the GORILLA GUIDE® to...

Modern Data Protection

How All Flash Storage Improves Data Resiliency

ED TITTEL

PURESTORAGE®

POWERED BY ActualTech MEDIA

the GORILLA GUIDE® to...

Modern Data Protection

By Ed Tittel

POWERED BY ActualTech MEDIA

www.actualtechmedia.com

PUBLISHER'S ACKNOWLEDGEMENTS

DIRECTOR OF CONTENT DELIVERY

Wendy Hernandez

CREATIVE DIRECTOR

Olivia Thomson

SENIOR DIRECTOR OF CONTENT

Katie Mohr

WITH SPECIAL CONTRIBUTIONS FROM PURE STORAGE

Roger Boss

SENIOR PRODUCT MARKETING MANAGER, SECURITY AND DATA PROTECTION SOLUTIONS

ABOUT THE AUTHOR

Ed Tittel is a 30-plus year veteran of the IT industry who writes regularly about cloud computing, networking, security, and Windows topics. Perhaps best known as the creator of the *Exam Cram* series of certification prep books in the late 1990s, Ed writes and blogs regularly for GoCertify.com, TechTarget, ComputerWorld, TekkiGurus, and other sites. For more information about Ed, including a resume and list of publications, please visit EdTittel.com.

ENTERING THE JUNGLE

CALLOUTS USED IN THIS BOOK

SCHOOL HOUSE

In this callout, you'll gain insight into topics that may be outside the main subject but are still important.

FOOD FOR THOUGHT

This is a special place where you can learn a bit more about ancillary topics presented in the book.

BRIGHT IDEA

When we have a great thought, we express them through a series of grunts in the Bright Idea section.

DEEP DIVE

Takes you into the deep, dark depths of a particular topic.

EXECUTIVE CORNER

Discusses items of strategic interest to business leaders.

DEFINITION

Defines a word, phrase, or concept.

GPS

We'll help you navigate your knowledge to the right place.

KNOWLEDGE CHECK

Tests your knowledge of what you've read.

WATCH OUT!

Make sure you read this so you don't make a critical error!

PAY ATTENTION

We want to make sure you see this!

TIP

A helpful piece of advice based on what you've read.

INTRODUCTION

The Importance of Reducing Risk and Business Resilience

Welcome to The Gorilla Guide To…® Modern Data Protection.

Cyber attacks. Product updates and maintenance. Natural events. Rogue admins or simple human error. Losing access to your data puts your business at risk. In an age where access to data on premises or in the cloud powers nearly every business event, uninterrupted access to data is essential for business survival..

This guide aims to provide you—executives, business and organization stakeholders, and IT managers and technical leads—with the information you need to make sense of modern data protection. Because users are more inclined to conduct business in a variety of digital forms than ever before, the data they wish to access jumps to the forefront of what lets organizations and businesses function. That's because data is what drives applications and services. Likewise, data is where insights and value ultimately originate, and data is what organizations must manage, control, and—most importantly—protect to meet legal, regulatory, and compliance requirements.

When it comes to data resiliency, the name of the game is to reduce operational risk and get back online ASAP. That's because time is money when it comes to downtime. Use of proper, modern tools and technologies also means that automation vastly speeds recovery times. In fact, automation also supports regular practice and testing for business continuity to ensure that recovery time objective (RTO) and recovery point objective (RPO) requirements are realistic and attainable. That's why they're key metrics when designing any kind of resiliency plan. Indeed, automation is also key to constant monitoring efforts—such as watching for bulk file encryption and out-of-the-ordinary data access and usage—that permits lightning fast recognition of and response to ransomware attacks as they begin. Legacy models for backup and recovery don't cut it anymore, because they're not able to provide the flexibility, resiliency, and recovery speed (or failover to avoid recovery) that modern businesses need.

In Chapter 1 we examine the basics of backup and recovery so that readers can understand how these activities—and the platforms and solutions that support them—provide data protection and the ability to recover from interruption or disaster to get the organization back to "business as usual."

CHAPTER 1

Setting the Stage for Protection and Recovery

IN THIS CHAPTER:

- Legacy systems establish a base
- The ransomware menace
- Flash-based technologies rule

Although the concept of backup and recovery is neither novel nor new, organizational requirements for fast data restore in the wake of an unplanned interruption are stronger than ever. As more organizations build their business and competitive edge atop a data foundation, rapid access to data for resiliency, disaster recovery and data reuse becomes imperative.

In response, backup and recovery techniques, which have evolved and adapted many times since their introduction six decades ago, are poised to undergo a new and significant transformation, driven hard by cloud and flash storage technologies.

But while data protection has always been vital to organizations, it doesn't always get enough attention. One reason is that the IT technologies used just two or three years ago simply can't deliver the rapid recovery scenarios necessary in today's reality. Companies were disinclined to invest in the latest and greatest tools when they

provided no real improvement to or benefits for recovery performance. Things have changed radically of late, as new challenges and technologies to address them force companies to rethink their data protection priorities and investments.

The approach to backup and recovery has changed profoundly considering the "big data" phenomenon, as companies seek to do more with backup data. It's not enough for data simply to get backed up—that same data can actually provide a competitive edge in the marketplace. A proliferation of technologies, in cloud compute and cloud storage, flash storage, and storage efficiency algorithms, plus artificial intelligence (AI) and machine learning (ML; together; AI/ML) has significantly altered the capabilities and intelligence that IT can provide by leveraging backup data.

AI CHANGES EVERYTHING

According to a March 2023 IDC forecast, worldwide spending on AI—including hardware, software, and services for AI-oriented systems and capabilities—will meet or exceed US$154B in 2023. That's an increase of nearly 27% vis-à-vis 2022. IDC says that number is expected to exceed US$300B by 2026.

Indeed, putting data to better use for internal customers is an enormous advantage. But another reason data protection enjoys elevated attention comes from the pressures that business leaders feel from growing risks of ransomware, breaches, hacks, and extended outages.

At the same time, virtualization and cloud technologies (especially with increasing adoption of and investment in hybrid multi-clouds) are changing IT's understanding of what data protection means and how it works. Assets may now be stored, backed up, and recovered

across a broad range of locations. Key concerns that emerge from virtualization also carry over to data protection, backup, and recovery. These concerns include:

- Snapshotting, checkpointing, tracking, and synchronizing relevant applications and storage assets and resources
- Performance and resource consumption monitoring, to achieve service-level agreements (SLAs) and other response time or user experience objectives, and maintain cost controls
- Enforcing security requirements for access controls, encryption for data in motion and at rest, and user and administrative privilege management, to guarantee the principle of least privilege and to audit access to sensitive data, use of administrative privileges, and more.

Organizations must understand that the best and final line of defense against ransomware is a known, good working backup (preferably, one that's tamper-proof), that can be quickly and reliably restored to bring your business back online.

And, finally, ransomware attacks underscore the vast importance of access to tamper-proof, clean, and malware-free backups. Indeed, paying attackers does not guarantee decryption will work. That's why the FBI says, "Pay no ransoms, period." Time considerations also argue against paying ransoms, because lengthening delays impose increased opportunity costs and reputation damage.

Legacy Systems Establish a Base

Early on, use of magnetic tape for back-up countered the high cost of hard disks, plus the intrinsic value of tape itself. But recovering from tape takes time; is prone to failure (tape is fragile and degrades as it ages); and can't meet aggressive RTOs.

Spinning disks are also limited. They have device identification issues, and repair and replacement challenges. Spinning disks fall prey to mechanical problems and failures, as a glance at Backblaze Hard Drive Stats[1] affirm. RAID arrays require trade-offs, where more redundancy, performance, and availability also add cost and complexity.

Because network-accessible hard disks often lack retention locks or version histories, they cannot act as "gold masters" for restores. Comparing recent snapshots to previous ones, you can identify changes and revert to previous versions. In fact, sudden changes in snapshot size, activity levels, and encryption activity can signal a ransomware attack. Other technologies offer improved protection, faster restores, and work from immutable (tamper-proof) snapshots guaranteed to be clean and virus-free. Such technologies are covered in more detail in Chapter 4 of this Guide.

Legacy backups also work per-server or per-department. This creates disparate, disjointed, and sometimes incompatible media best regarded as "backup silos." Backup silos are suboptimal in today's IT environments where customers seek to improve efficiency and IT admin productivity. Plenty of room for improvement!

[1] https://www.backblaze.com/b2/hard-drive-test-data.html

The Ransomware Menace

The threat level that ransomware poses is hard to overstate. Indeed, the numbers speak for themselves. The average ransom demand in such an attack in 2023 is US$1.5M.[2] In its 2023 Global Ransomware Report,[3] Fortinet discloses that over half of all global organizations report one or more such attacks in 2022, of which 71% paid at least some ransom. Also, paying attackers does not guarantee successful recovery nor quick resumption of operations: only 35% of those attacked recovered all their data.

Some sectors are more susceptible to ransomware attacks than others, particularly healthcare. An October 2023 Healthcare Dive brief reports costs of US$77.5B for downtime from 2016 through October 2023 (that's nearly US$0.85B a month, just for downtime not including ransoms or other recovery costs). Other oft-attacked sectors include critical infrastructure, government, healthcare, finance and more.[4] Ransomware is everywhere, attacks are inevitable, and recovery imposes significant costs. It's a clear call for prevention and protection.

A TYPICAL RANSOM SCENARIO

Let's examine a highly reported and well-understood ransomware attack to better grasp the typical sequence of events, their timeline, and costs and consequences. In May 2021, Colonial Pipeline—an Alpharetta, Georgia-based oil pipeline and fuel delivery company that supplies much of the eastern U.S. with refinery outputs such as gasoline, aviation fuel, and home heating oil—suffered a ransomware attack.

[2] Tech.co: Ransomware Statistics: Key Trends, Insights and Questions Answered

[3] Fortinent: The 2023 Global Ransomware Report

[4] Zscaler: 2023 ThreatLabz State of Ransomware

The attack proceeded in stages, through the in-house network. Within two hours of penetration, attackers exfiltrated as much as 100GB of data from the company's systems. Immediately thereafter, they infected as many systems with ransomware as they could, including those for billing and accounting.

This attack affected the company's digital systems (but not those that operate and track pipeline operations). For security reasons, Colonial shut its pipeline down to prevent further spread of the ransomware. The company went offline for six days while bringing its systems back up. At the same time, the company paid a US$4.4M ransom to the DarkSide cybercrime gang that originated the attack. Ultimately, the FBI tracked and recovered about 85% of the payment (63.7 bitcoin, worth US$2.3M at the time, of the total 75 transferred to the DarkSide drop-box).

Here are key take-aways from the Colonial Pipeline incident:

- As the attack got going, Colonial Pipeline failed to observe numerous tell-tale signs of intrusion and exfiltration.
- Only as systems were locked up and ransom demands appeared, did Colonial recognize the attack. Monitoring tell-tale activities could have raised an alarm far sooner (especially wholesale data encryption).
- Because it lacked secure, protected, and immutable backups for its locked-up systems, Colonial had to pay the ransom to restore its systems.

There is a silver lining to this decidedly dark tale, though. **The real key to avoiding ransomware's denial of service and access is the final ingredient listed here—namely, a secure, protected and immutable backup.** With the right tools and technology solutions in place, organizations need not fall prey to ransomware attacks. Nor should it take them six days to return to normal operations afterward. An important foundational technology for the right kind of backup and recovery is featured in the following section—namely flash-based storage.

FLASH-BASED TECHNOLOGIES RULE

Flash-based storage uses special non-volatile random access memory (NVRAM) chips to store information on silicon. It does not rely on conventional storage technologies such as solid state disks (SSDs), spinning hard disk platters (HDDs), or linear magnetic tape. Such flash-based technologies use their own, built-in optimization and monitoring tools to manage storage directly, rather than queueing up storage requests to a disk controller that then takes over and manages reads and writes independently. This provides considerable speed and control advantages. More on recommendations in Chapter 4.

CHAPTER 2

Beyond Backup and Recovery

IN THIS CHAPTER:

- The old backup and recovery mindset
- It's not about speeds and feeds
- Ransomware is on the rise

There's more to bringing back business operations than the mechanics of backup and recovery. Organizations must always meet or exceed their SLA recovery requirements. In fact, a more modern approach is emerging in the marketplace that goes beyond backup and recovery. Modern IT must be able to support a wide range of applications with diverse demands, including security and data protection applications, replication and snapshotting, transaction and activity logging, ransomware recovery, data analytics, AI/ML, and more.

Many experts believe that it truly is a matter of "when" more than "if" where ransomware attacks are concerned, no matter your industry, markets, customer base, or location. A strong and wide-ranging data protection strategy supports a trend toward data-driven outlooks to squeeze more value from data.

The Old Backup and Recovery Mindset

In the past, IT traditionally focused on RTOs and RPOs—long the standard by which IT success was measured. In many cases such measurements defined SLAs between the business and IT. Business managers, on the other hand, focus on availability, resilience, business intelligence, expansion, and time to market.

Two types of events within an IT ecosystem can affect SLAs and introduce risk into your business: business interruptions and declarations of disaster.

A business interruption is any event that might cause production or productivity to be hindered or come to a halt. Part of the definition is a "predetermined time," or an agreed-upon time interval during which the business can tolerate downtime for some particular function or service. When that interval is exceeded—and applications and data remain unavailable to users and customers—a business interruption turns into a declaration of disaster.

A typical DR plan outlines a process or procedure for declaring a disaster. Once one is declared, you pull the cord and it's all hands on deck. At that point, RPO becomes the focus, and the question becomes how long until "business as usual" can be resumed—however that's defined. Addressing the needs of the business requires asking the questions most important to the business in their terms—not IT's terms. This is what a modern enterprise needs, and what previous solutions lacked.

The pressures that business exerts on IT pave the way toward a new framework to meet business demands. It's a modern approach, not only in the technology it uses, but in its practical emphasis on meeting the needs of the business, first and foremost.

Simply put, the chief reason data gets protected is to facilitate its recovery and restoration in case of emergency or disaster. In IT terms, this is RTO. To the business, it may be a predetermined time, codified in an SLA. Even more narrowly, it might simply be called "availability."

The point here is that the realities of a business interruption mean different things to different groups within an organization. That said, the outcome should always answer the same question: How rapidly can we return to business as usual? The cost of not doing business in the wake of a business interruption or a disaster can be staggering. A failure can even impact the company's image and ultimately threaten the viability of the business itself. It's this threat, and its accompanying urgency, that drives organizations to undertake data center modernization. Indeed, modernization inevitably begins with every organization's crown jewels—namely, its data.

It's All About Recovery

When you look at solutions to help modernize your company's approach to data protection, it's essential to consider desired outcomes primarily in terms of recovery or resumption of activities.

Companies naturally want the fastest solution on the market. A flash storage system can provide the necessary recovery speed, but the data protection software used is every bit as important as the hardware used for storage.

Important: Recovery times matter more than backup times.

Indeed, the scale of restore is far greater than it was: it's possible you might need to restore an entire data center after a ransomware attack. That's a whole different story from old-fashioned views where you

may have needed to restore files and folders, or perhaps a few systems. For all these reasons, the notion of "time to data" is crucial when it comes to achieving recovery and meeting RPOs

To repeat for emphasis: It's not about speeds and feeds. It's about the whole business.

It's a Different World Now

As stated in the preceding chapter, ransomware is by far the most menacing threat looming over organizations, today and in the future. Modern backup, recovery, and data protection tools and platforms must be at home in such a world. That means capable and modern solutions must be able to back up to the cloud, recover from or to the cloud, and be ready to protect data wherever it lives. That means on-premises, at the edge, or in any and all of the clouds—private, public, and hybrid—that an organization uses. But it also makes the case for these special kinds of capability—namely, reliability, endurance, speed, and scale.

Ransomware on the Rise

Ransomware is now a clear and present danger for businesses and organizations of all kinds and sizes, public and private. It's also a major source of expense and financial loss for companies and organizations that suffer such attacks. A July report from U.K.-based tech.co,[5] forecast the 2023 global cost of ransomware over US $30B. The same report states "80% of people who pay a ransom will be attacked again."

[5] Ransomware Statistics 2023: Key Trends, Insights and Questions Answered

One of the most publicized ransomware attacks in 2023 was a social engineering attack on MGM Resorts International. MGM disclosed a cyber attack after guests reported issues related to room access, amenities and casino games that persisted for days. MGM's 8-K filing confirmed systems were offline, and that "swift response prevented threat actors from accessing any customer bank account numbers or payment card information." While MGM restored many of its systems and said affected operations have resumed as normal, remediation efforts proved costly, estimating the attack would cause a $100 million hit to its third-quarter results.

The only sure protection from ransomware is a known, good, tamper-proof backup that can restore systems to proper, uninfected, and unaffected operation quickly. This hinges on creating read-only backups that are carefully monitored, closely protected, and accessible only to highly trusted programs, users, and accounts.

Today's Recovery SLAs Are More Stringent

It's also the case that businesses and organizations bound to customers, clients, or end users by recovery SLAs find themselves increasingly hemmed in by shorter recovery windows, higher uptime requirements, shrinking budgets, and less tolerance for service delays and interruptions. Ultimately, this means that the backup, recovery, and data protection solutions they deploy to meet such requirements must be faster, more resilient, and better able to withstand the rigors and demands of today's world. While flash-based solutions were already attractive, as discussed in the preceding chapter, the pressure that stringent SLAs can exert make them both irresistible and necessary.

CHAPTER 3

Key Industry Trends

IN THIS CHAPTER:

- Data resilience and disaster recovery
- Hybrid data lives everywhere
- Storage in the cloud

The technology and industry landscape also defines how an organization should plan for and set recovery objectives. This landscape covers a variety of related tools and technologies, including data protection and disaster recovery. In the following sections, we'll review emerging considerations about which technology executives, stakeholders, and architects should be aware.

Data Resilience

Your business needs to be both resilient and agile, and preparing for the realities of recovery requires a future-proofed IT infrastructure built on modern data protection. A business that is both resilient and agile:

- **Expects the unexpected**
- **Is prepared to meet the challenges of change**
- **Can adapt to change as quickly and successfully as possible**

Applied to data, resilience means remaining available, no matter what, to both users and essential applications. This is especially true for the apps that a company relies on for modern data analytics. And it's about more than improving business continuity, although that's critical to maintaining robust data protection, too.

According to research conducted by Bredin, 67% of companies see data security, together with compliance, as the biggest risk to building a digital business. And 60% of those companies said they plan to invest in data security **this year** to reduce risks.

Building resilience and agility—not only for data but also for the business overall—requires future-proofing critical IT infrastructure and having a modern data protection strategy. This can help you safeguard your application data in a constantly changing threat environment, while also keeping it available enough so you can move fast and recover the data you need, when you need it.

A modern data protection architecture can help your business better meet the challenge of pervasive and costly threats but also keep you agile:

- **Before an attack:** By backing up your files through frequent snapshots, controlling access to secure files and data, investing in training, and more
- **During an attack:** By letting go of data protection infrastructure built on legacy architectures that can't handle the stress of a ransomware attack (and, if your backup systems and data are compromised, you may need to reinstall and reconfigure your backup tools before you even try to recover your data)
- **After an attack:** With the ability to recover, fast, in the wake of an attack (a modern data protection approach emphasizes simplified and secure ransomware backups that can't be modified, deleted, or encrypted).

Cloud technology allows organizations to virtualize data center resources. Given an efficient backup and recovery program, data can reside in some cloud instance. From there, it can be used by locally re-hosted applications or accessed remotely via wide-area or metropolitan-area links from servers in business data centers or user facilities. The affordability of clouds has never been better.

HYBRID DATA LIVES EVERYWHERE

In a May 20, 2023, blog post, "Cloud Adoption Statistics for 2023,"[6] Hosting Tribunal reported that "Organizations leverage almost five different cloud platforms on average" including both private and public clouds. They also assert that 94% of enterprises use one or more cloud services. When organizations seek to combine on-premises systems and assets with their cloud-based counterparts, the result is what's called a *hybrid cloud*.

A hybrid cloud brings all systems together in as seamless and flexible an environment as an organization might be willing to design (see **FIGURE 1**). Tying these environments together involves major effort and complexity. IT often chooses to rebuild or replace legacy tools and applications that cannot accommodate the APIs and connections a hybrid cloud needs to operate. Where "remove and replace" with something (hybrid) cloud-native is infeasible, an organization faces diminished flexibility when positioning and running workloads. In a hybrid cloud environment, managing data flows and workload placement become essential. Organizations may choose to run certain workloads on-premises and limit data outflows into the cloud, because they wish to keep them confidential or to meet specific compliance and governance requirements. Other workloads may get positioned in the cloud, or placed at the edge to provide users with the best experience (low latency, high performance, better response times, and so forth).

[6] https://hostingtribunal.com/blog/cloud-adoption-statistics/#gref

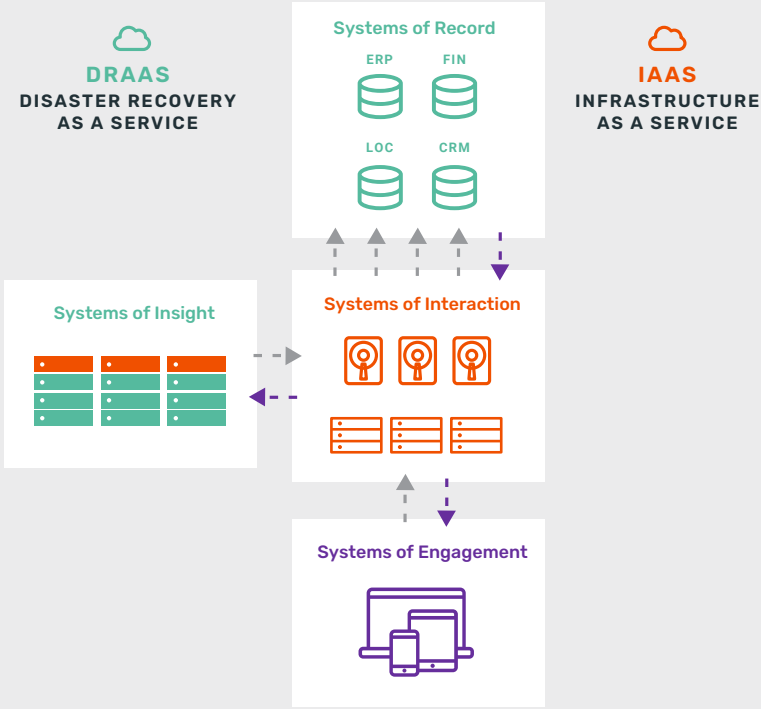


FIGURE 1: In a hybrid cloud model the corporate data center obtains select services from a cloud service provider (for example, Disaster Recovery as a Service or DRaaS) and additional compute, network, or storage resources on an as-needed basis from other clouds

In a hybrid cloud environment, organizations must juggle storage costs across a variety of forms. They must weigh the fully burdened costs of on-premises storage against the consumption and usage costs of cloud storage. The best economics are possible only when organizations can trade off unit costs for on-premises storage against comparable unit costs for cloud storage. Not surprisingly, in

some cases this leads to "reverse migration" where data stored in the cloud moves back into a data center because it's shown to be more cost efficient to house that data locally.

With the hybrid cloud in the picture, backup and recovery, along with BC/DR, become more fluid and complex. Organizations must understand the cost, security, and performance implications of creating and implementing hybrid cloud solutions. Only if solutions make sense should they be deployed. And if they do get deployed, they must be monitored and managed carefully to avoid unwanted or unexpected costs that sometimes loom large on the cloud side. Capacity planning and consumption planning turn out to be two sides of the same coin, where one side is in the data center and the other side is in the cloud.

Disaster Recovery

As with data protection, disaster recovery has on-premises and cloud- based options, including Disaster Recovery as a Service (DRaaS) offerings. On-premises, organizations must assess RTO and RPO intervals to choose among storage options that include tape, conventional hard disks, and flash elements. Tape makes sense only for long RTOs and RPOs, simply because of the time involved in reading from that linear medium (and, if necessary, transporting that data across a network with all its latency considerations). Hard disk drives (HDDs) remain the most common choice in many data centers, but can impose delays because of data access times and lack of scalability.

The advantages of architecting a modern approach to DR using flash are many. For one thing, all-flash performance can help firms struggling to meet data recovery standards as defined in IT SLAs. All-flash

storage brings the speed of data and system recovery and restore up to the speed of data backup, which has been the focus of most improvements in data protection over the past 20 years.

By enabling multiple uses for backup data, flash extends its value beyond risk reduction to providing additional business value. If consumed properly, backup data confers both cost containment and improved productivity. Business management prefers IT initiatives or strategies that deliver value in multiple domains, so a modern data protection strategy based on flash storage is more likely to gain budget approval.

And, finally, all modern approaches must consider cloud-centered technologies. Solutions that make backup data portable to and from the cloud—as do flash storage technologies—offer two key benefits. First, they provide necessary separation between original and backup copies of data to ensure resilience in the face of facility and regional disasters. Second, they provide a way to leverage cloud economics and flexibility needed to reduce cost and complexity in backup infrastructures. Indeed, DRaaS options are often highly attractive because of their cloud capabilities (they can back up and restore on-premises and cloud based assets, and handle hybrid cloud scenarios, including failover within or across multiple clouds).

By contrast, older, legacy storage technologies like tape and conventional hard disks, are typically too slow and cumbersome to meet today's RTOs and RPOs. Anything under four hours for either objective is unattainable with either tape or spinning disk.

Likewise, the following issues all lean heavily in favor of flash-based storage:

- **Device size.** Because the volume inside data center racks determines how much storage a rack can accommodate, compactness of flash devices puts them ahead of both tape and disk drives.

- **Heat output.** Flash devices run much cooler than spinning drives.
- **Energy consumption.** Flash devices require less power than either tape or spinning drives.
- **Failure rates.** Flash devices fail less often (lower mean time between failures, or MTBF) and have far lower bit error rates than either tape or spinning drives.
- **Media management.** Flash devices are accessible as soon as they're installed and connected. Tape devices use machines or human operators to swap media. Disk arrays impose performance penalties when drives fail, and require complex operations to integrate replacement drives while rebuilding an array.

Thus, even though flash devices remain more expensive per storage unit, many organizations have found their other characteristics (especially support for shorter RPO and RTO intervals, plus energy and space savings) compelling enough to make flash an essential technology. And with flash prices continuing to decline (Inc. says such costs have bottomed out, but that flash memory should remain at or near current levels until mid-2024 as undersupply lifts market costs), the tilt toward flash storage continues to increase.

STORAGE IN THE CLOUD

There are multiple storage scenarios to consider when the cloud comes into play. These may be characterized as follows:

- **Storage in the cloud.** This means consuming storage space in the cloud provider's environment, usually through some kind of vendor- and platform-neutral storage layer software and its supported APIs. Costs are incurred for space consumed and for activity involved in accessing its contents.

- **Recovery in the cloud (failover).** This means using a cloud provider's facilities, including storage, to recreate a data center using cloud-based infrastructure. Storage is only a part of this kind of operation, albeit an essential one. Costs are incurred for instantiating entire IT infrastructures, including compute, memory, and network resources, as well as storage. Large recurring costs for storage consumption are typical for such scenarios.
- **Recovery from the cloud (failback).** When a data center comes back online, a cutover from the cloud environment used for failover is required. This takes time and involves substantial egress charges from the cloud provider, plus charges for resource usage and consumption while the failback process is underway. This activity can be costly, too.
- **Using Kubernetes-based containers in the cloud.** Because this approach best uses efficient, flash-based object storage, it typically consumes fewer resources (compute, networking, and memory) and less storage space (thanks to deduplication and compression) than file- or block-based alternatives. This translates further into lower overall costs and faster, more responsive containerized apps and services. Flash is the best technology currently available for cloud-native Kubernetes-based containers, apps, and services. As discussed in the section on hybrid clouds, calculating cloud costs can be messy and complex. Flash helps simplify those calculations and keep those costs down.

There's considerable effort involved in translating between cloud-based storage costs and storage costs for data center-based systems and assets. The cloud's consumption-based, pay-as-you-go cost model depends on periodic billing based on storage space consumed, plus related usage costs for accessing and manipulating its contents and for sending and receiving related network traffic. Storage space costs are constant, while usage-based costs vary by activity and traffic levels.

The data center's fully burdened cost model is a whole different animal. It includes:

- **Capital expenditures** to acquire equipment, usually expressed through amortization and depreciation chargebacks to users. In addition to storage devices themselves, this gear includes the racks in which they're mounted, the devices and cables that connect them to the network, servers and stations to configure and manage storage, heating and cooling systems, backup power facilities to keep data centers running when local power fails, and so on.
- **Property costs** to house equipment, which may involve either capital expense (for company-owned property) or operational expense (for leased property).
- **Overhead costs** for planning, financing, purchasing and contract negotiation, and so forth.
- **Lifecycle management costs** to cover service and support contracts, and the necessary maintenance and upkeep for physical equipment and facilities.
- **Operation costs**, which include power needed to run the equipment, lights that are turned on, either more power or circulating water to keep the data center cool, and so on. Also, usage and subscription costs for cloud-based services that the data center may consume—such as Storage as a Service—also count.
- **Staffing costs**, which include the fully burdened costs of employees who work in the data center, the IT staff that uses and manages local data stores and related systems, their managers, and so forth.

Putting all these costs together and working out unit costs for storage for cloud and data center use takes time and effort. But it provides a helpful and rational basis for comparison, so that managers and stakeholders can define policies to guide storage placement and usage, and so IT can monitor, manage, and tweak them as needed.

PITFALLS IN MAKING CLOUD STORAGE WORK

Implementing data redundancy in the cloud can be a challenge, especially when it comes to balancing costs against capability, and meeting RTOs and RPOs. Organizations must test and pilot prospective implementations before making any commitments, primarily because testing and practice are the only benchmarks that provide necessary insight into what happens when a real outage or disaster strikes.

It's also essential to set the right priorities for recovery. These priorities will drive the choice of a cloud positioning and recovery strategy that is sensible, affordable, and workable. Key items include line-of-business applications, email, CRM, e-commerce, and website presence. In short, anything that can materially affect business activity and capability had better be on that list, and its RTOs and RPOs tested against what on-premises versus in-the-cloud recovery can deliver.

Beware of Ransomware!

Cybersecurity Ventures reports the costs of ransomware attacks are growing furiously. From $8B in 2018 to $11.5B in 2019 to $20B in 2021, the organization predicts ransomware will cost victims an astonishing $265B annually by 2031 (presuming 30% CAGR), with ransomware attacks occurring every 2 seconds (same source).

This omnipresent threat demands that organizations do everything they can to monitor and secure their systems and data against ransomware attacks.

Be aware that ransomware attacks seek to encrypt backup storage as well as primary storage holdings and filesystems. That dictates either an immutable, tamper-proof backup store, or physical isolation of backups from the network (and attack).

Automation Is Key

SLAs emphasize specific metrics that must be continuously monitored and compared to guaranteed levels or minima. Because time plays a vital role in certain SLA metrics, it is vital to respond to deviations below the minimum acceptable level in milliseconds. Thus, for example, consider uptime SLAs, which commonly fall in the four-nines to seven-nines (99.99% to 99.99999%) range. (Uptime specifies the percentage of time a system or service stays up and running over some total time interval.) Five-nines (99.999%) uptime allows *five minutes* of downtime in any given month (except February, which gets only 4.5 minutes because it's so short). Seven-nines drops that to 3 seconds per month. Given requirements to meet such SLAs, speedy response is the name of that game. This makes automation essential, to read and respond to SLA metrics in tens to hundreds of milliseconds. Humans are lucky to respond to alarms or alerts in under a minute, which is unfortunately 600 to 6,000 times slower. Without a well-orchestrated and automated set of tools and responses, organizations cannot hope to honor SLAs. Well-tested automation also eliminates human errors or mistakes, which can slow responses even more.

RECOVERY TIME AND RECOVERY POINT OBJECTIVES

RTOs and RPOs assign hard and fast timing requirements for system availability. An RTO defines how long a computer, system, network, or application can be unavailable after a failure or disaster strikes (see **FIGURE 2**). Essentially, RTO states how long it should take to return something (computer, system, network, or application) to working condition when it fails or becomes unavailable. RTOs may be measured in seconds to days, but it's important to understand that choosing an RTO conveys how long the organization can go without the asset to which that objective applies before it causes intolerable or unaffordable damage.

RTOs should prioritize applications by their importance and potential losses that result from their absence. Resources must match the resulting intervals involved. RTOs under 10 minutes usually require failover services. Up to four hours, some kind of near-real-time failover is required. A four-hour RTO usually leaves enough time to perform bare-metal recovery and restore working applications and data access on-premises using a designated recovery team on staff. An RTO of eight or more hours usually permits IT to delegate recovery to a service provider.

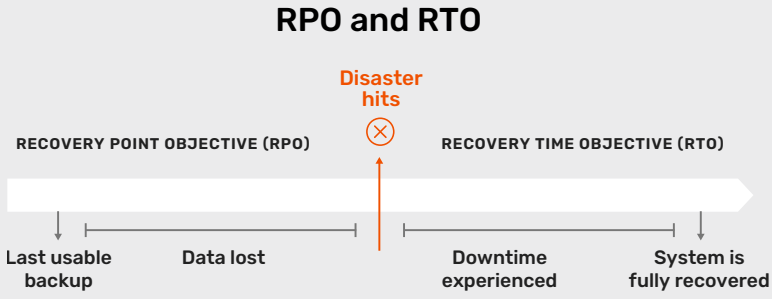


FIGURE 2: The differences between RTO and RPO

RPO, on the other hand, refers to how much data can be lost before the organization experiences intolerable or unaffordable losses. An RPO is expressed as a time interval that stretches from the time of failure or disaster back in time to the most recent backup that precedes it. Thus, for example, an organization that backs up all or most of its data once a day (every 24 hours) should be prepared to sustain a loss of 24 hours' worth of data. For some applications this is tolerable; for others it could be catastrophic.

MIND THOSE RTO AND RPO GAPS

When contrasted with typical RPOs for "normal" disaster recovery, ransomware's added complications can impact SLAs. This increases the risk for loss, reputational damage, regulatory penalties (when breaches also occur as they often do), and more. If everything worked as it should, transactions would never go missing, or might involve only minutes of downtime. When ransomware pops up, that interval—the gaps that RTO and RPO are meant to cover—can last for days or weeks. According to ESG, this is "…why executive management sees ransomware as a top business issue and not just an IT issue."

As with RTOs, shorter RPOs involve more expensive, time-sensitive technologies. RPOs under 10 minutes usually require real-time mirroring or replication solutions, so that another site maintains a live, available copy of all covered applications, data, and so forth, subject only to whatever data deltas (explained earlier in Chapter 1) apply to the link between the primary and secondary sites. Four-hour RPOs generally require snapshot replication (from primary to secondary),

whereas eight-hour RPOs often work with existing backup solutions so long as they don't affect the performance of production systems significantly.

When RTO and RPO are both near zero, organizations must invest in technologies that combine continuous replication failover services. This gets your priority systems and services as close to 100% availability as the applicable SLAs from your service providers and your local infrastructure will allow. A five-nines SLA may sound like a good idea, but can come at a formidable cost. Only you and your organization can know if that cost is justified, so consider potential losses and impacts carefully.

MANAGING CHANGING PRIORITIES

Establishing business continuity and supporting disaster recovery (BC/DR) is never a matter of "one and done." Regular practice is needed to make sure that your IT staff, stakeholders, and other players can execute their roles to meet applicable RTOs and RPOs. Over time, business conditions will change, as will available technologies. These changes make BC/DR a perpetually moving target: priorities will shift, which makes recalculating RPOs and RTOs vital. New technologies can change what's possible—or at least what's affordable—and will require occasional reworking of the organization's technology investments and service purchases.

Best practice dictates checking disaster recovery regularly, and enacting full-blown recovery drills anywhere from annually to quarterly. Other tests should occur more frequently, including convening the DR team monthly to keep team members in sync. It's a good idea to bring backup personnel into the meeting once every two months, to convey current practices to those who must step in when primary team members might be unavailable. In general, intervals between tests should reflect ongoing changes in how the organization

conducts itself, and how often network configurations, staff, technology tools and plat- forms, and compliance requirements (don't forget those!) change.

Protect from Cyber Threats for Business Resiliency

Ransomware poses profound risks to business. These go beyond downtime and associated opportunity costs to existential threats that might severely diminish or even close down a line of business (or the entire enterprise). This makes data protection and the ability to recover operations within a predetermined interval (as set by SLAs, RPOs, and RTOs) absolutely essential from a business standpoint. Indeed, this is why you will often hear that "ransomware is not just an IT problem: It's a *business* problem." And indeed, locking employees, contractors, partners, and customers out of your infrastructure is a huge business problem from the moment it starts until the moment things get back to normal.

In view of the requirements for true business resiliency, and a measured and predictable "return to normal" process in the face of threats and disruptions, companies must look for the following characteristics from any capable data protection and recovery solution:

- **Ongoing monitoring, automated and rapid notification: Attacks can occur at any time, through any potential point of access or ingress.** Continuous monitoring of network traffic and system behaviors can illuminate potentially dangerous activity or behavior. Automation supports rapid notification and pre-emptive responses (e.g., blocking external connections, shutting down affected file systems, and so forth) that permit companies to block attacks before they can compromise entire systems or networks. A proactive approach to threat management helps limit damage and speed recovery from the get-go.

- **Immutable backups: Increasingly, ransomware targets backups first and foremost when it encrypts files to keep victims from using them to recover and resume operations.** An immutable backup cannot be changed, so it cannot be encrypted, either. This is a huge step toward undoing the effects of ransomware.
- **Recovery inside predictable time limits: It's vital to test and verify that restore operations work, and that they complete within the time limits imposed by RTOs and SLAs.** The only way to ensure things work as planned, is to put those plans in action and time them to completion. In fact, regular practice will help companies recognize and accommodate changes, and prepare them for a real disruption when it occurs.
- **Accommodate recovery siting: Recovering a data center is a vastly different experience from recovering a server.** Ditto for recovering a distribution application or service that runs in a hybrid or multi-cloud environment. Recovery could easily involve failover (and then later, failback) situations where an immutable copy gets recovered in a location different from the one that went down. Working through all the details involved, and timing things, is likewise essential to handling "the real thing" should it occur.
- **Scope and scale for recovery.** Companies will need to make sure their recovery regime covers multiple scenarios. These could range from single site, server, or service attacks to disruptions on multiple fronts. Here, again, practice with internal red teams or external white hat attackers must often be a part of breaking security to see how resilient things really are, and how long it takes to work through the recovery process. This presents tremendous opportunities to plan and practice for trouble under controlled circumstances where learning, not loss or damage, becomes a positive outcome.

Ultimately, ensuring that the right processes, people, and tools are in place to handle recovery means that a business can manage the otherwise untenable risks that ransomware and other attacks might pose. Only by taking a "check, verify, adjust, and repeat" approach can companies ready themselves to get past the likelihood where statistics say there's a "reasonable likelihood" that an attack will occur sooner rather than later.

CHAPTER 4

Pure Storage Modern Data Protection

IN THIS CHAPTER:

- Understanding business-centric IT
- The importance of availability
- The value of rapid recovery

The combination of requirements for storage and backup calls for agile, secure, and integrated storage solutions, with cloud-native capabilities. Enter Pure Storage®, a global technology company headquartered in Santa Clara, California.

Pure Storage develops flash-based storage solutions for data center use, delivered on-premises, in the cloud or as-a-service. Pure Storage uses custom-built devices called DirectFlash Modules or DFMs for its fastest storage solutions. These DFMs work with custom Purity software built into its customized, high-speed device controllers. That's how a Pure Storage-based design outperforms other forms of storage for backups and data protection.

Pure Storage solutions deliver serious benefits over legacy storage and hybrid offerings. Those include:

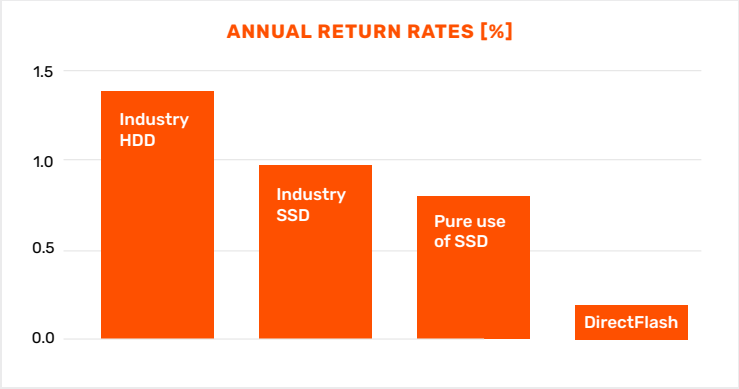
- **Enhanced Reliability and Endurance:** DFM devices use every bit of flash capacity: there's no over-provisioning, no internal data relocation (e.g. garbage collection), and no need for DRAM to cache data location maps. DFM controllers handle this independently. This approach increases reliability and reduces write amplification for improved endurance.
- **Energy Efficiency:** DFMs deliver storage density two to three times that of conventional SSDs, and consume less energy (39-54% fewer watts) per terabyte of storage. This makes data centers less energy intensive, not only because of reduced consumption, but also because DFM devices are smaller. They require less space and cooling that increases sustainability and lowers the carbon footprint. Looking ahead, Purity expects to increase its density advantage to 10x, which will further multiply these advantages.
- **Speed and Simplicity:** Because Purity software handles most of the data management, DFMs are simpler than SSDs. This makes them more robust. From a user perspective, their performance is more consistent and predictable. Also, a single DFM can make over 100 concurrent data transfers. This means DFMs in Pure Storage systems deliver better performance per device than systems built around conventional SSDs. As an example, a FlashArray// XL™ system (using Purity software and DFM devices) can deliver 36GB/s of random I/O. A Pure Storage FlashBlade// S™ (using conventional SSDs and controllers) delivers 6GB/s of streaming throughput; that's 6x more.

The superior speed and scalability of Purity-DFM flash storage make it an increasingly popular and common choice for backup and recovery—especially for recovery or failover, where every second counts. Although it remains more expensive than conventional SSDs and hard disks, Purity-DFM flash storage is better suited for multi-purpose

uses. Better yet, it is no longer cost-prohibitive to deploy in volume. It's also the case that flash storage helps future-proof modern IT platforms. Incorporating such storage not only helps organizations meet today's increasingly aggressive demands, it also lets their IT platforms scale linearly and grow to meet future demands.

DirectFlash Improves Reliability

5x over conventional SSDs, 7x over nearline HDD



- Low failure rates result in higher efficiency.
- Failure rate does not depend on capacity, which is critical to scaling.
- DMs also have a longer lifetime compared to HDD and SSDs.
- Pure1 telemetry drives continued improvement through field visibility

FIGURE 3: DirectFlash delivers the highest reliability and longest lifetime

The primary hardware-based product lines include:

- FlashBlade//S for unstructured data is a Unified Fast File and Object storage platform that serves as the storage foundation for high-performance and high-throughput access to file and object workloads.
- FlashArray//C™, based on QLC flash devices, provides an all-flash storage infrastructure that eliminates imaging delays and accelerates business-critical application performance and response.
- FlashArray//X, based on high-end NVMe flash devices, also provides an all-flash storage infrastructure that eliminates imaging delays and accelerates business-critical application performance and response.
- FlashStack®, developed in partnership with Cisco, provides a complete compute, network, and storage solution for modern IT infrastructures.
- The Pure//E Family provides all-flash storage at the cost of disk devices, but without their shortcomings in speed, energy efficiency, scale, reliability, and endurance. This family of products includes both FlashArray//E™ and FlashBlade//E™ offerings.
- AIRI®, developed with NVIDIA to extend NVIDIA's DCX A100 systems with FlashBlade devices, supports modern AI infrastructures that accelerate end-to-end GPU workflows.

In addition, Pure Storage offers software and related services and solutions. These include:

- Pure as-a-Service™, software-defined storage that works on-premises and in the cloud; it unifies your environment with a single subscription and one set of storage services. Supports both AWS and Microsoft Azure public clouds.

- Portworx®, cloud-native, Kubernetes-based data services platform that offers persistent storage, data protection and security, disaster recovery, cross-cloud and data migrations, and automated capacity management for Kubernetes-based applications. Works on any cloud; used at many Fortune 100 companies.
- Pure1®, smart AI-based storage management solution that can forecast application and infrastructure needs and provide continuous monitoring and proactive issue resolution. In October 2023, Pure Storage announced Pure Protect™ //DRaaS with energy efficiency guarantees for its Evergreen portfolio, and AI-powered services for storage management.
- Pure Storage Evergreen® storage subscription services permit rapid upgrades and storage expansion without disruption. Evergreen lets customers benefit from the latest software, hardware, and flash features with guarantees for data reduction, energy efficiency, data resilience, and data protection.
- Pure Cloud Block Store™, easy data mobility, data protection, and consistent, flexible block storage that runs natively on AWS and Azure clouds. Provides enterprise-grade data services for high availability, data reduction, and replication for applications in one or more public clouds, with a consistent experience across hybrid and multi-cloud environments.
- The Purity Operating Environment runs on Pure Storage's own FlashBlade and FlashArray solutions (and its Cloud Block Store environment) to provide secure, highly scalable, and user-friendly storage and data management in hybrid cloud environments. Purity includes core data protection, data reduction, data access, and data management features across the platform at no additional charge.

The Importance of Availability

From a business perspective, availability is making applications, data, and processes accessible to users or to consumers. This means that however the organization defines such access, it allows "business as usual" to proceed unhampered and unhindered.

When key applications or data become inaccessible, the ripple effect can be horrendous across the board. Thus, the primary purpose of modern data protection is to ensure data accessibility, availability, and resilience in the face of interruptions, outages, or attacks. It's what lets organizations get back to business as usual as quickly as possible.

The best way to start this journey is to keep your data from being lost in the first place. Building data resiliency into your environment starts with a highly available infrastructure. One example comes from the Pure Storage FlashArray with Purity ActiveCluster™. Its active-active synchronous replication provides a transparent, automatic, and non-disruptive failover between sites, as well.

Another example is Pure Storage FlashBlade, Unified Fast File and Object platform. FlashBlade's scale-out metadata architecture can handle tens of billions of files and objects with maximum performance and rich data services. Purity//FB supports cloud mobility with object replication and disaster recovery with file replication.

This kind of availability and performance is a good option to consider for the business continuity component of a data protection strategy.

Synchronous and Asynchronous Replication with Pure Storage

Features like ActiveCluster and ActiveDR are included at no charge with the Purity operating environment.

Features like ActiveCluster and ActiveDR are free features of the Purity operating system. Customers who could not typically afford this can explore deploying synchronous and asynchronous replication because added costs apply only to hardware.

This can make deployment costs easier to justify to executive management.

Unified Fast File and Object

Many of Pure's Data Protection offerings take advantage of the company's Unified Fast File and Object (UFFO) platform. Even though the term may be new to some, UFFO storage is quickly becoming the only category of storage that can address modern digital transformation data requirements in many use cases, including data protection.

Organizations today need a storage solution that addresses modern data requirements, provides simplicity and multi-dimensional performance, and enables consolidation of key unstructured data workloads. These capabilities decisively eliminate storage silos, and provide profound returns on investment.

It's not just that the storage platform can store both file and object data, but it delivers outstanding performance for both. Its characteristics are best understood as follows:

- **Multi-Dimensional Performance** means very high throughput and IOPS with low latency to support multiple workloads simultaneously, including those with small or large files, sequential or random I/O access, batched or real-time jobs, and large numbers of files.
- **Intelligent Architecture** means that the storage system is built from the ground up to truly leverage the performance and efficiencies of flash storage. It is also simple to deploy, manage, and upgrade without requiring constant tuning. A modern storage solution must be simple enough so that its operation doesn't overwhelm storage admins. In fact, it relieves them of the mundane tasks involved in managing networking complexities when deploying the system, volumes, cluster pairs, aggregates, and flash caches or configuring replication.
- **Cloud-Ready** refers to its cloud-like agility, flexibility, and consumption choices with on-premises management and control.
- **Always Available** describes the capability of going beyond traditional platform resiliency. Maintenance operations, software upgrades, and capacity expansions are completed without disruption. The software design makes it possible for Pure Storage solutions to deliver high availability over multiple years and upgrade scenarios.
- **Dynamic Scalability** refers to the ability to seamlessly scale not only capacity but also performance, metadata, number of files and objects, and more.
- **Multi-Protocol Support** means that a single platform provides native file and native object protocol support without compromising performance or any functionality.

The Value of Rapid Recovery

Aside from a complete facility disaster or outright failure, one common issue that arises in most day-to-day operations is restoration of a corrupt, lost, or otherwise damaged file, directory, volume, or virtual machine.

When a business interruption hits your organization, the last thing business leaders want to hear is, "We're working on it." What they want to know is when business as usual can go on. That's why recovery is such a critical component of your data protection strategy and is what success hinges on. Remember: Backup is critical, but the ability to quickly restore will define your success. Tech-speak no longer cuts it in the current age—talking in terms of availability and "business as usual" presents a business-centric approach that puts you in the good graces of the powers-that-be.

Keeping availability high, however, means keeping your infrastructure humming along. If you want to eliminate bottlenecks associated with traditional purpose-built backup appliances (PBBAs) using spinning disk, or if you're still using tape as your primary backup destination, an all-flash solution is essential to bring your operations into the modern era.

And Pure is a leader in this space. The company first introduced FlashBlade in 2016, and since then has added significant features that provide compelling reasons to buy it when looking to upgrade your data protection. Note that there is no silver bullet when it comes to backup and recovery: It's hard to architect a fast and resilient infrastructure. Going with FlashBlade, however, is a good way to start when modernizing a data protection strategy.

Why? FlashBlade requires no changes to your existing data protection software or the standard IT processes. Its flexibility allows your IT teams to offer a wide range of recovery options and multiple tiers of service.

Pure Storage systems work with leading data protection software solutions, your IT teams can continue to service mission-critical recovery and compliance requirements. They can continue to protect their most essential data using preferred solution providers, including Commvault, Veeam, Cohesity, Rubrik and Veritas.

Where SafeMode Snapshots Come into Play

Pure Storage introduced SafeMode™ Snapshots in 2019. This technology comes as a built-in feature on all Pure Storage systems, including FlashBlade and FlashArray. SafeMode snapshots create read-only snapshots of backup data and associated metadata catalogs after a full backup is performed. Organizations can recover data directly from these snapshots to guard against ransomware or insider attacks (think: rogue administrator), and against backups and key data assets. SafeMode Snapshots confer the following advantages:

- **Enhanced protection.** SafeMode-secured snapshots cannot be deleted, modified, or encrypted. In fact, only authorized employees of your organization working with Pure Technical Support can configure the feature, modify governing policy, or delete such snapshots manually.
- **Backup integration.** SafeMode uses the same snapshot process irrespective of the backup solution or native utility used to manage data protection processes.
- **Flexibility.** Admins can set their own snapshot cadence and deletion schedules for themselves, as part of flash system setup and maintenance.
- **Rapid Restore.** The underlying hardware delivers a massively parallel architecture. Its elastic performance scales with data to speed both backup and recovery.

Immutable snapshots such as SafeMode snapshots provide tremendous reassurance. Because they cannot be deleted, modified, or encrypted, they're essentially immune to ransomware attacks. Policy changes for making, storing, and deleting SafeMode snapshots occur only with vendor participation and permission, so an attacker's typical "next move" when stymied—namely, to use privileged access to reset permissions, then wreak havoc—is blocked in advance.

Becoming Cloud-Ready

If hybrid or multi-cloud isn't in your strategy today, it almost certainly will be in the future. Therefore, choosing a cloud-ready solution is just smart planning. Cloud storage is already in use today for a variety of purposes that include backup, DR, and long-term retention. If such capabilities aren't already in your portfolio, you should be considering them soon.

As you consider your storage needs, and the options available to help meet them, it's crucial to look for APIs and services that permit easy integration of storage as a standard service layer. These should be available in any and all of the clouds, both public and private, that you use, and should likewise be available in your on-premises applications and services. Containerized applications and services built within the Kubernetes ecosystem can deliver such capabilities both easily and readily.

SIMPLIFYING CLOUD DATA PROTECTION

Purity CloudSnap skips complexities and challenges inherent in typical cloud data protection solutions. Built right into Purity OS (v5.2 and higher) and its FlashArrays, CloudSnap provides intelligent, efficient data transfer into and out of the cloud. It offers rapid on-premises recovery so organizations can honor stringent SLAs and compliance requirements. See the "CloudSnap for AWS: Effortless Cloud Data Protection" case study for a walk-through based around Amazon Web Services.

Maximizing Data Agility

To begin with, it's interesting to consider what the term "agile data" really means. Let's start by considering the state of protected data in the past. Most of the time, it sat in a proprietary format on some cold device (tape or some other removable media). Thus, the most an organization could do with such protected data was report on files stored, creation and modification dates, and so on. But when that protected data reached maturation, its life and usefulness expired. If the media couldn't be reclaimed and re-used, it was best destroyed for security's sake.

Today, that's no longer the case because data is used for much more than recovery. Data agility is a way of extracting more value from backup data by using it in multiple ways, rather than letting it rot

away on a hard drive under a mountain. For example, that protected data could be used to create a virtual lab for DevOps. That's a better use of that data, isn't it?

In the same vein, a flash-based storage strategy breaks down silos within backup data. Thus, it permits other workloads, ranging from analytics to application testing and development, to use that data when it isn't needed for business recovery. This flexibility is significantly different from purpose-built backup appliances (PBBAs), designed for and dedicated exclusively to backup and recovery.

Doing More with Your Data

The most interesting outcome from exercising data agility is that you will find more and better ways to use your data, as you and your staff become more familiar with what it can tell you about your business. Backup/ recovery data becomes more accessible and usable when it can be accessed more easily and quickly. This in turn leads to more frequent and varied use of that data as analysts, researchers, and data scientists begin to understand what treasures are now at their disposal.

Organizations often start out with one or two pilot programs, but soon find themselves using the data for more complex and sophisticated purposes, particularly for training or improving AI- or ML-based systems and applications. Ultimately, this means organizations can take advantage of what they know, and what they learn, to do more for the business, and to be more efficient, productive, and innovative. It's a stealth benefit for modernization.

If this is appealing, consider that FlashBlade can become your data hub, serving DevOps needs and more, including as a target for analytics, AI, and ML, for training data, and for historical data analysis. And all this happens without an impact on production. For security, Pure provides data-at-rest encryption for all arrays to add an extra layer of protection against the bad guys. It provides more peace of mind for the business.

Data agility makes a world of difference to a business-centric approach. There's much more value to be gained by using assets than from allowing them to quietly expire and then get retired. It can also help IT by eliminating multiple silos, complexities, and additional cost burdens that would otherwise eat into annual IT budgets.

Why Choose Pure Storage?

Pure Storage has been named a Leader by Gartner for Primary Storage (10 consecutive years) and Distributed File Systems and Object Storage (three consecutive years).

Pure Storage believes a sustainable tech infrastructure is necessary to address climate change. Pure's ESG program (Environmental, Social, and Governance) does this by building products that use less power and space, and reduce e-waste—consuming between 2 to 5 times less power than SSDs, and 5 to 10 times less than hard disks.

Thus, Pure Storage has the systems, software, skills, and knowledge to help enterprises implement the best and most cost-effective storage for cloud-native containerized apps and services, especially those within the Kubernetes ecosystem. In addition, the company's partnerships with Cisco and NVIDIA provide it with added synergies that organizations may find useful.).

Appreciating ESG

As the acronym indicates, ESG covers a broad range of topics and concerns. Pure Storage takes these things seriously, and provides voluntary reporting that complies with the following standards and regimes:

- **Global Reporting Initiative (GRI) Universal Standards**
- **Sustainability Accounting Standards Board (SASB) Hardware Industry Standards and Software & IT Services Standards**
- **UN Sustainable Develop Goals (UN SDGs) to adopt best practices while meeting stakeholder expectations**

Pure Storage also monitors the ESG regulatory environment and reporting requirements to make sure it recognizes new and evolving standards, requirements, and recommendations. For more info, see the Pure Storage Annual ESG Report for 2023.

According to this study from the World Economic Forum, digital electronics put down a sizable carbon footprint. Across all categories, such electronics account for between 4% and 5% of all carbon emissions. Other studies say data centers account for 1% to 2% of global power production, of which data storage accounts for 20% to 25% of that consumption. Those numbers are going no place but up.

Today, most data stored in data centers—over 80%—sits on conventional hard disks. Pure Storage flash-based systems consume 50% to 80% less power (2x to 5x less) than SSD-based systems. They consume 80% to 90% less power (5x to 10x less) than hard disks. Thus, replacing hard disks in data centers with Pure Storage flash-based systems can lower overall data center power usage by up to 20%.

Similar savings on data center cubic volume (space) and e-waste are also on, along with lower labor costs and increased reliability. This is a big step forward from a sustainability standpoint.

Pure Storage looks at data storage and management as a key enabling technology for business and innovation, not a commodity. That's why it provides a single consolidated, consistent, and well-orchestrated data storage platform to meet the full range of customers' needs. Here, sustainability, efficiency, reliability, and endurance all play important roles, because buyers are more actively seeking solutions that consume less power and space, produce less e-waste, and enjoy better durability and reliability. As companies and organizations around the world undertake their own sustainability journeys, Pure Storage intends to help them find success.

Doing Good by Choosing Well

From an ESG standpoint, Pure Storage supports those who might prefer to choose storage technologies based on energy and floor space considerations. For example, Meta's own analysis showed Pure Storage solutions (for a 100PB installation) had an 80% lower TCO than other alternatives. Savings come primarily from storage infrastructure efficiency that requires significantly less power and floor space to deploy.

CHAPTER 5

Real Customers, Real Cases

IN THIS CHAPTER:

- Silicon Labs
- Meta Platforms
- Making the most of flash storage

As evidenced by an audited NPS score of 82, Pure Storage prides itself on being a customer-first organization. The driving force is to deliver to its customers a better storage platform along with an enhanced, more modern experience for all those who work with and use its products, software, and services. In the sections that follow, you'll have a chance to read some customer stories and to learn more about best use cases where Pure Storage technology is particularly apt and helpful.

Silicon Labs Automates E-Designs

Founded in Austin, Texas, in 1996 Silicon Laboratories Inc.—Silicon Labs—is a technology organization that seeks to create and disseminate secure, intelligent wireless technology to make the world a more and better connected place. It designs and builds semiconductors and other silicon-based devices. The company also offers an integrated

hardware and software platform that includes a widely used developer-friendly toolbox aimed at electronics design engineers and manufacturers, especially for the Internet of Things (IoT) global infrastructure.

Silicon Labs uses an electronic design automation (EDA) process to create new chips and devices. It generates millions of small files and folders for each such project. With a heavy emphasis on metadata (data that describes files, objects, and other data: literally "data about data") EDA workloads pose data throughput challenges for conventional storage systems. Silicon Labs put Pure Storage FlashBlade storage to work to provide fast, efficient data storage for EDA and other workloads.

Indeed, Silicon Labs needed the right storage to speed up its design process, which involve huge volumes of unstructured data. It wanted to run its EDA workloads with its other workloads on a single system to keep things simple. Given its many locations, some lacking IT staff entirely, Silicon Labs sought something easy to manage, with great support from a committed and capable vendor. The company picked Pure Storage.

Using a FlashBlade-based solution, Silicon Labs found itself able to do more, and do it more quickly. It also reported that the Pure Storage systems were easy to deploy and manage, even when on-site support wasn't available. Best of all the company was able to consolidate all its workloads onto a single platform, including EDA, with improved performance and productivity. Among the tangible benefits realized were:

- **40x faster for some functions as compared to legacy storage.**
- **System scaled to handle business design workloads.**
- **Footprint reduced by 85%, to simplify and contract data center operations.**

Silicon Labs' Jud Baron, R&D infrastructure architect, summed things up nicely when he said: "With Pure Storage FlashBlade, we consolidated our massive EDA workload with our many business and manufacturing workloads while cutting cost per terabyte by 50%, and reducing rack space by 85%." That's a big set of wins that other businesses should be able to capitalize on as well.

Meta Powers AI's Future

By now, readers should recognize Meta Platforms as a major global social media and networking presence. If not, think of its best-known subsidiary, Facebook, and the pieces should fall into place quickly. Facebook was founded in 2004 at Harvard College in Cambridge, Massachusetts. Today, Facebook serves over 3 billion monthly active users, of which two-thirds log in daily. Like many other major global software companies, Meta is making major forays into artificial intelligence. That's where Pure Storage comes into this picture.

Meta's AI Research SuperCluster (AIRSC) is a world-class AI supercomputer. It's designed to train next-generation AI models on petabytes of data at great speed and great scale. Using Pure Storage FlashArray and FlashBlade, Meta's SuperCluster can accommodate its GPU and storage needs while keeping operational costs under control. Managing massive AI workloads requires a system that can handle petabytes of data at supercomputer speeds. Equally important, however, Meta wanted to minimize the power needed for storage so that it could use more power-hungry, high-speed GPUs.

When Meta put FlashBlade and FlashArray to work, the AIRSC could support next-generation AI research and modeling trained on petabytes of data. It was also able to unify data delivery to scale across its most data-intensive jobs. Ultimately, this work serves Meta in learning how to deliver more content its users wish to see. Using the Pure Storage systems, Meta designed and built this unique AI research

cluster in under 24 months. It learned that flash storage required less maintenance than disk, and enjoyed lower operational costs as a result. And, as planned, lower power consumption for flash-based storage let Meta funnel more power to its GPUs, and enjoy faster completion time for workloads.

GRAPHICAL PROCESSING UNITS (GPUS)

GPUs are high-powered processing platforms originally designed to calculate, render, and image complex 2D and 3D graphics data. Today, their most important (and expensive) role is to provide massively parallel, special-purpose number- and data-crunching services to handle complex workloads for artificial intelligence and machine learning purposes. The best-known AI GPUs come from companies like NVIDIA, AMD, Google, Intel, and others. A single high-end AI-capable GPU can exceed US$70,000 on today's market (e.g. NVIDIA HD100 80GB HBM2E).

As Vivek Pai, the AIRSC storage lead at Meta relates: "We contacted a number of storage vendors of both disk and flash to evaluate their highest performance, highest density offerings. From a combination of performance and power and cost, we ended up selecting Pure Storage." That's a winning endorsement that stresses Pure Storage systems' simplicity, power efficiency, and speed.

Make the Most of Flash Storage

This Gorilla Guide has covered a broad range of topics, including the state of the storage market, the challenges that face organizations seeking to modernize, the impact of ransomware on the rise, and the insight, capability, and solutions that Pure Storage can bring to bear for organizations seeking to build modern, flexible, and adaptable IT infrastructures.

One message should be crystal clear from this screed—that old ways and tools for doing backup and recovery can't keep up with today's challenging environments and their associated RPOs and RTOs. As the cloud continues to change everything, there's more data to acquire, store, and protect than ever before, and that data lives everywhere from the edge to the core to the cloud.

As data remains the crown jewel of organizations everywhere, it's increasingly essential to make sure it's protected and resilient, and that it can be put to new, agile uses in all kinds of interesting and informative ways. By keeping data safe, secure, and easy to recover, organizations can respond effectively to business interruptions and even disasters, whether man-made (often by ransomware) or resulting from nature's wrath. Such organizations are the ones that will survive in the face of inevitable outages and interruptions. In fact, such organizations may not even notice or care when the power fails or an intruder gets in because they'll be ready for whatever happens next!

Is your organization ready? If you're unsure, please check out what Pure Storage can do to help. For more information, visit Pure's Modern Data Protection Solutions or send an email to info@purestorage.com.

ABOUT PURE STORAGE

Pure Storage® all-flash storage solutions help mitigate ransomware at each phase of the attack lifecycle by enabling organizations to build a data resilient architecture and restore systems faster than any other disk and hybrid solution.

Pure data-protection solutions work before an attack by providing the ability to build a high-speed security analytics architecture that helps identify indicators of compromise before an attack is launched. During an attack, Pure SafeMode Snapshots ensure attackers can't alter or destroy your backup data—even if admin credentials are compromised. Lastly, after an attack, Pure provides the fastest recovery speeds available—up to petabytes per day at scale so you can restore systems in hours, rather than weeks. And our Ransomware Recovery SLA guarantees you'll have a clean system for restore should you be hit by a ransomware attack.

Pure data-protection safeguards are included at no extra charge, are easy to deploy and work with your existing backup software, so defending against attacks is simple and fast.

Our commitment to providing true storage as-a-service gives customers the agility to meet changing data needs at speed and scale, whether they are deploying traditional workloads, modern applications, containers, or more. Pure believes it can make a significant impact in reducing data center emissions worldwide through its environmental sustainability efforts, including designing products and solutions that enable customers to reduce their carbon and energy footprint. And with the highest Net Promoter Score in the industry, Pure's ever-expanding list of customers are among the happiest in the world.

LEARN MORE >

ABOUT ACTUALTECH MEDIA

ActualTech Media, a Future company, is a B2B tech marketing company that connects enterprise IT vendors with IT buyers through innovative lead generation programs and compelling custom content services.

ActualTech Media's team speaks to the enterprise IT audience because we've been the enterprise IT audience.

Our leadership team is stacked with former CIOs, IT managers, architects, subject matter experts and marketing professionals that help our clients spend less time explaining what their technology does and more time creating strategies that drive results.

If you're an IT marketer and you'd like your own custom Gorilla Guide® title for your company, please visit actualtechmedia.com.