commvault
solving forward®

A CommVault® White Paper

Simpana® Software—
A Unique Cloud-Enabled Platform

From a single console, Simpana® software seamlessly extends local backup and archive to a new tier of cloud-based storage.Virtually unlimited storage capacity from our integrated cloud storage partners helps users dramatically reduce costs and operational complexity while improving long-term compliance.

Contents

commvault®
solving forward®

What is Cloud Storage?

Cloud storage is offered by many companies in different packages, but all are a form of secure storage that you can write to and read from without having to actively manage. You no longer have to worry about RAID levels, parity, disk swaps, space management, replication, error condition monitoring and reporting, or any similar problems that you'd normally have to worry about when managing your own storage.

In a cloud storage model, your monthly fee is based on the amount of actual data present at the provider's site.

Cloud storage is not a disk array that you own, lease, or manage. It is not necessarily a virtual LUN from a larger disk array. A cloud storage service offers you an API through which you can send and receive data. They promise, via service level agreements (SLAs), to receive and store the data you've sent to them, and to send it back to you when you request it. As long as they meet the SLAs to which you've agreed, how or where they store the data is up to the cloud vendor.

'How' cloud storage is priced actually helps to understand how different it is than anything previously offered in the storage market. If you need to store 5TB of backup data in your own data center, you would purchase or lease a tape library, VTL, or disk system capable of holding at least 5TB. (Usually you would buy much larger than that to allow for growth.) If you were to use a traditional outsourcing model, you would do the same planning and configuration. You would contract with the outsourcing company to create and manage a disk or tape system capable of storing at least 5TB. But cloud pricing is very different. In a cloud storage model, your monthly fee is based only on the amount of actual data present at the provider's site (and sometimes also how many GBs you transfer to/from their site). It supports "burst pricing," which means if you need to store 20TB this month but only 3TB next month, you can do that with no configuration changes, or anything like that. All you do is pay for 17 less Terabytes next month than you did this month.

Data Types Appropriate for the Public Cloud

First, let's discuss what about cloud storage makes it really unique. It's affordable, it's totally dynamic (use as much or as little as you want with no configuration changes), and it's reliable.

However, since the data is ultimately going to be accessed over the Internet, it does not offer the same performance characteristics of local storage—especially when it comes to random writes to larger data sets. This excludes databases and anything that writes to disk like a database. The throughput characteristics of the Internet are also such that you wouldn't necessarily want to use cloud storage for large data sets that are constantly sent to and read from storage *en masse*. That excludes workloads with large working sets such as video production create, store, read, then delete large data sets every single day.

So, which workloads are left that have an access pattern that is appropriate for storage? Well, there are two: backups and what is now called "persistent data." Backup systems are well known to everyone; and, "persistent data" is all of the various data sets you have that you have to keep around (often for legal reasons) that you'd really rather not actively manage.

Backup systems do create a large amount of data each day, but the amount of data created each day is only a small portion of its overall storage requirement. For example, a shop that creates 1-2TB of backup data per night might need the capacity to store 30-180TB of data, depending on their retention requirements. Chances are, they also will write Terabytes upon Terabytes of data to the backup system and only occasionally read a few Gigabytes of backup. Add to that the fact that most restores are single-file restores where throughput is irrelevant. As long as the file comes back sometime in the next few hours, the user is typically ecstatic. As will be mentioned later, there may be some backup types that aren't well suited to cloud backup. A large database with a recovery time objective (RTO) too aggressive to be met via the Internet is one such example. But, the bulk of both backup and restore requirements can be easily met with cloud storage.

One pundit uses an amusing synonym for persistence to help define persistent data. He calls it *annoying* data; that is, the data that you have to keep around but you're probably never going to read—kind of like an annoying relative that comes for 3 days and stays for 3 weeks! You can't delete this data any more than you can kick out the annoying relative —you've got to put it/them somewhere. The annoying relative gets introduced to the local motel, and your persistent data gets introduced to the cloud storage provider. That way you can get the relative (or your data) back if you ever actually want to get them back, but you don't have to clean up after them, listen to them snoring, or monitor their RAID group for bad disks.

Environment Characteristics

Considering these two data types (backups and persistent data), which environments would most benefit from using cloud storage?

The first environment that would obviously benefit from storing backups to the cloud are small-to-medium sized businesses. The cloud offers them a very affordable and reliable place to store their backups—and it happens to be offsite. They do not create the amount of backups that enterprises create, so there are few bandwidth concerns for backups or restores. Depending on the size of the business, you could restore the entire data center within a few hours—especially since many small businesses have more download than upload capacity.

What about enterprises? Is their data appropriate for the cloud? Obviously, a data center creating many Terabytes of backups per night is going to have a challenge providing enough bandwidth to get those backups transmitted to a cloud storage provider. Therefore, cloud storage is perhaps not the best place to store backups for very large data centers. However, these same companies often have many small remote sites that also need to be backed up, and each of those remote sites has the same backup and recovery characteristics as the small-to-medium sized business environment previously discussed. These types of backup and recovery requirements are perfect for the cloud.

The cloud offers [small-to-medium sized business] a very affordable and reliable place to store their backups—and it happens to be offsite.

Now, what type of company would store their persistent data in the cloud? The type of company that would find this most attractive is a company that has identified a significant portion of their storage being wasted on the aforementioned "annoying" data. Such businesses want to move this data off to a different tier of storage because this data is simply just taking up room in data center and offering very little value to the company. They don't want to manage it; they want to stop monitoring it; and they can't delete it either. It's difficult to say how big your persistent data needs to get before the cloud is the best option, but if you related to the previous sentence then perhaps the cloud is the place to put your persistent data.

Bandwidth Considerations

Bandwidth is not free, and the bandwidth requirements of moving your data to the cloud must be examined along with the costs of meeting those requirements. These must be duly considered in any decision to move to the cloud. However, depending on the application, additional bandwidth may not be needed. Consider the following example.

A hypothetical company decides that a T-3 line is needed to meet the bandwidth needs of their business. However, they are a traditional "8 to 5" operation that has almost no need for bandwidth at night. This means that this customer could use all available bandwidth of the T-3 to send backups or persistent data to the cloud at no additional cost to their environment. (This assumes, of course, that there will not be significant numbers of restores or retrievals of persistent data during business hours.)

This example is obviously not true of everyone, but hopefully it makes an important point. If the data types and access patterns of a certain data set are appropriate for the cloud, it is possible to begin using the cloud without significant additional bandwidth costs.

Inhibitors to Adoption

There are a number of objections that come up when people are considering moving some data to the cloud. Each of these objections must be addressed and overcome by any cloud vendor you are considering. It is important to understand that not all cloud vendors are the same. They don't cost the same, they don't offer the same SLA's, they don't backup your data the same way, they don't all have the same levels of security, and they all have different levels of reliability.

The first objection that many people have is cost. They look at cloud storage costs that range from $0.15/GB to $0.75/GB per month and think, "Gee! I could do it much cheaper than that." However, rarely is that the case and any good cloud storage provider can help you do a full TCO analysis of doing it yourself versus storing the data in the cloud.

These TCO analysis models take into consideration a number of important factors. Let's consider a few.

Data protection

Remember that to have one usable Gigabyte of protected data, you need close to 3GB of actual disk (or more), depending on the raid level you choose. Even the least expensive RAID levels (parity-based protection, not mirroring) have a 15-20% overhead. Add to that a second replicated copy of the data (which any good cloud storage provider would have) and you end up needing close to 3GB of disk to store 1GB of data. Add to that hot spares and other additional components needed to ensure availability, and the cost goes up again.

Power and cooling

Disk drives run around the clock causing your power meter to constantly spin. And that electricity creates heat which must be cooled. Power and cooling requirements of disk drives are actually the primary cost factors in many data centers.

Technical refreshes

Once every few years your equipment will wear out and need to be replaced. If you're managing your own storage, the process of spinning up to new storage and migrating the data from one platform to the other is all yours. The cost, and all the downtime associated with it, is all yours, too.

Day-to-day monitoring and management

One of the biggest costs of managing your own storage is just that—managing your own storage. Disk arrays need to be monitored—by people. These people need to respond at all hours of the day and night to do the things that need to be done to make sure that the disk systems continue to run. The people required to monitor and manage the systems are not cheap.

Engineering and planning

Unless you are the smallest of companies, you can't simply go to your favorite tech store and replace your old systems with new systems. You need senior people to take a look at the options available to your environment and to evaluate all of those options, and to then work with prospective vendors to design the next generation storage system. The cost of these infrastructure engineering people often goes ignored when considering the costs of managing your own storage.

Perhaps after fully accounting for all of these costs, $.15-$.20/GB per month might not sound all that expensive.

One final consideration with regards to cost is that nothing is free. Remember that everything you ask the vendor to provide costs them money and will be reflected in the costs to you. For example, there are vendors who store only one copy of your data; they are obviously cheaper than those who store multiple copies. There are those who use storage systems that have built-in snapshot capabilities to protect from accidental deletion; those storage systems and those snapshots cost extra money, as well. You may want some insight into how your data is being protected and managed, and you may want to monitor that on a regular basis. Monitoring software and systems cost money. Remember to take all of these items into consideration in considering the costs of different vendors.

The next big objection that people have to overcome when considering cloud storage is that of security, and anyone considering storing their data into cloud must address this concern. This is another area where all vendors are not created equal. It is well advised to retain the services of a security expert when evaluating cloud storage providers. Make sure that this expert is completely happy with all of the answers with any cloud storage provider you are considering.

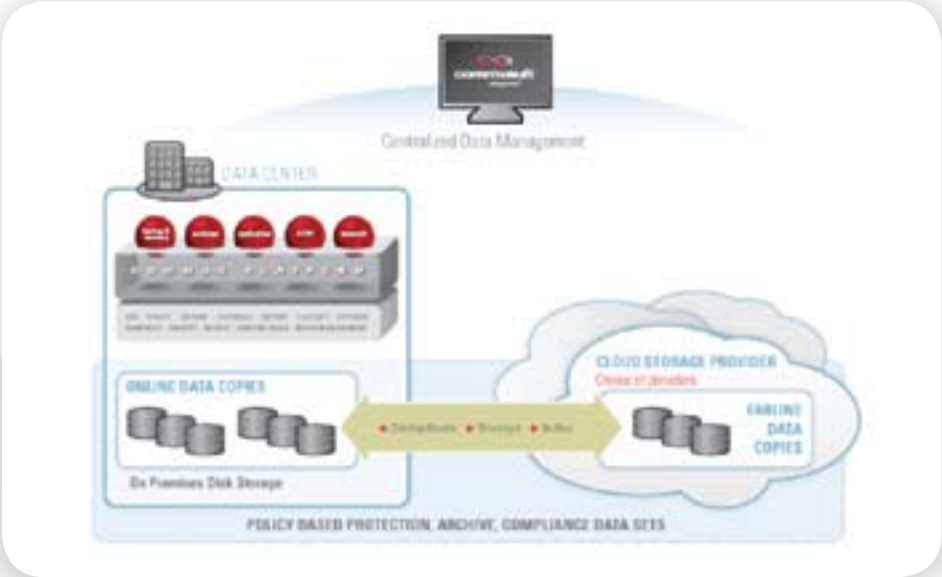
Reliability must also be addressed. Examine the history of the reliability of any company that you are considering. Remember again that all cloud storage providers are not created equal and just because some cloud storage providers have had major outages—even data loss—does not mean that all cloud storage is unreliable or unsafe.

The final objection that people often have to storing their data into cloud is about actually getting it there. For various reasons, including security concerns, cloud storage providers do not use the same storage protocols that you are used to in your on-premises environment. You'll learn about things like REST over HTTP and you may find yourself asking "how do I do that?" One compelling answer to that question is utilizing CommVault® technology and its Simpana® software cloud offering.

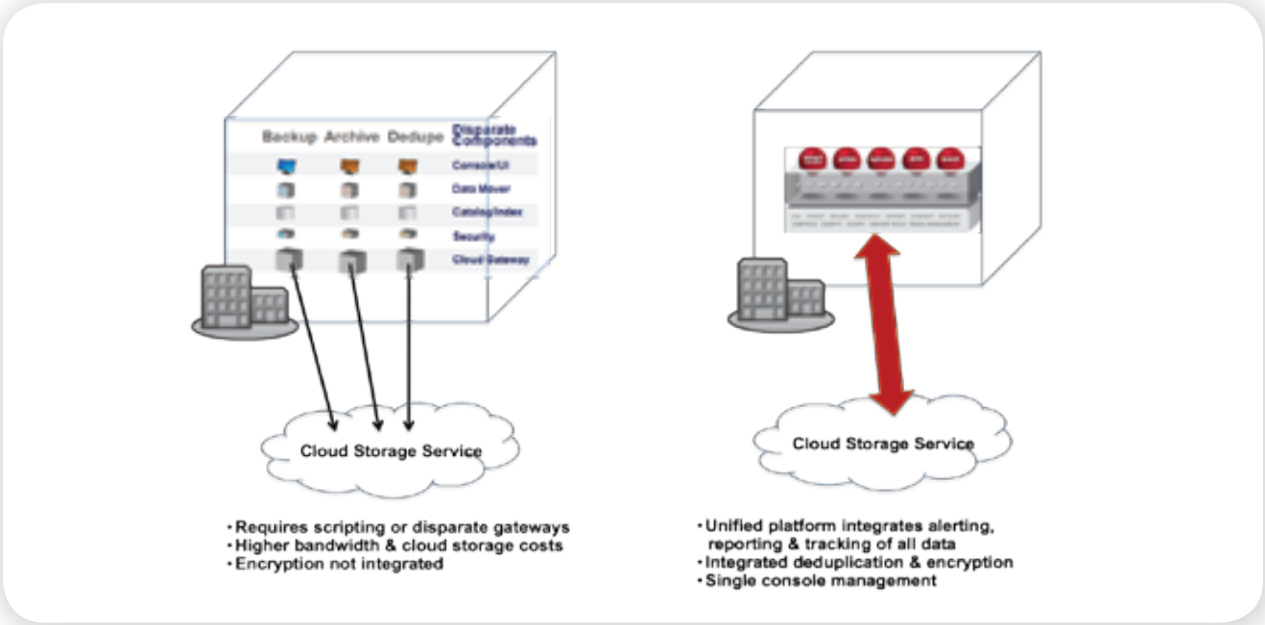
CommVault Simpana Software: Extending Data and Information Management to Cloud-based "Farline™" Storage

The CommVault Simpana software platform is built on a single codebase and provides a single data and information management foundation with modules for Backup & Recovery, Archive, Replication, Resource Management, and Search. Simpana software delivers data and information management across both physical and cloud-based storage, all managed from a single console. Combined with the virtually unlimited capacity of cloud storage, users now have a powerful solution for lowering operating costs, protecting remote data, providing disaster recovery for critical data, and improving long-term compliance management.

With a native cloud connector built in (leveraging REST over HTTP), Simpana software delivers seamless integration with cloud storage along with virtually any heterogeneous storage target including disk, tape, CAS, VTL or optical storage device. In addition, CommVault Simpana software delivers broad coverage of platforms including Microsoft, Unix, Linux, Netware and Mac for a broad set of applications and file types such as Exchange, SharePoint, SQL, Active Directory, Oracle, Notes, GroupWise, and more.



This unified approach to data and information management revolutionizes your ability to leverage the economics of cloud-based storage. By delivering one of the first REST-enabled solutions for data-aware movement into cloud storage, Simpana software manages alerting, reporting, tracking and data verification just as if cloud-based storage was located in your own data center. This singular solution eliminates the need for gateway appliances or scripting–reducing the cost and complexity of cloud storage integration.



Secure Farline™ Data Storage

CommVault technology lets you configure the right data encryption approach to meet your cloud storage requirements. You can protect data starting from within your datacenter or remote office with in-stream encryption that protects data from unauthorized access. Data that's "at-rest" in the cloud can also be secured using encryption, protecting it from exposure to anyone outside your organization.

Because the encryption keys are generated and kept in-house, you maintain full control. Your data is secure even from unauthorized access by your own service provider's IT team.

Minimized Network and Cloud-based Capacity

CommVault Simpana software delivers a variety of data reduction capabilities to minimize the amount of data sent to and stored in the cloud by up to 75 percent. During on-premises backup or archive jobs, Simpana software's block-based deduplication reduces data volumes by filtering out redundant data before it is stored on disk or sent to the cloud. Additional data management approaches such as incremental backups and data compression at the source can further reduce the amount of data transmitted and stored in the cloud.

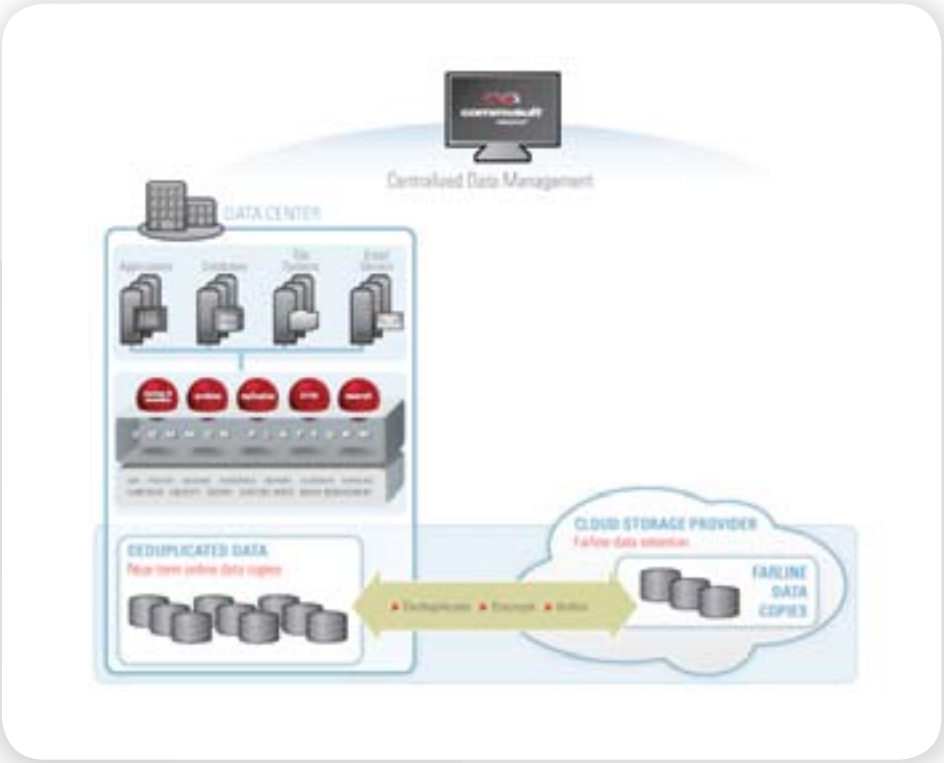
Increased Choice for Farline™ Storage

CommVault continues to work with an ecosystem of cloud storage service providers to maximize customer flexibility when it comes to choosing who to partner with for Farline storage. This means you can choose the storage provider that best meets your particular data retention, data recovery, and budgetary requirements. This open approach ensures that additional cloud-storage vendors will continue to be added in the future to increase the choices available with CommVault software's approach to cloud-based storage.

Leveraging CommVault and Cloud-based Farline™ Storage: Five Ways to Contain Costs, Maximize Flexibility and Retain Your Sanity

Tiering Data to Cloud-based Storage

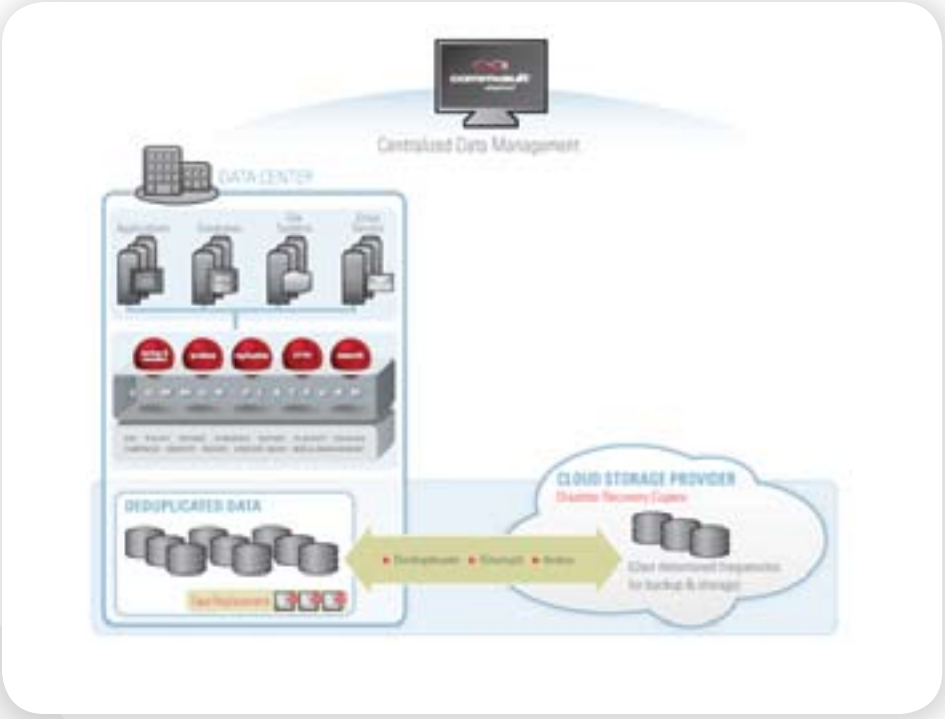
As data volumes continue to increase, many companies find themselves bumping up against the capacity, cooling or power limitations of their existing data centers. Meanwhile, government compliance and regulatory requirements mandate keeping ever-growing amounts of data, sometimes indefinitely. This 3-way balancing act between maintaining capacity, meeting compliance and addressing cost requires a flexible multi-tier approach to data and information management that extends to cloud-based Farline storage.



Cloud-based Disaster Recovery

Many small-and-medium sized enterprises often have the same data protection and recovery requirements for their data that larger enterprises have for theirs. However, they often lack the personnel, facilities, systems and budgets needed for disaster recovery.

Farline storage now enables them to protect their mission-critical data by securely storing it offsite, protecting them disruptive manmade and natural events. CommVault software enables them to configure how frequently they want to copy data to the cloud, and to quickly and easily retrieve it back to any alternate CommVault software-managed location.

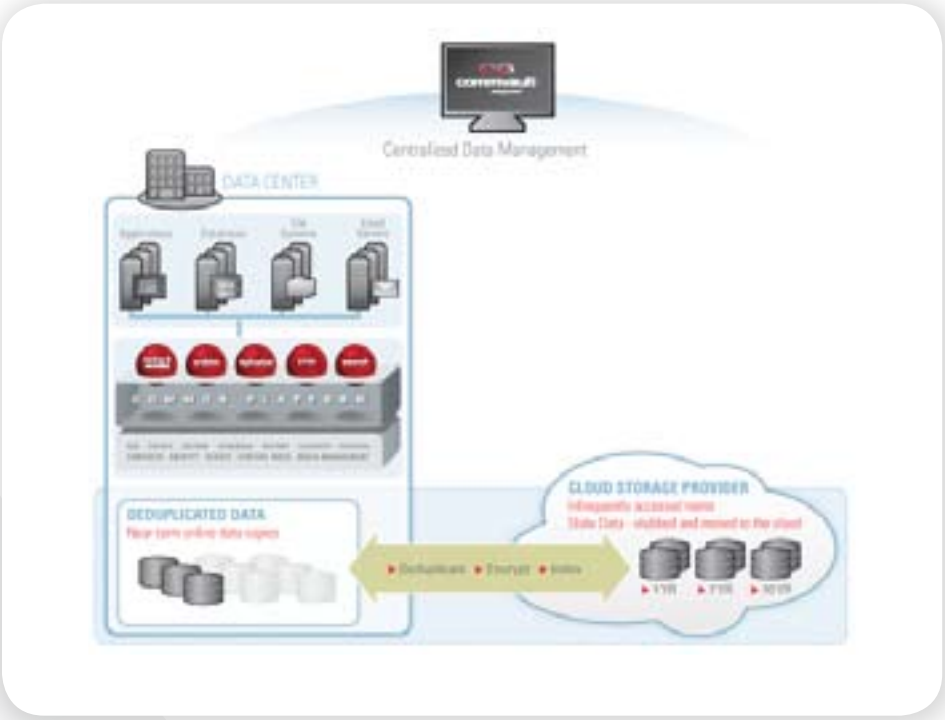


Archiving Stale Data to Farline™ Storage in the Cloud

Stale data can be found throughout your enterprise; from your central data center to remote office; and the workstation to the laptops of end-users worldwide. By some estimates, almost 50 percent of unstructured data is stale, meaning it hasn't been accessed in over a year.

While Simpana software enables you to take advantage of Farline storage for storing stale data, it unlocks benefits that range across different storage tiers:

- Archiving stale data frees up valuable capacity on your production storage tiers, allowing you to extend the useful life of your existing storage, and minimize the need for incremental purchases.
- Reduce the size of your backup datasets and backup windows by eliminating the need to backup infrequently accessed data.

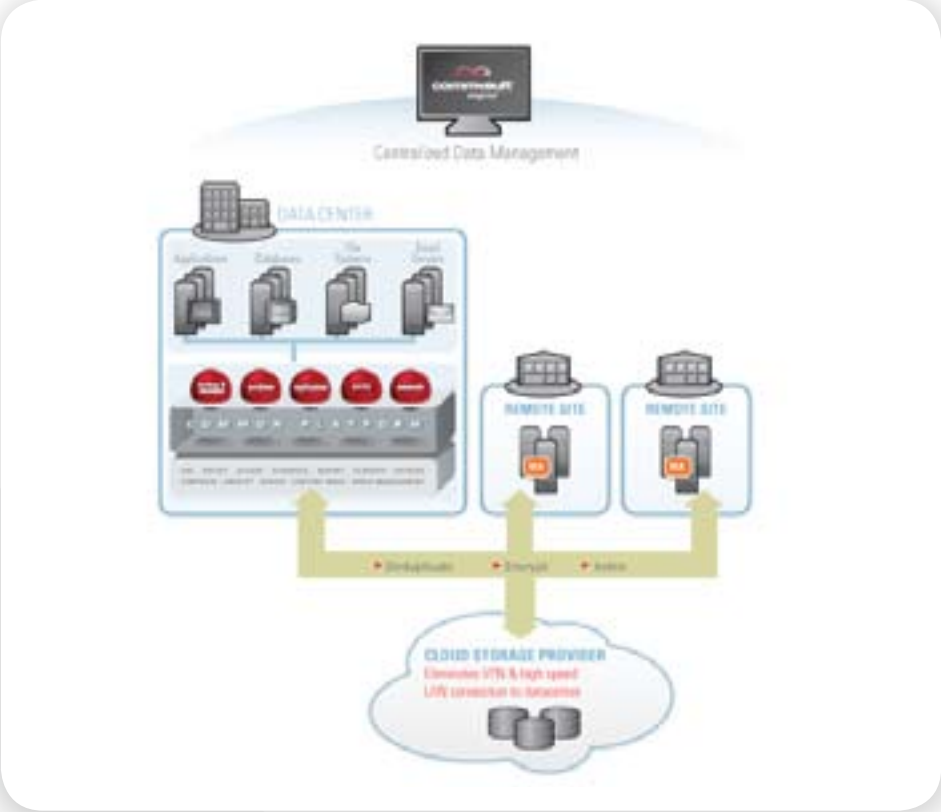


Protecting Remote-Office Data

CommVault Simpana software makes it easier to protect data that resides outside of the Data Center by storing it directly in the cloud. This enables the central IT team to control the movement and management of remote data along with defining the appropriate data retention and recovery policies.

Data from remote offices can be backed up directly to cloud-based storage eliminating the need to migrate the data to the Data Center first, and thereafter migrating the data to the cloud. As data ages past your retention guidelines it can be automatically deleted in the cloud, creating ongoing savings in capacity utilization charges.

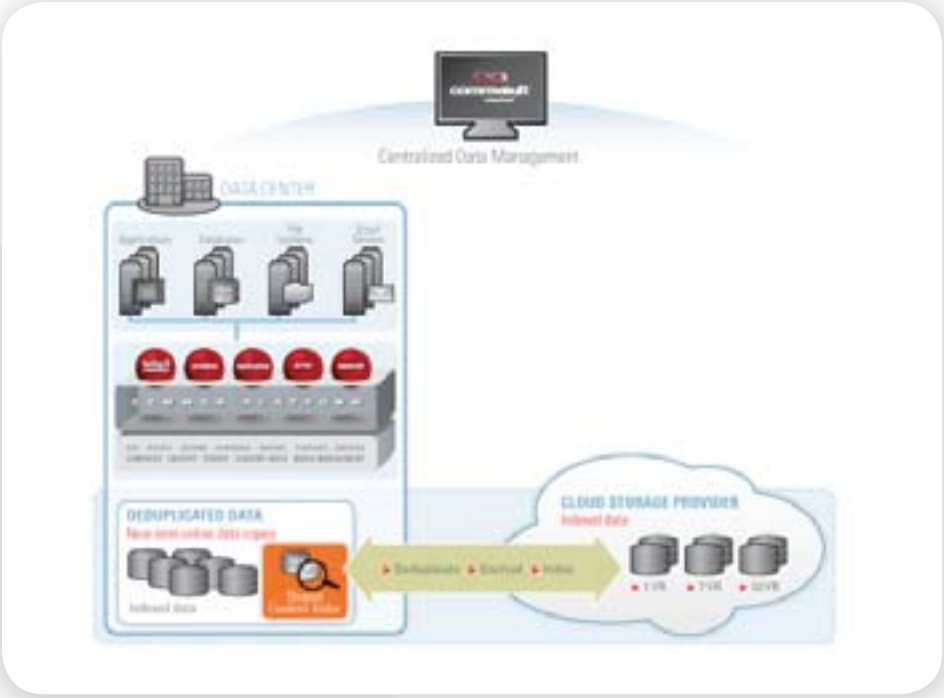
Because the data is managed just the same as if it were stored in the core Data Center, the Simpana Storage Resource Management (SRM) module can be easily used to monitor and analyze data across the enterprise regardless of whether it stored in the cloud, in the Data Center or in remote offices.



Search/eDiscovery

As data and information management expands beyond your physical infrastructure into the cloud, your legal and reporting requirements continue to grow as well. Simpana software offers four key benefits for Search/eDiscovery.

- First, indices of all data retained can be kept on-premises. This enables you to retain control of the most critical and sensitive aspects of information management, and ensures that content indexes are accessible only to designated personnel within your organization.
- Next, since the indices are searchable locally, there is no latency with regards to data that may be retained in the cloud over a number of years or even decades. This reduces the amount of time and data required by your legal and/or IT teams.
- Third, only the specific data required for eDiscovery requests is restored back from the cloud. This reduces your network load, the time needed for data restore, and minimizes the data retrieval costs charged by your cloud-storage vendor.
- Finally, global indexing of all relevant data, from the Data Center to remote sites, mobile users, and cloud-based data. This ensures that you not only have a global view of all your data, but that you also avoid the legal and financial risks associated with incomplete responses to eDiscovery requests.



Key Takeaways/Considerations:

Cloud-based data storage represents the "fourth wave" for storage technology, following on the heels of Direct-Attached-Storage (DAS), Network Attached Storage (NAS), and Storage Area Networks (SAN). As the marketplace continues to evolve, we find that different customers leverage cloud-based storage for different reasons. Some decide that they don't want to add to their Data Center footprint due to high fixed-cost or ongoing operational expenses. Others want to avoid the complexity of efficiently managing their storage as their datasets continue to grow in size. Regardless of organizational size and motivation, there are still four key aspects to consider when moving to cloud-based Farline storage:

Security

Your data represents the lifeblood of your business and losing data, especially to theft, is one of the fastest ways to lose the faith that customers have put in your company. Ask yourself whether your approach to cloud-based storage ensures data security when:

- Data is in transit, both to and from the cloud?
- Data is at-rest, even from service-provider personnel?

Portabillity

One of the key advantages of a cloud-based infrastructure is the ability to rapidly respond to shifting requirements, while driving down overall costs. So, can you:

- Easily move data back from the cloud if required?
- Migrate data if needed between cloud-based storage providers, to ensure the best selection of price and performance?

Accessibility

Getting fast reliable access to data is often critical to maximizing revenue opportunities, lowering support costs and increasing company productivity. When deciding on the right approach for your cloud-based storage approach, how will you:

- Restore your data back as fast as possible, directly from any physical or cloud-based storage tier?
- Configure your data management policies so that your most frequently accessed data is most easily and quickly retrieved when required?
- Match your network bandwidth capacities to your data's RTO (recovery time objective) requirements?

Manageability

Storing data in the cloud doesn't always reduce the operational requirements. In some cases, it can increase them, thereby negating the expected savings from cloud-based storage. When moving to take advantage of cloud-based computing, it important to evaluate your alternatives holistically by asking questions like:

- How do you intend to back up your data?
- Will you need to Archive data to the cloud, including setting up automated retention and deletion policies?
- Is global reporting for all data (physical and in-the-cloud) easily configurable or will it require a separate technology stack?
- Can you easily and securely extend your cloud-based data management to include Search/eDiscovery?

Determining Which Data Stays On-premises and Which Moves to the Cloud

As organizations evolve their data retention strategies, some will choose to keep their most valuable data "in-house" and move the rest to the cloud. So a key question is how do you determine which data is most valuable to your organization? One such framework is presented below as a suggestion. It uses a combination of factors including, (a) access requirements, (b) latency requirements, and (c) corporate requirements to help you identify your most valuable data.

Category	Example Questions	Impact
Access	■ How recently was the data accessed? ■ How often was the data required over the last 12 months? ■ How many end-users/applications required access to this data in the last 12 months?	Operational
Latency	■ How quickly will this data need to be restored? ■ What downstream applications/processing are dependent on the data?	Operational/Economic
Corporate	■ Is there certain data that need to be pulled in/put on Legal Hold for an eDiscovery request? ■ What data contains corporate trade secrets or IP? ■ What data might be considered highly sensitive, such as legal communication, or social security numbers?	Strategic/Legal

A Common Sense Approach to Cloud-based Storage

CommVault software provides an evolutionary approach to data and information management based on our unique Simpana software architecture and platform. We enable you to seamlessly leverage the cost-efficiency and flexibility of cloud-based storage without losing the ability to monitor, manage, protect, and secure your data, no matter where it resides.

Our approach to cloud-based storage enables you to seamless move data to and from the cloud using an automated policy-based approach. Integrated deduplication minimizes the amount of data in-transit and at-rest to reduce storage costs and shrink your data processing windows. Integrated encryption secures your data from unauthorized access with the ability to keep your encryption keys in-house.

Looking to take advantage of cloud-based storage? You're not alone. The Simpana software platform, combined with our expanding array of cloud storage provider partners, arms you with an extendable approach to Farline storage to help reduce storage costs, increase your organizational flexibility, and retain control of your data from cradle to grave.

About CommVault

A singular vision—a belief in a better way to address current and future data management needs—guides CommVault in the development of Singular Information Management® solutions for high-performance data protection, universal availability and simplified management of data on complex storage networks. CommVault's exclusive single-platform architecture gives companies unprecedented control over data growth, costs and risk. CommVault Simpana software was designed to work together seamlessly from the ground up, sharing a single code and common function set, to deliver superlative backup and recovery, archive, replication, search and resource management capabilities. More companies every day join those who have discovered the unparalleled efficiency, performance, reliability, and control only CommVault can offer. Information about CommVault is available at www.commvault.com. CommVault's corporate headquarters is located in Oceanport, New Jersey, in the United States.

commvault
solving forward®

www.commvault.com ▪ 888.746.3849 ▪ E-mail: info@commvault.com
CommVault Worldwide Headquarters ▪ 2 Crescent Place ▪ Oceanport, NJ 07757 ▪ 888-746-3849 ▪ Fax: 732-870-4525
CommVault Regional Offices: United States ▪ Europe ▪ Middle East & Africa ▪ Asia-Pacific ▪ Mexico & Latin America ▪ Canada ▪ India ▪ Oceania