



SOLUTION BRIEF

Clumio for Apache Iceberg on AWS

Iceberg-aware, air-gapped backup and recovery for AWS
data lakehouses

Modern AI and analytics initiatives rely on data lakehouses built on Apache Iceberg, one of the fastest-growing table formats on Amazon S3. But while Iceberg offers powerful versioning for analytics, native tools were not designed for enterprise-grade resilience. Organizations face real risks from data loss, corruption, account compromise, and compliance gaps—especially as data volumes and AI workloads scale.

Clumio for Apache Iceberg on AWS delivers the industry's first Iceberg-aware, air-gapped cyber resilience solution, providing purpose-built protection, rapid recovery, and long-term snapshot retention for your most valuable data assets.

THE CHALLENGE

Iceberg-based data lakes help accelerate AI and analytics, but their native protection mechanisms fall short for cyber resilience:

- **Snapshots are not air-gapped** and live in the same AWS account—leaving them vulnerable to ransomware, account compromise, or accidental deletion.
- **Retention is limited**, as accumulating snapshots degrade performance and force teams to prune history, impacting compliance.
- **Restores from generic S3 backups are slow and manual**, requiring a complex rebuild of metadata, manifests, and table state.
- **Data corruption or integrity issues** can cascade into failed AI models and analytics pipelines.

CLUMIO FOR APACHE ICEBERG ON AWS

Clumio provides a fully managed, serverless SaaS platform purpose-built to protect Iceberg-based data lakehouses. It understands Iceberg metadata, schema evolution, and snapshot lineage, helping support fast, reliable, and transactionally consistent recovery—without the need for manual reconfiguration.

Clumio protects both self-managed Iceberg tables in AWS Glue Data Catalog and fully managed Amazon S3 Tables, helping organizations safeguard the full lifecycle of their Iceberg data on AWS.

KEY BENEFITS

Iceberg-Aware Backup and Recovery

- Captures the full Iceberg table state—including metadata and manifests—to support fast, reliable, and transactionally consistent restores. Recovery options include point-in-time, snapshot-level, cross-region, cross-account, and in-place restore.

Air-Gapped, Immutable Protection

- Backups are isolated from the production AWS account in an immutable vault, helping protect against ransomware, accidental deletion, and account-level compromise.

Unlimited, Long-Term Retention

- Retain Iceberg snapshots for years without impacting live data lake performance. This helps organizations meet strict governance and regulatory requirements.

Enterprise Scale and Automation

- A serverless architecture with an incremental-forever backup approach scales on demand to support petabytes of data and billions of objects.

Tasks 3 Alerts

+ Filter ▾

Task type	Status	Source	Submit time	Start time	Duration	Action
Iceberg S3 tables SecureVault backup Task ID: 12345678	10% done	cse-prod AWS / ... / Apache Iceberg	02/12/2021 14:04	02/12/2021 14:08	1 min	retry
Iceberg Glue table SecureVault backup Task ID: 12345678	Completed	cse-prod AWS / ... / Apache Iceberg	02/12/2021 13:16	02/12/2021 14:04	47 sec	
Iceberg Glue table Snapshot restore Task ID: 12345678	60% done	cse-prod AWS / ... / Apache Iceberg	02/12/2021 13:13	02/12/2021 13:38	7 min	retry
Iceberg S3 tables Snapshot restore Task ID: 12345678	Completed	cse-prod AWS / ... / Apache Iceberg	02/12/2021 13:11	02/12/2021 13:33	7 min	

Restore table

Restore from

Snapshot ID
a81b0728-0ede-11ee-b9c6-0a69b620

Snapshot time
01/16/2024 12:00 am

Catalog type
S3 Tables

Table name
deployment_table

Table bucket
clumio-s3-bucket

Namespace
Database-prod-321

Size
20MiB

Schema
[View schema](#)

Restore to

Source
Restore to source table

Custom
Restore to a different account, region and to a specified table

ⓘ Restoration updates the metadata version without overwriting the current metadata.

AWS account	prod-my-aws-account
AWS region	US West (Oregon)
Catalog type	S3 Tables
Table bucket ARN	Select a table bucket ARN
Namespace name	Select a namespace name
Restored table	<input checked="" type="radio"/> New table <input type="text" value="Table name"/> <input type="radio"/> Existing table
Table optimization	<input type="checkbox"/> Compaction <small>Combine small data files into larger, more efficient files to optimize table performance.</small> <input type="checkbox"/> Snapshot retention <small>Optimize table storage by removing old snapshots from metadata overhead and expiring files that are no longer needed.</small>

[Back](#) [Cancel](#) [Restore](#)



WHAT ANALYSTS ARE SAYING

“

“The explosion of data lakehouse environments for AI has created a blind spot in many enterprises’ data protection strategies. Apache-aware data protection such as Commvault’s Clumio for Apache Iceberg directly addresses emerging governance risks with air-gapped, transactionally aware recovery that is becoming an imperative now. It’s a timely solution for a fast-rising market vulnerability.”

Archana Venkatraman
Senior Research Director, IDC

WHY CLUMIO

- Industry-first air-gapped, Iceberg-aware cyber resilience
- Protects Glue-managed Iceberg tables and Amazon S3 Tables
- Air-gapped, immutable backups outside your AWS account
- Point-in-time and snapshot-level recovery
- Unlimited retention without performance impact to live data lake
- Serverless, consumption-based SaaS—no infrastructure to manage
- Proven at hyperscale, including protection of S3 datasets with tens of billions of objects

GET STARTED

- Request a demo: <https://www.commvault.com/clumio/request-demo>
- Start in AWS Marketplace: <https://aws.amazon.com/marketplace/pp/prodview-hdul2n6jr5dou>

To learn more, visit [commvault.com](https://www.commvault.com)