

Next-Generation Backup and Recovery for MongoDB

Datos IO provides the industry's first cloud-scale, application-centric, data management platform enabling organizations to protect, mobilize, and monetize all their application data across private cloud, hybrid cloud and public cloud environments.

To learn more, visit www.datos.io



Key Benefits

Application-Centric

- Application and cluster-consistent point-in-time backups
- Scalable versioning supports large clusters
- Flexible deployment with API-based architecture and native UI interface

Recover in Minutes, Not Hours

- Single-click fully orchestrated recovery
- Recovery to same or different size clusters
- Granular recovery for fastest RTO

Increase Ops Efficiency

- Semantic deduplication cuts backup storage requirements up to 80%
- Automated test/dev refresh
- Use backup copies for migrations, test/dev and database cloning



The Challenge

Businesses are in the midst of a digital transformation journey. According to research from IDC, 70 percent of CIOs have a cloud-first strategy. They want to harness the power of the cloud to drive growth by delivering new customer-centric products and services, while also driving greater operational efficiency. To handle the data requirements of these modern high-volume, high-ingestion rate and real-time applications, enterprises are turning to scalable, non-relational databases such as MongoDB rather than traditional scale-up database and storage approaches.

However, this fundamental shift raises critical issues in the lifecycle of data management and data protection. Traditional backup and recovery products were originally designed for small-scale databases, tape-based storage media, and legacy on-premises architectures. This leaves modern applications built on non-relational databases and Big Data filesystems exposed to data loss and downtime.



The Solution: Datos IO RecoverX

Datos IO RecoverX is the industry's first and only scale-out data protection software solution to deliver scalable and reliable backup and recovery for modern applications built on MongoDB databases. With RecoverX, enterprises can deploy business-critical applications on MongoDB and be confident in the recoverability of data and the ability to maintain high application uptime.



Features and Benefits

Datos IO RecoverX is built to address the data protection needs of modern, cloud-native applications deployed on MongoDB. Unique capabilities include:

Scale-Out Architecture

Datos IO RecoverX is founded upon Consistent Orchestrated Distributed Recovery™ (CODR™), Datos IO's cloud-first, scale-out data management architecture that enables customers to meet their data protection requirements for MongoDB. CODR™ uses elastic compute services that can be auto-scaled with load, and remove the dependency on media servers. CODR™ also transfers data in parallel to and from file-based and object based secondary storage for multiple use cases, including data protection and test/dev refresh. To simplify the data recovery process and to avoid vendor lock-in, protected data is stored in the database native format.

Continuous Backup

By using native application intelligence, RecoverX creates point-in-time consistent backup copies of MongoDB collections (both sharded and unsharded) at user-specified intervals; a concept called cluster-consistent versioning. Datos IO RecoverX can produce these cluster-consistent versions across all shards without quiescing the MongoDB database. Backups can be generated at a user-specified time interval and at any granularity (collection-level or entire database), providing operational ease of use to database

administrators. And with RecoverX, backup operations are resilient to failovers (primary switch) and failures (node).

Fully Orchestrated and Granular Recovery

Datos IO RecoverX provides fully orchestrated, any-point-in-time recovery. Granular recovery can be selected based upon either time or query for optimal RTO/RPO and to support governance requirements like GDPR. MongoDB collections can be recovered directly back into the same MongoDB database (operational recovery). They can also be recovered to a different MongoDB database instance (e.g. test/dev refresh) with a different topology where the number of nodes on the destination cluster differs from the node count of the source cluster.

RecoverX supports all combinations of recovery—sharded to unsharded, unsharded to sharded, sharded to sharded—thus reducing the operational burden of refreshing test/dev clusters for continuous development DevOps environments. Further, the recovery process deals only with the logical data, making it three times faster than traditional approaches. During recovery, the data is directly transferred from secondary storage into target databases, resulting in the lowest possible RTO.

Semantic Deduplication

RecoverX includes semantic deduplication, an industry-first capability that reduces the cost of storing backups of distributed databases over their retention period. These space-efficient backups dramatically reduce the overall storage footprint resulting in up to an 80% reduction in backup storage requirements.

Infrastructure and Storage Independence

Datos IO RecoverX is elastic-compute software that can be deployed on a physical server, a virtual machine, or any cloud compute instance (e.g. Amazon EC2). Data can be backed up to any NFS or object storage on-premises or in a public cloud (e.g. Amazon S3). In addition to CLIs and RESTful APIs, customers can use the RecoverX consumer-grade UI to manage their data protection environment.



Datos IO for MongoDB: Compatibility Matrix

MongoDB Database Supported	MongoDB 3.0, 3.2, 3.4		
Deployment	On-premise	AWS	Google Cloud
Secondary Storage Type Supported	On-premise NFS Storage	AWS S	Google Cloud Storage
Datos IO Software Node	RHEL/Centos 6.x		
	8-core, 32GB memory	EC2 M4.2xlarge or above	Standard 8 vCPU, 30GB RAM
	256GB Local Storage (SSD)		



Datos IO for MongoDB: Reference Architecture

