Druva Special Edition

Cloud Data Protection

For dummies®

A Wiley Brand

Understand modern data protection

Deploy cloud backup and disaster recovery

Access and analyze information

Brought to you by

druva

Faithe Wempen

Letter to the Reader

Dear Reader,

In 2008, the universe placed a unique challenge in my path that would irrevocably change my life and the data protection industry as we know it. The challenge is this: How do we enable organizations to do more with their critical business information? From endpoints to servers to cloud applications, all governed by legacy silos across the globe, we thought there had to be a better way to manage and act on this data.

Druva was born in the cloud to do just that. Built on the public cloud free of legacy solutions, we're pioneering new opportunities for how enterprises protect, preserve, and discover their information. We call this Cloud Data Protection.

With this book, we hope that you are able to effectively integrate the cloud into your long-term technology vision and fundamentally change the way you view your data.

Most sincerely,

Jaspreet Singh
Founder and CEO,
Druva

Cloud Data Protection

Druva Special Edition

by Faithe Wempen

for
dummies
A Wiley Brand

Cloud Data Protection For Dummies®, Druva Special Edition

Publisher's Acknowledgments

Introduction

Your company owns a lot of data. A *lot.* And it's probably scattered all across the country, or all over the world, on hundreds of servers, desktops, laptops, tablets, and cloud storage applications, so it's difficult to get at quickly.

As a business leader, you might think wistfully of the days when all the data was on the main server, and everyone sat quietly at their desks and accessed it. Protecting data was simpler back then, and so was analysis. When everything was stored in one place, you didn't have to worry about sensitive data going out on inadequately secured mobile devices or how to comply with data privacy and legal requirements. When data was all behind the same firewall, antivirus programs could watch out for malware-causing email attachments. Life was simpler back then.

For better or for worse, though, today's business world is mobile, decentralized, and constantly in motion. No amount of IT regulation is going to make the company safe from threats like malware, data leaks, data regulatory lapses, storage corruption, and other potential hazards. Your employees need the freedom to use cloud applications, mobile devices, Wi-Fi hotspots in foreign countries, and all the other decentralized services.

What if you could allow your employees to keep doing what works for them, but your critical business data would be fully protected, and visible all the time to those who need it? What if the collection of *all* data, across *all* devices, happened non-intrusively and automatically, without the need for IT intervention? What if you could maintain historical archives of all data for any amount of time? What if you could instantly view, search, and roll back to earlier versions quickly and easily whenever data disasters like a server failure, or malware, struck a system?

Those aren't theoretical what-ifs, but reality, when you implement a cloud data protection system. In the rest of this book, I fill in the details about that.

How This Book Is Organized

As with other *For Dummies* books, this book doesn't assume that you'll begin on page one and read straight through to the end. Each chapter is written to stand alone, with enough contextual data provided so that you can understand the content.

Chapter 1: What Is Data Protection?

This chapter explains some basic data protection concepts. You learn what your company's most valuable resource is (*Hint:* It isn't your people), the difference between data and information, and how the classic silo-based model of data storage that grows organically in most companies isn't necessarily the best model.

Chapter 2: Modern Data Protection with the Cloud

In this chapter, you learn the basics of how cloud data protection works. You investigate the benefits of centralized access, searches, backups, and analysis, and look at some of the benefits you get with cloud data protection that aren't available with other approaches.

Chapter 3: Backup and Disaster Recovery

You probably already have some sort of backup system in place, but is it the best one for your needs? This chapter addresses traditional backups and their limitations, and then compares them to the cloud data protection approach for both servers and end-user data.

Chapter 4: Accessing and Analyzing Data

This chapter explains how traditional data analysis falls short by looking at an example scenario where an executive must gather data from multiple branches for a legal inquiry. I show you how a cloud-based system makes a lengthy and arduous task into a simple matter of a few mouse clicks.

Chapter 5: Minimizing Data Risks

This chapter looks at the many threats to your data's safety and integrity, from unhappy employees to insidious malware such as ransomware. You learn how cloud data protection can help minimize those risks, as well as assist in preparing for legal and regulatory challenges.

Chapter 6: Ten Reasons to Use Cloud Data Protection

This quick and easy list outlines ten benefits of cloud data protection compared to traditional data protection models.

Foolish Assumptions

This book assumes that you are a decision-maker responsible for the health and profitability of a medium-to-large business. It further assumes that you aren't fully satisfied with your company's data protection performance. Maybe you wish you had easier access to the data from branches, or from individual user devices. Or perhaps you are worried about what potentially expensive shortcomings might be discovered if your information systems were audited for regulatory compliance. At any rate, I assume that you're interested in hearing about a better, more cost-effective way to protect your data.

However, the book *doesn't* assume that you know anything about computer hardware or software. This isn't a book for techies. If you like what you read here, and you want to learn more about cloud data protection, Druva will be glad to connect your tech people to its experts, who will answer your team's in-depth questions about how things work.

Icons Used in This Book

The Remember icon marks important facts that are worth adding to your memory. Look for this icon next to key paragraphs containing essential facts.

The Tip icon identifies useful information that can help you deliver better results — or simply make your life a little less complicated.

TIP

The Warning icon guides you around common pitfalls and dangers. Be safe.

WARNING

Where to Go from Here

Just start reading! You can use my description of the chapters in this Introduction. If you are in a hurry and just want the "elevator pitch" on cloud data protection, start with Chapter 6! Then review the earlier chapters at your leisure to learn the details behind Chapter 6's main points.

Chapter **1**
What Is Data Protection?

In this chapter, you find out what *data protection* is and learn the difference between data and information. I explain the traditional data protection model that many companies arrive at organically as they grow — the silo approach — and point out some of its limitations and drawbacks. Then I introduce you to a more modern and desirable model — the cloud.

Your Most Valuable Asset

Picture, if you will, the worst disaster you can imagine that could possibly befall your company. What would be the most difficult situation to recover from? Your headquarters burning down? Your entire senior management team resigning? An international media blitz that spotlights a major flaw in your flagship product? As bad as any of that would be, companies have bounced back from all these situations.

No, the worst thing that could happen to your company would be the loss of all its digitally stored data. Customer contact databases, product inventories, personnel and payroll data, the company website, email systems . . . all wiped out. Go ahead, shudder at the thought. You know it's true.

I tell you this not to scare you, but to point out something that you already instinctively know: *Data is your company's most valuable asset.* Your company's data has been gathered over the course of many years, with the contributions of hundreds or thousands of people. Replacing it could take years, and cost millions of dollars.

Data is critical for three main reasons:

» **The negative consequence of losing it:** Customer, employee, product, and sales data is all critical. Safeguarding it allows the company to continue with "business as usual," without any interruptions.

» **The risk of lost revenue, lost productivity, and even lawsuits if sensitive data is not properly safeguarded:** Knowing your data — what, where, and who has access to it — is essential to responsible data stewardship. Whenever customers trust you with their personal data, your company had better make sure it doesn't end up in the wrong hands. If government regulations are involved in the use and storage of that data, the potential consequences are even more dire.

» **The positive benefit of being able to analyze company data for better decision-making:** Data protection enables you to move beyond "business as usual" into new strategies and ventures that your company's own data suggests. Many businesses become so mired in day-to-day operations that they forget what a goldmine they have in their data files. Mining that data to find trends and opportunities can make the difference between good and great strategic planning.

Safeguarding your data

Cloud data protection provides three key benefits: improving backup and disaster recovery, making it easier to access and analyze data, and minimizing the risks of data loss and improper data handling. Each of these topics is covered in an upcoming chapter. As Figure 1-1 shows, backup, recovery, and disaster readiness are crucial to your cloud data protection efforts.

Realistically, your company is likely never going to lose *all* its digital data. You have backups, after all, and built-in redundancy in your most critical data storage areas. You have an IT department, with plans in place for disaster recovery.
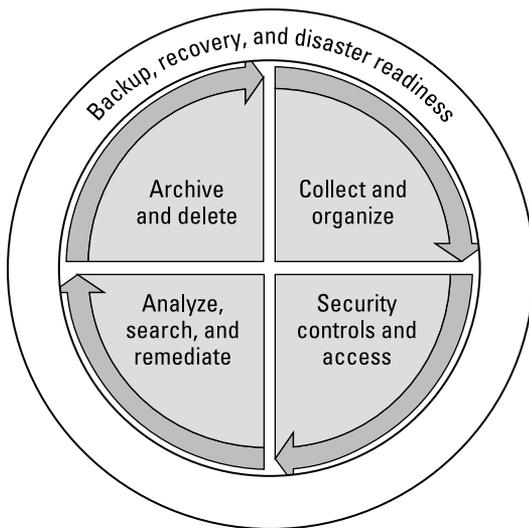
FIGURE 1-1: Cloud data protection.

WARNING

However, safeguarding data isn't nearly as simple — or as foolproof — as it used to be. New threats are springing up every month, including some dangers that IT departments of a decade ago could not even conceive. A company whose data protection scheme is still relying on simple server-to-tape backups or other older technologies can quickly find itself in a world of trouble when faced with these new threats.

To name just a few examples:

>> New threats involving ransomware, where malware infects one or more systems and holds the data for ransom until you pay

>> A disgruntled employee making malicious changes to a database that aren't discovered until much later, after they've caused the company significant embarrassment

>> The irreplaceable data that exists only on individual managers' cellphones and tablets, outside the company's backup system

>> Adhering to global data regulations, like the European Union's General Data Protection Regulation (GDPR) or the Health Information Protection and Accountability Act (HIPAA), that pose significant fines for sensitive data being compromised and not properly managed

Making the most of your data

Of course you want your data to be safe. That's a given. But if you're forward-thinking, you also want your company's data to be readily available for search and analysis. *All* the data, from the product catalog stored on a virtual machine in the basement of your headquarters to the contact list on a salesperson's phone as he or she travels to an international meeting. After all, what good is all that data if you can't readily access it?

Think about your company's current data storage and data protection systems. Your company owns a lot of data, and if you could mine it effectively, you would have a powerful tool for responding to compliance and legal inquiries as well as business decision-making. Unfortunately, though, most companies don't do anything with data other than hoard it, which creates inherent risks to the business as well. There is no central repository that an analyst or executive could query to get data because the data is decentralized, stored on dozens or hundreds or thousands of individual computing devices.

To make the most of your data, it ought to be:

» Available 24/7 to anyone who needs it

» Historically cached, so that trends over time can be identified

» Fully and automatically indexed for ease of access

» Tagged and categorized with metadata, to simplify searches and filtering

» Version-managed, so that duplicate or outdated copies don't create confusion

» Easily and reliably curated, so that data can be purged from multiple locations to ensure privacy, data integrity, and regulatory compliance

Does your current data protection model offer those capabilities?

Models of Data Protection

Every company's data protection systems are different, but they all fall into one of two basic structures:

» **Silo:** Each branch, department, or even individual worker maintains a separate storage system. The systems may be networked, but the data's physical location is still an important factor in locating and using it.

» **Federated:** All data is available anywhere. The physical location of the data is irrelevant, because anyone with the appropriate permissions may access it from anywhere in the world, at any time.

The silo approach

The *silo approach* to data protection is the traditional, classic approach. A company typically arrives at this structure organically as it grows. For example, a small business may begin with one server and one location, and then add branch offices. Each branch office needs a server, so now there are two data silos, each self-contained. Then each salesperson needs on-the-road access, so a shared online drive is created. Now there are three silos. Then Human Resources wants its own separate server, citing privacy concerns. Now there are four silos. Each salesperson gets a tablet and a phone. Now each employee's individual Microsoft or Google account is its own mini-silo. You get the idea — before long, a company's data systems are a jagged skyline of individual data silos, none of which are more than tangentially connected to one another.

This approach works all right for the most part, in the way that small-town governments work all right without any state or federal oversight. Workers grind through their daily activities, and money gets made, and minor problems get discovered and solved. Maybe a server crashes and is restored from a backup. Maybe someone's iPhone dies and the data has to be reloaded. Maybe the HR director's laptop gets a virus and a couple of personnel records have to be restored from an archive. But life goes on.

The silo approach has some drawbacks and limitations, though. This approach, while it is "basically okay" for daily operations, has some endemic problems, such as these:

» **No easy way to locate data:** Because there are so many places data can be stored, quickly locating a given piece of data can be difficult.

- » **Multiple file versions:** In the course of everyday business, people email files back and forth, upload and download to shared drives, and even transfer files wirelessly from phone to phone.

- » **Rampant duplication:** When each individual silo is backed up separately, your data has a great deal of redundancy.

- » **Inconsistent levels of protection:** Without one central data store, it isn't a sure thing that all the data — in all the various silos — is adequately protected in the event of a large or small data loss.

- » **Nobody watching for problems:** Without a central repository of data, you have no way to monitor data across the organization.

- » **Complicated governance:** If your company needs to ensure compliance with internal or external regulations, apply a retention policy, or place a certain user's data on legal hold, having a variety of data repositories presents huge complications.

REMEMBER

The traditional "silo" approach has some significant limitations. An organization may have grown up with this approach organically, but at some point (and that point is actually sooner than most organizations assume it is), the silo model ceases to be sufficient or a good value.

The federated approach

Taking a federated approach to data protection helps overcome many of the drawbacks of traditional silo systems outlined in the preceding section, and a cloud system is one of the best ways of achieving this federation.

A *cloud* is a secure computing environment that users access online. The cloud can offer storage, processing power, applications — nearly anything that a physical computer can offer, but can scale on demand to meet any storage and computational needs. Cloud-based data protection has many advantages over the traditional silo approach, addressing most of the problems I point out in the preceding section.

The word *cloud* is actually a very good metaphor, because a cloud is:

>> Somewhere "out there" in the wide-open space of the online world, rather than tied to one physical location

>> Viewable and accessible from many locations at once

>> Variable in size and shape, with no fixed borders

Similarly, cloud storage is *not tied to a physical location.* This is a very attractive feature for companies with multiple branch offices all over the country or all over the world. Whereas a particular silo exists in a specific office, which might or might not be open when you need its data, a cloud service is always equally close at hand, no matter where you are.

In silo-based storage, if a company needs to gather files for a legal request, someone scours each individual system to locate the required files, including physically obtaining and copying devices that may be deployed remotely. In contrast, when using cloud to federate data en masse, multiple parties can have immediate access to the most up-to-date data at a moment's notice. Chapter 4 contains a use case that illustrates this advantage in more detail.

TIP If you store all your data in a cloud, rather than in individual silos, you can work with that pool of data as a *federated whole.* You can analyze it for trends, opportunities, and potential threats, you can back it up, you can search it, and you can expand or contract the available space to meet the company's needs.

Cloud storage is *infinitely scalable.* With cloud-based storage, you never run out of storage space or processing power. The provider allocates or contracts resources as needed automatically, making your compute and storage pool fully *elastic.* When your company is smaller and needs less, you pay less. As it grows and needs more, your cost increases only incrementally.

REMEMBER Don't worry if you don't quite understand clouds fully just yet. Chapter 2 explains how it all works in more detail. For now, just know that a cloud-based data protection system makes all data available online, all the time, regardless of its original storage location, and that makes for more flexible, secure, and easy-to-manage data.

Chapter **2**

Modern Data Protection with the Cloud

In this chapter, I explain how a cloud-based data protection system works, and how it can benefit an organization. You also find out the differences between software, platforms, and infrastructures as cloud services. If you're still a little fuzzy on this whole "cloud" concept, this chapter should help clear up the confusion.

This chapter also outlines the basic concepts involved in cloud-based data protection. You learn how a cloud system's ability to retain a single secondary copy of each file pays off in terms of search, backup, archiving, governance, analysis, and risk awareness.

Understanding Cloud Services

As I explain in Chapter 1, a *cloud* is a secure computing environment that's based online. Many types of cloud services exist, from simple applications to full-featured software environments that support every aspect of a company's IT activities.

A company (or individual, for that matter) can employ a cloud service in any of these ways:

» **Software as a service (SaaS)** allows IT to provide access to applications from the cloud, rather than running them on their own internal systems. Most services also provide file storage.

» **Platform as a service (PaaS)** provides a cloud-based operating system on which companies can deploy their own applications, which they can then offer to their customers.

» **Infrastructure as a service (IaaS)** provides the cloud resources that enable platform customers to create and control their own platforms.

Your employees are probably already using one or more *software as a service (SaaS)* applications for handling business data. It's common for end-users to store and share files using systems like OneDrive, Dropbox, and Google Drive, and to use online applications like Microsoft Office Online. Although these systems are handy, and can help end-users be more productive in their daily work, they don't do much for a company's overall data protection. In fact, when users store files in individual online silos like these, they can hinder the company's efforts to keep track of data (unless you're using a data protection platform that is aware of SaaS applications and can interface with them).

Many cloud-based services are more complex than simple applications, though — they are entire integrated cloud stacks, typically referred to as being *cloud-native.* The integrated data protection solution that Druva offers, for example, tightly binds elements of all three layers to optimize how data is captured and stored, how it is made available, and how it can be interconnected with other systems*.*

These integrated cloud systems, in turn, work with an *infrastructure as a service (IaaS)* provider that supplies and maintains the hardware. As with all computing systems, clouds rely on underlying hardware to operate. An infrastructure provider maintains huge rooms full of servers in warehouse-sized datacenters, all secured and protected using the latest and most powerful technologies. These servers combine their resources to create a giant pool of computing resources, which can then be allocated dynamically to various cloud-native systems as needed. So even though

hardware is involved in cloud storage, it isn't your responsibility. You don't have to buy the hardware, maintain it, protect it, or worry about it.

REMEMBER There's a lot more to know about cloud services, but the preceding explanation should be enough to help you follow along with the rest of this book. Just know that the kind of enterprise-wide data protection service that this book outlines is a type of cloud service and platform, and that it offers a suite of integrated tools on a common underlying data set to accomplish a goal while freeing you from having to think about hardware.

Cloud Data Protection Basics

In the silo approach to data protection, which you can learn about in Chapter 1, a company's data is scattered across various locations and storage types with no central oversight. The data typically falls into these basic categories:

» **Datacenters:** The main servers at the main location. A few decades ago, it was assumed that all the company's important data would be stored here, but that's no longer the case in most organizations. The central datacenter may use a combination of physical servers and virtual machines.

» **Remote offices or regional datacenters:** Often geographically distributed, operations at these locations are often challenged by limited IT infrastructure, poor network connectivity, and little or no on-site IT staff.

» **Endpoints:** These are individual user devices such as desktops, laptops, tablets, and phones. Endpoints may be corporate or personally owned, and can be highly mobile, putting data at an increased risk of breach or loss.

» **SaaS applications:** The rapid adoption of applications like Microsoft Office 365, Google Apps, and Box has challenged IT by introducing even more data sources to protect and manage.

» **IaaS:** Some organizations already rely on cloud providers for services like storage and database access, often without realizing that the data will be without the same comprehensive data protection they would demand if the data were onsite.

Just as these disparate data silos create inefficiencies in the data protection process, so does the lack of a single control plane. Multiple data sources don't simply mean more repositories to protect and manage — they also mean more data sources to monitor for compliance, and more places to search across when doing electronic discovery.

So what's the solution? In a word: *federation.* You bring together copies of the data from all sources, and use metadata and full-text indexing to turn it into meaningful, searchable data.

At this point, maybe you're thinking something like this: *It's a nice idea, but moving everything to the cloud would cause a massive disruption that we're not prepared for. There's no way my branch offices are going to give up their servers. And I'll have to pry that iPhone out of my assistant's cold, dead hands.*

But wait — nobody said anything about giving up your current methods. A cloud-based data protection platform works *with* your current data systems; it doesn't replace them across the board. It mirrors, indexes, and tracks all your existing content so that the content can be accessed and acted upon, from the cloud, at any time.

Over time, you might find that certain data is better off stored solely or primarily in the cloud, where it's safe and always available, but you don't need to make any dramatic changes to take advantage of cloud-based data protection.

Centralized access and visibility

When your organization implements cloud-based data protection, the provider collects and consolidates raw and extended metadata from various sources, such as file servers, virtual machines, endpoints, and SaaS applications, to create a single secondary copy in the cloud. This copy can be used for multiple purposes, and is automatically updated by the system to capture changes.

That paragraph contained a lot of data, so I'll consider it piece-by-piece:

» **The provider creates a single copy of each file in the cloud.** That means if your organization had multiple identical copies scattered across a dozen devices, the cloud stores

only one copy. Eliminating duplication in this way is called *deduplication.* It's a key factor in minimizing your storage volume and the associated cost.

» **The copy in the cloud is secondary.** That means the primary copy — the original — remains where it started. The process of storing the secondary copy in the cloud is invisible to the end-user and the existing processes.

» **The platform collects and generates metadata.** Metadata means *data about data.* The metadata for a file includes its type, size, owner, date created, date modified, and any keywords or subject lines. Metadata is important for searching and categorizing files so you can easily find what you need.

» **The data can be accessed for multiple purposes.** For example, suppose HR wants to ensure that Social Security numbers, which pose a risk to the business, aren't stored anywhere that could result in a potential breach of that data. Cloud data protection services enable the business to search across the full set of data to identify the applicable files and the systems where they are stored, so that appropriate security measures can be put into place.

Centralized backups and archives

In the traditional silo approach, the various locations and media each might have its own backup system. For example, each branch office might have its own backup software for its servers, running incremental backups each night and full backups once a month. In addition, each endpoint device might have its own separate backup, such as employee iPhones backing up to iCloud and laptops using a third-party backup utility to back up data to a branch server. Although such systems may provide basic data protection, they're difficult to administer, and nearly impossible for IT staff to adequately monitor.

The federated system I describe earlier in the chapter not only simplifies and unifies backup; it also makes it more complete and comprehensive. Redundant backup copies aren't necessary anymore, because those golden secondary copies of all files are already on the cloud, and the cloud itself is well protected with built-in redundancy by the infrastructure provider. All your system and device data is synchronized at regular intervals of your choosing, so it's never out of date. I talk more about backups in Chapter 3.

The same goes for long-term data storage (commonly called *archiving*). Beyond the backups needed to guard against data loss, your company may need to archive previous versions of files, perhaps for years or even decades, for regulatory and internal compliance, legal and e-discovery support, and historical analysis. Archived data is traditionally stored on disks or tapes, but these media are unreliable in the long run, subject to damage and deterioration. The same cloud system that retains the current golden secondary copies of all files can also easily be configured to store as much historical data as you need. Chapter 3 delves into this subject in greater detail.

Centralized searches

Looking for a particular contract or spreadsheet? With cloud data protection, you can instantly search the full text of all files in the entire enterprise for that particular contract number, client name, or keyword. The files may have been copied from various locations or storage types, but all of that is irrelevant to your search.

When a secondary copy of a file is collected or modified, the platform updates its metadata in real-time, including a full-text index. That means it creates an index of every instance of every word in every file. So even if you only remember one word that was in the file, you can still find it. (Of course, the search results may be quite a long list if you're looking for a common word.)

Comprehensive indexing and meta-tagging comes in handy not only for run-of-the-mill internal searches, but for more formal situations, such as when doing electronic discovery, or *e-discovery* (the process of locating and analyzing data that is relevant to a legal action). I talk more about e-discovery and other kinds of searching in Chapter 4.

Centralized governance

In the context of data protection, *governance* means applying policies that affect the way the data is stored, managed, protected, and made available. Every business has its own policies and principles on these matters. Some of these policies are internal decisions, but many are mandated by outside agencies, government regulations, court rulings, and industry standards. I look at governance in more detail in Chapter 4.

When all enterprise data is federated, governance can also be centralized. Executives don't have to worry about inconsistent policies across branches, or inconsistently applied corporate policies, causing embarrassment, inconvenience, legal challenges, or worse. Companies can:

>> **Increase visibility** into business data dispersed across laptops, desktops, servers, mobile devices, and cloud applications

>> **Ensure compliance** with all applicable regulations for auditing, investigation, and e-discovery needs

>> **Fine-tune security and privacy settings** for different kinds of data, different data residency requirements, and different user roles

>> **Protect data on remote devices** with geo-location, remote wipe, enforced encryption, and other security features

Centralized analysis and risk awareness

When your company's data is completely visible to those who need to see it, it opens up a whole new world of analysis capabilities. Companies can identify trends and patterns that may point to threats, risks, and opportunities that might not otherwise be apparent.

Many enterprises analyze data in a reactive way, when legal or compliance teams inquire after an "uh-oh" event or some other data loss incident occurs. But with a federated cloud data protection system, proactive compliance monitoring can be applied to assess data at rest, where it resides on devices or services. Full-text data indexing means companies can look deeper, beyond metadata, and dive into the data in the files. They can identify files that contain intellectual property, personal health information, personally-identifiable information, or other data buried within unstructured data sources that shouldn't be allowed to escape into the wild.

TIP

Data can also be monitored for anomalies using automated templates, so it doesn't require painstaking human analysis. For example, a template can trigger a warning if a large number of files are renamed or deleted within a certain time period, which could indicate a malware attack or rogue employee activity.

Cloud-Native Technology

Cloud-native means your technology is built from the ground up to leverage the benefits of the cloud. A true cloud-native data protection approach doesn't simply cobble a cloud connection onto existing management systems. A cloud-native approach takes advantage of the unique features of a cloud to manage data in smarter, tighter, more efficient ways than were ever possible with traditional data protection systems.

Because it does not require an intermediary layer between older deployments and a gateway appliance, a cloud-native approach eliminates bottlenecks, boosting both performance and availability. This type of approach also offers versatility, allowing enterprises to use it as a convergence point for other important activities. It doesn't use cloud storage as the technological equivalent of a warehouse; rather, the data can be used for many purposes, including backup, disaster recovery, and legal investigations.

REMEMBER In this chapter, you learned how cloud data protection works, and how it both simplifies and enhances an enterprise's ability to protect, archive, search, govern, and analyze its data. The upcoming chapters look at some of these benefits in more detail, starting with backup and disaster recovery in Chapter 3.

Chapter **3**

Backup and Disaster Recovery

As Chapter 2 shows, a cloud data protection system provides benefit in three main ways: protection/backup, governance, and risk management. I address each of these in its own chapter, starting with this one.

I don't have to sell you on the value of having a comprehensive IT disaster recovery plan that includes reliable, recent backups, right? Being prepared for recovery of all types is just good corporate hygiene, a prophylaxis against any number of data loss horror scenarios. But how to best generate reliable, usable backups and be fully prepared for disaster recovery? Your current methods may not represent the best protection and the best value.

In this chapter, I compare the traditional silo-based backup approaches that many companies have taken to the federated approach offered by cloud data protection. This data collection can also be used for data governance and analysis, as I explain in Chapter 4.

Traditional Backups and Their Limitations

The classic approach to backups has tended to mirror the classic approach to data storage itself, which I describe in Chapter 1. It's based on discrete elements, or "silos." Each server, desktop, laptop, and mobile device has some sort of backup mechanism, but they aren't necessarily integrated, or even harmonized. For example, the central office might have a powerful backup appliance in its datacenter that handles all the servers and desktop PCs in the building, but each branch office might have its own completely different system that doesn't interface with the main one.

As you can imagine, this hodge-podge approach to data protection is an administrative nightmare. But because backup approaches grow organically along with the storage media, your IT people might not notice that it's creeping up until the situation is really unwieldy. First it's a couple of other small backups besides the datacenter. Then it's one per branch. Then it's one plus a few extra devices . . . and before you know it, you need full-time people at every branch who do nothing but back up and secure data. And even then, there's no foolproof way to ensure that nothing falls through the cracks.

REMEMBER

Some legacy backup systems that protect multiple data silos purport to be "cloud-based," but they are simply traditional hardware-based backup systems that tie into the cloud. You get some cloud benefits from them but miss out on others.

For example, some legacy backup vendors require the use of a gateway appliance to link on-site systems with cloud systems. Although this approach enables enterprises to govern the flow of the data to the cloud, it creates other problems. The gateway becomes a bottleneck both from a networking and a reliability standpoint; if the gateway fails, backups to the cloud don't happen, nor can recoveries. In addition, such a hardware-based solution has no facility for ensuring deduplication of data across sites; it actually sends more data than is necessary for secondary storage.

Another common option is the hosted model, in which the cloud service provider essentially duplicates the on-premises architecture of the client. The advantage of this route is primarily an accounting one: switching from capital expenditure (CapEx) to operational expenditure (OpEx). This method, though in

the cloud, still requires on-request capacity expansion, load-balancing, redundancy, and all the other on-premises considerations, typically at a much higher total cost of ownership.

The Cloud-Native Backup

As Chapter 2 shows, *cloud‑native* means your technology is built from the ground up to leverage the benefits of the cloud. With a cloud‑native approach, backup can be just one component of a full data protection system. Backup is federated across all the different data sources, the same way data aggregation is. Backup is fast and lightweight on company resources because cloud‑native models don't suffer from single‑point bottlenecks. Data can be gathered en masse, globally, with zero performance degradation.

Here are some of the major benefits to the cloud approach to backup:

» **Stronger, more flexible data protection:** With cloud-native backup, you go far beyond simply moving raw data files from one storage medium to another. Cloud data protection is intelligent. The cloud storage system automatically catalogs and tracks extended metadata, file movement, file name changes, and file data modifications. It can even provide early identification of ransomware by looking at patterns of file changes.

» **Global availability and consistency:** Centralized management of backup eliminates the use of legacy silos and manual errors, making it easier for administrators to manage policies and restore data. No matter what data needs to be restored, you never have to wonder where it is backed up or who is in charge of it. The data is simply *there.*

» **Reduced infrastructure:** You no longer need to plan for, acquire, update, and manage hardware. Cloud providers manage this all for the organization and provide on-demand elasticity and scale.

» **Higher levels of security and compliance:** Cloud-native backup eliminates worrisome "what-ifs" from the security picture, providing higher levels of security overall. Cloud providers undergo significant security audits and generally must meet far more stringent requirements than the average business data center. Additionally, the global reach

of the cloud makes it straightforward to comply with data residency requirements. And, stored data is protected with sophisticated encryption and data scrambling, and only your trusted representatives can decrypt it.

>> **Cost savings:** Although backup and disaster recovery are must-have services, there's no reason for them to be in-house services. A company can achieve significant savings by outsourcing these to the cloud. Just imagine that for a moment — never having to buy and maintain backup hardware, or allocate personnel for performing backups and preparing for data disasters. How much could you save in hardware, IT personnel hours, and maintenance? Likely at least 50 percent.

Types of Data Protection

Enterprises typically are concerned with protecting two kinds of data via backups:

>> **Servers:** The entire content of a server is backed up as a whole. If that server becomes unavailable for any reason, whether it's a dead hard drive or the datacenter burning down, IT can re-create that server in its entirety by restoring that backup data to a fresh computer.

>> **User data:** Data identified as valuable data is copied to a backup location to preserve a secondary copy of it in the event that something happens to the primary (original) files. Keep in mind that end-user data is often spread out over multiple mobile endpoints like smartphones, tablets, and laptops.

>> **Cloud workloads:** Data generated by SaaS, IaaS, or PaaS applications, hosted by one or more cloud service providers, is unified and either mirrored at different cloud locations or replicated on physical media (an increasingly obsolescent technique).

Protecting servers

When backing up servers, you consider the original location of the data. That's because if you ever need to restore that data, the first question you'll need to answer is "Which server was it that failed?" Be warned, the answer can become complex if you're

using a variety of cloud-service environments to maintain your databases! Based on that answer, you'll find and restore the corresponding backup.

Servers are backed up as overall images of the hard drives at a certain moment in time, called *snapshots.* When something goes wrong with a server, you can restore to any available snapshot. When you restore, you can restore the entire snapshot — which wipes out whatever is on the hard drive currently, replacing it with the image from the saved snapshot — or you can recover individual files and folders.

A snapshot-based recovery system has many advantages. For one thing, it's quick — very quick, because the operating system doesn't have to work with a file system. It just dumps the data into the hard disk, exactly as it was when the snapshot was taken. If you are restoring individual files, it's even quicker, because you can avoid restoring data that was not lost. Being able to easily manage both scenarios — regardless of the location of your data — is important.

Disaster readiness

In the case of a major failure, where hardware is damaged, or a datacenter goes down, you won't be able to restore the snapshot back to the original location, and business operations will be disrupted. A cloud data protection system provides mechanisms for immediate failover to minimize downtime, using cloud-based disaster recovery (DR), and instead quickly spin up a virtual instance of that server using its snapshot at a moment's notice to get business operations back up and running. This is particularly vital when data is distributed across a variety of cloud-service environments like EC2, EBS, RDS, and Redshift.

Restoring from a previous snapshot can also recover a system from viruses, ransomware, and other malware that a server may have recently become infected with. That's because you're taking the server back to an earlier point in time, before the infection occurred. It's like time travel, but for computer data. By retaining multiple snapshots for each server, you can go back to a variety of different points in time.

Druva's product, designed to back up and protect endpoints, servers, and cloud-service workloads, unifies backup, disaster recovery, and archiving in the cloud, thus removing the burden of having to protect legacy infrastructure. The product works

with physical servers and virtual machines, and allows IT departments to quickly replicate and spin up new systems for testing and development, all from within the cloud.

Realistically, the likelihood of needing a particular snapshot drops dramatically after 30 days or so. For maximum protection and flexibility, or to meet regulatory requirements, your company might want to maintain months or even years of snapshots, but storage space costs money.

Cloud-native products save you money while keeping your options open by storing the newest snapshots in fast-access cloud areas, and older snapshots in more economical cloud areas that are slower to access. Data that's 30 days old or younger is considered *hot data* and is stored where it is easily accessed. Data between 30 and 90 days old is considered *warm* and is stored in mid-tier storage. Data that's more than 90 days old is considered *cold data* and is stored in the least expensive archival areas. Data moves seamlessly between storage locations according to pre-configured policies; data can easily be retrieved at any time from any location.

Protecting user data

Recovering user data is a more granular operation than the wholesale wipe-and-replace of an entire server snapshot. You can recover specific files or groups of files, in addition to the entire device snapshot, and you can choose what historical versions of them you want.

Druva's product manages and ensures end-user data availability and data governance across endpoints and cloud applications. Data from cloud applications like Office 365 and devices like laptops and phones is backed up. Policies are created by IT that manage the file types and locations that are backed up, and users can further refine these settings. If a device is lost or stolen, Druva's product enables you to remotely wipe any sensitive data that is on it, as well as geo-locate the missing device.

Some of the ways that a cloud data protection system can help with user data backup and restore include:

>> **Frequent, automated backups:** Time-indexed backups of user data and user-specific system and app settings can be configured to occur at regular intervals (such as every few

minutes or hours), so a system can be easily returned to its original state when problems occur.

>> **Immediate data access:** Because it is cloud-based, administrators and end-users have immediate access to data from anywhere.

>> **Comprehensive coverage:** Data can be backed up across *all* end-user data storage locations, including mobile devices and cloud applications.

>> **User-friendly interface:** IT admins can sign in to an easy-to-use control panel to configure backups and restore files and settings, with zero impact to users.

>> **End-user configurability:** End-users can take control of their own backups by adding folders and self-selecting the data for backup. End-users can also self-restore data and settings.

>> **Anomaly tracking:** A robust protection platform tracks data anomalies (strange things), alerting IT to potential ransomware attacks.

Protecting cloud workloads

The challenge is less about physically protecting data and more about ensuring its fast recovery. Top-tier cloud service providers use automatic, redundant systems that make actual data loss unlikely. However, if you lose business because of slow return-to-operation (RTO) performance, the data might as well have been lost.

The keyword for protecting cloud workloads, generated from SaaS office tools on laptops to IaaS and PaaS database environments on VMs, is management. A viable data protection and management solution must provide:

>> Centralized administration across multiple cloud regions and accounts

>> Automated DR testing

>> Customizable scheduling and retention periods

>> Cross-region restoration of snapshots and machine images

In addition, capabilities such as fast global deduplication can be make-or-break features.

Six Backup Musts

I believe that cloud–native backup systems are the way to go, but some universal best practices for enterprise–level backups apply to any technology. Here's a quick list of important "musts" to consider, no matter what kind of backup and disaster recovery systems you use:

» **Use off-site storage.** If you store your backups in the same building as the computers they protect, and then the building is destroyed, your backups aren't going to do you much good. A cloud-based backup system is, by definition, off-site. Depending on your industry and the regulations that govern it, local residency laws may require you to store your backups in the same region as your business. A cloud-based system can be easily configured to conform to such a policy.

» **Cover your remote sites and users.** Important data isn't simply stored in your datacenter. Review your backup plan to make sure that the solution you choose deploys automatically to *all* end-users who need protection.

» **Review the scope of your backup plan.** You're probably protecting data files and email, but what about other user-specific data sets such as profiles, system and app settings, and personal folders? Make sure that you are backing up the important configuration and application files that your users need to be productive, not only their data files.

» **Assess backup frequency.** How often are you backing up? Every day? Every eight hours? Consider whether you may need a more aggressive schedule, at least for certain key users, to minimize productivity-killing data losses.

» **Validate your retention policy.** How long are you retaining your backups? Review your policies, and if needed, adopt a longer retention policy to meet internal objectives, especially for key people and departments. You may need a longer retention period to comply with legal regulations as well; make sure your legal department is looped in on this decision.

» **Reassess policies periodically.** Although the preceding measures might provide sufficient protection for the foreseeable future, I recommend that you revisit your backup policies approximately every six months to ensure that they meet your organization's needs.

Chapter **4**
Accessing and Analyzing Data

I f you've read Chapter 1, you know that data is your most valuable resource. If you don't have a way to efficiently locate and extract the data you need, in a timely manner, you aren't making the most of your resources.

In this chapter, you learn how a cloud data protection system can make it easier to access and analyze data across your entire organization.

What Is Governance?

Every business has policies and procedures that determine how it (and its representatives) behave. Some of these policies are internally generated, but many of them are mandated by outside agencies, government regulations, court rulings, and industry standards.

In basic terms, *data governance* is the process of applying a company's policies and procedures to data protection. It includes the techniques and policies that measure and control how data systems are managed. It ensures that IT assets are implemented

according to agreed-upon policies and procedures, and makes sure they're properly controlled and maintained.

Data governance has become quite a bit harder in recent years, because today data is everywhere and anywhere, across devices and cloud services. Analysts predict that by 2020, more than 50 percent of all corporate data will reside outside the datacenter. This means that IT and security teams face greater challenges than ever to ensure data security, compliance with regulations, and business continuity.

Some of the most important aspects of data governance include:

>> Making data visible and accessible wherever it is, across cloud services and mobile devices, which may be anywhere in the world

>> Ensuring compliance with any applicable regulations about data storage, access, and management

>> Responding quickly to compliance inquiries, internal investigations, and other legal department requests

>> Identifying and taking action on identified data risks

>> Controlling who has access to important files or data

>> Guarding against privacy breaches and data leaks involving personal or sensitive data

>> Setting policies governing data retention and version management that balance data storage costs with protecting the company's interests

>> Monitoring for early warning signs of malware infection or other anomalies

>> Reviewing activity logs and other forensic tools to learn why problems occurred and strategize to prevent them in the future

Using a data protection system that centralizes access to all company data can make a dramatic difference in the ease and accuracy of all these activities.

The Traditional Approach to Governance

With traditional data storage and data protection systems, data governance is often messy and inconsistent, with policies for data access, backups, and privacy set by individual IT managers at branch offices. Even if you receive directives from HQ regarding data protection, you have no guarantee that things won't fall between the cracks when more than one person is implementing the directives.

To illustrate the limitations of traditional data governance, here's a scenario: You have a dozen branch offices, each with its own servers that hold data about orders and inventory. Each employee has a desktop PC that accesses the servers, and each salesperson has a tablet and a smartphone. Important data is spread out across multiple servers and devices, and there is no central gathering point.

Now suppose that a VP wants copies of all orders for the past year involving a certain hazardous material. She sends an email to the branch managers. Each branch manager has an assistant who locates the files. Most of the files are on the branch file servers, but a couple of new orders still exist only on a salesperson's tablet. Three or four days go by before the VP receives all the files from the branches.

The VP's assistant starts going through the files and finds that several of the orders are duplicates, or that one is an earlier version of another. The assistant needs another day to sort through them and remove duplication.

The VP then compares the sales orders to a report from the warehouse and finds that some orders are missing from one of the branches. She contacts the branch and finds that the assistant who prepared the paperwork for some orders failed to appropriately categorize the orders as including those hazardous materials. Because document categorization is done manually at that branch, the mistake was easy to make.

But wait — the problem gets worse. The reason the VP needed this data is that the company has a pending lawsuit involving a hazardous materials sale. The legal department needs to put a

hold on the collected sales data. Because the VP had to tell the branch managers about the lawsuit to emphasize the importance of thorough compliance with her request, the gossip mills at the branches are churning, and wild speculation about the lawsuit abounds.

Sounds like quite a mess, doesn't it? This is exactly the kind of all-too-common scenario that companies find themselves in when data has no central management point for data: no corporate-level data protection system.

The Cloud Approach to Governance

With cloud data protection, the previous story would have gone very differently. A cloud system can:

» Centralize organizational visibility into business data dispersed across many storage locations

» Manage governance over business data for compliance auditing, investigation, and e-discovery needs

» Automate compliance monitoring for proactive identification of potential data risks across disparate data sources, devices, and users

» Defensibly delete data that's stored in the cloud, within the backup snapshots, and on the primary device, if required

Consider this version of the story instead: The VP signs in to the cloud data protection system and uses a targeted search involving keywords and file dates to pull a list of all the orders in the last 12 months involving that particular hazardous material. She then emails the data to the legal department. The entire process might take at most about 30 minutes. The branch offices never even know it's happening. *The end.*

That's what happens when you have a federated repository of data, with real-time, full-text indexing, searchable across all platforms and locations. You get the data you need, when you need it, without involving anyone who doesn't need to know about it. A web-based control panel interface enables authorized users to search across all stored data, as well as to view aggregate data about the data storage systems and examine data access trends.

Having ready access to your company's data, without having to beg IT people or branch managers for it, can be a powerful tool for corporate decision-making and regulatory/legal compliance, as well as a time-saver. (It's quite a money-saver too, since time is money, especially when it's the time of highly paid executives.)

Those two versions of the data-gathering story point out some of the many advantages of a cloud-based data system, including these:

>> **Federated repository:** Because all data from all sources across the company is mirrored into a centralized cloud-based repository, the VP didn't have to ask the branches for anything.

>> **Instant access:** Access to the data was nearly instantaneous. The VP didn't need to wait for anyone to dig up and provide data.

>> **Deduplication:** Because the cloud system compares files and eliminates duplicates, there was no concern about duplicate orders.

>> **Version and life cycle management:** Because the cloud system manages versions, you don't need to be concerned about different versions of the same order. Outdated versions of files can be deleted or archived to get them out of the way so they don't cause confusion.

>> **Full-text indexing:** Because all files are fully indexed automatically when they are placed in the cloud, orders can't slip through the cracks because of errors in categorization.

>> **Silent operation:** Because the VP can discover data silently, without involving the branches, no one gossips or speculates about what's going on. The branch managers needn't take time out of their busy schedules to round up the data, and there's no risk that necessary data will be manipulated or deleted.

>> **Easier compliance with legal demands:** Because the cloud system has tools in place to assist with legal and regulatory inquiries and holds, the legal department doesn't need to keep coming back to managers for more data as an investigation proceeds. Chapter 5 looks at legal and regulatory demands in more detail.

REMEMBER

Remember, having a single secondary-copy repository doesn't require you to change your work practices on your primary systems. A cloud data protection system allows a company to keep handling data using its existing processes and systems. The cloud system runs in the background, looking at the data and collecting it in a cloud repository. All the data visibility and access benefits I discuss in this chapter are available using that secondary data in the cloud, without interfering with your current business workflow. The same is true for the risk management capabilities of the cloud, discussed in Chapter 5.

Chapter **5**

Minimizing Data Risks

I t's a scary world out there. Threats to your company's data, as well as its overall well-being, loom on every side. Ransomware and other malware, rogue employees stealing data for a competitor, disgruntled staff intentionally deleting or corrupting data, careless workers leaking customer data . . . the list goes on and on.

In this chapter, you learn how a cloud data protection system can help protect your organization against many types of threats, providing an insurance policy against the potential data disasters your company may face.

Guarding against Malware

Malware is software that is written to cause problems with a computer system — usually to steal private data, delete valuable data, hijack the computer's processing capabilities, or disable systems. It's a constant threat to nearly every operating system. An organization can put up firewalls and use anti-malware monitoring software, but the occasional rogue infection still sneaks through. When this happens, having a recent clean backup to roll back to can be the difference between a slight inconvenience and the complete breakdown of business-as-usual for a week, or maybe longer.

Ransomware is an especially nasty, and increasingly prevalent, type of malware. An easy, low-risk way for criminals to exploit almost any network intrusion, ransomware encrypts a computer's data, and prevents owners from accessing their own data or computer systems until they pay a "ransom" to obtain a decryption key. And for those who do pay the ransom, there's no guarantee their data will be made available to them afterward. No industry is immune from ransomware attacks, but some, such as healthcare, have been especially hard hit. For example, a 2016 ransomware attack at Maryland's MedStart Health hospital network forced ten of its hospitals to operate without access to their central networks for more than a week. With ransomware attacks on the rise, organizations of all sizes have found themselves vulnerable and struggling to reduce risk or respond to an attack.

Most infections involve an unsuspecting individual clicking a tainted email or attachment. Systems with out-of-date or misconfigured software can also be compromised to spread ransomware. Windows systems have been the main target in the past, but this is rapidly changing, with Macs, Linux systems, and smartphones increasingly affected. No platform is safe. The widespread use of mobile devices has also escalated the risk of malware attacks. Although many companies protect data at the corporate firewall, employees often operate outside it, using laptops, tablets, phones, and other personal devices while on the move.

WARNING

Organizations may be tempted to cross their fingers and hope they won't be targeted, but the chances of ransomware or other malware attacks are high, with serious consequences. In addition to paying a stiff ransom, victims may suffer costly business downtime, and in some industries like healthcare, ransomware attacks must be reported as breaches of the Health Information Protection and Accountability Act (HIPAA), with associated fines and penalties.

Prevention techniques are useful, but limited. End-user awareness and smart browsing practices are important, as is regularly updating security, antivirus, and anti-malware software, as well as operating systems. However, such prevention provides a weak and variable level of protection. Malware writers are getting smarter every day, coming up with new workarounds, so a new vulnerability is always around the corner.

REMEMBER

A comprehensive data protection plan is based on a robust backup system, like the one that a cloud data protection system provides. Automated and time-indexed snapshot backups of data across servers, laptops, and cloud apps can enable the restoration of data to its original state, and as a result, organizations can access their data from any point in time prior to the attack. A zero-latency, time-indexed file system means that you can instantly jump back to the state of your data (in the small or large sense) to a time before the infection occurred.

Checking for Data Leaks

A *data leak* is a situation in which sensitive data gets into the wrong hands. A data leak can be the result of a rogue employee intentionally over-sharing the data, but it more often is caused by carelessness than malice. For example, suppose your R&D manager copies the technical drawings for your new, top-secret product, the Pear 8, to his laptop, to work on at home over the weekend. Then he stops by his favorite coffee shop and checks his email on the shop's Wi-Fi network. He clicks the wrong button when logging into the Wi-Fi, making the files on his hard drive accessible to others. Somebody takes advantage of that mistake, and copies all his data files to a thumb drive. When the opportunistic hacker checks them out later, he sees that he has the plans for the Pear 8, and posts them to Instagram. There goes your dramatic new-feature reveal for the Pear 8 that your marketing department had been planning for months.

REMEMBER

A cloud data protection system can help you guard against data leaks by monitoring file activities. For example, you could set the folders or drives that R&D uses to be monitored, so that you can generate a report showing who used which files and when. Such a report would have let you know that the R&D manager was copying important files to his laptop. No, it isn't a "data police" thing, designed to get the R&D manager in trouble. Rather, it's to help the IT department know where to beef up security measures. If it is known that he is taking important files home with him, your IT security people could equip his laptop with a higher level of security, such as an encryption algorithm that prevents files from being read if they are transferred to unauthorized devices.

Monitoring for Anomalies

An *data anomaly* is a situation where non-typical file activities are occurring. Anomalies can be either human- or machine-generated, and they are often early warning signs that something is about to go wrong. For example, a disgruntled employee getting ready to quit and go work for a competitor might transfer a large batch of important client files to his personal computer, or ransomware might encrypt thousands of files. If you can catch these activities before they come to full fruition, you might be able to minimize the damage.

TIP

Druva can be configured to identify and report a variety of anomalous conditions across your entire data collection that may signal something for your IT people to take a second look at. In many cases, early anomaly detection can prevent something much more serious and damaging from occurring.

Ensuring Legal and Regulatory Compliance

A data protection system should make it easy for the company to comply with all laws and policies regarding their data, and to react swiftly when issues arise that require the collection of data for a legal matter or a regulatory challenge. This isn't just theoretical capability. The world is getting more and more regulation-bound and litigious every day. According to the Norton Rose Fulbright Annual Litigation Trends Survey, 37 percent of large organizations reported having 20 or more pending lawsuits, and 42 percent of companies overall reported more than six lawsuits filed within the last 12 months.

Data privacy and access

In the past few decades, governing bodies have tightened data privacy regulations to protect both businesses and individuals. These regulations have greatly increased the complexity of data protection. Regulations are often different between countries, as well, so a multi-national company may have to conform to different privacy regulations depending on which country a branch office is based in. Similarly, certain regulations may specify that data be available only to workers in particular roles or regions.

TIP

These regulations require a granular level of data protection that most IT organizations do not have in place today, and that legacy, silo-centric models cannot accommodate. With a cloud data protection system, however, this type of compliance becomes a simple matter that is invisible to the end-user. You configure the system to meet your regulatory needs, and it just *works*.

Full-text data indexing means that organizations can look deeper into their data, to identify files that contain intellectual property (IP), personal health Information (PHI), personally-identifiable Information (PII), or other data buried within unstructured data sources that you don't want escaping into the wild. The cloud system can automatically scan aggregated data and alert the organization as necessary. IT staff can also select from preconfigured templates for common regulatory definitions (such as HIPAA, GDPR, or GLBA). The system applies these business rules when scanning the federated data set, making automatic a process that was once quite time consuming.

Compliance, legal, and IT teams see a single dashboard displaying the potential data risks by cloud application, by user, and by device. Administrators can drill down into the levels of detail as necessary to better assess and remediate those risks.

Data collection and retention

When a company's only data oversight is within the individual data silos, it's at the mercy of the individual employees who manage those silos. If one of them makes a mistake that results in a legal or regulatory penalty, it's the company that pays. It follows, then, that anything the company can do to shift the oversight and responsibility for compliance issues upward, to a central point of contact, can help minimize risk.

When you use a cloud data protection system, on the other hand, you get:

» **Legal hold management:** Your IT and InfoSec teams can work in partnership to provide legal teams access to manage legal holds on aggregated data that is held in place for the period of the legal matter.

» **Detailed auditing:** The system keeps full audit trails of user and administrative actions. It integrates logging data from cloud-based services to provide a composite of user data interactions.

- » **Federated searches:** Organizations can search across multiple metadata fields and parameters to identify data across the content base. This helps identify potential risks when facilitating investigation or compliance requests.

- » **E-discovery connectivity:** Some cloud-based systems, such as Druva, provide secure, direct connectivity with third-party e-discovery platforms to ingest the data without having to move it to an intermediary server.

- » **Defensible deletion:** Combined with federated search, or a unique file identifier, data that needs to be purged from systems can be quickly identified and deleted from the primary source and the data set stored in the cloud.

Chapter **6**

Ten Reasons to Use Cloud Data Protection

With the huge increase of data sprawl brought on by the mobile workforce and cloud services apps, managing a company's data has become much more complicated than it was just a few years ago. Cloud data protection systems can help navigate this complexity, providing a central access point, greater data visibility, less legal and regulatory risk, and cost savings.

Here is a summary of what you've learned about cloud systems in this book: ten important reasons to consider cloud data protection for your enterprise.

Assurance of Comprehensive Data Collection

With individual data silos, you always have the nagging feeling (probably based in reality) that you're missing some data when you try to build a comprehensive collection. And even when you finally get a complete collection, it's immediately out of date.

A cloud platform like Druva can reach into all the places where enterprise data hides, from servers and desktops to tablets, smartphones, and cloud apps like Office 365 and Box. You'll know you are getting *all* the data, and that it is continuously being updated.

Simplifies Backup and Recovery

Traditional backup systems can be slow, cumbersome, and expensive, both in terms of the hardware and the administrative cost. Wouldn't it be nice if backup just happened, and you didn't have to think about it?

A cloud data protection system makes backup "happen" like that. Because it stores a single secondary copy of all data in the cloud infrastructure, backup hardware is no longer needed locally. And because it constantly updates accordingly to a configurable schedule, it virtually eliminates the whole concept of performing a traditional backup. Backups, including historical ones from any point in time, are always available, from anywhere in the world.

Works across Locations Worldwide

Speaking of "anywhere in the world," that's what you get with a cloud data protection system. Because the cloud is location-agnostic, so is your access to it. Server crash? Your IT folk can restore its data just as easily from a taxi in Tangiers as from a desk in the datacenter. Need to quickly look up the details of a contract that someone in your Brussels office has on his device? It's just as simple to find as the contracts from the main office.

Some kinds of data are regulated, such that it must be stored physically in a certain region. Cloud data protection systems can work with that easily, because of the distributed physical architecture of the cloud infrastructure. Need your data hosted in a Germany-based datacenter? No problem.

Easy to Analyze Data for Trends

Because a cloud system indexes the full text of each data file, organizations can identify trends in data usage. A single dashboard provides access to the entire data protection system, with summaries of activity by service, by user, by device, and by date. Is a particular word coming up more frequently in a certain branch's internal memos than it used to, like *near-miss* or *downtime*? Is the sales team in a particular region writing fewer contracts after a new regulation went into effect? Having full, direct access to company data can help executives answer these and thousands of other questions relevant to top-level decision-making.

Makes Malware/Ransomware Recovery Easier

As you learn in Chapter 5, malware is a very real, insidious threat to your business's data. Educating users and putting robust anti-malware tools in place can help with prevention, but they can't help you once an infection occurs. To recover from a malware attack, you need recent, comprehensive backups from the time before the attack occurred. Any backup system can provide that, but the cloud-based data system, because of its more flexible backup capability, provides the most reliable, most thorough, most recent backups from which to recover.

Early Warning for Potential Data Access Anomalies

There are classic early warning signs when most types of threats occur. For example, ransomware may start renaming or encrypting files en masse, or a hacker bent on harming the company may delete thousands of database records at once. Constantly monitoring the company's data systems manually for such threats is inefficient, tiresome, and not cost-effective. However, with Druva, IT professionals can be alerted immediately and automatically whenever unusual activity is detected.

Ensures Compliance with Regulations

As I mention in Chapter 5, relying on individual IT departments to manage data protection compliance issues can put a company at risk because people make mistakes. The more people who are involved in managing data compliance, the greater the potential compounds. When you're dealing with regulatory and legal issues, though, a mistake isn't just a simple "oops." A company can be subject to significant penalties, fines, and legal judgments because of a simple oversight.

To make sure that all data storage conforms to applicable regulations about privacy, confidentiality, and retention, relying on a cloud data protection system is a much smarter strategy. The cloud system can be set up to monitor for certain types of legal and regulatory compliance issues, and can send alerts whenever adjustments should be made. In addition, depending on the type of data and your industry, cloud providers should provide organizations with SOC2, FedRAMP, HIPAA, and other compliance audits or certificates to ensure their holding data meets compliance mandates.

Makes E-Discovery Quicker and Easier

When a company's data is relevant to a pending lawsuit, a legal hold must be placed on the applicable data, and it must be made available to legal teams for e-discovery. As you learned in the story I tell in Chapter 4, digging up the needed data can be a time-consuming task when data is distributed across many storage locations, but with cloud data protection, data is proactively collected, making the entire process quicker and easier.

Invisible to End-Users

A cloud data protection system shouldn't force end-users to change the way they operate. Instead, it should work with existing systems, offering its benefits silently and behind the scenes. A cloud data protection system creates a single secondary backup copy of all data, drawing from the existing storage devices that

the end-users are already working with. Any queries, investigations, or legal holds put on those secondary copies are completely invisible to anyone who doesn't need to know about them.

Saves Money Compared to Other Options

Cloud-based data protection can save the average company 65 percent over a three-year period in overall expenses related to data protection, compared to traditional systems like hardware-based backup devices and silo-centric data analysis tools. So not only is the cloud better, but it's also a better value. Those cost savings come from areas such as:

» Not having to buy and maintain backup and other data system hardware and software

» Not having to store physical backup media off-site, and deliver it to or from that site daily

» Not having to allocate IT staff to manage multiple data systems hardware and media

» Quicker recovery in the event of a server or website crash, meaning less business interruption and less revenue lost

» Better managerial decision-making due to increased data visibility

» Intercepting potentially expensive data threats such as ransomware earlier in the process

» Lessened likelihood that the company will be subject to fines, penalties, or legal judgments resulting from data handling failures

Your journey to Data Protection for the Cloud Era starts now.

Your business can protect and manage enterprise data across endpoint, data center, and cloud workloads — in no time.

Druva.com

Data protection for the cloud era

Enterprises are making the shift to the cloud as they seek greater efficiency and agility in their business. They understand that only the cloud can deliver the visibility, reliability, governance, and cost savings that are needed to drive today's enterprise forward. This book shows you how to do just that — with a cloud data protection system.

Inside…

- Understand the basics
- Ensure data availability
- Achieve greater visibility
- Understand data risks
- Increase agility and control
- Make compliance easier
- Reduce costs

druva

Faithe Wempen is a veteran Dummies author with more than 150 books to her credit. She teaches about computer technology online and in the classroom.

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

WILEY END USER LICENSE AGREEMENT

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