems must include redundant, hot plug everything, enable on-line firmware upgrades, offer data protection that can ride through multiple simultaneous storage media failures without data loss or application impact, provide replication to enable disaster recovery (DR) configurations, support online capacity expansion and reconfiguration, and be able to accommodate failover requirements. Advanced features that minimize recovery point and recovery time objectives (RPO/RTO), such as continuous data protection (CDP) and stretch clusters, can be very valuable for mission-critical applications.

## **Scalability**

Mixed workload consolidation obviously requires greater capacity in most cases than a dedicated application deployment. Size a system by taking into account any pertinent data reduction ratios, data protection overhead, and capacity expectations due to business growth, snapshot and clone utilization, or the impact on capacity due to the use of other data services (replication, thin provisioning, etc.). Mixed workload consolidation can bring more efficiencies on a larger scale, but enterprises also need to be aware of the size of fault domains in their primary storage environments. Generally, a system should support at least several hundred TBs of raw capacity. Take into account how a system scales. Can controller performance be increased over the lifetime of the system as more capacity is added? How is capacity added? Systems that can add more disks into a shelf, more disk shelves, and offer the opportunity to cluster to hit higher throughput and capacity levels offer more flexibility. When considering capacity, be sure to work with effective capacity projections that take storage efficiency technologies like deduplication and others into account and have been adjusted to allow for data protection overhead.

When evaluating scalability, if floor space utilization is a concern also take storage density into account. Although new introductions by AFA vendors are always increasing storage density, IDC has noted that vendors that use CFMs have tended to support higher densities than those that use SSDs.

## **Data Services**

When an AFA is purchased for only a single application, customers most often use the data services in the application itself. For example, when a high performance Oracle database is placed on an AFA, Oracle's facilities for volume management, compression, encryption, replication, and failover are often used. When multiple workloads are consolidated onto an AFA, many enterprises prefer using array-based utilities in a consistent manner across all workloads. Issues like manageability, flexibility and how the data services perform at scale can become important considerations.

Snapshot and clone performance is a case in point. On legacy HDD-based arrays, performance tended to degrade rapidly as more snapshots and/or clones were created, retained, and/or used. These performance impacts were so great that administrators used these tools only sparingly in performance-sensitive environments. But some AFAs implement new, flash-optimized designs that allow extremely rapid creation, long term retention of very space-efficient copies, and the simultaneous use of hundreds or thousands of snapshots and/or clones without performance degradation in either the source or target volumes. Once administrators trust a system to deliver this order of scalable high performance, this opens up the opportunity to modify existing workflows in provisioning, data protection, recovery, and test and development operations that can have a materially positive business impact.

If tasks like compression can be offloaded from the hosts to the array, this can help reduce the CPU count and significantly lower software licensing costs. Performing a CPU-intensive task like compression on the host often requires customers to purchase additional CPU cores and in some cases additional servers. Each of those must be licensed to use the application. By offloading to the array, customers can generally always drive the same level of application performance with fewer CPUs and pay expensive software licensing fees on fewer cores.

Most primary applications require some form of DR. Mission-critical applications may require that DR be managed to very stringent RPOs and RTOs. When this is the case, generally some form of replication is required. Synchronous replication will keep two arrays in perfect lockstep data-wise, but the network hop to the remote system can negatively impact application performance. Asynchronous and periodic, snapshot-based replication can be used to keep two arrays in very close synchronization without impacting primary application performance. Some systems support a form of metro or stretch clustering that can be built on top of synchronous replication, providing immediate and fully transparent recovery from a primary site failure with no application service impact and no data loss. AFAs that offer all options provide the most flexibility in allowing customers to configure the DR solution that best meets their requirements for performance, recovery and cost.

## **Integration**

As mature products, today's enterprise-class storage arrays offer many integration points with existing infrastructure and applications. As data continues to grow at 45-50% per year for most enterprises, administrators are increasingly turning to automation to help them more reliably manage their environments. All AFA features and functions should be programmatically accessible via either a command line interface (CLI) or through external APIs, making it easier to leverage them to establish or improve datacenter workflows. Explore the ability of any candidate AFAs to be integrated into batch, monitoring, management, data protection, and DR workflows using standardized APIs.

## **Violin Memory and the AFA Market**

In 2009, Violin Memory introduced the first storage array that consistently used flash as a persistent data storage medium. Initially targeting dedicated application deployment in the enterprise, Violin positioned its products primarily for use with high performance databases and other tier 0 applications that required extremely high performance. In this first wave of AFA market development, enterprises were driven almost exclusively by performance requirements regardless of cost. As a first-mover with an extremely flash-optimized architecture that used CFMs instead of SSDs, Violin did extremely well in this wave of market development, which lasted until about 2012, amassing a number of bellwether Fortune 500 customers.

As customers started to become more comfortable using flash in enterprise production environments, many wanted to move more applications onto it. Price was the key obstacle keeping customers from deploying more flash, and vendors responded by integrating storage efficiency technologies like compression and deduplication into their systems to reduce the effective \$/GB cost. While the generally very reducible data sets in virtual environments can experience data reduction ratios of 10:1 or higher, customers are consistently achieving data reduction ratios in the 4:1 to 6:1 range across mixed workloads in common use. Violin did not correctly anticipate the need for inline data reduction, and stumbled a bit during this phase that lasted into 2014.

As the AFA market continued to develop and the effective \$/GB cost of flash closed in on that of 15K RPM HDDs, more customers began to think strategically about the use of flash across all primary application environments. Awareness of the magnitude of flash's secondary economic benefits and the TCO comparison with HDD-based arrays was increasing, and this drove entry into the next wave of market development. Customers needed higher capacity systems that offered the data services and integration points to be able to effectively replace primary storage arrays that were being used for dense mixed workload consolidation across multiple primary application tiers, and vendors responded. During this second phase, Violin pivoted its strategy and entered into an intense period of R&D to produce an AFA platform specifically designed to become the enterprise storage workhorse of the future. This product, called the Flash Storage Platform (FSP), became generally available in February 2015.

In this third wave of AFA market development, customers will be evaluating AFA offerings on their ability to support densely consolidated mixed workloads, and this will be the competitive battleground for vendors in this space going forward. Customers should be evaluating AFAs on their ability to cost-effectively replace legacy HDD-based systems that are running all or most of an enterprise's primary workloads across multiple tiers, and should expect these systems to drive lower TCOs based around the secondary economic benefits of flash deployment at scale. IDC concurs with Violin's view that AFAs will ultimately become the enterprise storage workhorses for primary workloads, and believes that by 2019 it will be a rare customer that buys an HDD-based system for primary storage.

For vendors, the game changes significantly with each new phase, and success in a prior phase does not necessarily guarantee success in the next one. Customers will still buy some small AFAs for dedicated application environments, but are well advised to consider the ramifications of an "all flash for primary storage" strategy even as they do that. Many enterprises are already pursuing such a strategy but are nevertheless letting the technology refresh cycle for legacy equipment determine how quickly they will move in that direction. A small number of brand name customers have in fact already moved most of their primary application environments to AFAs. Violin has identified the mixed workload consolidation requirements of this next wave, and closely matched the capabilities of the FSP to them, positioning itself well against the competition going forward.

## The Flash Storage Platform (FSP)

Violin's FSP is an all-flash primary storage array designed to run any and all primary applications with flash-optimized performance and enterprise-class availability and reliability. Building on an operating environment (Concerto OS 7) that Violin has been shipping for more than a year, the FSP product line can provide over 2M IOPS at sub-millisecond latencies, scales to support well more than two PBs of effective storage capacity today, offers availability features that rival those of the best enterprise-class primary storage arrays, and includes a full complement of data services that are integrated into a single operating environment. The FSP's data services include inline compression and deduplication, as well as other storage efficiency technologies that deliver a lower effective cost per GB than

comparably configured HDD-based systems, making good on Violin's claim to deliver a flash-optimized, enterprise-class storage platform for primary environments at an effective cost below traditional disk.

Many of the technologies culminating in the FSP have been proven out in production use during the last six years across Violin's more than 500 customers which include bellwether names in financial services, healthcare, e-commerce, telecommunications, government, insurance, education, and manufacturing. It is worth noting that those Violin customers that are already well on the way to the all-flash datacenter for primary storage view their strategy as a competitive differentiator, and do not like to discuss it publicly to so as preserve this advantage.

Violin is one of only two vendors in the industry who ship a primary storage AFA based entirely on CFMs rather than SSDs (IBM is the other vendor). Instead of buying SSDs off the shelf, Violin buys NAND flash directly from Toshiba to build Violin Intelligent Memory Modules (VIMMs). Each VIMM is a field replaceable unit (FRU) that today provides 1.1TB of raw flash capacity, and this dense packaging supports up to 23.4TB/U in a compact 3U system. CFM-based designs offer advantages over most SSD-based systems in terms of storage density and in their ability to deliver consistent performance as a system is scaled across its entire range of throughput. While these issues were not as important when smaller capacity AFAs were purchased for dedicated application use, they are extremely important for customers considering AFAs as direct replacements for their primary enterprise storage arrays that will deploy flash at scale.

### **Flash Fabric Architecture**

IDC has written recently about flash-optimized storage architectures and how these designs specifically impact the performance, resiliency, scalability and density provided by AFAs targeted for use in enterprises for mixed workload consolidation (see *Evolving Flash Optimized Storage Architectures*, IDC #256994, June 2015). Violin is the only vendor that has built an entirely integrated hardware, firmware and software-based platform targeted for use with primary enterprise applications. IDC believes that, going forward, this unique approach will differentiate Violin from other players in the industry in terms of performance, scalability, density and cost unless those players pursue a similar product strategy.

### **The Benefits of CFMs**

Instead of packaging flash media into a form factor that can be accessed just like a standard HDD, CFMs pack flash media onto a PCB. A flash translation layer (FTL) enables transparent access for all existing applications to the storage capacity on the board. Most vendors today use SSDs in the 1TB or 2TB class (Nimbus Data actually uses a 4TB SSD), but at the system level vendors that use CFMs generally achieve higher densities. Both Violin and IBM deliver industry-leading storage densities on their systems (23.3TB/U and 28.5TB/U of raw capacity respectively), and these engineered PCBs are more efficient in terms of power consumption than SSDs. Higher densities also translate to more compact systems that require less rack and floor space.

The other key advantage of CFMs is that they enable the system to know the individual state of all flash cells in a system, and this information is used to schedule garbage collection and other flash management operations more efficiently and with better granularity than SSD-based systems can. When it comes to garbage collection, SSDs operate somewhat as a black box, providing state information at the device level (as opposed to the individual flash cell level). Garbage collection operates at the device level with SSDs, and this higher granularity makes for less efficient free space

management. When AFAs are operated at low levels of utilization (as many are today), this difference is generally immaterial, but when systems are operated at the higher levels of utilization that dense mixed workload consolidation will drive, this is expected to make a material difference in a system's ability to deliver consistent latencies across its entire throughput range. Violin's choice to use a CFM-based design bodes well for this future.

Violin's flash management layer interacts directly with the flash media without the added overhead of legacy protocols like SATA and SAS that SSDs use. This again supports better efficiencies, helping to deliver higher performance from less infrastructure than SSD-based systems can.

## **Performance**

Violin's Flash Fabric Architecture is a well-balanced design for an AFA with inline data reduction. Able to process over 2M IOPS on the front end, on the back end it can sustain over 6M IOPS. For systems that will be chunking large block I/Os into smaller sizes for deduplication, the system back end must be able to sustain a multiple of the front end IOPS to be able to provide consistent performance at higher levels of utilization. FSPs support either Fibre Channel (FC) or iSCSI host connection through multiple interfaces, and offer dedicated 10/40 GbE connections for remote replication.

A resilient mesh of thousands of flash dies, organized into VIMMs, delivers consistent flash latencies even as throughput scales and garbage collection operations are scheduled and occur. All data is wide striped across vRAID groups that consist of five VIMMs for balanced performance. A vRAID group is a protection unit that can sustain multiple flash cell failures without data loss, and it uses a patented, flash-optimized version of RAID that does not suffer any performance degradation in the event of even an entire VIMM failure. With Violin's efficient XOR implementation and flash latencies, data can be reconstructed from the remaining VIMMs in the vRAID group with no impact to latency. Only one VIMM is allowed to perform garbage collection at a time in each vRAID group, leaving the other four VIMMs able to service I/Os. Given the vRAID data layout, no data is ever inaccessible during garbage collection, and given flash latencies and their RAID controller assist, data reconstruction does not impact the system's ability to continue to deliver sub-millisecond performance. This is why Violin refers to its architecture as "non blocking."

Violin's need for garbage collection is minimized through a variety of write minimization techniques, including its flash-optimized vRAID implementation, inline compression and deduplication, and write coalescing. The system keeps hot and cold data together in the same pages as much as possible, and the "cold" pages are not moved as often to minimize garbage collection overhead. Through communication with the application, the RAID controller is able to eliminate data that is no longer being used from being rewritten into new pages. Both of these latter features also contribute to a minimization of garbage collection. Many (although not all) of these techniques can be performed much more efficiently when the system has visibility at the flash cell instead of just the SSD level.

## **Endurance**

Enterprise environments are write-intensive, and early in the development of the AFA market there were concerns about flash's ability to meet standard enterprise life cycle requirements of four to five years. Most AFA vendors have perfected a number of techniques to manage flash media endurance well beyond the five year mark. Violin in fact warrants its flash media for the life of any system under warranty, regardless of how old the system is. Violin employs a number of techniques to deliver endurance, including the write minimization techniques mentioned above, wear leveling techniques that spread the write load across each vRAID group evenly (wide striping), and scheduling data

movement during garbage collection operations to ensure that the write wear generated by "hot spots" does not hit the same cells over and over again. These methods are sufficiently effective that Violin offers a lifetime flash endurance guarantee on its cost-effective MLC flash for any system under maintenance – one of the contributing factors to Violin's very aggressive under \$1.50/GB price for effective flash capacity today (assuming a 6:1 data reduction ratio across mixed workloads).

Over-provisioning ratios impact the endurance of flash media, with higher levels of overprovisioning producing both higher endurance and higher effective \$/GB cost. Violin shares a global pool of over-provisioned capacity across the entire array, and the performance numbers cited earlier as well as the \$1.50/GB mentioned above both reflect the FSP's default over-provisioning level of 16%. Violin is rare among AFA vendors in allowing customers to easily change the over-provisioning level in the field to "tune" the system to handle more write-intensive workloads while maintaining enterprise-class endurance. Administrators can increase the over-provisioning level to as much as 50% if they want to but this will have an impact on the effective \$/GB cost of Violin flash capacity.

### ***Availability and Reliability***

Violin employs a flash-optimized RAID implementation (vRAID) that can sustain the loss of one entire VIMM per vRAID group or up to 16 VIMMs across the entire system without impacting performance or causing data loss. Individual flash cells that fail or are marked as "about to fail" based on monitoring are just replaced on the fly with new cells from the over-provisioning pool – it's extremely rare that an entire VIMM would fail. Each FSP contains four hot spare VIMMs though, which can immediately and transparently be swapped in to replace a failed VIMM if such a failure should occur. All VIMMs are triple ported to provide maximum accessibility and minimum latency in Violin's PCIe-based mesh (note that SSDs are only dual ported).

Violin uses a dual controller architecture, with each controller being the primary owner for a designated set of LUNs in the array. In the event of a controller failure, there is no disruption of application services since LUNs owned by the failed controller are transparently switched over to the remaining controller, but there can be a performance impact. Each VIMM can drive a maximum of around 200K IOPS, but the 2M IOPS number quoted earlier assumes that each VIMM in such a fully loaded system is on average only providing roughly 40K IOPS (a maximally configured FSP supports up to 60 VIMMs and four hot spares). In a maximally loaded FSP, a controller failure can impact overall performance by up to 40% (worst case) but the impact will vary based on load. Systems loaded at lighter than maximum levels will see a smaller impact, and in cases where the system may only be operating at up to 60% utilization, no impact. Controllers are FRUs which can be quickly and non-disruptively replaced in the field.

Replication is a requirement for a primary storage system that will host multiple workloads, and Violin supports both synchronous and asynchronous replication options. Synchronous replication can be used to keep data in perfect synchronization in two locations to enable recovery from the loss of an entire system at one site without any data loss. Violin supports stretch clusters that use this synchronous replication capability to recover immediately and transparently without any application impact or data loss from a primary system failure as well. Note that uniquely in the AFA market, this stretch cluster capability does not require any additional products, resulting in a simpler, more cost-effective zero data loss solution than competitors that require external or third party products. Asynchronous replication can be used to set up DR configurations that can span longer distances than synchronous replication can realistically support (due to network latencies and the associated application performance impacts). Violin also substantially reduces the amount of IP bandwidth required for replication through the use of multiple technologies including deduplication, compression and WAN optimization.

Uniquely Violin supports a CDP capability as one of the data services on the FSPs. CDP journals all writes to a log, and allows customers to retroactively select any point in time within the log to create a snapshot which can be used for recovery or other purposes. Log size determines the length of the CDP window. A common way in which CDP is used is to allow an administrator to create the optimum recovery point to resolve a data corruption problem that minimizes data loss after the fact (a snapshot can be retroactively created at any point before the data corruption occurred and used for recovery). This recovery capability can also be used in stretch clusters on one or both arrays.

When RPO is a critical concern, CDP is probably the most space-efficient and simple approach to keeping many granular near term recovery points available (over the last several hours, depending on write volume). Violin also supports space-efficient snapshots that can be instantly created and large numbers of them (up to 1,000 per LUN) can be retained without any performance impact. Snapshots can be used for off-host backup or other administrative purposes, or they can be turned into clones to feed a variety of administrative operations that require read/write copies. They are also a more efficient way to retain less granular recovery points over longer periods outside of the CDP window.

Violin has also implemented specific data integrity enhancements. There are multiple CRC checks to ensure that there is no silent data corruption on an initial write. Data in flash cells is monitored to ensure that even cold data is not retained in the same cell too long – older data is scheduled into garbage collection as necessary to address this issue. The flash and RAID controllers cooperate together, performing a continuous background process that scrubs data, evaluating data integrity and performing error correction as necessary. This is the process that also looks for flash cells that may be about to go bad (mentioned earlier). When these are found, that information is given to the flash controller and those cells are proactively removed from use during garbage collection operations (the flash controller schedules all free space management operations).

Finally, all components are hot pluggable and field replaceable – controllers, CFMs (VIMMs), power supplies, and fans. All firmware, including VIMM and controller firmware, is upgradable without disrupting application services. Systems can be expanded and/or reconfigured online without causing any application downtime. The FSP is designed for five nines plus availability, and has been delivering at least that in the field to date (Violin monitors all deployed systems for uptime data among customers that allow this).

## ***Scalability and Capacity***

Interestingly, Violin provides performance and capacity numbers for their systems both in raw and usable numbers – one of the few vendors to do this. Their data sheets show raw capacity, usable capacity (after adjusting for formatting [over provisioning] and RAID overhead) and effective capacity (assuming data reduction ratios of 6:1, which may admittedly be a bit on the high side for many mixed workloads). The FSP's GUI Symphony, however, displays actual achieved data reduction ratios.

There are two FSP models: the 7300 and the 7700. Note that both are dual controller systems, with the 7300 using less powerful controllers. The 7300 is a smaller platform with lower entry price points, supporting up to 1M IOPS with 250 microsecond latencies with data reduction turned off and up to 250K IOPS with sub-millisecond latencies with data reduction (defined as deduplication) turned on. In a 3U form factor, the 7300 packs up to 70TB raw, 44TB usable, and over 200TB effective. 7300E models provide lower performance and capacity at lower price points (still in a 3U package). The 7700 supports up to 2M IOPS with 350 microsecond latencies and data reduction off and packs up to 700TB raw into a 36U form factor. A maximally configured 7700 can accommodate well over 2PB of effective capacity (assuming Violin's standard adjustments for formatting and data reduction).

## *Data Services*

Violin has a single, integrated operating environment (Concerto OS 7) that manages all aspects of system functionality from the low level flash management all the way up through the data services. An integrated approach is simpler and more efficient than approaches which use separate operating environments, regardless of how they are managed, for the flash management and the data services. Today, Violin is the only CFM-based AFA vendor whose solutions are targeted at the primary storage market with this level of integration.

Violin offers a broad set of data services that include snapshots, clones, CDP, encryption, volume prioritization (for QoS), and synchronous and asynchronous replication. Violin also includes inline data reduction (compression and deduplication), selectable at the individual volume level, as well as other storage efficiency technologies such as thin provisioning and space-efficient snapshots and clones. This selectability is important in systems targeted for mixed workload consolidation, since some enterprises will have some applications where features like encryption and data reduction cannot be used due to regulatory or business restrictions. Systems that offer selectability allow customers to flexibly meet these requirements while still consolidating multiple workloads. Note that there is a performance difference between applications using data reduction and those that are not, even though both easily stay under the 1 millisecond latency parameter on FSP arrays.

## *Integration and Management*

The FSP's primary management is delivered through Violin's Symphony Management software, included standard in all FSP purchases. All capabilities required for the day-to-day ongoing management of the array are accessible through Symphony. Symphony is very popular among the installed base as it provides a single pane of glass through which Violin customers can manage all of their installed Violin arrays. Management tools like Symphony that provide a centralized point for multi-array management make the administration of large Violin installations simpler and easier. Symphony provides a real-time, enterprise-level view of a customer's Violin installed base with graphical representations of LUN types, capacity sizes, performance data (including historicals) and data reduction results, to name just a few. Symphony is built on top of a REST API framework, which is also accessible to customers.

To fully address customer pain points, enterprise storage arrays are expected to have solution-level integration with applications prevalent in today's datacenters. Violin offers host agent support for many enterprise applications, including Microsoft SQL Server, Microsoft Exchange, Oracle databases, MySQL databases, VMware, and IBM DB2, that enable application-consistent snapshot creation to be integrated into data protection workflows. VMware integration points include VMware APIs for Data Protection (VADP) and VMware APIs for Array Integration (VAAI). FSPs can also be incorporated into DR workflows managed by VMware Site Recovery Manager (SRM) or can be deployed with VMware vSphere Metro Storage Cluster. Support for SNMP allows the FSPs to come under the unified monitoring umbrella of a number of different systems management platforms available in the market. To address enterprise customers interested in private cloud infrastructures, Violin also supports the Cinder service within the OpenStack IaaS architecture.

## **FUTURE OUTLOOK**

---

When markets go through periods of rapid evolution that introduce new requirements, not all vendors successfully make the transition to the next wave. In IDC's view, AFA vendors that understand the

broad requirements for primary storage and dense enterprise workload consolidation and build their platforms to meet these requirements are the ones that will succeed in this third wave of market development going on now. Their platforms are the ones that the future "all flash datacenter" will leverage, ultimately moving IT infrastructure to a higher level of density and therefore performance.

IDC has discussed the concept of IT infrastructure density, along with the IT infrastructure balance point model, in *Managing the IT Infrastructure Balance Point for Improved Performance and Efficiency*, IDC #256754, June 2015. Briefly, the document summarizes a model whereby the four major components of IT infrastructure – processors, storage, networking and applications – should stay within a range of balance relative to each other, defined as the efficiency zone, even as the performance capabilities of each increase over time. As each of the four components evolves, the IT infrastructure balance point changes. IT organizations should strive to maintain the balance point between the four components within the efficiency zone to achieve higher utilization and therefore better performance. Over the last decade, the significant lag in storage performance, limited by spinning disk technology, has moved the balance point way out of the efficiency zone. It is the performance of flash that has allowed IT organizations to re-position the balance point back within the efficiency zone, and it is one of the key reasons for the extremely rapid adoption of flash storage technology in the datacenter.

Violin's architecture includes the features enterprises looking to densely consolidate mixed workloads require. Their choice of a CFM-based design allows them to achieve efficiencies that SSD-based systems may not be able to when operating at the high levels of system utilization that dense workload consolidation will require. CFM-based architectures also appear to have a storage density advantage that will support lower floor space consumption and drive cost advantages. Although today's 4TB SSDs are roughly on par with the densest CFMs for primary storage use (IBM is at 5.7TB per board), it will be more difficult for drive vendors to increase this on the same schedule they have in the past, while vendors leveraging CFMs already have plans to double and even quadruple densities in the same footprint within the next 18 months. IDC expects the density advantage of CFM-based systems to increase over the next two years.

Simpler is better, and AFA vendors that have not already created a well-integrated design that uses a single operating environment to manage all aspects of the array and deliver all data services are moving in that direction. Violin is already there. At this point, almost all new Violin sales are of the FSP, so the point of product comparison enterprises should use for Violin is based around the FSP, not prior platforms.

As with most emerging technology vendors, enterprises often want to understand a vendor's financials to assess their continued viability. Violin has reduced expenses in each of the last six consecutive quarters and today, has \$130 million in cash in the bank with a burn rate of somewhere between \$10 million and \$15 million per quarter. As a public company they have the ability to raise additional funding if required through either share or bond holders. In calendar Q215, over 75% of the company's product revenue had already transitioned to the FSP where they have significant advantages over the competition, and in Q315 this is expected to move to 100%. The company is expected to meet and surpass its highest historical quarterly revenues by the end of this year (surpassing a \$100 million annual run rate). FSP product revenue has grown at over 100% quarter over quarter since its release, and the company has announced four Fortune 1000 wins in the recent quarter (Q215) and expect to close an additional three Fortune 100 accounts by the end of the current quarter (Q315). Violin also has a strong pipeline.

## CHALLENGES/OPPORTUNITIES

---

Violin has challenges to address – some imaginary and some real. Competitors have often in the past claimed that Violin's operating environment (Concerto OS 7) was based on FalconStor IP, intimating that a software layer built initially for spinning disk environments could not be efficient in all-flash configurations. The truth is that in 2010, Violin did license source code for FalconStor's software that gave them access to FalconStor's deduplication and replication engines. Steve Dalton, Violin's longtime senior vice president of engineering, knew from his 25 years of experience in the storage industry that building reliable engines to handle that functionality at scale was difficult, and wanted to spend Violin engineering time flash-optimizing deduplication and replication engines that had already been proven reliable in thousands of enterprise deployments.

Since that original one-time license agreement, Violin has built a heavily flash-optimized layer around the core FalconStor deduplication and replication engines – well over 80% of the code in Concerto OS 7 is Violin, not FalconStor, developed code that is specifically tuned for its CFM-based architecture. This strategy allowed Violin to get to market more quickly with reliable, scalable and mature deduplication and replication capabilities that were well integrated into the Concerto operating environment, all managed under a single unified management interface that can be accessed either through the Symphony GUI, a REST API, or the CLI. This is a good strategy and strong story – Violin needs to make its customers aware of this. Vendors that had (or still have) to spend time building their own deduplication and replication engines from scratch are behind Violin in terms of engine maturity and capability.

Today, the FSP provides significant IOPS at sub-millisecond latencies, and delivers overall performance headroom that effectively addresses the most demanding dense mixed workload consolidation requirements. There may eventually come a point at which more granular and definitive QoS capabilities are required to prevent "noisy neighbor" issues from arising even on extremely high performance and scalable AFAs. Violin should consider adding QoS capabilities to its system in the future to ensure it can continue to meet SLOs on an application-by-application basis at very high rates of system utilization. Such capabilities could include the ability to define minimum, maximum, and burst IOPS capabilities at the application level while enforcing admission control.

As more enterprises commit to an "all flash datacenter" strategy over time, Violin's architecture is already at the point where the market is moving. The FSP is ready for dense mixed workload consolidation today, while other vendors are still trying to resolve the last few issues in creating a fully featured, fully integrated solution. As the effective \$/GB cost of flash comes down and enterprises come to more fully appreciate the TCO advantages of all-flash configurations, the rate at which enterprises move to AFAs will accelerate. Although the information cannot necessarily be shared publicly, Violin already has a number of accounts that are well along the road to the "all flash datacenter." These enterprises and their stories provide proof points today that will become more common in the future, helping others to understand the real benefits of the move to all flash. If this information can be shared during the sales process, Violin prospects should definitely take the time to talk to some of these accounts. Violin has unveiled more of these accounts to IDC than other vendors have, and even though the number is small in the grand scheme of things today, it is an installed base that differentiates Violin.

## CONCLUSION

---

With the AFA market's move to a competitive battleground based around dense mixed workload consolidation, a new game is afoot. The rate at which enterprises can increase infrastructure density, remove time-consuming administrative tasks like storage performance tuning, improve reliability, and reap the secondary economic benefits of flash deployment at scale will be directly related to the real-world capabilities of their flash-based storage platforms. Those that evaluate vendors' product offerings based around past performance and superseded requirements will be doing themselves a disservice.

Today, Violin differentiates itself around three primary product areas: architecture, the maturity of its storage efficiency technologies, and the level of integration it has achieved. The company has chosen a CFM-based architecture, and is leveraging the advantages of the visibility this provides at the individual cell level to drive higher performance, better efficiency, and lower cost – all capabilities that will become extremely important as customers look to achieve consistent performance out of their AFAs even as systems scale to very high levels of utilization. The maturity and scalability of its deduplication and replication engines, two features critical to enterprises looking to perform dense consolidation of mission-critical primary workloads across multiple application tiers, are ahead of those of its competitors who had to build these engines from the ground up and have at most only several years of production run time against them. And finally, complex systems operate more efficiently when their major components are well integrated and operate in the right balance. Violin's strategy to use a single operating environment to manage all aspects of the FSP, enabling information exchange between well integrated and balanced components to drive higher levels of performance, scalability and efficiency, has reached fruition in the platform today, while other vendors discuss that kind of integration as a roadmap capability.

With a leadership team that is correctly focused on dense mixed workload consolidation as the true path to the "all flash datacenter," an enterprise-class array that delivers on those capabilities today, and strong financials, Violin should be on the short list of any enterprise looking to deploy flash storage for primary application environments at scale.

## About IDC

International Data Corporation (IDC) is the premier global provider of market intelligence, advisory services, and events for the information technology, telecommunications and consumer technology markets. IDC helps IT professionals, business executives, and the investment community make fact-based decisions on technology purchases and business strategy. More than 1,100 IDC analysts provide global, regional, and local expertise on technology and industry opportunities and trends in over 110 countries worldwide. For 50 years, IDC has provided strategic insights to help our clients achieve their key business objectives. IDC is a subsidiary of IDG, the world's leading technology media, research, and events company.

## Global Headquarters

5 Speen Street  
Framingham, MA 01701  
USA  
508.872.8200  
Twitter: @IDC  
[idc-insights-community.com](http://idc-insights-community.com)  
[www.idc.com](http://www.idc.com)

---

### Copyright Notice

External Publication of IDC Information and Data – Any IDC information that is to be used in advertising, press releases, or promotional materials requires prior written approval from the appropriate IDC Vice President or Country Manager. A draft of the proposed document should accompany any such request. IDC reserves the right to deny approval of external usage for any reason.

Copyright 2015 IDC. Reproduction without written permission is completely forbidden.

