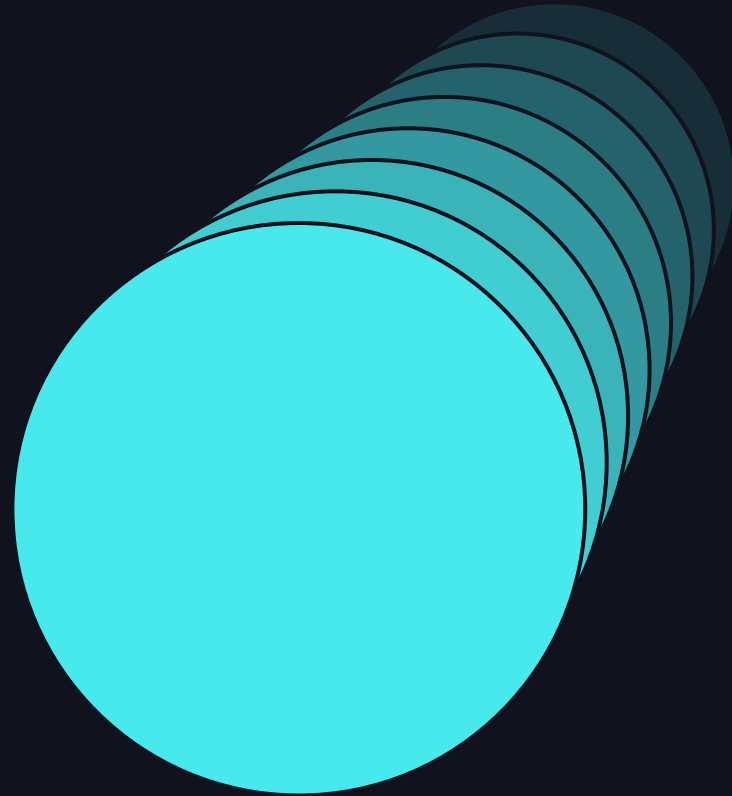


What's New with Data Sharing and Collaboration

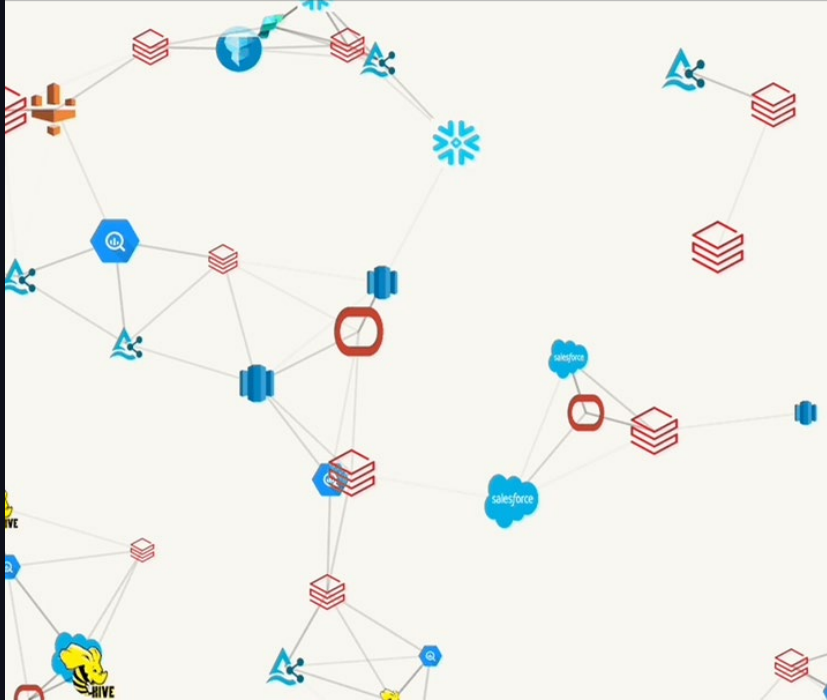


Tao Tao | Director, Engineering
Giselle Goicochea | Sr. Product Marketing Manager

Product safe harbor statement

Information is provided to outline Databricks' general product direction and is for informational purposes only. Customers who purchase Databricks services should make their purchase decisions relying solely upon services, features, and functions that are currently available. Unreleased features or functionality described in forward-looking statements are subject to change at Databricks discretion and may not be delivered as planned or at all

Secure, flexible and open sharing unlocks new value for your data



Gartner

CDOs who have successfully executed **data sharing initiatives** are **1.7X more effective in showing business value** and return of investment from their data analytics strategy.

Data sharing and collaboration is complex and expensive

Every vendor/partner has a different delivery method

Data must be moved and copied across companies

Different companies use different platforms

Can't share non-structured data assets



Slow onboarding of new data



Data duplication, stale data, and privacy risks



Difficult to collaborate across platforms



Can't unlock AI use-cases

Sharing and
collaboration:
an open approach
for maximal
reach and impact

Unified data sharing and collaboration

Unified mechanism to
onboard
Data and AI

Secure live sharing of
data and AI
assets

Open and
cross-platform
collaboration

Share tables,
volumes, AI models,
and notebooks



Faster data
onboarding and
innovation



Drive innovation with
no replication
or data movement



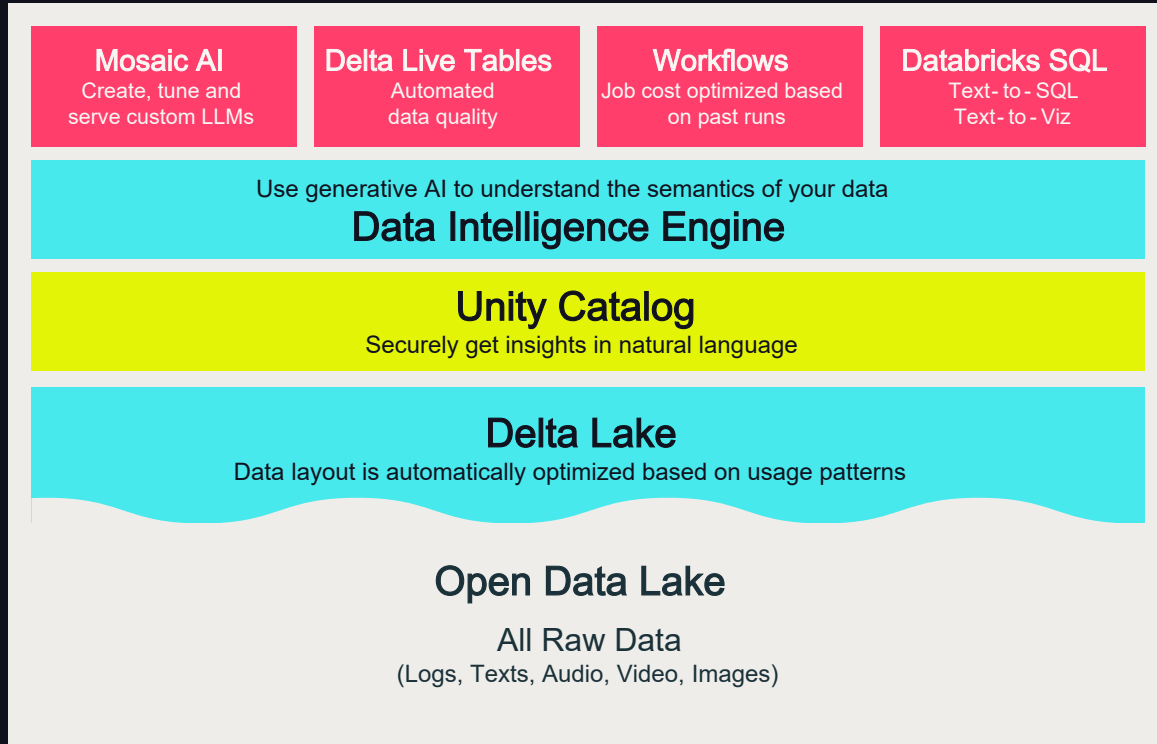
Access to an open,
cross-platform
ecosystem



Unlock all data and AI
use cases
for anyone

The most secure,
flexible and open
ecosystem for
sharing &
collaboration

Databricks Data Intelligence Platform



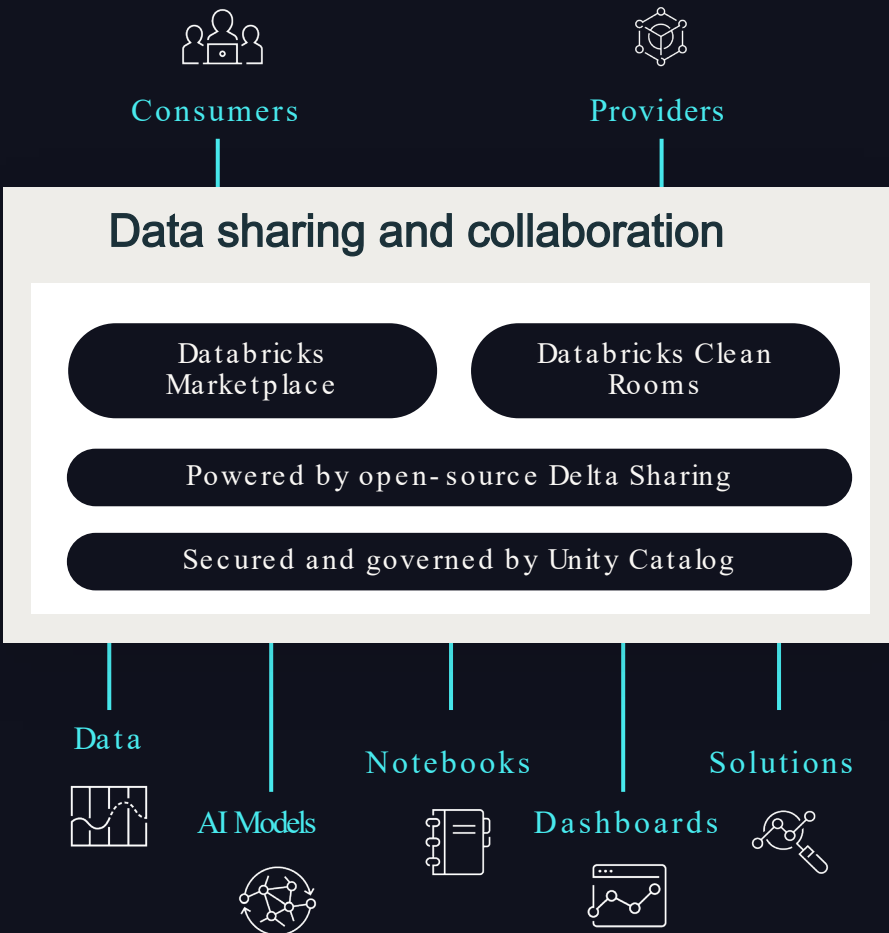
Databricks Data Intelligence Platform

Built for sharing + collaboration

Share data + AI in a unified way

Open ecosystem for massive reach

Secure and privacy-safe collaboration



Data collaboration drives **business impact**



Internal Sharing (Data Mesh)

Share data across divisions in different clouds & regions



Peer-to-Peer Sharing

Collaborate on data and AI with partners on any platform



3rd-Party Data Licensing

License data, AI models, and packaged solutions



Privacy - Safe Collaboration

Share data and AI with trusted partners in data clean room



Data collaboration drives **business impact**



SaaS Application Sharing

Expand your lakehouse to access all your data across SaaS applications

AVEVA

Amperity &

 theTradeDesk®

 tetrascience



Customers across industries are using Databricks for collaboration

Advertising & Marketing

- Customer 360 and CDP
- Identity Resolution
- Campaign planning and activation
- Measurement & attribution

Healthcare & Life Sciences

- R&D Optimization
- Commercial effectiveness
- Streamline Operations
- Digital supply chain

Financial Services

- Market data
- Regulatory & reporting
- Fraud detection & prevention
- Credit risk assessment

Retail & Consumer Goods

- Supply chain optimization
- Customer 360
- Retail media
- Inventory optimization

Industrials & Manufacturing

- Asset optimization
- Predictive Maintenance
- Customer 360
- Inventory optimization



Delta Sharing

The first open protocol for cross-cloud, cross-platform data sharing



Delta Sharing ecosystem incredible growth

16,000+

data recipients using
Delta Sharing

(+ **300%** YoY in monthly
active shares)

40%

of shares from our
customers are to non-
Databricks users



New Strategic Delta Sharing Partners

Industry-leading partners adopting Databricks for collaboration

ACXION

Amperity &

ATLASSIAN

AVEVA

Epsilon®

healthverity®

S&P Global

shutterstock®

T Mobile™

theTradeDesk®

+ a b | e a u®
from Salesforce

tetrascience

S&P Global

Enhancing data distribution capabilities with Delta Sharing

“We are excited to continue to work with Databricks to enhance our data distribution capabilities through their Delta Sharing platform which provides an open ecosystem, making a number of S&P Global content sets more seamlessly accessible and available to our clients. This expands our collaboration with Databricks which began with S&P Global Capital IQ Workbench, leveraging their technology to create a collaborative analytics notebook environment for our users.”

—David Coluccio, Head of Distribution Solutions, S&P Global Market Intelligence



Shutterstock's image datasets now available on the Databricks Marketplace

“Shutterstock is bringing **its vast collection of nearly a billion creative content assets to the Databricks Marketplace**, a platform renowned for fostering open data and AI collaboration. This integration provides unparalleled access to our extensive library of ethically-sourced visual content, propelling responsible AI and ML initiatives forward across various industries. **We are excited to add Delta Sharing as a method to deliver data.** Customers utilizing our rich dataset on Databricks can tap into new opportunities, catalyze product innovations, and secure a competitive advantage.”

—Aimee Egan, Chief Enterprise Officer, Shutterstock





Strategic Collaboration for Industrial AI Outcomes

“AVEVA is delighted to embark on this transformative journey with Databricks. Our combined strengths will empower our customers with a comprehensive solution that seamlessly integrates industrial software and powerful analytics, enabling enterprises to thrive in an increasingly data-driven world.”

—Rob McGreevy, Chief Product Officer, AVEVA





Driving insights with a flexible, open ecosystem

"Atlassian Analytics recently launched Data Shares, leveraging Delta Sharing from Databricks, to boost flexibility and accelerate customers' time - to - insight . Whether users choose to work within Atlassian Analytics or continue using dashboards they're already familiar with, Delta Sharing's open ecosystem of connectors, including Tableau, PowerBI, and Spark, enables customers to easily power their environments with data directly from the Atlassian Data Lake."

—Suresh Raman, Head of Product, Data Platform, Atlassian



Insights powered by flexible + open ecosystem

Create data share

Data share name

Select product instances

Configure which product instances' data will be accessible to your third-party tools through this data share. You may want to exclude product instances that contain sensitive or confidential data.

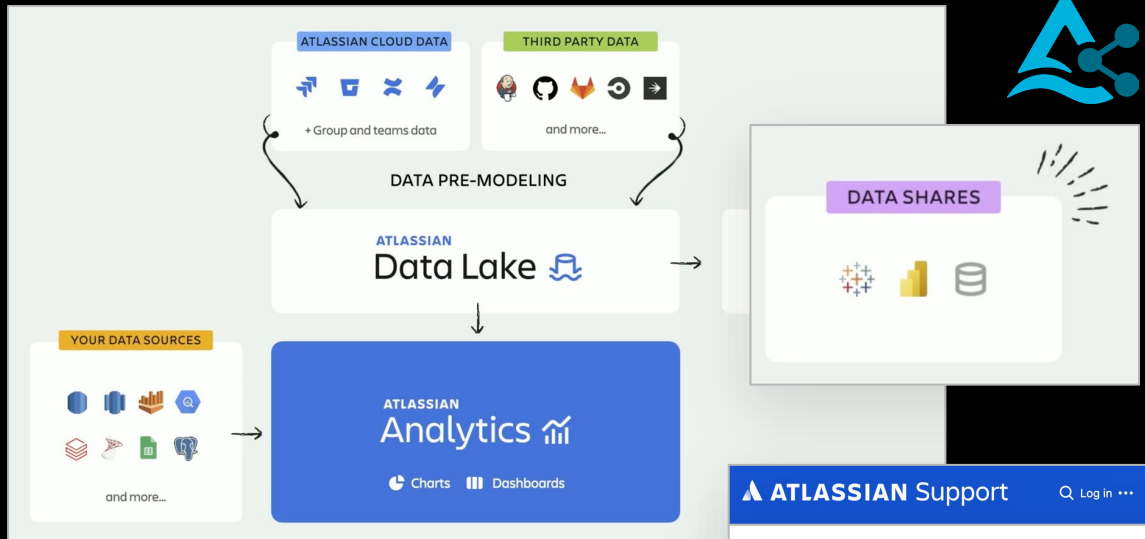
Filter

Type a name All products Can't find your product?

Product instances

- Engineering**
engineering.atlassian.net
 - Jira products**
 - Jira
 - Jira Service Management
 - Jira Work Management
 - Jira Product Discovery
- Marketing**
marketing.atlassian.net
 - Jira products**
 - Jira
 - Jira Service Management
 - Jira Work Management
 - Jira Product Discovery

Cancel **Next**



[Atlassian Team 24 Keynote](#)

ATLASSIAN Support Log in

What is a data share?

A data share is a read-only collection of tables with your organization's data from the Atlassian Data Lake. Organization admins can create data shares and credentials for those data shares in Atlassian Analytics. Those credentials can then be used to connect the data share to your organization's environments or third-party tools that support Databricks Delta Sharing protocol. [Learn more about Delta Sharing.](#)



Sharing and Marketplace Partner Ecosystem

Industry-leading partners adopting Databricks for collaboration



Databricks Marketplace Growth

2,000+
LISTINGS

320% YoY growth

500
Listings
May 2023

Recent Launches

Views, AI model and volume sharing

Private exchanges

Solution accelerators

Provider usage analytics

Non-Databricks clients

**What the
ecosystem is
saying...**

What the ecosystem is saying...



“**Shutterstock** is bringing its vast collection of **nearly a billion creative content assets to the Databricks Marketplace**, a platform renowned for fostering open data and AI collaboration. This integration provides unparalleled access to our extensive library of ethically-sourced visual content, propelling responsible AI and ML initiatives forward across various industries”

—Aimee Egan,
Chief Enterprise Officer, Shutterstock



“**T-Mobile** is excited to launch on Databricks Marketplace. **Databricks customers will now be able to tap into our capabilities to elevate their marketing strategies, optimize customer engagement, and drive business growth** through a deeper understanding of their customers and prospects”

—Grant Reis
SVP and Chief Data & AI Officer, T-Mobile



“We are excited to deepen our partnership with Databricks and **expand our presence through the Databricks Marketplace**. This collaboration **empowers our clients** —including pharmaceutical companies, government agencies, and healthcare organizations —with **unparalleled ease of access to comprehensive, de-identified healthcare data from the nation’s largest real-world healthcare data ecosystem”**

—Andrew Kress,
CEO, HealthVerity



Data Providers across Industries

Financial Services



Health and Life Sciences



Retail, Consumer and Media



What our
customers are
saying...

Moving at the speed of business



Challenge

Deutsche Börse Group is an international exchange organization and market infrastructure provider. Offering widely used data products, they faced challenges in a highly regulated environment with sharing data across siloed businesses and environments. Internal data processing and sharing from on - prem and hybrid cloud environments was complex and time - consuming, with data often being replicated repeatedly.

Solution

Deutsche Börse uses Databricks to streamline data processing across multiple cloud environments (GCP and Azure) with a unified platform that extends access to more users and empowers them to collaborate with streamlined data sharing. As a data provider on **Databricks Marketplace** , they also provide data products directly to customers with sample notebooks and pre - built code. Processes that used to take days can now be completed in a few minutes.

Impact

2- 4x faster insights
average time savings to extract insight from dataset vs. on - prem solution

4- 8x faster publishing
average time savings to publish product on Databricks Marketplace vs. previous solution

“Historically, each of our businesses have grown on their own tech stack, and we pursue a multicloud strategy. To connect them without having to move to a single cloud environment and locking ourselves into one vendor, we use **Delta Sharing** .”

—Abed Shareef, Vice President of Big Data and Advanced Analytics, **Deutsche Börse**



Harnessing the power of healing

ontada

Challenge

As a McKesson Corporation business, **Ontada** is part of a broader oncology ecosystem using iKnowMed, an oncology EHR system that provides clinical support for 2,400+ providers. With a growing volume of data they faced challenges and complexity sharing data with partners across their organization and maintaining strict governance of sensitive and highly regulated clinical data and protected health information (PHI).

Solution

Ontada selected Databricks to help build a next - generation data platform to focus on data consolidation for cost savings and efficiency. **Delta Sharing** enabled the company to share data efficiently, speeding up delivery of both raw and curated data, as well as simplified and strengthened the company's governance, access, filtering and audit processes.

Impact

8x faster insights
to complete data requests
to about an hour.

25% more efficient
of employees' time saved
by helping streamline
compliance oversight

“With Databricks and Delta Sharing, we have a comprehensive end - to - end ecosystem that enables us to gain deep insights in the oncology realm.”

—Kiran Reddy Nomula, Director of Data Management, **Ontada**



Banking for the future



Challenge

Coastal Community Bank based in WA with 40K customers had a vision to offer BaaS (Banking-as-a-Service) and support more customers through their CCBX partner network. Apart from scaling, they needed to streamline, modernize operations for complex legacy banking infrastructure, processes and provide a better customer experience.

Solution

Delta Sharing allowed CCB to scale and modernize by connecting data almost instantly with myriad partner network, as well as providing simplified governance. This enabled them to provide personalized financial products, better risk assessment, automated reporting, as well as adherence to regulatory requirements.

Impact

6M+ customers are now supported on partner network, scaling from 40K

10 minutes to start ingesting data from ONE, an emerging fintech startup, a CCB partner

“With Delta Sharing, we don’t have to move data around anymore. Delta Sharing is so easy as it’s wrapped within the parameters of the platform. No multiple handoffs, or only seeing siloed views.”

—**Barb MacLean**, Head of Technology, Coastal Community Bank



Cost - efficient blockchain solution



Challenge

Allium needed to build a cost - efficient data storage and sharing solution that met their customers' needs across financial products, including managed blockchain databases, enriched data schemas, and real - time notification

Solution

The combination of Delta Sharing with Cloudflare R2 has provided Allium with a cost - effective and secure data sharing solution, taking advantage of **zero egress costs** since there is no need for complex replications or vendor lock - in.

Impact

\$645K / year
savings due to minimizing storage and egress costs

1PB / mo.
data transfer monthly in last quarter + growing

“Combining both Delta Sharing and Cloudflare R2 together allows us to deliver data to our customers reliably and cost - effectively. We deliver the highest quality blockchain data to our customers in their preferred environment, while minimizing our storage and egress costs, saving up to \$645K per year.”

—Ethan Chan, Co- Founder and CEO of Allium



Knowledge Check #1

Which of the following sharing scenario is supported by Delta Sharing?

- A/ Access data from third-party data vendors
- B/ Collaborate with peers, suppliers, and customers
- C/ Access your own data from SaaS platforms
- D/ All of above

Knowledge Check #1

Correct answer:

D/ All of the above

Did you get it right?

What's New with Delta Sharing

Data sharing and collaboration

Databricks Marketplace

Databricks Clean Rooms

Powered by open-source Delta Sharing

Secured and governed by Unity Catalog

What's New with Delta Sharing



Share All Your Data and AI Assets: AI Models, Volumes, Streaming Tables, Materialized Views, Lakehouse Federation



Zero Egress Fee with
Cloudflare R2 Integration



New Delta Sharing
Tableau Connector

AI Model Sharing

Train your models in one place and share anywhere

Share AI models across cloud/region for collaboration, monetization or better inference latency and resources

Deploy shared models using Databricks Data Intelligence Platform AI capabilities for real-time or batch inference

Public Preview

Catalog Explorer > Shares >

llama_3_share

Comment:

Owner: eng-ml-productionization-team@databricks.com

Created on: 04/18/2024 17:39

Assets Recipients

Show alias

Status	Type	Name
		models.meta_llama_3_70b
		models.meta_llama_3_70b_instruct
		models.meta_llama_3_8b
		models.meta_llama_3_8b_instruct

Volume Sharing

Share large collection of data in any format, unlock ML/AI workloads

Share data in any format (images, audio, video, PDFs, etc.) along with tabular data, without the need for replication

Unlock collaboration on multimodal data to accelerate ML/AI workloads

General Availability

Catalog Explorer > Shares >

volume_share

Comment:

Owner: Admins

Created on: 06/08/2024 20:41

Assets Recipients

Show alias

Status	Type	Name
		schema_24.vol_documents
		schema_24.vol_images
		schema_24.vol_videos

Cross-Platform View Sharing

Expand view sharing to anyone in or outside of Databricks with fine-grained access control

Easily organize and share data using views with fine-grained access control at row and column levels

Reach all data recipients in Databricks and across-platforms

Private Preview

```
-- provider creates view with row filtering
> CREATE VIEW acme.default.my_view AS
  SELECT * FROM acme.default.my_table
  WHERE country = current_recipient('country');

-- provider creates view with column masking
> CREATE VIEW acme.default.my_view AS
  SELECT
    CASE
      WHEN current_recipient('country') = 'US' THEN pii
      ELSE 'REDACTED'
    END AS pii
  FROM acme.default.my_table;

-- set recipient property and share view to the recipient
> ALTER SHARE my_share ADD VIEW acme.default.my_view;
> ALTER RECIPIENT cdc SET PROPERTIES ('country' = 'US');
> GRANT SELECT ON SHARE my_share TO RECIPIENT cdc;
```

Streaming Table and Materialized View Sharing

Share fresh data from Databricks SQL and Delta Live Table infrastructure - free data pipelines

Enable easy sharing of streaming ingestion and transformation, with no need for extra data pipelines

Same fine-grained access control with better performance and lower cost for recipients with Materialized Views

Private Preview

Catalog Explorer > Shares >

streaming_share

Comment:

Owner: Admins

Created on: 06/06/2024 11:35

Assets Recipients

Filter assets...

Status	Type	Name
		2024.raw_records
		2024.weekly_records

Sharing for Lakehouse Federation

Share cross-platform data directly from where it is stored, with no data movement or copying

Support for many data sources: Snowflake, Amazon Redshift, Azure Synapse, Google BigQuery, MySQL, PostgreSQL, Microsoft SQL Server

Share external cross-platform data sources without replication

Coming Soon

Catalog Explorer > Shares >

lakehouse_federation_share

Comment:

Owner: Admins

Created on: 06/09/2024 22:06

Assets Recipients

Filter assets... Show alias

Status	Type	Name
		bigquery_schema.employee_contact (via BigQuery)
		snowflake.diamonds (via Snowflake)

Cloudflare R2 Storage Integration



New partnership provides cost-savings using data hosted on Cloudflare R2 storage platform

Zero egress fees cross regions

No need for costly replication

No vendor lock-in

General Availability

Catalog Explorer > cloudflare-r2-catalog > cloudflare-r2 >

cloudflare-r2-catalog.cloudflare-r2.dataset-volume ☆

Overview Details Permissions

Name	dataset-volume
Catalog Name	cloudflare-r2-catalog
Schema Name	cloudflare-r2
Volume Type	EXTERNAL
Storage Location	r2://tao-tao-dais-demo-bucket@c48df378c07a2e713d4d61c040341a8a.r2.cloudflarestorage.com





Catalog Explorer >

Delta Sharing

dais24-demo-sea-to-sky-bank

Shared by me Shared with me

Create shares to share data assets across organizations, and manage access to shares with recipients.

Shares Recipients

Sta...	Name	Authentication type	Created at	Owner	Comment
	fintech_co	DATABRICKS	2024-06...	tao.tao@databricks...	

Tableau Delta Sharing Connector



Real-time data sharing accelerates data-driven decisions and innovation

Expand the reach of your data by meeting end-users where they are in Tableau

Easily consume live data in Databricks data ecosystem directly from Tableau

General Availability

The screenshot shows the Tableau Exchange interface for the 'Delta Sharing by Databricks' connector. At the top, there is a search bar and a 'SIGN IN' link. Below the search bar are navigation tabs: 'Accelerators', 'Dashboard Extensions', 'Viz Extensions (Beta)', and 'Connectors'. The main content area features the Databricks logo and the connector name 'Delta Sharing by Databricks' with a 'Download' button. The page is divided into sections: 'Description', 'Installation', 'Tech Specifications', 'Version', 'Works with', and 'Resources'. The 'Description' section explains that Delta Sharing is an open protocol for secure real-time exchange of large datasets. The 'Installation' section provides a four-step guide for downloading and configuring the connector. The 'Resources' section includes links for 'Installation Guide', 'Driver Download', 'Developer Website', 'Privacy Policy', and 'Terms of Service'. A 'Try Tableau Now' button is located at the bottom right of the page.

tableau exchange SIGN IN

Accelerators Dashboard Extensions Viz Extensions (Beta) **Connectors**

Delta Sharing by Databricks [Download](#)
by Databricks

Description

Delta Sharing is an open protocol for secure real-time exchange of large datasets, which enables organizations to share data in real time regardless of which computing platforms they use. It is a simple **REST protocol** that securely shares access to part of a cloud dataset and leverages modern cloud storage systems, such as S3, ADLS, or GCS, to reliably transfer data. With Delta Sharing, a user accessing shared data can directly connect to it through pandas, Tableau, Apache Spark, Rust, or other systems that support the open protocol, without having to deploy a specific compute platform first. Data providers can share a dataset once to reach a broad range of consumers, while consumers can begin using the data in minutes.

Installation

1. Download the Connector file (.taco).
2. Move the .taco file here: For Desktop: Windows: C:\Users[Windows User]\Documents\My Tableau Repository\Connectors macOS: ~/Users/[user]/Documents/My Tableau Repository/Connectors For Prep: Windows: C:\Users[Windows User]\Documents\My Tableau Prep Repository\Connectors macOS: ~/Users/[user]/Documents/My Tableau Prep Repository/Connectors
3. Start Tableau and under Connect, select the "Delta Sharing by Databricks" connector.
4. Fill out the dialog box with the connection information: Endpoint URL = The endpoint for delta sharing server Bearer Token = The credentials used to access delta sharing server. (Leave blank if no authentication needed) Or Click Upload share file and upload your config.share file that you get from people who share data with you.

Tech Specifications

Version
1.0.0

Works with
Tableau 2024.1 and later

Resources

- [Installation Guide](#)
- [Driver Download](#)
- [Developer Website](#)
- [Privacy Policy](#)
- [Terms of Service](#)

Download and start your **free trial** of Tableau today.
[Try Tableau Now](#)



Connect

Search for Data

Tableau Server

To a File

Microsoft Excel

Text file

JSON file

PDF file

Spatial file

Statistical file

More...

To a Server

Oracle

Amazon Redshift

Delta Sharing by Databricks

Databricks

More... >

Saved Data Sources

Sample - Superstore

World Indicators

Sort by [Name \(a-z\)](#) ▾

Installed Connectors (1)

Delta Sharing by Databricks

Additional Connectors (0) ⓘ

Delta Sharing by Databricks

Sharing Server Credentials

Endpoint URL

Bearer Token

or

Upload share file

Get Data



An open, flexible way to
share all your data + AI assets
across platforms, clouds and regions

What's New with Marketplace

Data sharing and collaboration

Databricks
Marketplace

Databricks Clean
Rooms

Powered by open-source Delta Sharing

Secured and governed by Unity Catalog

What's New with Marketplace



Discover, Procure and
Monetize all Data and AI



Private Exchange to Reach
Targeted Consumers

Marketplace Beyond Tabular Data

Discover, monetize, and integrate AI models

 John Snow Labs

Clinical Text Summarization

This LLM can be used to summarize clinical documentation such as clinical notes, medical encounters, etc.

 Databricks

Meta Llama 3 Models

The Llama 3 - Large language models (LLMs) with 8, 70 billion parameters.
Keywords: Llama3, LLM

Multimodal data for training, fine-tuning and evaluation

 Shutterstock

Free Sample Dataset - People Detection (200 High-Res Images ...

A free sample dataset consisting of 200 images and accompanying metadata sourced from our +550 million image...

 Soundsnap

[SAMPLE] Music Data for Machine Learning (ML) | 50,000 Music File...

The number one music dataset in the world. 50,000 professional music track in all genres with human crafted ...

Faster time-to-insight with pre-built industry solution

 LiveRamp

LiveRamp Identity Resolution Accelerator

Enhance data accuracy with our technology; use LiveRamp for precise customer profiling and improved insights.

 Datavant

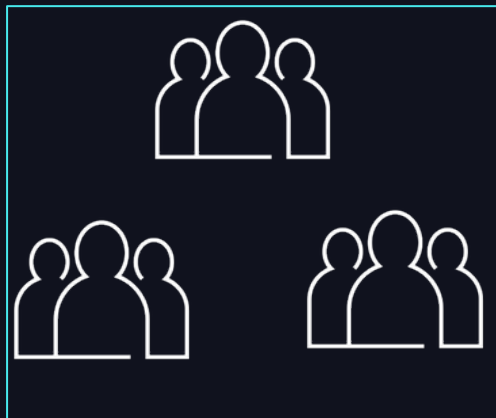
Datavant Connect

Datavant's connect technology for tokenizing and transforming tokens

Private Exchange

Curate and distribute data and AI products for targeted consumers to discover and use

GA in Nov 2023



Internal Teams (Data Mesh)

Curate and share assets across business units in different clouds, regions, or governance domains



External Customers and Partners

Customize listings for targeted customer and partner cohorts, e.g. industry collaboration





Databricks Marketplace

powered by DELTA SHARING

Learn

My requests

Provider console

Search for products

Product

Provider

Category

Model task

Free

Private exchange

Reset filters



No data products

Didn't find what you're looking for? Let us know.

Suggest a product

What's New with Clean Rooms

Data sharing and collaboration

Databricks Marketplace

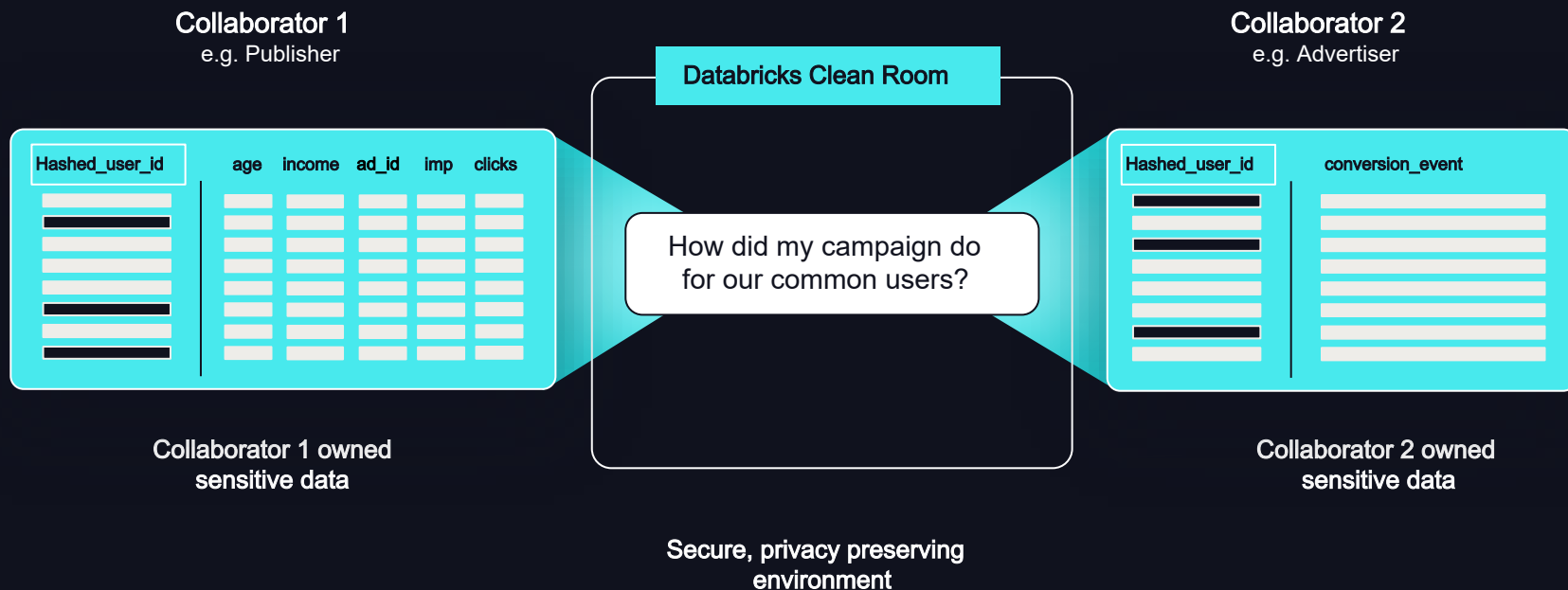
Databricks Clean Rooms

Powered by open-source Delta Sharing

Secured and governed by Unity Catalog

Databricks Clean Rooms

Privacy-safe environments for multiple organizations to run computations on joint data



Use cases for data clean rooms

From cross-joins to
building AI models on
sensitive data



Advertising and Retail

Shopper behavior analysis | Targeted advertising
Demand forecasting | Audience segmentation |
Inventory analysis and reconciliation



Finance

Regulatory risk and compliance | Fraud mitigation |
Targeted financial product development



Health and Sciences

Genomic target identification | Risk scoring
Drug discovery | Population health
Observational research | Targeted recruitment

Databricks Clean Rooms



Any Language
and Workload



Any Cloud
and Platform



Operating at
Any Scale

Any Language and Workload

Leverage the full power of Delta Sharing and Notebooks

Share any format of data and AI models
in a privacy-safe way

Run SQL or Python for complex compute
and ML/AI workloads

Share IP-sensitive code as Private
Libraries

0/ Import Measurement Co's private library

```
python_wheel_name = 'currency_identity_package-0.0.1-py3-none-any.whl'  
python_wheel_path = f'/Volumes/dcr_measurement_co/media_measurement/wheels/{python_wheel_name}'
```

```
%pip install $python_wheel_path  
dbutils.library.restartPython()
```

1/ Use Databricks Widget to capture params provided in job run

```
dbutils.widgets.text('report', 'no_report')  
dbutils.widgets.text('report_group_by', 'no_report_group_by')
```

```
report = dbutils.widgets.get('report')  
report_group_by = dbutils.widgets.get('report_group_by')
```

2/ Load tables into local dataframe

```
currency_df = spark.table('dcr_measurement_co.currency_identity_package')  
publisher_df = spark.table(f'dcr_measurement_co.publisher_{report_group_by}')  
combined_df = currency_df.join(publisher_df, 'publisher')
```

3/ Combine data for analysis using Measurement Co's Private Library

```
import currency_identity_package.functions as currency  
combined_df = currency.resolve_and_combine(nielsen_df, publisher_df)  
combined_df.createOrReplaceTempView('combined_tbl')
```

4/ Define report UDFs

- Function name can be passed in as a parameter when running a job.
- Defining functions is not required, however, it's a common way to add flexibility while also reducing operational overhead

```
# Example 'canned' report  
def rf_by_platform():  
    return spark.sql("""select platform_type, platform, count(distinct person_id) as reach, sum(impressions) as imp,  
                           int(imp/reach) as avg_freq from combined_tbl group by platform_type, platform""")  
  
# Example 'dynamic' report  
def dynamic_rf_report(report_group_by):  
    return spark.sql("""select {}, count(distinct person_id) as reach, sum(impressions) as imp, int(imp/reach) as avg_freq  
                           from combined_tbl group by {} having reach >= 50""")  
    .format(report_group_by, report_group_by)
```

Any Cloud and Any Platform

Meet your collaborators and their data wherever they are

Create Clean Room on any cloud or region,
regardless of collaborators' home regions

Securely collaborate on cross-platform data
with no replication

* Clean Room name





Select a cloud and region

This will be the cloud and region for your secure central clean room station.

AWS GCP Azure

us-west-2

Data assets (tables, views and volumes)

Status	Name	Managed by
<input checked="" type="checkbox"/>	 customers_snowflake.default.customer_loyalty (via  Snowflake)	cpg_company (You)
<input checked="" type="checkbox"/>	 retailer.audience_product_team.audience_graph	retailer
<input checked="" type="checkbox"/>	 retailer.audience_product_team.ad_science_private_library	retailer

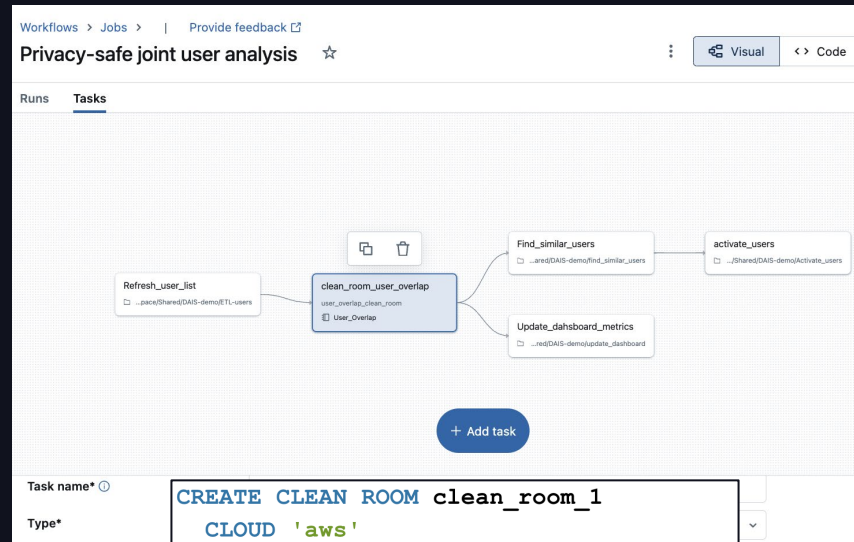
Operation at Scale

Ready for complex workload

APIs, SQL Commands, and Built-in workflow orchestration for automation

Easy access to approved Clean Room outputs in Notebooks and home catalogs

Multi-party collaboration at any trust level




More Exciting Information on Clean Rooms:

- Thursday Keynote with Matei

DATA+AI SUMMIT

JOIN US AT
DATA + AI SUMMIT

BY  databricks



Jamie Couture
/Sr. Dir, Software
Engineering,
Nasdaq


Zahoor Valani
/VP, Engineering
Databricks

Delta Sharing Unlocks
the Value of Your Data
to Partners and
Customers

JUNE 13

DATA+AI SUMMIT

JOIN US AT
DATA + AI SUMMIT

BY  databricks



Tao Tao
/Director, Engineering
Databricks


William Chau
/Staff Software Engineer
Databricks

Getting Started
with
Databricks
Clean Rooms

JUNE 13

DATA+AI SUMMIT

JOIN US AT
DATA + AI SUMMIT

BY  databricks



Bill Oshada
/Product Specialist
Databricks

Bhavin Kukadia
/Solutions Architect
Databricks

Best Practices
for Architecting
Data Collaboration
At-Scale Across Clouds,
Regions, Platforms

JUNE 13

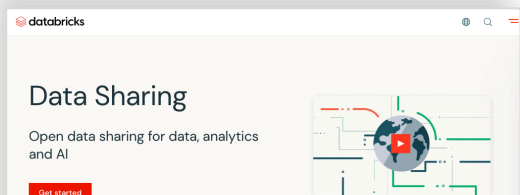
Getting Started



Data Sharing and Collaboration

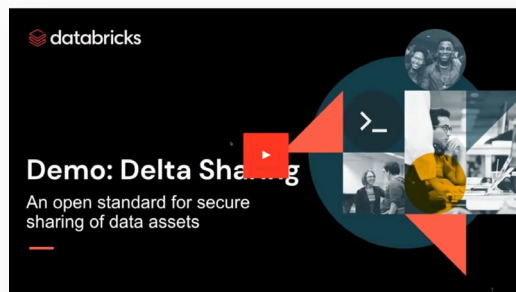
Getting started with Databricks

Get started with Data Sharing

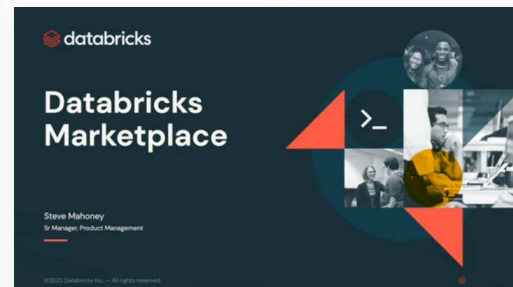


Databricks and the Linux Foundation developed [Delta Sharing](#) to provide the first open source approach to data sharing across data, analytics and AI. Customers can share live data sets — as well as models, dashboards and notebooks — across platforms, clouds and regions without dependencies on specific data-sharing services, including Databricks. This flexibility is achieved with strong security and governance by integrating Delta Sharing with [Databricks Unity Catalog](#) to govern, track and audit access to shared data sets.

Learn more about Delta Sharing



Learn more about Databricks Marketplace



Data Sharing on Databricks

Sharing and collaboration across clouds, platforms, and regions

O'Reilly technical guide



New data sharing eBook



Data sharing for industries

[Download best practices ebook →](#)



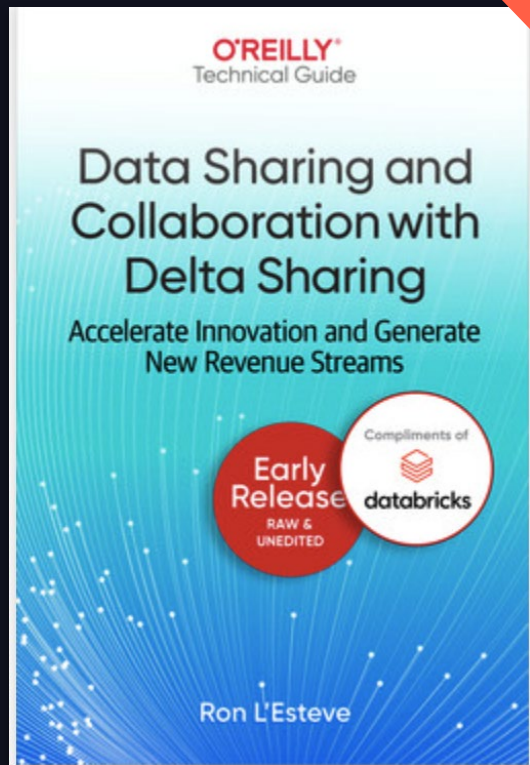
O'Reilly's Guide to Delta Sharing

Practical guidance to help accelerate innovation

Explore the power of data sharing and how Databricks' open source approach to data sharing helps customers share data, analytics and AI across platforms, clouds and regions with strong security and governance.

Chapter Summary:

- Current data sharing landscape, benefits and challenges
- Delta Sharing
- Databricks Marketplace
- Databricks Clean Rooms
- Crafting a data collaboration strategy
- Driving excellence with best practices



NEW

[Download now](#)



NEW

Accelerate Industry Innovation with Data Sharing

Best practices to increase collaboration- across clouds, platforms or regions

This industry-specific eBook is aimed at executive leaders and articulates the value they can achieve with data sharing and collaboration.

CDOs promoting data sharing have 3x greater economic performance, as compared to their peers.

Learn more about:

- Industry-specific use cases
- Customer stories and proof points
- Partners to help get started with Databricks



[Download eBook →](#)

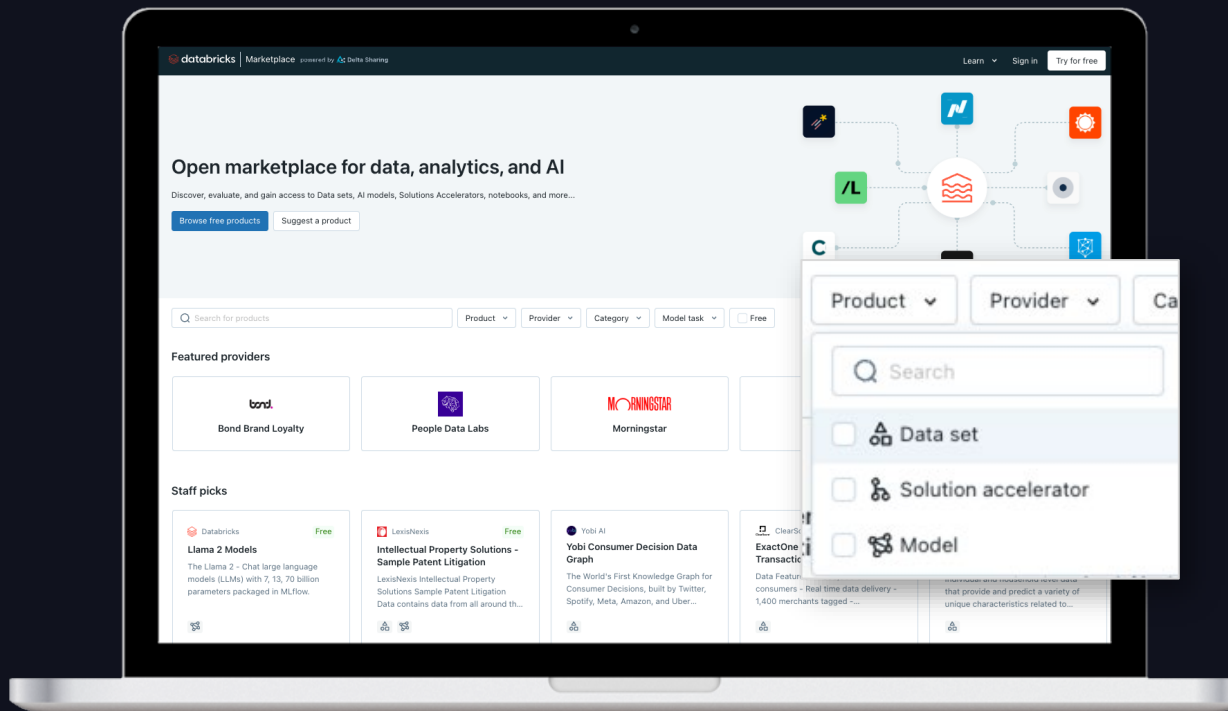
[Share on social media →](#)

[Sales email template →](#)



Explore Databricks Marketplace!

Open marketplace for all your data, analytics, and AI



Explore
Databricks Marketplace:
marketplace.databricks.com

Learn more:
databricks.com/product/marketplace



Join the Data Provider Partner Program



Access the broad, open ecosystem of data consumers with Databricks and monetize data assets all from a single platform.

Our partners can leverage the Databricks Data Intelligence Platform to:

- Reach more customers
- Reduce time and costs
- Provide a best - in- class customer experience for all their data sharing needs

Databricks natively integrates with Delta Sharing and is endorsed by a broad ecosystem of open source and commercial clients and data vendors.

Apply at [Databricks Partner Program](#)



Databricks Clean Rooms

How are you solving sensitive data use cases today?

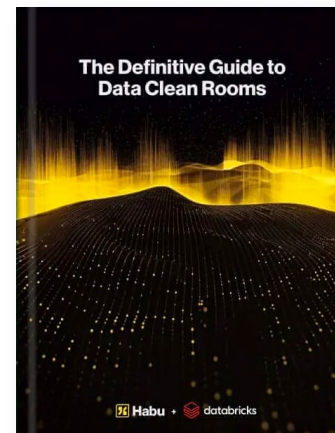
Join our private preview

The screenshot shows a web page titled "Request Access to Databricks Clean Rooms Private Preview". It features a "Join Waitlist" form with fields for First Name, Last Name, Company Email, Company Name, Job Title, Phone Number, and Country (set to United States). A "Sign Up" button is at the bottom. The page also includes a brief description of Databricks Clean Rooms and a list of key benefits: "Your language and workload of choice", "Scalable, multiparty collaboration", and "Interoperable".

Learn more about Databricks Clean Rooms

The screenshot shows the Databricks Clean Rooms landing page. It features a "Submit your interest" button and a "Key benefits" section with three icons: a gear for "Flexible", a group of people for "Scalable, multiparty collaboration", and a network diagram for "Interoperable". Below this is a "Use Cases" section. The page also includes a brief description of Databricks Clean Rooms and a list of key benefits.

Read our definitive technical guide



Q&A



DATA+AI SUMMIT