Devising a Plan to Implement Hybrid Cloud Storage

Companies that jump into a hybrid cloud deployment too quickly can make some grave errors. Planning properly, as well as understanding the benefits and drawbacks of going hybrid, can prevent cloud catastrophe.
Hybrid Clouds: Complicated, but Worth It

Preparing to implement a hybrid cloud is a lot like planning a wedding. First, you have to make sure it’s what you really want; then, you need a checklist, and even then, one mistake can ruin the entire affair.

On your wedding day, that fatal misstep that could manifest itself as a cheesy DJ who—against your urging—insists on playing the “Chicken Dance” and “Achy Breaky Heart.” In a hybrid cloud storage deployment, a mission-critical error probably looks more like a lot of wasted time, effort and money. But unlike your line dance-riddled wedding, your boss probably won’t forget if you muck up the company’s hybrid cloud storage project. How do you guard against cloud failure when it seems like there’s the potential for mistakes at every corner?

Start by arming yourself with knowledge about the two most common slip-ups IT shops make when transitioning to a hybrid cloud storage environment: Companies sometimes fail to examine what kind of public cloud suits their needs or figure out what on-premises storage they should work with. Luckily, now you know better and you won’t fall victim to those big blunders. Next, determine how you’ll migrate data to the hybrid cloud. Will you do it manually or with a gateway? Last, consider your steps against our checklist to make sure you’ve ticked all the appropriate boxes.

With this three-part guide to transitioning to a hybrid cloud storage environment, you can be sure wedding your IT shop to the cloud will go smoothly. But, for rogue DJ wrangling, you’re on your own.

Margaret Jones
Site Editor, SearchCloudStorage.com
Avoid These Two Hybrid Cloud Mistakes

Hybrid cloud storage has become increasingly popular, but there are some potentially costly errors businesses should look out for when they implement a hybrid cloud.

Hybrid clouds can significantly reduce capital expenditures by shrinking the amount of on-premises storage hardware, software and infrastructure companies need. In turn, this reduces on-premises operating expenditures. A hybrid approach to cloud storage can significantly reduce the amount of data that is stored in the public cloud because it deduplicates and compresses the data on premises before migrating it to the cloud. Reducing the amount of data stored in a public cloud subsequently reduces monthly costs.

Unfortunately, there are two costly and aggravating mistakes that IT shops can make: choosing the wrong kind of cloud and using the wrong on-premises storage.

Picking the Wrong Public Cloud

The first common mistake is choosing the wrong type of public cloud storage. There are six types of public cloud storage:

1. Block storage, which is local embedded disk or SAN storage for applications in the cloud that require higher performance.
2. File or NAS storage is for applications that need NFS or SMB protocols.
3. Object storage used for active archiving.
4. Object storage used for cool archiving.
5. Object storage used for cold archiving.
6. Tape storage—typically LTFS—which is also for cold archiving.

Each type of cloud storage has distinctive performance characteristics and costs, and choosing the wrong type of cloud storage can have disastrous consequences. For example, block storage has the lowest latency and the
TOP TWO MISTAKES

EDITOR’S NOTE

AVOID THESE TWO HYBRID CLOUD MISTAKES

EVALUATE WHICH HYBRID CLOUD INTEGRATION APPROACH MAKES SENSE

PREPARATION FOR A HYBRID CLOUD IMPLEMENTATION

DEVISING A PLAN TO IMPLEMENT HYBRID CLOUD STORAGE

highest IOPS and throughput, but it also has the highest storage cost. It can cost as much as 30 times more than active or cool archive storage. Choosing block cloud storage when object cloud storage will do the job is a very costly mistake.

A similar cost issue can occur if a shop inappropriately selects cold archive cloud storage. Cold archive storage is affordable; usually less than 1 cent per gigabyte per month. But if users need access to the data in that cold archive, they may run into some problems. First, it takes a long time to retrieve the data from the cold archive. The first byte of data can take five hours to retrieve. In addition, there are transit fees: The cloud storage service provider charges customers for reading more than a very small percentage of data from the archive. These fees can be as much as 12 times the storage costs.

Avoiding this mistake requires accurately matching the characteristics of the data to where it will be stored. How frequently will users access the data? What are the performance requirements for reads? What are the data retention requirements? How much data will be kept on premises versus in public cloud storage? Answers to these questions also affect the second common mistake.

PICKING THE WRONG ON-PREMISES STORAGE

The second most common hybrid cloud storage mistake is selecting the wrong on-premises storage. There are four primary ways to deploy hybrid cloud storage systems:

1. Implement a primary NAS or SAN storage system that replicates snapshots or tiers data to the public storage cloud based on policy. When tiering, it leaves a stub locally that makes it appear that the public cloud storage data is still local.

2. Utilize a gateway or cloud integrated storage (CIS). The CIS looks like local NAS or SAN storage. It caches the data locally while it moves all or most data to the public cloud based on policies. It also leaves a stub that makes data in the public cloud appear to be local.

3. Install an on-premises object storage system that either provides the same de facto interface as public cloud storage or extends to it. When the
on-premises object storage utilizes the same interface as the public cloud storage, applications can write to either—or both—based on their requirements. When the on-premises object storage system treats the public storage cloud or clouds as an extension or remote target of the object store, it replicates data to the public cloud based on policy, similar to a NAS or SAN tiering storage to the cloud. If the public cloud uses the same object storage software, then it can become a geographic extension of the on-premises object storage.

4. Continue to use the current NAS or SAN storage system and utilize archiving or backup software that copies data based on policy to the public cloud. Archiving software can also delete local copies of the data based on policy.

Every one of these options has pros and cons and works best with different uses. Picking the wrong one can have severe consequences. CIS systems tend to be quite cost-effective, for example. Some public cloud storage service providers include them for zero or limited additional monthly cost, which can be a great deal. It can also be quite costly if the amount of data cached locally is less than what applications need. When that happens, the CIS constantly pulls data from the public cloud back to the on-premises storage. There is a large performance penalty from the internet and additional latency penalty for data rehydration. There is also a high likelihood companies will have to pay transit fees to the service provider for reading out the data from the public cloud.

Disaster recovery (DR) can be problematic for the CIS and tiering storage system options. Data in the public cloud cannot be read directly without reading it through the CIS or an on-premises cloud tiering storage system. That means a duplicate of the CIS or cloud tiering storage system must be made available in the cloud provider’s facility, or at the DR facility. Several CIS and tiering storage system providers now offer software variations that can run as a virtual machine in the cloud or DR provider’s facilities. Regardless, the additional hardware and software variations add to the cost.

Object storage can be one of the simpler integrations between on-premises storage and a public cloud; however, object storage is not
TOP TWO MISTAKES

known for high performance. To avoid excessive user complaints about on-premises performance, it is imperative to make sure the object storage system’s performance matches. Additionally, object storage systems use the standard Amazon Web Services Simple Storage Service interface, but not all S3 interfaces are the same. Many interfaces are a subset of S3. An application designed for the S3 interface must be certified to work with the “subset” that the on-premises object storage uses as well as the one in the public cloud. Otherwise, administrators should expect irritation, aggravation and stress. Troubleshooting this problem takes time, effort and labor.

Backing up or archiving to the cloud can be a significant cost saver, but it can also cause intense heartburn. Sending backups to the cloud is fairly simple, but recovering them may not be. Typically a backup requires a media server to recover and restore the data. Most hybrid clouds include one or more media servers on premises. That simplifies recoveries and restores on premises, and it makes them much faster than attempting to recover and restore from the public cloud. But when the data is recovered and restored in the cloud, it still requires a physical or virtual media server in the public cloud. If there is no media server in the cloud, then there are no recovery and restores in the cloud. In addition, if the recoveries and restores are coming from one of the variations of object storage archive in public cloud storage, do not expect fast recoveries.

Archiving to create a hybrid cloud is often complicated. The on-premises source storage and the public storage cloud are totally ignorant of each other. Applications and users probably need to know where their data currently resides to be able to access it. Some archiving software will leave a stub; however, links can break. Users might become annoyed or angry at not being able to find their data. Most archiving software can help locate data with admins’ help, but troubleshooting is often a time-consuming exercise.

Just like the public clouds, it is crucial to match the characteristics of the data stored on premises to the ability of the on-premises storage systems to meet them. Shops can avoid mistakes by spending time and effort upfront doing the groundwork first. —Marc Staimer
Evaluate Which Hybrid Cloud Integration Approach Makes Sense

If your organization is like most, chances are you’re deploying at least some of your applications and data in a public cloud. Depending on your specific usage scenarios, you may now be wondering whether your environment might benefit from hybrid cloud storage. By supplementing data center storage with cloud resources, you can ensure your production apps have ready access to the most important primary data while offloading less frequently accessed data to low-cost, elastic cloud storage. There are several integration approaches you can use to achieve the benefits of hybrid cloud storage.

To determine which hybrid cloud integration approach makes the most sense for your organization, you must consider case requirements, the organization’s appetite for change and innovation, and the IT expertise and resources you have in-house to plan and implement a hybrid deployment.

DIY Migration to Cloud Storage
You might choose a do-it-yourself approach to hybrid cloud integration that uses REST-based APIs—based on HTTP storage protocols—to enable both data center and cloud apps to directly access cloud-based object storage, such as Amazon Simple Storage Service or OpenStack Swift. To optimize data placement, you can create hybrid cloud apps that migrate data between on- and off-premises storage to push less-active primary data to the cloud. This approach is not for everyone because it requires that apps be built or rearchitected for the

There are several integration approaches you can use to achieve the benefits of hybrid cloud storage.
cloud, and it takes some effort from the development team. If you have the expertise and resources in-house, this method gives you total flexibility to tailor the hybrid environment to your specific needs and uses.

**CLOUD GATEWAY MIGRATION APPROACH**

Most organizations are more likely to benefit from an on-premises gateway that is designed to deliver local performance for production apps and enable cost-effective capacity scaling in the cloud. This approach works well for traditional apps that don’t make sense to rebuild for the cloud, and it does not require upfront development. Some gateway products are based on a freestanding physical or virtual appliance; others integrate the gateway functionality with a general-purpose storage system. These offerings tend to look like a traditional storage array to workloads. They usually provide a large, local cache for performance along with a gateway to cloud capacity. Most products support multiple public clouds, and, in many cases, they enable multiprotocol—object, block and file—support.

Cloud storage gateways enable a variety of uses, including backup to cloud object storage, failover and recovery, active or long-term archiving, file sync and share, and data stores for cloud-based analytics. They may even serve as a replacement for local primary storage in a remote or branch office deployment, with local caching for performance and automatic syncing back to the cloud. Ctera Networks is an example of a company with offerings that meet many of these cases.

**CLOUD GATEWAY ALTERNATIVE CHECKLIST**

If you are looking for an alternative to cloud gateways, the product should include network and capacity optimization, including deduplication and compression; data encryption in transit and at rest; integration with the on-premises authentication system, such as Active Directory or LDAP; capacity management; cost controls; and nonproprietary migration of data out of the public cloud.

Some gateway offerings include intelligent cloud tiering as a fully integrated capability. One example is Microsoft StorSimple, a...
cloud-integrated storage appliance that can automatically move and manage data across tiers. You can use such products to automatically archive inactive primary data to a cloud tier, and to store snapshots in the cloud. Automated tiering functionality eliminates the need for an off-site, secondary storage infrastructure, but still provides a cost-effective way to manage data over its lifecycle.

**CLOUD GATEWAY EVALUATION**

As you consider gateway offerings and functionality, ask vendors about their caching or tiering algorithms to ensure they can meet the performance needs of your specific uses. Look for products that store the current working set on premises on solid-state drives and actively, but transparently, push less frequently accessed data to a spinning disk tier, and ultimately to a tier in the cloud. More advanced platforms offer storage policies that let you better tailor the caching or tiering algorithms to your needs.

This is an exciting and fast-moving market, and new architectural approaches are constantly emerging. It is worth considering one or more software-defined storage offerings—such as Hedvig, Nexenta or Qumulo—to control and manage hybrid storage across the cloud and on-premises infrastructure. Other vendors are turning the traditional hybrid cloud infrastructure stack on its head. One vendor example is Velostrata, which allows users to move compute to the cloud while retaining primary storage on-site. This helps deliver cost-effective performance to production apps.

Given the popularity of the public cloud and the innovative ways to connect it to on-premises apps and infrastructure, the time for hybrid cloud storage has come. Take the time to evaluate the various hybrid cloud integration options and decide which approach best fits your requirements. —Jeff Byrne
Preparation for a Hybrid Cloud Implementation

Before embarking on a hybrid cloud storage deployment, you must consider the big picture. What do you want to accomplish beyond what you’re already doing on premises or in the public cloud? Is a hybrid cloud architecture a good fit for your organization’s planned uses? Are you familiar with other companies that have successfully deployed a hybrid cloud for a similar set of usage scenarios?

If you’re happy with the answers to these questions, then you’re ready to prepare your on-premises environment for a hybrid cloud implementation. Consider the following items as part of your preparation checklist:

1. Map out the primary uses. This will help you decide which scenarios to pursue initially in the hybrid cloud.

2. Select the specific apps you will run to support those uses. It’s important to understand whether the applications themselves are designed for a hybrid cloud implementation. For example, a modern disaster recovery-as-a-service (DRaaS) app will likely have been created with the hybrid cloud in mind. This can reduce, or at least simplify, the infrastructure-level integration you’ll need to do, at least for your cloud-ready apps.

3. Determine and document your requirements for key service-level agreement (SLA) metrics. This covers SLA areas such as storage performance, availability, security and quality of service, as well as network bandwidth and latency. Evaluate integration alternatives against those requirements as well.

4. Evaluate candidate public cloud providers for ease of hybrid cloud integration. Check out factors such as virtual private network connectivity, infrastructure-as-a-service (IaaS) offerings,
service and usage terms, and corresponding costs. Ensure the provider’s IaaS offerings are sufficient for your needs and that the expected benefits significantly outweigh the anticipated costs of your hybrid cloud deployment.

It’s important to understand whether the applications themselves are designed for a hybrid cloud implementation.

5. **Narrow your integration choice to one or two options.** The best hybrid cloud implementation choices might include purpose-built software, cloud-enabled apps with built-in integration capabilities or cloud storage gateway functionality.

6. **Take stock of your current infrastructure.** Decide what additional systems and software you’ll need to make your hybrid cloud a reality. For example, if you choose a gateway approach, you’ll need to invest in an on-premises appliance or gateway-enabled array, whereas a do-it-yourself integration plan might benefit from some type of software-defined storage layer. These are two very different integration methods, and the latter will likely require lower Capex but higher development and integration costs.

7. **Assess your storage management needs.** Determine how your current management tools and processes will need to adapt to the hybrid cloud project.

8. **Scope out the resources and skill sets needed to implement and maintain each integration option on your short list.** This will help you to decide which ones you’ll be able to accomplish and support within a given budget. If you move scenarios or storage capacity to the cloud, you’ll need to consider the effect on your current staff. For example, a cloud deployment may offload storage administration, freeing up your current staff for new tasks or responsibilities. In the case of larger deployments, you may need to reassign IT resources or invest in retraining.

9. **Consider your future growth and case requirements.** This will ensure your chosen hybrid cloud implementation approach will accommodate long-term needs. For example,
an application-driven integration scheme that works well for DRaaS might not be optimal to support primary storage.

Once you’ve made it through these checklist items, you should have a pretty good feel for what approach, or set of approaches, will best satisfy your needs. At this point, you should be ready to move forward with a hybrid cloud deployment to support your initial uses.

As you gain experience with your initial foray into the hybrid cloud, you can apply what you’ve learned from those first one or two cases to further improve and increase the value of your deployment. As hybrid cloud becomes more popular, keep an eye out for new functionality and services that can ease the integration and management of your environment and make it more effective. While hybrid cloud is far from a panacea, it will continue to prove its value for an expanding set of cases. —Jeff Byrne
JEFF BYRNE is a senior analyst and consultant with Taneja Group.

MARC STAIMER is the founder, senior analyst and CDS at Dragon Slayer Consulting in Beaverton, Ore. The consulting practice has focused on the areas of strategic planning, product development and market development.