

Brought to you by:



Unified Fast File & Object (UFFO) Storage

for
dummies[®]
A Wiley Brand

Understand the
evolving data landscape



Meet the demands of
modern unstructured data



Optimize your storage
infrastructure



Lawrence C. Miller

Erin Poulson

Pure Storage Edition

About Pure Storage

Pure Storage gives technologists their time back. Pure delivers a modern data experience that empowers organizations to run their operations as a true, automated, storage as-a-service model seamlessly across multiple clouds. Pure helps customers put data to use while reducing the complexity and expense of managing the infrastructure behind it. And with a certified customer satisfaction score in the top one percent of B2B companies, Pure's ever-expanding list of customers are among the happiest in the world.

www.purestorage.com.



Unified Fast File & Object (UFFO) Storage

Pure Storage Edition

**by Lawrence C. Miller
and Erin Poulson**

for
dummies[®]
A Wiley Brand

Unified Fast File & Object Storage For Dummies®, Pure Storage Edition

Published by
John Wiley & Sons, Inc.
111 River St.
Hoboken, NJ 07030-5774
www.wiley.com

Copyright © 2022 by John Wiley & Sons, Inc.

No part of this publication may be reproduced, stored in a retrieval system or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, scanning or otherwise, except as permitted under Sections 107 or 108 of the 1976 United States Copyright Act, without the prior written permission of the Publisher. Requests to the Publisher for permission should be addressed to the Permissions Department, John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030, (201) 748-6011, fax (201) 748-6008, or online at <http://www.wiley.com/go/permissions>.

Trademarks: Wiley, For Dummies, the Dummies Man logo, The Dummies Way, Dummies.com, Making Everything Easier, and related trade dress are trademarks or registered trademarks of John Wiley & Sons, Inc. and/or its affiliates in the United States and other countries, and may not be used without written permission. Pure Storage and the Pure Storage logo are registered trademarks of Pure Storage. All other trademarks are the property of their respective owners. John Wiley & Sons, Inc., is not associated with any product or vendor mentioned in this book.

LIMIT OF LIABILITY/DISCLAIMER OF WARRANTY: WHILE THE PUBLISHER AND AUTHORS HAVE USED THEIR BEST EFFORTS IN PREPARING THIS WORK, THEY MAKE NO REPRESENTATIONS OR WARRANTIES WITH RESPECT TO THE ACCURACY OR COMPLETENESS OF THE CONTENTS OF THIS WORK AND SPECIFICALLY DISCLAIM ALL WARRANTIES, INCLUDING WITHOUT LIMITATION ANY IMPLIED WARRANTIES OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. NO WARRANTY MAY BE CREATED OR EXTENDED BY SALES REPRESENTATIVES, WRITTEN SALES MATERIALS OR PROMOTIONAL STATEMENTS FOR THIS WORK. THE FACT THAT AN ORGANIZATION, WEBSITE, OR PRODUCT IS REFERRED TO IN THIS WORK AS A CITATION AND/OR POTENTIAL SOURCE OF FURTHER INFORMATION DOES NOT MEAN THAT THE PUBLISHER AND AUTHORS ENDORSE THE INFORMATION OR SERVICES THE ORGANIZATION, WEBSITE, OR PRODUCT MAY PROVIDE OR RECOMMENDATIONS IT MAY MAKE. THIS WORK IS SOLD WITH THE UNDERSTANDING THAT THE PUBLISHER IS NOT ENGAGED IN RENDERING PROFESSIONAL SERVICES. THE ADVICE AND STRATEGIES CONTAINED HEREIN MAY NOT BE SUITABLE FOR YOUR SITUATION. YOU SHOULD CONSULT WITH A SPECIALIST WHERE APPROPRIATE. FURTHER, READERS SHOULD BE AWARE THAT WEBSITES LISTED IN THIS WORK MAY HAVE CHANGED OR DISAPPEARED BETWEEN WHEN THIS WORK WAS WRITTEN AND WHEN IT IS READ. NEITHER THE PUBLISHER NOR AUTHORS SHALL BE LIABLE FOR ANY LOSS OF PROFIT OR ANY OTHER COMMERCIAL DAMAGES, INCLUDING BUT NOT LIMITED TO SPECIAL, INCIDENTAL, CONSEQUENTIAL, OR OTHER DAMAGES.

For general information on our other products and services, or how to create a custom *For Dummies* book for your business or organization, please contact our Business Development Department in the U.S. at 877-409-4177, contact info@dummies.biz, or visit www.wiley.com/go/custompub. For information about licensing the *For Dummies* brand for products or services, contact BrandedRights&Licenses@Wiley.com.

ISBN: 978-1-119-79599-5 (pbk); ISBN: 978-1-119-79600-8 (ebk)

Publisher's Acknowledgments

Some of the people who helped bring this book to market include the following:

**Project Manager and
Development Editor:**
Carrie Burchfield-Leighton
Sr. Managing Editor: Rev Mengle

Acquisitions Editor: Ashley Coffey
**Business Development
Representative:** Molly Daugherty

Table of Contents

INTRODUCTION	1
About This Book	1
Foolish Assumptions.....	2
Icons Used in This Book.....	2
Beyond the Book.....	3
CHAPTER 1: Surveying the Modern Data Landscape.....	5
Explaining the Evolution of Modern Unstructured Data	5
Understanding the Limitations of Legacy Storage Infrastructure.....	7
Recognizing Key Market Drivers and Trends	10
CHAPTER 2: Exploring UFFO Storage and How Pure Storage Can Help.....	13
The Need for UFFO Storage	13
Delivering the Key Requirements of Modern Data Storage.....	15
Introducing Pure Storage FlashBlade	16
CHAPTER 3: Examining UFFO Use Cases and Success Stories.....	19
Applying Advanced Data Analytics	19
Modernizing Data Protection.....	21
Electronic Design Automation	24
High-Performance Computing.....	25
AI.....	27
CHAPTER 4: Ten Reasons You Need a UFFO Storage Platform	29
Manages Data Growth	29
Accelerates Modern Data and Applications.....	30
Enables Better and Faster Insights.....	30
Eliminates Data Silos.....	31
Delivers a Cloud-Like Experience	31
Supports Fast Object Use Cases.....	31
Improves Data Protection and Business Continuity	32
Eliminates Downtime.....	32
Maximizes ROI	33
Increases Operational Efficiency	33

Introduction

Once upon a time, cellphones were just cellphones. They stored contacts, made calls, and sent texts. Basic cellphones did the trick. Today, we ask a lot more from our phones. We want them to give us directions, take photos and video, browse the web, act as our wallets, and stream television shows. Simple flip phones can't do that, but smartphones can — by consolidating a ton of features into one powerful device.

A similar evolution is happening in the world of data storage. According to many estimates, 80 percent of global data will be unstructured by 2025. This data isn't the structured data of the past or the kind of data that legacy storage hardware can handle. Unstructured data includes millions of images, audio, and video files created every day. It's the telemetry and streaming data from smart appliances, cars, and Internet of Things (IoT) sensors.

To store, process, and analyze all this data — hello, digital transformation — you have to ask more from data storage. This demand has created a new category: unified fast file and object (UFFO). Think of UFFO as the smartphone of storage — powerful, efficient, simple, and modern — making a range of legacy storage infrastructure obsolete. And, like smartphones, UFFO will be something organizations can't live without.

About This Book

In this book, you get an idea of the vast potential for unstructured data — and also the infrastructure challenges that come with wrangling it. You discover why UFFO is the answer to addressing modern application and data storage requirements. Then, we give you some examples of how other organizations are utilizing UFFO to capitalize on modern unstructured data and how you can, too. Finally, you look at Pure Storage's UFFO storage platform: Pure Storage FlashBlade.

Unified Fast File & Object Storage For Dummies, Pure Storage Edition, consists of four chapters that explore the following:

- » The modern data landscape and the limitations and challenges of legacy storage infrastructure (Chapter 1)

- » Why UFFO meets modern data storage requirements, and how Pure Storage FlashBlade can work for you (Chapter 2)
- » Common UFFO use cases and customer success stories (Chapter 3)
- » Ten reasons your organization needs a UFFO storage platform (Chapter 4)

Each chapter is written to stand on its own, so if you see a topic that piques your interest, feel free to jump ahead to that chapter. You can read this book in any order that suits you (although we don't recommend upside down or backwards).

Foolish Assumptions

It's been said that most assumptions have outlived their usefulness, but we assume a few things nonetheless! For this book, we assume that you have at least a basic understanding of business and storage concepts and the intersection of the two. If you're in a role that has anything to do with storage digital transformation, unstructured data, modern applications, and business/technology, this book is for you.

Icons Used in This Book

Throughout this book, we use special icons occasionally to call attention to important information. Here's what to expect:



REMEMBER

This icon points out the information you should commit to your nonvolatile memory or your noggin.



TECHNICAL
STUFF

If you seek to attain the seventh level of NERD-vana, perk up! This icon explains the jargon beneath the jargon.



TIP

Tips are appreciated, but never expected — and we sure hope you appreciate these useful nuggets of information.



WARNING

These alerts point out the stuff your mother warned you about (well, probably not), but they do offer practical advice.



CASE STUDY

The information alongside these icons gives you examples of successful UFFO storage customer stories.

Beyond the Book

There's only so much we can cover in this short book, so if you find yourself at the end of this book wondering, "Where can I learn more?", check out the following resources:

- » **Website:** The Pure Storage website has many white papers, technical briefs, analyst reports, and more. To find out more about UFFO storage, check out www.purestorage.com/uffo. For details on Pure's UFFO platform, head to www.purestorage.com/flashblade.
- » **Blog:** The Pure Storage blog is full of articles about current topics, events, and news. Read more at blog.purestorage.com.

IN THIS CHAPTER

- » Understanding the growth and value of modern unstructured data
- » Recognizing the limits of legacy storage architectures
- » Looking at market trends and drivers

Chapter 1

Surveying the Modern Data Landscape

With the emergence of technologies such as Internet of Things (IoT), artificial intelligence (AI), and data analytics, unstructured data in the form of files and objects has grown exponentially. Legacy storage architectures make it difficult to manage and access this data, and evolving market trends and drivers are rapidly redefining needs and reshaping storage infrastructure requirements across all industries. In this chapter, you discover why modern data challenges require a new approach to storage and why your organization needs a unified fast file and object (UFFO) storage platform.

Explaining the Evolution of Modern Unstructured Data

Digital transformation is revolutionizing the way you work. *Digital transformation* refers to the process of using technology to rethink, streamline, and automate traditional business operations through digital methods. From human resources, accounting, marketing, operations, and IT to the way you manage the data that powers all those functions, virtually every area of your business can benefit from digital transformation.

It can be as simple as converting all your paper customer files to digital files and managing them through an online sales and marketing platform. Or it can be as complex as designing cloud-native applications from the ground up to reach customers who increasingly expect easy, personalized, anytime-anywhere engagement with your business from any device.

Digital transformation means different things to different people, but its essence lies in the power of leveraging new technologies to radically boost performance and business agility. There are some recurring themes of what exactly that technology is doing to bring about digital transformation. Common elements of digital transformation include

- » Automation
- » Cloud integration
- » Virtualization and containerization
- » Optimized user interfaces (UI) and user experiences (UX)
- » Data analytics
- » Infrastructure consolidation

As advanced technologies and tools, such as AI, machine learning (ML), data analytics, IoT, data protection, and ransomware mitigation and recovery, have become increasingly common, the amount of data businesses must collect, store, analyze, and manage has exploded. Most of this data is unstructured and brings with it the need for a new approach to storage, one that legacy storage infrastructure can't provide.

Unlike structured data, such as Excel files or SQL databases, unstructured data, in the form of files and objects, doesn't fit neatly into formatted tables. This data includes

- » Text and documents that require context from which to process and extract data, such as notes from a customer service rep in a call center
- » Audio data
- » Visual data, such as images and video
- » Rich data, such as weather data and spatial analysis data
- » IoT data, like sensor data, ticker info, and more

- » Data generated by social media activity, including user activity, sentiment analysis of comments, ad clicks, and demographics
- » Device and network data, such as telemetry and location data

The challenges in storing, managing, and analyzing unstructured data come from, well, its lack of structure. Data from disparate applications and systems can often be stored in siloed locations across your infrastructure. The information isn't managed or accessed in any unified way. By nature, it's difficult to know what's relevant. It can sometimes even be difficult to find.

But here's the thing: Finding a way to efficiently store, manage, access, and use that unstructured data could be the *very* key to whether your business lives or dies in today's super-competitive landscape. That data contains immense value. When it's analyzed and correlated, it can unlock surprising insights into customer preferences and behavior, production bottlenecks, hidden product design flaws, ineffective sales methods, workflow and process issues, and so much more.

Organizations are using those insights to redefine customer service, create new revenue streams, attract and retain new target audiences, streamline processes, and ultimately gain an edge and differentiate their offerings from the other guys.

Getting maximum value out of unstructured data is all about getting deep into its many layers. Fast, efficient analysis is the key. That means consolidating data from multiple sources and locations. Legacy storage systems just can't deliver the flexibility, speed of access, or easy, unified management you need to adequately handle unstructured data and make it deliver the necessary insights.

Understanding the Limitations of Legacy Storage Infrastructure

Storing, searching, and analyzing unstructured data — and extracting value from it — is much more complex than working with structured data. Legacy storage infrastructure was largely designed for efficiently storing data but not for addressing the demands of modern unstructured data, and it wasn't equipped to handle the massive increase in data volume. That means legacy systems are unable to simultaneously offer easy scalability *and* the

fast, flexible search and access capabilities businesses need today. At best, they can do one or the other, and not always well, at that.

Legacy storage infrastructure for unstructured data typically falls into one of two categories:

- » Very slow but highly scalable — mainly for archive or “cold” storage of data that’s rarely (or never) accessed
- » Fast but intended for delivery of smaller data sets — mainly production workloads that require speedy access

Typically, organizations have deployed a siloed legacy storage infrastructure for similar types of applications or use cases that need to be supported. For example, one silo may be for historical business data (such as analytics), another for live application data (such as transactional data), and another for archiving information (such as check images or historical customer statements). Backups are another type of unstructured data silo.

Each of these applications has different performance requirements, and each silo is effective for addressing that specific type of application and workload using a variety of compute and storage architectures. For example, some data may be stored on servers or in direct-attached storage (DAS) and some applications may use scale-out network-attached storage (NAS). Generally, any application using object storage has its own dedicated storage system, shown in Figure 1-1, and the backup environment typically has its own backup appliance. Many of these storage silos can only offer sluggish performance because they’re filled with unstructured data that is stored “just in case” and is unlikely to see the light of day again.

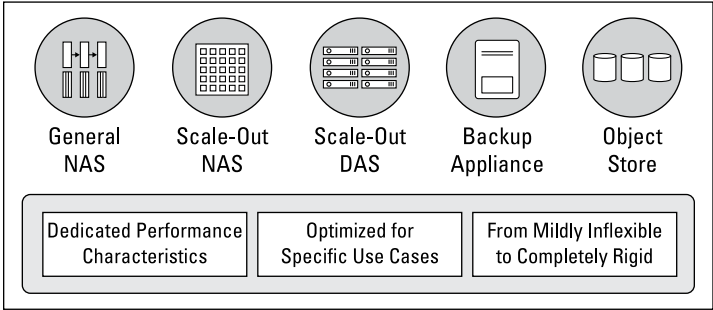


FIGURE 1-1: Each use case requires its own dedicated storage silo in a legacy infrastructure approach.



WARNING

These legacy data silos have become a problem for organizations. Even when organizations standardize their architecture with a single vendor, each silo is custom-made and unrelated to the others. They don't share or take into account data residing across silos. That makes them extremely difficult to manage and maintain — it can take an entire team of storage admins to deal with all the different silos. Legacy storage infrastructures also present other issues:

- » **Stranded capacity:** An organization may have 100 terabytes (TB) of extra capacity in its object platform but may be running out of space on its file platform. That 100 TB of extra capacity on the object platform can't be used to address the shortage on the file platform.
- » **Not fast enough:** Limited data sharing and reuse becomes increasingly problematic for organizations that need to correlate data in real time. Legacy storage infrastructure is often single-purpose and too slow for modern data and applications.
- » **Siloed data:** Silos with server-based storage require compute resources to be added every time storage capacity is added. This requirement can be expensive and inefficient, given that the extra compute is often not needed.

While some legacy storage solutions can deliver high performance for a specific type of file data, no single legacy system can deliver the high performance and scalability organizations need for a diverse set of data and applications.

These days, most organizations recognize the problems of siloed storage infrastructure and how it limits the value they can extract from that data. Simply attempting to consolidate data on legacy infrastructure, however, isn't the answer, because

- » Separate NAS silos have been optimized for specific workloads, and meeting the performance requirements for different workloads is challenging with existing storage infrastructure.
- » Different protocols are used in individual silos, such as Network File System (NFS) for Linux systems, Server Message

Block (SMB) for Windows systems, and Amazon Simple Storage Service (S3) for cloud-based object storage. Legacy storage infrastructure doesn't support all these protocols natively.

- » Storage management within and across different silos is complex.
- » Forklift upgrades are costly and time-consuming, requiring unacceptable downtime for the business.

Recognizing Key Market Drivers and Trends

Three critical data trends drive the need for a new approach to data storage:

- » Organizations are generating more unstructured data than ever before.
- » Storage admins need a simple, efficient way to manage that data.
- » Data search and access must be fast and easy to allow organizations to extract real-time insights for competitive advantage.

Organizations need a different kind of infrastructure. The evolution of modern data delivery is driving these business trends:

- » **Technology:** Infrastructure needs to meet the requirements of all workloads, including modern applications and large volumes of unstructured data.
- » **Cultural:** Storage admins need a consolidated storage platform for less complexity and improved ease of management.
- » **Organizational:** Businesses need infrastructure that can adequately address the storage and data access needs of today while also preparing for future capabilities.

These business and data trends are shown in Figure 1-2.

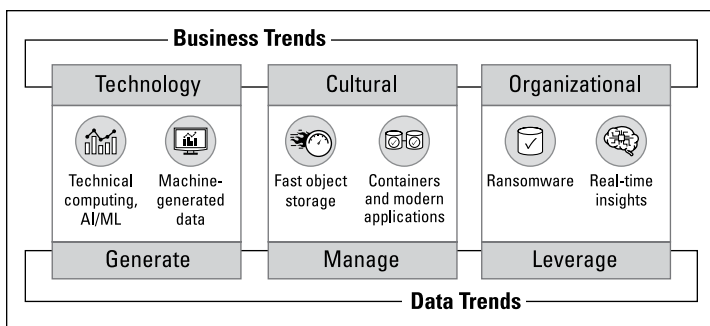


FIGURE 1-2: Key market drivers for modern data delivery.

Advanced applications promise to lead to exciting breakthroughs in product development, new revenue models, more personalized customer service, and more. The trick, however, is finding the right platform for your unstructured data, which generally makes up more than two-thirds of an organization’s entire data, which builds an infrastructure that allows you to leverage your data efficiently and discover innovative insights from it.

File storage has typically been associated with high-performance unstructured workloads. While object storage has traditionally been treated as a convenient way to archive large amounts of data, the growth in cloud-native apps that use object storage as their default storage has created a new market for higher-performing object storage. Fast object storage combines the ease of access and scale-out convenience of traditional object storage with the speed and performance needed to support the demanding workloads of cloud-native applications. The current market trends mean that modern storage needs to deliver *both* in a single platform.



REMEMBER

Today’s organizations need a storage solution that addresses modern data requirements and provides the simplicity and multi-dimensional performance required to enable the consolidation of key unstructured data workloads. A unified fast file and object (UFFO) storage platform fills that need, and we dive deeper into UFFO in Chapter 2.

IN THIS CHAPTER

- » Getting a UFFO storage primer
- » Examining how UFFO meets key requirements for modern data storage
- » Consolidating fast file and object workloads with FlashBlade

Chapter 2

Exploring UFFO Storage and How Pure Storage Can Help

In this chapter, you discover what unified fast file and object (UFFO) storage is and how it meets organizations' modern data storage needs. You find out about the key characteristics of UFFO storage and then take a look at how Pure Storage FlashBlade delivers a UFFO storage platform that can work for you.

The Need for UFFO Storage

Decades ago, it was virtually impossible for an organization to correlate all its unstructured data, gathered from so many disparate sources, and extract any actionable insights. It was simply beyond human capacity, but today's technology is changing that.



REMEMBER

UFFO is a new category of data storage, much like a smartphone is a category of mobile devices. A comparison is shown in Figure 2-1. UFFO consolidates a range of traditionally separate storage architectures based on file and object protocols into a single high-performance scale-out storage platform. It integrates the speed and

searchability of file storage with the supreme scalability of object storage. With UFFO, you can eliminate siloed data warehouses as well as all network-attached storage (NAS), traditional object stores, and direct-attached storage (DAS).

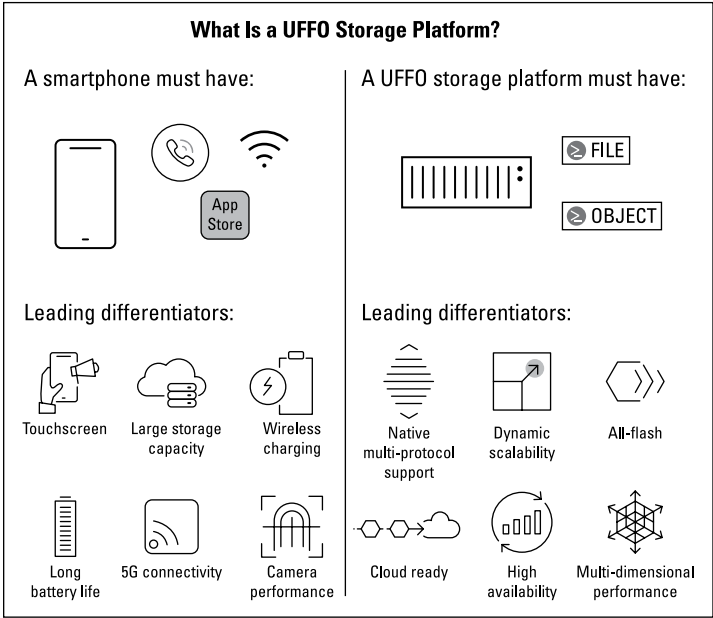


FIGURE 2-1: A view of a UFFO storage platform.

UFFO provides the right balance between performance, simplicity, and consolidation for unstructured data management and storage for today’s most demanding applications, including large-scale analytics and artificial intelligence (AI) and machine learning (ML) workloads. A UFFO storage platform can match the processing power of the most advanced chipsets to deliver maximum throughput.

With UFFO, you get multi-dimensional performance with high throughput, regardless of workload, file size, protocol, or file and object counts. The unified platform easily scales to support multiple applications, unlike single-use silos that separate file and object data. The platform also allows you to fine-tune utilization by decoupling compute and storage and allowing them to be scaled independently on demand.



File storage is the data storage format most people are familiar with — data is stored and managed as files in folders that are organized within a hierarchical file directory. Every time you access files on a hard drive, your local computer, or a shared drive hosted on a server, you're dealing with file storage. *Object storage* is a data storage architecture in which data is stored and managed as discrete, self-contained units called objects. Each object contains a unique identifier or key, which allows it to be found no matter where it's stored on a distributed system.

Delivering the Key Requirements of Modern Data Storage

To deliver the key requirements of modern data, a UFFO storage platform must have the following characteristics:

- » **Multi-dimensional performance:** This means high throughput and performance to meet the demands of any type of workload, including workloads with any file size (large or small); sequential or random input/output (I/O); batched or real-time jobs; and large numbers of files.
- » **Intelligent architecture:** UFFO is built for and maximizes the efficiencies of flash technology with simple deployment, management, and upgrades, and no need for constant tuning.
- » **Cloud-ready:** The platform has cloud-like agility, flexibility, and consumption choices that include on-premises control.
- » **Dynamic scalability:** UFFO delivers seamless scaling of capacity, performance, metadata, numbers of files, and objects.
- » **Always available:** This goes beyond traditional platform resiliency with a unique software design that enables non-disruptive upgrades and advanced data protection capabilities.
- » **Native, multi-protocol support:** This provides outstanding performance and functionality without compromise for either file or object storage protocols.

With these inherent characteristics, UFFO storage can be the smart foundation of your modern applications, which need to be

- »» Extremely flexible and dynamic and optimized to run multiple workloads
- »» Distributed and easily scalable across your entire infrastructure
- »» Automated and orchestrated through application programming interfaces (APIs)
- »» Resilient to always-evolving security threats and unexpected downtime due to outages
- »» Developed to process and use data in real time



REMEMBER

The most important takeaway here is that a UFFO storage platform enables your advanced applications and analytics solutions that turn unstructured data into more than a load of information that's simply archived for compliance reasons or an unexamined by-product of an IoT device. UFFO storage empowers you to derive real-time insights that can truly transform your business.

Introducing Pure Storage FlashBlade

Architected to address the demands of modern applications and modern data, Pure Storage FlashBlade delivers the simplicity and multi-dimensional performance that enables the consolidation of key unstructured data workloads on a single platform.

FlashBlade is different from traditional unstructured data storage systems that were either scalable and slow or fast but only meant for small data sets. FlashBlade's architecture is massively parallel to concurrently support tens of thousands of clients and built for flash, for the highest performance, and it manages both file and object data without a hassle. Scaling out for improved capacity and performance is a matter of simply adding more blades.

With FlashBlade, you get powerful support for data-intensive analytics and data protection use cases, such as rapid restore and ransomware remediation. You also get support for a wide variety of other modern use cases, including AI/ML, cloud-native applications and architectures, and more.

FlashBlade’s unique blade design includes integrated compute and capacity for storage. The solution’s software is designed for immense but simple scalability. And the solution is powered by an innovative approach to networking powered by a converged fabric that consolidates front-end, back-end, and control networks and delivers an aggregate bandwidth of up to 1.5 terabytes (TB) per second. All of this allows you to modernize your applications, meet the always-evolving demands of today’s DevOps ecosystems, and eliminate bottlenecks for faster time-to-market and greater return on investment (ROI).



REMEMBER

Like a smartphone is a specific category of mobile phones, UFFO is a defined category of storage. Pure Storage FlashBlade is an optimal choice as a UFFO storage platform. Just like realizing all your smartphone needs can come from a single mobile device, the same is true when deploying FlashBlade for your unstructured storage needs. For a further breakdown of this analogy, check out Figure 2-2.

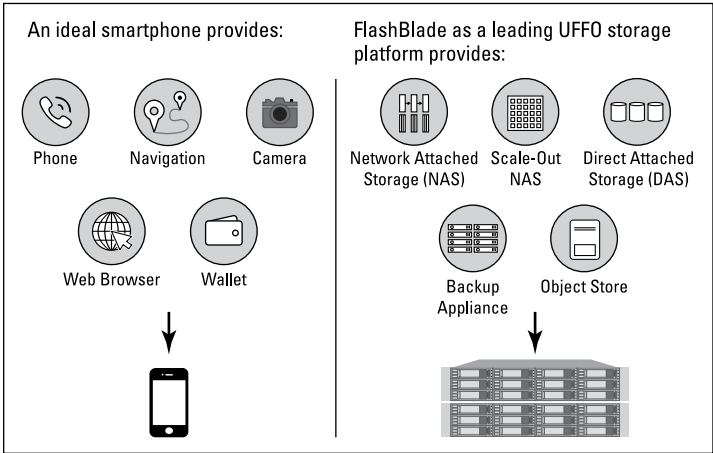


FIGURE 2-2: FlashBlade offers a consolidated storage platform.

FlashBlade delivers cloud-like simplicity and agility with consistent high performance and control. The primary features of FlashBlade include

- » **All-flash performance:** FlashBlade fast file and object storage goes beyond traditional scale-out NAS. It delivers massive throughput and parallelism with consistent multi-dimensional performance.

- » **Agile scale-out architecture:** FlashBlade's scale-out metadata architecture handles tens of billions of files and objects with maximum performance and rich data services. You can also add blades to scale capacity and performance up to 150 blades and support tens of thousands of attached compute cores.
- » **Simplified workload consolidation:** Simplify storage management with Pure1, the AI-driven cloud-based management platform. Deploying, updating, and managing FlashBlade is simple with automated APIs. You get high-performance native NFS, SMB, and S3 protocol support for all your modern file and object workloads.
- » **Modern enterprise-ready availability and uptime:** Enterprise features such as dynamic global erasure coding protect against concurrent blade failures, enable faster rebuild compared to the rigid traditional RAID, and proactively perform data scrubbing to minimize the threat of data loss.
- » **Support for cloud mobility, disaster recovery and easy updates:** Object replication helps support cloud mobility and file replication supports disaster recovery. Non-disruptive updates help you expand flash capacity, upgrade controllers, replace failed components, and update software without downtime.

Check out Chapter 3 for a peek at FlashBlade in action.

IN THIS CHAPTER

- » Gaining deep insights with analytics
- » Ensuring your data is protected
- » Supporting advanced chip design
- » Serving up data for high-performance computing
- » Using artificial intelligence to streamline business processes

Chapter 3

Examining UFFO Use Cases and Success Stories

In this chapter, you explore some of the most promising, exciting use cases for unified fast file and object (UFFO) storage to give you a sense of how tapping unstructured data can drive your business forward. We also share a few success stories from Pure Storage customers along the way.

Applying Advanced Data Analytics

Next-level intelligence gleaned from the analysis of unstructured data is driving some incredible breakthroughs across industries — from retail and healthcare to logistics and entertainment. These data-driven, immersive experiences are must-haves for the modern consumer. But analytics pipelines are too often starved by data that gets isolated in silos created by traditional, inflexible infrastructures.

Data warehouses were built to scale as a single source of data for analytics. But, when they're built on an inflexible scale-up architecture, they introduce another limitation. Flexibility becomes constrained by the need for forklift upgrades in order to expand capacity. Data lakes have introduced a scale-out model, too, offering more elastic scalability to meet the challenge of efficiently storing massive amounts of unstructured data. But these architectures are also limiting because they restrict the ability to deploy compute and storage resources separately. Both approaches are limited by the fact that they may not have the performance needed for real-time access to multiple analytics applications.

Modern, cloud-native applications are multi-protocol and built around a highly distributed and decentralized scale-out microservices architecture. While a data lake may meet the capacity requirements for these applications, a modern infrastructure also needs to address consistent performance with a high variety of data profiles. You need to consolidate data from multiple sources to fully leverage analytics without having to duplicate, move, and manage data across multiple silos, leading to greater inefficiencies.



REMEMBER

Data lakes and warehouses can store massive amounts of unstructured data, but these architectures can limit flexibility. A UFFO storage platform allows organizations to scale-out compute and storage independently of each other to support modern analytics pipelines and gain better insights. It also features highly parallel, multi-dimensional performance to offer more consistent speed at any scale. You can confidently adapt to the growth of new services, data, and users.



CASE STUDY

PURE PAYS BIG DIVIDENDS FOR A FINANCIAL SERVICES FIRM

Interesting things can happen when the worlds of fast-changing information technology and stability-conscious banking intersect. A major banking, investment, and insurance firm operating more than 130 branches and 1,200 automated teller machines (ATMs) was an early adopter of Pure Storage FlashBlade and UFFO storage. Business-critical applications at the firm run on Pure arrays, including an

in-house developed e-commerce suite, proprietary applications for bank tellers and ATMs, back-end transaction processing, PostgreSQL, Oracle and SQL databases, and Microsoft Exchange.

Consolidating multiple workloads

A recent addition to the firm's workload portfolio has come from a sharp increase in the use of Splunk log collection and analysis. At first, the e-commerce team was uploading all their logs for troubleshooting. The firm's security department soon became the dominant user of Splunk, which is now the primary tool for detecting any anomalies in their environment, ingesting around 600 gigabytes (GB) of data a day. The IT department also uses Splunk to perform health checks and troubleshoot problems in its infrastructure. Splunk on FlashBlade returns answers quickly, which benefits the infrastructure team.

Ease of use and availability

Pure eliminated interruptions and complex product updates with non-disruptive upgrades (NDUs). Pure replaced the aging legacy storage model with a subscription-based storage experience that keeps data in place as customers grow and stay modern. Pure's Evergreen Storage program combines a 100 percent non-disruptive product architecture with a flexible buying program. This combination is designed to free customers from legacy storage and complex, risky, and expensive forklift upgrades and data migrations every three to five years.

Modernizing Data Protection

These days, data protection has to be an organization's top priority. The current climate combined with the value of data to an organization makes *not* having the latest technology to safeguard data against loss, corruption, or security threats too risky. From ransomware mitigation and recovery, backup and restore, and data reuse, data protection should be a holistic approach to keeping data secure and highly available for its products, services, and operations.

The traditional approach focuses on redundancy: You can't afford to have your data lost, corrupted, or compromised, so you keep

reliable backup copies of your data. A modern approach is more proactive, going further to include

- » A platform that addresses the challenge of storage-management complexity
- » Innovative data storage technologies to enable high-throughput, low-latency data transfer
- » Predictive analytics to identify and remediate data storage issues before they occur
- » The ability to rapidly restore data in the event of a ransomware attack

With businesses more data dependent than ever before and with encryption ransomware attacks on systems and data becoming common events, being able to recover systems and data at speed and scale are more important than ever.

Traditional disk-based legacy storage systems may be more cost effective, but when hundreds of terabytes or petabytes need to be restored as quickly as possible, you could be faced with weeks or months before your data is fully restored.



REMEMBER

By incorporating a UFFO platform, built on flash, together with a data protection application, you can achieve faster writes and reads — resulting in high-speed restores at scale.

But data protection is also (and perhaps, most importantly) about meeting the service-level agreements (SLAs) required by your organization. Your recovery point objectives (RPOs) — the maximum acceptable period of time in which data may be lost — and recovery time objectives (RTOs) — the maximum acceptable period of time it takes to recover data — are critical to your business.

How do you determine RPO and RTO? Say your finance department is willing to lose a day's worth of data in the event of a ransomware attack, but can only afford to be offline for 4 hours, then your RPO is 24 hours, and your RTO is 4 hours. To achieve your RPO and RTO for the finance department, you need to back up your financial data once a day, and your data protection technology must be able to fully restore your finance data within 4 hours.



REMEMBER

Most organizations have an SLA gap; only 20 percent of organizations believe they can meet their recovery objectives. Pure FlashBlade delivers petabyte-scale recovery performance of 270 terabytes (TB) per hour, which enables your organization to restore more than 6 petabytes (PB) of data a day. And with immutable SafeMode snapshots, backup data can be restored safely and efficiently in the event of a ransomware attack.



CASE STUDY

LOCKING DOWN AND SAFEGUARDING A CITY'S DATA

It was 5:00 a.m., and the Director of Operations for IT and Innovation of a major city in North America woke up to an alarm of a different kind. He'd been alerted to suspicious server logins. It was the first clue that the city was experiencing a cyberattack. Within hours, city servers and computers were shut down to mitigate damage, and the city declared a state of emergency. The IT team began to cleanse the city's data while replacing its outdated storage infrastructure with Pure Storage to get systems back online quickly so citizens could access safely and securely the services they needed.

Ransomware attack disrupts citizen services

On the day of the attack, city IT officials shut down 470 servers and virtual machines, and thousands of computers, bringing the city to a halt. This quick thinking stopped the malware from spreading and encrypting all the city's data — ultimately preventing the attackers from extorting a ransom. But it also took key services, from paying taxes to routine court procedures, offline. Over 4,000 city employees and nearly 400,000 residents were impacted.

Secure, tamper-proof backups provide peace of mind

The city was already looking to update its storage when the attack took place. The decision to replace the entire storage footprint with Pure was a simple one. Pure Storage FlashArray and FlashBlade provide both primary storage and disaster recovery, along with data protection software from Veeam. FlashArray replicates data to a second data center while FlashBlade, with the SafeMode snapshots feature, protects the city's data backups from further ransomware attacks with immutable copies and early-detection capabilities.

With Pure, the city can now retrieve data quickly and safely with minimal disruption to services.

Electronic Design Automation

Semiconductor chips are what make the devices we use every day possible. From smartphones and wearables to smart home automation, there's an ever-increasing demand for products that are faster, cheaper, smaller, and have longer battery life. These requirements push chip designers to pack more functionality into silicon and move to more advanced technology nodes to manage power and costs. These changes have led to huge increases in product complexity, which demands more analysis before designs can be sent to manufacturing. Compounding these challenges are shrinking development schedules that continue with only a modest increase (if any) in resources and headcount.



CASE STUDY

PURE ACCELERATES EDA

A global technology company that designs and manufactures semi-conductors and devices to power Internet of Things (IoT) infrastructures, industrial automation, and consumer and automotive markets worldwide was getting held back. The culprit? An aging processing infrastructure. The costs were high, and performance was suffering. It was time for a new solution.

Getting the fastest chips to market as fast as possible

The company needed faster development to accelerate time-to-market for its newest chipset. It also wanted a better price and performance solution to support its heavy metadata workloads. Today, FlashBlade delivers true linear performance for the customer's EDA workloads that's more than 2.7 times faster than its existing infrastructure from scratch. Rapid File tools are almost 200 times faster. FlashBlade delivered significantly lower latency for all the company's users, with overall job times improving by at least 40 percent.

Reducing costs and avoiding forklift upgrades for the long haul

With FlashBlade as a UFFO storage platform enabling consolidation, the company is now consuming less than a third of its previous physical data center footprint, which has resulted in huge savings. The customer views its FlashBlade deployment as a ten-year investment that will help avoid costly and high-risk forklift upgrades.

For chip developers to leverage thousands of powerful servers and extremely fast networks, modern electronic design automation (EDA) workflows demand storage with much higher performance. This includes throughput, Input/Output Operations Per Second (IOPS), latency, fast deletes, and capacity. EDA development environments also tend to have high levels of concurrency and parallelism. Compute environments often have tens to hundreds of thousands of cores. Many jobs access the same project folders and perform huge amounts of concurrent I/O operations. In these circumstances, storage can often become a performance bottleneck with a significant impact on EDA job run times. Legacy storage architectures simply can't keep up.



REMEMBER

A critical capability of storage in EDA and high-performance computing environments is the ability to scale storage performance linearly to mitigate the effect of storage-related slowdowns. Additionally, EDA workloads are known to be metadata intensive in the front-end part of the design cycle. For advanced node designs, there are massive increases in throughput requirements for the back-end phase, as well.

High-Performance Computing

High-performance computing (HPC) uses clusters of high-performance computers, or *supercomputers*, to perform computationally intensive operations. Data volumes are exploding in HPC applications, and the demand for accelerated access to fuel chip development and artificial intelligence (AI) and machine learning (ML) workloads are at a trajectory never experienced before in the modern era. This trend has led to massive data bottlenecks and delayed product development times.

FLASHBLADE EXPANDS SUPERCOMPUTING POWER



CASE STUDY

The High Performance Computing Centre (HPCC) at a large university is a global institution with a strong multidisciplinary research culture. The HPCC worked with Pure Storage to fulfill new requirements and cater to a growing demand for supercomputing resources.

(continued)

(continued)

With Pure's FlashBlade and Evergreen delivering significant improvements, researchers have been able to accelerate their investigations and process multiple streams of high-volume data without disruptive, costly upgrades.

Expanding the university's supercomputing infrastructure

The HPCC processes data for projects ranging from genome sequencing to weather analysis, and more. The team required a storage architecture that could scale linearly and easily up to at least a million IOPS, given that it can run up to eight different workloads at any point in time. This led to the deployment of FlashBlade.

Cataloging botanical diversity

Researchers from the university's School of Biological Sciences were among the first to leverage FlashBlade. Their goal was to analyze the genomes of over 1,000 plant species. Conducting the first project of this scale globally, the research team worked with DNA samples to map the genetic codes, also known as assemblies, of every species in their study. To build each assembly, the teams compared DNA sequences for overlaps and predicted the genes in the sequences, piece by piece. With over a trillion possible combinations, genome assembly is one of the most computationally intensive processes the HPCC supports.

FlashBlade: A game-changer for genetic research

A key priority from the start was the ability to run multiple analyses concurrently. Looking at one genome a week would take 20 years to cover 1,000 species, so doing things in parallel and moving to FlashBlade was important in enabling them to speed up the process. With FlashBlade allowing them to run up to four jobs in parallel, the team has since crossed the project's halfway mark, completing more than 550 assemblies in just 18 months, compared to less than 100 assemblies within the same time span if examined one by one.

The team hopes to expand the study to other areas and dive deeper into specific species and groups. The data will also help other researchers studying environmental conservation and climate change as well as the plants' potential for medical and pharmaceutical use.

AI

AI is a branch of computer science concerned with creating self-learning systems that can perform tasks that normally require human intelligence. From self-driving cars to predicting the future, AI is revolutionizing the ways in which you can use data to shape your world. AI workloads process massive amounts of data requiring high-performance, architecturally optimized data storage, such as a UFFO platform.

AI can power your data analytics pipeline in a number of ways — from processing unstructured data like streaming video feeds to powering advanced predictive analytics tools. On the flip side, big data analytics pipelines can leverage AI to efficiently process large quantities of unstructured data spread across multiple siloed environments. But for AI to improve and thrive as it's meant to, it needs healthy data pipelines, flexibility, and scalability — something that brittle, legacy architectures can't easily accommodate.



TECHNICAL
STUFF

ML and deep learning (DL) are subsets of AI that are commonly referred to in the broader context of AI. The difference between AI, ML, and DL is the scope of specificity:

- » AI deals broadly with building computer systems that can learn on their own.
- » ML is a subset of AI that deals with algorithms that can self-learn from training data.
- » DL is a subset of ML that deals with neural networks — specifically those that have more than one “hidden” layer between the input and output layers of the network.

IT PIONEER AIMS HIGH AND FUELS ITS JOURNEY WITH PURE



CASE STUDY

A pioneering IT company has been helping its customers tackle their most ambitious, groundbreaking projects. With world-class AI and ML researchers and an HPC infrastructure, the tech company makes it possible and profitable for clients to achieve their goals.

(continued)

(continued)

Data is at the center of every project. The tech company's researchers must be able to access, analyze, and move data at will as it runs its experiments. Having maxed out capacity and performance on both cloud and direct-attached storage, the company turned to Pure Storage.

Fueling AI/ML research with big data

The tech company's AI and ML solutions are making self-driving cars better at detecting cyclists and pedestrians; preventing fraud by spotting anomalies among billions of financial transactions; and sparking innovation in logistics, healthcare, telecom, and beyond.

The company needed high-performing infrastructure, but its direct-attached storage was maxed out and couldn't keep up with the data-intensive workloads. The hard drives were burning out left and right from high I/O requirements of its applications. It needed the flexibility and performance to handle parallel workloads across its hybrid cloud environment.

Delivering faster transformative results to clients

With FlashBlade, the tech company combines the flexibility and abstraction of cloud storage with on-premises control and performance. That makes it easy to collaborate with clients on large-scale AI projects.

By using Amazon S3-compliant data stores on FlashBlade, the company can easily move data back and forth, accessing client files with low latency. Faster performance means the company can run multiple experiments simultaneously, allowing researchers to work unimpeded by resource constraints and deliver results faster.

By using FlashBlade, research teams have helped clients analyze Securities and Exchange Commission (SEC) data in three days instead of ten, calculate delivery routes for thousands of trucks in two-thirds less time, and detect billing anomalies with up to 93 percent accuracy for telco and utility companies.

IN THIS CHAPTER

- » Addressing data growth and complexity
- » Enabling real-time insights
- » Delivering a cloud-like experience
- » Supporting modern applications and use cases
- » Maximizing ROI and operational efficiency

Chapter 4

Ten Reasons You Need a UFFO Storage Platform

Deploying a unified fast file and object (UFFO) storage platform is a critical step in future-proofing your organization's infrastructure for the unstructured data explosion. Not all technologies and architectures are created equal. Many simply aren't built to handle the demands of modern data and applications.

In this chapter, we share a few reasons why you need a UFFO storage platform and highlight what Pure Storage has to offer with FlashBlade.

Manages Data Growth

Unstructured data in both enterprise data centers and the public cloud is growing exponentially. By 2025, IDC says worldwide data will grow to 175 zettabytes. That's almost triple from 2020. Many enterprises estimate their file and object data requirements will triple within the next three years. Without the right storage platform, the volume and variety of unstructured data can quickly become a bottleneck.



REMEMBER

Consolidation and scalability are key to storing and analyzing unstructured data with ease. Pure Storage FlashBlade brings the performance of flash to unstructured data and scales to support petabytes of data, billions of files and objects, and thousands of cores.

Accelerates Modern Data and Applications

With nearly 50 billion connected Internet of Things (IoT) devices — roughly six devices for every person on the planet — most data today is machine-generated. This trend creates new opportunities but also new challenges. How applications are deployed and how they're underpinned by infrastructure platforms to store, manage, and leverage massive amounts of data are key to deriving value from all the noise. Without the right architecture, it's more difficult to accelerate mission-critical applications and databases to support the multitude of use cases for modern data.



REMEMBER

FlashBlade can handle the complexities of modern unstructured data, striking a balance between simplicity, performance, and agility to meet the needs of data-intensive applications. It can support your current and future data growth with a scale-out architecture and the ability to consolidate different data types and workloads.

Enables Better and Faster Insights

Data's value is in the insights you extract from it. Rich data analytics from connected devices, ubiquitous internet, and artificial intelligence (AI)/machine learning (ML) are all making the gathering and analyzing of unstructured data an essential endeavor. From advanced analytics to business intelligence, the possibilities are there — but, the right underlying infrastructure is critical to take this on.



REMEMBER

Only multi-dimensional storage performance with high throughput can drive better data-driven decision-making in modern organizations. FlashBlade enables faster insights by delivering high performance regardless of workload, file size, protocol, or file and object count.

Eliminates Data Silos

Single-use silos for different types of file and object data inhibit collaboration, sharing, and accessibility. This practice quickly creates major inefficiencies and complexity across the enterprise. When traditional storage architectures lead to silos, you're left losing out on the potential of unstructured data.



TIP

The right data services platform coupled with the right storage solution breaks down silos with simplified storage management. With a solution like FlashBlade, you can eliminate silos created by traditional approaches to scaling out — including direct-attached storage clusters, federated scale-up storage arrays, and legacy scale-out storage systems.

Delivers a Cloud-Like Experience

DevOps teams are all about agility, but legacy approaches to capacity planning, hardware provisioning, performance optimization, and storage-management processes can slow them down. These unwieldy processes ultimately reduce agility, decrease productivity, and increase time-to-market.



TIP

A subscription-based model such as Pure-as-a-Service can replace aging legacy storage infrastructure capacity planning and procurement cycles, which in turn can help you deliver an agile, cloud-like experience for your users. The result: You gain simplicity and flexibility for cloud-native applications on-premises.

Supports Fast Object Use Cases

Object storage has moved from being a low-cost, low-performance commodity to a high-performance solution that powers modern use cases, including the IoT, media and entertainment, electronic design automation (EDA), and intensive data storage and analytics. With a UFFO storage platform, it's possible to combine the fast access of file storage with the scalability of object storage.



TIP

FlashBlade's UFFO storage platform can store vast quantities of unstructured data, offering the speed associated with flash storage technology in a scale-out architecture. Check out Chapter 3 to discover more about UFFO use cases and how Pure Storage helps real-world customers achieve better business outcomes.

Improves Data Protection and Business Continuity

Modern data protection that addresses current ransomware risks and critical business continuity needs is more critical than ever. Recent ransomware attacks demonstrate the potentially disastrous supply chain consequences of downtime.



REMEMBER

FlashBlade helps organizations secure their data and ensure business continuity with robust security and data protection capabilities, including file replication, snapshots, files system rollback, Kerberos authentication, SafeMode snapshots, audit logging, and always-on encryption.

Eliminates Downtime

Application and system downtime is becoming increasingly expensive in today's highly competitive, "always-on" business environment. Storage updates and upgrades are often cumbersome and fraught with risk — potentially bringing an entire business to a halt for days.



TIP

With an always-modern infrastructure, you can leverage non-disruptive upgrades and all-inclusive software licenses without repurchasing existing storage. Pure's Evergreen Storage subscription model combines a non-disruptive product architecture with a flexible buying program that frees customers from complex, risky, and expensive forklift upgrades and data migrations every three to five years.

Maximizes ROI

Storage consumes a significant portion of every organization's IT budget as data performance and capacity requirements continue to grow. Legacy buying models are risky and unsustainable because they require significant capital investments every year to simply keep up with business demands.



REMEMBER

FlashBlade provides a future-proof foundation with an infrastructure that adapts to evolving business needs, keeps data in place as customers grow, and stays modern. Pure as-a-Service, Pure's subscription-based consumption model, helps you maximize your return on investment (ROI) by only purchasing what you need when you need it.

Increases Operational Efficiency

Storage management often requires a delicate (and time-consuming) balancing act to achieve the optimum performance and capacity levels that business users require. Decoupling compute and storage can ease management and help you fine-tune utilization by enabling them to be scaled independently on demand.



TIP

Increase your operational efficiency with FlashBlade. It's simple to use, scales out, and consumes only what's necessary to deliver true investment protection.

DRIVEN BY DATA



Official Team Partner



AMG
PETRONAS
FORMULA ONE TEAM

purestorage.com/f1 →

©2021 Pure Storage, Inc. All right reserved.

These materials are © 2022 John Wiley & Sons, Inc. Any dissemination, distribution, or unauthorized use is strictly prohibited.

Get ready for the challenge of modern unstructured data

Your data has immense potential. It can help you transform operations and delight customers like never before. Here's the catch: You can only realize the value of that data if you can quickly and efficiently manage, analyze, and access it. That requires a completely new approach — a unified fast file and object (UFFO) storage platform that legacy storage infrastructure can't deliver. This book explains UFFO storage, how other companies are using it, and why your organization may not survive without it.

Inside...

- Discover UFFO storage
- Learn why legacy storage can't keep up
- Evaluate an optimal UFFO solution
- Explore real-life customer use cases
- See ten reasons why UFFO is for you



Lawrence C. Miller has worked in IT in various industries for more than 25 years. He is the co-author of *CISSP For Dummies* and has written more than 150 *For Dummies* books on numerous technology and security topics. **Erin Poulson** is a long-time freelance writer who focuses on business and technology.

Go to **Dummies.com**[™]
for videos, step-by-step photos,
how-to articles, or to shop!

ISBN: 978-1-119-79599-5
Not For Resale



for
dummies[®]
A Wiley Brand

WILEY END USER LICENSE AGREEMENT

Go to www.wiley.com/go/eula to access Wiley's ebook EULA.