

Compliments of **IBM**

IBM Limited Edition

# Software-Defined Storage

FOR  
**DUMMIES**<sup>®</sup>  
A Wiley Brand

## *Learn to:*

- Control storage costs
- Eliminate storage bottlenecks
- Use IBM Spectrum Storage to solve storage management challenges

Neal Ekker





# *Software-Defined Storage*

FOR  
**DUMMIES**<sup>®</sup>  
A Wiley Brand

*IBM Limited Edition*

**by Neal Ekker**

FOR  
**DUMMIES**<sup>®</sup>  
A Wiley Brand

## Software-Defined Storage For Dummies®, IBM Limited Edition

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

Copyright © 2017 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. IBM and the IBM logo are registered trademarks of International Business Machines Corporation. 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: THE PUBLISHER AND THE AUTHOR 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 WARRANTIES OF FITNESS FOR A PARTICULAR PURPOSE. NO WARRANTY MAY BE CREATED OR EXTENDED BY SALES OR PROMOTIONAL MATERIALS. THE ADVICE AND STRATEGIES CONTAINED HEREIN MAY NOT BE SUITABLE FOR EVERY SITUATION. THIS WORK IS SOLD WITH THE UNDERSTANDING THAT THE PUBLISHER IS NOT ENGAGED IN RENDERING LEGAL, ACCOUNTING, OR OTHER PROFESSIONAL SERVICES. IF PROFESSIONAL ASSISTANCE IS REQUIRED, THE SERVICES OF A COMPETENT PROFESSIONAL PERSON SHOULD BE SOUGHT. NEITHER THE PUBLISHER NOR THE AUTHOR SHALL BE LIABLE FOR DAMAGES ARISING HEREFROM. THE FACT THAT AN ORGANIZATION OR WEBSITE IS REFERRED TO IN THIS WORK AS A CITATION AND/OR A POTENTIAL SOURCE OF FURTHER INFORMATION DOES NOT MEAN THAT THE AUTHOR OR THE PUBLISHER ENDORSES THE INFORMATION THE ORGANIZATION OR WEBSITE MAY PROVIDE OR RECOMMENDATIONS IT MAY MAKE. FURTHER, READERS SHOULD BE AWARE THAT INTERNET WEBSITES LISTED IN THIS WORK MAY HAVE CHANGED OR DISAPPEARED BETWEEN WHEN THIS WORK WAS WRITTEN AND WHEN IT IS READ.**

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](mailto:info@dummies.biz), or visit [www.wiley.com/go/custompub](http://www.wiley.com/go/custompub). For information about licensing the *For Dummies* brand for products or services, contact [BrandedRights&Licenses@Wiley.com](mailto:BrandedRights&Licenses@Wiley.com).

ISBN: 978-1-119-14981-1 (pbk); ISBN: 978-1-119-14982-8 (ebk)

Manufactured in the United States of America

10 9 8 7 6 5 4 3 2 1

## Publisher's Acknowledgments

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

**Project Editor:** Carrie A. Burchfield

**Editorial Manager:** Rev Mengle

**Acquisitions Editor:** Steve Hayes

**Business Development Representative:**  
Sue Blessing

**Production Editor:** Antony Sami

# Table of Contents

## **Introduction . . . . . 1**

About This Book .....	1
Icons Used in This Book.....	2
Beyond the Book.....	2

## **Chapter 1: Projecting the Future of Data Storage . . . . . 3**

Making Connections between Business and Storage Technology.....	3
Looking at the Current and Future Drivers of Storage Solutions.....	6
Lower storage-related costs.....	6
Manage storage more efficiently.....	7
Move data around, easily and often.....	8
Protect your data as effectively as possible .....	8
Easily handle all varieties of data.....	9
Derive the most business value possible from your data assets.....	10

## **Chapter 2: What Is Software-Defined Storage? . . . . . 11**

Defining SDS.....	11
Storage Virtualization.....	12
Looking at the Key Benefits of SDS.....	12
Agility .....	13
Control .....	13
Efficiency.....	14

## **Chapter 3: Introducing IBM Spectrum Storage through Storage and Data Control . . . . . 15**

IBM Spectrum Control.....	16
IBM Spectrum Protect .....	18
Efficiency.....	18
Scalable performance.....	19
Advanced data protection.....	19
Flexible restores .....	20
Multi-site replication .....	20

**Chapter 4: Introducing the Infrastructure Members  
of the IBM Spectrum Storage Family .....21**

IBM Spectrum Virtualize .....	21
Storage virtualization .....	23
Compression .....	23
Encryption .....	24
Automated tiering.....	24
Replication, snapshots, and mirroring .....	25
Thin provisioning .....	26
IBM Spectrum Accelerate .....	26
IBM Spectrum Scale .....	29
Manage the information life cycle .....	33
Ensure data availability, reliability, and integrity .....	34
Support Hadoop.....	35
Enable OpenStack cloud SDS .....	35
IBM Spectrum Archive .....	36

**Chapter 5: Ten Use Cases for Software-Defined  
Storage .....39**

Partnering with Flash .....	39
Managing from the Cloud.....	40
Enabling Hybrid Cloud .....	40
Protecting Data from the Cloud .....	41
Archiving Data to the Cloud .....	41
Simplifying Big Data Analytics.....	42
Unifying File and Object Storage.....	43

# Introduction

---

**T**he business of the future — your business — will be a “thinking” business, or what IBM calls a cognitive business. A cognitive business uses its data assets to gain competitive advantage.

The technology that helps you do all the things you want to do with your data assets is called *information technology* (IT). One of the main components of business or enterprise IT is data storage. A real driving force within IT today and within business itself stems from the need to gain and maintain competitive advantage by deriving greater value from data assets. Simply storing data isn’t enough. Making faster, more informed business decisions; staying ahead of the competition; creating better customer experiences; implementing the best possible fraud protection and data security — these and many similar requirements are the drivers of the new era of cognitive business.

To help them derive more value from their data assets, many businesses around the globe have already begun deploying a new type of enterprise data storage called *software-defined storage* (SDS). This technology essentially transforms traditional, manually operated storage into “thinking” systems that automatically monitor your data, move it appropriately, and perform a long list of powerful functions that ultimately lower the cost and complexity of storage while making your data more valuable to you than ever before.

## About This Book

*Software-Defined Storage For Dummies*, IBM Limited Edition, tells the story of this paradigm shift in data management technology. This book introduces IBM Spectrum Storage, a leading family of SDS solutions, and provides plenty of information and ideas on how these “thinking” systems from IBM can help your cognitive business be a market leader.

## Icons Used in This Book

Throughout this book, I occasionally use special icons to call attention to important information. No smiley faces winking at you or any other cute little emoticons, but you'll definitely want to take note. Here's what you can expect.



This icon points out information that may well be worth committing to your nonvolatile memory — your gray matter — along with anniversaries and birthdays.



You won't find a map of the human genome or the blueprints for IBM's Watson here, but the content under this icon does provide useful explanations of the jargon beneath the jargon.



Thank you for reading, hope you enjoy the book, please take care of your writers. Seriously, this icon points out useful nuggets of information.



Proceed at your own risk . . . well, okay — it's actually nothing *that* hazardous. These helpful alerts offer practical advice to help you avoid potentially costly mistakes.

## Beyond the Book

Although this book is chock full of information, there's only so much I can cover in these pages! So, if you find yourself at the end of this book thinking “gosh, this was an amazing book, where can I learn more?” just go to [ibm.com/spectrum/storage](http://ibm.com/spectrum/storage). There, you can learn more about SDS and IBM Spectrum Storage.

If you want to find out more about how SDS fits into the bigger picture of software-defined infrastructure, visit [ibm.biz/SoftwareDefined](http://ibm.biz/SoftwareDefined).

Finally, if you're new to cognitive businesses, visit [ibm.com/cognitive](http://ibm.com/cognitive) to learn how IBM's cognitive technologies can help your business understand, reason, and learn.



## Chapter 1

---

# Projecting the Future of Data Storage

.....

### *In This Chapter*

- ▶ Recognizing the connections between business and data storage
  - ▶ Introducing the drivers of storage solutions
- .....

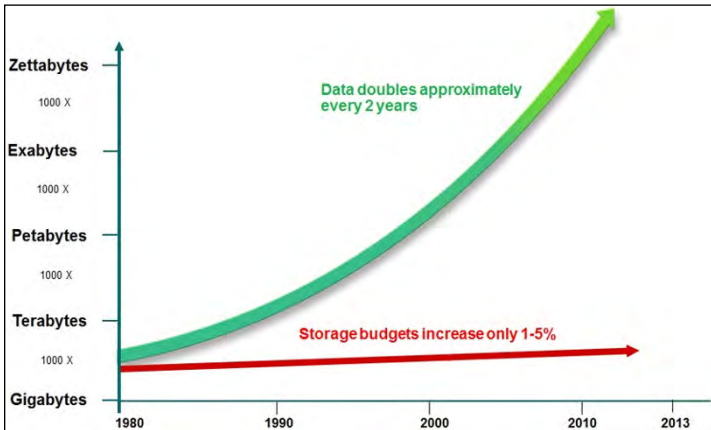
**I**n a very real way, the future of enterprise data storage mirrors the future of business itself. In this chapter, I provide some insight into the business and general IT reasons why software-defined storage (SDS) can be valuable to you and your business.

## *Making Connections between Business and Storage Technology*

Most often, when you consider the future of data storage, you focus on the explosive growth of data over the past several decades. Alarming numbers come to mind. For example, most industry analysts agree that the amount of information created and stored by businesses around the world is doubling every two years or so. At the same time, data storage budgets are hardly growing at all, as seen in Figure 1-1.

The collision of these trends most often gets blamed for what is happening today in enterprise storage and what will happen tomorrow. But in fact, these data growth and IT budget trends may be leading away from the real driver of

the future of storage. Consider this: Tape-based storage offers plenty of capacity to meet any foreseeable storage need at a very small cost per gigabyte (GB). If the challenge was simply to store all the proliferating data, you could do it at quite a reasonable cost.



**Figure 1-1:** Storage requirements are growing much faster than storage budgets.

Instead, the future of storage and of business itself lies in the relentless requirement to derive ever more value from data assets, no matter their volume or velocity.

What does this phrase mean — “derive value from data”? Think about an online shopper visiting a retail website. If a purchase is made, this eCommerce firm acquires the basic purchase data — what was bought, when, how much it cost, and so on.

What if this firm stored this information and kept track of other purchases made by similar buyers? At a later date, if this information were retrieved and run through certain kinds of data analytics software, the retail firm could discover useful trends — what demographics related most strongly to which products, for example, or which times of the day, month, or year saw spikes in buying activity.

From this simple analysis and the resulting trends, the retail business might manage its product supply chain more efficiently, decide to stop selling some items, and instead push others more strongly, and ultimately make better business decisions that could drive down costs and increase sales and profits.

Same data. More value.

But this analogy is just getting started. Imagine capturing information about each product web page a buyer visits. If you could analyze this data fast enough, you might post an advertisement for matching shoes right when the buyer is looking at a particular pair of pants. Increased sales! Greater value from available information.



To post targeted ads on a web page while a buyer is still on the page — essentially in “real-time” — requires the underlying IT infrastructure to operate extremely fast and efficiently. Data must be captured, stored, moved, processed, and secured in milliseconds. This is the future of information technology.

And I’m not done yet, by any means. Not only does the online retailer need to accept, verify, and process credit card payment information as quickly as possible in order to provide the best possible customer experience, but also what about fraud? Can the retailer’s IT systems search, load, and analyze enough information within the few seconds of the payment process to detect a stolen credit card?

Think of the value upon value added of data sets used not only for original transactions but also for additional marketing, fraud prevention, supply chain management, and even corporate strategic planning. Thousands of businesses worldwide derive this much value from their data, and more, every second. But the requirements on their data storage systems hardly stop there.

Every business wants to gain the most value it can from its available data as easily and as inexpensively as possible. Everyone wants more and more for less and less. SDS evolved exactly to provide this kind of IT infrastructure magic.

# Looking at the Current and Future Drivers of Storage Solutions

I'd like to drill down a little deeper into more specifically what enterprises of all types and sizes need from their data storage solutions. As you look out toward the future of storage, this section gives you six general business and IT requirements that enterprise data storage must address. These are, in fact, the drivers of the evolution of SDS technology.

## Lower storage-related costs

Every day, approximately 15 petabytes (a petabyte is 1,000 times 1,000 GBs) of new information is generated worldwide. According to Google CEO Eric Schmidt, every two days people now create as much information as they did from the dawn of civilization up until 2003. At the same time, storage budgets are increasing only one to five percent annually; thus the gap between data growth and storage spending is widening. This gap between storage growth and budget, as well as the relentless desire to gain ever more value from data assets, creates tremendous data storage challenges for enterprises and IT departments everywhere. Simply put, storage needs to be less expensive in order to keep up with demand.

In fact, technology advancements constantly offer new and better ways to lower storage-related costs. Here are a few examples:

- ✔ All three of the current mainstream data storage media — tape, hard disk drives, and “flash” solid state storage — are able to store more data at less cost every year.
- ✔ Technologies that store data more efficiently or reduce the amount of redundant data stored are getting better and growing much more common. These are called, collectively, *data reduction technologies*, and they can significantly reduce storage costs.

- ✓ Some storage media, such as flash and tape, use much less electricity, need less cooling, take up less data center floor space, or in other ways reduce storage operational expenses.

The topic of storage costs is extensive and complex; these are just a few of its many ingredients. But every storage cost reduction option requires some amount of intelligent and effective system monitoring and management. To fully realize the benefits offered by these and other storage cost reduction technologies, SDS must become part of the solution.

## *Manage storage more efficiently*

To get the most you can from storage cost reduction technologies and strategies, you must manage storage systems as efficiently as possible. In fact, managing data storage more efficiently leads to many additional cost savings all on its own. For example, imagine storing data that's accessed quite often on storage such as flash that offers the capability to help derive more value from these active data sets because of its particular performance characteristics. Then at the same time you would store less active data on storage media such as disk or even tape that aren't high performance but do offer certain cost advantages. This is called *storage tiering*, and it's a powerful technique to both lower costs and increase performance. But to do it well requires extra storage monitoring and management.

Labor costs are always a significant line item in any IT budget. What if you needed fewer IT staff because your storage systems could do some tasks automatically and perhaps more efficiently? For example, imagine that your storage system could perform storage tiering without manual intervention? You might reduce IT staff and save money, or refocus your valuable IT engineers on developing and implementing more and better ways to gain value from your data assets.



Managing data storage more intelligently, effectively, and efficiently goes hand in hand with reducing storage costs. This is exactly what SDS is designed to do.

## *Move data around, easily and often*

In order to do storage tiering well, your IT staff or your intelligent storage system must move data from one storage medium to another, perhaps quite often. Think of all the businesses around the world that must do month-end accounting and billing. Sales transactions, customer information, and expense data accumulate over the weeks and then for a few hours at the end of the month, these data sets are intensely processed to reconcile the books and produce billing invoices, among other results. This is a classic business use case for storage tiering. You can reduce costs by storing the inactive business data on tape or disk during the month. When this same data becomes active, you move it to flash. Efficiency increases; costs go down.

This is just the tip of the data movement iceberg. Consider global enterprises with offices from Manhattan to Mumbai. Each office needs certain data close at hand; the corporation wants to collect and protect all of its data in one location. The solution? Make copies of the data and move them as needed, next door or around the world, to optimize business value.

But copying and moving data can be complex and costly, demanding lightning fast, extremely accurate communication and coordination — a perfect job for intelligent storage systems.

## *Protect your data as effectively as possible*

Breaches of data security cost worldwide businesses billions of dollars annually. In fact, some companies never fully recover from a customer information theft. According to IT analyst firm Enterprise Strategy Group (ESG), more than one-third of organizations cite information security initiatives as one of their top IT priorities.

Hackers are incredibly destructive, but data loss and data protection issues run much deeper than just attacks from outside your business. What happens when a hurricane floods your

data center? Or a disk drive burns up? Or a clumsy employee accidentally unplugs the wrong power cable? Data protection technologies and solutions are deeply woven into the basic fabric of all enterprise-grade computing environments, from the chip level all the way up to company-wide policies.

These days, data encryption starts on the medium where data is stored. All reputable enterprise IT systems come standard with redundant components and multiple data paths to eliminate any single point of failure and potential data loss. Duplicate or what are known as mirrored compute and storage systems are configured for crucial applications. Additionally, data is copied and kept as backups. Some copies are sent to distant IT facilities designed to take over business functions in case of a disaster at the primary site.

Effective data protection requires a lot of data movement, management, coordination, and system functionality. All this of course leads to increased complexity and cost, which makes data protection an especially strong candidate for intelligent, efficient, highly functional SDS.

## ***Easily handle all varieties of data***

Enterprise users and data consumers are churning out vast amounts of documents and rich media such as images, audio, and video. Managing the volume and complexity of this information is a significant challenge, as more and more applications feed on and increase this deluge of information, and as individuals and businesses collaborate for intelligence, analytics, and information sharing.

Information exists as more than one type and is stored and managed in several ways. Structured data refers to data that can be organized, for example, in the rows and columns of a database. Unstructured data refers to data that doesn't have a defined model or framework — multimedia files, for example.

In the past, business generated and utilized lots of structured data, but nowadays, and looking into the future, the majority of data being generated will be unstructured data. Social systems of engagement such as Facebook, Twitter, and YouTube generate massive quantities of unstructured data, as do the myriad of mobile applications. Then consider how

unstructured data volume will explode when the Internet of Things (IoT) really gets going, with hundreds of millions of appliances and devices such as your car or even your refrigerator broadcasting status and telemetry data in unstructured formats.

If you aren't already, it's very likely that soon your business will want to tap into these 21st century data sources to gain still more value from available information. But to do this in a very cost-effective and efficient manner, your storage system must be able to handle structured and unstructured data, bundles of content called files, and the new content containers called "objects." Traditional storage solutions will struggle, but SDS may give you the competitive advantage you've been looking for.

## *Derive the most business value possible from your data assets*

Combine all these present and future storage drivers — reduced cost, greater efficiency, agile data mobility, foolproof data protection, and comprehensive format capabilities — and you've got a fine start on your quest to transform your static, costly, one-dimensional data assets into the engines of competitive advantage for your business.

The devil is in the details. Think about storage performance, for example. You want to react to customer activities in time to influence buying decisions, not just detect but actually prevent fraud, pivot on the latest market trends, make well-informed business decisions quickly, and essentially gain insight in real-time. All this is already possible, but to do it in the ways storage once worked will be ghastly expensive and wretchedly complex.

Or consider the details of moving data between all your worldwide office locations and your several disaster recovery sites. Which data? When? How often? In the past you made such decisions based on only a few attributes of your data files. Now there is SDS from IBM with cognitive capabilities that monitors, collects, analyzes, and can act on over 50 individual file attributes. Just set the policies you want for managing your own particular files, based on these dozens of options, and then go meet with your next customer.



## Chapter 2

# What Is Software-Defined Storage?

---

### *In This Chapter*

- ▶ Defining software-defined storage
  - ▶ Looking at the benefits of SDS
- 

**T**his chapter explains what software-defined storage (SDS) is and how solutions such as IBM Spectrum Storage can help organizations address their data challenges today and for the future.

## *Defining SDS*

In Chapter 1, I introduce the business and technology drivers of enterprise data storage solution development and innovation. As I define SDS and discuss its various characteristics and benefits, keep these basic storage requirements or drivers in mind, because the storage solutions you want are the ones that best address these needs. Flip back to Chapter 1 for more info on these drivers.

Essentially, SDS addresses these requirements much more effectively than the traditional storage techniques and technologies that came before it. But exactly what is SDS?

For practical purposes, you can simply define SDS as storage virtualization integrated with a suite of storage services and functionality that are independent of the hardware being used. The differentiation between various vendor SDS offerings lies in the quality of the storage virtualization engine and the particular attributes of the accompanying storage features.

## *Storage Virtualization*

Storage virtualization simply implements a software program between applications and storage that offers a smiling face to the applications while essentially breaking the direct connection to the backend storage devices and systems, hiding all their individual characteristics, faults, changes, ugliness, and even obsolescence. The consequences are marvelously powerful.

With the storage virtualized, you can move data from one storage system or medium to another to optimize economics or performance or whatever you want, and the frontend applications never know, for the most part, and never need to change or care, unlike in the traditional model where any change to data location or attributes would require an update or change to the associated applications. You can forklift out unwanted storage systems, add new ones or new storage media such as flash, even go so far as change the location of your storage solution by moving it out onto the cloud, and all your applications “see” is that same storage-as-a-resource smiling face.

Your storage virtualization program or “engine” becomes a catalyst. It facilitates storage system change and catalyzes the implementation of the second half of the SDS definition — all those new services and features. Instead of buying whole new storage systems every time you want new, better, or more data storage features and capabilities, you can simply load some software with the virtualization engine and within minutes your data is encrypted on every virtualized storage device, for example, even though it never had this capability before. You can deduplicate; compress; copy, replicate, and mirror; move, migrate, and tier; even monitor, manage, and optimize your data assets and their value to your business without necessarily buying any new storage array, but instead simply leveraging the power of SDS.

## *Looking at the Key Benefits of SDS*

Enterprises today are recognizing many significant benefits from deploying SDS in their data centers. These include increased agility, control, and efficiency.

## *Agility*

Because your business is constantly changing and evolving, your IT environment will too. This means it's very important to have choices, options, and lots of flexibility in what you install or deploy in your data center(s). Also, even if your business doesn't grow, almost certainly the amount of data you accumulate will, which means your IT systems must be able to grow or "scale" easily. Flexibility and scalability are keys to enabling business "agility" — your ability to change and adapt to new conditions quickly and easily.

An SDS solution gives your company flexible deployment options by allowing you to use non-proprietary standard hardware and, in many cases, leverage existing storage infrastructure as part of your enterprise storage solution in the data center and/or in the remote IT services delivery model called "cloud."

Additionally, organizations can achieve massive scale with an SDS solution by adding individual, heterogeneous hardware components as needed to increase capacity and improve performance in the solution, while continuing to manage storage as a single enterprise-class storage system. As you scale out your storage infrastructure, performance and reliability continue to improve.

## *Control*

SDS can automatically apply the data and storage management rules or policies you choose across all your storage resources. This eliminates the inefficiencies caused by separate storage systems or "silos" that don't communicate or coordinate well. In fact, SDS can manage the entire life cycle of any unit of data, moving data when it's "young" and active to high performance storage and then as the data ages and becomes less active, moving it again and again based on the policies you set to maximize cost or performance or some other value. This is called Information Life Cycle Management (ILM).

In addition to ILM tasks, SDS can automatically and intelligently perform many other storage functions and services that otherwise would be done manually and cost more in time and labor. Such storage services include making quick copies

or “snapshots” of a data set, which can be stored in case of emergencies; copying or “replicating” data more completely to use for a variety of purposes, from application development and testing to disaster recovery; and of course storage tiering, where data access patterns are monitored and the data automatically moved from one storage medium to another. Plus of course, SDS comes with tools that provide a wide range of data and storage analytics to help you better plan your next purchase or update. In a nutshell, SDS enables you to put the right data in the right place, at the right time, with the right performance, and at the right cost — automatically.

## *Efficiency*

Adding SDS solutions, especially storage virtualization, to your overall storage solution can significantly increase its efficiency and thus lower storage costs. Storage virtualization can enable you to make all your different storage systems behave as if they are all part of one single system. The first benefit this brings is to significantly increase the efficiency and agility of data movement within your storage solution as a whole, which is one of our key storage objectives or drivers. Add storage tiering to the mix and data flows automatically between your various storage systems depending on the priorities you set. You can accelerate the performance of your business applications while also lowering IT costs by moving data to flash, for example, only when higher performance is needed, and then immediately off flash and into high capacity disks, tape, or even cloud storage when high performance is no longer needed.

But agile and efficient data movement is just the tip of the SDS-berg. After your entire storage solution is virtualized, you can much more easily add storage services and functionality and spread them across all your storage systems. Copy data management offers an excellent example of how this can work. In almost every data center the number of copies of individual data sets is proliferating. Copies are continually made and kept for regulatory, data protection, and disaster recovery, as well as for application development and testing purposes. Every copy uses up storage capacity, to the point where more capacity is dedicated to storing copies than for the production data sets themselves. Managing copy data, tracking copy use and requirements, and hunting down and deleting unused copies increases efficiency and lowers your storage capacity requirements and thus your storage costs.

## Chapter 3

---

# Introducing IBM Spectrum Storage through Storage and Data Control

.....

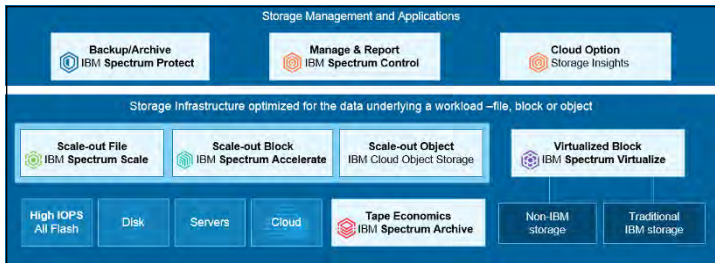
### *In This Chapter*

- ▶ Introducing the IBM Spectrum Storage family
  - ▶ Exploring IBM Spectrum Control and IBM Spectrum Protect
- .....

**I**BM was essentially offering software-defined storage (SDS) products and solutions long before anyone coined the name. For example, IBM SAN Volume Controller (SVC) is a well-known, highly successful storage virtualization product that has been deployed around the world in the most mission-critical environments for over 12 years. IBM General Parallel File System (GPFS) has a similar background in the high performance computing arena for file and object data. And of course, IBM Tivoli data and storage management products are well known.

Recently, IBM has invested heavily in SDS development and innovation. A significant portion of this effort and investment has gone into consolidating many of the previous standalone SDS-related solutions into one product family with a common user interface, and then of course continuing to innovate and extend new and better functionality and features to each and all. This new SDS product family is called IBM Spectrum Storage.

The IBM Spectrum Storage family offers a very wide range of SDS functionality and capabilities that address all your storage objectives. Figure 3-1 shows the six individual IBM Spectrum Storage family members organized by their basic functions: Storage Management and Applications, and Storage Infrastructure. In this chapter, I cover the first two members of the IBM Spectrum Storage family in the area of management and applications.



**Figure 3-1:** The IBM Spectrum Storage family.

## IBM Spectrum Control

IBM Spectrum Control is the member of the IBM Spectrum Storage family devoted entirely to improving and enhancing the management of your data and storage. IBM Spectrum Control is a storage resource management (SRM) suite that helps reduce the complexity of storage environments by enabling you to centralize, simplify, and automate routine tasks associated with data, storage systems, networks, and replication services. It helps standardize processes without requiring organizations to “rip and replace” their existing storage systems. IBM Spectrum Control

- ✔ Provides comprehensive storage system monitoring, automation, and analytics
- ✔ Helps automate common storage tasks, such as allocating or provisioning resources and coordinating and integrating all your storage assets
- ✔ Provides efficient storage infrastructure management for traditional, virtualized, cloud, and SDS environments

This solution provides new efficiencies in managing your storage environments — both at an operational level and in leveraging existing storage assets more effectively. The solution provides a single point of control for configuring, managing, and monitoring performance of storage-area network (SAN) storage devices. Storage administrators can configure storage devices from a single point, view the SAN status, and provide operational support. Plus, with IBM Cognos Business Intelligence integration, administrators can easily create custom reports on capacity and performance.



The key capabilities of IBM Spectrum Control include

- ✔ Single console for managing all types of data on disk, flash, file, and object storage systems, whether traditional, virtualized, or SDS
- ✔ Simplified visual administration tools — including an advanced web-based user interface, VMware vCenter plug-in, and Cognos Business Intelligence with pre-designed reports

IBM Spectrum Control helps reduce the costs of overall storage management with capabilities that save IT staff time and improve productivity, including

- ✔ Simplified inventory control, asset management, and reporting
- ✔ A single, integrated web-based administrative console designed to simplify the management of multiple storage devices
- ✔ Tools that enable IT staff to perform routine administrative tasks such as aggregation, grouping of devices, and policy-based actions from a single location

Thanks to these features and capabilities, IBM Spectrum Control

- ✔ Enables IT staff to proactively manage performance by setting thresholds based on performance metrics
- ✔ Reduces call volumes and improves diagnosis for help-desk staff

- ✓ Allows administrators to monitor metrics, such as I/O rates and cache utilization, to improve capacity management and planning
- ✓ Provides a central user interface for configuring and managing all supported devices on the SAN

## *IBM Spectrum Protect*

Protecting data as effectively as possible is one of the drivers of storage innovation and storage headaches. The concept of data protection encompasses a range of issues. You certainly don't want to lose your valuable data assets. You also don't want your data to become "corrupt" or inaccurate. You also need it to be available, even if your primary data center is not. And of course you want to prevent bad guys from stealing your data. It's a lot to ask from any product, but IBM Spectrum Protect delivers.

Data protection in the modern enterprise data center involves a number of tasks. To prevent loss of data, your storage system needs to make backup copies that remain readily available and can be used if systems fail or data sets become corrupt. Think of the coordination involved in replicating or copying data from heterogeneous storage systems, transmitting it to your disaster recovery site(s), and keeping what arrives there fairly current and accurate.

IBM Spectrum Protect provides just this level of coordination, and much more. It is a data protection platform that gives enterprises a single point of control and administration for backup and recovery. It enables reliable, scalable, and cost-effective backups and fast recovery for virtual, physical, and cloud environments in organizations of all sizes.

### *Efficiency*

To boost operational efficiency even further, IBM Spectrum Protect data deduplication, which deletes duplicate "chunks" (groups of characters) within a data set, enables organizations to handle massive volumes of data during a given backup window. Deduplication can help reduce costs by limiting the storage space required for backups and minimizing the impact on business applications.



For organizations with long-term data retention requirements, IBM Spectrum Protect provides highly efficient data compression capabilities that reduce backup infrastructure needs plus integration with cloud storage and tape systems for the lowest cost storage options. The solution enables easy migration of backup data between disk, tape, and cloud storage pools — so backups can be moved to lower-cost storage as they age.

## *Scalable performance*

IBM Spectrum Protect servers can expand to manage billions of objects per server, so there is less disruption and complexity as backup workloads grow. As data grows, IBM Spectrum Protect backups can be augmented with advanced agents and snapshots that can reduce backup and restore times for large applications and virtual machines. For example, IBM Spectrum Protect FlashCopy Manager can back up about 500 VMs in as little as 7 minutes.



IBM Spectrum Protect supports large, multi-petabyte environments, but small organizations can also benefit. Organizations with fewer than 50 managed servers or less than 100 terabytes of backup data can use IBM Spectrum Protect entry-level solution bundles.

## *Advanced data protection*

IBM Spectrum Protect enables advanced data protection for next-generation applications, including cloud, virtualized, and software defined environments. And data managed by IBM Spectrum Protect is easily replicated to remote recovery facilities for safekeeping.



Backups are important, but restores are essential. IBM Spectrum Protect enables fast, flexible system restores from primary and remote recovery sites. IBM Spectrum Protect helps recover individual items, complex systems, and entire data centers.

IBM Spectrum Protect and IBM Spectrum Protect FlashCopy Manager can back up and restore data residing on virtual machines without negatively impacting applications and operations that also rely on the virtual environment.

In addition, administrators using VMware tools can manage backups and restores in a familiar environment, without having to learn new data protection interfaces.

## *Flexible restores*

Flexible recovery options in IBM Spectrum Protect are designed to simplify common restore requests, such as

- ✔ Individual item-level recovery for Microsoft Exchange and SQL databases
- ✔ Full virtual machine instance restores
- ✔ Near-instant access to backup data for Windows and Linux Intel-based servers and Windows workstations
- ✔ Snapshot browsing and recovery
- ✔ Automatic restores from an alternate backup server, if the primary backup server is unavailable
- ✔ Simplified disaster-recovery audits and simulations, as well as guided recovery after disasters

## *Multi-site replication*

IBM Spectrum Protect includes enhanced features to simplify disaster recovery checks or audits and guide recovery from site-level disasters. Backup data can be replicated on an incremental, scheduled, or per-system basis from one IBM Spectrum Protect server to another. Replication is policy-driven, so on-site and off-site retention policies can be different. Plus, replication can be performed with deduplicated data, which improves network efficiency because less actual data must be transmitted. It can also be scheduled during relatively quiet times to help reduce the impact on production applications.

## Chapter 4

---

# Introducing the Infrastructure Members of the IBM Spectrum Storage Family

---

### *In This Chapter*

- ▶ Transforming SAN storage with IBM Spectrum Virtualize
  - ▶ Deploying scale-out block storage rapidly with IBM Spectrum Accelerate
  - ▶ Delivering global high-performance storage with IBM Spectrum Scale
  - ▶ Enabling multi-tiered tape-based data storage with Spectrum Archive
- 

**T**he IBM Spectrum Storage family of software-defined storage (SDS) solutions offers a variety of capabilities. In this chapter, we cover the four storage infrastructure areas.

## *IBM Spectrum Virtualize*

IBM Spectrum Virtualize, through the power of its storage virtualization capabilities, helps you address all the key storage objectives (see Chapter 2 for more info) — except the need to handle all varieties of data. Instead, it's focused on handling conventional structured data, often called *block storage*.

In typical enterprise data centers, storage capacity is often isolated or trapped in islands of disparate storage:

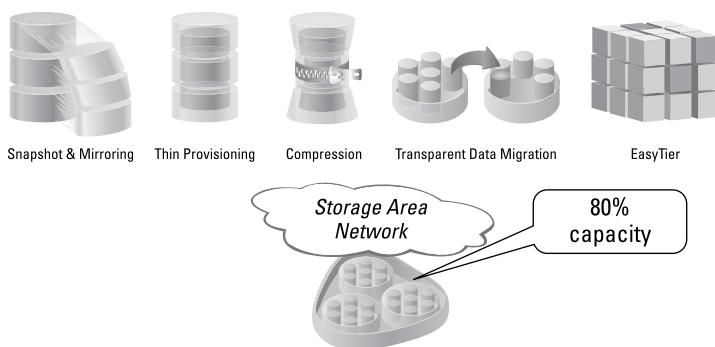
- ✔ Storage-area networks (SAN)
- ✔ Network-attached storage (NAS)
- ✔ Direct-attached storage (DAS)

Traditionally, different storage systems were provisioned for and owned by individual applications. This resulted in duplication of management points and poor capacity utilization because storage resources couldn't be moved as needed to other applications and workloads.

Storage virtualization like that provided by IBM Spectrum Virtualize transforms the traditional storage paradigm by providing a single management point for all the storage systems that you virtualize and thus enabling better utilization and performance from existing storage hardware through increased ability to move data as needed. Along with storage virtualization, IBM Spectrum Virtualize also provides many common enterprise storage services, such as

- ✔ **Snapshots, replication, and mirroring:** Features that make and store various types of copies of your data for protection purposes
- ✔ **Thin provisioning:** The strategy of allocating only the amount of capacity immediately needed, rather than allocating the amount you plan to use over some future time period
- ✔ **Compression:** A technology that reduces that amount of storage space used for data
- ✔ **Encryption:** A technology that helps prevent disclosure of information by making it unreadable with a secret “key” to unlock it
- ✔ **Transparent data migration across all storage vendor platforms:** Automated (the actual technology is called *IBM Easy Tier*) and manually initiated data movement between any of the storage systems you have virtualized

These storage services provided by IBM Spectrum Virtualize are illustrated in Figure 4-1. Together, they make your storage much more efficient and much easier to manage, update, and enhance with new features.



**Figure 4-1:** IBM Spectrum Virtualize maximizes storage efficiency.

Virtualizing storage helps make new and existing storage more effective. IBM Spectrum Virtualize includes many functions traditionally deployed separately in storage systems. By including these in virtualization software, IBM Spectrum Virtualize standardizes functions across virtualized storage for greater flexibility and lower costs.

Take a closer look at some of the key features and capabilities provided by IBM Spectrum Virtualize.

## *Storage virtualization*

*Storage virtualization* provides a software layer that helps simplify storage management. Among other benefits, data virtualization simplifies deployment of new applications and new storage tiers; eases movement of data among tiers; and enables consistent, easy-to-use optimization technologies across multiple storage tiers. IBM Spectrum Virtualize includes data virtualization technology to help insulate applications from physical storage. This enables applications to run without disruption, even when changes are made to the storage infrastructure.

## *Compression*

IBM Real-time Compression is designed to enable storing five times or more data in the same physical disk space by compressing the data. Unlike other approaches to compression, real-time compression is designed to be used with active

primary data such as production databases and email systems, which dramatically expands the range of candidate data that can benefit from compression.

Real-time compression transforms the economics of data storage. When applied to new or existing virtualized storage, it can significantly increase the usable capacity while maintaining application performance. This can help eliminate or drastically reduce costs for storage acquisition, rack space, and power and cooling and can extend the useful life of existing storage assets.

## ***Encryption***

*Encryption* is the process of encoding information using a secret key. After information has been encoded in this way, you must have the corresponding key to access it. Encryption is a fundamental technology used to help keep information secure.

IBM Spectrum Virtualize supports encrypting information as it's written to storage systems and then decrypting it as the information is read. This approach is designed to prevent unauthorized access to information in cases where a drive is removed from a storage system or when a system is sold or repurposed.

By performing encryption in the virtualization software, IBM Spectrum Virtualize can help protect stored information even on storage systems that themselves don't have the capability.

## ***Automated tiering***

Deploying tiered storage is an important strategy for controlling storage costs. Using this strategy, organizations consider the performance and cost-effectiveness of different types of storage to right-size the storage to meet business needs. However, management and functional differences between different types of storage — even from the same vendor — have traditionally made implementing tiered storage fairly complex, thereby limiting deployments.



Automated storage tiering with IBM Easy Tier can help improve performance at a lower cost by enabling more efficient use of multiple tiers of flash storage or disk drives. IBM Easy Tier automatically identifies more active data and moves that data to faster storage such as flash. This helps enterprises use flash storage for the data that will benefit the most — delivering the maximum benefit even from small amounts of flash storage capacity. IBM Easy Tier can use any supported flash storage to benefit any virtualized storage, delivering greater benefits from flash than tiering systems that are limited to just a single disk system.

## *Replication, snapshots, and mirroring*

Making copies of data can be done in several ways for many different purposes. The word *replication* is used to refer to complete copies of your entire data set. *Snapshots* are quick replicas (copies) created in your database or storage system at specific points in time, rather than just when the data arrives. Synchronous replication occurs immediately when the data arrives at your storage device. Asynchronous replication is done after the data is saved or written to the persistent storage medium. Sometimes this is also called *remote mirroring*, and it enables you to create a copy somewhere else to make a complete “mirror” of your application and its data for failover in time of disaster or site failure. IBM Spectrum Virtualize provides capabilities to perform and manage all these types of data protection functions and apply them across any virtualized system.



IBM FlashCopy is a technology within IBM Spectrum Virtualize designed to create snapshots of active data that can be used for data backups or for other purposes such as testing newly developed applications. IBM Spectrum Virtualize also supports remote mirroring to enable organizations to create and send copies of data to remote locations — including cloud data centers — for disaster recovery. Unlike traditional solutions, IBM FlashCopy and remote mirror can occur between any supported virtualized storage, offering greater flexibility.

## *Thin provisioning*

Traditionally, IT managers monitored the growth of each individual application within their enterprise and manually allocated certain hard disk drives that were dedicated only to accommodate the needs and growth of each specific application. This process ensured that each application would always have adequate storage available, but it often led to a lot of storage sitting idle until data growth caused them to be needed. Idle storage equals added cost and much lower efficiency.

*Thin provisioning* is an alternative where storage is allocated to an application only as it's needed. IBM Spectrum Virtualize automates both the monitoring of storage usage and the allocation of storage capacity resources, dramatically increasing storage utilization rates and efficiency, and reducing labor costs and the time needed to provision storage to a new or growing application.



IBM Spectrum Virtualize is available as software to run on Intel x86 servers and also is at the heart of IBM Storwize family, IBM FlashSystem V9000, IBM SAN Volume Controller, and VersaStack converged systems. It supports almost 400 different storage systems from a variety of vendors, so most likely it supports all the storage you already have.

## *IBM Spectrum Accelerate*

IBM Spectrum Accelerate lets you implement block storage using a different storage architecture from IBM Spectrum Virtualize. Conventional storage systems use storage computers called controllers to manage “boxes of disks” or storage enclosures. Usually, there are two or more controllers so that if one fails, the system can continue operating, though perhaps at lower performance levels. Instead, IBM Spectrum Accelerate uses a grid storage architecture in which essentially identical *building blocks* or *nodes* possess both controllers and storage.

In these highly parallel grid storage systems, adding another building block or node increases storage, controller, and networking capability, so you get more capacity and performance. The nodes “talk” to each other and share storage management



duties, plus they automatically spread or “stripe” the incoming data out across all the storage resources in all the nodes. Essentially, all the nodes in an IBM Spectrum Accelerate solution work together to create one storage system that automatically grows as you add more building blocks.

This has some powerful consequences. For example, no tuning is involved, because the system performs automating data distribution and load balancing. So the performance is at all times consistent and predictable, and without hotspots. The architecture and built-in mechanisms such as self-healing enable the system to offer high availability. Copying data is a non-issue for the same reason — as the data arrives, multiple copies are replicated across the entire system. If one node fails, the data is automatically redistributed and is available from other nodes. And the more nodes you add to your grid system, the more reliable and resilient — and faster — it naturally becomes.



Grid storage provides an excellent example of the benefits of SDS. And thanks to SDS like IBM Spectrum Accelerate, you can buy inexpensive commodity servers, load them with disks or solid-state drives (SSD), connect them with simple networking such as the standard Ethernet, and you have a very capable enterprise storage solution.



Like IBM Spectrum Virtualize, IBM Spectrum Accelerate offers a rich set of proven storage services features — and these are included at no extra charge. It integrates very well with the most popular application-side virtualization products on the market. For enterprises and service providers that seek to leverage heterogeneous server hardware and easily and quickly scale out cost-effective storage for cloud computing environments, IBM Spectrum Accelerate offers many advantages, including

- ✓ **Remote replication:** This is the mirroring mentioned above and used to keep a disaster recovery site filled with current data so that if the main data center fails, the remote site can take over with minimal business impact. IBM Spectrum Accelerate offers asynchronous and synchronous mirroring, and offline initialization.
- ✓ **Multi-tenancy and Quality of Service (QoS):** In the world of cloud computing, one storage device may support the

needs of multiple users or “tenants.” Managing resources in these situations and ensuring that all tenants get the storage capacity and service levels they expect requires extraordinary SDS capabilities, just the sort provided by IBM Spectrum Accelerate.

- ✔ **Snapshots:** Data protection is a top priority for any enterprise-grade storage solution. Coordinating snapshots, storing them with the least possible impact on capacity resources, and managing their use during system recoveries are valuable storage services provided by IBM Spectrum Accelerate.
- ✔ **Monitoring:** A key benefit provided by SDS involves the detailed or “granular” automatic collection of information about your storage system’s operation. This information is reported to IT managers and IBM Spectrum Accelerate uses it to perform other automated storage services such as rebuilding extra data copies after grid nodes fail.
- ✔ **Security:** In addition to the important cloud-related data protection capabilities cited above, IBM Spectrum Accelerate offers Lightweight Directory Access Protocol (LDAP) and role-based access.
- ✔ **Management tools:** IBM Spectrum Accelerate comes with an innovative user interface based on multiple patents. Importantly for hybrid cloud solutions, it includes IBM Hyper-Scale technologies that enable you to grow your storage system as large as you want and yet manage it from one pane of glass, and exercise exceptional flexibility moving volumes from one system to another.



When you add these storage services and other powerful features with the natural benefits of grid storage, the total package adds up to plenty of advantages from IBM Spectrum Accelerate:

- ✔ **Speed:** IBM Spectrum Accelerate enables rapid configuration and deployment of scale-out storage grids in automated environments in minutes rather than days. Built-in management tools and integration with open industry standards such as OpenStack make it a rapid cloud building block that supports automation and orchestration.
- ✔ **Agility:** IBM Spectrum Accelerate operates seamlessly across delivery models — on commodity-choice servers, on integrated hardware arrays, and on public cloud

infrastructure such as IBM SoftLayer and as a service on SoftLayer. This means you can right-fit storage deployment for different hardware, environments, and needs. Its reusable, portable licensing lets you save on future costs with flexible reuse across the IBM Spectrum Accelerate product family and IBM Spectrum Storage Suite, dynamically balance capital and operating expenses, and repurpose servers for optimal use.

- ✓ **Simplicity:** IBM Spectrum Accelerate offers tune-free performance and the ability to dynamically and non-disruptively add storage, managed easily with an intuitive management interface, plus an architecture that eliminates many traditional administrative chores.

IBM Spectrum Accelerate runs as a virtual machine on VMware vSphere, enabling you to build a server-based SAN from commodity hardware. It works by efficiently grouping virtual nodes with the underlying physical disks and spreading the data evenly across the nodes, creating a single virtual array. This solution cost-effectively uses your standard data center network for both inter-node and application host connectivity. IBM Spectrum Accelerate supports any hardware configuration and components that meet the minimum requirements and requires no explicit hardware certification. Scaling of nodes is linear and non-disruptive.



When you want to build a cloud storage solution with the confidence of a proven storage heritage and capabilities that can grow easily, cost effectively, and flexibly, IBM Spectrum Accelerate is an excellent choice. IBM Spectrum Accelerate also offers compelling combined private, public, and hybrid solutions, including business continuity, as a member of a family of products (IBM FlashSystem A9000 and A9000R, IBM XIV Storage System, and IBM Spectrum Accelerate) using the same technology and management system.

## *IBM Spectrum Scale*

IBM Spectrum Scale is the IBM Spectrum Storage family member devoted to helping you easily handle the several different types of data that your business activities will generate, such as data files and even the newer data “objects.” In the broadest sense, data exists in two basic forms — structured

and unstructured. Though IBM Spectrum Scale can certainly manage structured data used in traditional databases, it's especially good at managing unstructured data like the documents you build, the emails you send, and the photos you take.

Structured data, like your home address, for example, can naturally be managed using a tabular format — columns and rows. Decades ago, software called databases was developed to manage structured data and help us gain much more value from it quickly. By using databases, you could sort structured data in many different ways, search it quickly, even analyze it to discover important trends, as well as serve up the exact data that you want from within enormous data stores, almost instantaneously.

But the bits and bytes that make up your digitized family vacation photograph aren't so easily manipulated by using a database. Instead, all the data related to the image is simply collected and called a *file*. The chapters of this book are stored as files, as are the emails you sent today and the presentation you gave at the recent business meeting. To provide some structure to files so that they could be sorted, searched, and even analyzed in some ways, software called file systems was developed. File systems use a hierarchical organization structure to facilitate their operations, with files grouped into folders, into directories, and so on.

File systems are very common today and used to facilitate countless types of data processing. For example, to sequence a human genome, the data is separated into files — thousands and even millions of them for a single human being. One of the major advances enabling genomes to be sequenced in hours instead of months has been the ability of storage systems to feed many files simultaneously into powerful computing resources such as mainframes. The ancestor of IBM Spectrum Scale was developed to perform exactly this type of high performance file system management, which means that it has features and capabilities far beyond those of conventional file systems.

IBM Spectrum Scale started out as a high performance file system and has evolved into so much more. Today it's a full-featured set of file data management tools, including advanced storage virtualization, integrated high availability, automated tiered storage management, and high performance

configurations to effectively manage very large quantities of file data. IBM Spectrum Scale is designed to support a wide range of application workloads using a variety of access languages or “protocols” and has been proven extremely effective in very large, demanding environments.



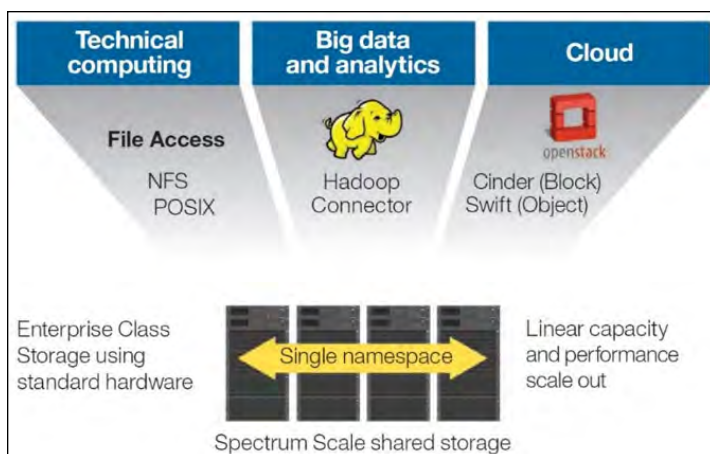
IBM Spectrum Scale is true SDS because it can be loaded onto any appropriate hardware. It gets its name from the fact that it uses a grid-like architecture similar to that employed by IBM Spectrum Accelerate. This means that you can very simply add more servers and storage and IBM Spectrum Scale will incorporate them automatically into a single storage resource and spread both management and data across all available components or nodes to maximize performance, efficiency, and data protection.

Unlike other storage solutions that must implement separate add-on systems to handle files, data objects, blocks, or big data analytics with Hadoop, all data is commonly stored, regardless of the way it is accessed, in IBM Spectrum Scale. Once part of the core file system, block, file, and object-based data can now be accessed and managed in essentially the same ways. And an impressively wide range of data storage services and features can be applied across all the data, as appropriate, including snapshots, information life cycle management, storage tiering, asynchronous or synchronous data replication, and some unique data protection strategies.

Both files and objects have information about themselves, called *metadata*, associated with the data they contain, but objects are characterized by their extended metadata. Each object is assigned a unique identifier that allows a server or end-user to retrieve the object without needing to know the physical location of the data. This approach is extremely useful for automating and streamlining data storage in cloud computing environments.

Object storage is often compared to valet parking at an upscale restaurant. When customers use valet parking, they exchange their car keys for a receipt. They don't know where their car will be parked or how many times an attendant might move the car while they're dining. In this analogy, an object is the car itself, and a storage object's unique identifier represents customers' valet receipts.

Unlike other SDS products, especially those in the object storage market, IBM Spectrum Scale offers native, high-performance and scalable access to block, file, and object data via almost all the standard object storage protocols, including OpenStack Swift, Amazon S3, CIFS, NFS, HDFS, or POSIX. With IBM Spectrum Scale, multiple systems and applications can share common pools of storage, allowing you to transparently administer the infrastructure without disrupting applications. This process is shown in Figure 4-2. IBM Spectrum Scale is platform-independent, so it can run on IBM Power Systems and on x86 systems, along with storage from IBM and other vendors. This flexibility can reduce costs and improve energy efficiency.



**Figure 4-2:** IBM Spectrum Scale provides a single SDS solution across a variety of application types.



Because of its high performance computing heritage, IBM Spectrum Scale offers many features found in no other enterprise-grade data management system. With IBM Spectrum Scale, you can

➤ **Accelerate file and object storage performance:**

IBM Spectrum Scale provides parallel access to data and shared storage, improving scalability for high-performance workloads. Data and metadata (the data that describes individual files) flow from the storage nodes managed by IBM Spectrum Scale to all the storage in parallel under the control of a sophisticated

distributed lock manager that prevents file corruption by limiting changes to one user at a time.

- ✔ **Simplify data management:** IBM Spectrum Scale includes integrated tools to help you control costs or manage growth to petabytes of data and billions of files. This is accomplished within a single overall resource pool or namespace that is easy to administer and can be scaled quickly, as desired, by simply adding more scale-out resources — eliminating the problem of “filer sprawl” caused by the need to add whole new file systems each time you need more capacity. IBM Spectrum Scale automatically spreads file data across multiple storage devices to most effectively utilize all available storage and deliver high performance. IBM Spectrum Scale has a common management interface that is easy to use and operations can be managed from any node in the IBM Spectrum Scale system or cluster, including standard file system administration functions such as user quotas, snapshots, and storage management.
- ✔ **Empower global collaboration:** IBM Spectrum Scale provides low latency access to your data from anywhere in the world with Active File Management (AFM) distributed caching and routing technology. AFM expands the IBM Spectrum Scale global namespace across geographical distances, providing fast read and write performance with automated namespace management from anywhere in the world. As data is written or modified at one location, all other locations get the same data with minimal delay. AFM leverages the inherent scalability of IBM Spectrum Scale, providing a high-performance, location-independent solution that masks failures and hides wide-area latencies and outages. These capabilities accelerate project schedules and improve productivity for globally distributed teams.

## *Manage the information life cycle*

IBM Spectrum Scale enhances information life cycle management (ILM) and lowers your data management costs significantly by using multiple tiers of storage, including tape and cloud storage. ILM is a comprehensive approach to managing the flow of data and associated metadata from creation and initial storage to the time when it becomes obsolete and is deleted. Policies cover all aspects of dealing with data,

starting with user practices, rather than just automating storage procedures. IBM Spectrum Scale enables more complex criteria for storage management than simply data age and frequency of access.

With the powerful policy-driven automation and tiered storage management provided by IBM Spectrum Scale, you can create optimized tiered storage pools by grouping storage media based on performance, locality, or cost characteristics. Data migrated to tape or cloud remains visible and directly available to end-users on any protocol. Migration policies that you establish and can change as needed transparently move data from one storage pool to another without changing file location in the directory structure.

For example, you can create a rule for thresholds that moves files out of the high performance pool if it's more than 80 percent full, thereby mitigating potential bottlenecks in the high performance pool. IBM Spectrum Scale ILM capabilities and benefits include

- ✔ Policy-driven automation and tiered storage management
- ✔ Flexibility to match the cost of storage to the value of data
- ✔ Storage pools/tiers that include high-performance SSDs, high-speed hard disk drives, and high-capacity hard drives
- ✔ Full integration with other IBM Spectrum Storage family members such as IBM Spectrum Archive, IBM Spectrum Control, and IBM Spectrum Protect.

### ***Ensure data availability, reliability, and integrity***

IBM Spectrum Scale provides easy system scalability, very high availability (the system almost never goes offline for any reason), and enterprise-grade reliability with no single-point-of-failure for storage infrastructures. You can configure IBM Spectrum Scale so file data remains available automatically even if a data center fails.



IBM Spectrum Scale RAID uses advanced erasure coding to disperse data across multiple drives and systems for reliability. This technology eliminates the cumbersome multi-day rebuild times that are common with today's multi-terabyte disk drives. In addition, IBM Spectrum Scale RAID ensures data integrity from the drive to the compute client with advanced checksum protection. This is critical because more and more data is kept for longer periods on larger devices. For additional reliability, IBM Spectrum Scale supports snapshots, synchronous and asynchronous replication, and asynchronous error diagnosis while affected input/output (I/O) operations continue.

## ***Support Hadoop***

Hadoop and its central technology, called MapReduce, is a popular big data analytics software product. IBM Spectrum Scale features a Hadoop connector that allows IBM Spectrum Scale to act as a drop-in replacement for the Hadoop Distributed File System (HDFS). In this way, commercial and open source analytics applications are transparently supported without change, including InfoSphere BigInsights, Apache Hadoop, Spark, and other big data analytics products. Unlike HDFS, IBM Spectrum Scale enables your existing computational infrastructure to perform big data analytics without having to purchase dedicated infrastructure. Data movement into and out of a dedicated analytics silo is eliminated, thereby speeding results. IBM Spectrum Scale goes beyond HDFS by providing enterprise-class data protection, efficiency, and full data life cycle management features.

## ***Enable OpenStack cloud SDS***

OpenStack is a family of open source software to deploy cloud services, virtual machines, and services quickly. OpenStack has multiple types of storage to support virtual machines from block devices to file services. IBM Spectrum Scale supports the entire portfolio of OpenStack storage, including Cinder, Swift, and Manilla. OpenStack gives IT architects building public, private, and hybrid clouds access to the features and capabilities of the industry's leading enterprise scale-out SDS. IBM Spectrum Scale unifies virtual machine (VM) images, block devices, objects, and files within a single

namespace no matter where data resides. IBM Spectrum Scale can add policy-based data placement to balance performance and cost by putting data in the best location, on the best tier, at the right time — all in software — on heterogeneous, commodity, industry-standard hardware.

## *IBM Spectrum Archive*

Digitized information has been stored on long ribbons of tape since the dawn of computing. Over the decades as new storage technologies have appeared, such as spinning hard disk drives and more recently flash solid-state storage, many industry analysts have predicted that tape storage would go extinct. But it hasn't. And thanks to SDS solutions like IBM Spectrum Archive, tape storage remains a cornerstone of storage architecture for many organizations and is a critical tool for emerging business solutions where cloud and analytics workloads are essential.

A key reason why tape storage continues to offer great value to many enterprises is the fact that the vast majority — perhaps as much as 80 percent — of the data your business generates will very seldom or never be accessed after a month or two. But for a variety of reasons, including governmental regulations and also opportunities to mine it later for valuable business insights, your data still must be stored, sometimes for decades. These seldom-used or “archived” data stores can become enormous and very expensive. Even very low cost disk storage can cost around six cents per month per gigabyte (GB) of archived data stored, but tape costs around two-tenths of a penny/month/GB — approximately 30 times less! This is why tape storage in general and IBM Spectrum Archive in particular remain extremely valuable to many enterprises.

Built on IBM Linear Tape File System (LTFS) technology, Spectrum Archive provides direct, intuitive, and graphical access to data stored on Linear Tape-Open (LTO) tape and IBM enterprise tape cartridges used in IBM tape drives and libraries. It eliminates the need for additional tape management and software to access data stored on tape.

IBM Spectrum Archive enables automatic data movement between flash/disk and tape to lower costs without the need for proprietary tape applications.



With IBM Spectrum Archive you can do the following:

- ✓ Access and manage files on tape as easily as if they were on hard disk
- ✓ Drag and drop files to and from tape without requiring device-specific software
- ✓ Share file data across platforms, similar to using a USB drive



LTFS is a self-describing tape format originally developed by IBM. The Storage Networking Industry Association (SNIA) has named IBM and Oracle as co-chairs of the LTFS Technical Work Group (TWG) working with other storage industry members, including HP and Quantum. The LTFS TWG is working toward adoption of an open LTFS industry standard recognized by the American National Standards Institute (ANSI) and International Organization for Standardization (ISO).

IBM Spectrum Archive is a significant step forward in moving tape storage away from its reputation as complex and difficult to use. Unlike disk storage, tape has previously required device-specific software to read, write, and manage data stored on its cartridges in libraries. But IBM Spectrum Archive sets a new standard for ease of use and portability for open-systems tape storage. Users can run any application designed for disk files against tape data without concern for the fact that the data is physically stored on tape.

IBM Spectrum Archive uses the file system's format and operating system (OS) resources to graphically display the contents of a tape cartridge in the OS graphical user interface (GUI) format, typically a folder/tree structure. The metadata of each cartridge, once mounted, is cached in server memory. Metadata operations, such as browse directory and filename search do not require tape movement.

Similar to using a USB drive or memory stick, IBM Spectrum Archive enables users to share data across platforms by simply dragging and dropping files. Access to data is faster and easier — users just load a tape into the drive and mount it into the file system, and files become visible as if they were on a disk. Because access is not through backup software, cartridges formatted with IBM Spectrum Archive can be exchanged more easily between users working in different

operating systems, using different software, and in different locations. Users' operational agility and efficiency is increased because using IBM Spectrum Archive doesn't require knowledge of tape.

IBM Spectrum Archive gives organizations an easy way to use cost-effective tape drives and libraries within a tiered storage infrastructure. By using tape libraries instead of disks for data that is stored for long-term retention, organizations can improve efficiency and reduce costs.

IBM Spectrum Archive helps organizations

- ✔ Create operational storage tiers with tape rather than storing static, unchanging files on costly disk storage
- ✔ Archive digital assets for the long term so assets can be referenced and monetized for years to come
- ✔ Create copies of data on operational storage, improving the efficiency and cost effectiveness of a tiered storage infrastructure



IBM Spectrum Archive seamlessly integrates with the scalability, manageability, and performance of IBM Spectrum Scale, enabling organizations to move beyond simply adding storage — to optimizing data management.

## Chapter 5

---

# Ten Use Cases for Software-Defined Storage

.....

### *In This Chapter*

- ▶ Looking at innovative uses for SDS in your organization
- .....

**T**his chapter introduces ten (okay, seven) different ways that enterprises of all sizes and types can leverage the power of IBM Spectrum Storage solutions and technologies to accomplish their storage objectives and solve real-world business challenges.

## *Partnering with Flash*

Before software-defined storage (SDS), the deployment of flash storage was sometimes challenging. For example, manually moving data sets to and from flash storage was slow and labor-intensive. The very notion of this chipped away at the intrinsic cost-related value of flash. But when you add intelligent, automatic storage tiering provided by IBM Spectrum Virtualize (covered in more detail in Chapter 4), data quickly moves around to the most advantageous storage medium, based on activity levels or other policies you choose. Suddenly the benefits of flash are magnified. Your applications get the performance they crave — but only when needed — and your budget gets a break when the high performance of flash isn't needed and disk or tape work fine.

## *Managing from the Cloud*

IBM Spectrum Storage offers options and capabilities to help you plug into the benefits raining from the cloud. An IT service that you might not imagine buying from the cloud is storage management. This is the realm of IBM Spectrum Control (see Chapter 3 for more info) and implementing cloud-based storage management offers the same types of advantages such as unlimited capacity and pay-as-you-go convenience that you'd expect from a more traditional cloud offering.

IBM Spectrum Control Storage Insights cloud-based solutions provide visibility to help you manage today's complex storage infrastructures. Combining IBM's proven data management leadership with proprietary analytics from IBM Research and Cloud delivery expertise, Storage Insights helps you quickly and easily lower costs by optimizing your storage environment.

You deploy the Storage Insights functionality from IBM's own public cloud service, called IBM SoftLayer. To run Storage Insights, you simply download a data collector, point the collector to your storage devices, and it will start sending data about your storage environment to the analytics engine in the cloud within minutes.

## *Enabling Hybrid Cloud*

Beyond the benefits inherent in its grid architecture, IBM Spectrum Accelerate provides features that make it an effective tool for building hybrid cloud storage. For example, by using tools that are members of the IBM Hyper-Scale family, you can automatically replicate data to off-site IBM Spectrum Accelerate or IBM XIV storage instances. You can use the Hyper-Scale toolset to scale your storage resources almost forever, to build a disaster recovery solution, and to enable hybrid cloud.



A hybrid cloud storage solution built by using IBM Spectrum Accelerate offers many advantages:

- ✔ You use and pay for the storage resources you need, when you need them.
- ✔ You don't pay for all the disks, tape, or flash needed to store this data in your own data center.
- ✔ Your on-premises storage may be mostly or entirely flash-based, under IBM Spectrum Accelerate, and provide extremely high performance with lower operational costs in power, cooling, floor space, and so on.

## *Protecting Data from the Cloud*

To help protect all your data onsite and off, reap the benefits of IBM Spectrum Protect delivered from the cloud (see Chapter 3 for more about IBM Spectrum Protect). Advanced cloud data protection capabilities available in IBM Spectrum Protect solutions include the following:

- ✔ A cloud portal that enables multi-tenancy and billing
- ✔ Cloud-based backup storage pools that can provide flexible capacity
- ✔ Automatic fail-over for system restores, which enables high availability of critical services
- ✔ A flexible deduplication architecture, which helps customers control costs
- ✔ An incremental “forever” design that eliminates the cost and disruption of periodic full backups

## *Archiving Data to the Cloud*

Believe it or not, one of the oldest storage mediums, tape, has actually found an effective new business use case in the most modern storage environment — the cloud. The IBM Spectrum Archive (covered in more detail in Chapter 4) cloud delivery model lowers costs while addressing the explosion of data volumes.



IBM Spectrum Archive provides hierarchical storage management for tape storage, including cloud-based tiers. Powerful software-defined, policy-driven automation dynamically moves data through tiered storage pools, based on performance or cost requirements. Thanks to IBM Spectrum Archive, data migrated to tape remains visible in your file system and is directly accessible to end users while lowering storage costs by up to 80 percent.

## *Simplifying Big Data Analytics*

If you're involved in the IT management or decision-making of an enterprise, you know that if you aren't already using some form of big data analytics, then sooner or later you will. Of course, SDS is playing a big role in big data analytics, which means there's an excellent use case to highlight with IBM Spectrum Scale.

Both the Hadoop Distributed File System (HDFS) and IBM Spectrum Scale provide the basic storage tools needed for analytics workloads, but that's where the similarities end. IBM Spectrum Scale includes File Placement Optimizer (FPO), a distributed computing architecture where each server is self-sufficient and utilizes local storage. Compute tasks are divided between these independent systems and no single server waits on another. IBM Spectrum Scale FPO provides higher availability through advanced clustering technologies, dynamic file system management, and advanced data replication techniques.

In addition, IBM Spectrum Scale offers a wide range of enterprise data storage features, such as snapshots, backup, archiving, tiered storage, data caching, Wide Area Network data replication, and management policies. IBM Spectrum Scale can be used by a wide range of applications running analytics workloads while also accessing other unstructured file data, an advantage, among many others, that HDFS just can't offer.



## *Unifying File and Object Storage*

IBM Spectrum Scale can be used in a tightly integrated solution with other SDS tools to tackle the unique challenges of object storage. For this task, you can add OpenStack Swift to the mix. Swift is an emerging open source object storage software platform that's widely used for cloud storage. The robustness and features of IBM Spectrum Scale combined with OpenStack Swift object extensions provide an enterprise-grade object store with high storage efficiency, tape integration, wide-area replication, transparent tiering, and snapshots — capabilities most object-based storage offerings can't match today. OpenStack on IBM Spectrum Scale delivers compelling efficiencies in a single unified storage solution that can support object and file access to the same data with robust and efficient IBM Spectrum Scale data protection. OpenStack Swift object storage on IBM Spectrum Scale can reduce the amount of raw storage you need to use compared to other object storage solutions.

Unifying your object and file storage with IBM Spectrum Scale offers significant cost and efficiency advantages for your overall storage solution. You can implement one storage architecture to address all your storage requirements. This single solution can bring your entire enterprise information store under one namespace; replicate it to any geographical location according to policies you set; support your entire range of applications — from databases to rich media file serving; and then make it all faster and lower cost.





## Reduce costs by using storage controlled with software

Traditional storage systems have become costly bottlenecks for enterprises struggling with ever-growing data challenges. This book explains how software-defined storage enables organizations to significantly reduce their storage costs while improving performance, reliability, and scalability with intelligent software that performs essential storage functions.

- **Increase flexibility** — *traditional storage systems limit your options and lock you in to a rigid and undadptable solution*
- **Simplify management** — *automated policy-driven storage management makes it easy to implement policies for information life cycle management and other storage administration tasks*
- **Empower global collaboration** — *low latency access to data from anywhere in the world to enable innovation and increase productivity*



Open the book and find:

- **Software-defined storage systems that meet your business needs**
- **How to resolve performance bottlenecks that exist in your storage infrastructure**
- **Turnkey software-defined storage solutions that are ready to deploy**
- **Ten use cases for software-defined storage**

**Making Everything Easier!™**

Go to **Dummies.com**<sup>®</sup>  
for videos, step-by-step examples,  
how-to articles, or to shop!

# **WILEY END USER LICENSE AGREEMENT**

Go to [www.wiley.com/go/eula](http://www.wiley.com/go/eula) to access Wiley's ebook EULA.