Diskover

MARKET BRIEF

Every file you send to AI has a price.

Curate first, and pay only for the data that matters.

Most AI projects point a model at a whole file share. You pay to store, move, and process every duplicate, old version, and temp file in it, even though the model never needed them.

60% of organizations went over AI budget on storage and compute (Snowflake)

$108B in AI investment wasted each year on weak data foundations

7x growth in AI data readiness spending, 2025 to 2029

20% of unstructured data is AI ready, per a Snowflake survey

At a glance

KEY CHALLENGES

You pay three times for data the model does not need: to store it, to copy it into the pipeline, and to process it. And storage keeps getting more expensive, with enterprise SSD contract prices up about 80% in Q1 2026 alone.

THE SOLUTION

Diskover indexes every file across NAS, cloud, object, and archive into one enriched metadata catalog. Teams curate a named dataset and deliver only that set to AI, so you stop paying for the rest.

WHY IT MATTERS

Every file you leave out is one you do not store, move, or process for AI. A curated dataset also gives better answers and keeps likely sensitive files away from the model.

Core capabilities

INDEX EVERY TIER

Distributed indexers read SMB, NFS, S3, Azure Blob, tape, and archive into one catalog across billions of files. Your data stays in place, so building the catalog adds no copies to pay for.

DESCRIBE EVERY FILE

Metadata such as owner, size, location, and last access date shows which files earn their place. Plugins and tags add business context from the systems you already run.

LEAVE OUT THE NOISE

Find duplicates, old versions, temp output, and likely sensitive files, and exclude them before delivery. You stop paying to store, move, and process files the model does not need.

ASK FOR THE FILES YOU NEED

Connect your own AI tool through the AI Connector and ask for the files a project needs in plain language. It returns matching files from the catalog and proposes tags. AI-assisted. Human-approved, every time.

CURATE AND DELIVER

Save the result as a named dataset with an owner and deliver it with DataDelivery. The pipeline receives only the files it needs, on the schedule you set.

KEEP THE EVIDENCE

Every dataset records which files went in, where they came from, and when. Traceable lineage helps teams answer audits, retention questions, and model reviews faster.

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

Diskover

MARKET BRIEF

How it works

1 CONNECT
ANY STORAGE

Point Diskover at your storage. Distributed indexers read metadata from each system you give them.

Example: a 2 PB NAS estate is indexed where it sits.

2 CATALOG
AUTOMATION

Every file lands in one searchable catalog with metadata such as owner, size, location, and last access date.

Example: a project share now shows its owner, age, and size.

3 CURATE
AI CONNECTOR OR SEARCH

Search the catalog or ask through the AI Connector. Leave out duplicates, old versions, and likely sensitive files.

Example: only current, owned files for one project make the dataset.

4 DELIVER
BY POLICY

Send only the curated dataset to your AI pipeline or lakehouse. The catalog records which files went in, and when.

Result: a cancer research center reclaimed 0.5 PB of leftover AI pipeline data and avoided $250K in storage overspend.

WORKS ACROSS Cloud · On-prem · Hybrid · NAS · SAN · Object · Tape · Archive · Data lakes · Lakehouses · AI/BI/ML

What you send to AI versus what it needs

Most AI projects start by pointing a model at a whole file share. You pay to store it, to copy it into the pipeline, and to process it, and the model reads the duplicates, drafts, and old versions as if they were all true.

Diskover reverses the order. It describes every file from metadata alone, lets teams decide what belongs, and delivers only that set. You pay for less data, and the model gets better data.

YOUR FILE SHARES →

WITHOUT DATA PREP
everything goes in

20% AI READY

WHAT YOU PAY FOR
Without data prep: storage, transfer, and processing for all of it.
Curated first: transfer and processing for only the files you need.
Only 20% of unstructured data is AI ready, per a Snowflake survey.

Common use cases

CURATE TRAINING DATA
Select the files that match a training need, such as recordings from a given condition or instrument, and leave out the rest, so the model learns from the right examples.

FEED THE LAKEHOUSE
Send a curated, described dataset to Snowflake or another lakehouse, so analysts and AI tools query one catalog instead of many shares.

CLEAR THE AI PIPELINE
Find the checkpoints, temporary output, and old training runs that collect around AI projects, and recover the premium storage they occupy.

Outcomes teams can feel

LOWER AI COSTS
Less data to store, move, and process. Diskover customers reclaim 30 to 50% of their storage by tiering and removing ROT data.

BETTER ANSWERS
The model reads current, owned files instead of duplicates and drafts, so its answers reflect the work your teams actually rely on.

LESS RISK IN THE MODEL
Likely sensitive files are flagged and left out before delivery, so they stay out of prompts, answers, and training runs.

GET STARTED

See what your AI data really costs.

Estimate it with the cost calculator in the eGuide, or size it with a 30-day proof of concept.

Book a demo

Get the eGuide

Sources: 60%, 20%: Snowflake (Omdia), March 2026. $108B: Hitachi Vantara, January 2026. 7x: Gartner, March 2026. SSD prices: TrendForce, June 2026.

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