

Cloud Storage

Deep Dive



Extending storage infrastructure into the cloud

Copyright © 2010 InfoWorld Media Group. All rights reserved.

Sponsored by





How to get the most from cloud-based storage

As organizations turn to the cloud to accommodate exponential data growth, dozens of providers offer services to match almost any mix of requirements

By David S. Linthicum

ACCORDING TO IDC, storage requirements grow from 50 to 100 percent annually at most global 2000 enterprises. Other analysts go even further. A study by Strategic Research Corp. predicted that enterprise storage demands will increase sevenfold over the next three years, while storage management resources will remain flat.

Until recently, the response to ballooning storage requirements has been to contact the usual storage technology vendor and buy yet another box for the data center. But eventually, to keep up, enterprises are going to have to turn to storage services in the cloud.

The long-term vision is to make cloud-based storage systems a seamless extension of on-premise storage systems, with the ability to move data reliably and securely between the two. This model is known as hybrid cloud computing, which mixes public cloud computing services with private, on-premise cloud systems. As it applies to storage, the hybrid cloud allows enterprises to incorporate public services as another tier in their overall storage strategy.

Unfortunately, workable hybrid architectures are not yet practical, in part because the protocols and standards to make cloud storage an extension of local storage are incomplete. Also, cloud storage costs are currently high when compared to those of traditional on-premise storage. But as the cloud matures, per-gigabyte costs will continue to drop and the technology will evolve into services that approach and sometimes surpass those of existing storage solutions.

Meanwhile, early adopters are turning to cloud providers for point solutions. Cloud storage can work as a kind of escape valve, providing temporary storage or an alternative to the capital expense of new hardware. Also, the cost and complexity of complying with

regulations such as HIPAA or SOX have compelled a number of enterprises to outsource the special handling of financial, medical, or other sensitive data to cloud storage services that guarantee compliance. And when building a business continuity strategy, a growing number of organizations are finding that cloud storage costs less than maintaining a warm site. Finally, some small businesses reluctant to invest in their own modern storage infrastructure are instead going “all in” and offloading all shared storage to the cloud.

Of course, there’s always a downside. With cloud storage, it’s all about trust: What about outages or the accidental destruction of data? What if the provider goes out of business? Moreover, your cloud storage service needs to know how to handle data covered by regulations, not to mention data security and privacy. And because many providers use proprietary approaches to storage, the portability of data may be an issue.

Each cloud storage provider offers a different mix of services. But there are some common patterns to this technology and it’s helpful to understand them before you determine whether or what type of cloud storage is right for you.

THE FUNDAMENTALS OF CLOUD STORAGE

Mission-critical storage is not to be confused with consumer-oriented backup services such as Mozy and Carbonite. Cloud storage providers that target IT offer many of the same storage and storage management capabilities as those developed by such premium storage vendors as EMC and NetApp. They also support advanced, well-known data center utilities such as SCP, FTP, SAMBA/CIFS, and RSYNC. They provide management consoles, deep encryption services, and the ability to support most compliance requirements around storage.



The primary advantage of leveraging cloud storage is on-demand expandability, which many call storage elasticity. Many IaaS (infrastructure as a service) providers that offer storage services sell cloud storage as “infinitely scalable.” While nothing is “infinitely scalable,” you pay only for what you use and can allocate resources to meet your storage requirements as you need them. This provides business agility as well, allowing IT to better adapt to changing market conditions.

DETERMINING COST AND UPTIME

Prices and terms vary for cloud storage. For example, at this writing, GoGrid allows you to store up to 10GB per month at no cost and charges 15 cents per gigabyte per month after that. On the other hand, Amazon S3 services charges 15 cents per gigabyte for the first 50TB, then 14 cents after that. Then there’s Google’s Cloud Storage Service, which charges 17 cents per gigabyte per month for simple storage, 10 cents per gigabyte for uploading data, and 15 to 30 cents per gigabyte for downloads.

Such permutations make cloud storage costs difficult to calculate and compare. You have to look at pricing on a case-by-case basis, considering that they are typically pennies apart, and adjust as your usage increases. So while Amazon S3 may be more cost effective than GoGrid when storing more than 50TB, GoGrid may be more cost effective when storing less.

You also need to consider data transfer costs. Either there is a cost to transfer the data to the cloud storage system over the Internet in one big upload, or cloud storage providers allow you to transfer data to them in more conventional ways, meaning tapes and disks sent via FedEx, which also incur service charges. Most enterprises will find that getting their data in the cloud costs thousands of dollars before the service is even up and running.

Another detail to consider is the SLA (service-level agreement), which also varies greatly from provider to provider. Most promise “five 9s” (99.999 percent) of uptime, with lower-cost services such as Amazon’s Reduced Redundancy Storage providing four nines (99.99 percent). Amazon’s S3 premium service actually offers ten nines (99.99999999 percent). The more nines, the more you pay. Many cloud storage services, such as those used primarily for client file backup, may offer no SLA whatsoever, but charge about \$5 per client.

A WALK THROUGH AMAZON S3

Amazon S3 seems to be setting the pace for cloud storage, with most cloud storage providers offering quick comparisons with S3, including either API or management console access to the storage services. Setting up and maintaining cloud storage is not as straightforward as many would imagine; it can be highly technical and involved, depending upon the storage provider and the use case.

So how do you get down to business with cloud storage? In the case of Amazon S3, you begin by creating a Bucket to store your data and deciding where that Bucket will physically reside -- an approach that allows you to reduce latency. Location is also important when considering compliance issues. For instance, in Europe, financial data cannot be transferred outside of the country of origin, or it must meet other regulations. S3 has physical locations in regions all over the world to provide its users with options that fulfill their requirements.

Once you set up a Bucket, you can upload Objects to the Bucket, where they can be accessed as if they were on a local mounted storage device. S3 Objects represent the data you store on S3, and they leverage a key (the number of the object) and a value (the data inside of the object). You can store an object on S3 using an API by specifying a key, its data, and the name of the bucket you want to put it in.

INTEGRATING CLOUD STORAGE

API access to storage services is a critical component. Many applications require access to storage services using an API that’s optimized for that specific storage system, either on-premise or cloud-based. The S3 API provides developers with SDKs for .Net and Java, along with libraries for additional platforms and languages. These interfaces typically use the Web services protocols REST and/or SOAP. At this writing, many cloud storage providers still don’t have public APIs, but most are moving in that direction.

Or you can simply drag and drop files from your local systems to your cloud storage provider and back again. All cloud storage providers offer a UI like this -- or they use native operating system tools to expose the remote storage systems as mounted file systems. If the cloud service can provide storage that looks like a logical drive, then you can typically use your own native file management tools.



To enable customers to integrate cloud storage with the data center, cloud storage providers need to support existing storage utilities. RSYNC, for instance, is a utility for Unix systems, which is able to synchronize files and directories between different storage locations. You should be able to manage your cloud storage with this utility as if it were a locally mounted drive, as well as use FTP and SCP (Secure Copy Protocol) for data transfer.

For cloud storage as an extension of the data center to succeed, those who currently manage the data center must also be able to manage their cloud storage systems using the same approaches and tools. Indeed, that should be a core criterion for selecting a cloud storage provider.

CLOUD STORAGE PROVIDERS AND USE CASES

The fastest-growing area of cloud computing is cloud storage. No wonder: Simple cloud storage services are fairly easy to build and launch. But the range of functionality is wide, and if you're in search of high availability or services for specialized needs, you'll want to turn to providers that have been around for a while and have a verifiable track record.

At the lowest level are the commodity services that offer a minimal SLA and a minimal price, such as Amazon's Reduced Redundancy Storage and Rackspace's Host Based Replication Services. This space will expand quickly and prices will fall as more providers move into the market.

At the other end of the spectrum, if you're looking for disaster recovery services, you'll find a number of providers optimized for that, with special features as part of their services. The brand names with major market share are familiar: Amazon, Rackspace, and GoGrid. Each has its own approach to disaster recovery, typically providing two levels of service: low and high availability (with low and high pricing).

On the high-availability side, you have offerings such as Enterprise Replication Services from Rackspace, which uses an enterprise-class, array-based, file-and-block-level replication technology. This solution supports high-end data replication services, including dedicated NAS and SAN devices, with replication services such as NetApp SnapMirror for NAS and EMC RPA technology for SANs.

Several third parties layer backup services onto other

cloud storage providers' offerings, adding new backup and recovery features. For example, Amanda Enterprise from Zamanda uses Amazon S3 to back up, archive, and retrieve data from anywhere on the Internet. In the future, we can expect third parties like this to add cutting-edge features, which will in turn be adopted by the major cloud storage providers.

Meanwhile, users need to understand the concerns surrounding audits, particularly such compliance issues as SOX or HIPAA. Several cloud storage providers offer a range of compliance services, from email archiving to special handling of electronic documents for vertical industries. Iron Mountain's cloud offerings, for example, address secure digital archiving, e-discovery services, and health information management.

Offloading data compliance to a service provider can be a cost-saving strategy. Do you really want your data center staffers worrying about the details of handling documents covered by complex regulations, or would you rather hand the whole mess to a provider? Obviously, trust in that provider is key, so it's essential for such sensitive material to be handled by a company with a proven track record. If your organization ever becomes entangled in litigation related to regulation, your own company, not your provider, is ultimately responsible.

CLOUD STORAGE STANDARDS

Missing from today's cloud storage are widely accepted standards, including standardized storage APIs -- the absence of which makes it quite difficult for customers to switch providers. Right now, once you've bound an application to a cloud storage service, you're pretty much stuck with that service unless you're willing to marshal the money and effort to move.

The Storage Networking Industry Association has released early drafts of a couple of documents created in the new Cloud Storage working group. The Cloud Storage Reference Model defines cloud storage elements, describing a logical view of their functions and capabilities. With this model as the foundation, the industry should be able to define the standards we need to move toward a standardized terminology for cloud storage services. The model also describes standard interfaces for cloud storage.

All that sounds good, but the day when such standards will have a real impact is a long way off. At this



point, what are the incentives for storage providers to foster portability? For the short term, you need to pick your cloud storage provider carefully, because you're likely to be married to it for a while.

Those charged with determining the value of this technology should consider opening trial accounts now

and testing the viability of cloud-base storage. As you calculate the effects of the ongoing explosion in storage requirements, you can no longer omit cloud-based storage from the mix. The agility that this technology provides to business is likely to be a key strategic advantage. 