

Rubrik Apache Iceberg Protection

Catalog-aware Backup and Recovery for Apache Iceberg tables on AWS

Apache Iceberg has become the layer that turns cheap object storage into a queryable lakehouse, making it the input layer for BI, real-time decisioning, and the AI/ML models enterprises now ship. It is the fastest-growing dataset on S3, yet it ships without built-in data protection and is rarely backed up any other way. Native Iceberg snapshots are metadata pointers — useless without the underlying data and metadata — and, like S3 versioning, they live in the same account as the data: not immutable, not off-site, and gone the moment a bad delete, a corrupted catalog, a compromised identity, or a rogue AI agent reaches them. And because they aren't catalog-aware, even when the files come back you're left with a pile of S3 objects, not a working table. How can you be sure a production Iceberg lakehouse comes back queryable — in hours, not months?



QUERYABLE RECOVERY, NOT JUST FILES BACK

Recover the whole table — data files, metadata, manifests, and the catalog pointer — automatically re-registered and immediately queryable in Athena, Spark, or Trino, with no manual catalog rebuild.



EFFICIENT AT PETABYTE SCALE

Back up only the latest compacted snapshot with forever-incremental change tracking and tiered storage, so protection stays efficient and affordable as the lake grows.



IMMUTABLE, AIR-GAPPED BACKUPS

Protect your most valuable data with immutable, out-of-band backups an attacker or a bad delete inside your account can't reach — kept in your own AWS account or Rubrik Cloud Vault.

RECOVER A WORKING, QUERYABLE TABLE

Catalog-Aware Recovery

Restore data files, metadata, and the catalog entry together as a consistent unit. Rubrik re-registers the table so it is queryable immediately in Athena, Spark, or Trino — with no post-recovery scripting.

Flexible, Granular Recovery

Recover in place, export to a new or existing table, or restore across account and Region for DR. Validate a recovery on an isolated branch before promoting it to production.

On-Demand Rollback Points

Take an on-demand backup before a risky schema change or large batch write for a clean, independent point to roll back to, without waiting for the next scheduled window.

EFFICIENT, AUTOMATED BACKUPS

SLA Policy-Driven Protection

Replace native snapshots and DIY scripts with a policy engine. Apply granular protection at the account, namespace, or individual table level, and new tables inherit protection automatically — no per-table tagging.

Automated Table Discovery

Automatically discover self-managed Iceberg on AWS Glue and fully managed Amazon S3 Tables through the existing cross-account role, eliminating protection gaps so no critical table is left unprotected.

Efficient, Compacted Backups

Back up the latest compacted snapshot instead of every one, and fetch only what changed using Iceberg's own manifest tracking after the first full — keeping backup windows and storage manageable at petabyte scale.

AIR-GAPPED SECURITY AND ENTERPRISE-GRADE

Air-gapped, Sovereign Storage

Keep immutable, air-gapped backups in your own AWS account (BYOS) or in Rubrik Cloud Vault, with customer-managed keys for server-side encryption — meeting stringent data-residency requirements.

Granular Access Controls

Onboard through the existing least-privilege cross-account IAM role and enforce role-based access controls, so users and Rubrik get only the access they need.

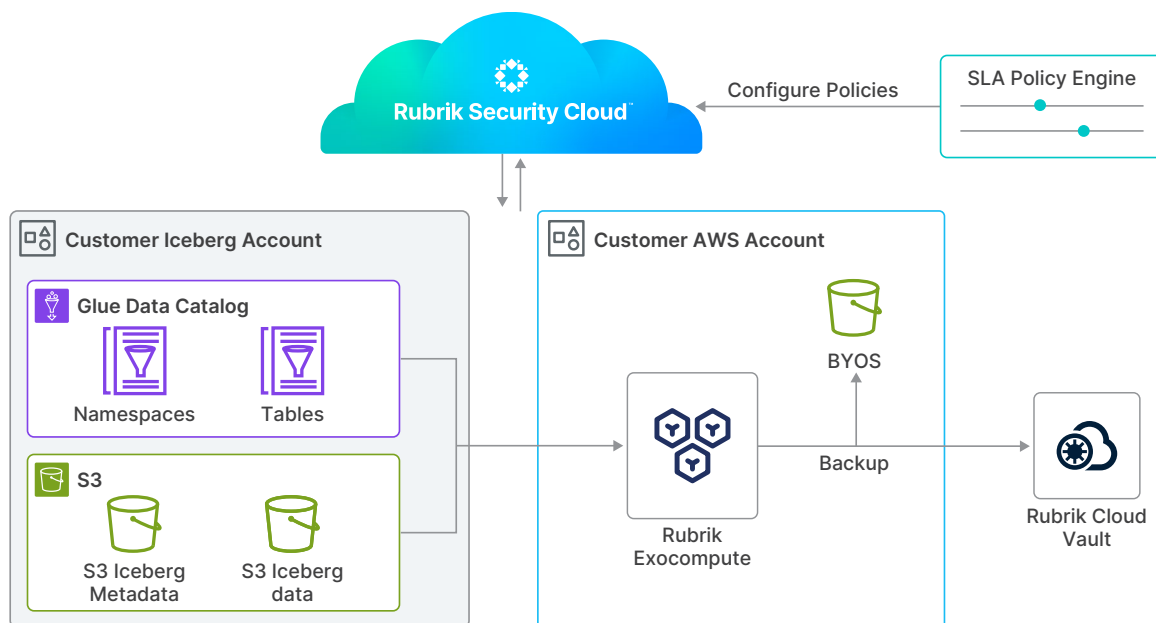
Unified Visibility & Reporting

Gain a single pane of glass for your lakehouse alongside AWS workloads (EC2, RDS, DynamoDB, S3), in the same compliance and audit reports, under existing CLU licensing.

HOW IT WORKS

Rubrik Apache Iceberg Protection runs from Rubrik Security Cloud (RSC) as the control plane. Rubrik's ephemeral Exocompute (EKS) runs inside your AWS environment and performs the backup and recovery work, reaching the protected account through a least-privilege cross-account — so credentials stay in the trusted Rubrik account and there is nothing new to stand up beyond existing AWS cloud native onboarding.

After the first full backup, Rubrik uses Iceberg's manifest-based change tracking to fetch only what changed. Exocompute pods transfer, index, and compress the changed metadata and data files, and the compacted backup written to an immutable destination you choose — your own AWS account or Rubrik Cloud Vault — for a logical air-gap. On recovery, Rubrik restores the data files, metadata, and catalog entry together. Exocompute pods read that snapshot's data and metadata from immutable storage. The data files and metadata are written back to the target S3 location — in place, overwriting the current data files, or as an export that creates a new table in an existing or new namespace. Rubrik then automatically re-registers the recovered branch in the catalog (Glue or S3 Tables), with no manual step, so the table is immediately queryable.



Global HQ

3495 Deer Creek Road
Palo Alto, CA 94304
United States

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

Rubrik (NYSE: RBRK) the Security and AI company, operates at the intersection of data protection, cyber resilience and enterprise AI acceleration. The Rubrik Security Cloud platform is designed to deliver robust cyber resilience and recovery including identity resilience to ensure continuous business operations, all on top of secure metadata and data lake. Rubrik's offerings also include Predibase to help further secure and deploy GenAI while delivering exceptional accuracy and efficiency for agentic applications.

For more information please visit www.rubrik.com and follow @rubrikinc on X (formerly Twitter) and Rubrik on LinkedIn. Rubrik is a registered trademark of Rubrik, Inc. All company names, product names, and other such names in this document are registered trademarks or trademarks of the relevant company.