IDC TECHNOLOGY SPOTLIGHT

Sponsored by: WANdisco

Mission-critical and business-critical applications are becoming distributed across private and public clouds, making data consistency a challenge. IT organizations need to take a proactive approach to business continuance. This IDC Technology Spotlight addresses how WANdisco's LiveData platform can help ensure data consistency and availability in a multicloud environment.

Ensuring Petabyte-Scale Data Consistency in a Multicloud Environment

July 2018

Written by: Phil Goodwin, Research Director

Introduction

AT A GLANCE

WHAT'S IMPORTANT

Legacy data protection plans are reaching a point where they will not be able to meet the data availability requirements of the most demanding applications.

IDC research finds that more than 70% of CIOs have a cloud-first strategy for application deployments. For most organizations, this includes legacy applications in the private cloud, cloud-native applications, and software-as-a-service (SaaS) applications. These applications may move and share data between both private clouds and public clouds for processing, analytics, data protection (e.g., backup as a service, disaster recovery as a service), and archiving. With most enterprise-class organizations deploying hundreds of applications, data is spread in a complex matrix across different platforms and geographies. In addition, IDC research indicates that IT organizations expect the number of applications to double in the next 12–24 months, making the problem only larger.

The service levels demanded by business units for data availability are constantly rising. The traditional method of data protection, which involves copying data on a scheduled basis and then restoring it after an event, is a reactive approach that soon will not be able to deliver the required data availability and consistency.

It is axiomatic to say that data growth continues to explode. IDC estimates that the global datasphere will reach 163ZB by 2025. Although not all this data must be captured and stored, we estimate the amount of data that IT organizations must manage will increase threefold from 2018 to 2025.

Organizations face an enormous challenge in ensuring data consistency and reliability because of significant new application deployments, diverse data repositories, stricter service-level requirements, and exponential data growth. We believe that IT organizations must evolve from reactive data protection practices to a proactive approach that delivers data consistency on a real-time basis.

Definitions

» **Recovery point objective (RPO):** The point in time that data is consistent and valid for recovery (RPO is really a measure of the amount of data loss tolerable in the event of system failure. For example, a 15-minute RPO would indicate that up to 15 minutes of processed data could be lost in the event of a system outage.)

» **Recovery time objective (RTO):** The total amount of time needed to restore data after a failure (For example, a four-hour RTO would indicate a tolerance of up to four hours to restore an application to production status following a system outage.)

» **Hybrid cloud:** A combination of on-premise compute/storage and public cloud–based compute/storage in delivery of an application environment

» **Multicloud:** A combination of two or more public cloud–based compute/storage resources in delivery of an application environment

Benefits

Recently, the concept of data platforms has been introduced to the storage industry. Data platforms break down storage management silos and make data always available. Benefits of data platforms include:

» **Comprehensive data inclusion.** Data platforms provide an enterprisewide view of data, irrespective of where the data is located.

» **Enterprise data policies.** Many SaaS applications have minimalist default data protection and retention policies. Data platforms can help bring the data from these applications into alignment with corporate data policies.

» **Proactive data availability.** Managing data replication by policy rather than platform helps improve data reliability and survivability, in contrast to backup/recovery, which is purely reactive. Most replications are technology specific (i.e., array or operating environment).

» **Management at scale.** Some organizations must deal with tens or hundreds of petabytes of data that demand automated management that can scale to such capacities.

» **Data for multiple use cases.** By managing data copies enterprisewide, data platforms make data available for multiple use cases to speed application development, testing, analytics, and so on.

» **Better resource utilization.** Unlike redundant and overlapping data replication tools, a data platform can offer overarching functionality and reduce the number of tools needed.

Trends

According to IDC research, 80% of new application deployments will include a cloud component. Most of these deployments will be either hybrid cloud or multicloud, meaning that data will be geographically dispersed. The temptation for IT organizations is to treat each of these situations uniquely and thus have siloed operations. These silos lead to personnel, process, and compute resource inefficiencies that increase the probability of data outages. If organizations can implement the capabilities that approach data availability as a platform holistically across the enterprise, the benefits can be significant.

In addition to the hybrid cloud environments, some organizations will prefer to have the data copied to more than one public cloud in a multicloud architecture. This may be for additional data protection, to avoid cloud vendor lock-in, or to make the data available to applications in different clouds. Thus data platforms must have the ability to migrate data around multiple public clouds simultaneously.

Considering WANdisco

WANdisco's Fusion LiveData platform significantly enhances data protection and availability, but it is not a backup product. It provides data management capabilities, but it is not a data management product. It provides data replication, but it is much more than a data copy product. Fusion brings together cloud data management concepts, data replication, and data protection capabilities to assist IT organizations with specific challenges and use cases. Its basic architecture is a shared pool of data across the enterprise with assured data consistency across all nodes. Specific use cases and capabilities include:

- **Disaster prevention.** Most importantly, we believe, Fusion is intended to help *prevent* the need for disaster recovery rather than merely facilitate recovery after a catastrophic event. Legacy solutions are designed to recover and restore data, which can take hours or days. The Fusion platform continuously synchronizes data in both hybrid cloud and multicloud environments.
- **Stringent service-level attainment.** The continuous synchronization of Fusion means that organizations can meet near-zero RPO and RTO data availability service levels. Because Fusion operates across the organization, this data availability is not limited to a few mission-critical applications with expensive high-availability infrastructure; rather, it is provided to all operations.
- **Simplified cloud migration.** Many organizations avoid cloud migration because of the time and risk involved. Fusion is designed to facilitate data migration to any supported public or private cloud without application downtime.
- **Data analytics.** Because data is immediately replicated and available, data analytics programs, such as Spark, can operate continuously against the latest version of data. Thus data-driven organizations can make decisions based on the most current information.
- **Hyperscale architecture.** Fusion scales out across on-premise and cloud environments to support petabyte-scale organizations.
- **Data discovery and experimentation in the cloud without risk.** Data can be replicated immediately so that copies of data can be used for data discovery in the cloud — using on-demand resources — with no risk of disruption to production systems.

Fusion is extensible to accommodate a variety of new and legacy data types. It provides some support for traditional SAN/NAS and NFS file systems but full support for newer data types such as Hadoop and Spark as well as public cloud environments, including Amazon S3 and AWS Snowball, Microsoft Azure HDInsight and Data Box, IBM OpenStack Swift, and Oracle BDCS. WANdisco has focused on providing complete data integrity and availability for growth platforms such as object storage and Hadoop and is leveraging the ability to add to its platform through plug-ins to extend those capabilities to the more traditional storage platforms.

Challenges

WANdisco Fusion is more radical in concept than it is in implementation. Fusion requires organizations to think differently about data management than its traditional meaning. Most organizations today manage data in silos under the direction of database administrators, application developers, and storage administrators. The Fusion difference in approach is not only technical but also organizational. Candidly, some organizations suffer from turf wars between these different groups. To get the most from Fusion, IT organizations must be willing to transform data availability management into a centralized function across operations. This does not mean that either data ownership or use needs to change. Done properly, implementation of a LiveData platform will allow different data users to complete their tasks more easily.

From a marketplace perspective, WANdisco will be challenged to properly position itself. As we noted previously, it is not a backup/recovery, replication, or data management company. However, the company will find itself competing with vendors in these categories. As an emerging vendor, WANdisco needs to clearly differentiate itself and help customers understand how it fits into the IT ecosystem.

Conclusion

Traditional backup/recovery, data protection, high availability, and disaster recovery methodologies are rapidly approaching a point at which they simply cannot continue to offer improved service-level delivery. The combination of rapid data growth, diverse application deployment platforms, and cloud computing will force organizations to rethink data availability strategies to meet evolving business requirements.

Rapid data growth, diverse application deployment platforms, and cloud computing will force organizations to rethink data availability strategies.

WANdisco's Fusion LiveData platform is, in many ways, a new product category designed to offer near-continuous data availability and assured data integrity. It offers the possibility of unifying data availability management even across petabyte-scale organizations. We believe that this approach is in alignment with the evolving requirements of the most demanding application environments. To the extent that WANdisco can address the challenges described in this paper, the company has a significant opportunity for success.

About the analyst:

***Phil Goodwin**, Research Director, Storage Systems and Software*

Phil Goodwin is a research director within IDC's Storage Systems and Software research practice. He provides detailed insight and analysis on evolving industry trends, vendor performance, and the impact of new technology adoption. Mr. Goodwin is responsible for producing and delivering timely, in-depth market research with a specific focus on data protection, business continuity and disaster recovery, and data availability.

IDC Custom Solutions

IDC Corporate USA
5 Speen Street
Framingham, MA
01701, USA
T 508.872.8200
F 508.935.4015
Twitter @IDC
idc-insights-community.com
www.idc.com