Enabling effective Hybrid Cloud Data Management

A Register Report

The public cloud has revolutionized enterprise computing. Businesses and other organizations have become used to virtually unlimited infrastructure resources, and the ability to manage and procure these in a transparent and predictable manner. This is as true of storage as it is of compute.

But not every workload or data set can be easily migrated to the cloud. Companies might have regulatory or security concerns or be conscious that their corporate applications can't easily be refactored for the cloud.

Moreover, they have become aware that many of the headaches associated with traditional infrastructure – data silos, data sprawl, vendor lock-in – can actually be reproduced in the cloud, potentially at a much larger scale. Meanwhile, ingress and egress charges for data can make it harder to build the architectures they desire, or switch providers when they wish, which can ultimately make the cloud a much more expensive option for the job in hand.

This paper examines these concerns, demonstrating why hybrid environments combining on-premises (on-prem) and cloud resources have a crucial role to play, particularly for medium sized organizations and those scaling up their operations. And it details the prerequisites that technology leaders should look for when seeking a partner that can deliver the best of on-prem storage in combination with the cloud.

MARKET OVERVIEW AND TECHNOLOGY CONSIDERATIONS

The world is creating a staggering amount of digital data. In 2021, IDC reported that 64.2ZB of data was created or replicated the previous year. However, only 2 percent at most of this data was actually "saved" beyond 2020 into 2021, with the remainder either deemed unworthy of retention or temporarily cached before being overwritten.

The analyst firm also predicted that data creation and replication will continue to grow at a CAGR of 23 percent until 2025, figures which include both business and consumer data. But it's important to grasp that enterprise volumes are growing twice as quickly as consumer,

64.2ZB	volume of data created or replicated in 2020
2%	volume of that data saved into 2021
23%	CAGR of data replication and creation volume 2020-2025

Source: IDC

due in part to increased business use of the cloud for storage. Likewise, cloud data creation is growing faster than the aggregate total which IDC terms the "DataSphere." And the amount of data stored in the cloud is growing faster still and is almost matched by the creation of data at the edge, the company says.

On-prem storage is here to stay

While the cloud is increasing its share of the DataSphere, it is also clear that there are plenty of organizations which are either in no hurry to shift their data off prem or will be happy to leave it there for the duration.

The most obvious reason for them to maintain on-prem storage is that they have mission critical applications and data they can't – or don't want – to transfer. In some cases, companies may simply have reservations about refactoring monolithic or aging applications to run in the cloud. But there are a range of other factors that give IT departments pause for thought when considering whether it makes sense for them to take the plunge.

Data governance, privacy and security

There may be regulatory reasons why organizations may not wish to trust their data – and its underlying storage infrastructure – to off-premises cloud environments. In Europe, for example, the GDPR imposes specific limits on how long data can be retained and how, and where, it can be stored or shared. Enterprises may conclude that holding data in the cloud is incompatible with their legal or contractual obligations, or that cloud providers' tools do not give them adequate control of their data, compared to on-prem tools.

Much depends on the nature of the data and workloads being hosted and the individual compliance requirements of the organization in question, but security remains an important consideration that influences tech leaders' thinking.

Legal obligations increasingly frame security concerns too. As cyber attacks become more of a problem, with ransomware at the top of the agenda, governments are becoming more involved in data protection. US authorities, for example, are imposing more stringent reporting requirements on organizations hit by ransomware and data breaches. Meanwhile, cyber attackers are known to be targeting more mainstream organizations alongside critical national infrastructure providers, corporates, and large enterprises.

Beyond legislative penalties, a data breach or cyber attack can have a catastrophic effect on an organization from a commercial and reputational standpoint - by disrupting revenue operations and causing customers to fret about their data being exposed for example, and leading shareholders question their support for the organization.

Much depends on the nature of the data and workloads being hosted and the individual compliance requirements of the organization in question, but security remains an important consideration that influences tech leaders' thinking.

Closer to the edge

The rise of edge computing also influences where companies choose to hold their data. Where exactly the edge lies might be hard to pinpoint, but in practice many organizations are looking to move compute resources closer to the source of the data. This can mean processing power and storage capacity moving out of the central datacenter to edge locations to perform transactions and analytics.

For example, organizations with fleets of distributed IoT devices may seek to store and process the data the devices produce at the edge, or even on the device itself. This can deliver bandwidth and latency benefits compared to transporting IoT data across the internal network, or from the edge, directly or via the internet for processing in the cloud.

Data gravity

Many of these concerns feed into the broader issue of data gravity. This is the notion that large data sets will pull other objects, such as applications and services, closer towards them, just as large celestial bodies attract smaller bodies.

Put another way, large data sets are hard to move. Once a large amount of data has coalesced in, say, an on-prem datacenter, moving it to the cloud will become increasingly difficult. The same applies in the other direction.

This will be partly due to technical issues. Replication and syncing raise the problem of copy sprawl, as well as latency concerns.

But moving data also involves hard financial costs in the shape of cloud vendors' ingress and egress fees. And the longer a data set remains in one place, the more inertia it accumulates, and the more intractable the problem becomes.

Cloud storage traditionally provides the "virtual device" to store data, but not necessarily the tools to manage and protect that data.

HOW STORAGE VENDORS ARE MEETING THE CLOUD

Storage vendors may see the rise of the cloud as either a challenge or an opportunity.

The promise of unlimited storage that is easy to manage, move, and manipulate might appear to be an existential threat to companies who sell on-prem infrastructure. These companies might withdraw into a defensive posture towards the cloud.

The ideal outcome is a hybrid cloud approach that delivers unified data management across on-prem and cloud infrastructure, without compromising on security, resilience, or cost.

But more forward-looking suppliers will view the cloud as a complement to on-prem infrastructure and an incentive to raise the bar for their own products, particularly when it comes to management.

That can mean embracing an approach that doesn't attempt to trap users and their data into siloes, but rather allows them to see and manage all the data and infrastructure they have available – whether that is on-prem or in the cloud, or both – within a single management console.

They will seek to help their customers move data across cloud and on-prem infrastructure. And they will aim to ease data management issues, by enabling on-premises and off-premises data to be viewed and managed consistently across solutions and multi-vendor clouds.

The ideal outcome is a hybrid cloud approach that delivers unified data management across on-prem and cloud infrastructure, without compromising on security, resilience, or cost.

Simple ways of scaling

This more open approach also feeds into how vendors help customers scale up their infrastructure. This includes making it easier to scale on-prem infrastructure up, or out, as the organization's development dictates, without hitting artificial architectural constraints.

But it also means being able to take advantage of the resources the cloud offers, as well as that on-prem infrastructure, while maintaining visibility and control across their entire data landscape.

Cloud storage, for example, may be the best solution for dealing with data from which the organization has extracted as much "value" as possible, which still needs to be retained though not necessarily on locally managed infrastructure.

This can be a prompt for vendors to give customers the ability to manage the processes needed to underpin storage (such as backup and data protection) from one consistent console rather than having to manage services using a different tool for each artificial "silo."

Cloud storage traditionally provides the "virtual device" to store data, but not necessarily the tools to manage and protect that data. So, data management providers need to deliver the tools and capability to do that across clouds and on-prem locations.

Business model considerations

In a fast-moving world, long-winded procurement processes can handicap corporate agility. They can also lead end users themselves to make their own cloud arrangements when what they really need is nothing more than a financial consumption model for IT aligned with their data usage.

Traditional on-prem CapEx procurement cycles can take months, meaning that by the time infrastructure is in place the world has already moved on. Moreover, by locking tech teams into a series of multi-year procurement and rollout cycles, it becomes harder to upgrade equipment or provision capacity quickly when demands change.

This doesn't just impact the customers' ability to respond to change. By postponing upgrades and tying up budget, it also hampers their ability to innovate and meet new business requirements.

It's not a one-way street

It's important to remember that it's not a one-way journey from the on-prem world to the cloud. Increasingly, startups, or individual departments or projects within larger organizations, will be "cloud-born." But as they grow, the teams behind them come to recognize the benefits of on-prem core or edge infrastructure, whether due to regulatory requirements, or the security and latency issues we've already discussed.

In either case, these customers do not just need easy movement, migration, and replication between cloud and on-prem infrastructure, but a unified cloud-like experience that encompasses both.

Because snapshots are read-only they provide near instant restores in the event of a ransomware or other cyberattack, or other data corruption incidents.

LENOVO'S STORAGE SOLUTIONS

Businesses and public sector organizations clearly face a challenge when it comes to managing their data, now and into the future. They may not want to move entirely to the cloud, but still want to take advantage of it when necessary. This means their on-prem infrastructure needs to be cloud-friendly. But it also needs to offer them the security, resiliency, and manageability which was previously reserved for top-tier on-prem enterprise storage systems.

Lenovo's DM Series of data management solutions is designed to bring enterprise class data infrastructure and features to medium sized organizations, providing simple starting blocks that can be scaled up easily, and which have built-in integration with cloud-based storage services.

Scalability, security, and replication in the cloud

While the DM Series brings enterprise storage features to mid-range, on-prem installations, it also provides API level integrations with the key enterprise cloud vendors, notably AWS, Microsoft, and IBM.

The base software layer for the DM Series Storage Manager Tools also sits in three public clouds. This means admins can transparently use on-prem and cloud resources for tiering and other storage management functions, with full visibility and automation, from a single management platform.

For example, cold data can be automatically tiered from a DM5100's SSDs to lower cost object storage in the cloud, to maximize on-prem performance. Critical data can be replicated to the cloud using Cloud Volumes, further enhancing data protection and security, in case of a disaster such as a ransomware attack, without the need for a separate on-prem site.

Automatically tiering or replicating into the cloud is as simple as adding Amazon account details into the management console and selecting the required capacity. Or organizations can set up their own S3 private cloud.

The DM series also features built in snapshotting. Because snapshots are read-only they provide near instant restores in the event of a ransomware or other cyberattack, or other data corruption incidents. Further security is provided with in-built encryption.

The result is a seamless experience that gives organizations the financial and technological flexibility of the cloud, combined with the control and security that comes with operating their own infrastructure on-prem.

The snapshot capability provides the basis for far more efficient replication between on-prem systems or the cloud, because only those blocks that have changed need to be migrated or synced.

In addition, the platform also provides integration with Veeam, Veritas and Commvault, meaning admins can manage their data protection and backup for multiple data and application types from a single platform.

TruScale

The DM Series is also available under Lenovo's TruScale Infrastructure-as-a-service program. This transforms procurement from a CapEx to an OpEx model, providing predictable costs and removing the headaches of extended, successive procurement cycles.

TruScale provides over-provisioned on-prem capacity which admins can tap into as needed, and the ability to burst into the cloud if they need further resources.

Just as important is the program's inclusion of upgrades and health checks. Software upgrades and patches are automated and applied instantly, eradicating the possibility of human error during patching or other routine management tasks which could leave vulnerabilities exposed.

With TruScale Infinite Storage, on-prem storage infrastructure is deployed with guaranteed technology upgrades that are transparent to the customer's operations and eliminate costly data migrations. In this way, organizations can focus on managing their data with the latest technology.

The result is a seamless experience that gives organizations the financial and technological flexibility of the cloud, combined with the control and security that comes with operating their own infrastructure on-prem.

CONCLUSION

The cloud can appear to be an easy solution to an organization's storage needs. But there are multiple reasons why mid-sized and large organizations will want to adopt a hybrid, multi-cloud approach to cloud and on-prem infrastructure and data management that gives them both the level of resource and the control of their data they need.

The alternative is reproducing all the drawbacks of legacy infrastructure – siloes, vendor lock-in, restrictive tooling – on a third-party provider's infrastructure, with the added danger of unlimited consumption.

A successful migration to effective hybrid cloud management starts with choosing a partner that gives the IT department flexibility in both their on-prem environment, and in the cloud, and a unified management experience across the two.

Sponsored by Lenovo

Lenovo Infrastructure Solutions Group (ISG)

Lenovo Infrastructure Solutions Group (ISG) is a smart infrastructure solutions provider to organizations of all sizes. Our technology and insight power the data-centered heart of smarter retail, smarter manufacturing, smarter cities, smarter healthcare, smarter finance, and beyond.

Through edge and cloud computing, analytics and artificial intelligence, data management and storage, and Infrastructure-As-a-Service, we deliver Smarter Technology for All. We're the only data center provider with end-to-end manufacturing. We own our entire supply chain for everything we build, to deliver a level of security and seamlessness that no one else can, anywhere in the world.