

## White Paper

# Modernizing Data Protection with Performance and Simplicity

Sponsored by: Pure Storage

Phil Goodwin  
October 2018

## INTRODUCTION

---

Digital transformation is fundamentally changing the way organizations view and manage their data. IDC research indicates that 60% of organizations will have embarked on a digital transformation effort by 2020. For many of them, the goal is becoming a data-driven organization. Data-driven organizations are those that are able to generate revenue and competitive advantage from the use of data. First and foremost for digital transformation, data must be accurate and available to the consumer of the data. Thus, we find that for many organizations, traditional notions of recovery point (RPO) and recovery time (RTO) measure in hours are unacceptable. Organizations are driving to serve data with zero downtime and zero data loss. Over time, we believe that organizations will consider data availability as the primary task, with data protection as one component of that task. Data protection (i.e., backup and recovery) will not be a separate task; rather, it will be a portion of the data availability scheme.

The backbone of data protection schemes for many organizations is the purpose-built backup appliance (PBBA). PBBAs were developed in response to the need for direct backup to disk rather than tape. The devices typically feature deduplication, which makes them very efficient for backup purposes, but require data "rehydration" in order for the data to be used. The rehydration process in recovering data can take an order of magnitude longer than backup workloads. PBBAs tend to be very efficient for streamlining backup and reducing the reliance on tape but were not designed to facilitate other workloads.

In addition, data-driven organizations are discovering a wealth of value in their backup data. No longer is this data intended to lie fallow as nothing more than an extra copy. This data can be used for analytics, test/dev, staging, ediscovery, and other uses – if only it can be made available. We call these secondary use cases because they use the second copy of data for additional workloads. While PBBAs will continue to offer value for backup/recovery, we find that many organizations are turning toward secondary storage systems. Secondary storage offers the advantages of PBBA as a data protection target but also unlocks the value of the data by making it available to other applications. We believe that organizations will leverage secondary storage for at least the following use cases:

- Data protection with fast data and virtual machine (VM) recovery
- Test/dev
- Analytics
- Artificial intelligence

## IN THIS WHITE PAPER

---

This white paper discusses the use cases for secondary storage systems as well as the benefits such systems present over traditional storage architecture. In addition, this document will examine Pure Storage's FlashBlade, a new scale-out file and object product built with a data hub architecture, enabling a single platform to consolidate numerous secondary use cases. In this document, we use the terms *secondary storage systems* and *data hub* interchangeably.

## SITUATION OVERVIEW

---

Traditional siloed data protection is becoming obsolete as IT organizations seek ways to both protect data and utilize the backup data for secondary uses. Consequently, secondary storage systems are emerging to address both data protection and secondary storage use cases. For many scenarios, where backup data needs to be put to work, flash-based secondary storage can offer a better cost profile as well as the quality of service (QoS) of flash.

Secondary storage is rapidly gaining attention because of its ability to solve multiple data management problems simultaneously. The traditional view of datacenter activities is that backup/recovery, test/dev, staging, and data analytics are separate tasks, usually deployed with separate infrastructure, staff, and business processes. For this reason, PBBA – under the purview of the backup team – has served only the needs of data backup. When DBAs or application developers need data sets for their own purposes, an entire separate provisioning process, often taking days or weeks, is needed to provision storage and make the data copy. This data copy is usually made to slower, older systems.

Modern secondary storage systems can deliver several benefits beyond traditional architectures, including:

- Serve as a backup target for fast backup
- Restore data rapidly in disaster recovery scenarios
- Serve multiple use cases simultaneously by making copies of data available to multiple applications
- Suited to serving file-based unstructured data
- Provide superior quality of service by addressing sequential and random-access data requests equally
- Speed up application delivery by reducing the business process for provisioning data copies from days or weeks to on-demand self-service
- Improve data copy quality by making up-to-date data immediately available for test/dev or analytics
- Reduce cost by consolidating purpose-built, single-use case storage systems
- Improve simplicity by reducing the number of devices that must be deployed and managed
- Provide data replication capabilities to the cloud

It is important to note that being called secondary storage does not imply secondary performance. Vendors are increasingly adding flash storage to their secondary storage arrays in an effort to deliver the QoS demanded of the various applications utilizing the storage. In some cases, the secondary storage is tightly associated with the primary storage to deliver seamless data management, rapid delivery, and assured recovery.

We believe a majority of applications will have a cloud component by 2020, including those currently deployed on-premise; the most common architecture will be hybrid cloud. For many applications, the primary cloud component may be data protection. In these cases, organizations utilize the cloud for offsite storage as a means to protect the data for disaster recovery or long-term archive. Rather than a separate function, this transfer of data to the cloud is a part of the data movement business process. As such, IT managers expect it to be both automated and seamless.

In addition to the hybrid cloud environment described previously, some organizations will prefer to have the data copied to more than one cloud in a multicloud architecture. This may be for additional data protection, to avoid cloud vendor lock-in, or to make the data available to applications in different clouds. Thus secondary storage systems must have the ability to service multiple public clouds simultaneously.

## CONSIDERING A DATA HUB FROM PURE STORAGE

---

Pure Storage has become well recognized in the industry for its high-performance all-flash primary storage systems. Recently, the company has entered the unstructured storage market with its FlashBlade product. FlashBlade is an all-flash solution optimized for serving file and object storage, built as a data hub to deliver high performance for numerous secondary use cases on a single platform.

FlashBlade has been designed to eliminate the need for single-purpose devices while serving data with the QoS assurances of primary storage. Typically, IT managers must make a trade-off between high-cost, high-performance storage and low-cost, low-performance storage. FlashBlade is intended to eliminate this trade-off.

As a data hub, FlashBlade has been designed to deliver on the needs for several use cases, including:

- **Data protection and rapid restore.** Data may be copied from a primary Pure Storage array to FlashBlade. Because of FlashBlade's all-flash scale-out architecture, data and virtual machine images can be restored in a fraction of the time needed for spinning disk arrays. Whereas some purpose-built devices are optimized for backup, FlashBlade is optimized for restore. From a service-level delivery perspective, this means that both RPO and RTO can be reduced because of more frequent data copies and faster data restores.
- **Test/dev.** Application developers can access the secondary copy of data to test against current data, eliminating the need to copy the data to an older storage system dedicated to test/dev workloads. This access is self-service, reducing the time and business process needed to provision and create a copy. This replication can include sandbox development or other areas where frequent access to current data is needed.
- **NAS and object storage.** FlashBlade is specifically designed to optimize serving of file-based data and object storage, including containers.
- **Data analytics.** The secondary data repository may be directly accessed by analytical applications, big data, and the like. Data is immediately accessible as it is refreshed from primary storage.
- **Cloud replication.** FlashBlade fits into both hybrid cloud and multicloud architectures by replicating data to a variety of public cloud repositories.
- **Cost efficiency.** FlashBlade takes a fraction of the space and environmental requirements of traditional spinning disk systems, saving money on deployment and maintenance costs.

FlashBlade is unique in its design to offer the performance of primary flash storage at the price and expandability of scale-out arrays. Pure Storage has also taken an interesting twist in its architecture for data protection. While many systems seek to reduce the backup window and optimize storage for backup, FlashBlade is designed to optimize data restore, which is the point at which organizations truly need the best performance.

## CHALLENGES/OPPORTUNITIES

---

In surveys conducted by IDC, organizations that opted not to use flash or to stop using flash for storage cited cost, lack of perceived benefits, and reliability issues as among the top reasons. While ongoing surveys indicate that these issues have been alleviated, Pure Storage and other vendors in the space still need to clearly communicate the benefits of their flash-based storage products. As a recent entrant into the secondary storage market, FlashBlade will need to prove its mettle in terms of performance and efficiencies designed to lower costs.

Yet data growth continues unabated. Whether structured or semi-structured, data is the new basis of competitive advantage. Storage is integral to an organization's data strategy as it is an active contributor to the process of reducing the cost and complexity associated with massive amounts of data. More importantly, data availability is foundational to digital transformation and to becoming a data-driven organization. Consequently, the need for data protection in varied application deployment models is driving IT organizations to buy best-in-class data protection solutions for specific environments.

## CONCLUSION

---

Data-driven organizations are moving away from siloed backup/recovery systems and realize that rapid data recovery and putting data to work are the key goals. Organizations are also realizing that buying single-purpose appliances, which house backup data simply waiting for something bad to happen, is an inefficient use of resources. Instead, devices that can help the organization simultaneously protect and utilize data offer better value for the money spent.

Secondary storage is a category gaining attention and traction in the market for its ability to deliver both data protection and secondary-use capabilities. Nevertheless, some systems utilize slower spinning disks for secondary systems. These systems take up more space and require more power and cooling than all-flash systems and can't deliver the QoS guarantees of all-flash systems. We believe that data-driven organizations will be less willing to make performance compromises for secondary use cases and that Pure Storage has positioned FlashBlade correctly to take advantage of this trend and that the company is at the forefront of flash-based secondary storage systems.

## About IDC

International Data Corporation (IDC) is the premier global provider of market intelligence, advisory services, and events for the information technology, telecommunications and consumer technology markets. IDC helps IT professionals, business executives, and the investment community make fact-based decisions on technology purchases and business strategy. More than 1,100 IDC analysts provide global, regional, and local expertise on technology and industry opportunities and trends in over 110 countries worldwide. For 50 years, IDC has provided strategic insights to help our clients achieve their key business objectives. IDC is a subsidiary of IDG, the world's leading technology media, research, and events company.

## Global Headquarters

5 Speen Street  
Framingham, MA 01701  
USA  
508.872.8200  
Twitter: @IDC  
idc-community.com  
www.idc.com

---

### Copyright Notice

External Publication of IDC Information and Data – Any IDC information that is to be used in advertising, press releases, or promotional materials requires prior written approval from the appropriate IDC Vice President or Country Manager. A draft of the proposed document should accompany any such request. IDC reserves the right to deny approval of external usage for any reason.

Copyright 2018 IDC. Reproduction without written permission is completely forbidden.

