



WHITE PAPER

Solving Data Management Challenges for NoSQL Databases

TABLE OF CONTENTS

3	INTRODUCTION
4	RUBRIK MOSAIC: HOW IT WORKS
5	RUBRIK MOSAIC: FEATURE DIFFERENTIATION
8	RUBRIK MOSAIC: IMPACT IN THE ENTERPRISE
10	CONCLUSION

INTRODUCTION

The non-relational NoSQL database movement has gone mainstream. Modern applications built on top of NoSQL databases—such as DataStax, Apache Cassandra, MongoDB, Amazon DynamoDB, Google BigTable, and others—are no longer experiments from Silicon Valley companies. Mainstream enterprises across all industries are now adopting these scale-out, non-relational databases for customer-facing applications.

Organizations invariably choose NoSQL databases for their unique capabilities, such as native database replication, developer-friendly usability, and built-in sharding capabilities for high-volume and large-scale data sets. In other words, they want their developers to do as much and move as fast as any startup without incurring the costs associated with typical relational databases such as Oracle. When it comes to mainstream adoption of NoSQL, organizations need to consider the following issues:

- Enterprise-grade data protection
- Business continuity
- Testing and development
- Migration to and across clouds
- Regulatory compliance

The hard reality is that organizations face new and immediate data protection challenges with these modern NoSQL data sources. Organizations have come to realize that native database replication is not a substitute for backup and certainly doesn't help meet test/dev and data mobility requirements.

For veteran infrastructure and IT operations professionals, who work on the frontier for those mainstream problems, the need for backups is nothing new. But backup and recovery for NoSQL data stores is a new challenge in itself. After all, NoSQL scale-out and distributed databases favor speed and scale over eventual consistency. Furthermore, these distributed databases have the following characteristics:

- Are deployed on shared-nothing architectures
- Can scale out horizontally or scale back based on dynamic application workloads
- Replicate every write across the database cluster
- Contain data sets that are highly pre-compressed

The result is that mainstream enterprise applications built on NoSQL databases need to view any downtime or data corruption as a threat to the business. A reliable NoSQL backup solution is now a necessity for enterprise organizations.

Rubrik Mosaic is a purpose-built software product that provides durable and reliable backup and recovery for NoSQL databases such as Apache Cassandra (including DataStax Enterprise) as well as MongoDB. Rubrik Mosaic offers application-consistent NoSQL protection, industry-leading features (such as semantic deduplication that drastically reduces backup storage costs for pre-compressed NoSQL data sets that are triple-replicated across database clusters), and modern deployment models (such as container, virtual machine, and compute instance) that reduce complexity.

In this paper, we first present the basics of the Rubrik Mosaic architecture. Then we cover the feature differentiations in these three areas:

- Application-consistent data protection
- Cloud-native backup
- Storage cost control

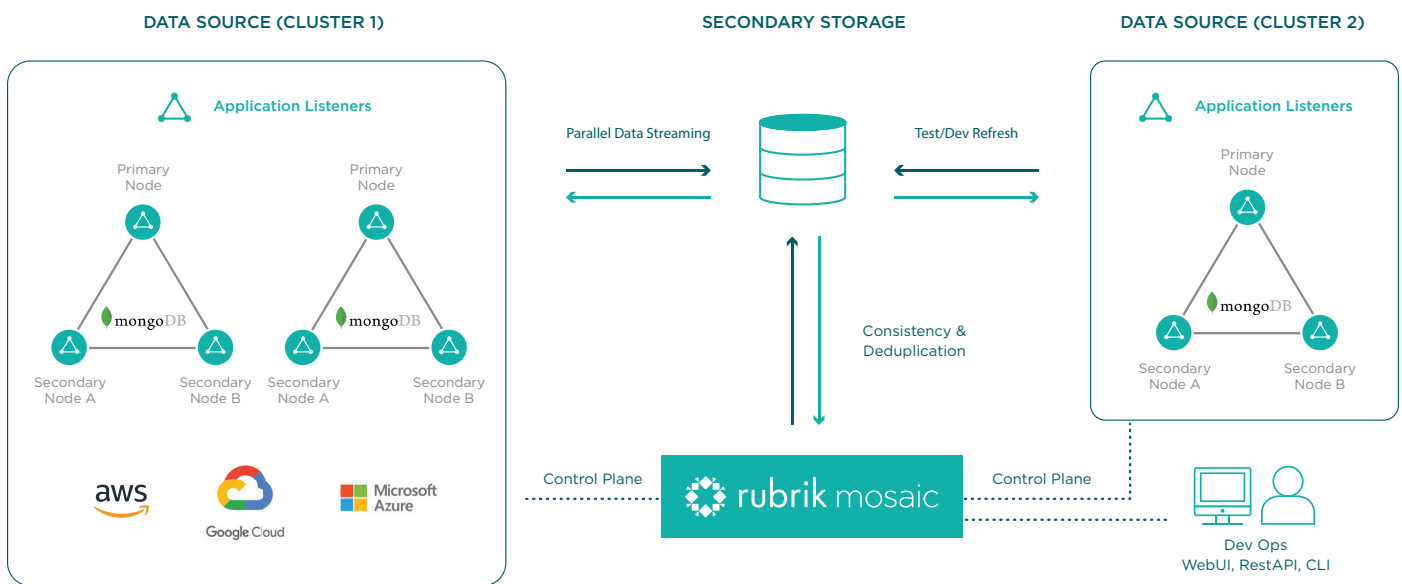
We close by discussing how modern application, database, and IT teams at well-known Fortune 500 and Global 2000 organizations use Rubrik Mosaic, both to protect their NoSQL applications and achieve their larger strategic priorities for data center modernization and digital transformation.

RUBRIK MOSAIC: HOW IT WORKS

Rubrik Mosaic is a purpose-built backup and recovery software designed specifically to solve the challenges of backing up modern NoSQL databases and big data file systems. Just as NoSQL databases operate differently from traditional scale-up relational databases, Rubrik Mosaic works differently from traditional backup-and-recovery solutions. In this section, we'll talk about the architecture of our software-only, scale-out data-management solution.

Reference Architecture

Rubrik Mosaic for MongoDB Backup and Recovery



First, Rubrik Mosaic is a software-only, control plane, data-protection solution that never stores any of your database data. Traditionally, storage companies sold backup and recovery software as a way to fill (and sell) secondary disk storage. Vendors were interested in creating as many secondary replicas of a production database system as they could in order to sell more storage. It was left to the operations and admin teams to figure out where the backup data ended up and how many duplicates they had on hand. This governance problem was incredibly difficult to deal with. Recovery was often a multistep process, and the vendor solution always involved purchasing more and more licenses, systems or both.

Unlike traditional appliance-based backup products, Rubrik Mosaic operates simply as a control plane software product, with a modern user interface, a scalable backup and versioning engine, orchestrated recovery, and industry-first semantic deduplication—all built natively into the software. Rubrik Mosaic is designed to help organizations perform backups to inexpensive commodity storage, either in the cloud (via Amazon Web Services, Google Cloud Platform, Microsoft Azure, or Oracle Cloud) or in their data center (via NFS or AWS S3-compatible object stores like Cloudian). Your production database streams data directly into your secondary backup storage of choice, backups are always stored as native formats, and Rubrik Mosaic handles the rest of the data-protection operations, with minimal impact to your production system performance.

Second, Rubrik Mosaic is designed to scale out, just like modern NoSQL databases scale out. Users have a choice of deploying it as a one-node, three-node, or five-node software cluster. This means that software can be both fault-tolerant and elastic—that is, it can be expanded or contracted on-demand as your production environment grows and your application demands increase or decrease. It also means that

Rubrik Mosaic can facilitate faster backup-and-recovery operations for large-scale NoSQL databases. Although Rubrik Mosaic does not hold data, as the source of truth for versions and deduplication it fully orchestrates application-consistent backups and all recoveries.

Third, Rubrik Mosaic was created to provide a single, customizable policy-driven dashboard, with all consistency, backup, and recovery options. When handling highly regulated data, or when meeting service level agreements (SLAs) that require rapid recovery of large datasets, Rubrik Mosaic makes the lives of application teams, DBAs, and DevOps staff much easier.

At the simplest level, Rubrik Mosaic accomplishes all this on the backup side with its lightweight, stateless NoSQL Service Managers, which are installed on the source database nodes automatically to track changed data and minimize the amount of data transferred back and forth. On the recovery side, Rubrik Mosaic accelerates matters by orchestrating the recovery of data stored as individual, deduplicated objects. You can restore an individual object directly from the secondary storage, or perform entire database restores in parallel to all nodes in a production cluster or an alternate cluster used for testing and development. Furthermore, you also have the flexibility to restore only data required to meet a specific query, data required to meet a specific time range, or data at the collection or column family level.

RUBRIK MOSAIC: FEATURE DIFFERENTIATION

Now that we've covered the architecture, let's move on to the specific features and functionality that differentiates Rubrik Mosaic from other solutions for NoSQL data protection.

You'll discover numerous benefits to using Rubrik Mosaic:

- Simplified application-consistent backup and recovery architecture
- Improved user experience resulting from our modern scale-out data-management software
- Ability to refresh production data for test/dev use cases
- Drastically reduced Recovery Point Objectives (RPO) and Recovery Time Objectives (RTO) for mission-critical production applications

Meet Rubrik Mosaic

Application-Consistent Backup and Recovery for NoSQL databases, Cassandra and MongoDB

- **Applicated-Consistency:** Deliver sharded cluster backups & repair-free recoveries → fast RTOs
- **Simplicity at Scale:** Eliminate complex home-grown / native backup tools → DBAs & DevOps
- **Semantic Deduplication:** Unleash up to 70%+ in secondary storage savings → Hard ROI



These benefits are the result of our feature differentiation. Rubrik Mosaic's distinct features fall into three categories:

- Application-consistent data protection
- Cloud-native data-management
- Storage footprint and cost controls

In each of these categories, we provide features that directly benefit infrastructure professionals, operations teams, and architects tasked with protecting NoSQL data across an enterprise. We explain how these features work and why they're far superior to built-in tools in databases or traditional storage-centric backup solutions.

APPLICATION-CONSISTENT DATA PROTECTION

Ensuring consistency has always been a challenge for NoSQL databases. Even as these products evolve and offer new settings, consistency guarantees come with a cost to read and write performance. For many admins running production systems, these tradeoffs are well understood.

In the backup and recovery world, however, consistency is expected, even from NoSQL. This is a different kind of consistency than the ACID transactions missing in most NoSQL products today. When infrastructure professionals refer to “application-consistent backups,” they mean that the backup will reflect the data as it was viewed by the application at a given point in time. This functionality is important for various legal, regulatory, and business continuity reasons. And for a larger NoSQL database that might have many nodes or many master nodes making writes, there is an innate complexity in guaranteeing application consistency.

Rubrik Mosaic guarantees this type of consistency and more. Two features fall under our “consistent data protection” umbrella:

- **Application consistent backup.** Rubrik Mosaic offers application-consistent backups by processing source data in secondary object storage. The initial backup occurs by taking a consistent, logical version of data stored in the production database and migrating it to an object or to file storage (the “version storage”). Rubrik Mosaic looks at all data migrated to the version storage and defines an application-consistent version based on the writes that have reached quorum across the nodes at the time the backup was taken. This orchestration is performed to minimize the chances that a point-in-time backup of a multinode NoSQL system would reflect inconsistent writes that have not fully replicated. Rubrik Mosaic then uses that consistent backup for recovery, providing application consistency without the need to perform repairs.
- **Faster incremental backups.** Once the initial full backup to version storage is performed, Rubrik Mosaic NoSQL Service Managers transfer only the incremental changes from the previous version. This approach significantly reduces the burden on the production system compared to snapshot-based or full-backup-based solutions. With far fewer full backups, chances are greatly reduced that a full backup will occur when nodes have failed or when data is inconsistent. This approach also makes it easier to meet RPO requirements measured in hours.

Rubrik Mosaic Product Benefits



App-Consistent Backups

Deliver app-consistent backups with a policy-driven approach. NoSQL database backups are always made available in native formats.



60–80% Hard Savings

Reap up to 80% savings by reducing secondary storage costs in the cloud or on-prem with patented semantic deduplication.



Development Efficiencies

Refresh production data with any-to-any topology restores. Address enterprise use cases—Compliance, Governance, GDPR, Ransomware, and DR.

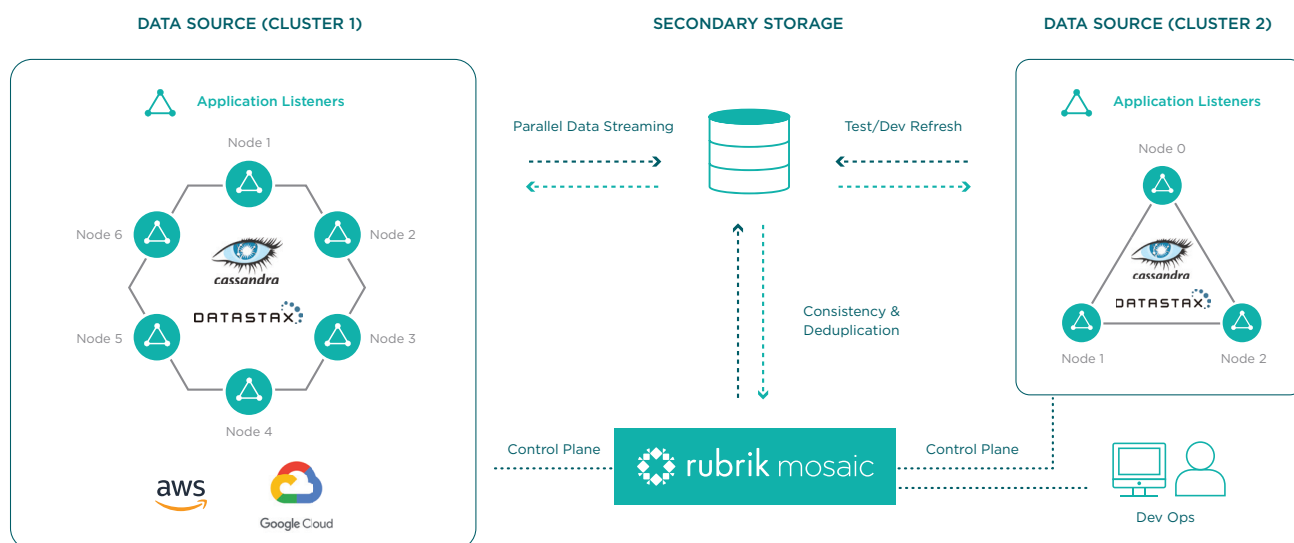
CLOUD-NATIVE DATA-MANAGEMENT SOFTWARE

Just as NoSQL databases have changed application development and data management, so too has cloud computing changed software and hardware infrastructure. Large IT organizations recognize the potential of the cloud, and whether they’re creating a private cloud, choosing a public cloud, or building a hybrid cloud strategy, it is essential that new components are cloud-compatible.

Rubrik Mosaic is not just cloud compatible, it’s cloud native. From the very start, Rubrik Mosaic was engineered to support cloud workloads and cloud archival storage, just as the NoSQL databases it protects were created around scale-out architecture. This innate cloud focus means Rubrik Mosaic can be installed in the cloud, protect MongoDB and Cassandra workloads in the cloud, and store backups in low-cost cloud archives. It’s right at home when NoSQL databases are taking full advantage of the cloud, whether they’re scaling out or scaling in.

This also means IT teams that run production MongoDB or Cassandra databases in the cloud will never have to worry about backup. Two features make this possible:

- **Elastic backup and recovery.** One benefit to NoSQL is the ability to scale out when demand is high—for example, during seasonal peaks in retail activity—and to scale back in when demand is low. Most vendor-provided and traditional backup offerings require the topology, or number of nodes and configuration, to be the same at the time of each backup. This requirement limits one of the greatest benefits of running in the cloud. Because Rubrik Mosaic orchestrates backups to create one consistent view based on a quorum of nodes, you can create incremental-forever backups for an individual Cassandra or MongoDB cluster as it scales out—say, from 11 to 13 nodes—or when it scales back in. This elasticity does not require teams to scale to a certain number of nodes for each backup. It also enables Rubrik Mosaic to recover to different topologies—say, to recover production data to a three-node test cluster, or to recover a seven-node Cassandra backup to a 13-node cluster that has grown to accommodate a spike in demand. Performing cross-topology restores in the cloud are easier than ever.



- **100% scale-out.** As a scale-out software solution, Rubrik Mosaic is built to scale to many nodes and to perform even when nodes fail. Unlike other vendor solutions, Rubrik Mosaic has no media server, no separate single master node, and no single point of failure. And as a solution built to archive to the cloud, it can easily let NoSQL ops and admin teams adopt easy-to-scale cloud storage, without any need to maintain a complete duplicate architecture for each of their backups or secondary architectures.

STORAGE FOOTPRINT AND COST CONTROLS

Managing backups for larger-scale, growing NoSQL systems can also involve managing lots of high-cost proprietary storage. With NoSQL databases, many copies of any one piece of data are already stored across a cluster. An individual key-value pair might be duplicated a half-dozen times to ensure it remains available should any one commodity disk fail, or any one node go offline; it also might accelerate read performance when there are many nodes that can retrieve a single piece of data.

With all this duplicate data characteristic of NoSQL databases, there is a need to minimize the storage and backup footprint, and to remove duplicates as quickly and efficiently as possible for RPO and RTO purposes. This area is where the following Rubrik Mosaic product features come into play:

- **CODR.** This is the “secret sauce” behind Rubrik Mosaic. Consistent Orchestrated Distributed Recovery is an engine that removes duplicate data blocks in version storage and greatly reduces the need for large-scale full backups. With CODR, versioning, events, notifications, timestamps, and alerts are all tracked, and all that data is used to minimize the number of data replicas stored in lower-cost object or file storage. For many production systems, CODR generates at least a 60% reduction in storage required to hold the initial backups—and a similar reduction in cost. And with Rubrik Mosaic’s incremental-forever approach, multiday retention policies can produce even greater savings.

The example below shows sample of storage savings that are achieved for operational recovery use case of Apache Cassandra database:

- Database size (logical): 10TB (10% daily insert rate)
- Replication factor: 3x
- Versioning frequency: Daily (7-day retention)

	Secondary Storage Required (Without Rubrik Mosaic)	Secondary Storage Required (With Rubrik Mosaic)
Day 1 (Initial Sync)	30 TB	10 TB
Day 2 (Incremental)	3 TB	1.0 TB
Day 3 (Incremental)	3.3 TB	1.1 TB
Day 4 (Incremental)	3.6 TB	1.21 TB
Day 5 (Incremental)	4.0 TB	1.33 TB
Day 6 (Incremental)	4.4 TB	1.46 TB
Day 7 (Incremental)	4.8 TB	1.61 TB
Total Storage	53.1 TB	17.7 TB (3x savings)

- **Advanced compaction.** As enterprises continue to scale their Cassandra database usage, the database logically stores the same data using different, compacted SSTables. Vendor-provided software backs up these new SSTables without realizing that the data it stores is already backed up—and this duplicated effort can add up quickly in larger deployments. Rubrik Mosaic identifies the compacted SSTables that hold already-existing, unchanged data and does not create any additional backup workload. This behavior is another way Rubrik Mosaic looks to reduce the storage footprint and costs for larger databases with more complex backup SLAs.

RUBRIK MOSAIC: IMPACT IN THE ENTERPRISE

As we discussed in the previous sections, Rubrik Mosaic is architected specifically for NoSQL data protection, with differentiated features that provide application-consistent data protection for NoSQL databases via cloud-native software. It also provides orchestrated recovery at scale and with controls that reduce storage footprint and cost.

These benefits make Rubrik Mosaic a better, more flexible backup and recovery solution for database admins and DevOps teams. But there is more to it than that. By removing the limitations and concerns of traditional or vendor-provided database native backup solutions, Rubrik Mosaic eliminates obstacles standing in the way of enterprise initiatives like large-scale digital transformation, adoption of DevOps principles, and more.

Large enterprises that have adopted Rubrik Mosaic often see two stages of benefits: improved data protection and the acceleration of a major initiative such as cloud-native or digital transformation. The next section details how two large enterprises—one a Fortune 100 retailer, the other a top-4 wireless provider in the United States—benefited from adopting Rubrik Mosaic.

SIMPLIFYING BACKUP AND EMBRACING MICROSERVICES IN THE CLOUD AT A FORTUNE 100 RETAILER

At one of the world's largest retailers, the application teams chose Apache Cassandra/DataStax Enterprise to build several exceptional mass consumer-scale applications. Those same NoSQL databases also enabled the application team to experiment with a new, more modern approach to application development.

The application and data teams saw tremendous potential in the microservices architecture adopted by many Silicon Valley companies. Specifically, they decided to deploy microservices to power different elements of their customer-facing home page, including the product catalog, catalog search, the shopping cart, and other standard features. They built all this using public cloud infrastructure to make it easier to scale out specific services during times of peak demand.

The challenge came when trying to protect this valuable data especially because the application is running natively in the cloud. DataStax OpsCenter presented them with two serious challenges:

- They needed to duplicate their production architecture for each microservice, giving them an enormous and complex backup architecture.
- It limited the elasticity of the cloud. With OpsCenter, they could not perform incremental backups as they scaled, and they could not recover to different topologies for different workloads. This limited their ability to scale different home page components differently, and it delayed their adoption of microservices.

The [enterprise applications and database teams recognized the power of Rubrik Mosaic](#) to solve these two problems. First, a three-node software cluster could replace the numerous backup clusters they had, drastically simplifying their backup and recovery experience. Second, with Rubrik Mosaic, they could back up from and recover to alternate topologies for each of their services. This solution enabled them to scale their production site without ever having to think about the impact on backup and recovery—and they could fully adopt a microservices architecture while meeting data-protection requirements.

REDUCING CLOUD STORAGE COSTS AND RECOVERY TIMES WITH RUBRIK MOSAIC FOR CLOUD BACKUP, AND IMPLEMENTING RECOVERY FOR MAXWELL HEALTH'S SAAS-BASED APPLICATION BUILT ON MONGODB DATABASES IN THE AWS CLOUD

Maxwell Health is an HR benefits-management service that simplifies the process of creating and managing employee health plans, benefits, and payroll for small and medium-sized businesses. It delivers a cloud-based SaaS platform for customers that includes a marketplace to make selecting benefits intuitive for employees. The cloud-based SaaS platform runs on an Amazon Web Services public cloud, and it uses MongoDB databases for critical data requirements.

Maxwell Health's challenges are summarized in two broad categories:

- The first challenge is backup and recovery of MongoDB in the AWS public cloud. Maxwell Health built its online business application using a microservice-based, cloud-native application architecture. The microservices are deployed natively in AWS and use underlying MongoDB databases to populate information for different customer-facing platforms. Maxwell Health's core online product is based on this web-scale, cloud-based application, and any data loss is detrimental to its business. Maxwell Health requires the ability to back up multiple MongoDB databases and recover from any data loss with an SLA of two hours.
- The second challenge is performing test-cluster refreshes using production data. Maxwell Health has multiple production and test clusters, and it implements continuous integration and continuous development (CI/CD) methodologies.

[Maxwell deployed Rubrik Mosaic](#) to meet its cloud backup and recovery requirements. Rubrik Mosaic was deployed within 15 minutes using the AMI (Amazon Machine Image) in the cloud as a complete solution.

With Rubrik Mosaic, Maxwell Health achieved the following improvements in performance:

- A 90% reduction in AWS S3 backup storage costs
- A 30% reduction in recovery time compared to a native MongoDB backup solution
- A 300% ROI in the first 12 months
- A four-hour backup SLA
- A two-hour recovery SLA

Because Rubrik Mosaic offers fully orchestrated recovery back to any MongoDB database cluster with any topology, Maxwell Health can also restore its production data into test clusters in a fully automated manner.

CONCLUSION

Over the past decade, next-generation applications with modern NoSQL databases have gone mainstream in enterprises. The benefits of MongoDB and DataStax/Apache Cassandra have allowed developers to bring powerful customer-facing applications into production faster than ever. Some of those applications still tackle trendy topics like the internet of things and “data lake” analytics, but others are at the center of mainstream enterprise use, such as customer analytics, content management, and e-commerce. Others, still, are a part of major IT programs to transform and modernize infrastructure.

Regardless of how you use your NoSQL database, Rubrik Mosaic offers you an enterprise-grade data-protection solution for these modern database platforms. With its application-consistent backups, cloud-native software-driven approach, and semantic deduplication and storage-related cost controls, Rubrik Mosaic provides a purpose-built data management solution for the next generation of databases.



Global HQ

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

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

Rubrik delivers a single platform to manage and protect data in the cloud, at the edge, and on-premises. Enterprises choose Rubrik's Cloud Data Management software to simplify backup and recovery, accelerate cloud adoption, and enable automation at scale. As organizations of all sizes adopt cloud-first policies, they rely on Rubrik's Polaris SaaS platform to unify data for security, governance, and compliance. For more information, visit www.rubrik.com and follow [@rubrikinc](https://twitter.com/rubrikinc) on Twitter.

20190730_v1