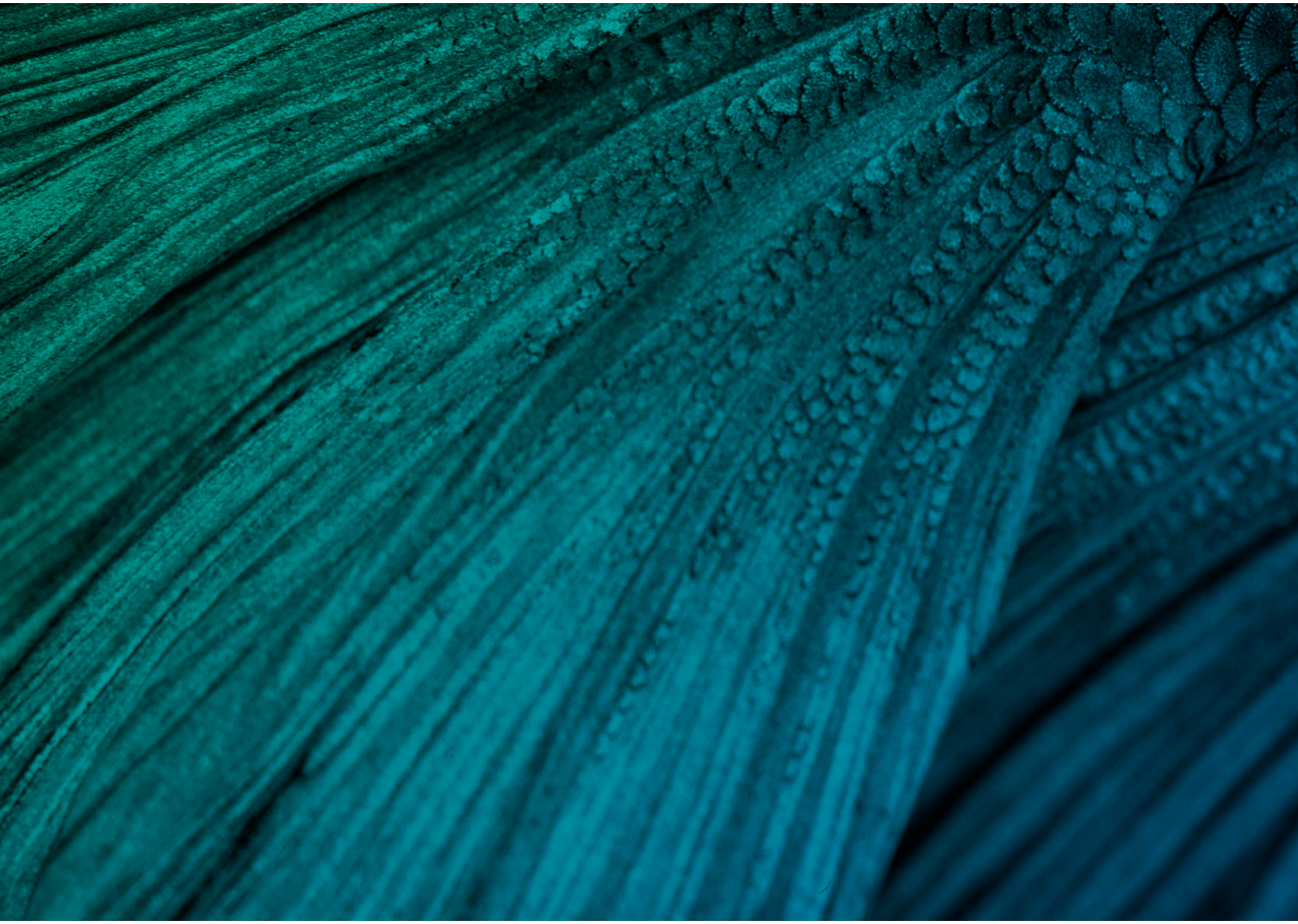
RUBRIK CONVERGED DATA MANAGEMENT

Technology Overview & How It Works



Contents

What We Solve

Instantaneous application recovery and data delivery remain elusive challenges for any IT organization. Businesses demand instant recovery from failures, quick access to test and development resources, and the latest data for business intelligence and analytics. To solve data management challenges, IT needs to integrate multiple points of technologies based on a legacy architecture:

- **Backup Software** is an application that captures data (files, databases, objects), creates supplementary copies, and transports to a backup storage target. Replication, sometimes included in backup software, generates and sends replica copies to a secondary target. In the event of a data loss, backup software orchestrates the recovery of the original content by using supplementary copies stored in the backup storage destination.

- **Backup Storage** provides an endpoint to store copies of data in active use. Disk and tape are often used as storage media. Additional tiers of backup storage, such as archival storage, are used to keep data copies that may not be in active use but will be needed for future use in analytics, record-keeping, compliance, etc.

This results in a brittle infrastructure with challenges such as:

- **Long Deployment Cycles:** Deployment takes months as IT migrates data and configures multiple components.

- **Management Complexity:** Using multiple point solutions results in architectural complexity and fragmented management. There may be separate interfaces to manage catalogs for each site, licensing becomes cumbersome to manage across different hardware and software components, etc.

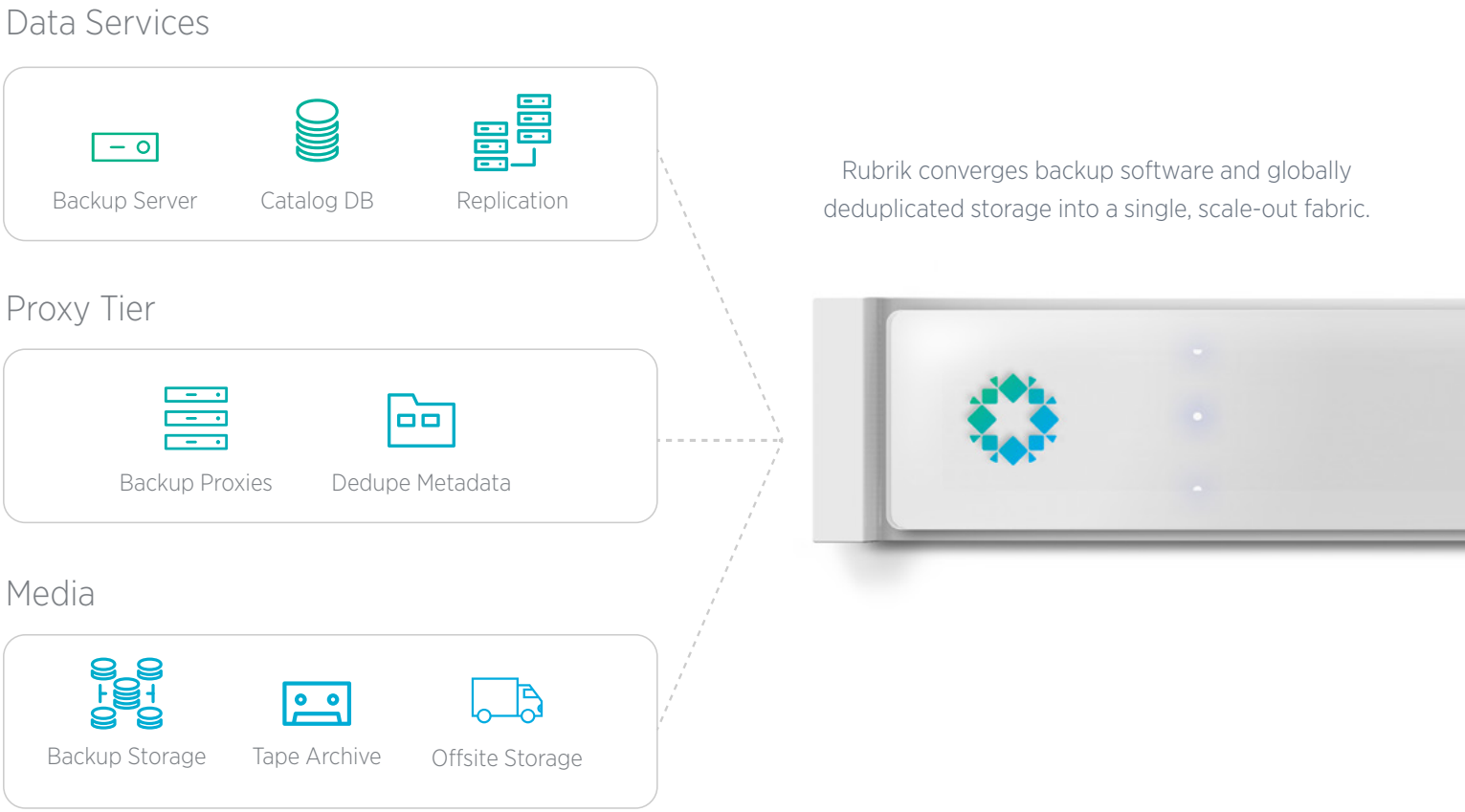
- **Inability to Meet SLAs:** The amount of data has increased dramatically while backup windows have shrunk. Data accessibility is hindered by the amount of time required to restore data from backup storage to production.

- **Lack of Scalability:** Inability to support demanding applications and big data growth beyond single digit petabytes of data.

Rubrik, the New Standard in Data Management

Rubrik redefines how data can be simply managed across data protection, disaster recovery, archival for compliance and long-term retention, application development, and data analytics.

We deliver the industry's first Converged Data Management Platform by combining backup software and globally deduplicated storage into a single, scale-out fabric. Rubrik horizontally scales to thousands of nodes in a single system. We package Rubrik with industry standard hardware and avoid usage of proprietary hardware.



The Rubrik Converged Data Management platform incorporates the following design principles:

- **Software Convergence:** We distill physically disparate components that comprise multi-tiered, legacy backup and recovery architecture into one single software.
- **Simplicity:** We solve for ease of use through simplicity. For example, we design our user interface to display only information that requires user attention, reducing cognitive overload.
- **Web-Scale:** We adopt the same web-scale technologies used by Google, Facebook, and Amazon, allowing our users to easily handle rapidly increasing volumes of information by adding more appliances to the cluster. Users avoid painful forklift upgrades and continue to easily manage Rubrik as a single system vs. multiple resource islands.
- **Efficiency:** We build intelligence into our software to help users efficiently manage data without incurring unnecessary costs (e.g., zero-byte cloning to save on storage capacity, sending only deduplicated data to the public cloud to reduce data transfer and storage) and labor (e.g., file search across a global index that spans private and public clouds).

- **Ecosystem Support:** We have designed our data platform to be vendor-agnostic and to work across modern data center applications and technologies.

Technology Overview

Rubrik Converged Data Management is a system that distributes data, metadata, and task management across the cluster in order to deliver predictive scalability and eliminate performance bottlenecks.

- **"The Core"** is the foundation of Rubrik and is comprised of the file system, metadata service, cluster management, and task framework.
- **"The Logic"** functions as the brains of Rubrik by organizing, removing redundancy, and making data available for search.
- **"The Interface"** provides a RESTful API-driven interface that interacts with users and supports virtualization, applications, and public cloud technologies.

The Core

Rubrik Cloud-Scale File System

Rubrik Cloud-Scale File System is a distributed file system built from the ground up to store and manage versioned data. We have designed this file system to be:

- **Fault Tolerant:** The system is resilient to multiple node and disk failures. We employ an intelligent replication scheme to distribute multiple copies of data throughout the cluster.
- **Flash-Optimized:** The system is built for a hybrid flash/disk architecture to maximize I/O throughput.
- **Storage Efficient:** The system utilizes zero-copy clones to make multiple copies of data from one "golden image."
- **Scale-out NFS server:** The system exposes itself as a scale-out NFS server to any host when a snapshot is mounted.

Rubrik Distributed Metadata System

Rubrik Distributed Metadata System operates alongside our Cloud-Scale File System, providing an index that can be accessed at high speeds. It delivers continuous availability, linear scalability, and operational simplicity with no single point of failure in the cluster. Our system is built to handle large amounts of data, distribute replicas of data across nodes (access to metadata is maintained even in the case of node failure), and provide low latency operations.

Rubrik Cluster Management

Rubrik Cluster Management manages the Rubrik system setup and ongoing system health. We use a zero-configuration multicast DNS protocol to automate appliance discovery – the cluster expands with minimal manual intervention with new nodes auto-discovering each other. Post system setup, it maintains the status of each node by performing health checks on individual nodes.

Rubrik Distributed Task Framework

Rubrik Distributed Task Framework is the engine that globally assigns and executes tasks across the cluster in a fault tolerant

and efficient manner. As a result, tasks are load balanced across the entire cluster, and tasks are distributed to the nodes that house the impacted data.

The Logic

Rubrik Data Management and Global Search

Rubrik Data Management serves as the "brains" of the system, enabling cradle-to-grave data lifecycle management from data ingest to archive and retirement. Fast, efficient data delivery is made possible by the ability to:

- store versions of data (we use a combination of full snapshot with forward incremental and reverse incremental copies)
- ensure data integrity (we build multiple checks within the file system and data management layers)
- apply content-aware global deduplication and compression (we intelligently apply data reduction at a global level while enabling fast data reconstruction)

The Interface

Programmatic Interface & Ecosystem Support

Our user interface is built on a RESTful API-driven framework with a HTML5 web user interface. Our UI is designed to provide ease-of-use and drive intuitive actions while reducing information overload for the user.

Rubrik is a vendor-agnostic platform with the ability to support any third-party ecosystem technology by building additional modules to the integration layer. This layer exposes the API set for building custom integration points into applications, hypervisors, containers, and protocols.

How It Works

Rack-and-Go System Setup

Once racked, Rubrik system setup is easily and quickly completed in 10-15 minutes. We invoke multicast DNS protocols to automatically discover and self-configure each of the nodes within the cluster. The user assigns IP addresses to each of the nodes (e.g., a r340 Hybrid Cloud Appliance has four nodes) and login credentials for the virtualized primary environment to be managed by Rubrik. To expand cluster size, the user simply assigns new IP addresses through the management dashboard. To reduce cluster size, the user selects the nodes to remove. Thereafter, the cluster automatically self-adjusts and re-balances to deliver fault tolerance against node and disk failures.

Automated Data Discovery

Once the user enters the credentials for its virtualized environment (e.g., vCenter username/password for VMware vSphere environments), Rubrik auto-discovers details of the entire virtualized environment, such as hosts and applications. Auto-discovery happens a variety of ways, depending on the user environment. Rubrik utilizes VMware APIs (vStorage APIs for Data Protection) to discover VMware environments. Support for additional virtualization hypervisors, containers, and applications will be rolled out in future releases.

Dynamic Policy Engine

From the list of discovered virtual machines (VMs), the user selects which VMs to protect and what SLA policies to apply for

recovery. To ease management, we have pre-configured SLA policies based on industry standards. The user has the flexibility to create new SLA domain policies by specifying the desired snapshot capture frequency and data retention policy. Users can select where data is stored, whether on-premise in the Rubrik r300 Series Hybrid Cloud Appliance or in a public cloud service (e.g., Amazon S3). The user simply slides the bar to the time at which data should be stored in the public cloud (e.g., 30 days). Rubrik provides a cost-effective alternative to tape for long-term data retention.

Rubrik allows users to intelligently and safely leverage the cloud. Only deduplicated data is transferred to the cloud. Data in-flight and at-rest in the cloud utilize military-grade AES 256-bit encryption.

Flash-Speed Data Ingestion

We have designed Rubrik as a high-speed data ingestion engine that can easily handle large volumes of data. Rubrik pioneers the usage of flash in backup and recovery, resulting in extremely fast data extraction and minimizing performance impact to the production environment. All data enters Rubrik through the flash tier. In addition, we have built an intelligent distributed workflow management system to maximize the number of parallel data streams processed. Since Rubrik is architected to be a web-scale system, performance for every dimension (such as network and disk throughput) increases predictably at a linear pace as more nodes are added to the cluster.

For VMware environments, we utilize VMware's Changed Block Tracking to identify and copy only the changed blocks from the previous operation. We apply intelligent global deduplication and compression before the data is stored in our cloud-scale file system. All metadata is stored in the flash tier for rapid access in a search pulldown. Data is distributed across multiple nodes to deliver a fault tolerant file system.

Easy, Fast Global File Search

Rubrik eliminates the file search complexity inherent in legacy backup and recovery solutions by introducing consumer-grade file search that delivers query results instantly. As the user types the query, Rubrik expedites the query by displaying suggested search results with auto-complete functionality. The user can instantly locate specific versions of files across all VMs.

Instant Recovery

By converging backup software and globally deduplicated backup storage into a single software fabric, Rubrik radically simplifies the recovery process. With just a click, users can instantly recover the VM by booting the virtual machine disk file (VDMK) directly on the Rubrik system. Rubrik serves as a storage endpoint for users to recover as many VMs as needed, eliminating the complexity and time wasted in transferring data back into the production system for recovery, thus providing a near zero RTO. Post-recovery, users can either choose to Storage vMotion the VMDK to the primary storage environment or continue using Rubrik as a storage endpoint. Rubrik's flash usage delivers fast IO performance. Writes and reads are gathered on the flash tier to deliver performance required by the recovered application.

Live Storage for DevOps

Rubrik pioneers the concept of Live Storage in which any copy of data can be mounted directly on Rubrik as a storage endpoint. As a result, Rubrik can be used to accelerate application development by providing multiple copies to developers from just one "golden image". Our Cloud-Scale File System has built-in native cloning capabilities to allow any number of mounts to be created without requiring additional storage capacity. Users can provision as many copies to developers as needed without impacting storage capacity and within a sandbox environment to prevent any network conflicts. As developers alter the provisioned data set, Rubrik stores the deltas by forking to a new branch. Our journaled Cloud-Scale File System provides an extremely efficient mechanism for accelerating and provisioning the latest data for application development. For medium-sized workloads, users receive all-flash performance comparable to a primary system of similar capacity. Rubrik intelligently allocates the flash

tier for all writes and hot reads when utilizing Live Storage.

The New Standard of Data Management

We have designed Rubrik to set the new standard for data management across backup and recovery and application development. Over time, we plan to build out deeper content discovery and hybrid cloud automation to continue delivering a comprehensive yet simple-to-use data management platform. Rubrik pioneers a radically innovative approach in data management by distilling formerly discrete, complex components that require manual stitching into a single, elegant software fabric packaged with industry standard hardware. As a result, our customers experience unprecedented simplicity, ease of use, and substantial cost savings.

Global HQ
299 South California Ave. #250
Palo Alto, CA 94306
United States

650–328–2745
info@rubrik.com
www.rubrik.com