PURE STORAGE PRESENTS

THE GORILLA GUIDE TO...®

SECOND EDITION

Rapid Restores with Flash and Cloud

A How-To Guide for Modern Enterprise Data Protection

David A. Chapa

INSIDE THE GUIDE:

- The changing face of data protection
- The real value of cloud-enabled backup and restore
- Why flash storage has become the new standard

HELPING YOU NAVIGATE THE TECHNOLOGY JUNGLE!

ActualTech Media
www.actualtechmedia.com

In Partnership With

PURESTORAGE®

THE GORILLA GUIDE TO...

Rapid Restores with Flash and Cloud

AUTHOR
David A. Chapa

EDITOR
Keith Ward, ActualTech Media

LAYOUT AND DESIGN
Olivia Thomson, ActualTech Media

Copyright © 2019 by ActualTech Media

All rights reserved. This book or any portion thereof may not be reproduced or used in any manner whatsoever without the express written permission of the publisher except for the use of brief quotations in a book review.
Printed in the United States of America.

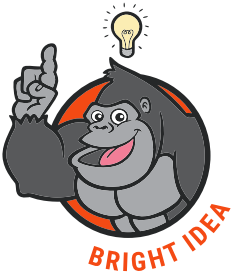
ACTUALTECH MEDIA
Okatie Village Ste 103-157
Bluffton, SC 29909
www.actualtechmedia.com

ENTERING THE JUNGLE

CALLOUTS USED IN THIS BOOK

The Gorilla is the professorial sort that enjoys helping people learn. In the School House callout, you'll gain insight into topics that may be outside the main subject but are still important.

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



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

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

Discusses items of strategic interest to business leaders.

ICONS USED IN THIS BOOK

DEFINITION

Defines a word, phrase, or concept.

KNOWLEDGE CHECK

Tests your knowledge of what you've read.

PAY ATTENTION

We want to make sure you see this!

GPS

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

WATCH OUT!

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

INTRODUCTION

Welcome to this Gorilla Guide covering enterprise data protection, backup, and recovery. The purpose of this guide is to provide insights and guidance to IT planners, storage architects, compliance officers, auditors, and backup administrators who are seeking to ensure and improve the data protection and recovery strategies currently being used to safeguard your organization's most irreplaceable asset: your data.

It's About More than Just Backup and Recovery

While the concept of backup and recovery is not new, organizational requirements for fast data restore in the wake of an unplanned interruption event have become more stringent than ever. As more organizations build their business and competitive edge on data, rapid access to data for disaster recovery and data reuse is 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, enabled by technologies such as cloud and flash storage.

Data protection has always been an imperative for organizations, but not always the first thought during budget cycles. One of the reasons is that the technology IT used 20, and even 10, years ago simply wasn't designed for the rapid recovery necessary today. Companies were less inclined to invest in the latest and greatest tape library when it provided no real improvement in recovery performance.

That's changed dramatically in the last few years, as companies are looking to do more with this backup data. It's not enough for that data to be just backed up—it can actually provide an edge in the marketplace.

The advancement of technologies like cloud compute and cloud storage, flash storage, and storage efficiency algorithms has significantly altered the solution opportunities IT may provide to its own internal customers by leveraging backup data.

While having more solutions to offer internal customers is an enormous advantage, the other reason data protection has been experiencing an elevated level of attention is the pressures the C-suite has been feeling as a result of the growing risks of ransomware, breaches, hacks, and extended outages.

The threats aren't only to data and IT operations, either—today, it's equally important to protect and guard the company brand and brand loyalty of its consumers. The risk of PR fallout is serious, and reputational damage—as companies who have experienced these types of attacks can attest—can be significant and long-lasting.

Mitigating risks, avoiding extended outages, and improving time to market are all things the business is looking for and more. We sit at this crossroads, with IT and business leaders converging to deliver the best solutions possible to meet the needs of the core business.

One important feature of this new era as it relates to backup and restore is the burgeoning use of flash storage, and the huge advantages it brings in terms of speed and latency.

The advantages of architecting a modern approach using flash are three-fold:

1. All-flash performance can help firms that are struggling to meet data recovery standards as defined in their IT Service-Level Agreements (SLAs). It provides a means to bring data and system recovery and restore up to similar performance with data backup, which has been the focus of most improvements in data protection over the past 20 years.

2. Modern approaches must consider cloud-centered technologies. Solutions that enable the portability of backup data and streamline

the placement of data into cloud storage repositories have two key benefits. First, they provide required separation between original and backup copies of data to ensure resilience in the face of facility and regional disaster events. They also provide a means to capture cloud economics and flexibility to reduce the cost and complexity of backup infrastructure.

3. The strategy breaks down the silos of backup data and permits its use by other workloads, ranging from analytics to application testing and development, when it isn't needed for business recovery. This is clearly differentiated from purpose-built backup appliances (PBBAs), for example, which were designed for and are dedicated to backup and recovery. By enabling multiple uses of backup data, the value of backup extends beyond the traditional focus on risk reduction and into the other domains of business value including cost containment and improved productivity. Business management prefers an IT initiative or strategy that delivers value in all three domains, so your modern data protection strategy is more likely to get the nod from those who hold the budgetary purse strings.

This book is intended to help the reader understand how a modern data protection architecture, and specifically solutions from Pure Storage, can help you realize your objectives. (Of course, those objectives vary based on the reader.)

A modern data protection architecture is built to restore rapidly and leverages flash storage as a key enabler, as it's significantly faster at delivering randomized read operations than hard disk storage systems, which tend to be optimized for sequential operations.

"Time to data" is the ultimate measure of disaster recovery strategy success. That means reducing the time required to restore data (one of the three basic components of time to data) by using technologies such as flash storage, augmented by data reduction (compression and deduplication) and having the ability to on-ramp data to the cloud

when appropriate. Significant restore time improvements have been demonstrated by users implementing a flash-based modern data protection strategy.

To the IT architect, of course, improved restore speed isn't the only advantage being sought from a data protection solution. From an overall IT strategy perspective, it's also about bending the storage capacity demand curve (how much space will backup data occupy on an increasingly scarce resource); enabling greater data "agility" (or reuse by other workloads) without creating yet another data silo; cloud strategies; and setting the stage for new processes that can extract the intrinsic value of data.

As stated earlier, compliance officers also have a stake in an improved backup and restore data protection process. They're under unprecedented pressure to work with IT to restore, search, and act on data in accordance with new privacy regulations, such as the General Data Protection Regulation (GDPR) enacted in 2018 by the European Union.

Many compliance-oriented users see the improvement of backup processes as a linchpin of building data stewardship excellence and ensuring legal/regulatory compliance.

A PAIN IN THE WALLET

Busting General Data Protection Regulations will hurt: Violators of GDPR may be fined up to €20 million, or up to 4% of the annual worldwide turnover of the preceding financial year, whichever is greater. Several large, multinational companies, including British Airways and Marriott International, have already been hit with fines.

Data protection is a central component of risk reduction strategy, but there's often a tendency among senior management to regard it simply as additional "insurance" and, as such, a low priority item in the corporate IT budget.

By integrating data security and regulatory compliance into the data protection/backup and restore process, the value of the strategy becomes easier to communicate and may increase the budgetary priority of the process.

To the backup administrator, there's always value in any effort to improve the backup process. Admins want backup to become as automated as possible. They want to consolidate the many proprietary tools and processes that have been developed and deployed over time to meet the requirements of different databases, applications, operating systems, file/object systems, or hardware platforms.

They're seeking solutions today that address the burgeoning requirements to protect data that's increasingly being hosted not only in on-premises storage infrastructure, but also in multiple public cloud services. Currently, orchestrating backups in a multi-cloud setting is a challenge that evokes dread among backup administrators.

A modern data protection strategy helps eliminate data silos in the backup and restore workflow, automate the process, remove some of the complexities in hardware introduced by other solutions, and deliver a truly scalable and cloud-ready fix to many issues that have plagued backup and restore for years.

This guide will provide an overview of solutions that may be deployed to help cure the doubts associated with your current backup and recovery strategy. In addition, this guide will offer use cases intended to survey and underscore the possibilities for leveraging Pure Storage FlashBlade to improve your data protection strategy and accelerate your journey to modernization. A good place to start is with a concise summary of the situation in most organizations today. That's the focus of Chapter 1.

CHAPTER 1

Disk-to-Disk-to-Tape: A Legacy System Past Its Prime

How We Got Here

Early on, the choice of magnetic tape as the exclusive medium for backup data was driven by the high cost of disk media, as well as by the intrinsic value of tape itself, including its portability (so that backup data could be transported to a safe offsite location) and the "air gap" that it provided for backup data (backup data was offline and out of reach for hackers and malware).

The disk-to-tape strategy for data protection persisted until the late 1990s, when distributed computing and lower-cost/higher-capacity disk drives appeared.

Distributed computing challenged tape-based backup in several ways. For one, sharing a tape system required connectivity between servers and their storage and the tape system, which ultimately became a key driver of storage area networks.

Adding to the complexity of physical cabling was the need to schedule (1) the use of the shared resource between multiple servers, and (2) the processing workloads of servers and networks to accommodate backup processing and backup data traffic. As server farms moved to 24x7 processing schedules, time grew scarce for tape backups, and new solutions were sought.

The desire to extract value from all data, including backup data, has ramifications for the way that we create and manage backups and further invalidates traditional backup-to-silo methodologies. The first solution was to write backup data to a cache of disks managed by its own server that emulated tape drives: a hybrid system referred to as a "virtual tape library" or VTL. VTLs, which leveraged a cache of disk drives to write data backup images, proliferated in the 1990s, offering a range of features and functions.

D2D2T

Over time, a strategy called Disk-to-Disk-to-Tape (D2D2T) became commonplace. The first "disk" referred to the storage hosting production data, the second disk comprised the cache in the VTL, and the tape referred to the tape library (**Figure 1**).

In 2009 a commercial photographer, Peter Krogh, published a book on digital asset management for photographers. In the book he outlined his methods for backup of his digital assets and the concept of 3-2-1 was invented—or, rather, formally published. While Krogh certainly wasn't the first one to use this method, he was the first to coin the term "3-2-1."

As Krogh explained in his book, the perfect backup was three copies of data, stored on two different sets of media, with one copy stored at an offsite location. This model has been adapted by IT, and works well. To create the perfect backup based on 3-2-1 means having backup data stored on two different sets of media (disk on the VTL and tape in the library), with one copy (tape) stored offsite as a hedge against disaster.

At the beginning of the new millennium, the D2D2T strategy underwent some changes. The advent of deduplication technology led to the implementation of data reduction to consolidate frequent backups and to conserve disk space.

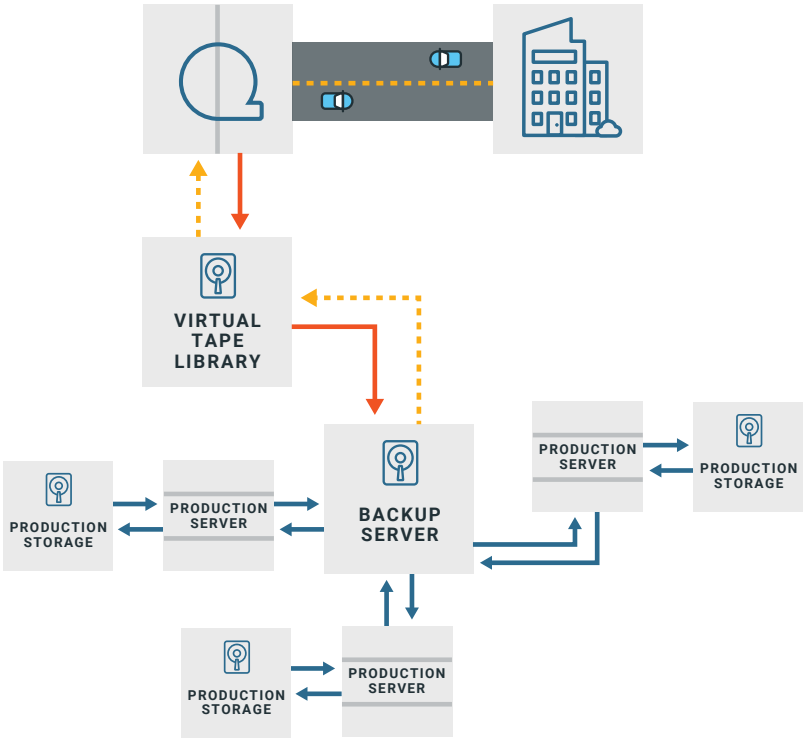


Figure 1: Disk-to-disk-to-tape (D2D2T)

Some firms saw this capability as an invitation to simplify backup by eliminating tape altogether, in some cases by leveraging a mirrored backup appliance located in an off-site, outsourced, data center, a service that was coming into vogue that, today, is called a cloud. Thus, D2D2C was born (**Figure 2**).

Cloud services were initially seen as a location to obtain "cheap-and-deep" data storage services, ideally suited to backup data. Many smaller and midsize firms simplified their 3-2-1 backups by eliminating on-site tape altogether and moving to public or private cloud storage repositories, while certain larger enterprises, such as those working in financial services, continued to leverage tape-based backup for on-premises.

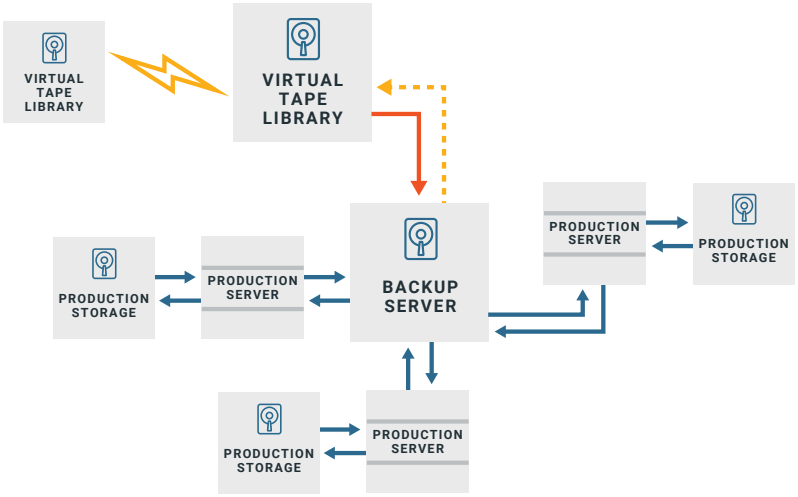


Figure 2: D2D2T morphs into VTL/backup server mirroring, eventually leveraging public clouds

The story hasn't ended there. Several changes are occurring today in the backup model. For one, backup devices are beginning to shed hard disk drive technology in favor of solid-state storage, which permits them to gain greater storage density with lower power consumption and much improved restore performance.

This reflects a change in priorities from backup speed, which was the focus of most development in the prior 20 years of purpose-built back-up appliances and protection process improvements, to restore speed.

With disk, backup speed—a write operation—was optimized; flash storage enabled significant improvements in restore speed, a function of its superior performance with random read operations. Specifically, with flash-based systems, you don't incur the read penalty when restoring from deduplicated backups on spinning disks, such as those within PBBAs.

Flash storage also offers some advantages in terms of data mirroring speed when compared to disk drive mirrors. This has also expedited

the replacement of Disk to Disk to Tape or Cloud as an overall data protection strategy, while uncovering new use cases for tape in highly specialized scenarios.

A More Modern Approach to Backup and Disaster Recovery

So, the many changes that have occurred in backup strategy have been driven by a combination of technological change, data growth, increasingly constrained operational windows, tighter service level requirements, and budgetary cost-cutting imperatives. The current state of backup remains in flux, which is why so many IT organizations are focused on modernizing their data protection approach and strategy as part of their overall digital transformation initiative.

That modernization starts with flash-based systems, which don't incur the read penalty when restoring from deduplicated backups on spinning disk, such as purpose built backup appliances.

Cloud and all-flash storage remain key guide rails of future backup models. For one thing, companies are using cloud services today in a different way than they were only a few years ago.

In some cases, cloud resources are being used as an adjunct to on-premises infrastructure or to provide specific services such as archival storage in a manner believed to be less expensive than do-it-yourself alternatives. In other cases, companies are instantiating entire workloads in the cloud and leveraging connectivity between geographically dispersed cloud facilities to provide failover and recovery services.

The latest analyses show some firms re-hosting workloads originally placed in public clouds on cloud-based infrastructures created in more private co-location and hosting facilities or in their own data centers. This trend, which is driven by cost considerations and other factors, might determine where and how backup data is stored and used.

It might be appropriate, for example, for data to be mirrored between primary and secondary storage so that both data and applications can simply failover from one environment to another if a disruption occurs.

The prerequisite of such an approach, however, is that the two storage devices (and their cloud hosts) be sufficiently separated by distance to avoid being consumed by the same disaster event.

With greater distance (more than about 50 miles) comes greater latency and "data deltas": differences between the original data and the backup copy. Synchronous replication must be replaced by asynchronous, leading to the potential loss of some data if a failover becomes necessary. The impact of data deltas will need to be considered in planning such failover strategies.

The advent of flash storage will also have a continuing impact on back-up and, especially, restore. As with most read-intensive workloads, the substitution of flash storage media for other media may be preferred for reasons of raw performance and throughput. With flash mirroring and replication, data can be replicated almost at the same time as it is recorded.

We'll look at the Pure solution in greater detail in subsequent chapters.

CHAPTER 2

Key Trends Fueling the Fire

A quick survey of trade press journals and the agendas of technology conferences confirms that the data protection landscape is under tremendous pressure to change. The reasons are several, and include the following:

Exponential Data Growth: According to industry analysts, data is growing at unprecedented rates. IDC projects that the total volume of new data that will be created by applications in 2020 will approach 60ZB. By 2024, they foresee data growing to approximately 160ZB, courtesy of trends in mobile commerce, Internet of Things (IoT), and other factors.

IBM Research published findings more than five years ago stating that humans are creating 2.1 quintillion bytes of data—or 2.1EB—every single day. And that's older information, before the explosion of data-creating monsters like IoT. This only adds to the dire need for more storage devices, and better, more efficient ways to protect and secure that data.

To illustrate this more simply, just take a look at the size of your smartphone backup and compare it to the backup of the phone you used in 2003 or 2004. Staggering differences indeed—and the same is true (only many orders of magnitude greater in size) in corporate data centers. Solutions must be found to host this data more cost effectively, a concern that's driving interest in using the cloud—especially for storing backup data.

Beyond Petabytes and Exabytes

A zettabyte is a unit of measurement equivalent to 1,000 exabytes or 1 million petabytes or 1 billion terabytes. Industry analysts began projecting zettabyte-sized data growth rates in the early 2000s. Current estimates of data growth from IDC are 60ZB by 2020 and 163ZB by 2024. While not all of this data will be created in your shop, it is likely that the data that is being generated and stored is increasing in volume and that this is putting a strain on your storage infrastructure capacity, both for production data storage and data copies that are made for purposes of backup and restore.

In addition to dealing with data storage capacity demand, zettabyte-level data growth is also compromising the ability of traditional data protection processes to enable the restoration of data for use by production systems following an interruption event in a timely way.

Many companies are struggling to make their data restore processes keep pace with the speed of business and constant growth of the amount of data. Tolerance for downtime is decreasing, and the need to improve data availability is paramount. Tried-and-true backup techniques are being reconsidered, and fresh approaches are being sought.

Privacy Issues: Data security concerns and new data privacy regulations are also encouraging changes in backup and restore strategies. Threats from ransomware and other malware are causing the discipline of data security to become merged with the discipline of data protection and backup. So, the latter must change to add protection and recoverability in the face of security threats as well as natural or manmade data access interruptions.

Moreover, the regulatory landscape is imposing new requirements to ensure the privacy of data that is collected and stored by companies, including data stored in archives and backups. An example is the General Data Protection Regulation.

The regulation, which impacts both companies located in the European Union (EU) and firms worldwide that do business with EU citizens, includes requirements to protect data and data privacy through a combination of data protection processes, pseudonymization of private data, and restrictions on the movement of certain data outside of geographical boundaries.

These regulations impose new requirements on backup and restore processes, including the need to redact or anonymize certain private data when recorded in a backup data set and the ability to tag certain backup data for deletion or restricted copy. GDPR also necessitates closer tracking of backup data so that it can be discovered as part of a Subject Access Request (SAR) and deleted if the EU citizen to whom the data refers exercises their "right to be forgotten."

In companies that use their backup solution to serve as a kind of archive, compliance means providing fast search, recovery and possibly deletion of data from their backup repositories.

For many organizations, the common strategy of keeping every bit of data forever, particularly in backup images, may undermine aspirations for achieving regulatory compliance in light of the new privacy regulations.

GDPR-like regulations are gaining traction in other countries and regions today, including some state governments in the United States, and with them will come a need to refine backup and recovery processes.

Re-Use of Backup Data: Another trend that's beginning to shape the backup landscape is the quest in many companies to find ways to re-use backup data when it isn't needed to recover from an interruption event.

Rather than writing backup data to an isolated data "silo," many organizations are seeking to place backups in shared volumes of storage media where the data can be directly used or copied readily to support such activities as application testing and development, patch testing, reporting, or analytics.

These additional uses of backup data, provided that they do not compromise the recoverability processes for which backups are created, are viewed as a great value-add to data protection.

The ability to repurpose backup data sets helps to offset the cost of undertaking a data protection process to enable a recovery process that in the best of circumstances will never need to be used. Additionally, it can drive down the cost of storage by reducing the amount of space that must be allocated for storing multiple copies of the same data to satisfy the demands of data consumers.

The Rise of Artificial Intelligence (AI)/Machine Learning (ML): Changes in the backup landscape are also being driven by the rapid rise of AI or ML. This new class of workload requires tremendous amounts of data, thousands-fold more than the amount typically used in operational datasets, in order to create intelligence to discern trends and relationships that have value to a business.

The Exploding AI/ML Market

According to Datanami, the artificial intelligence (AI) and machine learning (ML) fields are set to take off. ML spending is predicted to increase at a compound annual growth rate (CAGR) of 25% to 44% by 2024, depending on the source.

Datanami quotes IDC figures for AI growth, showing a predicted 37% CAGR through 2022.

The desire to extract value from all data, including backup data, has ramifications for the way that we create and manage backups and further invalidates traditional backup-to-silo methodologies. It also introduces requirements for fast access to data that often cannot be delivered by legacy devices; flash object storage is gaining ground as a preferred platform for serving active AI/ML analytics workloads.

Among many other applications, AI/ML is being used by many technology vendors to facilitate better data management by automating the placement of data on various parts of storage infrastructure and by providing data with protection, preservation, and privacy services that best fit the data—all according to data lifecycle policies.

These are just a few of the trends that are compelling organizations to reconsider their existing backup strategies. But this isn't the first time that data protection processes have found themselves under pressure to change.

CHAPTER 3

The Evolving Role of Flash in Backup and Recovery

In the previous chapter, we talked about the increasing role of flash in storage infrastructure and, specifically, in backup and recovery applications. We noted that several advantages of flash storage are making its use in backup devices preferable to hard disk storage. These include:

- The performance of flash storage is significantly faster than hard disk storage, especially in the case of random reads. Faster reads translate into faster restores of backup data following an interruption event.
- The physical attributes of flash memory make media much more compact than hard disk drives: hence, flash storage offers greater density and capacity than hard disk in the same physical space. Used in conjunction with a backup server or as a backup/ restore appliance, flash storage should provide a more economical and more capacious target for backup data writes.
- The lack of motorized spindles and other moving parts in flash storage reduces both the energy demand and the heat output of flash storage compared to HDD arrays, producing greater cost-efficiency.
- Flash storage can leverage protocols that extend the performance advantages of flash media to inter and intra-system data copy and replication/mirroring.
- Flash storage media is also less prone to bit errors than HDD media by at least an order of magnitude.

Initially, flash arrays found themselves deployed in enterprise data centers as repositories for hosting the data of the most demanding and valuable production data workloads. Pure Storage is well known in their invention of a successful enterprise class storage all-flash array aimed at mission critical workloads—the Pure Storage FlashArray™ is designed around effortless administgration, efficient storage, and a unique consumption model (Evergreen).

However, the company's innovation didn't end there. Pure Storage also envisioned bringing the power of all-flash platform to unstructured

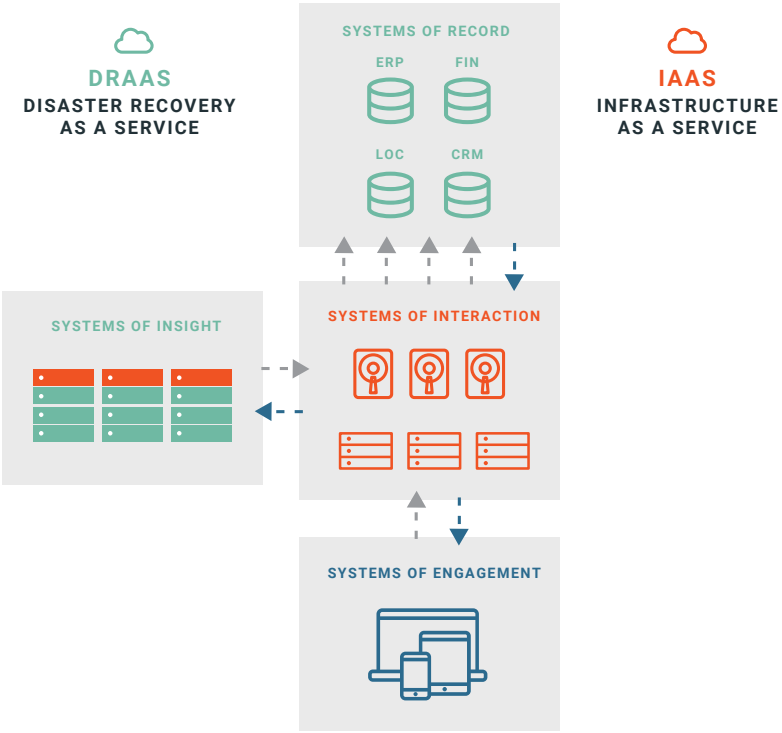


Figure 3: The hybrid cloud model: Corporate data center obtains select services from a cloud service provider (for example, Disaster Recovery as a Service or DRaaS) and additional compute, network, or storage resources on an as-needed basis from other clouds

file and object storage; as a result, the Pure Storage FlashBlade™ system was introduced to the market in 2016.

It's one of the industry's most advanced storage systems for unstructured data, and is central to Pure Storage's Data Hub strategy of helping customers unify their data on a single, optimized platform for modern workloads: from backup to analytics, test/dev, AI, and more.

The combination of FlashBlade's performance, scalability, and integration with public cloud storage sets the stage for a significant improvement in data protection, especially in the realm of data restore.

Pure's emphasis on front-end connectivity and partnerships with leading backup software vendors helps to make its solution a one-stop shop for data protection storage. Its support for a growing number of public and private cloud protocols and storage offerings enables it to scale effortlessly, as multi-cloud and hybrid cloud infrastructures are adopted in ever-increasing numbers (see **Figure 3**).

The eMoney Success Story

Moreover, this approach sets the stage for innovative approaches to data protection. Not surprisingly, it has been seized upon by companies such as eMoney Advisor and ServiceNow to improve their data availability strategies significantly.

eMoney Advisor, a cloud-based Software-as-a-Service (SaaS) platform delivering financial planning software tools to financial professionals, was acquired by Fidelity Investments in 2015, doubling its size. With more than 50,000 financial services professionals using its services to meet the needs of over two million clients, the company faced significant challenges including backup inefficiency.

The company's growth led to a data explosion and a commensurate slowdown in database backup speeds. Additionally, application software development work was limited by the ability to spin up database copies on legacy disk-based storage systems, with the latter creating

burdensome management complexity. The company's choice to solve the backup problem provided benefits in terms of the other challenges, too.

eMoney Advisor undertook a transformation of its legacy infrastructure, deploying FlashBlade products with the goal of ensuring 100% availability of software services and data to clients. Moving to the flash storage-enabled architecture proved its worth by delivering:

- Faster access to data and applications and a general improvement of the user experience for clients
- The ability to use snapshots to facilitate the testing and development of applications and databases
- A way to perform non-disruptive upgrades and capacity expansion without downtime
- Simplification of the management of applications and infrastructure so that IT staff time could be allocated to higher value projects

According to eMoney, the decision to invest in flash storage from Pure Storage has been a centerpiece of improved service levels generally. One spokesperson notes that "storage bottlenecks are non-existent," and clients have noticed and remarked about service improvement.

Behind the scenes, Pure Storage FlashBlade systems have made an important difference in mission-critical database backups; and more importantly, restores.

Using both third-party backup software and Pure snapshot technology, eMoney IT planners have several options for protecting data and for ensuring rapid restore and business recovery. In tests, database restore times came well within the window needed to meet Service-Level Agreements (SLAs).

Other features of FlashBlade have improved the automation and scripting of data movement and integration with VMware. Plus, richer

and more timely data analytics from improved data sharing are helping the company to create new revenue streams.

Although the acquisition of FlashBlade was premised originally on the improvements in database backup speeds, eMoney believes that the technology has also provided a means to achieve operational advantage.

The ServiceNow Solution

eMoney Advisor is not alone. ServiceNow is a juggernaut in SaaS space. SaaS customers tend to be very infrastructure savvy, because it directly impacts customer experience and operating costs. Thus, companies like ServiceNow need to ensure that they provide the most efficient infrastructure and processes, even in the face of massive volumes of active data.

Huge data growth presents many engineering challenges. For one, ServiceNow began to hit the limits of performance with traditional web scaling architectural models. It was deploying 2U servers with internal storage, with each backing up its database to the next. Despite excellent scaling standards, the company was filling up its public cloud data centers too rapidly.

Part of the problem was backup data, which was being generated at a rate of more than 490TB per hour. This process was needed to address the failure rates of the more than 30,000 hard disk drives used in more than 3,500 servers.

The combined disk storage repository produced more than 2,000 hard disk failures per year, and consumed a non-trivial amount of electrical power.

Further analyses revealed that a disproportionate amount of CPU cycles, network bandwidth, and storage capacity were being used to serve backup processes and backup data movement and storage. Plus, as databases grew, both backup and restore time frames lengthened—to the

point where planners determined that the company's business model was at risk.

They decided to change their architecture. Leaving behind server-centric scaling wasn't an easy process. Moving from internal disk and direct-attached storage to storage arrays involved high cost, new power requirements, and new management and administrative workloads.

SERVING SERVICENOW

ServiceNow is an innovative leader in cloud-based Software-as-a-Service (SaaS). After using Pure arrays to increase performance on critical database applications, ServiceNow Cloud Group chose FlashBlade to drive greater customer density per rack, which enables more flexible and dynamic infrastructure and helps increase profits.

Pure Storage FlashBlade appeared to be a better fit, especially given its physical dimensions and comparatively low power requirements. Products were brought in for testing: more specifically, to test their resilience. Once the engineers were satisfied with the product's durability, FlashBlade was deployed into production.

Eighteen months later, reporting at a Pure Storage user conference, ServiceNow explained its new architectural model, which delivered three times the storage density of the company's legacy infrastructure. Each rack comprised 30 1U servers with 1,000 processor cores and 1.5PB of effective flash storage capacity.

Moreover, server CPUs were no longer performing backup work, enabling their cores to be dedicated to improve database performance. The backup burden was shifted onto the FlashBlade systems, and the efficiencies and performance gains that accrued paid for the

transformation of the infrastructure from legacy servers with internal storage to servers and FlashBlade storage.

Uptime was the final success metric cited by ServiceNow. Spokespersons for the company hail Pure Storage for delivering the most non-disruptive support, expansion and upgrade services of any vendor with whom they have worked. This reliability, combined with the economics of Pure Storage at scale and fast networks, are game changers for SaaS providers.

The ServiceNow story combined with the eMoney case study demonstrate how flash storage technology is entering the disk-centric backup model to deliver significant improvements in durability, performance, and cost. This model, and its many advantages, further solidifies Pure as a key ingredient to enable strategic value when supporting the data demands of the core business. We'll look at this in the next chapter.

CHAPTER 4

Flash Forward to Pure Modern Data Protection

In the previous chapter, we discussed the emergence of flash, its use on-premises and in the cloud, and provided an overview of a Pure Storage data hub built with FlashBlade. Now we turn to modern data protection from a perspective that's often ignored, but should never be: a business-centric approach.

According to IDC, 90% of organizations lack confidence in their backup and recovery processes. Those numbers are staggering when considering the weight that business places on the shoulders of IT for its overall success. In fact, business imperatives now drive backup and recovery in a way they never have before.

What Is 'Business-Centric IT'?

"Business-centric" is a concept that always seems to get people talking, but what does it mean in practice? Although it may seem obvious that there's a focus on the business, that philosophy is slower to permeate some parts of the organization than others. When you create a business-centric approach, you're focusing on what's important to the business itself, not just what's important for IT success.

For example, if your company is in the hospitality industry, you probably depend significantly on online bookings, whether from your own website or partner websites. You also depend heavily on your point-of-sale system at the various properties under management for check-ins, guest billing, service orders, maintenance, and so on.

When those systems are inaccessible, down, or basically not available for customers to access, your business is in jeopardy. The best way to think about it isn't in terms of unstructured or structured data, but in terms of the key mission of the business: What is your company's mission? In other words, what is it that makes your organization successful?

An airline's chief mission, for instance, is safely moving passengers, crew, cargo, and planes from one destination to another. In order for that mission to be successful, what platforms, systems, applications, data, and so on is required? This information is then applied to a solution that will help meet those expectations.

All of this can be accomplished without using any deep technical jargon or terms, but in business terms. In the current era of modern business, IT and the stakeholders need to work much more closely to help the core business hit its business objectives.

The Old 'Backup and Recovery' Mindset

In the past, IT traditionally has been focused on recovery time objectives and recovery point objectives, known respectively as RTOs and RPOs (**Figure 4**); business, on the other hand, is focused on availability, business intelligence, expansion, and time to market. For years RTOs/RPOs have been the standard by which IT success was measured; often this measurement was an SLA between the business and IT.

There are two types of events that can occur within an IT ecosystem that affect SLAs. The first is a business interruption, and the second is a declaration of disaster.

A business interruption is any event that may cause production or productivity to be hindered or come to a halt. A typical disaster recovery (DR) plan outlines the process or procedure for declaration: You pull the trigger and it's all hands on deck to solve the problem. Part of that process or procedure is usually called "pre-determined time," or the

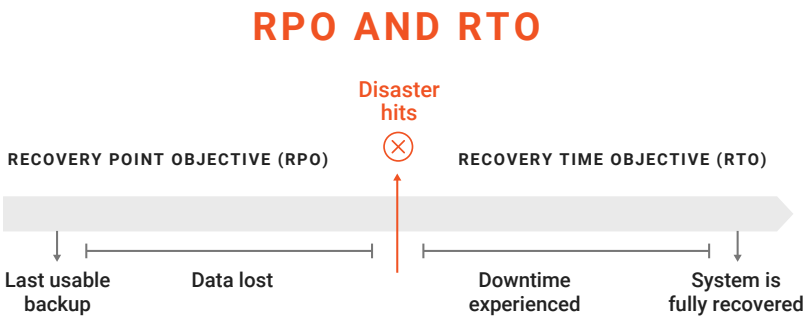


Figure 4: The differences between RTO and RPO

agreed to amount of time the business can tolerate downtime of a particular function or service. Beyond that time—when applications and data are no longer available to users and customers—is what turns a business interruption into a declaration of disaster.

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 terms. This is what the modern enterprise needs, and what previous solutions did not provide.

The pressures business exerts on IT is paving the way toward a new framework to meet these demands. It's a modern approach, not only in technology, but in practical application to meet the needs of the business.

Simply put, the chief reason data is protected is to be recovered in case of emergency or disaster. In IT terms, this is RTO; to the business, it may be a pre-determined time, agreed upon in an SLA; even more narrowly, it can simply be called "availability." The point here is that the realities of a business interruption mean different things to different groups within the organization, but the end point is always the same: How rapidly can we resume business as usual? The cost of not doing

business in the wake of a business interruption or a disaster can be staggering, which is driving the trends around data center modernization. That modernization starts with every organization's crown jewels: its data.

It's Not About 'Speeds and Feeds'

In a perfect world, there would be no need for restore, recovery, failover, or failback. As we don't live in that world, we always need to think about protection. When you look at solutions that will help modernize your company's approach to data protection (you should start replacing that phrase with "business protection"), it's important to first consider the desired outcomes in terms of recovery or resumption of activities.

Companies naturally want the fastest solution on the market, and a flash storage system will certainly provide the necessary speed—but you also need to consider the software driving it. Pure offers an optimal solution for modern data protection. Think of Pure as the high octane fuel needed to squeeze maximum performance from a race car: Yes, of course there are alternative fuels out there. But using a lower octane fuel will slow down your car. And the effect on your data center could be analogous, causing performance issues within the entire system.

As mentioned earlier, a solid-state storage solution increases backup performance (a write operation); but it's also significantly faster at delivering randomized read operations (restore) than hard disk storage. That "time to data" previously discussed is a crucial component when it comes to recovery.

To drive the point home: this isn't about speeds and feeds—instead, it's about the business, so let's look at the business-centric drivers to consider. They are:

1. Availability
2. Rapid Recovery

3. Cloud-Ready
4. Data Agility

Availability

Availability from the business perspective is about the applications, data, and processes being available to the user or to the consumer in whatever way that's defined by the business as "business as usual."

When key applications or data become inaccessible, the ripple effect can be horrendous across the board, which is why the whole purpose of modern data protection is to ensure data accessibility and availability.

The best way to start on this journey is to keep your data from being lost in the first place! And building data resiliency into your environment starts with a highly available infrastructure. One example of this is the Pure FlashArray with Purity ActiveCluster, which delivers six 9s of availability. Its active-active synchronous replication provides a transparent, automatic, and non-disruptive failover between sites.

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

Rapid Recovery

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

When a business interruption hits your organization, the last thing your business leaders want to hear is, "We're working on it." What they want to know is when things will be working again. That's why recovery is such a critical component of your entire data protection strategy, and is really what success hinges on. Remember: Backup is critical, but restorability defines your success.

Tech-speak no longer cuts it in the current age—talking in terms of availability and "business as usual" is the business-centric approach that will put you in the good graces of the powers-that-be. Keeping that availability high, however, means keeping your infrastructure humming along. If you want to eliminate the bottlenecks associated with traditional purpose-built backup appliances (PBBA) using spinning disk, or if you're still using tape as your primary backup destination, an all-flash solution is essential to moving 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 add up to a compelling reason to consider it when looking to up-level your data protection environment.

Note that there is no silver bullet when it comes to backup and recovery: It's hard to architect a highly performant and resilient infrastructure. Going with FlashBlade, however, is a good place to start when modernizing your data protection strategy.

Why? FlashBlade does not require you to change your existing data protection software or the processes on which your IT organization has standardized. Its flexible nature allows your IT teams to offer a wide range of recovery options and tiers of service.

Because Pure FlashBlade works with existing data protection software solutions, your IT teams can continue to service the mission-critical recovery and compliance requirements by protecting the most essential data using their preferred solution providers, including Commvault,, Veeam, and Veritas.

Cloud-Ready

If hybrid or multi-cloud isn't in your strategy today, it most certainly will be in the future—thus, choosing a cloud-ready solution is just wise

planning. Cloud storage is being used today for a variety of purposes, including backup, DR, and long-term retention.

If you're using Pure FlashArray or considering it in your ecosystem, then CloudSnap, a new feature included with the FlashArray, is something to consider. CloudSnap uses Pure's portable snapshot technology that was first introduced to allow users to create a portable snapshot, consisting of both meta and data encapsulated within this portable snapshot, and replicate it to FlashBlade or an NFS share.

Based on this technology, Pure has added another tier—cloud storage—to its support matrix. When it first introduced this support it was Amazon Simple Storage Service (Amazon S3) only, but in the third quarter of 2019, Microsoft Azure was added to the list, kicking off Pure's multi-cloud strategy.

Which Cloud?

There are many different types of clouds, and they can be confusing, so let's break them down.

- **Private cloud:** A cloud environment that exists only on-premises.
- **Public cloud:** A cloud like Amazon Web Services (AWS), Microsoft Azure, and Google Cloud Platform (GCP). It's fully offsite, with no on-premises components.
- **Hybrid cloud:** A mix of private and public cloud, with a component of each in use by an organization. An example would be Amazon Outposts. Typically, it involves use of just a single public cloud vendor.
- **Multi-cloud:** Using one or more public clouds, for example, Azure and GCP. This is sometimes confused with "hybrid cloud."

This built-in FlashArray technology can be looked at as a self-backup to the cloud: Just schedule it and forget it. And it can, if you'd prefer, run in parallel with your existing backup solution provider to add an additional layer of protection.

Another key benefit of this self-protecting solution is that only delta changes to your primary data are transferred, reducing even more data transfer to the cloud. And when recovery is necessary, CloudSnap is smart enough to know what's missing when it restores these multiple portable snapshots, only returning those missing pieces. It makes for an expeditious recovery process.

Between the integration of the backup solution providers who also support cloud and CloudSnap, IT teams have the flexibility and choice to determine the best possible data protection scenario for the business.

Data Agility

What does the term "agile data" really mean? Let's start by considering the state of our protected data in the past. Many times it sat in a proprietary format on a cold device (tape or some other removable media), and the most that could be done with that protected data was report on what files were stored, the dates they were created and modified, and so on. But when that protected data reached its maturation point, its life and usefulness expired.

That's no longer the case, as data is used for much more than just 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?

If this is appealing, consider that FlashBlade can become your data hub, serving DevOps needs and much more, including as a target for analytics, AI, and ML. And all this happens without impacting production.

As far as security's concerned, Pure provides data-at-rest encryption on all arrays, adding an additional layer of protection against the bad guys. It's more peace of mind for the business.

This makes a world of difference in a business-centric approach, by using assets instead of allowing them to quietly expire. It will also help IT by eliminating multiple silos, complexities, and additional cost burdens that would otherwise eat up the annual IT budget.

Real Customers, Real Successes

That was exactly the experience[1] of financial services company Man AHL, based in the United Kingdom. Man AHL manages more than $24 billion of other peoples' money, and has increasingly turned to machine learning to do it properly. When the company decided its legacy infrastructure could no longer keep up with its high-performance computing environment, it turned to Pure for help.

Man AHL went with an eight-blade FlashBlade configuration, and the results have been impressive. One researcher saw a 10x-to-20x improvement in throughput for Spark workloads, and overall productivity improved "significantly," according to the company.

One executive summed up Man AHL's success with Pure FlashBlade: "There are significant savings from the simplicity of operation. It just works ... We've essentially cut

our headcount for storage management from one-and-a-half to the equivalent of a half-time role. And going forward it'll approach zero. That means we can redeploy those assets to higher-priority activities."

As you can see, modernizing your approach to data protection by taking a business-centric viewpoint will not only help the business recognize the long-term value of the solutions, but provide IT much

[1] https://www.purestorage.com/pure-folio/showcase.html?item=/type/pdf/subtype/doc/path//content/dam/purestorage/pdf/Case%20Studies/Pure_Storage_Case_Study_Man_AHL.pdf

better communication and trust with the rest of the business as solutions are presented and recommended. Pure has a strong portfolio of solutions to help meet your key objectives of digital transformation and modernizing data protection, even to the point of business protection strategies.

CHAPTER 5

Implementing a Hybrid Cloud Backup Solution

In the previous chapter, we discussed the rationale for leveraging Pure FlashBlade and FlashArray technology from Pure Storage to implement a modernized strategy for data protection and availability from a business-centric approach.

As demonstrated by numerous case studies, doing so improves the efficiency of backup and the speed of restore processes over any other backup and restore methodology. This, in turn, delivers the kind of data recovery and business continuity solution required by today's more stringent SLAs.

Using a combination of Pure Storage on-premises in conjunction with the public cloud offers the advantages of performance, density, and cost benefits of flash storage; in addition, it positions you to leverage ongoing improvements in both flash media and interconnected technologies. The runway ahead for flash storage technology is long and robust, and it was inevitable that storage-intensive processes like backup and restore would eventually harness it.

Pure-powered solutions—both independent and integrated with the backup and restore solution providers—leverage cloud technology and architecture. Your public or private cloud service providers offer infrastructure and software services that are agile and easy to access and use, regardless of their proximity to your offices or data center.

Many organizations host their entire IT operation in the cloud, some host only select workloads there, and others leverage only select

Availability Percentage	Downtime Per Year	Downtime Per Month
99.9% (“three nines”)	8.77 hours	43.83 minutes
99.99% (“four nines”)	52.60 minutes	4.38 minutes
99.999% (“five nines”)	5.26 minutes	26.3 seconds
99.9999% (“six nines”)	31.56 seconds	2.63 seconds
99.99999% (“seven nines”)	3.16 seconds	262.98 milliseconds

Table 1: Downtime associated with common “nines” of availability

services and resources from the cloud. But one thing is beyond debate: Cloud is here to stay.

The New Normal

In fact, analysts are starting to talk about hybrid cloud environments as the new normal. Companies are distributing workloads across public and private clouds, in part to capture pricing advantages, but also to capitalize on the specialties of different providers. Many cloud service providers now offer Disaster Recovery-as-a-Service (DRaaS) or Backup-as-a-Service (BaaS) and enable the replication of backup data between multiple cloud sites to provide 99.9999% or better availability for data and workloads. See **Table 1**.

Many companies are drawn to the cloud to handle and host backups for purely economic reasons. By leveraging cloud services, these companies can shed local backup infrastructure and software, and remove the administration and supervision of backups from the duties of IT administrators who have better things to do with their time. In some cases, the combined equipment and labor cost savings more than pay for outsourcing backup to the cloud.

However, the real value of cloud-enabled backup is the possibilities it creates for delivering comprehensive recoverability for applications, especially those that were "born in the cloud." DR requires more than data recovery, though that's central to any successful recovery; recovering from a facility disaster or an outage event of regional impact requires application re-hosting and network re-direction, as well.

If workloads can be re-instantiated rapidly on cloud-based hosts and software-defined networks can be re-directed to the new locations of applications, the entire business can "fail over" to the cloud in the event of a major disaster event; and customers may not even know that any disruption has occurred.

Failover strategies date back to mainframe data centers, of course, but they always required two known data centers, each comparably equipped with processors and storage, and ongoing data mirroring between them.

The cost of maintaining two data centers and keeping all hardware, software, and data synchronized was huge, so only the firms with the deepest pockets could afford such a redundancy strategy. Many firms simply made copies of their data, the irreplaceable asset, and hoped for the best when it came to replacing infrastructure, cobbling together network resources, and getting critical systems back into production.

With cloud technology, data center resources are virtualized. Given an efficient backup and recovery program, data can be placed in a cloud where it can be used by locally re-hosted applications or accessed remotely via wide area or metropolitan area links by servers in business data centers or user facilities. The affordability of clouds has never been better.

A Unique Solution

Pure Storage is unique in its intent on making the journey to clouds easier for its customers. While the cloud is now 15 years in the market,

it's still evolving. It's only been in the past five years that larger firms have embraced the cloud, and many still prefer to split their IT operations between traditional local data centers and cloud-based services in hybrid cloud models.

There are a lot of reasons to embrace hybrid cloud. Outsourcing routine tasks such as backup to a cloud in order to free up local staff and resources to pursue other work of value to the business may make a lot of sense.

However, more often than not, the choice of a cloud service is guided less by business requirements than by technology constraints.

Pure Storage changes that dynamic, unifying clouds to empower enterprises to build a better world with data. Because in the end, it shouldn't matter whether you run an application in a private, colocated or public cloud.

If your primary data is stored on a Pure Storage FlashArray system, you can harness its native snapshot capabilities and export it to a FlashBlade system, without the need for additional software (**Figure 5** has a list of specifications). You can also use your existing backup and recovery solution to back up data from FlashArray to FlashBlade. Either way, as mentioned earlier, when building a highly performant, data resilient infrastructure, it's important to ensure all pieces either equal or exceed the requirement: In this case, flash-to-flash would certainly meet that objective.

In a heterogeneous storage environment, however, customers can utilize their existing backup and recovery solution to back up data to FlashBlade and off to cloud for long-term retention, based on the solution provider's feature set.

The configuration is simple. FlashBlade presents an S3 interface, which is configured as a backup target in the backup software. Typically, this backup target is added to a backup policy, which defines other parameters, such as retention periods and scheduling.

FlashBlade Specifications

SCALABILITY	• Start with 7 Blades and simply add blade to scale up to 75 Blades • Each blade adds capacity and performance
PERFORMANCE	• Up to 17 GB/s bandwidth with 15 blades* • Up to 7.5 NFS IOPS with 75 blades
CONNECTIVITY	• 8x 40 Gb/s or 32x 10Gb/s Ethernet ports / chassis • 2x FlashBlade External FabricModules (XFM) to scale up to 75 blades
PHYSICAL	• 4U per chassis • 1,800 watts per chassis (nominal at full configuration)

* Large block read IO with 3.1 compression

Figure 5: What Pure FlashBlade offers

Once a backup is initiated, data is backed up from the target storage via the backup server to FlashBlade.

This is a key advantage of Pure Storage's modern data protection strategy. To gain the benefits of FlashBlade, customers don't have to "rip and replace" their existing data protection solution or infrastructure. FlashBlade and can be easily deployed with many of the leading enterprise backup and recovery technologies, and with some minor configuration changes can be operational quickly.

In addition to Pure's hardware and software innovations, the company's continuous innovation is gaining a robust and exciting ecosystem of partners and technology support.

Moving forward, Pure's software offers a robust roadmap for the future. What started as a software-only adjunct to backup and restore that would facilitate the integration of cloud-based object storage into the backup and restore process has been integrated with FlashBlade to provide one of the fastest flash-based restore platforms in the world. It's conceivable that more public and private cloud services will be supported to help reduce the complexity and cost of modernizing a data protection strategy using cloud.

The Pure Storage vision extends beyond data storage and data protection to data management and value, too. In the next chapter, we'll look at some of the value-add that accrues from using the Pure Storage backup model and its supporting flash and service technologies to unlock the value of backup data.

CHAPTER 6

Warm Up Your Cold Data

As has been touched on previously, data protection vendors have long been trying to build a capability for re-using or repurposing backup data for other workloads, such as test/development, analytics, and AI. Why? Organizations realize their data can be used to establish competitive advantage, as well as running day-to-day business operations. Backup applications are well positioned to deliver on this. After all, they access all of an enterprise's data on a daily basis, so why not utilize data for other uses?

Newer entrants into the data protection market have attempted to combine the workloads mentioned earlier with backup and recovery. Mature data protection vendors have tried to extend their platforms to offer these capabilities. Both approaches have met with limited success. For the new vendors who are building new data protection offerings, there is a high degree of complexity in creating a product which supports the many operating systems, virtualization layers, and applications. For the mature data protection vendors, retrofitting these capabilities can be equally complex.

But by being able to restore data at rapid speed, it's now feasible to restore multiple terabytes of data, for example, in less than an hour, helping fuel test/dev or analytics workflows with the raw data that they need to help organizations to make use of it. This allows infrastructure teams to keep their existing data protection product or process in place (or choose a product that meets their data protection requirements), which minimizes disruption and complexity.

This is the problem a leading SaaS company was trying to solve, and it chose FlashBlade to solve it. This Pure customer performs more

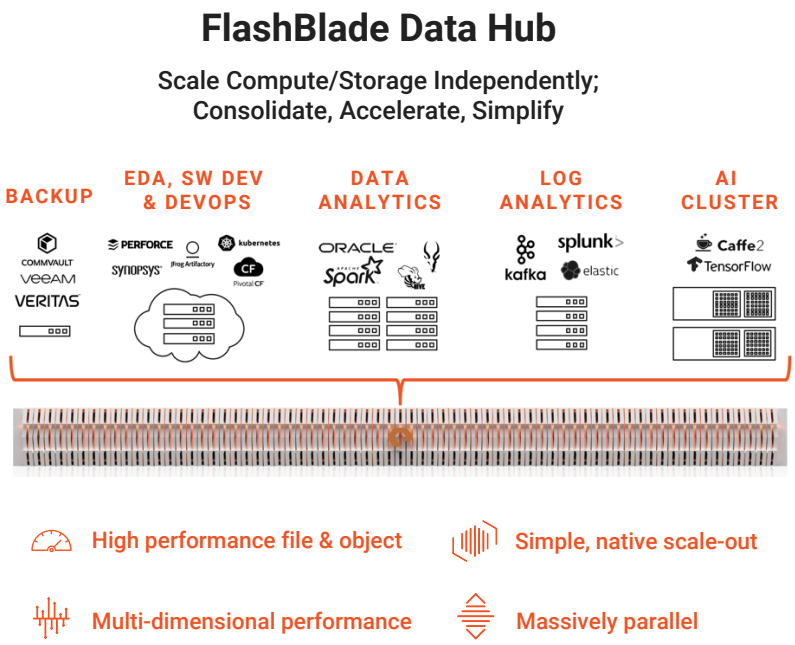


Figure 6: The Pure Storage FlashBlade data hub architecture

than 500 restores a month for its customers who wanted to repurpose their backup data for test/dev and analytics. FlashBlade was the only high-performance object store that could meet these aggressive re-store SLAs.

FlashBlade's unique architecture helps organizations collapse data silos and consolidate workloads with a Data Hub, providing a highly scalable and performant Object store, which is capable of meeting the performance requirements of large unstructured applications. See **Figure 6**.

What, exactly, is FlashBlade's unique architecture? Typically, storage has fixed compute power and memory, which means that on day one, when storage is provisioned (and is largely empty), read and write performance is as advertised. Consider, though, what happens on day 730. The storage is likely 50% full, but memory and compute remain

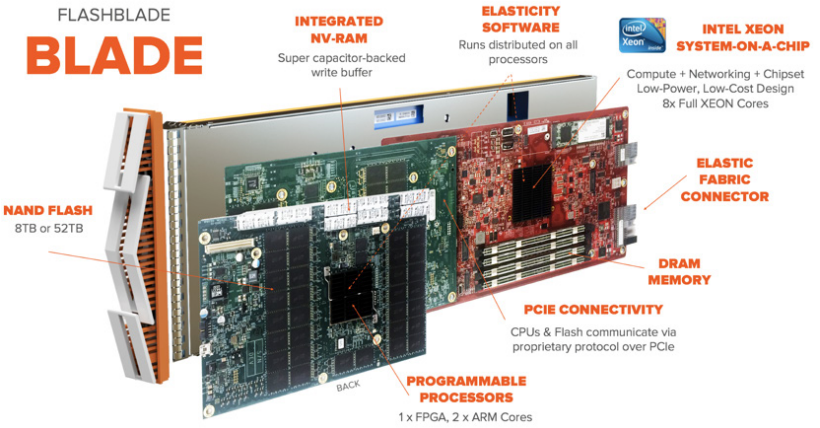


Figure 7: FlashBlade contains additional storage, as well as additional compute

the same. This can have severe consequences in some situations such as data recovery (look at restore performance for PBBAs). FlashBlade doesn't suffer from this because of its blade architecture. Each blade contains not only additional storage, but also additional compute (see **Figure 7**).

This means as storage expands, read and write performance increases (see **Figure 8**).

What does this mean for organizations? It means FlashBlade is built to handle large, unstructured workloads such as test/dev, analytics and data protection, helping customers to consolidate their unstructured workloads into a single storage layer. A great example of this is a leading cancer treatment organization in the United States. By deploying FlashBlade, this customer could not only guarantee a recovery SLA for their tier 1 application, but was able to reduce their overall TCO cost by 50% for their unstructured data storage requirements for their EPIC and object applications.

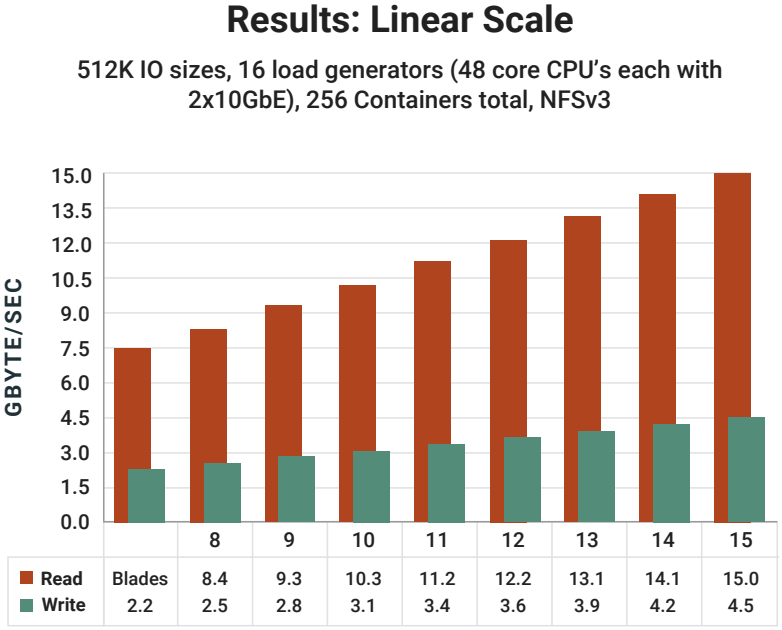


Figure 8: FlashBlade gets quicker as it gets bigger

Modernize—or Die

This Gorilla Guide has covered many topics, including the challenges of legacy backup and recovery; the changing nature of data; trends affecting the IT industry like public cloud storage and on-premises flash storage; and how Pure Storage has remained at the leading edge of these developments.

What should have come through loud and clear is that the old ways of doing backup and recovery are simply not up to today's challenging environments. It's a cloudy world, more data is being created than ever before, and that data can (and does) live anywhere and everywhere. It's also being touched by more and more devices than ever before.

It's obvious that to deal with those challenges and make sure that your data is not only properly protected and resilient, but working for you in profitable new ways, it's essential to rethink your backup and recovery

strategies. If you don't, it's guaranteed that you'll be left behind by the companies that have moved into the modern era, one which scales with their needs and keeps their data—wherever it is—safe, secure, and easily recoverable.

These are the businesses that can survive the inevitable outages that occur. Their operations may not even be interrupted when the power goes out or the hacker gets in. In other words, they're ready for whatever comes.

Are you? If not, check out what Pure Storage can do to help. For more information, please visit purestorage.com/cloudbackup or email info@purestorage.com.