Don’t Forget Data Protection When Selecting Cloud Providers

By George Crump

Storage Switzerland, LLC

The cloud is good at availability and data durability. If there is an outage cloud providers have an excellent track record of getting their services back online quickly and providing access to the latest copy of data.

What if though, the organization needs access to a previous copy of data?

Availability services are real-time, but previous versions are required to recover from cyber-attacks, rogue users and data corruption from misbehaved applications. While most cloud providers provide snapshot capabilities, they typically lack the ability to automatically create stand-alone, point in time copies of data, which are needed for recovery from human caused disasters.

The data protection services available for the cloud provider's environment are a critical factor, among others, in selecting that provider. Unfortunately, it is often overlooked or assumed to be present. In this eBook we will walk IT planners through the cloud selection process while focusing in on data protection as a key requirement.

CHAPTER 1 | WHAT TO LOOK FOR IN CLOUD PROVIDERS

The major cloud providers have a lot in common. They all provide infrastructure (compute, network and storage) on a pay as you need basis. The “as you go” consumption model is popular for organizations looking to reduce data center CapEx and better match the way they deliver services to their customers.

Within cloud provider offerings there exist unique offerings like a specialization on machine learning or image recognition. Most organizations will be more interested in the core capabilities of these services. The operative word in selecting a cloud provider is flexibility. The service should adjust to the organization’s needs as those needs evolve.

FLEXIBLE MACHINE TYPES

The needs of an organization change over time. A key requirement is flexible machine types. As new workload requirements arise, the organizations will want to create different types of virtual machines as needed. Google Cloud for example, provides many different machine types including; standard virtual machines, high memory virtual machines, high CPU virtual machines, shared core virtual machines, micro bursting virtual machines, memory optimized virtual machines and custom virtual machines. Additionally, GPUs can be added to any of these virtual machine types, for additional processing power.

SHORT-LIVED VIRTUAL MACHINES

Cloud native applications are written to take advantage of all available and assigned CPU resources. If a large processing job comes in, the organization may choose to assign thousands of processors to the task to reach completion sooner.

Most cloud providers require a minimum "buy-time" of at least an hour. That means if those thousands of processor complete the job in 10 minutes, the organization has to pay for 50 minutes of idle time. Google Cloud provides very granular, per second, processing rental.

POINT-IN-TIME DATA PROTECTION

Another key area, and one that this eBook will focus on, is data protection. Most cloud providers lack basic data protection. They can't create periodic stand-alone copies of data. In situations where the organization is dealing with cyber-attacks, rogue users, human errors and data corruption from misbehaved applications, these point in time, stand-alone copies are required.

Most cloud providers replicate data between storage systems within the primary cloud data center and then replicate it again to a remote data center. The problem is that the replication happens almost instantly. Data corruption caused by a cyber-attack or application coding error is immediately replicated to all the other storage systems. There is no "air-gap" between copies.

Snapshots provide some protection, although they are typically taken too frequently. The problem with snapshots is these copies are totally dependent on the primary copy being accessible.

The other problem with most cloud snapshot technologies is their use isn't integrated into the application. Whenever they have a change application owners must manually execute snapshots, or develop scripts to execute snapshots. For a couple of VMs this may be easy, but as they expand and as the environment grows it becomes very hard to maintain. In addition, since snapshot storage costs consume capacity, (which increases the monthly cloud bill). To manage costs application owners need to be aware of how many snapshots they keep and also remember to delete those snapshots or move the copies to an alternate location and manage that bucket.

Organizations that are moving some or most of their workloads to the cloud need a data protection solution that is similar to the capabilities of their on-premises backup without the management overhead of it and something an applications admin can take care of.

Focus on Flexibility of Consumption and Price Competitiveness.

Organizations should focus on flexibility of consumption and of course price competitiveness. They shouldn't be wooed by the largest provider or the one that dominates the headlines. IT needs to be careful though, not to overlook the importance of point-in-time data protection as part of their consideration. In the data protection area, there is a surprising amount of differentiation between the various cloud providers.

In our next chapter, we'll discuss the differences between the protection mechanisms that cloud providers use compared to stand alone backup solutions.

CHAPTER 2 | DO YOU NEED TO BACKUP CLOUD-NATIVE APPLICATIONS?

A common question from organizations with cloud native applications is:

Do they need to back those applications up?

The cloud, after all provides plenty of redundancy and all major cloud providers have multiple data centers at their disposal. Most providers have very respectable uptime track records to make sure applications continue to run through almost any disaster.

What most providers lack however is meaningful point-in-time protection to protect against data corruption, ransomware and rogue users.

WHAT IS HIGH AVAILABILITY?

High Availability (HA) is the process of making sure that if there is hardware or site failure, the application is rapidly recovered and put back into production. HA is typically achieved by replicating production data in real-time to another location. The objective is to make sure the secondary data copies are as closely in sync to the production data sets as possible. In the event of a server failure, storage system failure or even a data center failure, HA will instantiate the impacted applications at another site returning them to operation so quickly that users may never even realize there's been an outage.

THE PROBLEM WITH HIGH AVAILABILITY

High Availability's challenge is the speed at which it updates auxiliary copies, in real or near real-time. The speed of updates means that if a ransomware attack occurs, data is encrypted and then replicated to the alternate storage targets in near-real time. So as fast as the attack occurs, potential backup copies are also contaminated. The same concern holds true for a rogue user. If a user gains access and deletes a file share for example, those deletions are also executed on remote storage.

WILL SNAPSHOTS SAVE YOU?

Most cloud providers offer a snapshot technology that enables them to create a point-in-time copy but these copies are totally dependent on the primary storage platform. Cloud snapshots are also typically difficult to execute, often requiring application level triggering.

Snapshots are also difficult to manage. If there is a need to recover from a ransomware attack, determining which snapshot contains the best known good copy is difficult. Snapshots are often under application and user control, which means that a malicious user, or simply an ignorant one, can remove snapshots without IT authorization.

There is also a cost concern with snapshots. The snapshots are stored on the same storage as production data and the more frequently the organizations triggers the snapshots and the longer they keep them, the more production storage is consumed. Snapshot data also follows the same data protection process as its primary counterpart, which means the capacity that snapshot consumes locally is protected locally and replicated to alternate locations.

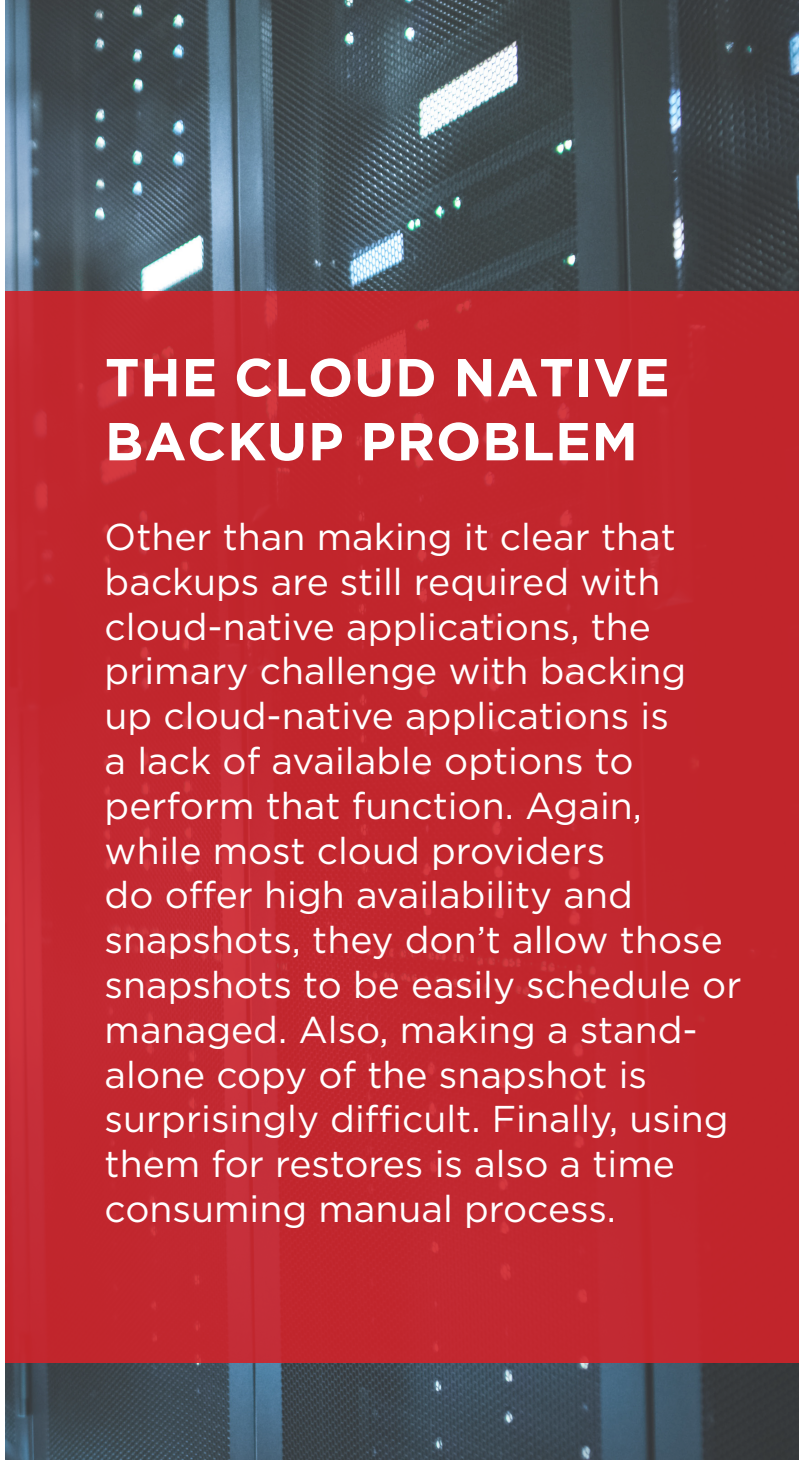
“Cloud snapshots are typically difficult to execute, often requiring application level triggering.”

WHAT IS A POINT-IN-TIME BACKUP?

A point-in-time backup is a standalone copy of data stored independently of production storage. If the volume containing production data is for some reason deleted then the point-in-time copy is still available but all data in snapshots is lost.

WHY BACKUPS SAVE YOU

As the name implies, a point-in-time backup is copied at a specific point in time and each copy is independent of production storage. Because of the independence of the copy, it can be made more difficult to access or it can even be marked read-only, making the backup copy more immune to a ransomware attack.



THE CLOUD NATIVE BACKUP PROBLEM

Other than making it clear that backups are still required with cloud-native applications, the primary challenge with backing up cloud-native applications is a lack of available options to perform that function. Again, while most cloud providers do offer high availability and snapshots, they don't allow those snapshots to be easily schedule or managed. Also, making a stand-alone copy of the snapshot is surprisingly difficult. Finally, using them for restores is also a time consuming manual process.

FULLY PROTECTING THE CLOUD

The cloud's natural HA capabilities are certainly important but the capability is mostly to protect the cloud provider from large scale disasters. Cloud HA doesn't typically protect user accounts from data corruption situations. Organizations with applications running natively in the cloud need to change their data protection orientation, they need to be less concerned about disaster recovery and more concerned about point-in-time protection. Organizations need to make it easy for application owners to recover from human errors that are typically not protected by a system's high availability. This is why IT needs to make point-in-time backup a priority for cloud-native applications.

CHAPTER 3 | WHAT TO LOOK FOR IN CLOUD NATIVE BACKUP

Users and software vendors realize the importance of protecting cloud-native applications. While cloud providers generally do an excellent job of protecting applications from a disaster they don't provide easy access to tools to provide point-in-time backups which are necessary to protect against user error, cyber-attack and application bugs. It is important to acknowledge that protecting those cloud-native applications is different than protecting traditional on-premises applications.

1

BUILT SPECIFICALLY FOR THE CLOUD Instantiating a legacy application in the cloud with all its years of code baggage is inefficient and complicates the process. Many of the capabilities that data protection requires already exists in the cloud, the software merely needs to leverage those capabilities to create point-in-time copies of data.

2

OFFERED AS A SERVICE The solution should be available through the provider's online marketplace and requires no additional infrastructure. Some cloud application protection solutions copy data outside of the cloud provider, either on-premises or to another cloud.

The copying out of data from the provider means that the organization has to physically create infrastructure in its data center or virtually create infrastructure in another cloud provider. In either case, the customer has to manage and pay for on-premises storage and compute for the backups, or it has to pay continually for backup storage and compute at another cloud provider.

Some vendors make a case that pulling data out of the cloud to another provider protects the customer from a cloud catastrophe. Given the infrastructure and multiple locations of major cloud providers like Google Cloud, it is highly unlikely that they would lose access to all their data centers for an extended period.

3

TAKE FULL ADVANTAGE OF WHAT THE PROVIDER OFFERS The solution should leverage the provider's capabilities to install directly from their marketplace with no additional requirements for installations. The cloud-native solution is ready to start capturing point-in-time backups almost instantly.

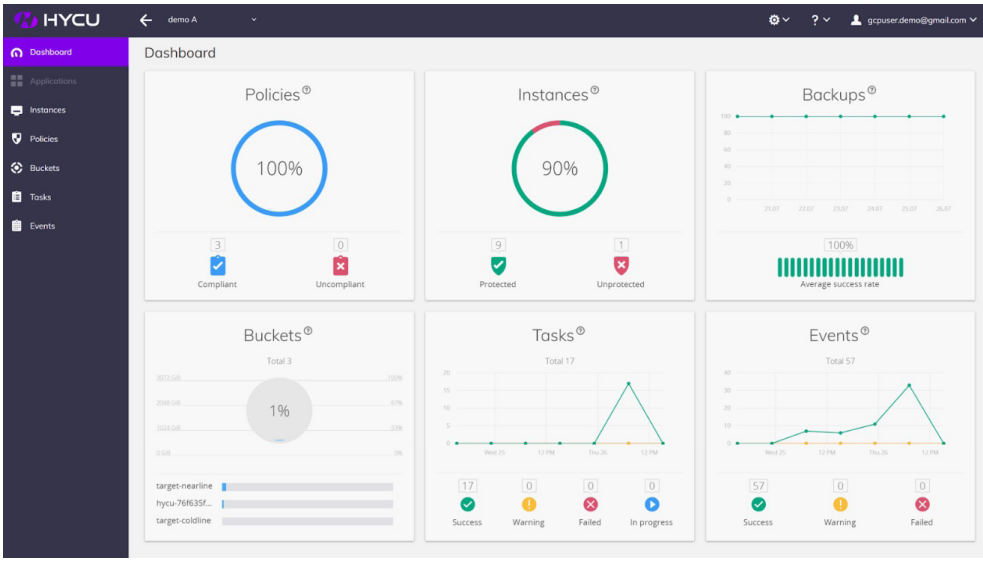
The solution should also leverage the provider's built-in snapshots. Google Cloud, for example, provides snapshot capabilities but accessing and automating them is difficult. The software should leverage Google's existing snapshot capabilities to gain zero-impact data protection points. It should then backup up those snapshots to Google Cloud Storage Buckets. It should also leverage Google's Identity and Access Management so the organization can control which users have access to sensitive backup information.

The goal of a cloud-native backup solution is to leverage the provider's snapshots for zero-impact data protection and then to eliminate the impact on production load by offloading the snapshot data to the protection software's storage bucket. If the software leverages available Google resources, then it doesn't need agents and is guaranteed application consistency. The software should also provide the ability to extract single files from these protection images.

Point-in-time backups of cloud-native applications is a must have for organizations looking to host production applications in the cloud

Finding the right backup solution is a challenge as legacy vendors cobble together solutions. IT planners need to look for solutions that are themselves cloud native and leverage cloud feature sets to deliver an easy to use but a reliable point-in-time backup.

CHAPTER 4 | HYCU FOR GOOGLE CLOUD PLATFORM



HYCU for Google Cloud Platform (GCP) is a native Google Cloud service designed to protect applications running in the Google Cloud. The software is available directly from the Google Marketplace. Instead of developing its own mechanisms for protecting data, HYCU for GCP uses the native Google Cloud snapshot capabilities to capture the first copy of data. It also fully integrates into Google Identity and Access Management (IAM).

HYCU manages the snapshot process and then leverages it to create a standalone copy of the data. During copy creation, it offloads the transfer of data from the production instance of the application, so the transfer does not impact application performance. Once the standalone copy is under HYCU management, the user can, through the HYCU interface, trigger either full Virtual Machine or single file recoveries.

HYCU WALK THROUGH

Getting HYCU up and running is straightforward.

- The user searches for the solution in the Google Marketplace.
- Once selected, the marketplace redirects you to a signup page using the administrator's Google Credentials.
- The HYCU service sends the administrator an email which takes them to a login screen.
- After authentication HYCU presents the administrator with the HYCU dashboard, which provides the protected status of each application in the environment.

Each instance or application is assigned a backup policy. There are several types of policies available, and users can define their own. The policy configuration screen is straightforward. The administrator sets how often the backup should occur and how long the backup should be retained. The backup policy will automatically backup to a default regional bucket.

The policy screen allows the administrator to setup custom parameters like custom buckets, for example, enabling exact control over where and how HYCU stores protected data copies. The administrator can also control the speed of a restore, where instead of recovering the data from the backup targets, it also keeps a configurable number of snapshot copies locally, which speeds recovery time. The local snapshots do cost money, so the organization needs to balance how many snapshots it keeps. Once administrators create a set of policies, they apply the appropriate policies to the organization's cloud application instances based on recovery point and time objectives and data retention requirements.

As companies migrate to the cloud, they can't afford to ignore the importance of a stand-alone, protected, point-in-time copy of data to safeguard against accidental user errors, rogue users or cyber-attacks. Most cloud services provide fundamental protection capabilities that require integration with the application or manual execution. Organizations like HYCU are automating the cloud protection process and making it easier to implement and manage, especially across a broad spectrum of applications. The availability of complete data protection solutions should be a crucial part of the cloud selection process.

ABOUT OUR PARTNER

HYCU makes it easy to thrive in a hyper-simple, hyper-converged world. The pioneering enterprise software company specializes in data backup, recovery and monitoring for hyper-converged infrastructures (HCI). Headquartered in Boston, Mass., HYCU harnesses 25 years of sophisticated IT experience, insights from over one million users, and work with more than 25,000 customers, more than 10 ISVs and 350 employees to create a deep and unrivaled well of industry expertise. The result is unsurpassed alignment with industry leaders and a formidable competitive advantage in the Enterprise Cloud space. HYCU's flagship product, a purpose-built backup and recovery solution for Nutanix, is acclaimed in the industry and features performance and value that are unmatched. To learn more about HYCU, visit www.hycu.com, follow @hycuinc and connect with us on LinkedIn.

THE FIRM

Storage Switzerland is the leading storage analyst firm focused on the emerging storage catergories of memory-based storage (Flash), Big Data, virtualiztion, and cloud computing. The firm is widely recognized for its blogs, white papers and videos on current approaches such as all-flash arrays, deduplication, SSD's, software-defined storage, backup appliances and storage networking. The name "Storage Switzerland" indicates a pledge to provide nuetral analysis or the storage marketplace. rather than focusing on a single vendor approach.

THE ANALYST

George Crump is the founder of Storage Switzerland, the leading storage analyst firm focused on the subjects of big data, solid state storage, virtualiztion, and cloud computing. He is widely recognized for his articles, white papers, and videos on such current approaches as all-flash arrays, deduplication, SSD's, software-defined storage, backup appliances, and storage networking. He has over 25 years experience designing storage solutions for data centers across the U.S.