White Paper

Maximizing Investments in Multicloud Data Management for Superior Results

Sponsored by: Dell Technologies

Phil Goodwin
October 2020

IDC OPINION

Data protection utilizing the cloud – to, from, and within – has been widely embraced throughout all industries and geographies. IDC research shows that 90% of organizations use the cloud as part of their data protection strategies. In most cases, this is a hybrid solution, which involves moving data from on-premises systems to the cloud. Increasingly, however, cloud data management and protection requirements are expanding to cloud-native data protection for SaaS and cloud-native application data – that is, software or applications built for and executed within a cloud environment, often based on Kubernetes containers. Because a majority of organizations have applications deployed in more than one cloud, IT managers are implementing multicloud data management systems, platforms, and strategies to ensure that valuable corporate data is properly protected and governed and that maximum value is generated from that data. Cloud data protection solutions offer on-demand resource availability and scaling plus the flexibility and agility to adapt to changing application requirements and deployments, whether from on premises to cloud or vice versa or across multiple clouds.

Moving data to the cloud yields important benefits. First, public cloud repositories are especially well suited for offsite data storage to ensure data survival in the event of any data loss at the data source, whether that source is a core or edge location. Using cloud-tiering technologies, a cloud environment simply becomes another tier of storage available to IT administrators. Second, when properly configured to separate the data and control paths, cloud repositories can be used as an "air gap" measure to ensure data recovery from ransomware and other malware attacks. Third, cloud repositories can leverage data protection data sets for value-added use cases such as DevOps (i.e., test/dev and staging) and analytics. The advantage of using backup sets for these use cases is that the data can be very fresh and be accessed without any impact to production systems.

Multicloud data management recognizes that data is spread across core, cloud, and edge repositories. IDC defines the "core" as traditional datacenters or on-premises private clouds, "cloud" as the public cloud, and "edge" as branch office/remote office – distributed compute and data, endpoint, and IoT devices. Hybrid cloud data protection is the seamless use of the cloud as a target for data moving from either the core or the edge. Multicloud refers to more than one public cloud.

Despite the benefits that cloud brings to data protection, it also has its limitations. Attempting to recall (restore) large volumes of data from cloud to either core or edge may be unacceptably slow due to limited network bandwidth. While deduplication and compression can help improve this restore performance, restoring tens, hundreds, or thousands of terabytes may not be feasible. In addition,

cloud egress charges may make such an operation feasible only as a last resort. To solve this problem, many organizations will deploy a purpose-built backup appliance (PBBA) on premises for rapid data restore with cloud tiering built in to get the best of both worlds.

To maximize investments in cloud data protection, IT organizations need to adopt technologies, architectures, and products that offer the greatest agility to meet the evolving needs of data protection and application deployments. IDC recommendations include:

- Use cloud tiering with source-side deduplication that seamlessly allows data to be backed up and recovered in the core, cloud, or edge.
- Interleave software, appliances, and virtual infrastructure to optimize recovery regardless of where the data lives.
- Deploy reliable backup infrastructure capable of 96% success on backup/recovery jobs.
- Deploy data protection solutions that enable the protection of traditional and cloud-native workloads.

SITUATION OVERVIEW

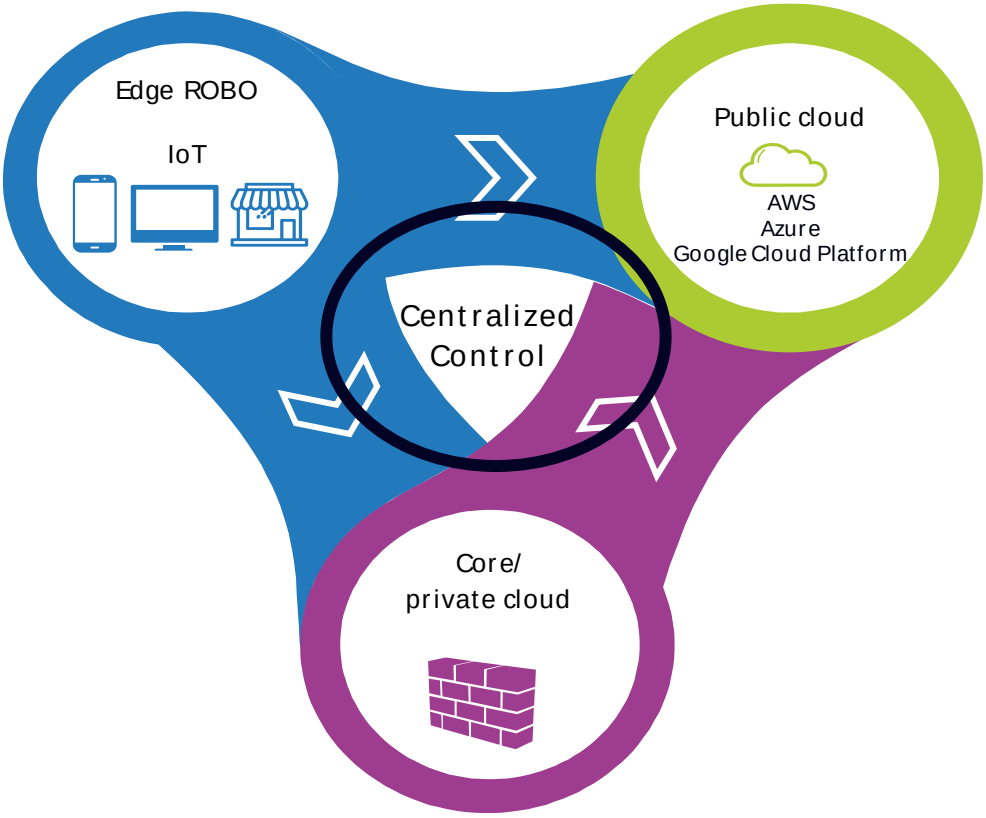
Applications being deployed across core, edge, and cloud locations with related data kept locally to the application has led to data sprawl and siloed data across many organizations. The data is often in different formats and managed by different groups. Too often, the management of the data is ad hoc on a silo-by-silo basis, making consistent management difficult. Data protection products may be purchased on a point solution basis without an overall strategy, leading to redundant tools and unnecessary labor to manage them. Policies are applied by the respective IT management teams, requiring manual effort to keep the policies synchronized — if they are synchronized at all.

Multicloud implementations complicate matters even further because many cloud applications are deployed by third-party SaaS providers. These providers have their own data management, protection, and retention schemes that usually do not match corporate standards and governance. For example, the provider schemes may be as minimal as 24-hour RPO, undefined RTO, and 30-day retention. For these situations, data protection cannot be an afterthought. Organizations must actively determine how to bring SaaS data under the corporate umbrella. This may involve using third-party data protection and replication tools to bring disparate data sets into a common repository.

Even as data proliferates and scatters across the organization's cloud implementations, the requirements to protect and recover it are becoming more stringent. In other words, IT organizations are expected to manage more data and recover it in less time with minimal data loss. Typical service-level requirements (SLRs) over the past several years have been 1-hour RPO and 4-hour RTO. Recently, 15-minute RPO and 1-hour RTO have become common. These SLRs simply cannot be met by most organizations without some data protection modernization. This will involve deploying several technologies, each with strengths for specific scenarios. Cloud can be used for data survival, data recovery, and restoring small quantities of data. Purpose-built backup appliances (PBBAs) can be deployed as physical devices at the core and edge and as virtual appliances in the cloud. PBBAs are well suited to restoring data rapidly, whether a few small files or hundreds of terabytes (see Figure 1).

FIGURE 1

Core, Cloud, and Edge with Central Control



Source: IDC, 2020

Moreover, because of the increasing threats associated with malware and ransomware, organizations are no longer viewing data protection, security, and governance as separate technologies. All three must be architected into a united solution where cyber-recovery is the imperative. Nevertheless, the governance capabilities must be flexible enough to address the regional differences regarding data sovereignty and privacy. Data protection products must be capable of identifying particular information, such as "right to be forgotten," so that the data in question is not inadvertently reintroduced to production systems during a restore operation.

FUTURE OUTLOOK

IDC estimates that as many applications will be deployed in the next five years as have been in the previous 40 years. In addition, 70% of CIOs have a "cloud first" mandate. Thus we expect the vast majority of new application deployments will be in the cloud, either as SaaS or as cloud native. Although many of these will be with the dominant hyperscalers, many will also be deployed in industry clouds, which are clouds catering to the needs of a specific industry (small to medium-sized financial services institutions, healthcare, etc.). As noted previously, IT organizations must take care to ensure that all corporate data is managed according to corporate requirements.

In addition, we expect a significant number of applications to be deployed at the edge, either in edge clouds or on IoT devices. These edge applications may generate both transient data and persistent data. Examples of transient data are device status, session information, and so on. This data need not be captured and protected. However, the persistent data generated by the application must be captured and protected according to SLRs and governance requirements. In some cases, the data will be captured, stored, and protected at the edge, and in other cases, the data will be captured and replicated or transferred from the edge to either the core or the cloud. Regardless of the process, the data protection and management software must be closely integrated with the application because only the application knows what data is transient or persistent.

Many of these new applications will be developed using containers. As with IoT applications (which may be container based), both the container and the data may be transient or persistent and data management and protection solutions must be designed to support this requirement. In most cases, that will involve integration with Kubernetes, which has become the dominant container orchestration platform. Yet, even though Kubernetes currently dominates, flexibility to support different orchestration tools will be important, whether those tools are for specialized environments or become an alternative to Kubernetes. When considering Kubernetes deployments, IT managers should also consider that Kubernetes is highly dynamic, undergoing frequent updates and modifications. As a result, entire systems must be recoverable to a specific state; simply recovering data or even containers may not be enough.

Among the greatest challenges facing IT organizations is cyberprotection and cyber-recovery. As a fundamental precept, we advise organizations to adopt a layered approach to data protection and backup. Of course, this starts with backing up data on a consistent basis, and our research shows that organizations that back up more frequently have lower risk of data loss. Organizations must also ensure an "operational air gap" between production systems and backup systems (i.e., have a physical break and/or separation of the control path and data path that cannot be compromised internally or externally) to keep backup systems from becoming compromised by malware or ransomware. PBBAs can be specially configured to create this air gap. Moreover, PBBAs can add a third layer of protection in the form of an immutable data copy that cannot be changed by malware. In addition, IT must ensure the fidelity of the vaulted data and ensure that it is free of malware, so you have safe and trusted backups to recover. As a fourth layer of protection, data encryption should be deployed for data that is at rest as well as in flight. This combination of backup copies, air gap copies, immutable copies, and data encryption offers organizations the greatest chance of recovering quickly from any cause of data loss with minimal loss of data.

As mentioned previously, hybrid and multicloud deployments are an IT fact of life and will only proliferate further. To address the data management and protection needs of these diverse repositories, consolidation of tools, processes, and policies will help apply consistency and reduce labor effort. While there may be instances where a point solution is required, being able to manage, govern, move, and protect data across core, cloud, and edge repositories will be advantageous. Moreover, we expect application evolution to cause deployments to morph in response to business requirements. Having a data management and protection strategy that can evolve with the requirements will be a "must have."

Considering Dell Technologies

The Dell EMC data protection portfolio consists of data protection appliances and software that allows organizations to protect and recover their critical data assets across edge, core, and cloud-based

infrastructure. With support for traditional (physical and virtual) and modern (containers, Kubernetes, and cloud-native apps) workloads, combined with integrated cyber-resiliency capabilities, these solutions provide organizations with the end-to-end data protection needed to ensure data is protected, secure, and recoverable.

Dell EMC PowerProtect Cyber Recovery delivers an air-gapped solution that strengthens and builds upon the base immutability offering through retention lock capabilities. In addition, running analytics on the data in the vault is an important component to enable a rapid recovery after an attack. Analytics help determine whether a data set is valid and usable for recovery (free from malware) or has been improperly altered or corrupted making it "suspicious" and potentially unusable. CyberSense analytics assist with assured data recovery because CyberSense reads through the backup set so that there is no need to restore data just to determine if it is clean, thereby avoiding the risk of opening harmful data in the vault. In addition, CyberSense can evaluate the full contents of the file, not just its metadata, to deliver superior analytics.

These solutions, available either in integrated appliances, software only, and hardware only or as comprehensive bundles, have the integrated features and technologies organizations need to satisfy the top drivers of data protection refresh uncovered in IDC's recent primary research in this area. Easy scalability into the petabyte range, assisted with inline compression and deduplication that can deliver up to 55:1 data reduction ratios; cloud tiering capabilities; integrated solid state storage options that speed both backup ingest and data recovery; and the flexibility to accommodate various deployment models (bare metal, virtual machines [VMs], and containers) enable these systems to deliver on the promise of modernized infrastructure to meet evolving data management and protection requirements.

Dell EMC data protection solutions are fully integrated with vSphere to enable VM admins to manage data protection directly from the native vSphere UI. With Dell EMC advanced VMware integration, VMware admins are empowered to more efficiently control their own data protection, resulting in faster backups and restores for virtualized mission-critical applications. In addition, PowerProtect Data Manager delivers support for vSphere 7 and Tanzu, paramount for protecting modern container-based applications and Kubernetes workloads.

CHALLENGES/OPPORTUNITIES

Dell Technologies arguably has one of the broadest and agile data protection portfolios in the industry. According to IDC, it was the market share leader (by revenue) for both data replication and protection software as well as purpose-built backup appliances in 2019. Dell's data protection platform can address the majority of enterprise data protection and cyber-recovery requirements, but such a broad range of products has its inherent challenges. For instance, some of Dell's products were first introduced well over a decade ago, well before the cloud became ubiquitous and containers were a serious application development choice. Dell introduced PowerProtect Data Manager to facilitate this evolution. Nevertheless, the company must continue to update not just product features but core architectural designs as well. Similarly, Dell must continually counter new entrants to the market that have the advantages of a single-purpose product focus without concern for prior architectures. Dell must deliver continued innovations in the area of Kubernetes, cyber-recovery, and significant VMware integration. Thus the key challenge for Dell is to remain competitive and innovative across numerous products and against dozens of competitors.

CONCLUSION

The emergence of cloud computing as a dominant application deployment architecture has yielded many benefits: on-demand resource availability, resource agility to meet changing requirements, lower up-front investment, and so on. However, it has also spawned a number of challenges: siloed data, fragmented platforms, inconsistent data governance, and redundant data protection products and schemes. When organizations have data on premises, in multiple public clouds, and at the edge, using point solutions for data management and protection can lead to significant inefficiencies in terms of both labor and redundant products.

To help address these multicloud challenges, organizations will adopt the broadest data protection capabilities possible for their situation. In most cases, the top priorities will be addressing ransomware and malware threats followed by better governance and data control. Being able to apply a layered approach that takes both proactive steps and offers response capabilities aimed at ensuring data recovery in the event of an attack is simply a best practice. Having a single, integrated data protection and management platform across core, cloud, and edge allows organizations to apply data management and protection policies across the enterprise to ensure consistent governance and service-level delivery. These solutions must protect not only current application data but future container and edge/IoT applications as well while adding value by making backup data available for secondary uses.

About IDC

International Data Corporation (IDC) is the premier global provider of market intelligence, advisory services, and events for the information technology, telecommunications and consumer technology markets. IDC helps IT professionals, business executives, and the investment community make fact-based decisions on technology purchases and business strategy. More than 1,100 IDC analysts provide global, regional, and local expertise on technology and industry opportunities and trends in over 110 countries worldwide. For 50 years, IDC has provided strategic insights to help our clients achieve their key business objectives. IDC is a subsidiary of IDG, the world's leading technology media, research, and events company.

Global Headquarters

5 Speen Street
Framingham, MA 01701
USA
508.872.8200
Twitter: @IDC
idc-community.com
www.idc.com