451 Research PATHFINDER REPORT

S&P Global Market Intelligence

Now a Part of

Balancing Hybrid Storage Technology as Part of a Hybrid Cloud Data Management Strategy

COMMISSIONED BY



SEPTEMBER 2020

©COPYRIGHT 2020 451 RESEARCH. ALL RIGHTS RESERVED.

About this paper

A Pathfinder paper navigates decision-makers through the issues surrounding a specific technology or business case, explores the business value of adoption, and recommends the range of considerations and concrete next steps in the decision-making process.

ABOUT THE AUTHOR



STEVEN HILL

SENIOR ANALYST, APPLIED INFRASTRUCTURE AND STORAGE TECHNOLOGIES

Steven Hill is a Senior Analyst of Storage technologies. He covers the latest generation of hyperconverged systems, hybrid storage, business continuity, disaster recovery, and unstructured data management and archiving technologies for enterprise customers.



Executive Summary

The flexibility, automation and dynamic scalability introduced by the cloud delivery model more than a decade ago has changed IT for the better; as a result, the adoption of hybrid infrastructure, both on-premises and off, is rapidly becoming the norm for business IT in every vertical market. A well-crafted approach to hybrid cloud offers IT professionals the freedom of choice to host applications and data on an ever-growing number of public and private platforms. However, that freedom also introduces a substantial amount of complexity when it comes to making intelligent decisions about how to best balance the use of on- and off-premises services and resources to ensure sufficient availability, scalability, performance, governance, security, protection and cost control.

Data storage plays a unique and important role in an IT ecosystem in that everything in computing literally begins and ends in some form of storage, and hybrid technology offers several new storage options for supporting application environments, shared datasets and the work products of employees and systems. Like the rest of business IT infrastructure, enterprise storage systems are undergoing an evolution. Flash-based solid-state drive (SSD) technology is grabbing all the headlines in the storage industry and is raising perfectly reasonable questions about the future of traditional disk-based storage. The goal of this paper is to examine the changing nature of business data, the current challenges of enterprise storage, and the evolving roles of SSD and hard disk drive (HDD) technologies in light of the steady trends of massive data growth, limited budgets, public cloud adoption and the advent of a new generation of data sources, such as those introduced by Internet of Things (IoT) initiatives.

Initial Observations

- **Business data is growing at an unprecedented pace.** The problem of unchecked data growth has been a constant refrain for decades, and we expect it to increase by about 25% in 2020.
- **Storage budgets aren't keeping up.** Our polling has shown that average storage spending will only increase by approximately 10% in the same time frame.
- Data is being kept longer for legal, compliance and ongoing analytics purposes. While it varies by vertical market, it's increasingly common to see companies archiving data for at least five to seven years, with several use cases extending well into multiple decades.
- There is continued interest in keeping data nearline for longer periods. Depending on the use case, there can be a number of benefits to keeping less-active secondary data nearline to on-prem systems for performance, compliance, security, and business continuity and disaster recovery (BC/DR) purposes.
- There remain fundamental differences between on- and off-premises hybrid cloud storage. Modern enterprise storage technology offers the same management flexibility and features of its public cloud counterparts, but maintaining critical data on-premises also addresses many of the concerns of data security, governance, sovereignty and compliance.
- **Data growth, cost and compliance continue to top the list of storage pain points.** While storage performance continues to be a key consideration for establishing data placement of high-performance primary applications, managing growth, cost and governance continues to top the list of enterprise storage challenges.



Finding a New Balance for Hybrid Cloud Adoption

The Emergence of the Hybrid Cloud Opportunity

Since the introduction of the first public cloud services over a decade ago, the entire IT industry has been in the process of understanding, evolving and assimilating the key principles of the cloud delivery model for technology services and resources. With the rise of the public cloud, the abstraction of services and the mobility of applications in the on-premises datacenter that started in the early days of server virtualization are now extended to a vast and growing number of public, on-demand compute, networking and storage resources. While these public resources are easy to consume and don't require any form of capital investment, it can be a mistake to assume that all public cloud services are identical to on-premises solutions without considering all the factors involved in the placement of workloads and data.

One of the most confusing early responses to the cloud was the idea that it had to be a binary choice, such as insisting on a 'cloud first' approach without determining whether the cloud was more efficient, effective or even capable of all an application's needs. The hybrid cloud delivery model is proving to be the logical progression for the consumption of dynamic datacenter services. There are still compelling arguments to be made for both on- and off-premises infrastructure, but the real power of a cloud model – both on- and off-premises – lies in the abstraction it offers to automate the delivery of high-level IT services for compute, network and storage on a more flexible and easier-to-consume basis. Figure 1 is based on 451 Voice of the Enterprise 2020 polling data and provides a picture of the challenges enterprises face when dealing with legacy applications.

Figure 1: Top reasons for modernizing on-premises critical workloads and applications

Source: 451 Research's Voice of the Enterprise: Digital Pulse, Workloads & Key Projects 2020

Q: You've indicated that your organization plans to modernize existing critical workloads/applications on-premises. What are the most important reasons for this choice? (n=189)





We contend that a hybrid model offers more options for modernizing enterprise applications and supports the flexibility to choose between on- and off-premises hybrid cloud resources based on the right combination of technical options, legal considerations, security requirements, application performance and data protection expectations. Perhaps most importantly, a hybrid cloud approach provides an environment where business IT development is no longer limited by the capabilities of its physical infrastructure, offering on-demand scalability to absorb production spikes, as well as offering access to advanced, cloud-based applications that may not even be available for on-premises systems. Perhaps the best way to look at the public cloud is as simply another useful tool for business IT that should only be used when it meets all the requirements of a given business application.

All that Glitters May Not be Gold

The other side of the hybrid discussion lies in defining and understanding the limitations of public cloud services. Because customers have little or no control over the design and operation of a cloud provider's infrastructure, they must depend on complex service-level agreements and good faith to ensure that issues like system availability, security, resilience and performance are continually met. Figure 2 highlights some of the key challenges and customer concerns about the adoption of off-premises cloud services for production applications.

Figure 2: Challenges and barriers to broader public and hosted cloud service adoption

Source: 451 Research's Voice of the Enterprise: Cloud, Hosting & Managed Services, Workloads & Key Projects 2019 Q: Which of the following challenges – if any – are the greatest barriers to broader implementation of IaaS/public cloud for production applications at your organization?



This chart graphically outlines customer concerns about business applications and data in the public cloud, but it also represents issues that can often be resolved by simply hosting data onpremises as part of a hybrid ecosystem while reserving public cloud adoption for applications that don't merit these concerns. Several of these considerations are in the process of being addressed for public cloud use cases, especially if an application is designed from the ground up to be cloud-native and capable of leveraging the specific capabilities of a single cloud vendor. Unfortunately, not all public cloud environments are identical, which presents even greater challenges based on the growing interest in hybrid multicloud applications.



Of course, this hasn't prevented public-cloud-based storage adoption from growing at an incredible pace across the board for a broad range of public-cloud-based applications, as well as use cases that don't have as serious an impact on critical applications, such as BC/DR, data backup and longer-term, object-based data archiving. We believe that the key takeaway is that it's important to build a hybrid cloud strategy that provides a seamless way to host data and applications both on- and off-premises based on the most efficient combination of cost, performance, security and management capabilities.

Bringing the Hybrid Cloud Home

The New Realities of On-Premises Storage

Many enterprise storage options now abound in the public cloud, albeit with some caveats. So, the next challenge in defining an effective hybrid cloud strategy lies in establishing what onpremises systems and capabilities are needed to complement storage services in the public cloud, as well as adapt to any specific, non-cloud business applications as necessary, be they legacy, cloud-native or beyond. There's little doubt that solid-state technology is making the greatest impact on the storage industry in decades, and enterprise-class SSDs now occupy the top of the storage performance spectrum for primary, highly transactional mission-critical applications. But individual device performance is only one part of the enterprise storage equation, and the actual performance of an enterprise storage system is determined by all of the factors in its design.

The architecture of storage systems has evolved substantially over the last decade, and today, most non-primary-storage vendors are moving toward a scale-out, software-defined model as opposed to the traditional scale-up monolithic storage of the past, which was based on proprietary hardware storage controllers that connected thousands of HDDs and spanned racks and racks of raw disk capacity. Today's software-defined-storage (SDS) environments are instead based on interconnected nodes of more generic x86-64-based server hardware, usually in the form of hyperconverged infrastructure (HCI) systems and sold as either validated, pre-integrated storage appliances or as do-it-yourself software that runs on existing server nodes. This scale-out approach offers linear scalability for processing, network and storage capacity by simply adding more nodes, but it also offers a more flexible approach for combining HDD and SSD technology in the same box. Unfortunately, this approach is also termed 'hybrid,' which can be a bit confusing, and it differs from hybrid cloud in that hybrid storage only addresses the internal composition of the storage node itself.

The value proposition of hybrid storage designs lies in the flexibility hybrid server nodes offer for optimizing the storage process itself. Modern x64 systems now offer a hardware combination of dozens of processor cores and threads, gigabytes of primary memory, terabytes of storage-class memory, or DC persistent memory and the ability to host both SSD and HDD media in the same chassis. It's this granularity of resources that combine with flexible storage software in a model that allows vendors and customers to optimize storage-specific nodes based on their intended use case by optimizing the combination of hardware and software toward raw capacity, performance, or even to match the specific data requirements of individual applications. But it's



also how a well-designed storage node's capabilities can exceed the sum of its parts – in that the open hardware environment allows vendors to increase overall system performance of both HDD and SDD nodes by making intelligent choices about how to leverage the far more expensive, high-performance components for I/O-intensive tasks to obtain an optimal combination of price, performance and capacity.

The simple truth is that SSD devices can offer 10-20x greater throughput than spinning disk, and somewhere on the order of 1,000x more input/output operations per second (IOPS), which makes them exceptionally well suited to highly transactional workloads. Of course, that would be the end of the conversation if there weren't a list of other key considerations for the adoption of a modern storage platform. In reality, most IT applications don't come anywhere near requiring that much performance, especially when measured against the substantial cost multipliers that exist between SSD and HDD devices. While performance still ranks relatively high in the list of enterprise pain points in the polling results shown in Figure 3, it rarely if ever ranks in the top three overall.

Figure 3: Top enterprise storage pain points

Source: 451 Research's Voice of the Enterprise: Storage, Budgets & Outlook 2020 (n=451) Q: What are your organization's top pain points from a storage perspective? (Please select all that apply.)



The shallonges of enterprise storage have been

The challenges of enterprise storage have been fairly consistent for decades, and they aren't necessarily tied to infrastructure choices but, rather, focus on the difficulty of scaling, managing and protecting the growing mountain of stored data. Fortunately, many next-generation SDS vendors are looking to tackle the challenges of data management and governance as part of a broader trend toward content-aware intelligent storage, especially to meet the growing challenges of unstructured data. But ultimately, it's the overall cost of storage that remains a major concern for enterprise storage customers.

Unfortunately, the cost of storage is being driven by relentless storage growth, which adds to every aspect of that overall cost. The challenge for IT professionals lies in making intelligent decisions when it comes to choosing next-generation hybrid storage. There are pros and cons to every form of storage, and it's up to us as technologists to understand all the variables and match the right storage to the right workloads. On-and off-premises hosting decisions should be based on the traditional storage criteria of price, performance and capacity, as well as the previously mentioned considerations of public cloud adoption listed in Figure 2.



Public services offer a broad range of cloud-hosted block, file and object options, each with its own set of costs, capabilities and characteristics for security, availability and latency – not to mention the costs of transporting large quantities of data between on- and off-premises storage platforms. While many cloud storage options today offer extremely attractive price points for basic storage, it's important to understand the potential hidden costs of some public cloud storage offerings – where there can be charges attached to data transfers, common API calls and egress fees. This can result in surprisingly high monthly charges for companies that aren't aware of their storage usage patterns, and it's worth noting that these types of micro-costs are not an issue with on-premises storage platforms.

The New Economics of On-Premises Hybrid Enterprise Storage

In the past, most enterprise storage was hosted on large and mostly proprietary monolithic storage systems. This was a necessity because of the limitations of the hardware at the time, and the prevailing trend was to centralize networked datacenter storage on large, highly resilient primary storage systems with redundant storage controllers and networking, and then protect all of it with equally large backup systems. While this was an improvement over direct-attached storage, block-based storage area network and file-based network-attached centralized storage, the monolithic model has been proven to be limited in terms of scale and cost.

This is where the adoption of the hybrid cloud model has made the greatest changes in enterprise storage; it's becoming commonplace for data of all sorts to be stored in a number of locations, both on-premises and off. Most next-generation storage systems are now capable of adapting to this new model of decentralized storage. Hybrid cloud capabilities are becoming table stakes for storage applications that are rapidly becoming less focused on the storage infrastructure where data resides and more focused on what that data is, to whom it belongs, how it can be accessed and how it needs to be handled based on a common set of policies, automation, and an awareness of that data's contents and business value.

Given a common understanding of the raw performance characteristics of flash storage compared to spinning disk, we polled our Voice of the Enterprise group to find their position on all-flash storage adoption, and the results in Figure 4 are telling. These results reflect our earlier data that shows cost and capacity continue to play a substantial role in the enterprise storage decision matrix, and that disk-based hybrid storage performance continues to be sufficient for a large number of enterprise and evolving nearline applications. This is an important distinction because hybrid storage systems that utilize limited flash in key areas while relying on HDD for raw capacity can accelerate overall node performance substantially by streamlining the metadata and storage management process while pooling the performance capabilities of multiple drives.



Figure 4: Reasons IT organizations aren't deploying all-flash storage

Source: 451 Research's Voice of the Enterprise: Storage, Workloads & Key Projects 2019 (n=174) Q: If your IT organization has not deployed all-flash storage arrays, which of the following have been the reasons?



The fact of the matter is HDDs have been the workhorse of the enterprise datacenter for decades, and still are; the vast majority of enterprise data still resides on spinning disk technology. And HDD development continues to advance: the first generation of heat-assisted write HDDs that are to be introduced this year will start at 20TB and are expected to reach up to 50TB in the standard 3.5" form factor by 2025. Research is also underway that focuses on continued capacity growth over the next decade, which bodes well for the continued price/ capacity-oriented value proposition of HDD for enterprise storage in the future.

In addition – and in response to requests from some of the larger cloud-scale operators – there is a new generation of HDDs available now that are based on a multi-actuator design. The first iteration is a dual-actuator drive that nearly doubles the IOPS capabilities of next-generation, enterprise-class disk drives, and there are plans to add even more actuators, which will allow vendors to scale IOPS linearly. This development also contributes to the ongoing value proposition of HDD technology as a tool to address the shift toward unstructured data that's making up a rapidly growing percentage of business information. While adding multiple actuators still doesn't make HDDs comparable to flash devices from an IOPS perspective, the impact of multiple-actuator technology provides a path for increasing HDD performance in the future.

On a grander scale, the growing adoption of the non-volatile memory express (NVMe) storage interface for flash devices is substantially shifting the drive-connectivity model for enterprise storage. The NVMe protocol offers an evolved command architecture for high-performance storage, is natively supported by system BIOS and leverages ultra-high performance, internal PCIe connectivity. NVMe has also been externalized via multiple NVMe-based network protocols like NVMe over fabrics, NVMe over TCP and NVMe over Fibre Channel, which makes NVMe extensible across most existing networks. This shift isn't lost on HDD vendors, and there are specifications for NVMe-based HDD technology under review now that are expected to further extend the value proposition of HDD for seamless, hybrid SSD/HDD node architectures. The first generation of enterprise-class NVMe-HDD devices is expected within the next few years.



Conclusions

If performance were the only factor, or even the top factor, in making enterprise storage decisions, this would have been a very short paper. The goal of this report is to provide insights into the current challenges and concerns of real-world, hybrid enterprise customers in order to highlight the right questions to ask when transitioning to the hybrid cloud model. Technology choices are rarely if ever binary, a fact that is borne out by the ongoing popularity of a hybrid approach to IT infrastructure. The hybrid cloud is based on the ability to seamlessly blend both on- and off-premises cloud services that give a customer the ability to choose application hosting based on their specific needs for cost, security, performance, scalability, resilience, management and other key business criteria. The same holds true for hybrid storage, where the flexibility of evolving SDS and HCI technologies allows the customer to seamlessly blend an optimized and easily scalable ratio of SSD and HDD resources to meet many of those same specific needs.

A well-designed hybrid cloud model can provide the best of both worlds, offering greater flexibility to optimize infrastructure choices at a low level if desired while abstracting and simplifying high-level services for easier consumption. Perhaps the greatest feature of the hybrid delivery model lies in the ability to automate much of the low-end effort required of earlier infrastructure systems, and that same automation is now being extended to address many of the new, storage-specific challenges of granular data management, governance, protection and security. In the past, enterprise storage was mostly a passive process that simply warehoused data until it eventually disappeared into the backup 'oubliette.' Today, businesses are starting to realize that there's ongoing value to be had from all the data we've been spending so much to store, and the need for simplified access to that information makes it even more worthwhile to keep data nearline for a longer time.

Customer Recommendations

We are firm believers in the future of hybrid technology, but the transition to hybrid requires both a philosophical and a technological shift toward a more intelligent infrastructure that delivers high-level services rather than raw capabilities. We also believe that storage will remain a somewhat separate challenge than either compute or networking in that it requires an additional level of considerations for ensuring security, protection and governance that continues for months or years after the data was created. While many of these considerations are common for most customers, there will always need to be provisions made for the specific needs of a given vertical market. But as we said earlier, the granular control offered by intelligent, next-generation hybrid storage can go a long way in addressing the evolving needs of business data management.

- **Understand the key requirements of your applications.** Whether it's a need for bandwidth, transactional speed, capacity or any other factor, the number of hardware options now available through a hybrid approach allows much greater optimization than traditional storage.
- **Plan for increased storage decentralization.** Part of the value proposition of hybrid storage lies in the ability to place data where it makes the most technological and business sense. A hybrid environment needs to be able to see and manage data regardless of physical location.



- Make intelligent choices on data and workload placement. Part of the workload placement challenge lies in deciding whether it's better to move the data to the application or the application to the data. The proximity of the application to the data is a major component in the performance calculation, and many times it comes down to simply having data in the right location, as well as on the most efficient platform for that application.
- **Optimize the movement of data.** As data grows, it becomes increasingly difficult and expensive to move. A hybrid approach can provide a model for better understanding the optimal location of data to reduce the amount of data that needs to be moved, as well as managing the number of copies of data that may exist on multiple storage platforms.
- Embrace content intelligence and automation to manage data growth. Automation is a huge benefit of the hybrid cloud model, and next-generation storage management is evolving to deal with the relentless growth of data, especially difficult-to-manage unstructured data. The storage process needs to be more active, and customers need to think in terms of deterministic workflows that allow data to follow a predictable and automated path throughout its lifecycle.



Ready to Learn More? Visit us at seagate.com.

PATHFINDER | BALANCING HYBRID STORAGE TECHNOLOGY AS PART OF A HYBRID CLOUD DATA MANAGEMENT STRATEGY



COMMISSIONED BY SEAGATE



S&P Global Market Intelligence

About 451 Research

451 Research is a leading information technology research and advisory company focusing on technology innovation and market disruption. More than 100 analysts and consultants provide essential insight to more than 1,000 client organizations globally through a combination of syndicated research and data, advisory and go-to-market services, and live events. Founded in 2000, 451 Research is a part of S&P Global Market Intelligence.

© 2020 451 Research, LLC and/or its Affiliates. All Rights Reserved. Reproduction and distribution of this publication, in whole or in part, in any form without prior written permission is forbidden. The terms of use regarding distribution, both internally and externally, shall be governed by the terms laid out in your Service Agreement with 451 Research and/or its Affiliates. The information contained herein has been obtained from sources believed to be reliable. 451 Research disclaims all warranties as to the accuracy, completeness or adequacy of such information. Although 451 Research may discuss legal issues related to the information technology business, 451 Research does not provide legal advice or services and their research should not be construed or used as such.

451 Research shall have no liability for errors, omissions or inadequacies in the information contained herein or for interpretations thereof. The reader assumes sole responsibility for the selection of these materials to achieve its intended results. The opinions expressed herein are subject to change without notice.



NEW YORK 55 Water Street New York, NY 10041 +1 212 505 3030

SAN FRANCISCO One California Street, 31st Floor San Francisco, CA 94111 +1 212 505 3030



LONDON 20 Canada Square Canary Wharf London E14 5LH, UK +44 (0) 203 929 5700

BOSTON 75-101 Federal Street Boston, MA 02110 +1 617 598 7200

