Diskover

DATASHEET

Data Mobility

Discover every share. Set the policy. Move data at scale.

DataDiscovery finds every share and bucket and exports file metadata to your lakehouse in Parquet. DataDelivery policies and agents move files from tier to tier.

Moving file data at scale is rarely a copy problem. It is a knowing problem, and it breaks down in three places.

The challenges and their impact

NO MAP OF THE ESTATE

- New shares and buckets appear faster than anyone can register them.
- Impact: tiering rules miss the data that costs the most.

SCRIPTS, NOT POLICY

- Tiering lives in cron jobs and one-off scripts that nobody owns.
- Impact: moves are slow, unverified, and hard to repeat.

METADATA IN SILOS

- File metadata never reaches the analytics stack.
- Impact: data teams rebuild file inventories again and again.

At a glance

THE SOLUTION

DataDiscovery indexes every NFS, SMB, and S3 source. DataDelivery policies and agents carry out the transfers, at up to 50 Gbps with a single agent. DataDiscovery shows what you have, and DataDelivery moves it.

HOW IT WORKS

DataDiscovery finds the shares and buckets on each server. A DataDelivery policy selects files by name, path, size, or age. Test it with a dry run, then let it copy, move, or sync on a schedule. Every run is recorded in the activity log.

WHY IT MATTERS

Tiering runs as policy instead of tickets. Every job is recorded in the activity log, so you know what ran and when. DataDiscovery metadata feeds your lakehouse, so data teams work from the current state of the estate.

Core capabilities

READ-ONLY SCANS

DataDiscovery workers mount volumes right before a scan, read-only by default. Credentials are encrypted and passed in memory, never written to disk.

POLICY CREATION

DataDelivery policies select files by name, path, size, or age, then copy, move, or sync them on a schedule. Need a one-off move? Run the same policy on demand.

DRY RUN FIRST

Dry-run any transfer first to see its scope and cost before the job runs. Adjust the policy until the scope is right, then run it with confidence.

HIGH-SPEED TRANSFERS

DataDelivery agents copy, move, and sync at up to 50 Gbps with a single agent. Several agents attached to the same storage systems share one job.

ACTIVITY LOG

Every DataDelivery job is recorded in the activity log, so you know what ran and when. No more guessing which script ran last night.

OPEN DATA OUT

DataDiscovery reports export chosen metadata fields to CSV or Parquet on a schedule. DataDelivery catalogs publish file metadata to Kafka topics.

About Diskover

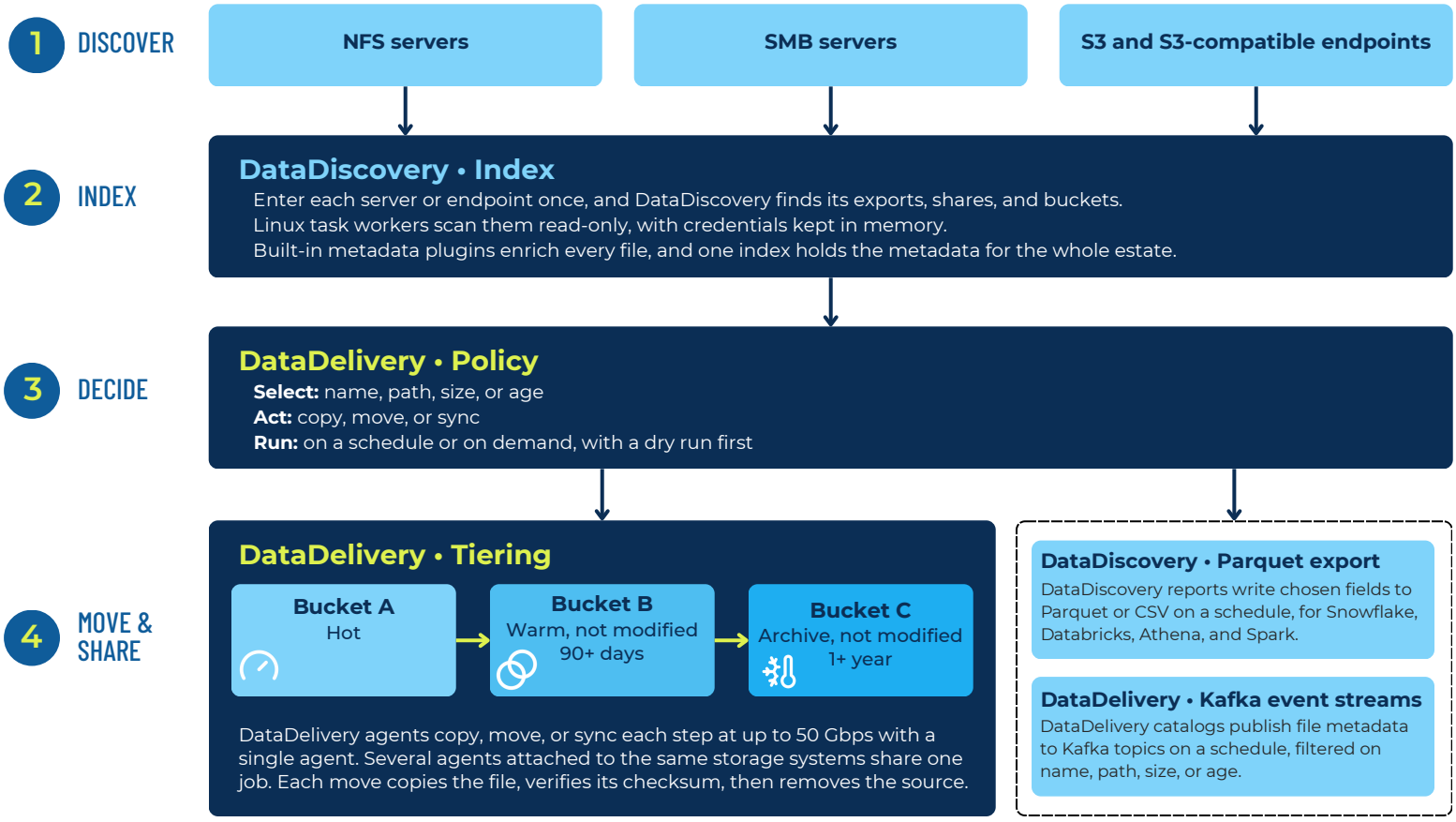
Diskover helps enterprises structure unstructured data across cloud, on-prem, and edge environments, making it easier to search, govern, optimize, and deliver file data for compliance, analytics, and AI-ready workflows.

Contact

diskoverdata.com
sales@diskoverdata.com
(800) 560-5853
San Jose, CA

How data moves

DataDiscovery maps and indexes your storage. DataDelivery policies and agents move the files. Today each product runs its own jobs. The COMING NEXT box shows how they will connect.



Technical specifications

DataDiscovery

SOURCES	NFS, SMB, S3, and S3-compatible storage
SCANNER	Linux task workers, read-only mounts, cache
SIGN-IN	LDAP/AD, Okta, and Entra ID
SCALE	1.4 billion files indexed in 2 hours
EXPORTS	CSV or Parquet reports on a schedule

DataDelivery

STORAGE	NFS, SMB, S3, Azure Blob, GCS, Wasabi, and PowerScale
POLICY ACTIONS	Copy, move, and sync
TRIGGERS	On a schedule or on demand, dry run first
THROUGHPUT	Up to 50 Gbps with a single agent
ROUTING	Single-agent or agent-to-agent, least-loaded agents first
SIGN-IN	Okta, Entra ID, and SAML
STREAMING	Kafka topics, through catalogs

COMING NEXT

Start DataDelivery transfers from DataDiscovery: select files, a whole search, or a policy, choose the destination, review the plan, and start the job. Mappings scale to hundreds of shares and buckets, and a Parquet ingester will write the full index as Parquet files for lakehouse loads.

About Diskover

Diskover helps enterprises structure unstructured data across cloud, on-prem, and edge environments, making it easier to search, govern, optimize, and deliver file data for compliance, analytics, and AI-ready workflows.

Contact

diskoverdata.com
sales@diskoverdata.com
(800) 560-5853
San Jose, CA

50 Gbps
data mobility with a single DataDelivery agent

1.4B
files indexed in 2 hours by DataDiscovery

30-50%
of storage reclaimed by tiering and removing ROT data

3–6
months to payback after deployment

Built for estates with billions of files

FAST & OPEN

DataDelivery works with the storage you already run. A high-performance engine with multipart uploads and per-agent concurrency control keeps large transfers fast. Standard APIs, webhooks, and role-based access control mean no lock-in, and it operates alongside movers you already have.

SECURE & YOURS

A DataDelivery controller runs as SaaS or hosted in your own environment, with lightweight agents close to the storage on Windows, Linux, or Mac. Access is browser-based, so end users install nothing. Traffic between DataDelivery agents and the controller is encrypted, and files are checksum verified by default. We never store your file contents.

WORKS ACROSS NFS · SMB · S3 · S3-compatible · Azure Blob · Google Cloud Storage · Wasabi · PowerScale · Lakehouses · AI/BI/ML

Outcomes teams can measure

SAVINGS THAT HOLD

Tier down what nobody uses, defer the next capacity purchase, and let policy keep your storage lean and clean.

DECISIONS YOU CAN DEFEND

Dry-run any change, so cost and scope are known before the job runs.

HOURS BACK TO YOUR TEAM

Rules and schedules replace ticket queues, spreadsheets, and one-off scripts.

PROTECTED & PROVABLE

Agent traffic is encrypted, files are checksum verified, and every job is recorded in the activity log.

Common use cases

TIERING ACROSS BUCKETS

Hot data lands in bucket A. One DataDelivery policy moves files not modified in 90 days to bucket B, and another moves them to bucket C after a year. Each move copies the file, verifies its checksum, then removes the source.

LAKEHOUSE ANALYTICS

Data teams query the file estate like any other table. Scheduled DataDiscovery reports export chosen fields, including plugin fields, to Parquet for Snowflake, Databricks, Athena, or Spark.

STORAGE REFRESH & MIGRATION

A refresh starts with knowing what you have. DataDiscovery inventories the estate, and DataDelivery carries only what matters onto the new platform, so years of unused files stay behind.

EVENT STREAMS WITH KAFKA

DataDelivery catalogs publish file metadata to Kafka topics on a schedule, filtered on name, path, size, or age, so downstream jobs pick up the matching files.

GET STARTED

Plan your first tiering policy.

Run a 30-day proof of concept on your own shares and buckets.

Scan to book a demo

About Diskover

Diskover helps enterprises structure unstructured data across cloud, on-prem, and edge environments, making it easier to search, govern, optimize, and deliver file data for compliance, analytics, and AI-ready workflows.

Contact

diskoverdata.com
sales@diskoverdata.com
(800) 560-5853
San Jose, CA

Revised November 2026