NetApp

EBOOK

Top 10 Tips and Tricks for **Storage in the Cloud**

NetApp

Index page

NEXT ❯

Introduction

No matter the size, most companies that are thinking of getting started in the cloud today are doing it for this main reason: storage. Storage is the most basic service the cloud provides, giving you a place where your data can be safely kept that doesn't require you to expend the money and resources required to house, maintain, and service in an on-premises data center.

Storage in the cloud offers great benefits, but there are a few gotchas that might catch a company up along the way. These are the tips and tricks to keep in mind when using storage in the cloud.

« BACK

NEXT »

Different Storage Types for Different Reasons

The best tip that anyone can give you about storage in the cloud is to know what the types of storage are used for. When you first look at the services offered by AWS and Azure, it can be a confusing jumble of letters and concepts, so your first step is to get to know them a little better.

There are three types of storage to consider in the cloud:

Block storage
Looks at data in whole chunks that it organizes along different sectors.

Object storage
Treats the data in use as objects, and organizes it by metadata.

File system storage
Organizes data in a hierarchy.

On AWS, block storage is offered by Amazon Elastic Block Store (Amazon EBS). Amazon EBS is used as the root volumes for Amazon Elastic Compute Cloud (Amazon EC2) instances, with an Amazon EBS volume tied to each Amazon EC2 instance. On Azure, Azure Virtual Machines are used with Azure Disk Storage which work the same way.

Object storage is offered on AWS in Amazon Simple Storage Service (Amazon S3) and on Azure in Blob storage. Object storage stores unstructured data in one place— buckets in Amazon S3 and blobs in Azure. This storage format tends to be the least expensive of the storage formats the public cloud offers, as it is generally slower, making it perfect for long-term storage of data that isn't accessed much.

On AWS file storage comes as Amazon Elastic File System (Amazon EFS). Azure's complement offering for file storage is Azure File Storage. These shared storage types don't require the use of compute such as Amazon EC2 and Azure Virtual Machines.

Know Your Protocols: NFS, CIFS, iSCSI

Protocols are how you connect to your storage, and there are different types of protocols that have different benefits. Knowing which to use for your workload is important.

Two very common file-sharing protocols are:

Network File Systems (NFS)
Used for Unix and Linux operating systems

Common Internet File System (CIFS)
a Microsoft-developed protocol based on the Server Message Block protocol (SMB).

There is also the Internet Small Computer Systems Interface, or iSCSI. Unlike CIFS or NFS, with iSCSI the client host can change blocks of data individually as opposed to having to work through the folder and file system used by CIFS and NFS. With workloads such as databases, this helps provide a higher level of performance.

The key to finding the right protocol for your business is to determine your own business priorities. Do you need the high performance of iSCSI or can you suffice with CIFS/NFS?

Reducing Footprint

Storage in the cloud offers one thing that on-premises storage has never been able to do: **limit the amount of resources it takes to house and maintain the physical storage environment itself.** All of those infrastructure costs and tasks are now the exclusive responsibility of the cloud provider. The result is that storage budgets that were once Capex costs for companies can turn into Opex costs.

The physical footprint is therefore dealt with, but what about the footprint of the data itself? With all of the required maintenance, housing, cooling, power, and other associated costs now shunted over to the cloud provider, how can companies using the cloud manage to reduce what they are still paying for?

Finding a way to cut down on how much storage you use isn't going to be easy if you look to the cloud service provider you're using for solutions. They aren't going out of their way to make sure you pay less every month. You need to use the kind of space-efficiency features that can make sure that only the minimal amount of storage is ever being used. In conjunction, NetApp's well-known data deduplication, data compression, and thin provisioning storage efficiencies can easily reduce cloud storage footprint by 50%, and in some cases, depending on the data type, by 70% or even higher.

Easily reduce cloud storage footprint by 50%, and in some cases, depending on the data type, by 70% or even higher.

BACK

NEXT

All Aboard: Cloud Onboarding

Getting to the cloud is the biggest challenge to an existing system. Finding all of your storage resources and bringing them to the new environment requires an effective transfer method, one that will keep in mind the amount of data you have to move and the costs and security of transporting it.

There are many ways to get data into the cloud, from the native tools provided by AWS and Azure, to open-source tools and third-party software. While every migration has multiple moving parts, data file migration is one of the more important. You'll want to make sure that through the migration, data is protected in transit, secure when it arrives in place, and able to be synced during the migration process and for future changes. This data movement means a real investment of both time and money.

There are some companies that need to move so much data in a cloud onboarding that it would take decades to move it all over the internet. In those cases, other options are available, such as AWS Snowmobile, which allows your data to be moved to a storage device located within a tractor trailer that can then transport it physically to the new repository. For other deployments, other tools can be used. For instance, existing NetApp storage users can take advantage of SnapMirror® or Cloud Sync technologies to seamlessly move their data to the cloud environment when using NetApp's Cloud Volumes.

Avoid Cost Sprawl

Lowering costs is the biggest draw that the cloud has, but what if you aren't paying attention to the way that you are using cloud resources? You may wind up paying for a lot of unnecessary services and resources—the situation known as cost sprawl.

Cost sprawl isn't new to storage experts. It's a problem that also plagued the on-prem data center. However, while the cloud makes it so easy to provision resources with just the click of a button, it's also a lot easier both to provision more resources than are necessary and to lose track of them.

Workloads can get abandoned, you may not be taking advantage of data tiering, and data may exist in duplicate, wastefully consuming more storage than you should have to be paying to maintain. Find a way to keep track of your storage resources to make sure they are working optimally and cost-effectively.

❮ BACK

NEXT ❯

Don’t Always Pay for Premium Storage: Tier It Instead

The biggest gotcha in the public cloud is that customers frequently pay too much for the type of storage they are using, when a cheaper format would suffice for their data usage needs. Data that is in use on a rolling basis is best used with systems like Amazon EBS and Azure Disks, but storage that isn’t going to be required as much can be tiered to the inexpensive object storage on Amazon S3 and Azure Blob Storage.

Both Amazon S3 and Azure Blob Storage are themselves tiered into several, increasingly slow and increasingly inexpensive storage levels. For data that may still need access on a regular basis, Amazon S3 Standard and the Azure Blob storage Hot Access tier are available. For more infrequently-needed data there is Amazon S3 Infrequent Access and Azure Blob Storage’s Cool Access Tier. For data that will need long-term warehousing that is hardly expected to be accessed at all, there is Amazon Glacier and Azure Archive.

Know which data you need to use the most and spend the right amount of money to store it, or else, you’ll pay a lot more than you should. Data tiering can easily save you as much as 10 times the cost of storing infrequently-used data on performant disks.

Protect Your Data, Protect Yourself

Data loss is always a threat. It could occur due to a user making an accidental deletion or modification, a malware attack, a natural disaster, or some kind of disk failure or outage. Even updates can cause a program to fail and harm your data. No matter what the cause, if that happens it means there will be a negative impact on the reliability of your business operations and your bottom line.

An important point to remember here is that it’s not just your data that you’ve decided to house in the cloud—it’s data that also belongs to all of your customers and users. In many cases there are compliance laws that require companies to guarantee the protection and availability of sensitive data. Companies need to put in place effective backup, snapshot, and disaster recovery solutions in order to ensure that data is safe no matter what kind of disaster is threatening.

Key to data protection is backup creation. Without a store of files that you can rely on to repair any loss of data in the primary environment you won’t be able to recover should something go wrong. Since many high-performance workloads will need to be backed up at an extremely regular basis, it’s highly important to configure a cost-effective and seamless way to backup your data. It’s also key to find a way to failover and restore your system in disasters so that users can have their data and use their data even as crisis events are unfolding. Backups need to be created regularly and automatically, copying data to and from on-prem storage, the cloud, and between both.

❮ BACK

NEXT ❯

The Old and the New: Hybrid Storage Environments

Another consideration that companies moving towards the cloud have is to set up a hybrid storage environment instead of going all in with the cloud or attempting to create a costly private cloud. This middle way or phased approach use both the public cloud resources and the traditional storage systems that companies have depended on for decades.

This combination allows the scalable and flexible resources in the cloud to be leveraged without forsaking the control and hands-on experience afforded by keeping existing on-prem systems. Hybridity also allows for seasonal or momentary spikes in usage or traffic that would put too much stress on a strictly physical data center storage. However, with architectures such as this in place, finding a way to orchestrate the movement of data and the management of all the disparate resources is key.

Hybrid architectures can be useful for transitioning to fully cloud-based architectures, or until physical storage devices can be allocated to expand existing data centers. The trick is finding the unifying element between the various environments. In hybrid cloud environments, it is imperative for the management platform to be versatile enough to handle data—no matter where the data is stored. In addition, complications can arise, when trying to manage data stored in a multicloud environment, such as using both AWS and Azure with an on-premises storage system. With the use of multiple storage systems, resources can also get out of sync or become lost through the process of deletion and continue to rack up charges. Finding a unified tool to manage and monitor these disparate storage systems ensures data synchronization and cut down their associated costs are the main challenges of unified storage management system.

Finding a unified tool to manage and monitor these disparate storage systems ensures data synchronization and cut down their associated costs are the main challenges of unified storage management system.

BACK

NEXT

Avoid Vendor Lock-In

Since a public cloud provider such as AWS offers on-demand usage and many companies offer services based on that usage, the cloud vendor can be seen as the business' silent partner. For every dollar they get from their customers, the public cloud provider gets a share for contributing to their infrastructure. This can lead to a situation where the cloud provider gains so much of a foothold of the company's business structure that the company can find its choices limited. In theory, a cloud provider could double their prices or change their policies overnight: such moves would have a direct impact on the client company's revenue. This is what is called vendor lock-in.

When it comes to vendor lock-in, you do have options. Just knowing that there is another cloud provider that you can use, can change your company positioning and approach in the cloud. Another option is to opt for a multicloud architecture.

Multicloud architectures protect the company from cloud cost overages. Multicloud deployments allow you to avoid getting locked in by one cloud service provider's rates and SLAs by keeping data housed with a second cloud service provider, giving you more flexibility and a wider range of pricing options.

However, while multicloud deployments can help avoid vendor lock-in, multicloud deployments also require a way to structure interoperability across the different infrastructures, so working in it is seamless for all your company's teams. When all storage operations are under a single umbrella, for example, it keeps access and management of the data consolidated. If you're going to try to structure a multicloud deployment, you'll need the governance tools that can control all of the storage in those disparate environments.

❮ BACK

NEXT ❯

Accessibility

While the cloud makes it easier than ever before to share your data, it's a scary thought to have all of your data accessible in such a way. Companies need to strike a balance between the amount of accessibility they allow for their data and their ability to control it. Luckily there are a number of tools and services that you can use to restrict, grant access, and even clone your data to limit and/or share it widely throughout your business.

AWS offers IAM policies and Azure relies on Azure Active Directory to make sure you can decide who can access your cloud deployment with secure control access. With these tools, different teams can be set up with different levels of access, depending on their respective duties. Also, security settings for different roles within your organization can be set using AWS Organizations and Azure Subscription and Service Management and Azure RBAC. For extra precaution, a single sign-in access with a range of verification methods can be set up using AWS Multi-Factor Authentication and Azure's Multi-Factor Authentication.

Companies need to strike a balance between the amount of accessibility they allow for their data and their ability to control it.

Conclusion

Storage in the cloud is a lot more complicated than it looks, but with a little help, most organizations can find ways to take advantage of it and start to reap the benefits. Now that you know what to look out for when you're using storage out in the cloud, keep in mind that NetApp cloud solutions offer great ways for you to leverage existing storage resources as you move into the cloud and begin to take advantage of all it has to offer. For each of these cloud tips, NetApp's Cloud Volumes offers an easy way for any enterprise to get more out of the cloud. With two service models—Cloud Volumes Service and Cloud Volumes ONTAP—you can decide how much control you want over how you deploy in the cloud.

Cloud Volumes Service is the new fully-managed file storage SaaS offering from NetApp, aimed specifically at high performance workloads and born-in-the-cloud enterprises. Offering multiprotocol file services support, Cloud Volumes Service can be used not only with AWS and Azure, but also with Google Cloud Platform. But for those businesses that want to retail full control of their storage and data management, Cloud Volumes ONTAP (formerly ONTAP Cloud) gives you the NetApp ONTAP experience built on AWS storage or Azure storage. For existing NetApp users, it's the premium, enterprise-grade data management platform that you know, ready to be used in the cloud.

❮ BACK

NEXT ❯

Refer to the Interoperability Matrix Tool (IMT) on the NetApp Support site to validate that the exact product and feature versions described in this document are supported for your specific environment. The NetApp IMT defines the product components and versions that can be used to construct configurations that are supported by NetApp. Specific results depend on each customer's installation in accordance with published specifications.

NA-287-0218

BACK