G00277799

Hype Cycle for Storage Technologies, 2015

Published: 13 July 2015

Analyst(s): Pushan Rinnen, Santhosh Rao

This Hype Cycle evaluates 36 storage-related hardware and software technologies in terms of their business impact, adoption rate and maturity level to help users decide where and when to invest.

Table of Contents

List of Tables

List of Figures

Analysis

What You Need to Know

This Hype Cycle provides a consolidated view of 36 storage-related hardware and software technologies and their relative positions on a single Hype Cycle curve. New to the Storage Hype Cycle this year is a technology at the Peak of Inflated Expectations and was created for the Server Hype Cycle last year: "integrated systems: hyperconvegence." Storage is an important component (and often a major differentiator) of the hyperconverged systems. Other technologies that are at the peak of the cycle include software-defined storage, integrated backup appliances, cloud-based backup services, virtual storage appliance and open-source storage.

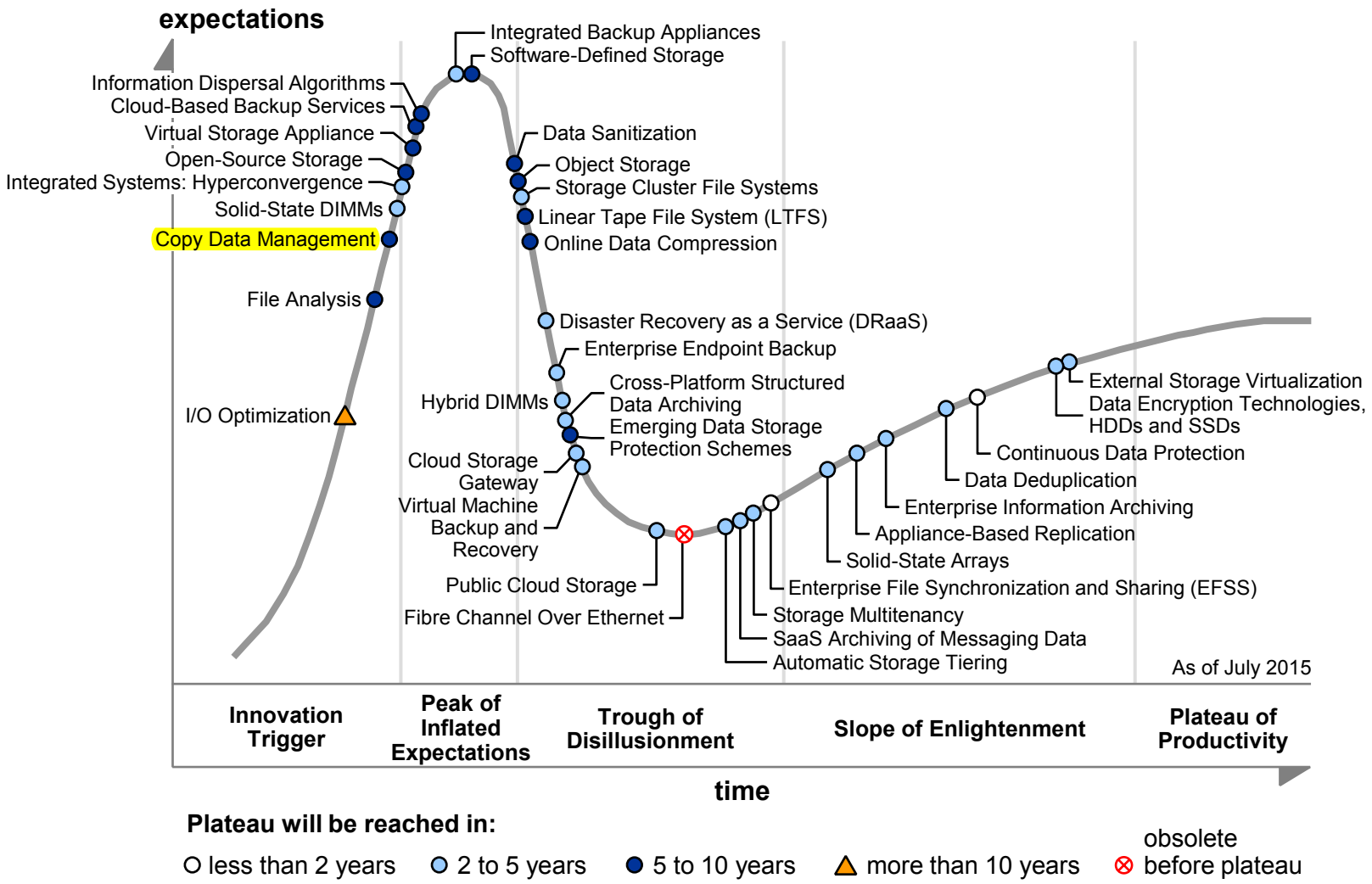
The technologies that have moved the farthest along the Hype Cycle curve in the past year include solid-state arrays, enterprise file synchronization and sharing (EFSS), cloud storage gateways, and integrated systems: hyperconverged. Solid-state arrays have taken the market by storm, and user adoption has grown tremendously in the past year. We have raised its business benefit to "transformational." While the EFSS technologies have evolved to the post-trough phase, becoming more mature, cloud storage gateways have moved toward the trough phase as product limitations are being discovered, triggering disillusionment. Hyperconveged integrated systems have moved to the pre-plateau phase, with 20% to 50% target market penetration.

Five technology profiles were retired this year due to their high maturity level and/or lack of future development (see list below). FCoE is listed as "obsolete before plateau," with a downgrade of the benefit rating to "low" due to lack of vendor support and operational complexities. The profile of read/write flash caching is merged with that of I/O optimization due to a large degree of overlap.

The Hype Cycle

This Hype Cycle illustrates the typical life cycle that a new storage technology goes through before it reaches mainstream adoption. This consolidated view provides users with a concise description of the technological landscape in the storage industry and helps clients understand the business benefits, risk areas and general adoption rate of each technology in order to shape their own plans for new technology adoption. Technologies that have fallen off the chart because of high maturity and widespread adoption may still be discussed in the Gartner IT Market Clock report. High-profile technologies with high and transformational business impacts tend to reach their Hype Cycle peak and trough quickly, while low-profile technologies or technologies with low or moderate business impact could move much more slowly on the curve and sometimes may never reach their peak before approaching the Trough of Disillusionment or becoming obsolete. Examples of fast-moving and high-impact technologies associated with storage include data deduplication, solid-state arrays and hyperconverged integrated systems. Examples of slow-moving technologies with moderate business impact include virtual storage appliance and external storage virtualization (see Figure 1).

Figure 1. Hype Cycle for Storage Technologies, 2015



Source: Gartner (July 2015)

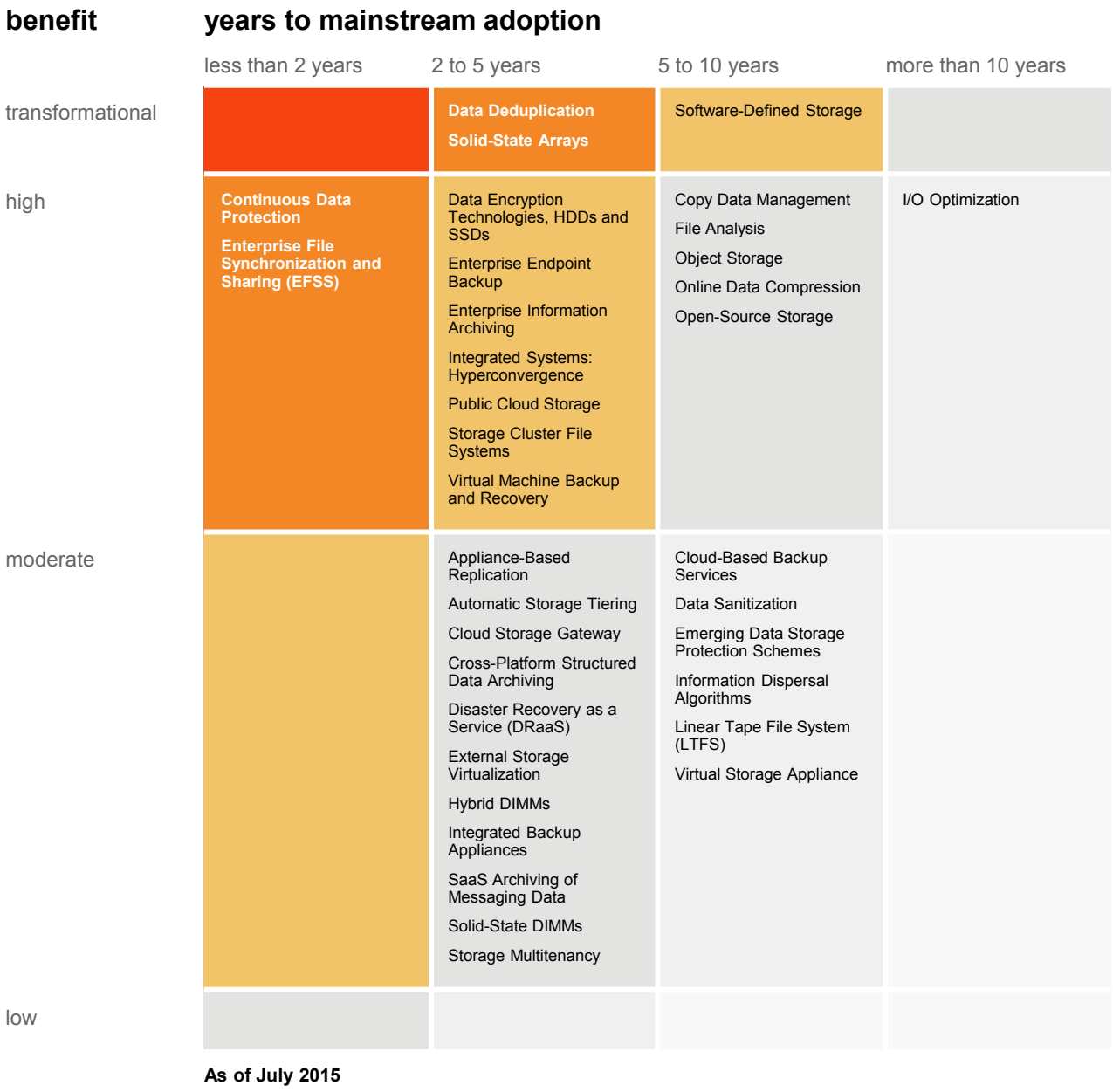
The Priority Matrix

The Priority Matrix maps the benefit rating for each technology against the length of time before Gartner expects it to reach the beginning of mainstream adoption. This alternative perspective can help users determine how to prioritize their storage hardware and storage software technology investments and adoption. In general, companies should begin in the upper-left corner of Figure 2, where technologies are rated transformational in business benefits and are likely to reach mainstream adoption quickly. These technologies tend to have the most dramatic impact on business processes, revenue or cost-cutting efforts. In the storage market, these transformational technologies include data deduplication and software-defined storage. This year, the business benefit for both solid-state arrays and software-defined storage has been raised from "high" to "transformational." Solid-state arrays now offer much richer data management in addition to other major benefits, including consistent performance without tuning, drastic reduction in storage footprint and high workload consolidation. With software-defined storage, decoupling storage software value from the underlying storage hardware has become instrumental in reducing cost in acquiring storage systems.

After these transformational technologies, users are advised to evaluate high-impact technologies that will reach mainstream adoption status in the near term and work downward and to the right from there. Examples of high-impact, near-term storage technologies include continuous data protection and enterprise file synchronization and sharing.

Figure 2 shows where the storage technologies evaluated in this year's Hype Cycle fall on the Priority Matrix. Note that this is Gartner's generic, cross-industry evaluation of the benefits of these technologies. Each organization's graph will differ based on its specific circumstances and goals.

Figure 2. Priority Matrix for Storage Technologies, 2015



Source: Gartner (July 2015)

Off the Hype Cycle

The following technologies have been removed from the Hype Cycle because they have reached the plateau and are considered "mature mainstream":

- Bare-metal restore
- Enterprise solid-state drives

- Server-based replication
- Storage resource management
- Thin provisioning

The profile of "read/write flash caching" has merged with "I/O optimization" due to a large degree of overlap.

On the Rise

I/O Optimization

Analysis By: Dave Russell; Arun Chandrasekaran

Definition: Input/output (I/O) optimization software aggregates server-side storage resources, such as memory, hard-disk drives or flash storage, across a single host or possibly multiple hosts, and delivers accelerated reads and, potentially, writes. In some cases, additional capabilities, such as fault tolerance, automated application tuning and other storage services can be delivered. I/O optimization software can run on a physical server, in a virtual machine (VM) or as part of a hypervisor.

Position and Adoption Speed Justification: The term "I/O blender" has been used to describe the circumstance where many VMs running on a single physical server make the I/O traffic appear very random. This randomness is caused by reads and writes that are being intermixed across several VMs, resulting in slower overall performance as storage array caching and prefetching algorithms become incapable of predicting what actions to take. To combat this, storage, virtualization and/or application administrators often overprovision storage to use more disk spindles, and use wide striping to leverage more disks. This results in lower storage utilization and higher costs.

Another issue can be that of an application running natively on a physical machine, and the underutilized resources (additional memory, hard disk and/or flash storage) not being shared to maximize value. I/O optimization software can help in both of these scenarios.

User Advice: With declining processor and flash storage costs, and the new memory capabilities that are becoming available, I/O optimization software has become a capability that application owners, server and virtualization teams, and storage administrators should evaluate for meeting their service-level objectives and/or their total cost of ownership (TCO) goals.

I/O optimization products are typically easy to deploy and easy to manage, and can offer real performance benefits and better resource utilization. Solutions can vary widely in terms of capabilities and hypervisor integration, and roadmaps are evolving, so due diligence is important in order to find the best solution for your requirements. Products can be workload-aware and self-optimizing, reducing the need for manual tuning and frequent administrative involvement. Going forward, the number and breadth of storage resources that can be optimized, the nondisruptive nature of deployment and the degree of automation will be the determining factors when selecting a product.

I/O optimization software works best with applications that generate many small block random writes. Virtual servers, virtual desktop infrastructure (VDI) and relational databases require a great deal of tuning and could benefit from this technology. I/O performance gains of 50% or more are common. The largest gains are typically achieved in VDI environments. Workloads that are bandwidth-constrained, such as backups, business analytics, Hadoop and many engineering applications that are characterized by large block sequential workloads, may not benefit from this technology.

I/O optimization capabilities could become embedded features in a hypervisor or some other solution, so validating additional benefits above what is available in other components is important.

Business Impact: I/O optimization products can improve storage response times and storage resource utilization, and may extend the usable life of previously installed infrastructure. Generally, these solutions are low cost compared to adding additional storage array capacity, and they are easily installed, making the time to value attractive. By better utilizing the existing infrastructure, I/O optimization can improve performance and help control costs.

Benefit Rating: High

Market Penetration: 1% to 5% of target audience

Maturity: Emerging

Sample Vendors: Atlantis Computing; CacheBox; Condusiv Technologies; Enmotus; Infinio; Intel; PernixData; Samsung; SanDisk; VMware; Western Digital

Recommended Reading: "Technology Overview for I/O Optimization Software"

"Cool Vendors in Storage Technologies, 2015"

"Innovation Insight: Separating Hype From Hope for Software-Defined Storage"

"Market Trends: The Next-Generation Server-Side SSD Landscape"

File Analysis

Analysis By: Alan Dayley

Definition: File analysis (FA) tools analyze, index, search, track and report on file metadata and, in some cases (e.g., in unstructured data environments), on file content. Usually offered as software, FA tools are provided by some vendors as features of storage arrays. FA tools differ from traditional storage reporting tools by reporting on simple file attributes and by providing detailed metadata and contextual information to enable better information governance and data management actions.

Position and Adoption Speed Justification: FA is an emerging technology that assists organizations in understanding the ever-growing repository of unstructured "dark" data, including file shares, email databases, SharePoint, etc. Metadata reports include data owner, location, duplicate copies, size, last accessed or modified, file types and custom metadata. Progressive and

cost-conscious organizations are moving past simply adding more storage capacity to address their storage problems and striving for a better understanding of their data. The desire to optimize storage cost, implement information governance and mitigate business risks (including security and privacy risks) are some of the key factors that impel the adoption of FA. The determination of file ownership and the enablement of more-accurate chargeback are also made available with FA.

User Advice: Organizations should use FA to better understand their unstructured data, including where it resides and who has access to it. Data visualization maps created by FA can be presented to other parts of the organization and be used to better identify the value and risk of the data, enabling IT, line of business and compliance organizations to make more-informed decisions regarding classification, information governance, storage management and content migration. Once known, redundant, outdated and trivial data can be defensibly deleted, and retention policies can be applied to other data.

Business Impact: FA tools reduce risk by identifying which files reside where and who has access to them, supporting remediation in such areas as the elimination of personally identifiable information, corralling and controlling intellectual property, and finding and eliminating redundant and outdated data that may lead to business difficulties, such as multiple copies of a contract. FA shrinks costs by reducing the amount of data stored. It also classifies valuable business data so that it can be more easily found and leveraged, and it supports e-discovery efforts for legal and regulatory investigations. In addition, FA products feed data into corporate retention initiatives by using file attributes.

Benefit Rating: High

Market Penetration: Less than 1% of target audience

Maturity: Emerging

Sample Vendors: Acaveo; Active Navigation; Bloomberg; CommVault Systems; DataGravity; HP (Autonomy); IBM-StoredIQ; STEALTHbits Technologies; Varonis; Veritas; Whitebox Security

Recommended Reading: "Market Guide for File Analysis Software"

"Save Millions in Storage Costs With These Effective Data Management Best Practices"

"Does File Analysis Have a Role in Your Data Management Strategy?"

"Best Practices for Data Retention and Policy Creation Will Lower Costs and Reduce Risks"

"Storage and Data Management Are Not the Same Thing"

Copy Data Management

Analysis By: Pushan Rinnen

Definition: Copy data management refers to products that use a live clone in a secondary storage system to consolidate, reduce and centrally manage multiple physical copies of production data

that is usually generated by different software tools and resides in separate storage locations. Such products have integration with host-side operating systems and applications, heterogeneous primary storage support, and additional data management capabilities for administrators and users such as granular file/object restore and user self-service recovery.

Position and Adoption Speed Justification: Many organizations have become acutely aware of the increasing cost of managing copy data, whose capacity is often significantly higher than production storage due to multiple copies for different use cases and less managed retention periods. Those copies could be snapshots, clones or replicas in primary storage arrays, and backup and remote replicas in various secondary storage (disk or tape). IT organizations have historically used different storage and software products to deliver backup, archive, replication, test/development, legacy application archiving and other data-intensive services with very little control or management across these services. This results in overinvestment in storage capacity, software licenses and operational expenditure costs associated with managing excessive storage and software. Copy data management facilitates the use of one copy of data for all of these functions via virtual copies, thereby dramatically reducing the need for multiple physical copies of data and enabling organizations to cut costs associated with multiple disparate software licenses and storage islands. The separation of the "golden image" from the production environment ensures minimum performance impact when copy data activities such as backup, replication or testing/development are performed; therefore, aggressive recovery point objectives (RPOs) and recovery time objectives (RTOs) can be achieved.

The concept of consolidating different copies is not new and has been tried by other vendors via different technologies with limited success due to product limitations and poor user education. For example, CommVault uses a single index engine, single pass data movement and a single storage repository for backup and archiving consolidation; and Catalogic Software (formerly a part of Syncsort) leverages NetApp's Data Ontap data protection/cloning mechanism to offer similar copy data management functions for NetApp customers. Catalogic recently expanded this functionality to VMware and IBM storage, whereas NetApp's Data Ontap 8.3 converges SnapVault and SnapMirror repositories into the same secondary storage system, which can also create thin clones for test and development. In the past few years, the emergence of the new architecture described in the first paragraph has gathered momentum as users start to understand and realize the benefit of copy data management. The main challenge faced by copy data management products is that they have to resonate with higher-level IT managers and architects, as adoption of such products to replace existing data protection, backup and disaster recovery (DR) solutions is usually a more strategic move and may disrupt the existing IT silos. In the case of test/development, copy data management improves the workflow process and operational efficiency by enabling database administrators and application developers more self-services capabilities.

User Advice: Copy data management is still an emerging concept with very few qualifying products in the market that can consolidate many types of copies. IT should look at copy data management as part of a backup modernization effort or when managing multiple application copies for testing/development has become costly, overwhelming or a bottleneck. Copy data management could also be useful for organizations that are looking for active access to secondary data sources for reporting or analytics due to its separation from the production environment. Due to the short

history of the new architecture and vendors, new use cases beyond the common ones are not field-proven and should be approached with caution.

Business Impact: The business impact of copy data management is threefold:

- It enables organizations to rethink and redesign their strategy for managing secondary copies of data to achieve operational efficiency, such as drastically enhanced RPOs and RTOs, and development cycles.
- It reduces the storage and management costs associated with various copies.
- It enables organizations to better leverage their secondary data for analytics, test/development and other non-mission-critical activities going forward.

Benefit Rating: High

Market Penetration: 1% to 5% of target audience

Maturity: Emerging

Sample Vendors: Actifio; Catalogic Software; Cohesity; Delphix; NetApp

Solid-State DIMMs

Analysis By: Michele Reitz

Definition: Solid-state dual in-line memory modules (SS DIMMs) are all-flash versions of nonvolatile DIMMs (NVDIMMs) that reside on the double data rate (DDR) DRAM memory channel and are persistent. These devices integrate NAND flash and a system controller chip. By sitting on the memory channel, they have key advantages over other types of solid-state drive (SSD) in terms of reduced write latency and increased input/output operations per second, bandwidth and scalability.

Position and Adoption Speed Justification: Solid-State DIMMs were introduced in 2014, when IBM debuted its eXFlash device (now owned by Lenovo), which is included in SanDisk's ULLtraDIMM line of products through a partnership with Diablo Technologies. Since DIMMs sit directly on the faster memory channel, rather than on the storage channel, they will not face storage channel bottlenecks of a traditional storage system. Because of this, these NAND flash-based SSDs can achieve drastically lower latencies (at least 50% lower) than any existing solid-state storage solution, and can be viable alternatives to DRAM memory, if the speeds are acceptable.

The NVDIMM Special Interest Group, a consortium within the Storage Networking Industry Association (SNIA), classifies two types of NVDIMM — NVDIMM-N and NVDIMM-F. Gartner classifies NVDIMM-N as a hybrid DIMM and NVDIMM-F as a solid-state DIMM.

Use of any solid-state DIMMs requires a mix or all of the following: support by the host chipset, optimization for the OS and optimization for the server hardware. As such, to achieve greater adoption, server, driver and OS support will need to extend beyond IBM and Supermicro on selected platforms. In addition, use cases for memory channel products will need to spread beyond

the extremely high-performance, high-bandwidth and ultra-low-latency applications for which they are attracting most interest today.

Despite these challenges, this technology will mature in the next two to five years as users will find the price/performance value proposition a good match for in-memory computing, cloud, virtualization, virtual desktops, big data and analytics applications. We have moved this technology profile up the curve closer to the peak due to two factors: first, the ongoing litigation that had challenged Diablo's market penetration has been ruled for Diablo, allowing them to resume their shipments going forward; and, second, Lenovo announced expanded support in May 2015, and we anticipate more vendors will follow.

User Advice: IT professionals should evaluate solid-state DIMMs for use as a new tier of storage, if ultra-low latency is important. They should use the significant improvement in flash response times — nearly half those of conventional SSDs — and the denser form factor to meet increased requirements for overall system storage.

IT professionals should analyze the roadmaps of the major server and storage OEMs, along with those of the SSD appliance vendors that will be launching DIMM-based storage systems, and weigh the benefits for their needs. They should be aware that servers, OSs and drivers will need to be customized to support these new types of DIMMs.

Business Impact: This technology's impact on users will be improved system performance overall. It will also offer an alternative to DRAM for certain in-memory computing applications that need support for large data stores and can sacrifice performance.

With DRAM prices at a premium compared with NAND flash, solid-state DIMMs will have a much-improved price/performance ratio when it meets users' performance needs.

NAND flash vendors should consider solid-state DIMMs to enhance their value proposition for commodity NAND flash and expand their addressable market.

Benefit Rating: Moderate

Market Penetration: Less than 1% of target audience

Maturity: Emerging

Sample Vendors: Diablo Technologies; IBM; SanDisk; Supermicro

At the Peak

Integrated Systems: Hyperconvergence

Analysis By: George J. Weiss; Andrew Butler

Definition: Hyperconverged systems are integrated systems that apply a modular building block approach to scale compute and direct-attached storage on commodity hardware by leveraging scale-out clusters of standardized building blocks.

Position and Adoption Speed Justification: Hyperconverged infrastructure is a nascent market segment that is gaining in mind share and adoption. By 2018, hyperconverged integrated systems (HCISs) will represent 30% to 35% of total converged infrastructure shipments by revenue from the current low-single-digit base in 2015. HCIS enables IT to start from a small base — in some cases a single node — and easily scale out as demand requires. The modular-building-block approach allows enterprises to take small incremental steps, rather than make the significant upfront investments required by traditional integrated systems. Traditional integrated systems typically have a proprietary chassis with fabric infrastructure, which tends to be costly. We expect that the hyperconverged infrastructure market will continue in the expansion phase with more new entrants, additional feature/function deliverables and broader vendor portfolios to address mixed workloads. As the growth of hyperconvergence begins peaking in the 2018 to 2020 period, the market will consolidate, and vendor survival and viability will be increasingly important issues in the selection process.

User Advice: IT leaders should recognize HCIS as an evolution within the broader category of integrated systems that lays the foundation to ease of use, simplicity, virtualization, cloud deployment and eventual bimodal implementations. IT should be able to harness its fundamental advantages in efficiency, utilization, agility, data protection, continued life cycle deployment and orchestration as part of a strategic data center modernization objective. Plan strategically but invest tactically in hyperconverged systems because the market is nascent and subject to turmoil. Plan for payback of two years or less to ensure financial success and investment value. Test scalability limits of solutions, as improvements will occur rapidly during this nascent period.

Business Impact: While integrated systems will be generally driven by new workloads and data center modernization initiatives, the hyperconverged portion has been jumpstarted primarily in the midsize part of the market. Use cases especially well-suited to HCIS include: virtual desktop infrastructure, server-based computing, data migration, private cloud, remote or branch office, relational databases, Hadoop and dedicated application infrastructure.

Benefit Rating: High

Market Penetration: 1% to 5% of target audience

Maturity: Adolescent

Sample Vendors: Atlantis; Dell; EMC; Gridstore; HP; Maxta; Nimboxx; Nutanix; Pivot3; Scale Computing; SimpliVity

Recommended Reading: "How to Evaluate Vendors in the Hyperconverged Space"

"Best Practices for Adopting Converged Infrastructure in Times of Market Turbulence"

"Market Trends: Hyperconverged Integrated Systems Meet Growing Acceptance From Midsize Enterprises, 2015"

"Plan Now for the Future of Converged Infrastructure"

"The Coming Converged Infrastructure Vendor Battle and What to Do About It"

"Deploying Hyperconverged Integrated Systems: Eight Great Use Cases"

Open-Source Storage

Analysis By: Arun Chandrasekaran

Definition: Open-source storage is core storage software that is used to create a storage array, as well as data protection and management software. It involves software abstracted from the underlying hardware for which the source code is made available to the public through free license. Similar to proprietary storage, open-source storage software supports primary, secondary and tertiary storage tiers, as well as heterogeneous management.

Position and Adoption Speed Justification: Although open-source storage has been around for a long time, it has been mainly relegated to file-serving deployments in small business environments. Products such as FreeNAS (TrueNAS for business environments), Openfiler and Amanda have been in use for many years. Recent innovations in multicore processors and CPU core density, combined with an innovative open-source ecosystem, are making open-source storage attractive for cloud and big data workloads and as a potential alternative to proprietary storage. As cloud computing, big data analytics and information archiving push the capacity, pricing and performance frontiers of traditional scale-up storage architectures, there has been renewed interest in open-source software (OSS) as a means to achieve high scalability in capacity and performance at lower acquisition costs.

The rise of open-source platforms such as Apache Hadoop and OpenStack, which are backed by a large, innovative community of developers and vendors, together with the entry of disruptive vendors such as Red Hat (Gluster, Ceph) and Intel (Lustre), is enabling enterprises to seriously consider open-source storage for use cases such as cloud storage, big data and archiving. More vendors are also bringing products to market based on the popular OpenZFS project. Even traditional, proprietary storage vendors such as EMC are now open sourcing key projects (such as ViPR). Gartner expects more open-source options to be available in near future.

User Advice: Although open-source storage offers a less-expensive upfront alternative to proprietary storage, IT leaders need to measure the benefits, risks and costs accurately. Some enterprise IT organizations overstate the benefits, and understate the costs and risks. Conversely, with the emerging maturity of open-source storage solutions, enterprise IT buyers should not overlook the value proposition of these solutions. IT leaders should actively deploy pilot projects, identify internal champions, train storage teams and prepare the overall organization for this disruptive trend. Although source code can be downloaded for free, it is advisable to use a commercial distribution and obtain support through a vendor, because OSS requires significant effort and expertise to install, maintain and support.

Customers deploying "open core" or "freemium" storage products need to carefully evaluate the strength of lock-in against the perceived benefits. This is a model in which the vendor provides proprietary software — in the form of add-on modules or management tools — that functions on top of OSS.

In most cases, open-source storage is not general-purpose storage. Therefore, choose use cases that leverage the strengths of open-source platforms — for example, batch processing or a low-cost archive for Hadoop and test/development private cloud for OpenStack — and use them appropriately. On-premises integration, management automation and customer support should be key priorities when selecting open-source storage solutions.

Business Impact: Open-source storage is playing an important role in enabling cost-effective, scalable platforms for new cloud and big data workloads. Gartner is seeing rapid adoption among technology firms and service providers, as well in research and academic environments. Big data and private cloud use cases in enterprises are also promising use cases for open-source storage, where Gartner is witnessing keen interest. As data continues to grow at a frantic pace, open-source storage will enable customers to store and maintain data, particularly unstructured data, at a lower acquisition cost, with "good enough" availability, performance and manageability.

Benefit Rating: High

Market Penetration: 1% to 5% of target audience

Maturity: Emerging

Sample Vendors: Basho Technologies; Cloudera; Hortonworks; IBM; Nexenta; OpenStack; Pivotal Labs; Red Hat; SwiftStack

Recommended Reading: "Market Guide for Open-Source Storage"

"IT Leaders Can Benefit From Disruptive Innovation in the Storage Industry"

"Should I Use Open Source in My Infrastructure?"

Virtual Storage Appliance

Analysis By: Arun Chandrasekaran; Dave Russell

Definition: A virtual storage appliance (VSA) is software that runs on a virtual machine (VM) to provide data management of a traditional storage solution and present block, file, and/or object access to users and applications. A VSA uses the host system's resources — such as memory, hard-disk drives (HDDs) or solid-state drives (SSDs) — for all storage capabilities, while providing data services and a management layer on its own or in conjunction with the hypervisor.

Position and Adoption Speed Justification: With the growth in server virtualization, there has been increased interest in reducing the storage total cost of ownership (TCO) and performance challenges that are common to many virtual environments. By placing storage close to the compute layer, VSAs reduce TCO by leveraging commodity direct-attached storage (DAS) and server resources that are embedded in the hosts and boost performance by eliminating the input/output (I/O) latency. VSA products are maturing, and they span primary storage, cloud tiering and backup use cases.

The number of VSA products has increased in the recent years, although most primary storage deployments tend to be among small or midsize businesses (SMBs) or remote offices, or are applied to specific use cases, such as server virtualization or virtual desktop infrastructures (VDIs) in enterprises. The growing use of public cloud storage is spawning additional investments in VSAs, as well as the market entry of VSAs that are focused solely on cloud environments.

User Advice: With growing vendor maturity and richer software features, today's VSAs offer more than just basic primary storage or backup/archiving targets. However, they are still often limited by scalability challenges, due to limited server-attached storage resources. Their initial inherent simplicity and cost-effectiveness makes them appealing to enterprises looking to simplify storage in remote office, hybrid cloud and VDI deployments.

VSA licensing costs tend to be more predictable, because licensing is often capacity-based. Although VSAs are supposed to be hardware- and hypervisor-agnostic, most VSAs support only VMware vSphere or, in some cases, Microsoft Hyper-V. VSAs are often deployed for niche use cases and, as a result, can be isolated from a management and integration standpoint.

Demarcation among the vendors — server, hypervisor, storage software and application providers — tends to be more complex, sometimes resulting in a lack of accountability. In such scenarios, end users should look at vendors with certified solutions and joint support contracts. More vendors are decoupling their appliances and are willing to offer VSAs, with a few deciding to offer free community editions that are limited by scale, making it easier for customers to "try and buy."

Business Impact: Virtual storage appliances can deliver cost savings and operational simplicity when deployed for appropriate use cases. They can reduce vendor lock-in through hardware abstraction and enable better utilization of hardware resources; however, the training and management, and especially the data services of the solution, can still be a form of lock-in. As hypervisors become more pervasive, Gartner expects more vendors to offer VSAs, which could unravel the system approach of storage megavendors and drive them toward an approach of storage software running on user-selected, precertified hardware.

Benefit Rating: Moderate

Market Penetration: 1% to 5% of target audience

Maturity: Adolescent

Sample Vendors: EMC; HP; NetApp; Nexenta; Panzura; StorMagic; TwinStrata; VMware

Recommended Reading: "IT Leaders Can Benefit From Disruptive Innovation in the Storage Industry"

"Innovation Insight: Separating Hype From Hope for Software-Defined Storage"

Cloud-Based Backup Services

Analysis By: Pushan Rinnen

Definition: Also known as backup as a service (BaaS), cloud-based backup services aim to replace or augment traditional on-premises backup with three main deployment models: (1) using local host agents to send backup data directly to the cloud data centers; (2) backing up first to a local device, which in turn sends backup data to the cloud either as another replica or as a lower tier; and (3) backing up data that is generated in the cloud.

Position and Adoption Speed Justification: Larger network bandwidth at more affordable prices, customers' increased comfort with cloud storage security, and their desire to achieve better operational efficiency than tape backup and to obtain additional protection than natively in the cloud are driving forces behind cloud-based backup.

The first deployment model described in the definition is the traditional online backup, increasingly used for endpoint backup and server backup for small businesses and small branch offices with limited amounts of data, aiming to eliminate the hassle of managing local backup. The down side of online backup is limited performance without a local backup copy. The second deployment model is also called hybrid cloud backup, where the local device offers much faster backup and restore capabilities due to the use of local networks instead of the Internet or WAN. It therefore can scale to a much larger server environment than the first scenario. All successful cloud server backup providers offer local device support. The more innovative solutions also offer integrated cloud backup and cloud disaster recovery (DR), where the backup copies stored in the cloud could be used to boot up standby virtual machines in the cloud for fast failover. Recently, major backup software ISVs have started enhancing their cloud connectors to copy local backup data to the cloud, either natively or via a cloud gateway, or to introduce backup application virtual editions to run on VMs in Web-scale public cloud to enable deduplication and in-cloud recovery. The third deployment model is still nascent: backup of data generated in the public cloud. Such cloud-born data include nontransient data created in public cloud compute instances, as well as cloud-native applications such as Google Apps, Microsoft Office 365 and Salesforce. While all public cloud providers offer various degrees of basic data protection functions within their own cloud, customers desiring additional backup functions not offered natively have to look for third-party backup tools.

While user interest in cloud backup is at an all-time high, there are still obstacles that prevent fast adoption. First, this is a highly fragmented market with many small service providers. And second, there is no industry benchmark or standard in terms of recovery point objective (RPO)/recovery time objective (RTO), retention and pricing.

User Advice: Cloud backup may not be less expensive than traditional backup, especially when retention period in the cloud is long. It could be much simpler than tape backup; but it could also be more complex than local disk backup at the adoption stage because of the additional consideration of networks, security and, possibly, a separate relationship with an IaaS storage provider. However, once implemented successfully, cloud backup can eliminate or offload daily management overhead for on-premises backup. Technological limitations make backing up to the cloud largely impractical for environments with 30TB or more of production server data. Gartner recommends the hybrid approach for environments with 500GB of daily incremental backup or restore workloads for most server environments today.

Business Impact: Cloud server backup is often used to replace traditional tape off-site backup, eliminating the daily operational complexities associated with tape backup and the management of removable media. Solutions offering integrated cloud backup and cloud DR provide small organizations with the business continuity benefits they couldn't afford before. Although the business impact for small businesses is high, its impact for large enterprises is low today.

Benefit Rating: Moderate

Market Penetration: 5% to 20% of target audience

Maturity: Emerging

Sample Vendors: Acronis; Asigra; Axcient; Barracuda Networks; Code42; Ctera Networks; Datacastle; Datto; Druva; EMC; HP Autonomy; Infrascale; Microsoft; Seagate Technology; Unitrends; Zerto; Zetta

Recommended Reading: "How to Determine If Cloud Backup Is Right for Your Servers"

"Exploring Common Cloud Backup Options"

"Pricing Differences Complicate the Adoption of Backup and Disaster Recovery as a Service"

Information Dispersal Algorithms

Analysis By: Valdis Filks

Definition: Information dispersal algorithms provide a methodology for storing information in pieces (i.e., dispersed) across multiple locations, so that redundancy protects the information in the event of localized outages, and unauthorized data access at a single location does not provide usable information. Only the originator or a user with a list of the latest pointers created by the original dispersal algorithm can properly assemble the complete information.

Position and Adoption Speed Justification: New commercial solutions have become available for the data center from large, established vendors and smaller startups for domestic use and file sync and share. The solutions are also built-in and are available in home consumer storage appliances. However, they differ from the presently prevailing centralized cloud storage offerings as these solutions are not centralized, but distributed and, similar to the Internet, have no central control or fault domain. The Information dispersal algorithm technology has been expanded to include peer-to-peer (P2P) file-sharing technologies and protocols, such as those based on the BitTorrent protocol, which has proved robust on the Internet. A variation is the open-source BitTorrent protocol used in P2P networks to store and recreate data among systems. This is an early cloud technology in which the data is truly dispersed, rather than stored in a small number of centralized, hyperscale or traditional data centers. Therefore, fault tolerance is provided by the nature of the design of these systems, and, due to the dispersal nature of the data, some protection is also provided by their design. Due to their innate design, many scale-out storage systems are implementing redundant array of independent disks (RAID) designs that disperse data among nodes within racks. This technology is expected to develop into geographically distributed, file-sharing nodes, blurring the lines between scale-out storage systems, information dispersal algorithms and cloud storage.

User Advice: Customers who are not satisfied with the centralized cloud storage offerings should investigate information dispersal algorithms as they reduce customer dependence on a few large hyperscale vendors and locations that still use the traditional centralized data center design. In many ways, these algorithms are tantamount to a form of encryption. The design, coding and testing of the attack resistance of dispersion algorithms is similarly complex to the design of encryption implementations, which have proved to be a difficult undertaking. Just as proprietary forms of encryption should not be considered as reliable as implementations based on well-proven algorithms and code, the robustness of proprietary dispersal algorithms — and especially their implementations — should not automatically be considered trusted code. Buyers that expect to rely on this technology for confidentiality control should seek evidence from the tester at high levels of testing and from peer review.

Business Impact: Information dispersal algorithms could eventually provide secure storage over the Internet and other public or private networks without the overhead and other costs of encryption and the need to have centralized hyperscale data centers, such as those from Amazon and Google. Use of the BitTorrent protocol has been political, because one of its early applications was to share copyrighted data via the Internet among home PCs. However, the protocol is content-neutral and simple to use. It could just as easily be used by software companies to distribute software, updates and any digital information that is stored and geographically dispersed among many nodes and computers in a network.

Open-source implementations are integrated into products by commercial companies as a new method to distribute and store digital data. This is one factor that increases the amount of unstructured data stored on the planet.

Benefit Rating: Moderate

Market Penetration: 1% to 5% of target audience

Maturity: Adolescent

Sample Vendors: Amplidata; BitTorrent; Caringo; Cleversafe; EMC; Security First Corp.; Symform; Tally-WariZen; Vivint

Recommended Reading: "Traditional Storage Vendors, Brands and Products Are No Longer Risk-Free"

"Increases in Disk Capacity Are Affecting RAID Recovery; Start to Purchase New RAID Technologies"

Integrated Backup Appliances

Analysis By: Pushan Rinnen

Definition: An integrated backup appliance is an all-in-one backup software and hardware appliance that combines the functions of a backup application server, media server (if applicable) and backup target device. The appliance is typically preconfigured and fine-tuned to cater to the

capabilities of the onboard backup software. It is a more simplified and easier-to-deploy backup solution than the traditional approach of separate software and hardware installations, but lacks flexibility on hardware choices and hardware scalability.

Position and Adoption Speed Justification: Integrated backup appliances have been around for many years without much fanfare. The current hype is driven by existing large backup software vendors that have started packaging their software in an appliance, and by innovative small vendors that are offering all-in-one appliances. In 2015, more backup software vendors introduced integrated appliances, fueling the hype to the peak. The momentum of integrated backup appliances is driven by the desire to simplify the setup and management of the backup infrastructure, as "complexity" is a leading challenge when it comes to backup management. Overall, integrated backup appliances have resonated well with many small and midsize customers that are attracted by the one-stop-shop support experience and tight integration between software and hardware. As the appliances scale up, they will be deployed in larger environments. There are generally three types of vendors selling integrated backup appliances. The first kind includes backup software vendors that package their software with hardware to offer customers integrated appliances. Examples include Arcserve, CommVault, Dell, EMC Avamar, Symantec, and Unitrends. Gartner evaluates this type of integrated appliance with backup software in the same Magic Quadrant because the key values reside on the backup software side. The second type integrates third-party backup software with its own hardware; examples include Fujitsu and STORServer. The third kind is a cloud backup provider that offers a customer an on-premises backup appliance as part of its cloud backup solution. Examples include Axcient, Barracuda Networks and Seagate Technology. The cloud backup appliances typically offer local storage to store backup data locally for faster backup and restore over the local-area network.

User Advice: Organizations should evaluate backup software functions first to ensure that their business requirements are met before making a decision about acquiring an integrated backup appliance or a software-only solution. Once a specific backup software product is chosen, deploying an appliance with that software will simplify operational processes and address any compatibility issues between backup software-only products and deduplication backup target appliances. If customers prefer deploying backup software only to gain hardware flexibility, they should carefully consider which back-end storage to choose — be it generic disk array/network-attached storage (NAS) or deduplication backup target appliances.

Business Impact: Integrated backup appliances ride the current trend of converged infrastructure and offer tight integration between software and hardware, simplify the initial purchase and configuration process, and provide the one-vendor support experience with no finger-pointing risks. On the down side, an integrated backup appliance tends to lack the flexibility and heterogeneous hardware support offered by backup-software-only solutions, which are often needed by large, complex environments.

Benefit Rating: Moderate

Market Penetration: 5% to 20% of target audience

Maturity: Adolescent

Sample Vendors: Actifio; Arcserve; Axcient; Barracuda Networks; CommVault; Dell; EMC Avamar; FalconStor Software; Fujitsu; Infrascale; Seagate Technology; Symantec; Unitrends

Recommended Reading: "Magic Quadrant for Enterprise Backup Software and Integrated Appliances"

"Magic Quadrant for Deduplication Backup Target Appliances"

Software-Defined Storage

Analysis By: Dave Russell; Arun Chandrasekaran

Definition: Software-defined storage (SDS) abstracts storage capabilities dynamically derived from physical or virtual devices and/or services to offer agility and deliver quality of service (QoS) while optimizing costs. Services are orchestrated via interoperable, programmable interfaces through the software layers that are separated into a control plane (management and policy implementation) and data planes (infrastructure and data transport), independent of where data is placed and stored to meet a defined policy or SLA.

Position and Adoption Speed Justification: While SDS is still largely a vision, it is a powerful notion that could revolutionize storage architectural approaches and storage consumption models over time. Foundational to SDS is the concept of abstracting and separating physical or virtual storage services via bifurcating the control plane (action signals) regarding storage from the data plane (how data actually flows). This is achieved largely through programmable interfaces (such as APIs), which are still evolving. SDS requests will negotiate capabilities through software that, in turn, will translate those capabilities into storage services that meet a defined policy or SLA. Whereas storage virtualization abstracts storage resources, which is foundational to SDS, the concepts of policy-based automation and orchestration, possibly triggered and managed by applications and hypervisors, are key to SDS.

The goal of SDS is to deliver greater business value than traditional implementations via better linkage of storage to the rest of IT, improved agility and cost optimization, with the higher-order capability being driven by policy management such that automation and storage administration are simplified, with less manual oversight being required.

User Advice: Gartner's opinion is that SDS is more than simply delivering software that runs on commodity hardware; that is, SDS is more than just repackaging of existing storage capabilities (see the "Virtual Storage Appliance" technology profile). Greater policy management extends the potential ecosystem of providers for data services, and could also reduce costs associated with storage hardware. That said, SDS is an evolving concept with much hype. Expect evolving vendor roadmaps, consolidation and acquisitions as leaders struggle to become relevant in an emerging marketplace. Today, interoperability across SDS products, or even other software-defined data center products, is minimal to nonexistent. Past multivendor storage initiatives have not fully delivered the desired results. Gartner views SDS as a continuum of storage capabilities, with some products realizing more of the SDS vision than others.

SDS is evolving something to keep abreast of the potential for cost improvement and enhanced manageability. Gartner recommends proof of concept (POC) implementations to determine suitability for broader deployment. Top reasons for interest in SDS, as gathered from interactions with Gartner clients, include:

- Improving the management and agility of the overall storage infrastructure through better programmability, interoperability, automation and orchestration
- Better linkage of storage to the rest of the IT
- Operating expenditure (opex) reductions by reducing the demands of administrators
- Capital expenditure (capex) reductions from more efficient utilization of proprietary storage systems and potentially the usage of commodity components to deliver storage services

Business Impact: SDS's ultimate endgame as part of a broader software-defined data center is ultimately very enticing in two ways. First, in the storage domain, the notion of optimizing storage hardware expenses via the broad deployment of underlying commodity hardware platforms under the direction of a robust, ideally heterogeneous, policy manager has great potential value. Second, in the data center as a whole, enabling multitenant data and workload mobility between servers, data centers and cloud providers without disrupting application and data services would be transformational. Agility-related benefits include workload optimization with potentially intelligent data placement (dynamically moving hot data closer to compute and cold data away from compute) and automated storage provisioning and decommissioning.

Benefit Rating: Transformational

Market Penetration: 1% to 5% of target audience

Maturity: Emerging

Sample Vendors: Atlantis Computing; Citrix; DataCore Software; EMC Software; IBM; Maxta; Nexenta; Red Hat; VMware

Recommended Reading: "Innovation Insight: Separating Hype From Hope for Software-Defined Storage"

"Cool Vendors in Storage Technologies, 2015"

"Should Your Enterprise Deploy a Software-Defined Data Center?"

"What Is the Value of a Software-Defined Data Center?"

"Solutions to Avoid SDS Repeating SRM Problems"

Data Sanitization

Analysis By: Philip Dawson; Rob Schafer

Definition: Data sanitization (formerly "data wiping") is the consistently applied, disciplined process of reliably and completely removing all data from a read/write medium so that it can no longer be read or recovered.

Position and Adoption Speed Justification: Growing concerns about data privacy and security, leakage, regulatory compliance, and the ever-expanding capacity of storage media are making robust data sanitization a core competency for all IT organizations.

This competency should be applied to all devices with storage components (such as PCs, mobile phones, tablets, and high-end printers and copiers) when they are being repurposed, returned to the supplier/lessor, sold, donated to charity or otherwise disposed of. Where organizations lack this robust data sanitization competency, it is often due to handling the various stages of the asset life cycle as isolated events, with little coordination between business boundaries (such as finance, security, procurement and IT). Thus, the personnel assigned to IT asset disposition (ITAD) are often different from those responsible for risk management and compliance, which can put the organization at risk of both internal and external noncompliance.

For mobile devices, a remote data-wiping capability is commonly implemented, triggered either by the user logging into a website or an administrator remotely invoking a mobile device manager (MDM). Although a remote capability like this should not be considered a fail-safe mechanism, reliability should be adequate for a significant majority of lost or stolen mobile devices. The degree to which various hardware storage technologies are reliably wiped varies according to organization type and device type.

User Advice: Follow a life cycle process approach to IT risk management that includes making an explicit decision about data sanitization and destruction, device reuse and retirement, and data archiving.

Implement policies that assign responsibility for all media carrying sensitive or regulated data — whether corporate or personal — to ensure that they are properly wiped or destroyed at the end of their production use.

Create appropriate data sanitization/destruction standards that provide specific guidance on the destruction process, based on data sensitivity.

Verify that your ITAD vendor consistently meets your data sanitization security specifications and standards.

Organizations that have yet to address CDs/DVDs and other portable data-bearing devices are even less prepared to deal with the implications of personal devices and plug-and-play storage.

Consider using whole-volume encryption for portable devices and laptops, and self-encrypting devices in the data center.

Organizations with storage devices containing highly sensitive and/or regulated data (e.g., in the financial and healthcare industries) should consider destroying the devices, either by mechanical

means or by using degaussing machines, rendering them permanently unusable and ensuring that the data is not recoverable.

Some undeleted but largely inaccessible data remains on most USB memory sticks, which is reason enough to forbid their use for sensitive, unencrypted files.

Buyers of any form of externally provisioned service should understand end-of-contract implications, and ask current and potential providers for an explanation of their storage reuse and account retirement practices.

Business Impact: At a relatively low cost, the proper use of encryption, wiping and, when necessary, destruction will help minimize the risk that proprietary and regulated data will leak.

By limiting data sanitization to encryption and/or software wiping, organizations can preserve the asset's residual market value; the *destruction* of data bearing devices within an IT asset typically reduces the asset's residual value (RV) to salvage, incurring the cost of environmentally compliant recycling.

The National Association for Information Destruction (NAID) supports best practices in data destruction services, and offers a list of service providers. See also the National Institute of Standards and Technology (NIST) December 2014 revision of its Special Publication 800-88: "Guidelines for Media Sanitization."

Benefit Rating: Moderate

Market Penetration: 20% to 50% of target audience

Maturity: Early mainstream

Sample Vendors: Blancco; DestructData; ITRenew; Kroll Ontrack; Tabernus

Recommended Reading: "Magic Quadrant for IT Asset Disposition, Worldwide"

"IT Asset Disposition Market Dynamics: The Global Secondary PC Market Challenge"

"Protecting Data on Decommissioned SSDs and HDDs"

"Best Practices for Data Destruction"

Sliding Into the Trough

Object Storage

Analysis By: Arun Chandrasekaran

Definition: An object is a logical collection of bytes with attributes and security policies. Unlike file- or block-based storage, object storage files, images and data blocks are stored as individual

objects or as elements of an object, and are associated with a unique identifier. Unique identifications are created via hash algorithms and maintained in an index database.

Position and Adoption Speed Justification: The first generation of object storage products emerged in early 2000 in the form of content-addressed storage (CAS) for compliance-centric archiving. With a new focus on cloud storage to emphasize data mobility and storage scalability, the next generation of object storage arrived prior to 2010. With data and metadata stored together, object storage provides easy manageability and migration for long-term storage. Although some cloud providers, such as Amazon and Facebook, built their own object storage software, storage system vendors are either building new products or repositioning their available products to target cloud service providers or enterprises planning to build private or hybrid clouds for collaboration, media streaming and data archiving.

Amazon S3 has achieved success as persistent storage with more than 2 trillion objects, and OpenStack (Swift) has become popular with large-scale installations at Disney, Rackspace and IBM. Broad adoption of second-generation commercial object storage has remained low so far; however, it is now accelerating, due to the need for cost-effective big data storage. The growing maturity of solutions from emerging vendors and refreshed products from large storage portfolio vendors is expected to further stimulate adoption from end users, as the addressable use cases for these products increase. More vendors are positioning object storage for file sync and share and distributed content use cases. These new use cases are starting to gain traction.

User Advice: IT leaders that require highly scalable, self-healing and cost-effective storage platforms for unstructured data should evaluate the suitability of object storage products. The common use cases that Gartner sees for object storage are archiving, cloud storage and content distribution. When building on-premises object storage repositories, customers should evaluate the product's API support for dominant public cloud providers, so that they can extend their workloads to a public cloud, if needed. Amazon's S3 has emerged as the dominant API, with more vendors starting to support OpenStack Swift API as well. Users should require that object storage vendors develop easy policies to migrate data from primary storage and use popular cloud APIs to minimize device and vendor lock-in and to facilitate easier migration.

Business Impact: Rapid growth in unstructured data (40% year over year) and the need to store and retrieve it in a cost-effective, automated manner will drive the growth of object storage. Object storage can be effective in big data environments, which scale from a few hundred terabytes to tens of petabytes. Because objects exist as self-describing logical entities coupled with metadata and controlled by policies, they can scale effectively. Object storage is also well-suited to multitenant environments that need stringent object-level security and rich metadata for easy automation and management. Object storage software, deployed on commodity hardware, is emerging as a threat to external controller-based (ECB) storage hardware vendors in big data environments with heavy volume challenges.

Benefit Rating: High

Market Penetration: 5% to 20% of target audience

Maturity: Adolescent

Sample Vendors: Basho Technologies; Caringo; Cleversafe; Cloudian; DataDirect Networks; EMC; Hitachi Data Systems; HP; Huawei; Imation; NetApp; OpenStack; Quantum; Red Hat; Scality; Tarmin Technologies; Western Digital

Recommended Reading: "Market Guide for Open-Source Storage"

Storage Cluster File Systems

Analysis By: Arun Chandrasekaran

Definition: Storage cluster file systems use one file system to cluster multiple storage nodes together, presenting a single image, a shared storage pool, and high bandwidth for multiple servers and clients. Data is distributed over multiple nodes in the cluster to gain high throughput and maintain data integrity.

Position and Adoption Speed Justification: The growing strategic importance of storing and analyzing large-scale, unstructured data is bringing scale-out storage architectures to the forefront of IT infrastructure planning. Storage vendors are continuing to develop cluster file systems to address high-latency and scalability limitations in traditional, scale-up network-attached storage (NAS) environments. This makes them suitable for batch and interactive processing and other high-bandwidth workloads. Although initial adoption was limited to academic high-performance computing (HPC) environments, commercial vertical industries — such as oil and gas, financial services, media and entertainment, life sciences, research and Web services — are beginning to use applications that require highly scalable storage bandwidth. This has led to an increase in market adoption for storage cluster file systems in the commercial sector.

Beyond the HPC use case, large home directories storage, rich-media streaming, backup and archiving are other common use cases for cluster file systems. Products from vendors such as Panasas, DataDirect Networks (DDN) and Intel are most common in HPC environments. Most leading storage vendors, such as EMC, HP, IBM and Dell, as well as emerging vendors, such as Red Hat, have enhanced their presence in this segment. A new breed of hyperconverged infrastructure vendors (e.g., Nutanix, SimpliVity and Scale Computing) are taking advantage of distributed file systems to offer converged computing and storage into an integrated appliance that can scale out in a near-linear manner.

Hadoop Distributed File System (HDFS), is starting to see wide enterprise adoption for big data, batch processing use cases and beyond. With the growing demand for high input/output operations per second (IOPS) and aggregated bandwidth for shared storage, cluster file systems are expected to see robust adoption in the future.

User Advice: Storage cluster file systems have been around for years, although vendor maturity varies widely. Users that need products that enable them to pay as they grow in a highly dynamic environment, or that need high bandwidth for shared storage, should put cluster file systems on their shortlists. Most commercial and open-source products specialize in tackling specific use cases and aren't mature enough to operate as general-purpose storage arrays. Evaluate your application

and input/output (I/O) requirements to select a pertinent cluster file system. Consider scalability, manageability, independent software vendor (ISV) support, and data protection and recovery features. There is little technical know-how regarding scale-out file systems in many enterprise IT organizations; hence, I&O leaders should allocate a portion of the storage budget to training.

Business Impact: Storage cluster file systems are competitive alternatives that scale storage bandwidth more linearly, surpassing expensive monolithic frame storage arrays in this capability. The business impact of storage cluster file systems is most pronounced in environments in which applications generate large amounts of unstructured data, and the primary access is through file protocols. However, they will also have an increasing impact on traditional data centers that want to overcome the limitations of dual-controller storage designs. Many storage cluster file systems will have a significant impact on private cloud services, which require a highly scalable and elastic infrastructure. IT professionals keen to consolidate file server or NAS filer sprawl should consider using cluster file system storage products that offer operational simplicity and nearly linear scalability.

Benefit Rating: High

Market Penetration: 5% to 20% of target audience

Maturity: Adolescent

Sample Vendors: Cloudera; Dell; EMC; Hortonworks; HP; Huawei; IBM; Intel; MapR Technologies; Nutanix; Panasas; Quantum; Red Hat; Symantec; Terascala

Recommended Reading: "Critical Capabilities for Scale-Out File System Storage"

"Who's Who in Open-Source, Scale-Out File System Storage"

Linear Tape File System (LTFS)

Analysis By: Alan Dayley; Dave Russell

Definition: In 2011, IBM delivered the linear tape file system (LTFS), which was donated to a standards organization in 2012. LTFS accelerates the seeking and transfer of data by making Linear Tape-Open (LTO) generation 5+ tapes, IBM TS1140 (or later) tapes and Oracle T10000C (or later) tapes self-defining, mountable and searchable by segmenting the tape into index and content partitions. The index partition holds the file metadata (based on XML schema) for the content partition, which holds the files.

Position and Adoption Speed Justification: A hierarchical directory structure provides folders on tape that can contain groups of related files to facilitate data sharing and the updating of specific files. The index data can be cached in supported tape libraries, so that the volume contents can be displayed without having to physically mount the volume. Each tape can appear as an extension of the file system to Windows, Linux and Mac OSX users (OS support depends on individual vendors). This makes tapes more portable, and archived data that is stored long term can be easily accessed without backup software — often without any specialty software at all.

LTFS further positions tape as a viable repository for "nearline" data, as well as the long-term preservation of data. Early adoption has been in vertical industries, such as media and entertainment, engineering/architecture firms and healthcare. As a mountable device with a readable directory, the structured e-discovery of tape contents is also facilitated. As a technology, LTFS has moved rapidly along the Hype Cycle curve; however, as expectations have diminished, LTFS has seen slow adoption. Tape vendors are attempting to overcome the growing customer perception that tape is going away, and they are trying to reposition tape as a viable, active archiving storage medium.

User Advice: Tape technologies, including LTO-6 and proprietary drives and tape formats from IBM and Oracle with LTFS and exploitation by a vendor ecosystem, combined with improved tape quality, should be evaluated as a storage medium for long-term archives with large quantities of data. In addition, LTFS and modern tape drives are well-suited to hybrid disk/tape environments, or even flash/disk/tape hierarchies, for cost-effectively storing large amounts of data, while maintaining transparent user and application access.

Business Impact: As a low-cost alternative to disks for nearline storage and long-term retention, LTFS and tape technology can provide significant cost savings, without sacrificing ease of access or reliability. This technology can be especially useful and cost-effective in large, petabyte-scale systems, with large file formats. As part of a nearline strategy, LTFS can be fronted with an integrated storage system with a thin layer of flash storage for expedited searches on the index and metadata. This would speed up query times, but there would still be some latencies in the retrieval times.

Benefit Rating: Moderate

Market Penetration: 1% to 5% of target audience

Maturity: Adolescent

Sample Vendors: Crossroads Systems; HP; IBM; Oracle; QStar Technologies; Quantum; Spectra Logic

Recommended Reading: "Evolving Best Practices for Backup, Archiving and Tape: Strategies for Alignment"

"Repository Strategies for Archived Data: The Blurred Line Between Hardware and Software"

"Tape's Role Is Changing From Data Protection to Active Archiving"

"IT Market Clock for Storage, 2014" "

Online Data Compression

Analysis By: Dave Russell

Definition: Online data compression encodes data using mathematical algorithms to reduce the number of bits needed to store an object, and decodes the data when it is retrieved. This analysis

deals with "lossless" compression schemes, meaning that the original data can be reconstructed in its entirety, exactly as it was, from the compressed data with no degradation. Run-length encoding, Lempel-Ziv and Huffman coding are three of the most popular algorithms in widespread use — sometimes with several of these techniques used in combination.

Position and Adoption Speed Justification: While compression algorithms have existed for over half a century, where, when and why compression is applied has changed over time. Compression has been in widespread use in tape technologies and backup software, and as server-installed software, for several decades. Since the early 2000s, compression has been used in backup appliances, such as virtual tape libraries and deduplication devices. These use cases often could tolerate the process and/or elapsed time demands that compression required.

In the last few years, compression has entered the primary storage market, and is often included in new hybrid, as well as 100%, flash arrays, to better-optimize the use of flash. Advancements in processor speed, overall cost improvements and especially the random access, nonmechanical nature of flash technology have accelerated compression usage for primary data.

User Advice: Online data compression offers favorable capacity savings with modest-to-no performance considerations (to compress data at the time of the write and to reinflate data during a read operation) for a large majority of workloads, and should be evaluated whenever available. The ability to apply online data compression to a greater number of use cases and workloads is increasing as the cost-per-CPU cycle declines and storage systems deploy more powerful processors and/or additional amounts of memory, both of which can accelerate the mathematical computations involved with compression algorithms. Compression should be considered a required capability for nonmechanical (e.g., memory- and flash-based) storage solutions, with vendors being asked "Why not?" versus "Why?" regarding the capability being offered. That said, the ability to selectively turn on or off compression for a volume, virtual machine, backup job, etc., is advantageous in case a workload is not compressible or a performance issue is encountered.

Despite the advantages that compression can offer, data reduction is not always achievable. Data that is previously encrypted or compressed may not exhibit any repeating patterns that compression algorithms can further reduce, thus minimizing or negating any benefit. In rare instances, attempting to compress previously compressed data can result in a slightly larger resultant; however, this is not common and the size difference is not great. Today, there is little harm caused by attempting to compress these workloads, as the processor or memory used for compression algorithms no longer requires the same percentage of resources compared to systems in the past.

Business Impact: Depending on performance considerations, type of data and retention periods, compression ratios can vary; however, typical results are usually in the 2-to-1 to 4-to-1 range, although Gartner generally guides clients to assume no higher than 2.5 to 1 for planning purposes. The positive impact of high data compression ratios on the need for additional storage purchases, operations, facility requirements and environmental costs will change the design of primary storage infrastructures, as well as backup/restore and archiving solutions.

Online data compression can actually improve performance. This is because with compression enabled, each input/output (I/O) operation could carry a higher effective data payload, thus, swapping hard disk I/O for processing cycles, an advantageous trade-off in cost and time.

Benefit Rating: High

Market Penetration: 20% to 50% of target audience

Maturity: Early mainstream

Sample Vendors: Atlantis Computing; Dell; EMC; Exablox; Hitachi Data Systems; IBM; Kaminario; Microsoft; NetApp; Nexenta; Nimble Storage; Nimbus Data; Nutanix; Oracle; Pure Storage; SimpliVity; SolidFire; Tegile; Tintri; Violin Memory

Recommended Reading: "Magic Quadrant for Solid-State Arrays"

"Critical Capabilities for Solid-State Arrays"

"Magic Quadrant for Enterprise Backup Software and Integrated Appliances"

"Best Practices for Repairing the Broken State of Backup"

Disaster Recovery as a Service

Analysis By: John P. Morency

Definition: Disaster recovery as a service (DRaaS) is a cloud-based recovery service in which the service provider is responsible for managing virtual machine (VM) replication, VM activation and exercise management. Over the past year, a growing number of service providers are now offering managed hosting services for those hybrid recovery configurations that are composed of both physical and virtual servers.

Position and Adoption Speed Justification: Over the past year, Gartner has seen a significant increase in the number of DRaaS offerings as well as a significant increase in the number of both production implementations (more than 21,000, a 75% increase over 2014). Initially, small organizations with less than 100 employees were DRaaS early adopters. The reason for the service uptake in smaller organizations was because these often lacked the recovery data center, experienced IT staff and specialized skill sets needed to manage a disaster recovery (DR) program on their own. This made managed recovery in the cloud an extremely attractive option. However, since the beginning of 2013, many large (1,000 to 5,000 employees) and very large (5,000-plus employees) enterprises have also begun initial piloting or have moved beyond the piloting stage to full production. Today, large and very large enterprises represent approximately 27% and 13% of the DRaaS installed base, respectively.

Recent reference to customer survey results (obtained during DRaaS Magic Quadrant research) showed both the positives and negatives of existing DRaaS offerings. Two of the top three service attributes found most attractive by the reference customers — support for both virtual and physical servers (i.e., hybrid configurations), and DRaaS contract flexibility — had not been significantly

differentiating service features previously. The providers found the market need for both virtual and physical server recovery configurations far more common than they previously thought. In addition to a broader level of provider support for hybrid configurations, the result revealed that service pricing and contract terms have improved considerably since the end of 2013. These changes also reflect the fact that, with well over 180 current DRaaS providers, there is no lack of industry competition.

Because of the growing number of production instances, rapidly falling service pricing and significant increases in service pilot evaluations being conducted in organizations of all sizes, Gartner has increased the Hype Cycle position of DRaaS to peak-trough midpoint.

User Advice: Clients should not assume that the use of cloud-based recovery services will subsume the use of traditional DR providers or self-managed DR any time in the near future. The key reasons being computing platform-specific recovery requirements, security concerns, active-active operations requirements and cost advantages of noncloud alternatives, among others. Therefore, it is important to look at DRaaS as just one possible alternative for addressing in-house recovery and continuity requirements.

Consider cloud infrastructure when you need DR capabilities for either Windows- or Linux-centric cloud-based applications, or when the alternative to a cloud-based recovery approach is the acquisition of additional servers and storage equipment for building out a dedicated recovery site. Additionally, because cloud services for enterprises are still rapidly evolving, carefully weigh the cost benefits against the service management risks as an integral part of your DR sourcing decision-making.

Business Impact: The business impact is moderate today. The actual benefits will vary, depending on the diversity of computing platforms that require recovery support and the extent to which service customers can orchestrate (and ideally fully automate) the recurring recovery testing tasks that need to be performed. An additional consideration is the extent to which the customer can transparently and efficiently use same provider cloud storage for ongoing data backup, replication and archival. The key challenge is ensuring that these services can be securely, reliably and economically used to complement or supplant the use of more traditional equipment subscription-based services or the use of dedicated facilities.

Benefit Rating: Moderate

Market Penetration: 5% to 20% of target audience

Maturity: Adolescent

Sample Vendors: Acronis; Allstream; Axcient; Bluelock; Carpathia CenturyLink; CenturyLink; Columbus Business Solutions; Databarracks; Dimension Data; EMC; Hosting; HP; iland; IBM Resiliency Services; IPR International; NTT Communications; Peak 10; Rackspace; Seagate Technology; StorageCraft; Sungard Availability Services; Verizon Terremark; VMware; Windstream; Zetta

Recommended Reading: "Magic Quadrant for Disaster Recovery as a Service"

"When to Use and Avoid Disaster Recovery as a Service"

"Pricing Differences Complicate the Adoption of Backup and Disaster Recovery as a Service"

"Critical Capabilities for Recovery as a Service"

Enterprise Endpoint Backup

Analysis By: Pushan Rinnen

Definition: Enterprise endpoint backup refers to backup products for laptops, desktops, tablets and smartphones that can recover corrupted or lost data residing on the devices, as well as personal configurations. Endpoint backup differs from file sync and share's versioning capabilities in that backup preserves secure, centrally managed copies that cannot be changed or deleted by end users and that it protects PC/laptop data in a more comprehensive way. However, mobile content protection is weaker due to lack of APIs from mobile OS providers.

Position and Adoption Speed Justification: Unlike server backup, which is a centrally managed, critical infrastructure component for all organizations to protect their data, endpoint backup has been often unmanaged, or less centrally managed via ad hoc policies. Organizations have been using different methods such as having users write to network file shares or backup to a local USB drive at home, none of which can be effectively enforced. Technology wise, early-generation PC backup products often had poor performance issues that impact user productivity and couldn't handle mobile aspects well. As employees become more mobile, laptop backup has been the driving force for organizations to adopt endpoint backup. Today, vendors have added more features to cater to the mobile nature of laptops, such as VPN-less backup over the Internet, cellular network awareness, and remote wipe. The old performance issues are tackled by the use of client-side deduplication in addition to incremental forever backups, near-continuous data protection (CDP) technologies, and CPU and network throttling, resulting in a more user-transparent and user-friendly experience.

The latest proliferation of tablets and smartphones as endpoint devices has complicated the situation for endpoint backup. First, the mobile OS platforms are typically sandboxed without APIs for third-party backup developers. Therefore all mobile device backup by third-party tools has very limited functionality. Second, the commingled state of personal and business content in the same app (such as camera) on the same mobile device makes it impossible so far to separate personal and business content. Backing up a camera app could drive up the storage cost substantially, not to mention the sensitive privacy issue when companies have access to private personal information. As a result, most organizations don't have any policy regarding mobile data backup. For more detailed analysis on mobile device backup, see technology profile "Backup Tools for Mobile Devices" in "Hype Cycle for Enterprise Mobile Security, 2015."

Overall, enterprise endpoint backup has a much lower adoption rate than server backup. However, organizations with a lot of intellectual properties or business data residing on endpoint devices have started adopting in earnest a centrally managed laptop backup to reduce legal and business risks. Those that have globally distributed offices and employees like to leverage Web-scale public cloud storage providers and backup-as-a-service providers that offer a multiple-country presence.

User Advice: Protecting endpoint user data must be part of a robust enterprise data protection and recovery plan. Although protecting mobile devices such as sandboxed tablets and smartphones continues to be a challenge, organizations should evaluate and deploy a laptop/PC backup solution, be it on-premises or in the cloud, to maintain control and prevent data loss or leakage, instead of depending on employees to create their own backup methods.

Business Impact: Endpoint backup and recovery have become increasingly important as the global workforce has become more mobile and is creating more business content on their various endpoint devices, including mobile devices. If employees don't back up their endpoint devices regularly (and many do not), companies may face significant risks when important or sensitive data is lost, stolen or leaked, including R&D setbacks, fines, legal actions and the inability to produce user data in a lawsuit. Based on Gartner's estimates, laptop/PC data loss as a result of lack of backup could cost an organization of 10,000 employees about $1.8 million a year.

Benefit Rating: High

Market Penetration: 5% to 20% of target audience

Maturity: Adolescent

Sample Vendors: Asigra; Code42; CommVault; Ctera Networks; Datacastle; Druva; EMC; HP; Infrascale; Seagate Technology

Recommended Reading: "How to Address Three Key Challenges When Considering Endpoint Backup"

"Critical Capabilities for Enterprise Endpoint Backup"

"Cloud File Sync/Share Is Not Backup"

Hybrid DIMMs

Analysis By: Michele Reitz

Definition: Hybrid dual in-line memory modules (hybrid DIMMs) are nonvolatile DIMMs that reside on the double data rate (DDR) DRAM memory channel, function as DRAM memory and focus on preserving data in case of power failure in critical applications. They integrate DRAM, NAND flash memory, a system controller chip and an ultracapacitor powerful enough to allow the module time to write all of the contents of the DRAM to the NAND flash memory when power is cycled, thereby providing persistent data storage.

Position and Adoption Speed Justification: Hybrid DIMMs are good alternatives to battery-powered backup systems or the super-capacitor-based DIMMs used to save current data in case of power failure. Hybrid DIMMs use the same industry-standard DRAM sockets and, with declines in NAND flash pricing, it is becoming economical to design systems with sufficient storage capacity to enable backup capability. Generally, there is a one-to-two ratio between DRAM and NAND storage capacity, but that can change, according to the application. Hybrid DIMMs also can be configured

for use as secondary storage, as long as they possess flash management and interface support for the host application.

The NVDIMM Special Interest Group, a consortium within the Storage Networking Industry Association (SNIA), classifies two types of NVDIMM: NVDIMM-N and NVDIMM-F. Gartner classifies NVDIMM-N as a hybrid DIMM and NVDIMM-F as a solid-state DIMM.

Industry support and standardization will be critical for adoption. Currently, only a few major OEMs are using the technology, with Supermicro providing support for x86 server architectures. In addition, hybrid DIMMs are available from only one major DRAM vendor — Micron (via AgigA Tech) — and from custom module providers like Viking Technology and a few other small module vendors. We expect other major memory vendors to enter the market, along with other custom module providers that are already involved in both DRAM and flash-based memory modules. We expect adoption of this technology to increase slowly in the next two years.

The continued lack of new introductions of hybrid DIMMs, lack of growing OEM support, together with slower-than-expected growth due to the slow migration of users to new technologies and a lack of education about the potential benefits of hybrid DIMMs, has limited penetration of this technology. We have, therefore, moved its position down closer to the Trough of Disillusionment on the Hype Cycle compared with last year.

User Advice: IT professionals should examine the roadmaps of major server and storage OEMs, as well as those of solid-state drive (SSD) appliance vendors, to see which will launch hybrid DIMM-based systems. They should ascertain whether hybrid DIMMs are supported by the OS and server they wish to use, as the latencies of hybrid DIMMs and all DRAM DIMMs require that servers, systems and OS timing routines are tuned properly.

IT professionals should educate themselves about the option to use hybrid DIMMs to meet their nonvolatile DIMM needs. Although the focus of hybrid DIMMs is DRAM backup, they also can be used as a new tier of storage with access times closer to DRAM, as they are significantly faster than conventional SSDs and have a denser form factor that allows for greater system capacities. For this use case, users should consider both hybrid and solid-state DIMMs.

Business Impact: Hybrid DIMMs have several advantages over conventional battery-powered backup DIMMs, including faster speed, lower maintenance costs, greater reliability, high availability and improved system performance. Currently, the cost premium over existing solutions is considerable, but it should drop as NAND flash pricing stabilizes and competition intensifies throughout the remainder of 2015 and into 2016.

Memory vendors should consider hybrid DIMMs as a way of adding value to what are essentially two commodity products — DRAM and NAND flash. By exploiting the attributes of these devices, they will not only enhance their own value proposition, but also expand their addressable market.

Benefit Rating: Moderate

Market Penetration: 1% to 5% of target audience

Maturity: Emerging

Sample Vendors: AgigA Tech; HGST; Micron; Viking Technology

Cross-Platform Structured Data Archiving

Analysis By: Garth Landers

Definition: Cross-platform structured data archiving software moves data from custom or commercially provided applications to an alternate file system or DBMS while maintaining data access and referential integrity. Reducing the volume of data in production instances can improve performance; shrink batch windows; and reduce storage acquisition costs, facilities requirements, the cost of preserving data for compliance when retiring applications and environmental footprints.

Position and Adoption Speed Justification: Structured data archiving tools have been available for more than a decade and have historically seen more adoption in larger enterprises. These products provide functionality to identify old or infrequently used application data and manage it appropriately. Although ROI can be very high, developing policies for retaining and deleting old application data is difficult and not seen as a priority. Organizations generally tend to add more database licenses or use native database capabilities, such as purging and partitioning, to address application growth. The technology has long been seen as a cost avoidance measure used to contain operational and capital expenditures related to data growth as well as improve factors like application performance. The market is changing and growing due to growth in data, application retirement, information governance and big data analysis opportunities.

Today's data archiving products are mature and will face challenges as various distributions of Hadoop add capabilities such as retention management. At the same time, new approaches to application retirement and curbing structured data growth are happening through areas such as virtualization and copy data management. These approaches, while immature when applied to these use cases, offer a less complex approach than the technology offered by leading offerings today. Although the market is seeing growth, awareness remains low. Application retirement continues to be a significant driver. Organizations are looking for ways to cut costs associated with maintaining no-longer-needed legacy applications while preserving application data for compliance or its historical value. Data center consolidations, including moving to the cloud, and mergers and acquisitions are contributing to the interest in structured data archiving solutions to reduce the number of enterprise applications.

Competition often comes from internal resources who want to build it themselves and from improvements in storage technology that transparently improve performance while reducing storage acquisition and ownership costs — more specifically, autotiering, SSDs, data compression and data deduplication. Do it yourself efforts typically lack appropriate governance controls such as secure access, data masking and retention management, including legal hold. The allure of tools that can support multiple applications and underlying databases and the added capabilities these tools provide for viewing data as business objects independent of the application are driving administrators to consider them as viable solutions. New capabilities — such as better search and reporting, integration with big data analysis tools, retention management, support for database partitioning, and support for SAP archiving — are broadening their appeal.

User Advice: The ROI for implementing a structured data archiving solution can be exceptionally high, especially to retire an application or to deploy a packaged application for which vendor-supplied templates are available to ease implementation and maintenance. Expect that the planning phase may take longer than the implementation. Among the roadblocks to implementation are requiring the consulting services, gaining application owner acceptance (especially through testing access to archived data), defining the archiving policies and building the initial business case. All vendors in this space can provide good references, and organizations should speak with references that have similar application portfolios and goals for managing their data. Enterprises should consider developing their own solutions when either the number of applications being retired is very low, data retention requirements are not very long (such as one to two years) or governance requirements such as audit or litigation are unlikely.

Business Impact: Creating an archive of less frequently accessed data and reducing the size of the active application database (and all related copies of that database) improves application performance and recoverability and lowers costs related to database and application license, server and infrastructure costs. Transferring old, rarely accessed data from a disk archive to tape can further reduce storage requirements. Most vendors in this space are supporting cloud storage as the repository for archived data. Retiring or consolidating legacy applications cuts the costs and risks associated with maintaining these systems. Overall, organizations can experience better information governance, including reduced risk associated with litigation.

Benefit Rating: Moderate

Market Penetration: 5% to 20% of target audience

Maturity: Mature mainstream

Sample Vendors: Delphix; HP; IBM; Informatica; OpenText; Solix Technologies; Teradata

Recommended Reading: "Magic Quadrant for Structured Data Archiving and Application Retirement"

"Application Retirement Drives Structured Data Archiving"

"Build a Leaner Data Center Through Application Retirement"

"Best Practices for Storage Management: Developing an Information Life Cycle Management Strategy"

Emerging Data Storage Protection Schemes

Analysis By: Stanley Zaffos

Definition: Emerging data storage protection schemes deliver higher mean time between data loss (MTBDL) than traditional redundant array of independent disks (RAID) and global sparing schemes. Commercial implementations of emerging data storage protection schemes include replacing the concept of spare disks with spare capacity, which enables lost data to be rebuilt at memory rather

than disk speeds, intelligent data rebuilds that only reconstruct data stored on failed disks or nodes, triple mirroring, erasure codes, and dispersal algorithms.

Position and Adoption Speed Justification: Each of the technologies enumerated above increases MTBDLs by reducing rebuild times or increasing the fault tolerance or resiliency of the data protection scheme. Erasure coding and dispersal algorithms, which add the physical separation of storage nodes to erasure coding, store blocks of data as systems of equations that can be transformed back into blocks of data on any storage media. This enables users to trade off data protection overheads (costs) against MTBDLs by allowing the user to specify the number of failures that can be tolerated during a data rebuild within an hard-disk drive (HDD) or solid-state drive (SSD) group. Because erasure coding and dispersal algorithms increase the number of overhead inputs/outputs (I/Os) needed to protect data, they are most commonly used in scale-out network-attached storage (NAS), storage area network (SAN), object storage systems, and/or archive and backup storage, rather than block storage systems supporting response-time-sensitive applications.

User Advice: Require vendors offering advanced data protection schemes to profile the performance/throughput of their storage systems supporting your workloads using the various protection schemes supported to better understand performance-overhead trade-offs. Request RAID rebuild time information from vendors for full disk rebuilds with 50% and 80% utilized/busy controller CPU and various RAID groups (3D+1P, 7D+1P) to size the likely rebuild window of vulnerability in a storage system supporting your production workloads. When rating bids, give extra weight to vendors that provide this information. Confirm that the choice of protection scheme does not limit the use of other value-added features, such as compression and deduplication, autotiering, or deduplication.

Business Impact: The deployment of advanced protection schemes will enable vendors and users to continue lowering storage costs by taking advantage of disk technology improvements that increase disk capacity faster than they increase disk manufacturing costs, and with the added benefit of being transparent to applications, policies and procedures. The business impact will grow as users grow their disk storage farms, because more disks equal more disk failures, which, when coupled with long rebuild times, increases the risk of data loss caused by multiple drive failures occurring during the rebuild window of vulnerability.

Benefit Rating: Moderate

Market Penetration: 5% to 20% of target audience

Maturity: Adolescent

Sample Vendors: Amplidata; Caringo; Cleversafe; DataDirect Networks (DDN); Dell; Dot Hill Systems; EMC; IBM; NEC; Panasas; Scality; SwiftStack

Recommended Reading: "Increases in Disk Capacity Are Affecting RAID Recovery; Start to Purchase New RAID Technologies"

"Technology Overview for Erasure Coding"

"Slow Storage Replication Requires the Redesign of Disaster Recovery Infrastructures"

Cloud Storage Gateway

Analysis By: Gene Ruth; Pushan Rinnen; Raj Bala

Definition: Cloud storage gateways connect on-premises storage to external cloud storage. They are supplied as hardware, software or as part of a storage array. Gateways connect to the cloud through Internet Protocol (IP) and integrate with on-premises IT infrastructures via standard file-and-block protocols. Gateways often have different architectures; however, they typically provide performance optimization, encryption and data footprint reduction.

Position and Adoption Speed Justification: Gateways are important elements and key enablers for connecting on-premises storage to off-site cloud storage services. This market is predominantly composed of startup vendors that closely partner with the cloud storage providers to offer comprehensive cloud storage solutions. In some cases, the cloud providers, such as Microsoft Azure and Amazon Web Services (AWS), offer their own gateways. Large vendors continue to increase participation. NetApp has purchased Riverbed's SteelStore product, and EMC has reintroduced the acquired TwinStrata as cloud array software for VMAX 3. However, the lack of cloud gateway functions on most mainstream storage arrays and the lack of integration with backup and archiving software present challenges for the wider adoption of cloud storage gateways.

Gateway implementations target small or midsize businesses (SMBs), branch-office support and non-mission-critical storage for large enterprises. User adoption is increasing, with most occurring in the SMB space; however, interest from enterprise customers for branch-office support and unstructured data applications is growing. The early adoption of public cloud storage gateway appliances focused on backup, archiving of stale data and disaster recovery. Gartner is seeing continued interest in technologies that support unstructured data delivery and enhance collaboration among distributed sites, such as support of file sync and share, or global file systems.

User Advice: Until cloud storage gateways are offered in a certified and turnkey manner from public cloud storage service providers or major storage hardware vendors with service partners, users should focus on proof of concept (POC) and branch-office support use cases, or they should apply gateways to non-mission-critical data. As the industry develops, cloud storage and gateways should be included in an overall long-term storage infrastructure strategy.

Businesses of all sizes should investigate total cost of ownership (TCO) versus traditional, on-premises storage; the impacts of cloud storage on IT operations; and the risks of moving data into the cloud. Organizations should negotiate with gateway appliance vendors on guarantees that ensure availability, data integrity, performance and financial objectives are met, and should expect partnership relationships between gateway vendors and public cloud storage service providers.

Business Impact: Gateways serve as a stopgap for bridging on-premises storage and cloud storage. They enable cloud storage to compete against on-premises primary storage arrays for workloads that are modestly transactional or involve unstructured file data. In addition, the gateway may replace existing secondary storage as a target for backup and/or archive data that is redirected into cloud storage. Gateways compete indirectly against asynchronous, remote replication tools

used for collaboration or disaster recovery, as well as some backup software, due to the data protection offered natively with cloud storage infrastructures.

These gateway appliances can provide customers that want to reduce in-house backup/disaster recovery processes, archives and unstructured data with compelling, cloud-based alternatives. Gateways have the potential to enable bursting in cloud computing environments that intend to move virtual machines (VMs) between private and public computing clouds. Gateways only partially address issues such as the unpredictability of monthly cloud storage costs, large workloads, storage SLAs and quality of service (QoS). Because many cloud storage gateway vendors are emerging startups with products that have not been widely proved, customers should be diligent when checking references.

Benefit Rating: Moderate

Market Penetration: 5% to 20% of target audience

Maturity: Adolescent

Sample Vendors: Amazon Web Services; Avere Systems; Ctera Networks; EMC; Hitachi Data Systems; Microsoft; Nasuni; NetApp; Panzura

Recommended Reading: "Cloud Storage Gateways: Enabling the Hybrid Cloud"

"Hybrid Cloud Storage Can Be an Antidote to Rapid Data Growth"

"How to Calculate the Total Cost of Cloud Storage"

Virtual Machine Backup and Recovery

Analysis By: Pushan Rinnen; Dave Russell

Definition: Virtual machine (VM) backup and recovery focus on protecting and recovering data from VMs, as opposed to the physical server they run on. Most backup software solutions have abandoned the traditional guest OS agent approach and adopted image-based backup that offers VM-level or application object recovery, leveraging hypervisor-native APIs. Other advanced features also emerged to leverage VMs' mobility, such as instant VM recovery from the backup copy or automated recovery testing.

Position and Adoption Speed Justification: Enterprise VM backup typically focuses on VMware and Hyper-V, as they are the most deployed hypervisors in the enterprise data centers. VMware-specific ISV backup tools have dwindled as some were acquired, leaving Veeam as the only major stand-alone vendor. Meanwhile, VMware has included the formerly chargeable vSphere Data Protection Advanced as a free standard function within vSphere 6.0, putting more pressure on competition, especially for small to midsize environments.

While almost all backup software leverages VMware VADP API for agentless, snapshot-based VM image backup, many traditional backup software requires installation of guest OS agents to do

granular item restore for applications such as Exchange and SharePoint. Other differentiators center on performance, scalability and usability, which are typically tied to the overall architecture of the backup solution.

VM-specific backup/recovery tools showed increased adoption during the past few years because they were early-to-market, cost-effective and simpler to manage than traditional comprehensive backup tools. Virtualization-focused solutions can also empower vSphere administrators to perform backup/recovery. However, the lack of physical server backup support from those tools forces users to maintain another backup application. Over time, the VM-specific backup/recovery solutions could continue to do well in highly virtualized environments with little need for tape, or could give way to more general solutions that offer wider platform support and greater scalability. The next frontier for VM backup and recovery will expand to the backup of containers like Docker, and the backup of VMs running in public cloud.

User Advice: Recoverability of the virtual infrastructure is a significant component of an organization's overall data availability, backup/recovery and disaster recovery plan. Protection of VMs needs to be taken into account during the planning stage of a server virtualization deployment, as virtualization presents new challenges and new options for data protection.

Evaluate application data protection and restoration requirements before choosing VM-level backup. Additionally, snapshot, replication and data reduction techniques, and deeper integration with the hypervisor provider should also be viewed as important capabilities. With hundreds to thousands of VMs deployed in the enterprise, and with 10 or more mission-critical VMs on a physical server, improved data capture, bandwidth utilization, and monitoring and reporting capabilities will be required to provide improved protection without complex scripting and administrative overhead.

Business Impact: As more production data is housed in or generated by VM environments, the need to protect data in these environments is increasing. VM backup and recovery solutions help recover from the impact of disruptive events, including user or administrator errors, application errors, external or malicious attacks, equipment malfunction, and the aftermath of disaster events. The ability to protect and recover VMs in an automated, repeatable and timely manner is important for many organizations.

Benefit Rating: High

Market Penetration: More than 50% of target audience

Maturity: Early mainstream

Sample Vendors: Acronis; Arcserve; Asigra; Axcient; Barracuda Networks; Catalogic Software; CommVault; Dell; EMC; FalconStor Software; HP; IBM; Infrascale; Microsoft; Quantum; Seagate Technology; StorageCraft Technology; Symantec; UltraBac Software; Unitrends; Veeam Software; Vision Solutions; VMware; Zerto

Recommended Reading: "Essential Practices for Optimizing VMware Backup"

"Magic Quadrant for Enterprise Backup Software and Integrated Appliances"

"Best Practices for Repairing the Broken State of Backup"

Public Cloud Storage

Analysis By: Gene Ruth; Pushan Rinnen; Raj Bala

Definition: Public cloud storage is infrastructure as a service (IaaS) that provides object, block or file storage services through a REST API using Internet Protocol (IP). The service is stand-alone, with no requirement for additional managed services. Service pricing is based on capacity, data transfer and/or services. It provides on-demand storage capacity elasticity and self-provisioning. Stored data exists in a multitenant environment, and users access that data through the Internet or dedicated network connectivity.

Position and Adoption Speed Justification: Cloud storage is available on a global basis, with providers offering a wide breadth of storage services and SLAs. Services target customers and workloads and are defined by their SLAs, global scale and pricing regimes. IaaS storage adoption is driven by the need for an easy way to provision storage when dealing with unpredictable workloads, to reduce storage costs and infrastructure refresh cycles of cold data, to offload on-premises infrastructure and to support collaborative applications. However, it is still inhibited by privacy, regulatory, economic, vendor and service provider credibility issues. Enterprise customers focus on hybrid solutions that bridge on-premises storage with public services, as they test the capability and prove the viability of public cloud storage for their use cases. The hybrid architecture is highly dependent on traditional applications' cloud connectors, common APIs and the cloud storage gateway market.

Regulatory and sovereignty concerns have contributed to differences in expectations among users in the U.S. and other geographic areas. In the U.S., a few larger service providers are solidifying their positions in the market and are continuing to establish their global presence and credibility. Other regional markets have been slower to adopt cloud storage and have fewer visible providers. Gartner expects adoption expansion to continue, as costs, legal concerns, security and infrastructure integration issues are sufficiently addressed to reduce the risk of entry by large enterprises. The limited number of enabling products that support hybrid infrastructures and the reluctance of IT organizations to use evolving storage service providers limits the growth of the market.

User Advice: Deploy cloud storage as an alternative to non-mission-critical storage services, such as archiving, file sharing and backup. These noncritical use cases allow for shortfalls in SLA compliance by service providers and facilitate expectation setting by clients. Cloud storage can be used to prestage data for transient cloud computing IaaS environments that depend on manipulating large amounts of data. Consider public cloud storage as an effective solution for providing data and data protection for branch offices and mobile users. For hybrid environments, include on-premises cloud storage gateway appliances that provide cache, data deduplication, thin provisioning, encryption and capabilities that address security and latency concerns.

When selecting a service provider, due diligence should include evaluating the client organization's sensitivities to SLA compliance, data sovereignty, bandwidth costs, usage rates and disaster

recovery requirements. Initially, considerable investments in time and money will be required to integrate cloud storage options into applications and environments.

Business Impact: The cost and agility expectations set by public cloud storage vendors are enabling in-house IT operations to change their storage infrastructure management procedures and storage infrastructure strategies. User demands for lower costs, more agility and operations that are more autonomic are influencing vendor R&D investments and cloud service offerings. Services that have already been influenced by user demands include backup, versioning, encryption and secure erasure.

To attract new customers, vendors continue to lower costs and offer pricing models that enable end users to align their storage costs with their usage rates and reduce costs in the short and long term. However, the cost of cloud storage may be unpredictable, and customers should be aware that, although operational costs may move into the cloud, management issues such as chargeback, asset management, billing, security and performance responsibilities remain with the customer.

Benefit Rating: High

Market Penetration: 5% to 20% of target audience

Maturity: Adolescent

Sample Vendors: Amazon; AT&T; Google; IBM; Microsoft; Rackspace; Verizon Enterprise Solutions

Recommended Reading: "Cloud Storage Gateways: Enabling the Hybrid Cloud"

"Hybrid Cloud Storage Can Be an Antidote to Rapid Data Growth"

"How to Calculate the Total Cost of Cloud Storage"

Fibre Channel Over Ethernet

Analysis By: Joe Skorupa; Roger W. Cox

Definition: **Definition**

Fibre Channel over Ethernet (FCoE) is an encapsulation of Fibre Channel (FC) frames over enhanced Ethernet networks. The goal of enhanced Ethernet is to provide a unified network that can transport multiple types of traffic, including FC storage, Internet Protocol (IP) and high-performance computing over Ethernet with a line speed of 10 Gbps or greater, eliminating storage network islands, such as FC, Ethernet and InfiniBand, as well as their associated discrete adapters and switches.

Position and Adoption Speed Justification: FCoE must run on an enhanced version of Ethernet called data center bridging (DCB). While DCB standards are complete, buyer and vendor support (particularly for management tools) is far from enthusiastic. It is possible to build an end-to-end multihop DCB environment, although this is rare outside the midmarket.

FCoE has suffered due to delayed standards, weak vendor support and strong alternatives. Operational, cultural and political concerns prevented core data center network consolidation. Additionally, other more cost-effective solutions — such as network-attached storage (NAS), Internet Small Computer System Interface (iSCSI) and hyperconverged systems with virtual storage area networks (SANs) — relegated FCoE to being one of multiple solutions for in-chassis and in-rack networking. FCoE has found acceptance only in the small or midsize business (SMB) space and in the broader market in converged infrastructure (where the providers embed the technology); adoption does not extend beyond the network edge.

FCoE adoption is unlikely to ever reach the Plateau of Productivity or maintain its current market share, although it will continue to see some limited adoption.

User Advice: When evaluating FCoE, carefully consider whether the risks are worth the limited benefits.

Users who want to deploy FCoE should look to their server or network vendors for rack-based and blade-server solutions. To maximize savings, seek servers with converged network adapters (CNAs) on the motherboard. Users should continue to deploy FC outside the rack, and, over time, should consider 10/40/100 Gigabit Ethernet (GbE) as a possible replacement, but only when 10/40/100 GbE with NAS or iSCSI is a viable, less expensive alternative to FC. In this case, plan to keep the core SAN as a separate, parallel network. Users should also consider management implications in IT operations:

- Who owns the FCoE network — storage department or network department?
- Who has priority for changes and upgrades in the FCoE network?
- Who owns change management?

Alternatively, use iSCSI or NAS as the mature solution in which consolidation onto 10 GbE is desired. Although the protocol overheads may be higher on the network, current-generation network interface cards (NICs) will offload TCP/IP and the iSCSI protocol from application servers.

Business Impact: Although CNAs are expensive, the primary benefit of FCoE deployment is cost savings from consolidation of host bus adapters (HBAs)/NICs and switches. Instead of separate and redundant sets for storage and networking, FCoE requires only one set. Because iSCSI can provide this benefit at a lower cost, the incremental benefit of FCoE is very low. The positive side for FCoE is that many of the driver and management stacks from FC can be preserved, and management products will likely support FC and FCoE at the same time. FCoE adoption implies an integration of the storage and data networks, which will impact organizational structure, job descriptions, SLAs and processes.

Benefit Rating: Low

Market Penetration: 5% to 20% of target audience

Maturity: Adolescent

Sample Vendors: Brocade; Cisco; Dell; Emulex; EMC; Hitachi Data Systems; HP; IBM; NetApp; QLogic

Recommended Reading: "Why, When and How to Converge Storage and Data Networks"

"Emerging Technology Analysis: Fibre Channel Over Ethernet, Networking and Communications"

"Myth: A Single FCoE Data Center Network = Fewer Ports, Less Complexity and Lower Costs"

"Use Top-of-Rack Switching for I/O Virtualization and Convergence; the 80/20 Benefits Rule Applies"

Automatic Storage Tiering

Analysis By: Stanley Zaffos

Definition: Automatic storage tiering transparently and dynamically moves pages of a logical volume or file between tiers of cache or storage. The movement of pages between tiers of cache or storage is done under user-defined policies and/or algorithms that manage page placement with the objectives of minimizing storage costs and environmental footprint, while meeting or exceeding performance and throughput service-level agreements. All page movements are transparent to applications except for their potential impact on performance and throughput.

Position and Adoption Speed Justification: Autotiering has made solid-state drive (SSD) usage in general-purpose storage systems commonplace. Users taking advantage of autotiering and SSDs are often able to decrease their storage acquisition and ownership costs by increasing their use of low-cost, high-capacity 7.2K-rpm hard-disk drives (HDDs) without sacrificing performance and throughput. The widespread usage of low-cost, high capacity 7.2K-rpm HDDs shrinks the number of HDDs in a storage system, which translates into lowering the frequency of repair activities by reducing the number of HDDs in the system. This also reduces its environmental footprint relative to a system configured to an equivalent capacity of 10K-rpm or 15K-rpm high-performance HDDs.

User Advice: Implement internal performance and capacity planning processes that use current and potential storage system suppliers' benchmarking expertise to optimally size each tier of cache or storage managed by the autotiering software. This increases the probability of proposed configurations meeting or exceeding performance SLAs, and gives storage suppliers at least partial ownership of any performance-related problems. Continually monitor performance and throughput, and resize cache and storage and adjust policies as indicated to accommodate changes in workload characteristics. Where performance needs can be met without the use of SSDs, request that vendors bid systems with and without autotiering software and SSDs to enable a quantitative cost-benefit analysis. When deploying autotiering software in a storage system that is being replicated to a disaster recovery site, take into account the impact of autotiering on replication software charges and failover performance at the disaster recovery site and the potential savings that archiving can deliver.

Business Impact: The value of autotiering will depend on the ratios of inactive, average, and high-performance or low-latency data in your storage system, and the number of applications with high input/output and/or low response time requirements. The value of autotiering increases with storage

system size, and, depending on workload characteristics, may further lower storage total cost of ownership (TCO) by improving staff productivity. Taken together, these benefits can help justify storage consolidation projects and the economies of scale that consolidation can deliver.

Benefit Rating: Moderate

Market Penetration: 20% to 50% of target audience

Maturity: Early mainstream

Sample Vendors: Dell; Dot Hill Systems; EMC; Hitachi Data Systems; HP; IBM; NetApp; Nimble Storage; Oracle; Tegile; Tintri; X-IO Technologies

Recommended Reading: "Overcome Disk Autotiering Problems With These Deployment Recommendations"

"How Much and What Type of Disk Storage Do IT Departments Need?"

"Where to Use SSDs in Your Storage Infrastructure"

"Solid-State Drives Will Complement, Not Replace, Hard-Disk Drives in Data Centers"

"Use SSDs, Rather Than Disk Striping, to Improve Storage Performance and Cut Costs"

SaaS Archiving of Messaging Data

Analysis By: Alan Dayley

Definition: SaaS archiving of messaging data involves email, instant messaging (IM) and social media. Compliance and regulatory requirements drive the retention of messaging data, with SaaS archiving increasingly becoming the repository of choice. The capture of messaging content occurs at the time of creation or as it enters the organization's communications systems, where it can be stored on immutable, write once, read many (WORM) storage.

Position and Adoption Speed Justification: SaaS archiving solutions are mature. Many users find the administration tools for SaaS archiving solutions more user-friendly than those available from on-premises solutions. As the journaling feature is turned on in the email administration console, capture is as simple as pointing the journaled email to the hosted provider's site. IM archiving is as mature as email, and it is often stored in an email format in the archive repository. Social media archiving is newer, and its capture is usually through APIs provided by the social media applications. Although social media data can be stored in an email format in the archive, the industry trend is to store it in native format.

Unlike backup or disaster recovery as a service, archive users are less concerned about latency and more interested in the accurate capture of metadata and the chain of custody of data; therefore, the speed of the Internet connections is not a major concern. This, coupled with prevalence of easy-to-use administrative and supervision tools, has led many organizations to choose a hosted solution,

enabling archive expenses to shift to an operating expenditure (opex) model and away from capital expenditures (capex).

As government and industry regulations proliferate, SaaS archiving vendors have been nimble at updating the compliance requirements of offered solutions. Most SaaS archiving vendors offer end users access to messaging data through a search interface or, in some cases, a native application folder view. Basic e-discovery capabilities of hosted solutions have received high marks from customers and are noted as another reason for adoption.

User Advice: Organizations in highly regulated industries will find SaaS message-archiving solutions to be mature, secure and reliable enough to meet the most stringent requirements. Organizations with message-archiving needs will find the hosted option easy to administer and attractively priced, and they will find that it offers an opportunity to optimize internal IT resources. Most organizations do not face internal or external requirements or regulations that require the data to reside on-premises, so the willingness to consider the cloud revolves primarily around company culture regarding risk, security, data sovereignty and costs.

When considering a solution, focus on indexing, search and discovery capabilities to ensure that your needs are met by the offering or through integration with a third-party e-discovery vendor. The migration of legacy email archives, including into and out of a hosted solution, can be expensive and should be scoped during the selection phase. Organizations should include in SaaS archiving contracts an exit strategy that minimizes costs, and they should remember that they, not the SaaS providers, own the data. When determining the costs versus benefits for SaaS archiving, include soft expenses associated with on-premises solutions for personnel and IT-involved discovery requests.

Business Impact: Organizations switch capex for opex costs when selecting a hosted archive solution. Pricing is typically based on a per-mailbox or a per-user basis, paid as a monthly subscription. IT departments are relieved of the responsibility for updating legacy, on-premises archive systems when hardware and software need to be refreshed. Compliance and legal personnel within organizations directly access the hosted solution without IT involvement and can more easily provide access to the hosted archive message data to outside parties, as required.

Benefit Rating: Moderate

Market Penetration: 5% to 20% of target audience

Maturity: Early mainstream

Sample Vendors: ArcMail; Bloomberg; Global Relay; Google; HP (Autonomy); Microsoft; Mimecast; Proofpoint; Smarsh; Sonian; Symantec

Recommended Reading: "Magic Quadrant for Enterprise Information Archiving"

"Critical Capabilities for Enterprise Information Archiving"

"How to Determine Whether Your Organization Needs Website Archiving"

"Five Factors to Consider When Choosing Between Cloud and On-Premises Email Archiving Solutions"

Storage Multitenancy

Analysis By: Stanley Zaffos

Definition: Storage multitenancy describes the security and quality of service (QoS) features that enable the secure sharing of a storage system between users running diverse workloads and still meet each user's service-level objectives. Multitenancy features generally include one or more of the following: LUN zoning and/or masking, logical partitioning, large primary or secondary cache configurations, autotiering, I/O prioritization and/or throttling, balanced host and back-end bandwidth, file virtualization, and clustered or distributed file support.

Position and Adoption Speed Justification: The widespread adoption of server virtualization, virtual desktop infrastructure (VDI), storage autotiering and the move to 24/7 operations are driving storage consolidation and the deployment of private cloud infrastructures. These trends, in turn, are shaping storage infrastructure refreshes that emphasize the adoption of high-end scale-up and scale-out storage systems with good multitenancy features.

User Advice: Multitenancy support should not be treated as a primary evaluation criterion, but a use case that affects the weighting of other more basic measures of storage system attractiveness, such as scalability and availability, performance/throughput, security features (such as partitioning and encryption), and ecosystem support. Where practical, users deploying multitenancy storage should test to scale, rather than rely on a simple proof of concept. Where impractical, users should obtain performance/throughput guarantees with meaningful remedies that are not punitive, yet enforceable, and span diverse workloads.

Business Impact: Good multitenancy features increase the appeal and probability of success of storage consolidation projects that can lower total cost of ownership (TCO) by reducing the total number of storage systems being managed. These projects also simplify disaster recovery testing by increasing the probability of applications being contained within a single storage system; increase the value of green storage technologies, such as thin provisioning and autotiering; and make the deployment of ever-larger storage systems practical.

Benefit Rating: Moderate

Market Penetration: 20% to 50% of target audience

Maturity: Early mainstream

Sample Vendors: Caringo; DataDirect Networks (DDN); EMC; Fujitsu; Hitachi Data Systems; HP; IBM; NetApp; Nimble Storage; Oracle; Scality; SolidFire; Tegile; X-IO Technologies

Recommended Reading: "Magic Quadrant for General-Purpose Disk Arrays"

Enterprise File Synchronization and Sharing (EFSS)

Analysis By: Monica Basso

Definition: Enterprise file synchronization and sharing (EFSS) refers to a range of on-premises or cloud-based capabilities that enable individuals to synchronize and share documents, photos, videos and files across multiple devices, such as smartphones, tablets and PCs. Sharing can happen within the organization or outside of it, such as with partners and customers. Security and collaboration capabilities are critical aspects for enterprises to adopt EFSS.

Position and Adoption Speed Justification: EFSS is maturing. Core capabilities include native apps for mobile devices, password protection and data encryption, and server integration (e.g., with SharePoint). Enhanced capabilities include native content creation, collaboration, digital rights management, cloud encryption key management (EKM) and modern user interface (UI). Basic EFSS capabilities are commodities, having increasingly been added to other products and services as feature extensions by a variety of IT vendors in multiple markets such as enterprise content management (ECM), collaboration, storage, backup and enterprise mobility management (EMM). Cloud storage providers such as Google, Microsoft and Dropbox accelerate the trend, with low-priced deals. Pure-play EFSS vendors ("destinations") are evolving onto two orthogonal paths: enterprise systems/resources integration and management, with support for federation, e-discovery, data governance and security; or modern collaboration and business enablement. Despite concerns that security and compliance may slow down adoption, EFSS investments continue growing, as organizations must balance IT control over bring-your-own (BYO) cloud services with user demand for modern productivity tools. Dropbox's presence drives interest, as does Microsoft Office 365 adoption.

User Advice: IT leaders responsible for digital workplace initiatives must introduce EFSS. To work more effectively, organizations must explore potential security risks of personal cloud services, as well as requirements for modern productivity tools from users. They should evaluate EFSS options and capabilities to enable secure mobile content sharing, collaboration and productivity, reducing potential risks. If email or legacy FTP services are used for file transfers, organizations should consider EFSS as an efficient way for employees to share data, rather than using personal services. Organizations looking for secure alternatives to personal cloud, while preserving user preferences, especially focusing on external collaboration, should consider public cloud offerings, such as Box or even the enterprise Dropbox offering. Organizations with tighter requirements on data control, as well as those with large investment in storage infrastructure, should focus on hybrid solutions (such as Citrix, Syncplicity, Egnyte) that grant more control over where data is housed and can make use of existing storage investments. Organizations with strong requirements for data protection, under strict regulations about data storage or with complex data manipulation requirements should focus on on-premises deployments (for example, WatchDox). Organizations looking for EFSS and requiring support for optimized file transfer among different systems should consider dedicated offerings, such as Accellion or Ctera.

Business Impact: Enterprise file sharing will enable higher productivity and collaboration for mobile workers who deal with multiple devices. Organizations investing in such capabilities will enable a more modern and collaborative real-time workplace, while reducing or avoiding the inherent

security/compliance threats of personal cloud services. Business benefits include increased productivity and cost savings.

Benefit Rating: High

Market Penetration: 5% to 20% of target audience

Maturity: Early mainstream

Sample Vendors: Accellion; Acronis; AirWatch; Boole Server; Box; Citrix; Ctera Networks; Dropbox; Egnyte; EMC; Google; Intralinks; Microsoft; ownCloud; Soonr; TeamDrive; Thru; Varonis; WatchDox

Recommended Reading: "How to Build EFSS Plans to Address Current and Future Business Requirements"

"Magic Quadrant for Enterprise File Synchronization and Sharing"

"Toolkit: Enterprise File Synchronization and Sharing RFI/RFP"

"Destinations and Wraparounds Will Reshape the Enterprise File Synchronization and Sharing Market"

"Enterprise File Sync and Share Will Shift Storage Management Burden From the Edge to Data Centers or Cloud"

"Secure Low-Cost Data Sharing and Collaboration With iPad"

Climbing the Slope

Solid-State Arrays

Analysis By: John Monroe; Joseph Unsworth

Definition: Solid-state arrays (SSAs) are a subcategory of the broader external controller-based (ECB) storage market. SSAs are scalable, dedicated solutions based solely on semiconductor technology for data storage that cannot be configured with hard-disk drives (HDDs) at any time. As distinct from solid-state drive (SSD)-only racks configured within ECB storage arrays, SSAs must be a stand-alone product denoted with a specific name and model number that typically includes an OS and data management software that are optimized for solid-state technology.

Position and Adoption Speed Justification: As enterprise SSD solutions have evolved based on greater clarity regarding customer application demands and more innovative hardware designs, the addition of a fully integrated system-level software stack — as part of the SSA solution — will be required to extend the advantages of NAND and guard against its weaknesses. SSAs reflect and embody this evolution, as new and established vendors marry hardware with software to deliver compelling, high-speed solutions, with complete suites of storage management capabilities. The aim is to provide new breeds of SSAs that not only can serve as replacements for high-

performance, mission-critical HDD storage, but also can be used for cost-effective capacity, with the operational benefits of flash storage.

The costs to conceive, achieve and maintain true differentiation in this dynamic and expanding market will be large, and startup vendors must obtain adequate funding. The fact that all the established server and storage companies either offer their own solutions, or engage in strategic partnerships with startup solution providers has accelerated market adoption, and will continue to foster the necessary software optimization finely tuned to flash memory, as opposed to legacy HDD technology.

User Advice: Vendors need to help IT professionals gain a clearer understanding of their data center needs by developing and deploying diagnostic software that can monitor evolving enterprise workloads. This will be an essential first step to determine the most cost-effective solutions, and provide the most suitable performance and high-availability benefits.

Users should demand that SSA vendors include the following as "must do" parts of their competitive strategy:

- Qualify and reliably integrate consumer-grade, multilevel cell flash to provide higher-capacity solutions that can be sold for less than $5 per raw (native, uncompressed) gigabyte.
- Provide total cost of ownership (TCO) analyses of performance efficiencies, energy savings and space conservation relating to specific application workloads.
- Embody flexibly selectable thin provisioning, in-line deduplication, in-line data compression, cross-application quality of service (to prevent inconsistent performance from the "noisy neighbor" effect), at-rest and on-the-fly encryption, high availability, and disaster recovery (by means of local/metro synchronous and remote asynchronous replication) in most, if not all, SSA configurations.
- Prove your company capable of providing competent and enduring global support and services, with aggressive seven-year warranty programs.

This list is ideal, but not (yet) pervasively supported.

Many SSA vendors like to quote "effective/usable" capacity with data compression and other storage management technologies. Customers should not take at face value vendor claims regarding effective/usable capacity as opposed to "raw/native" capacity. Not all data and workloads are equally compressible, and usable (or effective) capacity may change from application to application; customers' SSA capacity "mileage" certainly will vary. And some applications may simply require data acceleration, with minimal, or no, storage management.

IT professionals must weigh their actual needs against the cost and features of the available solutions when considering adoption. IT professionals should also remain cautious in deployment, and select only financially stable and proven system suppliers that have strong direct or indirect partnerships that can enable them to deliver competent and enduring support and services.

Business Impact: Compared with the legacy HDD-based and hybrid HDD-/SSD-based arrays, SSAs will continue to be relatively expensive on a dollar-per-gigabyte basis, but will deliver

competitive value on a dollar-per-transaction and TCO basis. The greatest opportunities are for deployment in online transaction processing, analytics and highly virtualized environments that include (but are not limited to) cloud data centers; database, data warehousing and high-performance computing environments; and hosted virtual desktop infrastructure environments.

Despite their relative expense, SSAs already have proved to be a dynamic subset of the ECB market arenas; SSA revenue grew by 130% in 2014, expanding from $621 million to $1.43 billion. Gartner predicts that the SSA share will grow to almost 30% (more than $7 billion) of the total ECB market, with a 37.7% compound annual growth rate (CAGR) from 2014 through 2019. Legacy ECB storage arrays, while still large at more than $17 billion, will display a negative 4.1% CAGR from 2014 through 2019.

The legacy ECB arrays and SSAs will reflect a symbiotic evolution as the markets move toward the all-flash data centers of the future. Efficient data management software that seamlessly integrates the benefits of variously configured storage tiers, and can be closely coupled with the host OS, increasingly will become an imperative measure for long-term success.

Benefit Rating: Transformational

Market Penetration: 5% to 20% of target audience

Maturity: Early mainstream

Sample Vendors: EMC; Fujitsu Eternus; HP; IBM; Kaminario; NetApp; Pure Storage; SolidFire; Tegile Systems; Violin Memory

Recommended Reading: "Market Share Analysis: SSDs and Solid-State Arrays, Worldwide, 2014"

"Competitive Landscape: Solid-State Arrays"

"Solid-State Drives Will Complement, Not Replace, Hard-Disk Drives in Data Centers"

"Where to Use SSDs in Your Storage Infrastructure"

"Prioritize Software Features When Buying SSD Arrays"

Appliance-Based Replication

Analysis By: Stanley Zaffos

Definition: Replication appliances, virtual or real, provide network-based, storage-vendor-neutral block-level and/or network-attached storage (NAS) replication services. These services can include local snapshots and/or clones and remote (synchronous or asynchronous) replication services. Appliance- or network-based replication solutions can span servers and storage systems. Integration into the storage infrastructure can be via software agents, storage area network (SAN) or direct storage system support.

Position and Adoption Speed Justification: Offloading replication services from storage systems into a network-based replication appliance provides operational and financial advantages, compared with software- and controller-based solutions. Operational benefits include preserving native storage system performance and the use of common replication services across multiple heterogeneous storage systems, which can simplify disaster recovery by creating a constant timeline or consistency group across multiple storage systems. Other potential operational advantages include the ability to include direct-attached storage (DAS) support via the use of software agents, and insulating disaster recovery integrity from imperfect software control procedures. Using replication appliances reduces the strength of storage vendor lock-ins, which can translate into lower storage ownership costs by keeping storage system acquisitions competitive. Despite these financial and operational advantages, market acceptance has been hampered by:

- A reluctance by end users to add another device to the input/output path
- Competition from storage virtualization appliances
- Storage array vendor lock-ins
- The channel constraints created by storage vendors and indirect channels protecting the vendor lock-ins that controller-based replication technologies create
- The increasing number of storage systems that use all-inclusive software-pricing models

User Advice: Users should consider the use of replication appliances when there:

- Is a need to create a constant timeline across multiple homogeneous or heterogeneous storage systems
- Is a problem with the usability or performance of the existing replication solution
- Are savings to be had from using a replication-appliance-based solution, rather than an existing or virtualization-appliance-based solution
- Is a need to preserve investments in existing storage systems
- Is a desire to pursue a dual-vendor strategy

Caution: Users should ensure that replication appliance microcode overhead and network latency do not create performance/throughput bottlenecks when protecting solid-state arrays (SSAs) and/or solid-state drives (SSDs).

Business Impact: Appliance-based replication services can:

Provide the benefits of storage-based replication solutions without the lock-ins that storage-system-based replication solutions create.

Delay storage system upgrades by offloading replication overhead from the storage system that lacks the compute power and bandwidth needed to limit the impact of replication services on native system performance.

Work with DAS, SANs and NAS.

Provide heterogeneous replication targets to allow lower-cost solutions.

Benefit Rating: Moderate

Market Penetration: 5% to 20% of target audience

Maturity: Early mainstream

Sample Vendors: DataCore Software; EMC; FalconStor; Hitachi Data Systems; Huawei; IBM; NetApp; Zerto

Recommended Reading: "Slow Storage Replication Requires the Redesign of Disaster Recovery Infrastructures"

Enterprise Information Archiving

Analysis By: Alan Dayley

Definition: Enterprise information archiving (EIA) solutions provide tools for capturing all or selected data in a distributed or centralized repository for efficient storage and access. EIA supports multiple data types (including email, file system, social media, Web, mobile and SharePoint). These tools provide access to archived data through a plug-in to the native application via a stub or pointer or via browser access, and some manage the data in place. EIA tools support operational efficiency, compliance, retention management and e-discovery.

Position and Adoption Speed Justification: The number of vendors offering EIA solutions has stabilized and, in some cases, there has been consolidation in the market. Driven by awareness created through Microsoft Office 365 adoption, archiving is becoming mainstream for meeting compliance and e-discovery needs for organizations implementing information governance programs, so market growth is continuing. SaaS for messaging data archiving, including email and social media, has also gained significant traction.

Support for the capture and supervision of social media has become a requirement in regulated industries. File archiving has slowed, with a focus on selective archiving for records. EIA products that support multiple content types are the norm. Many companies are looking to replace their archiving products with newer ones (particularly SaaS solutions), and many migration services are available. In addition, there is growing interest in managing the compliance and retention of data "in place," rather than moving it to a different repository.

The appetite for email-only archiving solutions remains; however, most organizations are looking to vendors with existing solutions or a roadmap for EIA products.

User Advice: As requirements to store, search and discover old data grow, companies should implement an EIA solution, starting with email as the first managed content type. Mailbox size for on-premises email implementations continues to grow, creating storage and compliance concerns.

Many organizations are looking to migrate to cloud email and productivity solutions, such as those offered by Microsoft and Google, and when migrating, associated compliance and regulatory retention requirements need to be considered. Organizations should ensure contractually that they have a reasonably priced process, as well as an option for removing data from an archive solution — namely from SaaS providers. Migrating personal stores, such as PSTs, to the archive should be part of the deployment of an email archive system.

Business Impact: EIA improves application performance, delivers improved service to users, and enables a timely response to legal discovery and business requests for historical information. Archived data can be stored in a less-expensive fashion, with the opportunity to take some data offline or delete it. Moving old data to an archive also reduces backup and recovery times by decreasing the active dataset.

Email remains the predominant content type archived as part of an EIA implementation. Archiving offered via SaaS is increasing in popularity, because of the benefits associated with offloading low-business-value tasks to a third party, as well as the reduced capital expense. SaaS-based message data archiving is leading the way and is currently priced on a per-user, per-month (PUPM) basis, with no storage overages. As cost structure and integration issues are ironed out, more file system data and application data will be archived in the cloud.

EIA is an important part of e-discovery, providing support for the Electronic Discovery Reference Model (EDRM). Legal hold, retention management, search and export features are used to meet discovery and compliance requirements. Supervision tools for sampling and reviewing messages are available with many EIA products, in response to requirements specific to the regulated portion of the financial industry. To meet the requirements of mobile workers, EIA offers a way for organizations to keep data compliant in an archive, while providing access via mobile devices.

Benefit Rating: High

Market Penetration: 20% to 50% of target audience

Maturity: Early mainstream

Sample Vendors: ArcMail; Barracuda Networks; Bloomberg; CommVault Systems; EMC; Global Relay; Google; Gwava; HP-Autonomy; IBM; MessageSolution; Metalogix Software; Microsoft; Mimecast; OpenText; Proofpoint; SilverSky; Smarsh; Sonian; Symantec; ZL Technologies

Recommended Reading: "Magic Quadrant for Enterprise Information Archiving"

"Critical Capabilities for Enterprise Information Archiving"

"How to Determine Whether Your Organization Needs Website Archiving"

"Five Factors to Consider When Choosing Between Cloud and On-Premises Email Archiving Solutions"

Data Deduplication

Analysis By: Dave Russell

Definition: Data deduplication is a unique form of compression that eliminates redundant data on a subfile level, improving storage utilization. Only one copy of the data is stored; all other redundant data will be eliminated, leaving only a pointer to the extraneous copies of the data. Compared to traditional compression, deduplication has a broader scope for comparing redundant data, such as across multiple users, VMs, backup jobs, etc., and examines data that has been written over a longer period of time and sometimes with greater levels of granularity.

Position and Adoption Speed Justification: This technology reduces the amount of physical storage required, significantly improving the economics of disk-, flash- or memory-based solutions for backup, archiving and primary storage. While deduplication has historically been used in backup activities (due to the repetitive nature of capturing largely unchanged data), it can be applied to long-term archiving and primary storage.

Deduplication has taken on a vital role in solid-state array (SSA) and hybrid flash array storage appliances in an effort to contain the cost of the flash solution while maximizing capacity. As such, nearly all flash storage and hybrid flash array devices possess some form of deduplication.

User Advice: Solutions vary in terms of where and when deduplication takes place, which can significantly affect performance and ease of installation. When used with backup, deduplication that occurs on a protected machine is referred to as "client side" or "source" deduplication, whereas deduplication that takes place *after* the protected machine — after the data is sent to the backup application — is considered "target side" deduplication. A distinction is made between solutions that deduplicate the data as it is processed (in-line deduplication) and products that have the data write directly to disk, as it would without deduplication, and then deduplicate it later, which is postprocessing or deferred deduplication. Deduplication solutions also vary in granularity, but 4KB to 128KB chunks (or segments) of data are typical. Some deduplication algorithms are content-aware, meaning that they apply special logic for further processing, depending on the type of application and data being stored, and/or can factor out metadata from an application, such as a backup program.

Gartner clients using deduplication for backup typically report seven to 25 times the reduction (a 7-to-1 to 25-to-1 ratio) in the size of data. Archiving deduplication ratios are often in the 3-to-1 to 10-to-1 range, and primary file data commonly yields 3-to-1 to 6-to-1 ratios. Restore performance can be negatively affected by deduplication, depending on the solution implemented.

Given the costs associated with flash storage, deduplication is an essential capability for improving the economics and wear endurance of flash, and should be considered a "must have" feature.

Business Impact: Deduplication improves the cost structure of storage solutions, as less storage needs to be purchased, deployed, powered and cooled. As a result, businesses may be able to use disk, flash or DRAM memory for more of their storage requirements and may retain data for longer periods of time, thus enabling faster recovery or read access, versus retrieval from slower media. The additional benefits of deduplication include its positive impact on disaster recovery (DR),

because less network connectivity is required, since each input/output (I/O) operation carries a larger data payload.

For a backup application, deduplication, along with replication, may mean that physical tapes do not need to be made at the primary location and manually taken off-site. For primary storage, deduplication can have a positive impact on improving overall performance, reducing the amount of cache consumed and reducing the overall input/output operations per second (IOPS).

Benefit Rating: Transformational

Market Penetration: 20% to 50% of target audience

Maturity: Early mainstream

Sample Vendors: Acronis; Actifio; Arcserve; Asigra; Atlantis Computing; Barracuda Networks; CommVault; Datacastle; Dell; EMC; Exablox; ExaGrid Systems; FalconStor; Hitachi Data Systems; HP; IBM; Infinio; Kaminario; NetApp; Nexenta; NEC; Nimbus Data; Nutanix; Oracle; Permabit; Pure Storage; Quantum; Seagate; SimpliVity; SolidFire; Symantec; Tegile; Tintri; Unitrends; Veeam; Violin Memory

Recommended Reading: "Magic Quadrant for Solid-State Arrays"

"'Critical Capabilities for Solid-State Arrays"

"Magic Quadrant for Enterprise Backup Software and Integrated Appliances"

"Best Practices for Repairing the Broken State of Backup"

Continuous Data Protection

Analysis By: Dave Russell

Definition: Continuous data protection (CDP) is an approach to recovery that continuously, or nearly continuously, captures and transmits changes to applications, files or blocks of data while journaling these changes. This capability provides the option to recover to many more-granular points in time to minimize data loss. Some CDP solutions can be configured to capture data either continuously (true CDP) or at scheduled times (near CDP).

Position and Adoption Speed Justification: The difference between near CDP and regular backup is that backup is typically performed once a day, whereas near CDP is often done every few minutes or hours, providing many more recovery options and minimizing any potential data loss. Several products also provide the ability to heterogeneously replicate and migrate data between two different types of disk devices, allowing for potential cost savings for disaster recovery solutions. Checkpoints of consistent states are used to enable rapid recovery to known good states (such as before a patch was applied to an OS, or the last time a database was reorganized) to ensure application consistency of the data and to minimize the number of log transactions that must be applied.

CDP and near-CDP capabilities can be packaged as server-based software, as network-based appliances that sit between servers and storage, and as part of a storage controller. To date, storage controllers offer near CDP only by way of the frequent use of snapshots, and do not allow for the capture, journaling and transmission of every write activity. The delineation between frequent snapshots (one to four per hour or less granularity) and near CDP is not crisp, and administrators often implement snapshots and CDP solutions in a near-CDP manner to strike a balance between resource utilization and improved recovery.

User Advice: Consider CDP for critical data where regular snapshots and/or backups do not enable meeting the required recovery point objectives (RPOs). Gartner has observed that true CDP implementations are most commonly deployed for files, email and laptop data. True CDP for databases and other applications is not common and has a much lower market penetration. Near CDP for applications might be more appropriate to ensure application consistency and to minimize the amount of disk and potential processor cycles required for the solution.

Many large vendors have acquired their current offerings, and only a few startup vendors remain, with most startups being acquired or having failed. Many backup applications, such as those from Arcserve, Asigra, CommVault, Dell and IBM, include CDP technology as an option to their backup portfolio. The market has since mostly adopted near-CDP solutions via more frequent, array-based snapshots, or as part of the backup application. The disk requirements and potential production application performance impact were among the main reasons for true CDP initially facing challenges. Later, as near CDP became more readily available, it satisfied most of the market's needs. Some backup vendors, such as Symantec, which was early to market in 2005 with its CDP in Backup Exec, have discontinued CDP features. EMC, with its RecoverPoint product, has been successful in selling a hybrid replication and true-CDP solution. Today, Catalogic Software and FalconStor Software are among the remaining "pure play" CDP providers in the market.

Business Impact: CDP can dramatically change the way data is protected, decreasing backup and recovery times, as well as reducing the amount of lost data, and can provide additional recovery points. Compared to traditional backup, which typically captures data once a day, the amount of data lost in a restore situation can be nearly 24 hours for backup versus minutes or a small number of hours with CDP. A recent, interesting development has been Oracle's introduction of the Zero Data Loss Recovery Appliance (ZDLRA) which captures data at the transaction-level and send data from the application server's memory buffer. ZDLRA provides real-time data protection for Oracle 11g databases and later.

Benefit Rating: High

Market Penetration: 20% to 50% of target audience

Maturity: Early mainstream

Sample Vendors: Actifio; Arcserve; Asigra; Catalogic Software; CommVault; Datacastle; DataCore Software; Dell; EMC; FalconStor Software; IBM; Microsoft; Oracle; Seagate Technology; Unitrends; Veeam Software; Vision Solutions

Recommended Reading: "Magic Quadrant for Enterprise Disk-Based Backup/Recovery"

"Best Practices for Repairing the Broken State of Backup"

Data Encryption Technologies, HDDs and SSDs

Analysis By: John Monroe

Definition: Software or hardware encryption can prevent unauthorized access to confidential data. When a storage device is configured with full-disk encryption (FDE) firmware, and the embedded encryption is enabled (turned on, "unlocked"), it is referred to as a "self-encrypting drive" (SED).

The showcase technology for standardized SEDs, commonly referred to as OPAL, was released in open source in 2009 by the Trusted Computing Group (TCG). All major manufacturers of mass-storage devices now support OPAL-based SEDs.

Position and Adoption Speed Justification: Most actively accessed digital data usually resides on hard-disk drives (HDDs), but increasingly, will reside on NAND-flash-based solid-state drives (SSDs) as well. When an HDD or SSD (or a tape) is enabled as an SED, unauthorized data retrieval by individuals who may have legitimate access to the physical storage devices holding confidential data, but not to the content itself, can be prevented. Encryption also prevents unauthorized access to data stored on laptops and other mobile devices that are lost or stolen, and it can be in the form of hardware or software encryption.

Regulatory initiatives requiring that data be protected, as well as the financial fallout from very public losses of unencrypted data by high-profile organizations, are accelerating the need for some kind of global data encryption in diverse storage markets. Financial losses are less about the loss and recovery of the physical device, and more about the data contained therein, and the chance of data loss or misuse by a company's own employees (current and former) has put a spotlight on the risk from within. Telecommuting and increased use of consumer technologies in business environments also put a company at greater risk for data loss and misappropriation.

While access control is in general use today, the real promise of storage data security lies in encryption on specific storage devices, primarily, on HDDs and SSDs, and secondarily, on tape drives and tapes, and optical drives and discs. Current overall adoption for primary storage, archive and backup is estimated at almost 35% (for corporate data). The vast majority of current encryption techniques are vendor-proprietary (not industry-standard), and reside on tapes.

Despite some incursion from SSDs, HDDs are the primary data storage elements in all computing systems and will remain so for many years to come. Only a small portion of the HDDs (and SSDs) shipped today for configuration in servers, external controller-based (ECB) storage systems, or PCs, contain any preloaded encryption technologies.

Although this is beginning to change with TCG standardization, users still face multiple issues in trying to implement encryption throughout their enterprises.

The ways that various components and users in enterprise environments interact in secure and unsecure ways will continue to be enormously complex. Some form of FDE will, within five years, reside natively on all HDDs and SSDs in all mobile, desktop and enterprise systems. However, it

may be implemented in diverse ways by different vendors in concert with other data security measures or other storage management software, or it may not be turned on at all, remaining unrecognized by the operating system and applications, but it will be there, available for use.

Standardized FDE will be a universal starting point, leading to many variations of more-comprehensive data security solutions that will benefit end users with a greater choice of features, functionality and location of deployment.

User Advice: Many Gartner clients complain that SEDs are difficult to obtain. If SEDs are desired, companies should negotiate server, ECB and PC supplier inventory guarantees in their purchase contracts.

The type and location of your data security rollout will depend on whether the data is at rest or in flight, as well as against whom you are securing the data. End users will want to weigh the risk of data loss, as well as the financial consequences of a data loss. Do not wait for an event-driven reaction to secure your data. Proactively securing your data will help ensure against a worst-case scenario and a financial impact that is likely to far surpass that of the data security purchase itself. Focus on key management, because it represents the true differentiator between vendors. Demand that your PC, server and ECB vendors integrate only FDE-embedded HDDs and SSDs to maximize the security of your data.

Business Impact: Data encryption in various forms will have a vast impact on minimizing the cost and consequences of data theft. Furthermore, storage devices with embedded FDE will play a key role in providing maximum data security in PCs, servers and ECBs.

This Hype Cycle applies principally to the manufacturing, supply side of storage devices that embed FDE. A second, related Hype Cycle discussion concerns more issues facing end users and companies that are seeking interoperable solutions based on SEDs.

Benefit Rating: High

Market Penetration: 20% to 50% of target audience

Maturity: Early mainstream

Sample Vendors: Dell; EMC; Fujitsu; Hitachi; HP; IBM; LSI; Marvell Semiconductor; NetApp; Oracle; Samsung; Seagate; Toshiba; Western Digital

Recommended Reading: "Protecting Sensitive Data on Decommissioned SSDs and HDDs"

"Secure Data Disposal for Smartphones and Tablets Is a Mandatory Ingredient for a Successful Mobile Device Program"

External Storage Virtualization

Analysis By: Roger W. Cox

Definition: Storage virtualization is a technology that sits between host servers and the external controller-based (ECB) storage system infrastructure. It provides a virtual view of the physical storage devices, and except for EMC's VPLEX, aggregates the devices into a common resource pool for presentation to the compute environment. Storage virtualization can be provided by hardware appliances or by software within an ECB storage system. Vendors sometimes use "storage hypervisor" to describe their storage virtualization offering.

Position and Adoption Speed Justification: Besides virtualizing multiple physical and, often, disparate ECB storage systems into a common storage pool, most storage virtualization solutions provide other services, such as common provisioning, including thin provisioning, as well as local and remote replication data services. Some even support storage efficiency features, such as tiered storage and data reduction.

Storage virtualization can be implemented as symmetrical or asymmetrical solutions. In a symmetrical (or in-band) approach, the layers of abstraction and processing used by the virtualization solution are inserted directly into the data path. In an asymmetrical (or out-of-band) approach, the abstraction and processing control lie outside the data path. Storage virtualization appliances, software-only offerings intended to create appliances and ECB storage array-based software solutions, employ the symmetrical implementation. Asymmetrical implementations, though, use storage area network (SAN) switches in conjunction with appliances.

EMC's VPLEX and IBM's SAN Volume Controller (SVC) represent symmetrical in-band storage virtualization appliances. The EMC Symmetrix Federated Tiered Storage option, available on its EMC VMAX 40K; Hitachi Universal Volume Manager software option, available on its Hitachi Unified Storage VM and Virtual Storage Platform G1000; IBM Storwize V7000/V5000; and NetApp FlexArray virtualization option, available on its FAS8000 series are examples of symmetrical in-band storage virtualization solutions that are incorporated within an ECB storage system. DataCore Software's SANsymphony-V10, and FalconStor Software Network Storage Server and FreeStor are available as a stand-alone, software-only symmetrical solutions. However, unlike other storage virtualization appliances, EMC's VPLEX enables separate virtual volumes hosted in different storage systems to be simultaneously mounted with read/write privileges, but relies on the virtualized ECB disk storage platforms for the above-referenced data service features.

Even though functionality and scalability have improved, and although the number of vendors supporting symmetrical in-band storage virtualization solutions has expanded since this technology was introduced in 1999, market traction remains muted relative to the overall size of the ECB storage system market. This minimal market traction can be attributed, in part, to the following:

- Some segments of the ECB storage system market, such as installations supporting mainframes and network-attached storage, are not broadly supported by storage virtualization solutions.
- Some silo applications, such as data warehousing and Microsoft Exchange, do not fit within the storage virtualization solution model.
- There are challenges associated with establishing or validating a compelling value proposition, compared with a conventional storage infrastructure composed of multiple ECB storage systems.

- Only four major storage vendors support an organic, symmetrical in-band storage virtualization solution — EMC, which launched Symmetrix Federated Tiered Storage for the VMAX 40K in May 2012 and its VPLEX in May 2010; Hitachi (Hitachi Data Systems), which launched its Universal Volume Manager in August 2004; IBM, which launched the SVC in June 2003 and the Storwize V7000 in October 2010; and NetApp, which released the V-Series in March 2005 (now withdrawn from NetApp's product card) and its FlexArray virtualization software option in February 2014.

Although it is a close race, current market momentum, from a vendor revenue perspective, favors the symmetrical in-band storage virtualization appliances from EMC (VPLEX) and IBM (SVC). However, symmetrical in-band storage virtualization solutions that are incorporated within an ECB storage system from Hitachi (Hitachi Data Systems) (Universal Volume Manager), IBM (Storwize V7000/V5000) and NetApp (FlexArray virtualization option) are gaining increasing market share as cost-effective migration tools.

Storage virtualization was referred to in the past as "appliance-based virtualization," "external or heterogeneous storage virtualization" or "heterogeneous external storage virtualization."

User Advice: Current offerings tend to attack different problems or segments of the market, so users should carefully examine what they want to accomplish before comparing products. Consider these devices for mass-volume migration between ECB storage systems, for management improvement, where the software tools are better than existing arrays, and for consolidation/transition to a single storage vendor when the user owns ECB storage systems from many vendors. Be aware that storage virtualization creates vendor lock-in, and most (but not all) disable the value-added data services and caching features of the ECB storage systems being virtualized. Therefore, cost justifications must be scrutinized in detail to guard against paying twice for the same features.

Business Impact: Storage virtualization eases migration from old to new ECB storage systems, and it enables temporary consolidation of older ECB storage systems prior to moving to a single-vendor solution. In addition, migrating existing ECB storage infrastructures to cloud paradigms is sparking interest in storage virtualization offerings as a bridge to a cloud environment by repurposing existing investments in a storage infrastructure. It can improve provisioning and other storage management to the extent that it provides better software tools, and it can sometimes reduce the cost of the back-end storage. However, storage administrators must still manage the ECB storage systems to perform basic low-level configuration tasks.

Benefit Rating: Moderate

Market Penetration: 5% to 20% of target audience

Maturity: Early mainstream

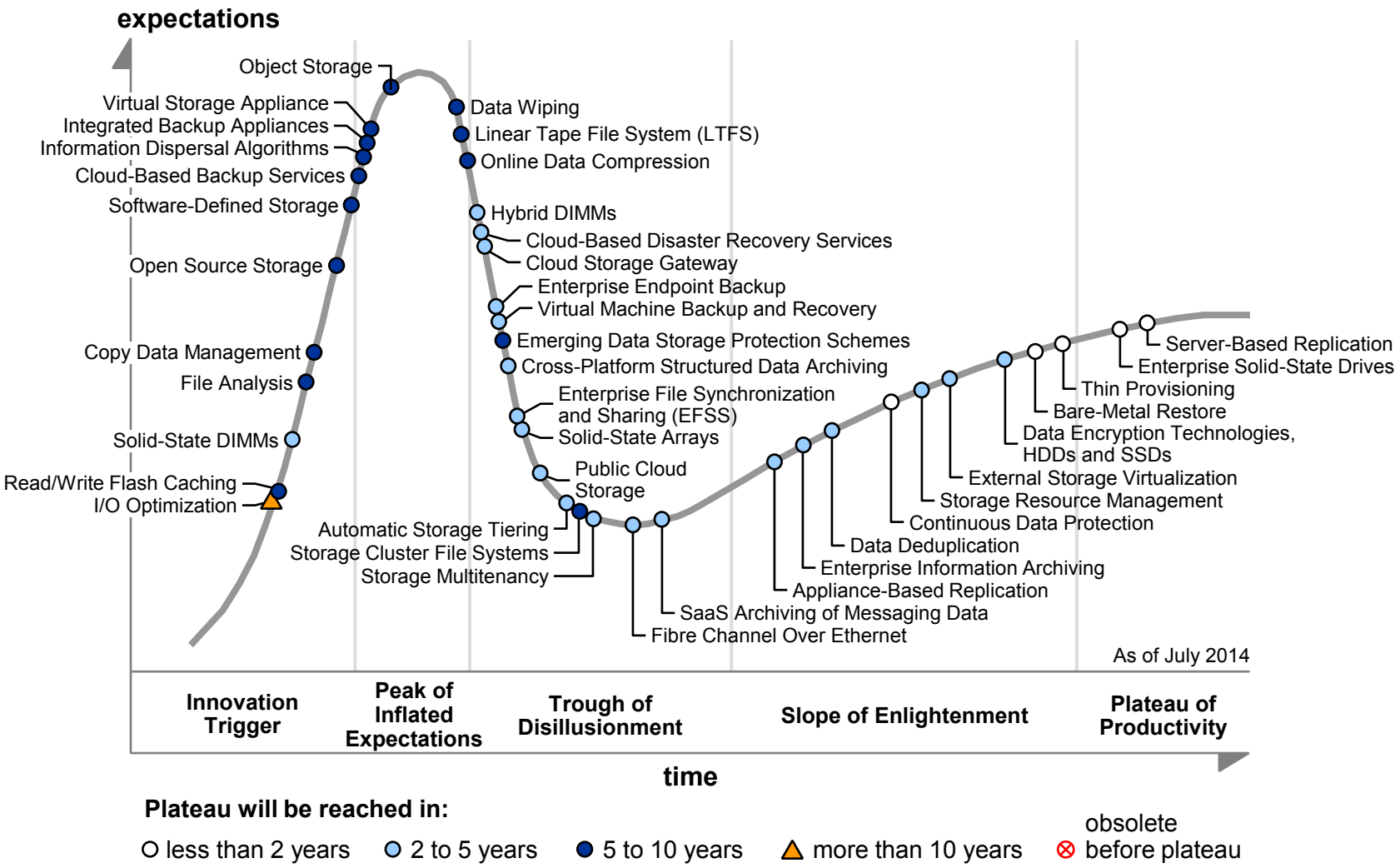
Sample Vendors: DataCore Software; EMC; FalconStor Software; Hitachi; IBM; NetApp

Recommended Reading: "Storage Virtualization"

"Decision Point for Server Virtualization Storage Selection"

Appendixes

Figure 3. Hype Cycle for Storage Technologies, 2014



Source: Gartner (July 2014)

Hype Cycle Phases, Benefit Ratings and Maturity Levels

Phase	Definition
Innovation Trigger	A breakthrough, public demonstration, product launch or other event generates significant press and industry interest.
Peak of Inflated Expectations	During this phase of overenthusiasm and unrealistic projections, a flurry of well-publicized activity by technology leaders results in some successes, but more failures, as the technology is pushed to its limits. The only enterprises making money are conference organizers and magazine publishers.
Trough of Disillusionment	Because the technology does not live up to its overinflated expectations, it rapidly becomes unfashionable. Media interest wanes, except for a few cautionary tales.
Slope of Enlightenment	Focused experimentation and solid hard work by an increasingly diverse range of organizations lead to a true understanding of the technology's applicability, risks and benefits. Commercial off-the-shelf methodologies and tools ease the development process.
Plateau of Productivity	The real-world benefits of the technology are demonstrated and accepted. Tools and methodologies are increasingly stable as they enter their second and third generations. Growing numbers of organizations feel comfortable with the reduced level of risk; the rapid growth phase of adoption begins. Approximately 20% of the technology's target audience has adopted or is adopting the technology as it enters this phase.
Years to Mainstream Adoption	The time required for the technology to reach the Plateau of Productivity.

Source: Gartner (July 2015)

Table 2. Benefit Ratings

Benefit Rating	Definition
Transformational	Enables new ways of doing business across industries that will result in major shifts in industry dynamics
High	Enables new ways of performing horizontal or vertical processes that will result in significantly increased revenue or cost savings for an enterprise
Moderate	Provides incremental improvements to established processes that will result in increased revenue or cost savings for an enterprise
Low	Slightly improves processes (for example, improved user experience) that will be difficult to translate into increased revenue or cost savings

Source: Gartner (July 2015)

Table 3. Maturity Levels

Maturity Level	Status	Products/Vendors
Embryonic	▪ In labs	▪ None
Emerging	▪ Commercialization by vendors ▪ Pilots and deployments by industry leaders	▪ First generation ▪ High price ▪ Much customization
Adolescent	▪ Maturing technology capabilities and process understanding ▪ Uptake beyond early adopters	▪ Second generation ▪ Less customization
Early mainstream	▪ Proven technology ▪ Vendors, technology and adoption rapidly evolving	▪ Third generation ▪ More out of box ▪ Methodologies
Mature mainstream	▪ Robust technology ▪ Not much evolution in vendors or technology	▪ Several dominant vendors
Legacy	▪ Not appropriate for new developments ▪ Cost of migration constrains replacement	▪ Maintenance revenue focus
Obsolete	▪ Rarely used	▪ Used/resale market only

Source: Gartner (July 2015)

Gartner Recommended Reading

Some documents may not be available as part of your current Gartner subscription.

"2015 Strategic Roadmap for Storage"

"Understanding Gartner's Hype Cycles"

GARTNER HEADQUARTERS

Corporate Headquarters
56 Top Gallant Road
Stamford, CT 06902-7700
USA
+1 203 964 0096

Regional Headquarters
AUSTRALIA
BRAZIL
JAPAN
UNITED KINGDOM

For a complete list of worldwide locations,
visit http://www.gartner.com/technology/about.jsp