rubrik

Rubrik Annapurna

The AI-ready unstructured data layer for enterprise Data Intelligence platforms. Catalog files in place. Deliver only what your AI pipeline calls for. Preserve provenance throughout.

OVERVIEW

Enterprises now hold more unstructured data than ever. It lives across cloud and on-premises NAS, AWS S3 and others, and most of it never reaches an AI pipeline. Less than 1 percent of unstructured data is ever used downstream for AI or analytics.[1]

The reason is not intent. It is the cost and complexity of moving the data. Activating an unstructured dataset today means copying it into staging, transforming it through ETL, and landing it in a lakehouse. You pay for storage, transformation, and egress at every hop, and the cost and complexity compound as the estate grows. By the time the data is ready, the analytical work has barely started.

This is the budget pattern data and analytics leaders recognize. In a representative program, the majority of the spend is consumed by plumbing (movement, storage, egress, transformation, cleaning) before the analytical output the business sees. The plumbing is not the value. But it is where the money goes.

KEY OUTCOMES

COST

Pipeline costs scale 1:1 with what AI actually pulls. Pay for the files in use, not for duplicating the full estate.

SPEED

Datasets that once needed months of assembly, ETL, and cleanup can return the same day.

TRUST

Provenance preserved per dataset. “Which files, on what date?” is a single query.

THE ANNAPURNA APPROACH

Rubrik Annapurna inverts the model. Instead of moving data to make it usable, it makes data usable where it already lives, and moves only what’s needed.

1. Catalog file metadata in place, across cloud and on-premises NAS, AWS S3 and others.
2. Query the catalog to find exactly the files you need, without copying anything.
3. Stage only the matched subset, and keep custody of what moved, so provenance, lineage, and versioning are preserved.

The headline is short: activate your unstructured data for AI. Move only what is needed.

HOW IT WORKS

- **Indexes in place.** Annapurna walks the source storage and reads metadata off every file. No bulk copy, no second system of record. The source data stays where it lives.
- **Publishes a queryable catalog.** Parsed file metadata is published as a queryable catalog, including source-system access controls captured and surfaced for downstream governance.
- **Stages only the matched subset.** When a query runs, Annapurna stages just the matched files into a managed object store (in the customer's S3, or locally for smaller datasets), positioned next to the AI pipeline.
- **Connects to your lakehouse.** Annapurna publishes to your Data Intelligence platforms such as Data Intelligence Catalog so data engineers query, stage, and mount from the Data Intelligence console. Native lakehouse connectors are on the roadmap.
- **Runs on Rubrik Security Cloud.** Deploys alongside existing storage and lakehouse environments. No new infrastructure.

Custody is retained throughout. Annapurna keeps the staged files in its managed object store, so provenance, lineage, and versioning are preserved per query and per dataset. Six months later, "exactly which files were in this dataset, and when?" is a single query against the staging store, not a forensic reconstruction.

WHERE ANNAPURNA FITS

- Multi-locality unstructured data estates spanning cloud and on-premises NAS, AWS S3 and others.
- Projects that depend on activating unstructured data for AI, ML, analytics, or regulatory reporting.
- Lakehouse environments where bronze ingest currently relies on bulk copies and bespoke ETL.
- Audit-sensitive workflows where dataset provenance and reproducibility are non-negotiable.

Activate your unstructured data for AI.

See Rubrik Annapurna live at Rubrik Forward on June 8,
or visit rubrik.com/annapurna to learn how Annapurna activates enterprise unstructured data for AI.

Source

1 DalleMule, L. and Davenport, T. "What's Your Data Strategy?" Harvard Business Review, May 2017.

rubrik

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 Operations 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.

brf-rubrik-annapurna / 20260609