rubrik

Rubrik Mosaic™: Powered by CODR™

BACKGROUND

Businesses are in the midst of a digital transformation journey. According to research from IDC, 70% of CIOs have a cloud-first strategy with the objective of harnessing the power of the cloud to drive growth by delivering new customer-centric products and services while also driving greater operational efficiency. As part of this journey, enterprises are now operating IT across a multi-cloud infrastructure, deploying applications on the best-suited cloud, whether private, public, or managed. More specifically, there are two fundamental shifts occurring:

- Next-generation applications, which are hyperscale and distributed, are being "born in the cloud"; that is, they are cloud-first applications. These applications are being deployed on next-generation distributed, non-relational databases such as Apache Cassandra, MongoDB, Apache HBase, and many others. As non-relational databases, they offer high availability, but compromise consistency.
- Traditional applications, originally designed and deployed on traditional datacenter infrastructure, are migrating to the cloud. These applications are still predominantly deployed on traditional, relational databases such as Microsoft SQL Server, but many of the workloads are moving to the cloud.

Like other platforms before it, operating in the cloud needs a data management strategy, but traditional strategies have not kept pace with this rapid change. To address the requirements brought on by multi-cloud, a new cloud-first approach to data management is required. Driven by the need to manage both traditional and next-generation mission-critical applications in a multi-cloud environment as well as the shift to a multi-cloud infrastructure, four new key data management requirements have emerged:

- A software-only deployment model is required to create flexibility of deployment as well as customer freedom to deploy the storage of their choice.
- Each cloud uses different infrastructure; the only common denominator is the "data" itself. There are no ESX virtual machines or SCSI logical unit numbers (LUNs) in the public cloud, and the only common thing that binds all of the clouds together is the data itself, which does not change representation across VMWare, Amazon Web Services (AWS), and Microsoft Azure and other clouds.
- The elastic nature of multicloud databases necessitates that data protection and management be highly available, scalable, and failure resilient.
- The eventually consistent nature of next-generation databases requires novel point-in-time techniques for consistent state across a cluster.

THE GENESIS OF RUBRIK MOSAIC

The data-management industry thus far has had a scale-up mindset, with relational databases and virtual machines being backed up to secondary storage through monolithic media servers. Worse, the media servers were not just a pass-through entity, but would convert the data into proprietary formats resulting in backup vendor lock-in — creating massive amounts of "dark" data for enterprises. At first glance, it was clear that such an architecture was not suitable to support the requirements of traditional scale-up databases, and certainly not the requirements of next-generation high-volume scale-out databases, for which failures are a norm, consistency is not guaranteed, clusters grow in size, and, most important where customers want to extract value from their secondary datasets.

With that challenge in mind, Rubrik Mosaic challenged the status quo. Starting from scratch, we set out to create a new cloud data management platform that could address the requirements of both traditional and next-generation applications, thereby enabling organizations to successfully manage all of their data across a multi-cloud environment.

INTRODUCING CONSISTENT ORCHESTRATED DISTRIBUTED RECOVERY

Consistent Orchestrated Distributed Recovery, or CODR, is the architecture upon which Rubrik Mosaic is built. The core principle of CODR is a scalable, application-centric view of data management and data protection that distinguishes it from conventional (media-server or LUN-based) approaches or virtualized (VM-based) approaches. The benefit of our application-centric approach is fine-grained and highly space-efficient data protection and data mobility that can span clouds over network links. Furthermore, such an application-centric view allows the CODR engine to enable rich data-management services (e.g., data governance, security, and masking), unlike VM-based or

LUN-based approaches, which treat data as an opaque object and have no application context.

CODR is built upon several additional distinguishing principles. Let's take a look at each one.

Elastic Compute Based

The CODR architecture is a compute-only data protection and mobility service that can be auto-scaled (elastic) up and down (scale out) depending on the application change rate — much like cloud compute and cloud storage services that are elastic in nature. The CODR architecture does not create its own storage or file-system; rather, it consumes storage as a service that is used for storing versions (backups) of databases. All of this results in significantly reduced infrastructure spending. By contrast, both conventional approaches (backup appliances) incur high fixed storage and compute costs that accrue even if the data protection service is not being used. As a result, these approaches are a complete misfit with the IT movement of moving applications to the cloud, and the purpose-built backup appliance (PBBA) market is declining much faster than before.

Semantic De-duplication

The CODR architecture introduces the industry-first *semantic deduplication*. There are three important reasons for the demise of traditional de-duplication techniques:

- Data formats are increasingly compressed, leading to the reduced effectiveness of de-duplication that relies on raw duplicated data within a content stream.
- In distributed database systems, data is replicated for availability, which results in multiple copies. These copies are not always exactly identical because ordering and flushing of updates differs across database nodes. The use of compressed formats complicates the challenge.
- Even if data were not compressed or replicated, the opportunity for de-duplication is also challenging because the average de-duplication fragment size is small, exponentially increasing the amount of metadata needed to keep track of de-duplication.

The goal of semantic de-duplication, therefore, is based on the following insight: even though the representations on disk are not physically identical due to compression or replication, they are semantically equivalent. Semantic deduplication identifies all semantically identical data fragments (such as a database column) so that you store only one copy of the data fragment in the secondary storage.

Parallel Streaming

Another important element of the CODR architecture is the direct parallel and streaming transfer of data from the application to the secondary storage (network storage and cloud storage). This results in customers having secondary storage that can provide low recovery time objectives (RTOs) and recovery point objectives (RPOs) for large data-sets, and high change rates as the data moves directly to secondary storage in parallel. By contrast, backup appliances and converged media server solutions can be deployed in only a single cloud and will be a choke point if the data to be protected is scaled out and globally distributed for performance and availability.

Globally Distributed Metadata Catalog

The CODR architecture not only enables rich data services but also uses a globally distributed catalog to make the services available across multiple locations. This means that not only can backups happen in any cloud, but the protected data is immediately visible to all clouds in the enterprise. As a result, in addition to being able to restore data to any cloud, rich data services can now be exposed to any location in the enterprise. This also completely eliminates any restriction on how data can move between clouds, providing complete cloud flexibility for data mobility requirements. Contrast this to traditional approaches in which both protection and recovery is uni-directional; for example, you can back up data from VMware to AWS but not the other way around. Similarly, in the same scenario, you can restore data from AWS to VMware but, again, the reverse direction is not possible.

Application-Consistent Versioning

CODR introduces the concept of versioning, wherein a version is defined to be a consistent view of a database. A version of a database is independent of the mechanism by which it is captured — it could be using snapshots, streaming logs, or asynchronous replication. As enterprises increase their pace to the cloud and everything becomes "as-a-service," versioning is likely to be abstracted away; consumers of a data protection service are going to be primarily focused on the promise of RTO and RPO rather than the "how" of the promise.

A MODULAR APPROACH DESIGNED FOR FLEXIBILITY

One of the key elements in the CODR architecture is the use of abstractions to isolate the architecture from the technical variations found in different data sources. One example is the use of the standard Open Database Connectivity (ODBC) API for accessing database management systems (DBMS). An application written using ODBC can be ported to other platforms, both on the client and server side, with few changes to the data access code. Similar other abstractions have allowed us to migrate between applications as well as versions of applications in a matter of weeks rather than months or years.

Additionally, the CODR architecture consists of two complementary software components. The first component is database specific and is referred to as *Application Listeners*. These listeners are lightweight, stateless, and integrate with the databases via standard APIs and stream data in parallel to and from secondary storage with no choke point.

The second database-independent component is the *Scale-Out Software Platform* that orchestrates data movement, computes consistent space-efficient backups, and orchestrates recovery operations while maintaining native formats. The Scale-Out Software Platform reads the data that the Application Listeners have transferred to the file or object in secondary storage and processes the data to make it ready for recovery.

Rubrik Mosaic isolates functionality of database-specific logic in the Application Listeners and uses abstractions in a manner such that the core CODR engine is independent of the application. This leads to a pluggable architecture in which the support for a new data store can be encapsulated in a plugin.

THE FOUNDATION OF RUBRIK MOSAIC DIFFERENTIATION

The CODR architecture is the technology foundation that enables Rubrik Mosaic current and future products to deliver unique benefits not found in any other solution. Let's look at a couple of these benefits.

Simple Deployment

You can deploy Rubrik Mosaic IO products in private, public, and hybrid environments with equal ease.

Lower Total Cost of Ownership (TCO)

CODR enables tremendous operational efficiency and cost savings on several fronts:

Elasticity
You can scale up and scale down Rubrik Mosaic software as needed depending upon workload requirements.

No Media Servers
The elimination of media servers dramatically reduces reliance on expensive, proprietary hardware.

No Proprietary Hardware or Appliances
Rubrik Mosaic requires no proprietary hardware or appliances. Instead, you simply take advantage of the massive economies of cloud-scale infrastructure.

Unparalleled Storage Efficiency
Semantic-deduplication delivers 10 times the storage efficiency of traditional de-duplication, across whatever cloud storage you choose to use, resulting in orders of magnitude cost savings.

No Cloud Lock-In
Rubrik Mosaic functions completely in the control plane, and all data protected or managed remains in native format. This eliminates any reliance on underlying infrastructure and provides complete cloud flexibility.

Rich, Granular Data Services
Another benefit of application centricity and preserving native formats is that data services can be delivered in-place at a granular, application-specific level, providing both flexibility as well as rich insight.

CONCLUSION

For the growing world of multi-cloud data sources to become a reliable component of modern businesses and institutions, Rubrik Mosaic has built the next-generation CODR architecture. CODR solves important technical challenges to provide cross-cloud data protection and data mobility for both traditional and next-generation applications.

rubrik

Global HQ
1001 Page Mill Rd., Building 2
Palo Alto, CA 94304
United States

1-844-4RUBRIK
inquiries@rubrik.com
www.rubrik.com

20190704_v2