PURE STORAGE PRESENTS

THE GORILLA GUIDE TO...®

Rapid Restores with Flash and Cloud

A How-To Guide for Modern Enterprise Data Protection

Jon Toigo

INSIDE THE GUIDE:

- The changing face of data protection
- The real value of cloud-enabled backup and restore
- Why Flash-to-Flash-to-Cloud is a better way to protect data

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
Jon Toigo

EDITOR
Keith Ward, ActualTech Media

LAYOUT AND DESIGN
Olivia Thomson, ActualTech Media

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

ENTERING THE JUNGLE

Introduction **6**

Chapter 1: Disk to Disk to Tape: The Gold Standard for Backup Under Pressure **10**

How We Got Here 14

Chapter 2: The Evolving Role of Flash in Backup and Recovery **20**

The eMoney Success Story 22

Chapter 3: From D2D2T to F2F2C: Clouds are the New Miracle Tape **27**

Addressing the Pain Points 31

Chapter 4: Implementing a Hybrid Cloud Backup Solution **34**

The New Normal 35

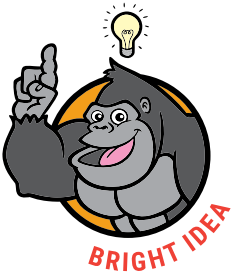
A Unique Solution 36

Chapter 5: Warm Up Your Cold Data **40**

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 that we hope will be useful to IT planners, storage architects, compliance officers, auditors, and backup administrators who are seeking to ensure and improve the data protection and recovery strategies that are currently being used to safeguard your organization's most irreplaceable asset: your data.

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 build their business and competitive edge on data, rapid access to data for disaster recovery and for 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 clouds and flash storage.

One of the key recent innovations in this area is the concept of "Flash-to-Flash-to-Cloud," or F2F2C.

The high-level advantages of F2F2C 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). F2F2C provides a means to bring data 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. True F2F2C enables the portability of backup data and streamlines the placement of data into cloud repositories using object

protocols such as Amazon S3. Using the cloud to store backup data provides both the required separation between original and backup copies of data to ensure resilience in the face of facility and regional disaster events, and also a means to capture cloud economics and flexibility to reduce the cost and complexity of backup infrastructure.

3. The F2F2C 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 F2F2C 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 to understand how the F2F2C architecture, and specifically with solutions from Pure Storage, can help you to realize your objectives. Of course, objectives vary based on the reader.

To the IT planner, the central rationale for undertaking data protection is evolving: while the entire ecosystem has been about optimizing for backup speeds and processes, it's really about rapid recovery. When business is offline, restore capability is the only thing that matters.

F2F2C architecture is built to restore rapidly and leverages flash storage as a key enabler as it is 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 object protocols. Significant restore time improvements have been demonstrated by users implementing an F2F2C strategy.

To the IT architect, of course, improved restore speed is not 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 re-use 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 above, 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 latest General Data Protection Regulation (GDPR) enacted in 2018 by the European Union.

With F2F2C, many compliance-oriented users see the improvement of backup processes as a linchpin of building data stewardship excellence and ensuring legal/regulatory compliance.

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.

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

An F2F2C 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 F2F2C and its components and will offer use cases intended to survey and underscore the possibilities for leveraging Pure Storage FlashBlade and ObjectEngine to improve your data protection strategy. A good place to start is with a concise summary of the situation in most organizations today. That is the focus of the next chapter.

CHAPTER 1

Disk to Disk to Tape: The Gold Standard for Backup Under Pressure

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 60 zettabytes (ZB). By 2024, they foresee data growing to approximately 160ZB, courtesy of trends in mobile commerce, Internet of Things, and other factors.

Beyond Petabytes and Exabytes

A zettabyte or ZB is a unit of measurement equivalent to 1000 exabytes or one million petabytes or one billion terabytes. Industry analysts began projecting ZB-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. Solutions must be found to host this data more cost effectively, a concern that is driving interest in using the cloud -- especially for storing backup data.

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 data growth; and, as tolerance for downtime is decreasing, 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 man-made 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 (GDPR) that took effect in 2018.

The regulation, which impacts both companies located in the European Union 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 requirements 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."

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 (GDPR) that took effect in 2018.

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 is beginning to shape the backup landscape is the quest in many companies to find ways to re-use backup data when it is not 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, scripting, 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.

The Rise of Artificial Intelligence (AI)/Machine Learning: Changes in the backup landscape are also being driven by the rapid rise of AI or machine learning. 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 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 life-cycle policies.

The above 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.

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.

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 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 24/7 processing schedules, time grew scarce for tape backups, and new solutions were sought.

The first solution was to write backup data to a cache of disk managed by its own server: 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.

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**).

At about the same time, the concept of 3-2-1 was invented: a perfect backup was three copies of 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 a facility disaster.

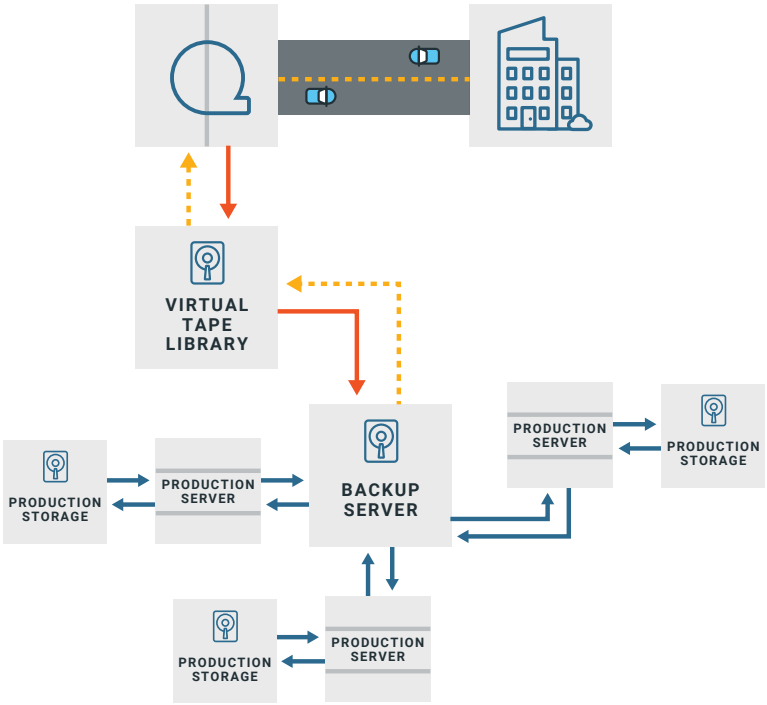


Figure 1: Disk-to-disk-to-tape (D2D2T).

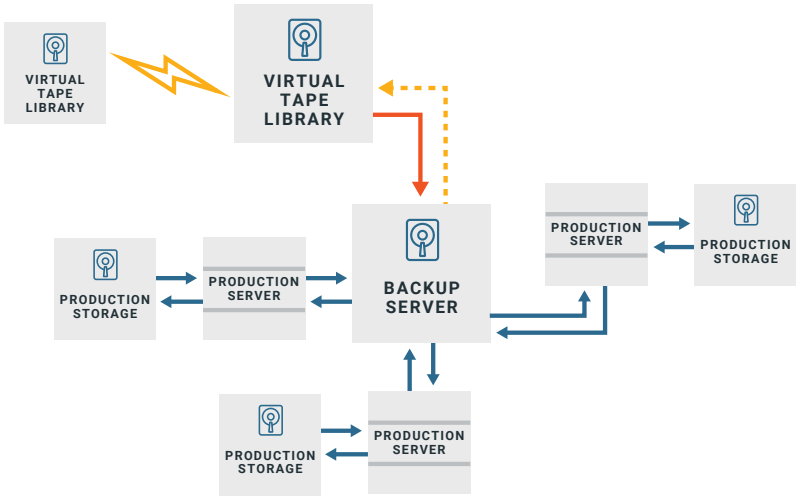


Figure 2: D2D2T morphs into VTL/backup server mirroring, eventually leveraging public clouds.

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.

Some firms saw this capability as an invitation to simplify backup by eliminating tape altogether, in some cases by leveraging a mirrored "backup server" (a backup appliance that no longer presented itself as a virtual tape library) located in an off-site, outsourced, data center 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 medium-sized 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.

The story hasn't ended there. Several changes are occurring today in the backup model. For one, backup servers 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 backup appliances and protection process improvements, to restore speed.

With flash-based systems, you don't incur the read penalty when restoring from deduplicated backups on spinning disks, such as those within PBBAs.

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.

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.

Clouds and solid-state 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 backup data to be mirrored between primary and secondary storage in a multi-cloud setting 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 backup repositories (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 backup 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.

This, in the final analysis, provides the context for Pure Storage's early leadership in the emerging flash-to-flash-to-cloud paradigm for data protection. We'll look at the Pure solution in greater detail in the next chapter.

CHAPTER 2

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 servers 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 the energy demand of flash storage compared to HDD arrays, producing greater cost-efficiency
- Flash storage can leverage new 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™ system,[1] with design principles centered on effortless administration, efficient storage and a unique Evergreen™ storage consumption model.

However, the company's innovation did not end there. Pure Storage also envisioned bringing the power of all-flash platform to unstructured file and object storage, and as a result, FlashBlade was introduced to the market in 2016.

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

Then, in 2018, Pure Storage acquired StorReduce, adding sophisticated deduplication technology to its object storage portfolio and expanding its public cloud integrations to meet the growing demand to manage unstructured data in multi-cloud environments. The StorReduce technology was subsequently renamed ObjectEngine.

The combination of FlashBlade, ObjectEngine, and the cloud 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

[1] Pure Storage, FlashArray, FlashBlade, ObjectEngine, and the P logo are trademarks or registered trademarks of Pure Storage, Inc. All other names may be trademarks of their respective owners.

number of public and private cloud protocols and storage offerings enables it to scale effortlessly as multi-cloud and hybrid cloud infrastructures are adopted en masse.

The eMoney Success Story

Moreover, this approach is facilitated by deduplication that reduces the backup data "footprint" and the cost of storage infrastructure and 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 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.

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.

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.

They're not alone. ServiceNow is a juggernaut in the Software-as-a-service (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.

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

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

Part of the problem was backup data, which was being generated at a rate of more than 490 TB 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 timeframes lengthened – to the point where planners determined that the company 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.

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 their new architectural model, which delivered three times the storage density of their legacy infrastructure. Each rack comprised 30 1U servers with 1,000 processor cores and 1.5 PBs of effective flash storage capacity.

Moreover, server CPUs were no longer performing backup work, enabling their cores to be dedicated to improved 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.

This reliability, combined with the economics of Pure Storage at scale and fast networks, are game-changers for SaaS providers.

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 extends into clouds. We will look at this in the next chapter.

CHAPTER 3

From D2D2T to F2F2C: Clouds are the New Miracle Tape

In the previous chapter, we discussed the emergence of flash-to-flash-to-cloud and provided an overview of a Pure Storage data hub built with FlashBlade and ObjectEngine. We recounted a pair of user case studies to underscore the process improvements and other benefits yielded by the adoption of Pure Storage flash array and appliance technology, and public and private cloud as components of a backup infrastructure.

Cloud is sometimes confused with a type of storage, such as tape, hard disk, optical disc or flash storage. In fact, clouds are a service delivery model. Within the cloud is a data center that uses much the same technology as businesses use in their own data centers.

When clouds first appeared on the IT landscape, they were locations where applications and/or computing, networking or storage resources could be accessed and leveraged on a subscription or pay-per-use basis. In other words, cloud service providers offered a method for outsourcing for companies that did not wish to host their own IT operating environment.

In other cases, cloud services were seen as "elbow room" for on-premises data centers, a source of additional compute or capacity during peak load times. The term "hybrid cloud" was introduced to describe an IT environment in which some infrastructure and workload was located on-premises, some workloads were delivered

via subscription cloud-based services, and some resources were acquired as needed in a sort of "build your base, buy your burst" model. An example is shown in **Figure 3**.

Cloud is sometimes confused with a type of storage, such as tape, hard disk, optical disc or flash storage. In fact, clouds are a service delivery model.

In other cases, organizations moved their workloads to the cloud for reasons ranging from the practical ("the local power grid is saturated, so we need to stand up new servers or storage outside our data center") to the philosophical ("IT is not a core competency of our widget company, so let's hire an expert to run it for us").

Today, it's not uncommon to see certain start-up firms creating their entire business operation in a public or private cloud environment, where *private cloud* is generally interpreted to mean the use of a service provider other than the largest "brand name" public service providers like Amazon Web Services (AWS), Google Cloud Platform (GCP), Microsoft Azure, etc. These businesses – and the application software and infrastructure they use – are sometimes referred to as "born in the cloud."

From the above, it's easy to see that cloud technology enables a number of implementation options, from total data center outsourcing, to selective application outsourcing, to selective infrastructure outsourcing. The one thing that cloud isn't is a kind of storage.

When obtaining storage-as-a-service from a cloud service provider, you're requesting access to a set of flash, disk, or tape resources that are deployed, managed and maintained by the cloud service provider. That said, both public and private clouds increasingly tend to use object-based storage, creating some challenges for older backup and restore workflows.

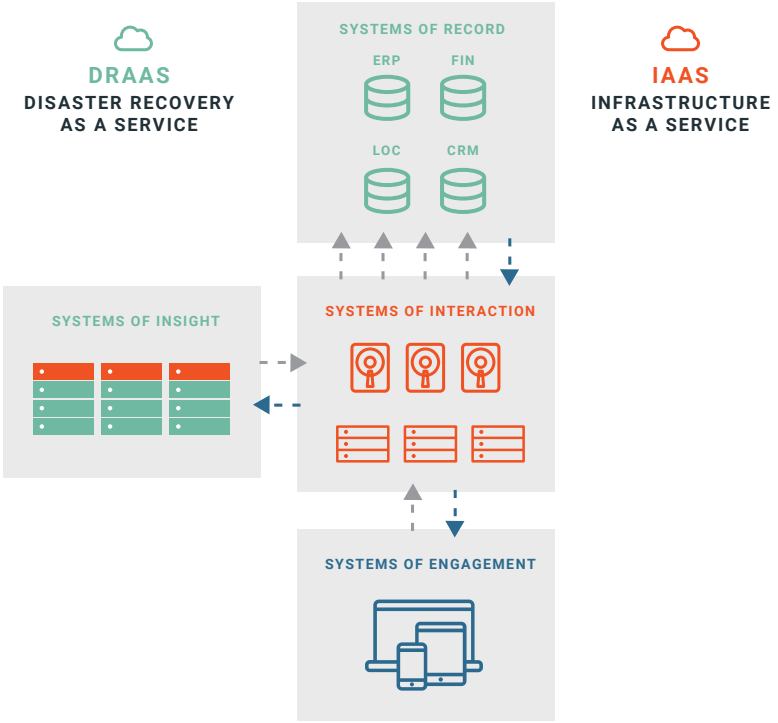


Figure 3: The hybrid cloud model: Corporate data center obtains select services from a cloud service provider (for example, Disaster Recovery as a Service or DRaaS) and additional compute, network, or storage resources on an as-needed basis from other clouds.

Applications that are "born in the cloud" are usually object storage ready. Pure Storage's Rapid Recovery solution, for instance, which combines FlashBlade and ObjectEngine, supports a growing list of cloud object protocols, beginning with Amazon S3. This has the added advantage of placing backup data in a format for ease of access and reuse by other applications, if desired.

With cloud storage in mind, Pure Storage designers have also built the rapid restore system to capitalize on their deduplication and transport parallelization technologies to optimize all transports and links both for rapid recovery on premises and across the public network to cloud service providers.

Clouds extend this robust backup/recovery story. Data can be backed up from non-flash production storage arrays and from flash production storage arrays on-premises (the first "F" of F2F2C) to a Pure Storage all-flash data hub (the second F) to enable rapid restore on-premises.

The contents of the FlashBlade/ObjectEngine rapid restore solution can then be copied to the public or private cloud (the C) to provide protection against facility- or milieu-level disasters without sacrificing rapid restore, whether to another location or in the cloud itself.

IDT Telecom provides a case study that may help to clarify some of these options. IDT Telecom is a pioneer in prepaid communication and payment services and international Voice over IP telephony. The company's IT planners prefer to leverage technology hosted in cloud services rather than purchasing, deploying, operating and maintaining the technology itself.

They sought to apply this preference to storage first as a cost-savings measure, eschewing a storage hardware vendor practice of forcing the customer to re-buy their maintenance agreement or refresh their products every three years.

Their product search led them to Pure Storage flash technology arrays, then to StorReduce, a company and technology for enabling "flash-to-flash-to-cloud." Pure Storage acquired StorReduce in 2018, rebranding the technology to ObjectEngine.

The combination of technologies enabled IDT to improve the quality of service they were delivering to customers, including a marked improvement in uptime and availability, by eliminating most of the downtime related to data access interruptions.

Pure Storage® is the early thought leader in the transformation of the disaster recovery emphasis moving from backup to rapid restore. This is happening with Pure Storage FlashBlade™ scale-out, network-attached file and object storage and the Pure Storage ObjectEngine™ platform.

Using their hardware and software elements, Pure has created a data hub that powers a wide range of workloads, including a data protection platform that works in concert with third-party backup vendors to store data safely in both public and private cloud.

Addressing the Pain Points

Flash-to-flash-to-cloud is more than a wordsmithing of a familiar backup meme. Advocates of this strategy have sought to directly tackle three issues that have challenged cloud-centric backup strategies in the past:

- How to move data from on-premises computing environments to the cloud
- How to copy block and file data to the cloud
- How to maintain the cost-efficiency of the strategy at scale

The challenges related to moving data from on-premises to cloud environments derive, according to F2F2C advocates, from legacy thinking about backup, which leveraged tape as a means to transport data to secure offsite storage.

This modality of backup data hosting produced slow restore speeds, because tape needed to be retrieved from offsite storage, delivered

to the original data center (or recovery site), loaded into a library, and their contents read in the proper sequence back onto production storage media.

The total time required to accomplish these steps, and the all-too-common hurdles of improper read sequences, could delay recovery by hours or days. (And this doesn't even factor considerations for long-term durability of tape.

As previously discussed, a cache of disk was inserted between the production data and the tape library in the hopes that it would expedite backup operations in the face of shrinking backup windows, and to facilitate rapid restore. However, scaling the disk cache using proprietary backup appliances proved to be another challenge.

Moreover, backup appliances were notoriously slow performers when it came to data restore speeds. While faster than tape with respect to backing up data, the restore speeds obtained from backup appliances tended to be about *ten times slower* than backup speeds.

Each appliance had a fixed capacity and required an entirely new appliance be deployed when the first appliance was fully populated with data. Each appliance also required individualized management and maintenance, increasing the complexity of the solution.

Moreover, backup appliances were notoriously slow performers when it came to data restore speeds. While faster than tape with respect to backing up data, the restore speeds obtained from backup appliances tended to be about *ten times slower* than backup speeds. The more backup data stored in the disk cache, the slower restore speeds became.

The verdict on backup appliances was that they misunderstood the primary goal of data availability, which was to restore data rapidly, not to back it up rapidly. Engineers at Pure Storage and elsewhere in the industry sought a solution that would provide greater speed in both backup *and* restore using low-cost flash and data reduction technology.

The last leg of F2F2C strategy involves moving data from the on-premises FlashBlade backup repository to the cloud itself; and, if desired, restoring data from the cloud to an on-premises FlashBlade system. In large part, the speed of data transfers to and from the cloud is throttled by wide area network link bandwidth. Connectivity must be sized to match service level agreements.

The good news is that placing what used to be "cold data" (siloed backup data) into scale-out, network-attached file and object storage such as FlashBlade can enable the use of the data by multiple workloads, including analytics and application testing and development. We will look at this in greater detail in Chapter 5.

CHAPTER 4

Implementing a Hybrid Cloud Backup Solution

In the previous chapter, we discussed the rationale for leveraging public and private clouds in conjunction with FlashBlade and ObjectEngine technology from Pure Storage to implement a modernized version of the disk-to-disk-to-tape strategy for data protection and availability called "flash-to-flash-to-cloud."

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.

F2F2C captures the performance, density, and cost benefits of flash storage, and positions you to leverage ongoing improvements in both flash media and interconnect 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.

F2F2C backup and restore, by definition, also leverages 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, others host select workloads there, and still others leverage only select services and resources from clouds. 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 and enable the replication of backup data between multiple cloud sites to provide five nines or better of availability for data and workloads.

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."

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." Disaster recovery requires more than data recovery, though that is 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 clouds, and 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 their intent on making the journey to clouds easier for their customers. While clouds are now 15 years in the market, they're still evolving. It's only been in the past five years that larger firms have embraced clouds, 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 wants to change that dynamic, and is seeking to unify the cloud and empower developers to build a better business with their data. It shouldn't matter whether you run an application in a private cloud, on-premises or off, or in a public cloud: technology should not impair your choices or stifle innovation. All that should matter is manageability, which should be seamless wherever workload or data are hosted.

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

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. You can also use your existing backup and recovery solution to back up data from FlashArray to ObjectEngine and FlashBlade. Either way, as mentioned earlier, this represents "flash-to-flash" in an F2F2C data protection strategy.

In a heterogeneous storage environment, however, customers can utilize their existing backup and recovery solution to back up data to ObjectEngine and FlashBlade.

The configuration is simple. ObjectEngine and FlashBlade present 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.

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

ObjectEngine data deduplication technology is leveraged to reduce the size of the payload that must travel over networks between flash systems. Pure Storage boasts data ingestion speeds of up to 25TB per hour, and data restore speeds of up 15TB per hour with ObjectEngine, depending on workload and assuming a data reduction rate of 10:1. The solution is designed for scaling up with the addition of nodes.

Data deduplicated by ObjectEngine is stored in a deduplicated format on FlashBlade. If a copy of the backup data is needed in the public cloud for compliance or other purposes, ObjectEngine will initiate the copy process of deduplicated backup data to the cloud. This represents "cloud" in an F2F2C data protection strategy, as mentioned earlier in this guide.

Likewise, recovery of data is also simple. The recovery process is initiated by the backup and recovery solution and can be directed either from FlashBlade or from the cloud. In both cases, ObjectEngine acts as the translator between the backup and recovery solution and the cloud or FlashBlade. Much like the backup process, the backup and recovery solution serves as the orchestrator for the recovery process (**Figure 4**).

This is a key advantage of Pure Storage's F2F2C data protection strategy. To gain the benefits of FlashBlade and ObjectEngine, customers don't have to "rip and replace" their existing data protection solution or infrastructure. FlashBlade and ObjectEngine can be easily deployed with many of the leading enterprise backup and recovery technologies, and with some minor configuration changes can be operational quickly.

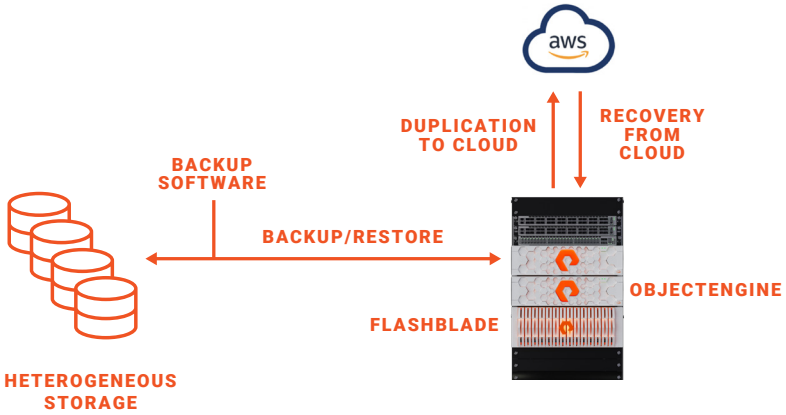


Figure 4: Pure storage's simple backup recovery process.

In addition to Pure Storage hardware and software innovations, the company's F2F2C innovation is gaining a robust and exciting ecosystem of partners and technology support.

Moving forward, ObjectEngine 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 F2F2C backup strategies.

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 5

Warm Up Your Cold Data

Up until now, our primary focus has been to familiarize you with the capabilities of a data protection strategy, F2F2C, that leverages flash technology, cloud services and object storage. The experience of Pure Storage developers and their customers confirms that moving data protection from legacy processes to a F2F2C model can help improve backup efficiency and restore speeds.

Pure's philosophy of F2F2C also helps decrease the complexity of the backup and restore service (and the data center in general) by offering a compelling alternative to disk-to-disk-to-tape architectures, such as dedicated backup appliances.

Inevitably, as data volumes increase, so does the need to store backups. As mentioned earlier, with a backup appliance strategy, this leads to the acquisition of additional appliances as your data continues to grow, creating "backup silos" that need to be managed (see below).

Overall, the backup appliance approach is fraught with inefficiencies and complexity, when customers are trying to reduce both of these challenges.

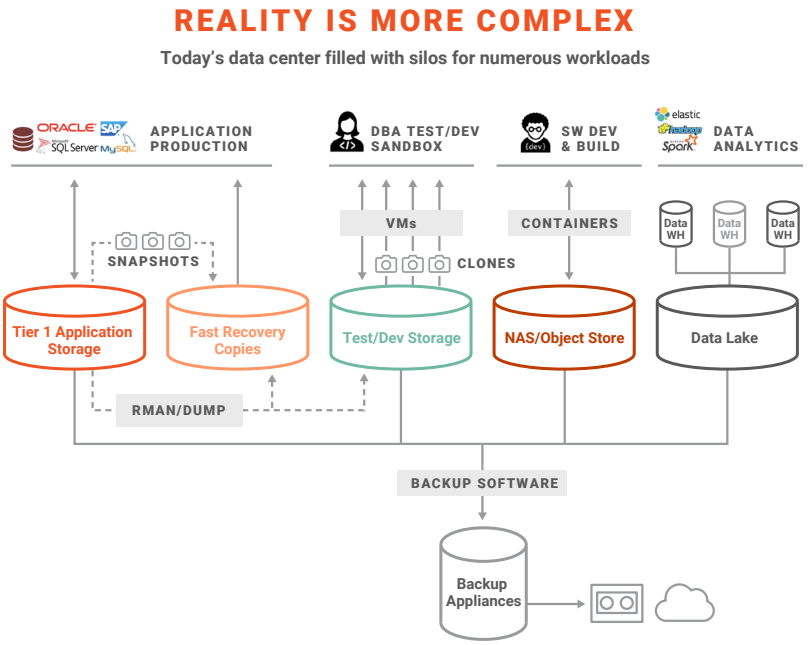


Figure 5: A typical data center, complex and difficult to back up and recover.

In addition, regarding storage efficiency, note that deduplication is achieved within the backup appliance itself, not across multiple backup appliances. Overall, the backup appliance approach is fraught with inefficiencies and complexity, when customers are trying to reduce both of these challenges (**Figure 5**).

Contrast this with the Pure Storage approach, where FlashBlade is an all-flash, scale-out, network attached file and object store platform: not only does it provide rapid data restore, but it also supports multiple use cases simultaneously for data reuse. These include machine learning, data lakes, analytics, and data warehouses, helping you warm up your cold data.

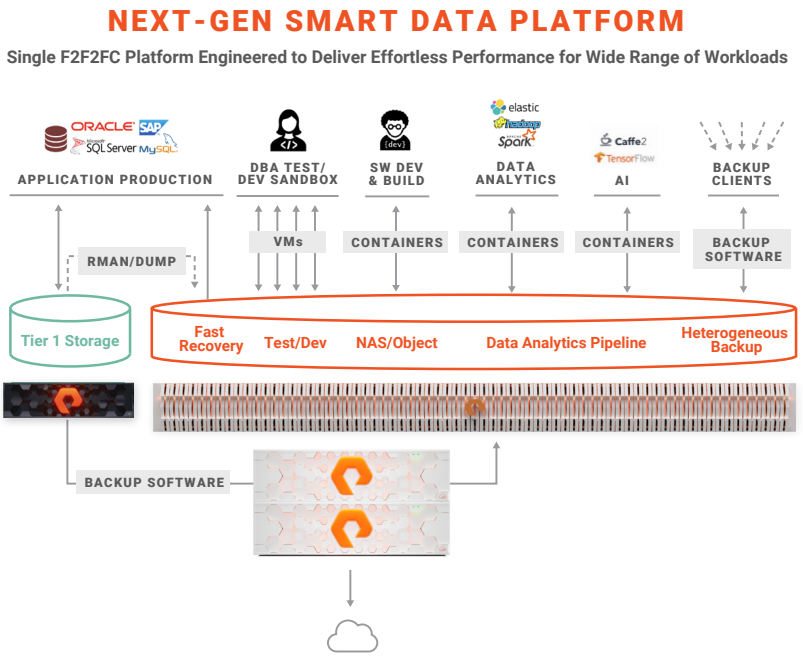


Figure 6: Backup and recovery streamlined with Pure Storage.

FlashBlade can consolidate applications onto a single storage layer (see below), and together with ObjectEngine, helps optimize backup storage footprint on-premises and/or in the cloud (**Figure 6**).

All of these capabilities suggest that the time is right to begin considering an F2F2C approach for data protection. For more information, please visit www.purestorage.com/cloudbackup or email info@purestorage.com.