HYBRID CLOUD STORAGE: EXTENDING THE STORAGE INFRASTRUCTURE FOR MAJOR COST SAVINGS

NOVEMBER 2013

Cloud computing has several clear business models. SaaS delivers software, upgrades and maintenance as a service, saving customers money by eliminating costs of ownership that the cloud provider now bears. Several technology factors contribute to SaaS' increasing popularity including protocol standardization, the ubiquity of Web browsing, access to broadband networks, and rapid application development. It's not perfect – people have legitimate concerns about data security, governance, vendor lock-in, and data portability, but based on its success, the advantages of SaaS seem to be outweighing its challenges. And the market segment is growing fast.

Another cloud computing model is IaaS, where the customer outsources the compute infrastructure to a cloud provider. This model is gaining traction, especially for application development and testing. App developers are able to take the capital they would otherwise have to spend on buying computing gear and target it to specific development projects that are underway in Internet data centers. The problem with IaaS is that cloud software development doesn't necessarily translate well into on-premises deployments and many developers prefer to develop SaaS.

Storage in the cloud is yet another business model with different dynamics. While SaaS and Iaas are strongly oriented towards cloud deployments, there are strong pressures driving cloud storage toward on-premises deployments. While storing data in the cloud for SaaS and IaaS computing is certainly important, the vast amount of data still resides on-premises where its growth is largely unchecked. If the cloud storage is going to succeed, it needs to become relevant to the people managing data in corporate on-premises data centers.

The Promise of Hybrid Cloud Storage

Running applications on-premise and using the cloud to store active files is not realistic due to latency and bandwidth issues. On-premise Tier 1 transactional applications with their heavy traffic, high churn rates and high performance I/O are completely out of the ballgame. However, Tier 2 applications like Exchange, SharePoint and office applications generate the majority of business data. This data quickly ages, resulting in vast stores of dormant data. This environment can strongly benefit from integrating on-premise storage with cloud storage.

The new hybrid cloud storage (HCS) model holds tremendous promise for cost savings around scalability and disaster recovery for busy Tier 2 applications. The idea of HCS is to integrate on-premise storage systems with online cloud storage tiers to extend the infrastructure across on-premises and cloud domains. This enables IT to significantly increase its ROI by lowering on-premises infrastructure costs for long-term storage of Tier 2 application data.

Why is this technology so welcome? Because IT is between a rock and a hard place with Tier 2 application data. The data is growing by leaps and bounds: reaching 60% or greater growth year after year — faster than any other type of business data. This data growth is extremely costly in terms of storage capacity, power and cooling capacity, and management overhead. Managing the physical plant is only part of the problem as best practices for data protection are also under heavy pressure. Without a different storage approach, many IT organizations are stuck with increasing their annual storage spending, including doubling down to pay for expensive remote DR sites to immediately failover and restore active application data.

HCS solves these challenges by turning the cloud into a functioning storage tier for an on-premise storage system. Hybrid cloud storage preserves on-premise processing speeds for the working set of data on-premises and adds the scalability of the cloud with its controllable costs. A common management interface unifies the on-premise and cloud environments, enabling IT and application admins to view stored data as an integrated unit.

HCS architecture consists of an on-premise storage system built for Tier 2 application requirements. Ideally the system contains SSD cache and SSD storage tiers for highly active data, and SAS disk for fast secondary tiers. The HCS storage system connects to the cloud and enables it to act as a third storage tier for aging application data. The application may freely access the data on the cloud without using backup catalogs or archive restore procedures.

The Problem with Latency

The trick to maintaining the cloud as an active storage tier is compensating for latency. Latency between on-premise and the cloud is a given at the present state of cloud development. The most common Web protocols including REST and SOAP introduce latency, although they are a vast improvement over direct file storage calls over the WAN. When active data is stored on the web and on-premise applications are attempting to interact with it, latency can play havoc with response times.

HCS vendors must compensate for latency in order to integrate cloud storage as an active tier. The industry is in the beginning of developing these capabilities and despite the challenges appears to be making real progress. Storage vendors are actively developing technology to counteract latency's effects because the business opportunity for integrating on-premise data with cloud data is so compelling. Vendors have been madly working from both sides of the divide: the on-premise side employs Web protocols with hybrid appliances and gateways, minimizes data with inline data dedupe and compression, optimizes the WAN, and runs snapshot technology in the cloud as well as the array. Several also map metadata across on-premises and cloud storage to facilitate fast restores from the cloud. Cloud service providers are also enabling the technology with encryption, data integrity scrubbing and in-cloud cloning and remote replication, making multiple cloud instances appear as a single cloud resource to the on-premises processes.

These integrated on-premise and cloud technologies offer acceptable performance levels for Tier 2 data that can be separated by activity and access frequency. They are not yet developed enough for fast I/O requirements on the Tier 1 application side, nor are they optimized for newly created data from Tier 2 applications. But they do add immense and cost-effective scalability to the production environment and its aging application data.

Key Capabilities

You won't find HCS systems on every street corner but they are fast gaining market traction. However, buyer beware of cobbled-together systems that do not integrate on-premises with cloud-

resident data. Real HCS systems are optimized to deliver high performance for working set data and treat the cloud as an additional tier for less-active data, not just a backup target. Look for the following capabilities in any HCS system:

- **Centralized management tools.** HCS enables centralized management for the same storage system whether tiers are on-premise or in the cloud. Management services run in both environments to monitor and share information to the central console. The cloud becomes a storage tier instead of a separate backup or archival target.

- **Efficient disaster recovery.** HCS systems establish efficient links with the cloud tier to enable fast data movement between the cloud and on-premise system. The cloud tier is optimized for disaster recovery, enabling customers to quickly restore data from the cloud without the expense of remote DR sites or massive tape vaults. (The HCS system should be geographically unbound so that the restore is not limited to a specific cloud data center.) Instead data transparently moves from Tier 3 onto on-premise storage, without requiring additional management or restore steps. In addition, HCS systems may be able to provide application-driven restores that prioritize the data that is downloaded so high-priority applications get up and running as quickly as possible. DR testing also benefits because the amount of data involved with application-driven restores is far less than traditional tape-driven restore processes.

- **High scalability.** HCS hugely expands on-premise scalability by allowing production environments to use cloud's scalability. The on-premise storage system should be high performance and scalable as well, preferably offering SSD and hard disk storage with dynamic tiering. This architecture reserves the on-premise system cache, SSD and disk tiers for frequently accessed data and the cloud tier for less active data that is immediately available to the application.

Vendor Landscape

There are a number of storage vendors that use the cloud as a backup target. This is a fine thing and we encourage backup admins to take advantages of cloud scalability and availability for that purpose. Archives have a similar value proposition. But HCS creates a whole new level of opportunity by enabling cloud storage for production environments.

Only a few storage vendors and their cloud partners have offerings that classify as HCS. How successful these vendors are over time will depend on the use cases they support, including backup, archiving, disaster recovery, primary storage, or all of them.

There is a lot of confusion around backup. HCS systems are intended to back themselves up to the cloud, as opposed to being used as a backup target for a backup server or software. While target usage has value, the data in the cloud is packaged inside a backup format and is not directly accessible by applications on-premises, which means the restore process is likely to be much longer.

DR is the next of kin to backup. HCS should provide location-independent DR so that recovery operations can be done from any other location. Application data is stored in the cloud where it is immediately available for fast restore. Both Microsoft StorSimple iSCSI systems (block) and TwinStrata CloudArray (file) are excellent in this respect, and Windows Azure and AWS EC2 both support the usage case.

Using cloud storage for primary data is the most advanced use case because it requires a high performance on-premise array that manages the latency between the on-premise and cloud domains. This is where HCS has the potential to have the deepest impact on storage costs by scaling storage across to the cloud instead of bulking it up on-premises. Nasuni has a good system for file-based data, and StorSimple with Windows Azure enables primary storage as well as backup, archiving and DR with its iSCSI and block storage for the enterprise. We expect to see numerous other entries as HCS gains more and more interest in the marketplace.

The cloud partner makes a big difference as well in the HCS solution. Amazon may be good for smaller business and provides its own software gateway to provide a way to move data between on-premise customers and AWS public cloud storage. However, enterprise customers will receive the greatest benefits from HCS even though enterprises have the most challenging environments. We find that StorSimple with Windows Azure is the best fit for the enterprise storage infrastructure. StorSimple provides a hybrid flash array with SAS disk for high performance on-premise processing, and extends storage tiering onto Azure for a highly scalable and economical storage expansion. Azure in turn provides highly scalable cloud infrastructure with cloud-based replication, geographically neutral thin restores, and data integrity scrubbing.

Taneja Group Opinion

One of HCS' most compelling propositions is that it is a natural intersection of cloud and on-premises computing, not a technology bleeding edge. With the ability for IT to leverage cloud scalability for production environments, we expect to see sizable growth in HCS offerings. It will be interesting to see how vendors with very large scale-out on-premise storage systems react, since HCS scales storage at a fraction of their cost.

We believe that development will accelerate in response to a market hungry for cloud scalability in production environments. Ongoing development will require continual attention to security and performance concerns, but customers do not have to wait for further development to invest in HCS. HCS is here today and offers strong benefits for disaster recovery and Tier 2 primary application environments.