# Software-Defined Storage

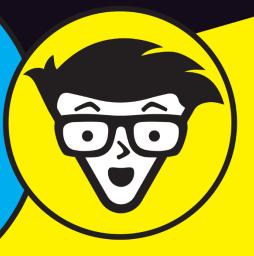## For dummies®
A Wiley Brand



Control
storage costs

—

Enable hybrid
cloud

—

Simplify storage
management

Chris Saul

# Software-Defined Storage

3rd IBM Limited Edition

by Chris Saul

for dummies®

A Wiley Brand

# Software-Defined Storage For Dummies®, 3rd IBM Limited Edition

## Publisher's Acknowledgments

# Table of Contents

# Introduction

New technologies and powerful global events are combining to change Information Technology (IT) in many ways. These days, most enterprises use resources from multiple cloud service providers. Business applications can reside in data centers or in the cloud — or both at the same time. Company employees work from home in greater numbers than ever before. And artificial intelligence (AI) technology is ever more prevalent, affecting many aspects of your life.

But none of these changes happen, no business turns a profit, and nobody buys a pair of socks online unless the underlying IT systems that move and store data operate effectively, efficiently, and securely. This is why the concept of *software-defined storage* (SDS) is so important. Today, SDS solutions are woven into the fabric of your life because they help enable so many things that we do. SDS can make online shopping safer, help companies offer better ways for employees to work from any location, accelerate research into new vaccines or climate change, and so much more. In the chapters ahead, you discover useful definitions of SDS and related technologies, find out how SDS works and what it does, and meet a family of SDS solutions that helps your business thrive in the 21st century.

## About This Book

*Software-Defined Storage For Dummies*, 3rd IBM Limited Edition, tells the story of how modern enterprises leverage the power of this rapidly evolving technology to help them achieve their business and IT objectives. This book introduces IBM Spectrum Storage, a leading family of SDS solutions. You also get plenty of information and ideas on how this highly capable software from IBM can help your business be a market leader.

## Icons Used in This Book

Throughout this book, you occasionally see special icons that call attention to important information. Here's what you can expect.

**REMEMBER**

This icon points out information that may well be worth giving your full attention.

**TIP**

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

**TECHNICAL STUFF**

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

# Beyond the Book

Although this book is chock full of information, there's only so much I can cover in these pages. If you find yourself at the end of this book thinking that you want to learn more, visit `ibm.com/storage`. There, you can find more information about SDS and IBM Spectrum Storage software.

Chapter **1**

# What Is Software-Defined Storage?

I n this chapter, I provide a simple and useful definition of software-defined storage (SDS), discuss the major functional types of this technology, and introduce the market-leading IBM Spectrum Storage family of SDS solutions.

## Defining SDS

In the most traditional IT architectures, computer programs or applications transform raw data into business value by process-ing data in various ways. But the raw data doesn't normally reside in the application; it's stored elsewhere in a storage device. The application in the traditional IT architecture requests or queries the needed data from the storage device, which responds by pro-viding the requested data.

But imagine placing software in the data path between the appli-cation and the storage device. From the application's perspective, nothing has changed — data is returned when queries are made. But from the storage perspective, everything has changed. This is the basis of SDS.

With a layer, or perhaps many layers, of software between applications and storage, the type of storage device may actually become somewhat irrelevant, and the devices themselves become interchangeable. The software maintains the location of each unit of data and can retrieve it, no matter where it physically resides. Separating or "uncoupling" the physical storage devices from the storage software offers many benefits and advantages. This arrangement also provides one of the most popular definitions for SDS — storage management software not dependent on any particular underlying storage devices.

But SDS has evolved into so much more. Another way to think about SDS is to imagine it as a layer of intelligence inserted at strategic points in the data path between applications and their data. With smart software setting up residence between the relatively passive physical storage media such as flash, disk, or tape and the rest of the IT infrastructure, a wide world of possibilities suddenly opens.

Imagine that you want to protect your data by making copies of it and sending those copies to other systems or locations. And, of course, you want to encrypt that data as it comes and goes and while it's hanging around. Imagine the cost savings that would accrue from monitoring the access patterns of your data and then storing only the most frequently accessed small portion of it on the fastest and most expensive storage, then sending the rest of it off to the public cloud or an object store or tape costing a small fraction of any other storage medium. You may even set up policies to retrieve data at a trigger event like evening backups, monthly reports, or for training an application based on artificial intelligence (AI). You want to get really smart? Imagine something that learns to predict what data you want and retrieves it ahead of time from that really inexpensive tape system so you don't have to wait an IT eternity for the appropriate tape cartridge to load.

**REMEMBER**

That's smart. That's SDS today. And that's the definition I use for SDS — the layer(s) of intelligence and functionality deployed between application hosts and physical storage devices.

This definition complements many of the other, more traditional SDS definitions. Once implemented, SDS is what responds to requests for data from applications, not the hardware that holds the actual bits and bytes. Instead, SDS manages the backend storage devices, retrieving data from them as requested, moving data to optimize business and IT objectives, encrypting it,

monitoring it . . . the list of activities is long and growing. So yes, under this definition almost any storage hardware — call it *commodity storage* — can be used behind smart SDS.

An IT strategy based on SDS leads to modern hybrid cloud storage environments becoming common today. Software separated from hardware allows the best configuration without compromises, and at the best available price. This approach is also beneficial for hardware upgrades over time, allowing enterprises to go through different hardware generations without paying for the software again.

# INTRODUCING CONTAINER-NATIVE STORAGE

As organizations continue to adopt hybrid cloud strategies, making sure that applications can be deployed in any environment becomes critically important. For example, hybrid cloud solutions can span on-premises traditional infrastructure, on-premises or hosted private clouds, and one or more public clouds. Nobody wants to rewrite applications based on where they're deployed — that's why container technology is now so important. Containers provide a lightweight environment that operates across all the different operating systems and deployment platforms. Container management software such as Kubernetes or Red Hat OpenShift provides orchestration of containers to simplify application deployment, scaling, and management in all manners of environments.

When you have an application, you need storage for its data. With the Container Storage Interface (CSI) application programming interface (API), containerized applications can use existing storage, including SDS. By using this approach, you can start to deploy containerized applications with the high-availability, management, and cost-saving benefits already inherent in your existing storage. Container-native storage (CNS) provides an alternative that's completely managed from within Kubernetes or OpenShift. In essence, CNS is SDS running in a container environment alongside applications. This approach means that Kubernetes administrators can use the same management tools for applications that they use for container-native storage software, which simplifies deployment. Container environments are designed to be easy to scale by adding or removing servers quickly and easily to increase or decrease storage capacity as needs change. CNS solutions scale in the same way, alongside the applications using them.

**REMEMBER**

SDS is just another application. It can be downloaded and deployed when you want, as you like, as long as you can accommodate its modest processing, networking, and storage capacity requirements. This accommodates another common definition of SDS — storage-related products that can be upgraded and/or downloaded and deployed in a way that's not hardware dependent and provides flexibility in deployment options.

The real issue with the definition of SDS is that once you begin to imagine it, the possibilities seem endless, perhaps even chaotic. SDS can provide almost any data service you want (it's rapidly moving toward that point). To simplify things and bring some coherence to SDS conversations, you discover useful divisions into which you can place various types of SDS products and tools in the later section "Discussing the Major Types of SDS."

# Discussing the Major Types of SDS

Think about what SDS does. One role that SDS tools and solutions can play is to monitor, assess, manage, and control the storage process and all its components. Perhaps most commonly, SDS performs a wide and ever-expanding range of particular tasks and services related to protecting, storing, moving, and providing access to data. And SDS products and solutions can give the basic organizational characteristics and capabilities to the underlying foundational storage function itself. From these simple observations, SDS can be placed into three basic categories.

**TIP**

SDS is easier to understand when you think of it as belonging to one of these three basic types — management and data discovery, data resilience, or storage infrastructure. But not just any SDS product you buy that provides management services will be compatible with any data protection solution or SDS operating as basic storage infrastructure — unless the products have been purposely and carefully designed and built to work well together.

## Management and data discovery

Do you know how much storage capacity your organization is using right now? How fast has it grown recently? What percentage of unused capacity do you have in reserve? What are the

access patterns? Do these patterns spike at the end of the month, or perhaps every workday morning? Honestly, where do each of your data sets and all their copies reside? Could you save money or increase application performance, staff productivity, or customer satisfaction if these data sets lived on different storage?

If you don't know the answers to these questions, your organization needs better management services from your SDS. These questions don't scratch the surface of what SDS management tools and solutions can do. This type of SDS is important, and yet many IT professionals don't even think about management and control when they consider SDS. Check out Chapter 3 for more info about storage management and data discovery solutions.

## Modern data resilience

The list of what you want to do to and with your data grows longer each day. But protecting your data from accidental or malicious corruption, loss, or attack tops almost everyone's list of priorities. SDS offers many powerful, efficient, and effective ways to provide data resilience and related services. In fact, data protection is so important that it demands and deserves its own category within the SDS discussion.

Data resilience strategies and solutions affect essentially every other aspect of your IT environment, including how you manage data growth and which solutions you choose to ensure that your business and its applications are always running and available to your customers.

**REMEMBER**

Data protection goes beyond simple backup and recovery. How well does your data protection solution reduce storage requirements? Is it easy to add tape and cloud storage resources for cost-effective data archival? Can it detect and protect your data from malware or other cyber attack? You can't simply call up your local cloud provider and turn on all the services you want with the performance and cost you need. In fact, effectively integrating public cloud resources can be complex and risky, depending on the quality of your SDS data protection and data services solutions. Head to Chapter 4 to realize the protection you can get from a modern data resilience solution.

# Basic storage infrastructure

All data is simply bits and bytes, but as soon as you gather enough bits and bytes, or your data is all related, such as the pixels of an image or video, it begins to exhibit certain characteristics, which you use for descriptive or organizational purposes. You start to refer to data as structured or unstructured, block, file, or object. Your structured data is the information spinning through databases and managed and manipulated in tabular form. But across the universe of information in general, around 80 percent of data is described as unstructured — documents, images, audio and video files, emails, and so forth. To make this data easier to manage, it's organized into files and uses special software called *file systems.* And finally, there are certain types of files that carry along with them extra information about themselves, or metadata, so they can be managed and searched without relying on traditional hierarchical organizational structures. These are known as objects. Each type of data — block, file, or object — needs its own special storage software, although a few products out there can effectively handle all three types.

This software sits closest to the storage hardware and storage media and helps form what might be called your basic storage infrastructure. It is often integrated right into the storage device itself as part of the controller functionality, but it doesn't need to be. These infrastructure-level solutions can operate just fine as legitimate SDS — downloadable, upgradable, and deployable simply as software not dependent on any particular hardware.

Whether deeply integrated with a particular device or uncoupled entirely, this infrastructure SDS has grown enormously powerful and capable over the past decade or so. Now this type of SDS can encrypt and copy data and send it out to the cloud as part of your hybrid cloud. If you want to know more about solutions that work for your basic foundational storage infrastructure functionality and capabilities, check out Chapter 5.

Chapter **2**

# Introducing IBM Spectrum Storage

IBM Spectrum Storage is a comprehensive family of software-defined storage (SDS) solutions that can help your business by improving the way you manage, protect, and store data. In this chapter, I provide a brief introduction to the family and also discuss some options for how you can acquire this software.

## Getting the Bigger Picture

IBM has consolidated existing solutions and continues to develop new technologies within a single, highly integrated SDS family with one consistent user experience. This SDS product family is called IBM Spectrum Storage.

IBM Spectrum Storage includes many individual SDS solutions. The family offers a wide range of SDS functionality and capabilities to address your storage objectives. Figure 2-1 shows the current members of the IBM Spectrum Storage family organized into SDS functional categories. You also see the icon that represents each solution.

IBM Spectrum Storage helps you manage all your data, of all types, wherever it resides, with a comprehensive portfolio of SDS applications. You can unify your storage across on-premises and hybrid cloud environments, leverage the power of the family to more easily and effectively implement and manage important business tools such as analytics and AI, and reduce costs while increasing business agility. IBM Spectrum Storage solutions also offer licensing options that provide many benefits and advantages over purchasing and deploying separate products from different vendors.

| Storage Management and Data Discovery | IBM Spectrum Control | |
| | IBM Storage Insights | |
| | IBM Spectrum Discover | |
| Modern Data Resilience | IBM Spectrum Archive | |
| | IBM Spectrum Protect | |
| | IBM Spectrum Protect Plus | |
| | IBM Spectrum Copy Data Management | |
| Storage Infrastructure | IBM Spectrum Virtualize | |
| | IBM Spectrum Scale | |
| | IBM Cloud Object Storage | |

**FIGURE 2-1:** The IBM Spectrum Storage family.

IBM Spectrum Storage solutions can be deployed in many ways. Some of them are integrated into IBM storage systems such as IBM FlashSystem so you get plenty of SDS features and functionality built right into your storage system. But they can also be implemented as software-only solutions running on a server in your data center. And many IBM Spectrum Storage offerings can run in the cloud so you gain their capabilities and features without requiring management and maintenance.

In the next three chapters, I introduce each of the current IBM Spectrum Storage family members and discuss what each does, why you might want to deploy that particular solution, and what benefits and advantages your enterprise might gain from doing so.

> Visit the IBM Spectrum Storage website at `www.ibm.com/products/software-defined-storage` to get the latest information.

**TIP**

# Acquiring IBM Spectrum Storage

You can acquire IBM Spectrum Storage SDS software in the conventional way: license whichever products you need. But in a rapidly changing world, "suite" collections provide greater flexibility and allow you to quickly switch from one product to another as your needs change.

## IBM Spectrum Storage Suite

Part of the challenge of managing your storage environment involves effectively and efficiently managing the SDS tools and solutions you employ. IBM addresses this issue by offering IBM Spectrum Storage Suite. IBM Spectrum Storage Suite gives you unlimited access to members of the IBM Spectrum Storage SDS family and IBM Cloud Object Storage software with licensing on a flat, cost-per-terabyte basis. It makes pricing easy to understand and predictable as capacity grows. Structured specifically to meet changing storage needs, the Suite is ideal for organizations just starting out with SDS, as well as those with established infrastructures that need to expand their capabilities. IBM Spectrum Storage Suite enables you to add software without raising costs because licenses are calculated based on the amount of storage capacity you're managing, not the number of software products you're using.

**TECHNICAL STUFF**

Containers are an open-source technology that lets applications be packaged with all the elements needed to run in any environment. Containers offer the versatility of virtual machines — but at a much smaller footprint and cost. Among their many benefits, containers allow existing or legacy applications to run in almost any host environment without being rewritten — an enormous advantage for enterprises with numerous applications, each composed of thousands of lines of code. Because of these benefits, containerization is a key enabling technology for flexibly delivering application workloads to private and public cloud. Tools that allow easy utilization of various container technologies and capabilities are woven into many of the IBM Spectrum Storage solutions, including IBM Storage Suite for IBM Cloud Paks.

# IBM Storage Suite for IBM Cloud Paks

To facilitate deployment of containerized workloads and development of new cloud-native applications, companies large and small are modernizing around software products such as Red Hat OpenShift. IBM has developed a series of software tools called IBM Cloud Paks designed to enhance and extend the functionality and capabilities of Red Hat OpenShift. IBM Cloud Pak solutions give enterprises a modular and easy way to bring important applications into modern, cloud-based environments.

To support IBM Cloud Paks, IBM has developed a version of the IBM Spectrum Storage Suite called IBM Storage Suite for IBM Cloud Paks. This particular version of the Suite brings enterprise data services to container environments with resources for file, block, and object storage that address almost any workload requirement.

The comprehensive suite is composed of a number of IBM and Red Hat SDS offerings, including both container-native and container-ready storage. Packaging these products together provides a convenient way to license the software you need to support container workloads. It provides a flexible menu of data management and storage solutions that help automate and enhance the integration of underlying storage and data management resources.

# Chapter **3**

# SDS for Storage Management and Data Discovery

The IBM Spectrum Storage family has many components. In this chapter, I introduce three members that provide various types of storage management and data discovery functionality. Figure 3-1 shows you the breakdown of these members, including the icons IBM uses to represent them.

| Storage Management and Data Discovery | IBM Spectrum Control | |
| --- | --- | --- |
| | IBM Storage Insights | |
| | IBM Spectrum Discover | |

**FIGURE 3-1:** IBM Storage Management and Data Discovery solutions.

## IBM Spectrum Control

Most likely, your data storage environment has been built up piecemeal. You have a variety of storage systems of different ages, from different vendors, each with its own user interface and management software. Within your complex storage ecosystem,

you must monitor the performance, capacity, and other characteristics of perhaps hundreds or even thousands of disks, flash drives, and even tape libraries. Accomplishing everything seems impossible, but it's just another day at the office for IBM Spectrum Control, a software-defined storage (SDS) solution that provides storage management services.

**REMEMBER** IBM Spectrum Control is a comprehensive, end-to-end, on-premises data and storage management solution that monitors, automates, and analyzes multi-vendor storage environments. It helps manage the performance, availability, and capacity utilization of storage systems, file systems, and databases.

IBM Spectrum Control provides a single point of control that helps administrators manage every aspect of the storage infrastructure — from the application hosts, through the storage network fabric, and down to the physical drives — across multisite storage environments, including both IBM and non-IBM systems.

You can use IBM Spectrum Control to perform everyday storage management tasks such as capacity provisioning, storage tier optimization, performance tuning, and data replication control and scheduling. It provides diagnostic capabilities to pinpoint resources impacted by an availability or performance issue, then generates timely alerts to enable event action.

IBM Spectrum Control's capabilities include

» Simplified inventory control, asset management, and reporting

» A single, integrated web-based administrative console designed to streamline 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

**TIP** Thanks to these capabilities, IBM Spectrum Control can benefit your organization in several ways:

» Reduce overall storage management costs by saving IT staff time and improving productivity.

- ›› Enable IT staff to proactively manage performance by setting thresholds based on performance metrics.
- ›› Reduce call volumes and improve diagnosis for help-desk staff.
- ›› Allow administrators to monitor metrics, such as I/O rates and cache utilization, to improve capacity management and planning.
- ›› Provide a central user interface for configuring and managing all supported devices on the SAN.

IBM Spectrum Control is on-premises software. But one of the characteristics of SDS is that software can run on-premises *or* in the cloud. IBM addresses this need with another member of the Spectrum Storage family: IBM Storage Insights. Storage Insights offers similar capabilities to IBM Spectrum Control, but from the secure IBM Cloud with a simple subscription plan.

# IBM Storage Insights

IBM believes that artificial intelligence (AI) can deliver significant benefits for data storage and management. As any storage administrator knows, managing large storage systems requires many hours of monitoring, analysis, decision making, and adjustment. When problems arise, troubleshooting complex storage infrastructure and implementing the most effective solutions can be challenging, to say the least.

To address these challenges and reduce both manual labor and mistakes, IBM offers IBM Storage Insights, an enterprise-proven, AI-enhanced, cloud-based system insights platform. Storage Insights helps monitor infrastructure, optimize storage, and resolve issues by analyzing root causes and predicting performance and capacity issues before they impact applications. This component can help you better understand trends in storage capacity and performance and expedite resolution when support is required.

IBM Storage Insights is a true SDS cloud-based solution within the management category that provides greater insights into block storage health, performance, capacity trends, costs, and utilization.

You can use IBM Storage Insights to perform the following:

» Keep an eye on storage health, performance, and capacity across the entire block storage environment.

» View 70+ metrics over years to see trends and compare them against best practices to identify anomalies before they impact applications.

» Resolve storage issues more quickly with automated log uploads and improved diagnostic data available to IBM support.

» View years of historical data with increased visibility into data growth rates and all available capacity to help take the guesswork out of capacity planning.

» Reduce costs by moving data to the most cost-effective storage tier with analytics-driven data management.

» Delay future purchases by identifying and reclaiming provisioned but unused storage.

**TECHNICAL STUFF**

Storage Insights monitors the health, capacity, and performance for all IBM block storage and external storage controlled by IBM Spectrum Virtualize (see Chapter 5 for more information) on a single pane of glass. Storage Insights Pro also handles IBM Spectrum Scale and IBM Cloud Object Storage (for more info, see Chapter 5).

**REMEMBER**

As a cloud-based service, IBM Storage Insights deploys in only a few minutes and is automatically updated by IBM with new functions without you taking any action.

Storage Insights provides proactive best practices and helps identify potential issues before they become problems, then speeds resolution when support is needed. This solution can help IBM storage customers enjoy faster resolution of issues, an enhanced user experience, higher systems availability, and the confidence of services delivered from a leading cloud environment.

# IBM Spectrum Discover

Mining the most business value possible from unstructured data assets isn't simple. Managing the sheer volume of unstructured data — and the systems needed to store it — is a major challenge. This is why IBM introduced IBM Spectrum Discover.

**REMEMBER**

IBM Spectrum Discover is a data catalog and policy engine that helps organize the AI information infrastructure. It automatically and continuously catalogs or indexes objects and files from multiple sources in real time for exabyte-scale, unstructured, heterogeneous storage.

Storage administrators often find that the information about their data (called *metadata*) doesn't provide an adequate view of storage consumption, data ownership, data content, and data quality needed for effective storage optimization. Basic system-level metadata is also inadequate for data scientists, business analysts, and knowledge workers who may spend much of their time searching files and objects that contain confidential or sensitive data.

**TECHNICAL STUFF**

Initial data analysis for the simplest project requires brutal, excruciating data wrangling. In fact, your data specialists can spend much of a project's time trying to parse these files and cross-reference data sources in order to build viable data sets.

As more and more data is required to feed AI-driven applications, and as companies start storing more data, the need to search and manage that data and use it to bring the highest value to the organization is becoming business-critical.

Enterprises need metadata management solutions that offer exceptional data visibility. Once organizations have a clear understanding of their unstructured data, they can more easily optimize storage systems, ensure that their unstructured data complies with government regulations and industry standards, and harness the value of unstructured data for competitive advantage and critical data insights.

Spectrum Discover connects data in real time, providing up-to-the-minute data information to two other members of the IBM Spectrum Storage family — IBM Cloud Object Storage and IBM Spectrum Scale (I cover these in Chapter 5) — to rapidly ingest,

consolidate, and index metadata for billions of files and objects. The richer metadata layer helps storage administrators, data stewards, and data scientists to efficiently manage, classify, and gain insights from massive amounts of unstructured data. You also get enhanced storage economics, improved data govern‐ance, and accelerated large‐scale analytics to create competitive advantage and speed critical research.

IBM Spectrum Discover's capabilities include

» Event-notifications and policy-based workflows to automate metadata ingestion and metadata indexing at petabyte-scale

» Fine-grained views of storage consumption based on a wide range of system and custom metadata

» Fast, efficient search through petabytes of data, resulting in highly relevant result sets for large-scale analytics

» Ability to quickly differentiate mission-critical business data from data that can either be deleted or moved to a cheaper, colder storage tier

» Policy-based custom tagging that enables organizations to classify and categorize unstructured data and align it with the needs of the business

**TECHNICAL STUFF**

IBM Spectrum Discover automatically captures system metadata from source storage systems, creates custom metadata based on user‐defined policies, and enables more advanced use cases such as deep content inspection and extraction of metadata from file headers and content using its Action Agent application program‐ming interface (API). The result is a rich layer of file and object metadata that's managed using one centralized solution.

**TIP**

Available as a VMware virtual appliance, IBM Spectrum Discover can be easily deployed and integrates with many IBM and non‐IBM data management tools.

# Chapter **4**
# SDS for Modern Data Resilience

Cybersecurity is the discipline of protecting against threats such as data fraud and cyberattacks. But what happens when cybersecurity efforts fail and digital systems are compromised? This is the domain of *cyber resilience,* which refers to the preparation organizations make to deal with threats and vulnerabilities, the defenses that have been developed, and the resources available for mitigating security failures after the fact. Cyber resilience capabilities are essential in IT systems, critical infrastructure, business processes, organizations, societies, and nation-states.

Cyber resilience involves making copies of data and applications that can be used to recover from cyber incidents. However, making these copies efficiently and accurately, storing them safely, and finding uncorrupted copies to use for system recovery can be complex and risky endeavors.

When enterprises begin to address the challenges of cyber resilience, they often turn to the IBM Spectrum Storage family. In this chapter, you discover four members that provide a wide range of modern data protection, cyber resilience, data archive, and copy management services essential to business and IT success. Figure 4-1 shows you the breakdown of these members and the icons IBM uses to represent them.

| Modern Data Resilience | IBM Spectrum Archive | |
|---|---|---|
| | IBM Spectrum Protect | |
| | IBM Spectrum Protect Plus | |
| | IBM Spectrum Copy Data Management | |

**FIGURE 4-1:** Modern Data Protection and Resilience solutions.

# 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 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 software-defined storage (SDS) solutions like IBM Spectrum Archive, tape storage remains a cornerstone of storage architectures for many organizations and is a critical tool for emerging business solutions where cloud and analytics workloads are common.

One reason why tape storage continues to offer great value is the fact that a majority of the data your business generates will seldom or never be accessed after a month or two. But that data must still be stored, sometimes for decades, often as part of government regulations but also for opportunities to mine it later for valuable business insights. These data stores can become enormous and expensive. Even low-cost disk storage can cost around several cents per month per gigabyte (GB) of archived data stored; meanwhile, tape costs a few tenths of a penny per month per GB. For the cost-savings alone, tape storage and IBM Spectrum Archive in particular remain extremely valuable to many enterprises.

Another reason why tape is still an important and often essential part of the enterprise data storage solution is its role in cyber resilience. A common challenge when storing data copies is protecting these copies from being corrupted by cyber intruders. Normally, if hackers can get to your data, they can get to the copies of your data. But this isn't true if the copies are removed from the network connections that act as pathways used by hackers and ransomware. A tape cartridge holding your valuable data can be physically removed from all network connections and literally put on the shelf.

One of the strongest ways to protect data is to make a copy and store that copy physically isolated from unsecured networks or servers where cyber attacks could start. With that separation, malware can't find or alter the data. This approach is often called an *air gap* because no electronic connection exists between servers and data copies. Removing a tape from a tape drive is one example of an air gap. Sometimes we speak of virtual air gaps where data copies are separated from servers not by a physical gap but by secure software.

**TECHNICAL STUFF**

Built on IBM Linear Tape File System (LTFS) technology, IBM 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 software to access data stored on tape.

IBM Spectrum Archive simplifies data movement between flash/disk and tape, helping to lower costs and provide air gap capabilities without the need for proprietary tape applications.

**REMEMBER**

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 flash or disk storage for data that is stored for long-term retention, organizations can improve efficiency and reduce costs.

IBM Spectrum Archive helps organizations perform the following:

» 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 from operational storage, improving the efficiency and cost effectiveness of a tiered storage infrastructure.

» Utilize the inherent air gap potential of tape cartridges to increase organizational cyber resilience.

**TIP**

For managing your digital files with the LTFS format, IBM Spectrum Archive offers three software solutions: Single Drive Edition, Library Edition (both Single Drive and Library Edition are available for free download), and the most complete version — IBM Spectrum Archive Enterprise Edition (EE).

To solve file storage capacity and economic problems, choose IBM Spectrum Archive EE in conjunction with Spectrum Scale's policy engine for automating movement and transparent retrieval. This solution is beneficial both in cost and automation.

# IBM Spectrum Protect

Organizations of all types and sizes are moving to multicloud architectures, and data protection is a key driver. By using public cloud resources, organizations can dramatically streamline data protection solutions. Resources can be purchased as needed, with essentially no capital outlays. Data can be copied to multiple cloud storage sites automatically — anywhere — lowering data loss risks and increasing application performance at branch offices and remote locations.

But connecting to the cloud, copying and moving massive data sets, maximizing network performance, leveraging the most cost-effective cloud and on-premises storage tiers, scheduling and monitoring and tracking backups — among many other tasks — can create complex and daunting challenges. These reasons, and many others, are why enterprises turn to IBM Spectrum Protect.

**REMEMBER**

IBM Spectrum Protect

>> Provides a wide range of backup, snapshot, archive, recovery, space management, bare metal system recovery, disaster recovery (DR), and data reuse capabilities

>> Can be downloaded and deployed by small to extremely large enterprises, protecting data on systems of all sizes from a single point of control

>> Enables advanced data protection for cloud, virtualized, and software defined environments and can recover myriad applications, databases, and file systems, including individual files

Backup to the cloud is simple, secure, and cost-effective with IBM Spectrum Protect implemented as an SDS solution. Container storage pools within IBM Spectrum Protect enable you to leverage object storage without additional hardware or gateways

on popular cloud environments such as IBM Cloud, Google Cloud Storage, Amazon S3, and Microsoft Azure. These container storage pools include in-line data deduplication and compression for efficient use of space and bandwidth as well as encryption to ensure that your data is secure.

IBM Spectrum Protect can help users substantially reduce backup infrastructure costs. Savings are typically found in storage media, backup servers, data center floor space, and power and cooling. Capabilities including high performance deduplication, compression, and incremental forever work together to reduce backup storage requirements. The solution's efficiency capabilities are enabled entirely in software — additional hardware-based appliances aren't needed for deduplication, encryption, network acceleration, or cloud access. For maximum cost flexibility, IBM Spectrum Protect enables a broad choice of storage options for backup data, including flash, disk, tape, and object storage. Plus, it can be used with tape to create air gap data protection to increase cyber resilience.

**TIP**

IBM Spectrum Protect provides prescriptive blueprints and configuration scripts designed to help reduce deployment time and guesswork by automating deployment steps and integrating best practices for small, midsized, and large environments. Simplified solution bundles such as IBM Spectrum Protect Suite provide easy-to-manage licensing options that include snapshots and popular backup agents for virtual environments, databases, email, and enterprise resource planning. 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 offerings.

# IBM Spectrum Protect Plus

Today's data centers are shifting away from manual tasks and toward automation, simplicity, and agility. IBM Spectrum Storage SDS solutions are leading this transformation with offerings such as IBM Spectrum Protect Plus — the next wave in IBM's data protection portfolio.

IBM Spectrum Protect Plus takes a modern approach to data protection by focusing on operational recovery and reuse of data instead of just being a backup solution.

IBM Spectrum Protect Plus helps organizations perform the following:

» Recovery, replication, and reuse for virtual machines (VMs), databases, applications, file systems, SaaS workloads, containers, and cloud environments.

» Deployment of the solution as a virtual appliance or as a container application, and the agentless architecture is easy to maintain.

» Automated service level agreement (SLA)-based policies for data protection processes, including operational backups, data replication, and data retention

» Role-based access control (RBAC) and application integration to secure self-service to improve the speed and efficacy of analytics, development, and testing for those who need fast access to the right data

IBM Spectrum Protect Plus consists of two components: the IBM Spectrum Protect server and the vSnap repository. Built-in deduplication and compression are both included in the vSnap repository, and when combined with incremental forever snapshots, IBM Spectrum Protect Plus can substantially reduce storage requirements. Snapshots are stored as native VM images and mounted for recovery to provide near instant access to protected data.

IBM Spectrum Protect Plus creates and maintains a global catalog of all copies of VMs, applications, and files that enables administrators to see what's protected, and more importantly, what isn't. When a data loss incident occurs, a powerful search capability and the global catalog help administrators quickly identify exactly what they want to use or access, instead of manually browsing through hundreds of objects and multiple point-in-time copies. Administrators can recover an entire VM or quickly recover individual files.

Cost-effective data retention and DR are achieved via data copy and archive to on-premises and cloud-based object storage as well as to IBM Spectrum Protect managed storage, including data

archival to physical and virtual tape. Cyber resilience is supported with object storage that uses write-once-read-many (WORM) times technology and the ability to air gap data on physical tape via integration with IBM Spectrum Protect.

In addition, IBM Spectrum Protect Plus simplifies protection for containerized workloads by providing a unified approach to data resilience for Red Hat OpenShift and Kubernetes environments. IBM Spectrum Protect Plus, combined with Red Hat OpenShift, supports true data resilience for containers running in Red Hat OpenShift and Kubernetes environments. Native integration with Kubernetes enhances developer productivity while ensuring complete data recovery with a modern data protection solution that manages both production data and operational metadata. Support for Red Hat OpenShift Container Storage and Common Storage Interface (CSI) snapshots provides the flexibility to use both existing and container-native storage.

IBM Spectrum Protect Plus also integrates with IBM Cloud Pak for Multicloud Management (MCM) to provide seamless data resilience and application management.

**REMEMBER**

The key to IBM Spectrum Protect Plus is its "pick, click, and done" simplicity. Configuration can be as easy as selecting the recovery service level you require: Pick from one of the predefined gold, silver, or bronze SLAs or custom create your own; then, apply that service level to one or more virtual machines; then, you're done.

The dashboard view gives at-a-glance status of the entire protection environment. You can see information about protection status, SLA compliance, and storage utilization all in one place. And, the dashboard helps you quickly identify missing data so you can make sure you're covered.

**TIP**

You don't have to be a backup expert to use IBM Spectrum Protect Plus. The software is simple to deploy and can be ready to use in minutes. You can use it as a stand-alone solution for agile and easy protection or integrate it as part of your IBM Spectrum Protect environment for a comprehensive data protection solution.

# IBM Spectrum Copy Data Management

Modern businesses can't sustain conventional approaches to data management that result in dozens of data copies created and stored throughout the IT environment, including the primary instance and the various backups, snapshots, replicas, and database dumps. And weighing heavy on the minds of most chief information officers (CIOs) and other business leaders today is the possibility of a security breach resulting in the loss of sensitive data. Copy data is one of the more susceptible vectors by which an organization could be attacked because it is all-too-often not tracked, sometimes not scrubbed for sensitive data, and not always secured like it should be. Enter IBM Spectrum CDM.

IBM Spectrum Copy Data Management (CDM) offers an alternative, more sustainable approach. The SDS capabilities of IBM Spectrum CDM enable enterprises to significantly improve overall data economics by creating efficient storage environments. IBM Spectrum CDM provides a leading-edge suite of data and storage services that can simplify copy management and help accelerate the agility and competitiveness of 21st century businesses. Deploying IBM Spectrum CDM can help strengthen your copy data security posture by employing robust data masking. This feature uses policy and role-based filtering to keep sensitive data away from prying eyes.

**REMEMBER**

The IBM Spectrum CDM suite of services helps enterprises

>> Track, create, refresh, and manage the use of data copies

These same common services can be applied to many different use cases within the data center, including DR, application development and testing, analytics, and archiving, among others. Using these common services across multiple use cases enables administrators to catalog and better manage the overall data environment, lowering both cost and complexity.

>> Make copies available to data consumers when and where they need them, without creating unnecessary copies or leaving unused copies on valuable storage

The solution runs out-of-band (outside of the production data path so that it doesn't create performance issues) to create a catalog of all storage, application, and VMware

environments. This catalog provides IT administrators with the ability to manage, orchestrate, and analyze data in order to unleash its power and meet business demands. Data consumers can use the self-service portal to create the copies they need, enabling business agility.

» Helps IT administrators automate and orchestrate otherwise complex data reuse scenarios

Its "Use Data" workflows allow administrators to quickly deliver data copies and eliminate manual processes that require application owners to submit a ticket to IT and wait for a response. Business operations such as DR can be automated so that they can be tested and validated every day. The power of IBM Spectrum CDM enables the resources used for these business operations to be brought up in a fenced/segregated environment, promoted to production quickly with the push of a button, or torn down and cleaned up after a test. Orchestration drives repeatability and auditability, allowing enterprises to leverage a single data copy for multiple purposes, reducing data sprawl and helping lower costs.

» Enables multicloud architectures

It not only helps you move data to the cloud, but also it can bring up live application environments that leverage the elastic compute infrastructure in the cloud. You can spin up workloads and then spin them back down reliably. This maximizes the economic benefit of the cloud by allowing you to use and pay for only the infrastructure you need.

# Chapter **5**

# SDS for Storage Infrastructure

In this chapter, I introduce the family members of IBM Spectrum Storage that provide the basic foundational storage infrastructure functionality and capabilities. As shown in Figure 5-1, I explain the three main components of storage infrastructure.

| Storage Infrastructure | IBM Spectrum Virtualize |  |
|---|---|---|
|  | IBM Spectrum Scale |  |
|  | IBM Cloud Object Storage |  |

**FIGURE 5-1:** Storage infrastructure solutions.

## IBM Spectrum Virtualize

In traditional enterprise data centers, storage capacity is often isolated or trapped in islands of disparate systems that are provisioned for and owned by individual applications. This results in duplication of management points and poor capacity utilization because storage resources can't be as needed to other applications and workloads.

For nearly two decades, IBM Spectrum Virtualize has been transforming inflexible, isolated storage systems into agile, powerful

storage resources through the "magic" of consolidation. Essentially, storage consolidation, also referred to as *virtualization*, is just a particular manifestation of the general SDS paradigm — software inserted in the data path between application hosts and their data. IBM Spectrum Virtualize is designed to simplify storage infrastructure while significantly reducing capital investments, operational expenses, and IT complexity by implementing a single management point for all the storage systems placed under its management. Its rich set of data services can be extended to over 500 heterogeneous storage systems, allowing organizations to simplify their IT operations by reducing the time spent on management and maintenance. Also, IBM Spectrum Virtualize helps to reduce storage cost with powerful data reduction capabilities that include block deduplication that works to minimize the number of data copies stored, plus hardware-accelerated data compression technology that provides consistent, high-performance results across all application workload patterns.



REMEMBER

IBM Spectrum Virtualize technology lies at the heart of IBM SAN Volume Controller and IBM FlashSystem, as well as VersaStack converged infrastructure solutions from IBM and Cisco. It delivers a full range of sophisticated storage functionality:

>> IBM Easy Tier artificial intelligence (AI)-driven automated tiering

>> Encryption to improve data security on existing storage

>> IBM FlashCopy and remote mirror for local and remote data replication

>> Support for using cloud storage to complement on-premises storage

>> IBM HyperSwap high-availability configurations for non-disruptive application and data mobility between data centers

>> Support for virtualized and containerized server environments, including VMware, Microsoft Hyper-V, IBM PowerVM, Kubernetes, and Docker

This foundational member of the IBM Spectrum Storage family integrates with virtualization tools such as VMware vCenter to improve agility with automated provisioning of storage and easy deployment of new storage technologies. It also enables supported storage to be deployed within Red Hat OpenShift, Kubernetes, and Docker container environments, accelerating deployment of persistent storage volumes with the Container Storage Interface (CSI) driver certified by Red Hat and IBM.

To further drive your IT transformation, IBM Spectrum Virtualize for Public Cloud offers multiple ways to create hybrid cloud solutions between on-premises storage and public cloud. It enables real-time, storage-based data replication and disaster recovery, as well as data migration between local storage, IBM Cloud, Amazon Web Services (AWS), and soon, Microsoft Azure. And thanks to its SDS nature, IBM Spectrum Virtualize for Public Cloud allows storage administration at a cloud service provider's site in the same way as on-premises, regardless of the type of storage.

# IBM Spectrum Scale

IBM Spectrum Scale is a high-performance data management solution that's been an industry leader in file management for over 25 years. The top 12 telecommunications companies in the world use this state-of-the-art SDS solution that offers a long list of data management and security features and has been deployed in demanding commercial and research environments, including two research supercomputers.

IBM Spectrum Scale is 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 large quantities of file data. It's designed to support a wide range of application workloads using a variety of access protocols and has been proven extremely effective in large, demanding environments.

Unlike other storage solutions that must implement separate add-on systems to handle files, data objects, or big data analytics with tools such as Spark or Hadoop, all data is stored in common in IBM Spectrum Scale, regardless of the way it's accessed.

After it is part of the core file system, file and object-based data can all 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.

IBM Spectrum Scale offers native, high-performance and scalable access to file and object data using almost all the standard storage

protocols, including OpenStack Swift, Amazon S3, CIFS, NFS, HDFS, and 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. IBM Spectrum Scale is platform–independent, so it can run on IBM Power Systems and on x86 machines, along with storage from IBM and other vendors. This flexibility can reduce costs and improve energy efficiency.

**REMEMBER**

IBM Spectrum Scale offers unique features. By using it, you can reap the following benefits:

» **Accelerate file and object storage performance.** IBM Spectrum Scale provides parallel access to data and shared storage, improving scalability for high-performance work-loads. Data and metadata 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 and manage growth to yottabytes of data and billions of files. This is accomplished within a single overall resource pool or namespace that's 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. It has a common management interface that's easy to use, and operations can be managed from any node in the IBM Spectrum Scale system or cluster, allowing 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 any distance, providing fast read and write performance with automated namespace management from anywhere. 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 network latencies and outages. These capabilities accelerate project schedules and improve productivity for globally distributed teams.

» **Enable multicloud architectures.** Multicloud solutions help safeguard against system failure, take advantage of features or pricing models from various providers to optimize solutions, and can place storage resources in different locations to minimize latency for end-users around the world. IBM Spectrum Scale provides a number of capabilities that help you build and manage multicloud environments. Transparent Cloud Tiering enables non-disruptive, intelligent policy-based migration of data between flash, disk, tape, and cloud storage tiers, allowing you to manage and access both on-premises data and cloud data via a unified view. Cloud Data Sharing allows you to import data from cloud storage into the IBM Spectrum Scale file system or export it from IBM Spectrum Scale to the cloud for access by other systems.

IBM Spectrum Scale can be loaded onto any appropriate hardware. It gets its name from the fact that it uses a massively parallel grid architecture for higher performance and greater system resilience. You can simply add more servers and storage, and IBM Spectrum Scale incorporates them automatically into a single storage resource and spreads both management and data across all available components or nodes to maximize performance, efficiency, and data protection.

# IBM Cloud Object Storage

Every time you post a photo to Facebook, hear a song on Spotify, or share a file on Box, you're using object storage. Objects are different from traditional data files because they include additional metadata that allows object storage systems to essentially eliminate hierarchical organization such as folders and directories. This powerful technology was designed to meet rapidly growing needs to store large amounts of unstructured data in a highly scalable, more reliable, efficient, and affordable manner than traditional storage technologies.

Object storage isn't a solution for high-performance applications such as online transaction processing, customer trend analysis during live browsing sessions, or real-time fraud detection and prevention. Instead, it works especially well for the massive data sets associated with global enterprises, research in genetics, astronomy, macro-economics, and the Internet of Things (IoT). Object storage is effective for all these use cases because of the ease of scalability and access with the efficiency to drive down storage costs.

**REMEMBER**

IBM Cloud Object Storage

>> Provides both on-premises and cloud-based object storage solutions

>> Delivers the capabilities required to provide continuous availability, protection, and access to data assets

>> Allows you to leverage your data when and where you need it to improve business processes, decision making, and responsiveness to regulatory/legal demands

>> Helps reduce storage costs while reliably supporting both traditional and emerging mobile, social, analytics, and AI application workloads

**TECHNICAL STUFF**

IBM Cloud Object Storage technology uses an innovative approach for cost-effectively storing large volumes of unstructured data while ensuring security, availability, and reliability. A unique Information Dispersal Algorithm separates data into unrecognizable slices that are distributed via network connections to storage nodes locally or across the world. With IBM Cloud Object Storage technology, transmission and storage of data are inherently secure. No complete copy of the data resides in any single storage node, and only a subset of nodes needs to be available in order to fully retrieve the data. This method eliminates the high overhead associated with Redundant Array of Inexpensive Disks (RAID)-based storage and the complexity of managing mirrors, replication, and disaster recovery required in a traditional data center.

Chapter **6**

# Ten Use Cases for IBM Spectrum Storage Solutions

IBM Spectrum Storage solutions are deployed to address many enterprise-grade use cases. In this chapter, you find out ten ways that enterprises of all sizes and types are leveraging the power of IBM Spectrum Storage (see Chapters 3, 4, and 5 for details of the different components I mention in this chapter) to accomplish their storage objectives and solve real-world business challenges.

## Managing the Hybrid Cloud

These days, hybrid cloud IT environments are common, but that doesn't mean they're easy. They're called *hybrid* because some applications and data remain on-premises in your company-owned data center, while the rest reside beyond the borders of your company, in the public cloud. Enterprises most often leverage the resources and benefits of several, sometimes many, cloud service providers at once. Weaving together IT infrastructure

elements that may be physically located around the planet brings many challenges, including compatibility issues, performance differences, security vulnerabilities, variations in versions and protocols, and the list goes on and on.

Software-defined storage (SDS) such as IBM Spectrum Scale and IBM Spectrum Virtualize (especially IBM Spectrum Virtualize for Public Cloud) were made for managing and enabling hybrid cloud IT environments. Inserting layers of intelligence between resources and requirements is a game changer:

>> Distance melts away.

>> Data moves transparently between clouds.

>> Compression and deduplication shrink data streams, making them more manageable and less expensive to move and store.

>> Encryption and multiple data protection strategies secure valuable corporate assets and personal information.

>> New and older technologies work together seamlessly.

Some folks might call this list magic. If it's not, it certainly rates as truly remarkable engineering.

IBM Spectrum Scale and IBM Spectrum Virtualize can be implemented as SDS solutions leveraging your existing IT infrastructure or deployed in various hardware-based systems such as IBM FlashSystem. IBM Spectrum Virtualize for Public Cloud is available on IBM Cloud and AWS. These solutions allow you to

>> Move data and applications between on premises and public cloud.

>> Implement new DevOps strategies.

>> Use public cloud for disaster recovery without the cost of a second data center.

>> Improve cyber resilience with "air gap" cloud snapshots.

>> Ensure comprehensive data protection and cyber resilience in hybrid cloud IT environments.

IBM Spectrum Protect Plus (also available through AWS and Microsoft Azure) provides modern data protection for physical, virtual, and cloud application workloads.

# Leveraging AI

Industry analysts predict that within a few years, the majority of enterprises will have implemented some form of artificial intelligence (AI) to improve productivity, manage risks, and drive cost reductions within their IT infrastructure. AI is poised to transform many aspects of your life. If you aren't already, your business organization or IT group will soon begin to investigate the benefits and advantages you may gain from AI. As you learn what AI is and what it can do right now, you may ask yourself these questions:

>> How do I actually go about deploying AI?

>> What would I use AI for within my IT infrastructure?

>> What benefits would AI bring?

IBM Spectrum Storage offers answers to those questions because IBM Spectrum Storage solutions are infused with AI. For example, IBM Storage Insights leverages the power of AI, and the only thing you need to do to gain its benefits is deploy an IBM storage solution. It collects data from many sources, both within your IT infrastructure and across the IT environments of literally thousands of other IBM customers worldwide, and then it constantly crunches this enormous amount of information, hunting for trends, best practices, recurring issues, and potential problems before humans can spot them. You and your IBM support team can then evaluate recommendations and plan how to implement them.

Similarly, IBM Spectrum Virtualize uses AI technology to assess workload trends over time and make decisions when deciding which data to move among storage tiers. In these cases, and many others, IBM Spectrum Storage adds the benefits of AI capability to storage, even the non-IBM storage you may already own in your data center. By providing advanced data management tools designed to be fast and efficient, IBM Spectrum Storage offerings can address the skills gaps that currently limit the adoption of AI across your organization.

# Enabling Agility

*Business agility* is the capability of your company to add, enhance, change, eliminate, or fix what you sell as quickly as possible to meet rapidly shifting market trends. This stage is where container technology enters the story. Designed to be flexible, lightweight, and portable, containers are used to run applications in everything from traditional and cloud data centers to cars, cruise ships, airport terminals, gateways, and the Internet of Things (IoT). By enabling applications to run in many different environments without costly add-ons or rewrites, containers enhance the agility of business and other use cases such as scientific research.

IBM Spectrum Storage solutions offer many ways to enable container technology and increase your business agility. One way to leverage many of these solutions together more cost-efficiently is by implementing IBM Storage Suite for IBM Cloud Paks (see Chapter 2 for more information). This integrated solution is designed specifically to work with container platforms such as Red Hat OpenShift and Kubernetes. And because it's all SDS, it can infuse plodding, outdated IT systems with leading-edge business agility — without costly infrastructure replacement.

**REMEMBER**

Business isn't the only place where containers enhance agility. Consider scientific research. In genomics, for example, applications have a short development life cycle and an even shorter shelf life. Environments are constantly evolving, with one tool being rapidly replaced by new variants, or something else entirely. With containers:

>> Research applications can be fully encapsulated independently of particular operating systems or other applications.

>> Users can run different software versions on the same host without worrying about conflicts.

>> New software packages and applications can be easily pushed out to compute nodes on demand, which potentially eases the application management burden for administrators.

# Accelerating Research

Research is a data-hungry business. So is AI. To visualize the texture of a virus cell wall or train an autonomous driving semi-truck, literally miles and miles of data is needed. Enormous data sets lead to sprawling file systems, which create many problems for industries ranging from trucking and telecommunications to healthcare and pharmaceuticals. Now more than ever, staff and work groups are scattered around the globe, and files need to travel quickly between them. Petabyte-scale data streams must be processed at high speed so new vaccines or truckloads of perishables can get to market quickly.

This is the realm of IBM Spectrum Scale. From the largest supercomputers to the smallest business edge location, IBM Spectrum Scale provides the high-speed data processing and comprehensive data management tools demanded in modern commercial, public, and research environments. It's called IBM Spectrum Scale because its massively parallel, grid-computing design paradigm allows it to easily scale from a single server to thousands of nodes worldwide. When it comes to the use cases that may define the 21st Century, IBM Spectrum Scale fits the bill.

# Improving Business Resilience

The majority of business applications are considered mission-critical. For these applications, organizations want always-on highly available IT infrastructure solutions that meet their particular business demands, regulatory requirements, and disaster recovery (DR) strategies. Because even an hour of downtime can cost thousands or even millions of dollars, as well as damage brand value, recovery point objectives (RPO) and recovery time objectives (RTO) for key business data need to be as close to zero as budgets will allow.

Public cloud-based business continuity solutions offer ways to minimize capital outlays while maintaining the accessibility of precious data. Production data can be stored on-premises to ensure security and improve system performance, while redundant infrastructure with near zero RTO/RPO can be provisioned in the cloud at little capital expense. Market-leading IBM Spectrum

Storage solutions offer a number of ways to implement cloud-based business continuity solutions.

For example, innovative business resilience solutions can be implemented using IBM Spectrum Virtualize and IBM Spectrum Virtualize for Public Cloud to support data mirroring between your on-premises systems and public cloud. With a copy of your data in the cloud, you have the choice of recovering quickly from the cloud or replicating data from the cloud to another data center.

# Multiplying Data Value

Business produces data. Every transaction, every item sold, every web page viewed by a potential customer generates information. In the past, data from business activity was often discarded — or at best, archived. Now, executives understand that data can offer great business value and competitive advantage after its original use. Data can be "reused," often many times, and through each reuse it produces more and more value for the enterprise.

You may not think of it this way, but copies made for data protection and backup purposes for disaster recovery are forms of data reuse. For example, existing data sets generated by business applications can be reused for testing purposes by development teams. Mining data assets for business insight has become extremely valuable. Moreover, data copies can also be used to facilitate reporting needs for a variety of reasons, including demonstrating regulatory compliance.

Thanks to their deep integration and compatibility, IBM Spectrum Storage family members can easily be combined to create solutions that address multiple challenges and requirements at once. Perhaps the most common set of IT requirements is for hybrid cloud, data reuse, and modern data protection capabilities. Almost every enterprise wants all three. The combination of IBM Spectrum Virtualize, IBM Spectrum Copy Data Management (CDM), and IBM Spectrum Protect Plus delivers on all these requirements, and much more. Used together, these components provide powerful data reuse, protection, efficiency, and hybrid cloud capabilities.

# Modernizing Data Protection

Data protection solutions need to address all your workloads. IBM's modern data protection portfolio provides recovery, replication, and reuse for VMs, databases, applications, file systems, SaaS workloads, containers, and cloud environments. These SDS-based data protection solutions are easy to deploy as virtual appliances or as container applications, and their agentless architectures are easy to maintain.

To help protect all your data onsite and off, you can leverage the benefits of both IBM Spectrum Protect and IBM Spectrum Protect Plus in hybrid cloud environments. They can utilize on-premises IBM Cloud Object Storage and public clouds to offload data for long-term data retention. IBM Spectrum Protect Plus can also use IBM Spectrum Protect as an Amazon cloud target repository for data offload.

**TIP**

For companies that want an additional level of data security, IBM Spectrum Protect Plus supports IBM Cloud Object Storage immutable storage features such as retention-enabled buckets. Objects stored in retention-enabled buckets can't be deleted or changed during a specified retention period. Retention periods can be predefined at the bucket level or set for each object using retention policies. And, for additional protection against cyber attacks, you can utilize tape storage to create impenetrable air gap barriers.

# 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, you will sooner rather than later. Of course, SDS is playing a big role in big data analytics, which means there's an excellent use case to highlight.

IBM Spectrum Storage software works together to provide complete analytics pipeline solutions. Consider autonomous vehicles, which generate vast amounts of data. As data streams from vehicles, IBM Cloud Object Storage provides the ideal repository for storing information in volume, potentially for a long period of time, as may be required by legislation. IBM Spectrum Discover

catalogs data as it's ingested and then data analysts use IBM Spectrum Discover to locate the specific data they need. That data is replicated to IBM Spectrum Scale, which provides the high performance necessary to support AI workloads. Finally, analytics results can be archived in IBM Cloud Object Storage for long-term retention.

To handle the growing volumes of data you need to analyze and the multiple analytics platforms that exist, you can choose IBM Spectrum Storage SDS solutions that are flexible and grow easily in capacity and performance. They support the latest open-source and commercial tools such as Spark, Hadoop, and the new AI/machine learning/deep learning frameworks.

# Making IT Transformation Easier

Everyone talks about IT transformation, but not everyone can accomplish it with the least business disruption and greatest cost-efficiency. To understand what IT transformation really means, take a look at your smartphone. Only a dozen years ago, smartphones were just hitting the marketplace — but what a transformation they caused. Today, you can do your personal banking through your phone, shop for almost anything, play nearly any music album ever recorded, even talk to your phone and ask it for the ballgame scores — none of which you could do back then.

The IT infrastructure — the hardware, software, networking, applications, ways of developing applications, ways that software works (for example, machine learning) — that existed when smartphones first hit the marketplace wouldn't get things done now. The problem is that there are still plenty of those older types of IT resources around and still in service. How do enterprises move from the old to the new without forklifting out the entire data center in one ugly-expensive convulsion?

IBM Spectrum Control can show the way. The challenge of survivable modernization is another use case for SDS in general and IBM Spectrum Storage in particular. Instead of throwing out IT components that may have years of good service left in them, you can repurpose them and extend their useful life with a simple software download. For example,

>> Containers can be hosted on servers that previously ran virtual machines or single applications before that.

>> Data can move in and out from multiple clouds to the same disk and tape storage.

IBM Spectrum Control or IBM Storage Insights can be deployed to inventory, assess, monitor, manage, and report on all the legacy resources and provide the foundation for implementing the new hybrid cloud architectures, containers, DevOps processes, and mobile apps that keep your enterprise competitive across today's business landscape.

**REMEMBER**

IT transformation doesn't have to be risky, expensive, and chaotic. You can gain control of it with IBM Spectrum Control and IBM Storage Insights.

IBM Spectrum Virtualize and IBM Spectrum Scale have also been used for many years as engines of IT transformation. With these SDS solutions, businesses of all sizes can improve IT system performance and efficiency, with the flexibility and scalability to support new-generation containerized workloads and hybrid cloud architectures.

# Cost-Effective Cold Data Storage

*Cold data* means data that's rarely — if ever — accessed. You may be surprised to learn that for many companies, cold data comprises 75 to 90 percent of their data. As a result, it's crucial that your storage strategy include a cost-effective cold storage solution. You may be keeping cold data for regulatory reasons because it may have value in the future or to enable big data analytics.

For cold data storage, IBM Spectrum Archive and tape offer the lowest total cost of ownership (TCO), especially for archival data, long-term retention, and when access time measured in minutes is acceptable. Using IBM Spectrum Archive allows cold data to be stored and efficiently managed in tape libraries, which uses far less floor space and energy than disk storage.

While structuring a cold data storage solution, IBM Spectrum Storage can simplify its implementation and provide the following advantages:

>> Avoiding the use of more costly primary storage solutions for cold data

>> Reducing the overall cost of data storage by investing on-tape technology with a lower TCO

>> Helping to meet governance and compliance requirements

# Meet your business needs with software-defined storage

Traditional storage systems can't keep up with the pace of change today. Users are modernizing applications and IT for new hybrid cloud deployments, while dealing with increased management complexity. This book explains how software-defined storage (SDS) enables organizations to become more agile while reducing storage costs, improving performance, reliability, and scalability with intelligent software that performs essential storage functions.

## Inside…

- Increase flexibility
- Simplify management
- Empower global collaboration
- Enable IT transformation
- Deploy turnkey SDS solutions
- Discover use cases for SDS

## IBM

**Chris Saul** is Program Director of Marketing for IBM Spectrum Storage. He's helped launch many IBM storage offerings, including IBM Spectrum Storage in 2015, Storwize in 2010, and SVC in 2003. He's worked for IBM since 1981 and held various positions associated with IBM storage software and systems.

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

## for dummies®
A Wiley Brand

Also available as an e-book

9 781119 806325

# WILEY END USER LICENSE AGREEMENT

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