

VMWARE VSAN AND RUBRIK MOSAIC FOR NOSQL DATA MANAGEMENT

Solution Overview

Contents

- NoSQL Has Gained Enterprise Momentum 1
- Data Management for Non-Relational NoSQL Databases 1
- About Rubrik Mosaic 2
- NoSQL Replication Is Not Backup 2
- NoSQL Data Management Challenges 3
- Why Rubrik Mosaic for NoSQL Data Protection on vSAN environments? 3
- Deployment Architecture 4
- Summary..... 5

Staying ahead of the competition means monetizing data, building apps that engage more customers, and creating new e-commerce applications that accelerate sales and reduce costs. These projects define the digital transformation that's fueling growth in many organizations. Success is measured by improved time to market, lower development and infrastructure costs, smoother app deployment and reduced reliance on legacy technology and vendor lock-in.

NoSQL Has Gained Enterprise Momentum

The digital transformation is driving a shift from monolithic apps deployed in traditional data centers supported by dedicated storage, to next-generation, cloud-native applications that support big data architectures, microservices and containers. These apps are increasingly born in the cloud and are designed to shift among cloud providers to circumvent vendor lock-in. Organizations have turned to DevOps, agile development, and the cloud as key parts of their digital transformation.

Along with the cloud, DevOps are aggressively adopting the new generation of non-relational or NoSQL databases like Apache Cassandra, DataStax Enterprise (DSE) and MongoDB. These databases have stepped-up performance and built-in availability with standard multi-node replication. NoSQL databases are redundant *and* they're fast. They are scalable, less complex and don't require the deep expertise that's needed to tune and scale traditional SQL databases. Their scale-out architecture and adaptability make them a natural fit for massive applications deployed in the cloud.

Data Management for Non-Relational NoSQL Databases

Data management for databases isn't what it used to be either. Virtualized data management software coupled with purpose-built backup storage hardware displaced tape over a decade ago. Today, that last-generation siloed compute and storage scheme is being replaced by a new generation of Hyper-Converged Infrastructure (HCI) deployed in the cloud or on-premises.

VMware vSAN™ is the market leader in Hyper-Converged Infrastructure (HCI), enables low cost and high-performance next-generation HCI solutions that virtualize physical infrastructure to help customers easily evolve their infrastructure without risk, improve TCO over traditional resource silos, and support new hardware, applications, and cloud strategies. The natively integrated VMware infrastructure combines radically simple VMware vSAN storage, the market-leading VMware vSphere Hypervisor, and the VMware vCenter Server unified management solution all on the broadest and deepest set of HCI deployment options.

VMware vSAN is a storage platform that leverages existing, on-premises, low-cost infrastructure to produce a high-performance, scalable storage system. The flexibility of direct-attached storage and the ability to deploy many HDD devices ensures a storage subsystem that can deliver optimal throughput. With high performance at

lower costs, organizations can now deploy critical databases and applications on VMware and vSAN platforms, and more importantly, there's an increasing trend of enterprises adopting and deploying NoSQL databases on the VMware vSAN platform^{1, 2}.

NoSQL databases, characterized by high performance and availability, end up sacrificing consistency to achieve the speed and redundancy. The fundamental characteristics of NoSQL databases also include a shared-nothing architecture, highly compressed data sets, and multiple replicas of the data. NoSQL supports very large data sets and elastic deployments. Given the hyper-scale, distributed nature of NoSQL databases, traditional backup and recovery products (volume-based backups, VM-based backups, et al.), are unable to protect these modern applications, leaving organizations exposed to data loss and extended downtime.

In addition to essential backup and recovery, NoSQL database deployments typically require additional data management capabilities that are missing from traditional backup products. The need to support test data with frequent refreshes as well as supporting data mobility among different clouds or among alternate database topologies have emerged as important requirements.

About Rubrik Mosaic

Rubrik Mosaic is an application-centric data management software that provides backup and recovery for NoSQL, non-relational enterprise databases such as Apache Cassandra, DataStax Enterprise, and MongoDB. Rubrik Mosaic delivers scalable backup, orchestrated and streaming recovery, and the industry's first semantic deduplication that reduces secondary storage costs up to 80% in the cloud or on-premises. Rubrik Mosaic includes any point-in-time backup and orchestrated recovery to safeguard against logical and human errors, malicious data corruption, application schema corruption, and other software errors.

Rubrik Mosaic is a software platform that can be deployed in any form factor, such as VMs on-premises, cloud-native instances in any cloud provider, or as Docker containers. Rubrik Mosaic does not store data but sends it directly from the database clusters to the backup storage target.

NoSQL Replication Is Not Backup

NoSQL databases trade consistency in order to get performance and availability. With native replication, one or more nodes can suffer hardware or network faults, or be taken offline while the database and application

¹ Deploying MongoDB on VMWare vSAN Environment:

<https://blogs.vmware.com/virtualblocks/2017/07/27/deploy-manage-run-mongodb-on-vmware-vsan/>

² DataStax Enterprise on VMWare vSAN Environment:

<https://www.vmware.com/content/dam/digitalmarketing/vmware/en/pdf/products/vsan/vmw-datstax-enterprise-cassandra-on-vsan.pdf>

continue to perform at scale. While replication solves the availability problem, it isn't designed to solve disaster recovery or human error, both of which are inevitable. Replication guarantees that data corruption, accidental or malicious deletion, ransomware, and other human-caused events will be replicated across all nodes. And NoSQL's eventual consistency makes it a challenge to get a durable point-in-time backup. With next-generation apps, recovery time becomes critical when application downtime means customers get locked out and business grinds to a halt.

NoSQL Data Management Challenges

DevOps and DBAs have also identified the need to leverage backups to accelerate the deployment of test data to accelerate application development. This introduces a new set of data management requirements that include the ability to query and quickly restore selected data, to mask sensitive PII and PCI data, and in particular, to restore data to dissimilar database topologies. Organizations are increasingly looking for an all-inclusive data management solution for large NoSQL databases.

Four key data management problems need to be solved in order to deploy NoSQL databases into a mission-critical app. First and foremost, the solution needs to be application-consistent in order to manage recovery point and recovery time objectives (RPO and RTO) for hundreds of terabytes and hundreds of nodes. Second, it needs to be simple to use as well as reside in any environment—on premises HCI or cloud-native, and it needs to be able to migrate data among clouds and then restore data to dissimilar databases. Third, the solution needs to deal with multiple copies by storing them as a single deduplicated data set in order to avoid out of control long-term storage and compliance costs. Finally, the solution needs to automate or orchestrate disaster recovery and the data duplication process in order to relieve operations and dev teams from the data management burden.

Rubrik Mosaic solves these problems with a cloud-native solution that was engineered to support massive cloud workloads and cloud archival storage, just as the NoSQL databases it protects were designed around scale-out architecture. This innate cloud focus means Rubrik Mosaic can be installed in the cloud, protect NoSQL workloads in the cloud, and store backup data in low-cost archives. By removing the limitations of traditional or vendor-provided native backup solutions, Rubrik Mosaic helps accelerate the digital transformation, adoption of DevOps, and more.

Why Rubrik Mosaic for NoSQL Data Protection on vSAN environments?

Rubrik and VMware have jointly undertaken this white paper to demonstrate vSAN as the HCI platform for NoSQL databases and Rubrik Mosaic for NoSQL backup/recovery and data management. Rubrik Mosaic has achieved VMware Ready status, VMware's highest level of endorsement to provide application-centric data protection for cloud-native applications and distributed NoSQL databases. Organizations can now leverage the

agility, flexibility, and scalability of deploying cloud-native applications on VMware vSAN platform with the continuous data protection capability of Rubrik Mosaic software. With this solution, organizations can scale business-critical applications on virtualized NoSQL databases and be confident in the recoverability of data and the ability to maintain high application uptime.

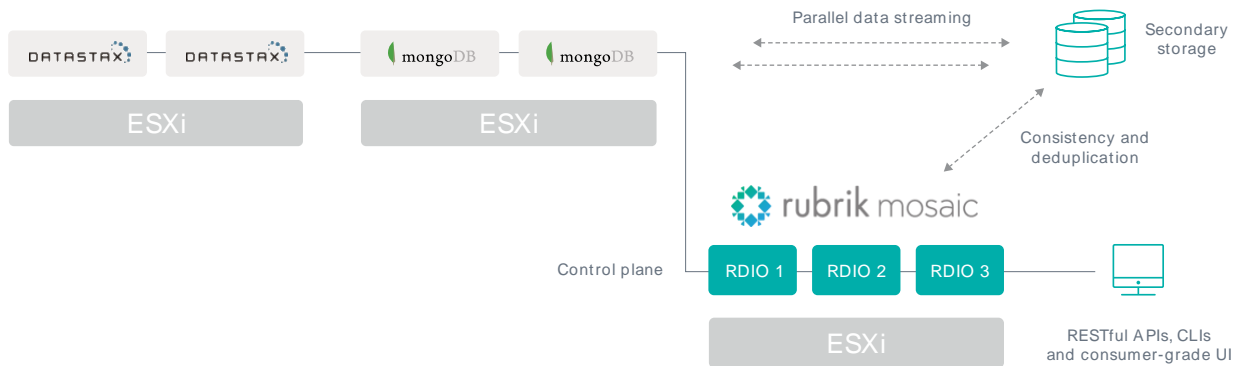


Figure 1: Reference deployment: Rubrik Mosaic in VMware environment

Deployment Architecture

Figure 1 depicts a typical reference architecture for Rubrik Mosaic in a VMware infrastructure (VMware ESX, VMware vSAN, etc.). Rubrik Mosaic relies on a shared storage backend for streaming the data from the NoSQL data stores onto secondary storage. The storage objects provided by vSAN will be mounted as a Virtual Machine Disk (VMDK) on all of the virtual machines running Rubrik Mosaic that acts as the control plane to perform all data management operations of backup/recovery, test/dev, data automation, and data mobility.

Rubrik Mosaic relies on a stable and reliable secondary storage backend to store the logical backups of the NoSQL databases. Rubrik Mosaic is a mostly stateless design, wherein the metadata and the catalog for the backup is itself backed up on the secondary storage provisioned for backup. Using the native replication support provided by vSAN, the cluster consistent, de-duplicated, logical backups produced by Rubrik Mosaic can be replicated to an offsite location at a reduced cost.

For cloud-native applications deployed on the VMware HCI platform, there are several configurations possible, all of which are customizable based on the workload requirements. The typical deployment will have a VMware vSAN-based object used to provide the storage object required for the individual NoSQL data stores. In such environments, Rubrik Mosaic can be deployed in one of the following models:

- Rubrik Mosaic sharing compute and storage in the same VMware vSAN HCI platform
 - The storage where the backups will reside can be located in either a different vSAN platform or dedicated secondary storage (NFS, Object Storage, et al.). This optimizes the failure modes between the backup storage and the database data.
- Rubrik Mosaic is deployed on a different VMware vSAN HCI cluster than the cluster where the primary NoSQL data store is deployed.
 - In this architecture, Rubrik Mosaic and the backup storage can share the same storage plan that is used for the compute cluster. This increases the overall performance by localizing the storage with the compute for Rubrik Mosaic.

Rubrik Mosaic can still be deployed to use external object storage as the backup target. Rubrik Mosaic also supports NFS or any other shared storage as backup targets. Any of the above deployments can make use of a storage object that provides multiple zone reliability and fault tolerance.

One key advantage of using vSAN-based storage as the backup target for Rubrik Mosaic is the ability to leverage the native replication of vSAN to provide tolerance against additional full disaster recovery scenarios.

Summary

Deploying and running NoSQL databases on the VMware vSAN platform is a robust enterprise solution. However, NoSQL databases being distributed in nature, being eventually consistent in nature, having highly compressed and large-scale datasets -- all pose a significant challenge to existing data protection practices. Application consistent backups for NoSQL databases on the vSAN platform are required. Additional requirements include a modern data management solution with orchestrated and streaming recovery operations, alternate database topology restores for test/dev and data automation needs, application-aware deduplication, and support for key test/dev and data mobility use cases. Rubrik Mosaic and VMware deliver all the necessary components of a modern data management solution for NoSQL databases when deployed on virtualized infrastructure.

