White Paper

Cloud-Native Data Protection for Modern Workloads

Sponsored by: Google Cloud

Phil Goodwin
January 2022

IDC OPINION

IDC forecasts that by 2025, 55% of organizations will have implemented a cloud-centric data protection strategy. We believe this to be the case because, even though organizations will continue to have workloads in the core (private or managed cloud), the preponderance of data and applications is shifting toward the cloud and edge. This shifting dynamic is driving senior IT management to approach data protection from a cloud-first perspective. We believe that data protection will extend to all reaches of the enterprise via management from the cloud using cloud-native platforms for self-service scenarios and using cloud service providers for services scenarios.

The move to cloud centricity is being driven by data location and new application deployment. IDC research finds that 50% of data is in the core, with the other 50% in the cloud, at the edge, or elsewhere. Currently, most organizations operate in a multicloud environment, including private cloud and more than one public cloud. Data is certainly growing at the core, but it is growing faster in the cloud and at the edge. Thus the volume of data in the core will soon be the minority of the data estate.

In addition, we estimate that 80% of new applications will be deployed in the cloud or at the edge, further driving data growth in those areas. These new applications will be predominantly cloud native and container based. Cloud-native, container-based applications will require a new approach to data protection, especially as stateful workloads in containers are deployed and become more mainstream. Cloud-based workloads will also include virtual machines (VMs), databases such as Oracle, and block-based applications such as SAP. Thus organizations must be prepared with an "all of the above" approach to data protection that can address traditional workloads, nontraditional workloads such as NoSQL data, SaaS applications, and containerized applications.

A key problem facing IT teams is that traditional data protection architectures are not designed to address the unique needs of containers. Specifically, older architectures are interval time based and focus only on recovering data and cannot restore Kubernetes container metadata. In contrast, container applications must be recovered as a system to a specific state at a point in time. Kubernetes, the dominant container orchestrator, is highly dynamic, and recovery products must be recovered in their entirety with the application and data. Moreover, microservices also must be addressed and recovered within the application context.

At the same time, data threats are expanding with more sophisticated ransomware and malware attacks. In fact, our surveys find that data security is often among the top priorities of IT teams. Data security and data protection can no longer be approached as separate tasks; they must be fully integrated and coordinated across the data estate. This includes the necessary disaster recovery (DR)

infrastructure and response needed to meet a variety of emergency scenarios, including ransomware. Use of the cloud as a backup and DR site can meet the needs of a secondary location while eliminating expensive dedicated secondary sites by leveraging on-demand cloud resources.

These emerging cloud-native data management platforms must also provide everyday data protection while driving ever better SLAs. These platforms should meet the recovery needs of containerized applications and facilitate disaster recovery in hybrid cloud and multicloud environments as well. To meet these stringent requirements, we believe organizations will begin turning to cloud-native data protection platforms that can simplify operations, reduce labor needed to manage infrastructure, and keep systems up to date against evolving threats.

SITUATION OVERVIEW

Data sprawl has become a leading challenge for IT organizations. As noted previously, data is now evenly spread over core (50%) and cloud and edge (50%), forcing IT and data managers to deal with data separated into silos. Our research shows that many organizations deal with between 14 and 20 separate silos, though in some cases, the number is substantially higher. Data becomes siloed due to data type, location, data owner, application deployment model, and other reasons.

Data silos cause numerous data management problems. First, individual silos may require different or unique management tools, including data protection that is deployed to address that specific environment. These multiple tools lead to overlapping product purchases, patching and maintenance vulnerabilities, IT team skills challenges, and labor inefficiencies. Second, silos cause governance challenges due to inconsistent policy application. Finally, silos also increase vulnerabilities of data exposure and mishandling or entry by internal or external bad actors for ransomware or data exfiltration exploitation.

Data sprawl is exacerbated by applications that are deployed in increasingly diverse ways. These applications will be predominantly cloud-native, container-based deployments relying on Kubernetes for orchestration or SaaS applications deployed and managed by third parties. Because of container implementation, DevOps teams are becoming more involved in the actual deployment of data protection products and strategies, making coordination between DevOps and ITOps an important activity.

FUTURE OUTLOOK

Because the preponderance of their application estate is rapidly shifting to the cloud, organizations are rapidly deploying data protection-as-a-service (DPaaS) solutions in the cloud. This includes backup as a service (BaaS), DR as a service (DRaaS) and archive as a service (AaaS). Collectively, DPaaS is growing at a 19.1% CAGR through 2025 and will exceed $18 billion by 2025 — the fastest-growing segment of the data protection market. This growth not only is indicative of the shift toward cloud-centric data protection but also illustrates the need for greater integration of data backup and disaster recovery operations. In this scenario, disaster recovery becomes an extension of data protection under the umbrella of business continuity. Indeed, the pervasive threat and impact of ransomware on an organization often require a DR plan and response to effectively recover. Moreover, best practices against ransomware include encrypting data at rest and in flight plus maintaining a copy in an immutable repository, such as a purpose-designed cloud repository, to ensure data integrity and recovery.

As organizations evolve their data protection strategies, solutions are getting closer to the cloud platform. Products with tighter integration with the cloud platform can offer simpler deployment, provisioning, and operations. Cloud solutions also offer the opportunity to leverage on-demand resources to optimize costs. Data is easily tiered to long-term archive for cost saving and data governance.

Cloud platform providers are actively developing an ecosystem of complementary products to satisfy the broad needs of organizations. These platforms can also facilitate management of data protection workloads across on-premises, cloud, and edge repositories, such as VMs as well as platforms from SAP and the like.

Considering Google Cloud

Google Cloud Backup and Disaster Recovery is based on the company's Actifio acquisition and the Actifio GO product. Actifio GO is a SaaS-based backup and DR for centralized, application-consistent data protection. The product supports Google Cloud Compute Engine and VMware Engine VMs, VMware, MySQL, Oracle, SQL Server, and SAP HANA. Features include incremental forever backup, rapid point-in-time recoveries, Persistent Disk snapshot orchestration, and geo-redundant Google Cloud Storage support. Organizations can also reuse cloud backups for additional use cases, such as test/dev and ransomware recovery. Actifio GO is available to protect newly migrated workloads in Google Cloud. Actifio GO can also be used to back up data for application deployed in the core (on premises) to Google Cloud.

In response to and in support of cloud-native applications, Google offers the Google Kubernetes Engine (GKE). The recent announcement of Backup for GKE is designed to deliver a simple cloud-native way for customers running stateful workloads to protect, manage, and restore container applications and data. Capabilities include incremental backups, application consistency support for multiregion (cross-region) backup, and selective (subcluster-level) backup and restore policies. This illustrates Google's effort to continue providing first-party integrated backup support for new products similar to cloud VM deployment solutions.

Google Cloud is establishing a broad ecosystem of first-party products as well as enabling ISV and systems integrator (SI) partners to offer freedom of choice for backup and DR. This includes support for backing up everything from Spanner to Cloud SQL. Organizations that want to use the cloud as a secure repository may do so using traditional cloud tiering. Organizations desiring a consistent cloud-centric experience with a traditional backup vendor can ensure their backups are being managed via the cloud. Many of these traditional backup vendors can be found in the Google Cloud Marketplace.

CHALLENGES/OPPORTUNITIES

Cloud-based data protection is highly dynamic, fragmented, and competitive. There are so many specialized use cases and requirements that no company, no matter how large, can address them all. Given all the development priorities of a cloud platform provider, data protection needs to be continually addressed proportionally to ensure continued capabilities.

To continue to meet customer requirements, Google must aggressively bring new capabilities to market to meet the majority of market needs. The company must balance this aggressive approach to the market with the need to create, foster, and grow a related ecosystem without alienating partners

that will supply critical specialized and differentiated capabilities. Google can continue to do this by maintaining open APIs and fostering relationships with ISV partners.

Although the company has an advantage of a fresh approach to protecting evolving workloads such as containers and without a legacy architecture to maintain, it still must deliver on the protection needs of those workloads. The majority of organizations are now multicloud (i.e., use more than one public cloud provider), and Google must deliver products that make IT buyers comfortable using Google tools in multicloud environments. It may also behoove Google to consider integrating Actifio GO with Backup for GKE in the future.

CONCLUSION

We believe the shift to cloud-centric data protection is going to accelerate in the coming years. Both data location trends and new application deployments are reaching a tipping point for cloud centricity. The additional complication of data protection across the core, cloud, and edge and the presence of data silos will further create the need for cloud solutions that allow central control of operations and consistency of data management policies.

Google Cloud, Backup for GKE, and Actifio GO are designed to support current and emerging application workload infrastructure to ensure that the infrastructure is fully protected. We believe that organizations will seek tighter integration with key cloud platforms that can offer the simplicity and consistency of data backup, disaster recovery, and cyber-recovery across core, cloud, and edge locations. Google Cloud's backup solutions are designed to give these organizations comprehensive solutions, whether they are protecting cloud-native, containerized workloads in GKE; traditional on-premises database workloads like Oracle; or newly migrated applications running in the cloud.

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

140 Kendrick Street
Building B
Needham, MA 02494
USA
508.872.8200
Twitter: @IDC
blogs.idc.com
www.idc.com

Copyright Notice

External Publication of IDC Information and Data – Any IDC information that is to be used in advertising, press releases, or promotional materials requires prior written approval from the appropriate IDC Vice President or Country Manager. A draft of the proposed document should accompany any such request. IDC reserves the right to deny approval of external usage for any reason.

Copyright 2022 IDC. Reproduction without written permission is completely forbidden.