Diskover

CUSTOMER PROFILE

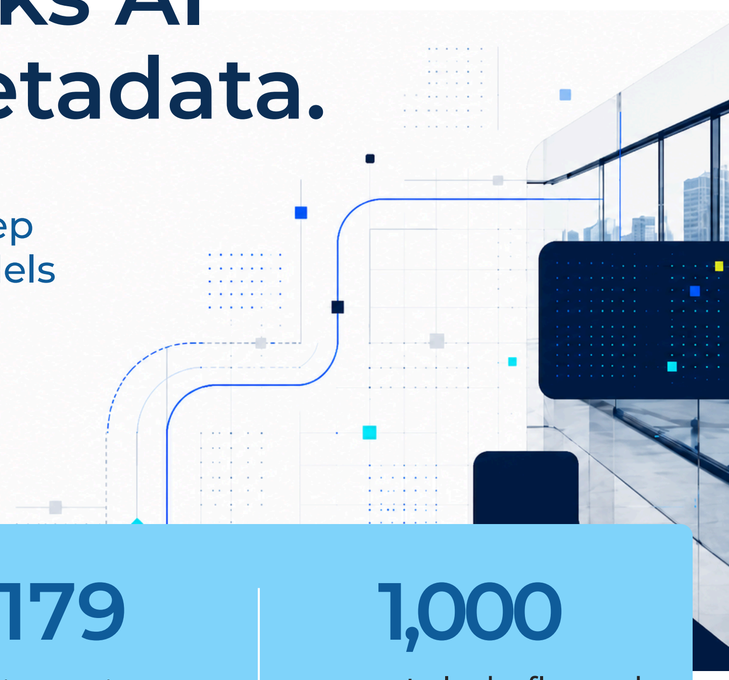
One of the world's largest vehicle makers, with an autonomous driving program and a planned 100-vehicle test fleet. Each vehicle records about 5 TB of camera, LiDAR, radar, and GPS data a day at remote test sites, and AI teams train perception models at headquarters.

The infrastructure team owns the storage and the network between sites. The data and AI architects decide which recordings the models need next.

CUSTOMER SUCCESS STORY

Global automaker picks AI training data from metadata.

How an autonomous driving program plans to keep test-fleet recordings at the edge and send its models only the data they lack.



500 TB	2,067	179	1,000
a day expected from the planned 100-vehicle fleet	recordings profiled in one demo scan	rain-session recordings named, under 9% of the set	suspect clocks flagged before training

At a glance

The automaker's default plan shipped every recording from every test site to headquarters, where the models are trained. With about 500 TB a day expected from a planned 100-vehicle fleet, the infrastructure team counted the cost in bandwidth, transfer windows, and duplicate storage. The AI team counted it in weeks spent waiting for data the model might not need.

Their question was simple: which recordings do our models still need? Diskover answered it from metadata. A plugin read each MCAP recording's sensors, capture time, and GPS position during the scan, a workflow step added the weather at capture, and dashboards showed the gaps by site, sensor, weather, and temperature.

On 2,067 public recordings, the demo named 179 recordings from a Boston rain session as candidates, found no snow or sub-zero driving among the recordings with weather, and flagged 1,000 suspect clocks before any training. The automaker is now a Diskover customer, working toward one Diskover across every site.

Facts

INDUSTRY
Automotive: autonomous driving and perception model training.

DATA
MCAP sensor recordings: camera, LiDAR, radar, GPS, and vehicle state.

SCALE
5 TB per vehicle per day, so about 500 TB a day for the planned 100-vehicle fleet.

USERS
Infrastructure, data, and AI teams at headquarters and the test sites.

LOCATIONS
Remote test sites, with model training at headquarters.

DEPLOYMENT STAGE
Diskover customer, with deployment in progress.

WHAT'S NEXT
Index each test site, enrich every recording, and copy only tagged recordings to the training store.

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

Diskover

CUSTOMER SUCCESS STORY

If I can transfer only the data containing rain back home, rather than all of this data, it makes a world of difference.

AI Data Architect
Global automaker
Autonomous driving program

CHALLENGES

EVERY RECORDING, EVERY BYTE
The default plan copied every recording from every test site to headquarters, at a cost in bandwidth, transfer windows, and duplicate storage.

ANSWERS INSIDE THE FILES
What a recording captured sits inside the file. To the file system, a recording is only a name, a size, and a date.

AI TEAMS WAITING
AI teams waited weeks for data the model might not need.

GROWTH MULTIPLIES THE COST
At 500 TB a day from a planned 100-vehicle fleet, a copy-everything plan turns into a bandwidth problem.

SOLUTIONS

READ EVERY RECORDING
An MCAP plugin writes fields such as sensors, capture time, and GPS position to the catalog during the scan. Fields are added as the team's questions change.

ADD THE WEATHER AT CAPTURE
A workflow step uses each recording's place and hour to add the weather the vehicle never recorded, such as rain, temperature, and wind.

APPROVED TAGS ONLY
Diskover proposes tags from the enriched fields, such as recordings captured in a wet hour, and applies them only after a person approves.

ONE DISKOVER, EVERY SITE
A central instance holds the catalog. Diskover indexes each site, metadata travels on every scan, and a policy copies only tagged recordings.

1 INDEX
Every test site indexed in place.
Site A · Site B · Site C · Site D

→

2 READ
MCAP fields written at scan time.
sensors · time · GPS · platform

→

3 ENRICH
Weather added from geolocation and time. Dashboards show gaps by site, sensor, and condition.

→

4 SELECT
Tags proposed, then approved.
wet hour + LiDAR + approved

→

5 DELIVER
Deliver only tagged recordings.
copy · keep · review

Recordings stay at the edge until a policy delivers them. Metadata travels on every scan.

DEMO RESULTS

GAPS ON ONE PAGE
One scan of 2,067 recordings showed coverage by site, sensor, weather, and temperature. Matched recordings held no snow or sub-zero driving.

CANDIDATES NAMED, FILE BY FILE
179 of 2,067 recordings, under 9% of the set, surfaced as rain candidates before any file moved.

PROBLEMS CAUGHT EARLY
1,000 suspect clocks and 20 recordings with no position were flagged before any training.

READY FOR EVERY SITE
The infrastructure team saw the transfer savings right away, and the AI architects asked how to bring their own extraction models.

GET STARTED

Ready to see it on your data?

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