Software-Defined Storage

For dummies®

A Wiley Brand

Control storage costs

—

Eliminate storage bottlenecks

—

Solve storage management challenges

Chris Saul

Software-Defined Storage

2nd IBM Limited Edition

by Chris Saul

for
dummies®
A Wiley Brand

Software-Defined Storage For Dummies®, 2nd IBM Limited Edition

Publisher's Acknowledgments

Table of Contents

Introduction

Data-driven enterprises are those organizations striving to gain all the value they can from their data assets. They use their data to improve and accelerate decision-making, enhance customer support and experience, create new revenue streams, and capture and keep competitive advantage in many different ways. Data-driven enterprises employ a number of different information technology (IT) tools and methods to derive more value from data. Cloud computing has recently become one of the most powerful, thanks to the capability of the cloud IT model to help lower costs and provide greater business agility. Enterprises of all types and sizes are leveraging the benefits of private cloud, public cloud, hybrid cloud, and now multicloud solutions and architectures.

Software-defined storage (SDS) is another expanding set of IT tools rapidly gaining popularity among data-driven enterprises. The term *software-defined storage* has been used so often to mean so many different things that its definition now is unclear and disputed. I provide a useful definition and discuss its attributes in the chapters ahead, but keep one overriding concept in mind — SDS is a toolset that helps you address the challenges of storing, moving, manipulating, protecting, and analyzing your data so you can better achieve your business objectives.

Most importantly, SDS is what enables the data-driven, multicloud enterprise.

About This Book

Software-Defined Storage For Dummies, 2nd IBM Limited Edition, tells the story of how data-driven organizations leverage the power of this rapidly evolving technology to help them achieve their business and IT objectives, and it 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, I occasionally use special icons to 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.

TECHNICAL STUFF

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.

TIP

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

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'd like to learn more, just go to `ibm.com/spectrumstorage`. There, you can find more information about SDS and IBM Spectrum Storage.

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 is 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 application and the storage device. From the application's perspective, nothing has changed — data is returned when que-ries 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 them to other systems or locations. And of course you want to encrypt it 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 tape costing a small fraction of any other storage medium. 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 the next generation of storage infrastructure. 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.

Also, SDS is just code. It can be downloaded and deployed when you want, as you like, as long as you can accommodate its modest processing and storage requirements. This accommodates another common definition of SDS — storage-related products that can be upgraded and/or downloaded and implemented separately from the underlying storage hardware.

The real issue with the definition of SDS is that once you begin to imagine, the possibilities seem endless, perhaps even chaotic. SDS might provide almost any data service we want (it's rapidly moving toward that point) or perform any circus trick we can think up. To simplify things and bring some coherence to SDS conversations, here are some useful divisions into which we can place various types of SDS products and tools.

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 — management, data protection and services, and infrastructure.

SDS for storage management

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 all these questions, your organization needs better management services from your SDS. And these questions are hardly the tip of the storage management iceberg. Consider the following:

>> Can you find any file you want, almost instantly, anytime?

>> Can you generate reports of all your storage activity in enough detail to be useful when you hold meetings with other company executives about significant storage investments?

>> When there's a problem, something within your storage infrastructure isn't performing as it should, can you identify it quickly and easily provide needed information to the customer support team of your storage vendor?

>> Do you feel confident that your storage configurations are optimized, based on industry standards and evolving real-world experience, to meet your IT and business objectives?

And that's not even mentioning the tangle of system integration and compatibility issues you may confront as you virtualize business-critical applications or try to utilize some newfangled technology such as containers — which leads to more questions: What about your data stored off in the cloud? Can you integrate the views of what's happening out there with what's happening in your on-premises storage systems?

These questions aren't even scratching 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. Instead, most often it's data services and data protection that come to mind.

SDS that provides modern data protection and services

The list of what you want to do to and with your data grows longer almost every 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 protection and related services. In fact, data protection has become so important that it demands and deserves its own category within the SDS discussion.

Data protection 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. For example, managing rapid data growth has a lot to do with how well your data protection solution performs at scale and how well the solution mitigates rising storage costs.

And data protection goes beyond simple backup and recovery. Data copies can cause plenty of headaches as well. In many data centers, more than half of overall storage capacity is used to store just the copies of data. What happens when application development and testing teams want a copy of the latest data set? Does anyone know where the latest copy lives? How does IT ensure users have access rights to copies of data? How long will it take to provide a copy cleansed of all sensitive information to the teams — a week or a few minutes?

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? It's not true that you can 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 — or not, depending on the quality of your SDS data protection and data services solutions.

SDS that operates as basic storage infrastructure

All data is simply bits and bytes. But as soon as you gather enough bits and bytes together, or your data is all related, such as the pixels of an image or video, then it begins to exhibit certain characteristics which we use for descriptive or organizational purposes. We 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. Across the universe of information in general around 80 percent of data is described as unstructured — documents, images, audio and video files, emails, telemetry from your car flying across the Internet of Things (IoT), your DNA after it's digitized, 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 that 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, though there are a few products out there that 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. It can monitor access patterns and move data sets to optimal storage media. It can compress data or recognize redundant information and store it only once in order to make storage much more efficient. This type of SDS can encrypt and copy data, send it out to the cloud, or manage many storage devices together in clusters and grids. It can even make tape behave like disk.

SDS is easier to understand when you think of it as belonging to one of these three basic types — management, data protection, 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.

Introducing IBM Spectrum Storage

IBM continues to invest in SDS development and innovation by consolidating previous stand-alone SDS-related solutions into one product family with a consistent user experience. This SDS product family is called IBM Spectrum Storage.

IBM Spectrum Storage is a leading SDS solution. The family offers a wide range of SDS functionality and capabilities to address your storage objectives. Figure 1-1 shows the current members of the IBM Spectrum Storage family organized into SDS functional categories. Visit the IBM Spectrum Storage website at `ibm.com/spectrumstorage` to get the latest information.

FIGURE 1-1: The IBM Spectrum Storage family.

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 multicloud environments, leverage the power of the family to more easily and effectively implement and manage important business tools such as analytics and artificial intelligence (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.

IBM Spectrum Storage solutions can be deployed in many different ways. Some of them are integrated into IBM storage systems such as IBM FlashSystem or the Storwize family so that 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 that you gain their capabilities and features, but you don't need to bother with the 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.

Chapter **2**

SDS for Storage Management

BM Spectrum Storage has many components. In this chapter, I explain the four members that provide various types of storage management functionality.

IBM Spectrum Control

Most likely, your data storage environment is a mongrel. 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 you must do seems impossible, but the truth is, it's just another day at the office for IBM Spectrum Control — the epitome of what is meant by software-defined storage (SDS) that provides storage management services. IBM Spectrum Control is a comprehensive, end-to-end 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 multi-site storage environments, including both IBM and non-IBM systems. You can use IBM Spectrum Control for performing 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 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

TIP 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. For even simpler implementation, IBM Storage Insights offers similar capability 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 storage systems come with IBM Storage Insights, IBM's enterprise-proven, AI-enhanced, cloud-based system insights platform to help you better understand trends in storage capacity and performance and expedite resolution when support is required. Storage Insights helps monitor infrastructure, optimize storage, and resolve issues by analyzing root causes and predicting performance and capacity issues before they impact applications.

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

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

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

» 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 on a single pane of glass. Storage Insights Pro also handles IBM Spectrum Scale and IBM Cloud Object Storage.

REMEMBER

As a cloud-based service, Storage Insights deploys in only a few minutes and is automatically updated by IBM with new functions without any action by you. Storage Insights provides proactive best practices and helps identify potential issues before they become problems, then speeds resolution when support is needed. Storage Insights 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. Storage administrators often find that the information about their data (also 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. This is why IBM introduced IBM Spectrum Discover.

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.

IBM Spectrum Discover is a sophisticated metadata management solution that provides data insight for petabyte-scale unstructured data storage. It connects easily to two other members of the IBM Spectrum Storage family — IBM Cloud Object Storage and IBM Spectrum Scale (I cover these in Chapter 4) — to rapidly ingest, consolidate, and index metadata for billions of files and objects, providing a rich metadata layer that enables storage administrators, data stewards, and data scientists to efficiently manage, classify, and gain insights from massive amounts of unstructured data. This enhances storage economics, improves data governance, and accelerates large-scale analytics to create competitive advantage and speed critical research.

IBM Spectrum Discover provides

>> 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

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 programming interface (API). The result is a rich layer of file and object metadata that is managed using one centralized solution.

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

IBM Spectrum Connect

IBM research confirms that today, more than three-quarters of enterprises have deployed some type of cloud computing capabilities and most organizations are already adopting multicloud strategies. But multicloud isn't easy. Each cloud you add brings with it connectivity and compatibility issues, not to mention plenty of management challenges. Enabling and simplifying multicloud deployment across your entire IBM storage environment is the focus of IBM Spectrum Connect.

Today's organizations demand easy and fast integration of storage in multiple cloud environments. In a single solution, IBM Spectrum Connect leverages existing IBM storage capabilities and empowers storage teams and other stakeholders by enabling provisioning, monitoring, automating, and orchestrating of IBM block storage in containerized, VMware, and Microsoft PowerShell environments.

IBM Spectrum Connect

» Manages the API dialogs for IBM storage systems from one place, providing a single pane of glass for orchestrating between multiple cloud platforms and IBM storage devices

» Enables container orchestration with Kubernetes in solutions such as IBM Cloud Private

» Simplifies the provisioning of storage for containers by defining policies by service level agreement (SLA) or by application workload

» Provides better storage visibility to improve troubleshooting in containerized environments

» Allows definition of easy-to-consume storage classes, such as by SLA or workload, simplifying self-service and automation of storage provisioning

>> Supports the IBM FlashSystem family, IBM Storwize family, IBM DS8000 series, IBM SAN Volume Controller, and IBM XIV, as well as IBM Spectrum Virtualize, IBM Spectrum Accelerate, and VersaStack converged infrastructure solutions from IBM and Cisco

>> Leverages existing IBM Storage capabilities and is provided at no additional charge to IBM storage system customers

TECHNICAL STUFF

IBM Spectrum Connect combines private cloud management capabilities with several enabling APIs. For example, IBM Storage Provider for VMware VASA (vStorage APIs for Storage Awareness) allows VMware vSphere vCenter to recognize and use the capabilities of various IBM storage systems, making it easier to allocate, maintain, and monitor the data stores for the VMs.

CHAPTER 2 **SDS for Storage Management** **17**

Chapter **3**

SDS for Data Protection and Services

T his chapter introduces the members of the IBM Spectrum Storage family that provide a wide range of modern data protection, archive, and copy management services essential to business and IT success.

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 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 seldom or never be accessed after a month or two. But for a variety of reasons, including government regulations and also opportunities to mine it later for valuable business insights, some data still must be stored, sometimes for decades. 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, but tape costs a few tenths of a penny per month per GB. For these reasons, 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, 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 and software to access data stored on tape.

IBM Spectrum Archive simplifies data movement between flash/disk and tape to lower costs 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

>> 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

TIP

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

IBM Spectrum Protect

Organizations of all types and sizes are moving to multicloud architectures for many reasons, and data protection may be top of the list. By using public cloud resources, organizations can dramatically streamline data protection solutions. Resources can be purchased only 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.

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. It can help protect data on systems of all sizes — from a single point of control. IBM Spectrum Protect can be downloaded and deployed to protect small to extremely large enterprises. It 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 enable you to leverage object storage without additional hardware or gateways on popular cloud environments such as IBM Cloud, 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. IBM Spectrum Protect high performance deduplication, compression, and incremental forever capabilities work together

to reduce backup storage requirements. The software'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.

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 rather than being only a backup solution.

IBM Spectrum Protect Plus protects virtual environments, specifically VMware ESXi and Microsoft Hyper-V. IBM Spectrum Protect Plus also supports file and database recovery, as well as multisite data replication for disaster recovery. IBM Spectrum Protect Plus data copies can be repurposed for multiple data copy use cases, such as provisioning application development/test environments and supporting analytics and reporting. The software can be up and running in less than an hour, compared to some alternatives that may take weeks to deploy and require costly professional services.

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. For long-term data retention, IBM Spectrum Protect Plus snapshots can be off-loaded to IBM Spectrum Protect and take advantage of all the storage media options provided, including disk, tape, cloud, or object storage.

IBM Spectrum Protect Plus creates and maintains a global catalog of all copies of virtual machines, applications, and files, enabling administrators to see just what is protected, and more impor-tantly, what isn't. When a data loss incident occurs, a power-ful 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.

REMEMBER

The key to IBM Spectrum Protect Plus is its "pick, click, and done" simplicity. Configuration can be as easy as selecting the recov-ery service level you require; picking from one of the predefined gold, silver, or bronze SLAs or custom creating your own; then applying that service level to one or more virtual machines — and you're done. If you miss any VMs, the dashboard view helps you quickly identify them so you can make sure you're covered. The solution has a sharp, modern look and feel, including a dashboard that gives at-a-glance status of the entire protection environ-ment. You can see information about protection status, SLA compliance, and storage utilization all in one place.

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 by installing a prebuilt virtual appliance. You can use it as a stand-alone solution for agile and easy VM and database protection or integrate it as part of your IBM Spectrum Protect environment for a comprehensive data protection solution.

IBM Spectrum Copy Data Management (CDM)

Modern businesses can't continue to sustain conventional approaches to creating and managing data copies. In many cases the traditional approach results 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.

IBM Spectrum CDM can help change this equation. 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.

IBM Spectrum CDM helps drive down cost and complexity by increasing operational and storage efficiency. Its 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.

IBM Spectrum CDM makes 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.

IBM Spectrum CDM 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 disaster recovery 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.

IBM Spectrum CDM also 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.

TIP

Weighing heavy on the minds of most 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. 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.

Chapter **4**

SDS for Infrastructure

n this chapter, I introduce the family members of IBM Spectrum Storage that provide the basic foundational storage infrastructure functionality and capabilities.

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 moved as needed to other applications and workloads.

For over 15 years, IBM Spectrum Virtualize has been transforming inflexible, isolated storage systems into agile, powerful storage resources through the magic of "virtualization." Essentially, storage virtualization is just a particular manifestation of the general SDS paradigm: software inserted in the data path between application hosts and their data. Through its "magic," IBM Spectrum

Virtualize provides a single management point for all the storage systems that it supports, which enables better utilization and performance, while extending a market-leading set of services and functionality to all those storage systems.

IBM Spectrum Virtualize capabilities include a wide range of data services that can be extended to over 440 different IBM and non-IBM storage systems: automated data movement, synchronous and asynchronous copy services either on-premises or to the public cloud, high availability configurations, storage tiering, and data reduction technologies, among many others.

IBM Spectrum Virtualize software delivers a full range of sophisticated storage functionality, including

>> IBM Easy Tier AI-driven automated tiering

>> Encryption to improve data security on existing storage

>> IBM FlashCopy and remote mirror for local and remote 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 with Kubernetes and Docker container environments.

REMEMBER

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 the public cloud. It enables real-time storage-based data replication and disaster recovery, as well as data migration between local storage and the IBM Cloud. 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.

TECHNICAL STUFF

Containers are an open source technology that lets software 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. As a result, containerization is a key enabling technology for flexibly delivering workloads to private and public cloud. Together with IBM Spectrum Connect (see Chapter 2), IBM Spectrum Virtualize enables all the storage systems under its management to become effective components in container environments, improving flexibility, simplifying deployment, and further lowering costs.

IBM Spectrum Virtualize offers 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 is available as software to run on Intel x86 servers and is also at the heart of IBM Storwize family, IBM FlashSystem 9100 and V9000, IBM SAN Volume Controller, and VersaStack converged infrastructure solutions from IBM and Cisco.

IBM Spectrum Scale

Today, your information ecosystem includes different data types and storage media. You have structured data with databases and transaction processing applications that consume and produce it. The majority of your data is unstructured; you need file systems to manage it. And you may also use object storage.

IBM Spectrum Scale is an SDS infrastructure solution designed to provide high-performance and highly functional data management for all the types of data that your business activities may generate, including structured data, unstructured data, and objects.

TECHNICAL STUFF

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.

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 is 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 via 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.

IBM Spectrum Scale offers unique features. Using it, you can

REMEMBER

>> **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 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. It 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, 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 geographical distances, 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 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. This means that you can 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. As true SDS, it is also available in a storage system (IBM Elastic Storage Server) or on IBM Cloud and Amazon Web Services.

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.

IBM Cloud Object Storage provides both on-premises and cloud-based object storage solutions. It delivers the capabilities required to provide continuous availability, protection, and access to data assets. It allows you to leverage your data when and where you need it to improve business processes, decision making, and responsiveness to regulatory/legal demands.

IBM Cloud Object Storage helps reduce storage costs while reliably supporting both traditional and emerging mobile, social, analytics, and AI application workloads. In fact, industry analysts

have confirmed that versus traditional storage, IBM Cloud Object Storage can save as much as 70 percent on overall storage-related costs.

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 RAID-based storage and the complexity of managing mirrors, replication, and disaster recovery required in a traditional data center.

All objects stored in IBM Cloud Object Storage are encrypted by default. Data at rest is encrypted with Advanced Encryption Standard (AES) 256-bit encryption and Secure Hash Algorithm (SHA)-256 hash, while data in motion is secured by using built-in carrier grade Transport Secure Layer (TLS)/Secure Socket Layer (SSL).

IBM Spectrum Accelerate

IBM Spectrum Accelerate lets you implement block storage using a different storage architecture from traditional systems. Conventional storage uses 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, CPU power, memory cache, 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 storage "tuning" is necessary, because the system performs automated data distribution and load balancing, so the performance is at all times consistent, predictable, and without hotspots. The architecture and built-in mechanisms such as self-healing enable the system to offer high availability. Protecting 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.

TIP

Like IBM Spectrum Virtualize (see the earlier section "IBM Spectrum Virtualize" for more info), IBM Spectrum Accelerate offers a rich set of proven storage services features — and these are included at no extra charge. It integrates 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:

» **Extraordinary simplicity:** The unique grid architecture provides advanced data distribution and self-healing capabilities that eliminate the need for planning and complex administration, enabling more efficient administration of enterprise-scale storage capacities at a fraction of the staffing and training requirements of traditional storage systems.

» **Remote replication:** Mirroring is traditionally 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 providing high availability for applications where continuous operation is crucial.

» **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.

» **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.

» **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:** IBM Spectrum Accelerate offers Lightweight Directory Access Protocol (LDAP), role-based access, and data at rest encryption.

» **Management at scale:** 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 help you grow your storage system as large as you want, manage it from one pane of glass, and exercise flexibility moving volumes from one system to another regardless of the cloud consumption model.

TECHNICAL STUFF

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.

TIP

Choose IBM Spectrum Accelerate when you want to build a cloud storage solution with capabilities that can grow easily, are cost effective, and provide flexibility. IBM Spectrum Accelerate also offers combined private, public, and hybrid cloud solutions as a member of two IBM Storage product families — IBM FlashSystem A9000 and A9000R and IBM XIV Storage Systems.

IBM Spectrum NAS

File systems were originally developed to give structure to unstructured data. By providing hierarchical organization in the form of files, folders, and directories, file systems made the bits and bytes making up images, documents, emails, and so forth much easier to store and find.

Network–attached storage (NAS) is dedicated file storage that enables multiple users and heterogeneous client devices to retrieve data from centralized storage. Users on a local area network (LAN) access the shared storage via a standard Ethernet connection. What most characterizes NAS is ease of access, high capacity, and fairly low cost. NAS devices provide infrastructure to consolidate storage in one place and to support tasks such as archiving and backup.

But a key liability of NAS is that each NAS system resides on the LAN as an independent network node, defined by its own unique Internet Protocol (IP) address. Essentially, to add more storage, performance, and capabilities, you must deploy an additional NAS system, and then another, when needed, and another, and so on. This isn't a model that easily grows/scales, and in this age of exploding unstructured data volumes, storage scalability is a must.

Enter IBM Spectrum NAS. It provides a highly scalable SDS solution for file systems. It can be loaded onto your choice of inexpensive servers ("commodity" x86 machines) to transform them into a group/grid/cluster of NAS storage devices or "nodes" all working together to provide one NAS solution. IBM Spectrum NAS manages the cluster of nodes, where each node has its own CPU, memory, storage, and cache with a single file system spread across all nodes. Each node adds access points, cache, storage capacity, and performance to the cluster. The IBM Spectrum NAS architecture is balanced by definition because every node plays an identical role in the storage cluster — no dedicated meta-data or special–purpose nodes exist. Users and applications can access the cluster through any node, spreading the load uniformly and eliminating bottlenecks. This architecture, along with data protection mechanisms such as erasure coding and snapshots, enhances reliability by avoiding single points of failure.

Thanks to its software-defined approach, IBM Spectrum NAS supports organizational evolution from a fixed, rigid, hardware-centric approach to one that is flexible and pragmatic. As the needs for storage or bandwidth grow, more nodes can be added to the cluster to immediately take on their share of the compute and storage load. This helps ensure that what works today will continue to work in the future.

The more nodes in a cluster, the greater that cluster's capacity and performance. More nodes mean more options for load rebalancing and for accommodating the self-healing process that takes place in the event of a node failure. Applications access the data through the IBM Spectrum NAS virtual file system, which allows multiple clients to interact with the same stored data. IBM Spectrum NAS allows multiple, separate domains and file systems to span a single storage cluster, reducing costs by simplifying administration and allowing more efficient use of storage resources, compared to more traditional NAS approaches.

IBM Spectrum NAS is designed for ease of deployment and management. Each node can be either "bare metal" or a virtual machine. A wizard-based guide allows for setup in 30 minutes, with non-disruptive network upgrades. Identical software per node further simplifies setup and maintenance.

Chapter **5**

Ten Use Cases for IBM Spectrum Storage Solutions

This chapter introduces ten — okay, only eight — different ways that enterprises of all sizes and types can leverage the power of IBM Spectrum Storage (see Chapters 2, 3, and 4 for details of the different components I mention in this chapter) solutions and technologies to accomplish their storage objectives and solve real-world business challenges.

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 our lives. If you aren't already, your business organization or IT group will soon begin to investigate the benefits and advantages you might gain from AI. As you learn what AI is and what it can do right now, one of the first questions you may ask is how do you actually go about deploying AI?

Then you may wonder what you would use AI for within your IT infrastructure, and what benefits it would bring.

IBM Spectrum Storage offers a simple and powerful solution: Overall, 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, then it constantly crunches this enormous amount of information, hunting for trends, best practices, recurring issues, potential problems long before any human could spot them. Then you and your IBM support team can evaluate recommendations and plan how to implement them, if you choose.

Similarly, IBM Spectrum Virtualize uses AI technology to assess workload trends over time and make decisions when deciding what data to move among 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.

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 is where a new technology called "containers" enters the story. The interest in container technologies is high and rising rapidly. Designed to be flexible, lightweight, and portable, containers will be used to run applications in everything from traditional and cloud data centers, to cars, cruise ships, airport terminals, and gateways to the Internet of Things (IoT). Containers enhance the agility of applications for business and for other use cases such as scientific research.

IBM Spectrum Connect provides the technologies within the IBM Spectrum Storage family that enable you to use containers to improve your business and IT agility. IBM Spectrum Connect leverages existing IBM storage capabilities and is provided at no additional charge to IBM customers. It works with tools and solutions such as Docker, Kubernetes, IBM Storage Enabler for Containers, and IBM Cloud Private to simplify container use.

And it provides better visibility into containerized environments to improve utilization and troubleshooting.

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

Modernizing for Better 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.

Likewise, for ultra-high-availability requirements, IBM HyperSwap in IBM Spectrum Virtualize and IBM Spectrum Accelerate supports dual active-active, potentially multi-site configurations with automated failover.

By deploying these IBM Spectrum Storage family members, you can leverage powerful business continuity solutions that help you ensure all business service-level agreements (SLAs) are met.

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 multicloud, 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 multicloud capabilities.

Partnering with Flash

Before 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, 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. Policy-based tiering in IBM Spectrum Scale provides a similar capability as files age. For more wholesale movement of data, the non-disruptive migration capabilities in IBM Spectrum Virtualize and IBM Spectrum Accelerate enable entire storage systems to be migrated to new environments.

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 offsite 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, which I cover in Chapter 4.

Modernizing Data Protection Using Hybrid Multicloud

Data protection solutions need to address all your workloads, including physical file servers, virtualized environments, and both physical and virtual databases and applications. Going further, 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, multicloud environments.

IBM Spectrum Protect and IBM Spectrum Protect Plus can both 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.

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.

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 time as may be required by legislation. IBM Spectrum Discover indexes data as it is 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 frameworks that exist, you can choose IBM Spectrum Storage SDS solutions that are flexible and grow easily in capacity and performance with support for the latest Open Source and commercial tools, such as Spark, Hadoop, and the new AI/machine learning/deep learning frameworks.

Meet your business needs with SDS systems

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

Inside…

- Increase flexibility
- Simplify management
- Empower global collaboration
- Resolve performance bottlenecks
- Deploy turnkey SDS solutions
- Discover use cases for SDS

IBM

Chris Saul is Marketing Manager for IBM Spectrum Storage. He's helped launch multiple 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 storage 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 554073

WILEY END USER LICENSE AGREEMENT