Diskover + snowflake®

CUSTOMER PROFILE

A global entertainment company valued at more than $180 billion, with studios, broadcast networks, streaming services, and theme parks serving audiences worldwide.

Its production, post-production, broadcast, and archive operations generate unstructured media at petabyte scale across six data centers and more than five file systems. Data platform teams support 30+ operational and business users who need to find, understand, and reuse that media for analytics and AI.

CUSTOMER SUCCESS STORY

260 PB of media. One catalog. Curated for Snowflake Cortex AI.

How a global entertainment leader turned its unstructured media estate into curated datasets for Snowflake and Cortex AI, and retired the platform that could not keep up.



260 PB	100B+	30+	1
of unstructured media indexed across 6 data centers	files cataloged with searchable metadata	data and operational users	metadata layer feeding Snowflake and Cortex AI

At a glance

The company holds 260 PB of unstructured media across production, post, broadcast, archives, and theatrical assets, spread over six data centers and more than five file systems. Feeding that estate into AI and BI pipelines ran through a general-purpose lakehouse platform that could not scale to unstructured data without heavy overhead, and preparing training datasets from raw media was a manual scripting job. Business users who wanted answers about the estate filed tickets and waited.

Diskover indexed every file across all six data centers into one global metadata catalog and enriched it with production context from the company's Autodesk production tracking system. Data teams tag, down-select, and curate datasets in the catalog with saved searches instead of scripts. The curated, tagged metadata streams into Snowflake, where Cortex AI lets operational and business users ask questions about the entire media estate in plain language.

The lakehouse workflow for unstructured media was decommissioned, and its license and overhead spend recovered. AI and BI pipelines are fed from 260 PB of media without manual preparation, ticket-driven queries have moved to self-service, and the company runs fewer tools and pipelines on one metadata layer across its content estate.

Facts

INDUSTRY
Global media and entertainment.

DATA
Unstructured media across production, post-production, broadcast, archives, and theatrical assets.

USERS
Data platform teams curating datasets, with 30+ users across 6 data centers querying the estate through Cortex AI.

DEPLOYMENT STAGE
Production. Snowflake and Cortex AI integrations live, previous unstructured-data pipeline decommissioned.

STORAGE INDEXED
260 PB across six data centers, on Dell PowerScale and NetApp file systems and AWS S3 object storage.

LOCATIONS
Studios, networks, and business units worldwide, served from six data centers.

WHAT'S NEXT
Expanding curated dataset coverage and conversational query patterns across more business functions.

DISKOVER WORKS ACROSS

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

Diskover

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

CUSTOMER SUCCESS STORY

Snowflake is built for data you can query. Media is 260 petabytes of files that nobody could query, so the first job was to turn every file into a record: what it is, which production it belongs to, its format, its status.

Once that catalog exists, curation is a search, not a script. The metadata that matters streams into Snowflake continuously, the file itself is pulled only when a pipeline needs it, and a business user asks Cortex AI a question instead of filing a ticket.

Olivier Rivard, VP of Product
Diskover

CHALLENGES

PETABYTES PAST THE PLATFORM
Unstructured media at 260 PB demanded overhead the existing lakehouse deployment could not absorb, so the AI and BI pipeline stalled at the source.

BUSINESS USERS BLOCKED
Operational and business teams could not query the media estate without filing a ticket or waiting for a script, so questions went unasked.

CURATION BY SCRIPT
Tagging, down-selecting, and preparing AI training datasets from raw media was a manual scripting burden carried by a small data platform team.

SIX SITES, NO SHARED VIEW
Media sat across six data centers and more than five file systems with no single inventory to curate from.

SOLUTIONS

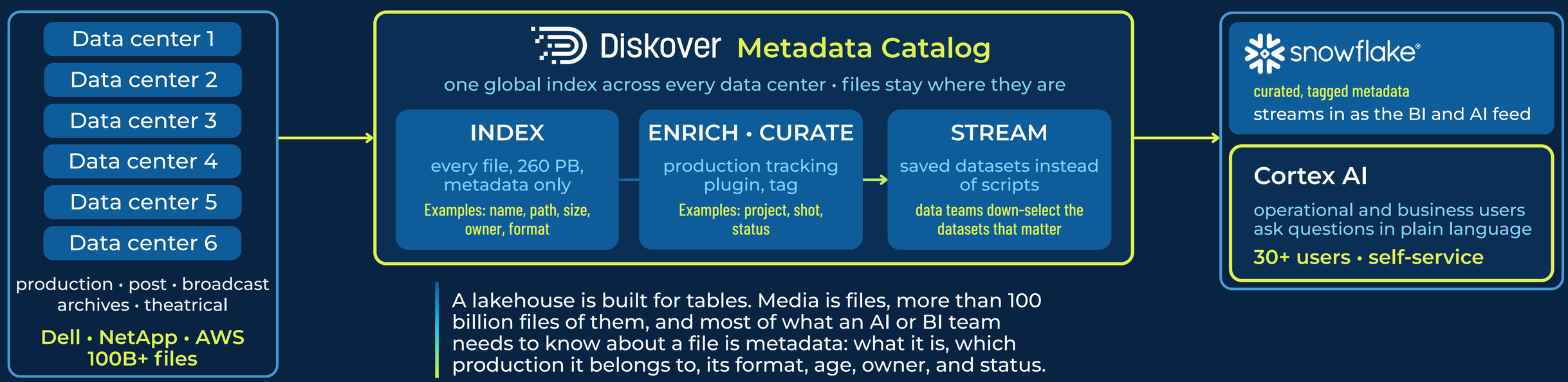
ONE CATALOG, 260 PB INDEXED
Diskover indexed every file across six data centers and more than five file systems into one global metadata catalog, with the files staying where they are.

TAG, DOWN-SELECT, CURATE
Saved searches and tags let data teams find the relevant slice of the estate and mark it as a dataset, replacing the scripts and the lakehouse workflow that did this before.

PRODUCTION CONTEXT IN THE INDEX
A plugin pulls project, shot, and status data from the company's Autodesk production tracking system into each file's record, so curation works on business meaning, not just paths.

SNOWFLAKE IN, CORTEX AI OUT
Curated, tagged metadata streams into Snowflake tables continuously, and file content is pulled on demand only for the datasets a pipeline needs. Operational and business users ask Cortex AI questions across the entire media estate in plain language.

260 PB of media, curated for AI.



A lakehouse is built for tables. Media is files, more than 100 billion files of them, and most of what an AI or BI team needs to know about a file is metadata: what it is, which production it belongs to, its format, age, owner, and status.

OUTCOMES

PIPELINE COSTS RECOVERED
The lakehouse workflow for unstructured media was decommissioned, and its license and overhead spend recovered.

ANSWERS WITHOUT TICKETS
30+ business and operational users get plain-language answers, and data teams stopped fielding ticket-driven queries.

AI-READY DATA AT SCALE
Curated metadata reaches Snowflake and Cortex AI from 260 PB of media without manual preparation.

FEWER TOOLS, ONE LAYER
One unified metadata layer across the content estate replaced separate tools and pipelines.

GET STARTED

Ready to see it on your estate?

Start with a 30-day proof of concept on your own data, and recover your investment in 3 to 6 months.

Diskover

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