The best storage products of 2016
It’s medal time—the winners of the 15th annual Products of the Year competition are...

SNAPSHOT 1
Security and collaboration power EFSS uptake

EDITOR’S NOTE / CASTAGNA
The storage world will be turned upside down in 2017, or not

CAPACITY GROWTH
The capacity growth survival guide

STORAGE REVOLUTION / TOIGO
‘Holistic’ data protection, a meme on the rise

SNAPSHOT 2
EFSS key to solving organizational challenges

HOT SPOTS / BUFFINGTON
Why you should leverage the cloud for data protection

NEXT MONTH IN STORAGE
Stay tuned for the solution to all data storage problems

READ-WRITE / TANEJA
Hyper-converged, not just for primary storage anymore
I PREDICT ... that the Chicago Cubs will win at least one more World Series over the next 108 years. If you don’t think I’m going out on a limb with that prognostication, think about how someone making the same prediction back in 1908 would’ve been laughed into Lake Michigan—after all, those Cubbies had Tinker to Evers to Chance and Mordecai “Three Finger” Brown on the mound.

We can meet up in a North Side Starbucks in 2124 to see how that prediction fared. Meanwhile, we have our hands full with a steady stream of 2017 predictions and self-serving surveys from data storage industry vendors that are just begging for some attention.

It’s an annual thing, as vendors try for a *vox populi* approach to marketing to persuade you that your peers—and you—are yearning for their technologies.

Here are some of my favorite vendor-generated divinations on the data storage industry, along with my probably much-less biased take on their predictions. As usual, no vendor names are revealed.

**DISASTER RECOVERY IS A DISASTER**

If you believe all the stats coming from data protection companies, you might as well kiss your company’s data goodbye. One report stated that “more than 90% of large businesses report major incidents occur at least several times a year, and nearly 60% report major incidents occur at least monthly.” That’s discouraging enough to make any backup admin consider a career change.

A somewhat cheerier but painfully obvious data storage industry prediction came from a leading backup app vendor: “In 2017, organizations will take ransomware more seriously and implement ways to rapidly identify compromised content and automate its recovery.”

Here’s another from the gee-no-kidding department of another major backup application vendor: “Threats from hacking, as well as the proliferation of botnets and malware (specifically ransomware) will keep IT managers up at night throughout 2017.” It looks like ransomware is shaping up as a major data protection theme for 2017.

Another vendor’s survey noted that 26% of responding
companies are using five to 10 different backup products. Maybe that is what’s making disaster recovery so confusing and hard to do. The survey seemed to drift into a game of liar’s poker with a question about how long it will take to recover from a server failure: 19% said less than an hour, while 55% said one to two hours. If that’s true, then I guess having major incidents occur every month is no big deal. One of the real shockers in this survey was the response to the question, “How critical is speed of backup and recovery of data for you?” Ninety-eight percent said it was very critical or important. I guess the remaining 2% were buffing up their resumes.

ALL-FLASH IN THE PAN

Data protection wasn’t the only data storage industry practice picking up predictions for the coming year. If it seems like flash has been the big story in storage for the past three or four years, it’s because it has. It’s hard to overstate flash’s role in the evolution of storage—but, of course, some vendors and storage swamis just can’t resist.

Take this bon mot from a leading analyst firm serving up a backhand slap to hard disks: “Anything in a data center that physically moves gets less useful and loses share of wallet.” I guess if you work in a data center you should freeze, or you’re liable to be considered “useless.” The flash phenomenon is beginning to soar to previously unimaginable heights, with Samsung’s 32 TB and Seagate’s 60 TB SSDs slated to ship in 2017. And you have to look high and low to find a pundit who might dare to say that all-flash arrays are still niche products. Perish the thought!

SOFTWARE IS THE NEW HARDWARE

Come on, data storage industry vendors, repeat after me: “I will stop saying ‘software-defined storage’ in 2017.” Please! But at least one WAN vendor is ignoring my plea with its prediction that “Everything becomes software-defined” this year. Of course, there are pretty strong arguments that everything has always been software-defined, but I get the point. It’s the old commodity hardware argument, and it’s a pretty good one except when good ol’ common off-the-shelf gear just isn’t up to the task.

WHAT’S REALLY GOING TO HAPPEN IN 2017?

The bottom line: Despite the threat of a new onslaught of disaster recovery surveys, DR is hard, and it’s likely to cost a company some cash. But you have to do it—and you have to make sure it will work when you need it.

Flash is wonderful. It’s also still pretty expensive, considering the alternatives. And it’s probably overkill for apps that just sip from the storage performance cup. That’s why hybrid systems will continue to be the most popular implementation for solid-state storage. And while it’s true that we’ve all turned into little data chipmunks, hoarding bits and bytes almost as fast as we create them, really huge and expensive flash drives will likely be appropriate for only the biggest big data prestidigitators around.

And sure, software will undoubtedly get better, but it just might need some help from some really cool hardware innovations once in a while.

RICH CASTAGNA is TechTarget’s VP of Editorial.
LIKE MANY FOLKS, I find myself put off by certain words that pop up often in advertisements. Holistic is one of those words. Usually associated with healthcare woo, holistic medicine probably started as a smart idea, instructing patients and providers to look beyond point solutions to consider the broader context for injuries or ailments. But somewhere along the way, it was seized upon by marketing folks to incorporate a lot of unscientific nonsense into claims about certain products or treatment modalities. A lot of today’s data protection marketing incorporates the “holistic” meme. Vendors are starting to disparage these so-called “point solutions” and offer suites of data protection tools under one brand name and management console. That’s holistic data protection, I guess.

HOLISTIC DATA PROTECTION
Of course, there is merit to the claims of an Arcserve or an Acronis about the need to have a free-standing instantiation of data protection services—set apart from services included in the software-defined stack that is siloed behind each deployed hypervisor and limited to storage systems operated by that stack. You can’t share capacity between software-defined storage delineated by different, and competing, hypervisor stacks, nor can you protect data residing on a “foreign” hypervisor’s storage. That is, unless you convert the workload and data into the native format of the protecting hypervisor.

Stand-alone data protection suite vendors want to address this issue by providing technology that’s blind to hypervisor silos and works with all kits and data, usually at a block level. Working with data at the block level anonymizes data and insulates us admins from infrastructure peculiarities. A block is a block. What makes such an approach “holistic” is this universality and also the ability to manage data protection processes from a single pane of glass. Throw in a mention or two regarding how comprehensive the data protection services are in a given software suite—or that they eliminate the need to buy continuous data protection (CDP) tools from one vendor, incremental snapshot tools from another and remote replication tools from a third—and you get another “holistic” brag to make.
MORE THAN DATA REPLICATION

Sometimes lost in the messaging is the fact that “holistic” data protection requires more than just data replication tools. Here are a couple of things to think about.

First, the data protection strategy you select needs to be granular. It has to match the specific protection and recovery requirements of data based on what that data is for and the business process it serves. For example, while mission-critical database transactions may require incremental and full snapshots and off-site replication to deliver an in-depth defense against logical and physical threats, this workload may also be a target for ne’er do wells who seek to encrypt that data and ransom it back. For that reason, you may also need CDP that can rewind I/O back to before the ransomware was introduced.

Conversely, archival data may not need much more than a weekly or monthly full backup because the data is seldom accessed and almost never updated. That said, depending on the nature of the archival data, it may be a candidate for encryption, another data protection service.

SECURITY: THE MISSING INGREDIENT

Given the different needs of different data, holistic data protection tools should include some means of characterizing and classifying data, as well as establishing a policy for protecting, preserving and ensuring the privacy of the data in a manner befitting its criticality or class.

While we are on the subject of security, late last year, Jeh Johnson, then Secretary of the U.S. Department of Homeland Security, identified phishing as the top threat to data. He cited events ranging from the WikiLeaks email disclosures during the election process; the 21 million employees and contractors who had their personal data stolen in the Office of Personnel Management hack; plus recent hacks at Sony Pictures, JPMorgan Chase & Co., eBay, Target and numerous banks worldwide. Moreover, roughly 97% of phish attacks contain a ransomware payload, according to a McAfee Labs Threats Report in 2015, reflecting what Hewlett Packard Enterprise Security Research identified in their Cyber Risk Report of 2016 as a trend toward the monetization of threats (malware writers seeking to produce revenue with their scripts, rather than just vandalize targets).

Given the increasing risk that it’ll be a security breach rather than a fire or flood that will take down your operating environment, shouldn’t holistic data management include security services? In most shops, and in most data protection tools, security is kept quite separate from disaster recovery (DR). About the only overlap is in the area of encryption and access control functionality, and neither of these are particularly effective against phishing or ransomware, by the way.

To call their wares “holistic data protection” tools, a vendor needs to cram a lot of functionality into a product. Perhaps one of the first areas that should be targeted for improvement is in the blending of security services with DR services. More to come.

JON WILLIAM TOIGO is a 30-year IT veteran, CEO and managing principal of Toigo Partners International, and chairman of the Data Management Institute.
Storage Products of the Year

Find out which 17 products from the past year are on the cutting edge of technology.

BY RODNEY BROWN, PAUL CROCETTI, GARRY KRANZ, SONIA LELII, DAVE RAFFO, CAROL SLIWA AND JAMES ALAN MILLER

THIS YEAR MARKS the 15TH edition of the Storage magazine and SearchStorage.com Products of the Year awards, which identify the best data storage products. The competition began with an invitation to technology vendors to nominate their products for consideration. Unlike other awards, we only recognize products introduced or significantly upgraded during the course of the previous year. This gives new or lesser-known products and startups a fairer chance against more established brands and products.

Technological innovation is a significant factor for our judges when adjudicating the nominees, but not the only one. A product wouldn’t do much to satisfy if it’s hard to use and implement! Our judges also graded on performance, manageability, how easily a product can be integrated and used, functionality and value. Our judging panel consisted of users, analysts and consultants, along with Storage magazine and SearchStorage.com writers and editors.

Categories included all-flash storage systems, backup and disaster recovery software and services, backup hardware, hybrid flash storage systems, server-based storage and SAN management tools.

Of the many products nominated in these six categories, only 17 earned gold, silver or bronze awards to become Products of the Year winners. We would like to congratulate each of the 17 winners on a job well done.
Zerto Virtual Replication 4.5

ZERTO, WHICH HAS earned a solid reputation for providing virtual machine-based disaster recovery, took the gold in the data protection software category with Virtual Replication 4.5.

“In many [disaster recovery as a service] DRaaS conversations, Zerto is the name to beat,” one judge wrote. “Combining its replication and recovery technologies with very smart and aggressive marketing/event presence, Zerto is becoming semi-synonymous with ‘DRaaS software’ as a category.”

Journal File Level Restore is a key addition to 4.5. The feature enables the restoration of files from a point in time seconds before a deletion, virus attack or data corruption.

Through compressed journaling of the changed blocks from the protected virtual machines (VMs), Zerto Virtual Replication 4.5 maintains a granularity of the data in increments of seconds up to the past two weeks.

The product is geared toward enterprises needing to protect multi-terabyte VMs with recovery point objectives of seconds, across hypervisors and clouds. It replicates block-level changes from the hypervisor, without performance impact.

Zerto Virtual Replication 4.5 allows customers to automatically replicate VM environments to a secondary site. The product can replicate between VMware and Microsoft Hyper-V environments, as well as between private cloud infrastructures and Amazon Web Services.

IT managers can restore specific files from a specific point in time without having to recover all of the VMs of the application. Journal File Level Restore applies to most common recovery uses—deleted files, file or data corruption, and an application or service outage because of configuration changes.

Zerto Virtual Replication 4.5 scored highly among judges for innovation and performance. Judges praised its features, including the journaling.

Zerto touted the product’s ability to recover entire sites, individual applications or VMs in five clicks for fully automated recovery and recovery time objectives of minutes.

Zerto Virtual Replication 4.5 is a stand-alone product. It is often used within industries such as healthcare and financial services that have long-term retention requirements.

Pricing starts at $750 per VM.
BACKUP AND DISASTER RECOVERY SOFTWARE AND SERVICES

Veeam Availability Suite v9

Veeam software won the silver award on the strength of the high-speed recovery provided by Veeam Availability Suite v9.

The vendor espoused the product’s **recovery time objectives of less than 15 minutes**, and proactive monitoring and alerting of issues before they affect operations. It features built-in VM replication with full failover orchestration.

**Veeam Availability Suite v9** works with any backup storage but tightly integrates with NetApp, EMC and Hewlett Packard Enterprise storage array snapshots to back up every 15 minutes without affecting performance.

Enhancements include the following:

- Scale-out Backup Repository that allows IT to create a single virtual pool of heterogeneous backup storage to which backups can be assigned.
- **Veeam Cloud Connect** Replication, which ensures availability of mission-critical applications with fully integrated, fast and secure cloud-based disaster recovery through a service provider.
- Veeam Explorer for Oracle provides transaction-level recovery of Oracle databases, including agentless transaction log backup.
- BitLocker, which automatically excludes deleted file blocks and useless files, enabling IT to save backup storage and network bandwidth to reduce costs.

Judges were impressed with the product’s functionality. “Veeam has parlayed its decade-dominance in VM protection into not only profitability and mindshare dominance in their core category, but an evolution of its product portfolio that far exceeds just VM backups,” one judge wrote.

**Veeam Availability Suite v8** won Gold in the Backup and Disaster Recovery Software and Services category of the 2015 Products of the Year.

Pricing for Veeam Availability Suite v9 starts at $1,150 per socket and requires a Microsoft Hyper-V or VMware vSphere environment and a Windows-based server.
BACKUP AND DISASTER RECOVERY SOFTWARE AND SERVICES

Datos IO RecoverX

Datos IO RecoverX is a multi-platform, multicloud, and infinite-scale data protection and recovery software product for next-generation workloads such as cloud applications, and nonrelational and cloud databases.

RecoverX combines backup, archive, replication and recovery software in one product. It provides semantic deduplication for storage space efficiency and it is purpose-built for nonrelational databases. It also enables scalable versioning for flexible backups and any-point-in-time recovery.

Datos IO promoted the product’s patented Consistent Orchestrated Distributed Recovery architecture, which is not dependent on media servers and transfers data in parallel to and from file-based and object-based secondary storage. The Datos IO architecture also delivers application-consistent backups in native formats on secondary storage, and it provides a consistent view of data even in the event of hardware failures. The application-consistent versioning feature allows users to recover data without running any database repairs, reducing application downtime.

Datos IO RecoverX targets a market that traditional backup and recovery products do not.

Judges scored RecoverX highly in innovation and performance.

Another judge said Datos IO charges premium pricing for RecoverX, but that’s because it pioneered the market of backup for nonrelational database management systems.

RecoverX can be deployed on servers on-premises, NFS-connected storage or in the public cloud (Amazon Web Services or Google Cloud Platform) using cloud native compute instances and object storage.

It is a stand-alone software-only data protection product, but it does support Apache Cassandra (2.0, 2.1), DataStax Enterprise (4.5, 4.6, 4.7, 4.8) and MongoDB (3.0, 3.2) as well as a wide variety of secondary storage options. Support for Amazon Simple Storage Service (S3) and Google Cloud storage is also available, with Microsoft Azure Blob Storage support on the roadmap.

RecoverX features a subscription-based capacity-tiered pricing model. The least-expensive configuration is a 1 TB subscription price for $5,250 per year.
**Unitrends Recovery Series**

**Though the Unitrends Recovery Series backup appliance, with its tiered flash storage, earned gold in the backup hardware category this year, the vendor began selling integrated appliances in 2003, long before the notion became popular.**

Recovery Series meets the category definition for its scalable family of high-performance appliances that offer end-to-end data protection and instant recovery. Judges ranked the product line high in ease of integration into the environment, as well as ease of use, manageability and functionality.

The appliances offer 1 TB to 182 TB of raw capacity, 8 GB to 256 GB of memory and two to 16 CPU cores. The appliances include a flash tier to improve backup performance and speed up recovery. The company first added flash to its Recovery Series in the enterprise-level appliances back in January 2015, an innovative step for backup appliance hardware at the time.

Unitrends appliances use small amounts of flash to improve performance. Along with the flash injection, the latest version of its appliances include a redesigned user interface that provides a dashboard view for managing backups, Recovery Assurance testing of backup, archive and replication data on physical Windows hardware, and RESTful APIs for integration with customer applications.

The product line also has built-in, adaptive deduplication that maximizes the use of backup storage. Unitrends bundles deduplication at no extra cost in the all-in-one appliance. It offers an optional pledge support plan for free hardware for life, with a free replacement appliance upgrade every four years.

Judges found the Recovery Series to be a solid product backup appliance line, giving it high scores for price and the updated user interface.

“A good story at a very good price,” one judge said.

Another added, “The key innovation was the update to the UI. It’s better than most of their competitors now.”

---

**BACKUP HARDWARE**

**Castagna: Will the storage world be turned upside down in 2017?**

**Toigo: ‘Holistic’ data protection, a meme on the rise**

**2016 Products of the Year**

**What’s powering EFSS uptake?**

**Capacity growth survival guide**

**EFSS solves a number of organizational challenges**

**Buffington: Why you should leverage the cloud for data protection**

**Taneja: Hyper-converged, not just for primary storage anymore**

**About us**
Actifio Virtual Data Pipeline 7.0

The Actifio Virtual Data Pipeline (VDP) 7.0 appliance is the latest iteration of a product that has expanded over the years to include more features and functionality. It uses data virtualization software that decouples data from storage, similar to how hypervisors decouple compute from servers.

VDP makes a gold copy of production application data, merges changes incrementally and creates point-in-time virtual copies. Actifio VDP 7.0 lets customers replace the numerous point tools that they use for backup, replication, array snapshots, dedupe, virtual machine databases, physical servers and the cloud.

Judges found the product expensive, but noted that it made up for the initial expense since customers would be consolidating multiple data protection point tools. They also noted that Actifio has some unique capabilities to store data in its native formats on object storage. VDP 7.0 allows data to be instantly accessed even from object storage in the cloud that keeps data in a native format. There is no conversion from proprietary backup of dedupe formats.

Other enhancements focused on lower recovery time objectives (RTOs) and recovery point objectives (RPOs). Upgrades include StreamSnap replication that drives off-site RPOs with incremental-forever capture, featuring compression, encryption, bandwidth throttling and multiple nodes of retaining recovery points.

VDP has streamlined transaction logs for Oracle and SQL Server. Another RTO enhancement is the Actifio Resiliency Director that delivers orchestrated and automated disaster recovery with an RTO of under an hour for virtual machine environments.

The new Actifio OnVault stores backups for long-term retention into decades if needed in the public cloud with object storage for Amazon Web Services, Google Cloud Nearline and IBM Cleversafe for private cloud.

Actifio VDP software can reside in one of two appliances. The first is the Actifio Sky, a virtual appliance that can run in vSphere, Hyper-V, Oracle Cloud and Microsoft Azure. It also comes in Actifio CDS hardware appliance for larger data center deployments.
Barracuda Backup 6.3

The Barracuda Backup 6.3 appliance stood out for its increased initial backup speed and recovery performance. The vendor claimed the new appliance is twice as fast as previous versions with large files and can be three times as fast for file server backups of small files with the Barracuda backup agent. It also upgraded the Barracuda 990 appliance from 36 TB to 48 TB of usable storage.

The latest version of the software includes cloud-to-cloud backup for Microsoft Office 365. Barracuda also introduced multistreaming in the latest version to complement existing recovery options for physical-to-physical, physical-to-virtual or virtual-to-virtual configurations. A new queuing system increases replication performance for high transactional environments with large amounts of replicated changed data.

Barracuda appliances include this software and built-in off-site replication to either Barracuda Cloud Storage or to another Barracuda Backup appliance in an off-site location or to external disk or tape. A single user interface can configure multiple site backups and cloud storage, so a storage expert is not needed on site.

Barracuda customers can get a cloud subscription for off-site vaulting for long-term retention for up to seven years. They can get an upgraded Barracuda 990 at no charge. The Instant Replacement program entitles eligible Barracuda customers to upgrade their hardware every four years.

For organizations that require longer retention or support for external storage, there is the Barracuda Backup Export Tool that’s designed to enable customers to export data to external disk or tape. Administrators have the option to export data to virtual tapes stored in Amazon Web Services S3 or Glacier for archiving by using an AWS storage gateway.
VirtualWisdom NAS Performance Probe

Virtual Instruments VirtualWisdom NAS Performance Probe is as named—a real-time, full-line rate performance and availability monitoring product for NAS infrastructure. That real-time feature is what Virtual Instruments claimed as its primary differentiator.

While Probe products have historically focused on Fibre Channel SAN monitoring, NAS systems also face performance issues. Those include things like rogue clients, noisy neighbors and bottlenecks in accessing metadata.

While real-time monitoring has been a staple of Fibre Channel systems, VirtualWisdom NAS Performance Probe brings that same feature to NFS traffic. Users can check on hundreds of different metrics the Probe provides for NAS monitors.

Virtual Instruments claimed NAS Performance Probe is vendor-independent and functions across domains. It comes in the form a 2U piece of network hardware or as a virtual appliance. The Probe supported NFS 3 out of the gate, with support for NFS 4, SMB and iSCSI scheduled for 2017.

One judge described VirtualWisdom NAS Performance Probe as “a high-end solution that can provide advanced analytics for IP-based NAS storage environments.” That judge also noted that it is so comprehensive in its analytics that it’s “likely overkill for many,” adding that for organizations with “mission-critical deployments on NAS, this will help enable better results.”

Previously, to get the kind of data Probe provides, a user would need to capture packet data and send it back to the NAS array vendor for analysis. Changing that to real-time information is just one of the reasons why VirtualWisdom NAS Performance Probe took gold this year.

(The core technology for VirtualWisdom NAS Performance Probe came to Virtual Instruments through its acquisition of Load DynamiX, which earned a gold medal in the 2013 Products of the Year, last March.)
Data Dynamics Inc. StorageX 7.8

Data Dynamics StorageX 7.8 is a policy-based tool for migration, consolidation, tiering and archiving of unstructured data in complex file storage environments. Data Dynamics claimed the software can reduce the time of an end-to-end file migration project and the total cost of storage implementation by 50%.

New features in version 7.8 include file security management that allows users to automatically map security identifiers from the source to the destination and visual data mobility planning. Another important new feature is a data protection module, which can replicate data from NFS or SMB storage sources. The module can support one-to-one or one-to-many replication topologies.

Prior to the upgrade to version 7.8, Data Dynamics primarily sold StorageX as a data migration software suite. With the new data protection features, Data Dynamics now considers it a more complete data management tool.

With the addition of the new features, one judge said StorageX 7.8 solves “several specific problems with an easy-to-manage approach that’s well-suited to large environments.” While that judge thought the latest version of StorageX was “not particularly innovative,” it was “effective and refined over multiple years of usage.”

Part of StorageX’s easy management comes from the inclusion of a unified dashboard that allows users to look into every type of storage in an environment with products from multiple vendors.

In addition, StorageX 7.8 now includes a software developer kit that has a REST API, so customers can create their own data mobility policies that allow them to manage the entire data lifecycle.

This is the third Product of the Year award for StorageX. Version 7.6 won gold for storage management software in 2015 and version 5.0 earned silver way back in 2004 under the product’s original vendor, NuView.
Major new features such as data deduplication and compression helped vault VMware Inc.'s vSAN 6.2 upgrade to the gold award in the server-based storage category.

The vSAN hyper-converged infrastructure software—formerly known as VMware Virtual SAN—was among the category leaders in judging in ease of integration, use and manageability. It also scored well for innovation and functionality.

The addition of inline data deduplication and compression was essential for VMware to compete in the hyper-converged market space. The data reduction technology can help users reduce their storage footprint by up to seven times on flash-based server nodes, according to VMware. The 6.2 version does not support inline dedupe and compression on vSAN hybrid nodes, which combine flash and HDDs.

Erasure coding is another new vSAN 6.2 feature that can help customers conserve all-flash storage capacity. With RAID-5 erasure coding, vSAN can withstand one host failure with four hosts deployed. RAID-6 erasure coding requires a minimum of six hosts and can tolerate two host failures.

The vSAN 6.2 update also added support for quality of service to manage and monitor the IOPS consumed by each VM and curb “noisy neighbor” problems that can have a negative impact on performance. The new version also features usability and management enhancements and supports business applications such as Microsoft SQL Server, MySQL, Oracle Database and SAP ERP.

VMware’s vSAN runs in the hypervisor and offers integration with the complete VMware software stack. Users can manage vSAN through the vSphere Web Client and apply storage policies on a per-VM basis to automate provisioning and balance resources. The product scales up to 64 nodes per cluster, and the maximum raw capacity exceeds 8 PB (petabytes).

List pricing for vSAN 6.2 starts at $2,495 per CPU. The cost for vSAN for Desktop begins at $50 per user. VMware introduced vSAN, then known as Virtual SAN 5.5, in 2014. The 6.2 version of vSAN is the fourth edition of the product.
Formation Data Systems FormationOne

Startup Formation Data Systems snared the silver award with a new software-defined storage product that supports block, file and object storage.

The FormationOne software scored highest for innovation and functionality among 12 server-based storage finalists, based on average scores from the competition’s judges.

CEO Mark Lewis, a former CTO at EMC, founded Formation Data Systems in September 2012. FormationOne’s distributed, fault-tolerant architecture underwent more than two years of customer testing before the product’s general release in March 2016.

The Formation software-defined storage runs on commodity x86 servers and dynamically tiers virtual volumes across SSDs and HDDs. The product’s unified connector design provides interfaces for block (iSCSI), file (NFS), object (Amazon S3) and Hadoop Distributed File System (HDFS).

“Functionality is impressive,” one judge commented.

One key feature, TimeLine, provides transactional data journaling (continuous data protection) with integrated snapshots and clones, and enables customers to recover data from any point in time. FormationOne’s quality of service, called PriorityOne, allocates and prioritizes I/O to specific application workloads—within configured performance thresholds—based on policies.

Additional storage capabilities include single-origin inline data deduplication to reduce the storage footprint and FireBreak anomaly detection analytics to spot performance problems and security issues. Virtual Storage Recapture discovers, pools and repurposes unused flash- and disk-based storage capacity in VMware virtual servers, VM-attached arrays and hyper-converged systems.

FormationOne software-defined storage uses a subscription-based pricing model. The annual price of $50 per terabyte per year includes all software plus maintenance. Discounts are available for high-capacity configurations. Customers are responsible for supplying x86 compute nodes. Formation’s compatibility list includes Cisco, Dell, Hewlett Packard Enterprise, Huawei, Lenovo, Quanta and Supermicro. The FormationOne software can scale out to hundreds or potentially thousands of nodes.

FormationOne software-defined storage runs on commodity x86 servers and dynamically tiers virtual volumes across SSDs and HDDs. The product’s unified connector design provides interfaces for block (iSCSI), file (NFS), object (Amazon S3) and Hadoop Distributed File System (HDFS).

“Functionality is impressive,” one judge commented.

One key feature, TimeLine, provides transactional data journaling (continuous data protection) with integrated snapshots and clones, and enables customers to recover data from any point in time. FormationOne’s quality of service, called PriorityOne, allocates and prioritizes I/O to specific application workloads—within configured performance thresholds—based on policies.

Additional storage capabilities include single-origin inline data deduplication to reduce the storage footprint and FireBreak anomaly detection analytics to spot performance problems and security issues. Virtual Storage Recapture discovers, pools and repurposes unused flash- and disk-based storage capacity in VMware virtual servers, VM-attached arrays and hyper-converged systems.

FormationOne software-defined storage uses a subscription-based pricing model. The annual price of $50 per terabyte per year includes all software plus maintenance. Discounts are available for high-capacity configurations. Customers are responsible for supplying x86 compute nodes. Formation’s compatibility list includes Cisco, Dell, Hewlett Packard Enterprise, Huawei, Lenovo, Quanta and Supermicro. The FormationOne software can scale out to hundreds or potentially thousands of nodes.
DataCore Hyper-converged Virtual SAN (PSP₅ Update)

DataCore Software Corp.’s Hyper-converged Virtual SAN update captured the bronze award in the server-based storage category as the PSP₅ update scored highest among 12 finalists in the areas of performance and functionality.

One judge called the PSP₅ update a “good evolution of the platform for multithreaded workloads.”

Technology optimizations in the Hyper-converged Virtual SAN’s Parallel I/O software can boost performance by up to 50% over prior releases when running on multicore x86 servers, according to DataCore. The Parallel I/O software, which reduces bottlenecks and enables multicore servers to use all available cores to execute multiple I/O threads, won the gold award for storage system software in 2015’s Products of the Year competition.

“Performance is dependent on whatever hardware you run it on,” one judge noted.

A key enhancement in the PSP₅ update of DataCore’s Hyper-converged Virtual SAN was the expansion of the maximum capacity of the high-speed cache from 1 TB to 8 TB per node. DataCore said the larger cache would serve up more I/O requests from memory and speed transactions and analysis.

The update also included support for faster 32 Gbps Fibre Channel host bus adapters (HBAs) from QLogic, deeper integration with VMware and Microsoft hypervisors, and VMware VVOLs certification. DataCore also provided more granular quality of service and access controls at the virtual disk level and improved performance monitoring tools to ease the identification of heavy workloads that can impact system response and capacity.

DataCore’s Hyper-converged Virtual SAN targets traditional hyper-converged uses, such as virtual desktop infrastructure and remote office branch office environments, as well as latency-sensitive enterprise applications, including databases and ERP systems. The minimum configuration for high availability is two nodes, and the maximum expansion is 64 nodes. DataCore’s Hyper-converged Virtual SAN can scale storage capacity independently from compute.

DataCore Hyper-converged Virtual SAN is priced at $5,000 per node. Current customers can upgrade to Version 10 PSP₅ at no charge under their existing software update service and support contracts.
Nimble Storage introduced the Predictive Flash Platform AF-Series all-flash arrays in 2016, after years of exclusively selling hybrid arrays that mix spinning disk with solid-state drives. Previously, Nimble tried to woo customers by adding an optional all-flash expansion shelf to its CS-Series hybrid arrays.

AF-Series Fibre Channel- or iSCSI-connected block storage arrays use 3D NAND triple-level cell (TLC) SSDs across four AF Series models: AF3000, AF5000, AF7000 and the flagship high-capacity AF9000.

AF-Series arrays share common data services with Nimble Hybrid Adaptive Flash arrays, allowing customers to build a mixed four-node cluster with unified management. A single all-flash AF9000 cluster can scale to more than 8 PB of effective capacity in 12U with Nimble’s proprietary inline data compression and deduplication, with rated throughput of 1.2 million IOPS.

As with its hybrid storage, Nimble all-flash arrays integrate the vendor’s cloud-based InfoSight Predictive Analytics. InfoSight correlates billions of data points to monitor the behavior of virtual machines to pinpoint potential issues across the infrastructure stack. Nimble claimed that its combination of predictive analytics and triple-parity RAID yields five nines of availability.

Nimble Timeless Storage guarantees AF-Series 3D TLC NAND flash media for seven years, and allows customers to upgrade to a higher performance flash controller after three years.

In rating the AF-Series entry, judges noted the flexibility that stems from Nimble’s flash fabric, which builds on its Cache Accelerated Sequential Layout (CASL) architecture for Adaptive Flash hybrid storage.

“(AF-Series) uses the Nimble architecture with flash, which complement each other. Good scalability,” one judge wrote.

The Nimble AF-Series all-flash arrays support VMware, Microsoft Azure, Docker containers and OpenStack cloud. They also support multi-tenancy with secure isolation and quality-of-storage management services. Integrated data protection allows you to move application-specific archive, backup and disaster recovery copies from Nimble all-flash primary storage to secondary tiers of Adaptive Flash arrays. Pricing starts at less than $40,000 for the Nimble AF1000 base building block.
**Tegile Systems IntelliFlash HD**

Tegile Systems’ silver-winning IntelliFlash HD 5U all-flash system is the fruit of the vendor’s OEM partnership with SanDisk (now part of Western Digital Corp.). The IntelliFlash HD combines Tegile’s 2U controller and IntelliFlash operating system with the SanDisk InfiniFlash IF100 all-flash chassis.

Tegile IntelliFlash HD incorporates two tiers of flash. A high-performance flash tier that logically partitions metadata and write logs is coupled with a dense flash layer to serve application and user data. Tegile software intelligence manages data placement across flash tiers based on an application’s performance requirements.

The IF100 enclosure supports 64 SanDisk SAS-connected PCIe NAND flash cards to cram 512 TB of raw capacity into 3U of rack space. The PCIe flash performance layer serves as a fast cache to support inline data deduplication across the array. Tegile combines the PCIe tier with a 2U box with slots of 24 enterprise MLC flash drives as a capacity tier.

Tegile IntelliFlash HD provides “similar innovation to Nimble, but with slightly lower density … better performance, [but trailing Nimble on] predictive analytics,” according to one judge.

Tegile also sells five internally developed all-flash storage array models. The IntelliFlash HD product targets medium and larger enterprises that want to collapse multiple workloads on a single all-flash platform.

The vendor claimed its IntelliFlash architecture provides flash storage at 50 cents per gigabyte. It lists big data analytics applications, data warehouses, image processing and private clouds as projected IntelliFlash HD use cases.

Tegile IntelliFlash supports unified storage with native block and file protocols with clones, inline data reduction, snapshots, replication and thin provisioning.

Effective capacity scales to 10 PB in 42U after inline data reduction is applied. Sustained rack performance is rated at 5 million IOPS with submillisecond latency.

List price starts at $425,000 for the IntelliFlash HD T10 KHD-100 model with 522 TB of effective all-flash storage. The T10KHD-200 and T10KHD-300 scale respectively to 1,044 TB and 1,566 TB.
**Hitachi Data Systems VSP F800**

**Bronze winner** Hitachi Data Systems’ (HDS) Virtual Storage Platform (VSP) F series marked the vendor’s first array platform to bundle both Hitachi Storage Virtualization Operating System (SVOS) and its dense custom-built flash module drives (FMDs).

The modular rack-mounted F800 is the high-capacity VSP model with a 512 GB and 448 TB of usable flash storage following deduplication. The system is available with various fixed capacity FMD configurations. It does not permit disk or SSDs to be used, part of HDS’ strategy to reduce host latency and boost throughput across all connection ports.

Hitachi VSP F800 is “based on strength of [HDS’] proven hardened array” one judge noted. Another judge said the manufacturer’s $69,860 list makes VSP F800 “a good value at the price.”

HDS engineered VSP F800 flash for high availability and top-end random performance of 1.4 million IOPS and 24 GBps of bandwidth, including Fibre Channel and iSCSI connectivity. Multiple applications can be consolidated on a single VSP F800 platform with HDS’ guaranteed performance and uptime. Data protection includes redundant snapshots, asynchronous and synchronous replication, and component monitoring.

Hitachi FMD devices use a parallel non-blocking architecture to maximize I/O to and from NAND flash. To minimize latency, the flash intelligence prioritizes application I/O higher than background tasks. Each FMD performs inline compression and distributes data reduction to reduce controller overhead.

The vendor claimed FMD storage is five times faster than traditional SSDs, with each module backed by a data availability guarantee for applications requiring high performance and uptime.

Hitachi Storage Advisor is a wizard-driven automated configuration and management tool geared to simplify VSP F800 installation. The management dashboard provides at-a-glance monitoring and phone-home capabilities. Integration is included for OpenStack, VMware and REST APIs for third-party management tools.

In addition, you can combine VSP F800 arrays as part of Hitachi’s Unified Compute Platform for organizations seeking a converged compute, network and storage stack. Other products in the VSP F Series include the VSP F400 and F600 all-flash storage.
The Zadara Storage Cloud is actually a double-hybrid. It meets the hybrid/all-disk storage category definition by mixing SSDs and HDDs, but can also serve as a hybrid cloud storage system.

Zadara Storage Cloud provides storage as a service for block, file or object storage either on premises, in the cloud or both. The storage system is managed and owned by Zadara, with customers paying a subscription for all-inclusive infrastructure—hardware, software, remote monitoring and management, and support. Pricing starts at $0.03 per gigabyte per month, including a 100% uptime service level agreement.

Zadara Storage Cloud is a touch of old and new technology. It is based on Zadara's tested Virtual Private Storage Array (VPSA) with block and file storage. New features include ZIOS Object Storage; Elastic Flash Cache; and support for 1.6 TB SSDs, VMware SRM and Zadara Container Services.

The object storage is Amazon S3 and OpenStack Swift compatible and gives Zadara Storage Cloud file, block and object storage on the same platform. The Elastic Flash Cache scales to 3.2 TB and provides speed to go with VPSA's capacity HDDs for bulk storage. ZSC allows Docker containers with dedicated compute and storage to run inside the array.

Zadara Storage Cloud offers a mix of SSDs with high-speed and 6 TB capacity SAS drives. Users can update their storage resources on the fly. They access an online portal, select options and click “submit” for quick access to additional resources.

VPSA consists of independent controller pairs, each with access to their own drives. Multiple tenants can run VPSA arrays or ZIOS Object Storage on the same physical machines without contention. The arrays are built on x86 hardware and run Zadara virtualization software.

One judge called Zadara Storage Cloud an “interesting business model with storage as a service on premises.” “There were significant improvements in this release,” another judge said. “Not a lot of innovation, but excellent additions.”

Zadara also impressed with its cost structure. “Priced right and nicely featured,” a judge said. “[It's] a good way to avoid Capex.”

Zadara Virtual Private Storage Arrays OPaaS (On-Premises as a Service) won bronze in storage software in the 2015 Products of the Year awards.
The upgraded Nimble storage Adaptive Flash Arrays (CS-Series) had enough impressive additions mixed in with its tried and true features to win the silver award.

While Nimble Storage spent a lot of time launching and pushing its all-flash storage arrays in 2016, the vendor remembered its disk roots. Nimble refreshed its CS hybrid array platform that mixes solid-state drives and spinning disk drives. The vendor added an entry-level product for under $59,000 while extending scale on the high end to 5.6 PB of effective storage with compression and 240,000 IOPS in a scale-out cluster.

Nimble increased the size of its flash cache in the upgraded systems and claimed its compression reduces dollars per usable gigabyte by up to 43%. The CS-Series supports Fibre Channel and iSCSI block storage.

The judges scored the CS platform high for ease of use and management and functionality. For instance, CS customers can perform their own firmware upgrades. Nimble’s analytics also won high marks.

“Some good innovation around the central analytics being collected and performed,” one judge said.

Nimble claimed its InfoSight Predictive Analytics can find 90% of its arrays’ problems before customers become aware of them, and data protection is integrated into the array. The CS-Series’ auto-performance adapts performance to meet the needs of each application.

Even before adding all-flash arrays, Nimble did a good deal of work to make SSDs perform better in its hybrid arrays. The CS hybrids allow customers to pin volumes to flash to give specific applications the same performance as they would on an all-flash array. That allows multiple service levels within an array.

Nimble’s CASL file system uses SSDs as a read cache and places persistent storage on hard disk drives. It improves write performance by sequencing random write data.

“Very high performance,” one judge said.

Nimble also won praise for its integration with hypervisors.

A judge summed up: “Very good hybrid-based on VM-centric ease of use.”

Nimble is no stranger to our Products of the Year winners list. Its CS700 won a bronze for hybrid/all-disk storage in 2014 and its AF-Series of all-flash arrays is the 2016 gold winner in all-flash storage.
Avera Systems Cloud-Core NAS (C2N)

AVERE SYSTEMS TOOK bronze award honors with its Cloud-Core NAS product that spans drive types, storage tiers and protocols.

NAS acceleration vendor Avera launched C2N to integrate private and public object storage with an organization’s NAS infrastructure. C2N is a NAS system with a file system also built for object storage. Customers can manage file and object storage on-premises and in the cloud as one logical storage pool through Avera’s global namespace.

Built on Avera’s FXT Edge filer hardware, C2N starts at 120 TB of capacity and expands in 80 TB increments to 5 PB. It includes an all-flash performance tier ranging from 14 TB to 480 TB of solid-state drive capacity. The goal of the appliance, which migrates data to the cloud, is to integrate object storage into an existing NAS setup without having to rewrite applications or train staff to manage new technology.

C2N supports NFS and SMB NAS protocols. For data protection, it uses N+4 erasure coding and triple replication. The appliance caches active data on internal SSD and dynamic RAM, and a C2N system can include up to 50 Edge Filers.

Avera pitched C2N as a lower-cost alternative to scale-out NAS and on-premises object storage.

List price starts at $99,500 for 120 TB. A sample private cloud configuration of three FXT 5200 filers, nine CX200 C2N nodes with 576 TB of usable capacity and disaster recovery across three sites would cost about $349,000, according to Avera.

One judge called C2N an “interesting move to finally produce an on-prem product to complement the caching FXT. Underlying object store [is] a good idea.”

Another added, “C2N is a major release. It is a high performance scale-out NAS, cloud integrated storage, global file system, high performance AFA or hybrid, and more—highly innovative.”

Although not all of our judges were sold on price, C2N’s performance and functionality impressed.

“The minimum config is expensive, but extremely capable,” one judge said.

The Avera FXT Series won a bronze in the disk systems category in 2009.
2016 Products of the Year finalists

**BACKUP AND DISASTER RECOVERY SOFTWARE AND SERVICES**
- Acronis Backup 12/Acronis Backup Advanced 11.7
- Catalogic ECX 2.4
- Commvault Data Platform V11
- Datos IO RecoverX
- Datto SIRIS 3
- Druva Phoenix
- HotLink Managed DRaaS
- Unitrends Enterprise Backup 9.0
- Veeam Availability Suite v9
- Zerto Virtual Replication 4.5
- Zetta Disaster Recovery 1.0

**BACKUP HARDWARE**
- Actifio Virtual Data Pipeline 7.0
- Barracuda Networks Backup 6.3
- NetApp AltaVault 4.2
- Quantum Corp. DXi6900-S
- Quorum OnQ 3.9
- Unitrends Recovery Series Appliances 9.0
- Western Digital HGST Active Archive Systems SA-1000

**SERVER-BASED STORAGE**
- Caringo SwarmNFS
- Cloudfishics Ignite
- DataCore Hyper-converged Virtual SAN (PSP5 Update)
- Datrrium DVX
- DriveScale System
- Formation Data Systems FormationOne
- Pivot3 vStac SX
- Portworx PX-Enterprise
- Red Hat Ceph Storage 2
- Scale Computing HC1150 Hybrid System
- SUSE Enterprise Storage 3
- VMware Virtual SAN 6.2

**STORAGE AND SAN MANAGEMENT TOOLS**
- Cisco Data Center Network Manager 10.0
- Data Dynamics Inc. StorageX 7.8
- NetApp ONTAP 9
- NetApp SANtricity 11.3
- ProphetStor Data Services Federator 4.0
- Virtual Instruments VirtualWisdom NAS Performance Probe

**STORAGE SYSTEMS: ALL-FLASH SYSTEMS**
- Hitachi Data Systems VSP F800
- IBM FlashSystem A9000
- IBM FlashSystem A9000R
- IBM Storwize V7000F
- NetApp SolidFire SF-Series 19210 Node
- Nimble Storage All Flash Arrays (AF-Series)
- Tegile System i320 HD

**STORAGE SYSTEMS: DISK AND HYBRID SYSTEMS**
- Avere Systems Cloud-Core NAS
- ClearSky Data Global Storage Network
- DataDirect Networks SFA14K Hybrid Solution
- Dell EMC SC Series with Storage Center Operating System 7 (SCOS 7)
- Hitachi Content Platform
- Nexsan Unity
- Nimble Storage Adaptive Flash Arrays CS-Series
- Quantum Xcellis Workflow Storage
- Qumulo Core 2.0
- Synology RackStation RS3617xs+
- Zadara Storage Cloud

---

About the Storage magazine/SearchStorage.com Products of the Year

Storage magazine and SearchStorage.com invited data storage product companies to nominate new or enhanced products for the 2016 Products of the Year awards. For previously available products, the upgrade must have incorporated significant new features. Products could be entered in six categories: backup and disaster recovery software and services, backup hardware, server-based storage, storage and SAN management tools, storage systems: all-flash systems, and storage systems: disk and hybrid systems. Products were judged by a panel of users, analysts, consultants, and Storage magazine and SearchStorage.com editors. Products were rated based on innovation, performance, ease of integration into environment, ease of use and manageability, functionality and value.
Snapshot 1

Better security and collaboration power enterprise file sync-and-share uptake

Which deployment options are you considering for enterprise file sync and share?*

- **48%** Hybrid of on- and off-premises
- **36%** On-premises, Private cloud
- **33%** Off-premises, Hosted (software as a service)
- **23%** Off-premises, Hosted
- **22%** Off-premises, Public cloud
- **16%** On-premises, Hosted

*MULTIPLE SELECTIONS ALLOWED

What are the most important factors driving your decision to deploy EFSS?*

- **52%** Increased need for collaboration tools
- **45%** Better file sharing security
- **40%** Simpler access to shared files
- **39%** Mobile access to documents
- **37%** Central management and control
- **27%** Discourage/prohibit unauthorized services
- **13%** Dissatisfaction with prior tools

*MULTIPLE SELECTIONS ALLOWED

Nearly half of those surveyed intend to purchase new or additional file sync-and-share services in the next nine to 12 months.

49%

Home
Castagna: Will the storage world be turned upside down in 2017?
Toigo: ‘Holistic’ data protection, a meme on the rise
2016 Products of the Year
What’s powering EFSS uptake?
Capacity growth survival guide
EFSS solves a number of organizational challenges
Buffington: Why you should leverage the cloud for data protection
Taneja: Hyper-converged, not just for primary storage anymore
About us
BUSINESS PRESSURES OUTSIDE the data center could lead to hasty decisions within, such as too quickly zeroing in on the cause of a storage performance slowdown. Hastiness could lead you to upgrade your storage network to all-new, faster connections, only to discover later that you haven’t solved anything. Why? Because the real problem was latency from too many servers banging on the system, rather than the speed of any individual connection.

Measuring application performance from end to end, preferably with a system that does historical tracking, is the only real way to discern exactly what’s causing a slowdown. You collect data over days, weeks or months to see how it’s moving over time, which can identify periodic spikes from regular events—from users logging in at 8 a.m. to weekly backups at midnight—and identify traffic consumption trends. That way, in addition to zeroing in on what’s negatively impacting storage today, you can detect trends and potential future problems, and maybe even fix them before they become a real issue.

The onset of storage performance troubles often goes hand-in-hand with initiatives to increase—or make the most of—data storage capacity. Fortunately, storage management products not only help solve and prevent pokey storage problems, they often include tools to help increase available capacity without negatively affecting performance.
**ZEROING IN ON CAPACITY**

*Storage management applications* help manage data storage capacity by identifying files that don’t need to be on a system (do your users really need to store a few terabytes of home movies in their work folder?), or by moving files that haven’t been opened for a while to older, slower storage. Also, when running low on capacity, you can search, move or delete large files or certain file types, such as .mpeg; compress data; and migrate it off primary storage to secondary, tertiary or other storage.

While you can manually search using tools in Windows, those built specifically for storage management from array vendors or a third party can do much more (see “Tools of the trade”). These include building flexible queries that run during off hours, as well as warning end users when files need moving before, for example, files over a certain size are automatically deleted. And while compression tools are also available in Windows—and may be included with a storage system or storage management application—care should be taken that compression overhead doesn’t negatively impact storage performance because of write delays. And, since a whole file must be decompressed and recompressed whenever even one byte is changed, the impact on overall file system performance can be disproportionate.

TIERING TOOLS can move files and leave pointers to the new location. Consequently, when a user clicks on a file in the old location, it is returned from the new one and opened. Alternatively, you can create a single namespace that concatenates several different storage systems, so the process of accessing files is completely transparent to users, no matter where they actually reside. Some tiering tools can even move files to the cloud, where costs are much lower, and return them as needed.

**TIERING AND PERFORMANCE**

*Caching and tiering* have something in common: If they’re sized large enough to provide space for the 10% of data typically active at any given time, all the data stored on the next tier down will effectively have the same performance as the top tier. That’s because all the data can be served from the fast 10%, with files being moved up from the next tier as needed.

This also works when there are more than two tiers. For example, a flash tier with a data storage capacity of 10 TB can front a fast hard drive tier with 100 TB, which can front a high-capacity tier with a petabyte of storage, which can then front a cloud tier with 10 PB. Applications then get the effective performance of the flash tier for the whole 11.1 PB of capacity in the storage infrastructure.

The key here is predictive software that can fetch a whole file when only a part is requested and monitor the overall system to ensure the data’s migrated up the chain as needed. Big arrays integrate these tools into their storage controllers, or you can purchase them separately in products such as DataCore SANsymphony.

**BEST PRACTICES**

Because some users dump gigabytes or terabytes of music, movies and more on company storage, you should
regularly accumulate trending data on the percent of overall storage in use to identify such drains on resources and move the data elsewhere. However it’s done, whether through the software included with your array, a separate storage management app such as Veritas InfoScale Operations Manager, or by tracking with a spreadsheet and the drive properties function in your server OS, it doesn’t matter as long as it’s done accurately and consistently.

In addition, having an “official” data policy in place—from encouraging users to keep music files off company servers to the archiving of old files—is a good starting place. A formal data policy can keep your backside covered when a data cleanup deletes files users thought would be kept indefinitely.

One simple way to avoid squabbles with users is to manually or (preferably) use tiering software to move old, unused, obsolete or inappropriate data from primary storage to secondary or tertiary storage. If your SAN or NAS doesn’t support this directly, you can buy auto-tiering software like SANsymphony separately. These programs let you use older storage as a secondary tier where you can move specific file types or files that haven’t been accessed for a while, keeping your new high-performance storage free for appropriate tasks.

**MAINTENANCE**

Storage, like other computer systems, tends to slow down as it gets older. A filer that was very fast when first installed and only had to support a few users and a few thousand files, may not perform as well with 12,000 users and 40,000,000 files—even if there appears to be sufficient disk space available. And that archive of a thousand CD-ROM images may only take up a few terabytes of disk space, for example, but the number and size of the files could cause trouble by slowing file system performance down and increasing the amount of system memory used by storage.

A monitoring tool that finds and notifies administrators of these kinds of issues can quickly pay for itself, and will often postpone the purchase of new hardware by regaining lost efficiency. Often included with storage systems, examples of **stand-alone monitoring tools** include SolarWinds Storage Resource Monitor and Cloudera Enterprise.

Beyond pruning files and reducing overhead, it’s possible to help an existing storage system deliver data faster with quicker response times without buying a new array. For instance, you can place an all-flash caching appliance, such as Permabit’s Enterprise Flash Caching Appliance or Cirrus’ Data Caching Server, between servers and storage. By caching most-used data and serving it up from the cache instead of old storage, the appliance transparently accelerates existing storage for relatively little cost.

Caching holds data temporarily, moving files in and out as they’re requested, while a tier 0 system adds a new permanent layer of faster storage. Tiering software can cause a **tier 0 layer** to act as a cache, but it isn’t automatic like a cache.

Depending on the SAN or NAS you have, you may also be able to accomplish this by buying an additional storage shelf or module and using the built-in tiering capability. And, for those willing to roll their own cache appliance
or tier 0, adding caching or auto-tiering software and a white-box all-flash storage server can accomplish the same thing, and often at much lower cost.

**EXPANDING EXISTING ARRAYS**

Adding modules, additional shelves of drives or new nodes in a cluster is an option for those with substantial investments in one vendor’s storage. However, while storage vendors often make discounts available to existing customers, and no new training is enticing, it may be the case that your existing array’s controller won’t support the faster interconnect speed of the new drive shelf, at least not without an additional hardware upgrade that requires taking storage controllers offline for hours and buying a new level of support.

It may be simpler and cheaper in the long run to add a new storage system, using storage management software to integrate the two systems. This has the added benefit of allowing the purchase of a highly rated system every few years, rather than locking you into a vendor that may not have kept up with the latest technologies.

Relatedly, while it may seem like buying partially populated storage systems with the intention of purchasing additional drives as prices drop, capacities increase and performance improves is an effective strategy for growing data storage capacity, it isn’t unless the times between buys are very short—quarterly or even monthly. Over longer intervals, it may become impossible to find drives to fit, say, a 2-year-old chassis, let alone one that’s 5 years old.

This is also true of systems that come fully loaded, with

**Tools of the trade**

**ALTHOUGH WINDOWS SERVER** and your existing storage may include storage management tools that measure application performance to help identify the source of any problems and increase data storage capacity, you may want to consider buying a third-party tool. Because even if some of your storage includes storage management software, it’s likely proprietary and will only help manage your current vendor’s storage arrays.

A third-party tool is capable of measuring and managing data across multiple vendors’ hardware, and can even help add additional tiers of storage or migrate some storage to the cloud. Examples of third-party storage management include DataCore SANsymphony, Veritas InfoScale Operations Manager and SolarWinds Storage Resource Monitor. Most vendors offer a 30-day trial or limited version you can try to see if their product meets your needs before buying.

There are also appliances that provide data management capabilities, from storage monitoring to acting as cloud storage gateways. These are often available for testing as virtual machines that can be installed, previewed and then purchased as a separate hardware appliance if you like the functionality.
capacity beyond what you buy initially. Here, you simply purchase software keys from the array vendor to unlock more storage as needed. However, by the time you require more capacity, you’re likely to need orders of magnitude more, rather than twice what you already have. In a typical capacity-on-demand system, you might buy 50 TB with a system that has 100 TB. By the time you’re ready to unlock the additional 50 TB, you may find you really need 250 TB, or 500 TB.

**CLUSTERED STORAGE: OPEN-ENDED SYSTEMS**

One answer for allowing expansion without having to maintain hardware compatibility is to cluster storage. This could be a system with software distributed over a number of storage nodes, think HyperGrid and Isilon. So when storage gets low, adding additional nodes not only increases data storage capacity, it improves performance as well, as reads and writes are then distributed over all the nodes in the cluster. And since many clustered products are hardware-agnostic, they allow you to purchase nodes as a complete bill of materials from one vendor or built with hardware available through many different vendors.

**HETEROGENEOUS STORAGE: FIGHTING VENDOR LOCK-IN**

If you buy a SAN or NAS from one of the main vendors, you get a system with lots of features and good performance for a reasonable cost. But two years later, you may go back to the same vendor, only to discover they’ve purchased a couple of competitors and are now pushing you toward a forklift upgrade to their newest system or, worse, no longer have the parts to expand your existing system—despite salesman assurances that parts would be available for at least five years.

There are a few ways to avoid this scenario. Buy a software-defined storage system, storage management system or appliance that allows you to create a single access point for multiple arrays from multiple vendors. These systems enable you add new storage systems from any vendor you like while getting maximum performance from new arrays, all while continuing to use the old ones.

**CONCLUSION**

It’s possible to keep your existing investments in storage active, allow for capacity increases and find and solve problems with storage and application performance, all while gathering data to plan for new growth without surprises. All it takes is measurement, analysis and tracking software, which will let you stay on top of the performance of your storage system.

This doesn’t have to be expensive, either in purchases or time spent. Having accurate data will not only point you toward possible inexpensive fixes, it can help justify the purchase of new storage when it’s really needed.

**LOGAN G. HARBAUGH** is a freelance reviewer and IT consultant who has worked in IT for more than 25 years.
How much file sync-and-share capacity are you buying?

- Less than 10 TB: 35%
- 1 Peta-byte (PB) or more: 5%
- 500 TB - 999 TB: 6%
- 200 TB - 499 TB: 8%
- 100 TB - 199 TB: 10%
- 50 TB - 99 TB: 10%
- 10 TB - 49 TB: 26%

Top 10 business challenges solved by EFSS*

- File sharing/syncing/centralizing content: 59%
- Enhance mobile experience/access: 40%
- Security/access controls: 39%
- Eliminate/reduce content silos/chaos: 31%
- Regulatory compliance discovery requirements: 28%
- Search, access, discovery of documents: 27%
- Effective content governance strategy: 27%
- Integration with core business apps: 25%
- Manage workflow/alerts/approvals: 19%
- M&A technology standardization: 19%

*Multiple selections allowed

Percentage of respondents who work in a regulated industry that has special compliance requirements for EFSS: 45%
Power of cloud lies in agility for data protection

You can do different things with cloud copies of data than tape copies, which is why—in the end—you need both.

There are lots of reasons to embrace cloud services as part of your enterprise data protection strategy. But unless you are an SMB in a nonregulated industry, eliminating your use of tape shouldn’t be one of them.

In ESG’s recent Data Protection Cloud Strategies report, although most respondents cited an intention to store data in a cloud for one to three years, they had to meet data retention mandates for five or more years. For most organizations, the longer they have to retain data, the more likely tape will continue to play a role in long-term data retention.

Nonetheless, even if you don’t reduce your usage of tape, that doesn’t make the cloud any less compelling for an enterprise data protection strategy. The power of the cloud lies in agility, not the race to bottom dollar per gigabyte stored.

When it comes to data protection and retention, the benefits of cloud-powered agility fall into a few camps, including lowering the potential amount of data that might be lost and being able to do more with secondary data copies. Let’s delve into these a bit.

Reduced data loss

While organizations often struggle with “cost of downtime,” it’s arguably easier to appreciate “cost of lost data” as a simpler and more compelling measurement of effort lost or repeated when data must be recreated. That’s because the time needed by workers to recreate their earlier efforts takes time away from new work they should be doing.

IT professionals, meanwhile, often cite recovery point objective (RPO) as a metric for data protection. This can be simplified by presuming “predictable data loss” is one half of the frequency of data protection.

For example, say an organization performs a daily data protection event (backup, snapshot, replication and so on) at midnight each evening. A server failure early in the morning would result in very little data loss, while a failure at 6 p.m. would result in a full business day’s worth of lost data (i.e., eight hours). So, to determine predictable data loss in this scenario, split the difference by assuming...
all failures happen at noon (midday). The result is a half-business day of predictable data loss when backup occurs on a nightly basis.

According to ESG research, 17% of organizations send data to a cloud on a daily, or nightly, basis as part of their enterprise data protection strategy. While typical of a tape backup process, it’s not the most effective way to take advantage of the agility of the cloud for data protection. Fortunately, users of cloud-based data protection send data to the cloud every two hours on average. For them, predictable data loss decreases from a half day to one hour, decreasing RPO from eight hours—a business day—to two hours in the process. That is real ROI!

**DATA AGILITY**

Many can simplify the topic of copy data management down to the rhetorical question: “What else can I do with the secondary and tertiary copies of my data?” After all, it’s not unreasonable to presume 10 or even more copies of secondary data exist, resulting from snapshots; replicas created for disaster recovery; and multiple backups across daily, weekly and monthly iterations. And while most of those copies and partial versions may be legitimized for the assured recovery or preservation of business data, they can also be cumbersome and expensive to store and maintain. As such, many organizations need to gain new “value” out of their enterprise data protection strategy beyond crisis preparation.

Some want to run reports or conduct analytics from secondary, otherwise dormant, data, for instance, while others will do patch testing or similar test/dev activities. You can often accomplish these efforts by simply harnessing secondary data storage through orchestration workflows, combined with noninvasive representation data-access methods that do not malign pristine copies that may still be required for data restoration.

While you can accomplish some of these goals through on-premises secondary storage, combining cloud-based compute (on-demand access) with cloud-based protection storage provides an economical and scalable (from instant-on to entirely dormant) set of business outcomes. Essentially, organizations that do the latter recognize the incremental business value of “warm” data within a cloud, something not typically attainable from the “cold” copies of data within tape cartridges.

This is the primary reason why businesses store data in clouds for the short term (one to three years), while still embracing tape for long-term (10 years-plus) retention. The warmer the data, the more agile it is and, therefore, the more useful that data is for purposes other than retention and protection objectives.

Now that we have a clearer understanding of why cloud(s) make sense as part of an enterprise data protection strategy, my next Hot Spots column will explore the key determinants between cloud storage, cloud backup services and disaster recovery services.

JASON BUFFINGTON is a senior analyst at Enterprise Strategy Group. He focuses primarily on data protection, as well as Windows Server infrastructure, management and virtualization. He blogs at CentralizedBackup.com and tweets as @Jbuff.
Hyper-converged, not just for primary storage

Architectural advances in the primary world have yet to make a big splash on the secondary side. That’s changing.

Primary data storage and primary workloads are well understood by now—their secondary counterparts, not so much. Consequently, the secondary side has been slow to benefit from the many architectural advances realized on the primary side over the last few years.

By convention, primary data storage directly interacts with an application when reading and writing data. Primary workloads, also by convention, are applications that directly interact with primary storage and often, but not always, support end users. Tier 1 or tier 2 applications such as ERP, customer relationship management, Microsoft Exchange and SQL Server, and so on are all considered primary applications. On the secondary side, we typically think of data protection, archiving, replication, data deduplication, compression, encryption, analytics and test/dev as individual secondary applications that work on and with data created by primary applications and stored in secondary storage.

The problem is this secondary world has been disparate, confusing and chaotic, made up of a large number of vendors who often solve only a single piece of the puzzle. In fact, it’s only recently that the industry’s started to see all these different workloads holistically through a common lens of “secondary” storage applications.

These definitions aren’t definitive—how could they be?—and the lines between primary and secondary storage and applications remain somewhat blurry. But they do serve as a basis for further discussion.

It all starts with primary storage

As with primary data storage, secondary storage is starting to undergo a tectonic shift. I’ll summarize what’s happening with primary storage first, as it applies to the revolution beginning with secondary storage.

The chaos created by silos of primary storage and their impact on cost, time and application performance has been well-documented. Thankfully, we’ve seen advances on this side brought about by using principles of virtualization; scale-out; convergence; and, more recently, hyper-convergence.
Virtualization abstracted compute hardware, changing the way we provision and manage compute resources. Convergence combined compute, storage, networks and server virtualization in a more integrated fashion to make the infrastructure stack easier to buy, provision and manage. And hyper-convergence flipped the industry on its ear by essentially melding compute, storage and storage networking (SAN, NAS) into a single entity and scale-out architecture.

Because hyper-convergence reduces so many of the pressure points IT has been struggling with over the past three decades, its growth shows no end in sight. Suffice to say, hyper-convergence reduces or eliminates the burden of provisioning or fine-tuning, managing, and arbitrating compute and storage resources, and will soon do the same for networking.

THE QUESTION IS ...
If hyper-convergence is so good for primary data storage, does it have a role on the secondary-storage side? While there has been plenty of innovation, the overarching answer is a loud “no.”

At a vision level, however, shouldn’t we be able to apply the principles of hyper-convergence, scale-out file systems, virtualization, software-defined storage (SDS) and more to gain some semblance of order in the secondary world, just as we did on the primary side? From the perspective of technology available, the answer seems to be a resounding “yes.”

The challenge is huge, though, and one has to wonder if it’s possible to develop a comprehensive product for all secondary workloads without starting from scratch. Before we address this, let us draw some boundaries and define what this new architecture for secondary workloads should achieve.

HYPER-CONVERGING SECONDARY STORAGE
We at Taneja Group have labeled this new architecture hyper-converged secondary storage. The reason for the name is simple: Products that are in this category use hyper-converged principles, but do so 100% in service of secondary workloads.

As you read the prerequisites we’ve developed for the new hyper-converged category of secondary storage, keep in mind that just because a product is missing a specific function today, doesn’t mean it wouldn’t belong in this category. The question you have to ask yourself is, “Is it architected so that this new functionality can be added without fundamentally redoing the product?” For no vendor may currently meet the full definition, some will come close and others may never make it. All of which is just as true of hyper-convergence on the primary-storage side.

In the end, you have to decide if what’s missing is important to your requirements or not. Taking this into consideration, here are the fundamental requirements for hyper-converged secondary storage:

- The storage must be “infinitely” scalable in a scale-out fashion, using a nodal architecture. Practically speaking, “infinite” means the same as a public cloud does today; i.e.,
it scales as far as you need to for a given set of applications, without performance dropping or latency increasing.

- **The storage can handle multiple workloads** with varying performance requirements without manual tuning. Hyper-convergence principles applied on the primary-application side are equally applicable here.

- **The storage is software-defined at the core**, separates the control plane and data plane and allows the use of most commercial off-the-shelf hardware. No hardware dependencies, it can run on-premises and in the cloud.

- **The storage tightly integrates** with [public or private clouds](#). This means having the ability to manage and protect data once it’s in the cloud. It must be able to use the cloud as a tier in a seamless fashion.

- **The storage can handle all secondary workloads**, today and in the future. Today, that means data protection, archiving, disaster recovery, replication, data migration, deduplication, compression, encryption, test/dev, copy data management (CDM) and analytics.

- **The storage supports multiple block, file and object protocols**, including iSCSI, Fibre Channel, NFS, SMB and REST at a minimum, and has the ability to store files and objects within the same storage pool.

- **The storage is based entirely on policy**. Set the policies for a workload at the outset, possibly using a predefined template and the system manages the entire data lifecycle and workflow thereafter—including spinning up and tearing down infrastructure resources—without operator involvement.

- **The storage has built-in quality of service (QoS)**. Since a multitude of secondary applications will, by definition, be running on the infrastructure, the system must allocate resources according to set policies, and there must be a way to ensure compliance.

- **The storage can support both physical and virtual workloads**, so the primary data source can be either.

- **The storage can index metadata and content, builds-in custom analytics**. It has a sophisticated search capability, including the ability to search on data within files. The results are available to applications via standard APIs.

- **The storage has one web-based management console** for the entire secondary infrastructure and is fully manageable from anywhere there’s internet—designed for a global namespace if the customer so chooses. And it has the ability to integrate with common management platforms such as VMware vRealize Automation.

- **The storage has self-healing architecture** that doesn’t require failed parts be replaced immediately, can deal with multiple disk and nodal failures without losing resiliency or availability with no requirement for data migration, and the level of resiliency required can be dialed by IT.
The storage has built-in data virtualization principles to ensure one data copy can serve many application workloads.

The storage has enterprise-grade security that includes encryption for data at rest and data in motion and access control. It may optionally include external key managers.

The storage offers recovery point objectives measured in seconds and minutes, with instantaneous recovery time objectives (no rebuilds and no rehydration).

There is a common misconception in the industry that a hypervisor is a prerequisite for hyper-convergence. Since hyper-convergence as a technology was first used on the primary side and applied to virtual workloads, the need for a hypervisor was evident. But as the industry moves toward containers, that need will disappear. The situation is no different on the secondary side.

As you can see, this is an extensive list of requirements. I believe, though, we are at the point when all the right technologies are available to implement this vision. Several vendors are very close, while others are scrambling to get there in the next 18 months or so, and some, depending on their current architectures, may never succeed.

HYPER-READY, OR NOT
At a casual level, many existing vendors will say they meet the above criteria when in fact they don’t. So let’s examine different categories of secondary products as they exist in the marketplace right now to learn more.

It is a fact of computer science that your architecture, once set, defines what you can or cannot do (effectively) in the future. In that vein, most vendors that started out producing a data protection product, or data protection and replication product, a few decades ago will likely not have a scale-out nodal structure, virtualization and many other technologies that surfaced in the past five years.

There is a common misconception that a hypervisor is a prerequisite for hyper-convergence.

If a vendor started out as a CDM player, for example, it is possible data would need to move to its repository before copies are managed. Data protection may or may not be an integral part of the product (that is, it may require a third-party data protection application) and the product may or may not scale-out.

The level of analytics available varies among vendors, with most still focused on storage utilization and data protection metrics. If a vendor only delivers archiving, it is likely a silo unto itself. Cloud integration may simply be a data transfer function. Most data deduplication products are just that: specialized devices for reducing copies of data. And they work in concert with a data protection application supplied by the same or a different vendor. Scale-out is present in some and not in other vendors’ products.
In the meantime, certain replication products only perform that one function and, hence, create a separate silo of functionality. Most lack true QoS capabilities. And while there may be a service-level agreement and policy-based control plane, there’s usually no way to determine compliance. By design, object storage purveyors have scale-out, single namespace, very secure and available platforms (due to erasure coding), but often lack integral file support and don’t support all the secondary workloads mentioned above.

Developed over the past three decades, the sweet spots for this arsenal of current secondary products vary all over the place. Most do what they were designed to do, and they do it well. But the world has changed, and data protection alone isn’t sufficient. We must rethink the whole paradigm of protection and availability in the realm of massive content, consumerization, compliance, social media, mobility, cloud and big data analytics.

Although it’s difficult to fundamentally transform a product architected two decades ago, vendors are desperately trying to do so, sometimes through development and often partnerships. How they get there, or whether they get there, remains to be seen.

At present, two companies started with visions closest to the one defined in this column. Cohesity meets most of the criteria, whereas Rubrik comes close but is missing some key pieces. CDM players such as Actifio and Catalogic also have much to offer, and will likely be reaching out to fill in the gaps. IBM Spectrum Copy Data Management utilizes Spectrum Protect, and IBM will (most likely) tightly integrate it more with IBM Cloud Object Storage to provide a scale-out capability. And lest we forget, both SimpliVity (recently acquired by Hewlett Packard Enterprise) and Scale Computing have publicly stated visions to apply hyper-convergence for both primary and secondary storage applications in a single infrastructure.

Led by public cloud vendors such as Amazon Web Services, the battle to transform primary data storage over the past decade has driven a staid industry to action. I believe hyper-convergence directly resulted from applying public cloud principles to on-premises infrastructures. I also believe we should apply a fresh approach based on hyper-convergence to the equally staid secondary-storage side of the data center. The need is even greater, as 80% of all data resides in secondary storage.

Either we choose to light secondary storage up and use it for better business decision-making or let it stagnate and become a bigger and murkier swamp. Yesterday’s architectures will not get us there. As with primary, a fresh approach based on hyper-convergence is needed.

ARUN TANEJA is founder and president at Taneja Group, an analyst and consulting group focused on storage and storage-centric server technologies.