

Next-Generation Backup and Recovery for Cassandra

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 Cassandra and DataStax 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 Cassandra and DataStax databases. With this solution, enterprises can deploy business-critical applications on Cassandra and DataStax 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 Cassandra and DataStax. Unique capabilities include:

Application-Consistent Backup and Recovery

By working at the application layer, Datos IO RecoverX provides a true point-in-time backup copy of eventually-consistent distributed databases — a concept that we refer to as **cluster-consistent versioning**. An application listener captures data before it is distributed on the cluster, ensuring consistency.

Unlike other solutions or script-based approaches, Datos IO RecoverX can produce this cluster-consistent version without quiescing the database, improving application performance and reducing downtime.

Datos IO RecoverX can scale to support very large clusters, and allows you to backup the database at any interval and at any granularity. The backups are cluster-consistent and application-consistent, incremental forever, and always maintained in native formats on backup or secondary storage.

Fully Orchestrated and Granular Recovery

Orchestrated and reliable recovery lets you restore for operational recovery and for test/dev use cases, both on-premises and across cloud boundaries. All recovery operations result in repair-free restores, leading to reduced application downtime.

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. Data can be recovered directly back into the same database (operational recovery), or can be recovered to a different 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.

Because the backups are deduplicated, the recovery process only deals with logical data, making it at least 3x faster than traditional approaches, which results in significant reduction in 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 Cassandra: Compatibility Matrix

Apache Cassandra	2.0, 2.1		
DataStax Enterprise	4.5, 4.6, 4.7, 4.8		
Deployment Type	On-premise	AWS Cloud	Google Cloud
Storage Type	NFS	AWS S3	Google Cloud Storage
RecoverX S/W Requirements (per node)	RHEL/Centos 6.x	RHEL/Centos 6.x	RHEL/Centos 6.x
	8-core physical or virtual machine	EC2 m4.2xlarge EC2 c3.2xlarge	Standard Compute Engine 8vCPUs, 30GB
	16GB Memory	EC2 m4.2xlarge EC2 c3.2xlarge	Standard Compute Engine 8vCPUs, 30GB
	128GB Local Storage (SSD)	128GB EBS or Local (SSD)	128GB (SSD)



Datos IO for Cassandra: Reference Architecture

