



CloudBerry Lab
#1 Cross Platform Cloud Backup

Archiving to the Cloud: Cold Storage



Through the digital transformation of businesses – moving from bricks and mortar, 9-to-5 businesses to ones that are available anytime, anywhere – has required a shift in the way organizations protect their data. Traditional backup storage models of on-premises tape and disk have given way to both cloud and hybrid-cloud backup storage, providing greater availability and accessibility of secure, protected data.

This backup shift has occurred, not just because of new cloud-based technologies, but also because the digital transformation has caused a proliferation of what data should be backed up, how long it needs to be retained, and how quickly it needs to be recovered. Because of the demand for business continuity as the norm for every business to remain competitive, backups that were once reserved only for critical data, applications, and systems are now utilized to protect every facet of an environment.

The challenge of backing up “everything”

This need to protect the entire environment can actually put organizations at risk; blindly backing up “everything” puts a material drain on budget, personnel, and time – all increasing the cost of data protection.

A more mature data protection strategy includes the following considerations:

- **Not all data needs to be backed up at the same frequency** – Some data sets (even ones that are critical to operations) may not change or change infrequently, making regular backups impractical.
- **Not all data needs to be recovered within the same timeframe** – Critical tier 1 systems may need to be up and running within minutes, while other data isn’t necessarily needed for hours or days after a loss.
- **Some data requires longer retention** – Your backups focused on disaster recovery may only have a 30 or 60-day retention. But, depending on your organization’s industry, varying compliance mandates, HR requirements, and legal needs around maintaining archives of data, specific data sets may be required to be kept for months or years.

When put in this proper perspective, and add in the fact that backup storage in the cloud comes with a cost, it becomes clear that not all data needs to be available and recoverable in the same measure – and, therefore, the value that “standard” (or hot storage) provides isn’t always appropriate – especially when considering data to be archived.

So, how can you archive backups to cloud storage cost-effectively and efficiently?



Going Cold: Cloud Backups and Storage Lifecycle Management

One of the reasons cloud storage has become so popular is its cost-effectiveness when compared to on-premises storage. But, as even cloud storage is becoming a commodity, organizations like yours are looking for ways to further optimize their budgets – but do so in a way that still gives them required levels of recoverability and retention.

Public cloud environments like Amazon S3, Microsoft Azure, and Google Cloud Storage have all realized that a single tier of storage no longer meets the backup, retention, recovery, and archiving needs of organizations with mature data protection strategies. What is needed are less expensive storage mediums that can cost-effectively house data while meeting retention and retrieval requirements of critical data, archive data, and everything in between.

Enter in three common tiers of storage: *hot, cool, and cold*.

Hot Storage

This is the traditional high-availability, low-latency, frequently accessed storage most organizations associate with cloud backup and recovery. This tier of storage is optimized for performance and durability, and usually contains the backup sets of data, applications, and systems needed to recover your organization's operations.

Cool Storage

This tier of storage is used for data that is needed, but changes infrequently or never. Examples of data perfect for cool storage include scientific data, telemetry data, security camera footage, customer service call recordings, and any other kind of larger data set that rarely (if ever) changes. Typically, data put into cool storage is accessed less than once monthly, and must meet minimum retention requirements (e.g. 30 days) or face an early deletion charge.

Cold Storage

This tier of storage is for inactive data that is needed for archival purposes. As a rule of thumb, the data sets that would sit in cold storage are ones your organization would expect to access less than once annually – email archives, or data retained for compliance, legal, or HR purposes. Cold storage is for data you may need in the future, but when you do need it, it's understood it may take some time to retrieve.



More Than Just “Cheaper” Storage

Cool and cold storage will always be less expensive than hot storage (for example, when this paper was written, Amazon pricing for their hot, cool, and cold storage options for the first 50 TB/month was \$.023, \$.0125, and \$.004/GB respectively). But, in order to achieve those lower price points there will be some tradeoffs:

- **Minimum Storage Duration** – Hot storage has no limits and is charged based on storage used within the billing period, but Cool and Cold storage normally have minimum durations for billing, so expect to leave your objects stored there for a minimum amount of time.
- **Lower Availability** – While seemingly small, Cool and Cold storage can offer lower SLAs than their Hot counterpart (e.g. 99% vs 99.95% respectively – a difference of about 3 and a half days of uptime a year)
- **Lower Durability** – Hot storage can be locally-, regionally-, and even geo-redundant (depending on the provider used). You should expect Cool and Cold storage to be limited in its redundancy.
- **Slower Access** – Hot storage, generally, provides millisecond-level response times. Cool and Cold storage are generally measured in sub-second response times, with some providers offering similar access times to that of their Hot storage offerings.
- **Slower Retrieval** – When in the middle of a recovery scenario, the amount of time it takes to obtain the needed data is critical. Cold storage is materially slower (e.g. measured in hours and not minutes) and is most definitely not appropriate for critical data and systems.
- **Higher Retrieval Costs** – When the time comes, and your organization needs data from Cool or Cold storage, the retrieval costs can be as much as double the cost of using Hot storage. Generally, retrieval of data from Cold storage can be accomplished using a number of retrieval options that materially speed up the retrieval process, but also further increase the cost of retrieval and, therefore, the use of Cold storage.

So, rather than simply looking online at the “price tag” of a Cool or Cold Storage offering, it’s important to understand the overall costs your organization will incur by using each storage tier – both at the point of backup and recovery – as well as determine if the performance required at the time of recovery is adequate to meet your recovery objectives. So, carefully consider the overall cost including amount stored, # of times/month data will be stored, # of retrivals, amount retrieved, # of transitions to other tiers of storage, and use of accelerated data transfers. In fact, most providers have extensive cost calculators available so you are keenly aware of the true cost of each tier of storage they offer.



Archiving to the Cloud: Cold Storage

The table below is a comparison of a few of the more prominent public cloud storage providers to demonstrate the differences in storage tiers.

	Hot	Cool	Cold
Amazon	Amazon S3	S3 Infrequent Access	Amazon Glacier
First Byte Latency	milliseconds	milliseconds	minutes or hours
Retention Minimum	None	30 days	90 days
Google	Cloud Storage	Nearline	Coldline
First Byte Latency	milliseconds	milliseconds	milliseconds
Retention Minimum	None	30 days	90 days
Microsoft	Azure Blob Hot Storage	Azure Blob Cool Storage	Azure Blob Archive
First Byte Latency	milliseconds	milliseconds	up to 15 hours
Retention Minimum	None	30 days	N/A

Controlling the Cold with Lifecycle Policies

All of the major cloud storage providers support lifecycle management of the data –using configured policies, data can be automatically moved from higher cost to lower cost tiers of storage based on age, as deleted based on expiration policies. For example, the Marketing department’s project files may sit in Hot storage for 180 days, and then be moved to Cool or Cold storage, as it’s unlikely they may need them (but desire to have backups accessible should they need to). That same backup containing the project files may be configured to expire after 2 years, at which time they it is deleted.

Proper Use of Cold Storage: Archiving

It's evident by the various storage tiers above that the providers have very different intentions for what types of data should be stored in each. Cold (and Cool) storage is meant to be a cost-effective alternative to store data that will be accessed less frequently, and has more lenient recovery time requirements.

In other words: *it's an archive.*

Any data that falls under HR, legal, or compliance requirements to be retained for an extended period of time normally needs to be archived - this can be anything from email, to file data, to application-specific data. Some backup solutions allow you to specify various retention times for a given backup data set. So, should you simply create a backup definition, configure it to be an archive, and directly backup to Cold Storage? You can, but you're creating more work for yourself.

Instead, leverage cold storage as part of an overall strategy that utilizes storage lifecycle management to automatically move specific backup data sets from Hot to Cool to Cold, as is appropriate - based on the business need for both recovery and retention of a given data set. That way, you can focus on protecting the business today, next month, and next year - all without having to create additional backup sets for each.

Staying Out In the Cold

Organizations no longer have simple retention requirements around the various types of data generated daily. This means backups - while certainly supporting lengthy retention times - must be looked at more holistically by including the backend storage tiers to attain a level of cost-effectiveness and efficiency at the time of backup, while at rest (over what can be years of time), and during a recovery event.

By leveraging Cold Storage as part of storage lifecycle management, organizations can focus not just on protecting their data, applications, and systems through the backup process, but ensure data is appropriately retained balancing IT budgets and IT performance.