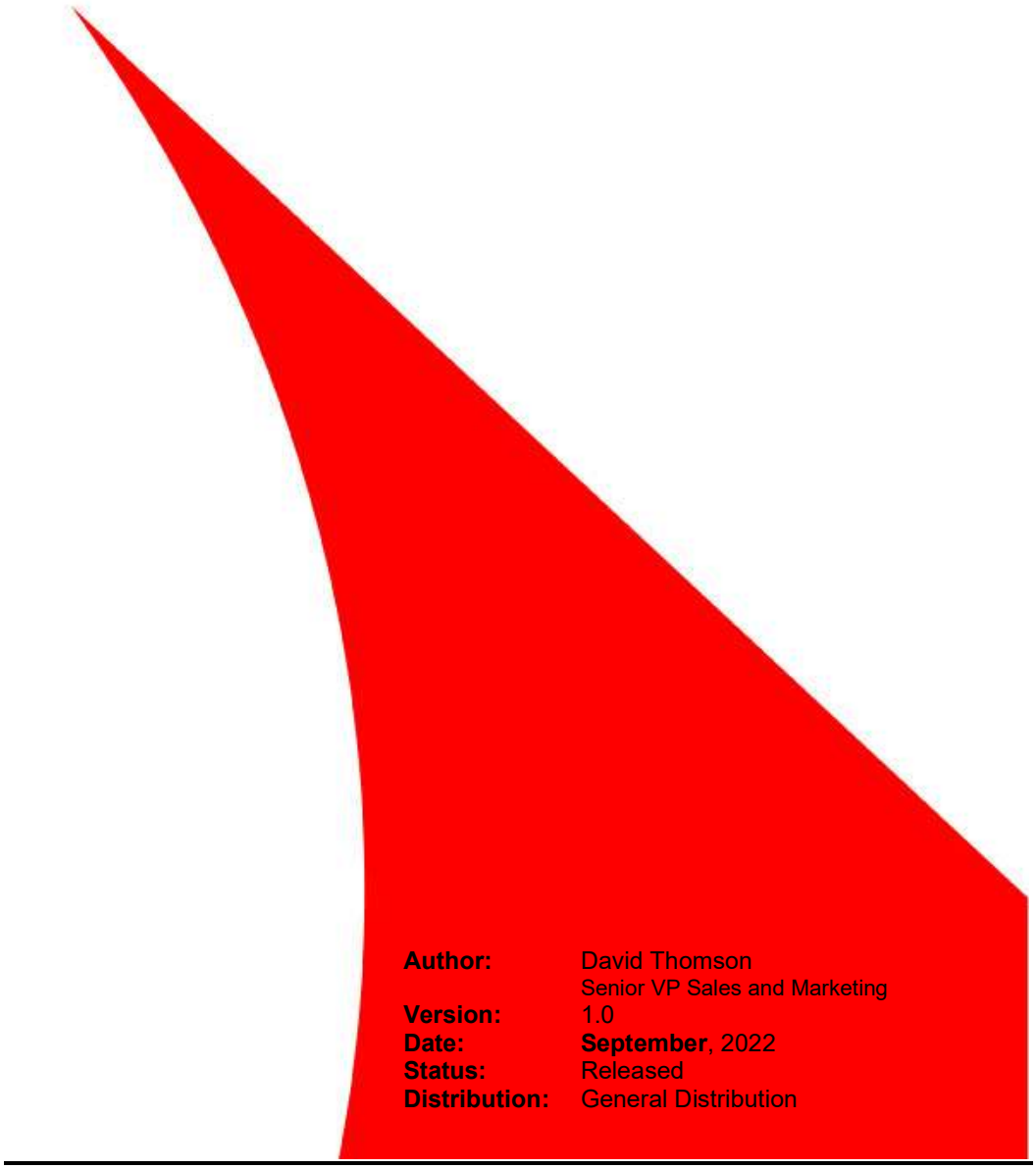
Why is Cloud Storage Relevant for your Archive Today?



Author: David Thomson
Senior VP Sales and Marketing
Version: 1.0
Date: **September**, 2022
Status: Released
Distribution: General Distribution

www.qstar.com

Introduction

It seems like Cloud Storage has been available forever, when in fact Amazon Web Services introduced AWS S3 in 2006. Since that time, we have seen a rapid increase in Cloud Storage offerings and organizations who offer and also use Cloud Storage.

Cloud Storage initially started as "online" or readily available content stored on disk-based systems. Perhaps the largest, AWS S3 will retrieve most content in under a second and so typically can be used as general storage for user or group content.

AWS understood that there are other use cases for data that does not need to be "instantly" accessible and in 2012 added Glacier (now called S3 Glacier Flexible Retrieval) as a new category of Cloud Storage. Designed specifically for archive and possibly backup data too, AWS reduced the cost of storage BUT increased the time taken to retrieve content to minutes or multiple hours. Finally, S3 Glacier Deep Archive, allows for lowest cost storage in the cloud with data retrieval within twelve hours. Google and Azure Clouds offer similar solutions as part of their "tiers" of Cloud approaches. Again, costs decline as time to data increases.

Any data stored to the Cloud does not need the organization to make additional copies as it is the responsibility of the Cloud provider to make copies (or backup data) and ensure data is always available. Cloud storage also means data is stored offsite and away from the main data center, which also secures data against local disasters. Naturally, many organizations still make their own copies as data loss can also mean the end of the organization.

Cloud Storage also provides a simpler method to connect, store and retrieve data for organizations that have a dispersed workforce. Traditional storage is accessed using file systems and remote users had to be allowed through firewalls to access to data. Cloud Storage relies on passwords and encryption to ensure any valid user has access to an organization's data, no matter where they reside.

Like anything, Cloud Storage has its issues, let's cover those downsides first.

Why not Cloud Storage?

Connectivity through the internet is essential for Cloud Storage. Without a good connection, Cloud Storage becomes unusable. What dictates "good" depends on individual businesses but typically requires a dedicated high-speed connection, so there are no peaks and troughs of connectivity speed. Connectivity resilience is also extremely important. There is no point paying for high-speed connection, if that connection repeatedly fails.

Cloud storage can be lower or higher cost that traditional storage depending on how it is architected and used. Using the wrong Cloud "tier" with the wrong data can end up costing organizations many times more than providing the storage themselves. Besides the cost of storing data (cost per TB per month), Cloud Storage providers typically also charge for uploading and / or downloading content (egress fees). Downloading, in particular, can be unpredictable and create unforeseen costs which occur only after the download. IT budgets can be compromised through poor decision making by lower-level employees.

Data privacy becomes more difficult to predict and some organizations do not allow data to be stored in some clouds or cloud regions. Data stored in some regions can have different implications on an organization based on the local data privacy laws of that country / region.

For many organizations access to Cloud storage is "open" in that many employees or applications can read or write data to Cloud. Typically, data written to Cloud is not protected through retention management or WORM (write once read many) techniques, which leaves data open to attack (deletion or modification) by third parties who find ways to circumvent data security initiatives.

Finally, for any organization with large capacities of data with one Cloud storage provider, changing vendors becomes very difficult. When buying next generation storage, organizations will replace and migrate data from old to new, which could be latest generation technology (disk to flash or NAS to scale-out NAS etc.) or vendor (better pricing, better support etc). Trying to migrate data at the petabyte level out of Cloud either to another Cloud storage provider or to next generation technology, becomes very time consuming and difficult to manage. The larger the capacities stored the larger the problem. Google – for example has a method for faster uploads of bulk data using a physical appliance that is filled and then shipped from an organization to Google's datacenter. They state, a 1 PB data transfer can be completed in just over 40 days using the (Google) Transfer Appliance, as compared to the three years it would take to complete an online data transfer over a typical (dedicated external) network (100 Mbps).

Finally, some (many?) cloud providers are working at a loss to encourage organizations to move data to their cloud in the hope that through economies of scale and decreasing cost of storage hardware, they will at some point in the future start to make profits. The industry has already seen some cloud providers closing their businesses and leaving organizations with little time to pull content out and store it elsewhere. Nirvanix closed in 2013 giving their users 2 weeks' notice of its closure after previously raising $70 million in venture capital.

So why Cloud Storage?

There are several key reasons Cloud Storage has become so popular.

A key benefit is only paying for capacity used. When buying traditional storage, you have to predict capacity requirements for multiple years in the future and so pay for capacity you are not using yet. Cloud storage provides "capacity on demand", you only pay for what you are using during that period. Scalability is effectively unlimited, so capacity planning (forecasting data growth within an organization) can be avoided.

This method of payment is also very different and for many organizations has significant advantages. Data storage becomes a monthly cost (OPEX) rather than a one-off payment (CAPEX), which can help businesses cash flow. Traditional storage can be purchased via rental or leasing to own converting CAPEX to OPEX expenditure, but the process adds more complication than the cost per TB per month model that Cloud storage offers.

By its very nature, cloud storage is designed for remote access. Today's workforce is significantly more dispersed than ever before. In the past those users relied on VPN (Virtual Private Network) technology to allow them access to data protected behind an organizations firewall, but many VPNs are either difficult to use or not very secure, allowing an all or nothing approach to connectivity. Cloud storage uses passwords and access control systems to allow granular access to data based on preset credentials.

The cloud provider provides guarantees on data accessibility and data protection. It is their responsibility to ensure data stored in their cloud is replicated and / or secured using backup processes. An effective cloud has data centers in multiple regions throughout a country or block of countries (e.g. European Union) so that a failure at one site does not impact their clients ability to access data as needed.

As mentioned in the introduction, Cloud Storage is not a "one-size fits all" as many offer a series of choices based on speed of access. The intention here is to replicate "classes" of storage seen in many organizations data center, from NAS storage to long-term deep archive based on tape. Many also allow for policies to be created to move data between these offerings to reduce cost on content that is not being accessed. In addition, enhanced features such as "object locking" or immutability adds additional features (at additional cost) for truly archive type data that will never be edited or updated. This feature can also be used for backups – where their content is required to be preserved for set periods, but then deleted when new backups are taken.

Data is protected for as long as the organization continues to pay their monthly storage fees. This means that the Cloud provider is responsible for all storage renewals and migrations, such as when hardware fails or is upgraded. Data migration from old to new is one of the hardest IT functions and takes careful planning to execute, including reports on data migration without corruption. It is also possible that Cloud providers can continue to innovate and add new services using new technology, currently unavailable. In addition, the cost per TB per month fees can and have decreased over time due to lower cost of disk storage and increased competition in the cloud storage space.

How QStar can Help

For organizations that choose to use Cloud Storage, QStar Technologies Archive Manager can provide file services (SMB and / or NFS) adding a NAS-like option for file-based archiving. Some Cloud Storage companies offer a very basic file archive option, with limited features. Archive Manager can create a secure archive environment using retention management options and roll-back file system features to help in the event of a ransomware attack or accidental or malicious deletes or overwrites. Other options such as encryption and compression are also available. For organizations that wish to secure their data even further and potentially save significant costs, QStar Archive Replicator allows content to be replicated to other sites and other technologies. For example, replicating data from Cloud storage to an on-prem Object Storage system or Tape library.

For more information see www.qstar.com or email sales@qstar.com.

Summary

Cloud Storage offers organizations many unique offerings over traditional storage technologies. For some organizations they are replacing NAS storage and others for active archive or deep archive functionality using OPEX rather than CAPEX to improve cash flow and provide unlimited growth capacities without the need for storage growth planning. Due to the rapid growth in this market, Cloud storage providers are undercutting each other by selling capacity under cost to gain more business, even if this is not sustainable for the long term.

Remote users or remote offices are automatically accepted as equals with employees working at the organization's headquarters or remotely, using standard Cloud API commands to store and retrieve content. Data security is provided by the Cloud provider along with data protection and long-term availability. Security conscious organizations, including government, military and critical services, may not be able to take advantage of highly scalable archive options due to the increased security risk.

For lower capacity users, there are few issues to worry about. For multiple PB users cloud storage can possibly present short-term issues based around unpredictability of costs because of access patterns and longer-term issues such as loss of control of data because of the length of time it would take (and cost) to migrate large capacities out of an existing cloud.

An increasing number of applications allow data to be stored directly to cloud using S3 compatible API commands, removing the need for costly file / NAS gateway solutions. However, gateways are still needed to create local stores of data or hybrid / multi-cloud solutions, that prevent "lock-in" by a Cloud Storage provider as the organization has options to delete and move on to new offerings without the costly (both in expense and time) migration process from one cloud offering to another.

Cloud Storage for your Archive Today	
Pros	**Cons**
Capacity on demand. Only pay for what you use today	Cost variability due to egress fees
OPEX – monthly costs	Security lower than on-prem
Unlimited capacity without planning	Data can be unmovable due to total capacity
Flexibility of offerings (access speeds) and many providers.	Requires fast, dedicated internet connections
Simple connectivity APIs connects to common applications	
Supports local and remote users equally	
Managed storage so no IT staff required to manage and protect	